



## WORKING PAPER SERIES

### **Opening the black box: structural factor models with large cross-sections**

*Mario Forni, Domenico Giannone, Marco Lippi and  
Lucrezia Reichlin*

Working paper 8

November 2007

[www.recent.unimore.it](http://www.recent.unimore.it)

# 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

November 12, 2007

## Abstract

This paper shows how large-dimensional dynamic factor models are suitable for structural analysis. We argue that all identification schemes employed in SVAR analysis can be easily adapted in dynamic factor models. Moreover, the “problem of fundamentalness”, which is intractable in structural VARs, can be solved, provided that the impulse-response functions are sufficiently heterogeneous. We provide consistent estimators for the impulse-response functions, as well as  $(n, T)$  rates of convergence. An exercise with US macroeconomic data shows that our solution of the fundamentalness problem may have important empirical consequences.

JEL subject classification : E0, C1

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

---

\*We would like to thank Manfred Deistler and Marc Hallin for helpful suggestions and participants to the conference Common Features in Rio, 2002, and to the Forecasting Seminar of the NBER Summer Institute, July 2002. M. Forni and M. Lippi are grateful to MUR (Italian Ministry of University and Research) for financial support. D. Giannone and L. Reichlin were supported by a PAI contract of the Belgian Federal Government and an ARC grant of the Communauté Française de Belgique

# 1 Introduction

Recent literature has shown that large-dimensional approximate (or generalized) dynamic factor models can be used successfully to forecast macroeconomic variables (Forni, Hallin, Lippi and Reichlin, 2005, Stock and Watson, 2002a, 2002b, Boivin and Ng, 2003, Giannone, Reichlin and Sala, 2005). These models assume that each time series in the dataset 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 highly singular, i.e. is driven by a very small number (as compared to the number of variables) of shocks (the “common shocks” or “common factors”). Indeed, evidence based on different datasets points to the robust finding that few shocks explain the bulk of dynamics of macro data (see Sargent and Sims, 1977 and Giannone, Reichlin and Sala, 2002 and 2005). If the common component of the variable to be predicted is large, a forecasting method based on a projection on linear combinations of these shocks performs well because, while being parsimonious, it captures the relevant comovements in the economy.

Here we argue that the scope of dynamic factor models goes beyond forecasting. Our aim is to open the black box of these models and show how statistical constructs such as factors can be related to macroeconomic shocks and their propagation mechanisms.

We define *macroeconomic* shocks those structural sources of variation that are cross-sectionally pervasive, i.e. that significantly affect most of the variables of the economy, as opposed to *idiosyncratic* sources of variation, that are specific to a single variable or a small group of variables, hence capturing both sectoral-local dynamics (let us say “micro” dynamics) and measurement error. Our aim is identification of the macroeconomic shocks and their dynamic effect on macroe-

conomic variables, whereas the idiosyncratic components are disregarded.

A key paper, in which the distinction between macroeconomic shocks and idiosyncratic sources of variation is systematically exploited for macroeconomic modeling is Sargent and Sims (1977), in which several models, both “Keynesian” and “classical”, are reformulated as factor models with a small number of macroeconomic shocks. More recent literature includes papers in which Dynamic Stochastic General Equilibria (DSGE), augmented with measurement errors, are estimated by maximum likelihood (augmenting a theory-based model with measurement errors goes back to Sargent, 1989; see also Altug, 1989, Ireland, 2004 and the literature mentioned therein; for an explicit link to factor models see Giannone, Reichlin and Sala, 2006, Boivin and Giannoni, 2006).

The approach we propose here is a combination of Structural vector Autoregression (SVAR) analysis and large-dimensional dynamic factor models. Precisely, the factor model is used to consistently estimate common and idiosyncratic components of macroeconomic variables. Then we apply SVAR analysis to identify the relationship between common components and macroeconomic shocks.

Our approach differs from error-augmented DSGE models in that we estimate the impulse-response functions of the macroeconomic variables to macroeconomic shocks without imposing any theory-based dynamic restriction. It has a close relationship to FAVAR models, in which a VAR is augmented with common factors (see Bernanke, Boivin and Elias, 2005). The link between factor models, FAVAR and VAR models has been studied in Stock and Watson (2005), who show how SVAR techniques can be used in a factor-model context. However, our analysis of the fundamentalness of the structural shocks in factor models, and the consequent motivation for an autoregressive approximation (see below and Section 3), is a distinctive feature of the present paper. An early work in which a large factor model is used for structural analysis is Forni and Reichlin (1998);

major differences with the present paper are the empirical focus and the proposed estimation procedure.

To give a brief outline of the structure of the paper, suppose that we are interested in key macroeconomic variables such as per-capita consumption, income and investment, denoted by  $c_t$ ,  $y_t$  and  $i_t$  (see our empirical exercise in Section 5). The macrovariables  $c_t$ ,  $y_t$  and  $i_t$  are embedded in a large macroeconomic dataset (the number of variables in our exercise is 89), and modeled as a common component, driven by structural macroeconomic shocks, plus an idiosyncratic component (variable specific shocks and measurement error). Under fairly general assumptions the common components can be estimated consistently (see Section 2).

The vector of the common components, call it  $\boldsymbol{\chi}_{nt}$ , has dimension  $n$ , the number of variables in the dataset, and rank  $q$ , the number of macroeconomic shocks (three in our exercise), and is therefore highly singular. A crucial step in our analysis is the dynamic specification of  $\boldsymbol{\chi}_t$  as a (singular) vector autoregression driven by the macroeconomic shocks. This implies assuming that the macroeconomic shocks are *fundamental* for the common components  $\boldsymbol{\chi}_{nt}$ . Section 3 is dedicated to showing that the fundamentalness problem, a weakness of SVAR analysis, finds a satisfactory solution within our approach (on the fundamentalness issue in SVAR models see Hansen and Sargent, 1991, Lippi and Reichlin, 1993 and 1994 and, more recently, Chari, Kehoe and Mcgrattan, 2005, Fernández-Villaverde, Rubio-Ramirez and Sargent, 2005, Giannone, Reichlin and Sala, 2006). Non fundamentalness of structural shocks is a consequence—this is the usual explanation—of the agents having an information set that is larger than the econometrician’s. We argue that in large-dimensional factor models, in which the number of observed variables is larger than the number of shocks (unlike in SVAR models), such “superior information” can occur only by a fluke (on the

importance of this feature for monetary models, see Bernanke and Boivin, 2003 and Giannone, Reichlin and Sala, 2002 and 2005).

Once the vector autoregressive specification for  $\boldsymbol{\chi}_{nt}$  has been motivated, we show that all the identification techniques developed in SVAR analysis, such as long-run or impact effects, can be successfully imported in the identification of structural macroeconomic shocks within large-dimensional dynamic factor models. Like in SVAR analysis, the structural shocks are obtained by linearly transforming the estimated residual vector  $\boldsymbol{v}_t$ , the key difference being that here the number of shocks  $q$  is smaller than the number of variables. Lastly, we can go back to the variables of interest and study their dynamic response to structural macroeconomic shocks. Section 5 analyses an empirical example on US macroeconomic data which revisits the results of King, Plosser, Stock and Watson (1991) in the light of our discussion on fundamentalness.

Section 4 studies consistency and rates of convergence for the estimators of the shocks and the impulse response functions.

## 2 The Large-Dimensional Dynamic Factor Model

The dynamic factor model used in this paper 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:

**Assumption 1.**  $\mathbf{x}_n^T$  is a finite realization of a real-valued stochastic process

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

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

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

are stationary, with zero mean and finite second-order moments  $\Gamma_k^x = \mathbf{E}[\mathbf{x}_{nt}\mathbf{x}_{n,t-k}']$ ,  $k \in \mathbb{Z}$ .

We assume that each variable  $x_{it}$  is the sum of two unobservable components, the *common component*  $\chi_{it}$  and the *idiosyncratic component*  $\xi_{it}$ . The common components are driven by  $q$  *common shocks*  $\mathbf{u}_t = (u_{1t} \ u_{2t} \ \cdots \ u_{qt})'$ . Note that  $q$  is independent of  $n$  (and small as compared to  $n$  in empirical applications). Precisely, defining  $\boldsymbol{\chi}_{nt} = (\chi_{1t} \ \cdots \ \chi_{nt})'$  and  $\boldsymbol{\xi}_{nt} = (\xi_{1t} \ \cdots \ \xi_{nt})'$ :

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

where:

**Assumption 2.**  $\mathbf{u}_t$  is a  $q$ -dimensional orthonormal white noise,  $B_n(L)$  is a nested sequence of one-sided  $n \times q$  absolutely summable matrix polynomials (infinite in general). Moreover, there exist an integer  $r \geq q$ , a nested sequence of  $n \times r$  matrices  $A_n$  and a one-sided absolutely summable  $r \times q$  matrix polynomial (infinite in general)  $N(L)$ , such that

$$B_n(L) = A_n N(L). \tag{2.2}$$

Defining the  $r \times 1$  vector  $\mathbf{f}_t$  as

$$\mathbf{f}_t = N(L)\mathbf{u}_t, \tag{2.3}$$

(2.1) can be rewritten in the static form

$$\mathbf{x}_{nt} = A_n \mathbf{f}_t + \boldsymbol{\xi}_{nt} \quad (2.4)$$

In the sequel, we shall use the term *static factors* to denote the  $r$  entries of  $\mathbf{f}_t$ , whereas the common shocks  $\mathbf{u}_t$  will be also referred to as *dynamic factors*.

The common shocks  $\mathbf{u}_t$  and  $B_n(L)$  are assumed to be *structural* sources of variation and impulse-response functions respectively. Therefore model (2.1), as specified in Assumptions 1 and 2 and the other assumptions below, is a *structural factor model*.

By contrast, the static factors  $\mathbf{f}_t$ , the matrix  $A_n$  and  $N(L)$  have no structural interpretation and are not unique. For, if  $\mathbf{g}_t = G\mathbf{f}_t$ , where  $G$  is  $r \times r$  and invertible, then  $\mathbf{x}_{nt} = [A_n G^{-1}]\mathbf{g}_t + \boldsymbol{\xi}_{nt}$ , with  $\mathbf{g}_t = [GN(L)]\mathbf{u}_t$ , is another static representation for  $\mathbf{x}_{nt}$ .

**Assumption 3.** (*Orthogonality of common and idiosyncratic components*) For all  $n$ , the vector  $\boldsymbol{\xi}_{nt}$  is stationary. Moreover,  $\mathbf{u}_t$  is orthogonal to  $\xi_{i\tau}$ ,  $i \in \mathbb{N}$ ,  $t \in \mathbb{Z}$ ,  $\tau \in \mathbb{Z}$ .

The assumption of orthogonality between common and idiosyncratic components has an economic justification. Interpreting the factor model as the joint model of the economy and the statistical agency, under reasonable hypotheses on the behavior of the statistical agency, the latter is orthogonal to the signal captured, in our framework, by the common shocks (see Sargent, 1989 for a discussion). Moreover, orthogonality between common and idiosyncratic components ensures that the entries of  $B_n(L)$  can be interpreted as impulse-response functions of the common shocks on the  $\chi$ 's and on the variables  $x_{it}$  themselves.

Some definitions are needed for the next two assumptions. Let  $\Gamma_k^\chi$  be the  $k$ -lag covariance matrix of  $\boldsymbol{\chi}_{nt}$ , and denote by  $\mu_j^\chi$  the  $j$ -th eigenvalue, in decreasing

order, of  $\Gamma_0^\chi$ . Moreover, let  $\Sigma^\chi(\theta)$  and  $\Sigma^\xi(\theta)$  be the spectral density matrix of  $\boldsymbol{\chi}_{nt}$  and  $\boldsymbol{\xi}_{nt}$  respectively, and denote by  $\lambda_j^\chi(\theta)$  and  $\lambda_j^\xi(\theta)$  their eigenvalues as functions of  $\theta \in [-\pi \ \pi]$ , in decreasing order.

