On the Estimation of Common Factors in the Presence of Block Structures*

Claudia Cicconi†

This version: December 2008

Abstract

This paper considers factor models in which the cross-correlation among time series is due to both common and block-specific factors. Under certain conditions, these models are asymptotically equivalent to approximate factor models and common factors can be consistently estimated by both principal components and quasi maximum likelihood. However, since local cross-correlation can slow down the law of large numbers, these two estimators may have poor properties in finite samples. We explicitly model the block structure of the data using a hierarchical model and show by Monte Carlo simulations that exact maximum likelihood can provide a better estimation of the common factors. The potential advantages of modeling the block structure are illustrated in an empirical application of the hierarchical model to sectoral business surveys. The common factor of the data is then used to forecast the manufacturing production index and its forecasting performance compared to that of other survey based indicators.

JEL codes: C01, C13, C33, C51, C53
Keywords: factor models, block structure, maximum likelihood estimation, principal components, business survey data, forecasting

---

*I wish to thank, without implication, Domenico Giannone for his precious contribution to this work and participants to the 5th EUROSTAT Colloquium on Modern Tools for Business Cycle Analysis for their useful comments.

†Univerisité Libre de Bruxelles (ECARES) and ISTAT (Italian National Statistical Institute), Via A. Depretis 74b, 00184 Rome. Tel: +39 06 4673 3120. E-mail: claudia.cicconi@istat.it.

1 Introduction

This paper considers factor models in which the cross-correlation among time series is due not only to common factors, like in standard factor models, but also to block-specific factors, i.e. factors whose influence is not completely unit-specific, nor common to all cross-sectional units, but common to groups of them.

Explicitly modelling the block structure of the data has revealed very useful in a variety of economic applications. Indeed, models with block-specific factors have been used, among others, by Forni and Reichlin (2001) to assess the scope for national output stabilization of European federal policies; by Kose, Otrock and Whiteman (2003) to characterize the world and regional business cycles; by Beck, Marcellino and Ubrich (2006) to analyze regional inflation dynamics within and across European countries; by Brooks and Del Negro (2004) to investigate the exposures of individual stocks to global, country-specific and industry-specific shocks and by Stock and Watson (2008) who model US building permits data as driven by a national factor, a regional factor, and a state-specific disturbance.

Moreover, data with a block structure is also very often used when the interest is on global factors without modelling the block-specific component. In many empirical applications, especially in macroeconomics and finance, approximate factor models are applied to heterogeneous blocks of data pooled together with the intent of extracting the few common factors driving the economy. Blocks can correspond to regions or countries and/or sectors of origin and/or groups of variables of different nature like prices, production indexes, GDP, stocks, interest rates, money aggregates[1].

If not modeled, the block structure would determine some cross-correlation in the idiosyncratic component. Though approximate factor models do not rule out cross-correlation in the idiosyncratic component, it has to be a mild cross-correlation to allow for identification and guarantee some estimation properties (see for example Chamberlain and Rotschild, 1983 and Forni, Hallin, Lippi and Reichlin, 2000). On the other hand, more recent results obtained by Boivin and Ng (2006), Onatski (2007) and Hallin and Liska (2008) suggest that modelling the block structure of the data can be relevant also when the interest is in global factors. Given the increasing amount of empirical applications involving data sets with block structures, it would be worth to investigate what kind of block structures can be accommodated for by approximate factor models and, in particular, what are the effects of block structures on the estimation of common factors. This is indeed the aim of this paper.

To simplify, let us suppose that our panel of time series is formed by q blocks of n^{b} time series ($n^{\mathrm{b}} < \infty$). Then, we can distinguish two cases: *(i)* $n^{\mathrm{b}}/n \nrightarrow 0$ when $n \to \infty$ and *(ii)* $n^{\mathrm{b}}/n \to 0$ when $n \to \infty$.

The first case reflects the standard approach adopted in practice to increase the number of series, which consists of increasing the level of disaggregation in each block while keeping the number of blocks fixed. In this case, approximate factor models are inadequate because the cross-correlation due to the block-specific factors does not vanish with the cross-section n. Block-specific factors take the form of 'weak' common factors which can only be distinguished with difficulty from 'strong' common factors and the principal component estimator is not consistent (see Onatski, 2007).

[1]Examples are Chamberlain and Rothschild (1983), Forni, Hallin, Lippi and Reichlin (2000); Stock and Watson (1989); Altissimo et al.(2001), Marcellino, Stock and Watson (2003), Angelini, Henry and Mestre (2001), Stock and Watson (2002a) and Bernanke and Boivin (2003) to quote only a few.

In the second case, on which the paper is focused, the number of time series increases as the number of blocks does, while the size of each block is fixed. This could happen, for example, when additional blocks of data are considered or when the level of disaggregation of the variable(s) according to which the blocks are formed increases (regions instead of countries, for example). In this case, the cross-correlation due to the block-specific factors dies out with the cross-section so that the model with block-specific factors is asymptotically equivalent to an approximate factor model. Then, even if the block-specific component is not explicitly modelled, both principal components and quasi maximum likelihood can provide consistent estimates of the common factors (see Doz, Giannone and Reichlin, 2006). However, since local cross-correlation in the idiosyncratic component can significantly slow down the law of large numbers (see Boivin and Ng, 2006), both principal components and quasi maximum likelihood estimators of the common factors may have poor properties in finite samples.

We model the block structure of the data through a hierarchical factor model and propose an exact maximum likelihood estimator of the common factors. The exact maximum likelihood estimator of the common factors is then compared to the principal components estimator (Stock and Watson, 2002) and to the quasi maximum likelihood estimator, recently proposed by Doz, Giannone and Reichlin (2006), under different forms of block structures. Our Monte Carlo simulation experiments show that while, as expected, the three estimators are asymptotically equivalent, the exact maximum likelihood estimators outperform the others in finite samples, suggesting that explicitly modeling the block structure of the data can indeed provide better estimates of the common factors.

The potential advantages of modeling the block structure are illustrated in an empirical application in which the model is applied to sectoral business survey data. The common factor of the data is estimated by exact maximum likelihood and it is used to forecast the manufacturing production index. Its forecasting performance is then compared to that of other survey based indicators by reproducing a pseudo real-time situation, which would fully exploit the timeliness of releases of business surveys with respect to preliminary official data on industrial production. The block-specific component accounts for a considerable amount of variations in the data (more than 20 per cent). The forecasts obtained by using our model with sector-specific factors are better than the average in terms of mean square forecast error. Moreover, we find that forecasting the manufacture production index by aggregating sectoral forecasts, obtained by using our estimated sector-specific factors, provides better results than simply forecasting the aggregate directly.

The remainder of the paper is organized as follows. Section 2 describes the model and some related literature; Section 3 presents the estimation method; Section 4 describes the Monte Carlo simulation experiment and presents the main results; Section 5 describes the results of the empirical application and Section 6 concludes.

2 The model with block-specific factors

In the hierarchical factor model the observed time series are decomposed in three orthogonal components: a common component, a block-specific component, which is driven by q block-specific factors, and an idiosyncratic component, which is completely unit specific. More formally,

$$\mathbf{x}_t = \Lambda^{c}\mathbf{f}_t^{c} + \Lambda^{b}\mathbf{f}_t^{b} + \mathbf{e}_t \tag{1}$$

where

- $\mathbf{x}_t = (x_{1t}, \cdots, x_{nT})$ is a $n \times 1$ stationary process;
- Λ^{c} is the $n \times r$ matrix of common factor loadings;
- $\mathbf{f}_t^{c}$, the vector of common factors, is a $r \times 1$ stationary process;
- $\Lambda^{b} = (\Lambda_k^{b})$, $k = 1, \cdots, q$, is the $n \times q$ matrix of block-specific factor loadings with

$$\Lambda_{ik}^{b} = \begin{cases} \textit{unconstrained} & \text{if } i \text{ is in block } k,\ i = 1, \cdots, n \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- $\mathbf{f}_t^{b}$, the vector of block-specific factors, is a $q \times 1$ stationary process;
- $\mathbf{e}_t = (e_{1t}, \cdots, e_{nt})$, the idiosyncratic component, is a $n \times 1$ stationary process;
- $E(\mathbf{e}_t \mathbf{e}_t^{'}) = \Psi$, a diagonal matrix;
- $\mathbf{f}_t^{c}$, $\mathbf{f}_t^{b}$ and $\mathbf{e}_t$ are orthogonal at all leads and lags;
- $E(\mathbf{f}_t^{c} \mathbf{f}_t^{c\prime}) = I_r$ and $E(\mathbf{f}_t^{b} \mathbf{f}_t^{b\prime}) = I_q$.

