Econometric Analysis of High Dimensional VARs Featuring a Dominant Unit*

Alexander Chudik†
European Central Bank, and CIMF

M. Hashem Pesaran‡
Cambridge University, and USC

April 1, 2011

Abstract

This paper extends the analysis of infinite dimensional vector autoregressive (IVAR) models proposed in Chudik and Pesaran (2011) to the case where one of the variables or the cross section units in the IVAR model is dominant or pervasive. It is an important extension from empirical as well theoretical perspectives. In the theory of networks a dominant unit is the centre node of a star network and arises as an efficient outcome of a distance-based utility model. Empirically, the extension poses a number of technical challenges that goes well beyond the analysis of IVAR models provided in Chudik and Pesaran. This is because the dominant unit influences the rest of the variables in the IVAR model both directly and indirectly, and its effects do not vanish as the dimension of the model (N) tends to infinity. The dominant unit acts as a dynamic factor in the regressions of the non-dominant units and yields an infinite order distributed lag relationship between the two types of units. Despite this it is shown that the effects of the dominant unit as well as those of the neighborhood units can be consistently estimated by running augmented least squares regressions that include distributed lag functions of the dominant unit and its neighbors (if any). The asymptotic distribution of the estimators is derived and their small sample properties investigated by means of Monte Carlo experiments.

Keywords: IVAR Models, Dominant Units, Star Networks, Large Panels, Weak and Strong Cross Section Dependence, Factor Models, Spatial Models.
JEL Classification: C10, C33, C51

*A preliminary version of this paper was presented at CEMMAP Workshop on "Cross Section Dependence", Institute of Fiscal Studies, London, at ECB Workshop on "International Linkages and the Macroeconomy: Applications of GVAR Modelling Approaches", at CIRANO-CIREQ Conference on "Econometrics of Interactions', Montreal, at the University of California, Riverside, and at Emory University. We are thankful to the participants at these venues, Sean Holly, Marco Lombardi and three anonymous referees for their helpful comments.

†European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany; e-mail: alexander.chudik@ecb.europa.eu. The views expressed in this paper are those of the authors and do not necessarily reflect those of the European Central Bank.

‡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

The econometric theory of vector autoregressive (VAR) models is well developed when the dimension of the model (N) is small and fixed whilst the number of time series observations (T) is large and expanding. This framework, however, is not satisfactory for many empirical applications where *both* dimensions N and T are large. Prominent examples include modelling of regional and national interactions, the panel data analysis of a large number of firms or industries over time. It is clear that without restrictions the parameters of the VAR model can not be consistently estimated in cases where N is relatively large, since in such cases the number of unknown parameters grows at a quadratic rate in N.

To circumvent this 'curse of dimensionality', several techniques have been suggested in the literature. These can be broadly characterized as: (*i*) data shrinkage, and (*ii*) parameter shrinkage. Factor models are examples of the former (see Geweke (1977), Sargent and Sims (1977), Forni and Lippi (2001), Forni et al. (2000), and Forni et al. (2004)). Spatial models, pioneered by Whittle (1954), and further developed by Cliff and Ord (1973), Anselin (1988), and Kelejian and Robinson (1995), and Bayesian type restrictions (e.g. Doan, Litterman, and Sims (1984)) are examples of the latter.

Chudik and Pesaran (2011) propose an alternative solution to the curse of dimensionality based on an *a priori* classification of the units into neighbors and non-neighbors.[1] Neighbors could be individual units or, more generally, linear combinations of the units (such as spatial or local averages). Based on this classification the coefficients corresponding to the non-neighboring units in the infinite dimensional VAR (IVAR) model are restricted to vanish in the limit as $N \to \infty$, whereas the neighborhood effects are left unrestricted. Such limiting restrictions on the parameters of the VAR model turns out to be equivalent to data shrinkage as $N \to \infty$. Chudik and Pesaran (CP) show that the properties of the IVAR model depend crucially on the degree of cross section dependence in the IVAR model. In the case where such dependencies are weak (in the sense formalized by Chudik, Pesaran, and Tosetti (2011)), CP establish that the IVAR model de-couples into separate individual regressions that can be estimated consistently. They also consider the case where the cross section units are strongly correlated, but confine their analysis to situations where the source of strong cross section dependence is external to the model and originates from a finite number of exogenously given common factors. For the latter case they propose a cross sectionally augmented least squares (CALS) estimator that they show to be consistent and asymptotically normal.

The present paper extends the analysis of CP to the case where one of the cross section units in the IVAR model is dominant or pervasive, in the sense that its direct or indirect effects on the rest of the system can lead to strong cross section dependence. An important example is the role of the US in the global economy. It is clear that in a multi-country analysis the US macro variables are likely to have pervasive effects on other variables in the global economy, considering that the US economy

[1] In this paper we assume that the network formation that underlies the classification of the units into neighbors and non-neighbors is given exogenously. However, recent theoretical developments in the analysis of economic and social networks can be used to relax this assumption, although such a task is beyond the scope of the present paper. Jackson (2008) provides a recent account of this literature.

accounts for more than a quarter of world output and has strong links to all major financial and capital markets. See, for example, Pesaran, Schuermann, and Weiner (2004). The dominance of the US economy raises not only the question of how to model the US macroeconomic variables, but also how to model the remaining economies. Another example could be the modelling of house prices in different regions in the UK, where developments in London tend to have widespread effects on all other regions. See, for example, Holly, Pesaran, and Yamagata (2011) for a recent application.

In the literature on social and economic networks, the star network represents an important example of a model with a dominant unit where all units are indirectly connected to each other through a single (dominant) unit at the center of the network. Such networks naturally arise as efficient outcomes of the distance-based utility model where costs of establishing links relative to benefits from the links fall in an intermediate range. See, for example, Proposition 6.1 in Jackson (2008).

Allowing for a dominant unit in the IVAR model is clearly important, but to date little is known about the estimation of such models. This paper contributes to the literature in this direction by: (*i*) deriving large N representations of cross section units, (*ii*) investigating the identification of parameters in such systems, and (*iii*) deriving asymptotic distribution of the proposed augmented least squares (ALS) estimators. This extension is not straightforward and involves several technical difficulties. The dominant unit influences the rest of the variables in the IVAR model both directly and indirectly, and its effects do not vanish as the dimension of the model (N) tends to infinity. The dominant unit acts as a dynamic factor in the regressions of the non-dominant units and induces infinite order distributed lag relations between the dominant and non-dominant units. Nevertheless, it is shown that the effects of the dominant unit as well as those of the neighborhood units can be consistently estimated by running ALS regressions that include distributed lag functions of the dominant unit. The asymptotic distribution of the estimators is derived and their small sample properties investigated by means of Monte Carlo experiments.

The remainder of this paper is organized as follows. Section 2 sets up the IVAR model with a dominant unit. Section 3 derives infinite order moving average or autoregressive approximations for the cross section units and discusses the conditions under which the IVAR model yields a dynamic factor model with the dominant unit acting as the factor. Section 4 considers the identification problem. The asymptotic distribution of the ALS estimator is derived in Section 5. Section 6 extends the analysis to the case where neighborhood effects and as well as a dominant unit are present. Section 7 allows for unobserved common factors. Section 8 investigates finite sample properties of the ALS estimator by means of Monte Carlo experiments. Section 9 provides some concluding remarks. Proofs and other technical details are given in the Appendix.

Notations: $\|\mathbf{A}\|_1 \equiv \max_{1\leq j\leq N} \sum_{i=1}^{N} |a_{ij}|$ denotes the column matrix norm of the $N \times N$ matrix $\mathbf{A}$, $\|\mathbf{A}\|_\infty \equiv \max_{1\leq i\leq N} \sum_{j=1}^{N} |a_{ij}|$ is the row matrix norm of $\mathbf{A}$. $\|\mathbf{A}\| = \sqrt{\varrho\left(\mathbf{A}'\mathbf{A}\right)}$ is the spectral norm of $\mathbf{A}$, where $\varrho\left(\mathbf{A}\right) = |\lambda_1\left(\mathbf{A}\right)|$ is the spectral radius of $\mathbf{A}$, and $\lambda_1\left(\mathbf{A}\right)$ is the largest eigenvalue (in absolute value) of $\mathbf{A}$.[2] All vectors are column vectors. The i^{th} row of $\mathbf{A}$ with its i^{th} element replaced by a 0 is denoted by $\mathbf{a}'_{-i} = (a_{i1}, a_{i2}, ..., a_{i,i-1}, 0, a_{i,i+1}, ..., a_{i,N})$. The i^{th} row of $\mathbf{A}$ with

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

its first and i^{th} elements replaced by 0 is denoted by $\mathbf{a}'_{-1,-i} = (0, a_{i2}, ..., a_{i,i-1}, 0, a_{i,i+1}, ..., a_{i,N})$. $\mathbf{a}_1 = (a_{11}, a_{21}, ..., a_{N1})'$ denotes the first column of $\mathbf{A}$. A matrix constructed from $\mathbf{A}$ by replacing its first column by a column vector of zeros is denoted as $\mathbf{A}_{-1}$. $\|x_t\|_{L_p}$ is L_p-norm of a random variable x_t, defined as $(E|x_t|^p)^{1/p}$. $a_n = O(b_n)$ denotes that the deterministic sequence $\{a_n\}$ is at most of order b_n. $x_n = O_p(y_n)$ states that random variable x_n is at most of order 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 are denoted by $\overset{d}{\to}$ and $\overset{p}{\to}$, respectively. Convergence in quadratic mean, and convergence in L_1 norm are denoted by $\overset{q.m.}{\to}$ and $\overset{L_1}{\to}$, respectively. We use K and ρ to denote positive real numbers that do not vary with N and/or T. $(N,T) \overset{j}{\to} \infty$ denotes joint asymptotics in N and T, with N and $T \to \infty$, in no particular order.

2 The IVAR Model with a Dominant Unit

Suppose we have T time series observations on N cross section units indexed by $i \in \mathcal{S}_{(N)} \equiv \{1, .., N\} \subseteq \mathbb{N}$. Both dimensions, N and T, are assumed to be large. For each point in time, t, and for each $N \in \mathbb{N}$, the N cross section observations are collected in the N dimensional vector, $\mathbf{x}_{(N),t} = \left(x_{(N),1t}, x_{(N),2t}, ..., x_{(N),Nt}\right)'$, and it is assumed that $\mathbf{x}_{(N),t}$ follows the VAR(1) model

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

where $\mathbf{\Phi}_{(N)}$ is an $N \times N$ matrix of unknown coefficients and $\mathbf{u}_{(N),t}$ is an $N \times 1$ vector of error terms. To distinguish high dimensional VAR models from the standard specifications we refer to the sequence of VAR models (1) of growing dimensions ($N \to \infty$) as the infinite dimensional VARs or IVARs for short.[3] The extension of the IVAR(1) to the p^{th} order IVAR model where p is fixed, is relatively straightforward and will not be attempted in this paper.

The explicit dependence of the variables and the parameters of the IVAR model on N is suppressed in the remainder of the paper to simplify the notations, but it will be understood that in general they vary with N, unless stated otherwise. In what follows we shall also focus on the problem of estimation of the parameters of individual units in (1). In particular, we consider the equation for the i^{th} unit that we write as

$$x_{it} = \sum_{j=1}^{N} \phi_{ij} x_{j,t-1} + u_{it}, \text{ for } t = 1, 2, ..., T. \tag{2}$$

Clearly, it is not possible to estimate all the N coefficients ϕ_{ij}, $j = 1, .., N$, when N and T grow at the same rate, unless suitable restrictions are placed on some of the coefficients. One such

[3] The sequence of models obtained from (1) for different values of N is compatible with both cases where $cov\left(x_{(N),it}, x_{(N),jt}\right)$ changes with N or is invariant to N. We allow for both possibilities since in some applications the covariance between individual units could change with the inclusion of a new unit - as it is likely to be the case when modelling firms or assets within expanding markets. For further details see Chudik and Pesaran (2011).

restriction is the 'cross section absolute summability condition',

$$\sum_{j=1}^{N} \left|\phi_{ij}\right| < K \text{ for any } N \in \mathbb{N} \text{ and any } i \in \{1, .., N\}\,, \tag{3}$$

which ensures that the variance of x_{it} conditional on information available at time $t - \ell$, for any fixed $\ell > 0$, exits for all N and as $N \to \infty$. The Lasso and Ridge shrinkage methods also use similar constraints.[4] Condition (3) implies that many of the coefficients are infinitesimal (as $N \to \infty$). However, assuming a mere existence of an upper bound K in (3) need not be sufficient to deal with the dimensionality problem and we impose additional restrictions below. We follow CP and suppose that in addition to (3), that for each $i \in \mathbb{N}$ it is possible to divide the units into 'neighbors' and 'non-neighbors'. But depart from CP by allowing one of the units, which we take to be the first unit without loss of generality, to be dominant or pervasive in the sense to be made precise below. Also given our focus, to simplify the analysis initially we set the neighborhood effects that do not relate to the dominant unit to zero. This restriction is relaxed in Section 6.[5]

ASSUMPTION 1 *(Coefficient matrix $\mathbf{\Phi}$)*

(a) (Dominant unit) There exists a constant $K < \infty$ (independent of i and N) such that $|\phi_{ii}| < K$, $|\phi_{i1}| < K$, for all $i \in \mathbb{N}$, and

$$\sum_{i=1}^{N} |\phi_{i1}| = O(N)\,. \tag{4}$$

(b) (Neighbors) There are no other neighbors other than the dominant unit which is a potential neighbor to all other units.

(c) (Non-neighbors) There exists a constant $K < \infty$ such that the coefficients corresponding to non-neighbors satisfy

$$\left\|\boldsymbol{\phi}_{-1}\right\|_{\infty} = \max_{j \in \{2,..,N\}} \left|\phi_{1j}\right| < \frac{K}{N}, \tag{5}$$

and

$$\left\|\boldsymbol{\phi}_{-1,-i}\right\|_{\infty} = \max_{j \in \{2,..,N\} \setminus \{i\}} \left|\phi_{ij}\right| < \frac{K}{N}, \tag{6}$$

for any $N \in \mathbb{N}$ and any $i \in \{2, 3, ..., N\}$, where $\boldsymbol{\phi}_{-1} = (0, \phi_{12}, \phi_{13}, ..., \phi_{1N})'$ and $\boldsymbol{\phi}_{-1,-i} = \left(0, \phi_{i2}, ..., \phi_{i,i-1}, 0, \phi_{i,i+1}, .., \phi_{iN}\right)'$.

The division of units in Assumption 1 imposes sufficient constraints that allows us to tackle the dimensionality problem. Consider the problem of estimating the unknown coefficients ϕ_{ii} and ϕ_{i1}.

[4] These 'data mining' methods attempt at estimating all the unknown coefficients of the i^{th} equation, ϕ_{ij}, $j = 1, .., N$, by minimizing $\sum_{t=1}^{T} u_{it}^2$ subject to $\sum_{j=1}^{N} \left|\phi_{ij}\right| \leq K$ (Lasso) or $\sum_{j=1}^{N} \phi_{ij}^2 \leq K$ (Ridge). But the outcome, perhaps not surprisingly, only yields a relatively small number of non-zero estimates. See Chapter 3.4.3 of Hastie, Tibshirani, and Friedman (2001) for detailed descriptions of Lasso and Ridge regression shrinkage methods.

[5] In a dynamic sense the lagged value of the i^{th} unit can also be viewed as the i^{th} neighbor, but we shall use the terminology of 'neighbors' for other units only.

We have

$$x_{it} = \phi_{ii} x_{i,t-1} + \phi_{i1} x_{1,t-1} + \sum_{j \neq 1,i} \phi_{ij} x_{j,t-1} + u_{it}, \tag{7}$$

for $i = 2, 3, ..., N$, and the estimation of the coefficients ϕ_{ii} and ϕ_{i1} depends on the stochastic behavior of the cross section average $\sum_{j \neq 1,i} \phi_{ij} x_{j,t-1}$, which captures the aggregate spatiotemporal impact of non-neighbors. CP shows that if $\{x_{it}\}$ is cross sectionally weakly dependent, then the aggregate impact of non-neighbors $\overset{q.m.}{\to} 0$ as $N \to \infty$ and therefore ignoring the non-neighbors would not be a problem for estimation of ϕ_{ii}. However, in our set-up, the unit 1 can potentially have a large impact on any of the remaining $N - 1$ units and therefore $\{x_{it}\}$ could be cross sectionally strongly dependent. In the case of strong cross section dependence, the aggregate impact of non-neighbors is $O_p(1)$, and it will not be possible to consistently estimate the coefficients ϕ_{ii} and $\phi_{i1}$by ignoring the non-neighborhood effects.

The coefficients in the first column of matrix $\mathbf{\Phi}$ correspond to the direct lagged impact of unit 1 on the rest of the system. The pervasive nature of unit 1 is characterized by (4), and represents an important departure from the set up in CP, where the influence of any of the cross section units on the rest of the system is restricted such that $\|\mathbf{\Phi}\| < K$. In this paper $\|\mathbf{\Phi}\|$ is allowed to be unbounded in N, but only through the dominant effects of unit 1.

Similar considerations also apply to contemporaneous dependence of the units through the error terms, $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$. Let

$$\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t, \tag{8}$$

where $\mathbf{R}$ is the $N \times N$ matrix of non-stochastic coefficients, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$ is an $N \times 1$ vector of random variables. This formulation is quite general and includes all models of spatial dependence considered in the literature, where it is assumed that $\mathbf{R}$ has bounded row and column matrix norms.[6] In the assumption below we relax this condition and allow for the sum of the coefficients in the first column of $\mathbf{R}$ to be unbounded in N.

ASSUMPTION 2 *(Error terms and contemporaneous dominance) The contemporaneous dependence of the errors $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$ in (1) is characterized by (8), where the individual elements of the double index array $\{\varepsilon_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ are distributed with mean 0, finite variances, and finite fourth moments uniformly bounded in $i \in \mathbb{N}$, and ε_{it} is independently distributed from $\varepsilon_{i't'}$ for any $(i,t) \neq (i',t')$. Consider the decomposition of* $\mathbf{R}$

$$\mathbf{R} = \mathbf{r}_1 \mathbf{s}_1' + \mathbf{R}_{-1}, \tag{9}$$

where $\mathbf{r}_1 = (r_{11}, r_{21}, ..., r_{N1})'$ is the first column of $\mathbf{R}$, $\mathbf{s}_1$ is an $N \times 1$ selection vector, $\mathbf{s}_1 = (1, 0, ..., 0)'$, and $\mathbf{R}_{-1}$ is obtained from $\mathbf{R}$ by replacing its first column with a vector of zeros. Assume that $r_{ii} = 1$ for all $i \in \mathbb{N}$ (without the loss of generality) and that there exists a constant $K < \infty$ (independent of i and N) such that

$$Var(\varepsilon_{it}) = \sigma_{\varepsilon i}^2 < K, \tag{10}$$

[6] See Pesaran and Tosetti (2011) for further details.

$$\|\mathbf{R}_{-1}\|_1 < K, \; \|\mathbf{R}_{-1}\|_\infty < K, \tag{11}$$

and

$$\|\mathbf{r}_{-1}\|_\infty = \max_{j\in\{2,..,N\}} |r_{1j}| < \frac{K}{N}, \tag{12}$$

for any $N \in \mathbb{N}$, where $\mathbf{r}_{-1} = (0, r_{12}, r_{13}, ..., r_{1N})'$ is the $N \times 1$ column vector constructed from the first row of $\mathbf{R}_{-1}$. In addition, $|r_{i1}| < K$, for all $i \in \mathbb{N}$, and

$$\sum_{i=1}^{N} |r_{i1}| = O(N). \tag{13}$$

Under this assumption the error of the first cross section unit acts as a (static) common factor for the rest of the units. Condition (13) allows for the first cross section unit to have a dominant effect on all the other cross section units. The boundedness of $\mathbf{R}_{-1}$ ensures that no other cross section units has a dominant effect on the rest of the units.

The above set up can be generalized to two or more dominant units so long as the number of such units is fixed and does not change with N. In this paper we focus on IVAR models with one dominant unit and assume that the dominant unit is known *a priori*. The problem of how to identify dominant units will be outside the scope of the present paper.

3 Large N Representations

The presence of a dominant unit in the IVAR model considerably complicates the analysis. This is because the effects of the dominant unit show up in all other units both contemporaneously as well as being distributed over time in the form of infinite order moving average or autoregressive representations. For empirical analysis it is important that conditions under which such infinite order processes can be well approximated by time series models with a finite number of unknown parameters are met. To this end we introduce a number of further assumptions restricting the behavior of $\mathbf{\Phi}$ and $\mathbf{R}$ for a finite N as well as when $N \to \infty$.

