The Generalized Dynamic Factor Model one-sided estimation and forecasting *

Mario Forni,
Universitá di Modena and CEPR

Marc Hallin
ISRO, ECARES, and Département de Mathématique
Université Libre de Bruxelles

Marco Lippi
Universitá di Roma La Sapienza

and

Lucrezia Reichlin
ECARES, Université Libre de Bruxelles and CEPR

January 31, 2003

Abstract

This paper proposes a new forecasting method that exploits information from a large panel of time series. The method is based on the generalized dynamic factor model proposed in Forni, Hallin, Lippi, and Reichlin (2000), and takes advantage of the information on the dynamic covariance structure of the whole panel. We first use our previous method to obtain an estimation for the covariance matrices of common and idiosyncratic components. The generalized eigenvectors of this couple of matrices are then used to derive a consistent estimate of the optimal forecast. This two-step approach solves the end-of-sample problems caused by two-sided filtering (as in our previous work), while retaining the advantages of an estimator based on dynamic information. The relative merits of our method and the one proposed by Stock and Watson (1999) are discussed.

JEL subject classification : C13, C33, C43. Key words and phrases : Dynamic factor models, principal components, time series, large cross-sections, panel data, forecasting.

1 Introduction

Economists and forecasters nowadays typically have access to information scattered through huge numbers of observed time series—aggregated and disaggregated, real and nominal variables. Intuition suggests that disregarding potentially useful information always produces suboptimal

*Research supported by a P.A.I of the Belgian Federal Government, an A.R.C. contract of the Communauté française de Belgique, and the Training and Mobility of Researchers Programme of the European Commission (Contract ERBFMRX-CT98-0213).

forecasts; the more scattered the information, the more severe this loss of forecasting efficiency. Yet, most multivariate forecasting methods in the literature are restricted to series of low dimension, and allow for incorporating only a limited number of key variables. Such methods are thus of little help in large panels of time series, where the cross-sectional dimension is often of the same order as, or even larger than the series lengths. The challenge for econometricians is to develop alternative techniques that are sufficiently powerful as to overcome this dimensionality problem, yet flexible enough to provide an adequate picture of economic reality.

Recently, Forni and Reichlin (1998), Stock and Watson (1999), Forni, Hallin, Lippi, and Reichlin (2000, 2001, 2003), Forni and Lippi (2001) have developed factor model techniques that are tailored to exploit a large cross-sectional dimension. Under such models, each time series in the panel is represented as the sum of two mutually orthogonal components : the *common* component, which is "strongly correlated" with the rest of the panel and has reduced stochastic dimension, and the *idiosyncratic* component. These idiosyncratic components are either mutually orthogonal or "mildly cross-correlated" across the panel. The common component is (non-parametrically) consistently estimated as both the size n of the cross-section and the series length T go to infinity. These results are obtained under conditions that look reasonable in empirical situations, whenever there are comovements between time series, as it is generally the case for macroeconomic data (for a documentation of this point, see Altissimo et al. 2001).

In a factor model, multivariate information can help forecasting the common component, while the idiosyncratic, being mildly cross-correlated, can be reasonably well predicted by means of traditional univariate methods (or methods based on low dimension models such as VARs). Therefore, the forecast of the future of any given series in the panel can be obtained as the sum of the forecast of the common component, where we exploit multivariate information, and the forecast of the idiosyncratic component, where multivariate information safely can be disregarded. The common component being of reduced stochastic dimension, its forecast can be expressed as a projection on the span of a small number of appropriately constructed aggregates, whereas each idiosyncratic component can be treated by means of standard univariate or low-dimensional forecasting methods.

The two methods proposed in the literature, Stock and Watson (1999) and Forni, Hallin, Lippi and Reichlin (2000) are both based on this general idea, but each of them presents a serious shortcoming. Forni, Hallin, Lippi and Reichlin (2000) base their estimation of the common and idiosyncratic components on the dynamic covariance structure of the data. This is a highly desirable feature, since economic time series in general are non-synchronized, and the leading variables should play a crucial role in the forecast of the lagging ones. Typically, provided that leading series are included in the panel, such methods should allow for forecasting even those lagging series that are unforecastable at univariate level. The estimator in Forni, Hallin, Lippi an Reichlin (2000), however, is derived from the spectral density of the data and, as a consequence, is based on a two-sided filtering of the observations (linear combination of present, past and future observations). This two-sidedness feature is not a problem when within-sample estimation of the common component is the objective; but it is most unpleasant in the forecasting context, since at the end of the sample future observations are not available.

Stock and Watson (1999), on the other hand, propose to estimate the common component by projecting onto the static principal components of the data. Their estimator relies on a one-sided filtering of the observations, which does not cause any problems at the end of the sample. However, being based on contemporaneous covariances only, it fails to exploit the potentially crucial information contained in the leading-lagging relations between the elements of the panel.

The method we propose in this paper aims at combining the advantages of both approaches

and consists of the following two steps. Firstly, using the dynamic techniques developed in Forni, Hallin, Lippi and Reichlin (2000), we obtain estimates of common and idiosyncratic variance-covariance matrices at all leads and lags as inverse Fourier transforms of the corresponding estimated spectral density matrices. Secondly, we use these estimates in the construction of the contemporaneous linear combinations of the observations having smallest idiosyncratic-common variance ratio. The resulting aggregates can be obtained as the solution of a *generalized principal component* problem (see Section 4). Our h-step ahead forecast is obtained as the projection of the h step ahead observation onto these estimated generalized principal components.

We prove, under a finite lag structure for the common components, that this forecast is a consistent estimator (for n and T going to infinity) of the optimal h-step ahead forecast. The same method can be applied to re-estimate the in-sample common component without making use of future observations, thus solving the end-of-sample problem arising in Forni, Hallin, Lippi and Reichlin (2000).

Both Stock and Watson's estimators and ours are one-sided linear combinations of the observations, but the weighting schemes used are different. Both provide a consistent forecast. Stock and Watson's predictor is simpler, while ours exploits more information. We discuss the relative performances of the two methods and argue that the frequency domain method should perform better when the various cross-sectional items differ significantly in the lag structure of the factor loadings, particularly if, in addition, there is substantial heterogeneity the fraction of total variance explained by the idiosyncratic components. We therefore recommend to use our frequency domain method when we expect *a priori* such features in the data set or whenever such features are detected via the static principal component method. These indications are confirmed by our simulation exercise.

The paper is organized as follows. In Section 2 we present the model. In Section 3 we illustrate the existing estimation methods. In Section 4 we present our proposed predictor. In Section 5 we proof consistency. In Section 6 we discuss the relative merits of the competing estimators. In Section 7 we present simulation results. Section 8 concludes. Some technical material is collected in the Appendix.

2 The model

In this paper we consider a specialization of the generalized dynamic factor model of Forni, Hallin, Lippi and Reichlin (2000) and Forni and Lippi (2001). Such models, and the one used here, differ from the traditional dynamic factor model of Sargent and Sims (1977) and Geweke (1977), in that the number of cross-sectional variables is infinite and the idiosyncratic components are allowed to be mutually correlated to some extent, along the lines of Chamberlain (1983) and Chamberlain and Rothschild (1983). Similar models have been recently proposed by Stock and Watson (1999) and Bai and Ng (2002).

Denote by $\mathbf{X}_n^T = (x_{it})_{i=1,\dots,n;\, t=1,\dots,T}$ an $n \times T$ rectangular array of observations. Throughout, we assume that

A1. $\mathbf{X}_n^T$ is a finite realization of a real-valued stochastic process $\mathbf{X} = \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ indexed by $\mathbb{N} \times \mathbb{Z}$, where the n-dimensional vector processes $\{\mathbf{x}_{nt} = (x_{1t} \ \cdots \ x_{nt})', \ t \in \mathbb{Z}\}$, $n \in \mathbb{N}$ are stationary, with zero mean and finite second-order moments $\mathbf{\Gamma}_{nk} = \mathrm{E}[\mathbf{x}_{nt}\mathbf{x}'_{n,t-k}]$, $k \in \mathbb{N}$.

The spectral techniques to be used in the sequel require in addition the following technical assumption

A2. For all $n \in \mathbb{N}$, the process $\{\mathbf{x}_{nt}, t \in \mathbb{Z}\}$ admits a Wold representation $\mathbf{x}_{nt} = \sum_{k=0}^{\infty} \mathbf{C}_k^n \mathbf{w}_{n,t-k}$, where the full-rank innovations $\mathbf{w}_{nt}$ have finite moments of order four, and the matrices $\mathbf{C}_k^n = (C_{ij,k}^n)$ satisfy $\sum_{k=0}^{\infty} |C_{ij,k}^n| k^{1/2} < \infty$ for all $n, i, j \in \mathbb{N}$.

Assumptions A1 and A2 jointly will be referred to as assumption A.

The process x_{it} is the sum of two unobservable components, the *common component* χ_{it} and the *idiosyncratic component* ξ_{it}. The common component is driven by a q-dimensional vector of *common factors* $\mathbf{f}_t = (f_{1t}\ f_{2t}\ \cdots\ f_{qt})$, which however are loaded with possibly different coefficients and lags:

$$\chi_{it} = b_{i1}(L) f_{1t} + b_{i2}(L) f_{2t} + \cdots + b_{iq}(L) f_{qt}.$$

Note that q is independent of n (and small as compared to n in empirical applications). More precisely, defining $\boldsymbol{\chi}_{nt} = (\chi_{1t}\ \ldots\ \chi_{nt})'$ and $\boldsymbol{\xi}_{nt} = (\xi_{1t}\ \ldots\ \xi_{nt})'$, our model is

$$\begin{aligned} \mathbf{x}_{nt} &= \boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt} \\ &= \mathbf{B}_n(L)\mathbf{f}_t + \boldsymbol{\xi}_{nt}, \end{aligned} \tag{2.1}$$

where the factors $\mathbf{f}_t$ follow a VAR scheme of the form

$$\mathbf{A}(L)\mathbf{f}_t = \mathbf{u}_t,$$

and

B1. (a) $\mathbf{B}_n(L) = \mathbf{B}_0^n + \mathbf{B}_1^n L + \ldots + \mathbf{B}_s^n L^s$ is a $n \times q$ polynomial of order s in the lag operator L, with $\mathbf{B}_s^m \neq \mathbf{0}$, for some m; (b) $\mathbf{A}(L) = \mathbf{I} - \mathbf{A}_1 L - \ldots - \mathbf{A}_S L^S$ a $q \times q$ polynomial of order $S \leq s+1$; (c) all solutions of $\det[\mathbf{A}(z)] = 0$, $z \in \mathbb{C}$, lie outside the unit ball;

B2. $\{\mathbf{u}_t = (u_{1t}\ \ldots\ u_{qt})', t \in \mathbb{Z}\}$, the vector of *common shocks*, is a q-dimensional orthonormal white noise process orthogonal to $\{\xi_{it},\ i = 1, \ldots, n,\ t \in \mathbb{Z}\}$.

Of course the matrices $\mathbf{B}_j^n$ are nested as n increases, so that $\mathbf{B}_s^n \neq \mathbf{0}$ for all $n > m$. Orthogonality between common and idiosyncratic components (Assumption B2) is a standard assumption in factor models literature. The assumption on the characteristic roots of $\mathbf{A}(L)$ guarantees the existence of the inverse operator $[\mathbf{A}(L)]^{-1}$. We shall return to condition B1(b) in the next section.

We will need the following additional assumptions. Let $\boldsymbol{\Sigma}_n^{\chi}(\theta)$, $\boldsymbol{\Sigma}_n^{\xi}(\theta)$, $\theta \in [-\pi, \pi]$, be the spectral density matrices of $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$, respectively, and λ_{nk}^{χ}, λ_{nk}^{ξ} the corresponding *dynamic eigenvalues*, namely, the mappings $\theta \mapsto \lambda_{nk}^{\chi}(\theta)$ and $\theta \mapsto \lambda_{nk}^{\xi}(\theta)$, where $\lambda_{nk}^{\chi}(\theta)$ and $\lambda_{nk}^{\xi}(\theta)$ stand for the k-th largest eigenvalues of $\boldsymbol{\Sigma}_n^{\chi}(\theta)$ and $\boldsymbol{\Sigma}_n^{\xi}(\theta)$, respectively. We assume:

C1. (a) $\lambda_{nq}^{\chi}(\theta) \to \infty$ as $n \to \infty$, θ-a.e. in $[-\pi\ \pi]$; (b) $\lambda_{nk}^{\chi}(\theta) > \lambda_{n,k+1}^{\chi}(\theta)$ θ-a.e. in $[-\pi\ \pi]$, $k = 1, \ldots, q$;

C2. There exists a real Λ such that $\lambda_{n1}^{\xi}(\theta) \leq \Lambda$ for any $\theta \in [-\pi\ \pi]$ and any $n \in \mathbb{N}$.

