Econometric Theory, **25**, 2009, 1319–1347. Printed in the United States of America.
doi:10.1017/S026646660809052X

OPENING THE BLACK BOX: STRUCTURAL FACTOR MODELS WITH LARGE CROSS SECTIONS

MARIO FORNI
Università di Modena e Reggio Emilia
and
CEPR

DOMENICO GIANNONE
ECARES, Université Libre de Bruxelles

MARCO LIPPI
Università di Roma "La Sapienza"

LUCREZIA REICHLIN
European Central Bank
ECARES, Université Libre de Bruxelles
and
CEPR

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

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

We thank, for suggestions and criticism, Manfred Deistler, Marc Hallin, two anonymous referees, and Pentti Saikkonen, co-editor of *Econometric Theory*. M. Forni and M. Lippi are grateful to the 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. Address correspondence to Marco Lippi, Dipartimento di Economia, Via Cesalpino 12, I-00161 Roma, Italy; e-mail: ml@lippi.ws.

 0266-4666/09 $15.00

that each time series in the data set can be expressed as the sum of two orthogonal components: the "common component," capturing that part of the series that comoves 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 data sets points to the robust finding that few shocks explain the bulk of dynamics of macro data (see Sargent and Sims, 1977; Giannone, Reichlin, and Sala, 2002; Giannone et al., 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, although 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 as 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 macroeconomic 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, and the literature mentioned therein; 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 factor augmented autoregression (FAVAR) models, in which a vector autoregression (VAR) is augmented with common factors (see Bernanke, Boivin, and Eliasz, 2005). The link between factor models, FAVAR models, and VAR models has been studied by 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 the discussion that follows 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 data set (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 data set, 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 $\boldsymbol{\chi}_{nt}$ 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, 1994; and, more recently, Chari, Kehoe, and McGrattan, 2005; Fernández-Villaverde, Rubio-Ramirez, and Sargent, 2005; Giannone et al., 2006). Nonfundamentalness 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; Giannone et al., 2002, 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. As 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. Last, we can go back to the variables of interest and study their dynamic response to structural macroeconomic shocks. Section 5 analyzes an empirical example on U.S. macroeconomic

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

data that 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 a 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. $\boldsymbol{x}_n^T$ is a finite realization of a real-valued stochastic process

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

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

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

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

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

$$\begin{aligned} \boldsymbol{x}_{nt} &= \boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt}, \\ \boldsymbol{\chi}_{nt} &= B_n(L)\boldsymbol{u}_t, \end{aligned} \tag{1}$$

where the following conditions hold.

Assumption 2. $\boldsymbol{u}_t$ is a q-dimensional orthonormal white noise, and $B_n(L)$ 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

$$B_n(L) = A_n N(L). \tag{2}$$

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

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

$$\boldsymbol{f}_t = N(L)\boldsymbol{u}_t, \tag{3}$$

(1) can be rewritten in the static form

$$\boldsymbol{x}_{nt} = A_n \boldsymbol{f}_t + \boldsymbol{\xi}_{nt}. \tag{4}$$

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

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

Obviously $\boldsymbol{x}_{nt}$ admits infinitely many different representations of the forms (1) and (4), with different dynamic and static factors. In particular, if H is an orthogonal $q \times q$ matrix, then $\boldsymbol{\chi}_{nt} = C_n(L)\boldsymbol{v}_t$, with $\boldsymbol{v}_t = H\boldsymbol{u}_t$, $C_n(L) = B_n(L)H'$ (the same applies to the static factors with H replaced by any invertible $r \times r$ matrix). In Sections 3 and 4.1 we discuss identification of $\boldsymbol{u}_t$, i.e., the conditions under which $B_n(L)$ and $\boldsymbol{u}_t$ can be determined among all alternative impulse-response functions and dynamic factors.

Assumption 3 (Orthogonality of common and idiosyncratic components). For all n, the vector $\boldsymbol{\xi}_{nt}$ is stationary. Moreover, $\boldsymbol{u}_t$ 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 (for a discussion, see Sargent, 1989). 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 χ's and on the variables x_{it} themselves.

Some definitions are needed for the next two assumptions. Let Γ_k^χ be the k-lag covariance matrix of $\boldsymbol{\chi}_{nt}$ and denote by μ_j^χ the jth eigenvalue, in decreasing order, of Γ_0^χ. 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 Γ_k^χ, μ_j^χ, etc., just defined, and of other scalars and matrices defined subsequently, 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 Γ_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 $n \to \infty$ we have $\lambda_q^{\chi}(\theta) \to \infty$ for θ almost everywhere (a.e.) in $[-\pi\ \pi]$.
(b) There exist constants $\underline{c}_j, \overline{c}_j,\ j = 1, \ldots, r$, such that $\underline{c}_j > \overline{c}_{j+1},\ j = 1, \ldots, r-1$, and

$$0 < \underline{c}_j < \liminf_{n\to\infty} n^{-1}\mu_j^{\chi} \le \limsup_{n\to\infty} n^{-1}\mu_j^{\chi} \le \overline{c}_j.$$

Assumption 5 (Nonpervasiveness of the idiosyncratic components). There exists a real $\mathcal{L}$ such that $\lambda_1^{\xi}(\theta) \le \mathcal{L}$ for any $n \in \mathbb{N}$ and θ a.e. in $[-\pi\ \pi]$. This obviously implies that $\mu_1^{\xi} \le \mathcal{L}$ for any $n \in \mathbb{N}$, μ_j^{ξ} being the jth eigenvalue of Γ_0^{ξ}.

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 nondeclining coefficients.

Some comments on our assumptions are in order.

1. Assumption 4(a) implies that the number q of dynamic factors and the common components χ_{it} are unique; i.e., a representation of the form (1)–(4) with a different number of dynamic factors or different common components is not possible (see Forni and Lippi, 2001).
2. Assumption 4(b) implies that the number r of static factors is unique; i.e., a static representation of the common components χ_{it} with a different number of static factors is not possible.
3. We define the static and dynamic rank of $\boldsymbol{f}_t$ as the rank of, respectively, its variance-covariance and spectral density matrix. By Assumption 4(a) the dynamic rank of $\boldsymbol{f}_t$ is q for θ a.e. in $[-\pi\ \pi]$. Assumption 4(b) entails that, for n sufficiently large, A_n has full rank r and that $\boldsymbol{f}_t$ has static rank r for any given t. Thus, for any given t, the space spanned by χ_{it}, $i \in \mathbb{N}$, coincides with the space spanned by the static factors $f_{jt},\ j = 1, \ldots, r$, 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 et al., 2005):

$$\boldsymbol{\chi}_{nt} = C_{n0}\boldsymbol{f}_t^* + C_{n1}\boldsymbol{f}_{t-1}^* + \cdots + C_{ns}\boldsymbol{f}_{t-s}^*, \tag{5}$$

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

$$\Theta(L)\boldsymbol{f}_t^* = (1 - \Theta_1 L - \cdots - \Theta_h L^h)\boldsymbol{f}_t^* = \boldsymbol{u}_t, \tag{6}$$

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

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

$$\boldsymbol{f}_t = \left(\boldsymbol{f}_t^{*'} \boldsymbol{f}_{t-1}^{*'} \dots \boldsymbol{f}_{t-s}^{*'}\right)', \qquad A_n = (C_{n0}\, C_{n1} \dots C_{ns}),$$

$$N(L) = (K(L) K(L) L \dots K(L) L^s)',$$

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

$$\boldsymbol{x}_{nt} = A_n \boldsymbol{f}_t + \boldsymbol{\xi}_{nt}. \tag{7}$$

The static rank of $\boldsymbol{f}_t$ is always $q(s+1)$. However, for (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}:

1. If no restrictions among the coefficients of the matrices C_{nj} hold (assume, e.g., that they are independently drawn from the same distribution), then (7) is a static representation of the model.
2. 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 (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 (7) is a static representation. But if the restriction $c_{i1} = a c_{i0}$ holds, then $r = 1$, $N(L) = 1 + aL$, $f_t = (1 + aL) u_t$, and $A_n = (c_{10} c_{20} \dots c_{n0})'$.

