Infinite-Dimensional VARs and Factor Models*

Alexander Chudik†
European Central Bank and CIMF

M. Hashem Pesaran‡
Cambridge University, CIMF and USC

January 22, 2010

Abstract

This paper proposes a novel approach for dealing with the 'curse of dimensionality' in the case of vector autoregressive (VAR) models with a large number of variables or units (N). It is assumed that each unit is related to a small number of neighbors and a large number of non-neighbors. The neighbors could be individual units or, more generally, linear combinations of units. The neighborhood effects are fixed and do not change with N, but the coefficients corresponding to the non-neighboring units are restricted to vanish in the limit as N tends to infinity. The conditions under which such an infinite-dimensional VAR (or IVAR) can be arbitrarily well characterized by a large number of finite-dimensional models are derived. Problems of estimation and inference in a stationary IVAR model with an unknown number of unobserved common factors are also investigated. A cross section augmented least squares (CALS) estimator is proposed and its asymptotic distribution is derived. Satisfactory small sample properties for the CALS estimator are documented by Monte Carlo experiments. An empirical illustration shows the statistical significance of dynamic spill-over effects in modelling of U.S. real house prices across the neighboring States.

Keywords: Large N and T Panels, Weak and Strong Cross Section Dependence, VARs, Spatial Models, Factor Models.

JEL Classification: C10, C33, C51

*We are grateful to Elisa Tosetti for helpful comments and suggestions. This version has also benefited from comments by the seminar participants at the University of California, San Diego, University of Southern California, Columbia, Leicester, European University Institute, McGill, Princeton, Pennsylvania, Stanford; as well as by the conference participants in the Nordic Econometric Meeting held at Lund University, and in the conference on the Factor Structures for Panels and Multivariate Time Series Data held at Maastricht University. We would also like to acknowledge Takashi Yamagata for carrying out the computation of the results reported in Section 6.

†European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany. E-mail: alexander.chudik@ecb.int.

‡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

Vector autoregressive models (VARs) provide a flexible framework for the analysis of complex dynamics and interactions that exist between economic variables. The traditional VAR modelling strategy postulates that the number of variables, N, is fixed and the time dimension, T, tends to infinity. But since the number of parameters to be estimated grows at a quadratic rate with N, the application of the approach in practice is often limited to a handful of variables. The objective of this paper is to consider VAR models where *both* N and T are large. In this case, parameters of the VAR model can no longer be consistently estimated unless suitable restrictions are imposed to overcome the dimensionality problem.

Two different approaches have been suggested in the literature to deal with this 'curse of dimensionality': (i) shrinkage of the parameter space, and (ii) shrinkage of the data. Spatial and/or spatiotemporal literature shrinks the parameter space by using *a priori* given spatial weights matrices that restricts the nature of the links across the units. Alternatively, prior probability distributions are imposed on the parameters of the VAR such as the 'Minnesota' priors proposed by Doan, Litterman, and Sims (1984). This class of models is known as Bayesian VARs (BVAR).[1]

The second approach is to shrink the data, along the lines of index models. Geweke (1977) and Sargent and Sims (1977) introduced dynamic factor models, which have more recently been generalized to allow for weak cross sectional dependence by Forni and Lippi (2001), Forni et al. (2000) and Forni et al. (2004). Empirical evidence suggests that few dynamic factors are needed to explain the co-movements of macroeconomic variables.[2] This has led to the development of factor-augmented VAR (FAVAR) models by Bernanke, Bovian, and Eliasz (2005) and Stock and Watson (2005), among others.

Applied researchers are often forced to impose arbitrary restrictions on the coefficients that link the variables of a given cross section unit to the current and lagged values of the remaining units, mostly because they realize that without such restrictions the model is not estimable. This paper

[1] Other types of priors have also been considered in the literature. See, for example, Del Negro and Schorfheide (2004) for a recent reference. In most applications, BVARs have been applied to relatively small systems (e.g. Leeper, Sims, and Zha (1996) considered 13- and 18-variable BVAR; a few exceptions include Giacomini and White (2006) and De Mol, Giannone, and Reichlin (2008)), with the focus being mainly on forecasting. Bayesian VARs are known to produce better forecasts than unrestricted VARs or structural models. See Litterman (1986) and Canova (1995) for further references.

[2] Stock and Watson (1999), Stock and Watson (2002), Giannone, Reichlin, and Sala (2005) conclude that only few, perhaps two, factors explain much of the predictable variations, while Bai and Ng (2007) estimate four factors and Stock and Watson (2005) estimate as much as seven factors.

proposes a novel way to deal with the curse of dimensionality by shrinking part of the parameter space in *the limit* as the number of variables (N) tends to infinity. An important example would be a VAR model where each unit is related to a small number of neighbors and a large number of non-neighbors. The neighbors could be individual units or, more generally, linear combinations of units (spatial averages). The neighborhood effects are fixed and do not change with N, but the coefficients corresponding to the remaining non-neighbor units are small, of order $O\left(N^{-1}\right)$. Such neighborhood and non-neighborhood effects could be motivated by theoretical economic considerations, or could arise due to the mis-specification of spatial weights.

Although under this set-up each of the non-neighboring coefficients is small, sum of their absolute values in general does not tend to zero and the aggregate spatiotemporal non-neighborhood effects could be large. This paper shows that under weak cross section dependence, the spillover effects from non-neighboring units are neither particularly important, nor estimable.[3] But the coefficients associated with the neighboring units can be consistently estimated by simply ignoring the non-neighborhood effects that are of second order importance in N. On the other hand, if the units are cross sectionally strongly dependent, then the spillover effects from non-neighbors are in general important, and ignoring such effects can lead to inconsistent estimates.

Another model of interest arises when in addition to the neighborhood effects, there is also a fixed number of dominant units that have non-negligible effects on all other units. In this case the limiting outcome is shown to be a dynamic factor model.[4] Accordingly, the paper provides a link between data and parameter shrinkage approaches to mitigating the curse of dimensionality. By imposing limiting restrictions on some of the parameters of the VAR we effectively end up with a data shrinkage. To distinguish high dimensional VAR models from the standard specifications we refer to the former as the infinite dimensional VARs or IVARs for short.

The paper also establishes the conditions under which the Global VAR (GVAR) approach proposed by Pesaran, Schuermann, and Weiner (2004) is applicable.[5] In particular, the IVAR featuring all macroeconomic variables could be arbitrarily well approximated by a set of finite-

[3] Concepts of strong and weak cross section dependence, introduced in Chudik, Pesaran, and Tosetti (2009), will be applied to VAR models.

[4] The case of IVAR models with a dominant unit is studied in Pesaran and Chudik (2010).

[5] GVAR model has been used to analyse credit risk in Pesaran, Schuermann, Treutler, and Weiner (2006) and Pesaran, Schuermann, and Treutler (2007). Extended and updated version of the GVAR by Dées, di Mauro, Pesaran, and Smith (2007), which treats Euro area as a single economic area, was used by Pesaran, Smith, and Smith (2007) to evaluate UK entry into the Euro. Global dominance of the US economy in a GVAR model is considered in Chudik (2008). Further developments of a global modelling approach are provided in Pesaran and Smith (2006). Garratt, Lee, Pesaran, and Shin (2006) provide a textbook treatment of GVAR.

dimensional small-scale models that can be consistently estimated separately in the spirit of the GVAR.

A second contribution of the paper is the development of appropriate econometric techniques for estimation and inference in stationary IVAR models with an unknown number of unobserved common factors. This extends the analysis of Pesaran (2006) to dynamic models where all variables are determined endogenously. A simple cross sectional augmented least-squares estimator (or CALS for short) is proposed and its asymptotic distribution derived. Small sample properties of the proposed estimator are investigated through Monte Carlo experiments. As an illustration of the proposed approach we consider an extension of the empirical analysis of real house prices across the 49 U.S. States conducted recently by Holly, Pesaran, and Yamagata (2009), and show statistically significant *dynamic* spillover effects of real house prices across the neighboring States.

The remainder of the paper is organized as follows. Section 2 introduces the IVAR model. Section 3 investigates cross section dependence in IVAR models. Section 4 focusses on estimation of a stationary IVAR model. Section 5 discusses the results of the Monte Carlo experiments, and Section 6 presents the empirical results. The final section offers some concluding remarks. Proofs are provided in the Appendix.

A brief word on notations: $|\lambda_1(\mathbf{A})| \geq |\lambda_2(\mathbf{A})| \geq ... \geq |\lambda_n(\mathbf{A})|$ are the eigenvalues of $\mathbf{A} \in \mathbb{M}^{n\times n}$, where $\mathbb{M}^{n\times n}$ is the space of real-valued $n \times n$ matrices. $\|\mathbf{A}\|_1 \equiv \max\limits_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|$ denotes the maximum absolute column sum matrix norm of $\mathbf{A}$, $\|\mathbf{A}\|_\infty \equiv \max\limits_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$ is the absolute row sum matrix norm of $\mathbf{A}$. $\|\mathbf{A}\| = \sqrt{\varrho(\mathbf{A}'\mathbf{A})}$ is the spectral norm of $\mathbf{A}$, $\varrho(\mathbf{A}) \equiv \max\limits_{1\leq i\leq n} \{|\lambda_i(\mathbf{A})|\}$ is the spectral radius of $\mathbf{A}$.[6] All vectors are column vectors and the i^{th} row of $\mathbf{A}$ is denoted by $\mathbf{a}_i'$. $a_n = O(b_n)$ denotes the deterministic sequence $\{a_n\}$ is at most of order b_n. $x_n = O_p(y_n)$ states that the random variable x_n is at most of order y_n in probability. $\mathbb{N}$ is the set of natural numbers, and $\mathbb{Z}$ is the set of integers. We use K and ϵ to denote positive fixed constants that do not vary with N or T. Convergence in distribution and convergence in probability is denoted by $\xrightarrow{d}$ and $\xrightarrow{p}$, respectively. Symbol $\xrightarrow{q.m.}$ represents convergence in quadratic mean. $(N,T) \xrightarrow{j} \infty$ denotes joint asymptotic in N and T, with N and $T \to \infty$, in no particular order.

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

2 Infinite-Dimensional Vector Autoregressive Models

Suppose we have T time series observations on N cross section units indexed by $i \in \mathcal{S}_{(N)} \equiv \{1,..,N\} \subseteq \mathbb{N}$. Individual units could be households, firms, regions, or countries. 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 $N \times 1$ vector $\mathbf{x}_{(N),t} = \left(x_{(N),1t},...,x_{(N),Nt}\right)'$, and it is assumed that $\mathbf{x}_{(N),t}$ follows the VAR(1) model

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

$$\mathbf{u}_{(N),t} = \mathbf{R}_{(N)}\boldsymbol{\varepsilon}_{(N),t}. \tag{2}$$

$\mathbf{\Phi}_{(N)}$ and $\mathbf{R}_{(N)}$ are $N \times N$ coefficient matrices that capture the dynamic and contemporaneous dependencies across the N units, and $\boldsymbol{\varepsilon}_{(N),t} = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$ is an $N \times 1$ vector of white noise errors with mean $\mathbf{0}$ and the covariance matrix, $\mathbf{I}_N$.

VAR models have been extensively studied when N is small and fixed, and T is large and unbounded. This framework, however, is not appropriate for many empirical applications of interest. This paper aims to fill this gap by analyzing VAR models where both N and T are large. The sequence of models (1) and (2) with $\dim(\mathbf{x}_{(N),t}) = N \to \infty$ will be referred to as the infinite-dimensional VAR model, or IVAR for short. The extension of the IVAR(1) to IVAR(p) where p is fixed, is relatively straightforward and will not be attempted in this paper.

The analysis of dependence over time is simplified by the fact that ordering of observations along the time dimension ($t = 1, 2, ..., T$) is immutable and the arrival of new observations *cannot* change past realizations, namely bygones are bygones. As a consequence for any given N, i, and j, the cross time covariance function, $cov\left(x_{(N),it}, x_{(N),j,t-\ell}\right)$, does not change with T and will depend only on ℓ if the time series processes are covariance stationary. However, since it can not be assumed that an immutable ordering necessarily exists with respect to the cross section dimension, addition of new cross section units to an existing set can potentially alter the pair-wise cross section covariances of all the units. For instance in models of oligopoly, where firms strategically interact with each other, new entries can change the relationship between the existing firms. Similarly, introduction of a new asset in the market can change the correlation of returns on the existing assets.

In what follows, to simplify the notations, the explicit dependence of $\mathbf{x}_t$ and $\mathbf{u}_t$ and the related parameter matrices on N will be suppressed with (1)-(2) written as

$$\mathbf{x}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{u}_t, \tag{3}$$

and

$$\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t. \tag{4}$$

Clearly, it is not possible to estimate all the N^2 elements of the matrix $\mathbf{\Phi}$ when both N and T are large. Only a small (fixed) number of unknown coefficients can be estimated per equation and some restrictions on $\mathbf{\Phi}$ must be imposed.

In order to deal with the dimensionality problem, we assume that for a given $i \in \mathbb{N}$, it is possible to classify cross section units *a priori* into 'neighbors' and 'non-neighbors'. No restrictions are imposed on neighbors, but the non-neighbors are assumed to have only negligible effects on x_{it} that vanish at a suitable rate with N. The number of neighbors of unit i, collected in the index set $\mathcal{N}_i$, is assumed to be small (fixed). Neighbors of unit i can have non-negligible effects that do not vanish even if $N \to \infty$. A similar classification is followed in the spatial econometrics literature, where the non-neighborhood effects are set to zero for all N and the non-zero neighborhood effects are often assumed to be homogenous across i. In this sense our analysis can also be seen as an extension of spatial econometric models.

Subject to the above classification, equation for the unit i can be written as

$$x_{it} = \underbrace{\sum_{j \in \mathcal{N}_i} \phi_{ij} x_{j,t-1}}_{Neighbors} + \underbrace{\sum_{j \in \mathcal{N}_i^c} \phi_{ij} x_{j,t-1}}_{Non-neighbors} + u_{it}. \tag{5}$$

The coefficients of the neighboring units, $\left\{\phi_{ij}\right\}_{j\in\mathcal{N}_i}$, are the parameters of interest and do not vary with N. The remaining coefficients, $\left\{\phi_{ij}\right\}_{j\in\mathcal{N}_i^c}$, tend to zero for each i as $N \to \infty$, where $\mathcal{N}_i^c \equiv \{1, .., N\} \setminus \mathcal{N}_i$ is the index set of non-neighbors. Note that the non-neighbors are unordered. More specifically,

$$\left|\phi_{ij}\right| \leq \frac{K}{N} \text{ for any } N \in \mathbb{N} \text{ and any } j \in \mathcal{N}_i^c. \tag{6}$$

Individually the coefficients of non-neighbors are asymptotically negligible, but as we argue below it is not clear if the same applies to their aggregate effects on the i^{th} unit, namely $\sum_{j\in\mathcal{N}_i^c} \phi_{ij} x_{j,t-1}$.

The bounds in (6) ensures that $\lim_{N\to\infty}\sum_{j=1}^{N}\left|\phi_{ij}\right| < K$. We refer to this as the 'cross section absolute summability condition', which is distinct from the absolute summability condition used in the time series literature where the same idea is applied to the coefficients of current and past innovations. A similar constraint is used in Lasso and Ridge regression shrinkage methods. The Lasso estimation procedure applied to (3) involves minimizing $\sum_{t=1}^{T}u_{it}^{2}$ for each i subject to $\sum_{j=1}^{N}\left|\phi_{ij}\right| \leq K$. Under the Ridge regression the minimization is carried out subject to the weaker constraint, $\sum_{j=1}^{N}\phi_{ij}^{2} \leq K$.[7] In application of shrinkage methods it is necessary that K is specified *a priori,* but no knowledge of the ordering of the units along the cross section dimension is needed. In our approach we do not need to specify the value of K.

Sum of the coefficients of the non-neighboring units, $\sum_{j\in\mathcal{N}_i^c}\phi_{ij}$, does not necessarily tends to zero as $N \to \infty$, which implies that the non-neighbors can have a large aggregate spatiotemporal impact on the unit i, as $N \to \infty$. The question that we address is whether it is possible to estimate neighborhood coefficients $\left\{\phi_{ij}\right\}_{j\in\mathcal{N}_i}$ without imposing further restrictions. As it turns out, the answer depends on the stochastic behavior of $\sum_{j\in\mathcal{N}_i^c}\phi_{ij}x_{j,t-1}$, which in turn depends on the strength of cross section dependence in $\{x_{it}\}$. If $\{x_{it}\}$ is weakly cross sectionally dependent then $\sum_{j\in\mathcal{N}_i^c}\phi_{ij}x_{j,t-1} \overset{q.m.}{\to} 0$, and the spillover effects from non-neighboring units are neither particularly important nor estimable. But the coefficients associated with the neighboring units can be consistently estimated by simply ignoring the non-neighborhood effects that are of second order importance in N. If on the other hand $\{x_{it}\}$ is strongly cross sectionally dependent, then $\lim_{N\to\infty}Var\left(\sum_{j\in\mathcal{N}_i^c}\phi_{ij}x_{j,t-1}\right)$ is not necessarily zero, and the spillover effects from non-neighbors are in general $O_p(1)$ and important. Therefore, ignoring the non-neighborhood effects can lead to inconsistent estimates. The concepts of weak and strong cross section dependence have been introduced in Chudik, Pesaran, and Tosetti (2009) and these concepts are applied to the IVAR model in the next section.

Our approach to dealing with the curse of dimensionality can be motivated with several examples. In economic applications interactions across agents often depends on the number of agents, with the degree of pair-wise interactions typically declining in the number of units. Consider, for example, the output and pricing decisions of firms in an industry with N firms. When N is small (cases of duopoly or oligopoly) pricing and output decisions are inter-related through the way firms

[7]See Section 3.4.3 of Hastie, Tibshirani, and Friedman (2001) for detailed description of the Lasso and Ridge regression shrinkage methods.

form expectations about the reactions of other firms, known as *conjectural variations.* But as N becomes large such conjectural variations become relatively unimportant and in the competitive case where N is sufficiently large conjectural variations are typically set to zero. Another important example is provided by the Arbitrage Pricing Theory (APT) originally developed by Ross (1976). Under approximate pricing the conditional mean returns of N risky assets, $\boldsymbol{\mu}_t$, is modelled in terms of a fixed number (k) of factor risk premia, $\boldsymbol{\lambda}_t$, and an $N \times 1$ vector of pricing errors, $\mathbf{v}_t$, namely

$$\boldsymbol{\mu}_t = \mathbf{B}\boldsymbol{\lambda}_t + \mathbf{v}_t,$$

where $\mathbf{B}$ is an $N \times k$ matrix of factor loadings. In the absence of arbitrage opportunities we must have $\mathbf{v}_t = 0$ when N is fixed, or $\mathbf{v}_t'\mathbf{v}_t = O_p(1)$ as $N \to \infty$. (see Hubermann (1982) and Ingersoll (1984)). It is clear that any pair-wise dependence of pricing errors must vanish as $N \to \infty$, otherwise there will be unbounded profitable opportunities. The third example relates to a multi-country DSGE model discussed in Chudik (2008). The country interactions need not be symmetric. Nevertheless, as long as foreign trade weights are granular, the equilibrium solution of such a multi-country DSGE model has a similar structure to the basic IVAR model set out in the paper. Neighbors in this setup could be identified in terms of the trade shares, for example. For instance, US would be Canada's neighbor considering that 80% of Canada's trade is with the US, although using the same metric Canada might not qualify as a neighbor of the US.