Assumption C1(b) requires that the first $q+1$ eigenvalues are distinct for almost all θ (note that $\lambda_{nj}^{\chi}(\theta) = 0$ for $j > q$ and all θ). It makes proofs easier while not causing a serious loss of generality. Assumptions C1(a) and C2 are needed to guarantee identification of the common and the idiosyncratic components (see Forni and Lippi, 2001). Note that condition C2 on the asymptotic behavior of $\lambda_{nk}^{\xi}(\theta)$ includes the case in which the idiosyncratic components are mutually orthogonal with an upper bound for the variances. Mutual orthogonality is a standard,

though highly unrealistic assumption in factor models; condition C2 relaxes such assumption by allowing for a limited amount of cross-correlation among the idiosyncratic components.

Finally, let $\mathbf{\Gamma}^{\chi}_{nk}$, $\mathbf{\Gamma}^{\xi}_{nk}$ be the k-lag covariance matrices of the vectors $\boldsymbol{\chi}_{nt}$, $\boldsymbol{\xi}_{nt}$ and let μ^{χ}_{nj}, μ^{ξ}_{nj} be the j-th eigenvalues of $\mathbf{\Gamma}^{\chi}_{n0}$, $\mathbf{\Gamma}^{\xi}_{n0}$, respectively. We assume that

D1. (a) $\lim_{n\to\infty} \mu^{\chi}_{nr} = \infty$, where $r = q(s+1)$; (b) $\mu^{\chi}_{nk} > \mu^{\chi}_{n,k+1}$, $k = 1, \ldots, r$;

D2. μ^{ξ}_{nn} is bounded away from zero as $n \to \infty$.

Thus D1(b) assumes that the first $r+1$ eigenvalues are distinct. D1(b) and D2 are assumed for mathematical convenience. Like C1(b), they do not rule out any empirically interesting cases. Assumption D1(a) rules out the case in which some of the elements in $\mathbf{f}_{t-k}$, $k = 0, \ldots, s$, are loaded only by a finite number of the x's. Note that C1(a) does not imply D1(a); for example, if $\chi_{1t} = u_{t-1}$ and $\chi_{it} = u_t$ for $i \geq 2$, C1(a) clearly holds, with $q = 1$, but D1(a) does not hold, since $\mu^{\chi}_{n1} \to \infty$ whereas μ^{χ}_{n2} is bounded as $n \to \infty$.

Letting $\mathbf{F}_t = (\mathbf{f}_t' \ \mathbf{f}_{t-1}' \ \ldots \ \mathbf{f}_{t-s}')'$ and $\mathbf{C}_n = (\mathbf{B}^n_0 \ \mathbf{B}^n_1 \ \cdots \ \mathbf{B}^n_s)$, the same model (2.1) can also be written under the form

$$\mathbf{x}_{nt} = \mathbf{B}_n(L)\mathbf{f}_t + \boldsymbol{\xi}_{nt} = \mathbf{C}_n\mathbf{F}_t + \boldsymbol{\xi}_{nt} \tag{2.2}$$

involving $r = q(s+1)$ common factors, loaded only contemporaneously. Equation (2.2) looks like a static factor model. However, the dynamic nature of (2.1) implies that $\mathbf{F}_t$ has a special structure: indeed, the spectral density matrix of $\mathbf{F}_t$ has rank q, which is smaller than r if $s > 0$. In the sequel we call 'static factors' the factors of the static representation (2.2), i.e. the r entries of $\mathbf{F}_t$, and 'dynamic factors' q the entries of $\mathbf{f}_t$.

3 Two competing estimation methods

Stock and Watson (1999) and Forni, Hallin, Lippi and Reichlin (2000) have proposed estimators for model (2.1). We will label them *static* and *dynamic* principal component method respectively.

In both cases the goal is either estimation of the common component or forecasting of the x's. Regarding forecasting, since both methods are based on the common-idiosyncratic decomposition, and since these two components are mutually orthogonal at any lead and lag, forecasting the x's can be split into two separate forecasting problems, for the common and for the idiosyncratic components, respectively. Moreover, since the idiosyncratic components are mutually orthogonal or only weakly correlated, their forecast can be obtained from standard univariate or low-dimensional multivariate methods. Thus the analysis is exclusively concentrated on forecasting the common component. Let us now present the two methods in detail.

Let $\mathbf{\Gamma}^T_{nk}$ be the k-lag sample covariance matrix of $\mathbf{x}_{nt}$ (i.e. $\mathbf{\Gamma}^T_{nk} = (n-k)^{-1}\sum_{t=k+1}^{T} \mathbf{x}_{nt}\mathbf{x}_{n,t-k}$), m^T_{nj} the j-th largest eigenvalue of $\mathbf{\Gamma}^T_{n0}$ and $\mathbf{S}^T_{nj}$ the corresponding row eigenvector. Moreover, let $\mathbf{M}^T_n$ be the $r \times r$ diagonal matrix the above eigenvalues, in descending order, on the diagonal, and $\mathbf{S}^T_n$ the $r \times n$ matrix $\left(\mathbf{S}^{T\prime}_{n1} \ \cdots \ \mathbf{S}^{T\prime}_{nr}\right)'$. The h-step ahead forecast of $\boldsymbol{\chi}_{nt}$ based on the static principal components is

$$\mathbf{\Gamma}^T_{nh}\mathbf{S}^{T\prime}_n(\mathbf{M}^T_n)^{-1}\mathbf{S}^T_n\mathbf{x}_{nT}. \tag{3.1}$$

Analogously, the in-sample estimator of the common component $\boldsymbol{\chi}_{nt}$ is

$$\mathbf{S}^{T\prime}_n\mathbf{S}^T_n\mathbf{x}_{nt}, \tag{3.2}$$

while the related estimate of the variance-covariance matrix of $\boldsymbol{\chi}_{nt}$, $\boldsymbol{\Gamma}^{\chi}_{n0}$, is

$$\mathbf{S}_n^{T\prime}\mathbf{M}_n^T\mathbf{S}_n^T. \tag{3.3}$$

The principal components are averages of the x's (with possibly negative weights). The intuition behind the method is that, by averaging along the cross-sectional dimension, the ξ's, which are poorly correlated, cancel out, whereas the χ's do not. Hence, the factor space, i.e. the space spanned by the components of $\mathbf{F}_t$, denoted by $\mathcal{G}(\mathbf{F},t)$, and the approximate factor space, i.e. the space spanned by the first r principal components, denoted by $\mathcal{G}(\mathbf{S}_n^T\mathbf{x}_n,t)$, tend to coincide as $n\to\infty$. An estimate of $\boldsymbol{\chi}_{nt}$ can then be obtained by projecting $\boldsymbol{\chi}_{nt}$ on the approximate factor space, at time t. Similarly, we can predict $\boldsymbol{\chi}_{n,T+h}$ by projecting $\boldsymbol{\chi}_{n,T+h}$ on the approximate factor space $\mathcal{G}(\mathbf{S}_n^T\mathbf{x}_n,T)$. Note that Assumption B1(b) implies that enlarging the projection space with past values of $\mathbf{F}_t$ does not improve the prediction.

The shortcoming of the static method is that it only exploits the information contained in the matrix $\boldsymbol{\Gamma}^T_{n0}$, whereas lagged covariances are ignored. Equivalently, the static method makes use of representation (2.2) without taking the dynamic structure of $\mathbf{F}_t$ into account (see the observation at the end of Section 2).

An attempt to overcome this problem is Forni, Hallin, Lippi and Reichlin (2000), who make use of the q largest *principal component series* or *dynamic principal components* in place of the usual principal components to approximate the common factor space.[1] The χ's are estimated by projecting on present, past and future values of such dynamic principal components.

Denote by $\boldsymbol{\Sigma}_n^T(\theta)=\left(\sigma_{ij}^T(\theta)\right),\ \theta\in[-\pi,\pi]$, a consistent periodogram-smoothing or lag-window estimator of the spectral density $\boldsymbol{\Sigma}_n(\theta)=(\sigma_{ij}(\theta))$ of $\mathbf{x}_{nt}$. Let $\lambda_{nj}^T(\theta)$ be $\boldsymbol{\Sigma}_n^T(\theta)$'s j-th largest eigenvalue and $\mathbf{p}_{nj}^T(\theta)=(p_{nj,1}^T(\theta)\ \dots\ p_{nj,n}^T(\theta))$ the corresponding row eigenvector. Defining

$$\underline{\mathbf{p}}_{nj}^T(L)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\left[\int_{-\pi}^{\pi}\mathbf{p}_{nj}^T(\theta)e^{ik\theta}d\theta\right]L^k, \tag{3.4}$$

the proposed estimator of $\boldsymbol{\chi}_{nt}$ is

$$\left[\tilde{\underline{\mathbf{p}}}_{n1}^T(L)\underline{\mathbf{p}}_{n1}^T(L)+\cdots+\tilde{\underline{\mathbf{p}}}_{nq}^T(L)\underline{\mathbf{p}}_{nq}^T(L)\right]\mathbf{x}_{nt}, \tag{3.5}$$

(tilde denoting conjugate and transpose) which is the projection of $\mathbf{x}_{nt}$ on the approximate factor space (for details on the approximate factor space and the projection see Forni, Hallin, Lippi and Reichlin, 2000). The corresponding estimates of the spectral density matrices of $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$ are

$$\boldsymbol{\Sigma}_n^{\chi T}(\theta) = \lambda_{n1}^T(\theta)\tilde{\mathbf{p}}_{n1}^T(\theta)\mathbf{p}_{n1}^T(\theta)+\cdots+\lambda_{nq}^T(\theta)\tilde{\mathbf{p}}_{nq}^T(\theta)\mathbf{p}_{nq}^T(\theta) \tag{3.6}$$

$$\boldsymbol{\Sigma}_n^{\xi T}(\theta) = \lambda_{n,q+1}^T(\theta)\tilde{\mathbf{p}}_{n,q+1}^T(\theta)\mathbf{p}_{n,q+1}^T(\theta)+\cdots+\lambda_{nn}^T(\theta)\tilde{\mathbf{p}}_{nn}^T(\theta)\mathbf{p}_{nn}^T(\theta), \tag{3.7}$$

respectively, and the estimates of the k-lag covariance matrices of $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$ are

$$\boldsymbol{\Gamma}_{nk}^{\chi T}=\int_{-\pi}^{\pi}e^{ik\theta}\boldsymbol{\Sigma}_n^{\chi T}(\theta)d\theta \quad\text{and}\quad \boldsymbol{\Gamma}_{nk}^{\xi T}=\int_{-\pi}^{\pi}e^{ik\theta}\boldsymbol{\Sigma}_n^{\xi T}(\theta)d\theta. \tag{3.8}$$

Equation (3.5) shows that, with this method, time filters are applied to the x's before averaging along the cross-sections. To get an intuition of why such filtering can be useful to approximate

[1] A comprehensive treatment of dynamic principal components can be found in Brillinger (1981).

the common factor space, assume $\chi_{1t} = u_t$, $\chi_{2t} = u_{t-1}$, $\text{var}(\xi_{1t}) = \text{var}(\xi_{2t}) = 1$, $\text{cov}(\xi_{1t}, \xi_{2t}) = 0$, so that the idiosyncratic to common variance ratio is 1 for both variables. Since the common shock u_t is serially independent, in the contemporaneous average $(u_t + u_{t-1} + \xi_{1t} + \xi_{2t})/2$, the idiosyncratic component is still as large as the common component. By contrast, if we can shift the x's over time before averaging, we can take $(x_{1t-1} + x_{2t})/2 = u_{t-1} + (\xi_{1t-1} + \xi_{2t})/2$, whose idiosyncratic variance is only $1/2$ of the common variance. We will come back to this point with greater detail in Section (6). Consistency results for the above estimators can be found in Forni, Hallin, Lippi and Reichlin (2000), while consistency rates are studied in Forni, Hallin, Lippi and Reichlin (2003).

