Opening the Black Box: Structural Factor Models with large cross-sections

Mario FORNI
Dipartimento di Economia Politica,
Università di Modena e Reggio Emilia and CEPR
Domenico GIANNONE
ECARES, Université Libre de Bruxelles
Marco LIPPI
Dipartimento di Scienze Economiche
Università di Roma La Sapienza,
Lucrezia REICHLIN*
European Central Bank, ECARES and CEPR

June 12, 2005

Abstract

This paper argues that large-dimensional dynamic factor models are suitable for structural analysis. We establish sufficient conditions for identification of the shocks and the associated impulse-response functions. Moreover, we propose a consistent method (and n, T rates of convergence) to estimate such shocks and response functions, as well as a bootstrapping procedure for statistical inference. The main features of the model, i.e. (i) a large panel on which to condition shocks estimation and (ii) a small number of macroeconomic shocks, allow us to recover the fundamental structural shocks by exploiting the cross-sectional dimension. The method therefore provides a solution for the "problem of fundamentalness" in structural VARs. We illustrate our method and ideas by revisiting the empirical analysis of King *et al.* (1991).

JEL subject classification : E0, C1

Key words and phrases : Dynamic factor models, structural VARs, identification, fundamentalness

*Forni and Lippi are grateful to MIUR (Italian Ministry of Education) for financial support. Giannone and Reichlin were supported by a PAI contract of the Belgian Federal Government and an ARC grant of the Communaute Francaise de Belgique

1 Introduction

Recent literature has shown that approximate (or generalized) dynamic factor models can be used successfully to forecast macroeconomic variables (Forni, Hallin, Lippi and Reichlin, 2002, Stock and Watson, 2002a, 2002b, Boivin and Ng, 2003, Giannone, Reichlin and Sala, 2005) if estimated on the basis of large panels of time series. These models assume that a given time series can be expressed as the sum of two orthogonal components: the "common component", capturing that part of the series which comove with the rest of the economy and the "idiosyncratic component" which is the residual. The vector of the common components is collinear, i.e. it can be written as a process driven by few shocks which generate common dynamics. Indeed, evidence based on different datasets points to the robust finding that few shocks explain the bulk of dynamics of macro data (cfr Sargent and Sims, 1977 and Giannone, Reichlin and Sala, 2002 and 2005).

If the common component is large, a forecasting model which is based on a projection onto some linear combination of these shocks does well because, while being parsimonious, it captures the essential comovements in the economy and can therefore account for information based on high dimensional datasets.

This paper argues that these models can be successfully used not only for forecasting, but also for the estimation and identification of macroeconomic shocks and their propagation mechanism. Our aim is to open the black box of factor models and show how statistical constructs such as factors can be related to economically meaningful shocks so as to use these models for structural analysis in an environment where economic agents have access to large information.

We define macroeconomic shocks as those sources of variations whose effect on all variables in the panel does not vanish as the cross-sectional dimension n becomes large. These shocks generate comovement between macro series (common shocks) and they are the ones we seek to identify. We disregard idiosyncratic shocks whose effect remains local and captures either micro dynamics or measurement error.

Our definition of macro shocks as cross-sectionally pervasive allows us to state precise conditions under which they can be distinguished from idiosyncratic shocks. These are the asymptotic conditions analyzed in Forni, Hallin, Lippi and Reichlin, 2000 and Forni and Lippi, 2001.

Under those conditions we show that, when the number of shocks is small relative to the number of variables and the panel is sufficiently heterogeneous in terms of its dynamic structure, the shocks can be recovered from the present and past observations of the data (the shocks are *fundamental*).

The two features of our framework – large information and small number of shocks relative to the number of variables – allow us to solve the problem that affects small models such as structural VARs where conditions for recovering fundamental shocks are unlikely to be met (on this point see Hansen and Sargent, 1991, Lippi and Reichlin, 1993 and 1994 and, more recently, Chari, Kehoe and Mcgrattan, 2005, Fernandez-Villaverde, Rubio-Ramirez and Sargent, 2005). Since our model can be estimated on the basis of large panels, we do not need to limit arbitrarily the size of the information set used to recover the shocks and we can base modelling on more realistic information assumptions than in VARS (on the importance of this feature for monetary models, see Bernanke and Boivin,2003 and Giannone, Reichlin and Sala, 2002 and 2005). Large information, by giving us a chance to capture heterogeneous dynamics, allows us to recover fundamental shocks and their impulses.

If fundamentalness is satisfied, restrictions based on economic theory allows us to recover the common shocks uniquely. Identification is then reached in two steps. First, identify a linear combination of the fundamental macroeconomic shocks. Second, chose a particular rotation of the latter with a structural interpretation. The economic assumptions, are imposed as in structural VARs (SVAR). Like in those models, identification is obtained by multiplying a vector of orthogonal shocks by an orthonormal matrix. The key difference is that in factor models, the number of shocks, say q, is no longer equal to the number of variables n. In VARs we estimate a reduced form and then we identify the shocks by imposing a number of economic restrictions that depends on n. In factor models, we impose technical restrictions to extract the q common shocks from the n variables and then we identify them by imposing a number of economic restrictions which depends on q. All the identification schemes proposed in the SVAR literature can be applied to structural factor models as well. Identification restrictions can be imposed on any variable in the panel, so that we can easily obtain overidentification by imposing zero long-run or impact effect on groups of variables, like for instance real variables, production indexes, interest rates and so on. Obviously, after identification and estimation, we can analyze the effect of a shock on many variables within a unified framework.

The identification analysis carried here should be distinguished from what studied in the traditional factor literature (see Sargent and Sims, 1977, Geweke, 1977, Geweke and Singleton, 1981, Altug, 1989, Sargent, 1989, Giannone, Reichlin and Sala, 2003). Since our model is approximate and feasible for large panels we need less stringent assumptions to identify the common from the idiosyncratic component (we do not need to impose cross-sectional orthogonality of the idiosyncratic residuals) and by analyzing identification as n increases we are able to provide a precise definition to macroeconomic shocks.

Finally, let us observe that our approach is different from the recently suggested FAVAR model (Bernanke, Boivin and Eliasz, 2005) which consists in augmenting the VAR by factors in order to condition on a large information set and then estimating shocks and impulses. While we focus on common shocks and their impulse response functions, the FAVAR method focuses on the idiosyncratic shocks for a block of variables of interest, i.e. on the residuals of a regression on the common factors.

A contribution of the paper is to show that the asymptotic identification assumptions for the common and idiosyncratic components are sufficient to estimate a linear combination of the shocks and their coefficients consistently and derive rates of convergence for time T and the cross-section n without imposing additional ad hoc technical assumptions. Therefore, we provide the complete theory for estimation and identification of common shocks and their propagation mechanisms when the factor model is based on a "large" panel of data.

The paper is organized as follows. In Section 2, we define the model and discuss the conditions needed to recover the macroeconomic shocks from the panel. Section 3 develops the structural analysis by showing conditions needed for recovering fundamental shocks and identify them uniquely. Section four proves consistency and rates for the estimation of the shocks and the impulse response functions. Section five analysis an empirical example on US macroeconomic data which revisits the results of King et al. (1991) in light of our discussion on fundamentalness.

2 The Model

In this paper we refer to the following dynamic factor model, which is a special case of the generalized dynamic factor model of Forni, Hallin, Lippi and Reichlin (2000) and Forni and Lippi (2001). Such model, and the one used here, differs 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), Chamberlain and Rothschild (1983) and Connor and Korajczyk (1988). Closely related models have been recently studied by Stock and Watson (2002a, 2002b), Bai and Ng (2002) and Bai (2003).

Denote by $\boldsymbol{x}_n^T = (x_{it})_{i=1,\dots,n;t=1,\dots,T}$ an $n \times T$ rectangular array of observations. We make two preliminary assumptions:

PA1. $\boldsymbol{X}_n^T$ is a finite realization of a real-valued stochastic process

$$\boldsymbol{X} = \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\,,\, x_{it} \in L_2(\Omega, \mathcal{F}, P)\}$$

indexed by $\mathbb{N} \times \mathbb{Z}$, where the n-dimensional vector processes

$$\{\boldsymbol{x}_{nt} = (x_{1t} \ \cdots \ x_{nt})',\ t \in \mathbb{Z}\}, \quad n \in \mathbb{N},$$

are stationary, with zero mean and finite second-order moments $\boldsymbol{\Gamma}_{nk} = \mathrm{E}[\boldsymbol{x}_{nt}\boldsymbol{x}'_{n,t-k}]$, $k \in \mathbb{N}$.

PA2. For all $n \in \mathbb{N}$, the process $\{\boldsymbol{x}_{nt}, t \in \mathbb{Z}\}$ admits a Wold representation $\boldsymbol{x}_{nt} = \sum_{k=0}^{\infty} C_k^n \boldsymbol{w}_{n,t-k}$, where the full-rank innovations $\boldsymbol{w}_{nt}$ have finite moments of order four, and the matrices $C_k^n = (C_{ij,k}^n)$ satisfy $\sum_{k=0}^{\infty} |C_{ij,k}^n| < \infty$ for all $n, i, j \in \mathbb{N}$.

We assume that 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 q *common shocks* $\boldsymbol{u}_t = (u_{1t}\ u_{2t}\ \cdots\ u_{qt})'$. Note that q is independent of n (and small as compared to n in empirical applications[1]). More precisely:

FM0. Defining $\boldsymbol{\chi}_{nt} = (\chi_{1t}\ \ldots\ \chi_{nt})'$ and $\boldsymbol{\xi}_{nt} = (\xi_{1t}\ \ldots\ \xi_{nt})'$, we have

$$\begin{aligned} \boldsymbol{x}_{nt} &= \boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt} \\ &= B_n(L)\boldsymbol{u}_t + \boldsymbol{\xi}_{nt}, \end{aligned} \tag{2.1}$$

where $\boldsymbol{u}_t$ is a q-dimensional orthonormal white noise vector.

We assume that

$$B_n(L) = A_n N(L),$$

where (i) $N(L)$ is an $r \times q$ absolutely summable matrix function of L, (ii) A_n is an $n \times r$ matrix, nested in A_m for $m > n$. Defining the $r \times 1$ vector $\boldsymbol{f}_t$ as

$$\boldsymbol{f}_t = N(L)\boldsymbol{u}_t, \tag{2.2}$$

(2.1) can be rewritten as

$$\boldsymbol{x}_{nt} = A_n \boldsymbol{f}_t + \boldsymbol{\xi}_{nt} \tag{2.3}$$

In our model, the common component is driven by q exogenous shocks, which we seek to identify, and can be written as a linear combination of r static factors $\boldsymbol{f}_t$. While q, the rank of the spectral density of the χ's (see the comment on Assumption FM2 below), denoted by $\Sigma_n^{\chi}(\theta)$, is determined by the common sources of exogenous variation in the model, the rank of their variance-covariance matrix r, depends on the degree of dynamic heterogeneity of the impulse responses to the q shocks.

In the sequel, we shall use the term *static factors* to denote the r entries of $\boldsymbol{f}_t$ and the term *dynamic factors* to mean the q entries of $\boldsymbol{u}_t$.

[1] On evidence on this point on the basis of different datasets, see Sargent and Sims, 1977 and Giannone, Reichlin and Sala, 2002 and 2005

FM1. The process $\boldsymbol{u}_t$ is orthogonal to $\xi_{it},\ i = 1, \ldots, n,\ t \in \mathbb{Z}$.