In some cases the strict division of individual units into neighbors and non-neighbors might be considered as too restrictive. In the assumption below we consider a slightly more general set up where the neighborhood effects are charachterized in terms of 'local' averages defined by $\mathbf{S}_i'\mathbf{x}_t$, where $\mathbf{S}_i$ is a known spatial or neighborhood weight matrix.

ASSUMPTION 1 *Let $\mathcal{K} \subseteq \mathbb{N}$ be a non-empty index set. For any $i \in \mathcal{K}$, the row i of coefficient matrix $\boldsymbol{\Phi}$, denoted by $\boldsymbol{\phi}_i'$, can be divided as*

$$\boldsymbol{\phi}_i' = \boldsymbol{\phi}_{ai}' + \boldsymbol{\phi}_{bi}', \tag{7}$$

where

$$\|\boldsymbol{\phi}_{bi}\|_\infty = \max_{j \in \{1,..,N\}} |\phi_{bij}| < \frac{K}{N}, \tag{8}$$

$$\phi_{ai} = \mathbf{S}_i \boldsymbol{\delta}_i, \tag{9}$$

$\|\boldsymbol{\delta}_i\| < K$, $\boldsymbol{\delta}_i$ is an $h_i \times 1$ dimensional vector containing the unknown coefficients to be estimated for unit i, which do not change with N, $h_i < K$, h_i is fixed and generally small, and $\mathbf{S}_i$ is a known $N \times h_i$'spatial' weight matrix such that $\|\mathbf{S}_i\|_1 < K$.

Assuming $\mathcal{K} \equiv \mathbb{N}$ and stacking (7)-(9) for $i = 1, 2, ..., N$, we have

$$\begin{aligned} \mathbf{\Phi} &= \mathbf{\Phi}_a + \mathbf{\Phi}_b, \\ &= \mathbf{D}\ \mathbf{S} + \mathbf{\Phi}_b, \end{aligned} \tag{10}$$

where $\mathbf{\Phi}_a = (\phi_{a1}, \phi_{a2}, ..., \phi_{aN})'$, $\mathbf{\Phi}_b = (\phi_{b1}, \phi_{b2}, ..., \phi_{bN})'$,

$$\underset{N\times h}{\mathbf{D}} = \begin{pmatrix} \boldsymbol{\delta}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\delta}_2' & & \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \boldsymbol{\delta}_N' \end{pmatrix}, \tag{11}$$

$h = \sum_{i=1}^{N} h_i$, and $\mathbf{S}$ is a known $h \times N$ matrix defined by $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2, ..., \mathbf{S}_N)'$. Note also that by assumption the individual elements of $\mathbf{\Phi}_b$ are uniformly $O\left(N^{-1}\right)$.

Example 1 *An example of $\mathbf{\Phi}_a$ is given by*

$$\mathbf{\Phi}_a = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 & 0 & \ldots & 0 & 0 & 0 \\ \phi_{21} & \phi_{22} & \phi_{23} & 0 & \ldots & 0 & 0 & 0 \\ 0 & \phi_{32} & \phi_{33} & \phi_{34} & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \phi_{N-1,N-2} & \phi_{N-1,N-1} & \phi_{N-1,N} \\ 0 & 0 & 0 & 0 & \ldots & 0 & \phi_{N-1,N} & \phi_{NN} \end{pmatrix}, \tag{12}$$

where the nonzero elements are fixed coefficients that do not change with N. This represents a bilateral spatial representation where each unit, except for the first and the last units, has one left and one right neighbor. In contrast the individual elements of $\mathbf{\Phi}_b$ are of order $O(N^{-1})$, in particular $\left|\phi_{bij}\right| < \frac{K}{N}$ for any $N \in \mathbb{N}$ and any $i, j \in \{1, .., N\}$. The equation for unit $i \in \{2, .., N-1\}$ can be *written as*

$$x_{it} = \phi_{i,i-1}x_{i-1,t-1} + \phi_{ii}x_{i,t-1} + \phi_{i,i+1}x_{i+1,t-1} + \boldsymbol{\phi}'_{bi}\mathbf{x}_{t-1} + u_{it}. \tag{13}$$

Section 3 shows that under weak cross section dependence of errors $\{u_{it}\}$*,* $\boldsymbol{\phi}'_{bi}\mathbf{x}_{t-1} \overset{q.m.}{\to} 0$*, while Section 4 considers problem of estimation of the individual-specific parameters* $\left\{\phi_{i,i-1}, \phi_{ii}, \phi_{i,i+1}\right\}$*. We refer to this model as a two-neighbor IVAR model which we use later for illustrative purposes as well as in the Monte Carlo experiments.*

Example 2 *As a simple example consider the model*

$$\mathbf{x}_t = \rho_x \mathbf{S}_x \mathbf{x}_{t-1} + \mathbf{u}_t, \tag{14}$$

$$\mathbf{u}_t = \rho_u \mathbf{S}_u \mathbf{u}_t + \boldsymbol{\varepsilon}_t, \tag{15}$$

where ρ_x *and* ρ_u *are scalar unknown coefficients, and* $\mathbf{S}_x$ *and* $\mathbf{S}_u$ *are* $N \times N$ *known spatial weights matrices. This model can be obtained from (1)-(2) by setting*

$$\mathbf{R} = \left(\mathbf{I} - \rho_u \mathbf{S}_u\right)^{-1}, \ \delta_i = \rho_x \text{ for } i \in \{1, .., N\}, \ \mathbf{S} = \mathbf{S}_x, \text{ and } \boldsymbol{\Phi}_b = \mathbf{0}.$$

3 Cross Sectional Dependence in Stationary IVAR Models

This section investigates the correlation pattern of $\{x_{it}\}$, over time, t, and along the cross section units, i. We follow Chudik, Pesaran, and Tosetti (2009) and define covariance stationary process $\{x_{it}\}$ to be cross sectionally weakly dependent (CWD), if for all weight vectors, $\mathbf{w} = (w_1, ..., w_N)'$, satisfying the 'granularity' conditions

$$\|\mathbf{w}\| = O\left(N^{-\frac{1}{2}}\right), \tag{16}$$

$$\frac{w_j}{\|\mathbf{w}\|} = O\left(N^{-\frac{1}{2}}\right) \text{ for any } j, \tag{17}$$

we have

$$\lim_{N\to\infty} Var\left(\mathbf{w}'\mathbf{x}_t\right) = 0, \text{ for all } t.$$

$\{x_{it}\}$ is said to be cross sectionally strongly dependent (CSD) if there exists a sequence of weight vectors, $\mathbf{w}$, satisfying (16)-(17) and a constant K such that

$$\lim_{N\to\infty} Var\left(\mathbf{w}'\mathbf{x}_t\right) \geq K > 0. \tag{18}$$

Necessary condition for covariance stationarity for fixed N is that all eigenvalues of $\mathbf{\Phi}$ lie inside of the unit circle. For a fixed N, and assuming that $\max_i |\lambda_i(\mathbf{\Phi})| < 1$, the Euclidean norm of $\mathbf{\Phi}^\ell$ defined by $\left[Tr\left(\mathbf{\Phi}^\ell\mathbf{\Phi}^{\ell\prime}\right)\right]^{1/2} \to 0$ exponentially in ℓ, and the process $\mathbf{x}_t = \sum_{\ell=0}^{\infty}\mathbf{\Phi}^\ell\mathbf{u}_{t-\ell}$ will be absolute summable, in the sense that the sum of absolute values of the elements of $\mathbf{\Phi}^\ell$, for $\ell = 0, 1, ...,$ converges. Observe that as $N \to \infty$, $Var(x_{it})$ need not necessarily be bounded in N even if $\max_i |\lambda_i(\mathbf{\Phi})| < 1$ (and bounded away from 1). For example, consider the IVAR(1) model with

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

and assume that $var(u_{it})$ is uniformly bounded away from zero as $N \to \infty$. It is clear that all the eigenvalues of $\mathbf{\Phi}$ are inside the unit circle if and only if $|\varphi| < 1$, regardless of the value of the neighborhood coefficient, ψ. Yet the variance of x_{Nt} increases in N without bounds at an exponential rate for $|\psi| > 1 - |\varphi|$.[8] Therefore, a stronger condition than stationarity for each N is required to prevent the variance of x_{it} from exploding as $N \to \infty$. This is set out in the following assumptions.

ASSUMPTION 2 *The elements of the double index process* $\{\varepsilon_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ *are independently distributed random variables with zero means and unit variances on the probability space* $(\Omega, \mathcal{F}, P)$.

ASSUMPTION 3 *(CWD Errors) Matrix* $\mathbf{R}$ *has bounded row and column matrix norms.*

[8] It can be shown that

$$Var\{x_{Nt}\} = \sum_{j=1}^{N}\psi^{2(N-j)}\sum_{\ell=0}^{\infty}\alpha^2_{N-j+1,\ell}\varphi^{2\ell},$$

where $\alpha_{k\ell} = \frac{1}{(k-1)!}\prod_{j=0}^{k-2}(\ell + k - 1 - j)$ for $k > 1$ and $\alpha_{1\ell} = 1$.

ASSUMPTION 4 *(Stationarity and bounded variances) There exists a real ϵ, in the range $0 < \epsilon < 1$, such that*[9]

$$\|\mathbf{\Phi}\| \leq 1 - \epsilon. \tag{19}$$

Remark 1 *Assumptions 2 and 3 imply $\{u_{it}\}$ is CWD, since for any weight vector, w, satisfying (16) we have $Var(\mathbf{w}'\mathbf{u}_t) \leq \|\mathbf{w}\|^2 \|\mathbf{R}\|_1 \|\mathbf{R}\|_\infty \to 0$ as $N \to \infty$. For future reference define covariance matrix $\mathbf{\Sigma} = Var(\mathbf{u}_t) = \mathbf{RR}'$ and denote the i^{th} diagonal element of $\mathbf{\Sigma}$ by $\sigma_{ii}^2 = Var(u_{it})$. Note also that $\|\mathbf{\Sigma}\| \leq \|\mathbf{R}\|_1 \|\mathbf{R}\|_\infty < K$, which as shown in Pesaran and Tosetti (2009) includes all commonly used processes in the spatial literature, such as spatial autoregressive and spatial error component models pioneered by Whittle (1954), and further developed by Cliff and Ord (1973), Anselin (1988), and Kelejian and Robinson (1995).*

Remark 2 *It is not necessary that proximity is measured in terms of physical space. Other measures such as economic (Conley (1999), Pesaran, Schuermann, and Weiner (2004)), or social distance (Conley and Topa (2002)) could also be employed. All these are examples of dependence across nodes in a physical (real) or logical (virtual) networks. In the case of the IVAR model, defined by (3) and (4), such contemporaneous dependence can be modelled through the $N \times N$ network topology matrix $\mathbf{R}$.*[10,11]

Remark 3 *The IVAR model when combined with $\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t$ yields an infinite-dimensional spatiotemporal model. The model can also be viewed more generally as a 'dynamic network', with $\mathbf{R}$ and $\mathbf{\Phi}$ capturing the static and dynamic forms of inter-connections that might exist in the network.*

Remark 4 *(Eigenvalues of $\mathbf{\Phi}$) Assumption 4 implies polynomial $\mathbf{\Phi}(L)$ is invertible (for any $N \in \mathbb{N}$) and*

$$\varrho(\mathbf{\Phi}) \leq 1 - \epsilon, \tag{20}$$

which is a sufficient condition for covariance stationarity. Assumption 4 also delivers a bounded variance for x_{it}, as $N \to \infty$.

[9] Our assumptions concerning coefficient matrix $\mathbf{\Phi}$ can be relaxed so long as they hold for all $N \geq N_0$ (where N_0 is a fixed constant that does not depend on N). But in order to keep notations and exposition simple, we simply state that Assumptions 1 and 4 hold for any value of N.

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

[11] It is also possible to allow for time variations in the network matrix, $\mathbf{R}$, to capture changes in the network structure over time. However, this will not be pursued here.

Proposition 1 *Consider model (1) and suppose that Assumptions 2-4 hold. Then for any arbitrary sequence of fixed weights* $\mathbf{w}$ *satisfying condition (16), and for any* $t \in \mathbb{Z}$*,*

$$\lim_{N \to \infty} Var\left(\overline{x}_{wt}\right) = 0. \tag{21}$$

Assumptions 2-4 are thus sufficient conditions for weak dependence. Proposition 1 has several interesting implications. Suppose that we can impose limiting restrictions given by Assumption 1.

Corollary 1 *Consider model (1) and suppose Assumptions 1-4 hold. Then, for any* $i \in \mathcal{K}$*,*

$$\lim_{N \to \infty} Var\left(x_{it} - \boldsymbol{\phi}_{ai}' \mathbf{x}_{t-1} - u_{it}\right) = 0. \tag{22}$$

Remark 5 *It is also possible to establish (22) under the following conditions:*

$$\|\boldsymbol{\phi}_{bi}\| = O\left(N^{-\frac{1}{2}}\right), \tag{23}$$

$$\|\boldsymbol{\Sigma}\| = O\left(N^{1-\epsilon}\right), \tag{24}$$

which are less restrictive than condition (8) and the Assumption 3 on the boundedness of the column and row norms of matrix $\mathbf{R}$*. These stronger conditions are needed for establishing the asymptotic properties of the CALS estimator to be proposed below in Section 4.*

3.1 IVAR Models with Strong Cross Sectional Dependence

The IVAR model can generate observations with strong cross section dependence if the boundedness assumption on the column and row norms of $\mathbf{R}$ and $\mathbf{\Phi}$ are relaxed. The analysis of this case is beyond the scope of the present paper and is considered in Pesaran and Chudik (2010). But even if the boundedness assumptions on $\mathbf{R}$ and $\mathbf{\Phi}$ are maintained, it is still possible for x_{it} to show strong cross section dependence if the IVAR model is augmented with common factors. The basic IVAR model, (3), can be augmented with exogenously specified common factors in a number of different ways. Here we consider two important possibilities. First, a finite number of common factors can be added to the vector of the error terms, defined by (4). This is equivalent to assuming that a finite number of the columns (or linear combinations of the columns) of $\mathbf{R}$ have unbounded norms. This compounding of the spatial (weak) cross section dependence with the strong factor dependence complicates the analysis unduly and will not be pursued here. A more attractive alternative would be to assume that

$$\mathbf{\Phi}(L)(\mathbf{x}_t - \boldsymbol{\alpha} - \mathbf{\Gamma}\mathbf{f}_t) = \mathbf{u}_t, \text{ for } t = 1, 2, ..., T, \tag{25}$$

where $\mathbf{\Phi}(L) = \mathbf{I} - \mathbf{\Phi}L$, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$ is an $N \times 1$ vector of fixed effects, $\mathbf{f}_t$ is an $m \times 1$ vector of unobserved common factors (m is fixed but otherwise unknown), $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_N)'$ is the $N \times m$ matrix of factor loadings, and as before $\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t$. Under this specification the strong cross section dependence of x_{it} due to the factors is explicitly separated from other sources of cross dependence as embodied in $\mathbf{\Phi}$ and $\mathbf{R}$.

4 Estimation of a Factor Augmented Stationary IVAR Model

We now consider the problem of estimation and inference in the case of the factor augmented IVAR model as set out in (25), as both N and T tend to infinity. We focus on parameters of the i^{th} equation and assume that $\boldsymbol{\phi}_i'$ (the i^{th} row of matrix $\mathbf{\Phi}$) can be decomposed as in Assumption 1. See (7)-(9). As an important example we consider the two-neighbor IVAR model defined in Example 1, where the parameters of interest is given by the elements of the i^{th} row of matrix $\mathbf{\Phi}_a$ given by (12). In what follows we set $\boldsymbol{\xi}_{it} = \mathbf{S}_i'\mathbf{x}_t$, where $\mathbf{S}_i$ is defined by (9), and note that it reduces to $(x_{i-1,t}, x_{it}, x_{i+1,t})'$ in the case of the two-neighbor IVAR model.

We suppose that the following assumptions hold.

ASSUMPTION 5 *(Available observations) Available observations are* $\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_T$ *with the starting values* $\mathbf{x}_0 = \sum_{\ell=0}^{\infty} \mathbf{\Phi}^{\ell}\mathbf{R}\boldsymbol{\varepsilon}_{-\ell} + \boldsymbol{\alpha} + \mathbf{\Gamma}\mathbf{f}_0$.

ASSUMPTION 6 *(Common factors) The unobserved common factors,* $f_{1t}, f_{2t}, ..., f_{mt}$, *are covariance stationary and follow the general linear processes:*

$$f_{st} = \psi_s(L)\,\varepsilon_{fst}, \ \textit{for } s = 1, 2, ..., m, \tag{26}$$

where $\psi_s(L) = \sum_{\ell=0}^{\infty} \psi_{s\ell}L^{\ell}$ *with absolute summable coefficients that do not vary with* N, *the factor innovations,* ε_{fst}, *are independently distributed over time with zero means and a constant variance,* $\sigma^2_{\varepsilon_{fs}}$, *that do not vary with* N. ε_{fst}*'s are also distributed independently of the idiosyncratic errors,* $\varepsilon_{it'}$, *for any* $i \in \mathbb{N}$, *any* $t, t' \in \mathcal{T}$, *and any* $s \in \{1, .., m\}$. $E(\mathbf{f}_t\mathbf{f}_t')$ *exists and is a positive definite* *matrix.*

ASSUMPTION 7 *(Existence of fourth order moments) There exists a positive real constant K such that $E\left(\varepsilon_{fst}^4\right) < K$ and $E\left(\varepsilon_{it}^4\right) < K$ for any $s \in \{1, .., m\}$, any $t \in \mathcal{T}$ and any $i \in \mathbb{N}$.*

ASSUMPTION 8 *(Bounded factor loadings and fixed effects) For any $i \in \mathbb{N}$, $\boldsymbol{\gamma}_i$ and α_i do not change with N, $\|\boldsymbol{\gamma}_i\| < K$, and $|\alpha_i| < K$.*

We follow Pesaran (2006) and introduce the following vector of cross section averages $\overline{\mathbf{x}}_{Wt} = \mathbf{W}'\mathbf{x}_t$, where $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N)'$ and $\{\mathbf{w}_j\}_{j=1}^N$ are $m_w \times 1$ dimensional vectors. Subscripts denoting the number of groups are again omitted where not necessary, in order to keep the notations simple. Matrix $\mathbf{W}$ does not correspond to any spatial weights matrix. It is any arbitrary matrix of pre-determined weights satisfying the following granularity conditions

$$\|\mathbf{W}\| = O\left(N^{-\frac{1}{2}}\right), \tag{27}$$

$$\frac{\|\mathbf{w}_j\|}{\|\mathbf{W}\|} = O\left(N^{-\frac{1}{2}}\right) \text{ for any } j. \tag{28}$$

Multiplying (25) by the inverse of polynomial $\boldsymbol{\Phi}(L)$ and then by $\mathbf{W}'$ yields

$$\overline{\mathbf{x}}_{Wt} = \overline{\boldsymbol{\alpha}}_W + \overline{\boldsymbol{\Gamma}}_W \mathbf{f}_t + \overline{\boldsymbol{\upsilon}}_{Wt}, \tag{29}$$

where $\overline{\boldsymbol{\alpha}}_W = \mathbf{W}'\boldsymbol{\alpha}$, $\overline{\boldsymbol{\Gamma}}_W = \mathbf{W}'\boldsymbol{\Gamma}$, $\overline{\boldsymbol{\upsilon}}_{Wt} = \mathbf{W}'\boldsymbol{\upsilon}_t$, and

$$\boldsymbol{\upsilon}_t = \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}^{\ell} \mathbf{u}_{t-\ell}. \tag{30}$$

Under Assumptions 2-3, $\{\mathbf{u}_t\}$ is weakly cross sectionally dependent and

$$\begin{aligned} \|Var(\overline{\boldsymbol{\upsilon}}_{Wt})\| &= \left\| \sum_{\ell=0}^{\infty} \mathbf{W}'\boldsymbol{\Phi}^{\ell}\boldsymbol{\Sigma}\boldsymbol{\Phi}'^{\ell}\mathbf{W} \right\|, \\ &\leq \|\mathbf{W}\|^2 \|\boldsymbol{\Sigma}\| \sum_{\ell=0}^{\infty} \left\|\boldsymbol{\Phi}^{\ell}\right\|^2, \\ &= O\left(N^{-1}\right), \end{aligned} \tag{31}$$

where $\|\mathbf{W}\|^2 = O\left(N^{-1}\right)$ by condition (27), $\|\boldsymbol{\Sigma}\| = O(1)$ by Assumption 3 (see Remark 1) and $\sum_{\ell=0}^{\infty} \left\|\boldsymbol{\Phi}^{\ell}\right\| \leq \sum_{\ell=0}^{\infty} \|\boldsymbol{\Phi}\|^{\ell} = O(1)$ under Assumption 4. This implies $\overline{\boldsymbol{\upsilon}}_{Wt} = O_p\left(N^{-\frac{1}{2}}\right)$ and the unobserved common factors can be approximated as

$$\left(\overline{\mathbf{\Gamma}}_W'\overline{\mathbf{\Gamma}}_W\right)^{-1}\overline{\mathbf{\Gamma}}_W'\left(\overline{\mathbf{x}}_{Wt}-\overline{\boldsymbol{\alpha}}_W\right)=\mathbf{f}_t+O_p\left(N^{-\frac{1}{2}}\right), \tag{32}$$

provided that the matrix $\overline{\mathbf{\Gamma}}_W'\overline{\mathbf{\Gamma}}_W$ is non-singular. It can be inferred that the full column rank of $\overline{\mathbf{\Gamma}}_W$ is important for the estimation of unit-specific coefficients. Pesaran (2006) shows that the full column rank condition is not, however, necessary if the object of the interest is the cross section mean of the parameters, $E(\boldsymbol{\delta}_i)$, as opposed to the unit-specific parameters, $\boldsymbol{\delta}_i$, which are the focus of the current paper.