The trouble with this estimation method is that the filters used in equation (3.5) are two-sided. This creates no problem in the central part of the sample. But at the end (or the beginning) of the sample, two-sided filters cannot be applied as they are. By truncating filters, the performance of (3.5) as an estimator of χ_{it} deteriorates as t approaches T. For the same reason, formula (3.5) cannot be used for prediction. Summing up, the dynamic principal component estimator is good only if we are interested in estimating the covariances or the historical values of the χ's. By contrast, if we are mainly interested in the most recent values of the χ's or in prediction, we are forced to use only one-sided filters or contemporaneous averaging.

However, although the two-sided filters appearing in (3.5) can no longer be used, the estimated covariances (3.8) can help to construct one-sided or contemporaneous averages that perform better than static principal components. Both the static and the dynamic method consist, as we have seen, of the construction of an approximate factor space and a projection on such a space. The basic idea of the present paper is that the estimates $\mathbf{\Gamma}_{n0}^{\chi T}$ and $\mathbf{\Gamma}_{n0}^{\xi T}$, obtained with the dynamic method, may produce both a better approximation of the factor space, and a better estimate of the projection coefficients. In particular, to construct the approximate factor space, we will make use of $\mathbf{\Gamma}_{n0}^{\chi T}$ and $\mathbf{\Gamma}_{n0}^{\xi T}$ to obtain contemporaneous averages of the x's that minimize the fraction of idiosyncratic variance contained in the aggregates.

4 A two-step one-sided estimator

Formally, we want to find r linear combinations $W_{nt}^{jT} = \mathbf{Z}_{nj}^T \mathbf{x}_{nt}$, where the weights $\mathbf{Z}_{nj}^T$ are defined as the solutions, for $1 \leq \ell \leq r$, of the problems

$$\begin{aligned} \mathbf{Z}_{n\ell}^T &= \text{Arg} \max_{\mathbf{a} \in \mathbb{R}^n} \mathbf{a} \mathbf{\Gamma}_{n0}^{\chi T} \mathbf{a}' \\ \text{subject to} \quad & \mathbf{a} \mathbf{\Gamma}_{n0}^{\xi T} \mathbf{a}' = 1 \\ & \mathbf{a} \mathbf{\Gamma}_{n0}^{\xi T} \mathbf{Z}_{nm}^{T'} = 0 \quad \text{for} \quad 1 \leq m < \ell \,,\, 1 \leq \ell \leq n, \end{aligned} \tag{4.1}$$

a prime denoting transpose. This problem is solved by the *generalized eigenvectors* $\mathbf{Z}_{nj}^T$, $j = 1, \ldots, n$, associated to the *generalized eigenvalues* ν_{nj}^T, of the couple of matrices $(\mathbf{\Gamma}_{n0}^{\chi T}, \mathbf{\Gamma}_{n0}^{\xi T})$, that is the vectors and scalars satisfying

$$\mathbf{Z}_{nj}^T \mathbf{\Gamma}_{n0}^{\chi T} = \nu_{nj}^T \mathbf{Z}_{nj}^T \mathbf{\Gamma}_{n0}^{\xi T} \qquad j = 1, 2, \ldots, n, \tag{4.2}$$

with the normalization $\mathbf{Z}_{nj}^T \mathbf{\Gamma}_{n0}^{\xi T} \mathbf{Z}_{nj}^{T'} = 1$ and $\mathbf{Z}_{nj}^T \mathbf{\Gamma}_{n0}^{\xi T} \mathbf{Z}_{nj}^{T'} = 0$ for $i \neq j$.[2] The linear combinations W_{nt}^{kT} are the *generalized principal components* of $\mathbf{x}_{nt}$.

[2]For this property of the generalized eigenvectors see, for instance, Anderson (1984) p. 513. However, a proof is provided in the Appendix, Lemma 6.1.

The forecast we propose for $\chi_{i,T+h}$ based on information available at time T is the estimated projection of $\chi_{i,T+h}$ onto the space spanned by the r aggregates W_{nT}^{kT}, $k = 1, \dots, r$, i.e.

$$\chi_{i,T+h|T}^{nT} = \sum_{j=1}^{n} K_{n,ij}^{Th} x_{jT}, \tag{4.3}$$

where

$$\left(K_{n,ij}^{Th}\right) = \mathbf{K}_n^{Th} = \mathbf{\Gamma}_{nh}^{\chi T} \mathbf{Z}_n^T \left(\mathbf{Z}_n^{T\prime} \mathbf{\Gamma}_{n0}^T \mathbf{Z}_n^T\right)^{-1} \mathbf{Z}_n^{T\prime}, \qquad \mathbf{Z}_n^T = (\mathbf{Z}_{n1}^{T\prime} \ \cdots \ \mathbf{Z}_{nr}^{T\prime})'.$$

The corresponding formula for the in-sample estimator is

$$\chi_{it}^{nT} = \sum_{j=1}^{n} K_{n,ij}^{T0} x_{jt}. \tag{4.4}$$

We refer to the estimators (4.3) and (4.4) as two-step estimators, the first step estimating the covariances $\mathbf{\Gamma}_{n0}^{\chi}$ and $\mathbf{\Gamma}_{n0}^{\xi}$, the second using such estimates to construct generalized principal components of the x's.

Note that an immediate consequence of (4.2) is that the vectors $\mathbf{Z}_{nj}^T$, $j = 1, \dots, n$, are also the generalized eigenvectors of the couple $(\mathbf{\Gamma}_{n0}^{\chi T} + \mathbf{\Gamma}_{n0}^{\xi T}, \mathbf{\Gamma}_{n0}^{\xi T}) = (\mathbf{\Gamma}_{n0}^T, \mathbf{\Gamma}_{n0}^{\xi T})$, with eigenvalues $\nu_{nj}^T + 1$. It follows that the generalized principal components reduce to the usual principal components in the special case $\mathbf{\Gamma}_{n0}^{\xi T} = \mathbf{I}_n$.

Moreover, it is easily seen that the generalized principal components are invariant with respect to linear transformations of the x's, i.e. the generalized principal components of $\mathbf{y}_{nt} = \mathbf{H}\mathbf{x}_{nt} = \mathbf{H}\boldsymbol{\chi}_{nt} + \mathbf{H}\boldsymbol{\xi}_{nt} = \boldsymbol{\alpha}_{nt} + \boldsymbol{\beta}_{nt}$ (say), with $\det(\mathbf{H}) \neq 0$, and those of $\mathbf{x}_{nt}$ are the same. To see this, consider that from (4.2) we get $\mathbf{Z}_{nj}^T \mathbf{H}^{-\mathbf{1}} \mathbf{\Gamma}_{n0}^{\alpha} = \mathbf{Z}_{nj}^T \mathbf{H}^{-\mathbf{1}} \mathbf{H} \mathbf{\Gamma}_{n0}^{\chi} \mathbf{H}' = \nu_{nj} \mathbf{Z}_{nj}^T \mathbf{H}^{-\mathbf{1}} \mathbf{H} \mathbf{\Gamma}_{n0}^{\xi} \mathbf{H} = \nu_{nj} \mathbf{Z}_{nj}^T \mathbf{H}^{-\mathbf{1}} \mathbf{\Gamma}_{n0}^{\beta}$, so that the generalized eigenvectors of the couple $(\mathbf{\Gamma}_{n0}^{\alpha}, \mathbf{\Gamma}_{n0}^{\beta})$ are $\mathbf{Z}_{nj}^T \mathbf{H}^{-\mathbf{1}}$ and the generalized principal components are $\mathbf{Z}_{nj}^T \mathbf{H}^{-\mathbf{1}} \mathbf{y}_{nt} = W_{nt}^{jT}$.

As a consequence, computing the generalized principal components is equivalent to computing the standard principal components of $\mathbf{y}_{nt} = \mathbf{H}\mathbf{x}_{nt}$, where $\mathbf{H}$ is such that the variance-covariance matrix of $\boldsymbol{\beta}_{nt}$ is the $n \times n$ identity matrix. In other words, the generalized principal component forecast can be obtained simply by applying a specific normalization and computing Stock and Watson's forecast (3.1) on the normalized x's. When the idiosyncratic variance-covariance matrix is diagonal, the normalization amounts to dividing each of the x's by the standard deviation of its idiosyncratic component. Such normalization is intuitively much more appealing than the usual one, which consists in dividing by the standard deviation of the variables.

5 Consistency

In this section we prove consistency of the two-step in-sample estimator, meaning convergence in probability of χ_{it}^{nT} to χ_{it} for each i as T and n tend to infinity, and convergence in probability of the two-step forecast $\chi_{i,T+h|T}^{nT}$ to the population projection of $\chi_{i,T+h}$ on the space $\mathcal{G}(\mathbf{F}, T)$. Stock and Watson (1999) prove consistency of their predictor for a model slightly different from the one we are analyzing here.

As in Section 3, denote by $\mathbf{\Sigma}_n^T(\theta) = \left(\sigma_{ij}^T(\theta)\right)_{1\leq i,j\leq n}$ any consistent estimator of the spectral density matrix $\mathbf{\Sigma}_n(\theta) = (\sigma_{ij}(\theta))_{1\leq i,j\leq n}$. Under Assumption A2 , for a given n and any $\epsilon > 0$,

$$\lim_{T\to\infty} \mathrm{P}\left[\max_{1\leq i,j\leq n} \sup_{\theta\in[-\pi,\pi]} |\sigma^T{}_{ij}(\theta) - \sigma_{ij}(\theta)| > \epsilon\right] = 0. \tag{5.1}$$

This is an easy consequence of Remark 1 to Theorem 10.4.1, Brockwell and Davis (1987), p. 353 (note that Remark 1 applies, *mutatis mutandis*, to Theorem 11.7.2, p. 447, which extends Theorem 10.4.1 to the multidimensional case).

Define

$$\check{\mathbf{\Sigma}}_n^{\chi}(\theta) = \lambda_{n1}(\theta)\tilde{\mathbf{p}}_{n1}(\theta)\mathbf{p}_{n1}(\theta) + \cdots + \lambda_{nq}(\theta)\tilde{\mathbf{p}}_{nq}(\theta)\mathbf{p}_{nq}(\theta)$$

and

$$\check{\mathbf{\Sigma}}_n^{\xi}(\theta) = \lambda_{n,q+1}(\theta)\tilde{\mathbf{p}}_{n,q+1}(\theta)\mathbf{p}_{n,q+1}(\theta) + \cdots + \lambda_{nn}(\theta)\tilde{\mathbf{p}}_{nn}(\theta)\mathbf{p}_{nn}(\theta)$$

where $\lambda_{nj}(\theta)$ and $\mathbf{p}_{nj}(\theta)$ are the population counterparts of $\lambda_{nj}^T(\theta)$ and $\mathbf{p}_{nj}^T(\theta)$ respectively.

Under Assumption C1(b), continuity of eigenvalues and of the first q eigenvectors[3] as functions of the entries of $\mathbf{\Sigma}_n^T(\theta)$ implies that (5.1) applies to the entries of $\mathbf{\Sigma}_{nt}^{\chi T}$ and $\check{\mathbf{\Sigma}}_{nt}^{\chi}$ respectively, i.e. that for a given n and any $\epsilon > 0$,

$$\lim_{T\to\infty} \mathrm{P}\left[\max_{1\leq i,j\leq n} \sup_{\theta\in[-\pi,\pi]} |\sigma_{ij}^{\chi T}(\theta) - \check{\sigma}_{ij}^{\chi}(\theta)| > \epsilon\right] = 0. \tag{5.2}$$

The same property holds for $\mathbf{\Sigma}_{nt}^{\xi T}$ and $\check{\mathbf{\Sigma}}_{nt}^{\xi}$. Thus $\mathbf{\Sigma}_{nt}^{\chi T}$ and $\mathbf{\Sigma}_{nt}^{\xi T}$ are consistent estimators of $\check{\mathbf{\Sigma}}_{nt}^{\chi}$ and $\check{\mathbf{\Sigma}}_{nt}^{\xi}$ respectively. Moreover, (5.2) implies that $\mathbf{\Gamma}_{nk}^{\chi T}$ and $\mathbf{\Gamma}_{nk}^{\xi T}$, as defined in (3.8), are consistent estimators, of

$$\check{\mathbf{\Gamma}}_{nk}^{\chi} = \int_{-\pi}^{\pi} e^{ik\theta}\check{\mathbf{\Sigma}}_n^{\chi}(\theta)d\theta \quad \text{and} \quad \check{\mathbf{\Gamma}}_{nk}^{\xi} = \int_{-\pi}^{\pi} e^{ik\theta}\check{\mathbf{\Sigma}}_n^{\xi}(\theta)d\theta, \tag{5.3}$$

respectively.