A similar model has been recently formalized by Ng, Moench and Potter (2008). For each observed variable i in block k they have

$$x_{it}^{k} = \lambda_{ik}(L)^{'} G_{kt} + e_{it}^{k} \tag{3}$$

$$G_{kt} = \gamma_k(L)\mathbf{f}_t + \mathrm{g}_{kt} \tag{4}$$

and therefore

$$x_{it}^{k} = \lambda_{ik}(L)^{'} \gamma_k(L)\mathbf{f}_t + \lambda_{ik}(L)^{'} \mathrm{g}_{kt} + e_{it}^{k} \tag{5}$$

where $\mathbf{f}_t$ is the $r \times 1$ vector of unobserved common factors, corresponding to our $\mathbf{f}_t^{c}$, g_{kt} is the $q_k \times 1$ vector of block-specific factors of the k-th block[2], corresponding to our f_{kt}^{b}, and e_{it}^{k} is the residual idiosyncratic component.
This model shows similarities and differences from our model. The main similarity is that we both take the block-structure as given, in the sense that it is completely suggested by the structure of the data. Concerning this point, the approach behind two models differs from Hallin and Liska (2008) where a theoretical framework for the identification of the block structure is developed. Another similarity is that both models, the model considered here and Ng, Moench and Potter (2008)'s, can be easily generalized to multiple levels models, i.e. to models in which the cross-sectional units are grouped according to more than one variable[3]. However the two models also differ in some characteristics. The first difference is that our model is static while their model is dynamic; the second is that in our model each block of data is driven by one block-specific factor while they allow for multiple block-specific factors in each block. The generalization of our model is straightforward and our preliminary results confirm the intuition that simplification can come without any loss of generality.
A third difference between the two models concerns the heterogeneity of the loadings on the common factors. In Ng, Moench and Potter (2008) the loading of unit i of block k on the common factors is $\lambda_{ik}^{'}\gamma_k$, implying that the responses of cross-sectional units in block

[2]In Ng, Moench and Potter (2008) the block-specific factors are the G_ts. We call their g_ts block-specific factors and the G_ts block-common factors.

[3]This is hardly possible in Hallin and Liska (2008) because of a problem of dimensionality. In fact, they assume a factor space that contains a total of 2^q factors where q is the number of blocks, thereby posing a limit to the number of blocks that can be taken into consideration in practice already when the cross-sectional units are partitioned along one dimension only.

k to the common factors can only differ to the extent that their exposure to the block-common factors G_t differs, whereas our Λ_i^{c} is completely unconstrained[4]. Not imposing restrictions on the loadings of the common factors has the advantage of allowing for more flexibility while the implied larger number of parameters to estimate has a very limited negative impact on computational time.

2.1 The model with block-specific factors and the approximate factor model

Let us now sort our time series so that the first n_1 belong to the first block, the next n_2 to the second block and so forth, till the last n_q, belonging to the q-th block ($n = \sum_k n_k$). Then, the decomposition of the covariance matrix of the observables implied by the hierarchical model is:

$$\Sigma = \Lambda^{\mathrm{c}}\Lambda^{\mathrm{c}\prime} + \Lambda^{\mathrm{b}}\Lambda^{\mathrm{b}\prime} + \Psi \tag{6}$$

where:

- $\Lambda^{\mathrm{c}}\Lambda^{\mathrm{c}\prime}$, the covariance matrix of the common component, is a full matrix;
- $\Lambda^{\mathrm{b}}\Lambda^{\mathrm{b}\prime}$, the covariance matrix of the block-specific components, is a block- diagonal matrix;
- Ψ , the covariance matrix of the idiosyncratic component, is a diagonal matrix.

Hence, the idiosyncratic component is completely unit-specific, the block-specific factors are only responsible for cross-correlation within-blocks, i.e. across units in the same block, while common factors are also responsible for cross-correlation across blocks.

If the block-specific component is not modeled, the idiosyncratic component will be $\widetilde{\mathbf{e}}_t = \Lambda^{\mathrm{b}}\mathbf{f}_t^{\mathrm{b}} + \mathbf{e}_t$, with block-diagonal covariance matrix $\widetilde{\Psi}$ with

$$\widetilde{\psi}_{ij} = \begin{cases} \psi_i + (\Lambda_{ik}^{\mathrm{b}})^2 & \text{if } i = j \\ \Lambda_{ik}^{\mathrm{b}}\Lambda_{jk}^{\mathrm{b}} & \text{if } i \neq j \text{ but } i \text{ and } j \text{ belong to the same block} \\ 0 & \text{if } i \neq j \text{ and } i \text{ and } j \text{ do not belong to the same block} \end{cases} \tag{7}$$

for all is in block k, $k = 1, \cdots q$.

However, if the size of the blocks does not grow as fast as the cross-section dimension n ($n_k/n \to 0$ as $n \to \infty$), the factor model with unmodeled block-specific component is asymptotically equivalent to an approximate factor model. In order to show that, we rely on the definition of the approximate factor model in Chamberlain and Rotschild (1983) according to which a factor model is an approximate factor model if (and only if) the largest eigenvalue of the covariance matrix of its idiosyncratic component is bounded, in the sense that goes to zero as n goes to infinity. Let $\lambda_{max}(\widetilde{\Psi})$ denote the largest eigenvalue of $\widetilde{\Psi}$. Then, if the model is not degenerate, $\lambda_{max}(\widetilde{\Psi})$ is larger than zero and bounded above by $\max_i \sum_{j=1}^{n} |\widetilde{\psi}_{ij}|$. Assume the maximum is achieved for $i = \bar{i}$ and that $\bar{i}$ belongs to block

[4]In this, we follow Kose, Otrok and Whiteman (2003)and Beck, Marcellino and Ubrich (2006), to cite only a few.

k. Then,

$$\begin{aligned}
0 \leq \lambda_{max}(\widetilde{\Psi})/n &\leq \sum_{j=1}^{n} |\widetilde{\psi}_{\bar{i}j}|/n = \\
&= \frac{\widetilde{\psi}_{\bar{i}}}{n} + \frac{1}{n}\sum_{j=1} n|\Lambda^{\mathrm{b}}_{\bar{i}k}\Lambda^{\mathrm{b}}_{jk}| \leq \\
&\leq \frac{\psi_{\bar{i}}}{n} + \frac{(\Lambda^{\mathrm{b}}_{\bar{i}k})^2}{n} + \frac{n_k}{n}|\Lambda^{\mathrm{b}}_{\bar{i}k}| \max_{j\neq\bar{i}} |\Lambda^{\mathrm{b}}_{jk}| = \\
&= \frac{\psi_{\bar{i}}}{n} + \frac{(\Lambda^{\mathrm{b}}_{\bar{i}k})^2}{n} + \frac{n_k}{n}|\Lambda^{\mathrm{b}}_{\bar{i}k}\Lambda^{\mathrm{b}}_{\bar{j}k}|.
\end{aligned}$$

Then, if the size of the blocks is bounded the largest eigenvalue of the covariance matrix of the idiosyncratic component is also bounded ($\lambda_{max}(\widetilde{\Psi})/n \to 0$) and the model is an approximate factor model.

From the estimation point of view, the asymptotic equivalence to the factor model implies that the common factors can be consistently estimated even if the block-structure is not modelled. However, when the cross-section is finite the two models differ and this may have an impact on the estimation of the common factors. The aim of this paper is indeed to evaluate such an impact.

3 The estimation method

The hierarchical factor model is estimated by exact maximum likelihood using the Expectation Restricted Maximization (ERM) algorithm - an EM algorithm in which the maximization step is performed under the desired constraints. With respect to the model described in the previous Section (see (1 and (2)), $\mathbf{f}_t^{\mathrm{c}}$, $\mathbf{f}_t^{\mathrm{b}}$ and $\mathbf{e}_t$ are assumed to be white noise processes. The main motivation behind such a choice is that we want to develop an estimation method for our simulation experiments where we focus our attention on the impact of the block structure on the estimation of the common factors, making all the rest as simple as possible. The dynamic version of the model can be estimated by maximum likelihood using the Kalman filter as in Doz, Giannone and Reichlin (2006) or by Monte Carlo Markov Chain as in Ng, Moench and Potter (2008).