In any case, with or without restrictions, existence of a static representation for model (5)–(6) is an immediate consequence of the following remark.

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

Proof. Let r be the dimension of $\mathcal{X}_t$. Stationarity of $\boldsymbol{\chi}_{nt}$, for all n, implies that r is independent of t. Without loss of generality we can assume that $\boldsymbol{f}_t = (\chi_{1t} \chi_{2t} \dots \chi_{rt})$ is a basis in $\mathcal{X}_t$, for a given t. Again, stationarity of $\boldsymbol{\chi}_{nt}$ implies that, for all t, $\boldsymbol{f}_t$ is a basis for $\mathcal{X}_t$ and that $\chi_{it} = a_i \boldsymbol{f}_t$, for all i, with a_i independent of t. Thus $\boldsymbol{\chi}_{nt} = A_n \boldsymbol{f}_t$, where A_n is $n \times r$ with a_j on its jth row. Moreover setting $N(L) = B_r(L)$, we have $\boldsymbol{f}_t = N(L)\boldsymbol{u}_t$. ∎

Model (5)–(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 (5)–(6). This is fairly obvious. If $\phi_j(L)$ is the least common multiple of the denominators of the entries in the jth column of $N(L)$, then $N(L) = N_1(L) N_2(L)^{-1}$, where $N_1(L)$ is an $r \times q$ moving average (MA) 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 assumption is equivalent to assuming (5) and (6).

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

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

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

$$\boldsymbol{f}_t - D_1\boldsymbol{f}_{t-1} - \cdots - D_m\boldsymbol{f}_{t-m} = R\boldsymbol{u}_t, \tag{8}$$

where R is an $r \times q$ matrix, so that $N(L) = (I - D_1L - \cdots - D_mL^m)^{-1}R$. This assumption implies, as shown in Proposition 2 in Section 3.2, that $\boldsymbol{u}_t$ is identified up to an orthogonal matrix. However, the VAR specification also implies that $\boldsymbol{u}_t$ belongs to the space spanned by present and past values of the variables χ_{it}, i.e., that $\boldsymbol{u}_t$ is fundamental for the χ'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. Fundamentalness in SVAR Analysis. 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 MA representation, i.e., that there exist a q-dimensional white noise $\boldsymbol{v}_t$ and an $n \times q$, one-sided, square-summable filter $K(L)$, such that

$$\boldsymbol{\mu}_t = K(L)\boldsymbol{v}_t. \tag{9}$$

If $\boldsymbol{v}_t$ belongs to the space spanned by present and past values of $\boldsymbol{\mu}_t$ we say that representation (9) is *fundamental* and that $\boldsymbol{v}_t$ 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 $\boldsymbol{v}_t$ 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 z. Then (see, e.g., Rozanov, 1967, Ch. 1, Sect. 10; Ch. 2, p. 76) we have the following result.

PROPOSITION F. *Representation (9) is fundamental if and only if the rank of $K(z)$ is q for all z such that $|z| < 1$.*

Assuming that (9) is fundamental, all fundamental white-noise vectors $\boldsymbol{z}_t$ are linear transformations of $\boldsymbol{v}_t$, i.e., $\boldsymbol{z}_t = C\boldsymbol{v}_t$ (see Proposition 2 in Section 3.2). Nonfundamental white-noise vectors result from $\boldsymbol{v}_t$ 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

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

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

Fundamentalness also plays an important role in 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 $\boldsymbol{w}_t$. The structural shocks are then obtained as a linear transformation $A\boldsymbol{w}_t$, 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), economic theory, in general, does not provide support for fundamentalness, so that all representations that fulfill the same economic statements but are nonfundamental 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 subsequently, fundamentalness is a nongeneric property for $n = q$, whereas it is generic for $n > q$. Thus the question "why assume fundamentalness?", which is legitimately asked when $n = q$, is replaced by "why should we care about nonfundamentalness?" when $n > q$.

An easy and effective illustration can be obtained assuming that $q = 1$ and that the entries of $K(L) = (K_1(L)K_2(L)\ldots K_n(L))'$ are polynomials whose degree does not exceed s, so that $K(L)$ is parameterized in $\mathbb{R}^{n(s+1)}$. In this case, if $n = q = 1$, nonfundamentalness 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 if nonfundamentalness holds for a point $\boldsymbol{\kappa}$ in the parameter space it holds also within a neighborhood of $\boldsymbol{\kappa}$. Existence of points in the parameter space for which nonfundamentalness holds is obvious; thus fundamentalness is nongeneric.

On the other hand, if $n > q = 1$, by Proposition F, nonfundamentalness 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). Fundamentalness is therefore generic.[1]

3.1.2. An Example of Structural Nonfundamentalness. The preceding discussion has a forceful macroeconomic counterpart. Let us first adapt to our framework the classic permanent-income consumption model, as used in Fernández-Villaverde, Rubio-Ramirez, Sargent, and Watson (2007) to illustrate nonfundamentalness. 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 labor 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 preceding equations we obtain

$$s_t - s_{t-1} = \sigma_u R^{-1}(1 - RL)u_t, \tag{10}$$

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), that is just a univariate autoregression, would produce an innovation that 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 (Sect. II of Fernández-Villaverde et al., 2007). 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 $\boldsymbol{u}_t$ for $\boldsymbol{\chi}_{nt}$.

1. As a rule nonfundamentalness 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, and 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 data set $x_{it} = \chi_{it} + \xi_{it}$, which is observed by the econometrician. The common components χ_{it} can be recovered using our large-dimensional factor-model techniques. Moreover, assuming for simplicity that $q = 1$, as 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 χ_{it} are driven by q shocks, a macroeconomic model that contains only q variables, suppose they are χ_{jt}, $j = 1, \ldots, q$, cannot ensure fundamentalness of $\boldsymbol{u}_t$, the reason being possible superior information of the agents with respect to present and past values of χ_{jt}, $j = 1, \ldots, q$. However, the informational advantage of the agents disappears if the econometrician observes a large set of additional macroeconomic variables. The generating process of χ_{jt}, $j = q+1, \ldots, n$, contains parameters that do not belong to the generating process of the first q and vice versa. Therefore, with all likelihood, their dynamic responses to $\boldsymbol{u}_t$ are sufficiently *heterogeneous*, with respect to the first q, to prevent the rank reduction that is, by Proposition (F), equivalent to nonfundamentalness.

3.1.3. Fundamentalness in the Dynamic Factor Model. Based on the preceding discussion we assume fundamentalness of $\boldsymbol{u}_t$ for χ_{it}, $i \in \mathbb{N}$.

PROPOSITION 1. *Under Assumptions 1, 2, and 4, fundamentalness of $\boldsymbol{u}_t$ for χ_{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 Assumptions 1–5, $\boldsymbol{u}_t$ belongs to the space spanned by present and past values of x_{it}, $i = 1, \ldots, \infty$, i.e., is fundamental for x_{it}, $i \in \mathbb{N}$.*

Proof. If $\boldsymbol{u}_t$ is fundamental for χ_{it}, $i \in \mathbb{N}$, then it is fundamental for $\boldsymbol{f}_t$; i.e., there exists a $q \times r$ filter $G(L)$ such that $\boldsymbol{u}_t = G(L)\boldsymbol{f}_t = G(L)N(L)\boldsymbol{u}_t$. As $\boldsymbol{u}_t$ 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 $S_n(L) = G(L)\left(A_n'A_n\right)^{-1} A_n'$, we have $S_n(L)\boldsymbol{x}_{nt} = S_n(L)\boldsymbol{\chi}_{nt} + S_n(L)\boldsymbol{\xi}_{nt}$. Now

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

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

The preceding proof also shows that fundamentalness of $\boldsymbol{u}_t$ for χ_{it}, $i \in \mathbb{N}$, is equivalent to fundamentalness of $\boldsymbol{u}_t$ 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 (5)–(6), if no restrictions hold among the coefficients of the matrices C_{nj}, we have $\tilde{N}(L) = (K(L)K(L)L \ldots K(L)L^s)'$, which has left inverse $(\Theta(L)0_q \ldots 0_q)$. 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 $\boldsymbol{f}_t$, let us consider the orthogonal projection of $\boldsymbol{f}_t$ on the space spanned by its past values:

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

$$\boldsymbol{f}_t = \text{Proj}(\boldsymbol{f}_t | \boldsymbol{f}_{t-1}, \boldsymbol{f}_{t-2}, \ldots,) + \boldsymbol{w}_t, \tag{11}$$

where $\boldsymbol{w}_t$ is the r-dimensional vector of the residuals. Under our assumptions, $\boldsymbol{w}_t$ has rank q. Moreover, by the same argument used to prove Proposition 2 (see Section 3.2), Assumption 7 implies that $\boldsymbol{w}_t = R\boldsymbol{u}_t$, where R is a maximum-rank $r \times q$ matrix.