Let us introduce some new notation. By w_{nt}^{kT} we denote the standardized version of W_{nt}^{kT}. Since

$$\mathbf{Z}_{nj}^T\mathbf{\Gamma}_{n0}^{\xi T}\mathbf{Z}_{nj}^{T'} = \mathbf{Z}_{nj}^T\mathbf{\Gamma}_{n0}^{\chi T}\mathbf{Z}_{nj}^{T'} + \mathbf{Z}_{nj}^T\mathbf{\Gamma}_{n0}^{\xi T}\mathbf{Z}_{nj}^{T'} = 1 + \nu_{nj}^T,$$

then $w_{nt}^{kT} = \mathbf{z}_{nj}^T\mathbf{x}_{nt}$, where $\mathbf{z}_{nj}^T = \mathbf{Z}_{nj}^T/\sqrt{1+\nu_{nj}^T}$. Note that since $\mathbf{Z}_{nj}^T\mathbf{\Gamma}_{n0}^T\mathbf{Z}_{nk}^{T'} = 0$ for $j \neq k$ (using the constraints of (4.1)), the vectors w_{nt}^{kT}, for $k = 1, 2, \ldots, r$, form an orthonormal system spanning a space of the same dimension as $\mathcal{G}(\mathbf{F}, t)$. Denote by $\mathbf{Z}_{nj}$, ν_{nj}, $\mathbf{K}_n^h$, $\chi_{i,T+h|T}^n$, etc., the objects playing the same roles as $\mathbf{Z}_{nj}^T$, ν_{nj}^T, $\mathbf{K}_n^{hT}$, $\chi_{i,T+h|T}^{nT}$, etc., but with respect to $\check{\mathbf{\Gamma}}_{n0}^{\chi}$ and $\check{\mathbf{\Gamma}}_{n0}^{\xi}$.

Proposition 5.1 *Suppose that Assumptions A, B, C and D hold for model (2.1). Then, for all $\epsilon > 0$ and $\eta > 0$, there exist $N_0 = N_0(\epsilon, \eta)$ and $T_0 = T_0(n, \epsilon, \eta)$ such that, for all $n \geq N_0$, all $T \geq T_0$, and all $1 \leq i \leq N_0$,*

$$\mathrm{P}\left[\left|\chi_{i,T+h|T}^{nT} - \chi_{i,T+h|T}\right| > \epsilon\right] \leq \eta. \tag{5.4}$$

[3]The somewhat inaccurate expression "continuity of the eigenvectors" stands for continuity of each matrix $\tilde{\mathbf{p}}_{nj}^T(\theta)\mathbf{p}_{nj}^T(\theta)$, for $j = 1, \ldots, q$.

The proof of Proposition 5.1 relies on the following lemmas.

Lemma 5.1 *Let* $\mathbf{a}_n = (a_{n1}, \ldots, a_{nn})$ *denote a triangular array of real numbers such that* $\lim_{n\to\infty} \sum_{i=1}^{n} a_{ni}^2 = 0$. *Then, under the assumptions of Proposition 5.1,* $\mathbf{a}_n \boldsymbol{\xi}_{nt} \to 0$ *in quadratic mean as* $n \to \infty$. *It follows that* $\mathbf{a}_n \mathbf{x}_{nt}$ *converges to* $\mathcal{G}(\mathbf{F}, t)$ *in quadratic mean.*

For a proof see e. g. Forni, Hallin, Lippi and Reichlin, 2000, Lemma 3, p. 551.

Lemma 5.2 *Let* $\mathcal{K}$ *denote a subspace of the Hilbert space* $\mathcal{H}$ *of centered, square-integrable random variables, with covariance scalar product. Assume that* $\mathcal{K}$ *is generated by a subset* $(v_1, \ldots, v_k)$ *of* k *linearly independent elements of* $\mathcal{H}$. *Let* $\{v_{n1}, \ldots, v_{nk}\ , n \in \mathbb{N}\}$ *be a sequence of* k*-tuples of* $\mathcal{H}$ *such that, denoting by* $\mathrm{proj}\,(\,.\,|\mathcal{K})$ *the projection onto* $\mathcal{K}$,

(i) $v_{nj} - \mathrm{proj}\,(v_{nj}|\mathcal{K})$ *converges to zero in quadratic mean as* $n \to \infty$,

(ii) the determinant of the covariance matrix $\left(\mathrm{Cov}\,(v_{ni}, v_{nj})\right)_{i,j=1,\ldots,k}$ *is bounded away from zero as* $n \to \infty$.

Then, the projection of $v \in \mathcal{H}$ *onto the space* $\mathcal{K}_n$ *spanned by* $\{v_{n1}, \ldots, v_{nk}\}$ *converges in quadratic mean, as* $n \to \infty$, *to the projection of* v *onto* $\mathcal{K}$.

For the proof see the Appendix.

Proof of Proposition 5.1. Lemmas 9.2 and 9.3 imply that ν_{nr} tends to infinity as $n \to \infty$. As a consequence, each of the r sequences $\{\mathbf{Z}_{nj}/\sqrt{1+\nu_{nj}}, n \in \mathbb{N}\}$, $j = 1, \ldots, r$, is a triangular array fulfilling the assumption of Lemma 5.1. Indeed, $\mathbf{Z}_{nj}$ is bounded in modulus, since $1 = \mathbf{Z}_{nj}\boldsymbol{\Sigma}_{n0}^{\xi}\mathbf{Z}_{nj}' \geq \mu_{nn}^{\xi}\mathbf{Z}_{nj}\mathbf{Z}_{nj}'$ where, in view of Assumption D2, μ_{nn}^{ξ} is bounded away from zero. Lemma 5.1 implies that $w_{nt}^{j} = \mathbf{Z}_{nj}\mathbf{x}_{nt}/\sqrt{1+\nu_{nj}}$ converge in quadratic mean , and therefore in probability, to the space $\mathcal{G}(\mathbf{F}, t)$ as $n \to \infty$ for any $j = 1, \ldots, r$. Moreover, by Lemma 5.2, $\chi_{i,T+h|T}^{n}$ converges to $\chi_{i,T+h|T}$ in mean square and therefore in probability. Thus, given $\epsilon > 0$ and $\eta > 0$, there exists $N_1(\epsilon, \eta)$, such that

$$\mathrm{P}(|\chi_{i,T+h|T} - \chi_{i,T+h|T}^{n}| > \epsilon) < \eta. \tag{5.5}$$

Under Assumption D1(b), the matrix $\mathbf{K}$ is a continuous function of the entries of the matrices $\boldsymbol{\Gamma}^{\chi}$ and $\boldsymbol{\Gamma}^{\xi}$. On the other hand, the matrices $\boldsymbol{\Gamma}_n^{\chi T}$ and $\boldsymbol{\Gamma}_n^{\xi T}$ converge in probability to $\check{\boldsymbol{\Gamma}}_n^{\chi}$ and $\check{\boldsymbol{\Gamma}}_n^{\xi}$, respectively, as $T \to \infty$, so that $\mathbf{K}_n^{Th}$ converges in probability to $\mathbf{K}_n^{h}$ for $T \to \infty$. This implies that, given n, $\epsilon > 0$ and $\eta > 0$, there exists $T_1(n, \epsilon, \eta)$ such that for $T > T_1$

$$\mathrm{P}\left(\sum_{j=1}^{n} |K_{n,ij}^{Th} - K_{n,ij}^{h}| > \epsilon\right) < \eta.$$

Moreover, given n and $\eta > 0$, let $M(n, \eta)$ be a positive real such that $\mathrm{P}(\max_{j=1,n} |x_{jt}| \geq M(n, \eta)) < \eta$. Then, given n, $\epsilon > 0$ and $\eta > 0$, there exists $T_2(n, \epsilon, \eta)$ such that for $T > T_2$

$$\mathrm{P}\left(\left|\chi_{i,T+h|T}^{n} - \chi_{i,T+h|T}^{nT}\right| > \epsilon\right) = \mathrm{P}\left(\left|\sum_{j=1}^{n}(K_{n,ij}^{Th} - K_{n,ij}^{h})x_{jt}\right| > \epsilon\right) < \eta. \tag{5.6}$$

To see this,

$$\begin{aligned}\mathrm{P}\left(\left|\textstyle\sum_{j=1}^{n}(K_{n,ij}^{Th} - K_{n,ij}^{h})x_{jt}\right| > \epsilon\right) \\ \leq \mathrm{P}\left(\textstyle\sum_{j=1}^{n}|(K_{n,ij}^{Th} - K_{n,ij}^{h})|M(n,\eta/2) > \epsilon \text{ AND } \max_{j=1,n}|x_{jt}| < M(n,\eta/2)\right) \\ +\mathrm{P}\left(\max_{j=1,n}|x_{jt}| \geq M(n,\eta/2)\right),\end{aligned}$$

so that (5.6) is obtained defining $T_2(n,\epsilon,\eta) = T_1(n,\epsilon/M(n,\eta/2),\eta/2)$. Lastly,

$$\mathrm{P}\left(\left|\chi_{i,T+h|T} - \chi_{i,T+h|T}^{nT}\right| > \epsilon\right) \leq \mathrm{P}\left(\left|\chi_{i,T+h|T} - \chi_{i,T+h|T}^{n}\right| > \epsilon/2\right) + \mathrm{P}\left(\left|\chi_{i,T+h|T}^{n} - \chi_{i,T+h|T}^{nT}\right| > \epsilon/2\right). \tag{5.7}$$

Defining $N_0(\epsilon,\eta) = N_1(\epsilon/2,\eta/2)$ and $T_0(n,\epsilon,\eta) = T_2(n,\epsilon/2,\eta/2)$, the conclusion follows from (5.5) and (5.6). QED

As the reader can easily check, the proof of Proposition 5.1 can be adapted with no difficulty to prove consistency of our two-step in-sample estimator (4.4). Moreover, Proposition 5.1 holds if the matrices $\mathbf{\Gamma}_{n0}^{\xi T}$ and $\check{\mathbf{\Gamma}}_{n0}^{\xi}$ are replaced by any other couple of positive definite symmetric matrices $\mathbf{D}_n^T$ and $\check{\mathbf{D}}_n$, provided that $\mathbf{D}_n^T$ converges to $\check{\mathbf{D}}_n$ in probability, and that the first eigenvalue of $\check{\mathbf{D}}_n$ is bounded, while the last one is bounded away from zero (indeed Lemmas 9.2 and 9.3 hold). This implies consistency of Stock and Watson's estimators (3.1) and (3.2), for which $\check{\mathbf{D}}_n = \mathbf{D}_n^T = \mathbf{I}_n$.

Finally, letting $\gamma_{ij,nk}^{\chi}$ and $\gamma_{ij,nk}^{\chi T}$ be the (i,j) entry of $\check{\mathbf{\Gamma}}_{nk}^{\chi}$ and $\check{\mathbf{\Gamma}}_{nk}^{\chi T}$ respectively, note that Lemma 5.2 implies that $\gamma_{ij,nk}^{\chi}$ converges to $\gamma_{ij,k}^{\chi} = \mathrm{E}(\chi_{it},\chi_{j,t-k})$ as $n \to \infty$. Using this result, the fact that $\mathbf{\Gamma}_{nk}^{\chi T}$ is a consistent estimator of $\check{\mathbf{\Gamma}}_{nk}^{\chi}$ for a given n as $T \to \infty$, and an analogue of (5.7), it is easily seen that $\gamma_{ij,nk}^{\chi T}$ is a consistent estimator of $\gamma_{ij,k}^{\chi}$, meaning that an analogue of Proposition 5.1 holds for $\gamma_{ij,nk}^{\chi T}$ and $\gamma_{ij,nk}^{\chi}$.