Let us first introduce some notations in order to write the likelihood of the model in a compact form. Let $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_t, \cdots, \mathbf{x}_T)'$ denote the $T \times n$ stacked vector of observed variables, $\Lambda = (\Lambda^{\mathrm{c}}, \Lambda^{\mathrm{b}})$ the $n \times (r+q)$ matrix of unrestricted factor loadings, $\mathbf{F} = (\mathbf{F}_1{}', \cdots, \mathbf{F}_t{}', \cdots \mathbf{F}_T{}')'$, where $\mathbf{F}_t = (\mathbf{f}_t^{\mathrm{c}\prime}, \mathbf{f}_t^{\mathrm{b}\prime})'$ the $T \times (r+q)$ stacked vector of common and block-specific factors and, finally, $\mathbf{E} = (\mathbf{e}_1, \cdots, \mathbf{e}_t, \cdots, \mathbf{e}_T)'$ the $T \times n$ stacked vector of idiosyncratic components.

The loglikelihood of the unrestricted model is then given by

$$\begin{aligned}
\mathcal{L}(\theta; \mathbf{X}, \mathbf{F}) = &-\frac{(n+r+q)T}{2}\log(2\pi) - \frac{T}{2}\log|\Psi| - \frac{T}{2}\mathrm{Tr}\left(\Psi^{-1}S_{XX}\right) - \\
&-\frac{T}{2}\mathrm{Tr}\left(\Psi^{-1}\Lambda S_{FF}\Lambda'\right) + \frac{T}{2}\mathrm{Tr}\left(\Psi^{-1}S_{XF}\Lambda'\right) + \\
&+\frac{T}{2}\mathrm{Tr}\left(\Psi^{-1}\Lambda S_{XF}{}'\right) - \frac{T}{2}\mathrm{Tr}\left(S_{FF}\right)
\end{aligned}$$

where θ collects the parameters of the model, Λ and Ψ, $S_{XX} = \frac{1}{T}\sum_t X_t X_t{}'$, $S_{XF} = S_{FX}{}' = \frac{1}{T}\sum_t X_t F_t{}'$ and $S_{FF} = \frac{1}{T}\sum_t F_t F_t{}'$.

Let us now denote with a R_k^{b} a $(q-1)\times q$ matrix obtained by inserting a column of zeros in the position k in an identity matrix of dimension $q-1$. Then, for all is in block k, we can define the $(q-1)\times(r+q)$ matrix $R_i = \begin{pmatrix}\mathbf{0}_{(q-1)\times r} & R_k^{\mathrm{b}}\end{pmatrix}$. The block-specific factor loadings in the hierarchical model satisfy the zero constraints in (2), which can be written in a matrix form as

$$R\mathrm{vec}(\Lambda) = 0_{n(q-1)\times 1}, \tag{8}$$

where R is the $n(q-1)\times n(r+q)$ block diagonal matrix with blocks R_i.

Then, the exact maximum likelihood estimates of the parameters can be obtained by maximizing the augmented loglikelihood of the model

$$\widetilde{\mathcal{L}}(\theta;\mathbf{X},\mathbf{F}) = \mathcal{L}(\theta;\mathbf{X},\mathbf{F}) - \varsigma' R\mathrm{vec}(\Lambda), \tag{9}$$

where ς is the $n(q-1)\times 1$ vector of Lagrange multipliers. This can be done applying the *ERM* algorithm. At each iteration l, we evaluate the conditional expectation of the loglikelihood given the current parameter estimates and the observed variables, $E(\widetilde{\mathcal{L}}\backslash\mathbf{X},\theta^{(l-1)})$, by computing $E(S_{XF}\backslash\mathbf{X},\theta^{(l-1)})$ and $E(S_{FF}\backslash\mathbf{X},\theta^{(l-1)})$, which are given, respectively, by the two sufficient statistics

$$S_{XF}^{(l)} = S_{FX}^{(l)}{}' = S_{XX}^{(l)}\Sigma^{(l-1)-1}\Lambda^{(l-1)};$$

$$S_{FF}^{(l)} = (I_{(r+q)} + \Lambda^{(l-1)'}\Psi^{(l-1)-1}\Lambda^{(l-1)})^{-1} + \Lambda^{(l-1)'}\Sigma^{(l-1)-1}S_{XX}^{(l)}\Sigma^{(l-1)-1}\Lambda^{(l-1)}$$

where $\Sigma^{(l-1)} := \Lambda^{(l-1)}\Lambda^{(l-1)\prime} + \Psi^{(l-1)}$.

Once the expected value of the augmented loglikelihood given the observed variables and the current estimate of the parameters is computed (E-step), we can maximize it (restricted maximization step) with respect to θ to obtain the updated parameter estimates

$$\theta^{(l)} = \mathrm{Argmax}_\theta \quad E(\widetilde{\mathcal{L}}(\theta)\backslash\mathbf{X},\theta^{(l-1)}).$$

In our particular case, the first order conditions give

$$\Lambda_i^{(l)} = S_{X_iF}^{(l)}(I_{r+q} - S_{FF}^{(l)}R_i{}'(R_i S_{FF}^{(l)}R_i{}')^{-1}R_i)S_{FF}^{(l)} \qquad i = 1,\cdots,n \tag{10}$$

$$\Psi^{(l)} = \mathrm{Diag}(S_{XX}^{(l)} - S_{XF}^{(l)}\Lambda^{(l)\prime} - \Lambda^{(l)}S_{XF}^{(l)}{}' + \Lambda^{(l)}S_{FF}^{(l)}\Lambda^{(l)\prime}). \tag{11}$$

We take the first r principal components of the observed variables as initial values for the r common factors and obtain the initial values for their factor loadings as orthogonal projections on the r common factors. We then compute $\dot{\mathbf{x}}_t = \mathbf{x}_t - \Lambda^{\mathrm{c}(0)}\mathbf{f}_t^{\mathrm{c}(0)}$ and for $k = 1,\cdots,q$ we set the initial value of the block-specific factor of block k equal to the first principal component of $\dot{\mathbf{x}}_i^k = \{\dot{x}_{it}\backslash i \text{ in block } k \text{ and } t = 1,\cdots,T\}$; the initial values of the non-zero block-specific factor loadings are then obtained as orthogonal projections on the corresponding block-specific factor, block by block.

We control for convergence by looking at the angle between $\hat{\mathbf{f}}_t^{\mathrm{c}(l)}$ and $\hat{\mathbf{f}}_t^{\mathrm{c}(l-1)}$ and stop the algorithm when it is less than 1e−4. The algorithm is repeated till convergence is achieved at iteration L and the maximum likelihood estimator of the common factors is defined as the expected value of the common factors given the data and the estimated parameters

$$\hat{\mathbf{f}}_{HIE,t}^{\mathrm{c}} = E(\mathbf{f}_t^{\mathrm{c}}\backslash\mathbf{X},\theta^{(L)}). \tag{12}$$

4 The small sample properties of the estimators of the common factors

As shown in Section 2, if the size of the blocks is bounded the cross-correlation due to block-specific factors can be accommodated by an approximate factor model. Then, both principal component and quasi maximum likelihood would provide consistent estimators of the common factors (see Doz, Giannone and Reichlin, 2006) that are asymptotically equivalent to the exact maximum likelihood estimator obtained by explicitly modelling the block structure. This makes modelling the block structure of the data asymptotically irrelevant. However, explicitly modelling the block-structure can make a difference in small samples. In fact, as the cross-correlation among the idiosyncratics due to the unmodelled block-structure slows down the law of large numbers (Bai and Ng, 2006), when n is finite the exact maximum likelihood estimator, obtained by fully specifying the block-structure, could provide a better approximation of the common factors than principal components and the quasi maximum likelihood estimator.

We analyze the small sample properties of these estimators through a set of Monte Carlo simulation experiments. We generate data from a static hierarchical model and estimate the common factors using three alternative estimators: the exact maximum likelihood estimator proposed in the paper, the quasi maximum likelihood estimator proposed by Doz, Giannone and Reichlin (2006) and the principal component estimator (Stock and Watson, 2002). Then, we compare them in terms of efficiency by looking at some trace statistics.

More precisely, we draw data from

$$\mathbf{x}_t = \Lambda^{c}\mathbf{f}_t^{c} + \Lambda^{b}\mathbf{f}_t^{b} + \mathbf{e}_t$$

where

- $\Lambda_{ik}^{b} = \begin{cases} \text{i.i.d. } \mathcal{N}(0,1) \text{ if } i \text{ is in block } k,\ i=1,\cdots,n,\ k=1,\cdots,q \\ 0 \quad \text{otherwise} \end{cases}$
- Λ_{ij}^{c} i.i.d. $\mathcal{N}(0,1)$ $i=1,\cdots,n$ and $j=1,\cdots,r$
- $\mathbf{f}_t^{c}$ i.i.d. $\mathcal{N}(0,I_r)$
- $\mathbf{f}_t^{b}$ i.i.d. $\mathcal{N}(0,I_q)$
- $\mathbf{e}_t$ i.i.d. $\mathcal{N}(0,\Psi)$, Ψ diagonal
- $\psi_i = \frac{\beta_i}{1-\beta_i-\gamma_i}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{r}\Lambda_{ij}^{c}f_{jt}^{c}\right)^2$ with β_i and γ_i i.i.d. $\mathcal{U}([u,\frac{2}{3}-u])$.