Model (2.2)-(2.3) provides a dynamic representation which is parsimonious and quite general. It accommodates for two interesting special cases.

Case 1. The finite-order moving average factor model

$$\boldsymbol{x}_{nt} = C_n(L)\boldsymbol{u}_t + \boldsymbol{\xi}_{nt}, \tag{2.4}$$

where $C_n(L) = C_0^n + C_1^n L + \cdots + C_s^n L^s$, so that the entries of $C_n(L)$ are polynomials of order s. This model can always be written in the form (2.2)-(2.3) by setting $r = q(s+1)$, $A_n = (C_0^n\ C_1^n\ \cdots\ C_s^n)$, $\boldsymbol{f}_t = (\boldsymbol{u}_t'\ \boldsymbol{u}_{t-1}'\ \ldots\ \boldsymbol{u}_{t-s}')'$ and $N(L) = (I_q\ \ I_q L\ \ldots\ I_q L^s)'$, where I_q is the q dimensional identity matrix[2].

Case 2.

$$\boldsymbol{x}_{nt} = C_n(L)\Psi(L)\boldsymbol{u}_t + \boldsymbol{\xi}_{nt}, \tag{2.5}$$

where $\Psi(L)$ is a $q \times q$ filter. In this case, the dynamic is infinite, but we are imposing a restriction on the heterogeneity of the impulse responses to the shocks. This model can be written in the form (2.2)-(2.3) by setting r and A_n as in Case 1, but $N(L) = (\Psi(L)'\ \ \Psi(L)'L\ \ldots\ \Psi(L)'L^s)'$[3].

In what follows we will define the conditions that allow us to extract the common component from the observables. This step is a prerequisite for the identification of the common shocks which will be analysed in the next Section.

Indicate by Γ_{nk}^{χ} and Γ_{nk}^{ξ} the k-lag covariance matrix of $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$ respectively. Denote by μ_{nj}^{χ} and μ_{nj}^{ξ} the j-th eigenvalue, in decreasing order, of $\boldsymbol{\Gamma}_{n0}^{\chi}$ and $\boldsymbol{\Gamma}_{n0}^{\xi}$ respectively.

FM2. There exists constants $\underline{c}_1, \overline{c}_1, ..., \underline{c}_r, \overline{c}_r$ such that

$$0 < \underline{c}_r \le \liminf_{n\to\infty} n^{-1}\mu_{nr}^{\chi} \le \overline{c}_r < ... < \underline{c}_1 \le \liminf_{n\to\infty} n^{-1}\mu_{nr}^{\chi} \le \overline{c}_1 < \infty$$

FM3. There exists a real Λ such that $\mu_{n1}^{\xi} \le \Lambda$ for any $n \in \mathbb{N}$.

[2]The model can also accommodate (2.4) with v linear restrictions on the coefficients, with $v < q(s+1)$. For example, suppose that $v = 1$ and that

$$A_n d = 0,$$

for all n, where d is a $q(s+1)$-dimensional vector. It is easily shown that (2.4) can be rewritten in the form (2.2)-(2.3) with a vector $\boldsymbol{f}_t$ of dimension $q(s+1) - 1$.

[3]This is the specification chosen by Quah and Sargent, 1993 for example.

FM3 limits the cross-correlation generated by the idiosyncratic shock. It 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 FM3 relaxes such assumption by allowing for a limited amount of cross-correlation among the idiosyncratic components.

FM2 entails that, for n sufficiently large, $A_n' A_n / n$ has full rank r. Consequently, regressing the observations $\boldsymbol{x}_t$ on the factor loadings A_n, it is possible to extract the r common factors $\boldsymbol{f}_t$. Precisely,

$$\left(A_n' A_n\right)^{-1} A_n' \boldsymbol{x}_t = \boldsymbol{f}_t + \left(A_n' A_n\right)^{-1} A_n' \boldsymbol{\xi}_t$$

and by assumption $FM3$, the last term converges to zero in mean square as $n \to \infty$ since

$$\mathrm{E}\left[\left(\left(A_n' A_n\right)^{-1} A_n' \boldsymbol{\xi}_t\right)^2\right] = \left(A_n' A_n\right)^{-1} A_n' \boldsymbol{\Gamma}_{n0}^{\xi} A_n \left(A_n' A_n\right)^{-1} \leq \frac{\Lambda}{n} \left(\frac{A_n' A_n}{n}\right)^{-1}$$

(see Forni, Hallin, Lippi and Reichlin, 2000).

Assumption FM2 implies the weaker condition that each common shock u_{it} is pervasive in the sense that it affects all items of the cross-section as n increases. Precisely, denoting by $\lambda_{nk}^{\chi}(\theta)$, $k = 1, 2, \dots, n$, the eigenvalues of the spectral density matrix $\Sigma_n^{\chi}(\theta)$, in decreasing order at each frequency, Assumption FM2 implies that $\lambda_{nq}^{\chi}(\theta) \to \infty$ as $n \to \infty$, for θ a.e. in $[-\pi\ \pi]$. Forni and Lippi (2001) show that the number q is unique, i.e. a representation (2.1)-(2.3) with a smaller number of dynamic factors is not possible. This notion of pervasiveness provide statistical content to the notion of a macroeconomic shock.

The example below will help understand the role of assumption FM2 in our analysis and the role played in the model by the parameters q and r. For a discussion of r and q in the context of a dynamic general equilibrium model, see also Giannone, Reichlin and Sala, 2003.

Example. Part A. Suppose that

$$\chi_{it} = a_i (1 - c_i L) u_t$$

Here we have one exogenous common shock and $q = 1$ no matter what is the value of the coefficients. The number of factors r, on the other hand, depends on the heterogeneity in the panel. We have:

$r = 1$ if $c_i = c$. In this case, $f_t = (1 - cL) u_t$, $(A_n)_{i.} = a_i$

$r = 2$ if for at least one i and j, $c_i \neq c_j$; $\boldsymbol{f}_t = (u_t, u_{t-1})'$, $(A_n)_{i.} = [a_i, a_i c_i]$ (i-th row of A_n)

Notice that if $r = 1$ we can only extract $f_t = (1 - cL)u_t$. We will return to this case in Section 3.

To understand the role of FM2, notice that:

If $r = 1$, necessary condition to satisfy FM2 is that the common shocks u_t is loaded by "almost all" the χ's with coefficients that do not decline as $n \to \infty$ (pervasiveness). More precisely, FM2 requires that there exist constants $\underline{a}$ and $\overline{a}$, such that for n large enough:

$$0 < \underline{a} \leq \frac{1}{n} \sum_{i=1}^{n} a_i^2 \leq \overline{a} < \infty; \tag{2.6}$$

If $r = 2$ FM2 is satisfied only if, in addition to (2.6), we have $c_i \neq c_j$ for infinitely many (i, j) (pervasiveness of heterogeneity). Precisely, it requires that there exist constants $\underline{\rho}$ and $\overline{\rho}$, such that for n large enough[4]:

$$0 < \underline{\rho} \leq \frac{1}{n} \sum_{i=1}^{n} a_i^2 c_i < \sqrt{\frac{1}{n} \sum_{i=1}^{n} a_i^2} \sqrt{\frac{1}{n} \sum_{i=1}^{n} a_i^2 c_i^2} < \overline{\rho} < \infty; \tag{2.7}$$

To identify the exogenous shocks u_t and the coefficients of the filter $B(L)$ we need more than the technical assumptions FM2-FM3. This requires to introduce informational and economic assumptions as it will be discussed in the next Section.

3 Identification of the shocks: structural factor models

3.1 Fundamentalness: n large and the role of information

The shocks u_t can be recovered from past x's only if the shocks can be recovered from the past and present of the observations. This is the *fundamentalness* condition F:

(F) $\boldsymbol{u}_t$ is fundamental for x_{it}, $i = 1, \ldots, \infty$; i.e. u_{ht}, $h = 1, \ldots, q$ belong to the linear space spanned by the present and the past of x_{it}, $i = 1, \ldots, \infty$.

[4] If $c_1 \neq c_2$, while all other are such that $c_i = c$, $r = 2$, then $|\rho| = 1$. In this case heterogeneity is not sufficiently pervasive and we can only recover u_t. FM2 rules out such situation.

For model (2.1)-(2.3), we require the existence of a $r \times n$ one-sided filter $S_n(L)$ such that $S_n(L)\boldsymbol{x}_n$ converges to $\boldsymbol{u}_t$ in mean square as n goes to infinity. This condition is stronger than what assumed so far since, in addition to identifying the unobserved common component from the data (conditions FM1-FM3), we need to extract the shocks from the common component and its past values (fundamentalness of the u's with respect to the χ's).

The requirement above is satisfied if:

(FM4) There is a left-inverse for $N(L)$, i.e. a $r \times q$ one-sided filter $G(L)$ exists such that $G(L)N(L) = I_q$.

Proposition 1

If FM0-FM4 are satisfied, $\boldsymbol{u}_t$ is fundamental for x_{it}, $i = 1, \ldots, \infty$.

Proof. Setting, $S_n(L) = G(L)\left(A_n' A_n\right)^{-1} A_n'$, where $G(L)$ satisfy FM4, we have:

$$S_n(L)\boldsymbol{\chi}_{nt} = G(L)\left(A_n' A_n\right)^{-1} A_n' A_n \boldsymbol{f}_t = G(L)\boldsymbol{f}_t = G(L)N(L)\boldsymbol{u}_t = \boldsymbol{u}_t$$

Moreover, $S_n(L)\boldsymbol{\xi}_{nt} = G(L)\left(A_n' A_n\right)^{-1} A_n' \boldsymbol{\xi}_t$ converges to zero in mean square by assumptions FM2 and FM3. Q.E.D.

Assumption FM4 is very mild in factor models with n large and sufficient dynamic heterogeneity ($r > q$). This is what makes factor models for large cross-sections particularly interesting for structural analysis. For example, for model (2.4) it easily seen that the filter $G(L) = (I_q \quad \mathbf{0}_q \ldots \mathbf{0}_q)$, where $\mathbf{0}_q$ is a $q \times q$ matrix of zeros, satisfies $FM4$. For (2.5), assumption FM4 is satisfied if $\Psi(L)$ is invertible. It this case, $G(L) = (\Psi(L)^{-1} \;\; \mathbf{0}_q \ldots \mathbf{0}_q)$.

If FM4 is satisfied, $G(L)$ can be approximated by a filter of finite order. Hence, $\boldsymbol{f}_t$ can be approximated by an autoregressive representation of order p. In what follow, we strengthen FM4 by assuming that factors have a VAR(1) representation:

(FM4)' The r-dimensional static factors $\boldsymbol{f}_t$ admit a VAR(1) representation

$$\boldsymbol{f}_t = F\boldsymbol{f}_{t-1} + R\boldsymbol{u}_t \tag{3.8}$$

where R is a constant matrix of dimension $r \times q$ with rank $q \leq r$.

The VAR(1) specification could be generalized to the case of a VAR(p) without introducing new theoretical difficulties. However, (3.8) provides a dynamic representation which is parsimonious and quite general. For example, in Case 1 of Section 2, FM4' is always satisfied. In Case 2, FM4' requires that $\Psi(L)^{-1}$ is of finite order k, not larger than $s+1$. More generally, if $r >> q$ (heterogeneity), the order of the autoregressive representation of the factors is very likely to be small. The intuition is that, when the panel dynamic is very heterogenous, we are likely to capture it by exploiting the cross-sectional information so that many lags on the factors are not needed.