6 Finite sample performances: A stylized example

We doubt that general theoretical results about the relative performances of our predictor and Stock and Watson's can be established, due to the extremely general nature of the model under study. However, some insight into their respective advantages can be obtained, partly by analyzing simple examples, partly through simulation. In this section we use a special, highly stylized, model to provide intuition on why and when our method should perform better than the static method. For the sake of simplicity, we concentrate on n-asymptotics, that is, we proceed as though T were infinite and therefore the covariance matrices of $\mathbf{x}_{nt}$ were known. In the next section we relax both the stylization and the $T = \infty$ assumption and estimate a fairly large class simulated models, with different T, n, q and a wide variety of dynamic loadings for the u's. Simulation results confirm the intuition based on the stylized model.

Let us assume that there is only one common shock, u_t, and two groups of variables, the first loading u_t with lag zero, the second with lag one. For convenience, we let n go to infinity by adding one variable to each group, so that n is even. Moreover, to simplify notation, we reorder the variables, for any given n, in such a way that the $m = n/2$ variables in the first group come first. Lastly, we assume that $\boldsymbol{\xi}_{nt}$ is a vector white noise, so that $\mathbf{\Sigma}_n^{\xi}(\theta) = \mathbf{\Gamma}_{n0}^{\xi}/2\pi$, and that $\mathbf{\Gamma}_{n0}^{\xi}$ is diagonal. We have

$$x_{it} = \chi_{it} + \xi_{it} = a_i u_t + b_i u_{t-1} + \xi_{it}, \tag{6.1}$$

where $b_i = 0$ for $i = 1,\ldots,m$, $a_i = 0$ for $i = m+1,\ldots,n$. The static form of (6.1) is

$$x_{it} = a_i F_{1t} + b_i F_{2t} + \xi_{it}, \tag{6.2}$$

where $F_{1t} = u_t$ and $F_{2t} = u_{t-1}$. We have one 'dynamic factor' in (6.1) and two 'static factors' in (6.2). As mentioned above, we assume $T = \infty$, so that, for a given n, the spectral densities and the covariances of $\mathbf{x}_{nt}$ are known.

Under our assumptions on the coefficients a_i and b_i, the spectral density matrix of $\mathbf{x}_{nt}$ is

$$\mathbf{\Sigma}_n(\theta) = \frac{1}{2\pi}\left[(a_1 \ \cdots \ a_m \ b_{m+1}e^{-i\theta} \ \cdots \ b_n e^{-i\theta})'(a_1 \ \cdots \ a_m \ b_{m+1}e^{i\theta} \ \cdots \ b_n e^{i\theta}) + \mathbf{\Gamma}^{\xi}_{n0}\right]. \quad (6.3)$$

The space spanned by the process $\{u_t, \ t \in \mathbb{Z}\}$ is approximated by the space spanned by present, past and future values of the first dynamic principal component of $\mathbf{x}_{nt}$, i.e. $\underline{\mathbf{p}}_{n1}(L)\mathbf{x}_{nt}$, where $\mathbf{p}_{n1}(\theta)$ is the eigenvector corresponding to the first eigenvalue of $\mathbf{\Sigma}_n(\theta)$ (see (3.4) for the relationship between the filter $\underline{\mathbf{p}}$ and the function $\mathbf{p}$).

In our example the filter $\underline{\mathbf{p}}_{n1}(L)$ can be determined by an elementary procedure. We first *align* the x's so that all of them load u_t:

$$\mathbf{y}_t = (x_{1t} \ \cdots \ x_{mt} \ x_{m+1,t+1} \ \cdots \ x_{n,t+1})'.$$

Secondly, let λ_n and $\mathbf{c}_n = (c_{n1} \ \cdots \ c_{nm} \ c_{n,m+1} \ \cdots \ c_{nm})$ be respectively the first eigenvalue and the corresponding normalized eigenvector of the variance-covariance matrix of $\mathbf{y}_t$, i.e.

$$(a_1 \ \cdots \ a_m \ b_{m+1} \ \cdots \ b_n)'(a_1 \ \cdots \ a_m \ b_{m+1} \ \cdots \ b_n) + \mathbf{\Gamma}^{\xi}_{n0}. \quad (6.4)$$

It easily seen that

$$\mathbf{p}_{n1}(\theta) = (c_{n1} \ \cdots \ c_{nm} \ c_{n,m+1}e^{i\theta} \ \cdots \ c_{nn}e^{i\theta}), \quad \underline{\mathbf{p}}_{n1}(L) = (c_{n1} \ \cdots \ c_{nm} \ c_{n,m+1}L^{-1} \ \cdots \ c_{nn}L^{-1}), \quad (6.5)$$

so that

$$\underline{\mathbf{p}}_{n1}(L)\mathbf{x}_{nt} = c_{n1}x_{1t} + \cdots + c_{nm}x_{mt} + c_{n,m+1}x_{m+1,t+1} + \cdots + c_{nn}x_{n,t+1}$$

(incidentally, note that even in this simple case, since the filter $\mathbf{p}_{n1}(L)$ contains negative powers of L, estimation of the factor using the dynamic method uses future values for some of the x's, this implying deterioration at the end of the sample). The estimated spectral density of $\boldsymbol{\chi}_{nt}$ is

$$\lambda_{n1}(\theta)\tilde{\mathbf{p}}_{n1}(\theta)\mathbf{p}_{n1}(\theta),$$

and the estimated $\mathbf{\Gamma}^{\chi}_{n0}$ is obtained by integrating the last expression, as in (3.8). Inspection of (6.3) shows that $\lambda_{n1}(\theta) = \lambda_n$ for any $\theta \in [-\pi \ \pi]$, so that, using (6.5), the estimated $\mathbf{\Gamma}^{\chi}_{n0}$, with the dynamic method, is

$$\begin{pmatrix} \lambda_n \mathbf{c}'_{n,[1,m]}\mathbf{c}_{n,[1,m]} & \mathbf{0}_m \\ \mathbf{0}_m & \lambda_n \mathbf{c}'_{n,[m+1,n]}\mathbf{c}_{n,[m+1,n]} \end{pmatrix},$$

where the vector $\mathbf{c}_{n,[1,m]}$ ($\mathbf{c}_{n,[m+1,n]}$) is obtained by selecting the components from 1 to m (from $m+1$ to n) of $\mathbf{c}_n$.

On the other hand, let μ_{nj} be the j-th largest eigenvalue of $\mathbf{\Gamma}_{n0}$, $\mathbf{S}_{nj}$ be the corresponding normalized row eigenvector and $\mathbf{S}_n = (\mathbf{S}'_{n1} \ \mathbf{S}'_{n2})'$. The estimated $\mathbf{\Gamma}^{\chi}_{n0}$, with the static method, is

$$\begin{pmatrix} \mu_{n1}\mathbf{S}'_{n1,[1,m]}\mathbf{S}_{n1,[1,m]} & \mathbf{0}_m \\ \mathbf{0}_m & \mu_{n2}\mathbf{S}'_{n2,[m+1,n]}\mathbf{S}_{n2,[m+1,n]} \end{pmatrix}$$

(see equation (3.3)).

Intuition on the difference between the two estimates can be obtained assuming $a_i = b_{m+i} = 1$ for $i = 1, \dots, m$, and $\sigma^2_{\xi_i} = 1$ for all i. In this case, denoting by $\mathbf{1}_\ell$ and $\mathbf{0}_\ell$ the $\ell \times \ell$ matrices whose entries are equal to 1 and 0 respectively, we obtain:

$$\mathbf{\Gamma}^\chi_{n0} = \begin{pmatrix} \mathbf{1}_m & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{1}_m \end{pmatrix} \text{ and } \mathbf{\Gamma}_{n0} = \begin{pmatrix} \mathbf{1}_m & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{1}_m \end{pmatrix} + \mathbf{I}_n.$$

Moreover, $\mu_{n1} = \mu_{n2} = m + 1$, $\mathbf{S}_{n1} = (1 \ \cdots \ 1 \ 0 \ \cdots 0)/\sqrt{m}$, $\mathbf{S}_{n2} = (0 \ \cdots \ 0 \ 1 \ \cdots 1)/\sqrt{m}$, $\lambda_n = 1 + n$, and, denoting by $\boldsymbol{\iota}_\ell$ the row ℓ-dimensional vector with all entries equal to 1, $\mathbf{c}_n = \boldsymbol{\iota}_n/\sqrt{n}$, $\mathbf{p}_{n1}(\theta) = (\boldsymbol{\iota}_m \ \ \boldsymbol{\iota}_m e^{i\theta})/\sqrt{n}$. Hence the estimates of $\mathbf{\Gamma}^\chi_{n0}$ with the dynamic and the static method are

$$\left(1 + \frac{1}{2m}\right) \begin{pmatrix} \mathbf{1}_m & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{1}_m \end{pmatrix} \text{ and } \left(1 + \frac{1}{m}\right) \begin{pmatrix} \mathbf{1}_m & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{1}_m \end{pmatrix}$$

respectively. Thus the dynamic estimation error is one half of the static. The reason for this result is easy to see. Static estimation uses two aggregates, the first averaging the x's from 1 to m, the second one averaging the x's from $m + 1$ to n, with equal weights $1/m$. By contrast, dynamic estimation aligns the variables, uses only one aggregate with equal weights $1/2m = 1/n$ and is therefore twice as efficient in annihilating the idiosyncratic component. The results are not very different if we allow for distinct values of a and b, the ratio between estimation errors remaining around 2.

Now let us slightly complicate the above example to analyze in-sample estimation of $\boldsymbol{\chi}_{nt}$. We assume that $m = 2s$ (so that $n = 4s$), $a_i = a$ for $i = 1, \dots, s$, $a_i = b$ for $i = s + 1, \dots, m$, $b_i = a$ for $i = m + 1, \dots, 3s$, and $b_i = b$ for $i = 3s + 1, \dots, n$, and $\sigma^2_{\xi_i}$ is such that all the x's have unit variance. Thus the coefficients a and b are equally represented within the group loading u_t and the group loading u_{t-1}. Figure 6.1 is obtained by letting a vary between .1 and .9, while $b = .6$ and $s = 5$ ($n = 20$). The solid, dashed and bold lines represent

$$\frac{\text{trace } \mathbf{\Gamma}^e}{\text{trace } \mathbf{\Gamma}^\chi_{20,0}}, \tag{6.6}$$

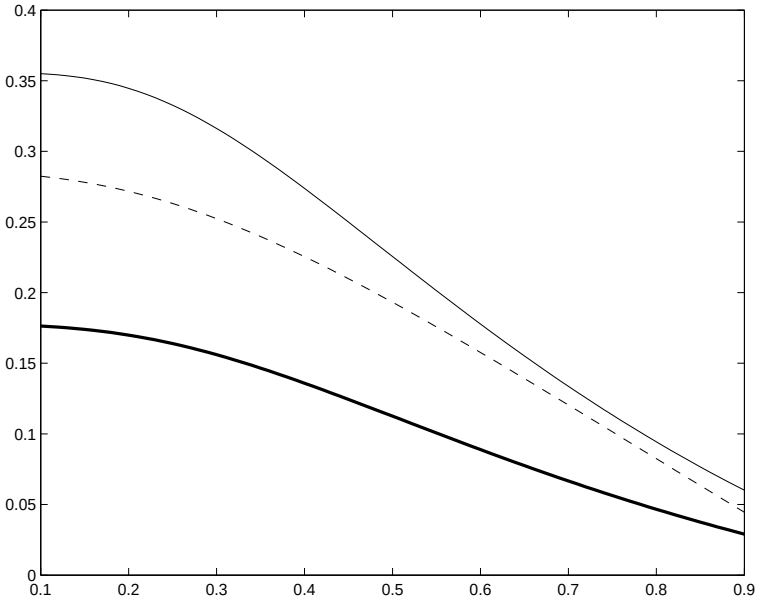
where $\mathbf{\Gamma}^e$ is the variance-covariance matrix of the difference between estimated and actual $\boldsymbol{\chi}_{nt}$, with the static method, the two-step method and the dynamic method, respectively. The plot shows that:

1. The dynamic method has a great advantage over the static method. The ratio between the traces varies around 2, which is, as we have seen, about the size of the estimation-error ratio when we estimate $\mathbf{\Gamma}^\chi_{n0}$ with the static and the dynamic methods.

2. A large part of this advantage is lost when we use the two-step estimator. Indeed, in the second step we can only use contemporaneous values of the x's. However, (I) the projection on the estimated factor space uses the covariance matrix estimated with the dynamic method, which is about twice as good as compared to what we use with the static method; (II) the estimated $\mathbf{\Gamma}^\chi_{n0}$ and $\mathbf{\Gamma}^\xi_{n0}$ are used to compute generalized principal components, so that the difference between a and b is optimally exploited.