To avoid heavy notation, indication of the dependence on  $n$  and  $T$  is kept to a minimum. In particular, dependence on  $n$  of  $\Gamma_k^\chi$ ,  $\mu_j^\chi$ , etc., just defined, and of other scalars and matrices defined below, is not made explicit. In the same way, reference to  $T$  and  $n$  will be avoided for estimated scalars and matrices. For example, the estimator of  $\Gamma_0^x$ , the covariance matrix of  $\boldsymbol{x}_{nt}$ , is denoted by  $\hat{\Gamma}_0^x$ .

**Assumption 4.** (*Pervasiveness of common dynamic and static factors*)

(a) As  $n \rightarrow \infty$  we have  $\lambda_q^\chi(\theta) \rightarrow \infty$  for  $\theta$  almost everywhere in  $[-\pi \ \pi]$ .

(b) There exists constants  $\underline{c}_j, \bar{c}_j$ ,  $j = 1, \dots, r$ , such that  $\underline{c}_j > \bar{c}_{j+1}$ ,  $j = 1, \dots, r-1$ , and

$$0 < \underline{c}_j < \liminf_{n \rightarrow \infty} n^{-1} \mu_j^\chi \leq \limsup_{n \rightarrow \infty} n^{-1} \mu_j^\chi \leq \bar{c}_j$$

**Assumption 5.** (*Non-pervasiveness of the idiosyncratic components*) There exists a real  $\mathcal{L}$  such that  $\lambda_1^\xi(\theta) \leq \mathcal{L}$  for any  $n \in \mathbb{N}$  and  $\theta$  a.e. in  $[-\pi \ \pi]$ . This obviously implies that  $\mu_1^\xi \leq \mathcal{L}$  for any  $n \in \mathbb{N}$ ,  $\mu_j^\xi$  being the  $j$ -th eigenvalue of  $\Gamma_{n0}^\xi$ .

Assumption 5 includes the case in which the idiosyncratic components are mutually orthogonal with an upper bound for the spectral densities (and therefore for the variances). Mutual orthogonality is the usual condition in finite-dimensional factor models. Assumption 3 relaxes such condition by allowing for a limited amount of cross-correlation among the idiosyncratic components. Assumption 4 (pervasiveness of the common factors) implies that each of the common shocks  $u_{jt}$  affects (almost) all the variables  $x_{it}$ ,  $i \in \mathbb{N}$ , with non-declining coefficients.

Some comments on our assumptions are in order:

(I) The number  $q$  of dynamic factors and the components  $\chi_{it}$  are identified. In particular, a representation of the form (2.1)-(2.4) with a different number of dynamic factors is not possible (see Forni and Lippi, 2001).

(II) Assumption 4(b) is necessary to identify  $r$ . In particular, a static representation of the common components  $\chi_{it}$  with a different number of static factors is not possible.

(III) We define the static and dynamic rank of  $\mathbf{f}_t$  as the rank of, respectively, its variance-covariance and spectral density matrix. By Assumption 4(a) the dynamic rank of  $\mathbf{f}_t$  is  $q$  for  $\theta$  a.e. in  $[-\pi \pi]$ . Assumption 4(b) entails that, for  $n$  sufficiently large,  $A_n$  has full rank  $r$  and that  $\mathbf{f}_t$  has static rank  $r$  for any given  $t$ . Thus, for any given  $t$ , the space spanned by  $\chi_{it}$ ,  $i \in \mathbb{N}$ , coincides with the space spanned by the static factors  $f_{jt}$ ,  $j = 1, \dots, r$ , and has therefore dimension  $r$ .

The following dynamic factor model has been often considered in the large-dimensional factor-model literature (see Stock and Watson, 2002a, 2002b, 2005, Bai and Ng, 2007, Forni, Hallin, Lippi and Reichlin, 2005):

$$\mathbf{x}_{nt} = C_{n0}\mathbf{f}_t^* + C_{n1}\mathbf{f}_{t-1}^* + \dots + C_{ns}\mathbf{f}_{t-s}^*, \quad (2.5)$$

where  $\mathbf{f}_t^*$  is  $q$ -dimensional and the matrices  $C$  are  $n \times q$  and nested, and that  $\mathbf{f}_t^*$  has the VAR representation:

$$\Theta(L)\mathbf{f}_t^* = (1 - \Theta_1L - \dots - \Theta_hL^h)\mathbf{f}_t^* = \mathbf{u}_t, \quad (2.6)$$

where  $\Theta(L)$  is  $q \times q$ . Using the definitions

$$\begin{aligned} \mathbf{f}_t &= (\mathbf{f}_t^{*'} \mathbf{f}_{t-1}^{*'} \dots \mathbf{f}_{t-s}^{*'})', & A_n &= (C_{n0} \ C_{n1} \ \dots \ C_{ns}), \\ N(L) &= (K(L) \ K(L)L \ \dots \ K(L)^{-1}L^s)', \end{aligned}$$

where  $K(L) = (\Theta(L)')^{-1}$ , we have  $\mathbf{f}_t = N(L)\mathbf{u}_t$  and

$$\mathbf{x}_{nt} = A_n\mathbf{f}_t + \boldsymbol{\xi}_{nt}. \quad (2.7)$$

The static rank of  $\mathbf{f}_t$  is always  $q(s+1)$ . However, in order for (2.7) to be a static representation of the model it is necessary that  $A_n$  be full rank, and this depends on the coefficients of the matrices  $C_{nj}$ :

(i) If no restrictions among the coefficients of the matrices  $C_{nj}$  hold (assume for example that they are independently drawn from the same distribution), then (2.7) is a static representation of the model.

(ii) If restrictions hold, such that  $A_n$  is not full rank, then  $r < q(s + 1)$  and obtaining a static representation requires further manipulation. For example, assume that  $q = 1$ ,  $s = 1$ , so that (2.5) can be written as  $\chi_{it} = c_{i0}u_t + c_{i1}u_{t-1}$ . If no restrictions hold among the  $c$ 's, then  $r = 2$  and (2.7) is a static representation. But if the restriction  $c_{i1} = ac_{i0}$  holds, then  $r = 1$ ,  $N(L) = 1 + aL$ ,  $\mathbf{f}_t = (1 + aL)\mathbf{u}_t$  and  $A_n = (c_{10} \ c_{20} \ \cdots \ c_{n0})'$ .

In any case, with or without restrictions, existence of a static representation for model (2.5)-(2.6) is an immediate consequence of the following remark:

**(R)** Assume that  $\boldsymbol{\chi}_{nt} = B_n(L)\mathbf{u}_t$ . Denoting by  $\mathcal{X}_t$  the space spanned by  $\chi_{it}$ ,  $i \in \mathbb{N}$ , if  $\mathcal{X}_t$  is finite dimensional, then  $\boldsymbol{\chi}_{nt}$  has a static representation  $\boldsymbol{\chi}_{nt} = A_n N(L)\mathbf{u}_t$ .

For, let  $r$  be the dimension of  $\mathcal{X}_t$ , for a given  $t$ . Stationarity of  $\boldsymbol{\chi}_{nt}$  implies that (i)  $r$  is independent of  $t$ , (ii)  $\mathcal{X}_t$  has a stationary basis  $\mathbf{f}_t = (f_{1t} \ f_{2t} \ \cdots \ f_{rt})$ , (iii)  $\chi_{it} = a_i \mathbf{f}_t$  with  $a_i$  independent of  $t$ . As  $\mathbf{f}_t \in \mathcal{X}_t$ ,  $\boldsymbol{\chi}_{nt} = B_n(L)\mathbf{u}_t$  implies that  $\mathbf{f}_t$  can be represented as  $N(L)\mathbf{u}_t$ .

Model (2.5)-(2.6) implies that the entries of  $N(L)$  are rational functions of  $L$ . Conversely, assuming that the entries of  $N(L)$  are rational functions of  $L$  implies that the model can be put in the form (2.5)-(2.6). This is fairly obvious. If  $\phi_j(L)$  is the least common multiple of the denominators of the entries in the  $j$ -th column of  $N(L)$ , then  $N(L) = N_1(L)N_2(L)^{-1}$ , where  $N_1(L)$  is a  $r \times q$  moving average and  $N_2(L)$  is  $q \times q$  with the polynomials  $\phi_j(L)^{-1}$  on the main diagonal and zero elsewhere. Thus the following is equivalent to assuming (2.5)-(2.6).

**Assumption 6.** *The entries of  $N(L)$  are rational functions of  $L$ .*

Note that if Assumption 6 holds for the vector  $\mathbf{f}_t$ , i.e. for a static representation, then it holds for all static representations.

Our last problem is a specification of  $N(L)$  which makes the model suitable for identification and estimation of the shocks  $\mathbf{u}_t$ . A standard solution is the assumption that  $N(L)$  results from inversion of a VAR, that is

$$\mathbf{f}_t - D_1\mathbf{f}_{t-1} - \dots - D_m\mathbf{f}_{t-m} = R\mathbf{u}_t,$$

where  $R$  is a  $r \times q$  matrix, so that  $N(L) = (I - D_1L - \dots - D_mL^m)^{-1}$ . This assumption implies, as shown in Proposition 2 below, that  $\mathbf{u}_t$  is identified up to a unitary matrix. However, the VAR specification also implies that  $\mathbf{u}_t$  belongs to the space spanned by present and past values of the variables  $\chi_{it}$ , i.e. that  $\mathbf{u}_t$  is fundamental for the  $\chi$ 's. This is the issue that will be thoroughly discussed in the next section.

## 3 Fundamentalness of the Structural Shocks

### 3.1 Response heterogeneity, $n$ large and fundamentalness

**3.1.1** Let us begin by briefly recalling some basic notions on fundamental representations of stationary stochastic vectors. Assume that the  $n$ -dimensional stochastic vector  $\boldsymbol{\mu}_t$  admits a moving average representation, i.e. that there exist a  $q$ -dimensional white noise  $\mathbf{v}_t$  and an  $n \times q$ , one-sided, square-summable filter  $K(L)$ , such that

$$\boldsymbol{\mu}_t = K(L)\mathbf{v}_t. \tag{3.8}$$

If  $\mathbf{v}_t$  belongs to the space spanned by present and past values of  $\boldsymbol{\mu}_t$  we say that representation (3.8) is *fundamental* and that  $\mathbf{v}_t$  is fundamental for  $\boldsymbol{\mu}_t$  (the condition defining fundamentalness is also referred to as the *miniphase assumption*; see e.g. Hannan and Deistler, 1988, p. 25). With no substantial loss of generality we can suppose that  $q \leq n$  and that  $\mathbf{v}_t$  is full rank. Moreover, for our purpose,

we can suppose that the entries of  $K(L)$  are rational functions of  $L$  and that the rank of  $K(z)$  is maximal, i.e.  $q$ , except for a finite number of complex numbers. Then (see e.g. Rozanov, 1967, Ch. 1, Section 10, and Ch. 2, p. 76):

**(F)** *Representation (3.8) is fundamental if and only if the rank of  $K(z)$  is  $q$  for all  $z$  such that  $|z| < 1$ .*

Assuming that (3.8) is fundamental, all fundamental white-noise vectors  $\mathbf{z}_t$  are linear transformations of  $\mathbf{v}_t$ , i.e.  $\mathbf{z}_t = C\mathbf{v}_t$  (see Proposition 2 below). Non fundamental white-noise vectors result from  $\mathbf{v}_t$  by means of linear filters that involve the so-called Blaschke matrices (see e.g. Lippi and Reichlin, 1994).