Below we will develop Example 1 of Section 1 to illustrate this discussion.

Example. Part B.

$$\chi_{it} = a_i(1 - c_i L)u_t$$

with $c_i > 1$ for all i, so that there are no invertible submatrices (the fundamentalness condition is violated for each row).

Nonetheless, if $c_i \neq c_j$ for at least a couple (i, j), then

$$\frac{a_i(1 - c_i L)a_j c_j - a_j(1 - c_j L)a_i c_i}{(a_j - c_i)a_i a_j} = 1.$$

Therefore we can set $s_h(L) = 0$ for $h \neq i, j$;

$$s_i(L) = \frac{a_j}{(a_j - a_i)a_i}; \qquad s_j(L) = \frac{a_i}{(a_j - a_i)a_j}.$$

Hence, fundamentalness of the whole system is insured by heterogeneity of the propagation mechanism of the common shocks, u_t. In this case, $r = 2$, $\boldsymbol{f}_t = (u_t, u_{t-1})'$ and FM4 is clearly satisfied with $G(L) = (1\ 0)$. The intuition is that, in this case, by exploiting the cross-sectional dimension, we are able to recover information on the dynamics of the panel and extract the lags of the common shocks.

The only case in which we have non-fundamentalness for the system is when $c_i = c$ for any i and $\mid c \mid > 1$. In this case, $r = 1$, $f_t = (1 - cL)u_t$ and FM4 is not satisfied.

This example illustrates three points:

- if $r >> q$ (heterogeneity), the whole system is likely to be fundamental, i.e. the $q \times 1$ common shocks are likely to be fundamental with respect to the $(r \times 1)$ common factors $\boldsymbol{f}_t$.
- fundamentalness of the whole n dimensional system does not imply fundamentalness of any $q \times q$ subsystem.

- the fundamentalness of any $r \times r$ subsystem in not insured since this requires that the variables in the chosen subsystem have heterogeneous propagation mechanism. This is why even in the case in which we want to focus on r variables for economic reasons, large n helps because it gives us a bigger chance to consider a sufficient number of variables with heterogenous impulses.

The previous discussion shows that, while recovering n fundamental shocks by estimating a dynamic system with n variables requires conditions that are unlikely to be met, when $n >> q$ the conditions to recover the q shocks are likely to be satisfied. We will discuss this point further in the next sub-section, in relation with the VAR literature.

Information (n large) helps provided that (a) by adding variables we do not increase the dimension of the shocks and (b) there is sufficient dynamic heterogeneity in the panel (lead-lag relations) so that $r >> q$. A factor model with large n and small q is therefore particularly interesting for structural analysis.

3.2 Economic conditions for shocks identification

Let us disregard the idiosyncratic component and concentrate on the common components

$$\boldsymbol{\chi}_{nt} = B_n(L)\boldsymbol{u}_t. \tag{3.9}$$

The model is not identified since (3.9) is equivalent to

$$\boldsymbol{\chi}_{nt} = C_n(L)\boldsymbol{v}_t \tag{3.10}$$

where $C_n(L) = B_n(L)H(L)$ and $\boldsymbol{v}_t = H(L^{-1})'\boldsymbol{u}_t$, and $H(L)$ is a $q \times q$ Blaschke matrix, i.e. $H(L)H(L^{-1}) = I_q$. However, if FM4 holds then the common shocks are fundamental. This implies that only static rotations of the common shocks are admissible, i.e. $H(L) = H$. Precisely:

Proposition 2 If

$$\boldsymbol{\chi}_{nt} = C_n(L)\boldsymbol{v}_t \tag{3.11}$$

for any $n \in \mathbb{N}$ where $\boldsymbol{v}_t$ is a q-dimensional fundamental orthonormal white noise vector, then representation (3.11) is related to representation (3.9) by

$$\begin{aligned} C_n(L) &= B_n(L)H \\ \boldsymbol{v}_t &= H'\boldsymbol{u}_t, \end{aligned} \tag{3.12}$$

where H is a $q \times q$ unitary matrix, i.e. $HH' = I_q$.

Proof. Projecting $\boldsymbol{v}_t$ entry by entry on the linear space $\mathcal{U}_t$ spanned by the present and the past of u_{ht}, $h = 1, \dots, q$ we get

$$\boldsymbol{v}_t = \sum_{k=0}^{\infty} H_k \boldsymbol{u}_{t-k} + \boldsymbol{r}_t, \tag{3.13}$$

where $\boldsymbol{r}_t$ is orthogonal to $\boldsymbol{u}_{t-k}$, $k \geq 0$. Now consider that $\mathcal{U}_t$ and the space spanned by present and past of the χ_{it}'s, call it $\mathcal{X}_t$, are identical, because the entries of $\boldsymbol{\chi}_{t-k}$, $k \leq 0$, belong to $\mathcal{U}_t$ by equation (3.9), while the entries of $\boldsymbol{u}_{t-k}$, $k \leq 0$, belong to $\mathcal{X}_t$ by condition FM4. The same is true for $\mathcal{X}_t$ and the space spanned by present and past of the v_{ht}'s, call it $\mathcal{V}_t$, so that $\mathcal{U}_t = \mathcal{V}_t$. Hence $\boldsymbol{r}_t = 0$. Moreover, serial non-correlation of the u_{ht}'s imply that $\sum_{k=1}^{\infty} H_k \boldsymbol{u}_{t-k}$ must be the projection of $\boldsymbol{v}_t$ on $\mathcal{U}_{t-1}$, which is zero because $\mathcal{U}_{t-1} = \mathcal{V}_{t-1}$. It follows that $\boldsymbol{v}_t = H_0 \boldsymbol{u}_t$. Orthonormality of $\boldsymbol{v}_t$ implies that H_0 is unitary $H_0 H_0' = I$. QED

Once fundamentalness is satisfied, identification consists in choosing H such that economically motivated restrictions on the matrix $B_n(L)H$ are fulfilled. For instance, identification can be achieved by maximizing or minimizing an objective function involving $B_n(L)H$ (see, for example, Giannone, Reichlin and Sala, 2005). An alternative is to impose zero restrictions either on the impact effects $B_n(0)H$ or the long-run effects $B_n(1)H_0$ or both. In this case we have to impose $q(q-1)/2$ restrictions (since orthonormality entails $q(q+1)/2$ restrictions). Notice that, once the conditions FM0-FM4 are satisfied, the number of economic identification restrictions we need to identify the shocks depend on q and not on n. This is an advantage for structural analysis, since, provided q is small, we need few restrictions for identification while we are not limited on the informational assumptions (size of the panel).

The connection with VAR analysis is immediate if we think of the χ_t as a vector of observables. To see this point, let us abstract from the idiosyncratic components and consider a block of q common components $\boldsymbol{\chi}_t = (\chi_{1t} \ \cdots \ \chi_{qt})'$. Assume:

$$\boldsymbol{\chi}_t = B(L)\boldsymbol{u}_t. \tag{3.14}$$

and that $B(L)$ is invertible so that $\boldsymbol{\chi}_t$ can be represented as a finite VAR process on q variables:

$$A(L)\boldsymbol{\chi}_t = \boldsymbol{u}_t.$$

Invertibility of $B(L)$ implies fundamentalness. Technically this requires that $\det B(L)$ has its roots outside the unit circle. This assumption is typically implicit in structural VAR analysis (for a critique of structural analysis in VARs based on this observation, see Hansen and Sargent, 1991 and Lippi and Reichlin, 1993 and 1994). Given this assumption, $\boldsymbol{u}_t$ is identified up to a $q \times q$ unitary matrix H

and economic analysis can be used to impose $q(q-1)/2$ restrictions to identify the shocks. Obviously if q is large, we need many of those restrictions.

Let us now go back to a system on n common components driven by q shocks. The discussion of Section 3.1 tells us that while there is no guarantee that the $q \times q$ subsystem of interest is fundamental, it is likely that the n system is, provided that $n >> q$ and $r >> q$. However, once a large factor model is estimated, we can focus on a q dimensional block of interest and fundamentalness can be checked empirically by looking at the roots of the determinant of the estimated $B(L)$ along the lines suggested by Lippi and Reichlin, 1991 and 1993. If the idiosyncratic components of the variables in the block are small, this will also tell us whether a VAR representation on those q variables is admissible. We illustrate this point in the empirical Section.

4 Estimation

Going back to equation (2.3) it is easily seen that the static factors $\boldsymbol{f}_t$ are identified only up to pre-multiplication by a non-singular $r \times r$ matrix. Hence we cannot estimate $\boldsymbol{f}_t$. However, we can estimate the common-factor space, i.e. we can estimate an r-dimensional vector whose entries span the same linear space as the entries of $\boldsymbol{f}_t$. Such vector can be written as $\boldsymbol{g}_t = G\boldsymbol{f}_t$, were G is a non-singular matrix.

The static factor space can be consistently estimated by the first r principal components of the panel $\boldsymbol{x}_{nt}$ as in Stock and Watson, 2001b and 2002. [5].

Precisely, the estimated static factors will be

$$\hat{\boldsymbol{g}}_t = \frac{1}{\sqrt{n}} W_n^{T\prime} \boldsymbol{x}_{nt}, \tag{4.15}$$

where W_n^T is the $n \times r$ matrix having on the columns the eigenvectors corresponding to the first r largest eigenvalues of the sample variance-covariance matrix of $\boldsymbol{x}_{nt}$, say Γ_{n0}^{xT}. We do not normalize the factors to have unit variance. The estimated variance-covariance matrix of $\hat{\boldsymbol{g}}_t$ is the diagonal matrix having on the diagonal the normalized eigenvalues of Γ_{n0}^{xT} in descending order, $\frac{1}{n}\Lambda_n^T = \frac{1}{n} W_n^{T\prime} \Gamma_{n0}^{xT} W_n^T$. The corresponding estimate of the common components is obtained by regressing $\boldsymbol{x}_{nt}$ on the estimated factors to get

$$\boldsymbol{\chi}_{nt}^T = W_n^T W_n^{T\prime} \boldsymbol{x}_{nt}. \tag{4.16}$$

Having an estimate of $\boldsymbol{g}_t$, we have still to unveil the leading-lagging relations between its entries, in order to find out the underlying dynamic factors (or, better,

[5] Alternative (n, T) consistent estimators proposed in the literature are Forni and Reichlin (1998), Boivin and Ng (2003) and Forni, Hallin, Lippi and Reichlin (2005).

a unitary transformation of such factors $\boldsymbol{v}_t = H\boldsymbol{u}_t$, with $HH' = I_q$). This can be done in our dynamic factor model by projecting $\boldsymbol{g}_t$ on its first lag. This approach is also followed in Giannone, Reichlin and Sala (2002, 2005).

4.1 Population formulas

By equation (3.8), any non-singular transformation of the common factors $\boldsymbol{g}_t = G\boldsymbol{f}_t$ has the VAR(1) representation

$$\boldsymbol{g}_t = GFG^{-1}\boldsymbol{g}_{t-1} + \boldsymbol{\epsilon}_t = D\boldsymbol{g}_{t-1} + \boldsymbol{\epsilon}_t. \tag{4.17}$$

Note that

$$D = \Gamma_1^g \left(\Gamma_0^g\right)^{-1}, \tag{4.18}$$

where $\Gamma_h^g = \mathrm{E}(\boldsymbol{g}_t\boldsymbol{g}'_{t-h})$, and

$$\mathrm{var}(\boldsymbol{\epsilon}_t) = \Gamma_0^g - D\Gamma_0^g D'. \tag{4.19}$$

By (3.8), the residual $\boldsymbol{\epsilon}_t$ can be written as

$$\boldsymbol{\epsilon}_t = GR\boldsymbol{u}_t = \left(GRH'\right) H\boldsymbol{u}_t = KMH\boldsymbol{u}_t, \tag{4.20}$$

where

(i) M is the diagonal matrix having on the diagonal the square roots of the first q largest eigenvalues of the variance-covariance matrix of $\boldsymbol{\epsilon}_t$, i.e. the matrix $GRR'G' = \Gamma_0^g - D\Gamma_0^g D'$, in descending order.