To get some insight into the orthogonal projection (11), consider again representation (5)–(6) with no restrictions. The static representation of the model has $r = q(s+1)$ and

$$\boldsymbol{f}_t = (K(L) K(L)L \ldots K(L)L^s)' \boldsymbol{u}_t.$$

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

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

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

Joining this observation with the usual approximation argument, a specification of $\boldsymbol{f}_t$ as in (8), even with m very small, does not seem to cause a dramatic loss of generality. In what follows we will adopt the VAR(1) specification that follows.

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

$$\boldsymbol{f}_t = D\boldsymbol{f}_{t-1} + R\boldsymbol{u}_t, \tag{12}$$

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

Under (12),

$$\boldsymbol{\chi}_{nt} = B_n(L)\boldsymbol{u}_t = A_n(I - DL)^{-1} R\boldsymbol{u}_t. \tag{13}$$

Note that assuming (12) (or a higher order autoregressive equation) is independent of the particular static factors we choose. For example, let $\boldsymbol{g}_t = G\boldsymbol{f}_t$, where G is $r \times r$ and invertible, be another basis in the space spanned by the χ_{it}'s. If $\boldsymbol{f}_t$ fulfills (12), then $\boldsymbol{g}_t = [GDG^{-1}]\boldsymbol{g}_{t-1} + [GR]\boldsymbol{u}_t$.

A convenient alternative formulation of Assumption 7′ is as follows.

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

$$\boldsymbol{f}_t = D\boldsymbol{f}_{t-1} + \boldsymbol{\epsilon}_t, \tag{14}$$

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

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

3.2. Alternative Fundamental Representations

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

PROPOSITION 2. *Consider the common components of model (1)*

$$\boldsymbol{\chi}_{nt} = B_n(L)\boldsymbol{u}_t \tag{15}$$

under Assumptions 1–7. If

$$\boldsymbol{\chi}_{nt} = C_n(L)\boldsymbol{v}_t \tag{16}$$

for any $n \in \mathbb{N}$, where the matrices $C_n(L)$ are nested and $\boldsymbol{v}_t$ is a q-dimensional fundamental orthonormal white-noise vector, then representation (16) is related to representation (15) by

$$\boldsymbol{u}_t = H\boldsymbol{v}_t,$$

$$B_n(L) = C_n(L)H',$$

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

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

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

where $\boldsymbol{r}_t$ is orthogonal to $\boldsymbol{v}_{t-k}$, $k \geq 0$. Now consider that $\mathcal{V}_t^-$ and the space spanned by present and past values of χ_{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 (16), whereas the entries of $\boldsymbol{v}_{t-k}$, $k \leq 0$, 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}, $i = 1, \ldots, q$, call it $\mathcal{U}_t^-$, so that $\mathcal{U}_t^- = \mathcal{V}_t^-$. Hence $\boldsymbol{r}_t = 0$. Moreover, serial noncorrelation of the v_{ht}'s implies that $\sum_{k=1}^{\infty} H_k \boldsymbol{v}_{t-k}$ is the projection of $\boldsymbol{u}_t$ on $\mathcal{V}_{t-1}^-$, which is zero because $\mathcal{V}_{t-1}^- = \mathcal{U}_{t-1}^-$. It follows that $\boldsymbol{u}_t = H_0\boldsymbol{v}_t$. Orthonormality of $\boldsymbol{u}_t$ implies that H_0 is orthogonal, i.e., $H_0H_0' = I$. ■

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 out along the same lines as standard SVAR analysis. Precisely, the following procedure is carried out.

Step A. We select the variables of interest, the first m with no loss of generality. Usually $m = q$.

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

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

Step 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 $\boldsymbol{u}_t = H\boldsymbol{v}_t$.

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

Assuming that the variables of interest have been selected, let us concentrate on step B. Denote by M^χ the $r \times r$ diagonal matrix having μ_j^χ as entry (j, j) and by W^χ the $n \times r$ matrix whose jth column is a unit-modulus eigenvector of Γ_0^χ corresponding to μ_j^χ, so that

$$W^{\chi\prime}\Gamma_0^\chi = M^\chi W^{\chi\prime}. \tag{17}$$

Then define

$$\boldsymbol{g}_t = \frac{1}{\sqrt{n}} W^{\chi\prime}\boldsymbol{\chi}_{nt}. \tag{18}$$

The entries of $\boldsymbol{g}_t$ are the first r (nonnormalized) principal components of $\boldsymbol{\chi}_{nt}$. Assumption 4(b) implies that for n large enough $\boldsymbol{g}_t$ 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 as follows.

1. By Assumption 7″, $\boldsymbol{g}_t = D\boldsymbol{g}_{t-1} + \boldsymbol{\epsilon}_t$, where, using $\Gamma_k^g = \mathrm{E}(\boldsymbol{g}_t\boldsymbol{g}_{t-k}') = (1/n)W^{\chi\prime}\Gamma_k^\chi W^\chi$,

$$D = \Gamma_1^g(\Gamma_0^g)^{-1} = \frac{1}{n}W^{\chi\prime}\Gamma_1^\chi W^\chi \left(\frac{M^\chi}{n}\right)^{-1}. \tag{19}$$

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

$$\Gamma^\epsilon = \Gamma_0^g - D\Gamma_0^g D' = \frac{M^\chi}{n} - D\frac{M^\chi}{n}D'. \tag{20}$$

3. Now let μ_j^ϵ, $j = 1, \ldots, q$, be the jth eigenvalue of Γ^ϵ, in decreasing order, $\mathcal{M}$ the $q \times q$ diagonal matrix with $\sqrt{\mu_j^\epsilon}$ as its (j, j) entry, K_j a unit-modulus column eigenvector corresponding to μ_j^ϵ, and $K = (K_1\ K_2\ \ldots\ K_q)$. Defining $\boldsymbol{v}_t = \mathcal{M}^{-1}K'\boldsymbol{\epsilon}_t$ and $\mathcal{K} = K\mathcal{M}$,

$$\boldsymbol{\chi}_{mt} = Q_m(I - DL)^{-1}\mathcal{K}\boldsymbol{v}_t = \sum_{h=0}^{\infty} Q_m D^h \mathcal{K}\boldsymbol{v}_{t-h} = C_m(L)\boldsymbol{v}_t, \tag{21}$$

where, using (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),

$$Q_m = \mathrm{E}(\boldsymbol{\chi}_{mt}\boldsymbol{g}_t')[\mathrm{E}(\boldsymbol{g}_t\boldsymbol{g}_t')]^{-1} = \sqrt{n}\mathcal{I}_m' W^\chi. \tag{22}$$

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

The proof of our consistency result will need that the first q eigenvalues of the matrix Γ^{ϵ} be distinct and asymptotically bounded away from zero (like the eigenvalues of Γ_0^{χ}/n; see Assumption 4(b)).

Assumption 7. There exist constants $\underline{d}_i$ and $\overline{d}_i$, $i = 1, \ldots, q$, such that $\underline{d}_i > \overline{d}_{i+1}$, $i = 1, \ldots, q-1$, and

$$0 < \underline{d}_i < \liminf_{n\to\infty} \mu_i^{\epsilon} \leq \limsup_{n\to\infty} \mu_i^{\epsilon} < \overline{d}_i.$$

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

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

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

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

4.2. Estimation

Let us start with some definitions and notation.