A fundamental white noise naturally arises with linear prediction. Precisely, the prediction error

$$\mathbf{w}_t = \boldsymbol{\mu}_t - \text{Proj}(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}, \boldsymbol{\mu}_{t-2}, \dots)$$

is white noise and fundamental for  $\boldsymbol{\mu}_t$ . As a consequence, when estimating an ARMA with forecasting purposes, the MA matrix polynomial is always chosen to be invertible, which implies fundamentalness.

Fundamentalness plays also an important role for the identification of structural shocks in SVAR analysis. SVAR analysis starts with the projection of a full rank  $n$ -dimensional vector  $\boldsymbol{\mu}_t$  on its past, thus producing an  $n$ -dimensional full rank fundamental white noise  $\mathbf{w}_t$ . The structural shocks are then obtained as a linear transformation  $A\mathbf{w}_t$ , the matrix  $A$  resulting from economic theory statements, which is tantamount to assuming that the structural shocks are fundamental. Fundamentalness has here the effect that the identification problem is enormously simplified. However, as pointed out in the literature mentioned in the Introduction (see also Section 3.1.2 below), economic theory, in general, does not provide support for fundamentalness, so that all representations that fulfill

the same economic statements but are non fundamental are ruled out with no justification.

Our main point is that the situation changes dramatically if structural analysis is conducted assuming that  $n > q$ . Precisely, as we see below, non fundamentalness is a generic property for  $n = q$ , while it is non generic for  $n > q$ . Thus the question “why assuming fundamentalness?”, which is legitimately asked when  $n = q$ , is replaced by “why should we care about non fundamentalness?” when  $n > q$ .

An easy and effective illustration can be obtained assuming that  $q = 1$ , that the entries of  $K(L) = (K_1(L) \ K_2(L) \ \cdots \ K_n(L))'$  are polynomials whose degree does not exceed  $s$ , so that  $K(L)$  is parameterized in  $\mathbb{R}^{n(s+1)}$ . In this case, if  $n = q = 1$ , non fundamentalness translates into the condition that at least one root of  $K_1(z)$  has modulus smaller than unity. Continuity of the roots of  $K_1(z)$  implies that non fundamentalness is generic, i.e. that if it holds for a point  $\kappa$  in the parameter space it holds also within a neighborhood of  $\kappa$ .

On the other hand, if  $n > q$ , by (F), non fundamentalness implies that the polynomials  $K_j(z)$  have a common root. As a consequence, their coefficients must fulfill  $n - 1$  *equality constraints* (see e.g. van der Waerden, 1953, p. 83). Non fundamentalness is therefore non generic.

**3.1.2** The discussion above has a forceful macroeconomic counterpart. Let us firstly adapt to our framework the classic permanent-income consumption model, as used in Fernández-Villaverde, Rubio-Ramirez, Sargent and Watson (2006) to illustrate non fundamentalness. With minor changes in notation:

$$\begin{aligned} c_t &= c_{t-1} + \sigma_u(1 - R^{-1})u_t \\ y_t - c_t &= -c_{t-1} + \sigma_u R^{-1}u_t, \end{aligned}$$

where  $c_t$  is permanent consumption,  $y_t$  is labour income,  $u_t$  is a white-noise process and  $R$  is a constant gross interest rate. The authors assume that the variable

$y_t - c_t$ , call it  $s_t$ , is observed by the econometrician whereas  $c_t$  is not. From the equations above we obtain

$$s_t - s_{t-1} = \sigma_u R^{-1}(1 - RL)u_t, \quad (3.9)$$

so that, as  $R > 1$ ,  $u_t$  is not fundamental for  $s_t$ . Therefore the VAR for  $s_t$  (the best the econometrician can do), which is just a univariate autoregression, would produce an innovation which is not the structural shock  $u_t$ . However, if the econometrician observes  $c_t$ , or  $y_t$ , or the value of the consumer's accumulated assets, then  $u_t$  becomes fundamental (p. 5). Precisely,  $u_t$  can be recovered using present and past values of  $s_t$  and another variable, whereas present and past values of  $s_t$  alone are not sufficient.

This extremely simple example contains all the elements we need to motivate fundamentalness of the structural shocks  $\mathbf{u}_t$  for  $\boldsymbol{\chi}_{nt}$ .

1. As a rule non fundamentalness arises when the econometrician's information set is smaller than the agent's (see Hansen and Sargent, 1980, 1991; also the learning-by-doing example in Lippi and Reichlin, 1993, can be reformulated in terms of information sets). In the permanent income model the agent observes permanent income whereas the econometrician does not.

2. However if any additional variable  $z_t = b(L)u_t$  is observed, then, by proposition (F),  $u_t$  is fundamental for the singular vector  $(s_t \ z_t)'$ , unless  $b(R^{-1}) = 0$ . For example, if  $z_t = \alpha(u_t - \beta u_{t-1})$ , then  $\beta \neq R$  is sufficient for fundamentalness of  $u_t$  for  $(s_t \ z_t)'$ .

3. In our framework, the agent still observes  $c_t$  and  $s_t$ , while the econometrician observes

$$x_{1t} = s_t + \xi_{1t},$$

i.e.  $s_t$  plus measurement error. However, we also assume that  $x_{1t}$  belongs to a large dataset  $x_{it} = \chi_{it} + \xi_{it}$ , which is observed by the econometrician. The

common components  $\chi_{it}$  can be recovered using our large-dimensional factor-model techniques. Moreover, assuming for simplicity that  $q = 1$ , like in the permanent-income example, the unique structural shock  $u_t$  is fundamental for the vector of the common components, unless *all the responses*  $b_i(L)$  fulfill the extremely unlikely constraint  $b_i(R^{-1}) = 0$ .

In general, if the variables  $\chi_{it}$  are driven by  $q$  shocks, a macroeconomic model that contains only  $q$  variables, suppose they are  $\chi_{jt}$ ,  $j = 1, \dots, q$ , cannot ensure fundamentalness of  $\mathbf{u}_t$ , the reason being possible superior information of the agents with respect to present and past values of  $\chi_{jt}$ ,  $j = 1, \dots, q$ . However, the informational advantage of the agents disappears if the econometrician observes a large set of additional macroeconomic variables. The generating process of  $\chi_{jt}$ ,  $j = q + 1, \dots, n$ , contains parameters that do not belong to the generating process of the first  $q$ , and viceversa. Therefore, with all likelihood, their dynamic responses to  $\mathbf{u}_t$  are sufficiently *heterogeneous*, with respect to the first  $q$ , to prevent the rank reduction which is, by Proposition F, equivalent to non fundamentalness.

**3.1.3** Based on the discussion above we assume fundamentalness of  $\mathbf{u}_t$  for  $\chi_{it}$ ,  $i \in \mathbb{N}$ .

**Proposition 1.** *Under Assumptions 1, 2 and 4, fundamentalness of  $\mathbf{u}_t$  for  $\chi_{it}$ ,  $i \in \mathbb{N}$ , is equivalent to left invertibility of  $N(L)$ , i.e. to the existence of a  $q \times r$  filter  $G(L)$  such that  $G(L)N(L) = I_q$ . Moreover, under 1, 2, 3, 4 and 5,  $\mathbf{u}_t$  belongs to the space spanned by present and past values of  $x_{it}$ ,  $i = 1, \dots, \infty$ , i.e. is fundamental for  $x_{it}$ ,  $i \in \mathbb{N}$ .*

*Proof.* If  $\mathbf{u}_t$  is fundamental for  $\chi_{it}$ ,  $i \in \mathbb{N}$ , then it is fundamental for  $\mathbf{f}_t$ , i.e. there exists a  $q \times r$  filter  $G(L)$  such that  $\mathbf{u}_t = G(L)\mathbf{f}_t = G(L)N(L)\mathbf{u}_t$ . As  $\mathbf{u}_t$  is a white noise,  $G(L)N(L) = I_q$ . Now assume that  $G(L)N(L) = I_q$ . Assumption 4 implies that  $A'_n A_n$  is full rank for  $n$  sufficiently large. Setting,

$S_n(L) = G(L) (A'_n A_n)^{-1} A'_n$ , we have  $S_n(L)\mathbf{x}_{nt} = S_n(L)\boldsymbol{\chi}_{nt} + S_n(L)\boldsymbol{\xi}_{nt}$ . Now

$$S_n(L)\boldsymbol{\chi}_{nt} = G(L) (A'_n A_n)^{-1} A'_n A_n \mathbf{f}_t = G(L)\mathbf{f}_t = G(L)N(L)\mathbf{u}_t = \mathbf{u}_t.$$

Therefore  $\mathbf{u}_t$  lies in the space spanned by present and past values of  $\boldsymbol{\chi}_{nt}$ . Moreover,  $S_n(L)\boldsymbol{\xi}_{nt} = G(L) (A'_n A_n)^{-1} A'_n \boldsymbol{\xi}_t$  converges to zero in mean square by assumptions 4 and 5. Q.E.D.

The proof above also shows that fundamentalness of  $\mathbf{u}_t$  for  $\chi_{it}$ ,  $i \in \mathbb{N}$  is equivalent to fundamentalness of  $\mathbf{u}_t$  for  $\boldsymbol{\chi}_{nt}$  for  $n$  sufficiently large. In view of Proposition 1, our fundamentalness assumption will be formulated as follows:

**Assumption 7.** (*Fundamentalness*) *There exists a  $q \times r$  one-sided filter  $G(L)$  such that  $G(L)N(L) = I_q$ .*

Obviously, if Assumption 7 holds for a particular static representation then it holds for all static representations.

Starting with representation (2.5)-(2.6), if no restrictions hold among the coefficients of the matrices  $C_{nj}$ , we have  $\tilde{N}(L) = (K(L) K(L)L \cdots K(L)L^s)'$ , which has left inverse  $(\Theta(L) 0_q \cdots 0_q)$ . Thus, as we can obviously expect, no restrictions implies “maximum heterogeneity” of the responses to the structural shocks and therefore fundamentalness. To see the effect of restrictions consider again the example with  $q = 1$  and  $\chi_{it} = c_{i0}u_t + c_{i1}u_{t-1}$  (see Section 2). If the restriction  $c_{i1} = ac_{i0}$  holds we have  $r = 1$  and  $N(L) = 1 + aL$ . In this extreme case Assumption 7, i.e.  $|a| < 1$ , is no less arbitrary as the fundamentalness assumption in VAR analysis. When restrictions hold but  $r > q$ , Assumption 7 rules out lower dimensional subsets of parameter space.

To introduce our last assumption, a VAR specification for  $\mathbf{f}_t$ , let us consider the orthogonal projection of  $\mathbf{f}_t$  on the space spanned by its past values:

$$\mathbf{f}_t = \text{Proj}(\mathbf{f}_t \mid \mathbf{f}_{t-1}, \mathbf{f}_{t-2}, \dots) + \mathbf{w}_t, \quad (3.10)$$

where  $\mathbf{w}_t$  is the  $r$ -dimensional vector of the residuals. Under our assumptions,  $\mathbf{w}_t$  has rank  $q$ . Moreover, by the same argument used to prove Proposition 2 (see the next subsection), Assumption 7 implies that  $\mathbf{w}_t = R\mathbf{u}_t$ , where  $R$  is a maximum-rank  $r \times q$  matrix.

