EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 998 / JANUARY 2009

INFINITE-DIMENSIONAL VARs AND FACTOR MODELS

by Alexander Chudik
and M. Hashem Pesaran

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 998 / JANUARY 2009

INFINITE-DIMENSIONAL VARs AND FACTOR MODELS[1]

by Alexander Chudik[2]
and M. Hashem Pesaran[3]

In 2009 all ECB publications feature a motif taken from the €200 banknote.

This paper can be downloaded without charge from http://www.ecb.europa.eu or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=1316219.

1 We are grateful to Elisa Tosetti for helpful comments and suggestions. This version has also benefited from comments by the seminar participants at University of California, San Diego, University of Southern California, Columbia, Leicester, McGill, Princeton, Pennsylvania, and Stanford. We would also like to acknowledge Takashi Yamagata for carrying out the estimation of the results reported in Section 6. The views expressed in this paper do not necessarily reflect the views of the European Central Bank or the Eurosystem.

2 University of Cambridge, Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom and European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany; e-mail: Alexander.Chudik@ecb.europa.eu; http://www.econ.cam.ac.uk/phd/ac474/

3 University of Cambridge, CIMF and USC; Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom; e-mail: mhp1@econ.cam.ac.uk; http://www.econ.cam.ac.uk/faculty/pesaran/

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Website
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

ISSN 1725-2806 (online)

CONTENTS

Abstract

This paper introduces a novel approach for dealing with the 'curse of dimensionality' in the case of large linear dynamic systems. Restrictions on the coefficients of an unrestricted VAR are proposed that are binding only in a limit as the number of endogenous variables tends to infinity. It is shown that under such restrictions, an infinite-dimensional VAR (or IVAR) can be arbitrarily well characterized by a large number of finite-dimensional models in the spirit of the global VAR model proposed in Pesaran *et al.* (JBES, 2004). The paper also considers IVAR models with dominant individual units and shows that this will lead to a dynamic factor model with the dominant unit acting as the factor. The problems of estimation and inference in a stationary IVAR with unknown number of unobserved common factors are also investigated. A cross section augmented least squares estimator is proposed and its asymptotic distribution is derived. Satisfactory small sample properties are documented by Monte Carlo experiments.

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

JEL Classification: C10, C33, C51

Non-technical summary

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

The objective of this paper is to analyze *large* linear dynamic systems of endogenously determined variables. In particular, we study VAR models where *both* the number of variables (N) and the number of time periods (T) tend to infinity. 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. This paper 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 endogenous 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 neighborhood effects are fixed and do not change with N, but the coefficients corresponding to the remaining units are small, of order $O\left(N^{-1}\right)$. 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 the case where the VAR contains neighborhood effects our specification would converge to a spatiotemporal as $N \to \infty$. Finally, when the VAR includes dominant units the limiting outcome will be a dynamic factor models. Such VAR models of growing dimension ($N \to \infty$) are referred in the paper as the infinite-dimensional VARs, or IVARs.

The analysis of the paper also formally establishes the conditions under which the Global VAR (GVAR) approach proposed by Pesaran, Schuermann and Weiner (*JBES*, 2004) is applicable. In particular, the IVAR featuring all macroeconomic variables could be arbitrarily well approximated by a set of finite-dimensional small-scale models that can be consistently estimated separately in the spirit of the GVAR.

Besides the development of an econometric approach for the analysis of groups that belong to a large interrelated system, the second main contribution of the paper is in considering the problems of the estimation and inference in stationary IVAR models with known as well as an unknown number of unobserved common factors. A simple cross sectional augmented least-squares estimator is proposed and its asymptotic distribution derived. Satisfactory small sample properties are documented by Monte Carlo experiments. As an illustration of the proposed approach we follow the recent empirical analysis of real house prices across the 49 U.S. States by Holly, Pesaran and Yamagata (2008) and show statistically significant dynamic spill over effects of real house prices across the neighboring states.

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, denoted as N, is fixed and the time dimension, denoted as T, tends to infinity. The number of parameters to be estimated grows at the quadratic rate with N and consequently the application of the approach in practice is often limited (by the size of typical datasets) to a handful of variables.

The objective of this paper is to analyze VAR models where *both* N and T tend to infinity. 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 the concept of spatial weights matrix, which links individual units with the rest of the system. Alternatively, one could use techniques whereby prior distributions are imposed on the parameters to be estimated. Bayesian VAR (BVAR) proposed by Doan, Litterman and Sims (1984), for example, use what has become known as 'Minnesota' priors to shrink the parameters space.[1] In most applications, BVARs have been applied to relatively small systems[2] (e.g. Leeper, Sims, and Zha, 1996, considered 13- and 18-variable BVAR), with the focus mainly on forecasting.[3]

The second approach to mitigating the curse of dimensionality is to shrink the data, along the lines of index models. Geweke (1977) and Sargent and Sims (1977) introduced dynamic factor models, which have more recently been generalized to allow for weak cross sectional dependence by Forni and Lippi (2001) and Forni *et al.* (2000, 2004). Empirical evidence suggests that few dynamic factors are needed to explain the co-movement of macroeconomic variables: Stock and Watson (1999, 2002), Giannoni, Reichlin and Sala (2005) conclude that only few, perhaps two, factors explain much of the predictable variations, while Bai and Ng (2007) estimate four factors and Stock and Watson (2005) estimate as much as seven factors. This has led to the development

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

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

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

of factor-augmented VAR (FAVAR) models by Bernanke, Boivin, and Eliasz (2005) and Stock and Watson (2005), among others.

This paper 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 endogenous 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 neighborhood effects are fixed and do not change with N, but the coefficients corresponding to the remaining units are small, of order $O\left(N^{-1}\right)$. 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. This set-up naturally arises in the context of global macroeconomic modelling. When all economies are small and open, using a multicountry DSGE model Chudik (2008) shows that the coefficients of the foreign variables in the rational expectations solution are all of order $O\left(N^{-1}\right)$. In the case where the VAR contains neighborhood effects our specification would converge to a spatiotemporal as $N \to \infty$. Finally, when the VAR includes dominant units the limiting outcome will be a dynamic factor models. Such VAR models will be referred as the infinite-dimensional VARs, or IVARs.

The analysis of the paper also 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. We apply the concept of strong and weak Cross Section (CS) dependence (introduced by Pesaran and Tosetti, 2007) in the context of IVARs and show that only strong CS dependence can be 'transmitted' through $O\left(N^{-1}\right)$ coefficients. This finding links our analysis to the factor models by showing that dominant unit becomes (in the limit) a dynamic common factor for the remaining units in a large system of endogenously determined variables. Static factor models are also obtained as a special case of IVAR. Last but not least, this paper formally establishes the conditions under which the Global VAR (GVAR) approach proposed by Pesaran, Schuermann and Weiner (2004) is applicable.[4] In particular, the IVAR featuring all macroeconomic variables could be arbitrarily well approximated by a set of finite-dimensional small-scale models that can be consistently estimated separately in the spirit of

[4]GVAR model has been used to analyse credit risk in Pesaran, Schuermann, Treutler and Weiner (2006) and Pesaran, Schuerman and Treutler (2007). Extended and updated version of the GVAR by Dees, di Mauro, Pesaran and Smith (2007), which treats Euro area as a single economic area, was used by Pesaran, Smith and Smith (2007) to evaluate UK entry into the Euro. Global dominance of the US economy in a GVAR model is explicitly considered in Chudik (2007). 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.

the GVAR.

Besides the development of an econometric approach for the analysis of groups that belong to a large interrelated system, the second main contribution of the paper is in considering the problems of the estimation and inference in stationary IVAR models with known as well as an unknown number of unobserved common factors. Our set-up 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 follow the recent empirical analysis of real house prices across the 49 U.S. States by Holly, Pesaran and Yamagata (2008) and show statistically significant dynamic spillover effects of real house prices across the neighboring states.

The remainder of this paper is organized as follows. Section 2 outlines IVAR model, introduces limiting restrictions, and provides few examples, which link IVAR with the literature. Section 3 investigates cross section dependence in IVAR models where key asymptotic results are provided. Section 4 focusses on estimation of a stationary IVAR. Section 5 presents Monte Carlo evidence and a spatiotemporal model of the US house prices is presented in Section 6. The final section offers some concluding remarks. Proofs are provided in the Appendix.

A brief word on notation: $|\lambda_1(\mathbf{A})| \geq ... \geq |\lambda_n(\mathbf{A})|$ are the eigenvalues of $\mathbf{A} \in \mathbb{M}^{n \times n}$, where $\mathbb{M}^{n \times n}$ is the space of real-valued $n \times n$ matrices. $\|\mathbf{A}\|_c \equiv \max\limits_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|$ denotes the maximum absolute column sum matrix norm of $\mathbf{A}$, $\|\mathbf{A}\|_r \equiv \max\limits_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$ is the absolute row-sum matrix norm of $\mathbf{A}$.[5] $\|\mathbf{A}\| = \sqrt{\varrho(\mathbf{A}'\mathbf{A})}$ 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] Row i of $\mathbf{A}$ is denoted by $\mathbf{a}_i'$ and the column i is denoted as $\mathring{\mathbf{a}}_i$. All vectors are column vectors. Row i of $\mathbf{A}$ with the i^{th} element replaced by 0 is denoted as $\mathbf{a}_{-i}'$. Row i of $\mathbf{A} \in \mathbb{M}^{n \times n}$ with the element i and the element 1 replaced by 0 is $\mathbf{a}_{-1,-i}' = (0, a_{i2}, ..., a_{i,i-1}, 0, a_{i,i+1}, ..., a_{i,iN})$. The matrix constructed from $\mathbf{A}$ by replacing its first column with a column vector of zeros is denoted by $\dot{\mathbf{A}}_{-1}$. Joint asymptotics in $N, T \to \infty$ are represented by $N, T \overset{j}{\to} \infty$. $a_n = O(b_n)$ states the deterministic sequence a_n is at most of order b_n. $x_n = O_p(y_n)$ states random variable x_n is at most of order y_n in probability. $\mathbb{N}$ is the set of natural numbers, and $\mathbb{Z}$ is the set of integers. We use K

[5] The maximum absolute column sum matrix norm and the maximum absolute row sum matrix norm are sometimes denoted in the literature as $\|\cdot\|_1$ and $\|\cdot\|_\infty$, respectively.

[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 $\mathbf{x}$.

and ϵ to denote large and small positive constants that do not vary with i, t, 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.

2 Infinite-Dimensional Vector Autoregressive Models

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

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

where one lag is assumed for the simplicity of exposition, $\mathbf{A}_0$, $\mathbf{A}_1$ and $\mathbf{A}_2$ are $N \times N$ matrices of unknown coefficients, and innovations collected into $N \times 1$ vector $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ are $IID\,(\mathbf{0}, \mathbf{I}_N)$. The model (1), for example, arises as the rational expectations solution of a multi-country DSGE model. (See, for example, Pesaran and Smith, 2006). Assuming matrix $\mathbf{A}_0$ is invertible, the reduced form of structural model (1) is:

$$\mathbf{x}_t = \boldsymbol{\Phi} \mathbf{x}_{t-1} + \mathbf{u}_t, \tag{2}$$

where the vector reduced-form errors $\mathbf{u}_t$ is given by the following 'spatial' model

$$\mathbf{u}_t = \mathbf{R} \boldsymbol{\varepsilon}_t, \tag{3}$$

$\boldsymbol{\Phi} = \mathbf{A}_0^{-1} \mathbf{A}_1$, and $\mathbf{R} = \mathbf{A}_0^{-1} \mathbf{A}_2$. Focus of this paper is on a sequence of reduced-form models (2) of growing dimension ($N \to \infty$), where the elements of $\boldsymbol{\Phi}$ and $\mathbf{R}$ (and hence the variance matrix of $\mathbf{u}_t$) depend on N. But to simplify the notations we do not show this dependence explicitly, although it will be understood throughout that all the parameters and the dimension of the random variables $\mathbf{x}_t$ and $\mathbf{u}_t$ vary with N, unless otherwise stated. The sequence of models (2) and (3) with $\dim(\mathbf{x}_t) = N$ growing will be referred to as the infinite-dimensional VAR(1) model.