3. These two improvements over the static method do not combine straightforwardly. Indeed, as inspection of Figure 6.1 shows, the minimum advantage of the two-step method lies near but not exactly at $a = .6$, where a and b are equal, so that (II) is lost.

Figure 6.1 Performance of static, two-step and dynamic estimators of χ_{nt}



Solid line: static estimator; dashed: two-step estimator; bold: dynamic estimator.

Summing up, the suggestion arising from our stylized model is that two features of the model underlying the data can be expected to be crucial for the relative performance of the competing estimators: (i) substantial hetrogeneity, among cross-sectional units, in the lag structure of the factor loadings, and (ii) sizable differences in the common-to-idiosyncratic variance ratios.

If one is interested in the estimation of variances and covariances of the common and the idiosyncratic components, we recommend using the dynamic method (formula (3.8)) whenever condition (i) seems to hold true. Similarly, if one is interested in the historical values of the common and the idiosyncratic components, formula (3.5) should be used.

On the other hand, if one is interested in forecasting or end-of-sample estimation, the two-step predictor of equation (4.3) may provide a substantial improvement over the static principal component predictor of equation (3.1) when (i) occurs, and in particular when (ii) occurs as well.

Different data sets may differ considerably regarding (i) and (ii). If we do not have any *a priori* information, we can get some indication about (i) simply be looking at the lagged cross-correlations of the x's. The static method, which is simpler, can be used for a preliminary assessment of (i) and (ii).

7 Finite sample performances: Simulation results

In this section we perform simulations on four models which differ by the degree of heterogeneity of the idiosyncratic variances and the dynamic structure of the common components.

The first model, M1, has one autoregressive factor, loaded only contemporaneously, and spherical idiosyncratic components. This is a case where, in principle, the static method should perform comparatively well. Models M2, M3 and M4 have a richer and heterogeneous dynamic structure—a feature which should favor the dynamic method. M2 has MA(3) loading filters,

two serially uncorrelated factors and diagonal idiosyncratic variance-covariance matrix. M3 and M4 have one autoregressive factor and the common components of different groups are shifted in time like in the stylized example of the previous section. The two models differ by the idiosyncratic components: in M3 they have different variances whereas in M4 the variance is the same. The comparison between the last two models should help us understanding the role of heterogeneity of the size of the idiosyncratic components. Finally, in M3 the idiosyncratic components are not mutually orthogonal (but condition C2 is still satisfied).

Model M1. More precisely, model M1 is

$$\begin{aligned} x_{it}^* &= \lambda_i f_t + \alpha c_i \epsilon_{it} \\ (1-aL) f_t &= u_t \end{aligned} \tag{M1}$$

where $a = 0.5$, the shocks u_t and ϵ_{it}, $t = 1, \ldots, T$, $i = 1, \ldots, n$ and the coefficients λ_i, $i = 1, \ldots, n$ are mutually independent standard normal variables, while the coefficients c_i are mutually independent, independent of the latter variables, and uniformly distributed on the interval $[0.1, 1.1]$ to avoid cases of nearly zero idiosyncratic components. The constant α is set so as to guarantee that the idiosyncratic-common variance ratio is equal to one on average (the same holds for all models below). Here $q = 1$ and $s = 1$.

Model M2. Model M2 is

$$x_{it}^* = \sum_{k=0}^{3} a_{ik} u_{1,t-k} + \sum_{k=0}^{3} b_{ik} u_{2,t-k} + \alpha c_i \epsilon_{it}. \tag{M2}$$

Again, a_{ik} and b_{ik}, $k = 0, 1, 2, 3$, $i = 1, \ldots, n$ and the shocks u_{1t}, u_{2t} and ϵ_{it} $t = 1, \ldots, T$, $i = 1, \ldots, n$ are standard normal variables while the c_i's are uniformly distributed on [0.1 1.1] as for (M1). Here $q = 2$ and $s = 3$.

Model M3. Here the observations are generated by the equation

$$x_{it}^* = \textstyle\sum_{k=l_i}^{l_i+2} \lambda_{k-l_i,i} f_{t-k} + \xi_{it}^* \tag{M3}$$

with

$$(1-aL) f_t = u_t$$

$$\xi_{it}^* = \alpha c_i (\epsilon_{it} + \epsilon_{i+1,t})$$

where $a = .5$, $l_i = 0$ for $1 \le i \le m$, $l_i = 1$ for $m+1 \le i \le 2m$ and $l_i = 2$ for $2m+1 \le i \le n$. In order for the three types to be equally present in the panel, we took $m = [n/3]$ (as usual, we denote by $[z]$ the largest integer less than or equal to z). Here $q = 1$ and $s = 5$. Note that ξ_{it}^* is positively correlated with $\xi_{i+1,t}^*$, but is orthogonal to $\xi_{i+k,t}^*$ at any lead and lag for $k > 1$.

Model M4. Here the observations are generated as in M3, but the idiosyncratic components are no longer cross-sectionally correlated ($\xi_{it}^* = \alpha c_i \epsilon_{it}$) and the coefficients c_i are set so as $\mathrm{var}(\lambda_i f_t)/\mathrm{var}(x_{it}^*) = .5$, so that the percentage of idiosyncratic variance is the same for all i.

Before estimation, all variables were taken in deviation from their sample means and divided by their standard deviations, i.e. spectral estimation was conducted on the standardized observations

$$x_{it} = (x_{it}^* - \bar{x}_i^*)/s_i, \tag{7.1}$$

where $\bar{x}_i^* = \sum_{t=1}^{T} x_{it}^*/T$ and $s_i^2 = \sum_{t=1}^{T} (x_{it}^* - \bar{x}_i^*)^2/(T-1)$.

An important empirical finding of our simulations is that, when the cross-sectional dimension n is large with respect to the period of observation T, forcing to zero the off-diagonal entries of the

estimated variance-covariance matrix $\mathbf{\Gamma}_{n0}^{\xi T}$ of the idiosyncratic components improves forecasting performance, even when the actual matrix is non diagonal. Our explanation for this somewhat counterintuitive result is the following. When computing $\mathbf{\Gamma}_{n0}^{\xi T}$, we unavoidably get some spurious large covariances, even when the true covariance is zero. When n increases and T is held fixed, the number of such errors increases as n^2, the order of the number of elements in the $n \times n$ matrices $\mathbf{\Gamma}_{n0}^{\xi T}$. On the other hand, by forcing to zero the off-diagonal entries of our estimated matrix, we ignore the true off-diagonal non-zero entries. Also in this case the error increases with n, but, owing to the boundedness of the eigenvalues, it increases only linearly in n. Therefore, we henceforth set to zero the off-diagonal entries of $\mathbf{\Gamma}_{n0}^{\xi T}$ before computing eigenvectors. Note that, as observed at the end of Section 5, replacing $\mathbf{\Gamma}_{n0}^{\xi T}$ with any symmetric positive semi-definite matrix with bounded eigenvalues does not affect consistency results.

We generated data from each model with $n = 20, 50, 100, 200$ and $T = 20, 50, 100, 200$. Then we computed the in-sample estimates and the forecasts with both the static and the two-step method. We estimated the spectral density matrix of the x's as

$$\mathbf{\Sigma}_n^T(\theta) = \frac{1}{2\pi} \sum_{k=-M}^{M} w_k \mathbf{\Gamma}_k^{nT} e^{-i\theta k}$$

where $w_k = 1 - \frac{|k|}{M+1}$ with window size $M = [T^{1/2}]$. The spectra were evaluated at 101 equally spaced frequencies in the interval $[-\pi, \pi]$, namely, at a grid of frequencies $\theta_h = \frac{2\pi h}{100}$, $h = -50, \ldots, 50$. We then computed the dynamic principal component decomposition, as explained in Section 3. In order to obtain $\mathbf{\Gamma}_{nk}^{\chi T}$ and $\mathbf{\Gamma}_{nk}^{\xi T}$ we used the inverse discrete Fourier transforms

$$\mathbf{\Gamma}_{nk}^{\chi T} = \frac{2\pi}{101} \sum_{h=-50}^{50} \mathbf{\Sigma}_n^{\chi T}(\theta_h) e^{i\theta_h k} \quad \text{and} \quad \mathbf{\Gamma}_{nk}^{\xi T} = \frac{2\pi}{101} \sum_{h=-50}^{50} \mathbf{\Sigma}_n^{\xi T}(\theta_h) e^{i\theta_h k},$$

with $\mathbf{\Sigma}_n^{\xi T}(\theta) = \mathbf{\Sigma}_n(\theta) - \mathbf{\Sigma}_n^{\chi T}(\theta)$. We assumed both the number q of dynamic factors and the number $r = q(s+1)$ of static factors to be known. Each experiment was replicated 1000 times.

We measured the performance of one-step-ahead forecasts by means of the criterion

$$\frac{\sum_{i=1}^{n} (\chi_{i,T+1|T}^{nT} - \chi_{i,T+1})^2}{\sum_{i=1}^{n} \sum_{t=1}^{T} \chi_{it}^2 / T}$$

and performance of within-sample estimates by means of

$$\frac{\sum_{i=1}^{n} \sum_{t=1}^{T} (\chi_{i,t}^{nT} - \chi_{it})^2}{\sum_{i=1}^{n} \sum_{t=1}^{T} \chi_{it}^2}.$$

Results for models M1, M2, M3 and M4 are shown in Tables 5.1, 5.2, 5.3 and 5.4 respectively, with part (a) devoted to forecasts and part (b) devoted to within-sample estimation. We report the average value of the criterion, along with the empirical standard deviation (in brackets), across the 1000 replications, for both the static and the two-step method.

Table 5.1a: Model M1, forecasting results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*static*	*two-step*	*static*	*two-step*	*static*	*two-step*	*static*
	0.9462	0.9492	0.9287	0.9325	0.9292	0.9323	0.9288	0.9321
$T = 20$	(1.2356)	(1.1961)	(1.2035)	(1.1598)	(1.2030)	(1.1519)	(1.1971)	(1.1470)
	0.8642	0.8606	0.8584	0.8529	0.8555	0.8486	0.8553	0.8488
$T = 50$	(1.1493)	(1.1430)	(1.1472)	(1.1288)	(1.1451)	(1.1190)	(1.1439)	(1.1174)
	0.7995	0.8018	0.7869	0.7881	0.7864	0.7864	0.7851	0.7818
$T = 100$	(1.0449)	(1.0425)	(1.034)	(1.0294)	(1.0321)	(1.0288)	(1.0326)	(1.0256)
	0.7833	0.7785	0.7753	0.7770	0.7731	0.7704	0.7721	0.7700
$T = 200$	(1.0723)	(1.0698)	(1.0584)	(1.0486)	(1.0576)	(1.0541)	(1.0577)	(1.0554)

Table 5.1b: Model M1, within-sample results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*static*	*two-step*	*static*	*two-step*	*static*	*two-step*	*static*
	0.1463	0.1759	0.1070	0.1139	0.0969	0.0969	0.0924	0.0888
$T = 20$	(0.1030)	(0.1447)	(0.0551)	(0.0665)	(0.0447)	(0.0508)	(0.0403)	(0.0427)
	0.0631	0.0878	0.0408	0.0506	0.0354	0.0399	0.0328	0.0345
$T = 50$	(0.0351)	0.0471	(0.0156)	0.0190	(0.0111)	0.0126	(0.0091)	0.0099
	0.0413	0.0649	0.0225	0.0327	0.0182	0.0233	0.0161	0.0186
$T = 100$	(0.0232)	(0.0299)	(0.0075)	(0.0097)	(0.0046)	(0.0057)	(0.0034)	(0.0039)
	0.0313	0.0544	0.0143	0.0244	0.0103	0.0155	0.0085	0.0111
$T = 200$	(0.0181)	(0.0225)	(0.0044)	(0.0058)	(0.0022)	(0.0028)	(0.0014)	(0.0016)

