Infinite Dimensional VARs and Factor Models*

Alexander Chudik†
Cambridge University

M. Hashem Pesaran‡
Cambridge University, CIMF and USC

November 19, 2007

Abstract

This paper introduces a novel approach for dealing with the 'curse of dimensionality' in the case of large linear dynamic systems. Restrictions on the coefficients of an unrestricted VAR are proposed that are binding only in a limit as the number of endogenous variables tends to infinity. It is shown that under such restrictions, an infinite-dimensional VAR (or IVAR) can be arbitrarily well characterized by a large number of finite-dimensional models in the spirit of the global VAR model proposed in Pesaran *et al.* (JBES, 2004). The paper also considers IVAR models with dominant individual units and shows that this will lead to a dynamic factor model with the dominant unit acting as the factor. The problems of estimation and inference in a stationary IVAR with unknown number of unobserved common factors are also investigated. A cross section augmented least squares estimator is proposed and its asymptotic distribution is derived. Satisfactory small sample properties are documented by Monte Carlo experiments. An empirical application to modelling of real GDP growth and investment-output ratios provides an illustration of the proposed approach. Considerable heterogeneities across countries and significant presence of dominant effects are found. The results also suggest that increase in investment as a share of GDP predict higher growth rate of GDP per capita for non-negligible fraction of countries and vice versa.

Keywords: Large N and T Panels, Weak and Strong Cross Section Dependence, VAR, Global VAR, Factor Models, Capital Accumulation and Growth

JEL Classification: C10, C33, C51, O40

---

*We are grateful to Elisa Tosetti for helpful comments and suggestions.
†Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, UK. E-mail: ac474@cam.ac.uk; http://www.econ.cam.ac.uk/phd/ac474/
‡Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, UK. E-mail: mhp1@econ.cam.ac.uk; http://www.econ.cam.ac.uk/faculty/pesaran/

1 Introduction

Following the seminal work of Sims (1980), vector autoregressive models (VARs) are widely used in macroeconometrics and finance. VARs provide a flexible framework for the analysis of complex dynamics and interactions that exist between variables in the national and global economy. However, the application of the approach in practice is often limited to a handful of variables which could lead to misleading inference if important variables are omitted merely to accommodate the VAR modelling strategy. Number of parameters to be estimated grows at the quadratic rate with the number of variables, which is limited by the size of typical data sets to no more than 5 to 7. In many empirical applications, this is not satisfactory. Some restrictions must be imposed for the analysis of large systems.

To deal with this so-called 'curse of dimensionality', two different approaches have been suggested in the literature: (*i*) shrinkage of the parameter space and (*ii*) shrinkage of the data. Parameter space can be shrunk by imposing a set of restrictions, which could be for instance obtained from a theoretical structural model, directly on the parameters. Alternatively, one could use techniques, where prior distributions are imposed on the parameters to be estimated. Bayesian VAR (BVAR) proposed by Doan, Litterman and Sims (1984), for example, use what has become known as 'Minnesota' priors to shrink the parameters space.[1] In most applications, BVARs have been applied to relatively small systems[2] (e.g. Leeper, Sims, and Zha, 1996, considered 13 and 18 variable BVAR) and the focus has been mainly on forecasting.[3]

The second approach to mitigating the curse of dimensionality is to shrink the data, along the lines of index models. Geweke (1977) and Sargent and Sims (1977) introduced dynamic factor models, which have been more recently generalized to allow for weak cross sectional dependence by Forni and Lippi (2001) and Forni *et al.* (2000, 2004). Empirical evidence suggest that few dynamic factors are needed to explain the co-movement of macroeconomic variables: Stock and Watson (1999, 2002), Giannoni, Reichlin and Sala (2005) conclude that only few, perhaps two, factors explain much of the predictable variations, while Stock and Watson (2005) estimate as much as seven factors. This has led to the development of factor augmented VAR (FAVAR) models by Bernanke, Boivin, and Eliasz (2005) and Stock and Watson (2005), among others.

This paper proposes to deal with the curse of dimensionality by shrinking the parameter space in *the limit* as the number of endogenous variables (N) tends to infinity. Under this set up, the infinite-dimensional VAR (or IVAR) could be arbitrarily well approximated by a set of finite-dimensional small-scale models that can be consistently estimated separately, which is in the spirit of global VAR (GVAR) models proposed in Pesaran, Schuermann and Weiner (2004, PSW). By imposing restrictions on the parameters of IVAR model that are binding only in the limit, we effectively end up with shrinking of the data. The paper thus provides a link between the two existing approaches to mitigating the curse of dimensionality in the literature and discusses the conditions under which it is appropriate to shrink the data using static or dynamic factor approaches. This also provides theoretical justification for factor models in a large systems with all variables being determined endogenously. We link our analysis to dynamic factor models by showing that dominant unit becomes (in the limit) a dynamic common factor for the remaining units in the system. Static factor models are also obtained as a special case of IVAR. In addition to the limiting restrictions proposed in this paper, other exact or Bayesian type restrictions can also be easily imposed.

One of the main motivations behind the proposed approach is to develop a theoretical econometric underpinning for global macroeconomic modelling that allows for variety of channels through which the

[1] Other types of priors have also been considered in the literature. See, for example, Del Negro and Schorfheide (2004) for a recent reference.

[2] Few exceptions include Giacomini and White (2006) and De Mol, Giannone and Reichlin (2006).

[3] Bayesian VARs are known to produce better forecasts than unrestricted VARs and, in many situations, ARIMA or structural models. See Litterman (1986), and Canova (1995) for further references.

individual economies are inter-related. The main problem is how to model a large number of endogenously related macroeconomic variables at regional, national, or global levels. Assuming coefficients corresponding to foreign countries in country specific equations are at most of order $O\left(N^{-1}\right)$ (possibly with the exception of a few dominant economies), enables us to consider asymptotic properties of IVARs.[4] This analysis is closely related to the GVAR model, originally developed by PSW and further extended by Dees, di Mauro, Pesaran and Smith (2007).[5] We show that the GVAR approach can be motivated as an approximation to an infinite-dimensional VAR featuring all macroeconomic variables. This is true for both, stationary systems as well as systems with variables integrated of order one, $I(1)$ for short.[6]

The main contribution of the paper is the development of an econometric approach for the analysis of groups that belong to a large interrelated system. It is potentially applicable to the analysis of large spatiotemporal data sets and networks, both with and without dynamic linkages. It is established that under certain granularity conditions on the coefficient matrices of IVAR and in the absence of common factors, in the limit as $N \to \infty$, cross sectional units de-couple from one another. This result does not, however, hold if there are dominant units or unobserved common factors. In such cases unit-specific VARs, conditioned on asymptotically exogenous cross sectional averages, need to be estimated. This is in line with the GVAR approach of PSW and formally establishes the conditions under which the GVAR approach is likely to work. We also consider estimation and inference in a stationary IVAR with an unknown number of unobserved common factors. Simple cross section augmented least squares estimator (or CALS for short) is proposed and its asymptotic distribution is derived. Small sample properties of the proposed estimator are investigated by Monte Carlo experiments, and an empirical application to real GDP growth and investment-output ratio across 98 countries over the period 1961-2003 is provided as an illustration of the proposed approach.

The remainder of this paper is organized as follows. Section 2 investigates cross sectional dependence in IVAR models where key asymptotic results are provided. Section 3 focusses on estimation of a stationary IVAR with (possibly) a fixed but unknown number of unobserved common factors. Section 4 presents Monte Carlo evidence. An empirical application to modelling the interactions of real GDP growth and investment-output ratios in the world economy is presented in Section 5. The final section offers some concluding remarks.

A brief word on notation: $\lambda_1(\mathbf{A}) \geq ... \geq \lambda_n(\mathbf{A})$ are the eigenvalues of $\mathbf{A} \in \mathbb{M}^{n\times n}$, where $\mathbb{M}^{n\times n}$ is the space of real-valued $n \times n$ matrices. $\left\Vert \mathbf{A}\right\Vert_c \equiv \max\limits_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|$ denotes the maximum absolute column sum matrix norm of $\mathbf{A}$, $\left\Vert \mathbf{A}\right\Vert_r \equiv \max\limits_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$ is the absolute row-sum matrix norm of $\mathbf{A}$.[7] $\left\Vert \mathbf{A}\right\Vert = \sqrt{\varrho\left(\mathbf{A}'\mathbf{A}\right)}$ is the spectral norm of $\mathbf{A}$, $\varrho\left(\mathbf{A}\right) \equiv \max\limits_{1\leq i\leq n} \left\{\left|\lambda_i\left(\mathbf{A}\right)\right|\right\}$ is the spectral radius of $\mathbf{A}$.[8] $\mathbf{A}^+$ is the Moore-Penrose inverse of $\mathbf{A}$. Row i of $\mathbf{A}$ is denoted by $\mathbf{a}_i'$ and the column i is denoted as $\mathring{\mathbf{a}}_i$. All vectors are column vectors. Row i of $\mathbf{A}$ with the i^{th} element replaced by 0 is denoted as $\mathbf{a}_{-i}'$. Row i of $\mathbf{A} \in \mathbb{M}^{n\times n}$ with the element i and

[4] See Chudik (2007b) for a theoretical micro-founded N-country DSGE model, where the orders of magnitudes of the coefficients in the equilibrium solution, as well as in the canonical system of linear rational expectation equation characterizing the equilibrium solution, are investigated.

[5] GVAR model has been used to analyse credit risk in Pesaran, Schuermann, Treutler and Weiner (2006) and Pesaran, Schuerman and Treutler (2007). Extended and updated version of the GVAR by Dees, di Mauro, Pesaran and Smith (2007), which treats Euro area as a single economic area, was used by Pesaran, Smith and Smith (2007) to evaluate UK entry into the Euro. Further developments of a global modelling approach are provided in Pesaran and Smith (2006). Garratt, Lee, Pesaran and Shin (2006) provide a textbook treatment of GVAR.

[6] An IVAR model featuring $I(1)$ variables is considered in a supplement available from the authors upon request. Further results for the IVARs with unit roots are also provided in Chudik (2007a).

[7] The maximum absolute column sum matrix norm and the maximum absolute row sum matrix norm are sometimes denoted in the literature as $\left\Vert\cdot\right\Vert_1$ and $\left\Vert\cdot\right\Vert_\infty$, respectively.

[8] Note that if $\mathbf{x}$ is a vector, than $\left\Vert\mathbf{x}\right\Vert = \sqrt{\varrho\left(\mathbf{x}'\mathbf{x}\right)} = \sqrt{\mathbf{x}'\mathbf{x}}$ corresponds to the Euclidean length of vector $\mathbf{x}$.

the element 1 replaced by 0 is $\mathbf{a}'_{-1,-i} = (0, a_{i2}, ..., a_{i,i-1}, 0, a_{i,i+1}, ..., a_{i,iN})$. Matrix constructed from $\mathbf{A}$ by replacing the first column vector by zero vector is denoted as $\dot{\mathbf{A}}_{-1}$. $\|x_t\|_{Lp}$ is L_p-norm of a random variable x_t. Joint asymptotics in $N, T \to \infty$ are denoted by $N, T \overset{j}{\to} \infty$. $a_n = O(b_n)$ states the deterministic sequence a_n is at most of order b_n. $x_n = O_p(y_n)$ states random variable x_n is at most of order y_n in probability. Similarly, we use little-o notation for deterministic sequences, $a_n = o(b_n)$ states a_n is of order less than b_n, and for random variables, $x_n = o_p(y_n)$ states x_n is of order less then y_n in probability. $\mathbb{R}$ is the set of real numbers, $\mathbb{N}$ is the set of natural numbers, and $\mathbb{Z}$ is the set of integers. Convergence in distribution and convergence in probability is denoted by $\overset{d}{\to}$ and $\overset{p}{\to}$, respectively. Symbol $\overset{q.m.}{\to}$ represents convergence in quadratic mean and $\overset{L_1}{\to}$ stands for convergence in L_1 norm.

2 Cross Sectional Dependence in IVAR Models

Suppose there are N cross section units indexed by $i \in \mathcal{S} \equiv \{1, .., N\} \subseteq \mathbb{N}$. Depending on empirical application, units could be households, firms, regions, countries, or macroeconomic indicators in a given economy. Let x_{it} denote the realization of a random variable belonging to the cross section unit i in period t, and assume that $\mathbf{x}_t = (x_{1t}, ..., x_{Nt})'$ is generated according to the following VAR(1) model:

$$\mathbf{x}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{u}_t, \tag{1}$$

where $\mathbf{\Phi}$ is $N \times N$ dimensional matrix of unknown coefficients, and $\mathbf{u}_t = (u_{1t}, ..., u_{Nt})'$ is the $N \times 1$ vector of disturbances. We refer to this model as infinite-dimensional VAR (denoted as IVAR) in cases where N is relatively large. Initially, we shall assume that the process has started from a finite past, $t \in \mathcal{T} \equiv \{-M+1, .., 0, ..\} \subseteq \mathbb{Z}$, M being a fixed positive integer. This assumption is relaxed in Subsection 2.1 for stationary models. Extension of the analysis to IVAR(p) models is relegated to Appendix B.

The objective is to study the correlation pattern of a double index process $\{x_{it}, i \in \mathcal{S}, t \in \mathcal{T}\}$, given by the IVAR model (1), over cross section units at different points in time, $t \in \mathcal{T}$. Unlike the time index t which is defined over an ordered integer set, the cross section index, i, refers to an individual unit of an unordered population distributed over space or more generally networks. Pesaran and Tosetti (PT, 2007, Definition 4) define process $\{x_{it}\}$ to be cross sectionally weakly dependent (CWD), if for all $t \in \mathcal{T}$ and for all weight vectors, $\mathbf{w}_t = (w_{1t}, ..., w_{Nt})'$, satisfying the following 'granularity' conditions

$$\|\mathbf{w}_t\| = O\left(N^{-\frac{1}{2}}\right), \tag{2}$$

$$\frac{w_{jt}}{\|\mathbf{w}_t\|} = O\left(N^{-\frac{1}{2}}\right) \text{ for any } j \leq N, \tag{3}$$

we have

$$\lim_{N \to \infty} Var\left(\mathbf{w}'_{t-1}\mathbf{x}_t \mid \mathcal{I}_{t-1}\right) = 0, \tag{4}$$

where $\mathcal{I}_t$ is the information set at time t. The concept of CWD can be articulated with respect to the conditional as well as the unconditional variance of weighted averages, if it exists. PT consider $\mathcal{I}_t$ containing at least $\mathbf{x}_t, \mathbf{x}_{t-1}, ...$ and $\mathbf{w}_t, \mathbf{w}_{t-1}, ...$.[9] For simplicity of exposition and without the loss of generality, time invariant non-random vector of weights $\mathbf{w}$ satisfying granularity conditions (2)-(3) is considered in this paper

[9] Note that in the context of IVAR model (1), if $\mathcal{I}_t$ contains at least $\mathbf{x}_t, \mathbf{x}_{t-1}, ...$ and $\mathbf{w}_t, \mathbf{w}_{t-1}, ...$, then

$$Var\left(\mathbf{w}'_{t-1}\mathbf{x}_t \mid \mathcal{I}_{t-1}\right) = Var\left(\mathbf{w}'_{t-1}\mathbf{u}_t \mid \mathcal{I}_{t-1}\right), \tag{5}$$

since $Var\left(\mathbf{w}'_{t-1}\mathbf{\Phi}\mathbf{x}_{t-1} \mid \mathcal{I}_{t-1}\right) = 0$ regardless of the coefficients matrix $\mathbf{\Phi}$. In this case, the process $\{x_{it}\}$ is CWD if and only if the errors $\{u_{it}\}$ are CWD.

only. Furthermore, in the context of a purely dynamic model such as the IVAR model (1) the concept of CWD is more meaningfully defined with respect to the initial values. Hence, unless otherwise stated in our analysis of cross section dependence, we take $\mathcal{I}_t$ to contain only the starting values, $\mathbf{x}(-M)$.[10]

Definition 1 *Dynamic double index process $\{x_{it}, i \in \mathcal{S}, t \in \mathcal{T}\}$, generated by the IVAR model (1), is said to be cross sectionally weakly dependent (CWD) at a point in time $t \in \mathcal{T}$, if for any non-random vector of weights $\mathbf{w}$ satisfying the granularity conditions (2)-(3),*

$$\lim_{N\to\infty} Var\left[\mathbf{w}'\mathbf{x}_t \mid \mathbf{x}(-M)\right] = 0 \tag{6}$$

We say that the process $\{x_{it}\}$ is cross sectionally strongly dependent (CSD) at a given point in time $t \in \mathcal{T}$ if there exists a weights vector $\mathbf{w}$ satisfying (2)-(3) and a constant K independent of N such that for any N sufficiently large,

$$Var\left[\mathbf{w}'\mathbf{x}_t \mid \mathbf{x}(-M)\right] \geq K > 0. \tag{7}$$

Consider now the following assumptions on the coefficient matrix, $\mathbf{\Phi}$, and the error vector, $\mathbf{u}_t$:

ASSUMPTION 1 *Individual elements of double index process of errors $\{u_{it}, i \in \mathcal{S}, t \in \mathcal{T}\}$ are random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. $\mathbf{u}_t$ is independently distributed of the starting values, $\mathbf{x}(-M)$, and of $\mathbf{u}_{t'}$, for any $t \neq t' \in \mathcal{T}$. For each $t \in \mathcal{T}$, $\mathbf{u}_t$ has mean and variance,*

$$\begin{aligned} E\left[\mathbf{u}_t \mid \mathbf{x}(-M)\right] &= E(\mathbf{u}_t) = \mathbf{0}, \quad (8) \\ E\left[\mathbf{u}_t\mathbf{u}_t' \mid \mathbf{x}(-M)\right] &= E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{\Sigma}_t, \quad (9) \end{aligned}$$

where $\mathbf{\Sigma}_t$ is an $N \times N$ symmetric, nonnegative definite matrix, with generic $(i,j)^{th}$ element $\sigma_{ij,t}$ and such that $0 < \sigma_{ii,t} < K < \infty$ for any $i \in \mathcal{S}$, where the constant K does not depend on N.

ASSUMPTION 2 *(Coefficients matrix $\mathbf{\Phi}$ and CWD $\mathbf{u}_t$)*

$$\left\|\mathbf{\Phi}\right\|_c \left\|\mathbf{\Phi}\right\|_r = O(1), \tag{10}$$

and

$$\varrho(\mathbf{\Sigma}_t) = o(N) \text{ for any } t \in \mathcal{T}. \tag{11}$$

Remark 1 *Assumption 1 and equation (11) of Assumption 2 imply $\{u_{it}\}$ is CWD. The initialization of a dynamic process could be from a non-stochastic point or could have been from a stochastic point, possibly generated from a process different from the DGP.*

Proposition 1 *Consider model (1) and suppose Assumptions 1-2 hold. Then for any arbitrary fixed weights $\mathbf{w}$ satisfying condition (2), and for any $t \in \mathcal{T}$,*

$$\lim_{N\to\infty} Var\left[\mathbf{w}'\mathbf{x}_t \mid \mathbf{x}(-M)\right] = 0 \tag{12}$$

Proof. The vector difference equation (1) can be solved backwards from $t = -M$, taking $\mathbf{x}(-M)$ as given:

$$\mathbf{x}_t = \mathbf{\Phi}^{t+M}\mathbf{x}(-M) + \sum_{\ell=0}^{t+M-1} \mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell} \tag{13}$$

[10] Instead of using $\mathbf{x}_{-M}$ to denote the vector of starting values, we use $\mathbf{x}(-M)$ in order to avoid possible confusion with the developed notation in Section 1.

The variance of $\mathbf{x}_t$ (conditional on initial values) is

$$\mathbf{\Omega}_t = Var\left[\mathbf{x}_t \mid \mathbf{x}(-M)\right] = \sum_{\ell=0}^{t+M-1} \mathbf{\Phi}^{\ell}\mathbf{\Sigma}_{t-\ell}\mathbf{\Phi}'^{\ell}, \tag{14}$$

and, using the Rayleigh-Ritz theorem,[11] $\left\|\mathbf{\Omega}_t\right\|_c$ is under Assumptions 1-2 bounded by

$$\left\|\mathbf{\Omega}_t\right\|_c \leq \sum_{\ell=0}^{t+M-1} \varrho\left(\mathbf{\Sigma}_{t-\ell}\right)\left(\left\|\mathbf{\Phi}\right\|_c \left\|\mathbf{\Phi}\right\|_r\right)^{\ell} = o\left(N\right) \tag{15}$$

It follows that for any arbitrary nonrandom vector of weights satisfying granularity condition (2),

$$\left\|Var\left[\mathbf{w}'\mathbf{x}_t \mid \mathbf{x}(-M)\right]\right\|_c = \left\|\mathbf{w}'\mathbf{\Omega}_t\mathbf{w}\right\|_c \leq \left\|\varrho\left(\mathbf{\Omega}_t\right)\left(\mathbf{w}'\mathbf{w}\right)\right\|_c = o\left(1\right) \tag{16}$$

where $\varrho\left(\mathbf{\Omega}_t\right) \leq \left\|\mathbf{\Omega}_t\right\|_c = o\left(N\right)$,[12] and $\mathbf{w}'\mathbf{w} = \left\|\mathbf{w}\right\|^2 = O\left(N^{-1}\right)$ by condition (2). ■

There are several interesting implications of Proposition 1. Consider following additional assumption on coefficients matrix $\mathbf{\Phi}$.

ASSUMPTION 3 *Let $\mathcal{K} \subseteq \mathcal{S}$ be a non-empty index set. Define vector $\boldsymbol{\phi}_{-i} = \left(\phi_{i1}, ..., \phi_{i,i-1}, 0, \phi_{i,i+1}, ..., \phi_{iN}\right)'$ where ϕ_{ij} for $i, j \in \mathcal{S}$ is the (i, j) element of matrix $\mathbf{\Phi}$. For any $i \in \mathcal{K}$, vector $\boldsymbol{\phi}_{-i}$ satisfies*

$$\left\|\boldsymbol{\phi}_{-i}\right\| = \left(\sum_{j=1, j\neq i}^{N} \phi_{ij}^2\right)^{1/2} = O\left(N^{-\frac{1}{2}}\right). \tag{17}$$

Remark 2 *An example of matrix $\mathbf{\Phi}$ that satisfy Assumptions 2 and 3 is*[13]

$$\phi_{ii} = O\left(1\right), \quad \left\|\boldsymbol{\phi}_{-i}\right\|_r = O\left(\frac{1}{N}\right) \text{ for any } i \in \mathcal{S}. \tag{18}$$

Namely, when the off-diagonal elements of $\mathbf{\Phi}$ are of order $O\left(N^{-1}\right)$. This special case is considered in Section 3 where we turn our attention to the problem of estimating infinite-dimensional VARs.

Remark 3 *Assumption 3 implies that for $i \in \mathcal{K}$, $\sum_{i=1, i\neq j}^{N} \phi_{ij} \leq \left\|\boldsymbol{\phi}_{-i}\right\|_c = O\left(1\right)$.*[14] *Therefore, it is possible for the dependence of each individual unit on the rest of the units in the system to be large. However, as we shall see below, in the case where $\{x_{it}\}$ is a CWD process, the model for the i^{th} cross section unit de-couples from the rest of the system as $N \to \infty$.*

[11] See Horn and Johnson (1985, p. 176)

[12] Spectral radius is lower bound for any matrix norm, see Horn and Johnson (1985, Theorem 5.6.9).

[13] Maximum absolute row sum matrix norm of N dimensional column vector $\mathbf{y} = (y_1, y_2, ..., y_N)'$ is

$$\left\|\mathbf{y}\right\|_r = \left\|\mathbf{y}'\right\|_c = \max_{1\leq j\leq N} \left|y_j\right|.$$

Similarly, maximum absolute column sum matrix norm of N dimensional column vector $\mathbf{y} = (y_1, y_2, ..., y_N)'$ is

$$\left\|\mathbf{y}\right\|_c = \left\|\mathbf{y}'\right\|_r = \sum_{j=1}^{N} \left|y_j\right|.$$

[14] Note that $\left\|\boldsymbol{\phi}_{-i}\right\|_c \leq \sqrt{N}\left\|\boldsymbol{\phi}_{-i}\right\|$. See Horn and Johnson (1985, p. 314). An example of vector $\boldsymbol{\phi}_{-i}$ for which $\lim_{N\to\infty} \sum_{i=1, i\neq j}^{N} \phi_{ij} \neq 0$ is when $\phi_{ij} = k/N$ for $i \neq j$ and any fixed non-zero constant k.

Corollary 1 *Consider model (1) and suppose Assumptions 1-3 hold. Then,*

$$\lim_{N\to\infty} Var\left[x_{it} - \phi_{ii}x_{i,t-1} - u_{it} \mid \mathbf{x}(-M)\right] = 0 \text{ for } i \in \mathcal{K}. \tag{19}$$

Proof. Assumption 3 implies that for $i \in \mathcal{K}$, vector $\boldsymbol{\phi}_{-i}$ satisfies condition (2). It follows from Proposition 1 that

$$\lim_{N\to\infty} Var\left[\boldsymbol{\phi}'_{-i}\mathbf{x}_t \mid \mathbf{x}(-M)\right] = 0 \text{ for } i \in \mathcal{K}. \tag{20}$$

System (1) implies

$$x_{it} - \phi_{ii}x_{i,t-1} - u_{it} = \boldsymbol{\phi}'_{-i}\mathbf{x}_t, \text{ for any } i \in \mathcal{S}. \tag{21}$$

Taking conditional variance of (21) and using (20) now yields (19). ■

2.1 Stationary Conditions for IVAR(1)

Conditions under which VAR model (1) (for a fixed N) is stationary are well known in the literature, see for instance Hamilton (1994, Chapter 10).

ASSUMPTION 4 *(Necessary condition for covariance stationarity) All eigenvalues of $\mathbf{\Phi}$, defined by λ that satisfy the equation $|\mathbf{\Phi}-\lambda\mathbf{I}_N| = 0$, lie inside of the unit circle. Furthermore, $\mathbf{\Sigma}_t = \mathbf{\Sigma}$ is time invariant.*

For a fixed N and assuming $\left\|Var\left[\mathbf{x}(-M)\right]\right\|_r$ is bounded, $\mathbf{x}_t$ converges under Assumption 4 in mean squared errors to $\sum_{j=0}^{\infty}\mathbf{\Phi}^j\mathbf{u}_{t-j}$ as $M \to \infty$, namely by assuming that the process has been in operation for sufficiently long time before its first observed realization, $\mathbf{x}_1$. Also for a finite N, when all eigenvalues of $\mathbf{\Phi}$ lie inside the unit circle, the Euclidean norm of $\mathbf{\Phi}$ defined by $\left[Tr\left(\mathbf{\Phi}^j\mathbf{\Phi}^{j\prime}\right)\right]^{1/2} \to 0$ exponentially in j and the process $\mathbf{x}_t = \sum_{j=0}^{\infty}\mathbf{\Phi}^j\mathbf{u}_{t-j}$ will be absolute summable, in the sense that the sum of absolute values of the elements of $\mathbf{\Phi}^j$, for $j = 0, 1, ...$ converge. It is then easily seen that $\mathbf{x}_t$ will have finite moments of order ℓ, assuming $\mathbf{u}_t$ has finite moments of the same order. In particular, under Assumptions 1, 4, and $\left\|Var\left[\mathbf{x}(-M)\right]\right\|_r < K$, as $M \to \infty$,

$$E(\mathbf{x}_t) = \mathbf{0}, \text{ and } \mathbf{\Omega} = Var(\mathbf{x}_t) = \lim_{M\to\infty}\mathbf{\Omega}_t = \sum_{j=0}^{\infty}\mathbf{\Phi}^j\mathbf{\Sigma}\mathbf{\Phi}'^j < \infty. \tag{22}$$

In the stationary case with $M \to \infty$, at any point in time t

$$\left\|Var(\mathbf{x}_t)\right\|_c = \left\|\sum_{j=0}^{\infty}\mathbf{\Phi}^j\mathbf{\Sigma}\mathbf{\Phi}'^j\right\|_c \leq \varrho(\mathbf{\Sigma})\sum_{j=0}^{\infty}\left\|\mathbf{\Phi}\right\|_c^j\left\|\mathbf{\Phi}\right\|_r^j, \tag{23}$$

Observe that under Assumption 4, $Var(x_{it})$ and $\left\|Var(\mathbf{x}_t)\right\|_c$ need not necessarily be bounded as $N \to \infty$, even if $\varrho(\mathbf{\Sigma}) = O(1)$.[15]

[15] For example, consider the IVAR(1) model with

$$\mathbf{\Phi} = \begin{pmatrix} 0.5 & 1 & 0 & \cdots & 0 \\ 0 & 0.5 & 1 & & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & & 0.5 & 1 \\ 0 & 0 & & 0 & 0.5 \end{pmatrix},$$

and assume that $var(u_{it})$ is uniformly bounded away from zero as $N \to \infty$. It is clear that all eigenvalues of $\mathbf{\Phi}$ are inside the unit circle, and in particular $\varrho(\mathbf{\Phi}) < 1 - \epsilon$, where $\epsilon > 0$ does not depend on N. Yet the variance of x_{it} increases in N without bounds.

Proposition 2 *Consider model (1) and suppose Assumptions 1-2, 4 hold, $\|Var[\mathbf{x}(-M)]\|_r < K$ and $M \to \infty$. Further assume that $\|\mathbf{\Phi}\|_c \|\mathbf{\Phi}\|_r < 1-\epsilon$ where $\epsilon > 0$ is arbitrarily small and does not depend on N. Then at any point in time $t \in T$ and for any set of weights, $\mathbf{w}$, satisfying condition (2),*

$$\lim_{N\to\infty} Var(\mathbf{w}'\mathbf{x}_t) = 0. \tag{24}$$

Proof. $\varrho(\mathbf{\Sigma}) = o(N)$ under Assumptions 2, and 4. Since $\|\mathbf{\Phi}\|_c \|\mathbf{\Phi}\|_r < 1-\epsilon$, $\sum_{j=0}^{\infty} \|\mathbf{\Phi}\|_c^j \|\mathbf{\Phi}\|_r^j = O(1)$. It follows from (23) that $\|Var(\mathbf{x}_t)\|_c = o(N)$. This establishes $\|Var(\mathbf{w}'\mathbf{x}_t)\|_c \leq \|\mathbf{w}'\mathbf{\Omega}\mathbf{w}\|_c = o(1)$, along similar lines used in establishing equation (16). ■

Hence, irrespective of whether the process is stationary and started in a finite or a very distance past, or it is non-stationary (including unit root or explosive) and has started from a finite past, the $\{x_{it}\}$ is CWD, if the conditions of Proposition 1 or 2 are satisfied. Of course, if the process is non-stationary $Var[\mathbf{x}_t \mid \mathbf{x}(-M)]$ exists only for $(t+M)$ finite.

2.2 Strong Dependence in IVAR models and Dominant Effects

PT introduced the concept of dominant effects. Assume covariance matrix $\mathbf{\Omega}_t = (\omega_{ijt})$ has m dominant columns, which, as defined in PT (2007, definition 14), means[16]

$$\begin{aligned} i)\ \sum_{i=1}^{N} |\omega_{ijt}| &= O(N) \text{ for } j \in \mathcal{S}(m) \\ ii)\ \sum_{i=1}^{N} |\omega_{ijt}| &= O(1) \text{ for } j \in \mathcal{S} \cap \mathcal{S}(m)^c \end{aligned}$$

where ω_{ijt} denotes the $(i,j)^{th}$ element of the covariance matrix $\mathbf{\Omega}_t$, and $\mathcal{S}(m)$ is m-dimensional subset of the set $\mathcal{S}$.[17] Observe that if the sum of the absolute elements of the column of $\mathbf{\Omega}_t$ corresponding to the individual unit j increases with N at the rate N, then the unit j has a strong relationship with all other units and thus the process $\{x_{it}\}$ is CSD.

Dominant effects could, for example, arise in the context of global macroeconomic modelling. Suppose that the equilibrium of a theoretical N-country macroeconomic DSGE model of the world economy is described by the following canonical system of the linear rational expectation equation

$$\mathbf{A}_0\mathbf{x}_t = \mathbf{A}_1 E_t \mathbf{x}_{t+1} + \mathbf{A}_2 \mathbf{x}_{t-1} + \boldsymbol{v}_t, \tag{25}$$

where $\mathbf{A}_0$, $\mathbf{A}_1$and $\mathbf{A}_2$ are $N \times N$ dimensional matrices of coefficients, and $\boldsymbol{v}_t = (v_{1t}, v_{2t}, ..., v_{Nt})'$ is a vector of serially uncorrelated (but in general cross sectionally correlated) structural disturbances. One lag and one endogenous variable per economy is assumed for the simplicity of exposition. Assuming that the set of monetary policy rules is such that there is a unique determinate equilibrium, then for a given N, the equilibrium solution has a VAR(1) representation with the reduced form given by the VAR model (1).[18] Strong cross sectional dependence could, for example, arise when the one or more absolute column sums of $\mathbf{A}_0$ in (25), are not bounded in N. Foreign trade (export and import) shares are one of the main determinants of the off-diagonal elements of $\mathbf{A}_0$ in a multicountry open economy DSGE model.[19] Examining the column-

[16] Clearly if $\mathbf{\Omega}_t$ has m dominant collumns, so does $\mathbf{\Omega}_t'$, that is $\mathbf{\Omega}_t$ has m dominant rows.