To get some insight in the orthogonal projection (3.10), consider again representation (2.5)-(2.6) with no restrictions. The static representation of the model has  $r = q(s + 1)$  and

$$\mathbf{f}_t = (K(L) \ K(L)L \ \cdots \ K(L)L^s)\mathbf{u}_t.$$

In particular, assuming that  $m \leq s$ ,  $\mathbf{f}_t$  has the AR(1) representation

$$\begin{pmatrix} \mathbf{f}_t^* \\ \mathbf{f}_{t-1}^* \\ \vdots \\ \mathbf{f}_{t-s}^* \end{pmatrix} = \begin{pmatrix} \Theta_1 & \Theta_2 & \cdots & \Theta_{s-1} & \Theta_s \\ I_q & 0_q & \cdots & 0_q & 0_q \\ 0_q & I_q & \cdots & 0_q & 0_q \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_q & 0_q & \cdots & I_q & 0_q \end{pmatrix} \begin{pmatrix} \mathbf{f}_{t-1}^* \\ \mathbf{f}_{t-2}^* \\ \vdots \\ \mathbf{f}_{t-s-1}^* \end{pmatrix} + \begin{pmatrix} I_q \\ 0_q \\ \vdots \\ 0_q \end{pmatrix} \mathbf{u}_t$$

where  $\Theta_j = 0_q$  if  $j > m$ . If  $m > s$  the order of the VAR is higher (but still finite).

Joining this observation with the usual approximation argument, a specification of  $\mathbf{f}_t$  as

$$\mathbf{f}_t = D_1\mathbf{f}_{t-1} + \cdots + D_h\mathbf{f}_{t-h} + R\mathbf{u}_t, \quad (3.11)$$

even with  $h$  very small, does not seem to cause a dramatic loss of generality. In the sequel we will adopt the VAR(1) specification:

**Assumption 7'.** (*Fundamentalness: VAR(1) specification*) *The  $r$ -dimensional static factors  $\mathbf{f}_t$  admit a VAR(1) representation*

$$\mathbf{f}_t = D\mathbf{f}_{t-1} + R\mathbf{u}_t \quad (3.12)$$

where  $D$  is  $r \times r$  and  $R$  is a maximum-rank matrix of dimension  $r \times q$ .

Under (3.12),

$$\boldsymbol{\chi}_{nt} = B_n(L)\mathbf{u}_t = A_n(I - DL)^{-1}R\mathbf{u}_t. \quad (3.13)$$

Note that assuming (3.12), or (3.11), is independent of the particular static factors we choose. For, let  $\mathbf{g}_t = G\mathbf{f}_t$ , where  $G$  is  $r \times r$  and invertible, be another basis in the space spanned by the  $\chi_{it}$ 's. If  $\mathbf{f}_t$  fulfills (3.12), then  $\mathbf{g}_t = [GDG^{-1}]\mathbf{g}_{t-1} + [GR]\mathbf{u}_t$ .

A convenient alternative formulation of Assumption 7' is

**Assumption 7''.** *The  $r$ -dimensional static factors  $\mathbf{f}_t$  admit a VAR(1) representation*

$$\mathbf{f}_t = D\mathbf{f}_{t-1} + \boldsymbol{\epsilon}_t \quad (3.14)$$

where  $D$  is  $r \times r$  and  $\boldsymbol{\epsilon}_t$  is a white noise of rank  $q$ .

### 3.2 Alternative fundamental representations

Our next result shows that if  $\boldsymbol{\chi}_{nt} = C_n(L)\mathbf{v}_t$  is a given fundamental representation, then  $\mathbf{u}_t$  can be obtained from  $\mathbf{v}_t$  my means of a static rotation.

**Proposition 2.** *Consider the common components of model (2.1)*

$$\boldsymbol{\chi}_{nt} = B_n(L)\mathbf{u}_t. \quad (3.15)$$

under Assumptions 1 through 7. If

$$\boldsymbol{\chi}_{nt} = C_n(L)\mathbf{v}_t \quad (3.16)$$

for any  $n \in \mathbb{N}$ , where the matrices  $C_n(L)$  are nested and  $\mathbf{v}_t$  is a  $q$ -dimensional fundamental orthonormal white noise vector, then representation (3.16) is related to representation (3.15) by

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

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

*Proof.* Projecting  $\mathbf{u}_t$  entry by entry on the linear space  $\mathcal{V}_t^-$  spanned by present and past values of  $v_{ht}$ ,  $h = 1, \dots, q$  we get

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

where  $\mathbf{r}_t$  is orthogonal to  $\mathbf{v}_{t-k}$ ,  $k \geq 0$ . Now consider that  $\mathcal{V}_t^-$  and the space spanned by present and past values of  $\chi_{it}$ ,  $i \in \mathbb{N}$ , call it  $\mathcal{X}_t^-$ , are identical, because the entries of  $\boldsymbol{\chi}_{t-k}$ ,  $k \leq 0$ , belong to  $\mathcal{V}_t^-$  by equation (3.16), while the entries of  $\mathbf{v}_{t-k}$ ,  $k \leq 0$ , belong to  $\mathcal{X}_t^-$  by assumption. The same is true for  $\mathcal{X}_t^-$  and the space spanned by present and past values of  $u_{ht}$ ,  $i = 1, \dots, q$ , call it  $\mathcal{U}_t^-$ , so that  $\mathcal{U}_t^- = \mathcal{V}_t^-$ . Hence  $\mathbf{r}_t = 0$ . Moreover, serial non-correlation of the  $v_{ht}$ 's imply that  $\sum_{k=1}^{\infty} H_k \mathbf{v}_{t-k}$  is the projection of  $\mathbf{u}_t$  on  $\mathcal{V}_{t-1}^-$ , which is zero because  $\mathcal{V}_{t-1}^- = \mathcal{U}_{t-1}^-$ . It follows that  $\mathbf{u}_t = H_0 \mathbf{v}_t$ . Orthonormality of  $\mathbf{u}_t$  implies that  $H_0$  is unitary  $H_0 H_0' = I$ . QED

## 4 Identification and Estimation

### 4.1 Variables of interest, identification

Proposition 2 has the consequence that structural analysis in large-dimensional factor models can be carried on along the same lines of standard SVAR analysis. Precisely:

(A) We select the variables of interest, the first  $m$  with no loss of generality.

Usually  $m = q$ .

(B) We determine a  $q$ -dimensional vector  $\mathbf{v}_t$ , which is fundamental for  $\chi_{it}$ ,  $i \in \mathbb{N}$ , and the corresponding representation  $\boldsymbol{\chi}_{mt} = C_m(L) \mathbf{v}_t$ .

(C) We assume that economic theory implies a set of zero and sign restrictions that uniquely determines the structural impulse-response function  $B_m(L)$  (just identification), i.e. that economic theory identifies a rotation  $H$  such that  $B_m(L) = C_m(L) H'$  and  $\mathbf{u}_t = H \mathbf{v}_t$ .

(D) We construct a consistent estimator  $\hat{B}_m(L)$  which is consistent with rate  $\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)$ .

Assuming that the variables of interest have been selected, let us concentrate on step B. Denote by  $W^\chi$  the  $n \times r$  matrix whose  $j$ -th column is the normalized eigenvector of  $\Gamma_0^\chi$ :

$$W^{\chi'}\Gamma_0^\chi = M^\chi W^{\chi'}, \quad (4.17)$$

where  $M^\chi$  is the  $r \times r$  diagonal matrix having  $\mu_j^\chi$  as entry  $(j, j)$ . Then define

$$\mathbf{g}_t = \frac{1}{\sqrt{n}} W^{\chi'} \boldsymbol{\chi}_{nt}. \quad (4.18)$$

The entries of  $\mathbf{g}_t$  are the first  $r$  (non-normalized) principal components of  $\boldsymbol{\chi}_{nt}$ . Assumption 4(b) implies that for  $n$  large enough  $\mathbf{g}_t$  is a basis for the space  $\mathcal{X}_t$ , thus a vector of static factors. A fundamental representation for  $\boldsymbol{\chi}_{mt}$  is now easily obtained:

(i) By Assumption 7'',  $\mathbf{g}_t = D\mathbf{g}_{t-1} + \boldsymbol{\epsilon}_t$ , where, using  $\Gamma_k^g = E(\mathbf{g}_t \mathbf{g}'_{t-k}) = \frac{1}{n} W^{\chi'} \Gamma_k^\chi W^\chi$ ,

$$D = \Gamma_1^g (\Gamma_0^g)^{-1} = W^{\chi'} \Gamma_1^\chi W^\chi \left( \frac{M^\chi}{n} \right)^{-1}, \quad (4.19)$$

(ii) Setting  $\Gamma^\epsilon = E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t)$ , we have

$$\Gamma^\epsilon = \Gamma_0^g - D\Gamma_0^g D' = \frac{M^\chi}{n} - D \frac{M^\chi}{n} D'. \quad (4.20)$$

(iii) Now let  $\mu_j^\epsilon$ ,  $j = 1, \dots, q$ , be the  $j$ -th eigenvalue of  $\Gamma^\epsilon$ , in decreasing order,  $\mathcal{M}$  the  $q \times q$  diagonal matrix with  $\sqrt{\mu_j^\epsilon}$  as its  $(j, j)$  entry,  $K$  the  $r \times q$  matrix with the corresponding normalized eigenvectors on the columns. Defining  $\mathbf{v}_t = \mathcal{M}^{-1} K' \boldsymbol{\epsilon}_t$  and  $\mathcal{K} = K\mathcal{M}$ ,

$$\boldsymbol{\chi}_{mt} = Q_m (I - DL)^{-1} \mathcal{K} \mathbf{v}_t = \left( \sum_{h=0}^{\infty} Q_m D^h \mathcal{K} \right) \mathbf{v}_t = C_m(L) \mathbf{v}_t, \quad (4.21)$$

where, using (4.17) and defining  $\mathcal{I}_m = (I_m \ 0_{m, n-m})'$  (the  $n \times m$  matrix with zero on the last  $n - m$  rows and  $I_m$  on the first  $m$  rows),

$$Q_m = E(\boldsymbol{\chi}_{mt} \mathbf{g}'_t) [E(\mathbf{g}_t \mathbf{g}'_t)]^{-1} = \sqrt{n} \mathcal{I}'_m W^\chi. \quad (4.22)$$

Note that  $\mathbf{g}_t$ ,  $Q_m$ ,  $D$ ,  $\mathcal{M}$ ,  $K$  and  $\mathbf{v}_t$ , all depend on  $n$ . Note also that  $\mathbf{v}_t$  is fundamental for  $\mathbf{g}_t$ , i.e. for all the  $\chi$ 's, not necessarily for  $\boldsymbol{\chi}_{mt}$ . In other words,  $\mathbf{v}_t$  can be linearly recovered using contemporaneous and past values of *all* the  $\chi$ 's, not necessarily the first  $m$  of the  $\chi$ 's (see Section 5.3 on this point).

The proof of our consistency result will need that the first  $q$  eigenvalues of the matrix  $\Gamma^\epsilon$  be distinct and asymptotically bounded away from zero (like the eigenvalues of  $\Gamma_0^x/n$ , see Assumption 4 (b)).

**Assumption 8.** *There exists constants  $\underline{d}_i$  and  $\bar{d}_i$ ,  $i = 1, \dots, q$ , such that  $\underline{d}_i > \bar{d}_{i+1}$ ,  $i = 1, \dots, q - 1$  and*

$$0 < \underline{d}_i < \liminf_{n \rightarrow \infty} \mu_i^\epsilon \leq \limsup_{n \rightarrow \infty} \mu_i^\epsilon < \bar{d}_i$$

Let us now briefly discuss Step C. The assumption of just identification can be formalized as follows:

- (i) Start with any representation  $\boldsymbol{\chi}_{mt} = S_1(I - S_2L)^{-1}S_3\mathbf{s}_t = S(L)\mathbf{s}_t$ , where  $\mathbf{s}_t$  is fundamental for the  $\chi$ 's.
- (ii) The restrictions implied by economic theory determine a *rule*, i.e. a *function*  $F$ , associating a unitary  $q \times q$  matrix with any triple  $S_1, S_2, S_3$ , such that

$$B_m(L) = S(L)F(S_1, S_2, S_3)', \quad \mathbf{u}_t = F(S_1, S_2, S_3)\mathbf{s}_t.$$

In particular, setting  $H = F(Q_m, D, \mathcal{K})$ , we have  $B_m(L) = C_m(L)H'$ ,  $\mathbf{u}_t = H\mathbf{v}_t$ .