Using (25), the equation for unit $i\in\mathcal{K}$ can be written as:

$$x_{it}-\alpha_i-\boldsymbol{\gamma}_i'\mathbf{f}_t=\boldsymbol{\delta}_i'\mathbf{S}_i'\left(\mathbf{x}_{t-1}-\boldsymbol{\alpha}-\mathbf{\Gamma}\mathbf{f}_{t-1}\right)+\zeta_{i,t-1}+u_{it}, \tag{33}$$

where

$$\zeta_{it}=\boldsymbol{\phi}_{ib}'\boldsymbol{\upsilon}_t=O_p\left(N^{-\frac{1}{2}}\right), \tag{34}$$

since by Assumption 1 $\boldsymbol{\phi}_{ib}$ satisfies condition (27). It follows from (29) that

$$\boldsymbol{\gamma}_i'\mathbf{f}_t-\boldsymbol{\phi}_{ia}'\mathbf{\Gamma}\mathbf{f}_{t-1}=\mathbf{b}_{i1}'\overline{\mathbf{x}}_{Wt}+\mathbf{b}_{i2}'\overline{\mathbf{x}}_{W,t-1}-\left(\mathbf{b}_{i1}+\mathbf{b}_{i2}\right)'\overline{\boldsymbol{\alpha}}_W-\mathbf{b}_{i1}'\overline{\boldsymbol{\upsilon}}_{Wt}-\mathbf{b}_{i2}'\overline{\boldsymbol{\upsilon}}_{W,t-1}, \tag{35}$$

where $\mathbf{b}_{i1}=\boldsymbol{\gamma}_i'\left(\overline{\mathbf{\Gamma}}_W'\overline{\mathbf{\Gamma}}_W\right)^{-1}\overline{\mathbf{\Gamma}}_W'$ and $\mathbf{b}_{i2}=-\boldsymbol{\delta}_i'\mathbf{S}_i'\mathbf{\Gamma}\left(\overline{\mathbf{\Gamma}}_W'\overline{\mathbf{\Gamma}}_W\right)^{-1}\overline{\mathbf{\Gamma}}_W'$. Substituting (35) into (33) yields

$$x_{it}=\boldsymbol{\delta}_i'\mathbf{S}_i'\mathbf{x}_{t-1}+\mathbf{b}_{i1}'\overline{\mathbf{x}}_{Wt}+\mathbf{b}_{i2}'\overline{\mathbf{x}}_{W,t-1}+c_i+u_{it}+q_{it}, \tag{36}$$

where $c_i=\alpha_i-\boldsymbol{\phi}_{ia}'\boldsymbol{\alpha}-\left(\mathbf{b}_{i1}+\mathbf{b}_{i2}\right)'\overline{\boldsymbol{\alpha}}_W$, and

$$q_{it}=\zeta_{i,t-1}-\mathbf{b}_{i1}'\overline{\boldsymbol{\upsilon}}_{Wt}-\mathbf{b}_{i2}'\overline{\boldsymbol{\upsilon}}_{W,t-1}=O_p\left(N^{-\frac{1}{2}}\right). \tag{37}$$

Consider now the following auxiliary regression based on (36):

$$x_{it}=\mathbf{g}_{it}'\boldsymbol{\pi}_i+\epsilon_{it}, \tag{38}$$

where $\epsilon_{it}=u_{it}+q_{it}$, $\boldsymbol{\pi}_i=\left(\boldsymbol{\delta}_i',\mathbf{b}_{i1}',\mathbf{b}_{i2}',c_i\right)'$ is the $k_i\times 1$ vector of coefficients associated with the regressors $\mathbf{g}_{it} = \left(\boldsymbol{\xi}'_{i,t-1}, \overline{\mathbf{x}}'_{Wt}, \overline{\mathbf{x}}'_{W,t-1}, 1\right)'$, and $k_i = h_i + 2m_w + 1$. The parameters of interest, $\boldsymbol{\delta}_i$, can now be estimated using the cross section augmented regression defined by (38). We refer to such an estimator of $\boldsymbol{\delta}_i$ as the cross section augmented least squares estimator (or CALS for short), and denote it by $\widehat{\boldsymbol{\delta}}_{i,CALS}$. We have

$$\widehat{\boldsymbol{\pi}}_i = \begin{pmatrix} \widehat{\boldsymbol{\delta}}_{i,CALS} \\ \hat{\mathbf{b}}_{i1} \\ \hat{\mathbf{b}}_{i2} \\ \hat{c}_i \end{pmatrix} = \left(\sum_{t=1}^{T} \mathbf{g}_{it}\mathbf{g}'_{it}\right)^{-1} \sum_{t=1}^{T} \mathbf{g}_{it} x_{it}. \tag{39}$$

Also using partitioned regression formula,

$$\widehat{\boldsymbol{\delta}}_{i,CALS} = \left(\mathbf{Z}'_i \mathbf{M}_H \mathbf{Z}_i\right)^{-1} \mathbf{Z}'_i \mathbf{M}_H \mathbf{x}_{i\circ}, \tag{40}$$

where

$$\mathbf{M}_H = \mathbf{I}_T - \mathbf{H}\left(\mathbf{H}'\mathbf{H}\right)^{+}\mathbf{H}', \;\; \mathbf{H} = \left[\overline{\mathbf{X}}_W, \overline{\mathbf{X}}_W(-1), \boldsymbol{\tau}\right], \tag{41}$$

$$\begin{aligned} \mathbf{Z}_i &= \left[\boldsymbol{\xi}_{i1}(-1), \boldsymbol{\xi}_{i2}(-1), ..., \boldsymbol{\xi}_{ih_i}(-1)\right], \\ \boldsymbol{\xi}_{ir}(-1) &= \left(\xi_{ir0}, ..., \xi_{i,r,T-1}\right)', \text{ for } r \in \{1,..,h_i\}, \end{aligned} \tag{42}$$

$\boldsymbol{\tau}$ is a $T \times 1$ vector of ones, $\overline{\mathbf{X}}_W = (\overline{\mathbf{x}}_{W1\circ}, ..., \overline{\mathbf{x}}_{Wm_w\circ})$, $\overline{\mathbf{X}}_W(-1) = [\overline{\mathbf{x}}_{W1}(-1), ..., \overline{\mathbf{x}}_{Wm_w}(-1)]$, $\overline{\mathbf{x}}_{Ws\circ} = (\overline{x}_{Ws1}, ..., \overline{x}_{WsT})'$, $\overline{\mathbf{x}}_{Ws}(-1) = (\overline{x}_{Ws0}, ..., \overline{x}_{Ws,T-1})'$, for $s \in \{1,..,m_w\}$, and $\mathbf{x}_{i\circ} = (x_{i1}, ..., x_{iT})'$,

For future reference also let $\mathbf{v}_{it} = \mathbf{S}'_i \boldsymbol{\upsilon}_t = \boldsymbol{\xi}_{it} - \mathbf{S}'_i \boldsymbol{\Gamma}\mathbf{f}_t - \mathbf{S}'_i \boldsymbol{\alpha}$,

$$\mathbf{Q} = [\mathbf{F}, \mathbf{F}(-1), \boldsymbol{\tau}], \tag{43}$$

and

$$\underset{(2m+1)\times(2m_w+1)}{\mathbf{A}} = \begin{pmatrix} 1 & \overline{\boldsymbol{\alpha}}'_W & \overline{\boldsymbol{\alpha}}'_W \\ 0 & \overline{\boldsymbol{\Gamma}}'_W & \mathbf{0}_{m\times m_w} \\ 0 & \mathbf{0}_{m\times m_w} & \overline{\boldsymbol{\Gamma}}'_W \end{pmatrix}, \tag{44}$$

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

First we consider the asymptotic properties of $\widehat{\boldsymbol{\pi}}_i$ (and $\widehat{\boldsymbol{\delta}}_{i,CALS}$) as $(N,T) \xrightarrow{j} \infty$, in the case where the number of unobserved common factors equals to the dimension of $\overline{\mathbf{x}}_{Wt}$ ($m = m_w$), and make the following additional assumption.

ASSUMPTION 9 *(Identification of $\boldsymbol{\pi}_i$) There exists T_0 and N_0 such that for all $T \geq T_0$, $N \geq N_0$ and for any $i \in \mathcal{K}$, $\left(T^{-1}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}$ exists, $\mathbf{C}_{(N),i} = E\left(\mathbf{g}_{it}\mathbf{g}_{it}'\right)$ is positive definite, and $\left\|\mathbf{C}_{(N),i}^{-1}\right\| < K$.*

Remark 6 *Assumption 9 implies $\overline{\boldsymbol{\Gamma}}_W$ is a square, full rank matrix and, therefore, the number of unobserved common factors is equal the number of columns of the weight matrix, $\mathbf{W}$ ($m = m_w$). In cases where $m < m_w$, full augmentation of individual models by (cross sectional) averages is not necessary.*

Theorem 1 *Let $\mathbf{x}_t$ be generated by model (25), Assumptions 1-9 hold, and $\mathbf{W}$ is any arbitrary (pre-determined) matrix of weights satisfying conditions (27)-(28), and Assumption 9. Then for any $i \in \mathcal{K}$ and as $(N,T) \xrightarrow{j} \infty$, $\widehat{\boldsymbol{\pi}}_i$ defined in equation (39) has the following properties.*

a)

$$\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i \xrightarrow{p} 0.$$

b) If in addition $T/N \rightarrow \varkappa$, with $0 \leq \varkappa < \infty$,

$$\frac{\sqrt{T}}{\sigma_{(N),ii}}\mathbf{C}_{(N),i}^{\frac{1}{2}}\left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right) \xrightarrow{D} N\left(\mathbf{0}, \mathbf{I}_{k_i}\right), \tag{45}$$

where $\sigma_{(N),ii}^2 = Var\left(u_{it}\right) = E\left(\mathbf{e}_i'\mathbf{R}\mathbf{R}'\mathbf{e}_i\right)$, and $\mathbf{C}_{(N),i}^{\frac{1}{2}}$ is the square root of the positive definite matrix $\mathbf{C}_{(N),i} = E\left(\mathbf{g}_{it}\mathbf{g}_{it}'\right)$. Also

c)

$$\mathbf{C}_{(N),i} - \widehat{\mathbf{C}}_{(N),i} \xrightarrow{p} 0, \text{ and } \sigma_{(N),ii} - \widehat{\sigma}_{(N),ii} \xrightarrow{p} 0,$$

where

$$\widehat{\mathbf{C}}_{(N),i} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}', \ \widehat{\sigma}_{(N),ii}^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{it}^2, \tag{46}$$

and $\widehat{u}_{it} = x_{it} - \mathbf{g}_{it}'\widehat{\boldsymbol{\pi}}_i$.

Remark 7 *Suppose that in addition to the assumptions of Theorem 1, the limits of* $\mathbf{C}^{-1}_{(N),i}$ *and* $\sigma^2_{(N),ii}$*, as* $N \to \infty$*, exist and are given by* $\mathbf{C}^{-1}_{(\infty),i}$*, and* $\sigma^2_{(\infty),ii}$*, respectively.*[12] *Then (45) yields*

$$\sqrt{T}\left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right) \xrightarrow{D} N\left(\mathbf{0}, \sigma^2_{(\infty),ii}\mathbf{C}^{-1}_{(\infty),i}\right). \tag{47}$$

Consider now the case where the number of unobserved common factors is unknown, but it is known that $m_w \geq m$. Since the auxiliary regression (38) is augmented possibly by a larger number of cross section averages than the number of unobserved common factors, we have potential problem of multicollinearity (as $N \to \infty$). But this does not affect the estimation of $\boldsymbol{\delta}_i$ so long as the space spanned by the unobserved common factors including a constant and the space spanned by the vector $(1, \overline{\mathbf{x}}'_{Wt})'$ are the same as $N \to \infty$. This is the case when $\overline{\boldsymbol{\Gamma}}_W$ has full column rank.

For this more general case we replace Assumption 9 with the following, and suppress the subscript N to simplify the notations.

ASSUMPTION 10 *(Identification of* $\boldsymbol{\delta}_i$*) There exists* T_0 *and* N_0 *such that for all* $T \geq T_0$, $N \geq N_0$ *and for any* $i \in \mathcal{K}$, $\left(T^{-1}\mathbf{Z}'_i\mathbf{M}_H\mathbf{Z}_i\right)^{-1}$ *exists,* $\overline{\boldsymbol{\Gamma}}_W$ *is a full column rank matrix,* $\boldsymbol{\Omega}_{vi} = E\left(\mathbf{v}_{it}\mathbf{v}'_{it}\right) = \sum_{\ell=0}^{\infty}\mathbf{S}'_i\boldsymbol{\Phi}^{\ell}\mathbf{R}\mathbf{R}'\boldsymbol{\Phi}'^{\ell}\mathbf{S}_i$ *is positive definite, and* $\left\|\boldsymbol{\Omega}^{-1}_{vi}\right\| = O(1)$.

Theorem 2 *Let* $\mathbf{x}_t$ *be generated by model (25), Assumptions 1-8, and 10 hold, and* $\mathbf{W}$ *is any arbitrary (pre-determined) matrix of weights satisfying conditions (27)-(28) and Assumption 10. Then for any* $i \in \mathcal{K}$*, and if in addition* $(N,T) \xrightarrow{j} \infty$ *such that* $T/N \to \varkappa$*, with* $0 \leq \varkappa < \infty$*, the asymptotic distribution of* $\widehat{\boldsymbol{\delta}}_{i,CALS}$ *defined by (40) is given by.*

$$\frac{\sqrt{T}}{\sigma_{ii}}\boldsymbol{\Omega}^{\frac{1}{2}}_{vi}\left(\widehat{\boldsymbol{\delta}}_{i,CALS} - \boldsymbol{\delta}_i\right) \xrightarrow{D} N\left(\mathbf{0}, \mathbf{I}_{h_i}\right), \tag{48}$$

where $\sigma^2_{ii} = Var\left(u_{it}\right)$, $\boldsymbol{\Omega}_{vi} = E\left(\mathbf{v}_{it}\mathbf{v}'_{it}\right)$ *and* $\mathbf{v}_{it} = \mathbf{S}'_i\boldsymbol{\upsilon}_t = \sum_{\ell=0}^{\infty}\mathbf{S}'_i\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell}$.

Remark 8 *As before, we also have*

$$\sqrt{T}\left(\widehat{\boldsymbol{\delta}}_{i,CALS} - \boldsymbol{\delta}_i\right) \xrightarrow{D} N\left(\mathbf{0}, \sigma^2_{(\infty),ii}\boldsymbol{\Omega}^{-1}_{v(\infty),i}\right),$$

where $\boldsymbol{\Omega}_{v(\infty),i} = \lim_{N\to\infty}\boldsymbol{\Omega}_{vi}$*, and* $\sigma^2_{(\infty),ii} = \lim_{N\to\infty}\sigma^2_{ii}$*, assuming limits exist.*

[12] Sufficient condition for $\lim_{N\to\infty}\mathbf{C}_{(N),i}$ to exist is the existence of the following limits (together with Assumptions 1-8): $\lim_{N\to\infty}\mathbf{S}'_i\boldsymbol{\alpha}$, $\lim_{N\to\infty}\mathbf{S}'_i\boldsymbol{\Gamma}$, $\lim_{N\to\infty}\mathbf{W}'\boldsymbol{\Gamma}$, $\lim_{N\to\infty}\mathbf{W}'\boldsymbol{\alpha}$, and $\lim_{N\to\infty}\sum_{\ell=0}^{\infty}\mathbf{S}'_i\boldsymbol{\Phi}^{\ell}\mathbf{R}\mathbf{R}'\boldsymbol{\Phi}'^{\ell}\mathbf{S}_i$.

5 Monte Carlo Experiments: Small Sample Properties of CALS Estimator

5.1 Monte Carlo Design

In this section we report some evidence on the small sample properties of the CALS estimator in the presence of unobserved common factors and weak error cross section dependence and compare the results with standard least squares estimators. Objectives of the experiments are twofold. First, we would like to investigate how well the CALS estimator performs in the presence of unobserved common factors. Second, we would like to find out the extent to which cross section augmentation affects the small sample properties of the estimator when the cross section dependence is weak, and therefore cross section augmentation is asymptotically unnecessary. The focus of our analysis will be on the estimation of the individual-specific parameters in an IVAR model that also allows for other inter-dependencies that are of order $O(N^{-1})$.

The data generating process (DGP) used is given by

$$\mathbf{x}_t - \boldsymbol{\gamma} f_t = \boldsymbol{\Phi} \left(\mathbf{x}_{t-1} - \boldsymbol{\gamma} f_{t-1} \right) + \mathbf{u}_t, \tag{49}$$

where f_t is the only unobserved common factor considered ($m = 1$), and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_N)'$ is the $N \times 1$ vector of factor loadings.