(ii) K is the $r \times q$ matrix whose columns are the eigenvectors corresponding to such eigenvalues.

(iii) H is a $q \times q$ unitary matrix;

By inverting the VAR we get

$$\boldsymbol{g}_t = (I - DL)^{-1}KMH\boldsymbol{u}_t.$$

On the other hand, by equations (2.1) and (2.3)

$$\boldsymbol{\chi}_{nt} = B_n(L)\boldsymbol{u}_t = A_n\boldsymbol{f}_t = A_nG^{-1}\boldsymbol{g}_t = Q_n\boldsymbol{g}_t, \tag{4.21}$$

where

$$Q_n = \mathrm{E}(\boldsymbol{\chi}_{nt}\boldsymbol{g}'_t) = \mathrm{E}(\boldsymbol{x}_{nt}\boldsymbol{g}'_t). \tag{4.22}$$

Hence, we have

$$\begin{aligned} \boldsymbol{\chi}_{nt} &= B_n(L)\boldsymbol{u}_t \\ &= Q_n(I - DL)^{-1}KMH\boldsymbol{u}_t \\ &= Q_n(I + DL + D^2L^2 + \cdots)KMH\boldsymbol{u}_t. \end{aligned} \tag{4.23}$$

4.2 Estimators

By substituting $\hat{\boldsymbol{g}}_t = \frac{1}{\sqrt{n}} W_n^{T\prime} \boldsymbol{x}_{nt}$ for $\boldsymbol{g}_t$, it is quite natural to estimate Q_n by $\frac{1}{\sqrt{n}} \Gamma_0^{xT} W_n^T$ (see equation (4.22)). Moreover, Γ_0^g, the variance-covariance matrix of $\boldsymbol{g}_t$, can be estimated by $\frac{1}{n} W_n^{T\prime} \Gamma_{n0}^{xT} W_n^T = \frac{1}{n} \Lambda_n^T$, and Γ_1^g by $\frac{1}{n} W_n^{T\prime} \Gamma_{n1}^{xT} W_n^T$, so that, basing on equation (4.18), we estimate D_n by $D_n^T = W_n^{T\prime} \Gamma_{n1}^{xT} W_n^T (\Lambda_n^T)^{-1}$. Finally, to estimate the eigenvectors and eigenvalues in K_n and M_n we estimate the variance-covariance matrix of $\boldsymbol{\epsilon}_t$ by $\Sigma_n^T = \frac{1}{n}(\Lambda_n^T - D_n^T \Lambda_n^T D_n^{T\prime})$ (see equation (4.19)).

Summing up, in analogy with (4.23) we propose to estimate the impulse-response functions by

$$B_n^T(L) = Q_n^T \left(I + D_n^T L + (D_n^T)^2 L^2 + \cdots \right) K_n^T M_n^T H, \tag{4.24}$$

where

(i) $Q_n^T = \frac{1}{\sqrt{n}} \Gamma_{n0}^{xT} W_n^T$, where Γ_{n0}^{xT} is the sample variance-covariance matrix of $\boldsymbol{x}_{nt}$ and W_n^T the $n \times r$ matrix having on the columns the eigenvectors corresponding to the first r largest eigenvalues of Γ_{n0}^{xT};

(ii) $D_n^T = W_n^{T\prime} \Gamma_{n1}^{xT} W_n^T (\Lambda_n^T)^{-1}$, where Γ_{n1}^{xT} is the sample covariance matrix of $\boldsymbol{x}_{nt}$ and $\boldsymbol{x}_{nt-1}$;

(iii) M_n^T is the diagonal matrix having on the diagonal the square roots of the first q largest eigenvalues of the the matrix $\frac{1}{n}(\Lambda_n^T - D_n^T \Lambda_n^T D_n^{T\prime})$, in descending order;

(iv) K_n^T is the $r \times q$ matrix whose columns are the eigenvectors corresponding to such eigenvalues.

(v) H is a unitary matrix to be fixed by the identifying restrictions.

In order to render operative the above procedure we need to set values for r and q. Unfortunately, there are no criteria in the literature to fix jointly q and r. Bai and Ng (2002) propose some consistent criteria to determine r. As regards the number of dynamic factors, we can follow a decision rule like that proposed in Forni, Hallin, Lippi and Reichlin (2000) i. e., we go on to add factors until the additional variance explained by the last dynamic principal component is less than a pre-specified fraction, say 5% or 10%, of total variance.

4.3 Consistency

Consistency of (4.24) as estimator of the impulse-response functions follows from the fact that as $n, T \to \infty$ the static factor space can be consistently estimated by principal components. Precisely:

Proposition 3 Under assumptions PA1-2, FM1-3, we have, as $\min(n,T) \to \infty$:

(i) $\sqrt{\delta_{nt}}|b_{ni}^T(L) - b_i(L)| = O_p(1), i = 1, ..., n.$

(ii) $n\mathrm{E}\left[(\hat{\boldsymbol{g}}_t - G_n^T \boldsymbol{f}_t)(\hat{\boldsymbol{g}}_t - G_n^T \boldsymbol{f}_t)'\right] = O(1).$

(iii) $G_n^T = \frac{1}{\sqrt{n}} W_n^{T'} W_n W_n' A_n$ is bounded and of full rank r.

where $\delta_{nt} = \min(n,T)$, $b_{ni}^T(L)$ and $b_i(L)$ denote the ith row of $B_n^T(L)$ and $B_n(L)$ respectively,

Proof. See the appendix.

Proposition 3 shows that consistency is achieved along any path for (n, T) with T and n both tending to infinity. The consistency rate is given by $\min\left(\sqrt{T}, \sqrt{n}\right)$. This implies that if the cross-section dimension n is large relative to the sample size T ($T/n \to 0$) the rate of consistency is $\sqrt{T}$, the same we would obtain if the common components were observed, i.e. if the variables were not contaminated by idiosyncratic component. On the other hand, if $n/T \to 0$, then the consistency rate is $\sqrt{n}$ reflecting the fact that the common components are not observed but have to be estimated.[6].

4.4 Standard errors and confidence bands

To obtain confidence bands and standard errors we propose the following bootstrap procedure.

First, compute $B_n^T(L)$ and $\boldsymbol{\chi}_t^T$ according to (**??**) and (4.24), and $\boldsymbol{\xi}_{nt}^T = \boldsymbol{x}_{nt}^T - \boldsymbol{\chi}_{nt}^T$.

Second, for each one of the estimated idiosyncratic components, estimate the univariate autoregressive models

$$a_j(L)\xi_{jt}^T = \sigma_j \omega_{jt}, \qquad j = 1, \dots, n,$$

whose the order can be fixed by the Schwarz criterion, and take the estimated coefficients $a_j^T(L)$ and σ_j^T and the unit variance residuals ω_{jt}^T. Third, generate new simulated series for the shocks, say $\boldsymbol{u}_t^*$ and ω_{jt}^*, $j = 1, \dots, n$, by drawing from the standard normal Use these new series to construct $\boldsymbol{\chi}_{nt}^* = B_n^T(L)\boldsymbol{u}_t^*$, $\xi_{jt}^* = a_j^T(L)^{-1}\sigma_j^T \omega_{jt}^*$, $j = 1, \dots, n$, and $\boldsymbol{x}_{nt}^* = \boldsymbol{\chi}_{nt}^* + \boldsymbol{\xi}_{nt}^*$.

Finally, compute new estimates of the impulse-response functions $B_n^*(L)$ starting from $\boldsymbol{x}_{nt}^*$.

By repeating the two last steps N times we get a distribution of estimated values which can be used to obtain standard errors and confidence bands. Note

[6]It should be pointed out that, under the model assumptions of Stock and Watson (2002a and 2002b) or Bai and Ng (2002), an alternative proof of consistency has been proposed by Giannone, Reichlin and Sala(2002).

that the estimates will in general be biased, since the estimation procedure involves implicitly the estimation of a VAR. An estimate of such bias is provided by the difference between the point estimate $B_n^T(L)$ and the average of the N estimates $B_n^*(L)$.

5 Empirical application

We illustrate our proposed structural factor model by revisiting a seminal work in the structural VAR literature, i.e. King *et al.* (1991, KPSW from now on). To this end, we constructed a panel of macroeconomic series including the series used by KPSW, with the same sampling period. Just like KPSW, we identify a long-run shock by imposing long-run neutrality of all other shocks on per-capita output. The data are well described by three common shocks, so that the comparison with the three-variable exercise of KPSW is particularly appropriate. Having the same data, the same identification scheme and the same number of shocks, different results can only be due to the additional information coming from the other series in the panel.

5.1 The data

The data set was constructed by downloading mainly from the FRED II database of the Federal Reserve Bank of St. Louis and Datastream. The original data of KPSW have been downloaded from Mark Watson's home page. We collected 89 series, including data from NIPA tables, price indeces, productivity, industrial production indeces, interest rates, money, financial data, employment, labor costs, shipments, and survey data. A larger n would be desirable, but we were constrained by both the scarcity of series starting from 1949 (like in KPSW) and the need of balancing data of different groups. In order to use Datastream series we were forced to start from 1950:1 instead of 1949:1, so that the sampling period is 1950:1 - 1988:4. Monthly data are taken in quarterly averages. All data have been transformed to reach stationarity according to the ADF(4) test at the 5% level. Finally, the data were taken in deviation from the mean as required by our formulas, and divided by the standard deviation to render results independent of the units of measurement. A complete description of each series and the related transformations is reported in the Appendix.

5.2 The choice of r and the number of common shocks

As a first step we have to set r and q. Let us begin with r. We computed the six consistent criteria suggested by Bai and Ng (2002) with $r = 1, \dots, 30$. The criteria IC_{p1} and IC_{p3} do not work, since they do not reach a minimum for $r < 30$; IC_{p2} has a minimum for $r = 12$. To compute PC_{p1}, PC_{p2} and PC_{p3}

we estimated $\hat{\sigma}^2$ with $r = 15$ since with $r = 30$ none of the criteria reaches a minimum for $r < 30$. PC_{p1} gives $r = 15$, PC_{p2} gives $r = 14$ and PC_{p3} gives $r = 20$. Below we report results for $r = 12$, $r = 15$ and $r = 18$, with more detailed statistics for $r = 15$. With $r = 15$, the common factors explain on average 79.7% of total variance. With reference to the variables of interest in KPSW, the common factors explain 85.6% of total variance for output, 84.4% for investment and 89.4% for consumption.

Regarding the choice of q, as explained above, for the sake of comparison we start with a strong preference in favor of $q = 3$. This choice is consitent with the decision rule proposed in Forni, Hallin, Lippi and Reichlin (2000), since, with Bartlett lag window size 18, the overall variance explained by the third dynamic principal component is larger than 10% (10.2%), whereas the variance explained by the fourth one is less than 10% (6.8%).