1. $\hat{\Gamma}_k^x = \frac{1}{T}\sum_{h=k+1}^{T} \boldsymbol{x}_{nt}\boldsymbol{x}'_{nt-k}$;
2. $\hat{\mu}_j^x$ is the jth eigenvalue of $\hat{\Gamma}_0^x$;
3. $\hat{M}^x$ is the $r \times r$ diagonal matrix with $\hat{\mu}_j^x$ as its entry (j, j);
4. $\hat{W}^x$ is the $n \times r$ matrix with the corresponding normalized eigenvectors on the columns.

The main motivation for using the static factors $\boldsymbol{g}_t$, as defined in (18), is that $\boldsymbol{g}_t$ can be approximated in probability by the sample principal components of $\boldsymbol{x}_{nt}$:

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

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 Γ_k^{χ}, M^{χ}, and W^{χ} 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 $\hat{B}_m(L) = \hat{C}_m(L)\hat{H}'$.

Let us now state our last assumption and the consistency result.

Assumption 8. Denote by $\gamma^x_{k,ij}$ and $\hat{\gamma}^x_{k,ij}$ the entries of Γ^x_k and $\hat{\Gamma}^x_k$, respectively. There exists a positive real ρ such that

$$T\,\mathrm{E}\left[(\hat{\gamma}^x_{k,ij} - \gamma^x_{k,ij})^2\right] < \rho,$$

for $k = 0, 1$ and for all positive integers T, i, and j.

Given i and j, convergence of $\mathcal{A}_{k,ij} = T\,\mathrm{E}\left[(\hat{\gamma}_{k,ij} - \gamma_{k,ij})^2\right]$ for $T \to \infty$ can be obtained under mild conditions on the autocovariances and fourth cumulants of the x's.[2] As a specification of such conditions would not be useful here, we simply assume convergence of $\mathcal{A}_{k,ij}$, which obviously implies that for some positive real numbers $\rho_{k,ij}$,

$$T\,\mathrm{E}\left[(\hat{\gamma}_{k,ij} - \gamma_{k,ij})^2\right] < \rho_{k,ij},$$

for all T. Assumption 9 requires the additional condition that the real numbers $\rho_{k,ij}$ have a common upper bound ρ.

PROPOSITION 3. *Denote by $b_{k,ij}$ the entries of the kth lag matrix coefficient of $B_m(L)$. Under Assumptions 1–9, for all $k \geq 0$, $i = 1, \ldots, m$, $j = 1, \ldots, q$,*

$$|b_{k,ij} - \hat{b}_{k,ij}| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right). \tag{23}$$

Proof. See the Appendix. ■

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 member of the infinite set of functions fulfilling the restrictions, (23) can be applied to $B_{3,3}(L)$.

In conclusion, (23) applies with just or partial identification. Overidentification is left to further research.

5. AN EMPIRICAL APPLICATION

We illustrate our structural factor model by revisiting an influential work in SVAR literature, namely, the three-dimensional SVAR estimated in King et al. (1991). The variables are U.S. 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 King et al. (1991) with the same sampling period. As we see subsequently, three common shocks, i.e., $q = 3$, are consistent with our data set. 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 King et al. (1991) allows a sensible and interesting comparison between our impulse-response functions and those found in King et al. (1991).

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 King et al. (1991) are available on Mark Watson's home page (www.princeton.edu/mwatson/public.html). We collected 89 series, including data from NIPA (National Income and Product Accounts) 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 (as in King et al., 1991) and the need to balance data of different groups. 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, because 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 $r = 15$ because with $r = 30$ none of the criteria reaches a minimum for $r < 30$. The criterion PC_{p1} gives $r = 15$, PC_{p2} gives $r = 14$, and PC_{p3} gives $r = 20$. Subsequently 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 the total variance for output, 84.4% for investment, and 89.4% for consumption. The 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 $r = 9$. Result 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 out the comparison between our results and those of King et al. (1991), does not conflict with available evidence.

5.3. Fundamentalness

We are interested in the impulse-response functions of per capita output, investment, and consumption, i.e., 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 3×3 matrix $C_3(L)$ is not necessarily fundamental.[3] 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 $\det(C_3(L))$ 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, because 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

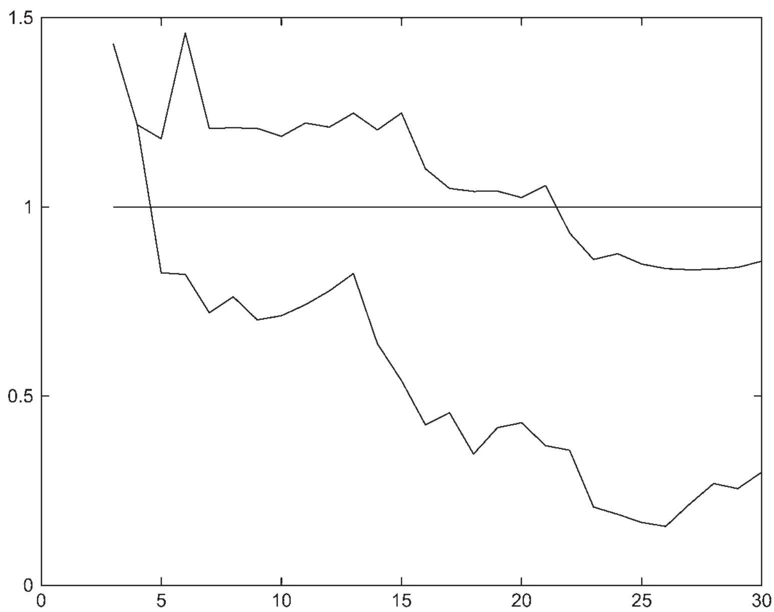

FIGURE 1. The moduli of the first and the second smallest roots as functions of r.

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

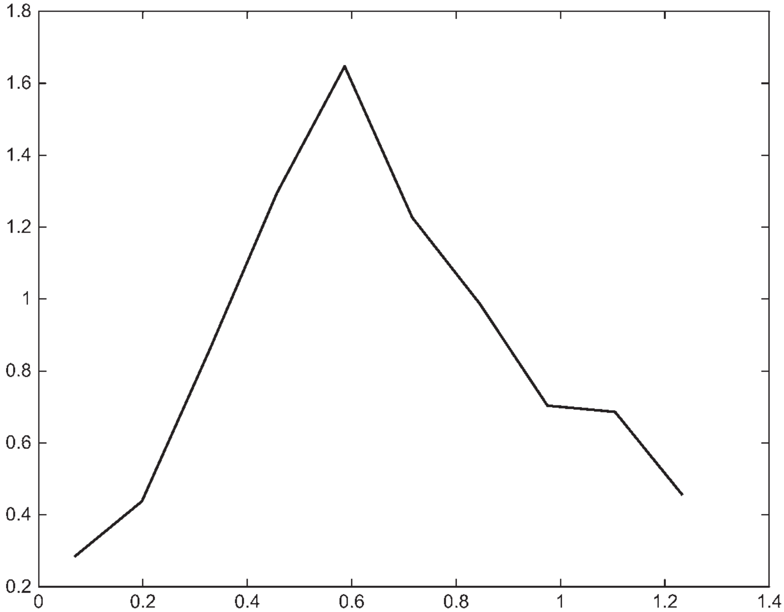

FIGURE 2. Frequency distribution of the modulus of the smallest root.

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

Figure 2 reports the distribution of the modulus of the smallest root for $r = 15$, across 1,000 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 favor nonfundamentalness 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 we argue in Section 5.4, nonfundamentalness of $C_3(L)$ explains some important differences between our structural impulse-response function and that of King et al. (1991).