To allow for neighborhood effects it is convenient to decompose $\boldsymbol{\Phi}$ into two components: a sparse $N \times N$ matrix matrix, $\mathbf{\Phi}_a$, with fixed elements (that do not vary with N) which captures the neighborhood effects, and a complement matrix, $\mathbf{\Phi}_b$, characterizing the remaining interactions, so that $\mathbf{\Phi} = \mathbf{\Phi}_a + \mathbf{\Phi}_b$. An example of $\mathbf{\Phi}_a$ is given by

$$\mathbf{\Phi}_a = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 & 0 & & 0 \\ \phi_{12} & \phi_{22} & \phi_{23} & 0 & & 0 \\ 0 & \phi_{32} & \phi_{33} & \phi_{34} & & 0 \\ 0 & 0 & \phi_{43} & \phi_{44} & \ddots & \\ & & & \ddots & \ddots & \phi_{N-1,N} \\ 0 & 0 & 0 & & \phi_{N-1,N} & \phi_{NN} \end{pmatrix}, \tag{4}$$

where the nonzero elements are fixed coefficients that do not change with N. This represents an 'approximate line' model where each unit, except the first and the last unit 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 $i, j \in \{1, .., N\}$ and any $N \in \mathbb{N}$. Equation for unit $i \in \{2, .., N-1\}$ can be written as

$$x_{it} = \phi_{i,i-1} x_{i-1,t-1} + \phi_{ii} x_{i,t-1} + \phi_{i,i+1} x_{i+1,t-1} + \boldsymbol{\phi}'_{bi} \mathbf{x}_{t-1} + u_{it}. \tag{5}$$

Next section shows that under weak CS dependence of errors $\{u_{it}\}$, $\boldsymbol{\phi}'_{bi} \mathbf{x}_{t-1} \overset{q.m.}{\to} 0$, and 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.

The above decomposition of matrix $\mathbf{\Phi}$ is a pivotal example of limiting restrictions developed in this paper. More generally, we have

$$\mathbf{\Phi} = \mathbf{D}\ \mathbf{S} + \mathbf{\Phi}_b, \tag{6}$$

where as before, the individual elements of matrix $\mathbf{\Phi}_b$ are (uniformly) of order $O\left(N^{-1}\right)$,

$$\mathbf{D} = \begin{pmatrix} \boldsymbol{\delta}'_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\delta}'_2 & & \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \boldsymbol{\delta}'_N \end{pmatrix}, \tag{7}$$

$\boldsymbol{\delta}_i$ is an $h_i \times 1$ dimensional vector containing the unknown coefficients to be estimated for unit $i \in \{1, .., N\}$, h_i is *bounded* in N, $h = \sum_{i=1}^{N} h_i$, and $\mathbf{S}$ is a known $h \times N$ matrix partitioned as $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2, ..., \mathbf{S}_N)'$, with $\mathbf{S}_i$ being the $N \times h_i$ selection matrix, which defines the neighbors for unit i as in the example above. $\mathbf{S}$ could also be related to a spatial weights matrix as in the following example.

Example 1 *Consider the following spatiotemporal model*

$$\mathbf{x}_t = \rho_x \mathbf{S}_x \mathbf{x}_{t-1} + \mathbf{u}_t, \tag{8}$$

$$\mathbf{u}_t = \rho_u \mathbf{S}_u \mathbf{u}_t + \boldsymbol{\varepsilon}_t, \tag{9}$$

where $\mathbf{S}_x$ and $\mathbf{S}_u$ are $N \times N$ spatial weights matrices. Spatiotemporal model (8)-(9) is a special case of the model (2)-(3) by setting

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

Remark 1 *Note, however, that not all types of structural models (1) have reduced forms that satisfy the restrictions given by (6). For example, consider the spatiotemporal model:*

$$\mathbf{x}_t = \rho \mathbf{S}_{\mathbf{x}} \mathbf{x}_t + \phi \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{10}$$

Assuming matrix $(\mathbf{I} - \rho_x \mathbf{S}_{\mathbf{x}})$ is invertible, the reduced form of spatiotemporal model (10) is model (2) with $\boldsymbol{\Phi} = \phi \left(\mathbf{I} - \rho_x \mathbf{S}_x\right)^{-1}$ and $\mathbf{R} = (\mathbf{I} - \rho_x \mathbf{S}_x)^{-1}$. For the known spatial weights matrix $\mathbf{S}_x$ and unknown parameters ρ_x and ϕ, the reduced form coefficient matrix $\boldsymbol{\Phi}$ cannot be decomposed as in equation (6), where the matrix $\mathbf{S}$ is assumed to be known, $\{\boldsymbol{\delta}_i\}$ and $\boldsymbol{\Phi}_b$ are unknown, and the elements of $\boldsymbol{\Phi}_b$ are uniformly $O\left(N^{-1}\right)$.

3 Cross Sectional Dependence in Stationary IVAR Models

Here we investigate the correlation pattern of $\{x_{it}\}$, over time, t, and across the cross section units, i. Unlike the time index t which is defined over an ordered integer set, the cross section index, i, refers to an individual unit of an unordered population distributed over space or more generally over networks. To avoid having to order the cross section units we make use of the concepts of weak and strong cross section dependence recently developed in Pesaran and Tosetti (2007, hereafter PT). A process $\{x_{it}\}$ is said to be cross sectionally weakly dependent (CWD) with respect to a pre-determined information set, $\mathcal{I}_{t-1}$, if for all weight vectors, $\mathbf{w}_t = (w_{1t}, ..., w_{Nt})'$, satisfying the 'granularity' conditions[7]

$$\begin{aligned} \|\mathbf{w}_t\| &= O\left(N^{-\frac{1}{2}}\right), & (11) \\ \frac{w_{jt}}{\|\mathbf{w}_t\|} &= O\left(N^{-\frac{1}{2}}\right) \text{ for any } j \leq N, & (12) \end{aligned}$$

we have

$$\lim_{N\to\infty} Var\left(\mathbf{w}'_{t-1}\mathbf{x}_t \mid \mathcal{I}_{t-1}\right) = 0, \text{ for all } t \in \mathcal{T}.$$

Since we will be dealing with stationary processes in what follows we confine our analysis to time invariant weight vectors, $\mathbf{w}$, and information sets, $\mathcal{I} = \emptyset$. Accordingly, we adopt the following concept.

Definition 1 *Stationary process $\{x_{it}, i \in \mathcal{S}, t \in \mathcal{T}, N \in \mathbb{N}\}$, generated by the IVAR model (2), is said to be cross sectionally weakly dependent (CWD), if for any sequence of non-random vectors of weights $\mathbf{w}$ satisfying the granularity conditions (11)-(12),*

$$\lim_{N\to\infty} Var\left(\overline{x}_{wt} \mid \mathcal{I}\right) = \lim_{N\to\infty} Var\left(\overline{x}_{wt}\right) = 0, \tag{13}$$

where $\overline{x}_{wt} = \mathbf{w}'\mathbf{x}_t$. $\{x_{it}\}$ is said to be cross sectionally strongly dependent (CSD) if there exists a sequence of weights vectors $\mathbf{w}$ satisfying (11)-(12) and a constant K such that

$$\lim_{N\to\infty} Var\left(\overline{x}_{wt}\right) \geq K > 0. \tag{14}$$

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

[7] Condition (12) is understood as

$$\frac{w_{jt}}{\|\mathbf{w}_t\|} \leq \frac{K}{\sqrt{N}}, \text{ for any } j \in \{1, .., N\} \text{ and any } N \in \mathbb{N},$$

where constant $K < \infty$ does not depend on N nor on j.

$\ell = 0, 1, ...$ converge. Observe that as $N \to \infty$, $Var(x_{it})$ need not necessarily be bounded in N if $\max_i |\lambda_i(\mathbf{\Phi})| < 1 - \epsilon$. For example, consider the IVAR(1) model with

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

and assume that $var(u_{it})$ is uniformly bounded away from zero as $N \to \infty$. It is clear all eigenvalues of $\mathbf{\Phi}$ are inside the unit circle if and only if $|\varphi| < 1$, regardless the value of the neighboring 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 is required to rule out variances of x_{it} exploding as $N \to \infty$. This is set out in the following assumptions.

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

$$E(\mathbf{u}_t) = \mathbf{0},\ E\left(\mathbf{u}_t\mathbf{u}_t'\right) = \mathbf{\Sigma},$$

where $\mathbf{\Sigma}$ is an $N \times N$ symmetric, nonnegative definite matrix, such that $0 < \sigma_{ii}^2 < K < \infty$ for any $i \in \mathcal{S}$ and $\sigma_{ii}^2 = Var(u_{it})$ is the i-th diagonal element of covariance matrix $\mathbf{\Sigma}$.

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

$$\|\mathbf{\Phi}\| < 1 - \epsilon, \tag{15}$$

$$\|\mathbf{\Sigma}\| = O\left(N^{1-\epsilon}\right), \tag{16}$$

where $\epsilon > 0$ is an arbitrarily small positive constant.

[8] It can be shown that

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

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

Remark 2 *Assumption 1 and equation (16) of Assumption 2 imply* $\{u_{it}\}$ *is CWD.*

Remark 3 *Condition (15) of Assumption 2 is a sufficient condition for covariance stationarity and also delivers bounded variance of* x_{it}*, as* $N \to \infty$*. Note that Assumption 2 also rules out cases where strong cross sectional dependence arises due to a particular unit (or units) since both* $\|\mathbf{\Phi}\|_c < \sqrt{N}\,\|\mathbf{\Phi}\| = O\left(\sqrt{N}\right)$ *and* $\|\mathbf{\Sigma}\|_c$ *cannot diverge to infinity at the rate* N*.*

Proposition 1 *Consider model (2) and suppose that Assumptions 1 and 2 hold. Then for any arbitrary sequence of fixed weights* $\mathbf{w}$ *satisfying condition (11), and for any* $t \in \mathcal{T}$*,*

$$\lim_{N\to\infty} Var\left(\overline{x}_{wt}\right) = 0. \tag{17}$$

Proposition 1 has several interesting implications. Suppose that unit i has a fixed number of neighbors, $j = 1, 2, .., p$, for which coefficients $\phi_{ij} = O(1)$, while the influence of each of the remaining units on the unit i through coefficient matrix $\mathbf{\Phi}$ is small. In particular, consider the following decomposition of the i^{th} row of matrix $\mathbf{\Phi}$, denoted as $\boldsymbol{\phi}_i'$, into a possibly *sparse* vector $\boldsymbol{\phi}_{ai}'$ and the remaining coefficients collected into vector $\boldsymbol{\phi}_{bi}'$:

ASSUMPTION 3 *Let* $\mathcal{K} \subseteq \mathcal{S}$ *be a non-empty index set. For any* $i \in \mathcal{K}$, $\boldsymbol{\phi}_i = \boldsymbol{\phi}_{ai} + \boldsymbol{\phi}_{bi}$*, where*

$$\|\boldsymbol{\phi}_{bi}\| = \left(\sum_{j=1}^{N} \phi_{bij}^2\right)^{1/2} = O\left(N^{-\frac{1}{2}}\right). \tag{18}$$

Remark 4 *Obvious examples of the decomposition of* $\boldsymbol{\phi}_i$ *is when* $\boldsymbol{\phi}_{ai}' = (0, ..., 0, \phi_{ii}, 0, ..., 0)$*, and* $\boldsymbol{\phi}_{bi}' = (\phi_{i1}, ..., \phi_{i,i-1}, 0, \phi_{i,i+1}, ..., \phi_{iN})$*, where* ϕ_{ii} *does not depend on* N*, and the left-right neigbourhood model where* $\boldsymbol{\phi}_{ai}' = (0, ..., 0, \phi_{i,i-1}, \phi_{ii}, \phi_{i,i+1}, 0..., 0)$ *and* $\boldsymbol{\phi}_{bi}' = (\phi_{i1}, ...\phi_{i,i-2}, 0, 0, 0, \phi_{i,i+2}..., \phi_{iN})$ *with* $\phi_{i,i-1}, \phi_{ii}$ *and* $\phi_{i,i+1}$ *being fixed parameters that do not vary with* N*.*

Remark 5 *As we shall see in Section 4, for estimation and inference the following slightly stronger condition on the row norm of* $\boldsymbol{\phi}_{bi}$ *will be needed.*

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

Remark 6 *Assumption 3 implies that for* $i \in \mathcal{K}$, $\sum_{j=1}^{N} \phi_{bij} \leq \|\boldsymbol{\phi}_{bi}\|_c = O(1)$*. Therefore, it is possible for the dependence of each individual unit on the rest of the units in the system to be large* *even if* $\boldsymbol{\phi}_{ai} = \mathbf{0}$*. However, as we shall see below, in the case where* $\{x_{it}\}$ *is a CWD process, vector* $\boldsymbol{\phi}_{bi}$ *does not play a role in the model for the* i^{th} *cross section unit as* $N \rightarrow \infty$*.*

Corollary 1 *Consider model (2) and suppose Assumptions 1-3 hold. Then,*

$$\lim_{N \to \infty} Var\left(x_{it} - \boldsymbol{\phi}'_{ai}\mathbf{x}_{t-1} - u_{it}\right) = 0, \text{ for } i \in \mathcal{K}. \tag{19}$$

Observe that if ϕ_{ii} is the only nonzero element of $\boldsymbol{\phi}_{ai}$, then the regression model for unit i completely de-couples from the rest of the system as $N \rightarrow \infty$, in the sense that

$$\lim_{N \to \infty} Var\left(x_{it} - \phi_{ii}x_{i,t-1} - u_{it}\right) = 0.$$

The above corollary in effect states that in econometric modelling of x_{it} one can ignore the effects of those cross section units that have zero entries in $\boldsymbol{\phi}_{ai}$ as N becomes large, so long as $\mathbf{x}_t$ is a CWD process.[9]

3.1 Contemporaneous Dependence: Spatial or Network Dependence

An important form of cross section dependence is *contemporaneous* dependence across space. The spatial dependence, pioneered by Whittle (1954), models cross section correlations by means of spatial processes that relate each cross section unit to its neighbor(s). Spatial autoregressive and spatial error component models are examples of such processes. (Cliff and Ord, 1973, Anselin, 1988, and Kelejian and Robinson, 1995). However, it is not necessary that proximity is measured in terms of physical space. Other measures such as economic (Conley, 1999, Pesaran, Schuermann and Weiner, 2004), or social distance (Conley and Topa, 2002) could also be employed. All these are examples of dependence across nodes in a physical (real) or logical (virtual) networks. In the case of the IVAR model, defined by (2) and (3), such contemporaneous dependence can be modelled through an $N \times N$ network topology matrix $\mathbf{R}$.[10,11] For example, in the case of a first order spatial

[9] Appropriate rates for $N,T \xrightarrow{j} \infty$ needed for inference about the nonzero parameters in $\boldsymbol{\phi}_{ai}$ are established Section 4.