5.3 Fundamentalness

Now let us focus on the 3×3 impulse-response function system for the three variables of KPSW, i.e. per capita consumption, per capita income and per capita investment. As observed at the end of Section 3, we can compute the roots of the determinant of this system to check whether it is invertible or not.[7] Figure 1 plots the moduli of the two smallest roots of the above determinant as a function of r, for r varying over the range 3-30. Note that for $r = 3$ all roots must be larger than one in modulus, since they stem from a three-variate VAR. This is in fact the case for $r = 3$ and $r = 4$, but for $r \geq 5$ the smallest root is declining and lies always within the unit circle. For $r \geq 22$ the second smallest root becomes smaller than one in modulus.

Figure 2 reports the distribution of the modulus of the smallest root for $r = 15$ across 1000 bootstrapping replications. The mean value is 0.71, indicating a non-negligible upward bias, since our point estimate for $r = 15$ is 0.54. We shall come back to the estimation bias below. Here we limit ourselves to observe that if the smallest root is overestimated on average, the true value could be even smaller than 0.54. Without any bias correction, the probability of an estimated value larger than one in modulus is less than 22%.

We conclude that the true, structural impulse-response function system for the common components associated with these three variables is probably non-fundamental. As a consequence, such impulse response functions, as well as the associated structural shocks, cannot be recovered by estimating a three-dimensional VAR.

[7]Note that these roots (and therefore fundamentalness) are independent of the identification rule adopted and the rotation matrix H.

Figure 1: **The moduli of the first and the second smallest roots as functions of r**

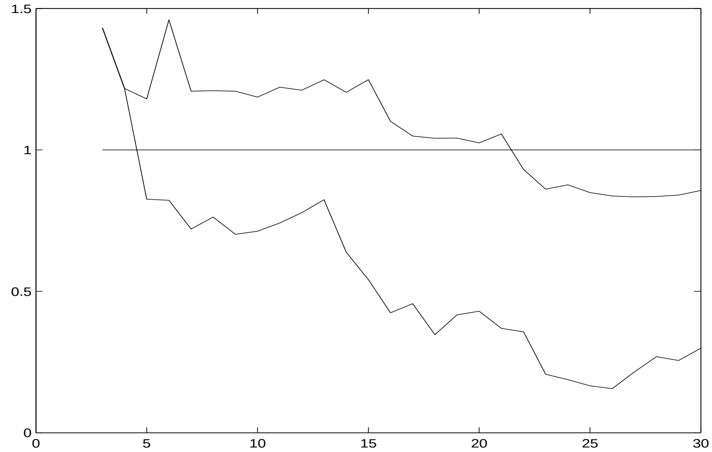
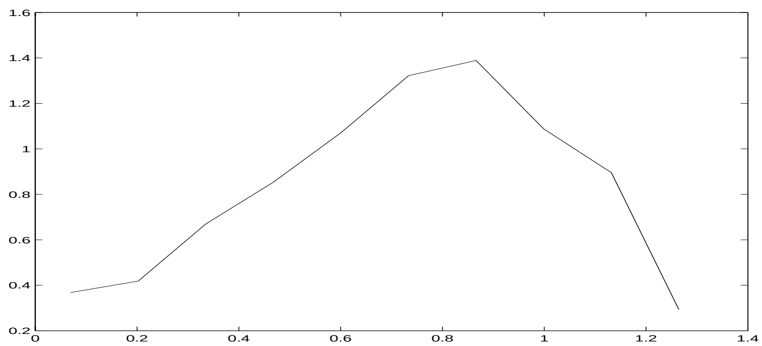


Figure 2: **Frequency distribution of the modulus of the smallest root**

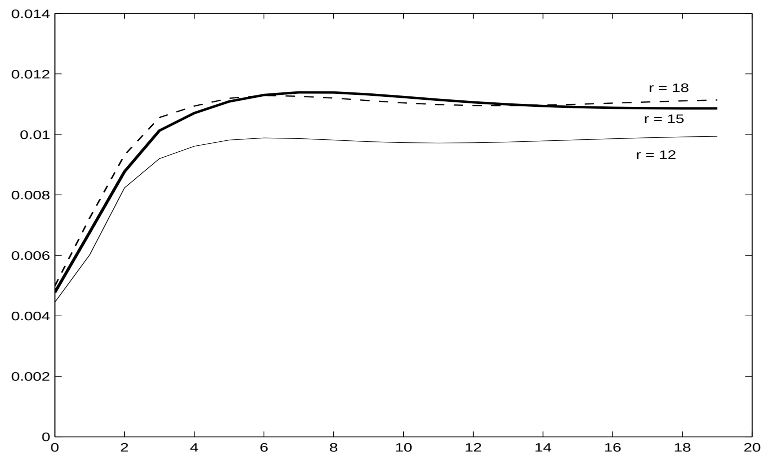


5.4 Impulse-response functions and variance decomposition

Coming to the impulse-response functions, as anticipated above we impose long-run neutrality of two shocks on per-capita output, like in KPSW. This is sufficient to reach a partial identification, i.e. to identify the long-run shock and its response functions on the three variables.

Figure 3 shows the response functions of per capita output for $r = 12, 15, 18$. The general shape does not change that much with r. The productivity shock has positive effects declining with time on the output level. The response function reach its maximum value after 6-8 quarters with only negligible effects after two years. It should be observed that this simple distribuited-lag shape is different from the one in KPSW, where there is a sharp decline during the second and the third year, which drives the overall effect back to the impact value.

Figure 3: **The impulse response function of the long-run shock on output for $r = 12, 15, 18$**



In Figure 4 we concentrate on the case $r = 15$. We report the response functions with 90% confidence bands for output, consumption and investment respectively. Confidence bands are obtained with the procedure explained above (with 1000 replications). The shapes are similar for the three variables, with a positive impact effect followed by important, though declining, positive lagged effects.

Note that confidence bands are not centered around the point estimate, especially for consumption, suggesting the existence of a non-negligible bias. This is not surprising, since formula (4.24) implicitly involves estimation of a VAR, where in addition the variable involved (the static factors) contain errors (a residual idiosyncratic term). Figure 5 shows the point estimate along with the mean of the bootstrap distribution for the output. Such a large bias is probably due to the small cross-sectional dimension. We have evidence of a much smaller bias for the larger data set of Giannone, Reichlin and Sala (2002). We do not make any attempt here to correct for the bias, but a procedure like the one suggested in Kilian (1998) could be appropriate.

Table 1 reports the fraction of the forecast-error variance attributed to the permanent shock for output, consumption and investment at different horizons. For ease of comparison we report the corresponding numbers obtained with the (restricted) VAR model and reported in Table 4 of KPSW.

At horizon 1, our estimates are smaller. The difference is important for consumption: only 0.30 according to the factor model as against 0.88 according to the KPSW model. But at horizons larger than or equal to 8 quarters our estimates are greater and the difference is very large for investment. At horizon 20 (5 years) the permanent shock explains 46% of investment variance according to KPSW as against 86% with the factor model. This result is interesting in that it

Figure 4: **The impulse response function of the long-run shock on output, consumption and investment for** $r = 15$

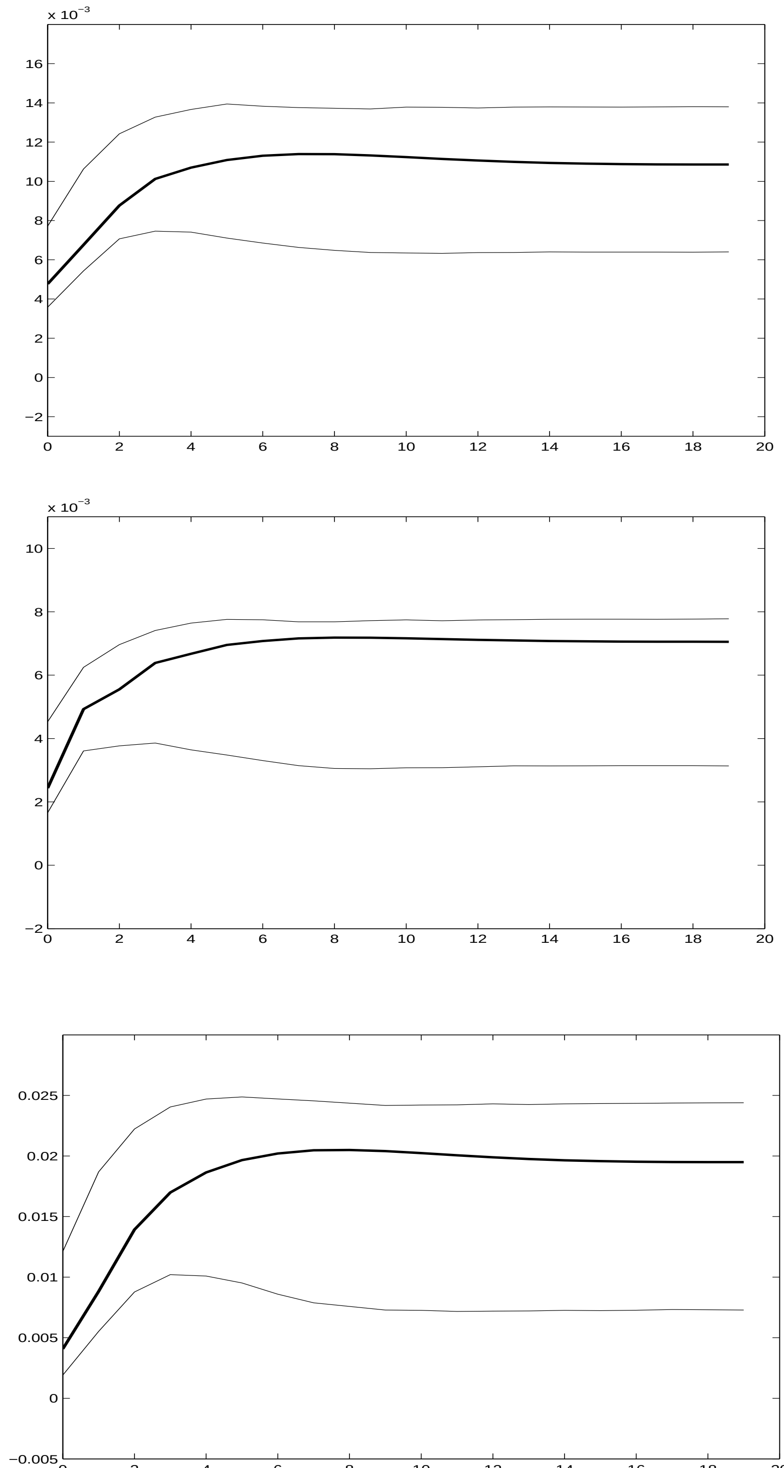
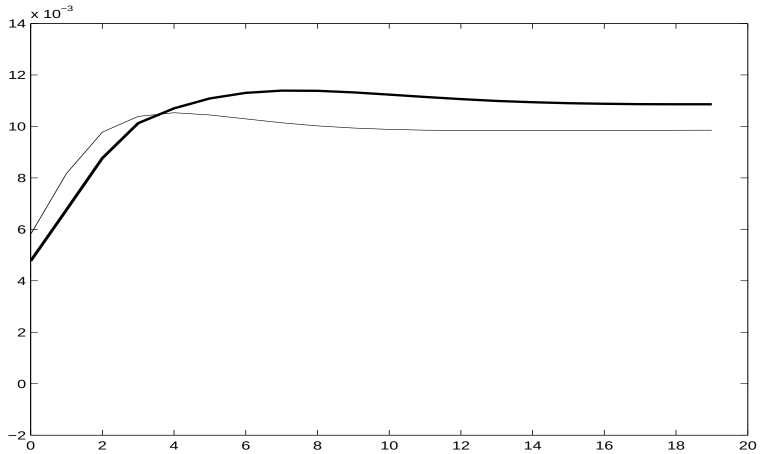


Figure 5: **Estimation bias**



solves a typical puzzle of the VAR literature: the finding that technological and other supply shocks explain a small fraction of investment variations even in the medium-long run.