ASSUMPTION 3 *(Starting values and stationarity) Available observations are $\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_T$ with $\mathbf{x}_0 = \sum_{\ell=0}^{\infty} \mathbf{\Phi}^\ell \mathbf{u}(-\ell)$, and there exists a real positive constant $\rho < 1$ (independent of N) such that for any $N \in \mathbb{N}$*

$$|\lambda_1(\mathbf{\Phi})| \leq \rho. \tag{14}$$

ASSUMPTION 4 *(Bounded variances and invertibility of large N ARMA representations) Similarly to (9) let*

$$\mathbf{\Phi} = \boldsymbol{\phi}_1 \mathbf{s}_1' + \mathbf{\Phi}_{-1}, \tag{15}$$

where $\mathbf{\Phi}_{-1}$ is obtained from $\mathbf{\Phi}$ by replacing its first column with a column of zeros and $\boldsymbol{\phi}_1$ is the first column of $\mathbf{\Phi}$. Assume that there exists a real positive constant $\rho < 1$ (independent of N) such that for any $N \in \mathbb{N}$:

$$\|\mathbf{\Phi}_{-1}\|_1 \leq \rho, \; \|\mathbf{\Phi}_{-1}\|_\infty \leq \rho, \tag{16}$$

and

$$\|\boldsymbol{\phi}_1\|_\infty = \max_{1\leq i\leq N} |\phi_{i1}| \leq \rho. \tag{17}$$

Furthermore,

$$\max{}_{i\in\mathbb{N}} |r_{i1}| \leq 1. \tag{18}$$

Remark 1 *Condition (14) of Assumption 3 is a well known sufficient condition for covariance stationarity for any fixed $N \in \mathbb{N}$. This condition, however, is not sufficient for $Var(x_{it})$ to remain bounded as $N \to \infty$. As shown in Chudik and Pesaran (2011), $\|\boldsymbol{\Phi}\| \leq \rho < 1$ would be sufficient for bounded variances (as $N \to \infty$), but in our set-up $\|\boldsymbol{\Phi}\|$ is* unbounded *due to the presence of a dominant unit in the IVAR model. Assumption 4 provides additional* sufficient *conditions for bounded variances (as $N \to \infty$) and also for the existence of an invertible large N $AR(\infty)$ and $MA(\infty)$ processes for the dominant unit.*

Using the notations introduced in Assumptions 2 and 4 (see equations (9) and (15)), model (1) can be written as

$$\begin{aligned} \mathbf{x}_t &= \left(\boldsymbol{\phi}_1\mathbf{s}_1' + \boldsymbol{\Phi}_{-1}\right)\mathbf{x}_{t-1} + \left(\mathbf{r}_1\mathbf{s}_1' + \mathbf{R}_{-1}\right)\boldsymbol{\varepsilon}_t, \\ &= \boldsymbol{\phi}_1 x_{1,t-1} + \boldsymbol{\Phi}_{-1}\mathbf{x}_{t-1} + \mathbf{r}_1\varepsilon_{1t} + \mathbf{e}_t, \end{aligned} \tag{19}$$

where

$$\mathbf{e}_t = \mathbf{R}_{-1}\boldsymbol{\varepsilon}_t. \tag{20}$$

Solving for $\mathbf{x}_t$ by backward substitution yields

$$\mathbf{x}_t = \boldsymbol{\Phi}_{-1}(L)\,\boldsymbol{\phi}_1 x_{1,t-1} + \boldsymbol{\Phi}_{-1}(L)\,\mathbf{r}_1\varepsilon_{1t} + \boldsymbol{\upsilon}_t, \tag{21}$$

where

$$\boldsymbol{\Phi}_{-1}(L) = \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}_{-1}^{\ell} L^{\ell}, \tag{22}$$

and

$$\boldsymbol{\upsilon}_t = \boldsymbol{\Phi}_{-1}(L)\,\mathbf{e}_t. \tag{23}$$

Lemma 1 *Suppose Assumption 2-4 hold. Then for any $N \times 1$ vector $\mathbf{a}$ satisfying condition $\|\mathbf{a}\| = O\left(N^{-1/2}\right)$ we have*

$$Var\left(\mathbf{a}'\boldsymbol{\upsilon}_t\right) = O\left(N^{-1}\right),$$

where $\boldsymbol{\upsilon}_t$ is defined by (23).

Lemma 1 establishes that $\boldsymbol{\upsilon}_t$ is cross sectionally weekly dependent (CWD), and in particular $\mathbf{a}'\boldsymbol{\upsilon}_t = O_p\left(N^{-1/2}\right)$ for any vector $\mathbf{a}$ satisfying $\|\mathbf{a}\| = O\left(N^{-1/2}\right)$. For the non-dominant units, $i > 1$, using (21) we have

$$x_{it} = d_i(L)\,x_{1,t-1} + b_i(L)\,\varepsilon_{1t} + \upsilon_{it}, \tag{24}$$

where $\upsilon_{it} = \mathbf{s}_i'\boldsymbol{\upsilon}_t$,

$$d_i(L) = \mathbf{s}_i'\boldsymbol{\Phi}_{-1}(L)\boldsymbol{\phi}_1, \tag{25}$$

$$b_i(L) = \mathbf{s}_i'\boldsymbol{\Phi}_{-1}(L)\boldsymbol{\phi}_1, \tag{26}$$

and $\mathbf{s}_i$ is an $N \times 1$ dimensional selection vector with $s_{ij} = 0$ for $j \neq i$ and $s_{ii} = 1$. In the case of the dominant unit ($i = 1$) equation (21) yields,

$$c(L)x_{1t} = b_1(L)\varepsilon_{1t} + \upsilon_{1t}, \tag{27}$$

where

$$b_1(L) = \sum_{\ell=0}^{\infty}\left(\mathbf{s}_1'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{r}_1\right)L^{\ell}, \tag{28}$$

$$c(L) = 1 - \phi_{11}L - \boldsymbol{\phi}_{-1}'\boldsymbol{\Phi}_{-1}(L)\boldsymbol{\phi}_1 L^2, \tag{29}$$

and $\upsilon_{1t} = \mathbf{s}_1'\boldsymbol{\upsilon}_t$. Note that υ_{1t} can be written as

$$\begin{aligned}
\upsilon_{1t} &= \sum_{\ell=0}^{\infty}\mathbf{s}_1'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{e}_{t-\ell} = e_{1t} + \sum_{\ell=1}^{\infty}\mathbf{s}_1'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{e}_{t-\ell} \\
&= e_{1t} + \mathbf{s}_1'\boldsymbol{\Phi}_{-1}\sum_{\ell=1}^{\infty}\boldsymbol{\Phi}_{-1}^{\ell-1}\mathbf{e}_{t-\ell}.
\end{aligned}$$

But $\mathbf{s}_1'\boldsymbol{\Phi}_{-1} = \boldsymbol{\phi}_{-1}'$, and

$$\sum_{\ell=1}^{\infty}\boldsymbol{\Phi}_{-1}^{\ell-1}\mathbf{e}_{t-\ell} = \sum_{\ell=0}^{\infty}\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{e}_{t-\ell-1} = \boldsymbol{\upsilon}_{t-1}.$$

Hence

$$\upsilon_{1t} = e_{1t} + \boldsymbol{\phi}_{-1}'\boldsymbol{\upsilon}_{t-1}. \tag{30}$$

Also it is easily seen that $e_{1t} = \mathbf{s}_1'\mathbf{R}_{-1}\boldsymbol{\varepsilon}_t = \mathbf{r}_{-1}'\boldsymbol{\varepsilon}_t$, and $\boldsymbol{\upsilon}_{t-1} = \sum_{\ell=1}^{\infty}\boldsymbol{\Phi}_{-1}^{\ell-1}\mathbf{R}_{-1}\boldsymbol{\varepsilon}_{t-\ell}$, where both of these composite variables have zero means and are uncorrelated. Therefore

$$Var(\upsilon_{1t}) = Var\left(\mathbf{r}_{-1}'\boldsymbol{\varepsilon}_t\right) + Var\left(\boldsymbol{\phi}_{-1}'\boldsymbol{\upsilon}_{t-1}\right) = O\left(N^{-1}\right), \tag{31}$$

where

$$Var\left(\mathbf{r}_{-1}'\boldsymbol{\varepsilon}_t\right), = \mathbf{r}_{-1}'Var(\boldsymbol{\varepsilon}_t)\mathbf{r}_{-1} \leq \|\mathbf{r}_{-1}\|^2\|Var(\boldsymbol{\varepsilon}_t)\|,$$

$\|\mathbf{r}_{-1}\|^2 \leq \|\mathbf{r}_{-1}\|_{\infty}\|\mathbf{r}_{-1}\|_1 = O\left(N^{-1}\right)$ by (12) of Assumption 2 , $\|Var(\boldsymbol{\varepsilon}_t)\| < K$ by condition (10) of Assumption 2, and $Var\left(\boldsymbol{\phi}_{-1}'\boldsymbol{\upsilon}_{t-1}\right) = O\left(N^{-1}\right)$ follows from Lemma 1 by setting $\mathbf{a} = \boldsymbol{\phi}_{-1}$ and noting that $\|\boldsymbol{\phi}_{-1}\| \leq \sqrt{\|\boldsymbol{\phi}_{-1}\|_{\infty}\|\boldsymbol{\phi}_{-1}\|_1} = O\left(N^{-1/2}\right)$ by condition (5) of Assumption 1. Therefore, since $E(\upsilon_{1t}) = 0$, then

$$\upsilon_{1t} = O_p\left(N^{-1/2}\right), \tag{32}$$

and equation (27) can be written as

$$c(L)\, x_{1t} = b_1(L)\, \varepsilon_{1t} + O_p\left(N^{-1/2}\right), \tag{33}$$

which is a large N ARMA(∞,∞) representation of the process for the dominant unit.

The next lemma establishes invertibility of polynomials $b_1(L)$ and $c(L)$.

Lemma 2 *Suppose Assumption 4 holds. Then inverses of the polynomials $b_1(L)$ and $c(L)$, defined by (28) and (29), respectively, exist for any $N \in \mathbb{N}$, and coefficients of polynomials $b_1^{-1}(L)$ and $c^{-1}(L)$ decay at an exponential rate uniformly in N. Also, there exist real positive constants $K < \infty$ and $\rho < 1$ such that*

$$|a_\ell| < K\rho^\ell, \textit{ for any } \ell \in \{0, 1, 2, ...\} \textit{ and any } N \in \mathbb{N}, \tag{34}$$

where

$$a(L) = \sum_{\ell=0}^{\infty} a_\ell L^\ell = b_1^{-1}(L)\, c(L). \tag{35}$$

It is worth noting that conditions $\|\mathbf{\Phi}_{-1}\|_\infty \leq \rho < 1$ and $\|\boldsymbol{\phi}_1\|_\infty \leq \rho < 1$ of Assumption 4 are sufficient to ensure that $c(L)$ is invertible and the coefficients of $c^{-1}(L)$ decay exponentially. On the other hand conditions $\|\mathbf{\Phi}_{-1}\|_\infty \leq \rho < 1$ and $\max_{i\in\mathbb{N}} |r_{i1}| \leq 1$, are sufficient in ensuring that $b_1(L)$ is invertible and the coefficients of $b_1^{-1}(L)$ decay exponentially. The exponential decay of the coefficients in these polynomials will be relevant for the selection of truncation lags in empirical applications as discussed below.

3.1 Large N AR and MA representations for the dominant unit

Multiplying both sides of (27) by $b_1^{-1}(L)$ we obtain

$$a(L)\, x_{1t} = \varepsilon_{1t} + \vartheta_{bt}, \tag{36}$$

where $\vartheta_{bt} = b_1^{-1}(L)\, \upsilon_{1t}$. By Lemma 2 the coefficients of $b_1^{-1}(L)$ decay exponentially and hence are absolute summable, and in view of (31) we have

$$Var(\vartheta_{bt}) = O\left(N^{-1}\right). \tag{37}$$

Also since $E(\vartheta_{bt}) = 0$, it follows that

$$\vartheta_{bt} = b_1^{-1}(L)\, \upsilon_{1t} = O_p\left(N^{-1/2}\right). \tag{38}$$

Using this result in (36) yields the following large N AR(∞) representation for the dominant unit,

$$a(L)\, x_{1t} = \varepsilon_{1t} + O_p\left(N^{-1/2}\right). \tag{39}$$

Similarly, multiplying both sides of (27) by $c^{-1}(L)$ we obtain

$$x_{1t} = a^{-1}(L)\varepsilon_{1t} + \vartheta_{ct}, \tag{40}$$

where $a^{-1}(L) = c^{-1}(L)b_1(L)$, and $\vartheta_{ct} = c^{-1}(L)\upsilon_{1t}$. Using similar arguments as in derivation of (37)

$$Var(\vartheta_{ct}) = O\left(N^{-1}\right), \tag{41}$$

and since $E(\vartheta_{ct}) = 0$, then

$$\vartheta_{ct} = c^{-1}(L)\upsilon_{1t} = O_p\left(N^{-1/2}\right), \tag{42}$$

and we have the following large N MA(∞) representation for x_{1t},

$$x_{1t} = a^{-1}(L)\varepsilon_{1t} + O_p\left(N^{-1/2}\right). \tag{43}$$

3.2 Large N representation for the non-dominant units $i > 1$

Consider now the equation for unit $i > 1$. Using (1) we have (noting that $u_{it} = r_{i1}\varepsilon_{1t} + e_{it}$)

$$x_{it} = \phi_{ii}x_{i,t-1} + \boldsymbol{\phi}'_{-1,-i}\mathbf{x}_{t-1} + \phi_{i1}x_{1,t-1} + r_{i1}\varepsilon_{1t} + e_{it}. \tag{44}$$

Multiplying both sides of (21) by $\boldsymbol{\phi}'_{-1,-i}$ yields

$$\boldsymbol{\phi}'_{-1,-i}\mathbf{x}_t = \boldsymbol{\phi}'_{-1,-i}\mathbf{\Phi}_{-1}(L)\boldsymbol{\phi}_1 x_{1,t-1} + \boldsymbol{\phi}'_{-1,-i}\mathbf{\Phi}_{-1}(L)\mathbf{r}_1\varepsilon_{1t} + \boldsymbol{\phi}'_{-1,-i}\boldsymbol{\upsilon}_t. \tag{45}$$

Substituting (45) in (44) and using (27) to eliminate ε_{1t} from (44) we have

$$x_{it} = \phi_{ii}x_{i,t-1} + \beta_i(L)x_{1t} + e_{it} + \zeta_{it}, \tag{46}$$

where

$$\beta_i(L) = \phi_{i1}L + \boldsymbol{\phi}'_{-1,-i}\mathbf{\Phi}_{-1}(L)\boldsymbol{\phi}_1 L^2 + \left[r_{i1} + \boldsymbol{\phi}'_{-1,-i}\mathbf{\Phi}_{-1}(L)\mathbf{r}_1 L\right]a(L), \tag{47}$$

and

$$\zeta_{it} = \boldsymbol{\phi}'_{-1,-i}\boldsymbol{\upsilon}_{t-1} - \left[r_{i1} + \boldsymbol{\phi}'_{-1,-i}\mathbf{\Phi}_{-1}(L)\mathbf{r}_1 L\right]\vartheta_{bt}. \tag{48}$$

Taking L_2-norm of (48) and using triangle inequality we obtain

$$\|\zeta_{it}\|_{L_2} \leq \left\|\boldsymbol{\phi}'_{-1,-i}\boldsymbol{\upsilon}_{t-1}\right\|_{L_2} + \left\|\left[r_{i1} + \boldsymbol{\phi}'_{-1,-i}\mathbf{\Phi}_{-1}(L)\mathbf{r}_1 L\right]\vartheta_{bt}\right\|_{L_2}. \tag{49}$$

But under condition (6) in Assumption 1, we have $\left\|\boldsymbol{\phi}_{-1,-i}\right\|_{\infty} = O\left(N^{-1}\right)$ uniformly in $i \in \{2, 3, ...\}$, which implies that $\left\|\boldsymbol{\phi}_{-1,-i}\right\| = O\left(N^{-1/2}\right)$, and it follows from Lemma 1 (by setting $\mathbf{a} = \boldsymbol{\phi}_{-1,-i}$) that

$$Var\left(\boldsymbol{\phi}'_{-1,-i}\boldsymbol{\upsilon}_{t-1}\right) = O\left(N^{-1}\right), \text{ uniformly in } i \in \{2, 3, ...\},$$

and since $E(\boldsymbol{\upsilon}_t) = \mathbf{0}$, then

$$\left\| \boldsymbol{\phi}'_{-1,-i} \boldsymbol{\upsilon}_{t-1} \right\|_{L_2} = O\left(N^{-1/2}\right), \text{ uniformly in } i \in \{2, 3, ...\}. \tag{50}$$

Also by (37) and noting that the coefficients of $\boldsymbol{\phi}'_{-1,-i} \boldsymbol{\Phi}_{-1}(L) \mathbf{r}_1$ decay exponentially to zero uniformly in $i \in \{2, 3, ...\}$ (see proof of Lemma 3 below) and $E(\vartheta_{bt}) = 0$, we have

$$\left\| \left[r_{i1} + \boldsymbol{\phi}'_{-1,-i} \boldsymbol{\Phi}_{-1}(L) \mathbf{r}_1 L \right] \vartheta_{bt} \right\|_{L_2} = O\left(N^{-1/2}\right), \text{ uniformly in } i \in \{2, 3, ...\}. \tag{51}$$

Using (50) and (51) in (49) and noting that $E(\zeta_{it}) = 0$, we have

$$Var(\zeta_{it}) = \|\zeta_{it}\|^2_{L_2} = O\left(N^{-1}\right), \text{ uniformly in } i \in \{2, 3, ...\}, \tag{52}$$

and

$$\zeta_{it} = O_p\left(N^{-1/2}\right), \text{ uniformly in } i \in \{2, 3, ...\}. \tag{53}$$

Hence, the large N representation of the process for the non-dominant unit $i > 1$ is given by

$$x_{it} = \phi_{ii} x_{i,t-1} + \beta_i(L) x_{1t} + e_{it} + O_p\left(N^{-1/2}\right). \tag{54}$$

The following proposition summarizes the main results derived in this section.

Proposition 1 *Let Assumptions 1-4 hold. Then, as $N \to \infty$, the dominant unit $i = 1$ can be modelled on its own as infinite AR or MA process and its large N AR and MA representations are given by (39) and (43), respectively. Large N representation for other units, $i > 1$, is given by (54), in which the dominant unit acts as an observed dynamic common factor.*

It is valid to exclude the contemporaneous values of x_{1t} from (54) if and only if $r_{i1} = 0$, for $i > 1$. However, $x_{1,t-1}$ enters the regression equation for the i^{th} unit even if $r_{i1} = \phi_{i1} = 0$. Note also that in general the polynomial $\beta_i(L)$ is of infinite order, and the errors, e_{it}, are serially uncorrelated but cross sectionally weakly dependent. The following lemma establishes that the coefficients in the polynomial $\beta_i(L)$ decline at a geometric rate (uniformly in i).

Lemma 3 *Suppose Assumption 4 holds. Then there exist real positive constants $K < \infty$ and $0 < \rho < 1$ such that*

$$|\beta_{i\ell}| < K\rho^\ell \text{ for any } \ell \in \{0, 1, 2, ...\}, \text{ any } N \in \mathbb{N} \text{ and any } i \in \{1, 2, ..., N\}, \tag{55}$$

where $\beta_{i\ell}$ is defined by the coefficients of polynomial $\beta_i(L) = \sum_{\ell=0}^{\infty} \beta_{i\ell} L^\ell$ in (47).

3.3 Large N representation for cross section averages

Cross section averages can also be used to capture the effects of the dominant unit, which acts as an observable common factor in the large N representation for the non-dominant units. Pesaran and Chudik (2011) show in the context of high dimensional VARs that the components with weak cross section dependence do not survive the cross section aggregation of a large number of units. Since the dominant unit in this section is the only source of strong cross section dependence, it is not surprising that there should be a relationship between the dominant unit, x_{1t}, and the cross section averages, $\overline{x}_{wt} = \mathbf{w}'\mathbf{x}_t$, where $\mathbf{w}$ is any weight vector satisfying the following granularity conditions:

$$\|\mathbf{w}\| = O\left(N^{-\frac{1}{2}}\right), \tag{56}$$

$$\frac{w_i}{\|\mathbf{w}\|} = O\left(N^{-\frac{1}{2}}\right) \text{ for any } i. \tag{57}$$

A simple example of granular weights are equal weights $w_i = N^{-1}$ for $i = 1, 2, ..., N$. In order to derive the relationship between the cross section averages and the dominant unit, multiply equation (21) by $\mathbf{w}'$ to obtain

$$\overline{x}_{wt} = \mathbf{w}'\mathbf{\Phi}_{-1}\left(L\right)\boldsymbol{\phi}_1 x_{1,t-1} + \mathbf{w}'\mathbf{\Phi}_{-1}\left(L\right)\mathbf{r}_1\varepsilon_{1t} + \mathbf{w}'\boldsymbol{\upsilon}_t, \tag{58}$$

in which $\mathbf{w}'\boldsymbol{\upsilon}_t = O_p\left(N^{-1/2}\right)$ by Lemma 1. Now substituting equation (36) for ε_{1t} and noting that $\mathbf{w}'\mathbf{\Phi}_{-1}\left(L\right)\mathbf{r}_1\vartheta_{bt} = O_p\left(N^{-1/2}\right)$, we obtain

$$\overline{x}_{wt} = \varphi_w\left(L\right)x_{1t} + O_p\left(N^{-1/2}\right), \tag{59}$$

where

$$\varphi_w\left(L\right) = \mathbf{w}'\mathbf{\Phi}_{-1}\left(L\right)\boldsymbol{\phi}_1 L + \left[\mathbf{w}'\mathbf{\Phi}_{-1}\left(L\right)\mathbf{r}_1\right]a\left(L\right), \tag{60}$$

and $a(L)$ is given by (35). Equation (59) shows that, as $N \to \infty$, $\overline{\mathbf{x}}_{wt}$ can be written as a distributed lag function of x_{1t}, and if $\varphi_w\left(L\right)$ is invertible, then the dominant unit, x_{1t}, can be approximated arbitrarily well by the cross section averages and their lags. Therefore, augmentation by cross section averages to take account of the effects of strong cross section dependence for the estimation of the dynamic coefficients ϕ_{ii}, for $i > 1$, should be asymptotically equivalent to the augmentation by the dominant unit and its lags. This equivalence property is also investigated in Monte Carlo experiments below. The idea of using cross section averages to take into account the effects of strong cross section dependence was originally proposed by Pesaran (2006) in the context of large heterogenous panels with unobserved common factors.

4 Identification

Note that the first two coefficients in $\beta_i(L) = \sum_{\ell=0}^{\infty}\beta_{i\ell}L^{\ell}$, as defined by (47), are (for $i = 2, 3, ..., N$)

$$\beta_{i0} = r_{i1}, \tag{61}$$

and

$$\beta_{i1} = \phi_{i1} + r_{i1}a_1 + \boldsymbol{\phi}'_{-1,-i}\mathbf{r}_1 a_0 = \phi_{i1} - r_{i1}\left(\boldsymbol{\phi}'_{-1}\mathbf{r}_1 + \phi_{11}\right) + \boldsymbol{\phi}'_{-1,-i}\mathbf{r}_1. \tag{62}$$

Also using $c(L) = \sum_{\ell=0}^{\infty} c_\ell L^\ell$ and $b_1(L) = \sum_{\ell=0}^{\infty} b_{1\ell} L^\ell$ given by (29) and (28), respectively, we have,

$$c_0 = 1,\ c_1 = -\phi_{11},\quad \text{and } b_{10} = 1,\ b_{11} = \boldsymbol{\phi}_{-1}' \mathbf{r}_1.$$

Hence, (using $a(L)$ in (35)) we have

$$a_0 = 1,\, a_1 = -\phi_{11} - \boldsymbol{\phi}_{-1}' \mathbf{r}_1, \tag{63}$$

The higher order lag coefficients, $\beta_{i\ell}$ and a_ℓ for $\ell = 2, 3, ...$, in general depend on all elements of $\mathbf{\Phi}$ and $\mathbf{r}_1$ and can be obtained similarly.