5.4. Impulse-Response Functions and Variance Decomposition

Coming to the impulse-response functions, as anticipated earlier we impose long-run neutrality of two shocks on per capita output, as in King et al. (1991). 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

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

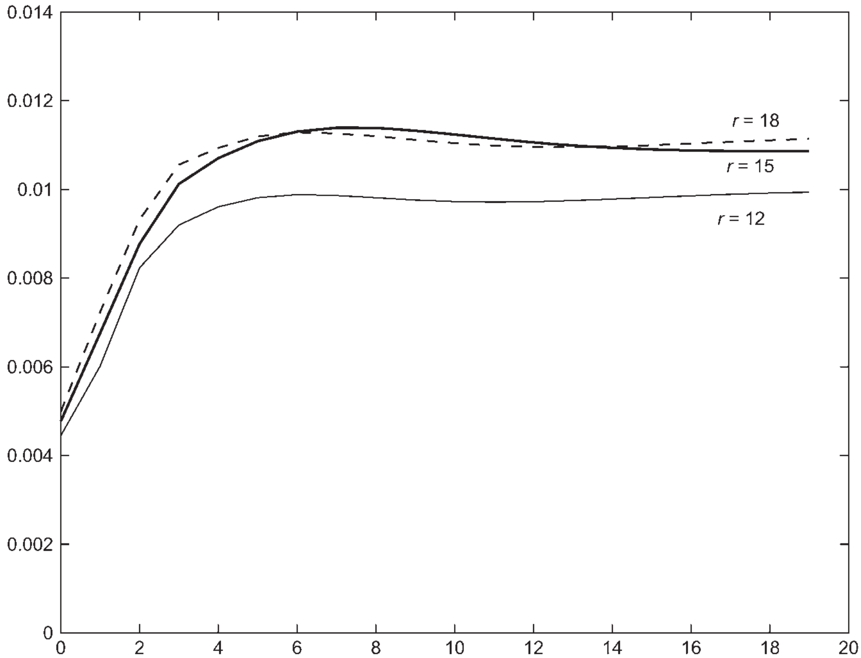

FIGURE 3. The impulse-response function of the long-run shock on output for $r = 12, 15, 18$.

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 King et al. (1991), where there is a sharp decline during the second and the third years, which drives the overall effect back to the impact value.

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 before). 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 King et al. (1991).

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 King et al. (1991) 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

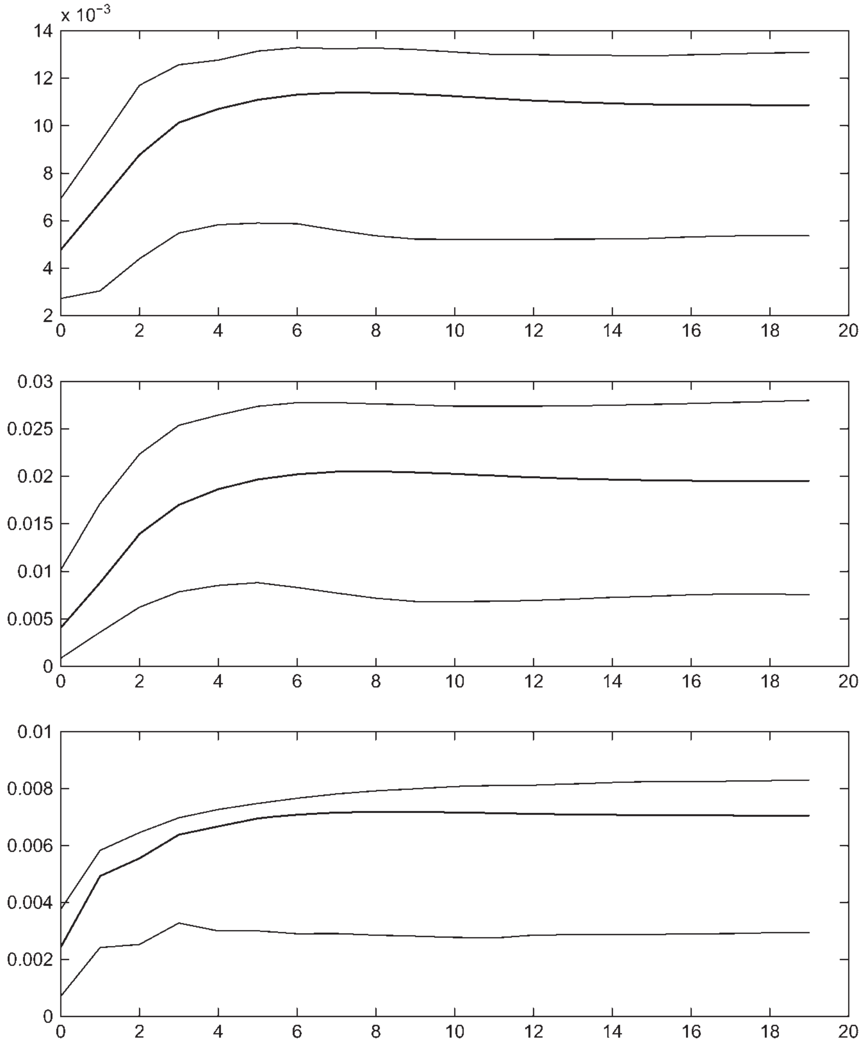

FIGURE 4. The plots represent the impulse-response functions of the long-run shock on output (top), consumption (middle), and investment (bottom) for $r = 15$.

according to King et al. (1991) 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), nonfundamentalness of the structural shocks for $(\chi_{1t}\ \chi_{2t}\ \chi_{3t})$, as opposed to funda-

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

TABLE 1. Fraction of the forecast-error variance due to the long-run shock

	Dynamic factor model			King et al. (1991)		
Horizon	Output	Consumption	Investment	Output	Consumption	Investment
1	0.37	0.30	0.07	0.45	0.88	0.12
	(0.20)	(0.26)	(0.16)	(0.28)	(0.21)	(0.18)
4	0.57	0.77	0.42	0.58	0.89	0.31
	(0.18)	(0.18)	(0.17)	(0.27)	(0.19)	(0.23)
8	0.78	0.87	0.72	0.68	0.83	0.40
	(0.13)	(0.13)	(0.13)	(0.22)	(0.18)	(0.18)
12	0.86	0.90	0.80	0.73	0.83	0.43
	(0.09)	(0.11)	(0.11)	(0.19)	(0.18)	(0.17)
16	0.89	0.91	0.83	0.77	0.85	0.44
	(0.08)	(0.11)	(0.10)	(0.17)	(0.16)	(0.16)
20	0.91	0.92	0.86	0.79	0.87	0.46
	(0.07)	(0.12)	(0.09)	(0.16)	(0.15)	(0.16)

mentalness of the King et al. (1991) shocks for $(x_{1t}\ x_{2t}\ x_{3t})$, appears to play a major role in explaining such different dynamic profiles.

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 allows 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 et al. (1991) for U.S. output, investment, and consumption. Using a large panel including such series, we estimate a factor model with three common shocks and apply the King et al. (1991) identification scheme. Two important outcomes are as follows.

1. The three-dimensional impulse-response function corresponding to output, investment, and consumption, implicit in our estimated factor model, is nonfundamental, an important difference with respect to the VAR estimated in King et al. (1991).
2. Comparing responses of the permanent shock in King et al. (1991) 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 King et al. (1991).

NOTES

1. For a general result see Anderson and Deistler (2008). The proof of their Proposition 1 can be adapted to show that for a generic rational $K(L)$, with given maxima for the orders of numerator and denominator polynomials, if $n > q \geq 1$, then $\text{rank}(K(z)) = q$ for all complex numbers z (no matter whether inside, on, or outside the unit circle).

2. Anderson (1971, Sect. 8.3), discusses the univariate case; see, in particular, Theorem 8.3.3, in which convergence is obtained under summability of the squared second moments and fourth cumulants. Analogous conditions can be obtained in the multivariate case (for formulas linking the second moments of estimated covariances to second and fourth cumulants of the x's, see Hannan, 1970, pp. 209–211).

3. Note that fundamentalness of $C_3(L)$ and of $C_3(L)H'$ are equivalent.

REFERENCES

Altug, S. (1989) Time-to-build and aggregate fluctuations: Some new evidence. *International Economic Review* 30, 889–920.

Anderson, B.D.O. & M. Deistler (2008) Properties of zero-free transfer function matrices. *SICE Journal of Control, Measurement, and System Integration*, forthcoming.