[17] If $\mathbf{\Omega}_t$ has m dominant columns, then by Theorem 15 of PT, first m largest eigenvalues of $\mathbf{\Omega}_t$ are $O(N)$ in general. By the same theorem, if first m largest eigenvalues of $\mathbf{\Omega}_t$ are $O(N)$, then $\mathbf{\Omega}_t$ has at least m dominant columns.

[18] The unique solution can be obtained using, for example, the results in Binder and Pesaran (1997).

[19] See, for example, Chudik (2007b) for details.

sum of the foreign trade share matrix provides additional useful information, besides the level of output, on the position of a country in the global economy. Not surprisingly, this column sum of the foreign trade share matrix is the largest for the US.[20] Figure 1 plots first six countries with the largest column-sums of the foreign trade-share matrix based on foreign trade data in 2006 for 181 countries in IMF DOTS database. The rising role of China in the global economy is nicely documented.

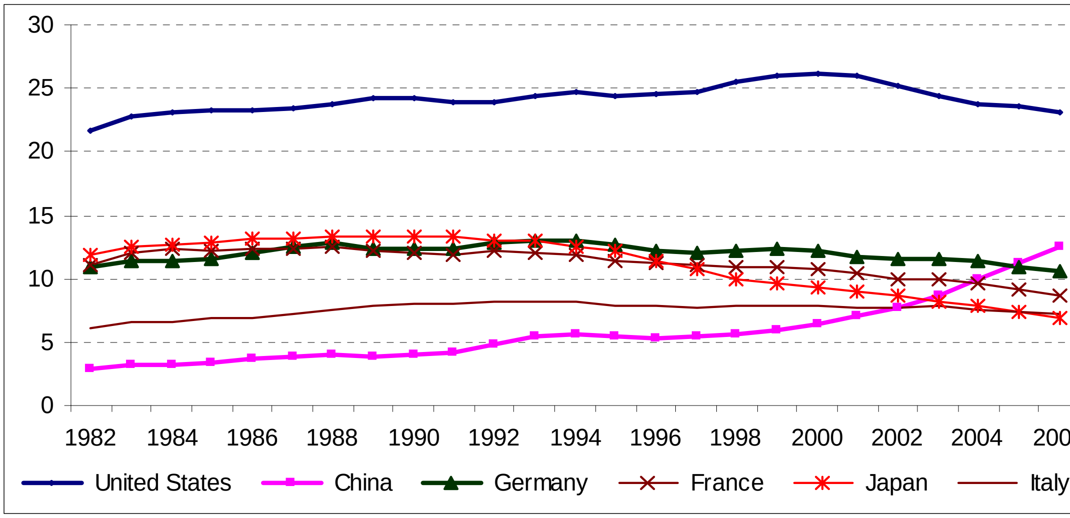

Figure 1: Column-sums of trade share matrix for the top 6 countries as of 2006 using three-year moving averages

Strong dependence in IVAR model (1) could arise as a result of CSD errors $\{u_{it}\}$, or could be due to dominant patterns in the coefficients of $\mathbf{\Phi}$, or both. Residual common factor models, where weighted averages of factor loadings do not converge to zero, is an example of CSD process $\{u_{it}\}$, see PT (2007, Theorem 16). Section 3 considers estimation and inference in a stationary CSD IVAR model where deviations of endogenous variables from unobserved common factors and deterministic trend follow an IVAR. An example of CSD IVAR model, where the maximum absolute column sum norm of matrix $\mathbf{\Phi}$ and/or matrix $\mathbf{\Sigma}$ is unbounded, is provided in the following subsection. The supplement available form the authors presents CSD IVAR model featuring both unobserved common factor f_t and coefficient matrices with unbounded maximum absolute column sum matrix norms, and $I(1)$ endogenous variables.

2.3 Contemporaneous Dependence: Spatial or Network Dependence

An important form of cross section dependence is *contemporaneous* dependence across space. The spatial dependence, pioneered by Whittle (1954), models cross section correlations by means of spatial processes that relate each cross section unit to its neighbour(s). Spatial autoregressive and spatial error component models are examples of such processes. (Cliff and Ord, 1973, Anselin, 1988, and Kelejian and Robinson, 1995). However, it is not necessary that proximity is measured in terms of physical space. Other measures such as economic (Conley, 1999, Pesaran, Schuermann and Weiner, 2004), or social distance (Conley and Topa, 2002) could also be employed. All these are examples of dependence across nodes in a physical (real) or logical (virtual) networks. In the case of the IVAR model, (1), such contemporaneous dependence can be modelled through an $N \times N$ network topology matrix $\mathbf{R}$ so that[21]

[20] Element (i, j) of foreign trade share matrix is constructed as the ratio of the sum of nominal exports from country i to country j and nominal imports from country j to country i on the aggregate foreign trade of country i (i.e. the sum of aggregate nominal exports and imports). Therefore, the row-sum of any row of trade share matrix is, by construction, equal one.

[21] A network topography is usually represented by graphs whose nodes are identified with the cross section units, with the pairwise relations captured by the arcs in the graph.

$$\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t, \tag{26}$$

where $\boldsymbol{\varepsilon}_t$ are $IID\left(\mathbf{0}, \mathbf{I}_N\right)$, $\boldsymbol{\varepsilon}_t$ is independent of $\boldsymbol{\varepsilon}_{t'}$ for all $t \neq t' \in \mathcal{T}$.[22] For example, in the case of a first order spatial moving average model, $\mathbf{R}$ would take the form

$$\mathbf{R}_{SMA} = \mathbf{I}_N + \rho_s \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 & 0 \end{pmatrix},$$

where ρ_s is the spatial moving average coefficient.

The contemporaneous nature of dependence across $i \in \mathcal{S}$ is fully captured by $\mathbf{R}$. As shown in PT the contemporaneous dependence across $i \in \mathcal{S}$ will be weak if the maximum absolute column and row sum matrix norm of $\mathbf{R}$ are bounded, namely if $\left\|\mathbf{R}\right\|_c \left\|\mathbf{R}\right\|_r < K < \infty$. It turns out that all spatial models proposed in the literature are examples of weak cross section dependence. More general network dependence such as the 'star' network provides an example of strong contemporaneous dependence that we shall consider below. The form of $\mathbf{R}$ for a typical star network is given by

$$\mathbf{R}_{Star} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ r_{21} & 1 & \cdots & 0 & 0 \\ r_{31} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & 1 & 0 \\ r_{N1} & 0 & \cdots & 0 & 1 \end{pmatrix},$$

where $\sum_{j=2}^{N} r_{j1} = O(N)$.

The IVAR model when combined with $\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t$ yields an infinite-dimensional spatiotemporal model. The model can also be viewed more generally as a 'dynamic network', with $\mathbf{R}$ and $\mathbf{\Phi}$ capturing the static and dynamic forms of inter-connections that might exist in the network.

2.3.1 IVAR Models with Strong Cross Sectional Dependence

Strong cross dependence could arise when matrix $\mathbf{\Phi}$ and/or $\mathbf{R}$ have dominant columns. In this subsection, we present a stationary IVAR model where the column corresponding to unit $i = 1$ in matrices $\mathbf{\Phi}$, and $\mathbf{R}$ is dominant (e.g. US in the world economy). In particular, consider the following assumptions.

ASSUMPTION 5 *Let* $\mathbf{\Phi} = \sum_{i=1}^{N} \mathring{\boldsymbol{\phi}}_i \mathbf{s}_{iN}' = \mathring{\boldsymbol{\phi}}_1 \mathbf{s}_{1N}' + \dot{\mathbf{\Phi}}_{-1}$ *where* $\mathring{\boldsymbol{\phi}}_i = \left(\phi_{1i}, ..., \phi_{Ni}\right)'$ *is the* i^{th} *column of matrix* $\mathbf{\Phi}$, $\mathbf{s}_{iN}$ *is an* $N \times 1$ *selection vector for unit i, with the* i^{th} *element of* $\mathbf{s}_{iN}$ *being one and the remaining zero.*[23] *Denote by* $\dot{\mathbf{\Phi}}_{-1}$ *the matrix constructed from* $\mathbf{\Phi}$ *by replacing its first column with a vector of zeros,*

[22] It is also possible to allow for time variations in the network matrix to capture changes in the network structure over time. However, this will not be pursued here.

[23] Subscript N is used to denote dimension of selection vector, because selection vectors of different dimensions are also used in the exposition below. Except selection vectors, subscripts to denote dimensions are omitted in order to keep the notation simple.

and note that $\dot{\mathbf{\Phi}}_{-1} = \sum_{i=2}^{N} \mathring{\boldsymbol{\phi}}_i \mathbf{s}'_{iN}$. *Suppose as* $N \to \infty$

$$\left\| \mathring{\boldsymbol{\phi}}_1 \right\|_r = O\left(1\right), \tag{27}$$

and

$$\left\| \dot{\mathbf{\Phi}}_{-1} \right\|_c \left\| \dot{\mathbf{\Phi}}_{-1} \right\|_r = O\left(1\right). \tag{28}$$

Further, let $\boldsymbol{\phi}_{-1,-i} = \left(0, \phi_{i2}, ..., \phi_{i,i-1}, 0, \phi_{i,i+1}, ..., \phi_{iN}\right)'$, *where* ϕ_{ij}, $i, j \in \mathcal{S}$, *are elements of matrix* $\mathbf{\Phi}$. *For any* $i \in \mathcal{S}$, *suppose that*

$$\left\| \boldsymbol{\phi}_{-1,-i} \right\|_r = O\left(N^{-1}\right). \tag{29}$$

ASSUMPTION 6 *(Stationarity) Process (1) starts from an infinite past (*$M \to \infty$*), and* $\left\| \mathbf{\Phi} \right\|_r < \rho < 1$.

ASSUMPTION 7 *The* $N \times 1$ *vector of errors* $\mathbf{u}_t$ *is generated by the 'spatial' model (26).* $\mathbf{\Sigma}_t = \mathbf{\Sigma} = \mathbf{R}\mathbf{R}'$ *is time invariant, where* $\mathbf{R} = \sum_{i=1}^{N} \mathring{\mathbf{r}}_i \mathbf{s}'_{iN} = \mathring{\mathbf{r}}_1 \mathbf{s}'_{1N} + \dot{\mathbf{R}}_{-1}$, *and* $\mathring{\mathbf{r}}_i = (r_{1i}, ..., r_{Ni})'$ *is the* i^{th} *column of matrix* $\mathbf{R}$. *Let*

$$\left\| \dot{\mathbf{R}}_{-1} \right\|_c \left\| \dot{\mathbf{R}}_{-1} \right\|_r = o\left(N\right), \tag{30}$$

$$\left\| \mathring{\mathbf{r}}_1 \right\|_r = O\left(1\right), \tag{31}$$

and

$$\left\| \mathbf{r}_{-1,-i} \right\| = o\left(1\right) \text{ for any } i \in \mathcal{S}, \tag{32}$$

where $\mathbf{r}_{-1,-i} = (0, r_{i1}, ..., r_{i,i-1}, 0, r_{i,i+1}, ..., r_{iN})'$ *and* r_{ij} *is generic* $(i,j)^{th}$ *element of matrix* $\mathbf{R}$.

Remark 4 *Assumptions 5 and 7 imply matrix* $\mathbf{\Phi}$ *has one dominant column and matrix* $\mathbf{R}$ *has at least one dominant column, but the absolute column sum for only one column could rise with* N *at the rate* N. *Equation (29) of Assumption 5 allows the equation for unit* i *to de-couple from units* $j > 1$, *for* $j \neq i$, *as* $N \to \infty$.

Remark 5 *Using the maximum absolute column/row sum matrix norms rather than eigenvalues allows us to make a distinction between cases where dominant effects are due to a particular unit (or a few units), and when there is a pervasive unobserved factor that makes all column/row sums unbounded. Eigenvalues of the covariance matrix* $\mathbf{\Omega}$ *will be unbounded in both cases and it will not be possible from the knowledge of the rate of the expansion of the eigenvalues of the matrix* $\mathbf{\Omega}$, $\mathbf{\Phi}$ *and/or* $\mathbf{R}$ *to known which one of the two cases are in fact applicable.*

Remark 6 *As it will become clear momentarily,* conditional *on* x_{1t} *and its lags, double index process* $\{x_{it}\}$ *become weakly dependent. We shall refer to unit* $i = 1$ *as the dominant unit.*

Remark 7 *Assumptions 6 and 7 imply system (1) is stationary for any* N, *and the variance of* x_{it} *for any* $i \in \mathcal{S}$ *is bounded as* $N \to \infty$.

Proposition 3 *Under Assumptions 5-7 and as* $N \to \infty$, *equation for the dominant unit* $i = 1$ *in the system (1) reduces to*

$$x_{1t} - \vartheta\left(L, \mathbf{s}_{1N}\right) \varepsilon_{1t} \overset{q.m.}{\to} 0, \tag{33}$$

where $\vartheta\left(L, \mathbf{s}_{1N}\right) = \sum_{\ell=0}^{\infty} \left(\mathbf{s}'_{1N} \mathbf{\Phi}^{\ell} \mathring{\mathbf{r}}_1\right) L^{\ell}$. *Furthermore, for any fixed set of weights* $\mathbf{w}$ *satisfying condition (2),*

$$x_t^* - \vartheta\left(L, \mathbf{w}\right) \varepsilon_{1t} \overset{q.m.}{\to} 0, \tag{34}$$

where $\vartheta\left(L, \mathbf{w}\right) = \sum_{\ell=0}^{\infty} \left(\mathbf{w}' \mathbf{\Phi}^{\ell} \mathring{\mathbf{r}}_1\right) L^{\ell}$ *and* $x_t^* = \mathbf{w}' \mathbf{x}_t$.

Proof. Solving (1) backwards yields

$$\mathbf{x}_t = \sum_{\ell=0}^{\infty} \left(\boldsymbol{\Phi}^{\ell}\mathbf{R}\right) \boldsymbol{\varepsilon}_{t-\ell}, \tag{35}$$

where $\mathbf{R} = \mathring{\mathbf{r}}_1 \mathbf{s}'_{1N} + \dot{\mathbf{R}}_{-1}$. Hence

$$x_{1t} - \sum_{\ell=0}^{\infty} \left(\mathbf{s}'_{1N}\boldsymbol{\Phi}^{\ell}\mathring{\mathbf{r}}_1\right) \varepsilon_{1,t-\ell} = \sum_{\ell=0}^{\infty} \left(\mathbf{s}'_{1N}\boldsymbol{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\right) \boldsymbol{\varepsilon}_{t-\ell}. \tag{36}$$

Under Assumptions 5-7,

$$\begin{aligned} \left\| Var\left(\sum_{\ell=0}^{\infty} \left(\mathbf{s}'_{1N}\boldsymbol{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\right) \boldsymbol{\varepsilon}_{t-\ell} \right) \right\|_c &= \left\| \sum_{\ell=0}^{\infty} \mathbf{s}'_{1N}\boldsymbol{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}'_{-1}\boldsymbol{\Phi}'^{\ell}\mathbf{s}_{1N} \right\|_c, \\ &\leq \left\| \mathbf{s}'_{1N}\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}'_{-1}\mathbf{s}_{1N} \right\|_c + \left\| \sum_{\ell=1}^{\infty} \mathbf{s}'_{1N}\boldsymbol{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}'_{-1}\boldsymbol{\Phi}'^{\ell}\mathbf{s}_{1N} \right\|_c. \end{aligned} \tag{37}$$

But

$$\left\| \mathbf{s}'_{1N}\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}'_{-1}\mathbf{s}_{1N} \right\|_c = \left\|\mathbf{r}_{-1}\right\|^2 = o(1), \tag{38}$$

where $\left\|\mathbf{r}_{-1}\right\|^2 = o(1)$ under Assumption 7. Set $\boldsymbol{\nu}'_{\ell} \equiv \mathbf{s}'_{1N}\boldsymbol{\Phi}^{\ell}$ and let $\boldsymbol{\nu}_{\ell,-1} = (0, \nu_{\ell 2}, ..., \nu_{\ell N})'$. Note that under Assumptions 5-6:

$$\begin{aligned} \left\|\boldsymbol{\nu}_{\ell}\right\|_c &\leq \rho^{\ell}, & (39) \\ \nu_{\ell 1} &= 0, & (40) \\ \left\|\boldsymbol{\nu}_{\ell,-1}\right\|_r &= O\left(N^{-1}\right), & (41) \end{aligned}$$

for $\ell = 0, 1, 2, ...$; result (39) follows from Assumption 6 by taking the maximum absolute row-sum matrix norm of $\boldsymbol{\nu}_{\ell} = \mathbf{s}'_{1N}\boldsymbol{\Phi}^{\ell}$,

$$\left\|\boldsymbol{\nu}_{\ell}\right\|_c \leq \left\|\mathbf{s}'_{1N}\boldsymbol{\Phi}\right\|_r \leq \rho^{\ell}. \tag{42}$$

Results (40)-(41) follow by induction directly from Assumptions 5-6. Using (39)-(41), we have

$$\begin{aligned} \left\| \sum_{\ell=1}^{\infty} \mathbf{s}'_{1N}\boldsymbol{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}'_{-1}\boldsymbol{\Phi}'^{\ell}\mathbf{s}_{1N} \right\|_c &= \left\| \sum_{\ell=1}^{\infty} \boldsymbol{\nu}'_{\ell}\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}'_{-1}\boldsymbol{\nu}_{\ell} \right\|_c, \\ &\leq \left\| \sum_{\ell=1}^{\infty} \nu_{\ell 1}^2 \mathbf{r}'_{-1}\mathbf{r}_{-1} \right\|_c + \left\| \sum_{\ell=1}^{\infty} \boldsymbol{\nu}'_{\ell,-1}\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}'_{-1}\boldsymbol{\nu}_{\ell,-1} \right\|_c, \\ &\leq \left\|\mathbf{r}_{-1}\right\|^2 \sum_{\ell=1}^{\infty} \rho^2 + \left\| \dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}'_{-1} \right\|_c \sum_{\ell=1}^{\infty} \left\|\boldsymbol{\nu}_{\ell,-1}\right\|_r \left\|\boldsymbol{\nu}_{\ell,-1}\right\|_c, \\ &= o(1), \end{aligned} \tag{43}$$

where as before $\left\|\mathbf{r}_{-1}\right\|^2 = o(1)$ under Assumption 7, $\sum_{\ell=1}^{\infty} \rho^2 = O(1)$ by Assumption 6, $\left\| \dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}'_{-1} \right\|_c = o(N)$ by Assumption 7, and $\sum_{\ell=1}^{\infty} \left\|\boldsymbol{\nu}_{\ell,-1}\right\|_r \left\|\boldsymbol{\nu}_{\ell,-1}\right\|_c = O\left(N^{-1}\right)$ by properties (39)-(41). Noting that $E\left[\sum_{\ell=0}^{\infty} \left(\mathbf{s}'_{1N}\boldsymbol{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\right) \boldsymbol{\varepsilon}_{t-\ell}\right] = 0$, equations (38), (43) establish

$$\sum_{\ell=0}^{\infty} \left(\mathbf{s}'_{1N}\boldsymbol{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\right) \boldsymbol{\varepsilon}_{t-\ell} \overset{q.m.}{\to} 0, \text{ as } N \to \infty. \tag{44}$$

This completes the proof of equation (33). To prove (34), we write

$$x_t^* - \sum_{\ell=0}^{\infty} \left(\mathbf{w}'\mathbf{\Phi}^{\ell}\mathring{\mathbf{r}}_1\right)\varepsilon_{1,t-\ell} = \sum_{\ell=0}^{\infty}\left(\mathbf{w}'\mathbf{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\right)\boldsymbol{\varepsilon}_{t-\ell}. \tag{45}$$

Since the vectors $\left\{\mathbf{w}'\mathbf{\Phi}^{\ell}\right\}$ have the same properties as vectors $\{\boldsymbol{\nu}_\ell\}$ in equations (39)-(41), it follows that (using the same arguments as above),

$$\sum_{\ell=0}^{\infty}\left(\mathbf{w}'\mathbf{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\right)\boldsymbol{\varepsilon}_{t-\ell} \overset{q.m.}{\to} 0, \text{ as } N \to \infty. \tag{46}$$

This completes the proof of equation (34). ■

The model for unit $i = 1$ can be approximated by an AR(p_1) process, which does not depend on the realizations from the remaining units as $N \to \infty$. Let the lag polynomial

$$a\left(L, p_1\right) \approx \vartheta^{-1}\left(L, \mathbf{s}_{1N}\right) \tag{47}$$

be an approximation of $\vartheta^{-1}\left(L, \mathbf{s}_{1N}\right)$. Then equation for unit $i = 1$ can be written as

$$a\left(L, p_1\right) x_{1t} \approx \varepsilon_{1t}. \tag{48}$$

The following proposition presents mean square error convergence results for the remaining cross section units.

Proposition 4 *Consider system (1), let Assumptions 5-7 hold and suppose that the lag polynomial $\vartheta\left(L, \mathbf{s}_{1N}\right)$ defined in Proposition 3 is invertible. Then as $N \to \infty$, equations for cross section unit $i \neq 1$ in the system (1) reduce to*

$$\left(1 - \phi_{ii}L\right) x_{it} - \beta_i\left(L\right) x_{1t} - r_{ii}\varepsilon_{it} \overset{q.m.}{\to} 0, \text{ for } i = 2, 3, ... \tag{49}$$

where $\beta_i\left(L\right) = \phi_{i1}L + \left[r_{i1} + \vartheta\left(L, \boldsymbol{\phi}_{-1,-i}\right)L\right]\vartheta^{-1}\left(L, \mathbf{s}_{1N}\right)$, and $\vartheta\left(L, \boldsymbol{\phi}_{-1,-i}\right) = \sum_{\ell=0}^{\infty}\left(\boldsymbol{\phi}'_{-1,-i}\mathbf{\Phi}^{\ell}\mathring{\mathbf{r}}_1\right)L^{\ell}$ for $i \neq 1$.

Proof.

$$x_{it} - \phi_{ii}x_{i,t-1} - \boldsymbol{\phi}'_{-1,-i}\mathbf{x}_{t-1} - \phi_{i1}x_{1,t-1} - r_{i1}\varepsilon_{1t} - r_{ii}\varepsilon_{it} = \mathbf{r}'_{-1,-i}\boldsymbol{\varepsilon}_t \tag{50}$$

Noting that $\mathbf{r}_{-1,-i}$ satisfies equation (32) of Assumption 7, and since $\left\|Var\left(\boldsymbol{\varepsilon}_t\right)\right\|_c = 1$ and $E\left(\boldsymbol{\varepsilon}_t\right) = 0$, then

$$\mathbf{r}'_{-1,-i}\boldsymbol{\varepsilon}_t \overset{q.m.}{\to} 0. \tag{51}$$

Considering (33) and noting that $\vartheta\left(L, \mathbf{s}_{1N}\right)$ is invertible, we have

$$\vartheta^{-1}\left(L, \mathbf{s}_{1N}\right) x_{1t} - \varepsilon_{1t} \overset{q.m.}{\to} 0.$$

This together with (34) implies

$$\boldsymbol{\phi}'_{-1,-i}\mathbf{x}_{t-1} - \vartheta\left(L, \boldsymbol{\phi}_{-1,-i}\right)\varepsilon_{1,t-1} \overset{q.m.}{\to} 0, \tag{52}$$

since $\boldsymbol{\phi}_{-1,-i}$ satisfies condition (2) under Assumption 5. Substituting these results into (50), equation (49) directly follows. ■

Remark 8 *Exclusion of x_{1t} from (49) is justified only if $r_{i1} = 0$.*

Remark 9 *Cross section unit 1 becomes (in the limit) a* dynamic common factor *for the remaining units in the system. Denote $f_t = x_{1t}$. Proposition 4 implies* [24]

$$(1 - \phi_{ii} L)\, x_{it} \approx r_{ii}\varepsilon_{it} + \beta_i (L)\, f_t \text{ for } i > 1. \tag{53}$$

Remark 10 *Conditional on x_{1t}, and its lags, the process $\{x_{it}\}$ become CWD.*

Remark 11 *For $\boldsymbol{\phi}_1 = \mathbf{0}$ and $\boldsymbol{\phi}_{-i} = \mathbf{0}$, we obtain from (49) the following static factor model as a special case*

$$(1 - \phi_{ii} L)\, x_{it} \approx r_{ii}\varepsilon_{it} + \left(\frac{r_{i1}}{r_{11}}\right) f_t, \text{ for } i > 1, \tag{54}$$

where $f_t = x_{1t}$.

3 Estimation of A Stationary IVAR

Let $\mathbf{x}_{it} = (x_{i1t}, x_{i2t}, ..., x_{ikt})'$ be a $k \times 1$ dimensional vector of k variables for unit (group) $i \in \{1, 2, ..., N\}$ and denote all the endogenous variables in the system by the $Nk \times 1$ vector $\mathbf{x}_t = (\mathbf{x}'_{1t}, \mathbf{x}'_{2t}, ..., \mathbf{x}'_{Nt})'$. Consider the case where k is fixed and relatively small and $(T, N) \overset{j}{\to} \infty$. This set-up corresponds, for example, to the panel of countries in the global economy, with a small set of key macroeconomic variables in each country, such as the GVAR model developed by Pesaran, Schuermann and Weiner (2004). Alternatively, $\mathbf{x}_{it}$ could refer to panels of firms or households with a finite number of variables per cross section unit. Other configurations of N, T and k could also be of interest. One possibility would be the case where N is fixed but $(k, T) \to \infty$, which corresponds to the case of one or more advanced open economies in data-rich ($k \to \infty$) environment. In this paper, we confine our analysis to the case where k is fixed. We shall assume the same number of variables per group only to simplify the exposition.[25]

Assume $\mathbf{x}_t$ is generated as:

$$\boldsymbol{\Phi}(L)\left(\mathbf{x}_t - \boldsymbol{\alpha} - \boldsymbol{\Gamma}\mathbf{f}_t\right) = \mathbf{u}_t, \tag{55}$$

for $t = 1, ..., T$, where $\boldsymbol{\Phi}(L) = \mathbf{I}_{Nk} - \boldsymbol{\Phi}L$, $\boldsymbol{\Phi}$ is $Nk \times Nk$ dimensional matrix of unknown coefficients, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1, \boldsymbol{\alpha}'_2, ..., \boldsymbol{\alpha}'_N)'$ is $Nk \times 1$ dimensional vector of fixed effects, $\boldsymbol{\alpha}_i = (\alpha_{i1}, ..., \alpha_{ik})'$, $\mathbf{f}_t$ is $m \times 1$ dimensional vector of unobserved common factors (m is fixed and relatively small), $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}'_1, \boldsymbol{\Gamma}'_2, ..., \boldsymbol{\Gamma}'_N)'$ is $Nk \times m$ dimensional matrix of factor loadings with

$$\boldsymbol{\Gamma}_i = \begin{pmatrix} \gamma_{i11} & \cdots & \gamma_{i1m} \\ \vdots & & \vdots \\ \gamma_{ik1} & \cdots & \gamma_{ikm} \end{pmatrix},$$

and $\mathbf{u}_t = (\mathbf{u}'_{1t}, \mathbf{u}'_{2t} ..., \mathbf{u}'_{Nt})'$ is the vector of error terms assumed to be independently distributed of $\mathbf{f}_{t'}$ $\forall t, t' \in \{1, .., T\}$.[26] Assumption of no observed common factors could be relaxed without major difficulties.[27]

[24] x_{1t} could be equivalently approximated by cross sectional weighted averages of $\mathbf{x}_t$ and its lags, namely $x_t^*, x_{t-1}^*, ...$

[25] Generalization of the model to allow for $k_i = O(1)$ number of variables per group is straightforward.

[26] One could also add observed common factors and/or deterministic terms to the equations in (55), but in what follows we abstract from these for expositional simplicity.

[27] Note that the presence of strictly exogenous observed common factors in the system (55) poses no difficulties as exogenous observed common factors could be directly included in the group-specific auxiliary regressions. In the case of endogenous observed common factors (such as the price of oil for example in the context of the global economy) an additional equation explaining the behavior of the endogenous common factors would be needed. This could be done by treating the endogenous observed common factors as additional dominant groups (allowing for k to differ across groups).

Elements of $k \times 1$ dimensional vector $\mathbf{u}_{it}$ are denoted by $\mathbf{u}_{it} = (u_{i1t}, ..., u_{ikt})'$. This set-up allows us to refer to the elements of vector $\mathbf{u}_t$ in two different ways: u_{lt} represents the l^{th} element of $\mathbf{u}_t$, while u_{irt} represents the r^{th} element of $\mathbf{u}_{it}$. It follows that $u_{lt} = u_{irt}$ for $l = (i-1)k + r$. Let us partition matrix $\mathbf{\Phi}$ into $k \times k$ dimensional sub-matrices, $\mathbf{\Phi}_{ij}$, so that

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{\Phi}_{11} & \cdots & \mathbf{\Phi}_{1N} \\ \vdots & & \vdots \\ \mathbf{\Phi}_{N1} & \cdots & \mathbf{\Phi}_{NN} \end{pmatrix}.$$

Define the following weighted averages

$$\underset{m_w \times 1}{\mathbf{x}_t^*} = \underset{m_w \times Nk}{\mathbf{W}'} \cdot \underset{Nk \times 1}{\mathbf{x}_t}, \tag{56}$$

where $\mathbf{W}' = (\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_N)$ and $\{\mathbf{W}_i\}_{i=1}^N$ are $m_w \times k$ sub-matrices. Matrix $\mathbf{W}$ is any matrix of pre-determined weights satisfying following granularity conditions[28]

$$\|\mathbf{W}\| = O\left(N^{-\frac{1}{2}}\right), \tag{62}$$

$$\frac{\|\mathbf{W}_i\|}{\|\mathbf{W}\|} = O\left(N^{-\frac{1}{2}}\right) \text{ for any } i \leq N \text{ and for any } N \in \mathbb{N}. \tag{63}$$

$\mathbf{x}_t^*$ is not necessarily the vector of cross sectional averages as there is no explicit restriction assumed on averaging different types of variables across groups. In empirical applications, cross sectional averages, however, are likely to be a preferred choice in the light of Assumption 17 below. In cases where the number of unobserved common factors (m) is lower than the number of endogenous variables per group (k), full augmentation by cross sectional averages ($m_w = k$) is not necessary for consistent estimation of $\mathbf{\Phi}_{ii}$.[29]

Subscripts denoting the number of groups are omitted in order to keep the notation simple. Note that as N is changing, potentially all elements of the $Nk \times Nk$ dimensional matrix $\mathbf{\Phi}$ are changing. Subscript for the number of groups in the system will only be used if necessary.

ASSUMPTION 8 *(No local dominance) Let $\mathbf{\Phi}_{-i} = (\mathbf{\Phi}_{i1}, \mathbf{\Phi}_{i2}, ..., \mathbf{\Phi}_{i,i-1}, \mathbf{0}_{k\times k}, \mathbf{\Phi}_{i,i+1}, ..., \mathbf{\Phi}_{iN})'$ where $\mathbf{\Phi}_{ij}$ are $k \times k$ dimensional sub-matrices of matrix $\mathbf{\Phi}$.*

$$\|\mathbf{\Phi}_{-i}\|_r = O\left(N^{-1}\right) \textit{ for any } i \leq N \textit{ and any } N \in \mathbb{N}. \tag{64}$$

[28] Granularity conditions (62)-(63) are equivalent to the following conditions (57)-(58)

$$\|\mathbf{W}\|_c \neq o(1) \tag{57}$$

$$\|\mathbf{W}\|_r = O\left(N^{-1}\right) \tag{58}$$

To see this, note that (63) implies

$$\frac{\|\mathbf{W}_i\|^2}{\|\mathbf{W}\|^2} = O\left(N^{-1}\right) \tag{59}$$

But $\|\mathbf{W}\|^2 = O\left(N^{-1}\right)$ by (62). Therefore

$$\|\mathbf{W}_i\|^2 = O\left(N^{-2}\right) \tag{60}$$

Since dimensions of $\mathbf{W}_i$ is $m_w \times k$, it follows that order of magnitude of $\|\mathbf{W}_i\|_M$ does not depend on a particular matrix norm $\|\cdot\|_M$ under consideration. Hence $\|\mathbf{W}_i\|_r^2 = O\left(N^{-2}\right)$. Note that condition (58) also establishes that

$$\|\mathbf{W}\|_c = O(1) \tag{61}$$

[29] For more details see Remark 14 below.

ASSUMPTION 9 *(Stationarity)*

$$\|\mathbf{\Phi}\|_r \|\mathbf{\Phi}\|_c < \rho < 1 \text{ for any } N \in \mathbb{N}. \tag{65}$$

ASSUMPTION 10 *(Sequence of diagonal sub-matrices of coefficient matrices* $\mathbf{\Phi}$ *as* $N \to \infty$*) Diagonal sub-matrices,* $\mathbf{\Phi}_{ii}$*, of matrix* $\mathbf{\Phi}$ *do not change with* $N \geq i$.

ASSUMPTION 11 *(Weakly dependent errors with finite fourth moments)* $\mathbf{u}_t$ *is independent of* $\mathbf{u}_{t'}$ *for* $\forall t \neq t'$. *Let* $E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{\Sigma}$, *and denote* $E(u_{nt}u_{ht}u_{st}u_{rt}) = \sigma_{nhsr}$.

$$\|\mathbf{\Sigma}\|_r = O(1) \text{ for any } N \in \mathbb{N}, \tag{66}$$

and

$$\sup_{N \in \mathbb{N},\ n,h \leq kN} \|\mathbf{\Psi}_{nh}\|_r = O(1), \tag{67}$$

where $kN \times kN$ *dimensional matrix* $\mathbf{\Psi}_{nh}$ *consists of elements* σ_{nhsr}, $s,r \in \{1,..,kN\}$.[30] $\mathbf{u}_{it} = (u_{i1t}, u_{i2t}, ..., u_{ikt})'$ *does not depend on* $N \geq i$. *For future reference, define* $k \times k$ *matrix* $\mathbf{\Sigma}_{ii} = E(\mathbf{u}_{it}\mathbf{u}_{it}')$.