## 4.2 Estimation

Let us start by some definitions and notation:

- (I)  $\hat{\Gamma}_k^x = \frac{1}{T} \sum_{h=k+1}^T \mathbf{x}_{nt}\mathbf{x}'_{nt-k}$ ,
- (II)  $\hat{\mu}_j^x$  the  $j$ -th eigenvalue of  $\hat{\Gamma}_0^x$ ,
- (III)  $\hat{M}^x$  the  $r \times r$  diagonal matrix with  $\hat{\mu}_j^x$  as its entry  $(j, j)$ ,

(IV)  $\hat{W}^x$  the  $n \times r$  matrix with the corresponding normalized eigenvectors on the columns.

The main motivation for using the static factors  $\mathbf{g}_t$ , as defined in (4.18), is that  $\mathbf{g}_t$  can be approximated in probability by the sample principal components of  $\mathbf{x}_{nt}$ :

$$\hat{\mathbf{g}}_t = \frac{1}{\sqrt{n}} \hat{W}^{x'} \mathbf{x}_t.$$

However, our consistency proof is not based on this result. Rather, we will directly deal with  $\hat{Q}_m$ ,  $\hat{D}$  and  $\hat{\mathcal{K}}$ , which are defined like  $Q_m$ ,  $D$  and  $\mathcal{K}$ , respectively, with  $\Gamma_k^x$ ,  $M^x$  and  $W^x$  replaced by  $\hat{\Gamma}_k^x$ ,  $\hat{M}^x$  and  $\hat{W}^x$  respectively. Therefore we can define  $\hat{C}(L) = \hat{Q}_m(I - \hat{D}L)^{-1}\hat{\mathcal{K}}$ ,  $\hat{H} = F(\hat{Q}_m, \hat{D}, \hat{\mathcal{K}})$  and the estimated impulse-response function  $\hat{B}_m(L) = \hat{C}_m(L)\hat{H}'$ . In the Appendix, under an additional technical assumption, we prove the following result:

**Proposition 3.** *For all  $k \geq 0$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, q$ ,*

$$|b_{ij,k} - \hat{b}_{ij,k}| = O_p \left( \max \left( \frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}} \right) \right). \quad (4.23)$$

In Section 5 we consider a case of partial identification, in which  $m = q = 3$ . The restrictions allow identification of the third column of  $B_3(L)$ , call it  $B_{3,3}(L)$ , i.e. the entries corresponding to the third common shock, but not of the whole  $B_3(L)$ . Quite obviously, taking as  $F$  any one of the infinite functions fulfilling the restrictions, (4.23) can be applied to  $B_{3,3}(L)$ .

In conclusion, (4.23) applies with just or partial identification. Overidentification is left to further research.

## 5 Empirical application

We illustrate our structural factor model by revisiting an influential work in SVAR literature, namely the three-dimensional SVAR estimated in King, Plosser, Stock

and Watson (1991) (KPSW henceforth). The variables are US per-capita output, investment and consumption, partial identification of the permanent shock and corresponding impulse-response functions being achieved by imposing long-run neutrality of the remaining shocks on output.

Our exercise is based on a panel of macroeconomic series including the three series used by KPSW, with the same sampling period. As we see below, three common shocks, i.e.  $q = 3$ , are consistent with our dataset. Moreover, upon estimation of the common components, the variance of the idiosyncratic components of output and investment accounts for about 15% of their total variance, the fraction falling to 10% for consumption. Thus, using the same identification restrictions applied in KPSW, allows a sensible and interesting comparison between our impulse-response functions and those found in KPSW.

## 5.1 The data

The data set is quarterly and is based on the FRED II database, Federal Reserve Bank of St. Louis, and Datastream. The original data of KPSW are available on Mark Watson's home page. We collected 89 series, including data from NIPA tables, price indices, productivity, industrial production indices, 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 make results independent of the units of measurement. A complete description of each series and the related transformations is available on request.

## 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 the total variance.

Regarding the variables of interest, the common factors explain 85.6% of total variance for output, 84.4% for investment and 89.4% for consumption. Bai and Ng estimators were criticized for easily overestimating the number of static factors when the idiosyncratic terms are strongly correlated. As a robustness check we therefore repeated our exercise with  $r = 9$ . Results are available upon request. The main conclusions do not change.

Regarding  $q$ , the criterion proposed by Hallin and Liška (2007), non-log criterion  $IC_1$ , for different choices of the parameters and the penalty functions, produces values of  $q$  within the range 2-5. Thus the value  $q = 3$ , necessary to carry on the comparison between our results and KPSW's, does not conflict with available evidence.

## 5.3 Fundamentalness

We are interested in the impulse-response functions of per-capita output, investment and consumption, that is, with no loss of generality, in the matrix  $C_3(L)H'$ . The question here is that although  $C_n(L)$ , which is  $n \times 3$ , is fundamental by assumption for  $n$  sufficiently large, the  $3 \times 3$  matrix  $C_3(L)$  is not necessarily

fundamental<sup>1</sup>. In other words, the common shocks can be recovered using contemporaneous and past values of the  $n$  common components, but we do not know whether the first three are sufficient.

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 the roots must be larger than unity 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 declines and lies always within the unit circle. For  $r \geq 22$  even the second smallest root becomes smaller than unity in modulus.

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

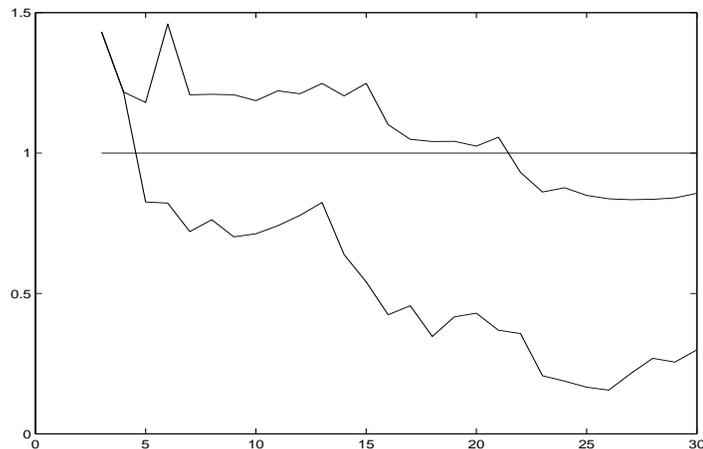
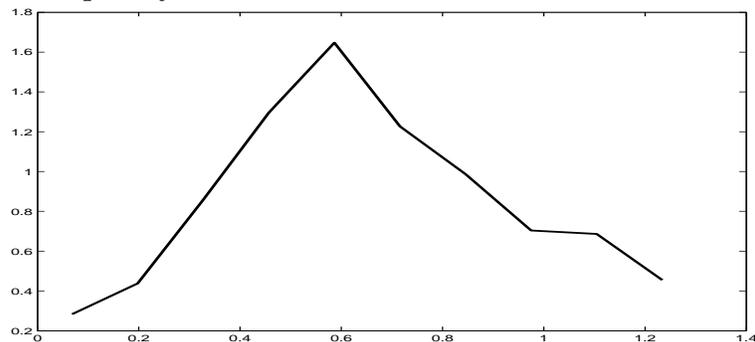


Figure 2 reports the distribution of the modulus of the smallest root for  $r = 15$ , across 1000 replications for a standard block bootstrap on the  $x$ 's; the length of the blocks was chosen to be equal to 22 quarters, large enough to retain the cyclical information in the series. The mean value is 0.66. The percentage of estimated values larger than one in modulus is 14.5.

Bootstrapping results strongly favour non fundamentalness of the structural impulse-response function  $C_3(L)H'$ . This implies that  $C_3(L)H'$  cannot be obtained by estimating a VAR for the three-dimensional vector  $(\chi_{1t} \chi_{2t} \chi_{3t})$ . As

<sup>1</sup>Note that fundamentalness of  $C_3(L)$  and of  $C_3(L)H'$  are equivalent.

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



we argue in Section 5.4, non fundamentalness of  $C_3(L)$  explains some important differences between our structural impulse-response function and KPSW's.

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

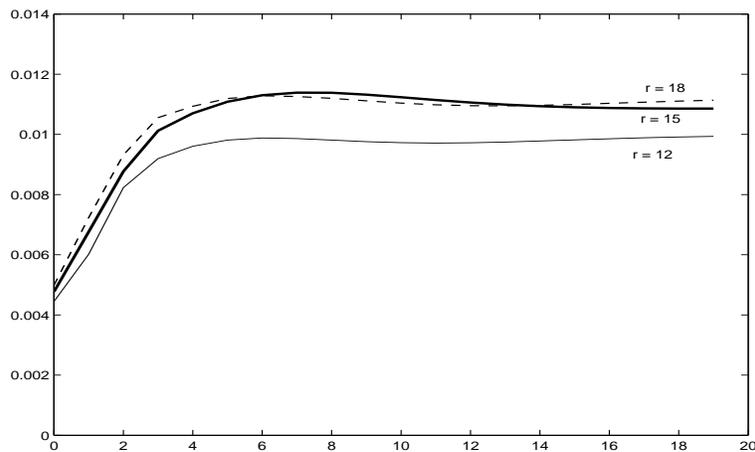
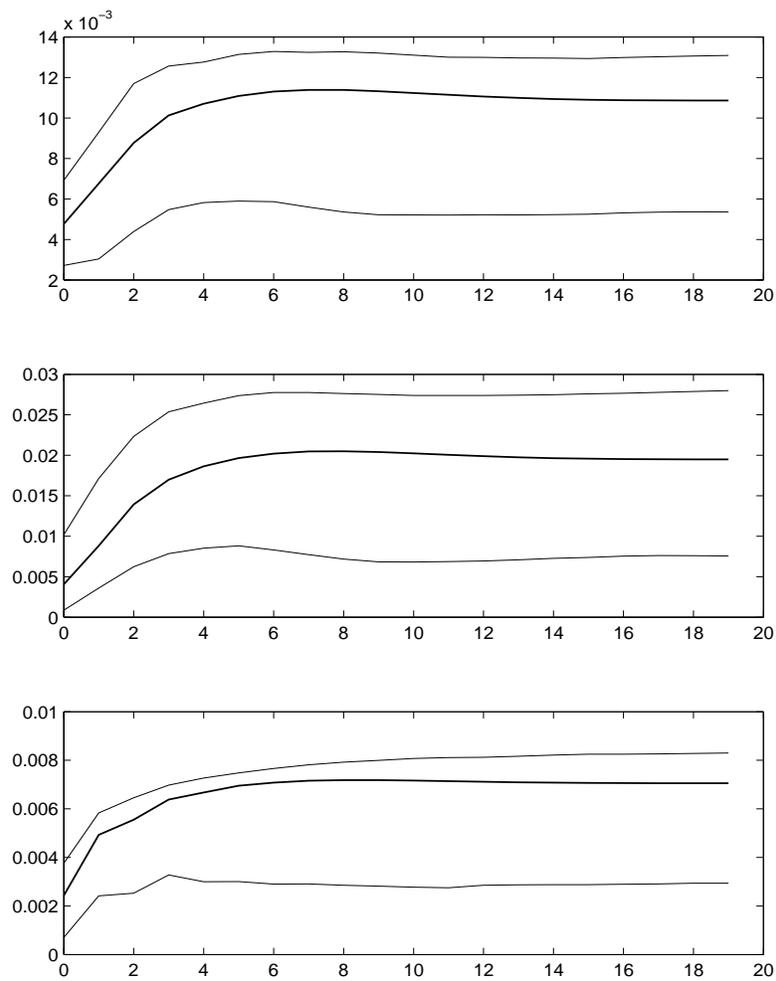


Figure 3 shows the response functions of per capita output for  $r = 12, 15, 18$ .