Anderson, T.W. (1971). *The Statistical Analysis of Time Series.* Wiley.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* 71, 135–171.

Bai, J. & S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.

Bai, J. & S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics* 25, 52–60.

Bernanke, B.S. & J. Boivin (2003). Monetary policy in a data rich environment. *Journal of Monetary Economics* 50, 525–546.

Bernanke, B.S., J. Boivin, & P. Eliasz (2005). Measuring monetary policy: A factor augmented autoregressive (FAVAR) approach. *Quarterly Journal of Economics* 120, 387–422.

Boivin, J. & M.P. Giannoni (2006). Has monetary policy become more effective? *Review of Economics and Statistics* 88, 445–462.

Boivin, J. & S. Ng (2003). Are more data always better for factor analysis? *Journal of Econometrics* 127, 169–194.

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

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

Chari, V.V., P.J. Kehoe, & E.R. McGrattan (2005). A Critique of Structural VARs Using Real Business Cycle Theory. Federal Reserve Bank of Minneapolis Working paper 631.

Connor, G. & R.A. Korajczyk (1988). Risk and return in an equilibrium APT. Application of a new test methodology. *Journal of Financial Economics* 21, 255–289.

Fernández-Villaverde, J., J.F. Rubio-Ramirez, T.J. Sargent, & M.W. Watson (2007) ABCs (and Ds) of understanding VARs. *American Economic Review* 97, 1021–1026.

Forni, M., M. Hallin, M. Lippi, & L. Reichlin (2000) The generalized dynamic factor model: Identification and estimation. *The Review of Economics and Statistics* 82, 540–554.

Forni, M., M. Hallin, M. Lippi, & L. Reichlin (2005) The generalized factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association* 100, 830–840.

Forni, M. & M. Lippi (2001). The generalized dynamic factor model: Representation theory. *Econometric Theory* 17, 1113–1141.

Forni, M. & L. Reichlin (1998). Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies* 65, 453–473.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D.J. Aigner & A.S. Goldberger (eds.), *Latent Variables in Socio-Economic Models*, pp. 365–383. North-Holland.

Giannone, D., L. Reichlin, & L. Sala (2002). Tracking Greenspan: Systematic and Nonsystematic Monetary Policy Revisited. CEPR Discussion paper 3550.

Giannone, D., L. Reichlin, & L. Sala (2005). Monetary policy in real time. In M. Gertler & K. Rogoff, (eds.), *NBER Macroeconomic Annual. 2004*, pp. 161–200. MIT Press.

Giannone, D., L. Reichlin, & L. Sala (2006). VARs, common factors and the empirical validation of equilibrium business cycle models. *Journal of Econometrics* 127, 257–279.

Hallin, M. & R. Liška (2007). The generalized dynamic factor model: Determining the number of factors. *Journal of the American Statistical Association* 102, 103–117.

Hannan, E.J. (1970). *Multiple Time Series*. Wiley.

Hannan, E.J. & M. Deistler (1988). *The Statistical Theory of Linear Systems.* Wiley.

Hansen, L.P. & T.J. Sargent (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control* 2, 7–46.

Hansen, L.P. & T.J. Sargent (1991). Two problems in interpreting vector autoregressions. In L.P. Hansen & T.J. Sargent (eds.), *Rational Expectations Econometrics*. pp. 77–119. Westview.

Ireland, P.N. (2004). A method for taking models to the data. *Journal of Economic Dynamics and Control* 28, 1205–1226.

King, R.G., C.I. Plosser, J.H. Stock, & M.W. Watson (1991). Stochastic trends and economic fluctuations. *American Economic Review* 81, 819–840.

Lippi, M. & L. Reichlin (1993) The dynamic effects of aggregate demand and supply disturbances: Comment, *American Economic Review* 83, 644–652.

Lippi, M. & L. Reichlin (1994) VAR analysis, nonfundamental representation, Blaschke matrices. *Journal of Econometrics* 63, 307–325.

Rozanov, Y. (1967). *Stationary Random Processes*. Holden-Day.

Sargent, T.J. (1989). Two models of measurements and the investment accelerator. *Journal of Political Economy* 97, 251–287.

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*, 45–109. Federal Reserve Bank of Minneapolis.

Stewart, G.W. & J. Sun (1990). *Matrix Perturbation Theory*. Academic Press.

Stock, J.H. & M.W. Watson (2002a). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20, 147–162.

Stock, J.H. & M.W. Watson (2002b). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.

Stock, J.H. & M.W. Watson (2005). Implications of Dynamic Factor Models for VAR Analysis. NBER Working papers 11467.

van der Waerden, B.L. (1953). *Modern Algebra*, vol. I. Frederick Ungar.

APPENDIX: Proof of Proposition 3

If A is a symmetric matrix we denote by $\mu_j(A)$ the jth 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. We will make use of the Weyl inequality (see, e.g., Stewart and Sun, 1990, Cor. 4.10, p. 203): If A and B are two $s \times s$ symmetric matrices, then

$$|\mu_j(A+B) - \mu_j(A)| \leq \sqrt{\mu_1(B^2)} = \|B\|, \qquad j = 1, \ldots, s. \tag{A.1}$$

PROPOSITION P.

(i) $\|\hat{Q}_m - Q_m \hat{\mathcal{J}}_r\| = O_p\left(\max\left((1/\sqrt{n}), (1/\sqrt{T})\right)\right)$,

(ii) $\|\hat{D}^k - \hat{\mathcal{J}}_r D^k \hat{\mathcal{J}}_r\| = O_p\left(\max\left((1/n), (1/\sqrt{T})\right)\right)$, *for all* $k \geq 0$,

(iii) $\|\hat{\mathcal{K}} - \hat{\mathcal{J}}_r \mathcal{K} \hat{\mathcal{J}}_q\| = O_p\left(\max\left((1/n), (1/\sqrt{T})\right)\right)$,

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. To see why we need $\hat{\mathcal{J}}_r$ and $\hat{\mathcal{J}}_q$ let us observe that the entries of $\boldsymbol{g}_t$ and $\hat{\boldsymbol{g}}_t$, and also the columns of K and $\hat{K}$, all result from taking eigenvectors and are therefore identified up to a sign (by Assumptions 4(b) and 8, no multiple eigenvalues can occur). Thus, e.g., $\hat{Q}_m$ converges to Q_m but only if we choose the right signs.

As a consequence, the raw impulse-response functions, i.e., the rows of $C_m(L)$ as defined in (21), and the entries of $\boldsymbol{v}_t$ are identified up to a sign:

$$\chi_{mt} = C_m(L)\boldsymbol{v}_t = \left[Q_m \hat{\mathcal{J}}_r\right]\left(I - \left[\hat{\mathcal{J}}_r D \hat{\mathcal{J}}_r\right] L\right)^{-1}\left[\hat{\mathcal{J}}_r \mathcal{K} \hat{\mathcal{J}}_q\right]\left[\hat{\mathcal{J}}_q \boldsymbol{v}_t\right]$$

$$= \check{Q}_m (I - \check{D}L)^{-1} \check{\mathcal{K}} \check{\boldsymbol{v}}_t = \check{C}_m(L)\check{\boldsymbol{v}}_t,$$

with $\check{Q}_m = Q_m \hat{\mathcal{J}}_r$, $\check{D} = \hat{\mathcal{J}}_r D \hat{\mathcal{J}}_r$, $\hat{\mathcal{K}} = \mathcal{J}_r \mathcal{K} \hat{\mathcal{J}}_q$, $\check{\boldsymbol{v}}_t = \hat{\mathcal{J}}_q \boldsymbol{v}_t$, $\check{C}_m(L) = C_m(L)\hat{\mathcal{J}}_q$.

On the other hand, $\boldsymbol{\chi}_{mt} = \check{C}_m(L)\check{\boldsymbol{v}}_t$ is obviously a fundamental representation. Therefore, by assumption, application of the zero and sign restrictions to $\boldsymbol{\chi}_{mt} = \check{C}_m(L)\check{\boldsymbol{v}}_t$ and to $\boldsymbol{\chi}_{mt} = C_m(L)\boldsymbol{v}_t$ gives the same result; i.e., setting $\check{H} = F(\check{Q}_m, \check{D}, \check{\mathcal{K}})$, we have

$$B_m(L) = C_m(L)H' = \check{C}_m(L)\check{H}'; \tag{A.2}$$

see the discussion in Section 4.1 (step C at the beginning and point 2 at the end).