[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 to capture changes in the network structure over time. However, this will not be pursued here.

moving average model, $\mathbf{R}$ would take the form

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

where ρ_s is the spatial moving average coefficient.

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

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

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

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

3.2 IVAR Models with Strong Cross Sectional Dependence

Strong dependence in IVAR model could arise as a result of CSD errors $\{u_{it}\}$, or could be due to dominant patterns in the coefficients of $\mathbf{\Phi}$, or both. Strong cross section dependence could also arise in the case of residual common factor models where the weighted averages of factor loadings do not converge to zero.[12] Section 4 considers estimation and inference in the case of stationary CSD IVAR models with unobserved common factors and/or deterministic trends. An example of a stationary IVAR model where the column corresponding to unit $i = 1$ in matrices $\mathbf{\Phi}$, and $\mathbf{R}$ is dominant is provided below.

The following assumption postulates that for any i, coefficient vector $\boldsymbol{\phi}_i$ can be decomposed into a sparse vector $\boldsymbol{\phi}_{ai} = (\phi_{i1}, 0, ..., 0, \phi_{ii}, 0, ..., 0)'$ and a vector $\boldsymbol{\phi}_{bi} = \boldsymbol{\phi}_{-1,-i}$ where $\boldsymbol{\phi}_{-1,-i} = \left(0, \phi_{i2}, ..., \phi_{i,i-1}, 0, \phi_{i,i+1}, ..., \phi_{iN}\right)'$.

ASSUMPTION 4 *Let $\mathbf{\Phi} = \sum_{i=1}^{N} \mathring{\boldsymbol{\phi}}_i \mathbf{e}_i' = \mathring{\boldsymbol{\phi}}_1 \mathbf{e}_1' + \dot{\mathbf{\Phi}}_{-1}$ where $\mathring{\boldsymbol{\phi}}_i = (\phi_{1i}, ..., \phi_{Ni})'$ is the i^{th} column of $\mathbf{\Phi}$, $\mathbf{e}_i$ is an $N \times 1$ selection vector for unit i, with the i^{th} element of $\mathbf{e}_i$ being one and the remaining elements zero. Denote by $\dot{\mathbf{\Phi}}_{-1}$ the matrix constructed from $\mathbf{\Phi}$ by replacing its first column with a vector of zeros, and note that $\dot{\mathbf{\Phi}}_{-1} = \sum_{i=2}^{N} \mathring{\boldsymbol{\phi}}_i \mathbf{e}_i'$. Suppose as $N \to \infty$*

$$\left\| \mathring{\boldsymbol{\phi}}_1 \right\|_r = O(1). \tag{20}$$

Further, suppose that

$$\left\| \boldsymbol{\phi}_{-1,-i} \right\|_r = O\left(N^{-1}\right) \text{ uniformly for all } i \in \mathbb{N}, \tag{21}$$

namely there exists a constant K such that

$$\left\| \phi_{-1,-i} \right\|_r < \frac{K}{N} \text{ for any } i \in \mathcal{S} \text{ and any } N \in \mathbb{N}.$$

ASSUMPTION 5 *(Stationarity) $\left\| \mathbf{\Phi} \right\|_r < \rho < 1$ for any $N \in \mathbb{N}$.*

ASSUMPTION 6 *The $N \times 1$ vector of errors $\mathbf{u}_t$ is generated by the 'spatial' model (3). $E\left(\mathbf{u}_t \mathbf{u}_t'\right) = \mathbf{\Sigma} = \mathbf{R}\mathbf{R}'$ is time invariant, where $\mathbf{R} = \sum_{i=1}^{N} \mathring{\mathbf{r}}_i \mathbf{e}_i' = \mathring{\mathbf{r}}_1 \mathbf{e}_1' + \dot{\mathbf{R}}_{-1}$, and $\mathring{\mathbf{r}}_i = (r_{1i}, ..., r_{Ni})'$ is the i^{th} column of matrix $\mathbf{R}$. Suppose as $N \to \infty$*

$$\left\| \dot{\mathbf{R}}_{-1} \right\|^2 = O\left(N^{1-\epsilon}\right), \tag{22}$$

$$\left\| \mathring{\mathbf{r}}_1 \right\|_r = O(1), \tag{23}$$

[12] See Pesaran and Toesetti (2007, Theorem 16).

and

$$\lim_{N\to\infty} \|\mathbf{r}_{-1,-i}\| = 0 \textit{ for any } i \in \mathbb{N}, \tag{24}$$

where ϵ *is an arbitrarily small constant,* $\mathbf{r}_{-1,-i} = (0, r_{i1}, ..., r_{i,i-1}, 0, r_{i,i+1}, ..., r_{iN})'$*, and* r_{ij} *denotes the* (i,j) *element of matrix* $\mathbf{R}$*.*

Remark 7 *Assumptions 4 and 6 imply matrix* $\mathbf{\Phi}$ *has one dominant column and matrix* $\mathbf{R}$ *has at least one dominant column, but the absolute column sum for only one column could rise with* N *at the rate* N*. Part (21) of Assumption 4 allows the equation for unit* $i \neq 1$ *to de-couple from the equations for units* $j \neq 1$*, for any* $j \neq i$*, as* $N \to \infty$*.*

Remark 8 *Using the maximum absolute column/row sum matrix norms rather than eigenvalues in principle allows us to make a distinction between cases where dominant effects are due to a particular unit (or a few units), and when there is a pervasive unobserved factor that makes all column/row sums unbounded. Eigenvalues of the covariance matrix* $\mathbf{\Omega}$ *will be unbounded in both cases and it will not be possible from the knowledge of the rate of the expansion of the eigenvalues of* $\mathbf{\Omega}$*,* $\mathbf{\Phi}$ *and/or* $\mathbf{R}$ *to known which one of the two cases are in fact applicable.*

Remark 9 *As it will become clear momentarily,* conditional *on* x_{1t} *and its lagged values, process* $\{x_{it}\}$ *become cross sectionally weakly dependent. We shall therefore refer to unit* $i = 1$ *as the dominant unit.*

Remark 10 *It follows that under Assumptions 5 and 6 the IVAR model specified by (2) and (3) is stationary for any* N*, and the variance of* x_{it} *will be uniformly bounded.*

Proposition 2 *Under Assumptions 4-6 and as* $N \to \infty$*, equation for the dominant unit* $i = 1$ *in the IVAR model defined by (2) and (3) reduces to*

$$x_{1t} - \vartheta(L, \mathbf{e}_1)\, \varepsilon_{1t} \overset{q.m.}{\to} 0, \tag{25}$$

where $\vartheta(L, \mathbf{e}_1) = \sum_{\ell=0}^{\infty} \left(\mathbf{e}_1' \mathbf{\Phi}^{\ell} \mathring{\mathbf{r}}_1\right) L^{\ell}$*. Furthermore, for any fixed sequence of weights* $\mathbf{w}$ *satisfying condition (11),*

$$\overline{x}_{wt} - \vartheta(L, \mathbf{w})\, \varepsilon_{1t} \overset{q.m.}{\to} 0. \tag{26}$$

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

$$a(L, p_1) \approx \vartheta^{-1}(L, \mathbf{e}_1) \tag{27}$$

be an approximation of $\vartheta^{-1}(L, \mathbf{e}_1)$. Then equation for unit $i = 1$ can be written as

$$a(L, p_1) x_{1t} \approx \varepsilon_{1t}. \tag{28}$$

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

Proposition 3 *Consider system (2), let Assumptions 4-6 hold and suppose that the lag polynomial $\vartheta(L, \mathbf{e}_1)$ defined in Proposition 2 is invertible. Then as $N \to \infty$, equations for cross section unit $i \neq 1$ in the IVAR model defined by (2) and (3) reduce to*

$$(1 - \phi_{ii}L) x_{it} - \beta_i(L) x_{1t} - r_{ii}\varepsilon_{it} \overset{q.m.}{\to} 0, \text{ for } i = 2, 3, ... \tag{29}$$

where

$$\beta_i(L) = \phi_{i1}L + \left[r_{i1} + \vartheta\left(L, \boldsymbol{\phi}_{-1,-i}\right) L\right] \vartheta^{-1}(L, \mathbf{e}_1),$$

and

$$\vartheta\left(L, \boldsymbol{\phi}_{-1,-i}\right) = \sum_{\ell=0}^{\infty} \left(\boldsymbol{\phi}'_{-1,-i} \boldsymbol{\Phi}^{\ell} \mathring{\mathbf{r}}_1\right) L^{\ell}, \text{ for } i > 1.$$

Remark 11 *Exclusion of the current value of x_{1t} from (29) is justified only if $r_{i1} = 0$. But even in this case x_{it} will depend on lagged values of x_{1t}.*

Remark 12 *Cross section unit* 1 *becomes (in the limit) a* dynamic common factor *for the remaining units in the IVAR model. Note that setting $x_{1t} = f_t$, (29) can be written as*[13]

$$(1 - \phi_{ii}L) x_{it} \approx r_{ii}\varepsilon_{it} + \beta_i(L) f_t, \text{ for } i > 1. \tag{30}$$

Remark 13 *Conditional on $\{x_{1t}, x_{1,t-1}, x_{1,t-2}, ...\}$, the process $\{x_{it}\}$ for $i > 1$ is CWD.*

[13] x_{1t} could be equivalently approximated by cross sectional weighted averages of $\mathbf{x}_t$ and its lags, namely $\overline{x}_{wt}, \overline{x}_{w,t-1}, ...$

Remark 14 *For $\boldsymbol{\phi}_1 = \mathbf{0}$ and $\boldsymbol{\phi}_{-i} = \mathbf{0}$, we obtain from (29) the following static factor model as a special case*

$$(1 - \phi_{ii}L)\, x_{it} \approx r_{ii}\varepsilon_{it} + \left(\frac{r_{i1}}{r_{11}}\right) f_t, \text{ for } i > 1, \tag{31}$$

where $f_t = x_{1t}$.

We now turn our attention to the problems of estimation and inference in IVAR models. In what follows we consider the relatively simple case where there are no dominant units, but allow for the possibility of unobserved common factors. The analysis of IVAR models featuring both unobserved common factors, $\mathbf{f}_t$, and $\mathbf{\Phi}$ matrices with unbounded maximum absolute column sum matrix norms is provided in a supplement, which is available from the authors on request.

4 Estimation of a Stationary IVAR

Assume $\mathbf{x}_t = (x_{1t}, ..., x_{Nt})'$ is generated according to the following factor-augmented IVAR(1):

$$\mathbf{\Phi}(L)(\mathbf{x}_t - \boldsymbol{\alpha} - \mathbf{\Gamma}\mathbf{f}_t) = \mathbf{u}_t, \tag{32}$$

for $t = 1, 2, ..., T$, where the vector of errors $\mathbf{u}_t$ is generated by spatial model (3), namely $\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t$, $\mathbf{\Phi}(L) = \mathbf{I}_N - \mathbf{\Phi}L$, $\mathbf{\Phi}$ is $N \times N$ dimensional matrix of unknown coefficients, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$ is $N \times 1$ dimensional vector of fixed effects, $\mathbf{f}_t$ is $m \times 1$ dimensional vector of unobserved common factors (m is fixed but otherwise unknown), $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_N)'$ is $N \times m$ dimensional matrix of factor loadings with its i-th row denoted as $\boldsymbol{\gamma}_i'$, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}..., \varepsilon_{Nt})'$ is the vector of error terms assumed to be independently distributed of $\mathbf{f}_{t'}$ $\forall t, t' \in \{1, .., T\}$.