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



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 by means of the block bootstrap technique mentioned above). The shapes are similar for the three variables, with a positive impact effect followed by important, though declining, positive lagged effects.

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, the difference being very large for investment: at horizon 20 (5 years) the permanent shock explains 46% of its forecast error variance according to KPSW as against 86% with the factor model. Thus 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, seems to find a solution in our factor model.

As the variance of the idiosyncratic components of output, investment and consumption does not exceed 15% of their total variance (see Section 5.2), non fundamentalness of the structural shocks for  $(\chi_{1t} \chi_{2t} \chi_{3t})$ , as opposed to fundamentalness of KPSW's shocks for  $(x_{1t} x_{2t} x_{3t})$ , appears to play a major role in explaining such different dynamic profiles.

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

Horizon	Dynamic factor model			KPSW vector ECM		
	Output	Cons.	Inv.	Output	Cons.	Inv.
1	0.37 (0.20)	0.30 (0.26)	0.07 (0.16)	0.45 (0.28)	0.88 (0.21)	0.12 (0.18)
4	0.57 (0.18)	0.77 (0.18)	0.42 (0.17)	0.58 (0.27)	0.89 (0.19)	0.31 (0.23)
8	0.78 (0.13)	0.87 (0.13)	0.72 (0.13)	0.68 (0.22)	0.83 (0.18)	0.40 (0.18)
12	0.86 (0.09)	0.90 (0.11)	0.80 (0.11)	0.73 (0.19)	0.83 (0.18)	0.43 (0.17)
16	0.89 (0.08)	0.91 (0.11)	0.83 (0.10)	0.77 (0.17)	0.85 (0.16)	0.44 (0.16)
20	0.91 (0.07)	0.92 (0.12)	0.86 (0.09)	0.79 (0.16)	0.87 (0.15)	0.46 (0.16)

## 6 Conclusions

We have argued that dynamic factor models are suitable for structural macroeconomic modeling and provide an interesting alternative to structural VARs.

As we have shown, a large panel with a small number of common shocks allow the econometrician to recover the structural shocks under a reasonable assumption on the heterogeneity of the impulse-response functions. Thus the fundamentalness problem, which has no solution in the VAR framework, where  $m$  shocks must be recovered using present and past values of  $m$  variables, becomes tractable when the number of variables exceeds the number of shocks.

Our empirical application revisits a SVAR estimated in King, Plosser, Stock and Watson (1991) for US output, investment and consumption. Using a large panel including such series, we estimate a factor model with three common shocks and apply KPSW's identification scheme. Two important outcomes are:

1. The three-dimensional impulse-response function corresponding to output, investment and consumption, implicit in our estimated factor model, is non fundamental, an important difference with respect to the VAR estimated in KPSW.
2. Comparing responses of the permanent shock in KPSW and the factor model, we find that long-run effects are much more important in the second. In particular, the long-run response of investment in the factor model is almost two times the one estimated in KPSW.

## Appendix

The following statement is proved below.

### Proposition P.

- (A)  $\|\hat{Q}_m - Q_m \hat{\mathcal{J}}_r\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right)$ ,  
 (B)  $\|\hat{D}^k - \hat{\mathcal{J}}_r D^k \hat{\mathcal{J}}_r\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ , for all  $k \geq 0$ ,  
 (C)  $\|\hat{\mathcal{K}} - \hat{\mathcal{J}}_r \mathcal{K} \hat{\mathcal{J}}_q\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ ,

where  $\hat{\mathcal{J}}_r$  and  $\hat{\mathcal{J}}_q$  are diagonal matrices,  $r \times r$  and  $q \times q$  respectively, depending on  $n$  and  $T$ , whose diagonal entries are equal either to 1 or  $-1$ .

Roughly speaking, Proposition P states that  $\hat{Q}_m$ ,  $\hat{D}^k$ ,  $\hat{\mathcal{K}}$  approximate  $Q_m$ ,  $D^k$ ,  $\mathcal{K}$  respectively, the reason for the presence of  $\hat{\mathcal{J}}_r$  and  $\hat{\mathcal{J}}_q$  being that we do not want to establish a rule to decide the sign of the eigenvectors of  $\hat{\Gamma}_0^x$  and  $\hat{\Gamma}^\epsilon$ . However,

$$\begin{aligned} \chi_{mt} = C_m(L)\mathbf{v}_t &= [Q_m \hat{\mathcal{J}}_r] \left( I - [\hat{\mathcal{J}}_r D \hat{\mathcal{J}}_r] L \right)^{-1} [\hat{\mathcal{J}}_r \mathcal{K} \hat{\mathcal{J}}_q] [\hat{\mathcal{J}}_q \mathbf{v}_t] \\ &= \check{Q}_m (I - \check{D}L)^{-1} \check{\mathcal{K}} \check{\mathbf{v}}_t = \check{C}_m(L) \check{\mathbf{v}}_t, \end{aligned}$$

which is obviously a fundamental representation. As a consequence, setting  $\check{H} = F(\check{Q}_m, \check{D}, \check{\mathcal{K}})$ , we have

$$B_m(L) = C_m(L)H' = \check{C}_m(L)\check{H}', \quad (6.24)$$

(see (ii) at the end of Section 4.1). On the other hand, Proposition P implies that

$$\|\hat{H} - \check{H}\| = \|F(\hat{Q}_m, \hat{D}, \hat{\mathcal{K}}) - F(\check{Q}_m, \check{D}, \check{\mathcal{K}})\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right),$$

this being a standard result under reasonable regularity assumptions for  $F$  (the usual identification schemes, with zero first-impact or long-run restrictions, produce functions  $F$  with elementary analytic entries). This result, combined with (6.24), implies Proposition 3.

To prove Proposition P we need an additional technical assumption.

**Assumption 9.** Denote by  $\gamma_{ij,k}^x$  and  $\hat{\gamma}_{ij,k}^x$  the entries of  $\Gamma_k^x, \hat{\Gamma}_k^x$  respectively. We require that  $\sqrt{T}\|\hat{\gamma}_k^x - \gamma_k^x\| = O_p(1)$  uniformly in  $n$ , i.e. that given  $\eta > 0$  there exists  $\delta(\eta)$  such that

$$P(\sqrt{T}|\gamma_{ij,k}^x - \hat{\gamma}_{ij,k}^x| > \delta(\eta)) < \eta,$$

for all  $i \in \mathbb{N}$  and  $j \in \mathbb{N}, k = 0, 1$ .

Moreover, we will make use of the following inequality, which is due to H. Weyl. If  $A$  is a symmetric matrix we denote by  $\mu_j(A)$  the  $j$ -th eigenvalue of  $A$  in decreasing order. Given a matrix  $B$ ,  $\|B\|$  denotes the spectral norm of  $B$ , thus  $\|B\| = \sqrt{\mu_1(BB')}$ , which is the euclidean norm if  $B$  is a row matrix. Let  $A$  and  $B$  be two  $s \times s$  symmetric matrices. Then (see e.g. Stewart and Sun, 1990, p. 203, Corollary 4.10):

$$|\mu_j(A + B) - \mu_j(A)| \leq \sqrt{\mu_1(B^2)} = \|B\|, \quad j = 1, \dots, s. \quad (6.25)$$

**Lemma 1.** Denoting by  $\mathcal{I}_m$  the  $n \times m$  matrix having the identity matrix  $I_m$  in the first  $m$  rows and 0 elsewhere (see Section 4.1),

- (i)  $\frac{1}{n}\|\hat{\Gamma}_k^x - \Gamma_k^x\| = O_p\left(\frac{1}{\sqrt{T}}\right), \quad k = 0, 1.$
- (ii)  $\frac{1}{\sqrt{n}}\|\mathcal{I}_m'(\hat{\Gamma}_0^x - \Gamma_0^x)\| = O_p\left(\frac{1}{\sqrt{T}}\right)$  for any (fixed)  $m$ .
- (iii)  $\frac{1}{n}\|\hat{\Gamma}_k^x - \Gamma_k^x\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right), \quad k = 0, 1.$
- (iv)  $\frac{1}{\sqrt{n}}\|\mathcal{I}_m'(\hat{\Gamma}_0^x - \Gamma_0^x)\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right)$  for any (fixed)  $m$ .

*Proof.* We have

$$\mu_1\left((\hat{\Gamma}_k^x - \Gamma_k^x)(\hat{\Gamma}_k^x - \Gamma_k^x)'\right) \leq \text{trace}\left((\hat{\Gamma}_k^x - \Gamma_k^x)(\hat{\Gamma}_k^x - \Gamma_k^x)'\right) = \sum_{i=1}^n \sum_{j=1}^n (\hat{\gamma}_{kij}^x - \gamma_{kij}^x)^2 = O_p\left(\frac{n^2}{T}\right),$$

where the last equality follows from Assumption 9. This proves (i). Similarly, we have

$$\text{trace}\left(\mathcal{I}_m'(\hat{\Gamma}_0^x - \Gamma_0^x)^2 \mathcal{I}_m\right) = \sum_{i=1}^m \sum_{j=1}^n (\hat{\gamma}_{0ij}^x - \gamma_{0ij}^x)^2 = O_p\left(\frac{n}{T}\right).$$

Statement (ii) follows. As for (iii), observe that  $\hat{\Gamma}_k^x - \Gamma_k^\chi = \hat{\Gamma}_k^x - \Gamma_k^x + \Gamma_k^\xi$  by Assumption 3, so that  $\frac{1}{n}\|\hat{\Gamma}_k^x - \Gamma_k^\chi\| \leq \frac{1}{n}\|\hat{\Gamma}_k^x - \Gamma_k^x\| + \frac{1}{n}\|\Gamma_k^\xi\|$ . The first term on the RHS is  $O_p\left(\frac{1}{\sqrt{T}}\right)$  by statement (i), whereas the second is bounded by  $\frac{1}{n}\mu_1^\xi$ , which is  $O\left(\frac{1}{n}\right)$  by Assumption 5. Statement (iv) is obtained in a similar way, using (ii) instead of (i) and the upper bound  $\frac{1}{\sqrt{n}}\mu_1^\xi$  instead of  $\frac{1}{n}\mu_1^\xi$ . *Q.E.D.*

**Lemma 2.**

(i)  $\frac{\hat{\mu}_j^x}{n} - \frac{\mu_j^\chi}{n} = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$  for any  $j$ .

(ii) There exists  $\bar{n}$  such that, for all  $n \geq \bar{n}$ ,  $\left(\frac{M^x}{n}\right)$  is invertible.

(iii) For any  $n \geq \bar{n}$  and  $\eta > 0$ , there exists  $\tau(\eta, n)$  such that, for  $T \geq \tau(\eta, n)$ ,  $\left(\frac{\hat{M}^x}{n}\right)$  is invertible with probability larger than  $1 - \eta$ ; moreover, if  $\left(\frac{\hat{M}^x}{n}\right)^{-1}$  exists for  $n = n^*$  and  $T = T^*$ , it exists for all  $n > n^*$  and  $T > T^*$ .

(iv)  $\left\|\frac{M^x}{n}\right\|$  and  $\left\|\left(\frac{M^x}{n}\right)^{-1}\right\|$ , which depend on  $n$ , are  $O(1)$ ;  $\left\|\frac{\hat{M}^x}{n}\right\|$  and  $\left\|\left(\frac{\hat{M}^x}{n}\right)^{-1}\right\|$ , which depend on  $n$  and  $T$ , are  $O_p(1)$ .