The coefficient β_i is the ratio between the variance of the idiosyncratic component and the variance of the observed variable x_{it}, while γ_i is the portion of the variance of x_{it} explained by sector-specific factors. In our simulations, β_i and γ_i are uniformly distributed with an average of 1/3, which means that each component in the model - the common, the block-specific and the idiosyncratic - accounts for one third of total variance on average. All of the blocks have the same size, n^{b}, kept fixed as n increases. Each block is driven by one block-specific factor, so that the number of block-specific factors is the same as the number of blocks, given by $q = n/n^{b}$.

We generate the model for different sizes of the cross-section - $n = 20, 50, 100, 150, 200, 250$ - and for sample sizes $T = 50, 100, 150$. Then, for each combination of n and T, we draw data 1000 times and compute the following estimates of the common factors:

- principal components (Stock and Watson, 2002): $\hat{\mathbf{f}}_{PC,t}^{c}$;

- quasi maximum likelihood (Doz, Giannone and Reichlin, 2006): $\hat{\mathbf{f}}^{c}_{DGR,t}$;

- exact maximum likelihood: $\hat{\mathbf{f}}^{c}_{HIE,t}$.

The first two estimation methods ignore the block-structure of the data: , the principal components estimator is a maximum likelihood estimator obtained under the assumption that the the idiosyncratic component is spherical and non cross-sectionally correlated, while the quasi maximum likelihood estimator by Doz, Giannone and Reichlin (2006), $\hat{\mathbf{f}}^{c}_{DGR,t}$, is obtained under the assumption that the idiosyncratic component is heteroscedastic but non cross-sectionally correlated.
Conversely, the exact maximum likelihood estimator, $\hat{\mathbf{f}}^{c}_{HIE,t}$, is obtained under the true model and it is not affected by any source of miss-specification.

Following Bai and Ng (2006) and Doz, Giannone and Reichlin (2006) we evaluate the performance of the three estimators by means of the following trace statistic:

$$TR_{(m)} = \frac{\text{Tr}\left(\mathbf{F}'\hat{\mathbf{F}}_{(m)}(\hat{\mathbf{F}}_{(m)}'\hat{\mathbf{F}}_{(m)})^{-1}\hat{\mathbf{F}}_{(m)}'\mathbf{F}\right)}{\text{Tr}(\mathbf{F}'\mathbf{F})}$$

where $\hat{F}_{(m)}$ is the estimate of the common factors obtained with the estimator m, $m =$ PC, DGR, HIE. Such a statistic can be seen as a multivariate version of the R^2 of the observed factors on the estimated factors: the closer to one, the better the approximation of the true common factors. In order to compare the performance of the three estimators, we compute the ratios $TR_{(HIE)}/TR_{(PC)}$ and $TR_{(HIE)}/TR_{(DGR)}$. Numbers higher than one indicate that the exact maximum likelihood estimates are more accurate than principal components and the quasi maximum likelihood estimates, respectively.

Table 1 shows the results in terms of medians across 1000 repetitions of our baseline Monte Carlo simulation ($r = 1$ and $n^{b} = 5$). The following results emerge: first, the approximation of the true common factors provided by the three estimators improves as the cross-section n increases; second, the exact maximum likelihood estimator always outperforms the other two; third, the exact maximum likelihood estimator is particularly good when n is small; fourth, the three estimators tend to perform similarly as n increases; fifth, the precision of the quasi maximum likelihood and of the principal components estimators reaches that of the exact maximum likelihood estimator only when the size of the cross-section is quite high; sixth, the effect of the sample size, T, is small and has some relevance only when n is very small; the number of iterations needed for the convergence of the ERM algorithm is quite high but it quickly decreases with n, presumably because of the improvement in the initial values used for the parameters.

Table 1. Monte Carlo experiment results*

Number of iterations of the ERM algorithm ($\hat{\mathbf{f}}^{c}_{(HIE)}$)

	$n=20$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$
$T=50$	61	36	27	23	21	20
$T=100$	58	34	26	22	20	19
$T=150$	55	33	25	22	20	18

$TR_{(HIE)}$

	$n=20$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$
$T=50$	0.92	0.96	0.97	0.97	0.97	0.97
$T=100$	0.93	0.97	0.98	0.98	0.99	0.99
$T=150$	0.94	0.97	0.98	0.99	0.99	0.99

$TR_{(HIE)}/TR_{(PC)}$

	$n=20$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$
$T=50$	1.082	1.025	1.011	1.007	1.005	1.004
$T=100$	1.076	1.023	1.011	1.007	1.005	1.004
$T=150$	1.076	1.023	1.010	1.007	1.005	1.004

$TR_{(HIE)}/TR_{(DGR)}$

	$n=20$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$
$T=50$	1.073	1.021	1.009	1.006	1.004	1.004
$T=100$	1.062	1.019	1.009	1.006	1.004	1.003
$T=150$	1.058	1.019	1.008	1.006	1.004	1.003

* Medians across 1000 experiments; $TR_{(m)} = \frac{\mathrm{Tr}\left(\mathbf{F}'\hat{\mathbf{F}}_{(m)}(\hat{\mathbf{F}}_{(m)}{}'\hat{\mathbf{F}}_{(m)})^{-1}\hat{\mathbf{F}}_{(m)}{}'\mathbf{F}\right)}{\mathrm{Tr}(\mathbf{F}'\mathbf{F})}$.
Data generating process: $n^{b}=5$, $r=1$, β_i and γ_i i.i.d. $\mathcal{U}([u, \frac{2}{3}-u])$.

If the data has a block structure, the convergence to the approximate factor model depends on the convergence of n^{b}/n to zero (see the Introduction and Section 2). This implies that the higher n^{b}, the better the relative performance of the exact maximum likelihood estimator for any given n and the higher n has to be for the three estimators to be equivalent. This is confirmed by the results shown in Table 2[5]: first, exact maximum likelihood always outperforms quasi maximum likelihood and principal components, independently of the block-size; second, as expected the larger the block-size the better the relative performance of the exact maximum likelihood estimator for any given n; third, the three estimators tend to perform similarly as n increases; fourth, the larger the block-size the higher n as to be in order to make the three estimators equivalent; finally, quasi maximum likelihood though outperformed by exact maximum likelihood always does better than principal components, independently of the block-size and of the cross-section dimension.

[5] Since the sample size T is not relevant, we only show the results for $T=100$. The full set of results is available upon request.

Table 2. The effect of the block-size*

$TR_{(HIE)}$

	$n=20$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$
$n^{b}=10$	0.94	0.97	0.98	0.98	0.99	0.99
$n^{b}=5$	0.93	0.97	0.98	0.98	0.98	0.99
$n^{b}=2$	0.92	0.96	0.98	0.98	0.98	0.99

$TR_{(HIE)}/TR_{(PC)}$

	$n=20$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$
$n^{b}=10$	1.286	1.035	1.014	1.009	1.006	1.005
$n^{b}=5$	1.076	1.023	1.011	1.007	1.005	1.004
$n^{b}=2$	1.033	1.012	1.006	1.004	1.003	1.003

$TR_{(HIE)}/TR_{(DGR)}$

	$n=20$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$
$n^{b}=10$	2.018	1.027	1.012	1.007	1.005	1.004
$n^{b}=5$	1.062	1.019	1.009	1.006	1.004	1.003
$n^{b}=2$	1.033	1.009	1.005	1.003	1.002	1.002

* Medians across 1000 experiments; $TR_{(m)} = \frac{\mathrm{Tr}\left(\mathbf{F}'\hat{\mathbf{F}}_{(m)}(\hat{\mathbf{F}}_{(m)}'\hat{\mathbf{F}}_{(m)})^{-1}\hat{\mathbf{F}}_{(m)}'\mathbf{F}\right)}{\mathrm{Tr}(\mathbf{F}'\mathbf{F})}$.

Data generating process: $r=1$, T=100, β_i and γ_i i.i.d. $\mathcal{U}([u, \frac{2}{3}-u])$.

In order to evaluate the negative effects of fitting a hierarchical model to data without a block structure, we simulate data from a static exact factor model and compare principal components, quasi maximum likelihood and 'exact' maximum likelihood obtained assuming that the data have a block-structure. The results in Table 3[6] show that 'exact' maximum likelihood always does better than principal components and, at worst, 0.2 per cent worse than quasi maximum likelihood. This is encouraging since it guarantees that, even if the model is an exact factor model and we erroneously fit a hierarchical factor model, we would get good estimates of the common factors[7].