Result (61) shows that the contemporaneous effects of the dominant unit on the rest of the units, r_{i1}, for $i > 1$, can be identified from β_{i0}, which can be consistently estimated using the unit-specific ALS regressions specified below. The own-lag effects of the non-dominant units, ϕ_{ii} (for $i > 1$), can also be consistently estimated.

But due to the feedback effects from non-dominant units, the own-lag effect of the dominant unit, ϕ_{11}, cannot be identified from a_1. To see this from (63) we note that $\phi_{11} = -a_1 + \boldsymbol{\phi}_{-1}' \mathbf{r}_1$, where $\boldsymbol{\phi}_{-1}' \mathbf{r}_1 = \Sigma_{i=2}^{N} \phi_{1i} r_{i1}$, $\max_{i>1} |\phi_{1i}| < KN^{-1}$, and the coefficients $r_{i1}, i > 1$, are fixed in N. Hence $\boldsymbol{\phi}_{-1}' \mathbf{r}_1$ is $O(1)$ and does not vanish as $N \to \infty$. Using the parameters from the large N representation for the non-dominant units we are able to identify r_{i1}. But due to the non-negligible lagged effects from the non-dominant units on the dominant unit, the parameters ϕ_{1i}, for $i > 1$ can not be identified when $N \to \infty$. As a result a consistent estimate of $\Sigma_{i=2}^{N} \phi_{1i} r_{i1}$ can not be obtained. Consequently, ϕ_{11} is not identified when $N \to \infty$. Accordingly, in the Monte Carlo experiments below, we shall only consider the estimation of β_{i0} and ϕ_{ii}.

5 Asymptotic Distribution of the Augmented Least Squares Estimator

5.1 Specification of Augmented Regressions

Based on the large N representation, (39), for the dominant unit, and the representation (54) for the non-dominant units ($i > 1$), we consider the following regressions:

$$x_{it} = \mathbf{g}_{it}' \boldsymbol{\pi}_i + \epsilon_{it}, \text{ for } i = 1, 2, ..., N, \tag{64}$$

where

$$\mathbf{g}_{it} = \begin{cases} (x_{1,t-1}, x_{1,t-2}, ..., x_{1,t-m})', & \text{for } i = 1 \\ (x_{i,t-1}, x_{1t}, x_{1,t-1}, ..., x_{1,t-m})' & \text{for } i > 1 \end{cases}, \tag{65}$$

$$\boldsymbol{\pi}_i = \begin{cases} -(a_1, a_2, ..., a_m)', & \text{for } i = 1 \\ (\phi_{ii}, \beta_{i0}, \beta_{i1}, ..., \beta_{im})' & \text{for } i > 1 \end{cases}, \tag{66}$$

$$\epsilon_{it} = \begin{cases} \psi_{m1t} + \vartheta_{bt} + \varepsilon_{1t}, & \text{for } i = 1 \\ \psi_{mit} + \zeta_{it} + e_{it} & \text{for } i > 1 \end{cases}, \tag{67}$$

and

$$\psi_{mit} = \begin{cases} -\sum_{\ell=m+1}^{\infty} a_\ell x_{1,t-\ell}, & \text{for } i=1 \\ \sum_{\ell=m+1}^{\infty} \beta_{i\ell} x_{1,t-\ell} & \text{for } i>1 \end{cases}. \tag{68}$$

Note that there are m regressors (and m unknown coefficients) in the regression for the dominant unit $i = 1$, and $m + 2$ regressors in the regressions for the non-dominant units, $i > 1$.

The error term ϵ_{it} in (67) is decomposed into three parts. The first term, ψ_{mit}, is due to the truncation of the infinite order lag polynomials $a(L)$ in the case of the dominant unit, and $\beta_i(L)$, for $i > 1$. Since the coefficients in these polynomials are absolutely summable, we have

$$\psi_{mit} \overset{q.m.}{\to} 0, \text{ as } m \to \infty,$$

for any $N \in \mathbb{N}$, any $i \in \{1, 2, ..., N\}$ and any $t \in \{1, 2, ..., T\}$. The second terms, ϑ_{bt} (in the case of the dominant unit), and ζ_{it}, for $i > 1$, are $O_p\left(N^{-1/2}\right)$. (See (38) and (53)). These terms arise from aggregation of weak dependencies in the individual-specific equations of the IVAR model, (1). The third term in (67) are serially uncorrelated errors, with ε_{1t} being orthogonal to e_{it} for any $i > 1$. Also as noted above, e_{it} is a cross sectionally weakly dependent process and as such ignoring it does not adversely impact the consistency of the estimators to be proposed here.

For future references, let

$$\mathbf{h}_{it} = \begin{cases} \left(\xi_{1,t-1}, \xi_{1,t-2}, ..., \xi_{1,t-m}\right) & \text{for } i=1 \\ \left(\xi_{i,t-1}, \xi_{1t}, \xi_{1,t-1}, ..., \xi_{1,t-m}\right) & \text{for } i>1 \end{cases}, \tag{69}$$

and

$$\mathbf{C}_i = E\left(\mathbf{h}_{it}\mathbf{h}_{it}'\right), \tag{70}$$

where

$$a(L)\,\xi_{1t} = \varepsilon_{1t}, \tag{71}$$

and

$$(1 - \phi_{ii}L)\,\xi_{it} = \beta_i(L)\,\xi_{1t} + e_{it}, \text{ for } i = 2, 3, ...N. \tag{72}$$

Process $\{\xi_{it}\}$ is the large N counterpart of $\{x_{it}\}$ in the following sense,

$$x_{it} - \xi_{it} = O_p\left(N^{-1/2}\right), \text{ for any } i \in \mathbb{N}. \tag{73}$$

Note that for any i, ξ_{it} is a linear stationary process with absolute summable autocovariances.

5.2 Consistency of the Augmented Least Squares Estimator

In what follows we focus on the estimation of the parameters of the non-dominant units, $i > 1$. The results for the dominant unit can be derived in a similar manner and to save space are not included . We denote the least squares estimator of the vector of unknown coefficients $\boldsymbol{\pi}_i$ as

$$\underset{m\times 1}{\widehat{\boldsymbol{\pi}}_1} = \begin{pmatrix} -\widehat{a}_1 \\ -\widehat{a}_2 \\ \vdots \\ -\widehat{a}_m \end{pmatrix} \text{ and } \underset{(m+2)\times 1}{\widehat{\boldsymbol{\pi}}_i} = \begin{pmatrix} \widehat{\phi}_{ii} \\ \widehat{\beta}_{i0} \\ \vdots \\ \widehat{\beta}_{im} \end{pmatrix}, \text{ for } i > 1,$$

where $\widehat{\phi}_{ii}$ refers to the augmented least squares (ALS) estimator of the own lag coefficient ϕ_{ii}, $\widehat{\beta}_{i\ell}$, $\ell = 0, 1, 2, ..., m$, denote the estimators of the first $m+1$ coefficients in $\beta_i(L)$, and $\widehat{a}_\ell$ for $\ell = 1, 2, ..., m$ denote the estimators of the corresponding coefficients in $a(L)$.

It is convenient to re-write (64) for $t = m+1, m+2, ..., T$ in matrix notations

$$\mathbf{x}_{i\cdot} = \mathbf{G}_i \boldsymbol{\pi}_i + \boldsymbol{\epsilon}_{i\cdot}, \text{ for } i > 1, \tag{74}$$

where

$$\underset{(T-m)\times(m+2)}{\mathbf{G}_i} = \begin{pmatrix} \mathbf{g}'_{i,m+1} \\ \mathbf{g}'_{i,m+2} \\ \vdots \\ \mathbf{g}'_{i,T} \end{pmatrix}, \underset{(T-m)\times 1}{\mathbf{x}_{i\cdot}} = \begin{pmatrix} x_{i,m+1} \\ x_{i,m+2} \\ \vdots \\ x_{i,T} \end{pmatrix}, \text{ and } \underset{(T-m)\times 1}{\boldsymbol{\epsilon}_{i\cdot}} = \begin{pmatrix} \epsilon_{i,m+1} \\ \epsilon_{i,m+2} \\ \vdots \\ \epsilon_{i,T} \end{pmatrix}. \tag{75}$$

Hence,

$$\widehat{\boldsymbol{\pi}}_i = \left(\mathbf{G}'_i \mathbf{G}_i\right)^{-1} \mathbf{G}_i \mathbf{x}_{i\cdot} \,. \tag{76}$$

In the general case where $\beta_i(L)$ is not a finite order polynomial the truncation lag m has to be selected depending on the available time series data, T, so that omission of the higher order lags of x_{1t} is asymptotically negligible. We use subscript T to denote this explicit dependence of the truncation lag on the available time series data in the remainder of this paper, namely we set $m_T = m(T)$, and consider the following assumptions on the relative expansion rates of N, T and m_T.

ASSUMPTION B1 $m_T^3/T \to \varkappa_1$, where $0 < \varkappa_1 < \infty$, as $T \to \infty$.
ASSUMPTION B2 $(N,T) \xrightarrow{j} \infty$ at any order.
ASSUMPTION B3 $(N,T) \xrightarrow{j} \infty$, and $T/N \to \varkappa_2$, where $0 < \varkappa_2 < \infty$.

Remark 2 *Assumption B1 gives a sufficient condition on the truncation lag m_T under which $\widehat{\boldsymbol{\pi}}_i$ is consistent and asymptotically normal. Assumption B1 can also be replaced by the following two conditions:*

$$m_T^2/T \to 0, \tag{77}$$

and

$$\lim_{T\to\infty} \rho^{m_T}\sqrt{T} = 0 \textit{ for any } 0 < \rho < 1. \tag{78}$$

Condition (78) ensures that m_T increases sufficiently rapidly so that the omitted variable problem from truncation of higher order lags is asymptotically negligible. Condition (77) ensures a sufficient degree of freedom to reliably estimate individual coefficients. Under Assumption B1 both of the above two conditions will be satisfied.

Identification of $\boldsymbol{\pi}_i$ requires invertibility of $\mathbf{G}_i'\mathbf{G}_i$, which is postulated in the following assumption.

ASSUMPTION 5 *There exist integers $T_0 \in \mathbb{N}$ and $N_0 \in \mathbb{N}$ such that for all $T \geq T_0$, and $N \geq N_0$, matrix $\mathbf{G}_i'\mathbf{G}_i$ is invertible.*

Let

$$\widehat{\mathbf{C}}_i = \frac{1}{T}\mathbf{G}_i'\mathbf{G}_i. \tag{79}$$

Substitute (74) in (76) to obtain

$$\begin{aligned}
\sqrt{T}\left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right) &= \widehat{\mathbf{C}}_i^{-1}\frac{\mathbf{G}_i'\boldsymbol{\epsilon}_{i\cdot}}{\sqrt{T}}, \\
&= \left(\widehat{\mathbf{C}}_i^{-1} - \mathbf{C}_i^{-1}\right)\frac{\mathbf{G}_i'\boldsymbol{\epsilon}_{i\cdot}}{\sqrt{T}} + \mathbf{C}_i^{-1}\frac{\mathbf{G}_i'\boldsymbol{\epsilon}_{i\cdot}}{\sqrt{T}} \\
&= \left(\widehat{\mathbf{C}}_i^{-1} - \mathbf{C}_i^{-1}\right)\frac{\mathbf{G}_i'\boldsymbol{\epsilon}_{i\cdot}}{\sqrt{T}} + \\
&\quad + \mathbf{C}_i^{-1}\left[\frac{\left(\mathbf{G}_i - \mathbf{H}_i\right)'\mathbf{e}_{i\cdot}}{\sqrt{T}} + \frac{\mathbf{H}_i'\mathbf{e}_{i\cdot}}{\sqrt{T}} + \frac{\mathbf{G}_i'\boldsymbol{\zeta}_{i\cdot}}{\sqrt{T}} + \frac{\mathbf{G}_i'\boldsymbol{\psi}_{i\cdot}}{\sqrt{T}}\right], \text{ for } i > 1,
\end{aligned} \tag{80}$$

where

$$\underset{(T-m_T)\times(m_T+2)}{\mathbf{H}_i} = \begin{pmatrix} \mathbf{h}_{i,m_T+1}' \\ \mathbf{h}_{i,m_T+2}' \\ \vdots \\ \mathbf{h}_{i,T}' \end{pmatrix}, \tag{81}$$

and

$$\underset{(T-m_T)\times 1}{\mathbf{e}_{i\cdot}} = \begin{pmatrix} e_{i,m_T+1} \\ e_{i,m_T+2} \\ \vdots \\ e_{iT} \end{pmatrix}, \underset{(T-m_T)\times 1}{\boldsymbol{\zeta}_{i\cdot}} = \begin{pmatrix} \zeta_{i,m_T+1} \\ \zeta_{i,m_T+2} \\ \vdots \\ \zeta_{iT} \end{pmatrix}, \underset{(T-m_T)\times 1}{\boldsymbol{\psi}_{i\cdot}} = \begin{pmatrix} \psi_{m_T,i,m_T+1} \\ \psi_{m_T,i,m_T+2} \\ \vdots \\ \psi_{m_T iT} \end{pmatrix}. \tag{82}$$

Note that $\boldsymbol{\epsilon}_{i\cdot} = \mathbf{e}_{i\cdot} + \boldsymbol{\zeta}_{i\cdot} + \boldsymbol{\psi}_{i\cdot}$, for $i > 1$, see (67).

We deal with the estimation of infinite order lag polynomials in a similar way as in Said and Dickey (1984) or Berk (1974) by selecting the truncation lag, m_T, as a suitable function of the sample size. However, the consistency and the asymptotic distribution of $\widehat{\boldsymbol{\pi}}_i$ does not automatically follow from the existing literature, since, aside from the issue of lag truncation, we also need to deal with the effects resulting from the aggregation of weakly cross sectionally dependent processes in the IVAR model (1), which is not straightforward. Consistency of $\widehat{\boldsymbol{\pi}}_i$ is established in the following theorem.

Theorem 1 *(Consistency) Suppose* $\mathbf{x}_t$ *is given by model (1) and Assumptions 1-5, B1, and B2 hold. Then*

$$\left\| \widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i \right\|_\infty \xrightarrow{p} 0, \text{ for any } i \in \mathbb{N}, \tag{83}$$

that is $\widehat{\boldsymbol{\pi}}_i$ *defined by equation (76) is a consistent estimator of* $\boldsymbol{\pi}_i$.

5.3 Asymptotic Distribution of $\widehat{\boldsymbol{\pi}}_i$

We continue to focus on the estimates $\widehat{\pi}_i$ for $i > 1$. Derivation of the asymptotic results for $\widehat{\pi}_1$ can be established in a similar manner.

Theorem 2 *(Asymptotic normality) Suppose* $\mathbf{x}_t$ *is given by model (1) and Assumptions 1-5, B1, and B3 hold. Then for any sequence of* $(m_T + 2) \times 1$ *dimensional vectors* $\mathbf{a}$ *such that* $\left\| \mathbf{a} \right\|_1 = O(1)$, *we have*

$$\sqrt{T} \frac{1}{\sigma_i} \mathbf{a}' \mathbf{C}_i^{\frac{1}{2}} \left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i \right) \xrightarrow{d} N(0,1), \text{ for any } i \in \{2, 3, ...\}, \tag{84}$$

where $\widehat{\boldsymbol{\pi}}_i$ *and* $\mathbf{C}_i$ *are defined by (76) and (70), respectively, and* $\sigma_i^2 = Var(e_{it})$. *Furthermore, for any sequence of* $m_T \times 1$ *dimensional vectors* $\mathbf{b}$ *such that* $\left\| \mathbf{b} \right\|_1 = O(1)$, *we have*

$$\sqrt{T} \frac{1}{\sigma_{\varepsilon 1}} \mathbf{b}' \mathbf{C}_1^{\frac{1}{2}} \left(\widehat{\boldsymbol{\pi}}_1 - \boldsymbol{\pi}_1 \right) \xrightarrow{d} N(0,1), \tag{85}$$

where $\widehat{\boldsymbol{\pi}}_1$ *and* $\mathbf{C}_1$ *are defined by (76) and (70), respectively, and* $\sigma_{\varepsilon 1}^2 = Var(\varepsilon_{1t})$.

6 Allowing for Neighborhood Effects as well as a Dominant Unit

The following assumption generalizes Assumption 1, and allows for neighborhood effects in addition to the effects that originate from the dominant unit.

ASSUMPTION 6 *(Neighbors and non-neighbors) Let*

$$\boldsymbol{\phi}_{-1} = \mathbf{S}_1 \boldsymbol{\delta}_1 + \boldsymbol{\phi}_{b1}, \tag{86}$$

and

$$\boldsymbol{\phi}_{-1,-i} = \mathbf{S}_i \boldsymbol{\delta}_i + \boldsymbol{\phi}_{bi}, \text{ for } i = 2, 3, ..., N \tag{87}$$

where $\mathbf{S}_i$, *for* $i = 1, 2, ..., N$, *are known* $k_i \times N$ *matrices that define* k_i *neighbors of unit* i, $k_i < K$, $\boldsymbol{\delta}_i$ *is* $k_i \times 1$ *vector of unknown parameters, and the coefficients corresponding to non-neighbors are characterized by vectors* $\boldsymbol{\phi}_{bi}$, $i = 1, 2, ..., N$, *and satisfy*

$$\left\| \boldsymbol{\phi}_{bi} \right\|_\infty < \frac{K}{N}, \tag{88}$$

for any $N \in \mathbb{N}$, *and any* $i \in \{1, 2, 3, ..., N\}$.

Let us denote the k_i neighbors of unit i by the $k_i \times 1$ vector $\mathfrak{n}_{it} = \mathbf{S}_i' \mathbf{x}_t$. The non-zero elements of $\mathbf{S}_i$ define the neighbors of unit i, and are closely related to the elements of the *adjancency matrix* in the literature on networks. See, for example, Chapter 2 in Jackson (2008). It is assumed that $\mathbf{S}_i$ is known and does not vary with N. A familiar example of $\mathbf{S}_i$ would contain only $(0,1)$ entries that selects individual units as neighbors of i. Other more general examples arise if a neighbor of i is defined by a (known) linear combination of other units in the IVAR model. The latter allows for identification of neighbors via suitable measures of economic proximity such as trade patterns, or commuting distances. Assumption 6 allows for both possibilities.

Some of the results derived in Section 3 are unaffected by the introduction of neighbors in the analysis. For example, the relationship (58) describing the cross section averages continues to hold. However, large N representations for individual cross section units would change. In the case of the dominant unit, we have

$$\begin{aligned} x_{1t} &= \phi_{11}x_{1,t-1} + \boldsymbol{\phi}'_{-1}\mathbf{x}_{t-1} + u_{1t} \\ &= \phi_{11}x_{1,t-1} + \underbrace{\boldsymbol{\delta}'_1\mathfrak{n}_{1,t-1}}_{Neighbors} + \underbrace{\boldsymbol{\phi}'_{b1}\mathbf{x}_{t-1}}_{Non-neighbors} + \varepsilon_{1t} + O_p\left(N^{-1/2}\right), \end{aligned} \tag{89}$$

where we have used equation (86) of Assumption 6, and $u_{1t} = \varepsilon_{1t} + O_p\left(N^{-1/2}\right)$ follows from Assumption 2. Inequality (88) implies that each of the non-neighbor coefficients in the vector $\boldsymbol{\phi}_{bi}$ is $O\left(N^{-1}\right)$ and therefore we can use relationship (58) to obtain

$$\boldsymbol{\phi}'_{b1}\mathbf{x}_t = \left[\boldsymbol{\phi}'_{b1}\boldsymbol{\Phi}_{-1}(L)\,\boldsymbol{\phi}_1\right]x_{1,t-1} + \left[\boldsymbol{\phi}'_{b1}\boldsymbol{\Phi}_{-1}(L)\,\mathbf{r}_1\right]\varepsilon_{1t} + O_p\left(N^{-1/2}\right). \tag{90}$$

Note that the coefficients in the polynomials on the right side of (90) decay exponentially to zero, uniformly in N. In particular, the following bounds can be derived:

$$\left|\boldsymbol{\phi}'_{b1}\boldsymbol{\Phi}^{\ell}_{-1}\boldsymbol{\phi}_1\right| \leq \left\|\boldsymbol{\phi}'_{b1}\right\|_\infty \left\|\boldsymbol{\Phi}^{\ell}_{-1}\right\|_\infty \left\|\boldsymbol{\phi}_1\right\|_\infty \leq K\rho^\ell,$$

and similarly

$$\left|\boldsymbol{\phi}'_{b1}\boldsymbol{\Phi}^{\ell}_{-1}\mathbf{r}_1\right| \leq \left\|\boldsymbol{\phi}'_{b1}\right\|_\infty \left\|\boldsymbol{\Phi}^{\ell}_{-1}\right\|_\infty \left\|\mathbf{r}_1\right\|_\infty \leq K\rho^\ell,$$

where $\left\|\boldsymbol{\Phi}^{\ell}_{-1}\right\|_\infty \leq \left\|\boldsymbol{\Phi}_{-1}\right\|^{\ell}_\infty \leq \rho^\ell$ under Assumption 4, $\left\|\boldsymbol{\phi}'_{b1}\right\|_\infty \leq N\left\|\boldsymbol{\phi}_{b1}\right\|_\infty \leq K$ under Assumption 6, and $\left\|\boldsymbol{\phi}_1\right\|_\infty$ together with $\left\|\mathbf{r}_1\right\|_\infty$ are bounded in N under Assumptions 1.a and 2, respectively. The following large N representation for the dominant unit can now be obtained by substituting equation (90) in (89):

$$c^*(L)\,x_{1t} = \boldsymbol{\delta}'_1\mathfrak{n}_{1,t-1} + b^*_1(L)\,\varepsilon_{1t} + O_p\left(N^{-1/2}\right), \tag{91}$$

where $c^*(L) = 1 - \phi_{11}L - \left[\boldsymbol{\phi}'_{b1}\boldsymbol{\Phi}_{-1}(L)\,\boldsymbol{\phi}_1\right]L^2$, and $b^*_1(L) = 1 + \boldsymbol{\phi}'_{b1}\boldsymbol{\Phi}_{-1}(L)\,\mathbf{r}_1$. The invertibility of $b^*_1(L)$, is established in the following lemma.