We consider two sets of factor loadings to distinguish the case of weak and strong cross section dependence. Under the former we set $\boldsymbol{\gamma} = \mathbf{0}$, and under the latter we generate the factor loadings γ_i, for $i = 1, 2, ..., N$, from a stationary spatial process in order to show that our estimators are invariant to possible cross section dependence in the factor loadings. Accordingly, the factor loadings are generated by the following bilateral Spatial Autoregressive Model (SAR) process

$$\gamma_i - \mu_\gamma = \frac{a_\gamma}{2} \left(\gamma_{i-1} + \gamma_{i+1} \right) - a_\gamma \mu_\gamma + \eta_{\gamma i},\ 0 < a_\gamma < 1, \tag{50}$$

where $\eta_{\gamma i} \sim IIDN \left(0, \sigma^2_{\eta\gamma} \right)$. As established by Whittle (1954), the unilateral SAR(2) scheme

$$\gamma_i = \psi_{\gamma 1} \gamma_{i-1} + \psi_{\gamma 2} \gamma_{i-2} + \eta_{\gamma i}, \tag{51}$$

with $\psi_{\gamma 1}=-2b_{\gamma}$, $\psi_{\gamma 2}=b_{\gamma}^{2}$ and $b_{\gamma}=\left(1-\sqrt{1-a_{\gamma}^{2}}\right)/a_{\gamma}$, generates the same autocorrelations as the bilateral SAR(1) scheme (50). The factor loadings are generated using the unilateral scheme (51) with 50 burn-in data points ($i=-49,...,0$) and the initialization, $\gamma_{-51}=\gamma_{-50}=0$. We set $a_{\gamma}=0.4$, $\mu_{\gamma}=1$, and choose $\sigma_{\eta\gamma}^{2}$ such that $Var\left(\gamma_{i}\right)=1$.[13] The common factors are generated according to the AR(1) process

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

with $\rho_{f}=0.9$.

In line with the theoretical analysis, the autoregressive parameters are decomposed as $\mathbf{\Phi}=\mathbf{\Phi}_{a}+\mathbf{\Phi}_{b}$, where $\mathbf{\Phi}_{a}$ captures own and neighborhood effects as in

$$\mathbf{\Phi}_{a}=\begin{pmatrix} \varphi_{1} & \psi_{1} & 0 & 0 & & 0 \\ \psi_{2} & \varphi_{2} & \psi_{2} & 0 & & 0 \\ 0 & \psi_{3} & \varphi_{3} & \psi_{3} & & 0 \\ 0 & 0 & \psi_{4} & \varphi_{4} & \ddots & \\ & & & \ddots & \ddots & \psi_{N-1} \\ 0 & 0 & 0 & & \psi_{N} & \varphi_{N} \end{pmatrix},$$

and the remaining elements of $\mathbf{\Phi}$, defined by $\mathbf{\Phi}_{b}$, are generated as

$$\phi_{bij}=\begin{cases} \lambda_{i}\omega_{ij} & \text{for } j\notin\{i-1,i,i+1\} \\ 0 & \text{for } j\in\{i-1,i,i+1\} \end{cases}, \text{ where}$$

$$\lambda_{i}\sim IIDU\left(-0.1,0.2\right) \text{ and } \omega_{ij}=\frac{\varsigma_{ij}}{\sum_{j=1}^{N}\varsigma_{ij}}, \tag{52}$$

with $\varsigma_{ij}\sim IIDU\left(0,1\right)$. This ensures that $\phi_{bij}=O_{p}(N^{-1})$, and $\lim_{N\rightarrow\infty}E\left(\phi_{bij}\right)=0$, for all i and j.

With $\mathbf{\Phi}_{a}$ as specified above, each unit i, except the first and the last, has two neighbors: the

[13] The variance of factor loadings is given by

$$\sigma_{\eta\gamma}^{2}=\frac{\left(1+\psi_{\gamma 2}\right)\left[\left(1-\psi_{\gamma 2}^{2}\right)-\psi_{\gamma 1}^{2}\right]}{\left(1-\psi_{\gamma 2}\right)}.$$

'left' neighbor $i-1$ and the 'right' neighbor $i+1$. The DGP for the i^{th} unit can now be written as

$$\begin{aligned}
x_{1t} &= \varphi_1 x_{1,t-1} + \psi_1 x_{2,t-1} + \boldsymbol{\phi}_{b1}' \mathbf{x}_{t-1} + \gamma_1 f_t - \left(\boldsymbol{\phi}_1' \boldsymbol{\gamma}\right) f_{t-1} + u_{1t}, \\
x_{it} &= \varphi_i x_{i,t-1} + \psi_i \left(x_{i-1,t-1} + x_{i+1,t-1}\right) + \boldsymbol{\phi}_{bi}' \mathbf{x}_{t-1} + \gamma_i f_t - \left(\boldsymbol{\phi}_i' \boldsymbol{\gamma}\right) f_{t-1} + u_{it},\ i \in \{2, .., N-1\}, \\
x_{Nt} &= \varphi_N x_{N,t-1} + \psi_N x_{N-1,t-1} + \boldsymbol{\phi}_{b,N}' \mathbf{x}_{t-1} + \gamma_N f_t - \left(\boldsymbol{\phi}_N' \boldsymbol{\gamma}\right) f_{t-1} + u_{Nt}.
\end{aligned}$$

To ensure that the DGP is stationary we generate $\varphi_i \sim IIDU\,(0.4, 0.6)$, and $\psi_i \sim IIDU\,(-0.1, 0.1)$ for $i \neq 2$. We choose to focus on the equation for unit $i = 2$ in all experiments and we set $\varphi_2 = 0.5$ and $\psi_2 = 0.1$. This yields $\|\boldsymbol{\Phi}\|_\infty \leq 0.9$, and together with $\left|\rho_f\right| < 1$ it is ensured that the DGP is stationary and the variance of x_{it} is bounded in N. The cross section averages, $\overline{x}_{wt}$, are constructed as simple averages, $\overline{x}_t = N^{-1} \sum_{j=1}^{N} x_{it}$.

The N-dimensional vector of error terms, $\mathbf{u}_t$, is generated using the following SAR model:

$$\begin{aligned}
u_{1t} &= a_u u_{2t} + \varepsilon_{1t}, \\
u_{it} &= \frac{a_u}{2}\left(u_{i-1,t} + u_{i+1,t}\right) + \varepsilon_{it},\ i \in \{2, .., N-1\} \\
u_{Nt} &= a_u u_{N-1,t} + \varepsilon_{Nt},
\end{aligned}$$

for $t = 1, 2, .., T$. We set $a_u = 0.4$ which ensures that the errors are cross sectionally weakly dependent, and draw ε_{it}, the i^{th} element of $\boldsymbol{\varepsilon}_t$, as $IIDN\left(0, \sigma_\varepsilon^2\right)$. We set $\sigma_\varepsilon^2 = N/tr\left(\mathbf{R}_u \mathbf{R}_u'\right)$ so that on average $Var(u_{it}) = 1$, where $\mathbf{R}_u = \left(\mathbf{I} - a_u \mathbf{S}\right)^{-1}$, and the spatial weights matrix $\mathbf{S}$ is

$$\mathbf{S} = \begin{pmatrix}
0 & 1 & 0 & 0 & & 0 \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & & 0 \\
0 & \frac{1}{2} & 0 & \frac{1}{2} & & 0 \\
 & & \ddots & \ddots & \ddots & \\
 & & & \frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & & 0 & 1 & 0
\end{pmatrix}. \tag{53}$$

In order to minimize the effects of the initial values, the first 50 observations are dropped. $N \in \{25, 50, 75, 100, 200\}$ and $T \in \{25, 50, 75, 100, 200\}$. For each N, all parameters were set at the beginning of the experiments and 2000 replications were carried out by generating new innovations

ε_{it}, η_{ft} and $\eta_{\gamma i}$.

The focus of the experiments is to evaluate the small sample properties of the CALS estimator of the own coefficient $\varphi_2 = 0.5$ and the neighboring coefficient $\psi_2 = 0.1$, in the case of the second cross section unit.[14] The cross-section augmented regression for estimating (ϕ_2, ψ_2) is given by

$$x_{2t} = c_2 + \psi_2 (x_{1,t-1} + x_{3,t-1}) + \varphi_2 x_{2,t-1} + \delta_{2,0}\overline{x}_t + \delta_{2,1}\overline{x}_{t-1} + \epsilon_{2t}. \tag{54}$$

We also report results of the Least Squares (LS) estimator computed using the above regression but without augmentation with cross-section averages. The corresponding CALS estimator and non-augmented LS estimator are denoted by $\widehat{\varphi}_{2,CALS}$ and $\widehat{\varphi}_{2,LS}$ (own coefficient), or $\widehat{\psi}_{2,CALS}$ and $\widehat{\psi}_{2,LS}$ (neighboring coefficient), respectively.

To summarize, we carry out two different sets of experiments, one set without the unobserved common factor ($\boldsymbol{\gamma} = \mathbf{0}$), and the other with the unobserved common factor ($\boldsymbol{\gamma} \neq \mathbf{0}$). There are many sources of interdependence between individual units: spatial dependence of innovations $\{u_{it}\}$, spatiotemporal interactions due to coefficient matrices $\mathbf{\Phi}_a$ and $\mathbf{\Phi}_b$, and finally in the case where $\boldsymbol{\gamma} \neq \mathbf{0}$ the cross section dependence also arises via the unobserved common factor, f_t, and the cross-sectionally dependent factor loadings, γ_i. Additional intermediate cases are also considered, the results of which are available in a Supplement from the authors, upon request.[15]

5.2 Monte Carlo Results

Tables 1 and 2 give the bias ($\times 100$) and RMSE ($\times 100$) of CALS and LS estimators as well as size and power of tests based on them at the 5% nominal level. Results for the estimated own coefficient, $\widehat{\varphi}_{2,CALS}$ and $\widehat{\varphi}_{2,LS}$, are reported in Table 1. The top panel of this table presents the results for the experiments with an unobserved common factor ($\boldsymbol{\gamma} \neq \mathbf{0}$). In this case, $\{x_{it}\}$ is CSD and the standard LS estimator without augmentation by cross section averages is not consistent. The bias of $\widehat{\varphi}_{2,LS}$ is indeed quite substantial for all values of N and T and the tests based on $\widehat{\varphi}_{2,LS}$ are grossly oversized. CALS, on the other hand, performs well for $T \geq 100$ and all values of N. For smaller values of T, there is a negative bias, and the test based on $\widehat{\varphi}_{2,CALS}$ is slightly oversized.

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

[15] The supplement presents the results for the experiments with all combination of zero and/or non-zero coefficient matrix $\mathbf{\Phi}_b$, zero or non-zero factor loadings γ, and low or high cross section dependence of errors ($a_u = 0.4$ or $a_u = 0.8$).

This is the familiar time series bias where even in the absence of any cross section dependence the LS estimator of the autoregressive coefficient is biased downward (when $\varphi_2 > 0$) in small T samples.

Moving on to the experiments without a common factor (given at the bottom half of the table), we observe that the LS estimator only slightly outperforms the CALS estimator. In the absence of common factors, $\{x_{it}\}$ is weakly cross sectionally dependent and therefore the augmentation with cross section averages is (asymptotically) innocuous. Distortions coming from cross section augmentation are in this case very small. Note that the LS estimator is not efficient because the residuals are cross sectionally dependent. Augmentation by cross-section averages helps to reduce part of this dependence. Nevertheless, the reported RMSE of $\widehat{\varphi}_{2,CALS}$ does not outperform the RMSE of $\widehat{\varphi}_{2,LS}$.

The estimation results for the neighboring coefficient, ψ_2, are presented in Table 2. These are qualitatively similar to the ones reported in Table 1. Cross section augmentation is clearly needed and very helpful when common factors are present. But in the absence of such common effects, the presence of weak cross section dependence, whether through the dynamics or error processes, does not pose any difficulty for the least squares and the CALS estimators so long as N is sufficiently large. Finally, not surprisingly, the estimates are subject to the small T bias irrespective of the size of N or the degree of cross section dependence.

Figure 1 plots the power of the CALS estimator of the own coefficient, $\widehat{\varphi}_{2,CALS}$, (top chart) and the neighboring coefficient, $\widehat{\psi}_{2,CALS}$, (bottom chart) for $N = 200$ and two different values of $T \in \{100, 200\}$. These charts provide a graphical representation of the results reported in Tables 1-2, and also suggest significant improvement in power as T increases for a number of different alternatives.

6 An Empirical Illustration: a spatiotemporal model of house prices in the U.S.

In a recent study Holly, Pesaran, and Yamagata (2009), hereafter HPY, consider the relation between real house prices, p_{it}, and real per capita personal disposable income y_{it} (both in logs) in a panel of 49 US States over 29 years (1975-2003), where $i = 1, 2, ..., 49$ and $t = 1, 2, ..., T$. Controlling for heterogeneity and cross section dependence, they show that p_{it} and y_{it} are cointegrated with coefficients $(1, -1)$, and provide estimates of the following panel error correction model:

$$\Delta p_{it} = c_i + \omega_i(p_{i,t-1} - y_{i,t-1}) + \delta_{1i}\Delta p_{i,t-1} + \delta_{2i}\Delta y_{it} + \upsilon_{it}. \tag{55}$$

To take account of unobserved common factors, HPY augmented (55) with simple cross section averages, $\Delta\bar{p}_t = \Sigma_{i=1}^{49}\Delta p_{it}/49$, $\Delta\bar{y}_t = \Sigma_{i=1}^{49}\Delta y_{it}/49$, and $\bar{p}_{t-1} - \bar{y}_{t-1} = \Sigma_{i=1}^{49}(p_{i,t-1} - y_{i,t-1})/49$, and obtained common correlated effects mean group and pooled estimates (denoted as CCEMG and CCEP) of $\{\omega_i, \delta_{1i}, \delta_{2i}\}$ which we reproduce in the left panel of Table 3. HPY then showed that the residuals from these regressions, $\hat{\upsilon}_{it}$, display a significant degree of spatial dependence. Here we exploit the theoretical results of the present paper and consider the possibility that dynamic neighborhood effects are partly responsible for the residual spatial dependence reported in HPY. To this end we considered an extended version of (55) where the lagged spatial variable $\Delta p^s_{i,t-1} = \sum_{j=1}^{N} s_{ij}\Delta p_{j,t-1}$ is also included amongst the regressors, with s_{ij} being the (i, j) element of a spatial weight matrix, $\mathbf{S}$, namely

$$\Delta p_{it} = c_i + \omega_i(p_{i,t-1} - y_{i,t-1}) + \delta_{1i}\Delta p_{i,t-1} + \psi_i\Delta p^s_{i,t-1} + \delta_{2i}\Delta y_{it} + \upsilon_{it}. \tag{56}$$

Here we consider a simple contiguity matrix $s_{ij} = 1$ when the States i and j share a border and zero otherwise, with $s_{ii} = 0$. Possible strong cross section dependence is again controlled for by augmentation of the extended regression equation with $\Delta\bar{p}_t$, $\Delta\bar{y}_t$, and $\bar{p}_{t-1} - \bar{y}_{t-1}$. Estimation results are reported in the right panel of Table 3. The dynamic spatial effects are found to be highly significant, irrespective of the estimation method, increasing $\bar{R}^2$ of the price equation by 6-9%. The dynamics of past price changes are now distributed between own and neighborhood effects giving rise to much richer dynamics and spill over effects. It is also interesting that the inclusion of the spatiotemporal variable $\Delta p^s_{i,t-1}$ in the model has had little impact on the estimates of the coefficient of the real income variable, δ_{2i}.

7 Concluding Remarks

This paper has proposed restrictions on the coefficients of infinite-dimensional VAR (IVAR) that are binding only in the limit as the number of cross section units (or variables in the VAR) tends to infinity to circumvent the curse of dimensionality. The proposed framework relates to the various approaches considered in the literature. For example when modelling individual households or firms, aggregate variables, such as market returns or regional/national income, are treated as exogenous. This is intuitive as the impact of a firm or household on the aggregate economy is small, of the order $O\left(N^{-1}\right)$. This paper formalizes this idea in a spatiotemporal context.

The paper establishes that in the absence of common factors and when the degree of cross section dependence is weak, then equations for individual units decouple as $N \to \infty$, and can be consistently estimated by running separate regressions. In the presence of observed and/or unobserved common factors, individual-specific VAR models can still be estimated separately if they are conditioned on the common factors. Unobserved common factors can be approximated by cross sectional averages, following the idea originally introduced by Pesaran (2006).

The paper shows that the global VAR approach of Pesaran, Schuermann, and Weiner (2004) can be motivated as an approximation to an IVAR model featuring all the macroeconomic variables. Asymptotic distribution of the cross sectionally augmented least-squares (CALS) estimator of the parameters of the unit-specific equations in the IVAR model is established both in the case when the number of unobserved common factors is known, and when it is unknown but fixed. Small sample properties of the proposed CALS estimator were investigated through Monte Carlo simulations, and an empirical illustration shows the statistical significance of dynamic spill-over effects in modelling of U.S. real house prices across the neighboring States.

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

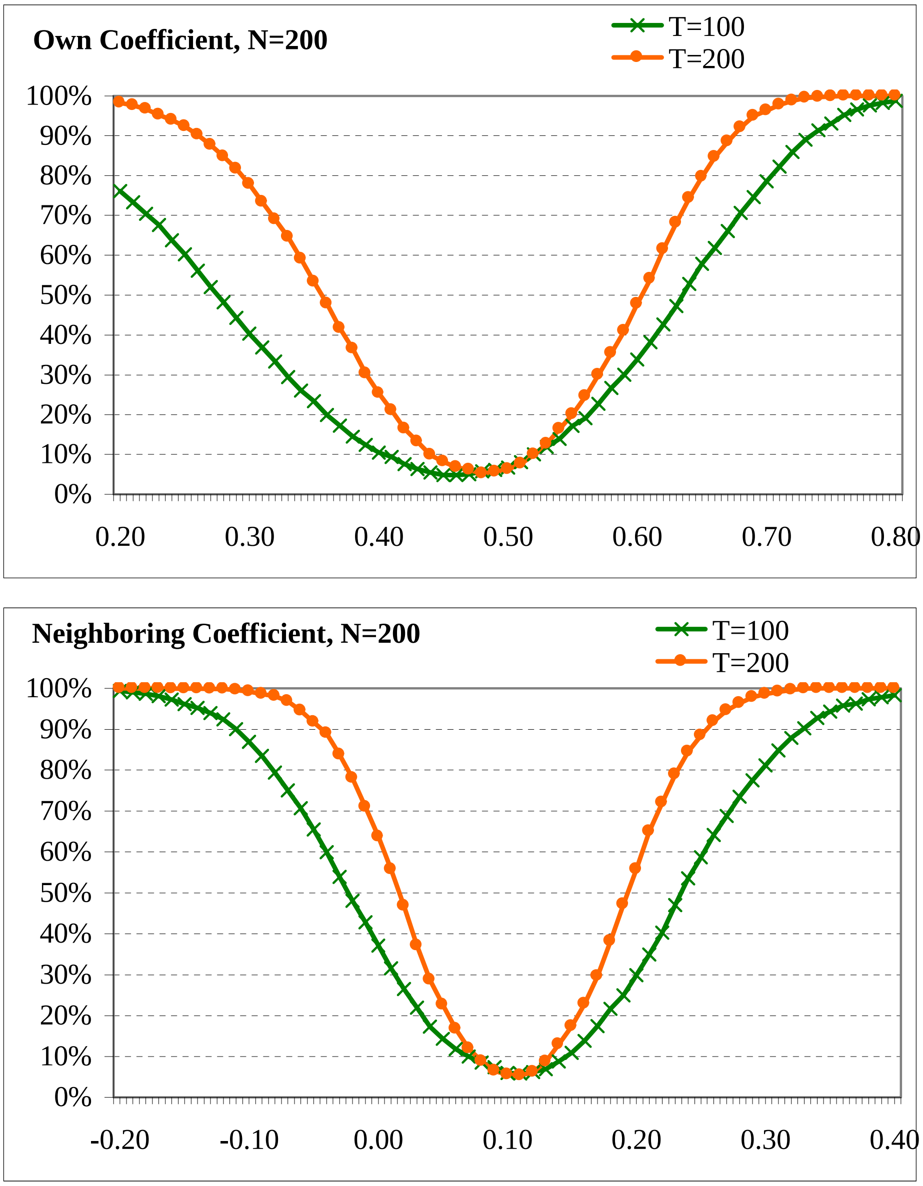

Figure 1: Power Curves for the CALS t-tests of Own Coefficient, φ_2 (the upper chart) and the Neighboring Coefficient, ψ_2 (the lower chart), in the Case of Experiments with $\boldsymbol{\gamma} \neq \mathbf{0}$.