6 Conclusions

In this paper we have argued that dynamic factor models are suitable for structural macroeconomic modeling and an interesting alternative to structural VARs when the objective is to analyse large panels of time series.

We have shown that large information and a small number of shocks generating the comovement of many variables, allow the econometrician to recover the fundamental shocks driving the economy under general conditions. They require that the structure of leads and lags is rich enough so that the cross-section can convey information on dynamic relations. These conditions - we also show - are more plausible that those needed in VAR analysis where the problem is to recover n shocks from n variables (the problem of fundamentalness highlighted by Hansen and Sargent, 1991, Lippi and Reichlin, 1993 and 1994 and, more recently, by Chari, Kehoe and Mcgrattan, 2005, Fernandez-Villaverde, Rubio-Ramirez and Sargent, 2005).

Having established sufficient conditions for identification, we have suggested a procedure for the estimation of the impulse response functions. Moreover, we have shown consistency of such a procedure and have suggested a bootstrapping method for the construction of confidence bands and inference purposes.

In the empirical application, we have revisited the seminal paper by King *et al.* (1991, KPSW). We have designed a large data set including output, consumption and investment (the data analysed by KPSW) on the same sample period.

Table 1: **Fraction of the forecast-error variance due to the long-run shock**

	Dynamic factor model			KPSW vector ECM		
Horizon	Output	Cons.	Inv.	Output	Cons.	Inv.
1	0.37	0.30	0.07	0.45	0.88	0.12
	(0.18)	(0.21)	(0.19)	(0.28)	(0.21)	(0.18)
4	0.57	0.77	0.42	0.58	0.89	0.31
	(0.12)	(0.12)	(0.19)	(0.27)	(0.19)	(0.23)
8	0.78	0.87	0.72	0.68	0.83	0.40
	(0.07)	(0.11)	(0.16)	(0.22)	(0.18)	(0.18)
12	0.86	0.90	0.80	0.73	0.83	0.43
	(0.05)	(0.11)	(0.16)	(0.19)	(0.18)	(0.17)
16	0.89	0.91	0.83	0.77	0.85	0.44
	(0.04)	(0.11)	(0.16)	(0.17)	(0.16)	(0.16)
20	0.91	0.92	0.86	0.79	0.87	0.46
	(0.03)	(0.11)	(0.16)	(0.16)	(0.15)	(0.16)

We have estimated a large factor model with a three-shock specification and, after having identified the shocks as in KPSW, we have analysed impulse response functions on the three variables of interest: output, consumption and investment. We find that the smallest root of the determinant of the impulse-response functions formed by the three variables sub-system is non-fundamental and therefore could have not be obtained by estimating a VAR on these three variables alone. These impulse response functions imply a larger effect of the permanent shock on output and investment than those found by KPSW.

Appendix 1: Proof of Proposition 3

Let $\mathbf{A}$ and $\mathbf{E}$ be two $n \times n$ symmetric matrices and denote by $\sigma_j(\cdot), j = 1, \dots, n$ the eigenvalues in decreasing order of magnitude. Throughout this section we will use the following inequalities due to Weyl (cfr. Stewart and Sun, 1990):

$$|\sigma_j(\mathbf{A}+\mathbf{E}) - \sigma_j(\mathbf{A})| \leq \sqrt{\sigma_1(\mathbf{E}^2)} \leq \sqrt{\text{trace}(\mathbf{E}^2)}$$

Denote by Λ_n and Λ_n^T, the $r \times r$ diagonal matrices having on the diagonal elements the first r largest eigenvalues of Γ_{n0}^{χ} and Γ_{n0}^{x}, respectively. Writing W_n and W_n^T for the $n \times r$ matrices having on the columns the corresponding eigenvectors, we have, by definition:

$$\Gamma_{n0}^{\chi} W_n = W_n \Lambda_n$$

$$\Gamma_{n0}^{xT} W_n^T = W_n^T \Lambda_n^T$$

Let us recall here our notation for the eigenvalues of the relevant matrices:

$$\mu_{nj}^{x} := \sigma_j(\Gamma_{n0}^{x}), \;\; \mu_{nj}^{xT} := \sigma_j(\Gamma_{n0}^{xT}), \;\; \mu_{nj}^{\chi} := \sigma_j(\Gamma_{n0}^{\chi}), \;\; \mu_{nj}^{\xi} := \sigma_j(\Gamma_{n0}^{\xi}), \;\; j = 1, ..., n$$

we have $\Lambda_n = \text{diag}(\mu_{n1}^{\chi}, ..., \mu_{nr}^{\chi})$ and $\Lambda_n^T = \text{diag}(\mu_{n1}^{xT}, ..., \mu_{nr}^{xT})$

Using the following non-singular transformation of the common factors, $\mathbf{g}_t = G_n \mathbf{f}_t$ where $G_n = \frac{1}{\sqrt{n}} W_n' A_n$, we have (cfr. Section 4.1):

$$Q_n = \frac{1}{\sqrt{n}} \Gamma_{n0}^{\chi} W_n, D_n = W_n' \Gamma_{n1}^{\chi} W_n \Lambda_n^{-1} \text{ and } \Sigma_n = \frac{1}{n}\Lambda_n - \frac{1}{n} D_n \Lambda_n D_n'$$

Lemma 1 Under assumptions PA1-2, FM1-3, as $n, T \to \infty$, we have:

(i) $\text{trace}\left[(\Gamma_{kn}^{xT} - \Gamma_{kn}^{x})^2\right] = O_p\left(\frac{n^2}{T}\right)$, $k = 0, 1$

(ii) $\frac{1}{n}\mu_{nj}^{xT} = \frac{1}{n}\mu_{nj}^{\chi} + O\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$ for $k = 1, ..., n$

Proof. By assumption PA2, there exists a positive constant $K \leq \infty$, such that for all $T \in \mathbb{N}$ and $i, j \in \mathbb{N}$

$$T\text{E}[(\hat{\gamma}_{0ij}^{xT} - \gamma_{0ij}^{x})^2] < K$$

as $T \to \infty$, where γ_{0ij}^{xT} and γ_{0ij}^{x} denote the i,jth entries of Γ_{0n}^{xT} and Γ_{0n}^{x} respectively.

We have:

$$\text{trace}\left[(\Gamma_{0n}^{xT} - \Gamma_{0n}^{x})^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n}(\gamma_{0ij}^{xT} - \gamma_{0ij}^{x})^2$$

Taking expectations, we obtain:

$$\mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(\gamma_{0ij}^{xT} - \gamma_{0ij}^{x})^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}\left[(\gamma_{0ij}^{xT} - \gamma_{0ij}^{x})^2\right] = O_p\left(\frac{n^2}{T}\right)$$

Result (i), for $k = 0$, follows from the Markov inequality. The result for $k = 1$ can be easily proved using the same arguments.

Turning to (ii), from the Weyl inequality, we have:

$$\left(\mu_{nj}^{xT} - \mu_{nj}^{x}\right)^2 \leq \text{trace}\left[(\Gamma_{0n}^{xT} - \Gamma_{0n}^{x})^2\right]$$

moreover, from assumption FM0-3:

$$\frac{1}{n}\mu_{nj}^{x} \leq \frac{1}{n}\mu_{nj}^{\chi} + \frac{1}{n}\mu_{n1}^{\xi} = \frac{1}{n}\mu_{nj}^{\chi} + O\left(\frac{1}{n}\right)$$

The desired result follows. *Q.E.D.*

Corollary 1 Under assumptions PA1-2, FM1-3, as $n, T \to \infty$, we have:

(i) $\frac{1}{n}\Lambda_n^T = \frac{1}{n}\Lambda_n + O_p(\frac{1}{\sqrt{T}}) + O_p\left(\frac{1}{n}\right)$

(ii) $W_n' W_n^T = I_r + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$

Proof. Result (i) trivially follows from Lemma 1. Turning to (ii), we have the following decomposition:

$$\frac{1}{n}\Lambda_n^T = \frac{1}{n}W_n^{T'}\Gamma_{n0}^{xT}W_n^T = \frac{1}{n}W_n^{T'}W_n\Lambda_n W_n' W_n^T + \frac{1}{n}W_n^{T'}\Gamma_{n0}^{\xi T}W_n^T + \frac{1}{n}W_n^{T'}\left(\Gamma_{n0}^{xT} - \Gamma_{n0}^{\chi}\right)W_n^T$$

From results Lemma 1 (i) we get:

$$\frac{1}{n}W_n^{T'}\left(\Gamma_{n0}^{xT} - \Gamma_{n0}^{\chi}\right)W_n^T \leq \frac{1}{n}\sqrt{\text{trace}\left[(\Gamma_{0n}^{xT} - \Gamma_{0n}^{x})^2\right]} = O\left(\frac{1}{\sqrt{T}}\right)$$

Moreover, $W_n^{T\prime}\Gamma_{n0}^{\xi T}W_n^T \leq \mu_{n1}^{\xi} = O_p(1)$ by assumption FM3. The desired result follows. *Q.E.D.*.

Lemma 2 Under assumption PA1-2, FM1-FM3, as $n, T \to \infty$, we have:

(i) $Q_{ni}^T - Q_{ni} = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$

(ii) $D_n^T - D_n = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$

(iii) $\Sigma_n^T - \Sigma_n = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$

where Q_{ni}^T and Q_{ni} denote the ith row of Q_n^T and Q_n, respectively.

Proof. Let us start from result (i). We have the following decomposition

$$Q_n^T = \frac{1}{\sqrt{n}}\Gamma_{n0}^{xT}W_n^T = \frac{1}{\sqrt{n}}\Gamma_{n0}^{\chi}W_n^T + \frac{1}{\sqrt{n}}\Gamma_{n0}^{\xi}W_n^T + \frac{1}{\sqrt{n}}\left(\Gamma_{n0}^{xT} - \Gamma_{n0}^{xT}\right)W_n^T$$

Write $\mathbf{1}_{ni}$ for the n dimensional vector with entries equal to zero at the ith position and zero for the rest. Consequently:

$$Q_{ni}^T = \mathbf{1}_{ni}'Q_n^T = \frac{1}{\sqrt{n}}\mathbf{1}_{ni}'\Gamma_{n0}^{xT}W_n^T = \frac{1}{\sqrt{n}}\mathbf{1}_{ni}'\Gamma_{n0}^{\chi}W_n^T + \frac{1}{\sqrt{n}}\mathbf{1}_{ni}'\Gamma_{n0}^{\xi}W_n^T + \frac{1}{\sqrt{n}}\mathbf{1}_{ni}'\left(\Gamma_{n0}^{xT} - \Gamma_{n0}^{xT}\right)W_n^T$$

Let us study separately each term of the right hand side. For the first term, Corollary 1 (ii), imply:

$$\frac{1}{\sqrt{n}}\mathbf{1}_{ni}'\Gamma_{n0}^{\chi}W_n^T = \frac{1}{\sqrt{n}}\mathbf{1}_{ni}'\Gamma_{n0}^{\chi}W_nW_n'W_n^T = Q_{ni}W_n'W_n^T = Q_{n1} + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

since $W_nW_n'A_n = A_n$ by Assumption FM0.