Lemma 4 *Suppose Assumption 4 hold and $\left\|\boldsymbol{\phi}_{b1}\right\|_1 \leq \rho < 1$. Then, polynomial $b^*_1(L)$ is invertible for any $N \in \mathbb{N}$, with coefficients that decay exponentially to zero, uniformly in N.*

Multiplying equation (91) by the inverse of $b_1^*(L)$, we obtain the following representation:

$$a^*(L)\,x_{1t} + \boldsymbol{\psi}_1'(L)\,\mathfrak{n}_{1,t-1} = \varepsilon_{1t} + O_p\left(N^{-1/2}\right), \tag{92}$$

where

$$\boldsymbol{\psi}_1(L) = -\left[b_1^*(L)\right]^{-1}\boldsymbol{\delta}_1, \text{ and } a^*(L) = \left[b_1^*(L)\right]^{-1}c^*(L).$$

Polynomial $a^*(L)$ is the counterpart of the polynomial $a(L)$ in the model without neighbors. Therefore, for N sufficiently large the process of the dominant unit can be approximated by an infinite order autoregressive distributed lag model in x_{1t} and its neighbors, $\mathfrak{n}_{1,t-1}$, with coefficients that decay exponentially, and hence can be appropriately truncated, in the same way as in Section 5.

Large N representation for the remaining units can be derived similarly. Equation for the unit $i > 1$ in the VAR model (1) with neighbors is

$$\begin{aligned} x_{it} &= \phi_{ii}x_{i,t-1} + \phi_{i1}x_{1,t-1} + \boldsymbol{\phi}_{-1,-i}'\mathbf{x}_{t-1} + u_{it} \\ &= \phi_{ii}x_{i,t-1} + \phi_{i1}x_{1,t-1} + \underbrace{\boldsymbol{\delta}_i'\mathfrak{n}_{i,t-1}}_{Neighbors} + \underbrace{\boldsymbol{\phi}_{bi}'\mathbf{x}_{t-1}}_{Non-neighbors} + r_{i1}\varepsilon_{1t} + e_{it}, \text{ for } i = 2,3,...,N, \end{aligned} \tag{93}$$

where we have used equation (87) of Assumption 6. Assumption about the vector $\boldsymbol{\phi}_{bi}$ for $i = 2,3,...,N$, is the same as assumption about the vector $\boldsymbol{\phi}_{b1}$, which allows us to use equation (90) again, but with $\boldsymbol{\phi}_{b1}$ replaced by $\boldsymbol{\phi}_{bi}$, namely

$$\boldsymbol{\phi}_{bi}'\mathbf{x}_t = \boldsymbol{\phi}_{bi}'\boldsymbol{\Phi}_{-1}(L)\,\boldsymbol{\phi}_1 x_{1,t-1} + \boldsymbol{\phi}_{bi}'\boldsymbol{\Phi}_{-1}(L)\,\mathbf{r}_1\varepsilon_{1t} + O_p\left(N^{-1/2}\right). \tag{94}$$

Substituting this equation for $\boldsymbol{\phi}_{bi}'\mathbf{x}_{t-1}$ in (93) and using (92) to eliminate ε_{1t}, we obtain the following large N representation for the unit $i = 2,3,...,N$,

$$x_{it} = \phi_{ii}x_{i,t-1} + \boldsymbol{\delta}_i'\mathfrak{n}_{i,t-1} + \beta_i^*(L)\,x_{1t} + \boldsymbol{\psi}_i'(L)\,\mathfrak{n}_{1,t-1} + e_{it} + O_p\left(N^{-1/2}\right), \tag{95}$$

for $i = 2,3,...$, where

$$\beta_i^*(L) = a^*(L)\left[r_{i1} + L\boldsymbol{\phi}_{bi}'\boldsymbol{\Phi}_{-1}(L)\,\mathbf{r}_1\right] + \phi_{i1}L + \left[\boldsymbol{\phi}_{bi}'\boldsymbol{\Phi}_{-1}(L)\,\boldsymbol{\phi}_1\right]L^2,$$

which is the counterpart of the polynomial $\beta_i(L)$ in the model without neighbors, and

$$\boldsymbol{\psi}_i(L) = -\left[r_{i1} + L\boldsymbol{\phi}_{bi}'\boldsymbol{\Phi}_{-1}(L)\,\mathbf{r}_1\right]\left[b_1^*(L)\right]^{-1}\boldsymbol{\delta}_1.$$

Note that in the case where the dominant unit does not have any neighbors, $\beta_i^*(L)$ reduces to $\beta_i(L)$. It is also worth noting that lagged values of the dominant unit's neighbors enter the large N representation for the non-dominant units. The following proposition summarizes the main results derived above.

Proposition 2 *(Large N representations in the presence of neighbors) Let Assumptions 1.a, 2-4,* *and 6 hold, and $\|\boldsymbol{\phi}_{b1}\|_1 \leq \rho < 1$. Then, the dominant unit has an infinite order distributed lagged representation in terms of its own lagged values as well as the lagged values of its neighbors, as given by (92). Large N representation for the non-dominant units $i > 1$ are given by (95), and in addition to its own lagged and own neighborhood effects, x_{it} also depends on infinite distributed lagged functions of the dominant unit and the dominant unit's neighbors. Only the contemporaneous value of the dominant unit is included in the large N representation of the non-dominant units.*

Asymptotic distribution for the estimation based on the corresponding ALS regressions in the presence of neighbors can be established in the same way as in Section 5 and will not be pursued here. We provide below some Monte Carlo evidence in case of neighbors present in the data generating process.

7 Allowing for Unobserved Common Factors

Unobserved common factors can be introduced into the IVAR model in a number of different ways. One possibility is to include unobserved common factors in the errors:

$$\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t + \boldsymbol{\Gamma}\mathbf{f}_t, \tag{96}$$

where $\boldsymbol{\Gamma}$ is an $N \times m_f$ matrix of factor loadings, and $\mathbf{f}_t$ is an $m_f \times 1$ vector of unobserved common factors, assumed to be covariance stationary with zero means and unit variances.[7] The number of factors, m_f, is finite and does not change with N. An alternative option is to introduce the unobserved common factors directly in the IVAR model, as in

$$\mathbf{x}_t - \boldsymbol{\Gamma}\mathbf{f}_t = \boldsymbol{\Phi}\left(\mathbf{x}_{t-1} - \boldsymbol{\Gamma}\mathbf{f}_{t-1}\right) + \mathbf{R}\boldsymbol{\varepsilon}_t. \tag{97}$$

Both these specifications lead to the same ALS regressions, namely that unit specific regressions must be augmented by a sufficient number of cross section averages and their lags in order to proxy for the effects of the unobserved common factors. But specification (97) is analytically simpler to work with and will be adopted in what follows. Let $\mathbf{z}_t = \mathbf{x}_t - \boldsymbol{\Gamma}\mathbf{f}_t$, and apply the large N representation results obtained so far to $\mathbf{z}_t = \boldsymbol{\Phi}\mathbf{z}_{t-1} + \mathbf{R}\boldsymbol{\varepsilon}_t$. In particular, abstracting from neighborhood effects, and using results (59) and (60) but applied to $\mathbf{z}_t$, we have

$$\bar{\mathbf{z}}_{Wt} = \boldsymbol{\varphi}_W\left(L\right) z_{1t} + O_p\left(N^{-1/2}\right), \tag{98}$$

where $\bar{\mathbf{z}}_{Wt} = \mathbf{W}'\mathbf{z}_t$ is an $m_w \times 1$ vector of cross section averages, obtained using the $N \times m_w$ matrix of granular weights $\mathbf{W}$, and

$$\boldsymbol{\varphi}_W\left(L\right) = \mathbf{W}'\boldsymbol{\Phi}_{-1}\left(L\right)\mathbf{r}_1 a\left(L\right) + \mathbf{W}'\boldsymbol{\Phi}_{-1}\left(L\right)\boldsymbol{\phi}_1 L.$$

[7] For the derivation of the asymptotic distribution of the corresponding ALS estimators, fourth moments would need to be bounded as well.

Equation (98) implies

$$\mathbf{A}(L)\begin{pmatrix} x_{1t} \\ \overline{\mathbf{x}}_{Wt} \end{pmatrix} = \left[\overline{\mathbf{\Gamma}}_W - \boldsymbol{\varphi}_W(L)\boldsymbol{\gamma}'_{1\circ}L\right]\mathbf{f}_t + O_p\left(N^{-1/2}\right),$$

where $\boldsymbol{\gamma}_{1\circ} = \left(\gamma_{11}, \gamma_{12}, ..., \gamma_{1m_f}\right)'$ is the first row of $\mathbf{\Gamma}$, and

$$\mathbf{A}(L) = \begin{pmatrix} \boldsymbol{\varphi}_W(L) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_w} \end{pmatrix}.$$

In cases where $\left[\overline{\mathbf{\Gamma}}_W - \boldsymbol{\varphi}_W(L)\boldsymbol{\gamma}'_{1\circ}L\right]$ is invertible, the unobserved common factors can be approximated arbitrarily well by the values of the dominant unit, cross section averages, $\overline{\mathbf{z}}_{Wt}$, and their lags,

$$\mathbf{f}_t = \mathbf{B}(L)\begin{pmatrix} x_{1t} \\ \overline{\mathbf{x}}_{Wt} \end{pmatrix} + O_p\left(N^{-1/2}\right),$$

where

$$\mathbf{B}(L) = \left[\overline{\mathbf{\Gamma}}_W - \boldsymbol{\varphi}_W(L)\boldsymbol{\gamma}'_{1\circ}L\right]^{-1}\mathbf{A}(L).$$

A similar relationship can also be derived in the presence of neighborhood effects and unobserved common factors. Derivations are available from the authors on request.

8 Monte Carlo Experiments

In this section we report some evidence on the small sample properties of the augmented least squares estimator $\widehat{\boldsymbol{\pi}}_i$. The data generating process (DGP) is given by the following stationary IVAR featuring the dominant unit and augmented by an unobserved common factor.

$$(\mathbf{x}_t - \boldsymbol{\gamma} f_t) = \mathbf{\Phi}(\mathbf{x}_{t-1} - \boldsymbol{\gamma} f_{t-1}) + \mathbf{u}_t, \tag{99}$$

where

$$\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t = \mathbf{r}_1\varepsilon_{1t} + \mathbf{e}_t, \tag{100}$$

which corresponds to model (1) augmented by one unobserved common factor f_t and residuals correspond to (8) and (20). Our focus is on estimation of the lagged own coefficient in equation for the non-dominant unit $i = 2$, namely ϕ_{22}, the lagged neighbor coefficient, ϕ_{23}, and $\beta_{20} = \beta^*_{20} = r_{21}$ in (61), when $\boldsymbol{\gamma} = \mathbf{0}$.[8] Corresponding ALS estimators for these coefficients are denoted by $\widehat{\phi}_{22}$, $\widehat{\phi}_{23}$, and $\widehat{\beta}_{20}$, respectively.

The elements of $\mathbf{\Phi}$ are generated so that unit 1 is dominant, and there are non-zero neighborhood

[8] Similar results are also obtained for other cross section units.

effects. To this end we first generate

$$\omega_{ij} = \begin{cases} \dfrac{\varsigma_{ij}}{\sum_{j \notin \{1,i,i+1\}} \varsigma_{ij}}, \text{ for } j \notin \{1,i,i+1\} \\ 0, \text{ for } j \in \{1,i,i+1\} \end{cases},$$

with $\varsigma_{ij} \sim IIDU(0,1)$. This ensures that $\omega_{ij} = O_p(N^{-1})$, and $\sum_{j=1}^{N} \omega_{ij} = 1$. Individual elements of $\mathbf{\Phi}$ are then generated as follows:

1. (Dominant Unit $i = 1$) $\phi_{11} = 0.7$, and $\phi_{1j} = \lambda_1 \omega_{1j}$ for $j = 2, 3, ..., N$, with $\lambda_1 = 0.1$.
2. (Unit $i = 2$) $\phi_{21} = 0.1$, $\phi_{22} = 0.5$, $\phi_{23} = 0.1$, and $\phi_{2j} = \lambda_2 \omega_{2j}$ for $j = 3, 4, ..., N$, with $\lambda_2 = 0.1$.
3. (Remaining units $i > 2$) $\phi_{ii} \sim IIDU(0.3, 0.5)$, $\phi_{i1} \sim IIDU(0, 0.1)$, $\phi_{i,i+1} \sim IIDU(-0.2, 0.2)$, and $\phi_{ij} = \lambda_i \omega_{ij}$ for $j \notin \{1, i, i+1\}$, where $\lambda_i \sim IIDU(-0.05, 0.15)$.

The focus parameters of the dominant unit 1, and unit $i = 2$ are fixed across all experiments. The remaining parameters are generated randomly. In all experiments $\mathbf{\Phi}$ is generated such that $\|\mathbf{\Phi}\|_\infty \leq 0.95$, which is a sufficient condition for stationarity of the IVAR model.

Two sets of factor loadings are considered, $\boldsymbol{\gamma} = \mathbf{0}$ (no unobserved common factor) and $\boldsymbol{\gamma} \neq \mathbf{0}$. Under the latter we set $\gamma_1 = 1$, $\gamma_2 = -0.5$, and the remaining factor loadings are generated randomly as $\gamma_i \sim 0.5\phi_{ii} + IIDN(1,1)$ for $i = 3, 4, ..., N$. The factor loadings are generated to depend on ϕ_{ii}, so that the robustness of the ALS estimator to this type of dependency can be evaluated. The common factor f_t is generated as

$$f_t = \rho_f f_{t-1} + \varepsilon_{ft},$$

where $\varepsilon_{ft} \sim IIDN\left(0, 1 - \rho_f^2\right)$, which yields $Var(f_t) = 1$. We choose relatively persistent common factor with $\rho_f = 0.9$. We set $e_{1t} = 0$ and generate the remaining error terms $\{e_{2t}, e_{3t}, ..., e_{Nt}\}$ from a stationary spatial process in order to show that our estimators are invariant to the weak cross section dependence of innovations. The following bilateral Spatial Autoregressive Model (SAR) is considered.

$$e_{it} = \frac{a_e}{2}\left(e_{i-1,t} + e_{i+1,t}\right) + \eta_{eit}, \tag{101}$$

where $\eta_{eit} \sim IIDN\left(0, \sigma_{\eta e}^2\right)$. As established by Whittle (1954), the unilateral SAR(2) scheme

$$e_{it} = \delta_{e1} e_{i-1,t} + \delta_{e2} e_{i-2,t} + \eta_{eit}, \tag{102}$$

with $\delta_{e1} = \alpha_e + \beta_e$, $\delta_{e2} = -\alpha_e \beta_e$, $\alpha_e = \left(1 - \sqrt{1 - a_e^2}\right)/a_e$, and $\beta_e^{-1} = \left(1 + \sqrt{1 - a_e^2}\right)/a_e$, generates the same autocorrelations as the bilateral SAR(1) scheme (101). The error terms are generated using the unilateral scheme (102) with 50 burn-in data points ($i = -49, -48, ..., 0$), and the initializations $e_{-51} = e_{-50} = 0$. The spatial AR parameter, a_e, is set to 0.4, which ensures that the process $\{e_{it}\}$ is cross sectionally weakly dependent. $\sigma_{\eta e}^2 = Var(\eta_{eit})$ is chosen so that the variance of errors

e_{it} is equal to 0.1.[9] $\varepsilon_{1t} \sim IIDN(0, 0.15)$ and $r_{11} = 1$, which implies that $Var(u_{1t}) = 0.15$. The second element of $\mathbf{r}_1$ in (100) is set to $r_{21} = 0.1$ and the remaining elements are generated as $r_{i1} \sim IIDU(0, 0.2)$ for $i = 3, 4, ..., N$.

We consider three different types of augmentation. In addition to the lagged neighbor unit 3, the regression for unit $i = 2$ is augmented by the following set of regressors: (i) the current and lagged values of the dominant unit, $\{x_{1,t-\ell}\}_{\ell=0}^{m_T}$, (ii) the simple cross section averages $\{\overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$, and (iii) $\{x_{1,t-\ell}, \overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$. In all three cases m_T is set to the integer value of $T^{1/3}$, which we denote by $[T^{1/3}]$.[10] For example, under case (i) the ALS regression for unit $i = 2$ is specified as:

$$x_{2t} = c_2 + \phi_{22}x_{2,t-1} + \phi_{23}x_{3,t-1} + \sum_{\ell=0}^{[T^{1/3}]} b_{1\ell}x_{1,t-\ell} + \epsilon_{2t}. \quad (103)$$

8.1 Monte Carlo results

We report results for experiments without the unobserved common factor first. Table 1 summarizes the results for the own coefficient $\hat{\phi}_{22}$, and Table 2 summarizes the results for the neighbor coefficient, ϕ_{23}. Each table gives the bias and the root mean squared error (RMSE) of the estimator as well as the empirical size and power of tests based on it. The results for $\hat{\phi}_{23}$ are a little better but overall similar to those for $\hat{\phi}_{22}$. The bias and RMSE of these estimators decline as N and T are increased irrespective of the augmentation procedure adopted. This is because in the absence of a common factor the dominant unit and the cross section averages are asymptotically equivalent and either set of variables (with long enough lags) are sufficient to deal with the cross section dependence and the omitted variable problems in the IVAR model. The augmentation by cross section averages has the advantage that it works regardless of whether strong cross section dependence is due to a dominant unit, or due to a different source such as an unobserved common factor. Full augmentation by the dominant unit as well as the cross section averages is not necessary in the absence of a common factor, and yields worse outcomes in terms of RMSEs. See the third panel of Tables 1 and 2.

The empirical size of the tests for values of $T > 50$ are also close to the 5 percent nominal level. For smaller values of T, however, there is a negative bias and the tests are oversized. This is the familiar time series bias where even in the absence of cross section dependence the LS estimators of autoregressive coefficients are biased in small T samples. But the size of the tests does not change much with N, which is in the line with the findings reported in CP. Overall, these findings suggest that N need not to be very large for the ALS estimator to work.

Results for $\widehat{\beta}_{20}$ are reported in Table 3. The top panel summarizes the results when the regression is augmented with $\{x_{1,t-\ell}\}_{\ell=0}^{m_T}$, as suggested by the theory. In this case the bias and RMSE of $\widehat{\beta}_{20}$ declines with N and T, and the empirical size is close to the nominal value of the test, very much in line with the results reported for $\hat{\phi}_{22}$ and $\hat{\phi}_{23}$. In contrast, the estimates at the bottom panel of Table 3 that are based on regressions augmented by $\{x_{1,t-\ell}, \overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$, behave less well and

[9] The variance of errors $\{e_{it}\}$ is given by $\sigma^2 = (1 + \delta_{e2}) \left[(1 - \delta_{e2}^2) - \delta_{e1}^2 \right] / (1 - \delta_{e2})$.

[10] $m_T = 2, 3, 4, 4, 5$ for $T = 25, 50, 75, 100, 200$, respectively.

for a given T the RMSEs deteriorate as N increases. The inclusion of cross section averages lead to a multicollinearity problem since $\{x_{1,t-\ell}\}_{\ell=0}^{m_T}$ and $\{\overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$ will be asymptotically equivalent. But this asymptotic multicollinearity problem does not affect the estimation of ϕ_{22} and ϕ_{23}.

Results for the experiments with the unobserved common factor are reported in Table 4 (own coefficient ϕ_{22}) and Table 5 (neighbor coefficient ϕ_{23}).[11] Theory suggests that augmentation by the dominant unit or by the cross section averages alone is not enough for consistent estimation in the presence of a dominant unit as well as a common factor, f_t. This is confirmed by the MC results in Tables 4 and 5, which indeed show substantial biases and significant size distortions in cases without the full augmentation (the empirical sizes are in the range $17\% - 70\%$ for $N = T = 200$). The ALS estimator based on the full augmentation is correctly sized for larger values of N and T and overall its performance is very similar to the experiments without the unobserved common factor.

9 Concluding Remarks

This paper has extended the analysis of infinite dimensional vector autoregressive (IVAR) models by Chudik and Pesaran (2011) to the case where one variable or a cross section unit is dominant in the sense that it has non-negligible contemporaneous and/or lagged effects on all other units as the cross section dimension rises without a bound. We showed that the asymptotic normality of the augmented least squares (ALS) estimator continues to hold once the individual auxiliary regressions are correctly specified. Satisfactory finite sample performance was documented by means of Monte Carlo experiments.

A number of applications of the IVAR model with a dominant unit have already been attempted in the literature. Holly, Pesaran, and Yamagata (2011) examine the diffusion of house prices across different regions in the UK and consider the possibility that London plays a dominant role in this process. Bussiere, Chudik, and Mehl (2011) investigate the functioning of the foreign exchange markets treating the US dollar as a dominant currency, and Chudik and Fratzscher (2011) present a study of global equity markets where US equity and money markets are treated as dominant.

The paper provides a general framework for dealing with the curse of dimensionality in large linear stationary dynamic models, when the dominant unit and the neighborhood patterns are given. Further work is clearly needed on identification of the dominant unit(s), patterns of spatial or network dependencies, and the role of unobserved common factors. These topics together with the extension of the analysis to nonstationary IVAR models must be left to future studies.

[11] Results for $\widehat{\beta}_{20}$ are not reported in this case since only in the absence of common factor, coefficient β_{20} corresponding to the contemporaneous value of the dominant unit equals r_{21}, as shown in equation (61).