Without major difficulties, one could also add observed common factors and/or additional deterministic terms to the equations in (32), but in what follows we abstract from these for expositional simplicity. System (32) models deviations of endogenous variables from common factors in a VAR. Alternatively, one could introduce common factors directly in the residuals. This extension is pursued in Pesaran and Chudik (2008), who focus on estimation of IVARs with dominant units.[14]

Define the following vector of weighted 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

[14] This extension is not straightforward as it introduces infinite-lag polynomials in the corresponding auxiliary cross-section augmented regressions for the individual units.

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[15]

$$\|\mathbf{W}\| = O\left(N^{-\frac{1}{2}}\right), \tag{33}$$

$$\frac{\|\mathbf{w}_j\|}{\|\mathbf{W}\|} = O\left(N^{-\frac{1}{2}}\right) \text{ for any } j \leq N. \tag{34}$$

We consider the problem of estimating the parameters of equation $i \in \mathbb{N}$ in a non-nested sequence of models (32) as both N and T tend to infinity, where $\mathbf{\Phi}$ can be decomposed as $\mathbf{\Phi} = \mathbf{D}\ \mathbf{S} + \mathbf{\Phi}_b$. See (6). As an important example we consider the two-neighbor IVAR model defined by (5). In the case of this model the vector of unknown coefficients of interest for the i^{th} equation is on the i^{th} row of $\mathbf{D}$, defined by (7) namely $\boldsymbol{\delta}_i = \left(\phi_{i,i-1}, \phi_{ii}, \phi_{i,i+1}\right)'$ for $i \notin \{1, N\}$, with $h_i = 3$, and the corresponding $N \times 3$ matrix $\mathbf{S}_i = (\mathbf{e}_{i-1}, \mathbf{e}_i, \mathbf{e}_{i+1})$ in $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2, ..., \mathbf{S}_N)'$, which selects the unit i and the left and the right neighbors of unit i.[16] In what follows we set $\boldsymbol{\xi}_{it} = \mathbf{S}_i'\mathbf{x}_t$, 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 *for any* $N \in \mathbb{N}$ *and* $i \in \{1, .., N\}$, unless otherwise stated.

ASSUMPTION 7 *(General limiting restrictions) The i^{th} row of matrix $\mathbf{\Phi}$ can be decomposed as,*

$$\boldsymbol{\phi}_i = \mathbf{S}_i \boldsymbol{\delta}_i + \boldsymbol{\phi}_{bi}, \tag{35}$$

where

$$\|\boldsymbol{\phi}_{bi}\|_r = \max_{j \in \{1,..,N\}} \left|\phi_{bij}\right| < \frac{K}{N}, \tag{36}$$

$\mathbf{S}_i$ is predetermined and known $N \times h_i$ dimensional matrix, $\|\mathbf{S}_i\|_c < K$, and $h_i < K$. The unknown coefficients and the fixed effects are bounded, namely $\|\boldsymbol{\delta}_i\| < K$ and $|\alpha_i| < K$. For any $i \in \mathbb{N}$, there exists constant $N_0 \in \mathbb{N}$ such that the vector of unknown coefficients $\boldsymbol{\delta}_i$ do not change with $N > N_0$.

[15] Condition (34) is understood as

$$\frac{\|\mathbf{w}_j\|}{\|\mathbf{W}\|} \leq \frac{K}{\sqrt{N}}, \text{ for any } j \in \{1, .., N\} \text{ and any } N \in \mathbb{N},$$

where constant $K < \infty$ does not depend on N nor on j.

[16] The first and the last unit has only one neighbor.

ASSUMPTION 8 *(Stationarity)* $\|\boldsymbol{\Phi}\| < \rho < 1$.

ASSUMPTION 9 *(Weakly dependent errors with finite fourth moments) Innovations* $\{\varepsilon_{jt}\}_{j=1}^{N}$ *are identically and independently distributed with mean* 0*, unit variances and finite fourth moments. Furthermore, matrix* $\mathbf{R}$ *has bounded row and column matrix norms.*

ASSUMPTION 10 *(Available observations) Available observations are* $\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_T$ *with the starting values* $\mathbf{x}_0 = \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}^{\ell} \mathbf{R} \boldsymbol{\varepsilon}(-\ell) + \boldsymbol{\alpha} + \boldsymbol{\Gamma} \mathbf{f}_0$.[17]

ASSUMPTION 11 *(Common factors) Unobserved common factors* $f_{1t}, ..., f_{mt}$ *follow stationary MA*(∞) *processes:*

$$f_{st} = \psi_s(L)\,\varepsilon_{fst}, \text{ for } s = 1, .., m, \tag{37}$$

where polynomials $\psi_s(L) = \sum_{\ell=0}^{\infty} \psi_{s\ell} L^{\ell}$ *are absolute summable,* $\varepsilon_{fst} \sim IID\left(0, \sigma^2_{\varepsilon_{fs}}\right)$*, and the fourth moments of* ε_{fst} *are bounded,* $E\left(\varepsilon^4_{fst}\right) < \infty$. ε_{fst} *is independently distributed of* $\boldsymbol{\varepsilon}_{t'}$ *for any* $t, t' \in \mathcal{T}$*, and any* $s \in \{1, .., m\}$*. Polynomials* $\psi_s(L)$ *and variances* $\sigma^2_{\varepsilon_{fs}}$ *, for* $s \in \{1, .., m\}$*, do not change with* N *and the covariance matrix* $E(\mathbf{f}_t \mathbf{f}_t')$ *is positive definite.*

ASSUMPTION 12 *(Bounded factor loadings)* $\|\boldsymbol{\gamma}_i\| < K$.

Remark 15 *(Eigenvalues of* $\boldsymbol{\Phi}$*) Assumption 8 implies polynomial* $\boldsymbol{\Phi}(L)$ *is invertible (for any* $N \in \mathbb{N}$*) and*

$$\varrho(\boldsymbol{\Phi}) < \rho < 1. \tag{38}$$

This is in line with the first part of Assumption 2 and is therefore sufficient for stationarity of $\mathbf{x}_t$ *for any* $N \in \mathbb{N}$*. Also, as noted in Section 3, this assumption rules out explosive variance of individual elements of the vector* $\mathbf{x}_t$ *as* $N \to \infty$*. Furthermore, since* $\|\boldsymbol{\Phi}\|_c \leq \sqrt{N}\|\boldsymbol{\Phi}\|$*, Assumption 8 rules out cases where* $\|\boldsymbol{\Phi}\|_c$ *diverges to infinity at the rate* N*. Hence the dominance of a particular unit or units due to the coefficient matrix* $\boldsymbol{\Phi}$ *is also ruled out by this assumption.*

Remark 16 *The spectral norm of covariance matrix* $E(\mathbf{u}_t \mathbf{u}_t') = \boldsymbol{\Sigma}$ *is bounded in* N *under Assumption 9, since* $\|\mathbf{R}\mathbf{R}'\| \leq \|\mathbf{R}\mathbf{R}'\|_r \leq \|\mathbf{R}\|_r \|\mathbf{R}\|_c$.[18] $\mathbf{u}_t$ *is therefore a cross sectionally weakly dependent process, which, as shown in Pesaran and Tosetti (2007), includes all commonly used spatial processes in the literature.*

[17] We use notation $\boldsymbol{\varepsilon}(-\ell)$ instead of $\boldsymbol{\varepsilon}_{-\ell}$ in order to avoid possible confusion with the notation used in previous sections.

[18] $\|\mathbf{A}\| \leq \sqrt{\|\mathbf{A}\|_c \|\mathbf{A}\|_r}$ for any matrix $\mathbf{A}$.

Multiplying system (32) by the inverse of polynomial $\boldsymbol{\Phi}(L)$ and then by $\mathbf{W}'$ yields

$$\overline{\mathbf{x}}_{Wt} = \overline{\boldsymbol{\alpha}}_W + \overline{\boldsymbol{\Gamma}}_W \mathbf{f}_t + \overline{\boldsymbol{\upsilon}}_{Wt}, \tag{39}$$

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

$$\boldsymbol{\upsilon}_t = \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}^{\ell} \mathbf{u}_{t-\ell}. \tag{40}$$

Under Assumption 9, $\{\mathbf{u}_t\}$ is weakly cross sectionally dependent and therefore

$$\begin{aligned} \left\| Var\left(\overline{\boldsymbol{\upsilon}}_{Wt}\right) \right\| &= \left\| \sum_{\ell=0}^{\infty} \mathbf{W}'\boldsymbol{\Phi}^{\ell} \boldsymbol{\Sigma} \boldsymbol{\Phi}'^{\ell} \mathbf{W} \right\|, \\ &\leq \left\| \mathbf{W} \right\|^2 \left\| \boldsymbol{\Sigma} \right\| \sum_{\ell=0}^{\infty} \left\| \boldsymbol{\Phi}^{\ell} \right\|^2, \\ &= O\left(N^{-1}\right), \end{aligned} \tag{41}$$

where $\left\| \mathbf{W} \right\|^2 = O\left(N^{-1}\right)$ by condition (33), $\left\| \boldsymbol{\Sigma} \right\| = O(1)$ by Assumption 9 (see Remark 16) and $\sum_{\ell=0}^{\infty} \left\| \boldsymbol{\Phi}^{\ell} \right\| \leq \sum_{\ell=0}^{\infty} \left\| \boldsymbol{\Phi} \right\|^{\ell} = O(1)$ under Assumption 8. This implies $\overline{\boldsymbol{\upsilon}}_{Wt} = O_p\left(N^{-\frac{1}{2}}\right)$ and the unobserved common factors can be approximated as

$$\left(\overline{\boldsymbol{\Gamma}}'_W \overline{\boldsymbol{\Gamma}}_W \right)^{-1} \overline{\boldsymbol{\Gamma}}'_W \left(\overline{\mathbf{x}}_{Wt} - \overline{\boldsymbol{\alpha}}_W \right) = \mathbf{f}_t + O_p\left(N^{-\frac{1}{2}}\right), \tag{42}$$

provided that the matrix $\overline{\boldsymbol{\Gamma}}'_W \overline{\boldsymbol{\Gamma}}_W$ is nonsingular. It can be inferred that the full column rank of $\overline{\boldsymbol{\Gamma}}_W$ is important for the estimation of unit-specific coefficients. Pesaran (2006) shows that the full column rank is, however, not necessary if the object of interest is a panel estimation of the common mean of the individual coefficients as opposed to the consistency of individual-specific estimates.