Table 1: MC results for the own coefficient φ_2.

	Bias ($\times 100$)					**Root Mean Square Errors** ($\times 100$)					**Size** (5% level, $H_0 : \varphi_2 = 0.50$)					**Power** (5% level, $H_1 : \varphi_2 = 0.60$)				
(N,T)	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
Experiments with nonzero factor loadings																				
	LS estimator not augmented with cross section averages, $\widehat{\varphi}_{2,LS}$																			
25	-4.34	5.46	7.34	8.76	11.02	25.78	18.82	17.48	17.49	17.63	13.20	21.15	27.70	33.25	45.55	16.85	19.80	23.55	29.20	42.00
50	-3.04	4.39	8.21	8.61	10.65	25.14	18.83	18.37	17.45	17.24	13.80	21.70	30.30	32.35	42.65	16.00	20.85	26.80	29.50	39.20
75	-2.94	4.58	6.99	8.36	10.59	25.20	18.50	17.55	17.47	17.55	13.85	20.40	27.50	32.05	43.60	16.45	18.45	24.60	31.10	41.85
100	-2.67	4.89	7.60	8.94	10.74	24.19	18.27	17.65	17.74	17.63	12.65	20.80	28.40	32.80	44.25	14.90	18.05	23.55	30.35	43.65
200	-3.71	4.11	7.79	8.84	10.26	24.59	18.90	17.98	17.62	17.18	13.50	21.30	28.75	32.70	42.85	15.30	20.95	26.35	29.00	41.45
	CALS estimator $\widehat{\varphi}_{2,CALS}$																			
25	-18.43	-7.63	-5.22	-3.38	-0.72	29.04	15.52	12.14	9.86	6.47	13.30	7.20	7.60	7.00	5.80	23.80	22.90	27.10	31.20	41.00
50	-19.65	-9.58	-5.69	-4.77	-1.76	29.40	16.73	12.44	10.13	6.56	14.35	9.45	7.95	6.15	5.30	24.85	26.90	28.10	33.50	45.15
75	-19.76	-9.36	-5.82	-4.41	-2.30	29.83	16.56	12.41	10.35	6.80	14.00	8.80	8.10	6.80	6.00	26.05	26.35	28.00	33.25	46.55
100	-20.19	-9.31	-6.00	-4.29	-2.30	29.94	16.82	12.60	10.18	6.83	13.20	9.35	8.70	6.50	6.55	25.85	26.80	28.40	31.50	47.80
200	-20.96	-10.17	-6.27	-4.85	-2.37	30.33	17.17	12.27	10.41	6.80	15.45	10.10	7.55	6.70	6.25	27.00	26.40	30.00	33.85	47.80
Experiments with zero factor loadings																				
	LS estimator not augmented with cross section averages, $\widehat{\varphi}_{2,LS}$																			
25	-13.10	-6.17	-3.89	-2.60	-1.59	24.20	14.80	11.45	9.35	6.67	8.40	6.95	6.65	5.35	5.45	15.80	19.25	21.70	25.00	42.60
50	-13.32	-6.16	-4.22	-2.76	-1.35	24.79	14.98	11.51	9.45	6.74	9.10	6.85	6.00	5.75	6.85	17.25	18.85	22.60	25.45	41.20
75	-12.74	-6.19	-4.14	-3.50	-1.37	24.15	14.97	11.60	9.97	6.56	8.35	7.15	6.80	6.25	5.20	17.35	19.70	22.90	27.65	42.25
100	-12.36	-5.65	-4.41	-3.15	-1.72	23.69	14.63	11.87	9.46	6.61	8.75	6.40	5.95	5.25	6.30	16.70	17.75	24.25	26.05	43.40
200	-13.30	-6.42	-4.54	-2.79	-1.46	24.49	15.03	11.43	9.47	6.62	8.95	7.35	5.20	5.00	6.30	17.30	20.80	24.15	25.10	41.00
	CALS estimator $\widehat{\varphi}_{2,CALS}$																			
25	-14.43	-6.22	-3.58	-2.21	-0.86	25.84	15.23	11.36	9.22	6.43	9.35	7.05	6.85	4.80	4.85	18.20	19.50	21.00	23.85	39.50
50	-15.38	-6.87	-4.54	-2.81	-1.10	27.13	15.61	11.77	9.48	6.69	10.90	7.65	6.35	5.80	6.40	19.45	20.20	23.85	25.95	39.90
75	-15.03	-6.93	-4.48	-3.64	-1.34	26.69	15.72	11.95	10.14	6.57	10.20	7.95	7.00	6.90	5.25	19.75	21.05	22.95	28.65	42.50
100	-14.90	-6.51	-4.78	-3.48	-1.75	26.49	15.40	12.15	9.70	6.63	10.55	7.10	7.15	5.65	6.00	18.90	18.35	24.65	27.85	42.80
200	-15.67	-7.51	-5.14	-3.23	-1.61	26.94	15.92	11.91	9.75	6.67	10.55	8.35	6.10	5.65	6.30	19.30	21.80	24.90	25.65	41.90

Notes: $\varphi_2 = 0.5$, $\psi_2 = 0.1$, $a_\gamma = a_u = 0.4$, and $Var(\gamma_i) = 1$. The DGP is given by 2-neighbor IVAR model (49) where the equation for unit $i \in \{2, .., N-1\}$ is $x_{it} = \varphi_i x_{i,t-1} + \psi_i (x_{i-1,t-1} + x_{i+1,t-1}) + \phi_{bi}' x_{t-1} + \gamma_i f_t - \phi_i' \gamma f_{t-1} + u_{it}$. The CALS estimator of the own coefficient φ_2 and the neighboring coefficient ψ_2 is computed using the following auxiliary regression, $x_{2t} = c_2 + \psi_2 (x_{1,t-1} + x_{3,t-1}) + \varphi_2 x_{2,t-1} + \delta_{2,0} \overline{x}_t + \delta_{2,1} \overline{x}_{t-1} + \epsilon_{2t}$. Estimators $\widehat{\varphi}_{2,LS}$ and $\widehat{\psi}_{2,LS}$ are computed from the auxiliary regressions not augmented with cross section averages. Unobserved common factor f_t is generated as stationary AR(1) process, and factor loadings and innovations $\{u_{it}\}$ are generated according to stationary spatial autoregressive processes. Please refer to Section 5 for detailed description of Monte Carlo design.

Table 2: MC results for the neighboring coefficient ψ_2.

	Bias ($\times 100$)					Root Mean Square Errors ($\times 100$)					Size (5% level, $H_0 : \psi_2 = 0.10$)					Power (5% level, $H_1 : \psi_2 = 0.20$)				
(N,T)	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**	**25**	**50**	**75**	**100**	**200**
Experiments with nonzero factor loadings																				
	LS estimator not augmented with cross section averages, $\widehat{\psi}_{2,LS}$																			
25	9.51	7.88	7.67	7.47	7.02	22.64	16.86	15.10	14.62	13.27	19.70	27.30	36.35	43.50	60.30	16.70	23.15	28.65	33.05	43.10
50	8.04	8.06	7.14	7.22	6.33	21.44	16.83	15.24	14.83	13.19	18.60	27.45	35.60	44.10	59.25	15.85	24.80	29.95	34.45	44.80
75	8.49	7.62	7.03	6.31	6.55	21.46	16.11	14.87	13.71	13.48	17.40	26.60	36.00	41.85	58.25	17.30	22.20	30.25	34.40	47.00
100	7.75	7.35	6.99	6.69	6.67	21.54	16.33	15.07	14.15	13.20	18.25	27.05	35.85	42.30	59.15	17.50	24.45	30.20	34.00	45.35
200	8.52	8.28	7.21	6.45	6.76	21.37	17.61	15.20	14.18	13.26	18.85	29.50	36.75	41.00	60.40	16.50	25.30	29.65	34.90	44.10
	CALS estimator $\widehat{\psi}_{2,CALS}$																			
25	2.31	1.70	1.18	1.17	0.78	18.01	10.69	8.20	6.99	4.93	9.10	6.75	5.75	6.70	7.30	12.65	15.75	21.85	28.10	52.80
50	2.20	1.52	0.88	0.89	0.43	16.96	10.56	8.39	6.79	4.73	8.15	6.85	7.40	5.95	6.05	11.60	16.00	24.60	29.55	57.35
75	2.71	1.35	0.87	0.52	0.64	17.27	10.37	8.25	6.82	4.79	8.05	6.75	6.05	5.70	6.25	11.45	15.60	23.35	30.70	53.85
100	1.82	1.26	0.71	0.55	0.58	17.00	10.74	8.23	6.72	4.60	8.10	6.90	6.75	5.45	5.25	11.00	17.00	24.20	30.50	53.90
200	2.55	1.22	1.14	0.72	0.46	18.06	10.55	8.30	6.88	4.57	9.00	6.45	6.50	6.00	5.55	11.95	17.80	21.95	29.90	55.70
Experiments with zero factor loadings																				
	LS estimator not augmented with cross section averages, $\widehat{\psi}_{2,LS}$																			
25	2.04	0.77	0.51	0.39	0.09	16.10	10.04	7.98	6.58	4.66	8.20	5.55	6.15	5.60	5.20	11.55	17.40	23.10	30.05	58.05
50	1.74	0.67	0.41	0.33	0.36	15.76	10.12	7.89	6.81	4.80	7.05	6.00	5.65	5.75	5.90	10.75	17.25	23.90	31.80	55.95
75	1.22	0.98	0.60	0.54	0.09	15.49	10.16	8.01	6.82	4.67	6.80	6.10	5.30	5.55	5.00	11.10	17.45	22.85	30.70	57.15
100	1.94	0.52	0.53	0.68	0.30	16.19	10.14	7.94	6.77	4.72	7.25	6.00	5.00	5.60	5.25	12.25	18.45	24.05	29.20	54.70
200	1.38	1.11	0.48	0.40	0.14	15.63	10.56	7.99	6.75	4.65	7.15	6.45	5.10	6.05	5.05	11.30	18.45	24.05	30.65	56.65
	CALS estimator $\widehat{\psi}_{2,CALS}$																			
25	2.15	1.02	0.66	0.58	0.23	17.64	10.47	8.29	6.70	4.63	8.25	6.35	5.85	5.00	4.70	11.80	17.45	23.55	28.20	56.35
50	1.92	0.91	0.47	0.43	0.46	16.88	10.48	8.08	6.92	4.84	7.50	6.70	5.90	5.90	5.75	10.00	16.30	23.30	29.95	54.30
75	1.15	1.05	0.64	0.61	0.13	16.48	10.43	8.12	6.98	4.69	7.00	6.60	5.40	5.60	5.30	11.45	16.20	22.25	30.20	56.20
100	2.07	0.58	0.58	0.82	0.38	17.40	10.42	8.18	6.84	4.77	8.00	6.15	5.80	5.60	5.50	12.20	17.65	23.70	27.50	53.50
200	1.31	1.17	0.54	0.43	0.14	16.80	10.90	8.13	6.87	4.68	7.10	7.40	5.40	5.95	4.85	11.70	16.85	23.05	30.05	57.35

See the notes to Table 1.

Table 3: Alternative Average Estimates of the Error Correction Models for House Prices Across 49 U.S. States over the Period 1975-2003

	Holly et al. (2009) regressions without dynamic spatial effects			Regressions augmented with dynamic spatial effects		
Δp_{it}	MG	CCEMG	CCEP	MG	CCEMG	CCEP
$p_{i,t-1} - y_{i,t-1}$	-0.105 (0.008)	-0.183 (0.016)	-0.171 (0.015)	-0.095 (0.009)	-0.154 (0.018)	-0.152 (0.018)
$\Delta p_{i,t-1}$	0.524 (0.030)	0.449 (0.038)	0.518 (0.065)	0.296 (0.060)	0.188 (0.049)	0.272 (0.082)
Δy_{it}	0.500 (0.040)	0.277 (0.059)	0.227 (0.063)	0.497 (0.040)	0.284 (0.059)	0.201 (0.088)
$\Delta p^s_{i,t-1}$	-	-	-	0.331 (0.066)	0.350 (0.085)	0.431 (0.105)
$\bar{R}^2$	0.54	0.70	0.66	0.60	0.79	0.72
Average Cross Correlation Coefficients ($\bar{\hat{\rho}}$)	0.284	-0.005	-0.016	0.267	-0.012	-0.016

Notes: MG, CCEMG and CCEP, respectively, stand for the Mean Group, the Common Correlated Effects Mean Group, and the Common Correlated Effects Pooled estimators defined in Pesaran (2006). Augmentation by simple cross section averages, $\Delta\bar{p}_t = \Sigma_{i=1}^{49}\Delta p_{it}/49$, $\Delta\bar{y}_t = \Sigma_{i=1}^{49}\Delta y_{it}/49$, and $\bar{p}_{t-1} - \bar{y}_{t-1} = \Sigma_{i=1}^{49}(p_{i,t-1} - y_{i,t-1})/49$, is used to deal with the possible effects of strong cross section dependence. Standard errors are in parentheses. $\bar{\hat{\rho}}$ denotes the average pair-wise correlation of the residuals from the cross-section augmented regressions across the 49 U.S. States.

References

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

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

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

Canova, F. (1995). Vector autoregressive models: Specification, estimation, inference, and forecasting. In M. Pesaran and M. Wickens (Eds.), *Handbook of applied econometrics: Macroeconomics*, Chapter 2. Oxford: Basil Blackwell.

Chudik, A. (2008). *Global Macroeconomic Modelling*. Ph. D. thesis, Trinity College, University of Cambridge.

Chudik, A., M. H. Pesaran, and E. Tosetti (2009). Weak and strong cross section dependence and estimation of large panels. ECB Working Paper No. 1100.

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

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*, 1–45.

Conley, T. G. and G. Topa (2002). Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics 17*, 303–327.

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

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics 146*, 318–328.

Del Negro, M. and F. Schorfheide (2004). Priors from general equilibrium models for VARs. *International Economic Review 45*, 643–673.

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

Dées, S., F. di Mauro, M. H. Pesaran, and L. V. Smith (2007). Exploring the international linkages of the Euro Area: A global VAR analysis. *Journal of Applied Econometrics 22*, 1–38.

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.

Garratt, A., K. Lee, M. H. Pesaran, and Y. Shin (2006). *Global and national macroeconometric modelling: A long run structural approach.* Oxford University Press.

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.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*, 1545–1578.

Giannone, D., L. Reichlin, and L. Sala (2005). Monetary policy in real time. In M. Gertler and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2004*, Volume 19, pp. 161–200. MIT Press.

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 (2009). A spatio-temporal model of house prices in the U.S. *forthcoming in Journal of Econometrics*.

Hubermann, G. (1982). A simple approach to the arbitrage pricing theory. *Journal of Economic Theory 28*, 183–191.

Ingersoll, J. (1984). Some results in the theory of arbitrage pricing. *Journal of Finance 39*, 1021–1039.

Kelejian, H. H. and D. P. 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.

Leeper, E. M., C. A. Sims, and T. Zha (1996). What does monetary policy do? *Brookings Papers on Economic Activity 2*, 1–63.

Litterman, R. (1986). Forecasting with bayesian vector autoregressions - five years of experience. *Journal of Business and Economics Statistics 4*, 25–38.

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 (2010). Econometric analysis of high dimensional VARs featuring a dominant unit. Mimeo, University of Cambridge.

Pesaran, M. H., T. Schuermann, and B. J. Treutler (2007). Global business cycles and credit risk. In M. Carey and R. Stultz (Eds.), *The risks of financial institutions*, Chapter 9. University of Chicago Press.

Pesaran, M. H., T. Schuermann, B. J. Treutler, and S. M. Weiner (2006). Macroeconomic dynamics and credit risk: A global perspective. *Journal of Money, Credit and Banking 38*, 1211–1262.

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

Pesaran, M. H., L. V. Smith, and R. P. Smith (2007). What if the UK or Sweden had joined the Euro in 1999? An empirical evaluation using a global VAR. *International Journal of Finance and Economics 12*, 55–87.

Pesaran, M. H. and R. Smith (2006). Macroeconometric modelling with a global perspective. *The Manchester School, Supplement*, 24–49.

Pesaran, M. H. and E. Tosetti (2009). Large panels with common factors and spatial correlation. Mimeo, University of Cambridge.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*, 341–360.

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.

Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics 44*, 293–335.

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

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. NBER Working Paper No. 11467.