Table 1: MC results for the own coefficient ϕ_{22} in experiments with zero factor loadings

	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0: \phi_{22} = 0.50$)					Power (5% level, $H_1: \phi_{22} = 0.70$)				
(N,T)	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	Augmentation by $\{x_{1,t-\ell}\}_{\ell=0}^{m_T}$ with $m_T = \left[T^{1/3}\right]$.																			
25	-18.17	-8.85	-6.09	-3.88	-1.76	30.71	17.59	13.46	10.87	7.14	11.90	8.05	6.25	5.80	5.60	33.30	45.35	57.15	65.95	91.25
50	-18.11	-9.01	-5.95	-4.36	-2.28	30.23	17.84	13.45	10.97	7.35	11.85	8.15	7.00	6.70	5.85	32.35	45.55	57.55	67.90	91.60
75	-16.30	-9.14	-5.07	-4.47	-1.74	28.62	18.49	13.17	11.35	6.86	8.80	9.45	6.55	7.45	4.85	31.15	44.20	52.60	67.45	91.45
100	-18.05	-7.84	-5.47	-4.58	-1.88	30.94	17.23	13.36	11.39	7.11	12.40	7.20	7.05	6.95	5.25	32.50	41.15	55.15	67.05	92.55
200	-17.94	-8.47	-5.17	-3.96	-2.21	30.20	17.56	13.49	10.78	7.27	11.75	8.35	7.65	6.65	6.20	33.75	43.45	55.55	66.15	91.35
	Augmentation by $\{\overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$ with $m_T = \left[T^{1/3}\right]$.																			
25	-17.57	-9.17	-6.51	-4.43	-2.42	29.63	17.60	13.64	10.99	7.32	10.95	8.15	7.40	7.55	6.30	33.05	47.35	60.25	68.85	92.80
50	-17.67	-8.96	-6.00	-4.46	-2.60	29.78	17.83	13.35	11.00	7.42	11.25	9.25	7.65	7.25	6.25	32.40	45.10	57.70	69.00	92.55
75	-16.13	-9.07	-5.21	-4.73	-1.97	28.55	18.13	13.14	11.37	6.88	9.40	8.80	6.55	7.20	5.05	29.75	44.60	53.90	68.65	92.05
100	-17.69	-8.03	-5.45	-4.67	-2.01	30.42	17.37	13.30	11.36	7.08	11.30	7.55	7.20	7.05	5.45	32.60	42.20	55.20	68.60	92.60
200	-17.69	-8.34	-5.19	-3.99	-2.28	29.82	17.41	13.45	10.81	7.27	11.00	8.30	7.35	6.45	6.20	32.85	43.45	54.95	65.55	91.60
	Augmentation by $\{x_{1,t-\ell}, \overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$ with $m_T = \left[T^{1/3}\right]$.																			
25	-21.78	-11.20	-7.70	-5.22	-2.64	34.75	19.63	14.80	11.67	7.51	13.90	9.95	8.20	7.65	6.75	32.95	47.80	59.40	68.50	92.60
50	-22.12	-10.93	-7.03	-5.27	-2.92	34.71	19.78	14.38	11.78	7.69	13.10	9.85	8.35	7.35	6.25	33.85	45.75	55.90	68.75	91.75
75	-20.10	-11.08	-6.42	-5.32	-2.29	32.92	20.26	14.23	11.94	7.17	11.20	10.45	7.25	7.45	5.30	32.00	45.25	53.55	68.45	92.00
100	-22.08	-9.92	-6.57	-5.36	-2.26	35.68	19.39	14.38	12.08	7.31	14.30	9.65	8.40	7.80	5.60	32.85	42.55	55.45	67.25	92.15
200	-21.29	-10.24	-6.44	-4.72	-2.61	34.30	19.37	14.70	11.43	7.54	13.05	9.20	8.45	7.05	6.80	33.00	43.85	55.20	65.75	91.85

Notes: $\phi_{22} = 0.5$, $\phi_{23} = 0.1$, and $\left[T^{1/3}\right]$ refers to the integer part of $T^{1/3}$, so that $m_T = 2, 3, 4, 4, 5$ for $T = 25, 50, 75, 100, 200$, respectively. The DGP is defined by (99) where the equation for unit $i \in \{2, 3, ..., N-1\}$ is $x_{it} - \gamma_i f_t = \sum_{j=1}^{N} \phi_{ij} \left(x_{j,t-1} - \gamma_j f_{t-1}\right) + r_{i1}\varepsilon_{1t} + e_{it}$, unit 1 is dominant, the neighbor of unit i are units $\{1, i, i+1\}$ for $i \in \{2, 3, ..., N-1\}$, and innovations $\{e_{it}\}$ are generated from spatial autoregressive model (102) and $e_{1t} = 0$. The ALS estimator of the coefficient ϕ_{22} is computed using the auxiliary regression, $x_{2t} = c_2 + \phi_{22}x_{2,t-1} + \phi_{23}x_{3,t-1} + \sum_{\ell=0}^{m_T} \mathbf{a}_{2\ell}\boldsymbol{\delta}_{t-\ell} + \epsilon_{2t}$, where three different augmentation schemes are considered for the vector $\boldsymbol{\delta}_t$: cross section averages, dominant unit $i = 1$, or both. See Section 8 for a detailed description of the Monte Carlo design.

Table 2: MC results for the neighbor coefficient ϕ_{23} in experiments with zero factor loadings

	Bias ($\times$100)					**Root Mean Square Errors** ($\times$100)					**Size** (**5% level,** $H_0 : \phi_{23} = 0.10$)					**Power** (**5% level,** $H_1 : \phi_{23} = 0.30$)				
(N,T)	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	Augmentation by $\{x_{1,t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	1.27	0.52	1.07	0.47	0.09	27.50	17.28	13.24	11.37	7.12	7.65	6.75	5.30	6.45	4.15	14.50	24.55	33.05	44.85	77.00
50	1.76	0.34	0.81	0.84	0.49	27.92	17.43	13.28	11.17	7.37	7.60	6.40	6.00	5.60	4.95	14.70	25.85	33.45	43.15	75.50
75	0.38	0.38	0.56	0.29	0.16	28.13	17.13	13.25	11.19	7.59	8.10	6.80	6.05	5.85	5.50	16.95	25.25	35.70	44.90	74.70
100	0.33	0.38	0.59	-0.09	0.26	27.57	16.75	13.21	11.22	7.38	7.95	6.80	5.25	5.20	4.50	15.65	24.55	34.70	47.15	75.50
200	1.66	1.05	0.47	0.19	0.00	27.74	17.09	13.63	11.12	7.52	7.80	7.05	6.30	5.95	5.35	15.40	22.70	34.05	46.60	77.15
	Augmentation by $\{\overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	-0.01	-0.59	0.05	-0.66	-1.02	27.98	17.33	13.41	11.28	7.16	8.00	6.90	5.80	6.85	4.30	16.10	27.45	35.05	47.70	81.10
50	1.23	-0.13	0.31	0.18	0.04	27.74	17.47	13.26	11.05	7.41	7.55	6.40	6.00	4.75	4.75	15.15	26.65	34.50	45.65	77.30
75	0.20	-0.10	0.30	0.04	-0.16	27.98	17.27	13.19	11.22	7.56	8.50	7.05	5.90	6.00	4.80	16.85	26.35	36.40	46.50	76.30
100	0.47	0.20	0.36	-0.28	0.05	27.54	16.64	13.18	11.21	7.34	7.60	6.80	4.75	5.25	4.30	15.05	24.30	35.25	48.40	76.50
200	1.71	0.89	0.40	0.07	-0.05	27.82	17.18	13.64	11.16	7.50	7.80	7.45	6.70	6.40	5.15	14.80	23.45	34.65	46.95	77.25
	Augmentation by $\{x_{1,t-\ell}, \overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	-0.19	-0.48	0.16	-0.48	-0.92	31.25	18.44	14.04	11.60	7.31	8.85	7.40	5.65	6.55	4.00	15.70	24.05	33.25	46.60	79.30
50	1.06	0.14	0.41	0.25	0.05	31.18	18.47	13.90	11.48	7.55	8.50	6.50	6.10	5.20	5.10	14.75	23.90	32.80	42.70	76.35
75	0.08	-0.13	0.28	-0.02	-0.15	31.40	18.43	13.82	11.57	7.68	8.95	6.95	6.35	6.15	4.90	15.65	24.45	34.15	44.15	74.15
100	0.44	0.28	0.51	-0.22	0.00	30.55	17.88	13.72	11.63	7.52	8.50	6.60	5.20	6.00	4.55	14.30	23.50	33.25	45.05	75.55
200	1.38	0.92	0.47	0.02	-0.01	30.63	18.30	14.38	11.41	7.62	7.60	7.95	7.10	5.55	4.80	14.90	22.05	32.15	44.05	75.35

See the notes to Table 1.

Table 3: MC results for the coefficient β_{20} in experiments with zero factor loadings

	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0 : \beta_{20} = 0.10$)					Power (5% level, $H_1 : \beta_{20} = 0.30$)				
(N,T)	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	Augmentation by $\{x_{1,t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	0.15	0.71	-0.07	-0.12	0.19	21.13	13.27	10.51	8.80	5.96	8.15	5.75	6.00	5.05	4.85	20.50	34.65	50.75	63.75	91.00
50	0.51	0.06	0.09	-0.04	-0.03	20.75	13.08	10.48	8.59	6.09	7.25	5.10	5.90	4.65	4.70	21.35	35.05	50.20	64.60	90.90
75	-0.87	0.10	0.08	0.07	0.11	21.39	13.39	10.21	8.50	5.90	7.90	6.20	4.70	4.50	4.95	22.65	36.65	49.95	63.75	91.60
100	-0.03	-0.14	-0.05	0.01	0.08	20.99	13.59	10.46	8.84	5.97	7.45	6.35	6.05	5.65	5.10	20.25	35.50	50.05	62.00	91.70
200	-0.89	-0.23	0.05	0.01	-0.09	21.31	13.02	10.49	8.79	5.96	7.75	5.70	4.90	5.25	5.45	22.55	35.90	48.10	64.35	91.85
	Augmentation by $\{x_{1,t-\ell}, \overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	-1.15	-0.22	0.10	0.21	0.59	32.17	20.63	16.31	13.81	9.83	6.45	5.20	4.65	4.60	5.20	12.70	17.80	22.90	30.20	51.60
50	-2.23	-1.13	0.12	-0.60	-0.03	38.27	24.19	19.12	16.12	11.25	6.15	4.90	5.05	4.60	5.20	12.20	15.50	18.50	24.40	43.25
75	-1.98	-0.72	-0.72	-0.37	-0.29	42.17	26.33	21.34	18.15	12.32	6.10	4.75	5.35	5.50	4.35	9.55	12.75	16.50	20.95	37.25
100	-0.96	-2.06	-0.55	-0.36	-0.54	46.54	30.22	23.68	19.52	13.62	7.00	5.45	5.25	4.80	5.00	8.80	12.25	15.40	17.30	32.30
200	-1.95	-0.13	-0.50	-0.82	-0.44	61.35	37.56	31.03	24.94	17.55	7.30	4.75	4.85	4.70	5.60	9.00	8.30	10.85	13.25	22.40

See the notes to Table 1.

Table 4: MC results for the own coefficient ϕ_{22} in experiments with nonzero factor loadings

	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0 : \phi_{22} = 0.50$)					Power (5% level, $H_1 : \phi_{22} = 0.70$)				
(N,T)	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	Augmentation by $\{x_{1,t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	-9.75	2.08	5.91	7.93	10.93	25.59	15.16	12.66	12.49	13.01	9.65	9.85	12.90	21.70	49.55	23.45	22.40	24.00	26.40	32.90
50	-10.19	1.38	5.60	7.80	10.71	25.75	14.84	13.18	12.55	12.66	8.20	8.65	15.65	20.95	47.25	24.55	23.60	24.90	25.05	33.60
75	-8.68	2.28	5.98	7.95	10.62	24.87	14.53	13.03	12.47	12.60	7.55	8.70	15.10	20.50	47.50	22.10	22.50	22.80	25.80	33.50
100	-9.73	2.09	5.63	8.33	11.03	25.76	15.24	13.02	12.70	12.95	9.65	9.95	14.45	22.55	48.85	24.25	23.30	24.80	23.75	31.60
200	-9.53	1.85	6.11	7.83	10.49	25.15	14.90	13.40	12.56	12.67	8.40	9.10	15.85	21.05	46.80	23.20	22.10	22.65	25.20	35.60
	Augmentation by $\{\overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	-10.93	0.18	3.91	5.55	8.55	26.48	15.28	12.33	11.38	11.00	8.85	8.20	10.70	14.10	34.45	25.20	24.90	27.55	31.10	42.50
50	-12.38	-1.19	2.53	4.69	7.41	27.20	15.35	12.71	11.20	10.08	9.20	7.50	10.75	12.60	27.10	25.80	28.65	31.90	34.20	49.75
75	-11.51	-0.76	2.58	4.55	7.20	26.27	15.07	12.29	10.97	9.96	7.65	7.40	8.85	11.95	25.60	24.25	27.00	30.80	35.20	50.10
100	-12.76	-1.50	2.06	4.63	7.33	27.65	15.66	12.15	10.91	9.96	9.20	7.75	8.60	11.85	25.85	27.05	29.25	32.05	34.65	49.60
200	-12.03	-1.65	2.43	4.25	6.69	27.40	15.63	12.39	10.88	9.77	9.35	7.25	8.15	11.65	24.10	24.85	29.50	32.00	36.35	52.80
	Augmentation by $\{x_{1,t-\ell}, \overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	-20.48	-7.63	-3.28	-1.31	1.83	33.61	18.14	12.62	10.40	7.32	11.40	8.15	6.05	6.05	7.90	32.60	39.25	46.20	55.60	77.60
50	-24.17	-10.79	-6.20	-4.10	-0.73	36.52	19.58	14.30	11.62	6.98	15.45	9.25	8.25	7.90	5.05	37.15	45.80	54.70	64.00	87.75
75	-23.39	-10.67	-6.59	-4.39	-1.25	36.05	19.50	14.18	11.35	6.98	14.00	9.90	7.45	7.00	4.95	35.60	45.85	54.25	66.85	89.25
100	-24.93	-11.49	-7.54	-4.35	-1.32	37.29	20.19	14.91	11.22	6.88	16.05	10.60	9.00	6.55	4.75	38.30	46.45	58.15	66.15	89.95
200	-24.48	-12.35	-7.25	-5.56	-2.51	36.85	20.66	14.94	12.04	7.68	15.00	11.15	9.30	7.70	7.05	34.90	48.35	56.25	68.25	90.90

See the notes to Table 1.

Table 5: MC results for the neighbor coefficient ϕ_{23} in experiments with nonzero factor loadings.

	Bias ($\times$100)					**Root Mean Square Errors** ($\times$100)					**Size (5% level,** $H_0: \phi_{23} = 0.10$)					**Power (5% level,** $H_1: \phi_{23} = 0.30$)				
(N,T)	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
	Augmentation by $\{x_{1,t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	-9.68	-12.16	-12.62	-13.06	-13.40	26.50	20.15	18.20	17.58	16.66	17.00	30.75	44.45	52.85	72.00	41.60	70.10	79.25	85.45	92.55
50	-10.43	-11.95	-12.89	-13.44	-13.84	26.43	20.08	18.29	17.77	16.77	17.90	31.70	43.50	54.25	72.95	43.80	69.15	80.65	86.20	93.60
75	-10.15	-12.68	-12.94	-13.55	-13.73	26.26	19.86	18.01	17.75	16.70	15.85	31.30	43.60	55.10	71.55	42.45	71.25	81.60	86.35	93.85
100	-9.75	-12.21	-12.78	-13.15	-13.74	26.50	19.93	18.36	17.33	16.75	16.00	31.05	43.25	53.60	71.40	40.95	69.30	80.90	86.35	93.50
200	-10.92	-12.62	-12.76	-13.16	-13.42	26.41	20.33	18.24	17.69	16.55	17.65	32.00	42.70	54.10	69.90	42.95	70.45	80.40	85.95	93.65
	Augmentation by $\{\overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	-3.81	-5.37	-5.73	-5.72	-6.00	29.47	19.30	16.12	14.15	11.86	10.15	10.60	12.40	15.15	25.75	21.25	37.60	51.85	62.75	84.05
50	-2.76	-3.89	-4.44	-4.62	-4.98	28.91	19.21	15.64	13.58	10.57	9.95	10.05	11.50	13.20	18.75	19.90	36.20	48.10	60.30	85.45
75	-2.10	-3.31	-4.21	-4.58	-4.78	28.08	17.96	14.98	13.41	10.55	7.95	8.15	10.45	13.05	19.75	17.10	33.40	47.75	60.00	84.35
100	-0.65	-3.02	-3.64	-3.96	-4.52	28.78	18.31	15.17	13.14	10.45	9.55	8.75	10.80	10.95	19.35	18.55	34.00	46.30	58.70	84.00
200	-2.69	-3.11	-3.54	-3.74	-3.90	29.15	18.36	14.95	12.97	10.17	9.30	9.10	9.95	10.90	17.50	18.75	33.40	46.20	57.50	82.30
	Augmentation by $\{x_{1,t-\ell}, \overline{x}_{t-\ell}\}_{\ell=0}^{m_T}$ **with** $m_T = \left[T^{1/3}\right]$.																			
25	-0.95	-2.09	-2.25	-2.33	-2.48	31.96	19.45	14.80	12.49	8.98	9.35	8.00	7.75	8.05	9.55	16.55	28.30	39.10	50.65	79.20
50	0.48	-0.35	-0.59	-0.87	-1.03	31.11	19.26	14.45	11.79	7.84	8.95	7.50	6.65	6.65	5.80	15.75	26.40	35.80	48.05	77.80
75	0.58	0.23	-0.32	-0.76	-0.87	30.96	17.82	13.84	11.54	7.68	8.70	6.05	5.75	5.70	5.30	14.35	22.40	33.75	46.15	78.75
100	2.52	0.46	0.27	0.00	-0.54	31.41	18.22	14.02	11.79	7.67	8.90	7.20	6.30	6.40	5.50	14.05	22.25	31.70	43.70	77.10
200	0.57	0.56	0.17	0.45	0.07	30.80	18.75	13.71	11.74	7.44	7.80	7.05	5.90	5.95	4.30	13.30	23.50	32.20	42.25	75.20

See the notes to Table 1.

A Supplementary Lemmas

Lemma A.1 *Let* $\psi(L) = \sum_{\ell=0}^{\infty} \psi_\ell L^\ell$, $\psi_0 = 1$ *and there exists a real positive constant* $0 < \rho < 1$ *such that* $|\psi_\ell| \leq \rho^\ell$ *for any* $\ell \in \mathbb{N}$. *Then there exists polynomial* $\theta(L) = \sum_{\ell=0}^{\infty} \theta_\ell L^\ell$ *such that* $\psi(L)\theta(L) = 1$,

$$|\theta_\ell| \leq \left(1 + \frac{\ell(\ell-1)}{2}\right)\rho^\ell \text{ for any } \ell \in \mathbb{N}, \tag{A.1}$$

and there also exist real constants $K < \infty$, *and* $0 < \rho_1 < 1$ *such that*

$$|\theta_\ell| \leq K\rho_1^\ell \text{ for any } \ell \in \mathbb{N}. \tag{A.2}$$

Proof. We have

$$\begin{aligned}
\theta_0 &= 1, \\
\theta_1 &= -\psi_1, \\
\theta_2 &= -\psi_1\theta_1 - \psi_2, \\
\theta_3 &= -\psi_1\theta_2 - \psi_2\theta_1 - \psi_3, \\
\theta_4 &= -\psi_1\theta_3 - \psi_2\theta_2 - \psi_3\theta_1 - \psi_4.
\end{aligned}$$

Note that

$$\begin{aligned}
|\theta_1| &= |\psi_1|, \\
|\theta_2| &\leq |\psi_1||\theta_1| + |\psi_2|, \\
|\theta_3| &\leq |\psi_1||\theta_2| + |\psi_2||\theta_1| + |\psi_3|, \\
|\theta_4| &\leq |\psi_1||\theta_3| + |\psi_2||\theta_2| + |\psi_3||\theta_1| + |\psi_4|,
\end{aligned}$$

and by recursive substitution

$$\begin{aligned}
|\theta_1| &= |\psi_1|, \\
|\theta_2| &\leq |\psi_1||\theta_1| + |\psi_2| = |\psi_1|^2 + |\psi_2|, \\
|\theta_3| &\leq |\psi_1||\theta_2| + |\psi_2||\theta_1| + |\psi_3| \leq |\psi_1|\left(|\psi_1|^2 + |\psi_2|\right) + |\psi_2||\psi_1| + |\psi_3|, \\
|\theta_3| &\leq |\psi_1|^3 + 2|\psi_2||\psi_1| + |\psi_3|, \\
|\theta_4| &\leq |\psi_1|^4 + 3|\psi_1|^2|\psi_2| + 2|\psi_1||\psi_3| + |\psi_2|^2 + |\psi_4|.
\end{aligned}$$

Suppose that $|\psi_i| \leq \rho^i$, for any $i \in \mathbb{N}$, and $0 < \rho < 1$. Then in general

$$\begin{aligned}
|\theta_s| &\leq \left(1 + \sum_{j=1}^{s-1} j\right)\rho^s, \\
|\theta_s| &\leq \left(1 + \frac{s(s-1)}{2}\right)\rho^s,
\end{aligned}$$

for any $s \in \mathbb{N}$. Choose a positive real constant $\epsilon > 0$ such that $\rho < 1 - \epsilon$. We have

$$
\begin{aligned}
|\theta_s| &\leq \left(1 + \frac{s(s-1)}{2}\right)(1-\epsilon)^s \left(\frac{\rho}{1-\epsilon}\right)^s, \\
|\theta_s| &\leq \left[\left(1 + \frac{s(s-1)}{2}\right)\rho_2^s\right]\rho_1^s,
\end{aligned}
$$

where $\rho_1 \equiv \rho/(1-\epsilon)$, $\rho_2 \equiv 1-\epsilon$, and note that $0 < \rho_1 < 1$, $0 < \rho_2 < 1$. Also,

$$
\left(1 + \frac{s(s-1)}{2}\right)\rho_2^s \to 0, \text{ as } s \to \infty,
$$

which implies existence of a real constant $K < \infty$ such that

$$
\left(1 + \frac{s(s-1)}{2}\right)\rho_2^s < K,
$$

for any $s \in \mathbb{N}$. It follows that $|\theta_s| < K\rho_1^s$, as desired. ■

Lemma A.2 *Suppose $\mathbf{x}_t$ is generated according to (1), and Assumptions 1-4, and B2 hold. Then*

$$
\max_{1 \leq i \leq N} E\left(x_{it}^2\right) < K, \tag{A.3}
$$

for any $N \in \mathbb{N}$, and any $t \in \mathbb{Z}$, where constant K does not depend on N.

Proof. Taking L_2-norm of (40) and using triangle inequality, we obtain

$$
\|x_{1t}\|_{L_2} = \|\xi_{1t} + \vartheta_{ct}\|_{L_2} \leq \|\xi_{1t}\|_{L_2} + \|\vartheta_{ct}\|_{L_2}, \tag{A.4}
$$

where $\xi_{1t} = a^{-1}(L)\varepsilon_{1t}$ (see (71)). Noting that $E(\vartheta_{ct}) = 0$, (41) implies

$$
\|\vartheta_{ct}\|_{L_2} = O\left(N^{-1/2}\right). \tag{A.5}
$$

Since the coefficients of $a^{-1}(L) = c^{-1}(L)\,b_1(L)$ are absolute summable (see Lemma 2), $E(\varepsilon_{1t}) = 0$, and $\sigma_{\varepsilon 1}^2 = Var(\varepsilon_{1t})$ is bounded under Assumption 2 (condition (10)), we have

$$
\|\xi_{1t}\|_{L_2} < K. \tag{A.6}
$$

Using (A.5) and (A.6) in (A.4), we obtain

$$
E\left(x_{1t}^2\right) = \|x_{1t}\|_{L_2}^2 < K < \infty, \tag{A.7}
$$

where K does not depend on N.