Using system (32), equation for unit i can be written as:

$$x_{it} - \alpha_i - \boldsymbol{\gamma}'_i \mathbf{f}_t = \boldsymbol{\delta}'_i \mathbf{S}'_i \left(\mathbf{x}_{t-1} - \boldsymbol{\alpha} - \boldsymbol{\Gamma} \mathbf{f}_{t-1} \right) + \zeta_{i,t-1} + u_{it}, \tag{43}$$

where

$$\zeta_{it} = \boldsymbol{\phi}'_{ib} \boldsymbol{\upsilon}_t = O_p\left(N^{-\frac{1}{2}}\right), \tag{44}$$

since the vector $\boldsymbol{\phi}_{ib}$ satisfies condition (33) under Assumption 7. It follows from equation (39) that

$$\boldsymbol{\gamma}_i'\mathbf{f}_t - \boldsymbol{\phi}_{ia}'\boldsymbol{\Gamma}\mathbf{f}_{t-1} = \mathbf{b}_{i1}'\overline{\mathbf{x}}_{Wt} + \mathbf{b}_{i2}'\overline{\mathbf{x}}_{W,t-1} - (\mathbf{b}_{i1}+\mathbf{b}_{i2})'\,\overline{\boldsymbol{\alpha}}_W - \mathbf{b}_{i1}'\overline{\boldsymbol{\upsilon}}_{Wt} - \mathbf{b}_{i2}'\overline{\boldsymbol{\upsilon}}_{W,t-1}, \tag{45}$$

where $\mathbf{b}_{i1} = \boldsymbol{\gamma}_i'\left(\overline{\boldsymbol{\Gamma}}_W'\overline{\boldsymbol{\Gamma}}_W\right)^{-1}\overline{\boldsymbol{\Gamma}}_W'$ and $\mathbf{b}_{i2} = -\boldsymbol{\delta}_i'\mathbf{S}_i'\boldsymbol{\Gamma}\left(\overline{\boldsymbol{\Gamma}}_W'\overline{\boldsymbol{\Gamma}}_W\right)^{-1}\overline{\boldsymbol{\Gamma}}_W'$. Substituting equation (45) into equation (43) yields

$$x_{it} = c_i + \boldsymbol{\delta}_i'\mathbf{S}_i'\mathbf{x}_{t-1} + \mathbf{b}_{i1}'\overline{\mathbf{x}}_{Wt} + \mathbf{b}_{i2}'\overline{\mathbf{x}}_{W,t-1} + u_{it} + q_{it}, \tag{46}$$

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

$$q_{it} = \zeta_{i,t-1} - \mathbf{b}_{i1}'\overline{\boldsymbol{\upsilon}}_{Wt} - \mathbf{b}_{i2}'\overline{\boldsymbol{\upsilon}}_{W,t-1} = O_p\left(N^{-\frac{1}{2}}\right). \tag{47}$$

Consider the following auxiliary regression based on the equation (46):

$$x_{it} = \mathbf{g}_{it}'\boldsymbol{\pi}_i + \epsilon_{it}, \tag{48}$$

where $\epsilon_{it} = u_{it} + q_{it}$, $\boldsymbol{\pi}_i = \left(c_i, \boldsymbol{\delta}_i', \mathbf{b}_{i1}', \mathbf{b}_{i2}'\right)'$ is $k_i \times 1$ vector of coefficients associated to the vector of regressors $\mathbf{g}_{it} = \left(1, \boldsymbol{\xi}_{i,t-1}', \overline{\mathbf{x}}_{Wt}', \overline{\mathbf{x}}_{W,t-1}'\right)'$, and $k_i = 1 + h_i + 2m_w$. Let $\widehat{\boldsymbol{\pi}}_i$ be the least squares (LS) estimator of $\boldsymbol{\pi}_i$:

$$\widehat{\boldsymbol{\pi}}_i = \left(\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\sum_{t=1}^{T}\mathbf{g}_{it}x_{it}. \tag{49}$$

We denote the estimator of coefficients $\boldsymbol{\delta}_i$ given by the corresponding elements of the vector $\widehat{\boldsymbol{\pi}}_i$ as the cross section augmented least squares estimator (or CALS for short), denoted as $\widehat{\boldsymbol{\delta}}_{i,CALS}$. Asymptotic properties of $\widehat{\boldsymbol{\pi}}_i$ (and $\widehat{\boldsymbol{\delta}}_{i,CALS}$ in the case where the number of unobserved common factors is unknown) are the objective of this analysis as N and T tend to infinity.

First we consider 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 13 *(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 $i \in \{1,..,N\}$, $\left(T^{-1}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}$ *exists, $\mathbf{C}_{iN} = E\left(\mathbf{g}_{it}\mathbf{g}_{it}'\right)$ is positive definite, and $\left\|\mathbf{C}_{iN}^{-1}\right\| = O(1)$.*

Remark 17 *Assumption 13 implies* $\overline{\mathbf{\Gamma}}_W$ *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 (32), Assumptions 7-13 hold, and* $\mathbf{W}$ *is any arbitrary (pre-determined) matrix of weights satisfying conditions (33)-(34) and Assumption 13. Then as* $N, T \xrightarrow{j} \infty$ *(in no particular order), the estimator* $\widehat{\boldsymbol{\pi}}_i$ *defined in equation (49) has the following properties.*

a)

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

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

$$\frac{\sqrt{T}}{\sigma_{ii,N}} \mathbf{C}_{iN}^{\frac{1}{2}} \left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i \right) \xrightarrow{D} N\left(\mathbf{0}, \mathbf{I}_{k_1} \right), \tag{50}$$

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

c)

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

where

$$\widehat{\mathbf{C}}_{iN} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_{it} \mathbf{g}_{it}', \quad \widehat{\sigma}_{ii,N}^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_{it}^2, \tag{51}$$

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

Remark 18 *Suppose that in addition to the assumptions of Theorem 1, the limits of* $\mathbf{C}_{iN}^{-1}$ *and* $\sigma_{ii,N}^2$ *, as* $N \to \infty$*, exist and are given by* $\mathbf{C}_{i\infty}^{-1}$*, and* $\sigma_{ii,\infty}^2$*, respectively.*[19] *Then (50) yields*

$$\sqrt{T} \left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i \right) \xrightarrow{D} N\left(\mathbf{0}, \sigma_{ii,\infty}^2 \mathbf{C}_{i,\infty}^{-1} \right). \tag{52}$$

[19] Sufficient condition for $\lim_{N\to\infty} \mathbf{C}_{iN}$ to exist is the existence of the following limits (together with Assumptions 7-12): $\lim_{N\to\infty} \mathbf{S}_i' \boldsymbol{\alpha}$, $\lim_{N\to\infty} \mathbf{S}_i' \mathbf{\Gamma}$, $\lim_{N\to\infty} \mathbf{W}' \mathbf{\Gamma}$, $\lim_{N\to\infty} \mathbf{W}' \boldsymbol{\alpha}$, and $\lim_{N\to\infty} \sum_{\ell=0}^{\infty} \mathbf{S}_i' \mathbf{\Phi}^{\ell} \mathbf{R} \mathbf{R}' \mathbf{\Phi}'^{\ell} \mathbf{S}_i$.

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 (48) 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 observation has no bearings on estimates 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 $\overline{\mathbf{\Gamma}}_W$ has full column rank. Using partition regression formula, the cross sectionally augmented least squares (CALS) estimator of $h_i \times 1$ dimensional vector $\boldsymbol{\delta}_i$ in the auxiliary regression (48) is

$$\widehat{\boldsymbol{\delta}}_{i,CALS} = \left(\mathbf{Z}'_i \mathbf{M}_H \mathbf{Z}_i\right)^{-1} \mathbf{Z}'_i \mathbf{M}_H \mathbf{x}_{i\circ}, \tag{53}$$

where $\mathbf{x}_{i\circ} = (x_{i1}, ..., x_{iT})'$, $\mathbf{Z}_i = \left[\boldsymbol{\xi}_{i1}(-1), \boldsymbol{\xi}_{i2}(-1), ..., \boldsymbol{\xi}_{ih_i}(-1)\right]$, $\boldsymbol{\xi}_{ir}(-1) = \left(\xi_{ir0}, ..., \xi_{i,r,T-1}\right)'$ for $r \in \{1, .., h_i\}$, $\mathbf{M}_H = \mathbf{I}_T - \mathbf{H}\left(\mathbf{H}'\mathbf{H}\right)^{+}\mathbf{H}'$, $\mathbf{H} = \left[\boldsymbol{\tau}, \overline{\mathbf{X}}_W, \overline{\mathbf{X}}_W(-1)\right]$, $\boldsymbol{\tau}$ is $T \times 1$ dimensional vector of ones, $\overline{\mathbf{X}}_W = (\overline{\mathbf{x}}_{W1\circ}, ..., \overline{\mathbf{x}}_{Wm_w\circ})$, $\overline{\mathbf{X}}_W(-1) = [\overline{\mathbf{x}}_{W1}(-1), ..., \overline{\mathbf{x}}_{Wm_w}(-1)]$, $\overline{\mathbf{x}}_{Ws\circ} = (\overline{x}_{Ws1}, ..., \overline{x}_{WsT})'$ and $\overline{\mathbf{x}}_{Ws}(-1) = (\overline{x}_{Ws0}, ..., \overline{x}_{Ws,T-1})'$ for $s \in \{1, .., m_w\}$. Define for future reference vector $\mathbf{v}_{it} = \mathbf{S}'_i \boldsymbol{\upsilon}_t = \boldsymbol{\xi}_{it} - \mathbf{S}'_i \mathbf{\Gamma}\mathbf{f}_t - \mathbf{S}'_i \boldsymbol{\alpha}$, and the following matrices.

$$\mathbf{Q} = \left[\boldsymbol{\tau}, \mathbf{F}, \mathbf{F}(-1)\right], \tag{54}$$

and

$$\underset{(2m+1)\times(2m_w+1)}{\mathbf{A}} = \begin{pmatrix} 1 & \overline{\boldsymbol{\alpha}}'_W & \overline{\boldsymbol{\alpha}}'_W \\ 0 & \overline{\mathbf{\Gamma}}'_W & \mathbf{0}_{m\times m_w} \\ 0 & \mathbf{0}_{m\times m_w} & \overline{\mathbf{\Gamma}}'_W \end{pmatrix}, \tag{55}$$

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 $r \in \{1, .., m\}$.

For this more general case we replace Assumption 13 with the following (and suppress the subscript N to simplify the notations)

ASSUMPTION 14 *(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 $i \in \{1, .., N\}$, $\left(T^{-1}\mathbf{Z}'_i \mathbf{M}_H \mathbf{Z}_i\right)^{-1}$ *exists, $\overline{\mathbf{\Gamma}}_W$ is full column rank matrix, $\mathbf{\Omega}_{vi} = E\left(\mathbf{v}_{it}\mathbf{v}'_{it}\right) = \sum_{\ell=0}^{\infty} \mathbf{S}'_i \mathbf{\Phi}^{\ell} \mathbf{R}\mathbf{R}' \mathbf{\Phi}'^{\ell} \mathbf{S}_i$ is positive definite and $\left\|\mathbf{\Omega}_{vi}^{-1}\right\| = O(1)$.*

Theorem 2 *Let $\mathbf{x}_t$ be generated by model (32), Assumptions 7-12, and 14 hold, and $\mathbf{W}$ is any arbitrary (pre-determined) matrix of weights satisfying conditions (33)-(34). Then if in addition $N, T \xrightarrow{j} \infty$ such that $T/N \rightarrow \varkappa$, with $0 \leq \varkappa < \infty$, the asymptotic distribution of $\widehat{\boldsymbol{\delta}}_{i,CALS}$ defined by (53) is given by.*

$$\frac{\sqrt{T}}{\sigma_{ii}} \boldsymbol{\Omega}_{vi}^{\frac{1}{2}} \left(\widehat{\boldsymbol{\delta}}_{i,CALS} - \boldsymbol{\delta}_i \right) \xrightarrow{D} N \left(\mathbf{0}, \mathbf{I}_{h_i} \right), \tag{56}$$

where $\sigma_{ii}^2 = Var\left(u_{it}\right)$, $\boldsymbol{\Omega}_{vi} = E\left(\mathbf{v}_{it}\mathbf{v}_{it}'\right)$ 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 19 *As before, we also have*

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

where $\boldsymbol{\Omega}_{vi,\infty} = \lim_{N\rightarrow\infty} \boldsymbol{\Omega}_{vi}$, and $\sigma_{ii,\infty}^2 = \lim_{N\rightarrow\infty} \sigma_{ii}^2$.

Extension of the analysis to a IVAR(p) model is straightforward and it is relegated to a Supplement available from the authors upon request.