*Proof.* Setting  $A = \Gamma_0^\chi$ ,  $B = \hat{\Gamma}_0^x - \Gamma_0^\chi$  and applying (6.25) we get  $\frac{1}{n}|\hat{\mu}_j^x - \mu_j^\chi| \leq \frac{1}{n}\|\hat{\Gamma}_0^x - \Gamma_0^\chi\|$ , which is  $O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$  by Lemma 1 (iii). As for (ii), by Assumption 4 (b) there exists  $\bar{n}$  such that, for  $n \geq \bar{n}$ ,  $\frac{\mu_r^\chi}{n} > \underline{c}_r > 0$ , so that  $\det\left(\frac{M^x}{n}\right) \neq 0$ . Turning to (iii), setting  $A = \Gamma_0^x$ ,  $B = \hat{\Gamma}_0^x - \Gamma_0^x$  and applying Weyl inequality we get  $\frac{1}{n}|\hat{\mu}_r^x - \mu_r^x| \leq \frac{1}{n}\|\hat{\Gamma}_0^x - \Gamma_0^x\|$ , which is  $O_p\left(\frac{1}{\sqrt{T}}\right)$  by Lemma 1 (i). Now,  $\mu_r^x \geq \mu_r^\chi$ , since  $\Gamma_0^\xi$  is positive semi-definite, so that, for  $n \geq \bar{n}$ ,  $\frac{\mu_r^x}{n} > \underline{c}_r > 0$ . Hence  $\det\left(\frac{\hat{M}^x}{n}\right)$  is bounded away from zero in probability as  $T \rightarrow \infty$ . The last part of statement (iii) follows from the fact that the rank of the observation matrix  $\mathbf{x}_n^T$ , and therefore the rank of  $\hat{\Gamma}_0^x$ , is non-decreasing in  $n$  and  $T$ . As for (iv), observe that  $\left\|\frac{M^x}{n}\right\| = \frac{\mu_1^\chi}{n}$  and  $\left\|\left(\frac{M^x}{n}\right)^{-1}\right\| = \frac{n}{\mu_r^\chi}$ , which are asymptotically bounded by  $\underline{c}_1$  and  $\frac{1}{\underline{c}_r}$  by Assumption 4 (b). Boundedness in probability of  $\left\|\frac{\hat{M}^x}{n}\right\|$  and  $\left\|\left(\frac{\hat{M}^x}{n}\right)^{-1}\right\|$  then follow from statement (i). *Q.E.D.*

**Lemma 3.**

(i)  $\left\|W^{\chi'}\hat{W}^x\frac{\hat{M}^x}{n} - \frac{M}{n}W^{\chi'}\hat{W}^x\right\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ .

$$(ii) \|\hat{W}^{x'}W^\chi W^{\chi'}\hat{W}^x - I_r\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right).$$

(iii) There exist diagonal  $r \times r$  matrices  $\hat{\mathcal{J}}_r$ , depending on  $n$  and  $T$ , whose diagonal entries are equal to either 1 or  $-1$ , such that  $\|\hat{W}^{x'}W^\chi - \hat{\mathcal{J}}_r\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ .

*Proof.* We have  $\|W^{\chi'}\hat{W}^x \frac{\hat{M}^x}{n} - \frac{M^x}{n}W^{\chi'}\hat{W}^x\| = \|\frac{1}{n}W^{\chi'}(\hat{\Gamma}_0^x - \Gamma_0^x)\hat{W}^x\| \leq \frac{1}{n}\|\hat{\Gamma}_0^x - \Gamma_0^x\|$ . Statement (i) then follows from Lemma 1 (iii). As for (ii), set

$$\begin{aligned} a &= \hat{W}^{x'}W^\chi W^{\chi'}\hat{W}^x = \hat{W}^{x'}W^\chi W^{\chi'}\hat{W}^x \frac{\hat{M}^x}{n} \left(\frac{\hat{M}^x}{n}\right)^{-1}, \\ b &= \hat{W}^{x'}W^\chi \frac{M^x}{n} W^{\chi'}\hat{W}^x \left(\frac{\hat{M}^x}{n}\right)^{-1} = \frac{1}{n}\hat{W}^{x'}\Gamma_0^x \hat{W}^x \left(\frac{\hat{M}^x}{n}\right)^{-1}, \\ c &= \frac{1}{n}\hat{W}^{x'}\hat{\Gamma}_0^x \hat{W}^x \left(\frac{\hat{M}^x}{n}\right)^{-1} = \frac{\hat{M}^x}{n} \left(\frac{\hat{M}^x}{n}\right)^{-1} = I_r. \end{aligned}$$

We have  $\|a - c\| \leq \|a - b\| + \|b - c\|$ . Both terms are  $O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ , the first by statement (i), the second by Lemma 1 (iii). Turning to (iii), let us denote by  $\hat{w}_j^x$  and  $w_j^x$  the  $j$ -th column of  $\hat{W}^x$  and  $W^x$  respectively. By taking a single entry of the matrix on the LHS of statement (i) we get

$$\frac{1}{n}(\hat{\mu}_j^x - \mu_i^x)w_j^{\chi'}\hat{w}_i^x = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right),$$

$i \leq r, j \leq r$ . Now, for  $j \neq i$ ,  $\frac{1}{n}(\hat{\mu}_j^x - \mu_i^x)$  is bounded away from zero in probability, since  $\mu_i^x$  and  $\mu_j^x$  are asymptotically distinct by Assumption 4 (b), while  $\hat{\mu}_j^x$  tends to  $\mu_j^x$  in probability by Lemma 2 (i). Hence the off-diagonal terms of  $\hat{W}^{x'}W^\chi$  are  $O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ . Turning to the diagonal terms, let us first observe that  $\hat{w}_i^{x'}W^\chi W^{\chi'}\hat{w}_i^x = 1 + O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$  by statement (ii). But

$$\hat{w}_i^{x'}W^\chi W^{\chi'}\hat{w}_i^x = (\hat{w}_i^{x'}w_i^x)^2 + \sum_{\substack{j=1 \\ j \neq i}}^r (\hat{w}_i^{x'}w_j^x)^2 = (\hat{w}_i^{x'}w_i^x)^2 + O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right).$$

Hence  $(1 - |\hat{w}_i^{x'}w_i^x|)(1 + |\hat{w}_i^{x'}w_i^x|) = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ , so that  $1 - |\hat{w}_i^{x'}w_i^x| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ . *Q.E.D.*

**Proof of Proposition P (A).** Set

$$\begin{aligned} a &= Q_m \hat{\mathcal{J}}_r = \sqrt{n}\mathcal{I}'_m W^\chi \hat{\mathcal{J}}_r, \text{ where } \hat{\mathcal{J}}_r \text{ has been defined in Lemma 3 (iii),} \\ b &= \sqrt{n}\mathcal{I}'_m W^\chi W^{\chi'}\hat{W}^x = \sqrt{n}\mathcal{I}'_m W^\chi W^{\chi'}\hat{W}^x \frac{\hat{M}^x}{n} \left(\frac{\hat{M}^x}{n}\right)^{-1}, \\ c &= \sqrt{n}\mathcal{I}'_m W^\chi \frac{M^x}{n} W^{\chi'}\hat{W}^x \left(\frac{\hat{M}^x}{n}\right)^{-1} = \frac{1}{\sqrt{n}}\mathcal{I}'_m \Gamma_0^x \hat{W}^x \left(\frac{\hat{M}^x}{n}\right)^{-1}, \end{aligned}$$

$$d = \frac{1}{\sqrt{n}} \mathcal{I}'_m \hat{\Gamma}_0^x \hat{W}^x \left( \frac{\hat{M}^x}{n} \right)^{-1} = \sqrt{n} \mathcal{I}'_m \hat{W}^x = \hat{Q}_m.$$

Firstly observe that  $\|\sqrt{n} \mathcal{I}'_m W^\chi\| = \|\mathcal{I}'_m A_n \Gamma_0^f \left( \frac{M^\chi}{n} \right)^{-1}\|$  is  $O(1)$ , since  $\|\mathcal{I}'_m A_n \Gamma_0^f\|$  is  $O(1)$  and  $\left\| \left( \frac{M^\chi}{n} \right)^{-1} \right\|$  is  $O(1)$  by Lemma 2 (iv). Hence we can apply Lemma 3 (iii) to get  $\|a - b\| = O_p \left( \max \left( \frac{1}{n}, \frac{1}{\sqrt{T}} \right) \right)$ , and Lemma 3 (i) to get  $\|b - c\| = O_p \left( \max \left( \frac{1}{n}, \frac{1}{\sqrt{T}} \right) \right)$ . Finally, Lemma 1 (iv) ensures that  $\|c - d\| = O_p \left( \max \left( \frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}} \right) \right)$ .  
*Q.E.D.*

**Proof of Proposition P (B).** We have  $\hat{D} = \frac{1}{n} \hat{W}^{x'} \hat{\Gamma}_1^x \hat{W}^x \left( \frac{\hat{M}^x}{n} \right)^{-1}$  and  $\hat{\mathcal{J}}_r D \hat{\mathcal{J}}_r = \frac{1}{n} \hat{\mathcal{J}}_r W^{x'} \Gamma_1^x W^\chi \left( \frac{M^\chi}{n} \right)^{-1} \hat{\mathcal{J}}_r = \frac{1}{n} \hat{\mathcal{J}}_r W^{x'} \Gamma_1^x W^\chi \hat{\mathcal{J}}_r \left( \frac{M^\chi}{n} \right)^{-1}$ , where  $\hat{\mathcal{J}}_r$  has been defined in Lemma 3 (iii). Set

$$\begin{aligned} a &= \hat{D} = \frac{1}{n} \hat{W}^{x'} \hat{\Gamma}_1^x \hat{W}^x \left( \frac{\hat{M}^x}{n} \right)^{-1}, \\ b &= \frac{1}{n} \hat{W}^{x'} \Gamma_1^x \hat{W}^x \left( \frac{\hat{M}^x}{n} \right)^{-1} = \frac{1}{n} \hat{W}^{x'} W^\chi W^{x'} \Gamma_1^x W^\chi W^{x'} \hat{W}^x \left( \frac{\hat{M}^x}{n} \right)^{-1}, \\ c &= \frac{1}{n} \hat{\mathcal{J}}_r W^{x'} \Gamma_1^x W^\chi \hat{\mathcal{J}}_r \left( \frac{\hat{M}^x}{n} \right)^{-1}, \\ d &= \hat{\mathcal{J}}_r D \hat{\mathcal{J}}_r = \frac{1}{n} \hat{\mathcal{J}}_r W^{x'} \Gamma_1^x W^\chi \left( \frac{M^\chi}{n} \right)^{-1} \hat{\mathcal{J}}_r = \frac{1}{n} \hat{\mathcal{J}}_r W^{x'} \Gamma_1^x W^\chi \hat{\mathcal{J}}_r \left( \frac{M^\chi}{n} \right)^{-1}. \end{aligned}$$

By Lemma 1 (i)  $\|a - b\|$  is  $O_p \left( \frac{1}{\sqrt{T}} \right)$ ; by Lemma 3 (iii)  $\|b - c\|$  is  $O_p \left( \max \left( \frac{1}{n}, \frac{1}{\sqrt{T}} \right) \right)$ ; by Lemma 2 (i)  $\|c - d\|$  is  $O_p \left( \max \left( \frac{1}{n}, \frac{1}{\sqrt{T}} \right) \right)$ . This proves the statement for  $k = 1$ . Observing that  $\hat{\mathcal{J}}_r^2 = I_r$ , the extension to the case  $k > 1$  is straightforward.  
*Q.E.D.*

**Lemma 4.**

- (i)  $\|\hat{\Gamma}^\epsilon - \hat{\mathcal{J}}_r \Gamma^\epsilon \hat{\mathcal{J}}_r\| = O_p \left( \max \left( \frac{1}{n}, \frac{1}{\sqrt{T}} \right) \right)$ , where  $\hat{\mathcal{J}}_r$  has been defined in Lemma 3 (iii).
- (ii)  $\hat{\mu}_j^\epsilon - \mu_j^\epsilon = O_p \left( \max \left( \frac{1}{n}, \frac{1}{\sqrt{T}} \right) \right)$   $j = 1, \dots, r$ .
- (iii)  $\|\hat{\mathcal{M}} - \mathcal{M}\| = O_p \left( \max \left( \frac{1}{n}, \frac{1}{\sqrt{T}} \right) \right)$ .
- (iv)  $\mathcal{M}^{-1}$  exists for  $n$  sufficiently large and its norm is  $O(1)$  as  $n \rightarrow \infty$ ; moreover,  $\|\hat{\mathcal{M}} \mathcal{M}^{-1} - I_q\| = O_p \left( \max \left( \frac{1}{n}, \frac{1}{\sqrt{T}} \right) \right)$ .
- (v) There exist diagonal  $q \times q$  matrices  $\hat{\mathcal{J}}_q$ , depending on  $n$  and  $T$ , whose diagonal entries are either equal to 1 or  $-1$ , such that  $\|K' \hat{\mathcal{J}}_r \hat{K} - \hat{\mathcal{J}}_q\| = O_p \left( \max \left( \frac{1}{n}, \frac{1}{\sqrt{T}} \right) \right)$ .