For the second term, we have:

$$\frac{1}{\sqrt{n}}\mathbf{1}_{ni}'\Gamma_{n0}^{\xi}W_n^T \leq \frac{1}{\sqrt{n}}\sqrt{\mathbf{1}_{ni}'\Gamma_{n0}^{\xi}\mathbf{1}_{ni}}\sqrt{W_n^{T\prime}\Gamma_{n0}^{\xi}W_n^T} \leq \frac{1}{\sqrt{n}}\mu_{n1}^{\xi} = O_p\left(\frac{1}{\sqrt{n}}\right)$$

from assumption FM3.

Writing w_{jh}^T for the entry of W_n^T in the jth row and the hth columns, the third term can be written as:

$\frac{1}{\sqrt{n}} \left| \mathbf{1}'_{ni} \left(\Gamma^{xT}_{n0} - \Gamma^{xT}_{n0} \right) W^T_n \right| \leq \frac{1}{\sqrt{n}} \sum_{h=1}^r \left| \sum_{j=1}^n (\gamma^{xT}_{0ij} - \gamma^x_{0ij}) w^T_{jh} \right|$

$\leq \frac{1}{\sqrt{n}} \sum_{h=1}^r \sqrt{\sum_{j=1}^n (\gamma^{xT}_{0ij} - \gamma^x_{0ij})^2} \sqrt{\sum_{j=1}^n (w^T_{jh})^2} = \frac{1}{\sqrt{n}} \sum_{h=1}^r \sqrt{\sum_{j=1}^n (\gamma^{xT}_{0ij} - \gamma^x_{0ij})^2}$

since W^T_n is orthonormal. Because $\mathrm{E}\left[\sum_{j=1}^n (\gamma^{xT}_{0ij} - \gamma^x_{0ij})^2\right] = O_p\left(\frac{n}{T}\right)$, from the Markow inequality, we get

$$\frac{1}{\sqrt{n}} \mathbf{1}'_{ni} \left(\Gamma^{xT}_{n0} - \Gamma^{xT}_{n0} \right) W^T_n = O_p\left(\frac{1}{\sqrt{T}}\right)$$

This proves result (i).

Turning to (ii), we have:

$$\frac{1}{n} D^T_n \Lambda^T_n = \frac{1}{n} W^{T\prime}_n \Gamma^{xT}_{n1} W^T_n = \frac{1}{n} W^{T\prime}_n \Gamma^{\chi}_{n1} W^T_n + \frac{1}{n} W^{T\prime}_n \Gamma^{\xi}_{n1} W^T_n + \frac{1}{n} W'_n (\Gamma^{xT}_{n1} - \Gamma^x_{n1}) W_n$$

¿From result (ii) of Corollary 1, we have:

$$\frac{1}{n} W^{T\prime}_n \Gamma^{\chi}_{n1} W^T_n = \frac{1}{n} (W^{T\prime}_n W_n) W'_n \Gamma^{\chi}_{n1} W_n (W'_n W^T_n) = \frac{1}{n} D_n \Lambda_n + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

since $W_n W'_n A_n = A_n$ by Assumption FM0.

By assumptions PA1-2 and FM3, $W^{T\prime}_n \Gamma^{\xi}_{n1} W^T_n = O_p(1)$. Moreover, Lemma 1 (i) implies that: $\frac{1}{n} W'_n (\Gamma^{xT}_{n1} - \Gamma^x_{n1}) W_n = O_p(\frac{1}{\sqrt{T}})$. Result (ii), hence, follows from Corollary 1 (i) and Assumption FM2.

Finally, result (iii) is an immediate consequence of Lemma 1 (i) and result (ii) above.
Q.E.D.

Proof of Proposition 2

Note that the matrix Σ_n is of fixed dimension r. Because of continuity of the eigenvalues and eigenvectors with respect to the matrix entries, by Lemma 2 (iii) and the continuous mapping theorem we have

$$M^T_n = M_n + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \quad \text{as } n, T \to \infty$$

and

$$K_n^T = K_n + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \quad \text{as } \ n, T \to \infty$$

Continuity of the matrix product (notice that D_n has fixed dimension r), implies:

$$\left(D_n^T\right)^h = (D_n)^h + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \quad \text{as } \ n, T \to \infty$$

Result (i) is hence an immediate consequence of Lemma 2 (i) and (ii).

For result (ii) and (iii), since $W_n W_n' A_n = A_n$ by Assumption FM0, we have:

$$\begin{aligned} \hat{\boldsymbol{g}}_t = \tfrac{1}{\sqrt{n}} W_n^{T\prime} \boldsymbol{x}_{nt} &= \tfrac{1}{\sqrt{n}} W_n^{T\prime} A_n \boldsymbol{f}_t + \tfrac{1}{\sqrt{n}} W_n^{T\prime} \boldsymbol{\xi}_{nt} \\ &= W_n^{T\prime} W_n G_n \boldsymbol{f}_t + \tfrac{1}{\sqrt{n}} W_n^{T\prime} \boldsymbol{\xi}_{nt} \end{aligned}$$

From FM3:

$$\mathrm{E}\left(\frac{1}{n} W_n^{T\prime} \boldsymbol{\xi}_{nt} \boldsymbol{\xi}_{nt}' W_n^T\right) \le \frac{1}{n} \mu_{n1}^{\xi} = O_p\left(\frac{1}{n}\right) \ \text{ as } n, T \to \infty$$

Denote $G_n^T = W_n^{T\prime} W_n G_n$. Assumption FM3 implies that G_n has full rank n. The desired result then follows from Corollary 1 (ii).

Q.E.D.

Appendix 3: Data description and data treatment

	Database	Original Source	Variable Description	ID Code in the Database	Units	Orig. Freq.	Seas. Adj.	Treatment
1	MW	Citibase	Per Capita Real Consumption Expenditure					DLOG
2	MW	Citibase	Per Capita Gross Private Domestic Fixed Investment					DLOG
3	MW	Citibase	Per Capita Private Gross National product					DLOG
4	MW	Citibase	Per Capita Real M2 (M2 divided by P)					DLOG
5	MW	Citibase	3-Month Treasury Bill Rate					D
6	MW	Citibase	Implicit Price Deflator for Private GNP					DDLOG
7	Fred II	BEA	Real Gross Domestic Product, 1 Decimal	GDPC1	Bil. of Ch. 1996 $	Q	YES	DLOG
8	Fred II	BEA	Real Final Sales of Domestic Product, 1 Decimal	FINSLC1	Bil. of Ch. 1996 $	Q	YES	DLOG
9	Fred II	BEA	Real Gross Private Domestic Investment, 1 Decimal	GPDIC1	Bil. of Ch. 1996 $	Q	YES	DLOG
10	Fred II	BEA	Real State & Local Cons. Expend. & Gross Inv., 1 Dec.	SLCEC1	Bil. of Ch. 1996 $	Q	YES	DLOG
11	Fred II	BEA	Real Private Residential Fixed Investment, 1 Dec.	PRFIC1	Bil. of Ch. 1996 $	Q	YES	DLOG
12	Fred II	BEA	Real Private Nonresidential Fixed Investment, 1 Dec.	PNFIC1	Bil. of Ch. 1996 $	Q	YES	DLOG
13	Fred II	BEA	Real Nonresidential Inv.: Equipment & Software, 1 Dec.	NRIPDC1	Bil. of Ch. 1996 $	Q	YES	DLOG
14	Fred II	BEA	Real Imports of Goods & Services, 1 Decimal	IMPGSC1	Bil. of Ch. 1996 $	Q	YES	DLOG
15	Fred II	BEA	Real Federal Cons. Expend. & Gross Investment, 1 Dec.	FGCEC1	Bil. of Ch. 1996 $	Q	YES	DLOG
16	Fred II	BEA	Real Government Cons. Expend. & Gross Inv., 1 Dec.	GCEC1	Bil. of Ch. 1996 $	Q	YES	DLOG
17	Fred II	BEA	Real Fixed Private Domestic Investment, 1 Decimal	FPIC1	Bil. of Ch. 1996 $	Q	YES	DLOG
18	Fred II	BEA	Real Exports of Goods & Services, 1 Decimal	EXPGSC1	Bil. of Ch. 1996 $	Q	YES	DLOG
19	Fred II	BEA	Real Change in Private Inventories, 1 Decimal	CBIC1	Bil. of Ch. 1996 $	Q	YES	NONE
20	Fred II	BEA	Real Personal Cons. Expenditures: Nondurable Goods	PCNDGC96	Bil. of Ch. 1996 $	Q	YES	DLOG
21	Fred II	BEA	Real State & Local Government: Gross Investment	SLINVC96	Bil. of Ch. 1996 $	Q	YES	DLOG
22	Fred II	BEA	Real Personal Consumption Expenditures: Services	PCESVC96	Bil. of Ch. 1996 $	Q	YES	DLOG
23	Fred II	BEA	Real Personal Cons. Expenditures: Durable Goods	PCDGCC96	Bil. of Ch. 1996 $	Q	YES	DLOG
24	Fred II	BEA	Real Personal Consumption Expenditures	PCECC96	Bil. of Ch. 1996 $	Q	YES	DLOG
25	Fred II	BEA	Real National Defense Gross Investment	DGIC96	Bil. of Ch. 1996 $	Q	YES	DLOG
26	Fred II	BEA	Real Federal Nondefense Gross Investment	NDGIC96	Bil. of Ch. 1996 $	Q	YES	DLOG
27	Fred II	BEA	Real Disposable Personal Income	DPIC96	Bil. of Ch. 1996 $	Q	YES	DLOG
28	Fred II	BEA	Personal Cons. Expenditures: Chain-type Price Index	PCECTPI	Index 1996 = 100	Q	YES	DDLOG
29	Fred II	BEA	Gross Domestic Product: Chain-type Price Index	GDPCTPI	Index 1996 = 100	Q	YES	DDLOG
30	Fred II	BEA	Gross Domestic Product: Implicit Price Deflator	GDPDEF	Index 1996 = 100	Q	YES	DDLOG
31	Fred II	BEA	Gross National Product: Implicit Price Deflator	GNPDEF	Index 1996 = 100	Q	YES	DDLOG
32	Fred II	BEA	Gross National Product: Chain-type Price Index	GNPCTPI	Index 1996 = 100	Q	YES	DDLOG
33	Fred II	BLS	Nonfarm Business Sector: Unit Labor Cost	ULCNFB	Index 1996 = 100	Q	YES	DLOG
34	Fred II	BLS	Nonfarm Business Sector: Real Compensation Per Hour	COMPRNFB	Index 1992 = 100	Q	YES	DLOG
35	Fred II	BLS	Nonfarm Bus. Sector: Output Per Hour of All Persons	OPHNFB	Index 1992 = 100	Q	YES	DLOG
36	Fred II	BLS	Nonfarm Business Sector: Compensation Per Hour	COMPNFB	Index 1992 = 100	Q	YES	DLOG
37	Fred II	BLS	Manufacturing Sector: Unit Labor Cost	ULCMFG	Index 1992 = 100	Q	YES	DLOG
38	Fred II	BLS	Manufacturing Sector: Output Per Hour of All Persons	OPHMFG	Index 1992 = 100	Q	YES	DLOG
39	Fred II	BLS	Business Sector: Output Per Hour of All Persons	OPHPBS	Index 1992 = 100	Q	YES	DLOG
40	Fred II	BLS	Business Sector: Compensation Per Hour	HCOMPBS	Index 1992 = 100	Q	YES	DLOG
41	Fred II	St.	Louis St. Louis Adjusted Reserves	ADJRESSL	Bil. of $	M	YES	DLOG
42	Fred II	St. Louis	St. Louis Adjusted Monetary Base	AMBSL	Bil. of $	M	YES	DLOG
43	Fred II	Moody's	Moody's Seasoned Aaa Corporate Bond Yield	AAA	%	M	NO	D
44	Fred II	Moody's	Moody's Seasoned Baa Corporate Bond Yield	BAA	%	M	NO	D
45	Fred II	FR	Bank Prime Loan Rate	MPRIME	%	M	NO	D
46	Fred II	FR	3-Month Treasury Bill: Secondary Market Rate	TB3MS	%	M	NO	D
47	Fred II	FR	Currency in Circulation	CURRCIR	Bil. of $	M	NO	DD4LOG
48	Fred II	FR	Currency Component of M1	CURRSL	Bil. of $	M	YES	DDLOG
49	Fred II	BLS	CPI for All Urban Consumers: All Items Less Food	CPIULFSL	Ind. 1982-84 = 100	M	YES	DDLOG
50	Fred II	BLS	Consumer Price Index for All Urban Consumers: Food	CPIUFDSL	Ind. 1982-84 = 100	M	YES	DDLOG
51	Fred II	BLS	CPI For All Urban Consumers: All Items	CPIAUCSL	Ind. 1982-84 = 100	M	YES	DDLOG
52	Fred II	BLS	CPI: Intermediate Materials: Supplies & Components	PPIITM	Index 1982 = 100	M	YES	DDLOG
53	Fred II	BLS	Producer Price Index: Industrial Commodities	PPIIDC	Index 1982 = 100	M	NO	DDLOG
54	Fred II	BLS	PPI: Fuels & Related Products & Power	PPIENG	Index 1982 = 100	M	NO	DDLOG
55	Fred II	BLS	PPI Finished Goods: Capital Equipment	PPICPE	Index 1982 = 100	M	YES	DDLOG
56	Fred II	BLS	Producer Price Index: Finished Goods	PPIFGS	Index 1982 = 100	M	YES	DDLOG
57	Fred II	BLS	Producer Price Index: Finished Consumer Goods	PPIFCG	Index 1982 = 100	M	YES	DDLOG
58	Fred II	BLS	Producer Price Index: Finished Consumer Foods	PPIFCF	Index 1982 = 100	M	YES	DDLOG
59	Fred II	BLS	PPI: Crude Materials for Further Processing	PPICRM	Index 1982 = 100	M	YES	DDLOG
60	Fred II	BLS	Producer Price Index: All Commodities	PPIACO	Index 1982 = 100	M	NO	DLOG
61	Fred II	FR	Commercial and Industrial Loans at All Comm. Banks	BUSLOANS	Bil. of $	M	YES	DLOG
62	Fred II	FR	Total Loans and Leases at Commercial Banks	LOANS	Bil. of $	M	YES	DLOG
63	Fred II	FR	Total Loans and Investments at All Commercial Banks	LOANINV	Bil. of $	M	YES	DLOG
64	Fred II	FR	Total Consumer Credit Outstanding	TOTALSL	Bil. of $	M	YES	DLOG
65	Fred II	FR	Real Estate Loans at All Commercial Banks	REALLN	Bil. of $	M	YES	DLOG
66	Fred II	FR	Other Securities at All Commercial Banks	OTHSEC	Bil. of $	M	YES	DLOG
67	Fred II	FR	Consumer (Individual) Loans at All Comm. Banks	CONSUMER	Bil. of $	M	YES	DLOG
68	Fred II	BLS	All Employees: Construction	USCONS	Thous.	M	YES	DLOG
69	Fred II	BLS	Total Nonfarm Payrolls: All Employees	PAYEMS	Thous.	M	YES	DLOG
70	Fred II	BLS	Employees on Nonfarm Payrolls: Manufacturing	MANEMP	Thous.	M	YES	DLOG
71	Fred II	BLS	Unemployed: 16 Years & Over	UNEMPLOY	Thous.	M	YES	DLOG
72	Fred II	BLS	Civilian Unemployment Rate	UNRATE	%	M	YES	DLOG
73	Fred II	BLS	Civilian Participation Rate	CIVPART	%	M	YES	DLOG
74	Fred II	BLS	Civilian Labor Force	CLF16OV	Thous.	M	YES	DLOG
75	Fred II	BLS	Civilian Employment: Sixteen Years & Over	CE16OV	Thous.	M	YES	DLOG
76	Fred II	BLS	Civilian Employment-Population Ratio	EMRATIO	%	M	YES	DLOG