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. The focus of our analysis will be on the estimation of the individual-specific parameters in an IVAR model that also allows for other interactions that are of order $O(N^{-1})$. The data generating process (DGP) is given by

$$\mathbf{x}_t - \boldsymbol{\gamma} f_t = \boldsymbol{\Phi} \left(\mathbf{x}_{t-1} - \boldsymbol{\gamma} f_{t-1} \right) + \mathbf{u}_t, \tag{57}$$

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 the cross section dependence of the factor loadings. The following bilateral Spatial Autoregressive Model (SAR) is considered.

$$\gamma_i - \mu_\gamma = \frac{a_\gamma}{2}\left(\gamma_{i-1} + \gamma_{i+1}\right) - a_\gamma \mu_\gamma + \eta_{\gamma i}, \tag{58}$$

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

$$\gamma_i = \psi_{\gamma 1}\gamma_{i-1} + \psi_{\gamma 2}\gamma_{i-2} + \eta_{\gamma i}, \tag{59}$$

with $\psi_{\gamma 1} = -2b_\gamma$, $\psi_{\gamma 2} = b^2_\gamma$ and $b_\gamma = \left(1 - \sqrt{1 - a^2_\gamma}\right)/a_\gamma$, generates the same autocorrelations as the bilateral SAR(1) scheme (58). The factor loadings are generated using the unilateral scheme (59) with 50 burn-in data points ($i = -49, ..., 0$) and the initializations $\gamma_{-51} = \gamma_{-50} = 0$. We set $a_\gamma = 0.4$, $\mu_\gamma = 1$, and choose $\sigma^2_{\eta\gamma}$ such that $Var\left(\gamma_i\right) = 1$.[20] The common factors are generated according to the AR(1) process

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

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 $\mathbf{\Phi}_a$ capture own and neighborhood effects as in

$$\mathbf{\Phi}_a = \begin{pmatrix} \varphi_1 & \psi_1 & 0 & 0 & & 0 \\ \psi_2 & \varphi_2 & \psi_2 & 0 & & 0 \\ 0 & \psi_3 & \varphi_3 & \psi_3 & & 0 \\ 0 & 0 & \psi_4 & \varphi_4 & \ddots & \\ & & & \ddots & \ddots & \psi_{N-1} \\ 0 & 0 & 0 & & \psi_N & \varphi_N \end{pmatrix},$$

[20] The variance of factor loadings is given by

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

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

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

$$\lambda_i \sim IIDU\left(-0.1, 0.2\right) \text{ and } \omega_{ij} = \frac{\varsigma_{ij}}{\sum_{j=1}^{N} \varsigma_{ij}}, \tag{60}$$

with $\varsigma_{ij} \sim IIDU\left(0, 1\right)$. This ensures that $\phi_{bij} = O_p(N^{-1})$, and $\lim_{N\to\infty} E\left(\phi_{bij}\right) = 0$, 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 'left' neighbor $i-1$ and the 'right' neighbor $i+1$. The DGP for the i^{th} unit can now be written as

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

To ensure the DGP is stationary we generate $\varphi_i \sim IIDU\left(0.4, 0.6\right)$, and $\psi_i \sim IIDU\left(-0.1, 0.1\right)$ 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 $\left\|\mathbf{\Phi}\right\|_r \leq 0.9$, 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 $Var(u_{it})$ is 1, where $\mathbf{R}_u = (\mathbf{I} - a_u\mathbf{S})^{-1}$, and the spatial weights matrix $\mathbf{S}$ is

$$\mathbf{S} = \begin{pmatrix} 0 & 1 & 0 & 0 & & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & & 0 \\ & & \ddots & \ddots & \ddots & \\ & & & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & & 0 & 1 & 0 \end{pmatrix}. \tag{61}$$

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 $\psi_2 = 0.1$. The cross-section augmented regression for estimating coefficients $\{\phi_2, \psi_2\}$ in the case of the second cross section unit is given by (similar results are also obtained for other cross section units)

$$x_{2t} = c_2 + \psi_2 \left(x_{1,t-1} + x_{3,t-1}\right) + \varphi_2 x_{2,t-1} + \delta_{2,0}\overline{x}_t + \delta_{2,1}\overline{x}_{t-1} + \epsilon_{2t}. \tag{62}$$

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 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 latter case with $\boldsymbol{\gamma} \neq \mathbf{0}$ the cross section dependence also arises via the unobserved common factor f_t and cross-sectionally dependent factor loadings. Additional intermediate cases are also considered, the results of which are available in a Supplement from the authors on request.[21]

[21] Supplement presents experiments with all combination of zero or nonzero coefficient matrix $\mathbf{\Phi}_b$, zero or nonzero

5.2 Monte Carlo Results

Tables 1-2 give the bias ($\times 100$) and RMSE ($\times 100$) of CALS and LS estimators as well as size and power of tests 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 with 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. This is the familiar time series bias where even in the absence of cross section dependence the LS estimator of autoregressive coefficients will be biased 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 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) redundant. 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 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 estimates 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 suggest significant improvement in power as T increases for a number of different

factor loadings γ, and low or high cross section dependence of errors ($a_u = 0.4$ or $a_u = 0.8$).

alternatives.

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

In a recent study Holly, Pesaran and Yamagata (2008, 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:

$$\Delta p_{it} = c_i + \omega_i(p_{i,t-1} - y_{i,t-1}) + \delta_{1i}\Delta p_{i,t-1} + \delta_{2i}\Delta y_{it} + \upsilon_{it}. \tag{63}$$

To take account of unobserved common factors HPY augmented (63) with cross section averages 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 (63) 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, **S,** namely

$$\Delta p_{it} = c_i + \omega_i(p_{i,t-1} - y_{i,t-1}) + \delta_{1i}\Delta p_{i,t-1} + \psi_i \Delta p^s_{i,t-1} + \delta_{2i}\Delta y_{it} + \upsilon_{it}. \tag{64}$$

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 cross section averages. 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 $\Delta p_{i,t-1}^{s}$ 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 bind only in the limit as the number of cross section units (or variables in the VAR) tends to infinity to circumvent the curse of dimensionality. The proposed framework relates to the various approaches considered in the literature. For example when modelling individual households or firms, aggregate variables, such as market returns 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 spatio-dynamic context.

It was established that, under certain conditions on the order of magnitudes of the coefficients in a large dynamic system, and in the absence of common factors, equations for individual units decouple as $N \to \infty$ and can be estimated separately. In the presence of a dominant economic agent or unobserved common factors, individual-specific VAR models can still be estimated separately if conditioned upon observed and unobserved 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 GVAR approach can be motivated as an approximation to an IVAR featuring all macroeconomic variables. This is true for stationary models as well as for systems with integrated variables of order one. Asymptotic distribution of the cross sectionally augmented least-squares (CALS) estimator of the parameters of the unit-specific models was established both in the case where the number of unobserved common factors is known and in the case where it is unknown but fixed. Small sample properties of the proposed CALS estimator were investigated through Monte Carlo simulations, and an empirical application was provided as an illustration of the proposed approach.

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

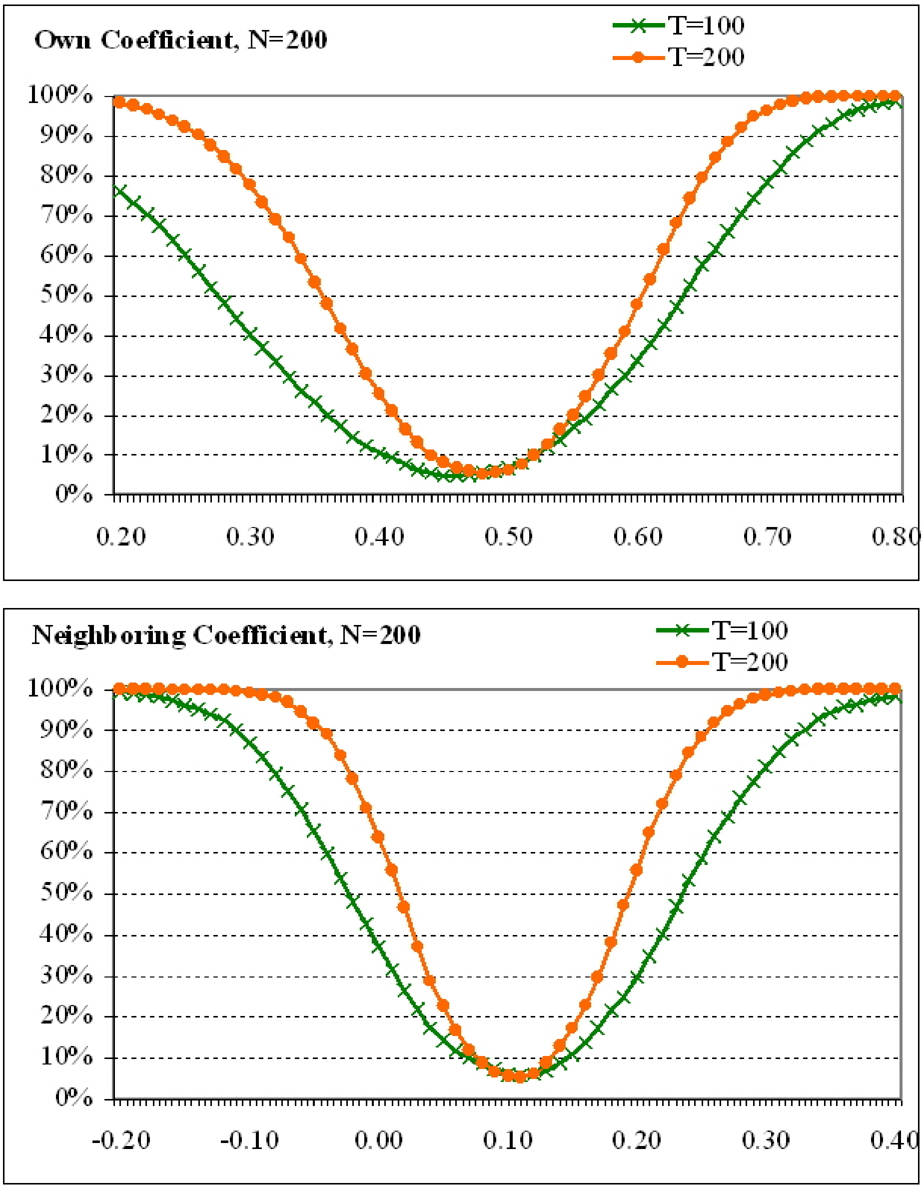

Figure 1: Power Curves for the CALS t-tests of Own Coefficient, φ_2 (the upper chart) and the Neighboring Coefficient, ψ_2 (the lower chart), in the Case of Experiments with $\boldsymbol{\gamma} \neq \mathbf{0}$.

Table 1: MC results for the own coefficient φ_2.

	Bias (×100)					Root Mean Square Errors (×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 (57) 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 $\widehat{\psi}_2$ 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 (×100)					Root Mean Square Errors (×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.* (2008) 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 stands for Mean Group estimates. CCEMG and CCEP signify the Common Correlated Effects Mean Group and Pooled estimates defined in Pesaran (2006). 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

[1] Anselin, L., 1988, Spatial econometrics: Methods and models. Kluwer Academic Publishers, Dordrecht, the Netherlands.

[2] Bai, J. and S. Ng, 2007, Determining the number of primitive shocks in factor models. Journal of Business and Economic Statistics, 25, 52-60.

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

[4] Conley, T.G., 1999, GMM estimation with cross sectional dependence. Journal of Econometrics, 92, 1-45.

[5] Conley, T.G. and G. Topa, 2002, Socio-economic distance and spatial patterns in unemployment. Journal of Applied Econometrics, 17, 303-327.

[6] Canova, F., 1995, Vector autoregressive models: Specification, estimation, inference, and forecasting, chapter 2 in: M.H. Pesaran and M. Wickens, (Eds.), Handbook of applied econometrics: Macroeconomics. Basil Blackwell, Oxford.

[7] Chudik, A., 2007, GVAR approach and the dominance of the US economy. Mimeo, University of Cambridge.

[8] Chudik, A., 2008, Equilibrium of a large multicountry DSGE models. Mimeo, University of Cambridge.

[9] Cliff, A. and J.K. Ord, 1973, Spatial autocorrelation. Pion, London.

[10] Davidson, J., 1994, Stochastic limit theory. Oxford University Press.

[11] Dees, S., di Mauro, F., Pesaran, M.H. and L.V. Smith, 2007, Exploring the international linkages of the Euro Area: A global VAR analysis. Journal of Applied Econometrics, 22, 1-38.

[12] Del Negro, M. and F. Schorfheide, 2004, Priors from general equilibrium models for VARs. International Economic Review, 45, 643-673.

[13] De Mol, C., Giannone, D. and L. Reichlin, 2006, Forecasting using a large number of predictors: Is bayesian regression a valid alternative to principal components? CEPR Discussion Paper No. 5829, September 2006.

[14] Doan, T., Litterman, R. and C. Sims, 1984, Forecasting and conditional projections using realistic prior distributions. Econometric Reviews, 3, 1-100.

[15] Forni, M. and M. Lippi, 2001, The generalized dynamic factor model: Representation theory. Econometric Theory, 17, 1113-41.

[16] Forni, M., Hallin, M., Lippi, M. and L. Reichlin, 2000, The generalized dynamic factor model: Identification and estimation. Review of Economics and Statistics, 82, 540-554.

[17] Forni, M., Hallin, M., Lippi, M. and L. Reichlin, 2004, The generalized dynamic factor model: Consistency and rates. Journal of Econometrics, 119, 231-255.

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

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

[20] Giacomini, R., and H. White, 2006, Tests of conditional predictive ability. Econometrica, 74, 1545–1578.

[21] Giannoni, D., Reichlin, L. and L. Sala, 2005, Monetary policy in real time. CEPR Discussion Paper No. 4981, March 2005.

[22] Holly, S., Pesaran, M.H. and T. Yamagata, 2008, A spatio-temporal model of house prices in the U.S., forthcoming in Journal of Econometrics.

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

[24] Leeper, E.M., Sims, C.A. and T. Zha, 1996, What does monetary policy do? Brookings Papers on Economic Activity, 2, 1-63.

[25] Litterman, R., 1986, Forecasting with bayesian vector autoregressions - five years of experience. Journal of Business and Economics Statistics, 4, 25-38.

[26] Pesaran, M.H., 2006, Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica, 74, 967-1012.

[27] Pesaran, M.H., 2007, A simple panel unit root test in the presence of cross section dependence. Journal of Applied Econometrics, 22, 265-312.