*Proof.* We have  $\hat{\Gamma}^\epsilon = \frac{\hat{M}^x}{n} - \hat{D} \frac{\hat{M}^x}{n} \hat{D}'$  and  $\hat{\mathcal{J}}_r \Gamma^\epsilon \hat{\mathcal{J}}_r = \hat{\mathcal{J}}_r \left( \frac{M^\chi}{n} - D \frac{M^\chi}{n} D' \right) \hat{\mathcal{J}}_r = \frac{M^\chi}{n} - \hat{\mathcal{J}}_r D \hat{\mathcal{J}}_r \frac{M^\chi}{n} \hat{\mathcal{J}}_r D' \hat{\mathcal{J}}_r$ . Statement (i) then follows from Lemma 1 (i) and Proposition

P (B). As for (ii), notice first that the eigenvalues of  $\hat{\mathcal{J}}_r \Gamma^\epsilon \hat{\mathcal{J}}_r$  are identical to those of  $\Gamma^\epsilon$ . Hence setting  $A = \Gamma^\epsilon$ ,  $B = \hat{\Gamma}^\epsilon - \hat{\mathcal{J}}_r \Gamma^\epsilon \hat{\mathcal{J}}_r$  and applying (6.25) we get  $|\hat{\mu}_j^\epsilon - \mu_j^\epsilon| \leq \|\hat{\Gamma}^\epsilon - \hat{\mathcal{J}}_r \Gamma^\epsilon \hat{\mathcal{J}}_r\|$ . Statement (ii) then follows from (i). Turning to (iii), we have  $\hat{\mathcal{M}}^2 - \mathcal{M}^2 = (\hat{\mathcal{M}} - \mathcal{M})(\hat{\mathcal{M}} + \mathcal{M})$ . As the second factor is asymptotically bounded away from zero by Assumption 8, the result follows from statement (ii). Statement (iv) follows from the fact that  $\mu_q^\epsilon > \underline{d}_q > 0$  for  $n$  sufficiently large by Assumption 8 and statement (iii). Finally, result (v) is obtained following the lines of Lemma 3, with Assumption 8 ensuring asymptotically distinct eigenvalues instead of Assumption 4 (b). *Q.E.D.*

**Proof of Proposition P (C).** Let us denote by  $\hat{\mathcal{N}}$  the diagonal matrix having on the diagonal the smallest  $r - q$  eigenvalues of  $\hat{\Gamma}^\epsilon$  and by  $\hat{K}_\perp$  the  $r \times (r - q)$  matrix having on the columns the corresponding eigenvectors, so that  $\hat{\Gamma}^\epsilon = \hat{K} \hat{\mathcal{M}}^2 \hat{K}' + \hat{K}_\perp \hat{\mathcal{N}} \hat{K}_\perp'$ . As  $\mu_j^\epsilon = 0$  for  $j > q$  by Lemma 4 (ii), the second term is  $O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ . Hence by Lemma 4 (i),  $\|\hat{K} \hat{\mathcal{M}}^2 \hat{K}' - \hat{\mathcal{J}}_r K \mathcal{M}^2 K' \hat{\mathcal{J}}_r\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ , where  $\hat{\mathcal{J}}_r$  has been defined in Lemma 3 (iii). Postmultiplying by  $\hat{K} \mathcal{M}^{-1}$ , which is  $O(1)$  by Lemma 4 (iv), we get

$$\|\hat{K} \hat{\mathcal{M}}^2 \mathcal{M}^{-1} - \hat{\mathcal{J}}_r K \mathcal{M}^2 K' \hat{\mathcal{J}}_r \hat{K} \mathcal{M}^{-1}\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right).$$

The desired result is obtained by applying Lemma 4 (iv) and Lemma 4 (v) and observing that  $\hat{\mathcal{J}}_q \mathcal{M}^{-1} = \mathcal{M}^{-1} \hat{\mathcal{J}}_q$ , where  $\hat{\mathcal{J}}_q$  has been defined in Lemma 4 (v). *Q.E.D.*

## References

- [1] Altug, S. (1989). Time-to-Build and Aggregate Fluctuations: Some New Evidence, *International Economic Review* **30**, 889-920.
- [2] Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions, *Econometrica* **71**, 135-171.
- [3] Bai, J., and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191-221.
- [4] Bai, J. and S. Ng (2007). Determining the Number of Primitive Shocks in Factor Models, *Journal of Business & Economic Statistics* **25**, 52-60.
- [5] Bernanke, B.S., and J. Boivin (2003). Monetary Policy in a Data Rich environment, *Journal of Monetary Economics* **50**, 525-46.
- [6] Bernanke, B.S., J. Boivin and P. Elias (2005). Measuring Monetary Policy: A Factor Augmented Autoregressive (FAVAR) Approach, *Quarterly Journal of Economics* **120**, 387-422.
- [7] Boivin, J. and S. Ng (2003). Are more data always better for factor analysis?, *Journal of Econometrics* **127**, 169-94.
- [8] Boivin, J. and M.P. Giannoni, 2006. Has Monetary Policy Become More Effective?, *The Review of Economics and Statistics* **88**, 445-62.
- [9] Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica* **51**, 1281-1304.
- [10] Chamberlain, G., and M. Rothschild (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets, *Econometrica* **51**, 1305-1324.
- [11] 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.
- [12] 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**, 255-89.
- [13] Fernández-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.
- [14] 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**, 540-54.
- [15] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized factor model: one-sided estimation and forecasting, *Journal of the American Statistical Association* **100**, 830-40.

- [16] Forni, M., and M. Lippi (2001). The generalized dynamic factor model: representation theory, *Econometric Theory* **17**, 1113-41.
- [17] Forni, M. and L. Reichlin (1998). Let's get real: a factor analytical approach to disaggregated business cycle dynamics, *Review of Economic Studies* **65**, 453-73.
- [18] 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*. Amsterdam: North Holland.
- [19] Giannone, D., L. Reichlin and L. Sala (2002). Tracking Greenspan: Systematic and Nonsystematic Monetary Policy Revisited, CEPR Discussion Paper no. 3550.
- [20] Giannone, D., L. Reichlin and L. Sala (2005). Monetary Policy in Real Time, in M. Gertler and K. Rogoff, Eds., *NBER Macroeconomic Annual, 2004*, MIT Press.
- [21] Giannone, D., L. Reichlin and L. Sala (2006). VARs, common factors and the empirical validation of equilibrium business cycle models, *Journal of Econometrics* **127**, 257-79.
- [22] Hallin, M. and R. Liška (2007). The generalized dynamic factor model: determining the number of factors, *Journal of the American Statistical Association* **102**, 103-17.
- [23] Hannan, E.J. and M. Deistler (1988). *The Statistical Theory of Linear Systems*. New York: Wiley.
- [24] Hansen, L.P. and T.J. Sargent (1980). Formulating and estimating dynamic linear rational expectations models, *Journal of Economic Dynamics and Control* **2**, 7-46.
- [25] 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, 77-119.
- [26] Ireland, P.N. (2004). A method for taking models to the data, *Journal of Economic Dynamics and Control* **28**, 1205-26.
- [27] King, R.G., C.I. Plosser, J.H. Stock and M.W. Watson (1991). Stochastic Trends and Economic Fluctuations, *American Economic Review* **81**, 819-40.
- [28] Lippi, M., and L. Reichlin (1993). The dynamic effects of aggregate demand and supply disturbances: Comment, *American Economic Review* **83**, 644-52.
- [29] Lippi, M., and L. Reichlin (1994). VAR analysis, non fundamental representation, Blaschke matrices, *Journal of Econometrics* **63**, 307-25.
- [30] Rozanov, Yu. (1967). *Stationary Random processes*. San Francisco: Holden Day.
- [31] Sargent, T.J. (1989). Two Models of Measurements and the Investment Accelerator, *The Journal of Political Economy* **97**, 251-87.

- [32] 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*, Minneapolis: Federal Reserve Bank of Minneapolis.
- [33] Stewart, G.W., and J. Sun (1990). *Matrix Perturbation Theory*. San Diego: Academic Press.
- [34] Stock, J.H., and M.W. Watson (2002a). Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics* **20**, 147-62.
- [35] 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**, 116779.
- [36] Stock, J.H., and M.W. Watson (2005). Implications of Dynamic Factor Models for VAR Analysis, NBER Working Papers no. 11467.
- [37] van der Waerden, B.L. (1953). *Modern Algebra*, Vol. I. New York: Frederick Ungar.

## RECent Working Papers Series

The most RECent releases are:

- No. 8 OPENING THE BLACK BOX: STRUCTURAL FACTOR MODELS WITH LARGE CROSS-SECTIONS (2007)  
M. Forni, D. Giannone, M. Lippi and L. Reichlin
- No. 7 DYNAMIC FACTOR ANALYSIS OF INDUSTRY SECTOR DEFAULT RATES AND IMPLICATION FOR PORTFOLIO CREDIT RISK MODELLING (2007)  
A. Cipollini and G. Missaglia
- No. 6 LEADING INDICATOR PROPERTIES OF US HIGH-YIELD CREDIT SPREADS (2007)  
A. Cipollini and N. Aslanidis
- No. 5 THE VANISHING BEQUEST TAX. THE COMPARATIVE EVOLUTION OF BEQUEST TAXATION IN HISTORICAL PERSPECTIVE (2007)  
G. Bertocchi
- No. 4 THE DISTRIBUTION OF THE GENDER WAGE GAP IN ITALY: DOES EDUCATION MATTER? (2007)  
T. Addabbo, D. Favaro and S. Magrini
- No. 3 MIGRANT NETWORKS: EMPIRICAL IMPLICATIONS FOR THE ITALIAN BILATERAL TRADE (2007)  
M. Murat and B. Pistoresi
- No. 2 MIGRANT BUSINESS NETWORKS AND FDI (2007)  
M. Murat and S. Flisi
- No. 1 THE MAXIMUM LQ-LIKELIHOOD METHOD: AN APPLICATION TO EXTREME QUANTILE ESTIMATION IN FINANCE (2007)  
D. Ferrari and S. Paterlini

The full list of available working papers, together with their electronic versions, can be found on the RECent website: <http://www.recent.unimore.it/workingpapers.asp>