Now suppose $i > 1$. Subtracting (72) from (46) yields

$$
(1 - \phi_{ii}L)\,x_{it} = (1 - \phi_{ii}L)\,\xi_{it} + \beta_i(L)\,\vartheta_{ct} + \zeta_{it}, \tag{A.8}
$$

where $\vartheta_{ct} = x_{1t} - \xi_{1t}$ (see (40) and (71)), and ζ_{it} is given by (48). $|\phi_{ii}| \leq \rho < 1$ by condition (16) of Assumption 4, and therefore the polynomial $(1 - \phi_{ii}L)$ is invertible for any $i \in \{2, 3, ...\}$. Multiplying (A.8) by $(1-\phi_{ii}L)^{-1}$, taking L_2 norm and using triangle inequality yields

$$\begin{aligned}\|x_{it}\|_{L_2} &= \left\|\xi_{it}+(1-\phi_{ii}L)^{-1}\beta_i(L)\vartheta_{ct}+(1-\phi_{ii}L)^{-1}\zeta_{it}\right\|_{L_2}\\ &\leq \|\xi_{it}\|_{L_2}+\left\|(1-\phi_{ii}L)^{-1}\beta_i(L)\vartheta_{ct}\right\|_{L_2}+\left\|(1-\phi_{ii}L)^{-1}\zeta_{it}\right\|_{L_2}\end{aligned}$$

But the coefficients of $(1-\phi_{ii}L)^{-1}$ and $\beta_i(L)$ are absolute summable, see Lemma 3. Using (41) and (52), noting that $E(\vartheta_{ct})=0$, and[12]

$$\|\xi_{it}\|_{L_2}<K,\text{ for any } N\in\mathbb{N},\text{ and any } i=\{2,3,...,N\}, \tag{A.9}$$

$$\left\|(1-\phi_{ii}L)^{-1}\beta_i(L)\vartheta_{ct}\right\|_{L_2}=O\left(N^{-1/2}\right),\text{ and }\left\|(1-\phi_{ii}L)^{-1}\zeta_{it}\right\|_{L_2}=O\left(N^{-1/2}\right),$$

we obtain

$$E\left(x_{it}^2\right)=\|x_{it}\|_{L_2}^2<K\text{ for any } N\in\mathbb{N}\text{ and any } i=\{2,3,...,N\}. \tag{A.10}$$

Results (A.7) and (A.10) establish (A.3), as desired. ■

Lemma A.3 *Suppose $\mathbf{x}_t$ is generated according to (1), and Assumptions 1-4, B1, and B2 hold. Then there exists a constant $K<\infty$, which does not dependent on N, $m_T\in\mathbb{N}$, $i,j\in\{1,2,...,N\}$, and $s\in\{1,2,...,m_T\}$, such that*

$$E\left(\frac{1}{T}\sum_{t=m_T+1}^{T}x_{it}x_{j,t-s}-E\left(\xi_{it}\xi_{j,t-s}\right)\right)^2\leq\frac{K}{T}, \tag{A.11}$$

where ξ_{it}, for $i\in\{2,3,...\}$, is defined by equation (72) and ξ_{1t} is defined by (71).

Proof. (A.11) can be established in a similar way to the proof of equations (2.10) and (2.11) in Berk (1974). ■

Lemma A.4 *Suppose Assumptions 1-4, B1, and B2 hold. Then for any $p,q\in\{0,1,2,...\}$, any $i\in\{2,3,...\}$, any $N\times 1$ dimensional vectors $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\mathbf{a}$, such that $\|\boldsymbol{\eta}\|_1=O(1)$, $\|\boldsymbol{\theta}\|=O(1)$ and $\|\mathbf{a}\|=O(1)$, we have*

$$\frac{1}{T}\sum_{t=m_T+1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}-E\left(\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}\right)\overset{L_1}{\to}0, \tag{A.12}$$

$$\frac{1}{T}\sum_{t=m_T+1}^{T}\varepsilon_{1,t-p}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-q}\overset{L_1}{\to}0, \tag{A.13}$$

$$\frac{1}{T}\sum_{t=m_T+1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\mathbf{a}'\boldsymbol{\varepsilon}_{t-q}-E\left(\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\mathbf{a}'\boldsymbol{\varepsilon}_{t-q}\right)\overset{L_1}{\to}0, \tag{A.14}$$

$$\frac{1}{T}\sum_{t=m_T+1}^{T}\xi_{1,t-p}e_{it}\overset{L_1}{\to}0, \tag{A.15}$$

and

$$\frac{1}{T}\sum_{t=m_T+1}^{T}\xi_{i,t-1}e_{it}\overset{L_1}{\to}0, \tag{A.16}$$

[12]Result (A.9) follows from definition of stationary process ξ_{it} (given by (72)) by noting that $Var(e_{it})$ is bounded under Assumption 2 (conditions (10) and (11)), coefficients in polynomial $\beta_i(L)$ are absolute summable (see Lemma 3) and that (A.6) holds.

where the convergence is uniform in p, and $\boldsymbol{\upsilon}_t$ is defined by (23).

Proof. Let $T_N = T(N)$ and $m_{T_N} = m(T_N)$ be any increasing integer valued functions of N satisfying Assumptions B1 and B2. Define the following two-dimensional array[13]

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

and the non-stochastic array

$$c_{Nt} = \frac{1}{T_N},$$

for any $t \in \mathbb{Z}$, and any $N \in \mathbb{N}$. Consider now the triangular array $\left\{\{\kappa_{Nt}/c_{Nt}, \mathcal{F}_{Nt}\}_{t=-\infty}^{T_N}\right\}_{N=1}^{\infty}$, where $\{\mathcal{F}_{Nt}\}$ denotes an array of σ-fields that is increasing in t for each N, and κ_{Nt} is measurable with respect to $\mathcal{F}_{Nt}$. Using the independence of $\mathbf{e}_t = \mathbf{R}_{-1}\boldsymbol{\varepsilon}_t$ and $\varepsilon_{1t'}$ for any $t, t' \in \mathbb{Z}$ (see Assumption 2), we have

$$\begin{aligned} E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{N,t-n}\right) &= E\left(\sum_{\ell=0}^{\infty}\boldsymbol{\theta}'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{e}_{t-q-\ell}\varepsilon_{1,t-p} \mid \mathcal{F}_{N,t-n}\right), \\ &= \begin{cases} 0 & \text{for } p < n \\ \sum_{\ell=\ell_1(n,q)}^{\infty}\boldsymbol{\theta}'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{e}_{t-q-\ell}\varepsilon_{1,t-p} & \text{for } p \geq n \end{cases}, \end{aligned}$$

where

$$\ell_1(n,q) = \max\{n-q, 0\}.$$

Also,

$$\begin{aligned} \sup_{p\in\{0,1,\ldots\}} E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-n}\right)\right]^2\right\} &= \sigma_{\varepsilon 1}^2 \sum_{\ell=\ell_1(n,q)}^{\infty}\boldsymbol{\theta}'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{R}_{-1}E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\mathbf{R}_{-1}'\boldsymbol{\Phi}_{-1}'^{\ell}\boldsymbol{\theta}, \\ &\leq \varsigma_{nq}, \end{aligned}$$

where

$$\varsigma_{nq} = \sigma_{\varepsilon 1}^2 \left\|Var\left(\boldsymbol{\varepsilon}_t\right)\right\| \left\|\mathbf{R}_{-1}\right\|^2 \left\|\boldsymbol{\theta}\right\|^2 \sum_{\ell=\ell_1(n,q)}^{\infty}\left\|\boldsymbol{\Phi}_{-1}\right\|^{2\ell}.$$

Condition (11) of Assumption 2 implies $\|\mathbf{R}_{-1}\| \leq \sqrt{\|\mathbf{R}_{-1}\|_1 \|\mathbf{R}_{-1}\|_\infty} = O(1)$, $\sigma_{\varepsilon 1}^2 < K$ and $\|Var(\boldsymbol{\varepsilon}_t)\| < K$ by condition (10) of Assumption 2, and $\|\boldsymbol{\Phi}_{-1}\| \leq \sqrt{\|\boldsymbol{\Phi}_{-1}\|_1 \|\boldsymbol{\Phi}_{-1}\|_\infty} \leq \rho < 1$ under Assumption 4, condition (16). Since also $\|\boldsymbol{\theta}\| = O(1)$, it follows that (for any fixed $q \in \mathbb{N}_0$)

$$\varsigma_{0,q} < K \text{ and } \varsigma_{nq} \to 0 \text{ as } n \to \infty.$$

Therefore, the array $\{\kappa_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm, which establishes uniform integrability. Furthermore, using Liapunov's inequality, the two-dimensional array $\{\kappa_{Nt}, \mathcal{F}_{Nt}\}$ is L_1-mixingale with respect to the non-stochastic array $\{c_{Nt}\}$. Noting that

$$\lim_{N\to\infty}\sum_{t=m_{T_N}+1}^{T_N} c_{Nt} = \lim_{N\to\infty}\sum_{t=m_{T_N}+1}^{T_N}\frac{1}{T_N} = \frac{T_N - m_{T_N}}{T_N} = 1 < \infty, \tag{A.17}$$

[13] Note that κ_{Nt} is also a function of p and q but we ommit these subscripts to simplify notations.

and

$$\lim_{N\to\infty}\sum_{t=m_{T_N}+1}^{T_N} c_{Nt}^2=\lim_{N\to\infty}\sum_{t=m_{T_N}+1}^{T_N}\frac{1}{T_N^2}=\frac{T_N-m_{T_N}}{T_N^2}=0, \tag{A.18}$$

it follows that the array $\{\kappa_{Nt},\mathcal{F}_{Nt}\}$ satisfies conditions of a mixingale weak law,[14] which implies $\sum_{t=m_{T_N}+1}^{T_N}\kappa_{Nt}\overset{L_1}{\to}0$, uniformly in p since the upper bound ς_{nq} does not depend on p. This completes the proof of result (A.13).

Result (A.14) is established in a similar fashion as result (A.13). This time we define

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

for any $t\in\mathbb{Z}$, and any $N\in\mathbb{N}$. Again let $\{\mathcal{F}_{Nt}\}$ denote array of σ-fields that is increasing in t for each N and κ_{Nt} is measurable with respect to $\mathcal{F}_{Nt}$. We have

$$E\left(\frac{\kappa_{Nt}}{c_{Nt}}\mid\mathcal{F}_{N,t-n}\right)=\begin{cases}\sum_{\ell=\ell_2(p,n)}^{\infty}\boldsymbol{\theta}'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{R}_{-1}\left[\boldsymbol{\varepsilon}_{t-p-\ell}\mathbf{a}'\boldsymbol{\varepsilon}_{t-q}-E\left(\boldsymbol{\varepsilon}_{t-p-\ell}\mathbf{a}'\boldsymbol{\varepsilon}_{t-q}\right)\right] & \text{for } q\geq n\\ 0 & \text{for } q<n\end{cases}, \tag{A.19}$$

where

$$\ell_2(p,n)=\max\{n-p,0\}.$$

Define

$$z_{tpq\ell}=\left(\boldsymbol{\theta}'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{R}_{-1}\boldsymbol{\varepsilon}_{t-p-\ell}\right)\left(\mathbf{a}'\boldsymbol{\varepsilon}_{t-q}\right). \tag{A.20}$$

Using (A.20) in (A.19), we obtain

$$E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}}\mid\mathcal{F}_{N,t-n}\right)\right]^2\right\}=\begin{cases}\sum_{\ell=\ell_2(p,n)}^{\infty}\sum_{h=\ell_2(p,n)}^{\infty}\left[E\left(z_{tpq\ell}z_{tpqh}\right)-E\left(z_{tpq\ell}\right)E\left(z_{tpqh}\right)\right] & \text{for } q\geq n\\ 0 & \text{for } q<n\end{cases}. \tag{A.21}$$

Note that

$$E\left(z_{tpq\ell}\right)=\begin{cases}0 & \text{for } \ell\neq p-q\\ \boldsymbol{\theta}'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{R}_{-1}Var\left(\boldsymbol{\varepsilon}_t\right)\mathbf{a} & \text{for } \ell=p-q\end{cases}.$$

This implies that

$$\sum_{\ell=\ell_2(p,n)}^{\infty}E\left(z_{tpq\ell}\right)=\begin{cases}\boldsymbol{\theta}'\boldsymbol{\Phi}_{-1}^{p-q}\mathbf{R}_{-1}Var\left(\boldsymbol{\varepsilon}_{t-q}\right)\mathbf{a} & \text{for } p-q\geq\max\{p-n,0\}\\ \mathbf{0} & p-q<\max\{p-n,0\}\end{cases}.$$

But

$$\begin{aligned}\left\|\boldsymbol{\theta}'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{R}_{-1}Var\left(\boldsymbol{\varepsilon}_{t-q}\right)\mathbf{a}\right\| &\leq \|\boldsymbol{\theta}\|\left\|\boldsymbol{\Phi}_{-1}\right\|^{\ell}\left\|\mathbf{R}_{-1}\right\|\left\|Var\left(\boldsymbol{\varepsilon}_{t-q}\right)\right\|\|\mathbf{a}\|\\ &< K,\end{aligned}$$

where as before $\|\boldsymbol{\theta}\|=O(1)$, $\|\mathbf{a}\|=O(1)$, $\|\boldsymbol{\Phi}_{-1}\|\leq\sqrt{\|\boldsymbol{\Phi}_{-1}\|_1\|\boldsymbol{\Phi}_{-1}\|_\infty}\leq\rho<1$ (by condition (16) of Assumption 4), $\|\mathbf{R}_{-1}\|\leq\sqrt{\|\mathbf{R}_{-1}\|_1\|\mathbf{R}_{-1}\|_\infty}=O(1)$ (by condition (11) of Assumption 2) and $\|Var(\boldsymbol{\varepsilon}_{t-q})\|=O(1)$ (by condition (10) of Assumption 2). It follows that for $q\geq n$,

$$\sup_{p\in\{0,1,2,\ldots\}}\sum_{\ell=\ell_2(p,n)}^{\infty}E\left(z_{tpq\ell}\right)\sum_{h=\ell_2(p,n)}^{\infty}E\left(z_{tpqh}\right)<K. \tag{A.22}$$

[14] See Theorem 19.11 in Davidson (1994).

Using similar arguments (and noting that the fourth moment of ε_{it} is uniformly bounded in i), it can be shown that

$$\sup_{p\in\{0,1,2,...\}} \sum_{\ell=\ell_2(p,n)}^{\infty} \sum_{h=\ell_2(p,n)}^{\infty} E\left(z_{tpq\ell} z_{tpqh}\right) < K \text{ for } q \geq n. \tag{A.23}$$

Results (A.21), (A.22) and (A.23) now establish the existence of a non-stochastic array, ς_{nq}, such that

$$\sup_{p\in\{0,1,2,...\}} E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{N,t-n}\right)\right]^2\right\} < \varsigma_{nq},$$

where for a fixed $q \in \{0, 1, 2, ...\}$,

$$\varsigma_{0q} < K \text{ and } \varsigma_{nq} \to 0 \text{ as } n \to \infty.$$

Therefore, the array $\{\kappa_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm, which establishes uniform integrability. Furthermore, using Liapunov's inequality, the two-dimensional array $\{\kappa_{Nt}, \mathcal{F}_{Nt}\}$ is L_1-mixingale with respect to the non-stochastic array $\{c_{Nt}\}$. Noting that equations (A.17)-(A.18) hold, it follows that the array $\{\kappa_{Nt}, \mathcal{F}_{Nt}\}$ satisfies conditions of a mixingale weak law,[15] which implies $\sum_{t=m_{T_N}+1}^{T_N} \kappa_{Nt} \overset{L_1}{\to} 0$, uniformly in p since the upper bound ς_{nq} does not depend on p. This completes the proof of (A.14).

Results (A.15) and (A.16) can also be established in the similar fashion as result (A.13), but this time we define $\kappa_{Nt} = \frac{1}{T_N}\xi_{1,t-p}e_{it}$ to establish result (A.15), and $\kappa_{Nt} = \frac{1}{T_N}\xi_{i,t-1}e_{it}$ in order to establish result (A.16). Result (A.14) can be established in the same way as Lemma 1 in Chudik and Pesaran (2011). This completes the proof. ■

Lemma A.5 *Let assumptions 1-4, B1, and B3 hold. Then for any $i \in \{1, 2, 3, ...\}$, any $j \in \{2, 3, ...\}$, any $p, q \in \{0, 1, 2, ...\}$, and any $N \times 1$ dimensional vectors $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, such that $\|\boldsymbol{\eta}\|_1 = O(1)$ and $\|\boldsymbol{\theta}\|_\infty = O(N^{-1})$,*

$$\frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q} - \sqrt{\varkappa_2} E\left(\sqrt{N}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}\right) \overset{L_1}{\to} 0, \tag{A.24}$$

$$\frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \upsilon_{1,t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q} - \sqrt{\varkappa_2} E\left(\sqrt{N}\upsilon_{1,t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}\right) \overset{L_1}{\to} 0, \tag{A.25}$$

$$\frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \varepsilon_{1,t-p}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-q} \overset{L_1}{\to} 0, \tag{A.26}$$

$$\frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \varepsilon_{1,t-p}\upsilon_{1,t-q} \overset{L_1}{\to} 0, \tag{A.27}$$

$$\frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}e_{i,t-q} - \sqrt{\varkappa_2} E\left(\sqrt{N}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}e_{i,t-q}\right) \overset{L_1}{\to} 0, \tag{A.28}$$

and

$$\frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \upsilon_{1,t-p}e_{j,t-q} \overset{L_1}{\to} 0, \tag{A.29}$$

where the convergence is uniform in p, $\boldsymbol{\upsilon}_t$ is defined by equation (23), $\mathbf{e}_t$ is defined by (20), and $\varkappa_2 = lim(T/N)$ as $(N,T) \overset{j}{\to} \infty$.

[15] See Theorem 19.11 in Davidson (1994).

Proof. We have

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}=\sqrt{\frac{T}{N}}\left(\frac{1}{T}\sum_{t=m_T+1}^{T}\left(\sqrt{N}\boldsymbol{\theta}\right)'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}\right), \tag{A.30}$$

where

$$\left\|\sqrt{N}\boldsymbol{\theta}\right\|=\sqrt{N\left\|\boldsymbol{\theta}\right\|_{\infty}\left\|\boldsymbol{\theta}\right\|_{1}}=O\left(1\right). \tag{A.31}$$

Using now result (A.12) of Lemma A.4 yields

$$\frac{1}{T}\sum_{t=m_T+1}^{T}\mathbf{b}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}-E\left[\mathbf{b}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}\right]\overset{L_1}{\to}0\text{ uniformly in }p, \tag{A.32}$$

under Assumptions B1 and B2, where $\mathbf{b}=\left(\sqrt{N}\boldsymbol{\theta}\right)'$, and $\left\|\mathbf{b}\right\|=O\left(1\right)$ by (A.31). Multiplying (A.32) by $(T/N)^{1/2}$, and noting that Assumption B3 is a special case of Assumption B2, where $(N,T)\overset{j}{\to}\infty$ at any rate, and that under Assumption B3,

$$\sqrt{\frac{T}{N}}\to\sqrt{\varkappa_2}<\infty,$$

we obtain

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}-\sqrt{\varkappa_2}E\left(\sqrt{N}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}\right)\overset{L_1}{\to}0\text{ uniformly in }p,$$

under Assumptions B1 and B3, as desired. This completes the proof of (A.24). Similarly, result (A.26) follows directly from result (A.13). Result (A.28) can also be established in a similar way by using (A.14) and noting that $e_{i,t-q}=\mathbf{a}'\boldsymbol{\varepsilon}_{t-q}$ for $\mathbf{a}=\mathbf{R}_{-1}'\mathbf{s}_i$ and that $\left\|\mathbf{R}_{-1}'\mathbf{s}_i\right\|\leq\sqrt{\left\|\mathbf{R}_{-1}\right\|_1\left\|\mathbf{R}_{-1}\right\|_{\infty}}=O\left(1\right)$ by condition (11) of Assumption 2.

To establish result (A.27), we make use of equation (30), which implies

$$\upsilon_{1t}=\mathbf{r}_{-1}'\boldsymbol{\varepsilon}_t+\boldsymbol{\phi}_{-1}'\boldsymbol{\upsilon}_{t-1}, \tag{A.33}$$

where $\mathbf{r}_{-1}'\boldsymbol{\varepsilon}_t=e_{1t}$ and the vector $\boldsymbol{\phi}_{-1}$ satisfies

$$\left\|\boldsymbol{\phi}_{-1}\right\|_{\infty}=O\left(N^{-1}\right), \tag{A.34}$$

by condition (5) of Assumption 1. Using result (A.26) for $\boldsymbol{\theta}=\boldsymbol{\phi}_{-1}$ we have

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\varepsilon_{1,t-p}\boldsymbol{\phi}_{-1}'\boldsymbol{\upsilon}_{t-q}\overset{L_1}{\to}0\text{ uniformly in }p, \tag{A.35}$$

for any $p,q\in\{0,1,2,...\}$, under Assumptions B1 and B3. Similarly, $\mathbf{r}_{-1}$ satisfies

$$\left\|\mathbf{r}_{-1}\right\|_{\infty}=O\left(N^{-1}\right), \tag{A.36}$$

by condition (12) of Assumption 2. Noting that $\boldsymbol{\upsilon}_t$ reduces to

$$\boldsymbol{\upsilon}_t=\sum_{\ell=0}^{\infty}\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{R}_{-1}\boldsymbol{\varepsilon}_t=\mathbf{I}_{-1}\boldsymbol{\varepsilon}_t\text{ for }\boldsymbol{\Phi}_{-1}=\mathbf{0}\text{ and }\mathbf{R}_{-1}=\mathbf{I}_{-1},$$

where $\mathbf{I}_{-1}$ is identity matrix with the first column replaced by zeros, result (A.26) implies (for $\boldsymbol{\theta}=\mathbf{r}_{-1}$,$\boldsymbol{\Phi}_{-1}=$

0 and $\mathbf{R}_{-1}=\mathbf{I}_{-1}$) that

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\varepsilon_{1,t-p}\mathbf{r}'_{-1}\boldsymbol{\varepsilon}_{t-q}\overset{L_1}{\to}0\text{, uniformly in } p, \tag{A.37}$$

for any $p,q\in\{0,1,2,...\}$, under Assumptions B1 and B3. (A.33), (A.35), and (A.37) now establish (A.27), as desired.