ASSUMPTION 12 *(Available observations) Available observations are* $\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_T$ *with the starting values* $\mathbf{x}_0 = \sum_{\ell=0}^{\infty} \mathbf{\Phi}^{\ell}\mathbf{u}(-\ell) + \boldsymbol{\alpha} + \mathbf{\Gamma}\mathbf{f}_0$.[31]

ASSUMPTION 13 *(Common factors)* $\|\boldsymbol{\alpha}\|_r = O(1)$ *and* $\boldsymbol{\alpha}_i$ *does not change with* $N \geq i$. *Unobserved common factors* $f_{1t}, ..., f_{mt}$ *follow stationary* $AR(\infty)$ *processes:*

$$f_{rt} = \psi_r(L)\varepsilon_{frt} \text{ for } r = 1,..,m, \tag{68}$$

where polynomials $\psi_r(L) = \sum_{\ell=0}^{\infty} \psi_{r\ell}L^{\ell}$ *are absolute summable,* $\varepsilon_{frt} \sim IID\left(0, \sigma^2_{\varepsilon_{fr}}\right)$ *and the fourth moments of* ε_{frt} *are bounded,* $E\left(\varepsilon^4_{frt}\right) < \infty$. ε_{frt} *is independently distributed of* $\mathbf{u}_{t'}$ *for any* $t \neq t'$ *and any* $r \in \{1,..,m\}$.

ASSUMPTION 14 *(Random factor loadings)*

$$\gamma_{isr} = \gamma_{sr} + \eta_{isr} \text{ for } s \in \{1,..,k\},\ r \in \{1,..,m\} \text{ and any } i \in \mathbb{N}, \tag{69}$$

where $\eta_{isr} \sim IID\left(0, \sigma^2_{\eta sr}\right)$, γ_{isr} *does not change with* $N \geq i$. η_{isr} *is independently distributed of* $\varepsilon_{r't}$ *and* $\mathbf{u}_t$ *for any* $s \in \{1,..,k\}$, $r, r' \in \{1,..,m\}$, $i \in \mathbb{N}$ *and any* $t \in \mathbb{Z}$. *Furthermore, the third and the fourth moments of* η_{isr} *are bounded.*

ASSUMPTION 15 *(Non-random factor loadings)* $\gamma_{isr} = O(1)$ *for any* $i \in \mathbb{N}$ *and any* $s \in \{1,..,k\}$, $r \in \{1,..,m\}$. γ_{isr} *does not change with* $N \geq i$.

ASSUMPTION 16 *Let* $\underset{m \times m}{\mathbf{\Gamma}_f}(0) = E(\mathbf{f}_t\mathbf{f}_t')$, $\underset{m \times m}{\mathbf{\Gamma}_f}(1) = E\left(\mathbf{f}_t\mathbf{f}_{t-1}'\right)$, $\underset{m_w \times m}{\mathbf{\Gamma}^*} = \mathbf{W}'\mathbf{\Gamma}$, $\underset{m_w \times 1}{\boldsymbol{\alpha}^*} = \mathbf{W}'\boldsymbol{\alpha}$, *and* $\underset{k \times k}{\mathbf{\Gamma}_{\boldsymbol{\xi}_i}}(0) = E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}'\right)$ *where*

$$\boldsymbol{\xi}_{it} = \mathbf{\Phi}_{ii}\boldsymbol{\xi}_{i,t-1} + \mathbf{u}_{it}, \tag{70}$$

[30] Please note that $\|\mathbf{\Sigma}\| \leq \|\mathbf{\Sigma}\|_r$ for any symmetric matrix since $\|\mathbf{A}\|^2 \leq \sqrt{\|\mathbf{A}\|_c \|\mathbf{A}\|_r}$ for any matrix $\mathbf{A}$. See for instance Horn and Johnson (1985).

[31] We use notation $\mathbf{u}(-\ell)$ instead of $\mathbf{u}_{-\ell}$ in order to avoid possible confusion with the notation used in Section 2.

for $t = 1, ..., T$ *with starting values* $\boldsymbol{\xi}_{i0} = \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}_{ii}^{\ell} \mathbf{u}_i (-\ell)$. *Matrix*

$$\underset{(1+k+2m_w)\times(1+k+2m_w)}{\mathbf{C}_i} = \begin{pmatrix} 1 & \boldsymbol{\alpha}_i' & \boldsymbol{\alpha}^{*\prime} & \boldsymbol{\alpha}^{*\prime} \\ \boldsymbol{\alpha}_i & \boldsymbol{\alpha}_i\boldsymbol{\alpha}_i' + \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0) + \boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_f(0)\boldsymbol{\Gamma}_i' & \boldsymbol{\alpha}_i\boldsymbol{\alpha}^{*\prime} + \boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_f(1)\boldsymbol{\Gamma}^{*\prime} & \boldsymbol{\alpha}_i\boldsymbol{\alpha}^{*\prime} + \boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_f(0)\boldsymbol{\Gamma}^{*\prime} \\ \boldsymbol{\alpha}^* & \boldsymbol{\alpha}^*\boldsymbol{\alpha}_i' + \boldsymbol{\Gamma}^*\boldsymbol{\Gamma}_f'(1)\boldsymbol{\Gamma}_i' & \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime} + \boldsymbol{\Gamma}^*\boldsymbol{\Gamma}_f(0)\boldsymbol{\Gamma}^{*\prime} & \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime} + \boldsymbol{\Gamma}^*\boldsymbol{\Gamma}_f(1)\boldsymbol{\Gamma}^{*\prime} \\ \boldsymbol{\alpha}^* & \boldsymbol{\alpha}^*\boldsymbol{\alpha}_i' + \boldsymbol{\Gamma}^*\boldsymbol{\Gamma}_f(0)\boldsymbol{\Gamma}_i' & \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime} + \boldsymbol{\Gamma}^*\boldsymbol{\Gamma}_f'(1)\boldsymbol{\Gamma}^{*\prime} & \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime} + \boldsymbol{\Gamma}^*\boldsymbol{\Gamma}_f(0)\boldsymbol{\Gamma}^{*\prime} \end{pmatrix},$$

is nonsingular for any $i \leq N$, $i, N \in \mathbb{N}$; *and also in the limit as* $N \to \infty$.

ASSUMPTION 17 *Matrix* $(\boldsymbol{\Gamma}^{*\prime}\boldsymbol{\Gamma}^*)$ *is nonsingular for any* $N \in \mathbb{N}$, *and also in the limit as* $N \to \infty$. *Furthermore,* $\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0)$ *is nonsingular and* $\boldsymbol{\alpha}^* \neq 0$.

Remark 12 *Assumption 8 implies*

$$\left\|\boldsymbol{\Phi}_{-i}\right\|^2 = O\left(N^{-1}\right) \text{ for any } i \leq N \text{ and any } N \in \mathbb{N}, \tag{71}$$

and matrix $\boldsymbol{\Phi}_{-i}$ *satisfies condition (62). Under Assumptions 8 and 11, besides* $\{1, \mathbf{f}_t\}$, *there are no additional common factors in the system (55). Locally dominant (spatial) dependence of errors* $\mathbf{u}_t$ *is not excluded in Assumption 11.*

Remark 13 *(Eigenvalues of* $\boldsymbol{\Phi}$ *and* $\boldsymbol{\Phi}_{ii}$*) Assumption 9 implies*

$$\varrho(\boldsymbol{\Phi}) < \rho < 1, \tag{72}$$

as well as

$$\varrho(\boldsymbol{\Phi}'\boldsymbol{\Phi}) = \left\|\boldsymbol{\Phi}\right\|^2 < \rho < 1. \tag{73}$$

This is because $\varrho(\boldsymbol{\Phi}) \leq \min\left\{\left\|\boldsymbol{\Phi}\right\|_r, \left\|\boldsymbol{\Phi}\right\|_c\right\}$ *and* $\left\|\boldsymbol{\Phi}\right\| \leq \sqrt{\left\|\boldsymbol{\Phi}\right\|_r \left\|\boldsymbol{\Phi}\right\|_c}$. *Furthermore, since* $\left\|\boldsymbol{\Phi}_{ii}\right\|_r \leq \left\|\boldsymbol{\Phi}\right\|_r$ *and* $\left\|\boldsymbol{\Phi}_{ii}\right\|_c \leq \left\|\boldsymbol{\Phi}\right\|_c$, *it follows from Assumption 9 that for any* $i \leq N$ *and for any* $N \in \mathbb{N}$:

$$\left\|\boldsymbol{\Phi}_{ii}\right\|_r \left\|\boldsymbol{\Phi}_{ii}\right\|_c < \rho < 1, \tag{74}$$

$$\varrho(\boldsymbol{\Phi}_{ii}) < \rho < 1, \tag{75}$$

$$\varrho(\boldsymbol{\Phi}_{ii}'\boldsymbol{\Phi}_{ii}) = \left\|\boldsymbol{\Phi}_{ii}\right\|^2 < \rho < 1, \tag{76}$$

and all roots of $|\mathbf{I}_k - \boldsymbol{\Phi}_{ii}L| = 0$ *lie outside the unit circle.*

Remark 14 *Assumption 16 implies* $\boldsymbol{\Gamma}^* = \mathbf{W}'\boldsymbol{\Gamma}$ *is a square, full rank matrix and therefore Assumption 16 implicitly assumes that the number of unobserved common factors is equal the number of columns of the weight matrix* $\mathbf{W}$ $(m = m_w)$. *In cases when* $m < m_w \leq k$, *full augmentation of group-specific VAR models by (cross sectional) averages is not necessary. There is a trade-off between,* (i) *possible inconsistency of estimates introduced by the situation where the number of (cross sectional) averages in the augmented group-specific VAR models is lower than the number of common factors, and* (ii) *possible loss of power by augmenting the group-specific regressions by more (cross sectional) averages than the true number of unobserved common factors.*

Remark 15 *Assumption 17 is used instead of Assumption 16 for the case when the number of unobserved common factors is unknown, but not greater than* m_w.

Remark 16 *Let* $cov(\mathbf{x}_{it}, \mathbf{x}_{j,t-q}) = \mathbf{\Gamma}_{ij}(q)$. *Note that for* $i = j$, $\mathbf{\Gamma}_{ii}(q) = cov(\mathbf{x}_{it}, \mathbf{x}_{i,t-q})$ *is the q-th order autocovariance of* $\mathbf{x}_{it}$. *For* $q = 0$, $\mathbf{\Gamma}_{ij}(0) = cov(\mathbf{x}_{it}, \mathbf{x}_{jt})$ *denotes cross sectional covariance of groups* i *and* j. *For* $i \neq j$ *and* $q \neq 0$, *we have temporal cross-covariance.*

Solving (55) backward and multiplying both sides by $\mathbf{W}'$ yields

$$\mathbf{x}_t^* = \boldsymbol{\alpha}^* + \mathbf{\Gamma}^*\mathbf{f}_t + \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^\ell \mathbf{u}_{t-\ell}. \tag{77}$$

Under Assumption 11, $\{\mathbf{u}_t\}$ is weakly cross sectionally dependent and therefore

$$\begin{aligned}\left\| Var\left(\sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^\ell \mathbf{u}_{t-\ell}\right)\right\| &= \left\|\sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^\ell \mathbf{\Sigma}\mathbf{\Phi}'^\ell \mathbf{W}\right\|, \\ &\leq \|\mathbf{W}\|^2 \|\mathbf{\Sigma}\| \sum_{\ell=0}^{\infty} \left\|\mathbf{\Phi}^\ell\right\|^2, \\ &= O\left(N^{-1}\right), \end{aligned} \tag{78}$$

where $\|\mathbf{W}\|^2 = O\left(N^{-1}\right)$ by condition (62), $\|\mathbf{\Sigma}\| = O(1)$ by Assumption 11 (weak dependence) and $\sum_{\ell=0}^{\infty} \left\|\mathbf{\Phi}^\ell\right\| = O(1)$ under Assumption 9 (see Remark 13). Therefore (78) implies

$$\mathbf{x}_t^* = \boldsymbol{\alpha}^* + \mathbf{\Gamma}^*\mathbf{f}_t + O_p\left(N^{-\frac{1}{2}}\right), \tag{79}$$

and the unobserved common factors can be approximated as

$$\left(\mathbf{\Gamma}^{*\prime}\mathbf{\Gamma}^*\right)^{-1} \mathbf{\Gamma}^{*\prime}\left(\mathbf{x}_t^* - \boldsymbol{\alpha}^*\right) = \mathbf{f}_t + O_p\left(N^{-\frac{1}{2}}\right), \tag{80}$$

provided that either Assumption 16 or 17 hold. It can be inferred that full column rank of $\mathbf{\Gamma}^*$ is an important assumption for the results derived in this section. In the case where $m < k$ (that is augmenting individual VAR models by $m \times 1$ dimensional vector of (cross sectional) averages could be satisfactory), augmenting the regression by $k \times 1$ dimensional vectors of (cross sectional) averages have asymptotically no bearings on estimates of coefficients $\mathbf{\Phi}_{ii}$ below. This is because in both cases, vectors of (cross sectional) averages span the same space asymptotically.

Consider following auxiliary VARX$^*(1,1)$ regression

$$\mathbf{x}_{it} = \mathbf{c}_i + \mathbf{\Phi}_{ii}\mathbf{x}_{i,t-1} + \mathbf{B}_{i1}\mathbf{x}_t^* + \mathbf{B}_{i2}\mathbf{x}_{t-1}^* + \boldsymbol{\epsilon}_{it}, \tag{81}$$

where $\mathbf{B}_{i1} = \mathbf{\Gamma}_i\left(\mathbf{\Gamma}^{*\prime}\mathbf{\Gamma}^*\right)^{-1}\mathbf{\Gamma}^{*\prime}$, $\mathbf{B}_{i2} = -\mathbf{\Phi}_{ii}\mathbf{B}_{i1}$, $\mathbf{c}_i = \boldsymbol{\alpha}_i - \mathbf{\Phi}_{ii}\boldsymbol{\alpha}_i + \left(\mathbf{B}_{i2} - \mathbf{B}_{i1}\right)\boldsymbol{\alpha}^*$, $\boldsymbol{\epsilon}_{it} = \mathbf{u}_{it} + \mathbf{h}_{it}$, and

$$\mathbf{h}_{it} = \mathbf{B}_{i1}\mathbf{W}'\left(\mathbf{x}_t - \boldsymbol{\alpha} - \mathbf{\Gamma}\mathbf{f}_t\right) + \left(\mathbf{B}_{i2}\mathbf{W}' + \mathbf{\Phi}_{-i}\right)\left(\mathbf{x}_{t-1} - \boldsymbol{\alpha} - \mathbf{\Gamma}\mathbf{f}_{t-1}\right).$$

For variable x_{irt}, $r \in \{1, .., k\}$, in system (81) we have

$$x_{irt} = \mathbf{g}_{it}'\boldsymbol{\pi}_{ir} + \epsilon_{irt}, \tag{82}$$

where $\boldsymbol{\pi}_{ir} = \left(c_{ir}, \phi_{r1}^{(ii)}, ..., \phi_{rk}^{(ii)}, b_{r1}^{(i1)}, ..., b_{rk}^{(i1)}, b_{r1}^{(i2)}, ..., b_{rk}^{(i2)}\right)' = \left(c_{ir}, \boldsymbol{\phi}_{ii,r}', \mathbf{b}_{i1,r}', \mathbf{b}_{i2,r}'\right)'$ is $(3k+1) \times 1$ dimensional vector of coefficients[32] and the vector of regressors is $\mathbf{g}_{it} = \left(1, \mathbf{x}_{i,t-1}', \mathbf{x}_t^{*\prime}, \mathbf{x}_{t-1}^{*\prime}\right)'$. Denote the corre-

[32] $\phi_{rn}^{(ii)}, b_{rn}^{(i1)}$, and $b_{rn}^{(i2)}$ denote element (r, n) of matrices $\mathbf{\Phi}_{ii}$, $\mathbf{B}_{i1}$, and $\mathbf{B}_{i2}$, respectively. $\boldsymbol{\phi}_{ii,r}', \mathbf{b}_{i1,r}'$, and $\mathbf{b}_{i2,r}'$ denote row r

sponding OLS estimator of $\boldsymbol{\pi}_{ir}$ as $\widehat{\boldsymbol{\pi}}_{ir}$:

$$\widehat{\boldsymbol{\pi}}_{ir} = \left(\sum_{t=1}^{T} \mathbf{g}_{it} \mathbf{g}_{it}' \right)^{-1} \sum_{t=1}^{T} \mathbf{g}_{it} x_{irt}. \tag{83}$$

We denote the estimator of $\boldsymbol{\phi}_{ii,r}$ given by the k-dimensional sub-vector of $\widehat{\boldsymbol{\pi}}_{ir}$ as the cross section augmented least squares estimator (or CALS for short), denoted as $\widehat{\boldsymbol{\phi}}_{ii,r,CALS}$. $\widehat{\boldsymbol{\phi}}_{ii,r,CALS}$ is equivalently defined using the partition regression formula in equation (103) under the Assumption 16, which ensures the invertibility of the matrix $\mathbf{C}_i$. Under the weaker Assumption 17, CALS is still well defined using the partition regression formula in equation (103). Asymptotic properties of $\widehat{\boldsymbol{\pi}}_{ir}$ (and $\widehat{\boldsymbol{\phi}}_{ii,r,CALS}$ in the case where the number of unobserved common factors is unknown) are the objective of this analysis as N and T tend to infinity. Following types of convergence for N and T are considered.

ASSUMPTION B1: $N, T \xrightarrow{j} \infty$ at any order.

ASSUMPTION B2: $N, T \xrightarrow{j} \infty$, and $T/N \to \varkappa < \infty$, where $\varkappa \geq 0$ is not necessarily nonzero.

Clearly, Assumption B2 is stronger than Assumption B1. Situations where the number of unobserved common factors equals m_w is analyzed first.

Theorem 1 *Consider model (55). Let $l = (i-1)k + r$, suppose Assumptions 8-13,16 hold and factor loadings are governed either by Assumption 14 or 15. Furthermore, let $\mathbf{W}$ be any arbitrary (pre-determined) matrix of weights satisfying conditions (62)-(63) and Assumption 16. Then for any $i \in \mathbb{N}$, and for any $r \in \{1, .., k\}$, the estimator $\widehat{\boldsymbol{\pi}}_{ir}$ defined by (83) has following properties.*

a) Under Assumption B1, $\widehat{\boldsymbol{\pi}}_{ir}$ is consistent estimator of $\boldsymbol{\pi}_{ir}$.

b) Under Assumption B2,

$$\sqrt{T} \left(\widehat{\boldsymbol{\pi}}_{ir} - \boldsymbol{\pi}_{ir} \right) \xrightarrow{d} N \left(\mathbf{0}, \sigma_{ll} \mathbf{C}_i^{-1} \right), \tag{84}$$

where $\sigma_{ll} = Var(u_{irt}) = Var(u_{lt})$, and $\mathbf{C}_i$ is positive definite matrix defined in Assumption 16.

c) Under Assumption B1, matrix $\mathbf{C}_i$ and scalar σ_{ll} can be consistently estimated by

$$\widehat{\mathbf{C}}_i = \frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_{it} \mathbf{g}_{it}', \text{ and } \widehat{\sigma}_{ll} = \frac{1}{T - 2k - 1} \sum_{t=1}^{T} \widehat{u}_{lt}^2, \tag{85}$$

respectively, where $\widehat{u}_{lt} = \widehat{u}_{irt} = x_{irt} - \mathbf{g}_{it}' \widehat{\boldsymbol{\pi}}_{ir}$.

d) Under Assumption B2,

$$\sqrt{T} \left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i \right) \xrightarrow{d} N \left(\mathbf{0}, \boldsymbol{\Sigma}_{ii} \otimes \mathbf{C}_i^{-1} \right), \tag{86}$$

where $\boldsymbol{\pi}_i = \left(\boldsymbol{\pi}_{i1}', \boldsymbol{\pi}_{i2}', ..., \boldsymbol{\pi}_{ik}' \right)'$, similarly $\widehat{\boldsymbol{\pi}}_i = \left(\widehat{\boldsymbol{\pi}}_{i1}', ..., \widehat{\boldsymbol{\pi}}_{ik}' \right)'$, and $\boldsymbol{\Sigma}_{ii} = E \left(\mathbf{u}_{it} \mathbf{u}_{it}' \right)$. Furthermore, $\boldsymbol{\Sigma}_{ii}$ can be consistently estimated by

$$\widehat{\boldsymbol{\Sigma}}_{ii} = \frac{1}{T - 2k - 1} \sum_{t=1}^{T} \widehat{\mathbf{u}}_{it} \widehat{\mathbf{u}}_{it}'. \tag{87}$$

Proof.

of matrices $\boldsymbol{\Phi}_{ii}$, $\mathbf{B}_{i1}$, and $\mathbf{B}_{i2}$, respectively.

b) Substituting for ϵ_{irt} in (83) allow us to write:

$$\begin{aligned}\sqrt{T}\left(\widehat{\boldsymbol{\pi}}_{ir}-\boldsymbol{\pi}_{ir}\right) &= \sqrt{T}\left(\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\sum_{t=1}^{T}\mathbf{g}_{it}\epsilon_{irt}, \\ &= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}h_{irt}+\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}u_{irt}. \quad (88)\end{aligned}$$

Under B2, equations (160)-(162) of Lemma 4 in Appendix imply

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}h_{irt}\xrightarrow{p}\mathbf{0}. \quad (89)$$

According to Lemma 5,

$$\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)\xrightarrow{p}\mathbf{C}_i, \quad (90)$$

if B1 holds, where $\mathbf{C}_i$ is nonsingular under Assumption 16. It therefore follows that under B2

$$\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}h_{irt}\xrightarrow{p}\mathbf{0}. \quad (91)$$

We can rewrite $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}u_{irt}$ as

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}u_{irt}=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{a}_{i}u_{irt}+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{g}_{it}-\mathbf{a}_i\right)u_{irt}, \quad (92)$$

where

$$\mathbf{a}_{it}=\begin{pmatrix}1\\ \boldsymbol{\alpha}_i+\boldsymbol{\xi}_{i,t-1}+\boldsymbol{\Gamma}_i\mathbf{f}_{t-1}\\ \boldsymbol{\alpha}^*+\boldsymbol{\Gamma}^*\mathbf{f}_t\\ \boldsymbol{\alpha}^*+\boldsymbol{\Gamma}^*\mathbf{f}_{t-1}\end{pmatrix}.$$

Lemma 6 implies (under B2)

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{g}_{it}-\mathbf{a}_{it}\right)u_{irt}\xrightarrow{p}\mathbf{0}. \quad (93)$$

It follows from Lemma 7 that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{a}_{it}u_{irt}\xrightarrow{D}N\left(\mathbf{0},\sigma_{ll}\mathbf{C}_i\right), \quad (94)$$

as $T\to\infty$ (Lemma 7 is standard time series result[33]). (90), (93) and (94) imply

$$\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}u_{irt}\xrightarrow{D}N\left(\mathbf{0},\sigma_{ll}\mathbf{C}_i^{-1}\right), \quad (95)$$

given B2. (91) and (95) establish (84).

[33] See for instance Hamilton (1994, Chapter 7 and Chapter 8).

a)

$$\left(\widehat{\boldsymbol{\pi}}_{ir}-\boldsymbol{\pi}_{ir}\right)=\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}h_{irt}+\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}u_{irt} \tag{96}$$

Consider now Assumption B1 where $N,T\xrightarrow{j}\infty$ at any order. Lemma 6 implies $\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{g}_{it}-\mathbf{a}_{it}\right)u_{irt}\xrightarrow{p}\mathbf{0}$. (94) establishes $\frac{1}{T}\sum_{t=1}^{T}\mathbf{a}_{it}u_{irt}\xrightarrow{p}\mathbf{0}$. Hence

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}u_{irt}\xrightarrow{p}\mathbf{0}. \tag{97}$$

Lemma 4, and Lemma 5 imply

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}h_{irt}\xrightarrow{p}\mathbf{0},\text{ and }\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\xrightarrow{p}\mathbf{C}_i, \tag{98}$$

respectively. Hence, from (97) and (98) it follows that $\widehat{\boldsymbol{\pi}}_{ir}\xrightarrow{p}\boldsymbol{\pi}_{ir}$.

c) Lemma 5 implies $\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'$ is a consistent estimator of $\mathbf{C}_i$ under Assumption B1. $\mathbf{x}_{it}-\mathbf{g}_{it}\widehat{\boldsymbol{\pi}}_{ir}\xrightarrow{p}\boldsymbol{\epsilon}_{it}$ (under B1) directly follows from the consistency of $\widehat{\boldsymbol{\pi}}_{ir}$. Since $\boldsymbol{\epsilon}_{it}=\mathbf{u}_{it}+\mathbf{h}_{it}$ and $\frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_{it}\mathbf{h}_{it}'\xrightarrow{p}\mathbf{0}$ under B1 by Lemma 1, it follows $\widehat{\sigma}_{ll}\xrightarrow{p}\sigma_{ll}$ for any $l\in\{1,...,Nk\}$.[34]

d) Consider now asymptotics B2 where $N,T\xrightarrow{j}\infty$, and $T/N\rightarrow\varkappa<\infty$. (91) and (93) imply

$$\sqrt{T}\left(\widehat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_i\right)=\mathbf{C}_i^{-1}\begin{pmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{a}_{it}u_{i,1,t}\\ \vdots\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{a}_{it}u_{i,k,t}\end{pmatrix}+o_p\left(1\right). \tag{99}$$

Define the $k\left(1+k+2m_w\right)\times 1$ dimensional vector $\boldsymbol{\zeta}_{it}=\left(\mathbf{a}_{it}'u_{i,1,t},...,\mathbf{a}_{it}'u_{i,k,t}\right)'$. Note that $\boldsymbol{\zeta}_{it}$ is a martingale difference process with finite fourth order moments and the variance matrix

$$\begin{aligned}E\left(\boldsymbol{\zeta}_{it}\boldsymbol{\zeta}_{it}'\right) &= \begin{pmatrix}E\left(u_{i1t}^2\right) & E\left(u_{i1t}u_{i2t}\right) & \cdots & E\left(u_{i1t}u_{ikt}\right)\\ E\left(u_{i2t}u_{i1kt}\right) & E\left(u_{i2t}^2\right) & & E\left(u_{i2t}u_{ikt}\right)\\ \vdots & & \ddots & \vdots\\ E\left(u_{ikt}u_{i1kt}\right) & E\left(u_{ikt}u_{i2t}\right) & \cdots & E\left(u_{ikt}^2\right)\end{pmatrix}\otimes E\left(\mathbf{a}_{it}\mathbf{a}_{it}'\right),\\ &= \boldsymbol{\Sigma}_{ii}\otimes\mathbf{C}_i.\end{aligned} \tag{100}$$

Using the central limit theorem (CLT) for martingale differences (see, for example, Hamilton, Proposition 7.9, p. 194) we have:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\zeta}_{it}\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Sigma}_{ii}\otimes\mathbf{C}_i\right). \tag{101}$$

[34] $\mathbf{u}_{it}\mathbf{h}_{it}'$ is also ergodic in mean and $\left\|E\left(\mathbf{u}_{it}\mathbf{h}_{it}'\right)\right\|=O\left(N^{-1}\right)\rightarrow 0$ under B1.

Expression (99) can be written as

$$\sqrt{T}\left(\widehat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_i\right) = \begin{pmatrix} \mathbf{C}_i^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_i^{-1} & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_i^{-1} \end{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{\zeta}_{it} + o_p(1),$$

$$= \left(\mathbf{I}_k \otimes \mathbf{C}_i^{-1}\right) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{\zeta}_{it} + o_p(1). \tag{102}$$

Thus

$$\sqrt{T}\left(\widehat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_i\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{ii} \otimes \mathbf{C}_i^{-1}\right),$$

since $\left(\mathbf{I}_k \otimes \mathbf{C}_i^{-1}\right)\left(\boldsymbol{\Sigma}_{ii} \otimes \mathbf{C}_i\right)\left(\mathbf{I}_k \otimes \mathbf{C}_i^{-1}\right) = \boldsymbol{\Sigma}_{ii} \otimes \mathbf{C}_i^{-1}$. Consistency of $\widehat{\boldsymbol{\Sigma}}_{ii}$ follows from part (c) of this proof.

■

Now we consider the case where the number of unobserved common factors is unknown, but it is known that $m_w \geq m$. Since the auxiliary regression (81) is augmented possibly by higher number of 'star' variables than the number of unobserved common factors, we have potential problem of (asymptotic) multicollinearity. But this has no bearings on estimates of $\boldsymbol{\Phi}_{ii}$ as long as the space spanned by unobserved common factors including a constant and the space spanned by $\mathbf{x}_t^*$ are the same (asymptotically). This is the case when $\boldsymbol{\Gamma}^*$ has full column rank, that is if Assumption 17 holds. Using partition regression formula, the cross sectionally augmented least squares (CALS) estimator of vector $\boldsymbol{\phi}_{ii,r} = \left(\phi_{r1}^{(ii)}, ..., \phi_{rk}^{(ii)}\right)'$ in the auxiliary regression (81) is

$$\widehat{\boldsymbol{\phi}}_{ii,r,CALS} = \left(\mathbf{X}_i'\mathbf{M}_Z\mathbf{X}_i\right)^{-1} \mathbf{X}_i'\mathbf{M}_Z\mathbf{x}_{ir\circ}, \tag{103}$$

where $\mathbf{x}_{ir\circ} = (x_{ir1}, ..., x_{irT})'$, $\mathbf{X}_i = [\mathbf{x}_{i1}(-1), \mathbf{x}_{i2}(-1), ..., \mathbf{x}_{ik}(-1)]$, $\mathbf{x}_{ir}(-1) = (x_{ir0}, ..., x_{i,r,T-1})'$, $\mathbf{M}_Z = \mathbf{I}_T - \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{+}\mathbf{Z}'$, $\mathbf{Z} = [\mathbf{X}^*, \mathbf{X}^*(-1)]$, $\mathbf{X}^* = \left(\mathbf{x}_{1\circ}^*, ..., \mathbf{x}_{m_w\circ}^*\right)$, $\mathbf{X}^*(-1) = \left[\mathbf{x}_1^*(-1), ..., \mathbf{x}_{m_w}^*(-1)\right]$, $\mathbf{x}_{r\circ}^* = (x_{r1}^*, ..., x_{rT}^*)'$ and $\mathbf{x}_r^*(-1) = \left(x_{r0}^*, ..., x_{r,T-1}^*\right)'$.

Define for future reference following matrices.

$$\mathbf{Q} = [\boldsymbol{\tau}, \mathbf{F}, \mathbf{F}(-1)], \tag{104}$$

and

$$\underset{(2m+1)\times 2m_w}{\mathbf{A}} = \begin{pmatrix} \boldsymbol{\alpha}^{*\prime} & \boldsymbol{\alpha}^{*\prime} \\ \boldsymbol{\Gamma}^{*\prime} & \mathbf{0}_{m\times m_w} \\ \mathbf{0}_{m\times m_w} & \boldsymbol{\Gamma}^{*\prime} \end{pmatrix}, \tag{105}$$

where $\mathbf{F} = (\mathbf{f}_{1\circ}, ..., \mathbf{f}_{m\circ})$, $\mathbf{F}(-1) = [\mathbf{f}_1(-1), ..., \mathbf{f}_m(-1)]$, $\mathbf{f}_{r\circ} = (f_{r1}, ..., f_{rT})'$ and $\mathbf{f}_r(-1) = (f_{r0}, ..., f_{r,T-1})'$ for $r \in \{1, .., m\}$. Furthermore, let

$$\boldsymbol{\upsilon}_t = \left(\boldsymbol{\upsilon}_{1t}', ..., \boldsymbol{\upsilon}_{Nt}'\right)' = \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell}, \text{ and } \boldsymbol{\upsilon}_t^* = \mathbf{W}'\boldsymbol{\upsilon}_t. \tag{106}$$

It follows that $\boldsymbol{\upsilon}_{it} = \sum_{\ell=0}^{\infty} \mathbf{S}_i'\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell}$, where $\mathbf{S}_i = (\mathbf{0}_{k\times k}, .., \mathbf{0}_{k\times k}, \mathbf{I}_k, \mathbf{0}_{k\times k}, .., \mathbf{0}_{k\times k})'$ is $kN \times k$ dimensional selection matrix for group i.