[6]Table 3 only shows results for $n^{b}=5$. We also computed exact maximum likelihood estimators assuming different values of the block-size obtaining similar results.

[7]This issue deserves further investigation. In particular, it would be interesting to investigate the case in which the data generating process is an approximate factor model to relax the assumption that the number of common factors is known. Both these two extensions are in our current research agenda.

Table 3. The effects of misspecification*

$TR_{(HIE)}$

	$n=20$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$
$T=50$	0.95	0.98	0.98	0.99	0.99	0.99
$T=100$	0.96	0.98	0.99	0.99	0.99	0.99
$T=150$	0.96	0.98	0.99	0.99	0.99	0.99

$TR_{(HIE)}/TR_{(PC)}$

	$n=20$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$
$T=50$	1.008	1.004	1.002	1.001	1.001	1.001
$T=100$	1.010	1.004	1.002	1.002	1.001	1.001
$T=150$	1.011	1.005	1.002	1.002	1.001	1.001

$TR_{(HIE)}/TR_{(DGR)}$

	$n=20$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$
$T=50$	0.998	0.999	1.000	1.000	1.000	1.000
$T=100$	0.999	1.000	1.000	1.000	1.000	1.000
$T=150$	0.999	1.000	1.000	1.000	1.000	1.000

* Medians across 1000 experiments; $TR_{(m)} = \frac{\mathrm{Tr}\left(\mathbf{F}'\hat{\mathbf{F}}_{(m)}(\hat{\mathbf{F}}_{(m)}'\hat{\mathbf{F}}_{(m)})^{-1}\hat{\mathbf{F}}_{(m)}'\mathbf{F}\right)}{\mathrm{Tr}(\mathbf{F}'\mathbf{F})}$.

Data generating process: exact factor model with heteroscedastic idiosyncratic component, $r = 1$, variance explained by the common component $= 1/2$.

$\hat{\mathbf{f}}^{c}_{t,(HIE)}$ obtained assuming $n^{b} = 5$

A further important issue to consider when dealing with maximum likelihood estimation of large factor models is that of computational time. Doz, Giannone and Reichlin (2006) have recently shown that maximum likelihood estimation is computationally feasible for factor models of large cross-sections. We find that the computational time for exact maximum likelihood estimation significantly increases with n but Table 4 shows that this is mainly due to the increase in the number of block-specific factors implicit in the increase in n for any given block-size. In fact, the dependence of computational time on the cross-section is worryingly strong only when the size of the blocks is small, i.e. when the number of blocks increases faster with n: an increase in n from 20 to 250 increases the computational time from around 0.15 to nearly 190 seconds when $n^{b} = 2$ and from 0.23 to only 27 seconds when $n^{b} = 10$. This is encouraging as it shows that computational time is not an obstacle to the applicability of this estimation method in all those cases in which the advantages of exact maximum likelihood estimation with respect to principal components and quasi maximum likelihood are higher, i.e. when the size of the blocks is larger[8].

[8]It would be worth investigating whether other algorithms than the ERM, like for example the gradient projection algorithm (Jamshidian, 2004), can provide faster convergence.

Table 4. Computational time, in seconds*

block-size: $n^{\mathrm{b}} = 10$

	$n = 20$ ($q = 2$)	$n = 50$ ($q = 5$)	$n = 100$ ($q = 10$)	$n = 150$ ($q = 15$)	$n = 200$ ($q = 20$)	$n = 250$ ($q = 25$)
$T = 50$	0.23	0.32	1.25	4.23	11.95	27.73
$T = 100$	0.23	0.30	1.22	4.54	11.51	27.10
$T = 150$	0.24	0.32	1.26	4.54	11.57	26.87

block-size: $n^{\mathrm{b}} = 5$

	$n = 20$ ($q = 4$)	$n = 50$ ($q = 10$)	$n = 100$ ($q = 20$)	$n = 150$ ($q = 30$)	$n = 200$ ($q = 40$)	$n = 250$ ($q = 50$)
$T = 50$	0.21	0.56	2.62	11.34	28.38	59.36
$T = 100$	0.20	0.54	2.53	10.78	27.13	56.54
$T = 150$	0.19	0.56	2.53	10.64	26.89	54.44

block-size: $n^{\mathrm{b}} = 2$

	$n = 20$ ($q = 10$)	$n = 50$ ($q = 25$)	$n = 100$ ($q = 50$)	$n = 150$ ($q = 75$)	$n = 200$ ($q = 100$)	$n = 250$ ($q = 125$)
$T = 50$	0.15	0.76	6.30	30.67	88.90	189.58
$T = 100$	0.14	0.75	6.29	29.67	87.90	186.10
$T = 150$	0.15	0.76	6.10	30.57	87.64	183.93

* Medians across 1000 experiments; number of block-specific factors in parenthesis; Data generating process: $r = 1$, β_i and γ_i i.i.d. $\mathcal{U}([u, \frac{2}{3} - u])$.

5 Empirical application

In our empirical application we use the hierarchical factor model to take account of sectoral heterogeneity in business survey data. We then use the estimated common factor to forecast the manufacturing production index and evaluate its forecasting performance against that of other survey based indicators. The main motivation behind this approach is that while business survey data is widely recognized as an important source of information, essentially because of their timeliness with respect to official data, these two aspects - sectoral heterogeneity and out of sample performance - have not been sufficiently taken into consideration in the existing literature, at least to our knowledge.

The choice of which survey based indicators to compare and the design of the forecast exercise reflect the aim of investigating on the following issues: *(i)* whether business survey data have an out of sample informational content; *(ii)* whether their timeliness plays a role in that; *(iii)* whether their out of sample informational content depends on the way indicators are computed; *(iv)* whether sectoral heterogeneity is important; *(v)* whether aggregating sectoral forecasts does provide more accurate results than just forecasting the aggregate.

In the next Subsection we will give a brief description of the data and illustrate their main characteristics as they emerge from our modelling framework. Then, we will describe the design of the forecast exercise and present the main results.

5.1 Data description

Our sectoral business survey data set consists of the balances of opinions of the five variables included in the Business Climate Indicator computed by the European Commission[9]: production expectations, order books, stock of finished products, export order books and production level.
Survey data is collected as individual qualitative assessments stating whether these variables are at a normal level, below or above it or whether they are equal, higher or lower than the previous period. For some variables, for example the stock of finished products, the firm is asked to make their assessment in terms of four categories instead of three. However, the balance of opinions, the official quantified measure of firms' qualitative assessments, is defined as the difference between the weighted percentage of answers in the two extreme categories[10]. The balances of opinions refer to the 14 subsections of the Italian manufacturing sector according to the Italian version of the NACE Rev1.1 classification (see Table A.1 in the Appendix). The data set spans the period from January 1991 through to January 2007.

On the basis of a preliminary principal components analysis, we set the number of common factors to one and assume that each block is driven by one sector-specific factor[11]. The variance decomposition of the data implied by the model shows that more than 20 per cent of the variance of sectoral survey data is explained by sector-specific shocks. Moreover, variance decomposition block by block shows that the contribution of the three components to the variability in the data is heterogeneous across sectors.

5.2 Forecast exercise. Design and main results

We use the common factor obtained by estimating the hierarchical model to forecast the annual growth rate of the manufacturing production index, $Y_t = 100 \times (y_t - y_{t-12})/y_{t-12}$, and compare its performance to that of other survey based indicators; we also compare two methods of obtaining aggregate forecasts: in the first one Y_t is forecasted directly as an aggregate, while in the second one Y_t is obtained by aggregating sectoral forecasts. The survey based indicators considered as competitors are

- the aggregate and sectoral balances of opinions of production expectations and production levels, BEXP and BLEV;
- the aggregate and sectoral Industrial Confidence Indicators, ICI;
- the aggregate and sectoral Italian Business Climate Indicators, IBCI;
- the common factor and the sector-specific factors obtained from the hierarchical model, CHIE and SHIE;
- the Principal Components Indicator (PCI), given by the first principal component of the data at a sectoral level (only aggregate);

[9]The Business Climate Indicator is the estimated first common factor of a factor model for European survey data. See European Commission (2000).

[10]See Malgarini and Martelli (2005) for a detailed description of the sample design and the weighting scheme adopted to aggregate Italian business survey data.

[11]The first principal component explains 38.8 percent of total variance of the data, the second 6.9 percent, the third 6.5, the fourth 3.8, etc.

- the first common factor of sectoral data estimated by quasi maximum likelihood (only aggregate), DGRI.