Last, 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 (A.2), implies Proposition 3.

The proof of Proposition P requires some intermediate results.

LEMMA 1. *Denoting by* $\mathcal{I}_m$ *the* $n \times m$ *matrix having the identity matrix* I_m *in the first* m *rows and zero 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) $\frac{1}{\sqrt{n}}\|\mathcal{I}_m'\left(\hat{\Gamma}_0^x - \Gamma_0^x\right)\| = O_p\left(\frac{1}{\sqrt{T}}\right)$ *for any (fixed)* m.

(iii) $\frac{1}{n}\|\hat{\Gamma}_k^x - \Gamma_k^\chi\| = O_p\left(\max\left((1/n), (1/\sqrt{T})\right)\right)$, $\quad k = 0, 1.$

(iv) $\frac{1}{\sqrt{n}}\|\mathcal{I}_m'\left(\hat{\Gamma}_0^x - \Gamma_0^\chi\right)\| = O_p\left(\max\left((1/\sqrt{n}), (1/\sqrt{T})\right)\right)$ *for any (fixed)* m.

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

Proof. We have

$$\mu_1\left((\hat{\Gamma}_k^x - \Gamma_k^x)(\hat{\Gamma}_k^x - \Gamma_k^x)'\right) \le \text{trace}\left((\hat{\Gamma}_k^x - \Gamma_k^x)(\hat{\Gamma}_k^x - \Gamma_k^x)'\right) = \sum_{i=1}^n \sum_{j=1}^n (\hat{\gamma}_{k,ij}^x - \gamma_{k,ij}^x)^2.$$

By Assumption 9,

$$\mathrm{E}\left[\sum_{i=1}^n \sum_{j=1}^n \left(\hat{\gamma}_{k,ij}^x - \gamma_{k,ij}^x\right)^2\right] < \frac{n^2\rho}{T},$$

for all positive integers T and $k = 0, 1$. Statement (i) follows using Chebychev's inequality. Similarly, we have

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

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 $\frac{1}{n}\left\|\hat{\Gamma}_k^x - \Gamma_k^\chi\right\| \le \frac{1}{n}\left\|\hat{\Gamma}_k^x - \Gamma_k^x\right\| + \frac{1}{n}\left\|\Gamma_k^\xi\right\|$. The first term on the right-hand side is $O_p\left(\frac{1}{\sqrt{T}}\right)$ by statement (i), whereas the second is bounded by $(1/n)\mu_1^\xi$, 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 $(1/\sqrt{n})\mu_1^\xi$ instead of $(1/n)\mu_1^\xi$. ∎

LEMMA 2.

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

(ii) There exists $\bar{n}$ such that, for all $n \ge \bar{n}$, $(\frac{M^\chi}{n})$ is invertible.

(iii) For any $n \ge \bar{n}$ and $\eta > 0$, there exists $\tau(\eta, n)$ such that, for $T \ge \tau(\eta, n)$, $\left(\frac{\hat{M}^x}{n}\right)$ 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) $\left\|\frac{M^\chi}{n}\right\|$ and $\left\|\left(\frac{M^\chi}{n}\right)^{-1}\right\|$, 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 (A.1) we get $\frac{1}{n}\left|\hat{\mu}_j^x - \mu_j^\chi\right| \le \frac{1}{n}\left\|\hat{\Gamma}_0^x - \Gamma_0^\chi\right\|$, which is $O_p\left(\max\left((1/n), (1/\sqrt{T})\right)\right)$ by Lemma 1(iii). As for (ii), by Assumption 4(b) there exists $\bar{n}$ such that, for $n \ge \bar{n}$, $\frac{\mu_r^\chi}{n} > \underline{c}_r > 0$, so that $\det\left(\frac{M^\chi}{n}\right) \ne 0$. Turning to (iii), setting $A = \Gamma_0^x$, $B = \hat{\Gamma}_0^x - \Gamma_0^x$ and applying Weyl inequality we get $\frac{1}{n}\left|\hat{\mu}_r^x - \mu_r^x\right| \le \frac{1}{n}\left\|\hat{\Gamma}_0^x - \Gamma_0^x\right\|$, which is $O_p\left(\frac{1}{\sqrt{T}}\right)$ by Lemma 1(i). Now, $\mu_r^x \ge \mu_r^\chi$, because Γ_0^ξ is positive semidefinite, so that, for $n \ge \bar{n}$, $\frac{\mu_r^x}{n} > \underline{c}_r > 0$. Hence $\det\left(\frac{\hat{M}^x}{n}\right)$ is bounded away from zero in probability as $T \to \infty$. The last part of statement (iii) follows from the fact that the rank of the observation matrix $\boldsymbol{x}_n^T$, and therefore the rank of $\hat{\Gamma}_0^x$, is nondecreasing in n and T. As for (iv), observe that $\left\|\frac{M^\chi}{n}\right\| = \frac{\mu_1^\chi}{n}$ and $\left\|\left(\frac{M^\chi}{n}\right)^{-1}\right\| = \frac{n}{\mu_r^\chi}$, 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). ■

LEMMA 3.

(i) $\|W^{\chi\prime}\hat{W}^x(\hat{M}^x/n)-(M^\chi/n)W^{\chi\prime}\hat{W}^x\| = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$.

(ii) $\|\hat{W}^{x\prime}W^\chi W^{\chi\prime}\hat{W}^x - I_r\| = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$.

(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* $\|\hat{W}^{x\prime}W^\chi - \hat{\mathcal{J}}_r\| = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$.

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

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

We have $\|a-c\|\le\|a-b\|+\|b-c\|$. Both terms are $O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$, 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^χ the jth column of $\hat{W}^x$ and W^χ, respectively. By taking a single entry of the matrix on the left-hand side of statement (i) we get

$$\frac{1}{n}\left(\hat{\mu}_j^x-\mu_i^\chi\right)w_j^{\chi\prime}\hat{w}_i^x = O_p\left(\max\left(\frac{1}{n},\frac{1}{\sqrt{T}}\right)\right),$$

$i\le r$, $j\le r$. Now, for $j\ne i$, $\frac{1}{n}\left(\hat{\mu}_j^x-\mu_i^\chi\right)$ is bounded away from zero in probability, because μ_i^χ and μ_j^χ are asymptotically distinct by Assumption 4(b), whereas $\hat{\mu}_j^x$ tends to μ_j^χ in probability by Lemma 2(i). Hence the off-diagonal terms of $\hat{W}^{x\prime}W^\chi$ are $O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$. Turning to the diagonal terms, let us first observe that $\hat{w}_i^{x\prime}W^\chi W^{\chi\prime}\hat{w}_i^x = 1+O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$ by statement (ii). But

$$\hat{w}_i^{x\prime}W^\chi W^{\chi\prime}\hat{w}_i^x = \left(\hat{w}_i^{x\prime}w_i^\chi\right)^2+\sum_{\substack{j=1\\ j\ne i}}^{r}\left(\hat{w}_i^{x\prime}w_j^\chi\right)^2 = \left(\hat{w}_i^{x\prime}w_i^\chi\right)^2+O_p\left(\max\left(\frac{1}{n},\frac{1}{\sqrt{T}}\right)\right).$$

Hence $\left(1-|\hat{w}_i^{x\prime}w_i^\chi|\right)\left(1+|\hat{w}_i^{x\prime}w_i^\chi|\right) = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$, so that $1-|\hat{w}_i^{x\prime}w_i^\chi| = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$. ■

Proof of Proposition P(i). Set

$a = Q_m\hat{\mathcal{J}}_r = \sqrt{n}\mathcal{I}_m'W^\chi\hat{\mathcal{J}}_r$, where $\hat{\mathcal{J}}_r$ has been defined in Lemma 3(iii),

$$b = \sqrt{n}\mathcal{I}_m'W^\chi W^{\chi\prime}\hat{W}^x = \sqrt{n}\mathcal{I}_m'W^\chi W^{\chi\prime}\hat{W}^x\frac{\hat{M}^x}{n}\left(\frac{\hat{M}^x}{n}\right)^{-1},$$