Theorem 2 *Consider model (55). Let $l=(i-1)\,k+r$, suppose Assumptions 8-13, 17, B2 hold and factor loadings are governed by Assumption 14 or 15. Furthermore, let $\mathbf{W}$ be any arbitrary (pre-determined) matrix of weights satisfying conditions (62)-(63) and Assumption 17. Then for any $i\in\mathbb{N}$, and for any $r\in\{1,..,k\}$, the cross sectionally augmented least squares estimator, $\widehat{\boldsymbol{\phi}}_{ii,r,CALS}$, defined by (103) has following properties.*

a)

$$\sqrt{T}\left(\widehat{\boldsymbol{\phi}}_{ii,r,CALS}-\boldsymbol{\phi}_{ii,r}\right)\xrightarrow{d}N\left(\mathbf{0},\sigma_{ll}\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}\left(0\right)\right), \tag{107}$$

where $\sigma_{ll}=Var\left(u_{irt}\right)=Var\left(u_{lt}\right)$ and $\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}\left(0\right)$ is autocovariance function of the process $\boldsymbol{\xi}_{it}$ defined in Assumption 16.

b)

$$\sqrt{T}\left(\widehat{\boldsymbol{\phi}}_{ii,CALS}-\boldsymbol{\phi}_{ii}\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Sigma}_{ii}\otimes\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}\left(0\right)\right), \tag{108}$$

where $\boldsymbol{\phi}_{ii}=vec\left(\boldsymbol{\Phi}_{ii}\right)$, similarly $\widehat{\boldsymbol{\phi}}_{ii,CALS}=\left(\widehat{\boldsymbol{\phi}}'_{ii,1,CALS},...,\widehat{\boldsymbol{\phi}}'_{ii,k,CALS}\right)'$, and $\boldsymbol{\Sigma}_{ii}=E\left(\mathbf{u}_{it}\mathbf{u}'_{it}\right)$.

Proof.

a) Vector $\mathbf{x}_{ir\circ}$ can be written, using system (55), as

$$\mathbf{x}_{ir\circ}=\boldsymbol{\tau}\left(\boldsymbol{\alpha}'_i-\boldsymbol{\alpha}'_i\boldsymbol{\Phi}'_{ii}\right)+\mathbf{X}_i\boldsymbol{\Phi}'_{ii}\mathbf{s}_{rk}+\mathbf{F}\boldsymbol{\Gamma}'_i\mathbf{s}_{rk}-\mathbf{F}\left(-1\right)\boldsymbol{\Gamma}'_i\boldsymbol{\Phi}'_{ii}\mathbf{s}_{rk}+\mathbf{u}_{ir\circ}+\mathbf{e}_{ir\circ}, \tag{109}$$

where $\mathbf{s}_{rk}$ is $k\times 1$ dimensional selection vector ($\mathbf{s}_{rj}=0$ for $j\neq r$ and $s_{rr}=1$), $e_{ir\circ}=\left(e_{ir1},...,e_{irT}\right)'$ and $e_{irt}=\mathbf{s}'_{rk}\boldsymbol{\Phi}'_{-i}\left(\mathbf{x}_{t-1}-\boldsymbol{\alpha}-\boldsymbol{\Gamma}\mathbf{f}_{t-1}\right)$. Substituting (109) into (103) and noting that by Lemma 10, equation (203),

$$\frac{\mathbf{X}'_i\mathbf{M}_Z\mathbf{Q}}{\sqrt{T}}=\frac{\mathbf{X}'_i\mathbf{M}_Z\left[\boldsymbol{\tau},\mathbf{F},\mathbf{F}\left(-1\right)\right]}{\sqrt{T}}=o_p\left(\sqrt{\frac{T}{N}}\right), \tag{110}$$

it follows

$$\sqrt{T}\left(\widehat{\boldsymbol{\phi}}_{ii,r}-\boldsymbol{\phi}_{ii,r}\right)=\left(\frac{\mathbf{X}'_i\mathbf{M}_Z\mathbf{X}_i}{T}\right)^{-1}\left[\frac{\mathbf{X}'_i\mathbf{M}_Z\left(\mathbf{u}_{ir\circ}+\mathbf{e}_{ir\circ}\right)}{\sqrt{T}}+o_p\left(\sqrt{\frac{T}{N}}\right)\right]. \tag{111}$$

According to Lemma 10 and equations (201)-(202),

$$\frac{\mathbf{X}'_i\mathbf{M}_Z\mathbf{X}_i}{T}\xrightarrow{p}\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}\left(0\right), \tag{112}$$

under Assumption B1 (and therefore also under Assumption B2), where $\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}\left(0\right)$ is nonsingular under Assumption 17.

Consider now the case when $N,T\xrightarrow{j}\infty$, and $T/N\rightarrow\varkappa<\infty$ (Assumption B2). Lemma 11, equation (215), implies

$$\frac{\mathbf{X}'_i\mathbf{M}_Z\mathbf{e}_{ir\circ}}{\sqrt{T}}\xrightarrow{p}0. \tag{113}$$

Furthermore, we have by Lemma 11, equation (216),

$$\frac{\mathbf{X}'_i\mathbf{M}_Z\mathbf{u}_{ir\circ}}{\sqrt{T}}=\frac{\boldsymbol{\Upsilon}'_i\mathbf{u}_{ir\circ}}{\sqrt{T}}+o_p\left(\sqrt{\frac{T}{N}}\right)+o_p\left(1\right), \tag{114}$$

where the $T \times k$ dimensional matrix $\boldsymbol{\Upsilon}_i = (\boldsymbol{\upsilon}_{i0}, \boldsymbol{\upsilon}_{i1}, ..., \boldsymbol{\upsilon}_{i,T-1})'$. We can re-write $\boldsymbol{\Upsilon}_i'\mathbf{u}_{ir\circ}/\sqrt{T}$ as

$$\frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{ir\circ}}{\sqrt{T}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\xi}_{i,t-1}u_{irt} - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\boldsymbol{\upsilon}_{i,t-1} - \boldsymbol{\xi}_{i,t-1}\right)u_{irt}. \tag{115}$$

Lemma 6, equations (174) implies

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\boldsymbol{\upsilon}_{i,t-1} - \boldsymbol{\xi}_{i,t-1}\right)u_{irt} \xrightarrow{p} 0. \tag{116}$$

Using Lemma 12, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\xi}_{i,t-1}u_{lt} \xrightarrow{D} N\left(0, \sigma_{ll}\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0)\right). \tag{117}$$

It follows from equations (114)-(117) that

$$\frac{\mathbf{X}_i'\mathbf{M}_Z\mathbf{u}_{ir\circ}}{\sqrt{T}} \xrightarrow{D} N\left(0, \sigma_{ll}\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0)\right). \tag{118}$$

Equations (111)-(113) and (118) now establish (107).

b) Consider the case when $N, T \xrightarrow{j} \infty$, and $T/N \to \varkappa < \infty$ (Assumption B2). Proof of (108) is similar to the proof of (86). Part (a) of this proof implies

$$\sqrt{T}\left(\widehat{\boldsymbol{\phi}}_{ii,CALS} - \boldsymbol{\phi}_{ii}\right) = \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}(0)\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\xi}_{i,t-1}u_{i,1,t} \\ \vdots \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\xi}_{i,t-1}u_{i,k,t} \end{pmatrix} + o_p(1) + o_p\left(\sqrt{\frac{T}{N}}\right). \tag{119}$$

Define $k^2 \times 1$ dimensional vector $\boldsymbol{\zeta}_{it} = \left(\boldsymbol{\xi}_{i,t-1}'u_{i,1,t}, ..., \boldsymbol{\xi}_{i,t-1}'u_{i,k,t}\right)'$. Notice that $\boldsymbol{\zeta}_{it}$ is a martingale difference sequence with finite fourth moments and variance

$$\begin{aligned} E\left(\boldsymbol{\zeta}_{it}\boldsymbol{\zeta}_{it}'\right) &= \begin{pmatrix} E\left(u_{i1t}^2\right) & E\left(u_{i1t}u_{i2t}\right) & \cdots & E\left(u_{i1t}u_{ikt}\right) \\ E\left(u_{i2t}u_{i1kt}\right) & E\left(u_{i2t}^2\right) & & E\left(u_{i2t}u_{ikt}\right) \\ \vdots & & \ddots & \vdots \\ E\left(u_{ikt}u_{i1kt}\right) & E\left(u_{ikt}u_{i2t}\right) & \cdots & E\left(u_{ikt}^2\right) \end{pmatrix} \otimes E\left(\boldsymbol{\xi}_{i,t-1}\boldsymbol{\xi}_{i,t-1}'\right). \\ &= \boldsymbol{\Sigma}_{ii} \otimes \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0) \end{aligned} \tag{120}$$

It follows from a vector martingale difference CLT (see for example Hamilton, 1994, Proposition 7.9) that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\zeta}_{it} \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{ii} \otimes \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0)\right). \tag{121}$$

Expression (119) can be written as

$$
\begin{aligned}
\sqrt{T}\left(\widehat{\boldsymbol{\phi}}_{ii,CALS}-\boldsymbol{\phi}_{ii}\right) &= \begin{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}(0) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}(0) & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}(0) \end{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\zeta}_{it} + o_p(1) + o_p\left(\sqrt{\frac{T}{N}}\right), \\
&= \left(\mathbf{I}_k \otimes \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}(0)\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\zeta}_{it} + o_p(1) + o_p\left(\sqrt{\frac{T}{N}}\right). \qquad (122)
\end{aligned}
$$

Thus

$$
\sqrt{T}\left(\widehat{\boldsymbol{\phi}}_{ii}-\boldsymbol{\phi}_{ii}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{ii}\otimes\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}(0)\right), \qquad (123)
$$

since $\left(\mathbf{I}_k \otimes \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}(0)\right)\left(\boldsymbol{\Sigma}_{ii}\otimes\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0)\right)\left(\mathbf{I}_k \otimes \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}(0)\right) = \boldsymbol{\Sigma}_{ii}\otimes\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}^{-1}(0)$. This completes the proof.

■

Extension of the analysis to a IVAR(p) model is straightforward and it is relegated to the appendix B.2. An alternative method for carrying out estimation and inference in panels with multifactor error structure is using the method of principal components to estimate the unobserved common factors, following for example the work of Stock and Watson (2002).[35] [36]

3.1 Selection of Weights W

As specified above the matrix of weights, $\mathbf{W}$, is common across groups. However, since any weights satisfying conditions (62)-(63) can be used in estimation of the vector $\widehat{\boldsymbol{\phi}}_{ii}$, the proposed CALS estimator allows to use different, group-specific, weights for estimation of the slope coefficients of the individual groups. Asymptotically, there is no difference between two sets of weights as long as both satisfy granularity conditions (62)-(63), Assumptions 16 or 17 and both are pre-determined. In small samples, selection of optimal weights could be an important issue. For a related discussion, see Pesaran (2006, Section 6).

3.2 Consistent Estimation of $\boldsymbol{\phi}_{ii}$ In the Case Where the Number of Common Factors Exceeds the Number of Endogenous Variables per Group

There is no restriction on the number of columns m_w of the weight matrix $\mathbf{W}$, besides being bounded in N. Theorems 1 and 2 apply also for the case where the number of unobserved common factors exceeds the number of endogenous variables, as long as matrix $\mathbf{W}$ satisfies Assumptions 16 or 17. However, when $m > k$, it is harder to justify the choice of the matrix $\mathbf{W}$, while for $m \leq k$, cross sectional averages seems a natural choice. This is because one would expect different type of variables to be affected by common factors in

[35] Geweke (1977) and Sargent and Sims (1977) introduced dynamic factor models, which were generalized to allow for weak cross sectional dependence in innovations by Forni and Lippi (2001) and Forni *et al.* (2000, 2004).

[36] System (55) implies

$$
\mathbf{x}_t - \boldsymbol{\alpha} = \boldsymbol{\Gamma}\mathbf{f}_t + \sum_{\ell=0}^{\infty}\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell}. \qquad (124)
$$

Process $\sum_{\ell=0}^{\infty}\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell}$ is serially correlated, but weakly cross sectionally dependent. We can estimate the common factors $\mathbf{f}_t$ (up to a linear transformation) by the first m principal components of

$$
\mathbf{x}_t^d = \mathbf{x}_t - \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t. \qquad (125)
$$

a heterogenous fashion. In the context of global modeling, one could expect group of advanced economies and the group of developing economies to be affected by global common factors in a heterogenous fashion, which would justify the use of weighted averages of advanced economies and weighted averages of developing economies instead of one cross sectional average per variable.

4 Monte Carlo Experiments: Small Sample Properties of $\widehat{\phi}_{ii,CALS}$

4.1 Monte Carlo Design

We assume one variable per cross section unit, $k = 1$ and $\mathbf{x}_t = (x_{1t}, ..., x_{Nt})'$, and a single unobserved common factor f_t $(m = 1)$. Vector of factor loadings is denoted by $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_N)'$. The data generating process (DGP) used is given by

$$\mathbf{x}_t - \boldsymbol{\gamma} f_t = \boldsymbol{\Phi} \left(\mathbf{x}_{t-1} - \boldsymbol{\gamma} f_{t-1} \right) + \mathbf{u}_t, \tag{126}$$

which corresponds to the system (55) with $\boldsymbol{\alpha} = \mathbf{0}$. The common factor is generated according to the following AR(1) process

$$f_t^{(r)} = \rho_f f_{t-1}^{(r)} + \eta_{ft}^{(r)}, \, \eta_{ft}^{(r)} \sim IIDN \left(0, 1 - \rho_f^2 \right),$$

where the superscript r denotes replications, $r \in \{1, .., R\}$. The unobserved common factor, f_t, is chosen to be relatively persistent, $\rho_f = 0.9$.

In order to construct the $N \times N$ dimensional matrix $\boldsymbol{\Phi}$, random vectors $\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2}, ..., \delta_{iN})'$ are generated first:

$$\delta_{ij} = \frac{\varsigma_{ij}}{\sum_{j=1}^{N} \varsigma_{ij}}, \tag{127}$$

where $\varsigma_{ij} \sim IIDU \left[0, 1 \right]$. Matrix $\boldsymbol{\Phi}$ is then constructed as follows (recall that sufficient condition for stationarity is $\left\| \boldsymbol{\Phi} \right\|_r < 1$):

1. (Diagonal elements) $\phi_{ii} = \phi = 0.5$ for any $i = 1, .., N$.
2. (Off-diagonal elements) For $\forall i \neq j$: $\phi_{ij} = \lambda_i \delta_{ij}$ where $\lambda_i \sim IIDU \left(-0.1, 0.3 \right)$. Case where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)' = \mathbf{0}$ (that is the matrix $\boldsymbol{\Phi}$ is diagonal) is considered as well.

The above parameters yield $\left\| \boldsymbol{\Phi} \right\|_r \leq 0.8$, which, together with $\left| \rho_f \right| < 1$, ensure that the DGP is stationary.

The star variables, x_{it}^* for $i = 1, 2, ..., N$, are constructed as simple arithmetic cross sectional average of remaining units. N-dimensional vector of error terms for the r-th replication, denoted as $\mathbf{u}_t^{(r)}$, is generated using the following Spatial Autoregressive Model (SAR):

$$\mathbf{u}_t^{(r)} = \rho_u \mathbf{S}_u \mathbf{u}_t^{(r)} + \boldsymbol{\varsigma}_t^{(r)}, \tag{128}$$

for $t = 1, .., T$, where the spatial weights matrix is

$$\mathbf{S}_u = \begin{pmatrix} 0 & 1 & & 0 \\ 1 & 0 & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & & 1 & 0 \end{pmatrix}. \tag{129}$$

Two scenarios are investigated; a low Cross Sectional Dependence (CSD) $\rho_u = 0.2$, and a high CSD case $\rho_u = 0.4$. Note that for $\rho_u = 0.5$, $\mathbf{u}_t^{(r)}$ given by the SAR model (83) would no longer be CWD. $\varsigma_{it}^{(r)}$, the i^{th}

element of $\boldsymbol{\varsigma}_t^{(r)}$, is drawn from $IIDN\left(0, \sigma_\varsigma^2\right)$. Variance of $\mathbf{u}_t^{(r)}$ can be written as

$$Var\left(\mathbf{u}_t^{(r)}\right) = \sigma_\varsigma^2 \mathbf{R}\mathbf{R}', \tag{130}$$

where $N \times N$ matrix $\mathbf{R} = \left(\mathbf{I}_N - \rho_u \mathbf{S}_u\right)^{-1}$. Variance σ_ς^2 is chosen so that the variance of $u_{it}^{(r)}$ on average is $1 - \phi^2$, that is [37]

$$\sigma_\varsigma^2 = \left(1 - \phi^2\right) \frac{N}{\sum_{i=1}^N \sum_{j=1}^N r_{ij}^2}, \tag{131}$$

where r_{ij} is the $(i,j)^{th}$ element of the matrix $\mathbf{R}$. The factor loadings, γ_i, $i = 1, 2, ..., N$, are generated from $IIDN(2, 1)$, which asymptotically yields an R^2 for the regression of x_{it} on f_t of around 80% on average. DGP without common factors, that is $\gamma_i = 0$ for $i = 1, 2, ..., N$, is also considered.

In order to minimize the effects of the initial values, the first 20 observations are dropped. $N \in \{1, 10, 25, 50, 75, 100, 200\}$ and $T \in \{25, 50, 75, 100, 200\}$. For each N, all parameters were set at the beginning of the experiments and 2000 replications were carried out. The focus of the experiments is to evaluate the small sample properties of $\widehat{\phi}_{ii,CALS}$ as the estimator of ϕ defined in (83) under alternative DGP's.[38] Since it does not matter which cross section unit i we look at, we chose $i = 1$, and, for the simplicity, we denote the corresponding CALS estimator of ϕ simply as $\widehat{\phi}_{CALS}$ in the exposition below (i.e. we drop the subscript 11). An intercept was included in all the individual regressions when estimating ϕ.

The same Monte Carlo experiments were also conducted for the principal component(s) augmented least squares estimator (or simply PCALS), denoted as $\widehat{\phi}_{ii,PCALS}$, which is similar to $\widehat{\phi}_{ii,CALS}$ defined in (103) but the first principal component of $\mathbf{x}_t^d$, defined in (125), is used instead of star variable x_{it}^*.

4.2 Monte Carlo Results

Tables 2-5 gives the bias (×100) and RMSE (×100) of $\widehat{\phi}_{CALS}$, as well as size ($H_0 : \phi = 0.5$) and power ($H_1 : \phi = 0.7$) at the 5% nominal level of tests based on $\widehat{\phi}_{CALS}$ for a number of different experiments with a relatively high spatial error dependence where $\rho_u = 0.4$. The different experiments relate to different treatment of factor loadings (zero or random) and the values chosen for the off-diagonal elements $\mathbf{\Phi}$ (zero or random). Experiments for the case of low spatial cross sectional dependence with $\rho_u = 0.2$ yield similar results and are reported in a Supplement available from the authors. The variance of $\widehat{\phi}_{CALS}$ is computed as $(2, 2)$ of the 4×4 matrix $\widehat{\sigma}_{11} \widehat{\mathbf{C}}_1^{-1}/T$ where $\widehat{\sigma}_{11}$ and $\widehat{\mathbf{C}}_1$ are defined in (85).

Estimated average pair-wise cross sectional correlation of $\{x_{it}\}$ across experiments is reported in Table 1. The degree of cross section correlation depends very much on the presence of common factor. In cases where $\boldsymbol{\gamma} \neq \mathbf{0}$, the average of pair-wise cross section correlations lie in the range 53% to 72%. Experiments with diagonal matrix $\mathbf{\Phi}$ ($\boldsymbol{\lambda} = \mathbf{0}$) have marginally lower average pair-wise cross section correlation as compared to the case where $\boldsymbol{\lambda} \neq \mathbf{0}$. The most noticeable difference is when $N = 10$ and there are no common factors (44% versus 36%). Considering the models without a common factor, it can be inferred from Table 1 that the average pair-wise cross correlation decreases as N increases, in line with our theory's prediction since the process $\{x_{it}\}$ is CWD for $\boldsymbol{\gamma} = \mathbf{0}$. For $N = 10$, average pair-wise cross correlation is still quite high, in the range of 35-46%, even without a common factor. This cross correlation is high predominantly due to the spatial dependence of errors as opposed to the cross correlations originating from the temporal cross dependence captured by the off-diagonal elements of $\mathbf{\Phi}$.

[37] Note that $Var\left(u_{it}^{(r)}\right) = \sigma_\varsigma^2 \sum_{j=1}^N r_{ij}^2$.

[38] Note that for $k = 1$, $r = 1$ and $\widehat{\boldsymbol{\pi}}_i = \widehat{\boldsymbol{\pi}}_{ir}$, hence we can drop the subscript r because there is only one endogenous variable per group.

Table 2 reports the summary results of the experiments carried out for the baseline case where $\boldsymbol{\gamma} \neq \mathbf{0}$ (common factor is present in the DGP) and $\boldsymbol{\lambda} \neq \mathbf{0}$ (matrix $\boldsymbol{\Phi}$ is not restricted to be diagonal). It is seen that the cross section augmented least squares estimator of ϕ, namely $\widehat{\phi}_{CALS}$, performs quite well for large T, even in the endogenous system with N being as small as 10. For small values of T, there is a negative bias, and the test based on $\widehat{\phi}_{CALS}$ is oversized. This is the familiar time series bias where even in the absence of cross section dependence the OLS estimator of ϕ will be biased downward for a finite T. When $N = 1$, and in all other regressions that are not augmented by star variables, the OLS estimator of ϕ, denoted as $\widehat{\phi}_{OLS}$ (again, we focus on the regression for unit $i = 1$, but drop the subscript for simplicity) shows considerable bias and relatively large RMSE due to the omition of f_t.[39] Power for alternatives in the interval $\langle 0, 1\rangle$ is reported in Figure 2.

Moving on to the experiments without a common factor but with $\boldsymbol{\lambda} \neq \mathbf{0}$, it can be seen from Table 3 that the performance of the LS estimators based on regressions with or without star variables is relatively good for N larger than 10. For $N = 1$, $\widehat{\phi}_{OLS}$ is consistent as $T \to \infty$. For a fixed $N > 1$, $\widehat{\phi}_{OLS}$ is inconsistent as $T \to \infty$. In particular, for $T = 200$ and $N = 10$ the tests based on the OLS estimator of ϕ are oversized and the OLS estimates based on the regressions without star variables exhibit a positive bias. Note that x_{it} is more persistent for $N = 10$ as opposed to the case where $N = 1$, simply because the maximum eigenvalue of the matrix $\boldsymbol{\Phi}$ is bounded by 0.8 (for $N > 1$), while it is $\phi_{11} = 0.5$ for $N = 1$. Interestingly, the inclusion of the star variables in the regressions result in a slight improvement of the size of the test when $N = 10$, while the RMSE is lower for the regressions without the star variables. As N increases, OLS estimators based on regressions without star variables achieve the correct size even for T as small as 25, while in the case of regressions with (redundant) star variables (redundant because common factor is absent), the test is slightly oversized for $T = 25$, while for $T > 25$, the size is relatively good. Comparing the results in Table 3 with those for the experiments where both $\boldsymbol{\gamma} = \mathbf{0}$ (no common factors) and $\boldsymbol{\lambda} = \mathbf{0}$ (matrix $\boldsymbol{\Phi}$ is diagonal) reported in Table 5, the size of the test based on $\widehat{\phi}_{OLS}$ is good for all values of N, including $N = 10$, as the maximum eigenvalue of $\boldsymbol{\Phi}$ in this case is 0.5. Adding the redundant star variables slightly biases the estimates downwards, especially for $T = 25$, and the RMSE is slightly higher.

Finally, in the case of the experiments with a common factor but a diagonal $\boldsymbol{\Phi}$ ($\boldsymbol{\gamma} \neq \mathbf{0}$, $\boldsymbol{\lambda} = \mathbf{0}$), it can be seen from Table 4 that adding star variables is crucial for a consistent estimation of ϕ. In this case the OLS estimator, $\widehat{\phi}_{OLS}$, is biased due to the omitted variable problem. Similar results are obtained in Table 2.

Finally, the alternative estimator, $\widehat{\phi}_{PCALS}$, which uses the first principal component of $\mathbf{x}_t^d$ defined in (125) instead of the star variable x_{it}^*, performs similarly. These Monte Carlo results are provided in a Supplement available from the authors on request.

[39] Due to the inconsistency of $\widehat{\phi}_{OLS}$, the tests based on $\widehat{\phi}_{OLS}$ also show substantial size distortions. To save space, these results are reported in a Supplement available from the authors on request.

Table 1: Average pair-wise cross section correlation (in %) of $\{x_{it}\}$ for experiments with high spatial cross section dependence.

	$\boldsymbol{\gamma} \neq \mathbf{0}$, $\boldsymbol{\lambda} \neq \mathbf{0}$ Baseline case					$\boldsymbol{\gamma} = \mathbf{0}$, $\boldsymbol{\lambda} \neq \mathbf{0}$ No common factor, $\boldsymbol{\Phi}$ is unrestricted				
	T					T				
N	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
10	66.39	69.40	70.26	70.84	71.95	40.74	41.94	42.89	43.18	43.66
25	54.86	59.18	61.07	62.11	64.04	17.40	17.92	18.15	18.13	18.33
50	59.15	64.72	66.85	67.98	69.83	9.63	10.04	10.28	10.34	10.52
75	58.02	63.60	66.09	67.33	69.36	6.22	6.38	6.51	6.49	6.57
100	56.02	60.54	62.55	63.60	65.42	4.91	5.05	5.07	5.17	5.21
200	53.78	59.40	61.68	62.95	64.56	2.48	2.58	2.64	2.64	2.68

	$\boldsymbol{\gamma} \neq \mathbf{0}$, $\boldsymbol{\lambda} = \mathbf{0}$ $\boldsymbol{\Phi}$ is diagonal					$\boldsymbol{\gamma} = \mathbf{0}$, $\boldsymbol{\lambda} = \mathbf{0}$ No common factor, $\boldsymbol{\Phi}$ is diagonal				
	T					T				
N	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
10	65.61	68.12	69.37	69.75	70.68	35.57	35.79	35.84	35.96	35.96
25	54.12	58.78	60.61	61.60	63.39	16.02	16.16	16.35	16.31	16.33
50	58.98	64.60	66.39	67.88	69.72	8.27	8.46	8.49	8.47	8.51
75	57.64	63.86	65.81	67.44	69.14	5.59	5.67	5.69	5.72	5.73
100	55.56	60.65	62.62	63.62	65.57	4.24	4.28	4.32	4.33	4.34
200	54.18	59.24	61.32	62.82	64.51	2.14	2.17	2.16	2.16	2.17

Notes: $\phi = 0.5$ and $\rho_u = 0.4$. Please refer to Section 4 for description of Monte Carlo design.

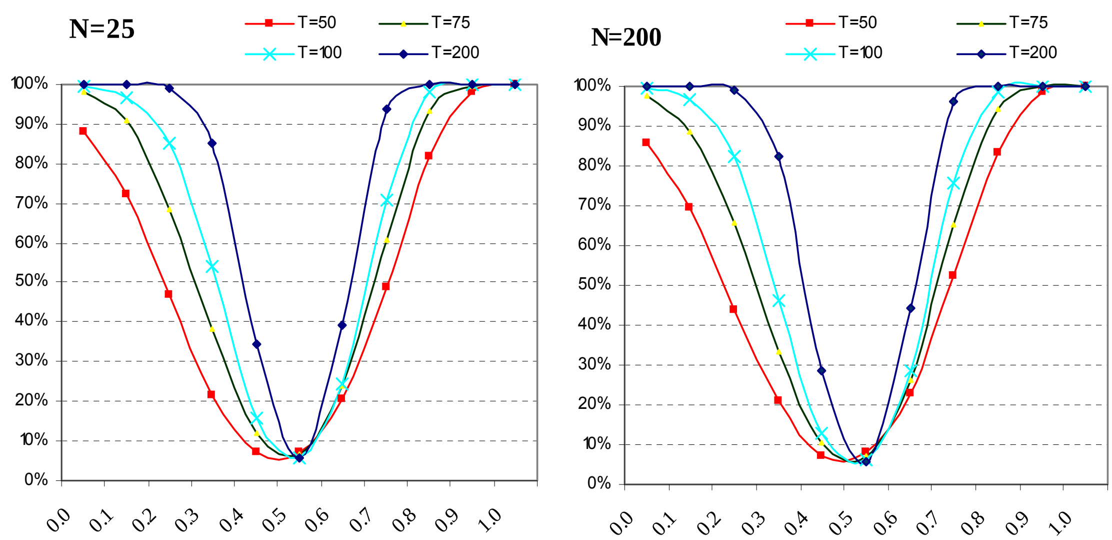

Figure 2: Power of the test based on estimator $\widehat{\phi}_{CALS}$ in the case with $\boldsymbol{\gamma} \neq \mathbf{0}$, $\boldsymbol{\lambda} \neq \mathbf{0}$, and high spatial dependence of errors ($\rho_u = 0.4$).

Table 2: Small sample properties of estimator $\widehat{\phi}_{CALS}$ in the case of high spatial cross section dependence of errors, nonzero factor loadings $\boldsymbol{\gamma} \neq \mathbf{0}$, and $\boldsymbol{\lambda} \neq \mathbf{0}$.

	Bias ($\times$100)					RMSE ($\times$100)				
	T					T				
N	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	$\widehat{\phi}_{CALS}$									
10	-15.06	-8.48	-5.65	-4.38	-2.68	25.21	15.72	11.93	9.91	6.87
25	-12.99	-6.44	-3.72	-2.29	-0.79	23.64	14.78	11.45	9.32	6.37
50	-15.58	-8.58	-6.18	-4.84	-2.95	25.07	15.84	12.17	10.53	7.02
75	-15.15	-8.51	-5.40	-4.49	-2.86	24.82	15.72	11.63	10.41	6.86
100	-15.16	-7.66	-4.91	-4.04	-2.23	25.10	15.09	11.57	9.97	6.70
200	-14.23	-7.31	-4.96	-3.46	-1.68	24.35	15.27	11.81	9.60	6.45
	$\widehat{\phi}_{OLS}$ (OLS regression without star variables)									
1	12.01	21.83	25.21	26.72	29.53	21.95	24.74	26.83	27.86	30.00
10	16.63	25.94	29.23	30.41	33.15	24.38	28.02	30.40	31.26	33.45
25	17.25	25.96	29.27	30.92	33.27	24.80	28.37	30.33	31.67	33.57
50	17.78	26.26	29.78	30.93	33.51	24.62	28.37	30.84	31.68	33.81
75	17.08	26.30	29.92	30.93	33.69	24.31	28.48	31.01	31.75	33.95
100	17.50	26.29	29.03	30.99	33.69	24.38	28.56	30.18	31.78	33.99
200	17.48	25.80	29.62	31.22	33.45	24.32	28.09	30.75	31.98	33.75

	Size ($\times$100)($H_0 : \phi = 0.5$)					Power ($\times$100)($H_1 : \phi = 0.7$)				
	T					T				
N	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	$\widehat{\phi}_{CALS}$									
10	10.85	8.60	7.15	6.60	7.40	38.40	56.60	67.75	79.35	96.65
25	8.65	7.10	6.55	5.85	5.80	34.30	48.75	60.60	70.75	93.90
50	10.30	9.05	7.60	8.35	6.95	38.15	54.25	70.20	79.95	97.85
75	10.05	8.25	6.80	7.70	6.05	38.10	55.65	68.30	77.65	97.20
100	10.00	7.70	6.35	6.85	6.20	37.55	51.90	66.15	77.25	96.20
200	9.50	8.00	7.05	6.15	5.80	35.35	52.35	65.15	75.70	96.10

Notes: $\phi = 0.5$ and $\rho_u = 0.4$. Estimator $\widehat{\phi}_{CALS}$ is defined in (83). Regression for unit $i = 1$ is used to estimate ϕ, but the subscript ii is dropped for the simplicity of exposition. Variance of $\widehat{\phi}_{CALS}$ is given by the element $(2, 2)$ of the 4×4 matrix $\widehat{\sigma}_{11}\widehat{\mathbf{C}}_1^{-1}/T$ where $\widehat{\sigma}_{11}$ and $\widehat{\mathbf{C}}_1$ are defined in (85). Please refer to Section 4 for description of Monte Carlo design.

Table 3: Small sample properties of estimator $\widehat{\phi}_{CALS}$ in the case of high spatial cross section dependence of errors and no common factors ($\boldsymbol{\gamma} = \mathbf{0}$, $\boldsymbol{\lambda} \neq \mathbf{0}$).