The forecast exercise consists in estimating the business survey indicators and computing the backcasts, the nowcasts and the forecasts[12] recursively, on rolling windows of ten years of data. In order to take into account the effects of new data releases, the forecasts are computed three times a month. The first forecasts are computed at the beginning of month T, when official data are available till month $T-3$ and survey data till month $T-1$; the second forecasts are computed in the middle of month T and take into account the release of preliminary data for the manufacturing production at $T-2$ taking place during the second week of the month; finally, the third forecasts are computed the last day of the month and take into account of release of survey data for the current month T taking place during the last week of the month.

Denoting as Π a generic business survey indicator, s the sector of activity, h the forecast horizon, T' the last available observation for the manufacturing production index and T^* the last available observation for survey data ($T' < T^* \leq T$), these are the competing forecasts we take into account

- the constant growth forecast, $\hat{Y}^{(c)}_{T'+h} = \hat{c}_{(c)} = \sum_{t=1}^{T'} Y_t/T'$;
- the autoregressive forecast, $\hat{Y}^{(AR)}_{T'+h} = \hat{c}_{(AR)} + \hat{\alpha}_{(AR)}(L)Y_{T'}$;
- the augmented autoregressive forecast, $\hat{Y}^{(AAR)}_{T'+h} = \hat{c}_{(AAR)} + \hat{\alpha}_{(AAR)}(L)Y_{T'} + \hat{\beta}_{(AAR)}(L)\Pi_{T^*}$;
- the forecast obtained by aggregating the sectoral forecasts $\hat{Y}_{T'+h,s} = \hat{c}_s + \hat{\alpha}_s(L)Y_{T',s} + \hat{\beta}(L)\Pi_{T^*} + \hat{\beta}_s(L)\Pi_{T^*,s}$, $s = 1, ..., 14$

where the length of the backward filters, αs and βs, is set according to the AIC. The prediction accuracy is evaluated by means of the Mean Squared Forecast Errors (MSFE) computed over the period March 2001 through January 2007.

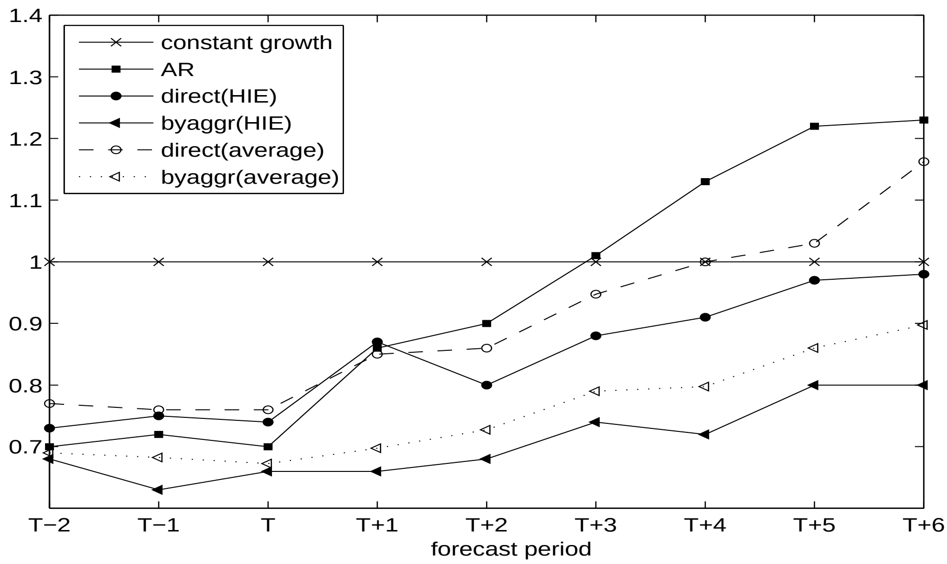
Figure 1 shows the results concerning the forecasts made at the beginning of the month using the common factor and the sector-specific factors obtained by modelling the block-structure of the data; the dashed and the dotted lines refer to the average performance of the forecasts obtained by using the aggregate survey based indicators (direct augmented forecasts) and of the forecasts computed by aggregation of sectoral forecasts (aggregated augmented forecasts), respectively[13]. The results are expressed in terms of relative MSFE with the constant growth forecasts used as benchmark, so that numbers below/above one indicate a better/worse performance than the constant growth forecasts. The following main conclusions emerge: the MSFE of the augmented forecasts obtained using the hierarchical indicators is lower than the average MSFE obtained using alternative indicators; the forecasting performance deteriorates with the forecast horizon; autoregressive forecasts deteriorate faster than those obtained by using the additional information coming from survey data; the forecasting performances are similar as far as backcasts ($T-2$ and $T-1$) and nowcasts (T) are concerned and more differentiated at longer horizons; direct augmented forecasts are more accurate than autoregressive forecasts only from two step ahead on; at all forecast horizons, the lowest MSFE is reached by aggregating sectoral forecasts: the gap in the MSFE with respect to direct forecasts is

[12] Hereafter, we sometime use the term 'forecasts' in its general sense, including backcasts, nowcasts and forecasts.

[13] See the Appendix for a detailed table of the results (Table A.2).

relevant from one step ahead on, it increases with respect to the autoregressive forecasts and it is constant with respect to the direct augmented forecasts.

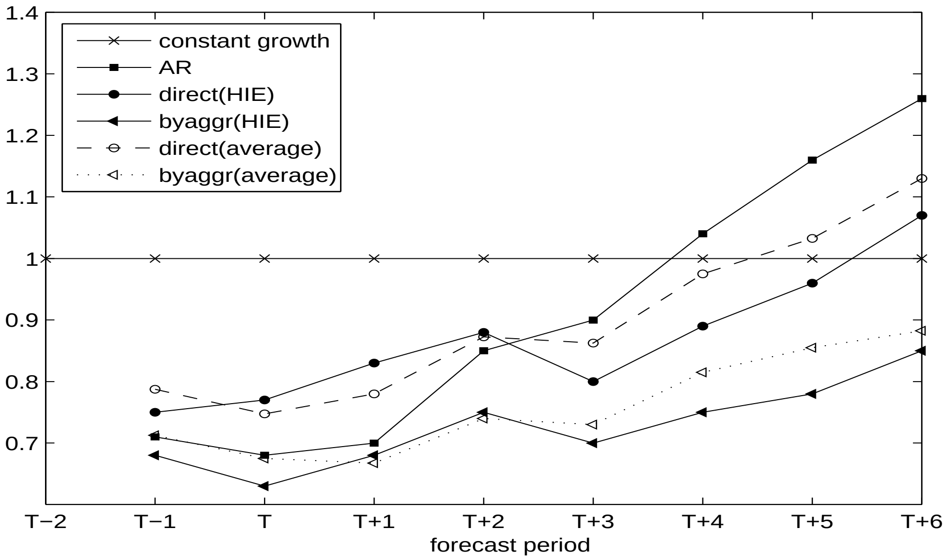
Figure 1 - Forecasts at the beginning of month T *



* direct(HIE): MSFEs of direct forecasts augmented with the hierarchical common factor; direct(average): average MSFEs of direct forecasts augmented with the hierarchical common factor; byaggr(HIE): MSFEs of forecasts obtained by aggregating sectoral forecasts augmented with the sector-specific factors; byaggr(average): average MSFEs of forecasts obtained by aggregating sectoral forecasts augmented with the sector-specific factors. Evaluation period: March-2001 to January-2007.

Figure 2 shows the results obtained by forecasting in the middle of the month after the release of manufacturing production index for $T-2$. The main conclusions emerged by analyzing forecasts made at the beginning of month T hold: forecasts deteriorate with respect to the constant growth forecasts as the forecast horizon increases and the aggregation of sectoral forecasts gives the best results at all forecast horizons (see Table A.2 in the Appendix for detailed results). By comparing Figure 1 and Figure 2 it also emerges that the release of official data on the manufacturing sector at $T-2$ has a negligible effect on the augmented forecasts while it significantly improves the autoregressive forecasts from one step ahead on. As a result, using survey indicators in augmented direct forecasts only provides some advantages over the autoregressive forecasts starting on three step ahead, while using them in augmented aggregated forecasts entails a lower MSFE already at two step ahead.