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

Appendix

A Lemmas and Proofs

Proof of Proposition 1. For any $N \in \mathbb{N}$, the variance of $\mathbf{x}_t$ is

$$\mathbf{\Omega} = Var\left(\mathbf{x}_t\right) = \sum_{\ell=0}^{\infty} \mathbf{\Phi}^{\ell} \mathbf{\Sigma} \mathbf{\Phi}'^{\ell}, \tag{57}$$

and under Assumptions 2-4

$$\left\|\mathbf{\Omega}\right\| \leq \left\|\mathbf{\Sigma}\right\| \sum_{\ell=0}^{\infty} \left\|\mathbf{\Phi}\right\|^{2\ell} < K. \tag{58}$$

Hence, it follows that for any arbitrary non-random vector of weights satisfying the granularity condition (16),

$$\left\|Var\left(\mathbf{w}'\mathbf{x}_t\right)\right\| = \left\|\mathbf{w}'\mathbf{\Omega}\mathbf{w}\right\| \leq \left\|\varrho\left(\mathbf{\Omega}\right)\left(\mathbf{w}'\mathbf{w}\right)\right\|, \tag{59}$$

where $\varrho\left(\mathbf{\Omega}\right) = \left\|\mathbf{\Omega}\right\| < K$, and $\mathbf{w}'\mathbf{w} = O\left(N^{-1}\right)$ by condition (16). Therefore, $\lim_{N\to\infty}\left\|Var\left(\mathbf{w}'\mathbf{x}_t\right)\right\| = 0$. ■

Proof of Corollary 1. Assumption 1 implies that for any $i \in \mathcal{K}$, vector $\boldsymbol{\phi}_{bi}$ satisfies condition (16). It follows from Proposition 1 that

$$\lim_{N\to\infty} Var\left(\boldsymbol{\phi}_{bi}'\mathbf{x}_t\right) = 0 \text{ for } i \in \mathcal{K}. \tag{60}$$

Also (1) implies that

$$x_{it} - \boldsymbol{\phi}_{ai}'\mathbf{x}_{t-1} - u_{it} = \boldsymbol{\phi}_{bi}'\mathbf{x}_{t-1}, \text{ for any } i \in \mathcal{K} \text{ and any } N \geq i. \tag{61}$$

Taking variance of (61) and using (60) now yields (22). ■

Lemma 1 *Suppose that Assumptions 2, 3 and 4 hold. Then for any $p, q \in \{0,1\}$ and for any sequences of non-random vectors $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, such that $\left\|\boldsymbol{\theta}\right\| = O(1)$ and $\left\|\boldsymbol{\varphi}\right\|_1 = O(1)$, as $(N,T) \xrightarrow{j} \infty$ we have*

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p} \xrightarrow{p} 0, \tag{62}$$

and

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q} - E\left(\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\right) \xrightarrow{p} 0, \tag{63}$$

where the process $\boldsymbol{\upsilon}_t$ is defined by (30). Furthermore, if $\left\|\boldsymbol{\theta}\right\| = O\left(N^{-\frac{1}{2}}\right)$ then

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_t \xrightarrow{p} 0, \tag{64}$$

and

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q} - E\left(\sqrt{N}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\right) \xrightarrow{p} 0. \tag{65}$$

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

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

where the subscript N is used to emphasize the number of cross section units,[16] and $\{\mathcal{F}_{Nt}\}$ denotes the array of σ-fields that is increasing in t for each N and κ_{Nt} is measurable with respect to $\mathcal{F}_{Nt}$. Let $\left\{\{c_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ be two-dimensional array of constants and set $c_{Nt} = \frac{1}{T_N}$ for all $t \in \mathbb{Z}$ and $N \in \mathbb{N}$. Note that

$$\begin{aligned} E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{N,t-n}\right)\right]^2\right\} &= \sum_{\ell=\mathsf{m}_{np}}^{\infty} \boldsymbol{\theta}'\boldsymbol{\Phi}^{\ell-p}\boldsymbol{\Sigma}\boldsymbol{\Phi}'^{\ell-p}\boldsymbol{\theta}, \\ &\leq \varsigma_n, \end{aligned} \tag{66}$$

where $\mathsf{m}_{np} = \max\{n, p\}$ and[17]

$$\varsigma_n = \sup_{N\in\mathbb{N}}\left\{\|\boldsymbol{\theta}\|^2\|\boldsymbol{\Sigma}\|\|\boldsymbol{\Phi}\|^{2\left(\mathsf{m}_{np}-p\right)}\sum_{\ell=0}^{\infty}\|\boldsymbol{\Phi}\|^{2\ell}\right\}.$$

Under Assumptions 2, 3 and 4, ς_n has the following properties

$$\varsigma_0 < K, \text{ and } \varsigma_n \to 0 \text{ as } n \to \infty. \tag{67}$$

By Liapunov's inequality, $E\left|E\left(\kappa_{Nt} \mid \mathcal{F}_{N,t-n}\right)\right| \leq \sqrt{E\left\{\left[E\left(\kappa_{Nt} \mid \mathcal{F}_{N,t-n}\right)\right]^2\right\}}$ (Theorem 9.23 of Davidson (1994)). It follows that the two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to the constant array $\{c_{Nt}\}$. Equations (66) and (67) establish array $\{\kappa_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm. This implies uniform integrability.[18] Note that

$$\lim_{N\to\infty}\sum_{t=1}^{T_N} c_{Nt} = \lim_{N\to\infty}\sum_{t=1}^{T_N}\frac{1}{T_N} = 1 < \infty, \tag{68}$$

$$\lim_{N\to\infty}\sum_{t=1}^{T_N} c_{Nt}^2 = \lim_{N\to\infty}\sum_{t=1}^{T_N}\frac{1}{T_N^2} = 0. \tag{69}$$

Therefore array $\left\{\{\kappa_{Nt}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ satisfies conditions of a mixingale weak law,[19] which implies $\sum_{t=1}^{T_N}\kappa_{Nt} \xrightarrow{L_1} 0$, i.e.:

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

as $(N,T) \xrightarrow{j} \infty$ at any rate. Convergence in L_1 norm implies convergence in probability. This completes the proof of the result (62). Under the condition $\|\boldsymbol{\theta}\| = O\left(N^{-\frac{1}{2}}\right)$, result (64) follows from result (62) by noting that $\left\|\sqrt{N}\boldsymbol{\theta}\right\| = O(1)$.

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

[17] We use submultiplicative property of matrix norms ($\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$ for any matrices $\mathbf{A}$, $\mathbf{B}$ such that $\mathbf{AB}$ is well defined) and the fact that the spectral matrix norm is self-adjoint (i.e. $\|\mathbf{A}'\| = \|\mathbf{A}\|$). Note also that Assumption 4 implies $\sum_{\ell=0}^{\infty}\left\|\boldsymbol{\Phi}^{\ell}\right\|^2 = O(1)$.

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

[19] Davidson (1994, Theorem 19.11).

Result (63) is established in a similar fashion. Consider the following two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_{N,t}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, defined by[20]

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

where as before $T_N = T(N)$ is any non-decreasing integer-valued function of N such that $\lim_{N\to\infty} T_N = \infty$. Set $c_{Nt} = \frac{1}{T_N}$ for all $t \in \mathbb{Z}$ and $N \in \mathbb{N}$. Note that

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

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

$$\begin{aligned} E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{N,t-n}\right)\right]^2\right\} &= \sum_{s=\mathfrak{m}_{pn}}^{\infty}\sum_{\ell=\mathfrak{m}_{qn}}^{\infty}\sum_{j=\mathfrak{m}_{pn}}^{\infty}\sum_{d=\mathfrak{m}_{qn}}^{\infty} E\left(\boldsymbol{\theta}_{s-p}'\mathbf{u}_{t-s}\boldsymbol{\varphi}_{\ell-q}'\mathbf{u}_{t-\ell}\boldsymbol{\theta}_{j-p}'\mathbf{u}_{t-j}\boldsymbol{\varphi}_{d-q}'\mathbf{u}_{t-d}\right) - \\ &\quad - \left(\sum_{s=\mathfrak{m}_{pn}}^{\infty}\sum_{\ell=\mathfrak{m}_{qn}}^{\infty} E\left(\boldsymbol{\theta}_{s-p}'\mathbf{u}_{t-s}\boldsymbol{\varphi}_{\ell-q}'\mathbf{u}_{t-\ell}\right)\right)^2. \end{aligned} \tag{70}$$

Using the independence of $\mathbf{u}_t$ and $\mathbf{u}_{t'}$ for any $t \neq t'$ (Assumption 2), we have

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

where

$$\varsigma_{a,n} = \sup_{N\in\mathbb{N}}\left\{\|\boldsymbol{\theta}\|\,\|\boldsymbol{\varphi}\|\,\|\boldsymbol{\Sigma}\|\,\|\boldsymbol{\Phi}\|^{\chi_1(p,n,q)}\sum_{\ell=0}^{\infty}\|\boldsymbol{\Phi}\|^{2\ell}\right\},$$

and $\chi_1(p,n,q) = \max\{0, q-p, n-p\} + \max\{0, p-q, n-q\}$. $\|\boldsymbol{\Sigma}\| = O(1)$ by Assumptions 2 and 3, $\sum_{\ell=0}^{\infty}\|\boldsymbol{\Phi}\|^{2\ell} = O(1)$ by Assumption 4, and $\|\boldsymbol{\theta}\| = O(1)$, $\|\boldsymbol{\varphi}\| \leq \|\boldsymbol{\varphi}\|_1 = O(1)$. $\varsigma_{a,n}$ has the following properties

$$\varsigma_{a,0} < K_a, \text{ and } \varsigma_{a,n} \to 0 \text{ as } n \to \infty. \tag{71}$$

Similarly, since by Assumption 2 $\mathbf{u}_t$ and $\mathbf{u}_{t'}$ are independently distributed for any $t \neq t'$, the first term on the right

[20] As before, $\{\mathcal{F}_{Nt}\}$ denotes the array of σ-fields that is increasing in t for each N and κ_{Nt} is measurable with respect to $\mathcal{F}_{Nt}$.

side of equation (70) is bounded by $\varsigma_{b,n}$:[21]

$$\varsigma_{b,n} = \sup_{N\in\mathbb{N}}\left\{\|\mathbf{B}\|\cdot\|\boldsymbol{\theta}\|^2\|\boldsymbol{\varphi}\|^2\sum_{\ell=\max\{p,q,n\}}\|\boldsymbol{\Phi}\|^{2(\ell-p)+2(\ell-q)}+2\varsigma_{a,n}^2+\right.$$
$$\left.+\|\boldsymbol{\theta}\|^2\|\boldsymbol{\Sigma}\|^2\|\boldsymbol{\varphi}\|^2\|\boldsymbol{\Phi}\|^{2\chi_2(p,n,q)}\left(\sum_{\ell=0}^{\infty}\|\boldsymbol{\Phi}\|^{2\ell}\right)^2\right\},$$

where $\chi_2(p,n,q)=\max\{0,n-p\}+\max\{n-q,0\}$, $\mathbf{B}$ is an $N\times N$ matrix with the (i,j) element given by $\|\boldsymbol{\Psi}_{ij}\|$, and $\boldsymbol{\Psi}_{ij}$ is an $N\times N$ matrix of fourth moments with its (n,s) element given by $E(u_{it}u_{jt}u_{nt}u_{st})$. It follows from Assumptions 2-4 that $\varsigma_{b,n}$ has following properties[22]

$$\varsigma_{b,0}<K_b, \text{ and } \varsigma_{b,n}\to 0 \text{ as } n\to\infty. \tag{72}$$

$E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}}\mid\mathcal{F}_{N,t-n}\right)\right]^2\right\}$ is therefore bounded by $\varsigma_n=\varsigma_{a,n}+\varsigma_{b,n}$. Equations (71) and (72) establish

$$\varsigma_0<K,\ \varsigma_n\to 0 \text{ as } n\to\infty. \tag{73}$$

By Liapunov's inequality, $E\left|E\left(\kappa_{Nt}\mid\mathcal{F}_{N,t-n}\right)\right|\leq\sqrt{E\left\{\left[E\left(\kappa_{Nt}\mid\mathcal{F}_{N,t-n}\right)\right]^2\right\}}$ (Theorem 9.23 of Davidson (1994)). It follows that the two-dimensional array $\left\{\{\kappa_{Nt},\mathcal{F}_{N,t}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, is L_1-mixingale with respect to a constant array $\{c_{Nt}\}$. Furthermore, (73) establishes that array $\{\kappa_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm. This implies uniform integrability.[23] Since also equations (68) and (69) hold, array $\left\{\{\kappa_{Nt},\mathcal{F}_{N,t}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ satisfies conditions of a mixingale weak law (Theorem 19.11 of Davidson (1994)), which implies $\sum_{t=1}^{T_N}\kappa_{Nt}\xrightarrow{L_1}0$, i.e.

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}-E\left(\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\right)\xrightarrow{L_1}0,$$

as $(N,T)\xrightarrow{j}\infty$. Convergence in L_1 norm implies convergence in probability. This completes the proof of result (63). Under $\|\boldsymbol{\theta}\|=O\left(N^{-\frac{1}{2}}\right)$, result (65) follows from result (63) by noting that $\left\|\sqrt{N}\boldsymbol{\theta}\right\|=O(1)$. ■

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

[22] Matrix $\mathbf{B}$ is symmetric by construction. Therefore $\|\mathbf{B}\|\leq\sqrt{\|\mathbf{B}\|_\infty\|\mathbf{B}\|_1}=\|\mathbf{B}\|_\infty$, where

$$\begin{aligned}
\|\mathbf{B}\|_\infty &= \max_{n\in\{1,..,N\}}\sum_{s=1}^{N}\|\boldsymbol{\Psi}_{ns}\| \\
&= \max_{n\in\{1,..,N\}}\sum_{s=1}^{N}\max_{i\in\{1,..,N\}}\sum_{j=1}^{N}\sum_{\ell=0}^{N}|r_{i\ell}r_{j\ell}r_{s\ell}r_{n\ell}| \\
&\leq \max_{n\in\{1,..,N\}}\sum_{s=1}^{N}\max_{i\in\{1,..,N\}}\sum_{j=1}^{N}\left(\sum_{\ell=0}^{N}|r_{i\ell}r_{j\ell}|\cdot\sum_{\ell'=0}^{N}|r_{s\ell'}r_{n\ell'}|\right) \\
&\leq \left(\max_{n\in\{1,..,N\}}\sum_{s=1}^{N}\sum_{\ell'=0}^{N}|r_{s\ell'}r_{n\ell'}|\right)\cdot\left(\max_{i\in\{1,..,N\}}\sum_{j=1}^{N}\sum_{\ell=0}^{N}|r_{i\ell}r_{j\ell}|\right) \\
&\leq \left\|\mathbf{R}\mathbf{R}'\right\|_\infty^2\leq\|\mathbf{R}\|_\infty^2\|\mathbf{R}\|_1^2<K
\end{aligned}$$

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

Lemma 2 *Suppose that* $\mathbf{x}_t$ *is generated by model (25), and that Assumptions 2 to 8 hold. Then as* $(N,T) \xrightarrow{j} \infty$, *for any* $p,q \in \{0,1\}$, *and for any sequence of non-random vectors* $\boldsymbol{\theta}$ *and* $\boldsymbol{\varphi}$ *with growing dimension* $N \times 1$ *such that* $\|\boldsymbol{\theta}\|_1 = O(1)$ *and* $\|\boldsymbol{\varphi}\|_1 = O(1)$, *we have*

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\mathbf{x}_{t-p} - E\left(\boldsymbol{\theta}'\mathbf{x}_{t-p}\right) \xrightarrow{p} 0, \tag{74}$$

and

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\mathbf{x}_{t-p}\boldsymbol{\varphi}'\mathbf{x}_{t-q} - E\left(\boldsymbol{\theta}'\mathbf{x}_{t-p}\boldsymbol{\varphi}'\mathbf{x}_{t-q}\right) \xrightarrow{p} 0. \tag{75}$$

Furthermore, for $\|\boldsymbol{\theta}\| = O(1)$ *and* $\|\boldsymbol{\varphi}\|_1 = O(1)$ *we have*

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\Gamma}\mathbf{f}_{t-q} \xrightarrow{p} 0, \tag{76}$$

where $\boldsymbol{\upsilon}_t$ *is defined in equation (30).*

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

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

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

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

where

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

$\|\boldsymbol{\theta}\|^2 = O(1)$, $\|\boldsymbol{\Phi}\| \le 1-\epsilon$ by Assumption 4, and $\|\boldsymbol{\Sigma}\| \le \sqrt{\|\boldsymbol{\Sigma}\|_1 \|\boldsymbol{\Sigma}\|_\infty} = O(1)$ by Assumption 3. Furthermore, since $\mathbf{f}_{t-q}$ is covariance stationary and $\|\boldsymbol{\varphi}'\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{\varphi}\| = O(1)$ (by condition $\|\boldsymbol{\varphi}\|_1 = O(1)$ and Assumption 8), we have

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

It follows that ς_n has following properties

$$\varsigma_0 < K \text{ and } \varsigma_n \to 0 \text{ as } n \to \infty.$$

Array $\{\kappa_{Nt}/c_{Nt}\}$ is thus uniformly bounded in L_2 norm. This proves uniform integrability of array $\{\kappa_{Nt}/c_{Nt}\}$. Furthermore, using Liapunov's inequality, two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$. Noting that equations (68) and (69) hold, it follows that the array $\{\kappa_{Nt}, \mathcal{F}_{Nt}\}$ satisfies conditions of a mixingale weak law, (cf Theorem 19.11 of Davidson (1994)), which implies $\sum_{t=1}^{T_N} \kappa_{Nt} \xrightarrow{L_1} 0$. Convergence in L_1 norm implies convergence in probability. This completes the proof of result (76).

Assumption 8 implies that sequence $\boldsymbol{\theta}'\boldsymbol{\alpha}$ (as well as $\boldsymbol{\varphi}'\boldsymbol{\alpha}$) is deterministic and bounded. Vector of endogenous variables $\mathbf{x}_t$ can be written as

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

Process $\mathbf{f}_t$ is independent of $\boldsymbol{\upsilon}_t$. Suppose $(N,T) \xrightarrow{j} \infty$. Processes $\{\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\}$ and $\{\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\}$ are ergodic in mean by Lemma 1 since $\|\boldsymbol{\theta}\| \leq \|\boldsymbol{\theta}\|_1 = O(1)$. Furthermore,

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\Gamma}\mathbf{f}_t - \boldsymbol{\theta}'\boldsymbol{\Gamma}E(\mathbf{f}_t) \xrightarrow{p} 0,$$

and

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\boldsymbol{\Gamma}\mathbf{f}_t\boldsymbol{\varphi}'\boldsymbol{\Gamma}\mathbf{f}_{t-q} - \boldsymbol{\theta}'\boldsymbol{\Gamma}E\left(\mathbf{f}_t\mathbf{f}_{t-q}'\right)\boldsymbol{\Gamma}'\boldsymbol{\varphi} \xrightarrow{p} 0,$$

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

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

by Assumption 8, condition$\|\boldsymbol{\theta}\|_1 = O(1)$ and condition $\|\boldsymbol{\varphi}\|_1 = O(1)$. Sum of bounded deterministic process and independent processes ergodic in mean is a process that is ergodic in mean as well. This completes the proof. ■

Lemma 3 *Let $\mathbf{x}_t$ be generated by model (25), Assumptions 1-8 hold and $(N,T) \xrightarrow{j} \infty$. Then for any $p,q \in \{0,1\}$, for any sequence of non-random weight matrices, $\mathbf{W}$, of growing dimension $N \times m_w$ satisfying conditions (27)-(28), and for any $i \in \mathcal{K}$,*

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathbf{W}'\boldsymbol{\upsilon}_{t-p} \xrightarrow{p} \mathbf{0}, \tag{77}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathbf{W}'\boldsymbol{\upsilon}_{t-p}\overline{\mathbf{x}}_{W,t-q} \xrightarrow{p} \mathbf{0}, \tag{78}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathbf{W}'\boldsymbol{\upsilon}_{t-p}x_{i,t-q} \xrightarrow{p} \mathbf{0}, \tag{79}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathbf{g}_{it}q_{it} \xrightarrow{p} \mathbf{0}, \tag{80}$$

where the process $\boldsymbol{\upsilon}_t$ is defined in equation (30), vector $\mathbf{g}_{it} = \left(1, \boldsymbol{\xi}_{i,t-1}', \overline{\mathbf{x}}_{Wt}', \overline{\mathbf{x}}_{W,t-1}'\right)'$ and q_{it} is defined in equation (37).

Proof. Let $\mathring{\mathbf{w}}_r$ for $r \in \{1,..,m_w\}$ denote the r^{th} column vector of matrix $\mathbf{W}$. Noting that $\left\|\sqrt{N}\mathring{\mathbf{w}}_r\right\| = O(1)$ by granularity condition (27), result

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathring{\mathbf{w}}_r'\boldsymbol{\upsilon}_{t-p} \xrightarrow{p} 0 \tag{81}$$

follows directly from Lemma 1, equation (64). This completes the proof of result (77).