	Database	Original Source	Variable Description	ID Code in the Database	Units	Orig. Freq.	Seas. Adj.	Treatment
77	EconStats	FR	Industrial Production: total	Index		M	YES	DLOG
78	EconStats	FR	Industrial Production: Manufacturing (SIC-based)	Index		M	YES	DLOG
79	Datastream	ISM	ISM Manufacturers Survey: Supplier Delivery Index	USNAPMDL	Index	M	YES	NONE
80	Datastream	ISM	Chicago Purchasing Manager Business Barometer	USPMCUBB	%	M	NO	NONE
81	Datastream	ISM	ISM Manufacturers Survey: New Orders Index	USNAPMNO	Index	M	YES	NONE
82	Datastream	ISM	ISM Manufacturers Survey: Employment Index	USNAPMIV	Index	M	YES	NONE
83	Datastream	ISM	ISM Manufacturers Survey: Production Index	USNAPMEM	Index	M	YES	NONE
84	Datastream	ISM	ISM Purchasing Managers Index (MFG Survey)	USNAPMPR	Index	M	YES	NONE
85	Datastream	BC	Manufacturing Shipments - Total	USMNSHIPB	Bil. of $	M	YES	DLOG
86	Datastream	BC	Shipments of Durable Goods	USSHDURGB	Bil. of $	M	YES	DLOG
87	Datastream	BC	Shipments of Non-Durable Goods	USSHNONDB	Bil. of $	M	YES	DLOG
88	Datastream	S&P	Standard & Poor's 500 (monthly average)	US500STK	Index	M	NO	DLOG
89	Datastream	FT	Dow Jones Industrial Share Price Index	USSHRPRCF	Index	M	NO	DLOG

Abbreviations:
MW: Mark Watson's home page (http://www.wws.princeton.edu/ mwatson/publi.html)
Fred II: Fred II database of the Federal Reserve Bank of St. Louis
BEA: Bureau of Economic Analysis
BLS: Bureau of Labor Statistics
FR: Federal Reserve Board
St Louis: Federal Reserve Bank of St. Louis
ISM: Institute for Supply Management
BC: Bureau of Census
S&P: Standard & Poors'
FT: Financial Times
Q: Quarterly
M: Monthly (we take quarterly averages)

References

[1] Altug, S. (1989). Time-to-Build and Aggregate Fluctuations: Some New Evidence, *International Economic Review*, **30**, pp.889-920.

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

[3] Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions, *Econometrica*, 71, 135-171.

[4] Bernanke, B. S., and J. Boivin (2003). Monetary Policy in a Data Rich environment, *Journal of Monetary Economics* **50**, pp. 525-546.

[5] Bernanke, B. S., J. Boivin and P. Eliasz (2005). Measuring Monetary Policy: A Factor Augmented Autoregressive (FAVAR) Approach, *Quarterly Journal of Economics* **120**, pp.387-422.

[6] Boivin, J. and S. Ng, (2003), Are more data always better for factor analysis?, NBER Working Paper no. 9829. *Journal of Econometrics* forthcoming.

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

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

[9] Chari, V. V., P. J. Kehoe and E. R. Mcgrattan (2005). A Critique of Structural VARs Using Real Business Cycle Theory, Federal Reserve Bank of Minneapolis Working no. 631.

[10] Connor, G. and R.A. Korajczyk (1988). Risk and return in an equilibrium APT. Application of a new test methodology. *journal of Financial Economics* **21**, pp.255-89.

[11] Fernandez-Villaverde, J., J. Rubio-Ramirez and T. J. Sargent (2005). A, B, C's (and D)'s for Understanding VARs. NBER Technical Working Papers no. 0308.

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

[13] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2002). The generalized factor model: one-sided estimation and forecasting. CEPR Discussion Paper no. 3701. *Journal of the American Statistical Association*, forthcoming.

[14] Forni, M., and M. Lippi (2001). The generalized dynamic factor model: representation theory. *Econometric Theory* **17**, 1113-41.

[15] 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.

[16] 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.

[17] Geweke J. F., and K. J. Singleton (1981). Maximum Likelihood "Confirmatory" Factor Analysis of Economic Time Series, *International Economic Review*, **22**, pp.37-54.

[18] Giannone, D., L. Reichlin and L. Sala (2002). Tracking Greenspan: Systematic and Nonsystematic Monetary Policy Revisited, CEPR Discussion Paper no. 3550.

[19] Giannone, D., L. Reichlin L. and L. Sala (2003). VARs, Factor Models and the Empirical Validation of Equilibrium Business Cycle Models, CEPR Discussion Paper no. 3701. *Journal of Econometrics* forthcoming.

[20] Giannone, D., L. Sala and L. Reichlin (2005). Monetary Policy in Real Time. In M. Gertler and K. Rogoff, Eds., *NBER Macroeconomic Annual, 2004*, MIT Press.

[21] Hansen, L.P., and T.J. Sargent (1991) Two problems in interpreting vector autoregressions. In *Rational Expectations Econometrics*, L.P. Hansen and T.J. Sargent, eds. Boulder: Westview, pp.77-119.

[22] Kilian, L. (1998). Small-Sample Confidence Intervals for Impulse Response Functions *Review of Economics and Statistics* **80**, pp.218-30.

[23] King, R.G., C. I. Plosser, J. H. Stock and M.W. Watson (1991). Stochastic Trends and Economic Fluctuations *American Economic Review* **81**, pp.819-40.

[24] Lippi, M., and L. Reichlin (1993). The dynamic effects of aggregate demand and supply disturbances: Comment. *American Economic Review* **83**, pp.644-52.

[25] Lippi, M., and L. Reichlin (1994). VAR analysis, non fundamental representation, Blaschke matrices. *Journal of Econometrics* **63**, pp.307-25.

[26] Quah, D., and Sargent, T. J. (1993) A Dynamic Index Model for Large Cross Sections, in J. Stock and M. Watson, Eds., *Business Cycle, Indicators and Forecasting*, University of Chicago Press and NBER, Chicago.

[27] Rudebush, G.D. (1998) Do measures of monetary policy in a VAR make sense? *International Economic Review* **39**, pp.907-31.

[28] Sargent, T. J. (1989). Two Models of Measurements and the Investment Accelerator *The Journal of Political Economy*, **97**, pp.251-287.

[29] Sargent, T.J., C.A. Sims (1977). Business cycle modelling 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.

[30] Stewart, G. W., and Ji-Guang Sun (1990), *Matrix Perturbation Theory.* Academic Press, Inc., San Diego.

[31] Stock, J.H., and M.W. Watson (2002a) Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics* **20**, 147-162.

[32] Stock, J.H., and M.W. Watson (2002b) Forecasting Using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association* **97**, pp.116779.