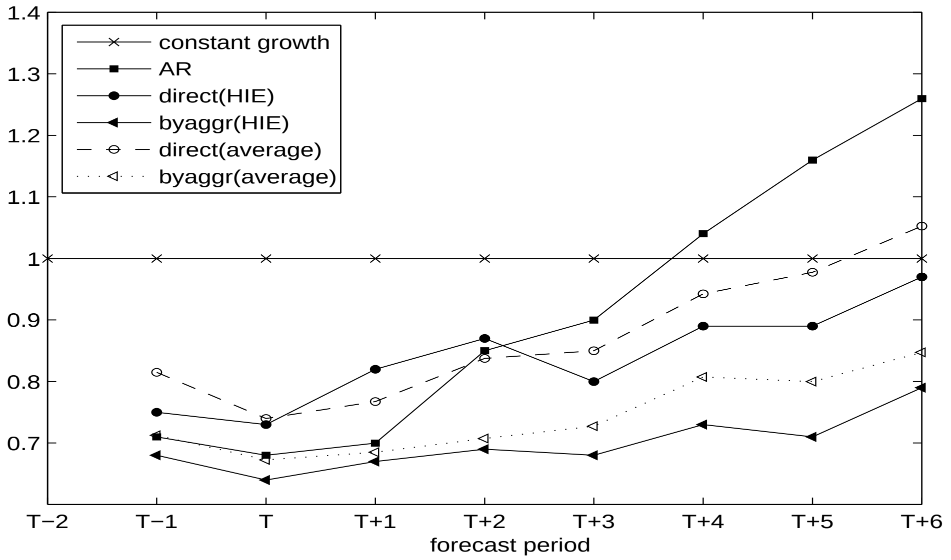
Figure 2 - Forecasts in the middle of month T *



* direct(HIE): MSFEs of direct forecasts augmented with the hierarchical common factor; direct(average): average MSFEs of direct forecasts augmented with the hierarchical common factor; byaggr(HIE): MSFEs of forecasts obtained by aggregating sectoral forecasts augmented with the sector-specific factors; byaggr(average): average MSFEs of forecasts obtained by aggregating sectoral forecasts augmented with the sector-specific factors. Evaluation period: March-2001 to January-2007.

Forecasts made at the end of the month benefit from one additional observation of survey data, that of the current month T. As shown in Figure 3, such an effect is nearly absent in backcasts and nowcasts while, though limited and not homogeneous across competing forecasts, it has some relevance for longer horizons (see Table A.4 in the Appendix for detailed results).

Figure 3 - Forecasts at the end of month T *



* direct(HIE): MSFEs of direct forecasts augmented with the hierarchical common factor; direct(average): average MSFEs of direct forecasts augmented with the hierarchical common factor; byaggr(HIE): MSFEs of forecasts obtained by aggregating sectoral forecasts augmented with the sector-specific factors; byaggr(average): average MSFEs of forecasts obtained by aggregating sectoral forecasts augmented with the sector-specific factors. Evaluation period: March-2001 to January-2007.

The analysis of Figures 1 to 3 focused on the forecasts obtained using our hierarchical indicators with particular attention to three aspects: the comparison with autoregressive forecasts, the comparisons between direct and aggregated augmented forecasts and the effect of data releases. The qualitative results on the ranking of these three alternative forecast methods - autoregressive, direct augmented and aggregated augmented - for hierarchical indicators hold when extended to the other survey based indicators. Such an analysis helped us answer some of the issues that motivated this forecasting exercise: survey based indicators help forecasting the manufacturing production index especially at longer longer horizons; the release of survey data has a limited effect on the augmented forecasts and it is concentrated at longer horizons; aggregating sectoral forecasts always provide better results than directly forecasting the aggregate.

Table 5 sheds light on two further issues: whether multivariate indicators should be preferred to univariate ones and whether aggregate indicators computed using sectoral data produce better direct forecasts than those computed using aggregate data. The Table shows the average relative MSFEs across forecasts augmented with univariate (BEXP and BLEV) and multivariate (ICIs, IBCIs, CHIE, SHIEs, PCIs and DGRI) indicators, and across direct forecasts augmented with indicators computed using sectoral data (CHIE and SHIEs, PCI and DGRI) and aggregate data (BEXP and BLEV, ICI and IBCI). Multivariate indicators outperform univariate ones and aggregate indicators computed using sectoral data outperform those computed with aggregate data. This result together with the good performance of aggregated versus direct augmented forecasts suggest that indeed sectoral heterogeneity represents a very important source of information in survey data.

Table 5. Characteristics of the indicators and forecasting performance*

Forecasts at the beginning of month T

	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
Univariate	0.75	0.73	0.72	0.77	0.81	0.92	0.99	1.04	1.12
Multivariate	0.73	0.71	0.71	0.76	0.76	0.83	0.84	0.88	0.94
Aggregate data	0.77	0.76	0.76	0.85	0.86	0.95	1.00	1.03	1.16
Sectoral data	0.77	0.76	0.75	0.81	0.78	0.88	0.93	0.95	0.97

Forecasts in the middle of month T

	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
Univariate	-	0.76	0.72	0.72	0.81	0.82	0.96	1.02	1.09
Multivariate	-	0.75	0.72	0.73	0.80	0.77	0.84	0.89	0.95
Aggregate data	-	0.79	0.75	0.78	0.87	0.86	0.98	1.03	1.13
Sectoral data	-	0.79	0.77	0.77	0.85	0.80	0.91	0.96	1.01

Forecasts at the end of month T

	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
Univariate	-	0.76	0.72	0.73	0.78	0.80	0.94	0.97	1.03
Multivariate	-	0.76	0.71	0.73	0.76	0.76	0.82	0.83	0.89
Aggregate data	-	0.82	0.74	0.77	0.84	0.85	0.94	0.98	1.05
Sectoral data	-	0.79	0.76	0.78	0.81	0.78	0.88	0.90	0.95

* Average relative MSFEs. Univariate: direct and aggregated forecasts augmented with univariate indicators (aggregate and sectoral BEXPs and BLEVs); Multivariate: direct and aggregated forecasts augmented with multivariate indicators (aggregate and sectoral ICIs, aggregate and sectoral IBCIs, CHIE, SHIEs, PCI, DGRI); Aggregate data: direct forecasts augmented with indicators based on aggregate survey data (aggregate BEXP and BLEV, aggregate ICI, aggregate IBCI); Sectoral data: direct forecasts augmented with indicators based on sectoral survey data (PCI, DGRI and CHIE).
Evaluation period: March-2001 to January-2007.

6 Conclusions

This paper investigates the finite sample properties of three alternative estimators of the common factors when the data has a block structure. It shows that in finite samples explicitly modeling the block structure of the data can improve the quality of the estimates of the common factors with respect to principal components and quasi maximum likelihood. How large n has to be to make the three estimators equivalent in terms of efficiency crucially depends on the size of the blocks: the larger the higher n has to be to make quasi maximum likelihood principal components as precise as exact maximum likelihood. On the other hand, it is shown that erroneously fitting a hierarchical factor model when the data does not have a block structure does not significantly worsen the quality of the estimates.

A further contribution of the paper consists of providing a comparative evaluation of the forecasting performance of business survey based indicators. The forecast exercise, performed on rolling windows of ten years, reaches the following conclusions: the out of sample informational content of survey data increases with the forecast horizon; the release of new survey data has a minor effect on the performance of augmented forecasts; at all forecast horizons, aggregating sectoral forecasts provide better results than directly forecasting the aggregate; multivariate survey based indicators perform better than univariate ones and, finally, indicators computed using sectoral data perform better than those obtained using aggregate data.

The analysis in the paper suggested several directions for further research in the field, some of which are already in our current research agenda.

References

[1] Beck Gunter W., K. Ubrich and M. Marcellino, 2006, Regional inflation dynamics within and across euro area countries and a comparison with the U.S., European Central Bank WP, n. 681;

[2] Boivin J. and S. Ng, 2006, Are more data always better for factor analysis?, Journal of Econometrics, Vol. 132;

[3] Brooks R. and M. Del Negro, 2004, A latent factor model with global, country and industry shocks for international stock returns, Federal Reserve Bank of Atlanta WP, n. 2002-23b

[4] Deroose S., P. Mills and B. Saint Aubin, 2001, Business climate indicator for the euro area, Paper presented for the CEPR/Banca d'Italia Conference on Monitoring the euro area business cycle, 6-7 September 2001, Rome;

[5] Doz C., D. Giannone and L. Reichlin, 2006, A quasi maximum likelihood approach for large approximate dynamic factor models, European Central Bank WP, N. 674;

[6] Doz C. and F. Lenglart, 2001, Dynamic factor analysis : estimation and test with an application to european business surveys, Paper presented for the CEPR/Banca d'Italia Conference on Monitoring the euro area business cycle, 6-7 September 2001, Rome;

[7] European Commission, 2000, Business climate indicator for the euro area, Presentation Paper (November 2000), available at http://europa.eu.int/comm/economy_finance/indicators/business_climate/2001/presentation_climate.pdf;