	Bias (×100)					RMSE (×100)				
	T					T				
N	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	$\widehat{\phi}_{CALS}$									
10	-14.88	-8.88	-6.15	-5.73	-4.04	24.36	15.92	12.20	10.72	7.39
25	-12.86	-5.96	-4.14	-2.85	-1.50	23.51	14.25	11.38	9.61	6.49
50	-12.52	-6.20	-4.54	-3.31	-1.78	23.10	14.43	11.40	9.30	6.44
75	-13.25	-6.34	-3.98	-3.45	-1.49	23.81	14.80	11.16	9.91	6.49
100	-13.47	-6.31	-3.67	-3.29	-1.23	23.88	14.81	11.05	9.63	6.49
200	-12.90	-6.58	-4.37	-3.04	-1.58	23.61	14.68	11.58	9.51	6.42
	$\widehat{\phi}_{OLS}$ (OLS regression without star variables)									
1	-9.75	-4.46	-3.26	-2.54	-1.17	20.61	13.61	10.90	9.22	6.50
10	-4.72	0.61	2.59	3.45	4.84	18.69	12.27	10.52	9.48	7.69
25	-8.39	-3.23	-0.92	-0.58	0.76	20.01	13.38	10.50	8.91	6.48
50	-7.94	-3.15	-1.38	-0.45	0.89	19.88	13.01	10.29	8.68	6.17
75	-8.34	-4.33	-1.56	-1.21	-0.08	19.77	13.66	10.02	9.07	6.21
100	-9.26	-4.32	-2.66	-1.69	-0.33	20.80	13.52	10.73	8.95	6.25
200	-9.59	-4.53	-2.89	-1.82	-0.69	20.73	13.42	11.05	8.95	6.37

	Size (×100)($H_0 : \phi = 0.5$)					Power (×100)($H_1 : \phi = 0.7$)				
	T					T				
N	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	$\widehat{\phi}_{CALS}$									
10	9.40	8.85	7.55	8.25	8.50	36.75	57.65	70.45	83.80	98.45
25	8.30	6.75	6.20	6.30	5.75	33.80	48.85	62.90	73.75	95.85
50	8.50	6.35	6.45	4.50	5.65	32.45	49.35	65.05	76.80	96.60
75	9.25	7.60	6.00	6.40	6.05	34.00	48.40	62.00	75.15	95.30
100	8.70	6.70	5.50	6.45	5.40	33.90	48.20	60.90	74.20	95.15
200	8.65	6.75	6.40	6.20	5.40	33.20	49.35	62.25	73.20	95.60
	$\widehat{\phi}_{OLS}$ (OLS regression without star variables)									
1	5.25	5.35	5.90	5.25	5.90	26.35	41.85	58.80	72.00	94.65
10	4.40	4.85	7.30	9.10	14.50	19.00	29.20	36.45	45.75	73.30
25	4.95	5.95	5.55	4.90	6.50	23.90	37.65	49.25	63.45	89.30
50	5.10	4.95	4.95	4.35	5.55	22.40	40.10	52.95	64.15	89.80
75	4.95	5.35	4.25	5.40	5.50	23.35	41.80	53.45	65.70	93.15
100	6.15	5.45	5.55	5.05	5.60	26.00	42.20	56.85	69.90	92.95
200	5.90	5.35	5.95	4.85	5.25	25.10	42.40	56.85	69.65	94.05

Notes: $\phi_{11} = 0.5$, $\rho_u = 0.4$ and $\boldsymbol{\gamma} = \mathbf{0}$. Estimator $\widehat{\phi}_{CALS}$ is defined in (83). Regression for unit $i = 1$ is used to estimate ϕ, but the subscript ii is dropped for the simplicity of exposition. Variance of $\widehat{\phi}_{CALS}$ is given by the element $(2, 2)$ of the 4×4 matrix $\widehat{\sigma}_{11}\widehat{\mathbf{C}}_1^{-1}/T$ where $\widehat{\boldsymbol{\sigma}}_{11}$ and $\widehat{\mathbf{C}}_1$ are defined in (85). Please refer to Section 4 for description of Monte Carlo design.

Table 4: Small sample properties of estimator $\widehat{\phi}_{CALS}$ in the case of high spatial cross section dependence of errors, nonzero factor loadings and diagonal matrix $\mathbf{\Phi}$.

	Bias (×100)					RMSE (×100)				
	T					T				
N	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	$\widehat{\phi}_{CALS}$									
10	-11.84	-4.29	-1.75	-0.67	1.78	23.36	13.89	10.65	9.22	6.59
25	-12.70	-5.91	-3.70	-1.80	-0.20	23.94	14.38	11.51	9.01	6.35
50	-14.31	-6.58	-4.64	-3.64	-1.39	24.65	14.70	11.47	9.70	6.35
75	-14.70	-7.07	-4.97	-3.73	-1.70	24.64	15.04	11.63	9.68	6.50
100	-15.02	-7.66	-4.98	-3.61	-1.76	24.66	15.41	11.81	9.63	6.43
200	-14.69	-7.15	-4.95	-3.73	-1.98	24.56	14.97	11.60	9.70	6.71
	$\widehat{\phi}_{OLS}$ (OLS regression without star variables)									
1	12.62	21.79	25.20	26.51	29.63	22.21	24.87	26.84	27.59	30.06
10	16.70	25.13	28.53	30.51	32.69	24.09	27.43	29.77	31.29	33.02
25	17.09	25.82	29.26	30.85	33.12	24.47	28.21	30.42	31.65	33.45
50	17.92	25.78	29.36	31.15	33.18	25.06	28.08	30.50	31.92	33.51
75	17.35	26.33	29.57	30.99	33.37	24.81	28.46	30.67	31.76	33.67
100	17.19	26.36	29.42	31.09	33.40	24.56	28.53	30.56	31.85	33.72
200	17.49	26.62	29.48	30.92	33.77	24.52	29.68	30.68	31.70	34.07

	Size (×100)($H_0 : \phi = 0.5$)					Power (×100)($H_1 : \phi = 0.7$)				
	T					T				
N	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	$\widehat{\phi}_{CALS}$									
10	8.70	6.60	5.75	5.85	7.75	31.95	42.25	52.80	62.95	85.80
25	9.30	6.80	7.25	4.85	5.10	32.60	45.90	61.65	70.25	92.35
50	10.00	6.90	6.55	6.40	5.40	36.80	49.80	63.65	76.40	95.20
75	9.80	7.40	5.90	6.00	5.80	36.05	50.90	65.85	77.20	95.35
100	10.45	7.80	6.80	5.75	4.95	36.35	53.10	65.50	75.30	96.05
200	9.75	7.60	6.15	6.10	6.10	36.65	51.30	64.75	77.05	96.05

Notes: $\phi = 0.5$, $\rho_u = 0.4$ and $\lambda_i = 0$ for $i \in \{1, .., N\}$. Estimator $\widehat{\phi}_{CALS}$ is defined in (83). Regression for unit $i = 1$ is used to estimate ϕ, but the subscript ii is dropped for the simplicity of exposition. Variance of $\widehat{\phi}_{CALS}$ is given by the element $(2, 2)$ of the 4×4 matrix $\widehat{\sigma}_{11}\widehat{\mathbf{C}}_1^{-1}/T$ where $\widehat{\sigma}_{11}$ and $\widehat{\mathbf{C}}_1$ are defined in (85). Please refer to Section 4 for description of Monte Carlo design.

Table 5: Small sample properties of estimator $\widehat{\phi}_{CALS}$ in the case of high spatial cross section dependence of errors, no common factors and diagonal matrix $\mathbf{\Phi}$.

	Bias (×100)					RMSE (×100)				
	T					T				
N	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	$\widehat{\phi}_{CALS}$									
10	-12.08	-5.97	-4.03	-3.14	-1.75	23.03	14.36	11.09	9.51	6.59
25	-12.38	-6.30	-4.14	-3.05	-1.47	23.18	14.59	11.19	9.62	6.26
50	-12.57	-6.25	-3.99	-3.26	-1.34	24.05	14.66	11.39	9.58	6.34
75	-12.16	-5.80	-4.32	-3.34	-1.46	23.17	13.98	11.48	9.45	6.24
100	-12.48	-6.05	-4.05	-2.78	-1.69	23.64	14.37	11.37	9.21	6.36
200	-11.90	-5.69	-4.52	-3.21	-1.51	23.14	14.17	11.49	9.50	6.39
	$\widehat{\phi}_{OLS}$ (OLS regression without star variables)									
1	-9.43	-5.30	-3.17	-2.55	-1.14	20.44	13.70	10.66	9.29	6.18
10	-9.94	-5.49	-3.10	-2.70	-1.08	21.29	13.44	10.60	9.28	6.24
25	-9.65	-5.20	-3.48	-2.46	-1.18	20.67	13.87	11.09	9.08	6.37
50	-9.86	-4.99	-3.21	-2.11	-1.04	21.31	13.58	10.63	9.02	6.43
75	-9.51	-4.77	-3.09	-2.41	-1.12	20.56	13.51	10.89	9.35	6.30
100	-9.92	-5.28	-3.19	-2.74	-1.22	20.78	13.85	10.98	9.43	6.45
200	-10.30	-5.24	-2.98	-2.53	-1.33	21.31	13.93	10.73	9.44	6.44

	Size (×100)($H_0 : \phi = 0.5$)					Power (×100)($H_1 : \phi = 0.7$)				
	T					T				
N	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	$\widehat{\phi}_{CALS}$									
10	8.20	6.90	5.15	5.50	5.80	31.75	48.25	63.05	74.95	95.25
25	8.50	6.70	6.10	6.25	4.85	32.15	48.80	63.00	72.50	95.40
50	10.30	7.25	6.05	5.95	4.90	32.65	48.75	62.00	75.35	94.65
75	8.20	5.30	7.00	5.00	5.35	32.05	47.35	63.25	75.05	96.00
100	8.50	6.20	5.90	5.10	4.65	32.20	46.65	62.15	73.40	96.55
200	8.85	6.20	6.80	5.85	5.35	31.50	47.05	64.10	73.85	95.80
	$\widehat{\phi}_{OLS}$ (OLS regression without star variables)									
1	5.15	5.90	4.30	5.70	4.35	25.80	44.65	58.95	72.20	95.60
10	6.20	5.00	5.15	5.40	4.90	26.90	46.55	58.80	72.85	94.95
25	5.10	5.40	6.05	5.05	5.30	26.90	45.15	60.20	72.10	95.30
50	6.60	5.20	5.60	5.50	5.05	27.30	44.95	59.05	70.75	94.45
75	4.95	4.85	5.65	5.05	5.30	24.80	44.10	59.30	70.75	95.55
100	5.20	5.85	5.55	5.90	5.15	26.45	44.15	58.65	73.00	94.50
200	5.85	5.85	5.40	5.50	5.75	27.55	45.40	58.90	71.20	95.25

Notes: $\phi = 0.5$, $\rho_u = 0.4$, $\boldsymbol{\gamma} = \mathbf{0}$ and $\boldsymbol{\lambda} = \mathbf{0}$. Estimator $\widehat{\phi}_{CALS}$ is defined in (83). Regression for unit $i = 1$ is used to estimate ϕ, but the subscript ii is dropped for the simplicity of exposition. Variance of $\widehat{\phi}_{CALS}$ is given by the element $(2, 2)$ of the 4×4 matrix $\widehat{\sigma}_{11}\widehat{\mathbf{C}}_1^{-1}/T$ where $\widehat{\sigma}_{11}$ and $\widehat{\mathbf{C}}_1$ are defined in (85). Please refer to Section 4 for description of Monte Carlo design.

5 An Empirical Application

In this section we consider an application of the IVAR methodology to one of the long standing questions in the growth literature: does higher investment as a share of GDP causes (predicts) higher growth, or is it higher growth that causes (predicts) a rise in investment-output ratio, or does the direction of causality (predictability) run both ways?[40]

In what follows we re-examine this relationship using data on output and investment from the Penn World Table Version 6.2 database. Specifically we measure output growth, denoted by $x_{1it} = \Delta y_{it}$, as the first difference of the logarithm of Real GDP per capita, Constant Prices: Laspeyres, and the investment-output ratio, denoted by $x_{2it} = inv_{it} - y_{it}$, computed as the logarithm of Investment Share of RGDPL. Out of 188 countries in this database, 98 have uninterrupted time series over the period 1961-2003, thus providing us with a balanced panel composed of $N = 98$ countries and $T = 43$ time periods. It will be assumed that the processes generating Δy_{it} and $inv_{it} - y_{it}$ are covariance stationary.[41]

Our theoretical framework allows us to (i) investigate the presence of dominant effects by testing the joint significance of the star variables in the relationships between x_{1it} and x_{2it}, (ii) test whether, after controlling for the dominant effects, idiosyncratic innovations to Δy_{it} can help forecast future values of $inv_{it} - y_{it}$, and *vice versa*. Granger causality tests not augmented with cross section averages applied to Δy_{it} and $inv_{it} - y_{it}$ are also conducted. We start by testing for the presence of dominant effects.

5.1 Presence of Dominant Effects

It is initially assumed that $\{\Delta y_{it}\}_{i=1}^{N}$ are endogenously determined in a VAR model as defined by (55) with up to one unobserved common factor where $\mathbf{x}_t = (\Delta y_{1t}, \Delta y_{2t}, ..., \Delta y_{Nt})'$. Under the assumptions of Theorem 1 the following cross section augmented regressions can be estimated consistently by least squares for each country i;

$$\Delta y_{it} = a_i + \sum_{\ell=1}^{p_i} b_{i\ell} \Delta y_{i,t-\ell} + \sum_{\ell=0}^{q_i} c_{i\ell} \Delta y^*_{i,t-\ell} + u_{it}, \text{ for } i = 1, 2, ..., N, \tag{132}$$

where Δy^*_{it} is a cross section average of output growth. Presence of dominant effects is tested by conducting Wald tests of the joint significance of the coefficients for the star variables in (132), namely

$$H_0^G : c_{i\ell} = 0 \text{ for } \ell \in \{0, ..., q_i\}. \tag{133}$$

Assuming (55) is the DGP and N is sufficiently large, the null hypothesis (133) holds only for countries with zero factor loadings.[42] The same exercise is carried out for the investment-output ratio based on the regressions

$$inv_{it} - y_{it} = d_i + \sum_{\ell=1}^{p_i} e_{i\ell} \left(inv_{i,t-\ell} - y_{i,t-\ell}\right) + \sum_{\ell=0}^{q_i} f_{i\ell} \left(inv^*_{i,t-\ell} - y^*_{i,t-\ell}\right) + v_{it} \tag{134}$$

[40] A survey of the recent literature on the relationship between growth and investment is provided in Bond *et al.* (2004, Section 2).

[41] This assumption is supported by the CIPS panel unit root test recently proposed in Pesaran (2007) that allows for cross section dependence. Using the CIPS test with one lag and individual-specific intercepts the null hypothesis of a unit root in output growth and log investment output ratios were rejected at 1% and 5% nominal levels, respectively.

[42] This directly follows from Theorem 1.

The joint null hypothesis of interest in this case is given by

$$H_0^{IY}: f_{i\ell} = 0 \text{ for } \ell \in \{0, ..., q_i\}. \tag{135}$$

Results for lag orders equal to one are reported. We also tried higher order lags but found the additional lags to be generally unimportant.[43] Two options are considered for the star variables: *i*) country specific simple cross section averages constructed as the arithmetic average of foreign variables, for example

$$y^*_{s,it} = (N-1)^{-1} \sum_{j=1, j\neq i}^{N} y_{jt},$$

and *ii*) country specific trade weighted cross section averages constructed for example as

$$y^*_{w,it} = \sum_{j=1}^{N} w_{ij} y_{jt}, \text{ with } w_{ii} = 0,$$

where w_{ij} is the share of country j^{th} trade in country i, estimated as the 1991-93 average based on foreign trade flows taken from IMF DOTS database.

The test results are summarized in Table 6, and give the fraction of countries for which the null hypothesis is rejected at the 5 and 10 percent nominal levels. Figure 3 plots the fraction of rejections as a function of the nominal size of the tests. Fractions of rejections are not very high, but well above the nominal levels of the underlying tests. For example, in the case of regressions augmented by simple cross section averages and using 10% nominal level, the fraction of rejections is 32.7% when testing H_0^G and 30.6% when testing H_0^{IY}. Higher fractions of rejections are noted for regressions augmented with trade weighted cross section averages. In particular, 40.8% of the tests reject H_0^G and 44.9% of the tests reject H_0^{IY}, using again the 10% nominal level in both cases. Trade weighted cross section averages seem to be more appropriate in small samples and will be used what follows.

Table 6: Fraction of the tests for which the null of no dominant effects was rejected (in %).

	Choice of Cross Section Averages			
	Simple ($y^*_{s,it}$)		Trade Weighted ($y^*_{w,it}$)	
Nominal level of tests:	5%	10%	5%	10%
Null hypothesis H_0^G	20.4	32.7	33.7	40.8
Null hypothesis H_0^{IY}	22.4	30.6	35.7	44.9

Notes: The underlying regressions are given by (132) and (134). Lag orders (p_i,q_i) are set equal to one.

[43] Sensitivity of the test results to the different choices of lag orders is investigated in a Supplement available from the authors on request.

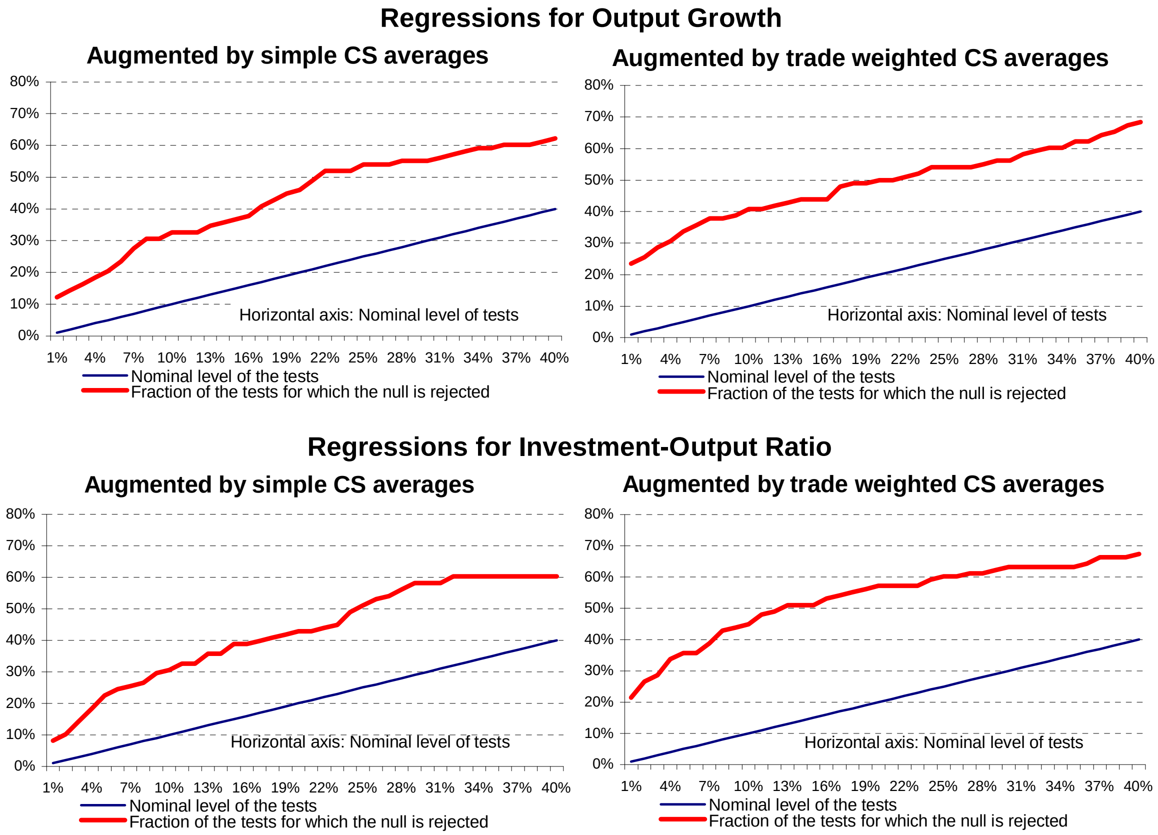

Figure 3: Fractions of countries for which the joint significance of star variables (the null hypotheses H_0^G and H_0^{IY}) was rejected.

5.2 Granger Non-causality Tests

5.2.1 Without cross section augmentation

Initially we construct 'Granger non-causality' tests without cross section augmentation using the following bivariate country-specific VAR models in $\mathbf{x}_{it} \equiv (\Delta y_{it}, inv_{it} - y_{it})'$,

$$\mathbf{x}_{it} = \mathbf{d}_i + \sum_{\ell=1}^{p_i} \mathbf{\Phi}_{\ell ii} \mathbf{x}_{i,t-\ell} + \boldsymbol{\varepsilon}_{it}. \tag{136}$$

Hypotheses of interest are:

1. $inv_{it} - y_{it}$ does not 'Granger cause' Δy_{it} (denoted by $inv_{it} - y_{it} \nrightarrow \Delta y_{it}$),

$$H_0^a : \phi_{12}^{(\ell,ii)} = 0 \text{ for } \ell \leq p_i, \tag{137}$$

2. Δy_{it} does not 'Granger cause' $inv_{it} - y_{it}$ (denoted by $\Delta y_{it} \nrightarrow inv_{it} - y_{it}$),

$$H_0^b : \phi_{21}^{(\ell,ii)} = 0 \text{ for } \ell \leq p_i, \tag{138}$$

where $\phi_{kj}^{(\ell ii)}$ denotes the generic element (k, j) of the matrix $\Phi_{\ell ii}$. The lag order $p_i = 1$ turned out to be sufficient again.[44] Country models (136) do not explicitly control for possible dominant effects.

5.2.2 With cross section augmentation

Suppose now that all the 196 variables in $\mathbf{x}_t = (\mathbf{x}'_{1t}, \mathbf{x}'_{2t}, ..., \mathbf{x}'_{Nt})'$ are endogenously determined within a 196×196 VAR(1) model (55) with up to two unobserved common factors. In this case the country-specific VAR models must be augmented with cross averages, namely we need to consider

$$\mathbf{x}_{it} = \mathbf{d}_i + \sum_{\ell=1}^{p_i} \mathbf{\Phi}_{\ell ii} \mathbf{x}_{i,t-1} + \sum_{\ell=0}^{q_i} \mathbf{B}_{\ell i} \mathbf{x}^*_{i,t-\ell} + \boldsymbol{\varepsilon}_{it} \tag{139}$$

The hypotheses of interest in this case are given by

1. lagged values of $inv_{it} - y_{it}$ are not significant in equations for Δy_{it},

$$H_0^c : \phi_{12}^{(\ell ii)} = 0 \text{ for } \ell \leq p_i \text{ in the regression (139).} \tag{140}$$

2. lagged values of Δy_{it} are not significant in equation for $inv_{it} - y_{it}$,

$$H_0^d : \phi_{21}^{(\ell ii)} = 0 \text{ for } \ell \leq p_i \text{ in the regression (139).} \tag{141}$$

5.2.3 Results of Granger non-causality tests

The results of Granger non-causality tests with and without cross section augmentations are summarized in Table 7 and Figure 4. Without CS augmentation the null hypothesis that $inv_{it} - y_{it}$ does not 'Granger cause' Δy_{it} is rejected in the case of 23.5% of countries at the 10% nominal level, while the hypothesis that Δy_{it} does not Granger cause $inv_{it} - y_{it}$ is rejected in the case of 36.7% of the 98 countries in the sample. With CS augmentation the rejection rates (again at the 10% nominal level) are 26.5% and 25.5%, respectively. In the case of $inv_{it} - y_{it} \nrightarrow \Delta y_{it}$ the test results do not seem to be much affected by cross section augmentation. But for the reverse hypothesis $\Delta y_{it} \nrightarrow inv_{it} - y_{it}$, the results seem to suggest that one would have somewhat exaggerated the importance of country-specific output growth for investment by ignoring the possible effects of common factors.

Table 7: Granger non-causality tests with and without cross section augmentation.

Fraction of tests for which the null hypothesis is rejected (in %)					
Without CS augmentation			**With CS augmentation**		
Nominal level of tests:	5%	10%		5%	10%
Hypothesis			Hypothesis		
H_0^a $(inv_{it} - y_{it} \nrightarrow \Delta y_{it})$	16.3	23.5	H_0^c	15.3	26.5
H_0^b $(\Delta y_{it} \nrightarrow inv_{it} - y_{it})$	25.5	36.7	H_0^d	18.4	25.5

Notes: Lag orders equal to one. Hypotheses H_0^a, H_0^b, H_0^c and H_0^d are defined in (137), (138), (140) and (141), respectively. Granger causality tests without cross section augmentation use bivariate VAR models (136). Granger causality tests with cross section augmentation use bivariate VAR models (139).

[44] Sensitivity of the test results to the different choices of lag orders is investigated in a Supplement available from the authors on request.

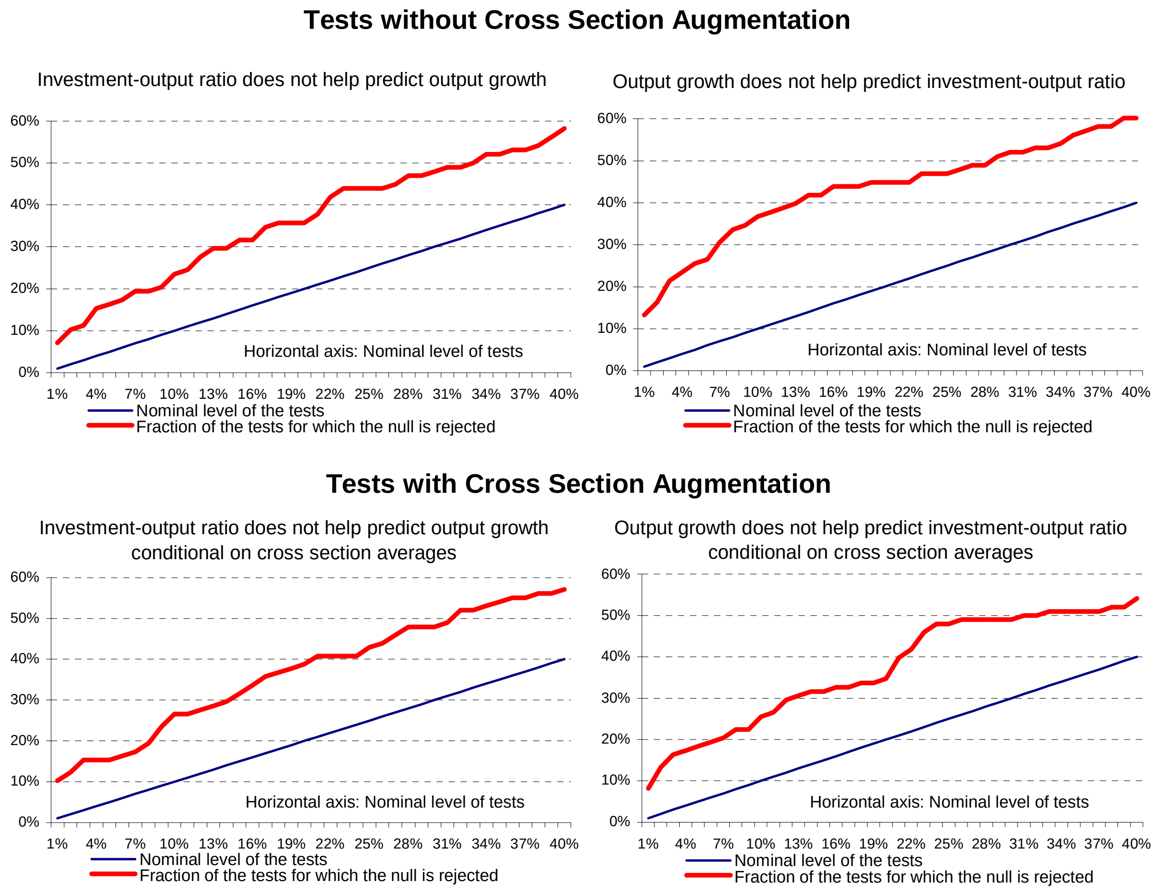

Figure 4: Granger non-causality tests with and without cross section augmentation.

5.3 Summary of Findings

Overall, our empirical analysis suggests that there are considerable heterogeneities across countries. There are statistically significant dominant effects in the relationships for output growth and investment-output ratios. Country-specific cross sectional averages play an important role in about half of the 98 economies under consideration. 'Granger causality' tests without cross section augmentation applied to Δy_{it} and $inv_{it} - y_{it}$ confirm that for a non-negligible fraction, but not the majority of, countries 'Granger causality' goes both ways. Controlling for the possible dominant effects, IY ratio still helps to predict future growth in non-negligible fraction (but not the majority) of counties.

6 Concluding Remarks

This paper proposes restrictions on the coefficients of infinite-dimensional VAR (IVAR) that bind only in the limit as the number of cross section units (or variables in the VAR) tends to infinity to circumvent the curse of dimensionality. The proposed framework relates to the various approaches considered in the literature. For example when modelling individual households or firms, aggregate variables, such as market returns, regional or national income, are treated as exogenous. This is intuitive as the impact of a firm or household on the aggregate economy is small, of order $O\left(N^{-1}\right)$. The paper formalizes this idea in a spatio-dynamic context. It is established that under certain conditions on the order of magnitudes of the coefficients in a large dynamic system, and in the absence of common factors, individual units de-couple from the other units in the system as $N \to \infty$ and can be estimated separately. In the presence of dominant economic agent or unobserved common factors, individual-specific VAR models can still be estimated separately conditioning on observed and unobserved common factors. Unobserved common factors can be approximated by cross sectional averages, following the idea originally introduced in Pesaran (2006).

The paper shows that the GVAR approach can be motivated as an approximation to an IVAR featuring all macroeconomic variables. This is true for stationary models as well as for systems with variables integrated of order one. Asymptotic distribution of the cross sectionally augmented least squares (CALS) estimator of the parameters of the unit-specific models are established both in the case where the number of unobserved common factors is known and when it is unknown but fixed. Small sample properties of the proposed CALS estimator are investigated by Monte Carlo simulations. The proposed estimation and inference techniques are applied to modelling real GDP growth and investment-output ratios.

Topics for future research could include estimation and inference in the case of IVAR models with dominant individual units, analysis of large dynamic networks with and without dominant nodes, and a closer examination of the relationships between IVAR and dynamic factor models.

Appendix

A Lemmas

Lemma 1 *Let Assumptions 8–9, 11, and B1 hold. Then for any $p, q \in \{0,1\}$ and for any $Nk \times 1$ dimensional vectors $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, such that $\|\boldsymbol{\theta}\| = O(1)$ and $\|\boldsymbol{\varphi}\|_c = O(1)$, we have*

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p} \xrightarrow{p} 0, \tag{142}$$

and

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q} \xrightarrow{p} E\left(\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\right), \tag{143}$$

where $\boldsymbol{\upsilon}_t$ is defined in (106). Furthermore, if also $\|\boldsymbol{\theta}\| = O\left(N^{-\frac{1}{2}}\right)$ then

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_{t} \xrightarrow{p} 0, \tag{144}$$

and

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q} \xrightarrow{p} E\left(\sqrt{N}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\right). \tag{145}$$

Proof. Let $T_N = T(N)$ be any increasing integer-valued function of N satisfying Assumption B1. Consider the following two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, defined by

$$\kappa_{Nt} = \frac{1}{T_N}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p},$$

where subscript N is used to denote the number of cross section units,[45] and $\{\mathcal{F}_t\}$ denotes an increasing sequence of σ-fields ($\mathcal{F}_{t-1} \subset \mathcal{F}_t$) such that $\mathcal{F}_t$ includes all information available at time t and κ_{Nt} is measurable with respect to $\mathcal{F}_t$ for any $N \in \mathbb{N}$. Let $\left\{\{c_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ be two-dimensional array of constants and set $c_{Nt} = \frac{1}{T_N}$ for all $t \in \mathbb{Z}$ and $N \in \mathbb{N}$. Note that

$$\begin{aligned} E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-n}\right)\right]^2\right\} &= \sum_{\ell=\mathfrak{m}_{np}}^{\infty} \boldsymbol{\theta}'\boldsymbol{\Phi}^{\ell-p}\boldsymbol{\Sigma}\boldsymbol{\Phi}'^{\ell-p}\boldsymbol{\theta}, \\ &\leq \zeta_n, \end{aligned} \tag{146}$$

where $\mathfrak{m}_{np} = \max\{n, p\}$ and[46]

$$\zeta_n = \|\boldsymbol{\theta}\|^2 \|\boldsymbol{\Sigma}\| \|\boldsymbol{\Phi}\|^{2(\mathfrak{m}_{np}-p)} \sum_{\ell=0}^{\infty} \|\boldsymbol{\Phi}\|^{2\ell}.$$

[45] Note that vectors $\boldsymbol{\upsilon}_t$ and $\boldsymbol{\theta}$ change with N as well, but subscript N is omitted here to keep the notation simple.

[46] We use submultiplicative property of matrix norms ($\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$ for any matrices $\mathbf{A}$, $\mathbf{B}$ such that $\mathbf{AB}$ is well defined) and the fact that the spectral matrix norm is self-adjoint (i.e. $\|\mathbf{A}'\| = \|\mathbf{A}\|$). Note also that by Assumption 9 (and Remark 13), $\|\boldsymbol{\Phi}\|^2 < \rho < 1$. This implies $\sum_{\ell=0}^{\infty} \|\boldsymbol{\Phi}^{\ell}\|^2 = O(1)$.

Using Assumptions 9, and 11, ζ_n has following properties[47]

$$\zeta_0 = O(1), \, \zeta_n \to 0 \text{ as } n \to \infty. \tag{147}$$

By Liapunov's inequality, $E\left|E\left(\kappa_{Nt} \mid \mathcal{F}_{t-n}\right)\right| \leq \sqrt{E\left\{\left[E\left(\kappa_{Nt} \mid \mathcal{F}_{t-n}\right)\right]^2\right\}}$ (Davidson, 1994, Theorem 9.23). It follows that two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to the constant array $\{c_{Nt}\}$. Equations (146) and (147) establish array $\{\kappa_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm. This implies uniform integrability.[48] Note that

$$\lim_{N\to\infty} \sum_{t=1}^{T_N} c_{Nt} = \lim_{N\to\infty} \sum_{t=1}^{T_N} \frac{1}{T_N} = 1 < \infty, \tag{148}$$

$$\lim_{N\to\infty} \sum_{t=1}^{T_N} c_{Nt}^2 = \lim_{N\to\infty} \sum_{t=1}^{T_N} \frac{1}{T_N^2} = 0. \tag{149}$$

Therefore array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ satisfies conditions of a mixingale weak law,[49] which implies $\sum_{t=1}^{T_N} \kappa_{Nt} \xrightarrow{L_1} 0$, i.e.:

$$\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\theta}' \boldsymbol{\upsilon}_{t-p} \xrightarrow{L_1} 0,$$

under Assumption B1. Convergence in L_1 norm implies convergence in probability. This completes the proof of the result (142). Under the condition $\|\boldsymbol{\theta}\| = O\left(N^{-\frac{1}{2}}\right)$, result (144) follows from result (142) by noting that $\left\|\sqrt{N}\boldsymbol{\theta}\right\| = O(1)$.