Let $\boldsymbol{\varphi}$ be any sequence of non-random $N \times 1$ dimensional vectors of growing dimension such that $\|\boldsymbol{\varphi}\|_1 = O(1)$. We have

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \mathbf{x}_{t-q} = \frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \left(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{f}_{t-q} + \boldsymbol{\upsilon}_{t-q}\right). \tag{82}$$

Since $\left\|\sqrt{N}\mathring{\mathbf{w}}_r\right\| = O(1)$ for any $r \in \{1, .., m_w\}$ by condition (27), we can use Lemma 1, result (65), which implies

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\upsilon}_{t-q} - E\left(\mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\upsilon}_{t-q}\right) \xrightarrow{p} 0. \tag{83}$$

Sequence $\{\boldsymbol{\varphi}'\boldsymbol{\alpha}\}$ is deterministic and bounded in N, and therefore it follows from Lemma 1, result (64), that

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\alpha} \xrightarrow{p} 0. \tag{84}$$

Similarly, Lemma 2 equation (76) implies

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \boldsymbol{\Gamma}\mathbf{f}_{t-q} \xrightarrow{p} 0. \tag{85}$$

Results (83), (84) and (85) establish

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T} \mathring{\mathbf{w}}_r' \boldsymbol{\upsilon}_{t-p} \boldsymbol{\varphi}' \mathbf{x}_{t-q} \xrightarrow{p} 0. \tag{86}$$

Result (78) follows from equation (86) by setting $\varphi = \mathring{\mathbf{w}}_l$ for any $l \in \{1, .., m_w\}$. Result (79) follows from equation (86) by setting $\boldsymbol{\varphi} = \mathbf{e}_i$ where $\mathbf{e}_i$ is $N \times 1$ dimensional selection vector for the i^{th} element.

Finally, the result (80) directly follows from results (77)-(79). This completes the proof. ■

Lemma 4 *Let $\mathbf{x}_t$ be generated by model (25), Assumptions 1-8 hold, and $(N,T) \xrightarrow{j} \infty$. Then for any sequence of non-random matrices, $\mathbf{W}$, of growing dimension $N \times m_w$ satisfying conditions (27)-(28), and for any $i \in \mathcal{K}$,*

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{g}_{it}\mathbf{g}_{it}' - \mathbf{C}_i \xrightarrow{p} \mathbf{0}, \tag{87}$$

where matrix $\mathbf{C}_i = E\left(\mathbf{g}_{it}\mathbf{g}_{it}'\right)$ and vector $\mathbf{g}_{it} = \left(\boldsymbol{\xi}_{i,t-1}', \overline{\mathbf{x}}_{Wt}', \overline{\mathbf{x}}_{W,t-1}', 1\right)'$.

Proof. Result (87) directly follows from Lemmas 1, 2 and 3. ■

Lemma 5 *Let $\mathbf{x}_t$ be generated by model (25), Assumptions 2-8 hold, and $(N,T) \xrightarrow{j} \infty$. Then for any sequence of non-random weight matrices, $\mathbf{W}$, of growing dimension $N \times m_w$ satisfying conditions (27)-(28), and for any fixed $p \geq 0$,*

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{W}'\boldsymbol{\upsilon}_{t-p} u_{it} \xrightarrow{p} \mathbf{0}, \tag{88}$$

where the process $\boldsymbol{\upsilon}_t$ is defined in equation (30). If in addition $T/N \to \varkappa$, with $0 \leq \varkappa < \infty$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathbf{W}'\boldsymbol{\upsilon}_{t-p} u_{it} \xrightarrow{p} \mathbf{0}. \tag{89}$$

Proof. Let $T_N = T(N)$ be any non-decreasing integer-valued function of N such that $\lim_{N\to\infty} T_N = \infty$ and $\lim_{N\to\infty} T_N/N = \varkappa < \infty$, where $\varkappa \geq 0$ is not necessarily nonzero. Define

$$\boldsymbol{\kappa}_{Nit} = \frac{1}{\sqrt{T_N}} \left\{ \mathbf{W}'\boldsymbol{\upsilon}_{t-p}u_{it} - E\left(\mathbf{W}'\boldsymbol{\upsilon}_{t-p}u_{it}\right) \right\}, \tag{90}$$

where the subscript N is used to emphasize the number of cross section units.[24] Let $\{\mathcal{F}_{Nt}\}$ denotes the array of σ-fields that is increasing in t for each N and κ_{Nt} is measurable with respect to $\mathcal{F}_{Nt}$. First it is established that for any fixed $i \in \mathbb{N}$, the vector array $\left\{\{\boldsymbol{\kappa}_{Nit}/c_{Nt}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is uniformly integrable, where $c_{Nt} = \frac{1}{\sqrt{NT_N}}$. For $p > 0$, we can write

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

where $\|\mathbf{W}\|^2 = O\left(N^{-1}\right)$ by condition (27), $\|\boldsymbol{\Sigma}\| = O(1)$ by Assumption 3, and $\sum_{\ell=0}^{\infty} \left\|\boldsymbol{\Phi}^{\ell}\right\|^2 = O(1)$ by Assumption 4. For $p = 0$, we have

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

where as before $\boldsymbol{\Psi}_{ii}$ is $N \times N$ symmetric matrix with the element (n, s) equal to $E(u_{it}u_{it}u_{nt}u_{st})$. Therefore for $p \geq 0$, the two-dimensional vector array $\{\boldsymbol{\kappa}_{Nit}/c_{Nt}\}$ is uniformly bounded in L_2 norm. This proves uniform integrability of $\{\boldsymbol{\kappa}_{Nit}/c_{Nt}\}$.

$$E\left|E\left(\boldsymbol{\kappa}_{Nit} \mid \mathcal{F}_{N,t-n}\right)\right| = \left\{\begin{array}{cc} \mathbf{0} & \text{for any } n > 0 \text{ and any fixed } p \geq 0 \\ \boldsymbol{\tau}_{m_w} c_{Nt} O(1) & \text{for } n = 0 \text{ and any fixed } p \geq 0 \end{array}\right., \tag{91}$$

and $\left\{\{\boldsymbol{\kappa}_{Nit}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$.[25] Note that

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

and

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

[24] Note that $\mathbf{W}$ and $\boldsymbol{\upsilon}_{t-p}$ change with N, but as before we ommit subscript N here to keep the notation simple.
[25] The last equality in equation (91) takes advatage of Liapunov's inequality. $\boldsymbol{\tau}_{m_w}$ is $m_w \times 1$ dimensional vector of ones.

Therefore for each fixed $i \in \mathbb{N}$, each of the m_w two-dimensional arrays given by the elements of vector array $\left\{\{\boldsymbol{\kappa}_{Nit}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ satisfies conditions of a mixingale weak law[26], which implies

$$\frac{1}{\sqrt{T_N}} \sum_{t=1}^{T_N} \mathbf{W}' \boldsymbol{\upsilon}_{t-p} u_{it} - \sqrt{T_N} E\left(\mathbf{W}' \boldsymbol{\upsilon}_{t-p} u_{it}\right) \xrightarrow{L_1} \mathbf{0}.$$

But

$$\left\| \sqrt{T_N} E\left[\mathbf{W}' \boldsymbol{\upsilon}_{t-p} u_{it}\right] \right\|_1 = \sqrt{T_N} \left\| E\left(\mathbf{W}' \mathbf{u}_t u_{it}\right) \right\|_1 = \sqrt{T_N} O\left(\frac{1}{N}\right) \to 0,$$

since $\lim_{N\to\infty} T_N/N = \varkappa < \infty$. Convergence in L_1 norm implies convergence in probability. This completes the proof of result (89).

Result (88) is established in a very similar fashion. Define new vector array $\mathbf{q}_{Nit} = \frac{1}{\sqrt{T_N}} \kappa_{Nit}$ where κ_{Nit} is array defined in (90) and $i \in \mathbb{N}$ is fixed. Let $T_N = T(N)$ be any non-decreasing integer-valued function of N such that such that $\lim_{N\to\infty} T_N = \infty$. Notice that for any fixed $i \in \mathbb{N}$, vector array $\left\{\{\sqrt{T_N}\mathbf{q}_{Nit}/c_{Nt}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is uniformly integrable because $\left\{\{\boldsymbol{\kappa}_{Nit}/c_{Nt}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is uniformly integrable. Furthermore, $\left\{\{\mathbf{q}_{Nit}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is L_1-mixingale with respect to the constant array $\left\{\frac{1}{\sqrt{T_N}} c_{Nt}\right\}$ since $\left\{\{\boldsymbol{\kappa}_{Nit}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is L_1 mixingale with respect to the constant array $\{c_{Nt}\}$. Note that

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

and

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

Therefore for any fixed $i \in \mathbb{N}$, a mixingale weak law[27] implies

$$\sum_{t=1}^{T_N} \mathbf{q}_{Nit} \xrightarrow{L_1} 0 \text{ as } N \to \infty. \tag{92}$$

Since also

$$E\left(\mathbf{W}' \boldsymbol{\upsilon}_{t-p} u_{it}\right) = O\left(N^{-1}\right),$$

it follows

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

as $N, T \xrightarrow{j} \infty$ at any rate. Convergence in L_1 norm implies convergence in probability. This completes the proof of result (88). ■

Lemma 6 *Let $\mathbf{x}_t$ be generated by model (25), Assumptions 1-8 hold and $(N,T) \xrightarrow{j} \infty$ such that $T/N \to \varkappa$, with $0 \leq \varkappa < \infty$. Then for any sequence of non-random matrices of weights $\mathbf{W}$ of growing dimension $N \times m_w$ satisfying conditions (27)-(28), and for any $i \in \mathcal{K}$, we have,*

[26] See Theorem 19.11 of Davidson (1994).
[27] See Theorem 19.11 of Davidson (1994).

a) under Assumption 9,

$$\frac{1}{\sigma_{ii}}\mathbf{C}_i^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{\mathbf{g}}_{it}u_{it}\xrightarrow{D}N\left(0,\mathbf{I}_{k_i}\right),\tag{93}$$

where $\mathbf{C}_i = E\left(\widetilde{\mathbf{g}}_{it}\widetilde{\mathbf{g}}_{it}'\right)$ *and* $\widetilde{\mathbf{g}}_{it} = \left(\boldsymbol{\xi}_{i,t-1}', \mathbf{f}_t'\overline{\boldsymbol{\Gamma}}_W', \mathbf{f}_{t-1}'\overline{\boldsymbol{\Gamma}}_W', 1\right)'$,

b) under Assumption 10,

$$\frac{1}{\sigma_{ii}\sqrt{T}}\boldsymbol{\Omega}_{vi}^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{v}_{i,t-1}u_{it}\xrightarrow{D}N\left(0,\mathbf{I}_{h_i}\right),\tag{94}$$

where matrix $\boldsymbol{\Omega}_{vi} = E\left(\mathbf{v}_{it}\mathbf{v}_{it}'\right)$ *and vector* $\mathbf{v}_{it} = \mathbf{S}_i'\sum_{\ell=0}^{\infty}\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell}$.

Proof. Let $\mathbf{a}$ be any $k_i \times 1$ dimensional vector such that $\|\mathbf{a}\| = 1$ and define

$$\kappa_{Nt} = \frac{1}{\sqrt{T_N}\sigma_{ii}}\mathbf{a}'\mathbf{C}_i^{-\frac{1}{2}}\widetilde{\mathbf{g}}_{it}u_{it},$$

where $T_N = T\left(N\right)$ is any non-decreasing integer-valued function of N such that $\lim_{N\to\infty} T_N = \infty$ and $\lim_{N\to\infty} T_N/N = \varkappa < \infty$, where $0 \leq \varkappa < \infty$. Array $\{\kappa_{Nt}, \mathcal{F}_{Nt}\}$ is a stationary martingale difference array.[28] Lemmas 1 and 2 imply $\mathbf{a}'\mathbf{C}_i^{-\frac{1}{2}}\widetilde{\mathbf{g}}_{it}$ is ergodic in variance, in particular

$$\frac{1}{T_N}\sum_{t=1}^{T_N}\mathbf{a}'\mathbf{C}_i^{-\frac{1}{2}}\widetilde{\mathbf{g}}_{it}\widetilde{\mathbf{g}}_{it}'\mathbf{C}_i^{-\frac{1}{2}}\mathbf{a}\xrightarrow{p}1.$$

$\widetilde{\mathbf{g}}_{it}$ and u_{it} are independent and the fourth moments of u_{it} are finite. Therefore $\mathbf{a}'\mathbf{C}_i^{-\frac{1}{2}}\widetilde{\mathbf{g}}_{it}u_{it}$ is ergodic in variance and

$$\sum_{t=1}^{T_N}\kappa_{Nt}^2\xrightarrow{p}1.\tag{95}$$

Furthermore, $E\left(\sigma_{ii}^{-1}\mathbf{a}'\mathbf{C}_i^{-1/2}\widetilde{\mathbf{g}}_{it}u_{it}\right)^4 = O\left(1\right)$ and therefore

$$\lim_{N\to\infty}\sum_{t=1}^{T_N}E\left(\kappa_{Nt}^4\right) = 0.$$

Using Liapunov's theorem (Theorem 23.11 of Davidson (1994)), Lindeberg condition[29] holds, which in turn implies

$$\max_{1\leq t\leq T_N}\left|\kappa_{Nt}\right|\xrightarrow{p}0 \text{ as } N\to\infty.\tag{96}$$

Results (95), (96) and the martingale difference array central limit theorem (Theorem 24.3 of Davidson (1994)) establish

$$\sum_{t=1}^{T_N}\kappa_{Nt} = \frac{1}{\sqrt{T_N}\sigma_{ii}}\mathbf{a}'\mathbf{C}_i^{-\frac{1}{2}}\sum_{t=1}^{T_N}\widetilde{\mathbf{g}}_{it}u_{it}\xrightarrow{D}N\left(0,1\right)\tag{97}$$

Since equation (97) holds for any $k_i \times 1$ dimensional vector $\mathbf{a}$ such that $\|\mathbf{a}\| = 1$, result (93) directly follows from equation (97) and Theorem 25.6 of Davidson (1994).

Result (94) can be established in the same way as the result (93), but this time we set $\kappa_{Nt} = \frac{1}{\sqrt{T_N}\sigma_{ii}}\mathbf{a}'\boldsymbol{\Omega}_{vi}^{-\frac{1}{2}}\mathbf{v}_{i,t-1}u_{it}$,

[28] As before, $\{\mathcal{F}_{Nt}\}$ denotes the array of σ-fields that is increasing in t for each N and κ_{Nt} is measurable with respect to $\mathcal{F}_{Nt}$.

[29] See Condition 23.17 of Davidson (1994).

where $\mathbf{a}$ is any $h_i \times 1$ dimensional vector such that $\|\mathbf{a}\| = 1$. ■

Lemma 7 *Let $\mathbf{x}_t$ be generated by model (25), and suppose Assumptions 1-8 hold and $(N,T) \xrightarrow{j} \infty$. Then for any arbitrary matrix of weights, $\mathbf{W}$, satisfying conditions (27)-(28), for any $p, q \in \{0,1\}$, and for any $i \in \mathcal{K}$,*

$$\frac{1}{T}\sum_{t=1}^{T} \overline{\boldsymbol{\upsilon}}_{W,t-p} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{98}$$

$$\frac{1}{T}\sum_{t=1}^{T} \overline{\boldsymbol{\upsilon}}_{W,t-p}\mathbf{f}'_{t-q} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{99}$$

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\upsilon}_{i,t-p}\overline{\boldsymbol{\upsilon}}'_{W,t-q} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{100}$$

$$\frac{1}{T}\sum_{t=1}^{T} \overline{\boldsymbol{\upsilon}}_{W,t-p}\overline{\boldsymbol{\upsilon}}'_{W,t-q} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{101}$$

$$\frac{\boldsymbol{\Upsilon}'_i\mathbf{Q}}{T} = o_p(1). \tag{102}$$

Furthermore,

$$\frac{\mathbf{H}'\mathbf{Q}}{T} = \mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{103}$$

$$\frac{\mathbf{Z}'_i\mathbf{H}}{T} = \frac{\mathbf{Z}'_i\mathbf{Q}}{T}\mathbf{A} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{104}$$

$$\frac{\mathbf{H}'\mathbf{H}}{T} = \mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{105}$$

$$\frac{\mathbf{H}'\mathbf{u}_{i\circ}}{T} = \mathbf{A}'\frac{\mathbf{Q}'\mathbf{u}_{i\circ}}{T} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{106}$$

where

$$\underset{T\times h_i}{\boldsymbol{\Upsilon}_i} = \left(\mathbf{v}_{i0}, \mathbf{v}_{i1}, ..., \mathbf{v}_{i,T-1}\right)', \tag{107}$$

$\mathbf{v}_{it} = \mathbf{S}'_i \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell}$, *$\mathbf{H}$ and $\mathbf{Z}_i$ are defined by (41) and (42), respectively, and $\mathbf{Q}$, $\mathbf{F}$ and $\mathbf{A}$ are defined in equations (43)-(44).*

Proof. Result (98) follows directly from equation (64) of Lemma 1 since the spectral norm of any column vector of the matrix $\mathbf{W}$ is $O\left(N^{-\frac{1}{2}}\right)$. Result (99) follows from result (98) by noting that $\mathbf{f}_t$ is independently distributed of $\overline{\boldsymbol{\upsilon}}_{W,t}$ and all elements of the variance matrix of $\mathbf{f}_t$ are finite. Furthermore, since (by Lemma 1) $\frac{1}{T}\sum_{t=1}^{T}\mathbf{v}_{it} \xrightarrow{p} 0$, equation (102) follows. Results (100) and (101) follows directly from equation (65) of Lemma 1 by noting that

$$\sqrt{N}E\left(\boldsymbol{\upsilon}_{i,t-p}\overline{\boldsymbol{\upsilon}}'_{W,t-q}\right) = O\left(\frac{1}{\sqrt{N}}\right) \tag{108}$$

as well as[30]

$$\sqrt{N}E\left(\overline{\boldsymbol{\upsilon}}_{W,t-p}\overline{\boldsymbol{\upsilon}}'_{W,t-q}\right) = O\left(\frac{1}{\sqrt{N}}\right). \tag{109}$$

In order to prove equations (103)-(106), first note that the row t of the matrix $\mathbf{H} - \mathbf{QA}$ is $\left(0, \overline{\boldsymbol{\upsilon}}'_{Wt}, \overline{\boldsymbol{\upsilon}}'_{W,t-1}\right)$.

[30] Results (108) and (109) are straightforward to establish by taking the row norm and by noting that the granularity conditions (27)-(28) imply $\|\mathbf{W}\|_{\infty} = O\left(N^{-1}\right)$.

Using results (98)-(101), we have

$$\frac{(\mathbf{H}-\mathbf{Q}\mathbf{A})'\mathbf{Q}}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix} 0 \\ \overline{\boldsymbol{\upsilon}}_{Wt} \\ \overline{\boldsymbol{\upsilon}}_{W,t-1} \end{pmatrix}\begin{pmatrix} 1, & \mathbf{f}_t', & \mathbf{f}_{t-1}' \end{pmatrix}\right] = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{110}$$

$$\frac{\mathbf{Z}_i'(\mathbf{H}-\mathbf{Q}\mathbf{A})}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\boldsymbol{\xi}_{i,t-1}\begin{pmatrix} 0 \\ \overline{\boldsymbol{\upsilon}}_{Wt} \\ \overline{\boldsymbol{\upsilon}}_{W,t-1} \end{pmatrix}'\right] = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{111}$$

$$\frac{\mathbf{H}'(\mathbf{H}-\mathbf{Q}\mathbf{A})}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix} \overline{\mathbf{x}}_{Wt} \\ \overline{\mathbf{x}}_{W,t-1} \end{pmatrix}\begin{pmatrix} 0 \\ \overline{\boldsymbol{\upsilon}}_{Wt} \\ \overline{\boldsymbol{\upsilon}}_{W,t-1} \end{pmatrix}'\right] = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{112}$$

$$\frac{(\mathbf{H}-\mathbf{Q}\mathbf{A})'(\mathbf{H}-\mathbf{Q}\mathbf{A})}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix} 0 \\ \overline{\boldsymbol{\upsilon}}_{Wt} \\ \overline{\boldsymbol{\upsilon}}_{W,t-1} \end{pmatrix}\begin{pmatrix} 0 \\ \overline{\boldsymbol{\upsilon}}_{Wt} \\ \overline{\boldsymbol{\upsilon}}_{W,t-1} \end{pmatrix}'\right] = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{113}$$

Equations (110)-(111) establish results (103) and (104). Note that

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

where the last equality uses equations (110) and (112). This completes the proof of result (105).