[8] Forni M. and L. Reichlin, 2001, Federal policies and local economies: Europe and the US, European Economic Review, Vol. 45;

[9] Gayer C. and J. Genet, 2006, Using factor models to construct composite indicators from BCS data. A comparison with European Commission confidence indicators, European Economy Economic Papers, N. 240, European Commission;

[10] Geweke John F. and Kenneth J. Singleton, 1980, Maximum likelihood confirmatory factor analysis of Economic Time Series, International Economic Review, Vol. 22;

[11] Ghahramani Z. and Geoffrey E. Hinton, 1996, Parameter estimation for linear dynamical systems. Technical report, Manuscript, University of Toronto, available at http://www.gatsby.ucl.ac.uk/ zoubin;

[12] Hallin M. and R. Liska, 2008, Dynamic factors in the Presence of Block Structure, EUI Working Papers, ECO 2008/22;

[13] Jamshidian M., 2004, On algorithms for restricted maximum likelihood estimation, Computational Statistics & Data Analysis, Vol. 45;

[14] Kose M. Ayan, C. Otrok and Charles H. Whiteman, 2003, International business cycles: world, region and country-specific factors, American Economic Review, Vol. 93;

[15] Malgarini M. and Bianca M. Martelli, 2005, Re-engineering the ISAE manufacturing survey, ISAE Working Paper Series, N. 47;

[16] Marcellino M., 2006, Dynamic factor models for survey-based confidence indicators. Final Report, mimeo, Report prepared for the European Commission;

[17] Ng S., E. Moench and S. Potter, 2008, Dynamic Hierarchical factor Models, mimeo, Columbia University;

[18] Onatski A., 2007, Asymptotics of the Principal Components Estimator of Large Factor Model with Weak Factors and i.i.d. Gaussian Noise, mimeo, Columbia University;

[19] Sargent Thomas J. and C. Sims, 1977, Business cycle modelling without pretending to have to much a-priori economic theory. In Christopher Sims, editor, New Methods in Business Cycle Research, Federal Reserve Bank of Minneapolis;

[20] Stock James H. and M. Watson, 1989, New indices of coincident and leading indicators, in O.J. Blanchard and S. Fischer (eds), NBER Macroeconomics Annual 1989, Cambridge: MIT Press;

[21] Stock James H. and M. Watson, 2002, Forecasting using principal components from a large numbers of predictors, Journal of the American Statistical Association, Vol. 97, n. 460;

[22] Stock James H. and M. Watson, 2002a, Macroeconomic forecasting using diffusion indeces, Journal of Business and Economic Statics, Vol. 20, pp. 147-162;

[23] Stock James H. and M. Watson, 2008, The evolution of national and regional factors in U.S. housing construction, prepared for the Volume *Essays in Volatility in Finance and Economics, Time Series, and Regional Economics: A Festschrift in Honor of Robert F. Engle.*

A Appendix

A.1 Sectoral disaggregation

Table A.1 - NACE.Rev1.1 subsections

Code	description
DA	Food products, beverages and tobacco
DB	Textiles and textile products
DC	Leather and leather products
DD	Wood and products of wood and cork (except furniture); articles of straw and plaiting materials
DE	Pulp, paper and paper products; recorded media; printing services
DF	Coke, refined petroleum products and nuclear fuel
DG	Chemicals, chemical products and man-made fibres
DH	Rubber and plastic products
DI	Other non metallic mineral products
DJ	Basic metals and fabricated metal products
DK	Machinery and equipment n.e.c.
DL	Electrical and optical equipment
DM	Transport equipment
DN	Other manufactured goods n.e.c.

A.2 Forecast results

Table A.2 - Forecasts at the beginning of month T *

	Non augmented forecasts								
	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
constant growth (MSFE)	22.89	21.32	22.32	22.43	22.50	22.81	22.57	22.57	22.89
AR	0.70	0.72	0.70	0.86	0.90	1.01	1.13	1.22	1.23
	Direct augmented forecasts								
	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
BEXP	0.74	0.75	0.75	0.89	0.91	1.09	1.24	1.30	1.36
BLEV	0.87	0.82	0.79	0.80	0.84	0.94	0.98	1.00	1.12
ICI	0.72	0.73	0.74	0.87	0.87	0.94	0.90	0.91	1.03
IBCI	0.75	0.74	0.76	0.84	0.82	0.82	0.88	0.91	1.14
CHIE	0.73	0.75	0.74	0.87	0.80	0.88	0.91	0.97	0.98
PCI	0.81	0.76	0.75	0.80	0.79	0.87	0.92	0.91	0.95
DGRI	0.78	0.76	0.75	0.77	0.75	0.90	0.96	0.96	0.97
	Aggregated augmented forecasts								
	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
BEXPs	0.67	0.68	0.66	0.72	0.77	0.88	0.94	0.99	1.02
BLEVs	0.70	0.68	0.68	0.68	0.71	0.78	0.78	0.86	0.99
ICIs	0.68	0.67	0.67	0.71	0.75	0.79	0.75	0.80	0.80
IBCIs	0.78	0.68	0.68	0.68	0.68	0.71	0.72	0.79	0.78
HIEs	0.61	0.66	0.66	0.66	0.68	0.74	0.72	0.80	0.80

* Relative MSFE. Evaluation period: March-2001 to January-2007.

Table A.3 - Forecasts in the middle of month T *

	Non augmented forecasts								
	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
constant growth (MSFE)	-	21.24	22.20	22.25	22.41	22.68	22.47	22.49	22.14
AR	-	0.71	0.68	0.70	0.85	0.90	1.04	1.16	1.26
	Direct augmented forecasts								
	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
BEXP	-	0.77	0.73	0.79	0.91	0.92	1.15	1.23	1.35
BLEV	-	0.86	0.77	0.76	0.85	0.84	0.96	1.02	1.10
ICI	-	0.75	0.73	0.83	0.89	0.85	0.92	0.96	1.08
IBCI	-	0.77	0.76	0.74	0.84	0.84	0.87	0.92	0.99
CHIE	-	0.75	0.77	0.83	0.88	0.80	0.89	0.96	1.07
PCI	-	0.81	0.76	0.74	0.83	0.77	0.91	0.94	0.98
DGRI	-	0.82	0.77	0.75	0.83	0.84	0.94	0.99	0.98
	Aggregated augmented forecasts								
	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
BEXPs	-	0.68	0.67	0.67	0.75	0.80	0.95	0.98	0.95
BLEVs	-	0.68	0.69	0.64	0.74	0.70	0.77	0.85	0.95
ICIs	-	0.67	0.68	0.67	0.73	0.73	0.82	0.82	0.82
IBCIs	-	0.68	0.66	0.69	0.74	0.69	0.72	0.77	0.81
HIEs	-	0.66	0.63	0.68	0.75	0.70	0.75	0.78	0.85

* Relative MSFE. Evaluation period: March-2001 to January-2007.

Table A.4 - Forecasts at the end of month T *

	Non augmented forecasts								
	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
constant growth (MSFE)	-	21.24	22.20	22.25	22.41	22.68	22.47	22.49	22.14
AR	-	0.71	0.68	0.70	0.85	0.90	1.04	1.16	1.26
	Direct augmented forecasts								
	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
BEXP	-	0.77	0.73	0.77	0.91	0.88	1.11	1.17	1.28
BLEV	-	0.86	0.80	0.78	0.80	0.84	0.94	0.96	1.02
ICI	-	0.75	0.71	0.76	0.87	0.85	0.90	0.91	0.98
IBCI	-	0.76	0.72	0.76	0.77	0.83	0.82	0.87	0.93
CHIE	-	0.75	0.73	0.82	0.87	0.80	0.89	0.89	0.97
PCI	-	0.81	0.77	0.77	0.79	0.78	0.87	0.89	0.92
DGRI	-	0.80	0.79	0.76	0.76	0.75	0.89	0.92	0.96
	Aggregated augmented forecasts								
	$T-2$	$T-1$	T	$T+1$	$T+2$	$T+3$	$T+4$	$T+5$	$T+6$
BEXPs	-	0.69	0.67	0.67	0.72	0.77	0.92	0.95	0.97
BLEVs	-	0.72	0.68	0.69	0.69	0.71	0.79	0.78	0.86
ICIs	-	0.72	0.65	0.69	0.72	0.74	0.81	0.77	0.79
IBCIs	-	0.72	0.69	0.69	0.70	0.69	0.71	0.70	0.77
HIEs	-	0.68	0.64	0.67	0.69	0.68	0.73	0.71	0.79

* Relative MSFE. Evaluation period: March-2001 to January-2007.