Result (143) is proved in a similar fashion. Consider the following two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, defined by[50]

$$\kappa_{Nt} = \frac{1}{T_N} \boldsymbol{\theta}' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\upsilon}_{t-q} - \frac{1}{T_N} E\left(\boldsymbol{\theta}' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\upsilon}_{t-q}\right),$$

where as before $T_N = T(N)$ is any increasing integer-valued function of N satisfying Assumption B1. Set $c_{Nt} = \frac{1}{T_N}$ for all $t \in \mathbb{Z}$ and $N \in \mathbb{N}$. Note that

$$\begin{aligned} E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-n}\right) &= E\left(\sum_{s=p}^{\infty} \boldsymbol{\theta}' \boldsymbol{\Phi}^{s-p} \mathbf{u}_{t-s} \sum_{\ell=q}^{\infty} \boldsymbol{\varphi}' \boldsymbol{\Phi}^{\ell-q} \mathbf{u}_{t-\ell} \mid \mathcal{F}_{t-n}\right) - E\left(\boldsymbol{\theta}' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\upsilon}_{t-q}\right), \\ &= \sum_{s=\mathfrak{m}_{np}}^{\infty} \sum_{\ell=\mathfrak{m}_{nq}}^{\infty} \left[\boldsymbol{\theta}' \boldsymbol{\Phi}^{s-p} \mathbf{u}_{t-s} \boldsymbol{\varphi}' \boldsymbol{\Phi}^{\ell-q} \mathbf{u}_{t-\ell} - E\left(\boldsymbol{\theta}' \boldsymbol{\Phi}^{s-p} \mathbf{u}_{t-s} \boldsymbol{\varphi}' \boldsymbol{\Phi}^{\ell-q} \mathbf{u}_{t-\ell}\right)\right]. \end{aligned}$$

Let $\boldsymbol{\theta}_s' = \boldsymbol{\theta}' \boldsymbol{\Phi}^s$ and $\boldsymbol{\varphi}_\ell' = \boldsymbol{\varphi}' \boldsymbol{\Phi}^\ell$.

[47] $\|\boldsymbol{\Sigma}\| = O(1)$ since $\|\boldsymbol{\Sigma}\| \leq \sqrt{\|\boldsymbol{\Sigma}\|_r \|\boldsymbol{\Sigma}\|_c}$, $\|\boldsymbol{\Sigma}\|_c = \|\boldsymbol{\Sigma}\|_r$ (because $\boldsymbol{\Sigma}$ is symmetric) and $\|\boldsymbol{\Sigma}\|_r = O(1)$ by Assumption 11. Furthermore, $\|\boldsymbol{\Phi}\|^2 < \rho < 1$ by Assumption 9 (and Remark 13).

[48] Sufficient condition for uniform integrability is $L_{1+\varepsilon}$ uniform boundedness for any $\varepsilon > 0$.

[49] Davidson (1994, Theorem 19.11).

[50] As before, $\{\mathcal{F}_t\}$ is an increasing sequence of σ-fields ($\mathcal{F}_{t-1} \subset \mathcal{F}_t$) such that $\mathcal{F}_t$ includes all information available at time t and κ_{Nt} is measurable with respect of $\mathcal{F}_t$ for any $N \in \mathbb{N}$.

$$
\begin{aligned}
E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-n}\right)\right]^2\right\} &= \sum_{s=\mathfrak{m}_{pn}}^{\infty}\sum_{\ell=\mathfrak{m}_{qn}}^{\infty}\sum_{j=\mathfrak{m}_{pn}}^{\infty}\sum_{d=\mathfrak{m}_{qn}}^{\infty} E\left(\boldsymbol{\theta}_{s-p}'\mathbf{u}_{t-s}\boldsymbol{\varphi}_{\ell-q}'\mathbf{u}_{t-\ell}\boldsymbol{\theta}_{j-p}'\mathbf{u}_{t-j}\boldsymbol{\varphi}_{d-q}'\mathbf{u}_{t-d}\right) - \\
&\quad - \left(\sum_{s=\mathfrak{m}_{pn}}^{\infty}\sum_{\ell=\mathfrak{m}_{qn}}^{\infty} E\left(\boldsymbol{\theta}_{s-p}'\mathbf{u}_{t-s}\boldsymbol{\varphi}_{\ell-q}'\mathbf{u}_{t-\ell}\right)\right)^2. \qquad (150)
\end{aligned}
$$

Using independence of $\mathbf{u}_t$ and $\mathbf{u}_{t'}$ for any $t \neq t'$ (Assumption 11), we have

$$
\begin{aligned}
\sum_{s=\mathfrak{m}_{pn}}^{\infty}\sum_{\ell=\mathfrak{m}_{qn}}^{\infty} E\left(\boldsymbol{\theta}_{s-p}'\mathbf{u}_{t-s}\boldsymbol{\varphi}_{\ell-q}'\mathbf{u}_{t-\ell}\right) &= \sum_{\ell=\max\{p,q,n\}}^{\infty} \boldsymbol{\theta}'\boldsymbol{\Phi}^{\ell-p}\boldsymbol{\Sigma}\boldsymbol{\Phi}'^{\ell-q}\boldsymbol{\varphi} \\
&\leq \zeta_{a,n},
\end{aligned}
$$

where

$$
\zeta_{a,n} = \|\boldsymbol{\theta}\|\,\|\boldsymbol{\varphi}\|\,\|\boldsymbol{\Sigma}\|\,\|\boldsymbol{\Phi}\|^{\chi_1(p,n,q)}\sum_{\ell=0}^{\infty}\|\boldsymbol{\Phi}\|^{2\ell},
$$

and $\chi_1(p,n,q) = \max\{0, q-p, n-p\} + \max\{0, p-q, n-q\}$. $\|\boldsymbol{\Sigma}\| = O(1)$ by Assumption 11, $\sum_{\ell=0}^{\infty}\|\boldsymbol{\Phi}\|^{2\ell} = O(1)$ by Assumption 9 (and Remark 13), $\|\boldsymbol{\theta}\| = O(1)$, $\|\boldsymbol{\varphi}\| \leq \|\boldsymbol{\varphi}\|_c = O(1)$, and $\zeta_{a,n}$ has following properties

$$
\zeta_{a,0} = O(1), \; \zeta_{a,n} \to 0 \text{ as } n \to \infty. \qquad (151)
$$

Similarly, using the independence of $\mathbf{u}_t$ and $\mathbf{u}_{t'}$ for any $t \neq t'$ (Assumption 11),[51] the first term on the right side of equation (150) is bounded by the following upper bound $\zeta_{b,n}$:

$$
\begin{aligned}
\zeta_{b,n} &= \sup_{i,r\in\mathcal{S}}\|\boldsymbol{\Psi}_{ir}\|\cdot\|\boldsymbol{\theta}\|^2\sum_{\ell=\max\{p,q,n\}}\|\boldsymbol{\Phi}\|^{2(\ell-p)}\left(\boldsymbol{\varphi}'\boldsymbol{\Phi}^{\ell-q}\boldsymbol{\tau}\right)^2 + 2\zeta_{a,n}^2 + \\
&\quad + \|\boldsymbol{\theta}\|^2\|\boldsymbol{\Sigma}\|^2\|\boldsymbol{\varphi}\|^2\|\boldsymbol{\Phi}\|^{2\chi_2(p,n,q)}\left(\sum_{\ell=0}^{\infty}\|\boldsymbol{\Phi}\|^{2\ell}\right)^2,
\end{aligned}
$$

where $\chi_2(p,n,q) = \max\{0, n-p\} + \max\{n-q, 0\}$. Note that

$$
\left(\boldsymbol{\varphi}'\boldsymbol{\Phi}^{\ell}\boldsymbol{\tau}\right)^2 \leq \left\|\boldsymbol{\varphi}'\boldsymbol{\Phi}^{\ell}\boldsymbol{\tau}\right\|_r^2 \leq \|\boldsymbol{\varphi}\|_c^2\left\|\boldsymbol{\Phi}^{\ell}\boldsymbol{\tau}\right\|_r^2 = O(1), \text{ for any } \ell \in \mathbb{N}, \qquad (152)
$$

where $\|\boldsymbol{\varphi}\|_c^2 = O(1)$, and $\left\|\boldsymbol{\Phi}^{\ell}\boldsymbol{\tau}\right\|_r^2 < K$ is established in Lemma 2 (constant K is independent of N and ℓ). It follows from equation (152) and Assumptions 9 and 11 that $\zeta_{b,n}$ has following properties

$$
\zeta_{b,0} = O(1), \; \zeta_{b,n} \to 0 \text{ as } n \to \infty. \qquad (153)
$$

$E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-n}\right)\right]^2\right\}$ is therefore bounded by $\zeta_n = \zeta_{a,n} + \zeta_{b,n}$. Equations (151) and (153) establish

$$
\zeta_0 = O(1), \; \zeta_n \to 0 \text{ as } n \to \infty. \qquad (154)
$$

By Liapunov's inequality, $E\left|E\left(\kappa_{Nt} \mid \mathcal{F}_{t-n}\right)\right| \leq \sqrt{E\left\{\left[E\left(\kappa_{Nt} \mid \mathcal{F}_{t-n}\right)\right]^2\right\}}$ (Davidson, 1994, Theorem 9.23). It

[51] $E\left(\boldsymbol{\theta}_{s-p}'\mathbf{u}_{t-s}\boldsymbol{\vartheta}_{\ell-q}'\mathbf{u}_{t-\ell}\boldsymbol{\theta}_{j-p}'\mathbf{u}_{t-j}\boldsymbol{\vartheta}_{d-q}'\mathbf{u}_{t-d}\right)$ is nonzero only if one of the following four cases: *i)* $s = \ell = j = d$, *ii)* $s = \ell$, $\ell \neq j$, and $j = d$, *iii)* $s = j$, $j \neq \ell$, and $\ell = d$, or *iv)* $s = d$, $d \neq \ell$, and $\ell = j$.

follows that two-dimensional array $\left\{\{\kappa_{Nt},\mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, is L_1-mixingale with respect to a constant array $\{c_{Nt}\}$. Furthermore, (154) establishes array $\{\kappa_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm. This implies uniform integrability.[52] Since also equations (148) and (149) hold, array $\left\{\{\kappa_{Nt},\mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ satisfies conditions of a mixingale weak law,[53] which implies $\sum_{t=1}^{T_N}\kappa_{Nt}\xrightarrow{L_1}0$, i.e.:

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\xrightarrow{L_1}E\left(\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\right),$$

under Assumption B1. Convergence in L_1 norm implies convergence in probability. This completes the proof of result (143). Under the condition $\|\boldsymbol{\theta}\|=O\left(N^{-\frac{1}{2}}\right)$, result (145) follows from result (143) by noting that $\left\|\sqrt{N}\boldsymbol{\theta}\right\|=O(1)$. ■

Lemma 2 *Let matrix $\mathbf{\Phi}$ satisfy Assumptions 8 and 9. Then there exist a constant $K<\infty$ independent of $N\in\mathbb{N}$ and $\ell\in\mathbb{N}_0$ such that*

$$\left\|\mathbf{\Phi}^{\ell}\boldsymbol{\tau}\right\|_r<K, \tag{155}$$

where $\boldsymbol{\tau}$ is $Nk\times 1$ dimensional vector of ones.

Proof. Let $\mathbf{S}_i=(\mathbf{0}_{k\times k},..,\mathbf{0}_{k\times k},\mathbf{I}_k,\mathbf{0}_{k\times k},..,\mathbf{0}_{k\times k})'$ be $kN\times k$ dimensional selection matrix for cross section unit i. Following equality holds

$$\mathbf{S}_i'\mathbf{\Phi}^{\ell}=\mathbf{\Phi}_{ii}^{\ell}\mathbf{S}_i'+\mathbf{\Phi}_{-i}'\sum_{s=1}^{\ell}\mathbf{\Phi}_{ii}^{s-1}\mathbf{\Phi}^{\ell-s}.$$

Denote $\mathbf{D}_\ell=\sum_{s=1}^{\ell}\mathbf{\Phi}_{ii}^{s-1}\mathbf{\Phi}^{\ell-s}$. Assumption 9 (and Remark 13) implies

$$\|\mathbf{D}_\ell\|\leq\sum_{s=1}^{\ell}\left\|\mathbf{\Phi}_{ii}^{s-1}\right\|\left\|\mathbf{\Phi}^{\ell-s}\right\|\leq\sum_{s=1}^{\ell}\ell\rho^{\ell-1}, \tag{156}$$

and $\|\mathbf{D}_\ell\|$ is uniformly bounded (in $N\in\mathbb{N}$ and $\ell\in\mathbb{N}_0$) by a constant $\rho/(1-\rho)^2$.

$$\begin{aligned}\left\|\mathbf{S}_i\mathbf{\Phi}^{\ell}\boldsymbol{\tau}\right\|_r &\leq \left\|\mathbf{\Phi}_{ii}^{\ell}\mathbf{S}_i'\boldsymbol{\tau}\right\|_r+\left\|\mathbf{\Phi}_{-i}'\mathbf{D}_\ell\boldsymbol{\tau}\right\|_r,\\ &\leq \left\|\mathbf{\Phi}_{ii}^{\ell}\mathbf{S}_i'\boldsymbol{\tau}\right\|\sqrt{k}+\left\|\mathbf{\Phi}_{-i}'\mathbf{D}_\ell\boldsymbol{\tau}\right\|\sqrt{k}.\end{aligned}$$

$\left\|\mathbf{\Phi}_{ii}^{\ell}\mathbf{S}_i'\boldsymbol{\tau}\right\|\leq\|\mathbf{\Phi}_{ii}\|^{\ell}\sqrt{k}=O(1)$ is uniformly bounded in $N\in\mathbb{N}$, $\ell\in\mathbb{N}_0$ and $i\leq N$ by Assumption 9 (and Remark 13). $\left\|\mathbf{\Phi}_{-i}'\mathbf{D}_\ell\boldsymbol{\tau}\right\|\leq\left\|\mathbf{\Phi}_{-i}'\right\|\|\mathbf{D}_\ell\|\|\boldsymbol{\tau}\|=O(1)$ is uniformly bounded (in $N\in\mathbb{N}$, $\ell\in\mathbb{N}_0$, and $i\leq N$) by equation (156), and Assumption 8. It follows that there exist a constant K independent of $N\in\mathbb{N}$ and $\ell\in\mathbb{N}_0$ such that

$$\left\|\mathbf{\Phi}^{\ell}\boldsymbol{\tau}\right\|_r\leq\sup_{\ell\in\mathbb{N}}\left\|\mathbf{\Phi}^{\ell}\boldsymbol{\tau}\right\|_r\leq\sup_{\ell\in\mathbb{N}}\sup_{i\leq N,\ N\in\mathbb{N}}\left\|\mathbf{S}_i\mathbf{\Phi}^{\ell}\boldsymbol{\tau}\right\|<K.$$

■

Lemma 3 *Consider model (55), let Assumptions 8, 9, 10–13 and B1 hold, and assume factor loadings are governed either by Assumption 14 or 15. Then for any $p,q\in\{0,1\}$ and for any $kN\times 1$ dimensional vectors*

[52] Sufficient condition for uniform integrability is $L_{1+\varepsilon}$ uniform boundedness for any $\varepsilon>0$.
[53] Davidson (1994, Theorem 19.11).

$\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, such that $\|\boldsymbol{\theta}\|_c = O(1)$ and $\|\boldsymbol{\varphi}\|_c = O(1)$, we have

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\mathbf{x}_{t-p} \xrightarrow{p} E\left(\boldsymbol{\theta}'\mathbf{x}_{t-p} \mid \boldsymbol{\Gamma}\right), \tag{157}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\mathbf{x}_{t-p}\boldsymbol{\varphi}'\mathbf{x}_{t-q} \xrightarrow{p} E\left(\boldsymbol{\theta}'\mathbf{x}_{t-p}\boldsymbol{\varphi}'\mathbf{x}_{t-q} \mid \boldsymbol{\Gamma}\right). \tag{158}$$

Furthermore, for $\|\boldsymbol{\theta}\| = O(1)$ and $\|\boldsymbol{\varphi}\|_c = O(1)$ we have

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\Gamma}\mathbf{f}_{t-q} \xrightarrow{p} 0, \tag{159}$$

where $\boldsymbol{\upsilon}_t$ is defined in (106).

Proof. Let $T_N = T(N)$ be any increasing integer-valued function of N such that Assumption B1 holds. Consider the following two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, defined by

$$\kappa_{Nt} = \frac{1}{T_N}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\Gamma}\mathbf{f}_{t-q},$$

where $\{\mathcal{F}_t\}$ denotes an increasing sequence of σ-fields ($\mathcal{F}_{t-1} \subset \mathcal{F}_t$) such that $\mathcal{F}_t$ includes all information available at time t and κ_{Nt} is measurable with respect to $\mathcal{F}_t$ for any $N \in \mathbb{N}$. Let $\left\{\{c_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ be two-dimensional array of constants and set $c_{Nt} = \frac{1}{T_N}$ for all $t \in \mathbb{Z}$ and $N \in \mathbb{N}$. Using submultiplicative property of matrix norm, and independence of $\mathbf{f}_t$ and $\boldsymbol{\upsilon}_{t'}$ for any $t, t' \in \mathbb{Z}$, we have

$$E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-n}\right)\right]^2\right\} \le \zeta_n,$$

where

$$\zeta_n = \|\boldsymbol{\theta}\|^2\|\boldsymbol{\Sigma}\|\|\boldsymbol{\Phi}\|^{2\max\{0,n-p\}}\sum_{\ell=0}^{\infty}\|\boldsymbol{\Phi}\|^{2\ell}E\left\{\left[E\left(\boldsymbol{\varphi}'\boldsymbol{\Gamma}\mathbf{f}_{t-q} \mid \mathcal{F}_{t-n}\right)\right]^2\right\}.$$

$\|\boldsymbol{\theta}\|^2 = O(1)$, $\|\boldsymbol{\Phi}\| < \rho$ by Assumption 9 (and Remark 13), $\|\boldsymbol{\Sigma}\| \le \sqrt{\|\boldsymbol{\Sigma}\|_c\|\boldsymbol{\Sigma}\|_r} = O(1)$ by Assumption 11, and

$$E\left\{\left[E\left(\boldsymbol{\varphi}'\boldsymbol{\Gamma}\mathbf{f}_{t-q} \mid \mathcal{F}_{t-n}\right)\right]^2\right\} = O(1),$$

since $\mathbf{f}_{t-q}$ is covariance stationary, $\mathbf{f}_t$ is independently distributed of factor loadings $\boldsymbol{\Gamma}$, and $\left\|E\left(\boldsymbol{\varphi}'\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{\varphi}\right)\right\| = O(1)$ by Assumption 14 (or 15). It follows that ζ_n has following properties

$$\zeta_0 = O(1) \text{ and } \zeta_n \to 0 \text{ as } n \to \infty.$$

Array $\{\kappa_{Nt}/c_{Nt}\}$ is thus uniformly bounded in L_2 norm. This proves uniform integrability of array $\{\kappa_{Nt}/c_{Nt}\}$. Furthermore, using Liapunov's inequality, two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$. Noting that equations (148) and (149) hold, it follows that the array $\{\kappa_{Nt}, \mathcal{F}_t\}$ satisfies conditions of a mixingale weak law,[54] which implies $\sum_{t=1}^{T_N}\kappa_{Nt} \xrightarrow{L_1} 0$. Convergence in L_1 norm implies convergence in probability. This completes the proof of result (159).

Assumption 13 implies that sequence $\boldsymbol{\theta}'\boldsymbol{\alpha}$ (as well as $\boldsymbol{\varphi}'\boldsymbol{\alpha}$) is deterministic and bounded. Vector of

[54] Davidson (1994, Theorem 19.11)

endogenous variables $\mathbf{x}_t$ can be written as

$$\mathbf{x}_t = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{f}_t + \boldsymbol{\upsilon}_t.$$

Process $\mathbf{f}_t$ is independent of $\boldsymbol{\upsilon}_t$. Consider now Assumption B1 where $N, T \xrightarrow{j} \infty$, at any rate. Processes $\left\{\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\right\}$ and $\left\{\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\right\}$ are ergodic in mean by Lemma 1 since $\|\boldsymbol{\theta}\| \leq \|\boldsymbol{\theta}\|_c = O(1)$. Furthermore,

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\boldsymbol{\Gamma}\mathbf{f}_t \xrightarrow{p} \boldsymbol{\theta}'\boldsymbol{\Gamma}E(\mathbf{f}_t),$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\boldsymbol{\Gamma}\mathbf{f}_t\boldsymbol{\varphi}'\boldsymbol{\Gamma}\mathbf{f}_{t-q} \xrightarrow{p} \boldsymbol{\theta}'\boldsymbol{\Gamma}E\left(\mathbf{f}_t\mathbf{f}_{t-q}'\right)\boldsymbol{\Gamma}'\boldsymbol{\varphi},$$

since $\mathbf{f}_t$ is covariance stationary $m \times 1$ dimensional process with absolute summable autocovariances ($\mathbf{f}_t$ is ergodic in mean as well as in variance), and

$$\begin{aligned} \left\|E\left(\boldsymbol{\theta}'\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{\varphi}\right)\right\| &= O(1), \\ \left\|E\left[\left(\boldsymbol{\theta}'\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{\varphi}\right)^2\right]\right\| &= O(1), \end{aligned}$$

by Assumption 14 (or 15). Sum of bounded deterministic process and independent processes ergodic in mean is a process that is ergodic in mean as well. This completes the proof. ■

Lemma 4 *Consider model (55). Suppose Assumptions 8-13 hold and factor loadings are governed either by Assumption 14 or 15. Then for any $p, q \in \{0, 1\}$, for any $Nk \times m_w$ dimensional arbitrary matrix of weights $\mathbf{W}$ satisfying conditions (62)-(63), and for any $r \in \{1, .., m_w\}$,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\left(\mathbf{x}_{t-p} - \boldsymbol{\alpha} - \boldsymbol{\Gamma}\mathbf{f}_{t-p}\right) \xrightarrow{p} 0 \text{ under Assumption B2,} \tag{160}$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\left(\mathbf{x}_{t-p} - \boldsymbol{\alpha} - \boldsymbol{\Gamma}\mathbf{f}_{t-p}\right)\mathbf{x}_{t-q}^* \xrightarrow{p} \mathbf{0} \text{ under Assumption B2,} \tag{161}$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\left(\mathbf{x}_{t-p} - \boldsymbol{\alpha} - \boldsymbol{\Gamma}\mathbf{f}_{t-p}\right)\mathbf{x}_{i,t-q} \xrightarrow{p} \mathbf{0} \text{ under Assumption B2,} \tag{162}$$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{h}_{it}' \xrightarrow{p} \mathbf{0} \text{ under Assumption B1,} \tag{163}$$

where $\mathring{\mathbf{w}}_r$ is the r^{th} column vector of matrix $\mathbf{W}$, and vectors $\mathbf{g}_{it}$ and $\mathbf{h}_{it}$ are defined in (82) and (81), respectively.

Proof.

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\left(\mathbf{x}_{t-p} - \boldsymbol{\alpha} - \boldsymbol{\Gamma}\mathbf{f}_{t-p}\right) = \frac{\sqrt{T}}{\sqrt{N}}\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\boldsymbol{\upsilon}_{t-p}$$

where $\boldsymbol{\upsilon}_t$ is defined in equation (106). Under Assumption B2,

$$\frac{\sqrt{T}}{\sqrt{N}} \to \varkappa < \infty. \tag{164}$$

Noting that $\left\| \sqrt{N} \mathring{\mathbf{w}}_r \right\| = O(1)$ by granularity condition (62),

$$\frac{\sqrt{N}}{T} \sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \xrightarrow{p} 0, \tag{165}$$

follows directly from Lemma 1, result (144). This completes the proof of result (160).

Let $\boldsymbol{\varphi}$ be any $Nk \times 1$ dimensional vector such that $\|\boldsymbol{\varphi}\|_c = O(1)$. We have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \left(\mathbf{x}_{t-p} - \boldsymbol{\alpha} - \boldsymbol{\Gamma} \mathbf{f}_{t-p} \right) \boldsymbol{\varphi}' \mathbf{x}_{t-q} = \frac{\sqrt{T}}{\sqrt{N}} \frac{\sqrt{N}}{T} \sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \left(\boldsymbol{\alpha} + \boldsymbol{\Gamma} \mathbf{f}_{t-q} + \boldsymbol{\upsilon}_{t-q} \right). \tag{166}$$

Since $\left\| \sqrt{N} \mathring{\mathbf{w}}_r \right\| = O(1)$ for any $r \in \{1, .., m_w\}$ by condition (62), we can use Lemma 1, result (145), which implies

$$\frac{\sqrt{N}}{T} \sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\upsilon}_{t-q} \xrightarrow{p} E\left(\mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\upsilon}_{t-q} \right) = 0, \tag{167}$$

under Assumption B1. Sequence $\{\boldsymbol{\varphi}' \boldsymbol{\alpha}\}$ is deterministic and bounded in N, and therefore it follows from Lemma 1, result (144), that

$$\frac{\sqrt{N}}{T} \sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\alpha} \xrightarrow{p} 0, \tag{168}$$

under Assumption B1. Similarly,

$$\frac{\sqrt{N}}{T} \sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\Gamma} \mathbf{f}_{t-q} \xrightarrow{p} 0, \tag{169}$$

under Assumption B1, by Lemma 3, result (159).

Results (167), (168) and (169) establish

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \left(\mathbf{x}_{t-p} - \boldsymbol{\alpha} - \boldsymbol{\Gamma} \mathbf{f}_{t-p} \right) \boldsymbol{\varphi}' \mathbf{x}_{t-q} \xrightarrow{p} 0. \tag{170}$$

Result (161) follows from equation (170) by setting $\varphi = \mathring{\mathbf{w}}_l$ for any $l \in \{1, .., m_w\}$. Result (162) follows from equation (170) by setting $\varphi = \mathbf{s}_l$ where $\mathbf{s}_l$ is $Nk \times 1$ dimensional selection vector for the l^{th} element, $l = (i-1)k + r$, and $r \in \{1, .., k\}$.

Finally, the result (163) directly follows from results (160)-(162). This completes the proof. ■

Lemma 5 *Consider model (55). Suppose Assumptions 8-13, B1 hold and factor loadings are governed either by Assumption 14 or 15. Then for any arbitrary matrix of weights* $\mathbf{W}$ *satisfying conditions (62)-(63),*

$$\frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_{it} \mathbf{g}_{it}' \xrightarrow{p} \mathbf{C}_i, \tag{171}$$

where vector $\mathbf{g}_{it}$ *is defined in (82), and matrix* $\mathbf{C}_i$ *is defined in Assumption 16.*

Proof. It follows from Lemma 3 that

$$\frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_{it} \mathbf{g}_{it}' \xrightarrow{p} E\left(\mathbf{g}_{it} \mathbf{g}_{it}' \mid \boldsymbol{\Gamma} \right),$$

under Assumption B1. Recall that $\mathbf{g}_{it}' = \left(1, \mathbf{x}_{i,t-1}', \mathbf{x}_t^{*\prime}, \mathbf{x}_{t-1}^{*\prime}\right)$. Note that for $q \in \{0,1\}$:

$$
\begin{aligned}
E\left(1\right) &= 1, \\
E\left(\mathbf{x}_{i,t-1} \mid \mathbf{\Gamma}\right) &= \boldsymbol{\alpha}_i, \\
E\left(\mathbf{x}_{t-q}^{*} \mid \mathbf{\Gamma}\right) &= \boldsymbol{\alpha}^{*}, \\
E\left(\mathbf{x}_{i,t-1}\mathbf{x}_{i,t-1}' \mid \mathbf{\Gamma}\right) &= \boldsymbol{\alpha}_i\boldsymbol{\alpha}_i' + \mathbf{S}_i'\sum_{\ell=0}^{\infty}\mathbf{\Phi}^{\ell}\mathbf{\Sigma}\mathbf{\Phi}'^{\ell}\mathbf{S}_i + \mathbf{\Gamma}_i\mathbf{\Gamma}_f\left(0\right)\mathbf{\Gamma}_i', \\
&\rightarrow \boldsymbol{\alpha}_i\boldsymbol{\alpha}_i' + \mathbf{\Gamma}_{\boldsymbol{\xi}_i}\left(0\right) + \mathbf{\Gamma}_i\mathbf{\Gamma}_f\left(0\right)\mathbf{\Gamma}_i' \text{ under Assumption B1}, \\
E\left(\mathbf{x}_{i,t-1}\mathbf{x}_{i,t-q}^{*\prime} \mid \mathbf{\Gamma}\right) &= \boldsymbol{\alpha}_i\boldsymbol{\alpha}^{*\prime} + \mathbf{S}_i'\sum_{\ell=0}^{\infty}\mathbf{\Phi}^{\ell}\mathbf{\Sigma}\mathbf{\Phi}'^{\ell+\max\{0,1-q\}}\mathbf{W} + \mathbf{\Gamma}_i\mathbf{\Gamma}_f\left(q-1\right)\mathbf{\Gamma}^{*\prime}, \\
&\rightarrow \boldsymbol{\alpha}_i\boldsymbol{\alpha}^{*\prime} + \mathbf{\Gamma}_i\mathbf{\Gamma}_f\left(q-1\right)\mathbf{\Gamma}^{*\prime} \text{ under Assumption B1}, \\
E\left(\mathbf{x}_{it}^{*}\mathbf{x}_{i,t-q}^{*\prime} \mid \mathbf{\Gamma}\right) &= \boldsymbol{\alpha}^{*}\boldsymbol{\alpha}^{*\prime} + \mathbf{W}\sum_{\ell=0}^{\infty}\mathbf{\Phi}^{\ell+\max\{0,q\}}\mathbf{\Sigma}\mathbf{\Phi}'^{\ell}\mathbf{W} + \mathbf{\Gamma}^{*}\mathbf{\Gamma}_f\left(q\right)\mathbf{\Gamma}^{*\prime}, \\
&\rightarrow \boldsymbol{\alpha}^{*}\boldsymbol{\alpha}^{*\prime} + \mathbf{\Gamma}^{*}\mathbf{\Gamma}_f\left(q\right)\mathbf{\Gamma}^{*\prime} \text{ under Assumption B1},
\end{aligned}
$$

where $\mathbf{S}_i = \left(\mathbf{0}_{k\times k}, .., \mathbf{0}_{k\times k}, \mathbf{I}_k, \mathbf{0}_{k\times k}, .., \mathbf{0}_{k\times k}\right)'$ is $kN \times k$ dimensional selection matrix for group i. ($\mathbf{I}_k$ starts in the row $(i-1)k+1$). Therefore $E\left(\mathbf{g}_{it}\mathbf{g}_{it}' \mid \mathbf{\Gamma}\right) = \mathbf{C}_i$. This completes the proof. ■

Lemma 6 *Consider model (55). Suppose Assumptions 8-13 hold and factor loadings are governed either by Assumption 14 or 15. Then for any arbitrary matrix of weights* $\mathbf{W}$ *satisfying conditions (62)-(63), and for any fixed* $p \geq 0$,

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{x}_{t-p}^{*} - \boldsymbol{\alpha}^{*} - \mathbf{\Gamma}^{*}\mathbf{f}_{t-p}\right)u_{irt} \xrightarrow{p} \mathbf{0} \text{ under Assumption B2,} \tag{172}
$$

$$
\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}_{t-p}^{*} - \boldsymbol{\alpha}^{*} - \mathbf{\Gamma}^{*}\mathbf{f}_{t-p}\right)u_{irt} \xrightarrow{p} \mathbf{0} \text{ under Assumption B1,} \tag{173}
$$

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{x}_{i,t-1} - \boldsymbol{\alpha}_i - \boldsymbol{\xi}_{i,t-1} - \mathbf{\Gamma}_i\mathbf{f}_{t-1}\right)u_{irt} \xrightarrow{p} \mathbf{0} \text{ under Assumption B2,} \tag{174}
$$

$$
\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}_{i,t-1} - \boldsymbol{\alpha}_i - \boldsymbol{\xi}_{i,t-1} - \mathbf{\Gamma}_i\mathbf{f}_{t-1}\right)u_{irt} \xrightarrow{p} \mathbf{0} \text{ under Assumption B1,} \tag{175}
$$

where $\boldsymbol{\xi}_{it}$ *is defined in equation (70).*

Proof. Let $T_N = T(N)$ be any increasing integer-valued function of N such that Assumption B2 holds. Let $l = (i-1)k + r$ and define

$$
\boldsymbol{\kappa}_{Nlt} = \frac{1}{\sqrt{T_N}}\left\{\left(\mathbf{x}_{t-p}^{*} - \boldsymbol{\alpha}^{*} - \mathbf{\Gamma}^{*}\mathbf{f}_{t-p}\right)u_{lt} - E\left[\left(\mathbf{x}_{t-p}^{*} - \boldsymbol{\alpha}^{*} - \mathbf{\Gamma}^{*}\mathbf{f}_{t-p}\right)u_{lt}\right]\right\}, \tag{176}
$$

where we use subscript N to denote the number of groups.[55] Let $\{\mathcal{F}_t\}$ denote an increasing sequence of σ-fields ($\mathcal{F}_{t-1} \subset \mathcal{F}_t$) with κ_{Nlt} measurable with respect of $\mathcal{F}_t$ for any $N \geq i$, $i \in \mathbb{N}$. First it is established that for any fixed $l \in \mathbb{N}$, vector array $\left\{\{\boldsymbol{\kappa}_{Nlt}/c_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is uniformly integrable, where $c_{Nt} = \frac{1}{\sqrt{NT_N}}$.