[28] Pesaran, M.H., and A. Chudik, 2008, Estimation of IVARs with dominant units. Under preparation.

[29] Pesaran, M.H., Schuermann, T. and B.J. Treutler, 2007, Global business cycles and credit risk, chapter 9 in: M. Carey and R.M. Stultz, (Eds.), The risks of financial institutions. University of Chicago Press.

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

[31] Pesaran, M.H., Schuermann, T., Treutler, B.J. and S.M. Weiner, 2006, Macroeconomic dynamics and credit risk: A global perspective. Journal of Money, Credit and Banking, 38, 1211-1262.

[32] Pesaran, M.H. and R. Smith, 2006, Macroeconometric modelling with a global perspective. The Manchester School, Supplement, 24-49.

[33] Pesaran, M.H., Smith, L.V. 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.

[34] Pesaran, M.H. and E. Tosetti, 2007, Large panels with common factors and spatial correlations. IZA Discussion Paper No. 3032, September 2007.

[35] 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. Federal Reserve Bank of Minneapolis, Minneapolis.

[36] Stock, J.H. and M.W. Watson, 1999, Forecasting inflation. Journal of Monetary Economics, 44, 293-335.

[37] Stock, J.H. and M.W. Watson, 2002, Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics, 20, 147-162.

[38] Stock, J.H. and M.W. Watson, 2005, Implications of dynamic factor models for VAR analysis. NBER Working Paper No. 11467, June 2005.

[39] 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

$$\mathbf{\Omega} = Var\left(\mathbf{x}_t\right) = \sum_{\ell=0}^{\infty} \mathbf{\Phi}^{\ell} \mathbf{\Sigma} \mathbf{\Phi}'^{\ell}, \tag{65}$$

and, $\|\mathbf{\Omega}\|$ is under Assumptions 1-2 bounded by

$$\|\mathbf{\Omega}\| \leq \|\mathbf{\Sigma}\| \sum_{\ell=0}^{\infty} \|\mathbf{\Phi}\|^{2\ell} = O\left(N^{1-\epsilon}\right). \tag{66}$$

It follows that for any arbitrary nonrandom vector of weights satisfying granularity condition (11),

$$\left\|Var\left(\mathbf{w}'\mathbf{x}_t\right)\right\| = \left\|\mathbf{w}'\mathbf{\Omega}\mathbf{w}\right\| \leq \left\|\varrho\left(\mathbf{\Omega}\right)\left(\mathbf{w}'\mathbf{w}\right)\right\|, \tag{67}$$

where $\varrho\left(\mathbf{\Omega}\right) = \|\mathbf{\Omega}\| = O\left(N^{1-\epsilon}\right)$, and $\mathbf{w}'\mathbf{w} = \|\mathbf{w}\|^2 = O\left(N^{-1}\right)$ by condition (11). This implies $\|Var\left(\mathbf{w}'\mathbf{x}_t\right)\| = O\left(N^{-\epsilon}\right)$ and $\lim_{N\to\infty} \|Var\left(\mathbf{w}'\mathbf{x}_t\right)\| = 0$. ■

Proof of Corollary 1. Assumption 3 implies that for $i \in \mathcal{K}$, vector $\boldsymbol{\phi}_{bi}$ satisfies condition (11). It follows from Proposition 1 that

$$\lim_{N\to\infty} Var\left(\boldsymbol{\phi}_{bi}'\mathbf{x}_t\right) = 0 \text{ for } i \in \mathcal{K}. \tag{68}$$

System (2) implies

$$x_{it} - \boldsymbol{\phi}_a'\mathbf{x}_{t-1} - u_{it} = \boldsymbol{\phi}_b'\mathbf{x}_{t-1}, \text{ for any } i \in \mathcal{S} \text{ and any } N \in \mathbb{N}. \tag{69}$$

Taking variance of (69) and using (68) now yields (19). ■

Proof of Proposition 2. Solving (2) backwards yields

$$\mathbf{x}_t = \sum_{\ell=0}^{\infty} \left(\mathbf{\Phi}^{\ell}\mathbf{R}\right) \boldsymbol{\varepsilon}_{t-\ell}, \tag{70}$$

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

$$x_{1t} - \sum_{\ell=0}^{\infty} \left(\mathbf{e}_1'\mathbf{\Phi}^{\ell}\mathring{\mathbf{r}}_1\right) \varepsilon_{1,t-\ell} = \sum_{\ell=0}^{\infty} \left(\mathbf{e}_1'\mathbf{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\right) \boldsymbol{\varepsilon}_{t-\ell}. \tag{71}$$

Under Assumptions 4-6,

$$\begin{aligned} Var\left(\sum_{\ell=0}^{\infty} \left(\mathbf{e}_1'\mathbf{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\right) \boldsymbol{\varepsilon}_{t-\ell}\right) &= \sum_{\ell=0}^{\infty} \mathbf{e}_1'\mathbf{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}_{-1}'\mathbf{\Phi}'^{\ell}\mathbf{e}_1, \\ &\leq \mathbf{e}_1'\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}_{-1}'\mathbf{e}_1 + \sum_{\ell=1}^{\infty} \mathbf{e}_1'\mathbf{\Phi}^{\ell}\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}_{-1}'\mathbf{\Phi}'^{\ell}\mathbf{e}_1. \end{aligned} \tag{72}$$

But

$$\lim_{N\to\infty}\left\|\mathbf{e}_1'\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}_{-1}'\mathbf{e}_1\right\| = \lim_{N\to\infty}\left\|\mathbf{r}_{-1}\right\|^2 = 0 \tag{73}$$

under Assumption 6. Set $\mathbf{a}_\ell' \equiv \mathbf{e}_1'\mathbf{\Phi}^\ell$ and let $\mathbf{a}_{\ell,-1} = (0, a_{\ell 2}, ..., a_{\ell N})'$. Note that under Assumptions 4-5:

$$\begin{aligned}\left\|\mathbf{a}_\ell\right\|_c &\le \rho^\ell, & (74)\\ a_{\ell 1} &= O(1), & (75)\\ \left\|\mathbf{a}_{\ell,-1}\right\|_r &= O\left(N^{-1}\right), & (76)\end{aligned}$$

for $\ell = 0, 1, 2, ...$ Result (74) follows from Assumption 5 by taking the maximum absolute row-sum matrix norm of $\mathbf{a}_\ell' = \mathbf{e}_1'\mathbf{\Phi}^\ell$.[22] Results (75)-(76) follow by induction directly from Assumptions 4-5. Using (74)-(76), we have

$$\begin{aligned}\left\|\sum_{\ell=1}^{\infty}\mathbf{e}_1'\mathbf{\Phi}^\ell\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}_{-1}'\mathbf{\Phi}'^\ell\mathbf{e}_1\right\| &= \left\|\sum_{\ell=1}^{\infty}\mathbf{a}_\ell'\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}_{-1}'\mathbf{a}_\ell\right\|,\\ &\le \left\|\sum_{\ell=1}^{\infty}a_{\ell 1}^2\mathbf{r}_{-1}'\mathbf{r}_{-1}\right\| + \left\|\sum_{\ell=1}^{\infty}\mathbf{a}_{\ell,-1}'\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}_{-1}'\mathbf{a}_{\ell,-1}\right\|,\\ &\le \left\|\mathbf{r}_{-1}\right\|^2\sum_{\ell=1}^{\infty}\rho^{2\ell} + \left\|\dot{\mathbf{R}}_{-1}\right\|^2\sum_{\ell=1}^{\infty}\left\|\mathbf{a}_{\ell,-1}\right\|_r\left\|\mathbf{a}_{\ell,-1}\right\|_c \qquad (77)\end{aligned}$$

where as before $\lim_{N\to\infty}\left\|\mathbf{r}_{-1}\right\|^2 = 0$ under Assumption 6, $\sum_{\ell=1}^{\infty}\rho^{2\ell} = O(1)$ by Assumption 5, $\left\|\dot{\mathbf{R}}_{-1}\right\|^2 = O\left(N^{1-\epsilon}\right)$ by Assumption 6, and $\sum_{\ell=1}^{\infty}\left\|\mathbf{a}_{\ell,-1}\right\|^2 \le \sum_{\ell=1}^{\infty}\left\|\mathbf{a}_{\ell,-1}\right\|_r\left\|\mathbf{a}_{\ell,-1}\right\|_c = O\left(N^{-1}\right)$ by properties (74)-(76). It follows that $\lim_{N\to\infty}\left\|\sum_{\ell=1}^{\infty}\mathbf{e}_1'\mathbf{\Phi}^\ell\dot{\mathbf{R}}_{-1}\dot{\mathbf{R}}_{-1}'\mathbf{\Phi}'^\ell\mathbf{e}_1\right\| = 0$. Noting that $E\left[\sum_{\ell=0}^{\infty}\left(\mathbf{e}_1'\mathbf{\Phi}^\ell\dot{\mathbf{R}}_{-1}\right)\boldsymbol{\varepsilon}_{t-\ell}\right] = 0$, we have

$$\sum_{\ell=0}^{\infty}\left(\mathbf{e}_1'\mathbf{\Phi}^\ell\dot{\mathbf{R}}_{-1}\right)\boldsymbol{\varepsilon}_{t-\ell} \overset{q.m.}{\to} 0, \text{ as } N\to\infty. \tag{78}$$

This completes the proof of equation (25). To prove (26), we write

$$\overline{x}_{wt} - \sum_{\ell=0}^{\infty}\left(\mathbf{w}'\mathbf{\Phi}^\ell\mathring{\mathbf{r}}_1\right)\varepsilon_{1,t-\ell} = \sum_{\ell=0}^{\infty}\left(\mathbf{w}'\mathbf{\Phi}^\ell\dot{\mathbf{R}}_{-1}\right)\boldsymbol{\varepsilon}_{t-\ell}. \tag{79}$$

Since the vectors $\left\{\mathbf{w}'\mathbf{\Phi}^\ell\right\}$ have the same properties as vectors $\{\mathbf{a}_\ell\}$ in equations (74)-(76), it follows that (using the same arguments as above),

$$\sum_{\ell=0}^{\infty}\left(\mathbf{w}'\mathbf{\Phi}^\ell\dot{\mathbf{R}}_{-1}\right)\boldsymbol{\varepsilon}_{t-\ell} \overset{q.m.}{\to} 0, \text{ as } N\to\infty. \tag{80}$$

This completes the proof of equation (26). ■

Proof of Proposition 3.

$$x_{it} - \phi_{ii}x_{i,t-1} - \boldsymbol{\phi}_{-1,-i}'\mathbf{x}_{t-1} - \phi_{i1}x_{1,t-1} - r_{i1}\varepsilon_{1t} - r_{ii}\varepsilon_{it} = \mathbf{r}_{-1,-i}'\boldsymbol{\varepsilon}_t \tag{81}$$

The vector $\mathbf{r}_{-1,-i}$ satisfies equation (24) of Assumption 6, $Var(\boldsymbol{\varepsilon}_t) = \mathbf{I}_N$, and $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$, which implies

$$\mathbf{r}_{-1,-i}'\boldsymbol{\varepsilon}_t \overset{q.m.}{\to} 0. \tag{82}$$

[22] $\left\|\mathbf{a}_\ell\right\|_c \le \left\|\mathbf{e}_1'\mathbf{\Phi}^\ell\right\|_r \le \left\|\mathbf{e}_1'\right\|_r\left\|\mathbf{\Phi}\right\|_r^\ell \le \rho^\ell$.