Table 5.2a: Model M2, forecasting results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*static*	*two-step*	*static*	*two-step*	*static*	*two-step*	*static*
	0.8901	0.9757	0.7773	0.8552	0.7227	0.7763	0.6911	0.7349
$T = 20$	(0.5266)	(0.5423)	(0.4790)	(0.4851)	(0.4276)	(0.4422)	(0.4132)	(0.4255)
	0.6514	0.7446	0.5025	0.5650	0.4613	0.4911	0.4412	0.4577
$T = 50$	(0.4360)	(0.4793)	(0.3446)	(0.3692)	(0.3165)	(0.3251)	(0.3078)	(0.3130)
	0.5385	0.6332	0.3944	0.4427	0.3552	0.3775	0.3402	0.3509
$T = 100$	(0.3844)	(0.4284)	(0.2895)	(0.3015)	(0.2692)	(0.2736)	(0.2645)	(0.2689)
	0.4949	0.5832	0.3660	0.4076	0.3278	0.3487	0.3127	0.3223
$T = 200$	(0.3367)	(0.3702)	(0.2852)	(0.2896)	(0.2694)	(0.2733)	(0.2673)	(0.2693)

Table 5.2b: Model M2, within-sample results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*static*	*two-step*	*static*	*two-step*	*static*	*two-step*	*static*
	0.4587	0.7163	0.3683	0.5476	0.3290	0.4536	0.3057	0.3933
$T = 20$	(0.1523)	(0.2457)	(0.1049)	(0.1705)	(0.0836)	(0.1340)	(0.0708)	(0.1075)
	0.2838	0.5632	0.1827	0.3106	0.1496	0.2109	0.1340	0.1672
$T = 50$	(0.0751)	(0.1536)	(0.0347)	(0.0759)	(0.0235)	(0.0430)	(0.0180)	(0.0284)
	0.2154	0.4861	0.1238	0.2110	0.0931	0.1303	0.0788	0.0953
$T = 100$	(0.0525)	(0.1234)	(0.0207)	(0.0435)	(0.0120)	(0.0205)	(0.0081)	(0.0122)
	0.1842	0.4336	0.0921	0.1613	0.0613	0.0909	0.0471	0.0598
$T = 200$	(0.0448)	(0.0998)	(0.0157)	(0.0279)	(0.0076)	(0.0117)	(0.0042)	(0.0059)

Table 5.3a: Model M3, forecasting results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*static*	*two-step*	*static*	*two-step*	*static*	*two-step*	*static*
$T = 20$	0.7644 (0.6472)	0.8645 (0.6793)	0.6680 (0.5975)	0.7748 (0.6102)	0.5713 (0.4619)	0.6383 (0.4517)	0.5292 (0.3919)	0.5499 (0.3522)
$T = 50$	0.5273 (0.5886)	0.5913 (0.5993)	0.4636 (0.5620)	0.4962 (0.5393)	0.3565 (0.3770)	0.3773 (0.3601)	0.2792 (0.2355)	0.2771 (0.2139)
$T = 100$	0.4482 (0.5312)	0.4966 (0.5367)	0.3935 (0.5136)	0.4094 (0.4806)	0.2873 (0.3470)	0.2957 (0.3210)	0.1958 (0.1758)	0.1943 (0.1678)
$T = 200$	0.4131 (0.4779)	0.4690 (0.5099)	0.3493 (0.4213)	0.3719 (0.4275)	0.2488 (0.2826)	0.2562 (0.2793)	0.1521 (0.1440)	0.1563 (0.1435)

Table 5.3b: Model M3, within-sample results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*static*	*two-step*	*static*	*two-step*	*static*	*two-step*	*static*
$T = 20$	0.4143 (0.2145)	0.9583 (0.4796)	0.3104 (0.1326)	0.6670 (0.2913)	0.2970 (0.1096)	0.5440 (0.2299)	0.3025 (0.0985)	0.4505 (0.1893)
$T = 50$	0.2285 (0.0900)	0.7279 (0.2752)	0.1444 (0.0433)	0.4083 (0.1223)	0.1335 (0.0317)	0.2753 (0.0791)	0.1326 (0.0268)	0.1856 (0.0525)
$T = 100$	0.1682 (0.0613)	0.6468 (0.2184)	0.0921 (0.0234)	0.3176 (0.0758)	0.0831 (0.0157)	0.1839 (0.0415)	0.0809 (0.0119)	0.1044 (0.0218)
$T = 200$	0.1386 (0.0469)	0.5959 (0.1748)	0.0662 (0.0151)	0.2646 (0.0473)	0.0554 (0.0092)	0.1344 (0.0222)	0.0501 (0.0059)	0.0647 (0.0096)

Table 5.4a: Model M4, forecasting results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*static*	*two-step*	*static*	*two-step*	*static*	*two-step*	*static*
$T = 20$	0.7649 (0.6897)	0.8531 (0.7121)	0.6885 (0.6447)	0.7820 (0.6610)	0.5855 (0.5072)	0.6424 (0.4965)	0.5373 (0.4042)	0.5395 (0.3727)
$T = 50$	0.5288 (0.5340)	0.5692 (0.5195)	0.4577 (0.4869)	0.4886 (0.4724)	0.3632 (0.3427)	0.3759 (0.3343)	0.2933 (0.2286)	0.2746 (0.1988)
$T = 100$	0.4745 (0.5263)	0.4919 (0.5242)	0.4192 (0.5171)	0.4184 (0.4944)	0.3061 (0.3459)	0.3025 (0.3374)	0.2027 (0.1810)	0.1908 (0.1719)
$T = 200$	0.4207 (0.4575)	0.4390 (0.4587)	0.3575 (0.4391)	0.3595 (0.4303)	0.2534 (0.2941)	0.2496 (0.2826)	0.1567 (0.1463)	0.1509 (0.1437)

Table 5.4b: Model M4, within-sample results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*static*	*two-step*	*static*	*two-step*	*static*	*two-step*	*static*
$T = 20$	0.3883 (0.1775)	0.8443 (0.3858)	0.3011 (0.1297)	0.6068 (0.2761)	0.2881 (0.1100)	0.4920 (0.2223)	0.2937 (0.1038)	0.4084 (0.1869)
$T = 50$	0.2284 (0.0641)	0.6099 (0.1763)	0.1487 (0.0393)	0.3566 (0.0986)	0.1359 (0.0317)	0.2404 (0.0678)	0.1324 (0.0283)	0.1668 (0.0476)
$T = 100$	0.1799 (0.0371)	0.5217 (0.1059)	0.1024 (0.0189)	0.2685 (0.0510)	0.0896 (0.0148)	0.1586 (0.0309)	0.0839 (0.0122)	0.0949 (0.0187)
$T = 200$	0.1594 (0.0262)	0.4734 (0.0690)	0.0811 (0.0111)	0.2216 (0.0283)	0.0656 (0.0080)	0.1172 (0.0150)	0.0561 (0.0059)	0.0611 (0.0079)

This is what we read from the tables.

1. For model M1 the competing methods have similar performances for all n and T.

2. The two-step method performs better than the static method for models with heterogeneous dynamics, i.e. M2, M3 and M4. The performance is considerably better for in-sample estimation.

3. Homogeneity in the dynamics (M4 versus M3) reduces somewhat the difference between the two methods, which however remains large, indicating that in this models a substantial gain is obtained simply from the estimation of the matrix used to project the χ's on the common factor space, whereas the advantage stemming from a better estimation of the space itself is smaller.

8 Summary and conclusions

This paper proposes a new forecasting method that exploits information from a large panel of time series. The method is based on the dynamic factor model proposed by Forni, Hallin, Lippi, and Reichlin (2000) and proceeds in two steps. In the first step, we estimate the lagged covariances of the common and idiosyncratic components using the frequency domain approach proposed by Forni, Hallin, Lippi and Reichlin (2000). In the second step we use information about the 'degree of commonality' of each variable to estimate the common factors and project the variables to be predicted onto the linear space spanned by these factors. We show that the projection converges to the optimal forecast as n and T go to infinity. Being a linear combination of the x's which does not involve future observations, the two-step predictor solves the end-of-sample problems caused by two-sided filtering in the estimation method of Forni, Hallin, Lippi and Reichlin (2000), while exploiting the advantages of dynamic information. Both theoretical arguments and simulations suggest that our predictor can provide a substantial improvement over the static principal component predictor when the various cross-sectional items differ significantly in the lag structure of the factor loadings, particularly if, in addition, there is substantial heterogeneity the fraction of total variance explained by the idiosyncratic components.

9 Appendix

Let $\mathbf{\Gamma}$ be a symmetric positive semidefinite $n \times n$ matrix and $\mathbf{D}$ a symmetric, positive definite $n \times n$ matrices. The real, non-negative numbers ν_j, and the real $1 \times n$ vectors $\mathbf{v}_j$ satisfying

$$\begin{aligned} \mathbf{\Gamma}\mathbf{v}_j' &= \nu_j \mathbf{D}\mathbf{v}_j', \\ \mathbf{v}_j\mathbf{D}\mathbf{v}_j' &= 1, \\ \mathbf{v}_j\mathbf{D}\mathbf{v}_m' &= 0, \quad j \neq m, \end{aligned} \tag{9.1}$$

are called the *generalized eigenvalues* and *generalized eigenvectors* of the couple $(\mathbf{\Gamma}, \mathbf{D})$ respectively (see e. g. Anderson, 1984, Section A2). Throughout, we assume that eigenvalues are ranked in decreasing order of magnitude.

Lemma 9.1 *Let* $\mathbf{\Gamma}$ *and* $\mathbf{D}$ *be as above and consider the* n*-tuple of maximization problems*

$$\begin{aligned} \mathbf{a}_j \quad &= \quad \operatorname{Arg}\max_{\mathbf{a}\in\mathbb{R}^n} \mathbf{a}\mathbf{\Gamma}\mathbf{a}' \\ \text{subject to} \quad & \quad \mathbf{a}\mathbf{D}\mathbf{a}' = 1, \\ & \quad \mathbf{a}\mathbf{D}\mathbf{a}_m' = 0 \quad \text{for} \quad 1 \leq m \leq j-1, \end{aligned} \tag{9.2}$$

$j = 1, \ldots, n$. *Denoting by* ν_j *and* $\mathbf{v}_j$ *the generalized eigenvalues and eigenvectors of the couple* $(\mathbf{\Gamma}, \mathbf{D})$, *the solutions of (9.2) are* $\mathbf{a}_j = \mathbf{v}_j$, *and* $\mathbf{a}_j \mathbf{\Gamma} \mathbf{a}_j' = \nu_j$.

Proof. The Lagrangian for $j = 1$ in (9.2) is $\mathbf{a}\mathbf{\Gamma}\mathbf{a}' - \lambda(\mathbf{a}\mathbf{D}\mathbf{a}' - 1)$, so that the first order conditions are

$$\mathbf{\Gamma}\mathbf{a}' = \lambda \mathbf{D}\mathbf{a}' \quad \text{and} \qquad \mathbf{a}\mathbf{D}\mathbf{a}' = 1.$$

These conditions are satisfied by the generalized eigenvectors of $\mathbf{v}_j$, $j = 1, \ldots, n$ only. Since $\nu_j = \mathbf{v}_j \mathbf{\Gamma} \mathbf{v}_j'$ is the value of the objective function, $\mathbf{a}_1 = \mathbf{v}_1$ solves the problem for $j = 1$.

The Lagrangian for $j = 2$ in (9.2) is $\mathbf{a}\mathbf{\Gamma}\mathbf{a}' - \lambda(\mathbf{a}\mathbf{D}\mathbf{a}' - 1) - \mu(\mathbf{a}\mathbf{D}\mathbf{v}_1')$, so that the first order conditions take the form

$$2\mathbf{\Gamma}\mathbf{a}' - 2\lambda \mathbf{D}\mathbf{a}' - \mu \mathbf{D}\mathbf{v}_1' = 0, \quad \mathbf{a}\mathbf{D}\mathbf{a}' = 1, \quad \text{and} \quad \mathbf{a}\mathbf{D}\mathbf{v}_m' = 0. \tag{9.3}$$

Premultiplying by $\mathbf{v}_1$ and taking (9.1) into account, we get $2\mathbf{v}_1\mathbf{\Gamma}\mathbf{a}' - \mu = 2\lambda \mathbf{v}_1 \mathbf{D}\mathbf{a}' - \mu = -\mu = 0$, so that (9.3) reduces to

$$\mathbf{\Gamma}\mathbf{a}' = \lambda \mathbf{D}\mathbf{a}', \quad \mathbf{a}\mathbf{D}\mathbf{a}' = 1, \quad \text{and} \quad \mathbf{a}\mathbf{D}\mathbf{v}_1' = 0.$$

These conditions are satisfied only by the generalized eigenvectors $\mathbf{v}_j$, $j = 2, \ldots, n$, so that (9.2) for $j = 2$ is solved by $\mathbf{a}_2 = \mathbf{v}_2$. The lemma follows recursively. QED

Lemma 9.2 *Given the integer* $k > 0$, *consider a sequence of real, symmetric, positive semi-definite* $n \times n$ *matrices* $\mathbf{\Gamma}_n$ *and a sequence of real, symmetric, positive definite* $n \times n$ *matrices* $\mathbf{D}_n$, $n = k, k+1, \ldots$, *and assume that*

(i) $\mathbf{\Gamma}_n$*'s k-th largest eigenvalue* μ_{nk} *diverges as* $n \to \infty$, *and*

(ii) $\mathbf{D}_n$*'s largest eigenvalue is bounded from above by* δ.