[55]Note that $\mathbf{x}_t^{*}$, $\boldsymbol{\alpha}^{*}$ and $\mathbf{\Gamma}^{*}$ change with N, but as before we ommit subscript N here to keep the notation simple.

System (55) implies

$$\mathbf{x}^*_{t-p} - \boldsymbol{\alpha}^* - \boldsymbol{\Gamma}^*\mathbf{f}_{t-p} = \sum_{\ell=0}^{\infty} \mathbf{W}'\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell-p}.$$

For $p > 0$, we can write

$$\begin{aligned}
\left\| E\left(\frac{\boldsymbol{\kappa}_{Nlt}\boldsymbol{\kappa}'_{Nlt}}{c^2_{Nt}}\right)\right\| &= N \cdot \left\| E\left[\left(\sum_{\ell=0}^{\infty}\mathbf{W}'\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell-p}u_{lt}\right)\left(\sum_{\ell=0}^{\infty}\mathbf{W}'\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell-p}u_{lt}\right)'\right]\right\|, \\
&= N\left\|\sigma_{ll}\sum_{\ell=0}^{\infty}\mathbf{W}'\boldsymbol{\Phi}^{\ell}\boldsymbol{\Sigma}\boldsymbol{\Phi}'^{\ell}\mathbf{W}\right\|, \\
&\leq N\sigma_{ll}\left\|\mathbf{W}\right\|^2\left\|\boldsymbol{\Sigma}\right\|\sum_{\ell=0}^{\infty}\left\|\boldsymbol{\Phi}^{\ell}\right\|^2, \\
&= O(1),
\end{aligned}$$

where $\left\|\mathbf{W}\right\|^2 = O\left(N^{-1}\right)$ by condition (62), $\left\|\boldsymbol{\Sigma}\right\| = O(1)$ by Assumption 11 (and footnote 30), and $\sum_{\ell=0}^{\infty}\left\|\boldsymbol{\Phi}^{\ell}\right\|^2 = O(1)$ by Assumption 9 (and Remark 13). For $p = 0$, we have

$$\begin{aligned}
\left\| E\left(\frac{\boldsymbol{\kappa}_{Nlt}\boldsymbol{\kappa}'_{Nlt}}{c^2_{Nt}}\right)\right\| &= \left\| N \cdot Var\left(\mathbf{W}'\mathbf{u}_t u_{lt} + \sum_{\ell=1}^{\infty}\mathbf{W}'\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell}u_{lt}\right)\right\|, \\
&\leq N\left(\left\|\mathbf{W}\right\|^2\left\|\boldsymbol{\Psi}_{ll}\right\| + \sigma_{ll}\left\|\mathbf{W}\right\|^2\left\|\boldsymbol{\Sigma}\right\|\sum_{\ell=1}^{\infty}\left\|\boldsymbol{\Phi}^{\ell}\right\|^2 + O\left(N^{-1}\right)\right), \\
&= O(1).
\end{aligned}$$

Therefore for $p \geq 0$, $\{\boldsymbol{\kappa}_{Nlt}/c_{Nt}\}$ is uniformly bounded in L_2 norm. This proves uniform integrability of array $\{\boldsymbol{\kappa}_{Nlt}/c_{Nt}\}$.

$$E\left|E\left(\boldsymbol{\kappa}_{Nlt} \mid \mathcal{F}_{t-n}\right)\right| = \begin{cases} \mathbf{0} & \text{for any } n > 0 \text{ and any fixed } p \geq 0 \\ \boldsymbol{\tau}_{m_w} c_{Nt} O(1) & \text{for } n = 0 \text{ and any fixed } p \geq 0 \end{cases}, \tag{177}$$

and $\left\{\{\boldsymbol{\kappa}_{Nlt}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$.[56] Note that

$$\lim_{N\to\infty}\sum_{t=1}^{T_N} c_{Nt} = \lim_{N\to\infty}\sum_{t=1}^{T_N}\frac{1}{\sqrt{NT_N}} = \lim_{N\to\infty}\sqrt{\frac{T_N}{N}} = \sqrt{\varkappa} < \infty,$$

$$\lim_{N\to\infty}\sum_{t=1}^{T_N} c^2_{Nt} = \lim_{N\to\infty}\sum_{t=1}^{T_N}\frac{1}{T_N N} = \lim_{N\to\infty}\frac{1}{N} = 0.$$

Therefore for each fixed $l \in \mathbb{N}$, each of the m_w two dimensional arrays given by the vector array $\left\{\{\boldsymbol{\kappa}_{Nlt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ satisfies conditions of a mixingale weak law[57], which implies

$$\frac{1}{\sqrt{T_N}}\sum_{t=1}^{T_N}\left(\mathbf{x}^*_{t-p} - \boldsymbol{\alpha}^* - \boldsymbol{\Gamma}^*\mathbf{f}_t\right)u_{lt} \xrightarrow{L_1} \sqrt{T_N}E\left[\left(\mathbf{x}^*_{t-p} - \boldsymbol{\alpha}^* - \boldsymbol{\Gamma}^*\mathbf{f}_{t-p}\right)u_{lt}\right].$$

[56] The last equality in equation (177) takes advatage of Liapunov's inequality. $\boldsymbol{\tau}_{m_w}$ is $m_w \times 1$ dimensional vector of ones.
[57] See Davidson (1994, Theorem 19.11).

But

$$\left\| \sqrt{T_N} E\left[\left(\mathbf{x}_{t-p}^{*}-\boldsymbol{\alpha}^{*}-\boldsymbol{\Gamma}^{*}\mathbf{f}_{t-p}\right)u_{lt}\right]\right\|_{c} = \sqrt{T_N}\left\|E\left(\mathbf{W}'\mathbf{u}_t u_{lt}\right)\right\|_c = O\left(\frac{\sqrt{T_N}}{N}\right),$$

and $\sqrt{T_N}/N \to 0$ under Assumption B2. Convergence in L_1 norm implies convergence in probability. This completes the proof of result (172).

Result (173) is proved in a very similar fashion. Define new vector array $\mathbf{q}_{Nlt} = \frac{1}{\sqrt{T_N}}\kappa_{Nlt}$ where κ_{Nlt} is array defined in (176). Let $T_N = T(N)$ be any increasing integer-valued function of N such that Assumption B1 holds. Notice that for any fixed $l \in \mathbb{N}$, array $\left\{\left\{\sqrt{T_N}\mathbf{q}_{Nlt}/c_{Nt}, \mathcal{F}_t\right\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is uniformly integrable because $\left\{\left\{\boldsymbol{\kappa}_{Nlt}/c_{Nt}, \mathcal{F}_t\right\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is uniformly integrable. Furthermore, $\left\{\left\{\mathbf{q}_{Nlt}, \mathcal{F}_t\right\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is L_1-mixingale with respect to the constant array $\left\{\frac{1}{\sqrt{T_N}}c_{Nt}\right\}$ since $\left\{\left\{\boldsymbol{\kappa}_{Nlt}, \mathcal{F}_t\right\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is L_1 mixingale with respect to the constant array $\{c_{Nt}\}$. Note that

$$\lim_{N\to\infty}\sum_{t=1}^{T_N}\frac{1}{\sqrt{T_N}}c_{Nt} = \lim_{N\to\infty}\sum_{t=1}^{T_N}\frac{1}{T_N\sqrt{N}} = \lim_{N\to\infty}\frac{1}{\sqrt{N}} = 0,$$

$$\lim_{N\to\infty}\sum_{t=1}^{T_N}\left(\frac{1}{\sqrt{T_N}}c_{Nt}\right)^2 = \lim_{N\to\infty}\sum_{t=1}^{T_N}\left(\frac{1}{T_N\sqrt{N}}\right)^2 = \lim_{N\to\infty}\frac{1}{T_N N} = 0.$$

Therefore for each fixed $l \in \mathbb{N}$, we apply a mixingale weak law[58], which implies

$$\sum_{t=1}^{T_N}\mathbf{q}_{Nlt} \xrightarrow{L_1} 0 \text{ as } N \to \infty,$$

under Assumption B1. Since

$$E\left[\left(\mathbf{x}_{t-p}^{*}-\boldsymbol{\alpha}^{*}-\boldsymbol{\Gamma}^{*}\mathbf{f}_{t-p}\right)u_{lt}\right] = O\left(N^{-1}\right),$$

it follows

$$\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}_{t-p}^{*}-\boldsymbol{\alpha}^{*}-\boldsymbol{\Gamma}^{*}\mathbf{f}_{t-p}\right)u_{lt} \xrightarrow{L_1} \mathbf{0},$$

under Assumption B1. Convergence in L_1 norm implies convergence in probability. Recall that any increasing integer-valued function of N such that $\lim_{N\to\infty}T_N = \infty$ was assumed. This completes the proof of result (173).

Roots of $|\mathbf{I}_k - \boldsymbol{\Phi}_{ii}L| = 0$ lies outside the unit circle under Assumption 9 (and Remark 13). System (55) implies[59]

$$\begin{aligned}
\mathbf{x}_{it}-\boldsymbol{\alpha}_i-\boldsymbol{\Gamma}_i\mathbf{f}_t &= \boldsymbol{\Phi}_{ii}\left(\mathbf{x}_{i,t-1}-\boldsymbol{\alpha}_i-\boldsymbol{\Gamma}_i\mathbf{f}_{t-1}\right)+\boldsymbol{\Phi}_{-i}\left(\mathbf{x}_{t-1}-\boldsymbol{\alpha}-\boldsymbol{\Gamma}\mathbf{f}_{t-1}\right)+\mathbf{u}_{it},\\
\left(\mathbf{I}_k-\boldsymbol{\Phi}_{ii}L\right)\left(\mathbf{x}_{it}-\boldsymbol{\alpha}_i-\boldsymbol{\Gamma}_i\mathbf{f}_t\right)-\mathbf{u}_{it} &= \boldsymbol{\Phi}_{-i}\left(\mathbf{x}_{t-1}-\boldsymbol{\alpha}-\boldsymbol{\gamma}f_{t-1}\right),\\
\left(\mathbf{I}_k-\boldsymbol{\Phi}_{ii}L\right)\left(\mathbf{x}_{it}-\boldsymbol{\alpha}_i-\boldsymbol{\Gamma}_i\mathbf{f}_t-\sum_{\ell=0}^{\infty}\boldsymbol{\Phi}_{ii}^{\ell}\mathbf{u}_{i,t-\ell}\right) &= \boldsymbol{\Phi}_{-i}\sum_{s=0}^{\infty}\boldsymbol{\Phi}^{s}\mathbf{u}_{t-\ell},\\
\mathbf{x}_{it}-\boldsymbol{\alpha}_i-\boldsymbol{\Gamma}_i\mathbf{f}_t-\boldsymbol{\xi}_{it} &= \sum_{\ell=0}^{\infty}\boldsymbol{\Phi}_{ii}^{\ell}\boldsymbol{\Phi}_{-i}\sum_{s=0}^{\infty}\boldsymbol{\Phi}^{s}\mathbf{u}_{t-\ell},
\end{aligned}$$

where $\boldsymbol{\xi}_{it} = \sum_{\ell=0}^{\infty}\boldsymbol{\Phi}_{ii}^{\ell}\mathbf{u}_{i,t-\ell}$. Let $l = (i-1)k + r$ and let $T_N = T(N)$ be now any increasing integer-valued

[58] See Davidson (1994, Theorem 19.11).

[59] Note that for any $N \in \mathbb{N}$, polynomial $(\mathbf{I}_N - \boldsymbol{\Phi}L)$ is invertible since $\rho(\boldsymbol{\Phi}) < 1$ by Assumption 9 and Remark 13.

function of N, such that Assumption B2 holds. Define

$$\boldsymbol{\kappa}_{Nlt} = \frac{1}{\sqrt{T_N}} \left(\sum_{\ell=0}^{\infty} \boldsymbol{\Phi}_{ii}^{\ell} \boldsymbol{\Phi}_{-i} \sum_{s=0}^{\infty} \boldsymbol{\Phi}^{s} \mathbf{u}_{t-\ell-1} \right) u_{lt}. \tag{178}$$

First it is established that for any fixed $l \in \mathbb{N}$, array $\left\{ \{\boldsymbol{\kappa}_{Nlt}/c_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty} \right\}_{N=i}^{\infty}$ is uniformly integrable, where $c_{Nt} = \frac{1}{\sqrt{NT_N}}$.

$$\begin{aligned}
\left\| E\left(\frac{\boldsymbol{\kappa}_{Nlt}\boldsymbol{\kappa}_{Nlt}'}{c_{Nt}^2} \right) \right\| &= \left\| N\sigma_{ll} \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}_{ii}^{\ell} \boldsymbol{\Phi}_{-i}' \left(\sum_{s=0}^{\infty} \boldsymbol{\Phi}^{s} \boldsymbol{\Sigma} \boldsymbol{\Phi}'^{s} \right) \boldsymbol{\Phi}_{-i} \boldsymbol{\Phi}_{ii}'^{\ell} \right\|, \\
&\leq N\sigma_{ll} \left\| \boldsymbol{\Phi}_{-i} \right\|^2 \left\| \boldsymbol{\Sigma} \right\| \sum_{\ell=0}^{\infty} \left\| \boldsymbol{\Phi}_{ii}^{\ell} \right\|^2 \cdot \sum_{s=0}^{\infty} \left\| \boldsymbol{\Phi}^{s} \right\|^2, \\
&= O(1).
\end{aligned} \tag{179}$$

Hence $\{\boldsymbol{\kappa}_{Nlt}/c_{Nt}\}$ is uniformly bounded in L_2 norm. This proves uniform integrability of array $\{\boldsymbol{\kappa}_{Nlt}/c_{Nt}\}$.

$$E\left| E\left(\boldsymbol{\kappa}_{Nlt} \mid \mathcal{F}_{t-n} \right) \right| = 0 \text{ for any } n > 0,$$

and $\left\{ \{\boldsymbol{\kappa}_{Nlt}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty} \right\}_{N=i}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$.[60] Note that

$$\lim_{N\to\infty} \sum_{t=1}^{T_N} c_{Nt} = \lim_{N\to\infty} \sum_{t=1}^{T_N} \frac{1}{\sqrt{NT_N}} = \lim_{N\to\infty} \sqrt{\frac{T_N}{N}} = \sqrt{\varkappa} < \infty,$$

$$\lim_{N\to\infty} \sum_{t=1}^{T_N} c_{Nt}^2 = \lim_{N\to\infty} \sum_{t=1}^{T_N} \frac{1}{T_N N} = \lim_{N\to\infty} \frac{1}{N} = 0.$$

Therefore for any fixed $\ell \in \mathbb{N}$, each of the k_i two dimensional arrays given in the vector array $\left\{ \{\boldsymbol{\kappa}_{Nlt}, \mathcal{F}_t\}_{t=-\infty}^{\infty} \right\}_{N=i}^{\infty}$ satisfies conditions of a mixingale weak law[61], which implies

$$\sum_{t=1}^{T_N} \boldsymbol{\kappa}_{Nlt} \xrightarrow{L_1} 0.$$

Convergence in L_1 norm implies convergence in probability. This completes the proof of result (174).

Proof of result (175) is identical to the proof of result (173), but this time we define array $\mathbf{q}_{Nlt} = \frac{1}{\sqrt{T_N}} \kappa_{Nlt}$ where κ_{Nlt} is array defined in (178). ■

Lemma 7 *Consider model (55). Suppose Assumptions 8-13, 16 hold, factor loadings are governed either by Assumption 14 or 15, and $l = (i-1)k + \ell$ where $i \leq N$, $\ell \in \{1, .., k\}$. Then for any arbitrary matrix of weights $\mathbf{W}$ satisfying conditions (62)-(63),*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{a}_{it} u_{lt} \xrightarrow{D} N\left(0, \sigma_{ll} \mathbf{C}_i\right) \ \text{as } T \to \infty, \tag{180}$$

[60] Using Liapunov's inequality, it follows from equation (179) that

$$E\left| E\left(\mathbf{s}_{jk}' \boldsymbol{\kappa}_{Nlt} \mid \mathcal{F}_t \right) \right| \leq c_{Nt} O(1)$$

for any $k \times 1$ dimensional selection vector $\mathbf{s}_{jk}$.

[61] See Davidson (1994, Theorem 19.11).

where $\mathbf{a}_{it}$ is defined in equation (92) and the matrix $\mathbf{C}_i$ is defined in Assumption 16.

Proof. Lemma 7 is well known time series result. Define

$$\mathbf{z}_{lt} = \mathbf{a}_{it} u_{lt} = \begin{pmatrix} 1 \\ \boldsymbol{\alpha}_i + \boldsymbol{\xi}_{i,t-1} + \boldsymbol{\Gamma}_i \mathbf{f}_{t-1} \\ \boldsymbol{\alpha}^* + \boldsymbol{\Gamma}^* \mathbf{f}_t \\ \boldsymbol{\alpha}^* + \boldsymbol{\Gamma}^* \mathbf{f}_{t-1} \end{pmatrix} u_{lt},$$

where additional subscript i is omitted for the vector $\mathbf{z}_{lt}$ since i is determined uniquely by subscript l ($l = (i-1)k + \ell$ where $i \in \{1,..,N\}$, $\ell \in \{1,..,k\}$). Denote elements of the $k \times 1$ dimensional vector $\mathbf{z}_{lt}$ as z_{jlt}, $j \in \{1,..,k\}$. We have:

1. $$E\left(\mathbf{z}_{lt}\mathbf{z}_{lt}'\right) = \sigma_{ll}\mathbf{C}_i,$$

   where matrix $\mathbf{C}_i$ is defined in Assumption 16.

2. Roots of $|\mathbf{I}_k - \boldsymbol{\Phi}_{ii} L| = 0$ lies outside the unit circle for any $i \in \{1,..,N\}$. $\mathbf{f}_t$ is stationary process with absolute summable autocovariances, and fourth moments of $\mathbf{u}_{it}$ and $\boldsymbol{\varepsilon}_{it}$ are finite. Therefore

   $$E\left(z_{hlt} z_{jlt} z_{slt} z_{nlt}\right) < \infty \text{ for } h, j, s, n \in \{1,..,k\}.$$

3. $(1/T)\sum_{t=1}^{T} \mathbf{z}_{lt}\mathbf{z}_{lt}' \xrightarrow{p} \sigma_{ll}\mathbf{C}_i$. This is a standard time series result[62].

Applying a vector martingale difference CLT yields (180) (see for example Hamilton, Proposition 7.9). ■

Lemma 8 *Consider model (55). Suppose Assumptions 8-13, B1 hold, and factor loadings are governed either by Assumption 14 or 15. Then for any arbitrary matrix of weights $\mathbf{W}$ satisfying conditions (62)-(63), and for $p, q \in \{0, 1\}$:*

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\upsilon}_{t-p}^{*} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{181}$$

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\upsilon}_{t-p}^{*}\mathbf{f}_{t-q}' = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{182}$$

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\upsilon}_{i,t-p}\boldsymbol{\upsilon}_{t-q}^{*\prime} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{183}$$

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\upsilon}_{t-p}^{*}\boldsymbol{\upsilon}_{t-q}^{*\prime} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{184}$$

$$\frac{\boldsymbol{\Upsilon}_i'\mathbf{Q}}{T} = o_p(1). \tag{185}$$

Furthermore,

$$\frac{\mathbf{Z}'\mathbf{Q}}{T} = \mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{186}$$

$$\frac{\mathbf{X}_i'\mathbf{Z}}{T} = \frac{\mathbf{X}_i'\mathbf{Q}}{T}\mathbf{A} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{187}$$

[62] See for example Hamilton (1994, Chapter 7 and 8).

$$\frac{\mathbf{Z}'\mathbf{Z}}{T} = \mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{188}$$

$$\frac{\mathbf{Z}'\mathbf{u}_{ir\circ}}{T} = \mathbf{A}'\frac{\mathbf{Q}'\mathbf{u}_{ir\circ}}{T} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{189}$$

where

$$\underset{T\times k}{\mathbf{\Upsilon}_i} = \left(\boldsymbol{v}_{i0}, \boldsymbol{v}_{i1}, ..., \boldsymbol{v}_{i,T-1}\right)', \tag{190}$$

$\boldsymbol{v}_{it}$ *is defined in (106), matrices* $\mathbf{Z}$ *and* $\mathbf{X}_i$ *are defined in (103), matrices* $\mathbf{Q}$*,* $\mathbf{F}$ *are defined in (104), and matrix* $\mathbf{A}$ *is defined in (105).*

Proof. Consider asymptotics given in Assumption B1. Lemma 1, equation (144), implies[63]

$$\sqrt{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{\ell=0}^{\infty}\mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-p} \xrightarrow{p} 0.$$

This establishes (181). Noting $\mathbf{f}_t$ is independently distributed of $\mathbf{u}_t$, and all elements of variance matrix of $\mathbf{f}_t$ are finite completes the proof of equation (182). Furthermore, since (by Lemma 1)

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{v}_{it} \xrightarrow{p} 0,$$

equation (185) follows. Using Lemma 1, equation (145), it follows

$$\frac{1}{T}\sum_{t=1}^{T}\left[\left(\sum_{\ell=0}^{\infty}\mathbf{S}_i'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-p}\right)\left(\sum_{\ell=0}^{\infty}\sqrt{N}\mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-q}\right)'\right] \to E\left(\left(\sum_{\ell=0}^{\infty}\mathbf{S}_i'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-p}\right)\left(\sum_{\ell=0}^{\infty}\sqrt{N}\mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-q}\right)'\right).$$

Note that

$$\begin{aligned}
\left\|E\left(\left(\sum_{\ell=0}^{\infty}\mathbf{S}_i'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-p}\right)\left(\sum_{\ell=0}^{\infty}\sqrt{N}\mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-q}\right)'\right)\right\|_r &= \left\|\sqrt{N}\sum_{\ell=0}^{\infty}\mathbf{S}_i'\mathbf{\Phi}^{\ell}\mathbf{\Sigma}\mathbf{\Phi}'^{\ell}\mathbf{W}\right\|_r, \\
&\leq \sqrt{N}\left\|\mathbf{S}_i'\right\|_r\left\|\mathbf{\Sigma}\right\|_r\left\|\mathbf{W}\right\|_r\sum_{\ell=0}^{\infty}\left\|\mathbf{\Phi}\right\|_r^{\ell}\left\|\mathbf{\Phi}\right\|_c^{\ell}, \\
&= O\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}$$

and

$$E\left(\left(\sum_{\ell=0}^{\infty}\mathbf{S}_i'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-p}\right)\left(\sum_{\ell=0}^{\infty}\sqrt{N}\mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-q}\right)'\right) \xrightarrow{p} \mathbf{0},$$

under Assumption B1, where $\left\|\mathbf{W}\right\|_r = O\left(N^{-1}\right)$, $\sum_{\ell=0}^{\infty}\left\|\mathbf{\Phi}\right\|_r^{\ell}\left\|\mathbf{\Phi}\right\|_c^{\ell} = O(1)$, $\left\|\mathbf{\Sigma}\right\|_r = O(1)$ and $\left\|\mathbf{S}_i'\right\|_r = 1$.

[63]Note that

$$\begin{aligned}
\left\|Var\left(\sum_{\ell=0}^{\infty}\mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-p}\right)\right\| &= \left\|\sum_{\ell=0}^{\infty}\mathbf{W}'\mathbf{\Phi}^{\ell}\Sigma\mathbf{\Phi}'^{\ell}\mathbf{W}\right\| \\
&\leq \left\|\mathbf{W}\right\|^2\left\|\mathbf{\Sigma}\right\|\sum_{\ell=0}^{\infty}\left\|\mathbf{\Phi}\right\|^{2\ell} \\
&= O\left(\frac{1}{N}\right)
\end{aligned}$$

This completes the proof of equation (183). Similarly, equation (184) follows from Lemma 1, equation (145), noting that $\|\mathbf{W}'\|_r = O(1)$.

In order to prove equations (186)-(189), first note that row k of the matrix $\mathbf{Z} - \mathbf{QA}$ is

$$[\mathbf{Z} - \mathbf{QA}]_k = \begin{pmatrix} \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{k-\ell} \\ \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{k-\ell-1} \end{pmatrix}'.$$

Using results (181)-(184), we can write

$$\begin{aligned}
\frac{(\mathbf{Z} - \mathbf{QA})'\mathbf{Q}}{T} &= \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix} \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell} \\ \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-1} \end{pmatrix}\begin{pmatrix} 1 & \mathbf{f}_t' & \mathbf{f}_{t-1}' \end{pmatrix}\right] = o_p\left(\frac{1}{\sqrt{N}}\right), &(191)\\
\frac{\mathbf{X}_i'(\mathbf{Z} - \mathbf{QA})}{T} &= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}_{i,t-1}\begin{pmatrix} \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell} \\ \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-1} \end{pmatrix}'\right] = o_p\left(\frac{1}{\sqrt{N}}\right), &(192)\\
\frac{\mathbf{Z}'(\mathbf{Z} - \mathbf{QA})}{T} &= \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix} \mathbf{x}_t^* \\ \mathbf{x}_{t-1}^* \end{pmatrix}\begin{pmatrix} \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell} \\ \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-1} \end{pmatrix}'\right] = o_p\left(\frac{1}{\sqrt{N}}\right), &(193)\\
\frac{(\mathbf{Z} - \mathbf{QA})'(\mathbf{Z} - \mathbf{QA})}{T} &= \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix} \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell} \\ \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-1} \end{pmatrix}\begin{pmatrix} \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell} \\ \sum_{\ell=0}^{\infty} \mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell-1} \end{pmatrix}'\right] = o_p\left(\frac{1}{\sqrt{N}}\right), &(194)
\end{aligned}$$

where

$$\mathbf{x}_t^* = \boldsymbol{\alpha}^* + \mathbf{\Gamma}^*\mathbf{f}_t + \sum_{\ell=0}^{\infty}\mathbf{W}'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell}, \tag{195}$$

$$\mathbf{x}_{it} = \boldsymbol{\alpha}_i + \mathbf{\Gamma}_i\mathbf{f}_t + \sum_{\ell=0}^{\infty}\mathbf{S}_i'\mathbf{\Phi}^{\ell}\mathbf{u}_{t-\ell}. \tag{196}$$

Equations (191)-(192) establish (186)-(187). Note that

$$\begin{aligned}
\frac{\mathbf{Z}'\mathbf{Z}}{T} &= \frac{\mathbf{Z}'(\mathbf{Z} - \mathbf{QA})}{T} + \frac{\mathbf{Z}'(\mathbf{QA})}{T},\\
&= \frac{\mathbf{Z}'(\mathbf{Z} - \mathbf{QA})}{T} + \frac{(\mathbf{Z} - \mathbf{QA})'\mathbf{Q}}{T}\mathbf{A} + \mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A},\\
&= \mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A} + o_p\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}$$

where we have used (191) and (193). This completes the proof of equation (188). Equation (189) follows from Lemma 6, equation (173). This completes the proof. ■

Lemma 9 *Consider model (55). Suppose Assumptions 8-13, 17, B1 hold and factor loadings are governed either by Assumption 14 or 15. Then for any arbitrary matrix of weights* $\mathbf{W}$ *satisfying conditions (62)-(63),*

$$\frac{\mathbf{Q}'\mathbf{Q}}{T} \xrightarrow{p} \mathbf{C}_{QQi},\ \mathbf{C}_{QQi} \text{ is nonsingular,} \tag{197}$$

$$\frac{\mathbf{\Upsilon}_i'\mathbf{\Upsilon}_i}{T} \xrightarrow{p} \mathbf{\Gamma}_{\boldsymbol{\xi}_i}(0), \tag{198}$$

where

$$\mathbf{C}_{QQi} = \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{\mathbf{f}}(0) & \mathbf{\Gamma}_{\mathbf{f}}(1) \\ \mathbf{0} & \mathbf{\Gamma}_{\mathbf{f}}(1) & \mathbf{\Gamma}_{\mathbf{f}}(0) \end{pmatrix},$$

matrix $\mathbf{Q}$ *is defined in equation(104), and matrix* $\boldsymbol{\Upsilon}_i$ *is defined in (190).*

Proof. Assumption 13 implies matrix $\mathbf{C}_{QQi}$ is nonsingular. Equation (197) directly follows from the ergodicity properties of the time-series process $\mathbf{f}_t$.

Consider asymptotics B1, where $N, T \xrightarrow{j} \infty$ at any rate. Lemma 1 implies $\boldsymbol{\upsilon}_{it}$ is ergodic in variance, hence

$$\frac{\boldsymbol{\Upsilon}_i' \boldsymbol{\Upsilon}_i}{T} \xrightarrow{p} E\left(\boldsymbol{\upsilon}_{i,t-1}\boldsymbol{\upsilon}_{i,t-1}'\right).$$

Note that[64]

$$\begin{aligned} E\left(\boldsymbol{\upsilon}_{it}\boldsymbol{\upsilon}_{it}'\right) &= \sum_{\ell=0}^{\infty} \mathbf{S}_i' \boldsymbol{\Phi}^{\ell} \boldsymbol{\Sigma} \boldsymbol{\Phi}'^{\ell} \mathbf{S}_i, \\ &= \sum_{\ell=0}^{\infty} \left(\boldsymbol{\Phi}_{ii}^{\ell} \mathbf{S}_i' + \sum_{n=1}^{\ell} \boldsymbol{\Phi}_{ii}^{n} \boldsymbol{\Phi}_{-i}' \boldsymbol{\Phi}^{\ell-n} \right) \boldsymbol{\Sigma} \left(\boldsymbol{\Phi}_{ii}^{\ell} \mathbf{S}_i' + \sum_{n=1}^{\ell} \boldsymbol{\Phi}_{ii}^{n} \boldsymbol{\Phi}_{-i}' \boldsymbol{\Phi}^{\ell-n} \right)', \\ &= \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}_{ii}^{\ell} \mathbf{S}_i' \boldsymbol{\Sigma} \mathbf{S}_i \boldsymbol{\Phi}_{ii}'^{\ell} + \sum_{\ell=0}^{\infty} \left(\boldsymbol{\Phi}_{ii}^{\ell} \mathbf{S}_i' \boldsymbol{\Sigma} \sum_{n=1}^{\ell} \boldsymbol{\Phi}'^{\ell-n} \boldsymbol{\Phi}_{-i} \boldsymbol{\Phi}_{ii}'^{n} \right) + \sum_{\ell=0}^{\infty} \left(\sum_{n=1}^{\ell} \boldsymbol{\Phi}_{ii}^{n} \boldsymbol{\Phi}_{-i}' \boldsymbol{\Phi}^{\ell-n} \boldsymbol{\Sigma} \boldsymbol{\Phi}_{ii}^{\ell} \mathbf{S}_i' \right) + \\ &\quad + \sum_{\ell=0}^{\infty} \left[\left(\sum_{n=1}^{\ell} \boldsymbol{\Phi}_{ii}^{n} \boldsymbol{\Phi}_{-i}' \boldsymbol{\Phi}^{\ell-n} \right) \boldsymbol{\Sigma} \left(\sum_{n=1}^{\ell} \boldsymbol{\Phi}'^{\ell-n} \boldsymbol{\Phi}_{-i} \boldsymbol{\Phi}_{ii}'^{n} \right) \right], \\ &= \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}_{ii}^{\ell} \mathbf{S}_i' \boldsymbol{\Sigma} \mathbf{S}_i \boldsymbol{\Phi}_{ii}'^{\ell} + O\left(\frac{1}{\sqrt{N}}\right). \end{aligned} \tag{199}$$

Therefore

$$E\left(\boldsymbol{\upsilon}_{it}\boldsymbol{\upsilon}_{it}'\right) \to \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0), \tag{200}$$

as $N \to \infty$. This completes the proof. ■

Lemma 10 *Consider model (55). Suppose Assumptions 8-13, 17, B1 hold and factor loadings are governed either by Assumption 14 or 15. Then for any arbitrary matrix of weights* $\mathbf{W}$ *satisfying conditions (62)-(63) and Assumption 17,*

$$\frac{\mathbf{X}_i' \mathbf{M}_Z \mathbf{X}_i}{T} = \frac{\mathbf{X}_i' \mathbf{M}_Q \mathbf{X}_i}{T} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{201}$$

$$\frac{\mathbf{X}_i' \mathbf{M}_Q \mathbf{X}_i}{T} \xrightarrow{p} \boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0), \tag{202}$$

$$\frac{\mathbf{X}_i' \mathbf{M}_Z \mathbf{Q}}{\sqrt{T}} = o_p\left(\sqrt{\frac{T}{N}}\right), \tag{203}$$

$$\frac{\mathbf{X}_i' \mathbf{M}_Z \mathbf{u}_{ir\circ}}{\sqrt{T}} = \frac{\boldsymbol{\Upsilon}_i' \mathbf{M}_Q \mathbf{u}_{ir\circ}}{\sqrt{T}} + o_p\left(\sqrt{\frac{T}{N}}\right), \tag{204}$$

where $\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0)$ *is defined in Assumption 16,* $\mathbf{M}_Z = \mathbf{I}_T - \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{+}\mathbf{Z}'$*, matrices* $\mathbf{Z}$*,* $\mathbf{X}_i$ *and* $\mathbf{Q}$ *are defined in equations (103)-(104) and matrix* $\boldsymbol{\Upsilon}_i$ *is defined in (190).*

[64] In equation (199), we mean that

$$\left\| E\left(\boldsymbol{\upsilon}_{it}\boldsymbol{\upsilon}_{it}'\right) - \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}_{ii}^{\ell} \mathbf{S}_i' \boldsymbol{\Sigma} \mathbf{S}_i \boldsymbol{\Phi}_{ii}'^{\ell} \right\| = O\left(\frac{1}{\sqrt{N}}\right)$$

Proof.

$$\frac{\mathbf{X}_i'\mathbf{M}_Z\mathbf{X}_i}{T}=\frac{\mathbf{X}_i'\mathbf{X}_i}{T}-\frac{\mathbf{X}_i'\mathbf{Z}}{T}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{+}\frac{\mathbf{Z}'\mathbf{X}_i}{T}. \tag{205}$$

Results (187)-(188) of Lemma 8 imply

$$\frac{\mathbf{X}_i'\mathbf{Z}}{T}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{+}\frac{\mathbf{Z}'\mathbf{X}_i}{T}=\frac{\mathbf{X}_i'\mathbf{Q}}{T}\mathbf{A}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^{+}\mathbf{A}'\frac{\mathbf{Q}'\mathbf{X}_i}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{206}$$

Using definition of the Moore-Penrose inverse, it follows

$$\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^{+}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)=\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right). \tag{207}$$

Multiply equation (207) by $\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\left(\mathbf{A}\mathbf{A}'\right)^{-1}\mathbf{A}$ from the left and by $\mathbf{A}'\left(\mathbf{A}\mathbf{A}'\right)^{-1}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}$ from the right to obtain[65]

$$\mathbf{A}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^{+}\mathbf{A}'=\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}. \tag{208}$$

Equations (208) and (206) imply

$$\frac{\mathbf{X}_i'\mathbf{Z}}{T}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{+}\frac{\mathbf{Z}'\mathbf{X}_i}{T}=\frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{X}_i}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{209}$$

Result (201) follows from (209) and (205).