Result (A.29) is established in a similarly way by making use of (A.28) and (A.33). For $\boldsymbol{\theta}=\boldsymbol{\phi}_{-1}$ (see (A.34)), (A.28) implies

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\boldsymbol{\phi}'_{-1}\boldsymbol{\upsilon}_{t-p}e_{j,t-q}-\sqrt{\varkappa_2}E\left(\sqrt{N}\boldsymbol{\phi}'_{-1}\boldsymbol{\upsilon}_{t-p}e_{j,t-q}\right)\overset{L_1}{\to}0\text{ uniformly in } p, \tag{A.38}$$

under Assumptions B1 and B3, where

$$E\left(\sqrt{N}\boldsymbol{\phi}'_{-1}\boldsymbol{\upsilon}_{t-p}e_{j,t-q}\right)=\begin{cases}\sqrt{N}\boldsymbol{\phi}'_{-1}\boldsymbol{\Phi}_{-1}^{q-p}E\left(\mathbf{e}_{t-q}e_{j,t-q}\right) & \text{for } q\geq p\\ 0 & \text{for } q<p\end{cases}, \tag{A.39}$$

$$\left\|E\left(\sqrt{N}\boldsymbol{\phi}'_{-1}\boldsymbol{\upsilon}_{t-p}e_{j,t-q}\right)\right\|_1\leq\sqrt{N}\left\|\boldsymbol{\phi}_{-1}\right\|_\infty\left\|\boldsymbol{\Phi}_{-1}\right\|_1^{q-p}\left\|E\left(\mathbf{e}_{t-q}e_{j,t-q}\right)\right\|_1=O\left(N^{-\frac{1}{2}}\right),$$

$\left\|\boldsymbol{\phi}_{-1}\right\|_\infty=O\left(N^{-1}\right)$ by condition (5) of Assumption 1, $\left\|\boldsymbol{\Phi}_{-1}\right\|_1^{q-p}\leq\rho^{q-p}\leq1$, for $q\leq p$, by condition (16) of Assumption 4,

$$\left\|E\left(\mathbf{e}_{t-q}e_{j,t-q}\right)\right\|_1\leq\left\|\mathbf{R}_{-1}\right\|_1\left\|\mathbf{R}_{-1}\right\|_\infty\left\|Var\left(\boldsymbol{\varepsilon}_t\right)\right\|_1,$$

$\left\|\mathbf{R}_{-1}\right\|_1\left\|\mathbf{R}_{-1}\right\|_\infty<K$ by condition (11) of Assumption 2, and $\left\|Var\left(\boldsymbol{\varepsilon}_t\right)\right\|_1<K$ by condition (10) of Assumption 2. For $\boldsymbol{\theta}=\mathbf{r}_{-1}$,$\boldsymbol{\Phi}_{-1}=\mathbf{0}$ and $\mathbf{R}_{-1}=\mathbf{I}_{-1}$, (A.28) implies

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\mathbf{r}'_{-1}\boldsymbol{\varepsilon}_{t-p}e_{j,t-q}-\sqrt{\varkappa_2}E\left(\sqrt{N}\mathbf{r}'_{-1}\boldsymbol{\varepsilon}_{t-p}e_{j,t-q}\right)\overset{L_1}{\to}0\text{ uniformly in } p, \tag{A.40}$$

under Assumptions B1 and B3, where

$$E\left(\sqrt{N}\mathbf{r}'_{-1}\boldsymbol{\varepsilon}_{t-p}e_{j,t-q}\right)=\begin{cases}\sqrt{N}\mathbf{r}'_{-1}\mathbf{R}_{-1}\mathbf{s}_j & \text{for } q=p\\ 0 & \text{for } q\neq p\end{cases}, \tag{A.41}$$

$$\left\|E\left(\sqrt{N}\mathbf{r}'_{-1}\boldsymbol{\varepsilon}_{t-p}e_{j,t-q}\right)\right\|_1\leq\sqrt{N}\left\|\mathbf{r}_{-1}\right\|_\infty\left\|\mathbf{R}_{-1}\right\|_1=O\left(N^{-\frac{1}{2}}\right), \tag{A.42}$$

$\left\|\mathbf{r}_{-1}\right\|_\infty=O\left(N^{-1}\right)$ and $\left\|\mathbf{R}_{-1}\right\|_1<K$ by Assumption 2 (see conditions (12) and (11), respectively). (A.38)-(A.42) establish (A.29), as desired.

Result (A.25) is also established by making use of equation (A.33). For $\boldsymbol{\theta}=\boldsymbol{\phi}_{-1}$ (noting that $\boldsymbol{\phi}_{-1}$ satisfies (A.34)) and for any vector $\boldsymbol{\eta}$ such that $\left\|\boldsymbol{\eta}\right\|_1=O\left(1\right)$, (A.24) implies

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\boldsymbol{\phi}'_{-1}\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}-\sqrt{\varkappa_2}E\left(\sqrt{N}\boldsymbol{\phi}'_{-1}\boldsymbol{\upsilon}_{t-p}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-q}\right)\overset{L_1}{\to}0\text{ uniformly in } p, \tag{A.43}$$

under Assumptions B1 and B3. Result (A.14) of Lemma A.4 implies, by setting $\mathbf{a}=\sqrt{N}\mathbf{r}_{-1}$ and noting that $\left\|\mathbf{a}\right\|=\sqrt{N}\left\|\mathbf{r}_{-1}\right\|=\sqrt{N}\sqrt{\left\|\mathbf{r}_{-1}\right\|_\infty\left\|\mathbf{r}_{-1}\right\|_1}=O\left(1\right)$ (see (A.36)) and $\left\|\boldsymbol{\eta}\right\|\leq\sqrt{\left\|\boldsymbol{\eta}\right\|_1\left\|\boldsymbol{\eta}\right\|_\infty}=O\left(1\right)$, we have

$$\frac{1}{T}\sum_{t=m_T+1}^{T}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-p}\sqrt{N}\mathbf{r}_{-1}\boldsymbol{\varepsilon}_{t-q}-E\left(\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-p}\sqrt{N}\mathbf{r}_{-1}\boldsymbol{\varepsilon}_{t-q}\right)\overset{L_1}{\to}0\text{ uniformly in } p, \tag{A.44}$$

under Assumptions B1, and B2. Using the same arguments as in (A.30), it follows from (A.44) that

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-p}\mathbf{r}_{-1}\boldsymbol{\varepsilon}_{t-q}-\sqrt{\varkappa_2}E\left(\boldsymbol{\eta}'\boldsymbol{\upsilon}_{t-p}\sqrt{N}\mathbf{r}_{-1}\boldsymbol{\varepsilon}_{t-q}\right)\stackrel{L_1}{\to}0 \text{ uniformly in } p, \tag{A.45}$$

under Assumptions B1, and B3. (A.33), (A.43) and (A.45) establish (A.25), as desired. This completes the proof. ■

Lemma A.6 *Suppose that Assumptions 1-4 hold. Then for any $i\in\mathbb{N}$, any $p\in\{0,1,2,...\}$, and any $N\times 1$ dimensional vector $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta}\|_\infty=O\left(N^{-1}\right)$,*

$$E\left(\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-1}\right)=O\left(N^{-1}\right), \tag{A.46}$$

and

$$E\left(\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\upsilon_{1t}\right)=O\left(N^{-1}\right), \tag{A.47}$$

where $\mathbf{s}_i$ is an $N\times 1$ dimensional selection vector with $s_{ij}=0$ for $j\neq i$ and $s_{ii}=1$, and $\boldsymbol{\upsilon}_t$ is defined by equation (23).

Proof. We have

$$\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-1}=\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\upsilon}_{t-1}'\boldsymbol{\theta}=\sum_{\ell=0}^{\infty}\mathbf{s}_i'\boldsymbol{\Phi}_{-1}^{\ell}\mathbf{R}_{-1}\boldsymbol{\varepsilon}_{t-p-\ell}\sum_{\ell=0}^{\infty}\boldsymbol{\varepsilon}_{t-1-\ell}'\mathbf{R}_{-1}'\boldsymbol{\Phi}_{-1}'^{\ell}\boldsymbol{\theta}. \tag{A.48}$$

Taking expectations of (A.48) and noting that $\boldsymbol{\varepsilon}_t$ is independently distributed of $\boldsymbol{\varepsilon}_{t'}$ for any $t\neq t'$, we obtain

$$\begin{aligned}E\left(\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-1}\right) &= \sum_{\ell=\max\{1,p\}}\mathbf{s}_i'\boldsymbol{\Phi}_{-1}^{\ell-p}\mathbf{R}_{-1}E\left(\boldsymbol{\varepsilon}_{t-\ell}\boldsymbol{\varepsilon}_{t-\ell}'\right)\mathbf{R}_{-1}'\boldsymbol{\Phi}_{-1}'^{\ell-1}\boldsymbol{\theta}\\ &\leq \|\mathbf{R}_{-1}\|_\infty\|\mathbf{R}_{-1}\|_1\|\boldsymbol{\theta}\|_\infty\|Var\left(\boldsymbol{\varepsilon}_t\right)\|_\infty\sum_{\ell=\max\{1,p\}}\|\boldsymbol{\Phi}_{-1}\|_\infty^{\ell-p}\|\boldsymbol{\Phi}_{-1}\|_1^{\ell-1},\end{aligned}$$

where $\|\mathbf{R}_{-1}\|_\infty\|\mathbf{R}_{-1}\|_1=O(1)$ by condition (11) of Assumption 2, $\|\boldsymbol{\theta}\|_\infty=O\left(N^{-1}\right)$, $\|E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t\right)\|_\infty=\|Var\left(\boldsymbol{\varepsilon}_t\right)\|_\infty=O(1)$ by condition (10) of Assumption 2, and $\|\boldsymbol{\Phi}_{-1}\|_\infty\leq\rho<1$, $\|\boldsymbol{\Phi}_{-1}\|_1^{\ell}\leq\rho<1$ by condition (16) of Assumption 4. It follows that $E\left(\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-1}\right)=O\left(N^{-1}\right)$, as required.

To establish result (A.47), we make use of equation (A.33). We have

$$E\left(\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\upsilon_{1t}\right)=E\left(\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\mathbf{r}_{-1}'\boldsymbol{\varepsilon}_t\right)+E\left(\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\phi}_{-1}'\boldsymbol{\upsilon}_{t-1}\right).$$

Noting that $\left\|\boldsymbol{\phi}_{-1}\right\|_\infty=O\left(N^{-1}\right)$ by condition (5) of Assumption 1, result (A.46) (for $\boldsymbol{\theta}=\boldsymbol{\phi}_{-1}$) implies $E\left(\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\phi}_{-1}'\boldsymbol{\upsilon}_{t-1}\right)=O\left(N^{-1}\right)$. Furthermore,

$$E\left(\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\mathbf{r}_{-1}'\boldsymbol{\varepsilon}_t\right)=\begin{cases}0 & \text{for } p>0\\ \mathbf{s}_i'\mathbf{R}_{-1}E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\mathbf{r}_{-1} & \text{for } p=0\end{cases},$$

where

$$\mathbf{s}_i'\mathbf{R}_{-1}E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\mathbf{r}_{-1}\leq\|\mathbf{R}_{-1}\|_\infty\|Var\left(\boldsymbol{\varepsilon}_t\right)\|_\infty\|\mathbf{r}_{-1}\|_\infty=O\left(N^{-1}\right),$$

using the same arguments as in the derivation of (A.46) and noting that $\|\mathbf{r}_{-1}\|_\infty=O\left(N^{-1}\right)$ by condition (12) of Assumption 2. It follows that $E\left(\mathbf{s}_i'\boldsymbol{\upsilon}_{t-p}\upsilon_{1t}\right)=O\left(N^{-1}\right)$, as required. ■

Lemma A.7 *Suppose $\mathbf{x}_t$ is given by model (1) and Assumptions 1-4, B1, and B2 hold. Then for any $i > 1$ we have,*

$$\left\| \widehat{\mathbf{C}}_i - \mathbf{C}_i \right\|_\infty \overset{p}{\to} 0,$$

where $\mathbf{C}_i$ and $\widehat{\mathbf{C}}_i$ are defined by (70) and (79), respectively.

Proof.

$$\left\| \widehat{\mathbf{C}}_i - \mathbf{C}_i \right\|_\infty = \max_{j \in \{1,..,m_T+2\}} \sum_{\ell=1}^{m_T+2} \left| \widehat{c}_{ij\ell} - c_{ij\ell} \right|, \tag{A.49}$$

where $c_{ij\ell}$ and $\widehat{c}_{ij\ell}$ denote the (j,ℓ)-th elements of $\mathbf{C}_i$ and $\widehat{\mathbf{C}}_i$, respectively. Liapunov's inequality and Lemma A.3 in Appendix establish

$$E \left| \widehat{c}_{ij\ell} - c_{ij\ell} \right| \leq \sqrt{E\left[\left(\widehat{c}_{ij\ell} - c_{ij\ell} \right)^2 \right]} \leq K \frac{1}{\sqrt{T}}, \tag{A.50}$$

where $K < \infty$ does not depend on N, $m_T \in \mathbb{N}$, and $j, \ell \in \{1, 2, ..., m_T + 2\}$. Taking expectations of both sides of (A.49) and making use of (A.50) yields

$$E \left\| \widehat{\mathbf{C}}_i - \mathbf{C}_i \right\|_\infty \leq K \left(\frac{m_T + 2}{\sqrt{T}} \right).$$

But under Assumption B1, $m_T^2/T \to 0$, and hence $\left\| \widehat{\mathbf{C}}_i - \mathbf{C}_i \right\|_\infty \overset{L_1}{\to} 0$. Convergence in L_1 norm implies convergence in probability. ■

Lemma A.8 *Suppose $\mathbf{x}_t$ is given by model (1) and Assumptions 1-5, B1 and B2 hold. Then for any $i > 1$ we have,*

$$\left\| \widehat{\mathbf{C}}_i^{-1} - \mathbf{C}_i^{-1} \right\|_\infty \overset{p}{\to} 0,$$

where $\mathbf{C}_i$ and $\widehat{\mathbf{C}}_i$ are defined by (70) and (79), respectively.

Proof. Let $p_c = \left\| \mathbf{C}_i^{-1} \right\|_\infty$, $q_c = \left\| \widehat{\mathbf{C}}_i^{-1} - \mathbf{C}_i^{-1} \right\|_\infty$, and $r_c = \left\| \widehat{\mathbf{C}}_i - \mathbf{C}_i \right\|_\infty$. Using the triangle inequality and the submultiplicative property of matrix norm $\|.\|_\infty$, we have

$$\begin{aligned} q_c &= \left\| \widehat{\mathbf{C}}_i^{-1} \left(\mathbf{C}_i - \widehat{\mathbf{C}}_i \right) \mathbf{C}_i^{-1} \right\|_\infty, \\ &\leq \left\| \widehat{\mathbf{C}}_i^{-1} \right\|_\infty r_c p_c, \\ &\leq \left\| \left(\widehat{\mathbf{C}}_i^{-1} - \mathbf{C}_i^{-1} \right) + \mathbf{C}_i^{-1} \right\|_\infty r_c p_c, \\ &\leq \left(p_c + q_c \right) r_c p_c, \end{aligned}$$

and (subtracting $r_c p_c q_c$ from both sides)

$$\left(1 - r_c p_c \right) q_c \leq p_c^2 r_c. \tag{A.51}$$

Note that $r_c \overset{p}{\to} 0$ by Lemma A.7, and $p_c = O(1)$ since ξ_{it}, for $i \in \{1, 2, ..., N\}$, is a stationary invertible process with absolute summable autocovariances. Therefore

$$\left(1 - r_c p_c \right) \overset{p}{\to} 1, \tag{A.52}$$

and

$$p_c^2 r_c \overset{p}{\to} 0. \tag{A.53}$$

Results (A.51)-(A.53) imply that $q_c \xrightarrow{p} 0$, as desired.[16] ■

Lemma A.9 *Suppose $\mathbf{x}_t$ is given by model (1) and Assumptions 1-4, B1 and B2 hold. Then for any $i > 1$ we have,*

$$\left\| \frac{\mathbf{G}_i' \boldsymbol{\psi}_{i\cdot}}{\sqrt{T}} \right\|_\infty \xrightarrow{p} 0,$$

where $\boldsymbol{\psi}_{i\cdot}$ is defined by (82), and $\mathbf{G}_i$ is defined by (75).

Proof. Each of the individual elements of $\mathbf{G}_i' \boldsymbol{\psi}_{i\cdot} / \sqrt{T}$ can be expressed as

$$\frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} x_{j,t-s} \psi_{m_T it},$$

for a suitable choice of $j \in \{1, i\}$, and $s \in \{0, 1, 2, ..., m_T\}$, where $\psi_{m_T it}$ is defined by (68). We have

$$\begin{aligned} E \left| \frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} x_{j,t-s} \psi_{m_T it} \right| &\leq \frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} E \left| x_{j,t-s} \psi_{m_T it} \right| \\ &\leq \frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \left[E \left(x_{j,t-s} \right)^2 E \left(\psi_{m_T it} \right)^2 \right]^{1/2} \\ &\leq \frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \max_{j \in \{1,2,...N\}} \left[E \left(x_{j,t-s}^2 \right) \right]^{1/2} \sum_{\ell=m_T+1}^{\infty} \left| \beta_{i\ell} \right| \left[E \left(x_{1,t-\ell}^2 \right) \right]^{1/2}, \quad \text{(A.54)} \end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality and the third inequality uses the triangle inequality, which implies $\left\| \psi_{m_T it} \right\|_{L_2} \leq \sum_{\ell=m_T+1}^{\infty} \left| \beta_{i\ell} \right| \left\| x_{1,t-\ell} \right\|_{L_2}$. But by Lemma A.2, $\max_{j \in \{1,2,...,N\}} E \left(x_{jt}^2 \right) < K$, and (A.54) now yields

$$E \left| \frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} x_{j,t-s} \psi_{m_T it} \right| \leq K \sqrt{T} \sum_{\ell=m_T+1}^{\infty} \left| \beta_{i\ell} \right|.$$

But using Lemma 3 (for $0 < \rho < 1$)

$$\sqrt{T} \sum_{\ell=m_T+1}^{\infty} \left| \beta_{i\ell} \right| \leq K \frac{\sqrt{T} \rho^{m_T+1}}{1-\rho},$$

and under Assumptions B1-B2, and noting that $K < \infty$ does not depend on N, or T, we have

$$\sqrt{T} \sum_{\ell=m_T+1}^{\infty} \left| \beta_{i\ell} \right| \to 0, \text{ as } T \to \infty,$$

and hence

$$\left\| \frac{\mathbf{G}_i' \boldsymbol{\psi}_{i\cdot}}{\sqrt{T}} \right\|_\infty \xrightarrow{L_1} 0.$$

Convergence in L_1 norm implies convergence in probability. ■

[16] Here we have used the fact that for any real constant $0 < \epsilon < 1$, the probability of $r_c p_c > \epsilon$ can be made arbitrarily small by choosing T sufficiently large, since $r_c p_c \xrightarrow{p} 0$.