Equation (92) (see proof or Lemma 5) implies

$$\frac{1}{T}\sum_{t=1}^{T}\overline{\boldsymbol{\upsilon}}_{W,t-p}u_{it} - E\left(\overline{\boldsymbol{\upsilon}}_{W,t-p}u_{it}\right) \xrightarrow{p} 0,$$

as $N, T \xrightarrow{j} \infty$ at any rate. Result (106) follows by noting that $\sqrt{N}E\left(\overline{\boldsymbol{\upsilon}}_{W,t-p}u_{it}\right) = O\left(N^{-\frac{1}{2}}\right)$. This completes the proof. ■

Lemma 8 *Let $\mathbf{x}_t$ be generated by model (25), suppose Assumptions 1-8, 10 hold, and $(N,T) \xrightarrow{j} \infty$. Then for any $i \in \mathcal{K}$, and for any arbitrary matrix of weights, $\mathbf{W}$, satisfying conditions (27)-(28) and Assumption 10, we have*

$$\frac{\mathbf{Q}'\mathbf{Q}}{T} \xrightarrow{p} \boldsymbol{\Omega}_Q, \tag{114}$$

$\boldsymbol{\Omega}_Q$ is non-singular, and

$$\frac{\boldsymbol{\Upsilon}_i'\boldsymbol{\Upsilon}_i}{T} - \boldsymbol{\Omega}_{vi} \xrightarrow{p} 0, \tag{115}$$

where

$$\boldsymbol{\Omega}_Q = \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{\mathbf{f}}(0) & \boldsymbol{\Gamma}_{\mathbf{f}}(1) \\ \mathbf{0} & \boldsymbol{\Gamma}_{\mathbf{f}}(1) & \boldsymbol{\Gamma}_{\mathbf{f}}(0) \end{pmatrix},$$

$\boldsymbol{\Gamma}_{\mathbf{f}}(\ell) = E(\mathbf{f}_t\mathbf{f}'_{t-\ell})$, $\boldsymbol{\Omega}_{vi} = E(\mathbf{v}_i\mathbf{v}'_i)$, *matrix* $\mathbf{Q}$ *is defined in equation (43), and matrix* $\boldsymbol{\Upsilon}_i = (\mathbf{v}_{i0}, \mathbf{v}_{i1}, ..., \mathbf{v}_{i,T-1})'$.

Proof. Assumption 6 implies matrix $\boldsymbol{\Omega}_Q$ is non-singular. Result (114) directly follows from the ergodicity properties of the covariance stationary time-series process $\mathbf{f}_t$.

Consider now asymptotics $N, T \xrightarrow{j} \infty$ at any rate. Lemma 1 implies that $h_i \times 1$ dimensional vector $\mathbf{v}_{it} = \mathbf{S}'_i\boldsymbol{\upsilon}_t$ is ergodic in variance, in particular $\frac{1}{T}\sum_{t=1}^{T}\mathbf{S}'_i\boldsymbol{\upsilon}_t\boldsymbol{\upsilon}'_t\mathbf{S}_i - E(\mathbf{S}'_i\boldsymbol{\upsilon}_t\boldsymbol{\upsilon}'_t\mathbf{S}_i) \xrightarrow{p} 0$.[31] This completes the proof. ■

Lemma 9 *Let* $\mathbf{x}_t$ *be generated by model (25), suppose Assumptions 1-8 and 10 hold, and* $(N,T) \xrightarrow{j} \infty$. *Then for any* $i \in \mathcal{K}$, *and for any arbitrary matrix of weights* $\mathbf{W}$ *satisfying conditions (27)-(28) and Assumption 10, we have*

$$\frac{\mathbf{Z}'_i\mathbf{M}_H\mathbf{Z}_i}{T} = \frac{\mathbf{Z}'_i\mathbf{M}_Q\mathbf{Z}_i}{T} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{116}$$

$$\frac{\mathbf{Z}'_i\mathbf{M}_Q\mathbf{Z}_i}{T} - \boldsymbol{\Omega}_{vi} \xrightarrow{p} 0, \tag{117}$$

$$\frac{\mathbf{Z}'_i\mathbf{M}_H\mathbf{Q}}{\sqrt{T}} = o_p\left(\sqrt{\frac{T}{N}}\right), \tag{118}$$

$$\frac{\mathbf{Z}'_i\mathbf{M}_H\mathbf{u}_{i\circ}}{\sqrt{T}} = \frac{\boldsymbol{\Upsilon}'_i\mathbf{M}_Q\mathbf{u}_{i\circ}}{\sqrt{T}} + o_p\left(\sqrt{\frac{T}{N}}\right), \tag{119}$$

where $\boldsymbol{\Omega}_{vi}$ *is defined in Assumption 10,* $\mathbf{M}_H$ *and* $\mathbf{Z}_i$ *are defined in (41) and (42), respectively,* $\mathbf{Q}$ *and* $\mathbf{F}$ *are defined by (43), and* $\boldsymbol{\Upsilon}_i = (\mathbf{v}_{i0}, \mathbf{v}_{i1}, ..., \mathbf{v}_{i,T-1})'$.

Proof.

$$\frac{\mathbf{Z}'_i\mathbf{M}_H\mathbf{Z}_i}{T} = \frac{\mathbf{Z}'_i\mathbf{Z}_i}{T} - \frac{\mathbf{Z}'_i\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{+}\frac{\mathbf{H}'\mathbf{Z}_i}{T}. \tag{120}$$

Results (104)-(105) of Lemma 7 imply

$$\frac{\mathbf{Z}'_i\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{+}\frac{\mathbf{H}'\mathbf{Z}_i}{T} = \frac{\mathbf{Z}'_i\mathbf{Q}}{T}\mathbf{A}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^{+}\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Z}_i}{T} + o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{121}$$

Using definition of the Moore-Penrose inverse, it follows

$$\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^{+}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right) = \left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right). \tag{122}$$

Multiply equation (122) by $\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}$ from the left and by $\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}$ from the right to obtain[32]

$$\mathbf{A}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^{+}\mathbf{A}' = \left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}. \tag{123}$$

[31] $\|\mathbf{S}_i\|_1 = O(1)$ by Assumption 1.

[32] Note that $plim_{T\to\infty}\frac{1}{T}\mathbf{Q}'\mathbf{Q}$ is nonsingular by Lemma 8, equation (114). $\mathbf{A}\mathbf{A}'$ is nonsingular, since matrix $\mathbf{A}$ has full row-rank by Assumption 10.

Equations (123) and (121) imply

$$\frac{\mathbf{Z}_i'\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{+}\frac{\mathbf{H}'\mathbf{Z}_i}{T}=\frac{\mathbf{Z}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{Z}_i}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{124}$$

Result (116) follows from equations (124) and (120).

Using (25) we have

$$\mathbf{Z}_i=\boldsymbol{\tau}\boldsymbol{\alpha}_i'\mathbf{S}_i+\mathbf{F}\left(-1\right)\boldsymbol{\Gamma}_i'\mathbf{S}_i+\boldsymbol{\Upsilon}_i. \tag{125}$$

Since $\mathbf{Q}=[\boldsymbol{\tau},\mathbf{F},\mathbf{F}\left(-1\right)]$, it follows

$$\frac{\mathbf{Z}_i'\mathbf{M}_Q\mathbf{Z}_i}{T}=\frac{\boldsymbol{\Upsilon}_i'\mathbf{M}_Q\boldsymbol{\Upsilon}_i}{T}=\frac{\boldsymbol{\Upsilon}_i'\boldsymbol{\Upsilon}_i}{T}+\frac{\boldsymbol{\Upsilon}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\boldsymbol{\Upsilon}_i}{T}. \tag{126}$$

Using equations (102), (114) and (115), result (117) follows directly from (126).

Results (103)-(105) of Lemma 7 imply

$$\frac{\mathbf{Z}_i'\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{+}\frac{\mathbf{H}'\mathbf{Q}}{T}=\frac{\mathbf{Z}_i'\mathbf{Q}}{T}\mathbf{A}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^{+}\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{127}$$

Substituting equation (123), it follows

$$\frac{\mathbf{Z}_i'\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{+}\frac{\mathbf{H}'\mathbf{Q}}{T}=\frac{\mathbf{Z}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{Q}}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{128}$$

Equation (128) implies

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

This completes the proof of result (118).

Results (104)-(106) of Lemma 7 imply

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

Substituting equation (123), it follows

$$\frac{\mathbf{Z}_i'\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{+}\frac{\mathbf{H}'\mathbf{Q}}{T}=\frac{\mathbf{Z}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{u}_{i\circ}}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{129}$$

Noting that $\mathbf{M}_Q\left(\boldsymbol{\tau}\boldsymbol{\alpha}_i'\mathbf{S}_i+\mathbf{F}\boldsymbol{\Gamma}_i'\mathbf{S}_i\right)=\mathbf{0}$ since $\mathbf{Q}=[\boldsymbol{\tau},\mathbf{F},\mathbf{F}\left(-1\right)]$, equations (129) and (125) imply

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

This completes the proof. ■

Lemma 10 *Let $\mathbf{x}_t$ be generated by model (25), and suppose Assumptions 1-8 and 10 hold, and $(N,T)\xrightarrow{j}\infty$. Then for any $i\in\mathcal{K}$, and for any arbitrary matrix of weights, $\mathbf{W}$, satisfying conditions (27)-(28) and Assumption 10, we* *have*

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\boldsymbol{\zeta}_i(-1)}{T} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{130}$$

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\mathbf{u}_{i\circ}}{\sqrt{T}} = \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{i\circ}}{\sqrt{T}} + o_p\left(\sqrt{\frac{T}{N}}\right) + o_p(1), \tag{131}$$

where matrices $\mathbf{M}_H$ *,and* $\mathbf{Z}_i$ *are defined in (41) and (42), respectively,* $\boldsymbol{\Upsilon}_i = (\mathbf{v}_{i0}, \mathbf{v}_{i1}, ..., \mathbf{v}_{i,T-1})$ *and vector* $\boldsymbol{\zeta}_i(-1) = (\zeta_{i,0}, ..., \zeta_{i,T-1})'$.

Proof.

$$\frac{\mathbf{Z}_i'\boldsymbol{\zeta}_i(-1)}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}_{i,t-1}\left(\boldsymbol{\phi}_{ib}'\sum_{\ell=0}^{\infty}\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell-1}\right)'\right],$$

$$\frac{\mathbf{H}'\boldsymbol{\zeta}_i(-1)}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix}\overline{\mathbf{x}}_{Wt}\\ \overline{\mathbf{x}}_{W,t-1}\end{pmatrix}\left(\boldsymbol{\phi}_{ib}'\sum_{\ell=0}^{\infty}\boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell-1}\right)'\right].$$

$\left\|\boldsymbol{\phi}_{ib}\right\|_{\infty} = O\left(N^{-1}\right)$ by Assumption 1, therefore result (130) directly follows from equations (111) and (112).

$$\begin{aligned}\frac{\boldsymbol{\Upsilon}_i'\mathbf{M}_Q\mathbf{u}_{i\circ}}{\sqrt{T}} &= \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{i\circ}}{\sqrt{T}} + \frac{\boldsymbol{\Upsilon}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{u}_{i\circ}}{\sqrt{T}},\\ &= \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{i\circ}}{\sqrt{T}} + o_p(1),\end{aligned} \tag{132}$$

where $\frac{\mathbf{Q}'\mathbf{u}_{i\circ}}{\sqrt{T}} = O_p(1)$, $plim_{T\to\infty}\frac{1}{T}\mathbf{Q}'\mathbf{Q}$ is non-singular by Lemma 8, and $\frac{\boldsymbol{\Upsilon}_i'\mathbf{Q}}{T} = o_p(1)$ by Lemma 7, equation (102). Substituting (132) into equation (119) implies result (131). This completes the proof. ■

Proof of Theorem 1.

a) Substituting for x_{it} in equation (39) yields

$$\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it} + \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}u_{it}\right). \tag{133}$$

With $N, T \xrightarrow{j} \infty$ in any order, Lemma 5 yields[33]

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}u_{it} \xrightarrow{p} \mathbf{0}. \tag{134}$$

Also using Lemmas 3 and 4 we have

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it} \xrightarrow{p} \mathbf{0}, \tag{135}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}' - \mathbf{C}_{(N),i} \xrightarrow{p} \mathbf{0}, \tag{136}$$

[33] $\frac{1}{T}\sum_{t=1}^{T}x_{j,t-1}u_{it} \xrightarrow{p} 0$ since x_{jt} is ergodic in mean by Lemma 2 and u_{it} is independent of $x_{j,t-1}$ for any $N \in \mathbb{N}$ and any $j \in \{1, .., N\}$. Furthermore, using similar arguments, $\frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t u_{it} \xrightarrow{p} \mathbf{0}$.

respectively. Assumption 9 postulates that the matrix $C_{(N),i}$ is invertible and $\left\|\mathbf{C}_{(N),i}^{-1}\right\|$ is bounded in N. It follows from equation (136) that

$$\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}-\mathbf{C}_{(N),i}^{-1}\xrightarrow{p}\mathbf{0}. \tag{137}$$

Result $\widehat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_i\xrightarrow{p}0$ directly follows from equations (134), (135) and (137).

b) Multiplying equation (133) by $\sqrt{T}$ yields

$$\sqrt{T}\left(\widehat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_i\right)=\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it}+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}u_{it}\right). \tag{138}$$

With $(N,T)\xrightarrow{j}\infty$ such that $T/N\rightarrow\varkappa<\infty$, Lemma 3 can be used to show that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it}\xrightarrow{p}\mathbf{0}. \tag{139}$$

Since $\left\|\mathbf{C}_{(N),i}^{-1}\right\|=O(1)$, equations (137) and (139) now yield

$$\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it}\xrightarrow{p}\mathbf{0}. \tag{140}$$

Lemma 5 establishes

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overline{\boldsymbol{v}}_{\mathbf{W},t-p}u_{it}\xrightarrow{p}\mathbf{0}\text{ for }p\in\{0,1\}. \tag{141}$$

It follows from equation (141) that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{g}_{it}-\widetilde{\mathbf{g}}_{it}\right)u_{it}\xrightarrow{p}0, \tag{142}$$

where $\widetilde{\mathbf{g}}_{it}=\left(\boldsymbol{\xi}_{i,t-1}',\mathbf{f}_t'\overline{\boldsymbol{\Gamma}}_W',\mathbf{f}_{t-1}'\overline{\boldsymbol{\Gamma}}_W',1\right)'$. Lemma 6 establishes that

$$\frac{1}{\sigma_{(N),ii}}\mathbf{C}_{(N),i}^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{\mathbf{g}}_{it}u_{it}\xrightarrow{D}N\left(\mathbf{0},\mathbf{I}_{k_i}\right), \tag{143}$$

Equations (137), (140), (142) and (143) imply result (45).

c) Lemma 4 establishes $\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'-\mathbf{C}_{(N),i}\xrightarrow{p}\mathbf{0}$. The estimated residuals from auxiliary regression (38) are equal to $\widehat{u}_{it}=u_{it}-\mathbf{g}_{it}'\left(\widehat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_i\right)$, which implies

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{it}^2=\frac{1}{T}\sum_{t=1}^{T}u_{it}^2-2\left(\widehat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_i\right)'\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}u_{it}+\left(\widehat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_i\right)'\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)\left(\widehat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_i\right), \tag{144}$$

where $\frac{1}{T}\sum_{t=1}^{T}u_{it}^2-\sigma_{(N),ii}^2\xrightarrow{p}0$, $\widehat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_i\xrightarrow{p}0$ is established in part (a) of this proof, $\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'-\mathbf{C}_{(N),i}\xrightarrow{p}\mathbf{0}$ is established in Lemma 4, and $\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}u_{it}\xrightarrow{p}\mathbf{0}$ is established in equation (134). This completes the proof.

■

Proof of Theorem 2. Vector $\mathbf{x}_{i\circ}$ can be written, using system (25), as

$$\mathbf{x}_{i\circ}=\boldsymbol{\tau}\left(\alpha_i-\boldsymbol{\delta}_i'\mathbf{S}_i'\boldsymbol{\alpha}\right)+\mathbf{Z}_i\boldsymbol{\delta}_i+\mathbf{F}\gamma_i-\mathbf{F}(-1)\boldsymbol{\Gamma}'\mathbf{S}_i\boldsymbol{\delta}_i+\boldsymbol{\zeta}_i(-1)+\mathbf{u}_{i\circ}, \tag{145}$$

where $\boldsymbol{\zeta}_i\left(-1\right) = \left(\zeta_{i0}, ..., \zeta_{i,T-1}\right)'$. Substituting equation (145) into the partition least squares formula (40) and noting that by Lemma 9,

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\mathbf{Q}}{\sqrt{T}} = o_p\left(\sqrt{\frac{T}{N}}\right), \tag{146}$$

it follows

$$\sqrt{T}\left(\widehat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i\right) = \left(\frac{\mathbf{Z}_i'\mathbf{M}_H\mathbf{Z}_i}{T}\right)^{-1}\left[\frac{\mathbf{Z}_i'\mathbf{M}_H\left(\mathbf{u}_{i\circ} + \boldsymbol{\zeta}_i\left(-1\right)\right)}{\sqrt{T}} + o_p\left(\sqrt{\frac{T}{N}}\right)\right]. \tag{147}$$

Lemma 9 also establishes that

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\mathbf{Z}_i}{T} - \boldsymbol{\Omega}_{vi} \xrightarrow{p} \mathbf{0}, \text{ as } N, T \xrightarrow{j} \infty \text{ at any rate,} \tag{148}$$

where $\boldsymbol{\Omega}_{vi} = E\left(\mathbf{v}_{it}\mathbf{v}_{it}'\right)$ is non-singular by Assumption 10.

Consider now asymptotics $N, T \xrightarrow{j} \infty$ such that $T/N \rightarrow \varkappa < \infty$. Lemma 10 establishes

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\boldsymbol{\zeta}_i\left(-1\right)}{\sqrt{T}} \xrightarrow{p} 0, \tag{149}$$

and

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\mathbf{u}_{i\circ}}{\sqrt{T}} = \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{i\circ}}{\sqrt{T}} + o_p\left(\sqrt{\frac{T}{N}}\right) + o_p\left(1\right), \tag{150}$$

where $\boldsymbol{\Upsilon}_i = \left(\mathbf{v}_{i0}, ..., \mathbf{v}_{i,T-1}\right)'$. Also from Lemma 6

$$\frac{1}{\sigma_{ii}\sqrt{T}}\boldsymbol{\Omega}_{vi}^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{v}_{i,t-1}u_{it} \xrightarrow{D} N\left(\mathbf{0}, \mathbf{I}_{h_i}\right). \tag{151}$$

The desired result (48) now follows from (147)-(151). ■