System (55) implies

$$\mathbf{X}_i=\boldsymbol{\tau}\boldsymbol{\alpha}_i'+\mathbf{F}\left(-1\right)\boldsymbol{\Gamma}_i'+\boldsymbol{\Upsilon}_i. \tag{210}$$

where $\boldsymbol{\Upsilon}_i$ is defined in (190). Since $\mathbf{Q}=[\boldsymbol{\tau},\mathbf{F},\mathbf{F}\left(-1\right)]$, it follows

$$\frac{\mathbf{X}_i'\mathbf{M}_Q\mathbf{X}_i}{T}=\frac{\boldsymbol{\Upsilon}_i'\mathbf{M}_Q\boldsymbol{\Upsilon}_i}{T}=\frac{\boldsymbol{\Upsilon}_i'\boldsymbol{\Upsilon}_i}{T}+\frac{\boldsymbol{\Upsilon}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\boldsymbol{\Upsilon}_i}{T}. \tag{211}$$

Noting (198), (185), and (197), result (202) follows directly from (211).

Results (186)-(188) of Lemma 8 imply

$$\frac{\mathbf{X}_i'\mathbf{Z}}{T}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{+}\frac{\mathbf{Z}'\mathbf{Q}}{T}=\frac{\mathbf{X}_i'\mathbf{Q}}{T}\mathbf{A}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^{+}\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{212}$$

Substituting equation (208), it follows

$$\frac{\mathbf{X}_i'\mathbf{Z}}{T}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{+}\frac{\mathbf{Z}'\mathbf{Q}}{T}=\frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{Q}}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{213}$$

Equation (213) implies

$$\frac{\mathbf{X}_i'\mathbf{M}_Z\mathbf{Q}}{\sqrt{T}}=\frac{\mathbf{X}_i'\mathbf{M}_Q\mathbf{Q}}{\sqrt{T}}+o_p\left(\sqrt{\frac{T}{N}}\right)=o_p\left(\sqrt{\frac{T}{N}}\right).$$

This completes the proof of result (203).

[65] Note that $plim_{T\to\infty}\frac{1}{T}\mathbf{Q}'\mathbf{Q}$ is nonsingular by Lemma 9, equation (197). $\mathbf{A}\mathbf{A}'$ is nonsingular, since matrix $\mathbf{A}$ has full row-rank by Assumption 17.

Results (187)-(189) of Lemma 8 imply

$$\frac{\mathbf{X}_i'\mathbf{Z}}{T}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{+}\frac{\mathbf{Z}'\mathbf{u}_{ir\circ}}{T}=\frac{\mathbf{X}_i'\mathbf{Q}}{T}\mathbf{A}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^{+}\mathbf{A}'\frac{\mathbf{Q}'\mathbf{u}_{ir\circ}}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right).$$

Substituting equation (208), it follows

$$\frac{\mathbf{X}_i'\mathbf{Z}}{T}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{+}\frac{\mathbf{Z}'\mathbf{Q}}{T}=\frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{u}_{ir\circ}}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{214}$$

Equations (214) and (210) imply, noting that $\mathbf{M}_Q\left(\boldsymbol{\tau}\boldsymbol{\alpha}_i'+\mathbf{F}\boldsymbol{\Gamma}_i'\right)=0$ since $\mathbf{Q}=[\boldsymbol{\tau},\mathbf{F},\mathbf{F}(-1)]$,

$$\begin{aligned}\frac{\mathbf{X}_i'\mathbf{M}_Z\mathbf{u}_{ir\circ}}{\sqrt{T}} &= \frac{\mathbf{X}_i'\mathbf{M}_Q\mathbf{u}_{ir\circ}}{\sqrt{T}}+o_p\left(\sqrt{\frac{T}{N}}\right),\\ &= \frac{\boldsymbol{\Upsilon}_i'\mathbf{M}_Q\mathbf{u}_{ir\circ}}{\sqrt{T}}+o_p\left(\sqrt{\frac{T}{N}}\right).\end{aligned}$$

This completes the proof. ■

Lemma 11 *Consider model (55). Suppose Assumptions 8-13, 17, B1 hold and factor loadings are given either by Assumption 14 or 15. Then for any arbitrary matrix of weights $\mathbf{W}$ satisfying conditions (62)-(63),*

$$\frac{\mathbf{X}_i'\mathbf{M}_Z\mathbf{e}_{ir\circ}}{T} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{215}$$

$$\frac{\mathbf{X}_i'\mathbf{M}_Z\mathbf{u}_{ir\circ}}{\sqrt{T}} = \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{ir\circ}}{\sqrt{T}}+o_p\left(\sqrt{\frac{T}{N}}\right)+o_p(1), \tag{216}$$

where $\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0)$ is defined in Assumption 16, matrices $\mathbf{M}_Z$ and $\mathbf{X}_i$ are defined in equations (103), and matrix $\boldsymbol{\Upsilon}_i$ is defined in (190) and vector $e_{ir\circ}$ is defined in equation (109).

Proof.

$$\begin{aligned}\frac{\mathbf{X}_i'\mathbf{e}_{ir\circ}}{T} &= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}_{i,t-1}\left(\mathbf{s}_{rk}'\sum_{\ell=0}^{\infty}\boldsymbol{\Phi}_{-i}\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell-1}\right)'\right],\\ \frac{\mathbf{Z}'\mathbf{e}_{ir\circ}}{T} &= \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix}\mathbf{x}_t^*\\ \mathbf{x}_{t-1}^*\end{pmatrix}\left(\mathbf{s}_{rk}'\sum_{\ell=0}^{\infty}\boldsymbol{\Phi}_{-i}\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell-1}\right)'\right].\end{aligned}$$

Note that $\boldsymbol{\Phi}_{-i}$ satisfy conditions $\left\|\boldsymbol{\Phi}_{-i}\right\|_r=O\left(N^{-1}\right)$, therefore result (215) directly follows from equations (192) and (193).

Note that

$$\begin{aligned}\frac{\boldsymbol{\Upsilon}_i'\mathbf{M}_Q\mathbf{u}_{ir\circ}}{\sqrt{T}} &= \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{ir\circ}}{\sqrt{T}}+\frac{\boldsymbol{\Upsilon}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{u}_{ir\circ}}{\sqrt{T}},\\ &= \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{ir\circ}}{\sqrt{T}}+o_p(1),\end{aligned} \tag{217}$$

since $\frac{\mathbf{Q}'\mathbf{u}_{ir\circ}}{\sqrt{T}}=O_p(1)$, $plim_{T\to\infty}\frac{1}{T}\mathbf{Q}'\mathbf{Q}$ is nonsingular by Lemma 9, equation (197), and $\frac{\boldsymbol{\Upsilon}_i'\mathbf{Q}}{T}=o_p(1)$ by Lemma 8, equation (185). Substituting (217) into equation (204) implies result (216). This completes the proof. ■

Lemma 12 *Consider model (55). Suppose Assumptions 8-13, hold and* $l = (i-1)k + \ell$ *where* $i \in \{1,..,N\}$, $\ell \in \{1,..,k\}$. *Then,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{\xi}_{i,t-1} u_{lt} \xrightarrow{D} N\left(0, \sigma_{ll}\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0)\right), \text{ as } T \to \infty. \tag{218}$$

Proof. Lemma 12 is well-known time series result. Proof of Lemma 12 is similar to the proof of Lemma 7. Denote

$$\mathbf{z}_{lt} = \boldsymbol{\xi}_{i,t-1} u_{lt},$$

where additional subscript i is omitted for the vector $\mathbf{z}_{lt}$ since i is determined uniquely by subscript l ($l = (i-1)k + \ell$ where $i \in \{1,..,N\}$, $\ell \in \{1,..,k\}$). We have:

1. $$E\left(\mathbf{z}_{lt}\mathbf{z}_{lt}'\right) = \sigma_{ll}\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0),$$

   where matrix $\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0)$ is defined in Assumption 16.

2. Roots of $|\mathbf{I}_k - \boldsymbol{\Phi}_{ii}L| = 0$ lies outside the unit circle for any $i \in \{0,1,...,N\}$, and fourth moments of $\mathbf{u}_{it}$ are finite. Therefore

   $$E\left(z_{hlt}z_{jlt}z_{slt}z_{nlt}\right) < \infty \text{ for } h,j,s,n \in \{1,..,k\}.$$

3. $(1/T)\sum_{t=1}^{T} \mathbf{z}_{it}\mathbf{z}_{it}' \xrightarrow{p} \sigma_{ll}\boldsymbol{\Gamma}_{\boldsymbol{\xi}_i}(0)$. This is a standard time series result.[66]

Applying a vector martingale difference CLT yields (218) (see for example Hamilton 1994, Proposition 7.9). ■

B Extension of the Analysis to IVAR(p) Models

B.1 Cross Sectional Dependence in IVAR(p) Models

Nature of cross section correlation pattern in IVAR(1) models at any given point in time, $t \in \mathcal{T}$, is investigated in Section 2. This subsection extends the analysis to IVAR(p) models, where p is finite. It is assumed that $\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{Nt})'$ is given by the following IVAR(p) model:

$$\mathbf{x}_t = \sum_{j=1}^{p} \boldsymbol{\Phi}_j \mathbf{x}_{t-j} + \mathbf{u}_t, \tag{219}$$

where $\boldsymbol{\Phi}_j$ for $j = 1,...,p$ are $N \times N$ dimensional matrices of coefficients, and as before, errors $\mathbf{u}_t = (u_{1t},...,u_{Nt})'$ satisfy Assumption 1. The process starts from a finite past, $t > -M$, M being fixed. This assumption is relaxed in Subsection B.1.1 below for stationary models.

We can rewrite the IVAR(p) representation (219) in the companion form

$$\begin{pmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+2} \\ \mathbf{x}_{t-p+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \dots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \mathbf{I}_m & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m & & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_m & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \vdots \\ \mathbf{x}_{t-p+1} \\ \mathbf{x}_{t-p} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \tag{220}$$

[66] See for example Hamilton (1994, Chapter 7 and 8).

which is an IVAR(1) model, but in the $Np \times 1$ vector of random variables $\mathbf{X}_t = (\mathbf{x}'_t, \mathbf{x}'_{t-1}, ..., \mathbf{x}'_{t-p+1})'$, namely

$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{U}_t, \tag{221}$$

where $\mathbf{\Phi}$ is now the $Np \times Np$ companion coefficient matrix,

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & ... & \mathbf{\Phi}_{p-1} & \mathbf{\Phi}_p \\ \mathbf{I}_m & \mathbf{0} & ... & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m & & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & ... & \mathbf{I}_m & \mathbf{0} \end{pmatrix}, \tag{222}$$

and $\mathbf{U}_t = (\mathbf{u}'_t, \mathbf{0}, ..., \mathbf{0})'$ is the $Np \times 1$ vector of error terms. $\mathbf{x}_t$ is then obtained as

$$\mathbf{x}_t = \mathbf{S}'_{1p}\mathbf{X}_t, \tag{223}$$

where $Np \times N$ selection matrix $\mathbf{S}_{1p} = \begin{pmatrix} \mathbf{I}_N, & \mathbf{0}_{N\times N}, & \mathbf{0}_{N\times N}, & ..., & \mathbf{0}_{N\times N} \end{pmatrix}'$. Thus extension of the results in Section 2 is in most cases straightforward.

Proposition 5 *Consider the IVAR(p) model (219) and suppose Assumptions 1-2 (on errors* $\mathbf{u}_t$ *and the* companion *coefficient matrix* $\mathbf{\Phi}$*) hold. Then for any arbitrary nonrandom vector of weights* $\mathbf{w}$ *satisfying condition (2), and for any* $t \in \mathcal{T}$*,*

$$\lim_{N\to\infty} Var\left[\mathbf{w}'\mathbf{x}_t \mid \mathbf{X}(-M)\right] = 0 \tag{224}$$

Proof. The vector difference equation (221) can be solved backwards from $t = -M$, taking $\mathbf{X}(-M)$ as given:

$$\mathbf{X}_t = \mathbf{\Phi}^{t+M}\mathbf{X}(-M) + \sum_{\ell=0}^{t+M-1} \mathbf{\Phi}^{\ell}\mathbf{U}_{t-\ell}.$$

The variance of $\mathbf{X}_t$ (conditional on starting values $\mathbf{X}(-M)$) is

$$\mathbf{\Omega}_t = Var\left[\mathbf{X}_t \mid \mathbf{X}(-M)\right] = \sum_{\ell=0}^{t+M-1} \mathbf{\Phi}^{\ell}\mathbf{S}_{1p}\mathbf{\Sigma}_{t-\ell}\mathbf{S}'_{1p}\mathbf{\Phi}'^{\ell},$$

and, using the Rayleigh-Ritz theorem,[67] $\|\mathbf{\Omega}_t\|_c$ is under Assumptions 1-2 bounded by

$$\|\mathbf{\Omega}_t\|_c \leq \sum_{\ell=0}^{t+M-1} \varrho\left(\mathbf{\Sigma}_{t-\ell}\right)\left(\|\mathbf{\Phi}\|_c \|\mathbf{\Phi}\|_r\right)^{\ell} = o(N). \tag{225}$$

It follows that for any arbitrary nonrandom vector of weights satisfying granularity condition (2),

$$\|Var\left(\mathbf{w}'\mathbf{x}_t\right)\|_c = \left\|\mathbf{w}'\mathbf{S}'_{1p}\mathbf{\Omega}_t\mathbf{S}_{1p}\mathbf{w}\right\|_c \leq \left\|\varrho\left(\mathbf{S}'_{1p}\mathbf{\Omega}_t\mathbf{S}_{1p}\right)\left(\mathbf{w}'\mathbf{w}\right)\right\|_c = o(1), \tag{226}$$

where $\varrho\left(\mathbf{S}'_{1p}\mathbf{\Omega}_t\mathbf{S}_{1p}\right) \leq \|\mathbf{S}_{1p}\|_r \|\mathbf{S}_{1p}\|_c \|\mathbf{\Omega}_t\|_c$,[68] $\|\mathbf{S}_{1p}\|_c = \|\mathbf{S}_{1p}\|_r = 1$, $\|\mathbf{\Omega}_t\|_c = o(N)$, and $\mathbf{w}'\mathbf{w} = \|\mathbf{w}\|^2 = O\left(N^{-1}\right)$ by condition (2). ■

[67] See Horn and Johnson (1985, p. 176)

[68] Spectral radius is lower bound for any matrix norm, see Horn and Johnson (1985, Theorem 5.6.9).

IVAR(p) model (219) is thus CWD under Assumptions 1-2 at any point in time $t \in \mathcal{T}$. Consider the following additional assumption on coefficients matrices $\{\mathbf{\Phi}_j\}_{j=1}^{p}$.

ASSUMPTION 18 *Let $\mathcal{K} \subseteq \mathbb{N}$ be a non-empty index set. Define vector $\boldsymbol{\phi}_{j,-i} = \left(\phi_{i1}^{(j)}, ..., \phi_{i,i-1}^{(j)}, 0, \phi_{i,i+1}^{(j)}, ..., \phi_{iN}^{(j)}\right)'$ where $\phi_{ik}^{(j)}$ for $i, k \in \mathcal{S}$ are elements of matrix $\mathbf{\Phi}_j$. For any $i \in \mathcal{K}$, vector $\boldsymbol{\phi}_{j,-i}$ satisfies*

$$\left\|\boldsymbol{\phi}_{j,-i}\right\| = O\left(N^{-\frac{1}{2}}\right) \text{ for } j \in \{1, .., p\}\,. \tag{227}$$

Corollary 2 *Consider model (219) and suppose Assumptions 1-2 (on errors $\mathbf{u}_t$ and the* companion *coefficient matrix $\mathbf{\Phi}$), and Assumption 18 hold. Then*

$$\lim_{N\to\infty} Var\left(x_{it} - \sum_{j=1}^{p}\phi_{ii}^{(j)}x_{i,t-1} - u_{it} \mid \mathbf{X}(-M)\right) = 0, \text{ for } i \in \mathcal{K}. \tag{228}$$

Proof. Assumption 18 implies that for $i \in \mathcal{K}$, vectors $\left\{\boldsymbol{\phi}_{j,-i}\right\}_{j=1}^{p}$ satisfy condition (2). It follows from Proposition 5 that for any $i \in \mathcal{K}$, and any $j \in \{1, .., p\}$:

$$\lim_{N\to\infty} Var\left(\boldsymbol{\phi}_{j,-i}'\mathbf{x}_t \mid \mathbf{X}(-M)\right) = 0. \tag{229}$$

System (219), and equation (229) establish

$$\lim_{N\to\infty} Var\left(x_{it} - \sum_{j=1}^{p}\phi_{ii}^{(j)}x_{i,t-1} - u_{it} \mid \mathbf{X}(-M)\right) = \lim_{N\to\infty} Var\left(\sum_{j=1}^{p}\boldsymbol{\phi}_{j,-i}'\mathbf{x}_t \mid \mathbf{X}(-M)\right) = 0, \text{ for } i \in \mathcal{K}.$$

■

B.1.1 Stationary Conditions for IVAR(p)

Conditions under which IVAR(p) model (219) is stationary for fixed N are well known in the literature, see for instance Hamilton (1994, Chapter 10).

ASSUMPTION 19 *(Necessary conditions for covariance stationarity) Let $\mathbf{\Sigma}_t = \mathbf{\Sigma}$ be time invariant and all the eigenvalues of the companion coefficients matrix $\mathbf{\Phi}$, defined by λ that satisfy the equation $|\mathbf{\Phi} - \lambda\mathbf{I}_{Np}| = 0$, lie inside of the unit circle.*

For fixed N and $\|Var[\mathbf{X}(-M)]\|_r < K$, we have under Assumptions 1, 19 and as $M \to \infty$ (following similar arguments as in subsection 2.1),

$$E(\mathbf{x}_t) = \mathbf{0}, \text{ and } Var(\mathbf{x}_t) = \sum_{\ell=0}^{\infty}\mathbf{D}_\ell\mathbf{\Sigma}\mathbf{D}_\ell' < \infty, \tag{230}$$

where $\mathbf{D}_\ell \equiv \mathbf{S}_{1p}'\mathbf{\Phi}^\ell\mathbf{S}_{1p}$. In the stationary case with $M \to \infty$, at the point in time t

$$\|Var(\mathbf{x}_t)\|_c = \left\|\sum_{\ell=0}^{\infty}\mathbf{D}_\ell\mathbf{\Sigma}\mathbf{D}_\ell'\right\|_c \leq \varrho(\mathbf{\Sigma})\sum_{\ell=0}^{\infty}\|\mathbf{D}_\ell\mathbf{D}_\ell'\|_c\,. \tag{231}$$

Observe that $Var(x_{it})$ for given i, as well as $\|Var(\mathbf{x}_t)\|_c$ need not necessarily be bounded as $N \to \infty$, under the Assumption 19, even if $\varrho(\mathbf{\Sigma}) = O(1)$.

Proposition 6 *Consider model (219) and suppose Assumptions 1-2 (on errors* $\mathbf{u}_t$ *and the* companion *coefficient matrix* $\mathbf{\Phi}$*), and Assumption 19 hold,* $\|Var[\mathbf{X}(-M)]\|_r < K$ *and* $M \to \infty$*. Furthermore, assume there exists arbitrarily small* $\epsilon > 0$ *that do not depend on* N *and such that* $\varrho(\mathbf{E}) < 1 - \epsilon$*, where*

$$\underset{p \times p}{\mathbf{E}} = \begin{pmatrix} \|\mathbf{\Phi}_1\| & \|\mathbf{\Phi}_2\| & ... & \|\mathbf{\Phi}_{p-1}\| & \|\mathbf{\Phi}_p\| \\ 1 & 0 & ... & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 1 & 0 \end{pmatrix}. \tag{232}$$

Then at any point in time $t \in \mathcal{T}$ *and for any nonrandom weights* $\mathbf{w}$ *satisfying condition (2),*

$$\lim_{N \to \infty} Var(\mathbf{w}'\mathbf{x}_t) = 0. \tag{233}$$

Proof. $\mathbf{D}_\ell = \mathbf{S}'_{1p}\mathbf{\Phi}^\ell \mathbf{S}_{1p}$ solves the following matrix difference equation

$$\mathbf{D}_\ell = \sum_{j=1}^{p} \mathbf{\Phi}_j \mathbf{D}_{\ell-j}, \tag{234}$$

with starting values $\mathbf{D}_0 = \mathbf{I}_N$, and $\mathbf{D}_k = \mathbf{0}_{N \times N}$ for $k = 1-p, ..., -2, -1$. Taking the spectral matrix norm of the both sides of equation (234), we have

$$\|\mathbf{D}_\ell\| \leq \sum_{j=1}^{p} \|\mathbf{\Phi}_j\| \|\mathbf{D}_{\ell-j}\|.$$

It follows that for any arbitrary small $\delta > 0$ that do not depend on N, there exists $\ell_0(\delta) \in \mathbb{N}$ that do not depend on N (recall p is finite) such that

$$\|\mathbf{D}_\ell\| < (\varrho(\mathbf{E}) + \delta)^\ell \text{ for any } \ell > \ell_0(\delta).$$

Choose $\delta = \frac{\epsilon}{2}$. Noting that the spectral norm is self-adjoint (see Horn and Johnson, 1985, p. 309),

$$\|\mathbf{D}_\ell \mathbf{D}'_\ell\| \leq \|\mathbf{D}_\ell\|^2 < \left(\varrho(\mathbf{E}) + \frac{\epsilon}{2}\right)^{2\ell} < \left(1 - \frac{\epsilon}{2}\right)^{2\ell} \text{ for any any } \ell > \ell_0\left(\frac{\epsilon}{2}\right).$$

This implies that $\sum_{\ell=0}^{\infty} \|\mathbf{D}_\ell \mathbf{D}'_\ell\| = O(1)$.[69] Since also $\varrho(\mathbf{\Sigma}) = o(N)$ under the Assumptions 2 and 19, it follows from (231) that $\|Var(\mathbf{x}_t)\| = o(N)$. This establishes $\|Var(\mathbf{w}'\mathbf{x}_t)\| \leq \left\|\mathbf{w}'\mathbf{S}'_{1p}\mathbf{\Omega}_t \mathbf{S}_{1p}\mathbf{w}\right\| = o(1)$, along similar arguments used in establishing equation (226). ■

[69] Recall $\|\mathbf{\Phi}\|^2 \leq \|\mathbf{\Phi}\|_c \|\mathbf{\Phi}\|_r = O(1)$ under Assumption 2.

$$\sum_{\ell=0}^{\infty} \|\mathbf{D}_\ell \mathbf{D}'_\ell\| \leq \sum_{\ell=0}^{\infty} \|\mathbf{D}_\ell\|^2 = \sum_{\ell=0}^{\ell_0} \left\|\mathbf{S}'_{1p}\mathbf{\Phi}^\ell \mathbf{S}_{1p}\right\|^2 + \sum_{\ell=\ell_0+1}^{\infty} \|\mathbf{D}_\ell\|^2.$$

But since $\ell_0 = \ell_0\left(\frac{\epsilon}{2}\right)$ does not depent on N, $\sum_{\ell=0}^{\ell_0} \left\|\mathbf{S}'_{1p}\mathbf{\Phi}^\ell \mathbf{S}_{1p}\right\|^2 \leq \sum_{\ell=0}^{\ell_0} \|\mathbf{\Phi}\|^{2\ell} = O(1)$ and $\sum_{\ell=\ell_0+1}^{\infty} \|\mathbf{D}_\ell\|^2 < \sum_{\ell=\ell_0}^{\infty} \left(1 - \frac{\epsilon}{2}\right)^\ell = O(1)$.

B.2 Estimation of a Stationary IVAR(p) Model

Extension of the estimation of IVAR(1) to IVAR(p) model is straightforward. Any IVAR(p) model can be written as an IVAR(1), see Appendix B.1, equations (219)-(223). If each of the coefficients matrices $\{\mathbf{\Phi}_j\}_{j=1}^{p}$ of IVAR(p) model satisfies Assumption 8 then so does the corresponding companion $Nkp \times Nkp$ dimensional matrix $\mathbf{\Phi}$ (see equation (222)). Note that $\mathbf{\Phi}$ does not satisfy Assumption 9. Therefore an alternative sufficient condition for absolute summability, such as

$$\sum_{\ell=0}^{\infty} \left\| \mathbf{\Phi}^{\ell} \mathbf{\Phi}'^{\ell} \right\| = O\left(1\right),$$

will be needed. One possibility is the condition presented in Appendix B.1.1 (see Proposition 6), that is

$$\varrho\left(\mathbf{E}\right) < 1 - \epsilon,$$

where arbitrarily small $\epsilon > 0$ does not depend on N and the $p \times p$ dimensional matrix $\mathbf{E}$ is defined in equation (232). Thus the analysis of IVAR(p) models is almost identical to that of IVAR(1) model above.

References

[1] Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht, the Netherlands.

[2] Bernanke, B.S., Bovian, J. and Eliasz, P. (2005), "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach", *Quarterly Journal of Economics*, 120, 387-422.

[3] Binder, M. and Pesaran, M.H. (1997), "Multivariate linear rational expectations models: characterization of the nature of the solutions and their fully recursive computation", *Econometric Theory*, 13, 877-888.

[4] Bond, S., Leblebicioglu, A. and Schiantarelli, F. (2004), "Capital Accumulation and Growth: A New Look at the Empirical Evidence", IZA Discussion Paper No. 1174, June 2004.

[5] Conley, T.G. (1999), "GMM Estimation with Cross Sectional Dependence", *Journal of Econometrics*, 92, 1-45.

[6] Conley, T.G., and Topa, G. (2002), "Socio-economic Distance and Spatial Patterns in Unemployment", *Journal of Applied Econometrics*, 17, 303-327.

[7] Canova, F. (1995), "Vector Autoregressive models: Specification, estimation, inference, and forecasting", Chapter 2 in Pesaran, M.H. and M. Wickens (eds.), *Handbook of Applied Econometrics: Macroeconomics*, Basil Blackwell, Oxford.

[8] Chudik, A. (2007a), "GVAR Approach and the Dominance of the US Economy", mimeo, University of Cambridge.

[9] Chudik, A. (2007b), "Macroeconomic Interdependence and The Curse of Dimensionality", mimeo, University of Cambridge.

[10] Cliff, A. and Ord, J.K. (1973), *Spatial Autocorrelation*, Pion, London.

[11] Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press.

[12] Dees, S., di Mauro, F., Pesaran, M.H. and Smith, L.V. (2007), "Exploring the International Linkages of the Euro Area: A Global VAR analysis", *Journal of Applied Econometrics*, 22, 1-38.

[13] Del Negro, M. and Schorfheide, F. (2004), "Priors from General Equilibrium Models for VARs", *International Economic Review*, 45, 643-673.

[14] De Mol, C., Giannone, D. and Reichlin, L. (2006), "Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components?", CEPR Discussion Paper No. 5829, September 2006.

[15] Doan, T., Litterman, R. and Sims, C. (1984), "Forecasting and Conditional Projections Using Realistic Prior Distributions", *Econometric Reviews*, 3, 1-100.

[16] Forni, M. and Lippi, M. (2001), "The Generalized Dynamic Factor Model: Representation theory", *Econometric Theory*, 17, 1113-41.

[17] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), "The Generalized Dynamic Factor Model: Identification and Estimation", *Review of Economics and Statistics*, 82, 540-554.

[18] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004), "The Generalized Dynamic Factor Model Consistency and Rates", *Journal of Econometrics*, 119, 231-255.

[19] Garratt, A., Lee, K., Pesaran, M.H. and Shin, Y. (2006), *Global and National Macroeconometric Modelling: A Long Run Structural Approach*, Oxford University Press.

[20] Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series", in: D.J. Aigner and A.S. Goldberger (eds.), *Latent Variables in Socio-Economic Models*, North-Holland, Amsterdam.

[21] Giacomini, R., and White, H. (2006), "Tests of Conditional Predictive Ability", *Econometrica*, 74, 1545–1578.

[22] Giannoni, D., Reichlin, L. and Sala, L. (2005), "Monetary Policy in Real Time", CEPR Discussion Paper No. 4981, March 2005.

[23] Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press.

[24] Horn, R.A., and Johnson, C.A. (1985), *Matrix Analysis*, Cambridge University Press.

[25] Kelejian, H.H., and Robinson, D.P. (1995), "Spatial correlation: A suggested Alternative to the Autoregressive Model", in Anselin, L., and Florax, R.J. (eds.), *New Directions in Spatial Econometrics*, 75-95, Springer-Verlag, Berlin.

[26] Leeper, E.M., Sims, C.A. and Zha, T. (1996), "What does monetary policy do?", *Brookings Papers on Economic Activity*, 2, 1-63.

[27] Litterman, R. (1986), "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience", *Journal of Business and Economics Statistics*, 4, 25-38.

[28] Pesaran, M.H. (2006), "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure", *Econometrica*, 74, 967-1012.

[29] Pesaran, M.H. (2007), "A Simple Panel Unit Root Test in the Presence of Cross Section Dependence", *Journal of Applied Econometrics*, 22(2), 265-312.

[30] Pesaran, M.H., Schuermann, T. and Treutler, B.J. (2007), "Global Business Cycles and Credit Risk", Chapter 9 in Mark Carey and Rene M. Stultz (eds.), *The Risks of Financial Institutions*, University of Chicago Press (also available at: http://fic.wharton.upenn.edu/fic/papers/05/p0514.html).

[31] Pesaran, M.H., Schuermann, T. and Weiner, S.M. (2004), "Modelling Regional Interdependencies Using A Global Error-Correcting Macroeconometric Model", *Journal of Business and Economics Statistics*, 22, 129-162.

[32] Pesaran, M.H., Schuermann, T., Treutler, B.J. and Weiner, S.M. (2006), "Macroeconomic Dynamics and Credit Risk: A Global Perspective", *Journal of Money, Credit and Banking*, 38, 1211-1262.

[33] Pesaran, M.H. and Smith, R. (2006), "Macroeconometric Modelling with a Global Perspective", *The Manchester School*, Supplement, 24-49.

[34] Pesaran, M.H., Smith, L.V. and Smith, R.P. (2007), "What if the UK had Joined the Euro in 1999? An empirical evaluation using a Global VAR", *International Journal of Finance and Economics*, 12, 55-87.

[35] Pesaran, M.H. and Tosetti, E. (2007), "Large Panels with Common Factors and Spatial Correlations", IZA Discussion Paper No. 3032, September 2007.

[36] Sargent, T.J., and Sims, C.A. (1977), "Business cycle modeling without pretending to have too much a-priori economic theory", in: C. Sims *et al.* (eds.), *New Methods in Business Cycle Research*, Federal Reserve Bank of Minneapolis, Minneapolis.

[37] Sims, C. (1980), "Macroeconomics and Reality", *Econometrica*, 48, 1-48.

[38] Stock, J.H. and Watson, M.W. (1999), "Forecasting Inflation", *Journal of Monetary Economics*, 44, 293-335.

[39] Stock, J.H. and Watson, M.W. (2002), "Macroeconomic forecasting using diffusion indexes", *Journal of Business and Economic Statistics*, 20, 147-162.

[40] Stock, J.H. and Watson, M.W. (2005), "Implications of Dynamic Factor Models for VAR Analysis", NBER Working Paper No. 11467, June 2005.

[41] Whittle, P. (1954), "On Stationary Processes on the Plane", *Biometrika*, 41, 434-449.