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press

$$c = \sqrt{n}\mathcal{I}_m' W^\chi \frac{M^\chi}{n} W^{\chi\prime}\hat{W}^x \left(\frac{\hat{M}^x}{n}\right)^{-1} = \frac{1}{\sqrt{n}}\mathcal{I}_m'\Gamma_0^\chi \hat{W}^x \left(\frac{\hat{M}^x}{n}\right)^{-1},$$
$$d = \frac{1}{\sqrt{n}}\mathcal{I}_m'\hat{\Gamma}_0^x \hat{W}^x \left(\frac{\hat{M}^x}{n}\right)^{-1} = \sqrt{n}\mathcal{I}_m'\hat{W}^x = \hat{Q}_m.$$

Let us begin by observing that $\left\|\mathcal{I}_m' W^\chi \left(M^\chi\right)^{1/2}\right\| = \|\mathcal{I}_m'\Gamma_0^\chi \mathcal{I}_m\|^{1/2}$ depends on the entries of the upper left $m \times m$ submatrix of Γ_0^χ and therefore does not depend on n. Denoting such fixed quantity by ζ, we have

$$\|\sqrt{n}\mathcal{I}_m' W^\chi\| = \left\|\sqrt{n}\mathcal{I}_m' W^\chi \left(\frac{M^\chi}{n}\right)^{1/2}\left(\frac{M^\chi}{n}\right)^{-1/2}\right\| \leq \zeta \left\|\left(\frac{M^\chi}{n}\right)^{-1/2}\right\|,$$

which is $O(1)$ by Lemma 2(iv). Hence $\|\sqrt{n}\mathcal{I}_m' W^\chi\|$ is $O(1)$, so that we can apply Lemma 3(iii) to get $\|a-b\| = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$ and Lemma 3(i) to get $\|b-c\| = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$. Finally, Lemma 1(iv) ensures that $\|c-d\| = O_p\left(\max\left((1/\sqrt{n}),\ (1/\sqrt{T})\right)\right)$. ■

Proof of Proposition P(ii). We have $\hat{D} = (1/n)\hat{W}^{x\prime}\hat{\Gamma}_1^x\hat{W}^x\left(\frac{\hat{M}^x}{n}\right)^{-1}$ (see (19)) and $\hat{\mathcal{J}}_r D\hat{\mathcal{J}}_r = (1/n)\hat{\mathcal{J}}_r W^{\chi\prime}\Gamma^\chi W^\chi\left(\frac{M^\chi}{n}\right)^{-1}\hat{\mathcal{J}}_r = (1/n)\hat{\mathcal{J}}_r W^{\chi\prime}\Gamma_1^\chi W^\chi\hat{\mathcal{J}}_r\left(\frac{M^\chi}{n}\right)^{-1}$, where $\hat{\mathcal{J}}_r$ has been defined in Lemma 3(iii). Set

$$a = \hat{D} = \frac{1}{n}\hat{W}^{x\prime}\hat{\Gamma}_1^x\hat{W}^x\left(\frac{\hat{M}^x}{n}\right)^{-1},$$

$$b = \frac{1}{n}\hat{W}^{x\prime}\Gamma_1^\chi\hat{W}^x\left(\frac{\hat{M}^x}{n}\right)^{-1} = \frac{1}{n}\hat{W}^{x\prime}W^\chi W^{\chi\prime}\Gamma_1^\chi W^\chi W^{\chi\prime}\hat{W}^x\left(\frac{\hat{M}^x}{n}\right)^{-1},$$

$$c = \frac{1}{n}\hat{\mathcal{J}}_r W^{\chi\prime}\Gamma_1^\chi W^\chi\hat{\mathcal{J}}_r\left(\frac{\hat{M}^x}{n}\right)^{-1},$$

$$d = \hat{\mathcal{J}}_r D\hat{\mathcal{J}}_r = \frac{1}{n}\hat{\mathcal{J}}_r W^{\chi\prime}\Gamma_1^\chi W^\chi\left(\frac{M^\chi}{n}\right)^{-1}\hat{\mathcal{J}}_r = \frac{1}{n}\hat{\mathcal{J}}_r W^{\chi\prime}\Gamma_1^\chi W^\chi\hat{\mathcal{J}}_r\left(\frac{M^\chi}{n}\right)^{-1}.$$

By Lemma 1(i) $\|a-b\|$ is $O_p\left(\frac{1}{\sqrt{T}}\right)$; by Lemma 3(iii) $\|b-c\|$ is $O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$; by Lemma 2(i) $\|c-d\|$ is $O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$. 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. ■

LEMMA 4.

(i) $\|\hat{\Gamma}^\epsilon - \hat{\mathcal{J}}_r\Gamma^\epsilon\hat{\mathcal{J}}_r\| = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$, *where* $\hat{\mathcal{J}}_r$ *has been defined in Lemma 3(iii).*

(ii) $\hat{\mu}_j^\epsilon - \mu_j^\epsilon = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$ $j = 1,\ldots,r$.

(iii) $\|\hat{\mathcal{M}} - \mathcal{M}\| = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$.

(iv) $\mathcal{M}^{-1}$ *exists for n sufficiently large, and its norm is* $O(1)$ *as* $n\to\infty$; *moreover,* $\|\hat{\mathcal{M}}\mathcal{M}^{-1} - I_q\| = O_p\left(\max\left((1/n),(1/\sqrt{T})\right)\right)$.

(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* $\|K' \hat{\mathcal{J}}_r \hat{K} - \hat{\mathcal{J}}_q\| = O_p\left(\max\left((1/n), (1/\sqrt{T})\right)\right)$.

Proof. We have $\hat{\Gamma}^\epsilon = (\hat{M}^x/n) - \hat{D}(\hat{M}^x/n)\hat{D}'$ (see (20)) and $\hat{\mathcal{J}}_r \Gamma^\epsilon \hat{\mathcal{J}}_r = \hat{\mathcal{J}}_r \left((M^\chi/n) - D(M^\chi/n)D'\right)\hat{\mathcal{J}}_r = (M^\chi/n) - \hat{\mathcal{J}}_r D \hat{\mathcal{J}}_r (M^\chi/n)\hat{\mathcal{J}}_r D' \hat{\mathcal{J}}_r$. Statement (i) then follows from Lemma 1(i) and Proposition P(ii). As for (ii), notice first that the eigenvalues of $\hat{\mathcal{J}}_r \Gamma^\epsilon \hat{\mathcal{J}}_r$ are identical to those of Γ^ϵ. Hence setting $A = \Gamma^\epsilon$, $B = \hat{\Gamma}^\epsilon - \hat{\mathcal{J}}_r \Gamma^\epsilon \hat{\mathcal{J}}_r$ and applying (A.1) we get $\left|\hat{\mu}_j^\epsilon - \mu_j^\epsilon\right| \le \left\|\hat{\Gamma}^\epsilon - \hat{\mathcal{J}}_r \Gamma^\epsilon \hat{\mathcal{J}}_r\right\|$. Statement (ii) then follows from (i). Turning to (iii), we have $\hat{\mathcal{M}}^2 - \mathcal{M}^2 = \left(\hat{\mathcal{M}} - \mathcal{M}\right)\left(\hat{\mathcal{M}} + \mathcal{M}\right)$. 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). ■

Proof of Proposition P(iii). 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 $O_p\left(\max\left((1/n), (1/\sqrt{T})\right)\right)$. Hence by Lemma 4(i), $\|\hat{K}\hat{\mathcal{M}}^2\hat{K}' - \hat{\mathcal{J}}_r K \mathcal{M}^2 K' \hat{\mathcal{J}}_r\| = O_p\left(\max\left((1/n), (1/\sqrt{T})\right)\right)$, 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 Lemmas 4(iv) and (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). ■

https://doi.org/10.1017/S026646660809052X Published online by Cambridge University Press