Lemma A.10 *Suppose $\mathbf{x}_t$ is generated according to (1) and Assumptions 1-4, B1 and B3 hold. Then for any $i > 1$,*

$$\left\| \frac{\mathbf{G}_i' \boldsymbol{\zeta}_{i\cdot}}{\sqrt{T}} \right\|_\infty \xrightarrow{p} 0, \tag{A.55}$$

where matrix $\mathbf{G}_i$ is defined by (75), and $\boldsymbol{\zeta}_{i\cdot}$ is defined by (82). Consider now the case where Assumption B3 is replaced by (weaker) Assumption B2, but the other assumptions are maintained. Then for any $i > 1$,

$$\left\| \frac{\mathbf{G}_i' \boldsymbol{\zeta}_{i\cdot}}{T} \right\|_\infty \xrightarrow{p} 0. \tag{A.56}$$

Proof. The first element of the $(m_T + 2) \times 1$ dimensional vector $\mathbf{G}_i' \boldsymbol{\zeta}_{i\cdot} / \sqrt{T}$ is

$$\frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} x_{i,t-1} \zeta_{it}. \tag{A.57}$$

Multiplying equation (27) by $c^{-1}(L)$ and substituting the outcome into equation (24) for $x_{1,t-1}$ yields the following relation for the non-dominant unit.

$$x_{it} = f_i(L) \varepsilon_{1t} + d_i(L) c^{-1}(L) \upsilon_{1,t-1} + \upsilon_{it}, \text{ for } i > 1, \tag{A.58}$$

where

$$f_i(L) = L d_i(L) c^{-1}(L) b_1(L) + b_i(L). \tag{A.59}$$

The process ζ_{it}, as defined in (48), can be written as,

$$\zeta_{it} = \boldsymbol{\phi}_{-1,-i}' \boldsymbol{\upsilon}_{t-1} - g_i(L) \upsilon_{1t}, \tag{A.60}$$

where

$$g_i(L) = \left[r_{i1} + \boldsymbol{\phi}_{-1,-i}' \boldsymbol{\Phi}_{-1}(L) \mathbf{r}_1 L \right] b_1^{-1}(L). \tag{A.61}$$

Coefficients in the polynomials $c^{-1}(L)$, $b_1(L)$, and $b_1^{-1}(L)$ are absolute summable (see Lemma 2). (B.2) implies absolute summability of the coefficients in $\boldsymbol{\phi}_{-1,-i}' \boldsymbol{\Phi}_{-1}(L) \mathbf{r}_1$, and using the same arguments as in proof of Lemma 3, we have

$$|d_{i\ell}| = \left\| \mathbf{s}_i' \boldsymbol{\Phi}_{-1}^{\ell} \boldsymbol{\phi}_1 \right\|_\infty < K \rho^\ell, \text{ and } b_{i\ell} = \left\| \mathbf{s}_i' \boldsymbol{\Phi}_{-1}^{\ell} \mathbf{r}_1 \right\|_\infty < K \rho^\ell. \tag{A.62}$$

It follows that polynomials $f_i(L)$, $d_i(L) c^{-1}(L)$, and $g_i(L)$ in (A.58) and (A.60) are absolute summable. Vector $\boldsymbol{\phi}_{-1,-i}$ satisfies $\left\| \boldsymbol{\phi}_{-1,-i} \right\|_\infty = O\left(N^{-1}\right)$ by condition (6) of Assumption 1 and result (A.26) of Lemma A.5 imply (for $\boldsymbol{\theta} = \boldsymbol{\phi}_{-1,-i}'$, and $p = q = 1$)

$$\frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \varepsilon_{1,t-1} \boldsymbol{\phi}_{-1,-i}' \boldsymbol{\upsilon}_{t-1} \xrightarrow{L_1} 0. \tag{A.63}$$

Result (A.27) of Lemma A.5 imply (by setting $p = 1$, and $q = 0$)

$$\frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \varepsilon_{1,t-1} \upsilon_{1t} \xrightarrow{L_1} 0. \tag{A.64}$$

Noting again that $\left\| \boldsymbol{\phi}_{-1,-i} \right\|_\infty = O\left(N^{-1}\right)$, result (A.46) of Lemma A.6 imply (for $i = 1$, $p = 2$, and

$\boldsymbol{\theta} = \boldsymbol{\phi}_{-1,-i}$)

$$E\left(\upsilon_{1,t-2}\boldsymbol{\phi}'_{-1,-i}\boldsymbol{\upsilon}_{t-1}\right) = O\left(N^{-1}\right). \tag{A.65}$$

(A.65) and result (A.24) of Lemma A.5 yields (for $\boldsymbol{\eta} = \mathbf{s}_1$, $\boldsymbol{\theta} = \boldsymbol{\phi}'_{-1,-i}$, $p = 1$, and $q = 2$)

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\upsilon_{1,t-2}\boldsymbol{\phi}'_{-1,-i}\boldsymbol{\upsilon}_{t-1} \xrightarrow{L_1} 0. \tag{A.66}$$

Result (A.47) of Lemma A.6 yields (for $p = 2$ and $i = 1$)

$$E\left(\upsilon_{1,t-2}\upsilon_{1t}\right) = O\left(N^{-1}\right). \tag{A.67}$$

(A.67) and result (A.25) of Lemma A.5 imply (for $\boldsymbol{\eta} = \mathbf{s}_1$, $p = 0$ and $q = 2$)

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\upsilon_{1,t-2}\upsilon_{1t} \xrightarrow{L_1} 0. \tag{A.68}$$

Similarly to (A.66) and (A.68), results (A.24) and (A.25) of Lemma A.5 can be used (for a suitable choice of $\boldsymbol{\eta}$, $\boldsymbol{\theta}$, p and q) to show that

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\upsilon_{i,t-1}\boldsymbol{\phi}'_{-1,-i}\boldsymbol{\upsilon}_{t-1} \xrightarrow{L_1} 0, \tag{A.69}$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}\upsilon_{i,t-1}\upsilon_{1t} \xrightarrow{L_1} 0, \tag{A.70}$$

where we have also used Lemma A.6 (for a suitable choice of p, i and $\boldsymbol{\theta}$), which implies

$$E\left(\upsilon_{i,t-1}\boldsymbol{\phi}'_{-1,-i}\boldsymbol{\upsilon}_{t-1}\right) = O\left(N^{-1}\right), \tag{A.71}$$

and

$$E\left(\upsilon_{i,t-1}\upsilon_{1t}\right) = O\left(N^{-1}\right). \tag{A.72}$$

Substituting equation (A.58) for $x_{i,t-1}$ and definition of ζ_{it} (see (A.60)) in (A.57), and using results (A.63), (A.64), (A.66), (A.68), (A.69) and (A.70) establish

$$E\left|\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}x_{i,t-1}\zeta_{it}\right| \to 0, \tag{A.73}$$

where we have used the fact that the coefficients of the polynomials $f_i(L)$, $d_i(L)c^{-1}(L)$, and $g_i(L)$ are absolute summable. Similarly to proof of result (A.73), Lemma A.5 can be used repeatedly for a suitable choice of p,q, $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ to show that

$$\max_{p\in\{0,1,2,\ldots,m_T\}} E\left|\frac{1}{\sqrt{T}}\sum_{t=m_T+1}^{T}x_{1,t-p}\zeta_{it}\right| \to 0, \tag{A.74}$$

where x_{1t} is given by (40). Results (A.73) and (A.74) complete the proof of (A.55) by noting that convergence in L_1 norm implies convergence in probability. Proof of result (A.56) can be constructed in the same way, but this time Lemma A.4 is used instead of Lemma A.5 and the expansion rates considered for N and T under Assumptions B1 and B2. ■

Lemma A.11 *Suppose* $\mathbf{x}_t$ *is generated according to (1), and Assumptions 1-4, B1 and B3 hold. Then for any* $i > 1$,

$$\left\| \frac{(\mathbf{G}_i - \mathbf{H}_i)' \mathbf{e}_{i\cdot}}{\sqrt{T}} \right\|_\infty \xrightarrow{p} 0, \tag{A.75}$$

where $\mathbf{G}_i$ *and* $\mathbf{H}_i$ *are defined by (75), and (81), respectively. Consider now the case where Assumption B3 is replaced by (weaker) Assumption B2, but the other assumptions are maintained. Then for any* $i > 1$,

$$\left\| \frac{(\mathbf{G}_i - \mathbf{H}_i)' \mathbf{e}_{i\cdot}}{T} \right\|_\infty \xrightarrow{p} 0. \tag{A.76}$$

Proof. Since $|\phi_{ii}| < 1$ by condition (16) of Assumption 4, the polynomial $(1 - \phi_{ii}L)^{-1}$ exists (for any $i = 2, 3, ..., N$). Multiplying equation (A.8) by $(1 - \phi_{ii}L)^{-1}$ yields

$$x_{it} - \xi_{it} = (1 - \phi_{ii}L)^{-1} \left[\beta_i(L)\vartheta_{ct} + \zeta_{it}\right], \text{ for } i = 2, 3, ..., N, \tag{A.77}$$

where ζ_{it} is given by (A.60). Under Assumptions B1 and B3, and using (A.77) and Lemma A.5 (results (A.28) and (A.29)), it can be shown that (using a suitable choice of p, q and vector $\boldsymbol{\theta}$, similarly as in the proof of Lemma A.10) for any $i > 1$ we have

$$\max_{j\in\{1,i\},\ p\in\{1,2,...,m_T\}} E\left| \frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} \left(x_{j,t-p} - \xi_{j,t-p}\right) e_{it} \right| \to 0, \tag{A.78}$$

and

$$E\left| \frac{1}{\sqrt{T}} \sum_{t=m_T+1}^{T} (x_{1t} - \xi_{1t})\, e_{it} \right| \to 0. \tag{A.79}$$

Noting that

$$\mathbf{g}_{it} - \mathbf{h}_{it} = \begin{cases} \left(x_{1,t-1} - \xi_{1,t-1}, x_{1,t-2} - \xi_{1,t-2}, ..., x_{1,t-m_T} - \xi_{1,t-m_T}\right) & \text{for } i = 1 \\ \left(x_{i,t-1} - \xi_{i,t-1}, x_{1t} - \xi_{1t}, x_{1,t-1} - \xi_{1,t-1}, ..., x_{1,t-m_T} - \xi_{1,t-m_T}\right) & \text{for } i > 1 \end{cases},$$

then (A.78)-(A.79) establish (A.75). Proof of (A.76) is identical, but this time Lemma A.4 is used instead of Lemma A.5, together with Assumptions B1 and B2. ■

B Proofs

Proof of Lemma 1.

$$\begin{aligned} Var(\mathbf{a}'\boldsymbol{\upsilon}_t) = \|Var(\mathbf{a}'\boldsymbol{\upsilon}_t)\| = & \left\| \sum_{\ell=0}^{\infty} \mathbf{a}'\boldsymbol{\Phi}_{-1}^{\ell} \mathbf{R}_{-1} Var(\boldsymbol{\varepsilon}_{t-\ell}) \mathbf{R}_{-1}' \boldsymbol{\Phi}_{-1}^{\ell\prime} \mathbf{a} \right\|, \\ \leq & \|\mathbf{a}\|^2 \|\mathbf{R}_{-1}\|^2 \sum_{\ell=0}^{\infty} \|\boldsymbol{\Phi}_{-1}\|^{2\ell} \|Var(\boldsymbol{\varepsilon}_{t-\ell})\|. \end{aligned} \tag{B.1}$$

But $\|\mathbf{R}_{-1}\|^2 \leq \|\mathbf{R}_{-1}\|_\infty \|\mathbf{R}_{-1}\|_1 = O(1)$ by condition (11) of Assumption 2,[17] $\|Var(\boldsymbol{\varepsilon}_{t-\ell})\| < K$ (for any $\ell = 0, 1, 2, ...$) by condition (10) of Assumption 2, $\|\mathbf{a}\|^2 = O(N^{-1})$, $\|\boldsymbol{\Phi}_{-1}\| \leq \sqrt{\|\boldsymbol{\Phi}_{-1}\|_1 \|\boldsymbol{\Phi}_{-1}\|_\infty} \leq \rho$ by

[17] We use the matrix norm inequality $\|A\| \leq \sqrt{\|A\|_1 \|A\|_\infty}$. See Horn and Johnson (1985) for details and other useful matrix inequalities.

condition (16) of Assumption 4 and $\sum_{\ell=0}^{\infty} \|\mathbf{\Phi}_{-1}\|^{2\ell} \leq \sum_{\ell=0}^{\infty} \rho^{2\ell} < K$. Hence, $\|Var(\mathbf{a}'\boldsymbol{\upsilon}_t)\| = O\left(N^{-1}\right)$, as required. ■

Proof of Lemma 2. Coefficients of the polynomial $c(L) = \sum_{\ell=0}^{\infty} c_\ell L^\ell$, as defined by equation (29), satisfy: $c_0 = 1$, and $|c_\ell| = \left|\mathbf{s}_1'\mathbf{\Phi}_{-1}^{\ell-1}\boldsymbol{\phi}_1\right| \leq \left\|\mathbf{\Phi}_{-1}^{\ell-1}\right\|_\infty \|\boldsymbol{\phi}_1\|_\infty$ for any $\ell \in \mathbb{N}$.[18] Conditions (16) and (17) of Assumption 4 postulate that $\|\mathbf{\Phi}_{-1}\|_\infty \leq \rho < 1$ and $\|\boldsymbol{\phi}_1\|_\infty \leq \rho < 1$, which implies that $|c_\ell| \leq \rho^\ell$ for any $\ell \in \mathbb{N}$. The invertibility of $c(L)$ and exponential decay of the coefficients in $c^{-1}(L)$ now directly follow from Lemma A.1. Exponential decay of the coefficients in $c^{-1}(L)$ is uniform in N, because ρ does not depend on $N \in \mathbb{N}$.

The coefficients of the polynomial $b_1(L) = \sum_{\ell=0}^{\infty} b_{1\ell} L^\ell$, as defined by equation (28), satisfy $b_{10} = 1$, and $|b_{1\ell}| = \left|\mathbf{s}_1'\mathbf{\Phi}_{-1}^{\ell}\mathbf{r}_1\right| \leq \left\|\mathbf{\Phi}_{-1}^{\ell}\right\|_\infty \|\mathbf{r}_1\|_\infty$ for any $\ell \in \mathbb{N}$. Conditions (16) and (18) of Assumption 4 imply $\left\|\mathbf{\Phi}_{-1}^{\ell}\right\|_\infty \|\mathbf{r}_1\|_\infty \leq \rho^\ell$, which establishes $|b_{1\ell}| \leq \rho^\ell$ for any $\ell \in \mathbb{N}$. The invertibility of $b_1(L)$ and exponential decay of the coefficients in $b_1^{-1}(L)$ now follows from Lemma A.1. Similarly to $c^{-1}(L)$, the coefficients of $b_1^{-1}(L)$ exponentially decay uniformly in $N \in \mathbb{N}$.

Noting that $|c_\ell| \leq \rho^\ell$ for any $\ell = 0, 1, 2, ..$, and that the coefficients of $b_1^{-1}(L)$ decay exponentially, it follows that the coefficients of $a(L) = b_1^{-1}(L)\, c(L)$ must also decay at an exponential rate. This completes the proof. ■

Proof of Lemma 3. Existence of real positive constants $K < \infty$ and $0 < \rho < 1$ (independent of N) such that $|a_\ell| < K\rho^\ell$ was established in Lemma 2. The coefficients of polynomials $\boldsymbol{\phi}_{-1,-i}'\mathbf{\Phi}_{-1}(L)\,\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_{-1,-i}'\mathbf{\Phi}_{-1}(L)\,\mathbf{r}_1$ satisfy:

$$\left\|\boldsymbol{\phi}_{-1,-i}'\mathbf{\Phi}_{-1}^{\ell}\boldsymbol{\phi}_1\right\|_\infty < K\rho^\ell, \text{ and } \left\|\boldsymbol{\phi}_{-1,-i}'\mathbf{\Phi}_{-1}^{\ell}\mathbf{r}_1\right\|_\infty < K\rho^\ell, \tag{B.2}$$

where $\left\|\boldsymbol{\phi}_{-1,-i}'\right\|_\infty = \sum_{j\neq 1,i} \left|\phi_{ij}\right| < K$ by (6) of Assumption 1, $\left\|\mathbf{\Phi}_{-1}^{\ell}\right\|_\infty \leq \rho^\ell < 1$ by (16) of Assumption 4, $\|\boldsymbol{\phi}_1\|_\infty \leq \rho < 1$ by (17) of Assumption 4, and $\|\mathbf{r}_1\|_\infty = \max_{i=1,..,N} |r_{i1}| \leq 1$ by (18) of Assumption 4. Result (55) now directly follows by noting that linear combinations and products of polynomials with exponentially decaying coefficients are also polynomials with exponentially decaying coefficients. ■

Proof of Lemma 4. Let us examine the polynomial $b_1(L)$ first.

$$\begin{aligned} b_1(L) &= \sum_{\ell=0}^{\infty} \mathbf{s}_1'\mathbf{\Phi}_{-1}^{\ell}\mathbf{r}_1 L^\ell = 1 + \left(\sum_{\ell=0}^{\infty} \boldsymbol{\phi}_{-1}'\mathbf{\Phi}_{-1}^{\ell}\mathbf{r}_1 L^\ell\right) L \\ &= 1 + \boldsymbol{\phi}_{-1}'\mathbf{\Phi}_{-1}(L)\,\mathbf{r}_1 L, \end{aligned}$$

where $\mathbf{s}_1'\mathbf{r}_1 = r_{11} = 1$, and $\mathbf{s}_1'\mathbf{\Phi}_{-1} = \boldsymbol{\phi}_{-1}'$. Under Assumption 4, $\boldsymbol{\phi}_{-1}$ is any vector that satisfies $\left\|\boldsymbol{\phi}_{-1}\right\|_c \leq \rho < 1$. The same condition is assumed to hold for the vector $\boldsymbol{\phi}_{b1}$, and therefore the invertibility of the polynomial $b_1^*(L) = 1 + \boldsymbol{\phi}_{b1}'\mathbf{\Phi}_{-1}(L)\,\mathbf{r}_1 L$ now directly follows from Lemma 2. ■

Proof of Theorem 1. Suppose $i > 1$. Taking maximum absolute row-sum matrix norms of both sides of equation (80), we have

$$\begin{aligned} \|\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\|_\infty \leq\ & \left\|\left(\frac{\mathbf{G}_i'\mathbf{G}_i}{T}\right)^{-1} - \mathbf{C}_i^{-1}\right\|_\infty \left\|\frac{\mathbf{G}_i'\boldsymbol{\epsilon}_{i\cdot}}{T}\right\|_\infty \\ & + \left\|\mathbf{C}_i^{-1}\right\|_\infty \left(\left\|\frac{(\mathbf{G}_i - \mathbf{H}_i)'\,\mathbf{e}_{i\cdot}}{T}\right\|_\infty + \left\|\frac{\mathbf{H}_i'\mathbf{e}_{i\cdot}}{T}\right\|_\infty + \left\|\frac{\mathbf{G}_i'\boldsymbol{\zeta}_{i\cdot}}{T}\right\|_\infty + \left\|\frac{\mathbf{G}_i'\boldsymbol{\psi}_{i\cdot}}{T}\right\|_\infty\right), \end{aligned}$$

[18] We use the submulplicative property of matrix norms, which states that for any matrix norm $\|.\|_M$ and any square matrices $\mathbf{A}$ and $\mathbf{B}$, we have $\|\mathbf{AB}\|_M \leq \|\mathbf{A}\|_M \|\mathbf{B}\|_M$.

where $\left\|\mathbf{C}_i^{-1}\right\|_\infty = O(1)$ since ξ_{it} is a stationary invertible process with absolute summable autocovariances. The desired result (83), for $i > 1$, now follows using Lemmas A.7-A.11 and noting that $\|\mathbf{H}_i'\mathbf{e}_{i\cdot}/T\|_\infty \xrightarrow{p} 0$ by results (A.15) and (A.16) of Lemma A.4. The consistency of $\widehat{\pi}_1$ can be established in a similar manner. ■

Proof of Theorem 2. Suppose $i > 1$.

$$\left\|\sqrt{T}\frac{1}{\sigma_i}\mathbf{a}'\mathbf{C}_i^{\frac{1}{2}}\left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right) - \frac{1}{\sigma_i}\mathbf{a}'\mathbf{C}_i^{-\frac{1}{2}}\frac{\mathbf{H}_i'\mathbf{e}_{i\cdot}}{\sqrt{T}}\right\|_\infty \leq \left\|\frac{1}{\sigma_i}\mathbf{a}'\mathbf{C}_i^{\frac{1}{2}}\right\|_\infty \cdot \cdot\left\|\sqrt{T}\left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right) - \mathbf{C}_i^{-1}\frac{\mathbf{H}_i'\mathbf{e}_{i\cdot}}{\sqrt{T}}\right\|_\infty, \tag{B.3}$$

where $\left\|\frac{1}{\sigma_i}\mathbf{a}'\mathbf{C}_i^{\frac{1}{2}}\right\|_\infty = O(1)$. Using (80) we have

$$\begin{aligned}\left\|\sqrt{T}\left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right) - \mathbf{C}_i^{-1}\frac{\mathbf{H}_i'\mathbf{e}_{i\cdot}}{\sqrt{T}}\right\|_\infty \leq & \left\|\left(\frac{\mathbf{G}_i'\mathbf{G}_i}{T}\right)^{-1} - \mathbf{C}_i^{-1}\right\|_\infty \left\|\frac{\mathbf{G}_i'\boldsymbol{\epsilon}_{i\cdot}}{\sqrt{T}}\right\|_\infty \\ & + \left\|\mathbf{C}_i^{-1}\right\|_\infty \left(\left\|\frac{\left(\mathbf{G}_i - \mathbf{H}_i\right)'\mathbf{e}_{i\cdot}}{\sqrt{T}}\right\|_\infty + \left\|\frac{\mathbf{G}_i'\boldsymbol{\zeta}_{i\cdot}}{\sqrt{T}}\right\|_\infty\right) \\ & + \left\|\mathbf{C}_i^{-1}\right\|_\infty \left\|\frac{\mathbf{G}_i'\boldsymbol{\psi}_{i\cdot}}{\sqrt{T}}\right\|_\infty \\ & \xrightarrow{p} 0,\end{aligned} \tag{B.4}$$

where the convergence follows from Lemmas A.7-A.11. Furthermore,

$$\frac{1}{\sigma_i}\mathbf{a}'\mathbf{C}_i^{-\frac{1}{2}}\frac{\mathbf{H}_i'\mathbf{e}_{i\cdot}}{\sqrt{T}} \xrightarrow{d} N(0,1) \tag{B.5}$$

is a standard time series result, which can be established using the martingale difference array central limit theorem (Theorem 24.3 of Davidson (1994)) in the same way as Lemma 6 of Chudik and Pesaran (2011). Equations (B.3)-(B.5) establish result (84), as desired. The asymptotic distribution of $\widehat{\pi}_1$ can be established in a similar manner. ■

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics 2*, 489–502.

Bussiere, M., A. Chudik, and A. Mehl (2011). Does the euro make a difference? Spatio-temporal transmission of global shocks to real effective exchange rates in an infinite VAR. ECB working paper No. 1292.

Chudik, A. and M. Fratzscher (2011). Identifying the global transmission of the 2007-09 financial crisis in a GVAR model. *European Economic Review 55*, 325–339.

Chudik, A. and M. H. Pesaran (2011). Infinite dimensional VARs and factor models. *forthcomming in the Journal of Econometrics*.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *Econometrics Journal 14*, C45–C90.

Cliff, A. and J. Ord (1973). *Spatial autocorrelation.* London: Pion.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press.

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projections using realistic prior distributions. *Econometric Reviews 3*, 1–100.

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

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004). The generalized dynamic factor model: Consistency and rates. *Journal of Econometrics 119*, 231–235.

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

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. Aigner and A. Goldberger (Eds.), *Latent variables in socio-economic models.* Amsterdam: North-Holland.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning.* Springer Series in Statistics.

Holly, S., M. H. Pesaran, and T. Yamagata (2011). Spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics 69*, 2–23.

Horn, R. A. and C. A. Johnson (1985). *Matrix Analysis.* Cambridge University Press.

Jackson, M. O. (2008). *Social and Economic Networks.* Princeton University Press.

Kelejian, H. and D. Robinson (1995). Spatial correlation: A suggested alternative to the autoregressive model. In L. Anselin and R. Florax (Eds.), *New directions in spatial econometrics*, pp. 75–95. Berlin: Springer-Verlag.

Pesaran, M. H. (2006). Estimation and inference in large heterogenous panels with multifactor error structure. *Econometrica 74*, 967–1012.

Pesaran, M. H. and A. Chudik (2011). Aggregation in large dynamic panels. IZA discussion paper No. 5478.

Pesaran, M. H., T. Schuermann, and S. Weiner (2004). Modelling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business and Economics Statistics 22*, 129–162.

Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlations. *Journal of Econometrics 161*, 182–202.

Said, E. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika 71*, 599–607.

Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a-priori economic theory. In C. Sims (Ed.), *New methods in business cycle research.* Minneapolis: Federal Reserve Bank of Minneapolis.

Whittle, P. (1954). On stationary processes on the plane. *Biometrika 41*, 434–449.