Then, the k-th largest generalized eigenvalue of $(\mathbf{\Gamma}_n, \mathbf{D}_n)$, ν_{nk}, *diverges as* $n \to \infty$.

Proof. Denote by $\mathbf{W}_n$ the $k \times n$ matrix whose rows are the k first eigenvectors of $\mathbf{\Gamma}_n$, and consider the representation

$$\mathbf{W}_n \mathbf{D}_n \mathbf{W}_n' = \mathbf{U}_n \mathbf{R}_n \mathbf{U}_n',$$

where $\mathbf{U}_n$ is a unitary $k \times k$ matrix and $\mathbf{R}_n = \text{Diag}(r_{n1}, \ldots, r_{nj})$ is the diagonal matrix of the k largest eigenvalues of $\mathbf{D}_n$, ordered in decreasing order of magnitude. Clearly, $r_{nj} \leq \delta$ for all j, whereas the rows of $\mathbf{R}^{-1/2}\mathbf{U}_n'\mathbf{W}_n$ satisfy the normalization and the orthogonality constraints of problem (9.2), since

$$\mathbf{R}^{-1/2}\mathbf{U}_n'\mathbf{W}_n\mathbf{D}_n\mathbf{W}_n'\mathbf{U}_n\mathbf{R}^{-1/2} = \mathbf{I}_k.$$

Hence, Lemma 9.1 implies that $\mathbf{R}^{-1/2}\mathbf{U}_n'\mathbf{W}_n$'s last row, namely, $\boldsymbol{\rho}_n \mathbf{U}_n' \mathbf{W}_n$, where $\boldsymbol{\rho}_n = (0 \ \cdots \ 0 \ r_{nk}^{-1/2})$, is such that $\boldsymbol{\rho}_n \mathbf{U}_n' \mathbf{W}_n \mathbf{\Gamma}_n \mathbf{W}_n' \mathbf{U}_n \boldsymbol{\rho}_n' \leq \nu_{nk}$. It follows that, denoting by $\mathbf{M}_n = \text{Diag}(\mu_{n1}, \ldots, \mu_{nk})$ the diagonal matrix of $\mathbf{\Gamma}_n$'s first k eigenvalues,

$$\nu_{nk} \geq \boldsymbol{\rho}_n \mathbf{U}_n' \mathbf{M}_n \mathbf{U}_n \boldsymbol{\rho}_n' \geq \frac{\mu_{nk}}{r_k} \geq \frac{\mu_{nk}}{\delta}.$$

The lemma follows. QED

Lemma 9.3 *Let* $\check{\mathbf{\Gamma}}_{n0}^{\chi}$ *and* $\check{\mathbf{\Gamma}}_{n0}^{\xi}$ *denote the common and idiosyncratic covariance matrices, as defined in (5.3). Then, under the assumptions of Proposition 5.1,*

(i) the $r = q(s+1)$ *largest eigenvalues of* $\check{\mathbf{\Gamma}}_{n0}^{\chi}$ *tend to infinity as* $n \to \infty$, *and*

(ii) the eigenvalues of $\check{\mathbf{\Gamma}}_{n0}^{\xi}$ *are bounded as* $n \to \infty$.

Proof. For any n-dimensional unit-modulus row vector $\mathbf{v}$, we have

$$\begin{aligned} \mathbf{v}\check{\mathbf{\Gamma}}_{n0}^{\xi}\mathbf{v}' &= \mathbf{v}\left[\int_{-\pi}^{\pi} \check{\mathbf{\Sigma}}_n^{\xi}(\theta) d\theta\right]\mathbf{v}' = \int_{-\pi}^{\pi} \mathbf{v}\check{\mathbf{\Sigma}}_n^{\xi}(\theta)\mathbf{v}' d\theta \\ &\leq \int_{-\pi}^{\pi} \lambda_{n,q+1}(\theta) d\theta = \alpha \text{ (say)} \end{aligned}$$

We have (see Lancaster and Tismenetsky, 1985, p. 301, Theorem 1) $\lambda_{n,q+1} \leq \lambda^{\chi}_{n,q+1} + \lambda^{\xi}_{n1}$. Thus, by Assumption C2, $\alpha \leq 2\pi\Lambda$. Part (ii) of the lemma follows. By an easier argument, C2 implies that $\mathbf{v}\mathbf{\Gamma}^{\xi}_{n0}\mathbf{v}' \leq 2\pi\Lambda$. Setting $\mathbf{A}_n = \mathbf{\Gamma}^{\xi}_{n0} - \check{\mathbf{\Gamma}}^{\xi}_{n0}$ and observing that $\mathbf{v}\mathbf{\Gamma}^{\xi}_{n0}\mathbf{v}'$ and $\mathbf{v}\check{\mathbf{\Gamma}}^{\xi}_{n0}\mathbf{v}'$ are non-negative, we obtain

$$|\mathbf{v}\mathbf{A}_n\mathbf{v}'| = |\mathbf{v}\mathbf{\Gamma}^{\xi}_{n0}\mathbf{v}' - \mathbf{v}\check{\mathbf{\Gamma}}^{\xi}_{n0}\mathbf{v}'| \leq 2\pi\Lambda.$$

Since

$$\mathbf{A}_n = \int_{-\pi}^{\pi} \left(\mathbf{\Sigma}^{\xi}_n(\theta) - \check{\mathbf{\Sigma}}^{\xi}_n(\theta)\right) d\theta = \int_{-\pi}^{\pi} \left(\check{\mathbf{\Sigma}}^{\chi}_n(\theta) - \mathbf{\Sigma}^{\chi}_n(\theta)\right) d\theta = \check{\mathbf{\Gamma}}^{\chi}_{n0} - \mathbf{\Gamma}^{\chi}_{n0},$$

it follows that

$$\check{\mathbf{\Gamma}}^{\chi}_{n0} + 2\pi\Lambda\mathbf{I}_n = \mathbf{\Gamma}^{\chi}_{n0} + [2\pi\Lambda\mathbf{I}_n + \mathbf{A}_n].$$

Since the matrix in square brackets is positive semi-definite, the result in Lancaster and Tismenetsky mentioned above implies that the eigenvalues of the sum on the left-hand side are larger than or equal to the corresponding eigenvalues of $\mathbf{\Gamma}^{\chi}_{n0}$. Denoting by $\check{\mu}^{\chi}_{nj}$, $j = 1, \ldots, n$, the eigenvalues of $\check{\mathbf{\Gamma}}^{\chi}_{n0}$, this entails that $\check{\mu}^{\chi}_{nr} + 2\pi\Lambda \geq \mu^{\chi}_{nr}$; part (i) of the lemma thus follows from Assumption D1(a).

Proof of Lemma 5.2. Since we are only interested in $\text{proj}(v|\mathcal{K}_n)$, we can assume with no loss of generality that $\text{var}(v_{nj}) = 1$ for all n and j. Let $\mathbf{v} = (v_1, \ldots, v_k)'$ and $\mathbf{v}_n = (v_{n1}, \ldots, v_{nk})'$. Consider the decomposition $\mathbf{v}_n = \mathbf{a}_n\mathbf{v} + \mathbf{R}_n$, of $\mathbf{v}_n$ into its (componentwise) orthogonal projection $\mathbf{a}_n\mathbf{v}$ onto $\mathcal{K}$ and the orthogonal complement. Assumption (i) implies that $\mathbf{R}_n \to 0$ in quadratic mean. Decomposing similarly v into

$$v = \mathbf{b}_n\mathbf{v}_n + s_n = \mathbf{b}_n\mathbf{a}_n\mathbf{v} + \mathbf{b}_n\mathbf{R}_n + s_n \quad \text{and} \quad v = \mathbf{b}\mathbf{v} + s,$$

where $\mathbf{b}_n\mathbf{v}_n$ and $\mathbf{b}\mathbf{v}$ denote the orthogonal projections of v onto $\mathcal{K}_n$ and $\mathcal{K}$, respectively, we obtain

$$\text{proj}\,(v|\mathcal{K}_n) - \text{proj}\,(v|\mathcal{K}) = \mathbf{b}_n\mathbf{v}_n - \mathbf{b}\mathbf{v} = (\mathbf{b}_n\mathbf{a}_n - \mathbf{b})\mathbf{v} + \mathbf{b}_n\mathbf{R}_n = s - s_n. \tag{9.4}$$

The assumption on $\text{var}(v_{nj})$ and Assumption (ii) imply that $\mathbf{b}_n$ is bounded. As a consequence $\mathbf{b}_n\mathbf{R}_n \to 0$ in q. m., hence $[(\mathbf{b}_n - \mathbf{b})\mathbf{v} - (s - s_n)] \to 0$ in q.m. But

$$|\text{cov}((\mathbf{b}_n - \mathbf{b})\mathbf{v}, s - s_n)| = |\text{cov}((\mathbf{b}_n - \mathbf{b})(\mathbf{v} - \mathbf{b}_n\mathbf{v}_n), s_n)| \leq \sqrt{\text{var}(v)\text{var}((\mathbf{b}_n - \mathbf{b})\mathbf{R}_n)}.$$

Convergence to zero of $\mathbf{R}_n$ and $\mathbf{b}_n\mathbf{R}_n$ implies convergence to zero of the right hand side. This implies that $\lim_{n\to\infty}(\mathbf{b}_n - \mathbf{b})\mathbf{v} = \lim_{n\to\infty}(s - s_n) = 0$. QED

References

[1] Altissimo, F., A. Bassanetti, R. Cristadoro, M. Forni, M. Lippi, L. Reichlin, and G. Veronese (2001). A real time coincident indicator of the Euro area business cycle, Mimeo (August), Banca d'Italia, Rome.

[2] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York.

[3] Bai, J. (2001). Inference on factor models of large dimension. *Econometrica*, to appear.

[4] Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191-221.

[5] Brillinger, D.R. (1981). *Time Series : Data Analysis and Theory*. Holden-Day, San Francisco.

[6] Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica* **51**, 1281-1304.

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

[8] Engle, R.F. and M.F. Watson (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association* **76**, 774-781.

[9] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized factor model : identification and estimation. *The Review of Economics and Statistics* **82**, 540-554.

[10] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2001). Coincident and leading indicators for the Euro area. *The Economic Journal* **111**, 62-85.

[11] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003). The generalized dynamic factor model : consistency and rates. *Journal of Econometrics*, to appear.

[12] Forni, M. and M. Lippi (2001). The generalized dynamic factor model : representation theory. *Econometric Theory* **17**, 1113-1141.

[13] Forni, M. and L. Reichlin (1998). Let's get real : a factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies* **65**, 453-473.

[14] Geweke, J. (1977). The dynamic factor analysis of economic time series. In D.J. Aigner and A.S. Goldberger, Eds., *Latent Variables in Socio-Economic Models*, North Holland, Amsterdam.

[15] Lancaster and Tismenetsky (1985). *The Theory of Matrices*, 2nd Edition, Orlando: Academic Press.

[16] Quah, D., and T.J. Sargent (1993). A dynamic index model for large cross sections. In J.H. Stock and M.W. Watson, Eds, *Business Cycles, Indicators, and Forecasting*, N.B.E.R. and University of Chicago Press, Chicago.

[17] Sargent, T.J. and C.A. Sims (1977). Business cycle modeling without pretending to have too much *a priori* economic theory. In C.A. Sims, Ed., *New Methods in Business Research*, Federal Reserve Bank of Minneapolis, Minneapolis.

[18] Stock, J.H. and M.W. Watson (1999). Diffusion indexes. Unpublished manuscript.