Result (25) and invertibility of polynomial $\vartheta(L, \mathbf{e}_1)$ implies

$$\vartheta^{-1}(L, \mathbf{e}_1) x_{1t} - \varepsilon_{1t} \overset{q.m.}{\to} 0. \tag{83}$$

Finally, since $\boldsymbol{\phi}_{-1,-i}$ satisfies condition (11) under Assumption 4, equation (26) implies

$$\boldsymbol{\phi}'_{-1,-i}\mathbf{x}_{t-1} - \vartheta\left(L, \boldsymbol{\phi}_{-1,-i}\right)\varepsilon_{1,t-1} \overset{q.m.}{\to} 0. \tag{84}$$

Substituting results (82)-(84) into (81) establishes equation (29). ■

Lemma 1 *Let Assumptions 8 and 9 hold and suppose $N, T \overset{j}{\to} \infty$ at any rate. Then for any $p, q \in \{0, 1\}$ and for any sequences of non-random vectors $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, such that $\|\boldsymbol{\theta}\| = O(1)$ and $\|\boldsymbol{\varphi}\|_c = O(1)$, we have*

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p} \overset{p}{\to} 0, \tag{85}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q} - E\left(\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\right) \overset{p}{\to} 0, \tag{86}$$

where the process $\boldsymbol{\upsilon}_t$ is defined in equation (40). Furthermore, if $\|\boldsymbol{\theta}\| = O\left(N^{-\frac{1}{2}}\right)$ then

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t} \overset{p}{\to} 0, \tag{87}$$

and

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q} - E\left(\sqrt{N}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\right) \overset{p}{\to} 0. \tag{88}$$

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 $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, 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,[23] and $\{\mathcal{F}_t\}$ denotes an increasing sequence of σ-fields ($\mathcal{F}_{t-1} \subset \mathcal{F}_t$) such that $\mathcal{F}_t$ includes all information available at time t and κ_{Nt} is measurable with respect to $\mathcal{F}_t$ for any $N \in \mathbb{N}$. Let $\left\{\{c_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ be two-dimensional array of constants and set $c_{Nt} = \frac{1}{T_N}$ for all $t \in \mathbb{Z}$ and $N \in \mathbb{N}$. Note that

$$\begin{aligned} E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-n}\right)\right]^2\right\} &= \sum_{\ell=\mathtt{m}\ np}^{\infty}\boldsymbol{\theta}'\boldsymbol{\Phi}^{\ell-p}\boldsymbol{\Sigma}\boldsymbol{\Phi}'^{\ell-p}\boldsymbol{\theta}, \\ &\leq \varsigma_n, \end{aligned} \tag{89}$$

[23]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.

where $\mathtt{m}_{np} = \max\{n, p\}$ and[24]

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

Using Assumptions 8 and 9, ς_n has the following properties

$$\varsigma_0 = O(1)\text{, and } \varsigma_n \to 0 \text{ as } n \to \infty. \tag{90}$$

By Liapunov's inequality, $E\left|E(\kappa_{Nt} \mid \mathcal{F}_{t-n})\right| \leq \sqrt{E\left\{[E(\kappa_{Nt} \mid \mathcal{F}_{t-n})]^2\right\}}$ (Davidson, 1994, Theorem 9.23). It follows that the two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to the constant array $\{c_{Nt}\}$. Equations (89) and (90) establish array $\{\kappa_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm. This implies uniform integrability.[25] Note that

$$\lim_{N\to\infty} \sum_{t=1}^{T_N} c_{Nt} = \lim_{N\to\infty} \sum_{t=1}^{T_N} \frac{1}{T_N} = 1 < \infty, \tag{91}$$

$$\lim_{N\to\infty} \sum_{t=1}^{T_N} c_{Nt}^2 = \lim_{N\to\infty} \sum_{t=1}^{T_N} \frac{1}{T_N^2} = 0. \tag{92}$$

Therefore array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ satisfies conditions of a mixingale weak law,[26] 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 (85). Under the condition $\|\boldsymbol{\theta}\| = O\left(N^{-\frac{1}{2}}\right)$, result (87) follows from result (85) by noting that $\left\|\sqrt{N}\boldsymbol{\theta}\right\| = O(1)$.

Result (86) is established in a similar fashion. Consider the following two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, defined by[27]

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

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

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

[24] 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 8 implies $\sum_{\ell=0}^{\infty} \left\|\boldsymbol{\Phi}^{\ell}\right\|^2 = O(1)$.

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

[26] Davidson (1994, Theorem 19.11).

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

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

$$E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-n}\right)\right]^2\right\} = \sum_{s=\mathrm{m}\ pn}^{\infty}\sum_{\ell=\mathrm{m}\ qn}^{\infty}\sum_{j=\mathrm{m}\ pn}^{\infty}\sum_{d=\mathrm{m}\ qn}^{\infty} E\left(\boldsymbol{\theta}_{s-p}'\mathbf{u}_{t-s}\boldsymbol{\varphi}_{\ell-q}'\mathbf{u}_{t-\ell}\boldsymbol{\theta}_{j-p}'\mathbf{u}_{t-j}\boldsymbol{\varphi}_{d-q}'\mathbf{u}_{t-d}\right) - \left(\sum_{s=\mathrm{m}\ pn}^{\infty}\sum_{\ell=\mathrm{m}\ qn}^{\infty} E\left(\boldsymbol{\theta}_{s-p}'\mathbf{u}_{t-s}\boldsymbol{\varphi}_{\ell-q}'\mathbf{u}_{t-\ell}\right)\right)^2. \tag{93}$$

Using the independence of $\mathbf{u}_t$ and $\mathbf{u}_{t'}$ for any $t \neq t'$ (Assumption 9), we have

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

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 Assumption 9, $\sum_{\ell=0}^{\infty}\|\boldsymbol{\Phi}\|^{2\ell} = O(1)$ by Assumption 8, $\|\boldsymbol{\theta}\| = O(1)$, $\|\boldsymbol{\varphi}\| \leq \|\boldsymbol{\varphi}\|_c = O(1)$, and $\varsigma_{a,n}$ has the following properties

$$\varsigma_{a,0} = O(1)\,, \text{ and } \varsigma_{a,n} \to 0 \text{ as } n \to \infty. \tag{94}$$

Similarly, using the independence of $\mathbf{u}_t$ and $\mathbf{u}_{t'}$ for any $t \neq t'$ (postulated in Assumption 9),[28] the first term on the right side of equation (93) is bounded by the following upper bound $\varsigma_{b,n}$:

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

where $\chi_2(p,n,q) = \max\{0, n-p\} + \max\{n-q, 0\}$, $\mathbf{B}$ is $N \times N$ matrix with the element (i,j) equal to $\|\boldsymbol{\Psi}_{ij}\|$, and $\boldsymbol{\Psi}_{ij}$ is $N\times N$ matrix of fourth moments with the element (n,s) equal to $E(u_{it}u_{jt}u_{nt}u_{st})$. It follows from Assumptions

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

8 and 9 that $\varsigma_{b,n}$ has following properties[29]

$$\varsigma_{b,0} = O(1), \text{ and } \varsigma_{b,n} \to 0 \text{ as } n \to \infty. \tag{95}$$

$E\left\{\left[E\left(\frac{\kappa_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-n}\right)\right]^2\right\}$ is therefore bounded by $\varsigma_n = \varsigma_{a,n} + \varsigma_{b,n}$. Equations (94) and (95) establish

$$\varsigma_0 = O(1), \ \varsigma_n \to 0 \text{ as } n \to \infty. \tag{96}$$

By Liapunov's inequality, $E\left|E\left(\kappa_{Nt} \mid \mathcal{F}_{t-n}\right)\right| \leq \sqrt{E\left\{\left[E\left(\kappa_{Nt} \mid \mathcal{F}_{t-n}\right)\right]^2\right\}}$ (Davidson, 1994, Theorem 9.23). It follows that the two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, is L_1-mixingale with respect to a constant array $\{c_{Nt}\}$. Furthermore, (96) establishes array $\{\kappa_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm. This implies uniform integrability.[30] Since also equations (91) and (92) hold, array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ satisfies conditions of a mixingale weak law,[31] 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 $N, T \xrightarrow{j} \infty$ at any rate. Convergence in L_1 norm implies convergence in probability. This completes the proof of result (86). Under the condition $\|\boldsymbol{\theta}\| = O\left(N^{-\frac{1}{2}}\right)$, result (88) follows from result (86) by noting that $\left\|\sqrt{N}\boldsymbol{\theta}\right\| = O(1)$.

■

Lemma 2 *Let $\mathbf{x}_t$ be generated by model (32), Assumptions 7-12 hold, and $N, T \xrightarrow{j} \infty$ at any rate. Then for any $p, q \in \{0, 1\}$, and for any sequences of non-random vectors $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ with growing dimension $N \times 1$ such that $\|\boldsymbol{\theta}\|_c = O(1)$ and $\|\boldsymbol{\varphi}\|_c = O(1)$, we have*

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}'\mathbf{x}_{t-p} - E\left(\boldsymbol{\theta}'\mathbf{x}_{t-p}\right) \xrightarrow{p} 0, \tag{97}$$

[29] Matrix $\mathbf{B}$ is symmetric by construction. Therefore $\|\mathbf{B}\| \leq \sqrt{\|\mathbf{B}\|_r \|\mathbf{B}\|_c} = \|\mathbf{B}\|_r$, where

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

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

and

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\mathbf{x}_{t-p}\boldsymbol{\varphi}'\mathbf{x}_{t-q}-E\left(\boldsymbol{\theta}'\mathbf{x}_{t-p}\boldsymbol{\varphi}'\mathbf{x}_{t-q}\right)\overset{p}{\to}0. \tag{98}$$

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

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\Gamma}\mathbf{f}_{t-q}\overset{p}{\to}0, \tag{99}$$

where $\boldsymbol{\upsilon}_t$ *is defined in equation (40).*

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 $\left\{\{\kappa_{Nt},\mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, defined by

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

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

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

where

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

$\|\boldsymbol{\theta}\|^2 = O(1)$, $\|\boldsymbol{\Phi}\|<\rho$ by Assumption 8, and $\|\boldsymbol{\Sigma}\|\leq\sqrt{\|\boldsymbol{\Sigma}\|_c\|\boldsymbol{\Sigma}\|_r}=O(1)$ by Assumption 9. Furthermore, since $\mathbf{f}_{t-q}$ is covariance stationary and $\|\boldsymbol{\varphi}'\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{\varphi}\|=O(1)$ (by condition $\|\boldsymbol{\varphi}\|_c=O(1)$ and Assumption 12), we have

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

It follows that ς_n has following properties

$$\varsigma_0=O(1)\text{ and }\varsigma_n\to0\text{ as }n\to\infty.$$

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 (91) and (92) hold, it follows that the array $\{\kappa_{Nt},\mathcal{F}_t\}$ satisfies conditions of a mixingale weak law,[32] which implies $\sum_{t=1}^{T_N}\kappa_{Nt}\overset{L_1}{\to}0$. Convergence in L_1 norm implies convergence in probability. This completes the proof of result (99).

Assumption 11 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

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

[32] Davidson (1994, Theorem 19.11)

Process $\mathbf{f}_t$ is independent of $\boldsymbol{\upsilon}_t$. Suppose $N,T \xrightarrow{j} \infty$ at any rate. 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 $\|\boldsymbol{\theta}\| \leq \|\boldsymbol{\theta}\|_c = O(1)$. 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 12, condition$\|\boldsymbol{\theta}\|_c = O(1)$ and condition $\|\boldsymbol{\varphi}\|_c = O(1)$. 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 (32), Assumptions 7-12 hold and $N,T \xrightarrow{j} \infty$ at any rate. Then for any $p,q \in \{0,1\}$, for any sequence of non-random matrices of weights $\mathbf{W}$ of growing dimension $N \times m_w$ satisfying conditions (33)-(34), and for any $r \in \{1,..,m_w\}$,*

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathbf{W}'\boldsymbol{\upsilon}_{t-p} \xrightarrow{p} \mathbf{0}, \tag{100}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathbf{W}'\boldsymbol{\upsilon}_{t-p}\overline{\mathbf{x}}_{W,t-q} \xrightarrow{p} \mathbf{0}, \tag{101}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathbf{W}'\boldsymbol{\upsilon}_{t-p}\mathbf{x}_{i,t-q} \xrightarrow{p} \mathbf{0}, \tag{102}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it} \xrightarrow{p} \mathbf{0}, \tag{103}$$

where the process $\boldsymbol{\upsilon}_t$ is defined in equation (40), 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 equation (47).

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 (33), result

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\boldsymbol{\upsilon}_{t-p} \xrightarrow{p} 0 \tag{104}$$

follows directly from Lemma 1, equation (87). This completes the proof of result (100).

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

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\mathbf{x}_{t-q} = \frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\left(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{f}_{t-q} + \boldsymbol{\upsilon}_{t-q}\right). \tag{105}$$

Since $\left\|\sqrt{N}\mathring{\mathbf{w}}_r\right\| = O(1)$ for any $r \in \{1,..,m_w\}$ by condition (33), we can use Lemma 1, result (88), which implies

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q} - E\left(\mathring{\mathbf{w}}_r'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\upsilon}_{t-q}\right) \xrightarrow{p} 0. \tag{106}$$

Sequence $\{\boldsymbol{\varphi}'\boldsymbol{\alpha}\}$ is deterministic and bounded in N, and therefore it follows from Lemma 1, result (87), that

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\alpha} \xrightarrow{p} 0. \tag{107}$$

Similarly, Lemma 2 equation (99) implies

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\boldsymbol{\Gamma}\mathbf{f}_{t-q} \xrightarrow{p} 0. \tag{108}$$

Results (106), (107) and (108) establish

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathring{\mathbf{w}}_r'\boldsymbol{\upsilon}_{t-p}\boldsymbol{\varphi}'\mathbf{x}_{t-q} \xrightarrow{p} 0. \tag{109}$$

Result (101) follows from equation (109) by setting $\varphi = \mathring{\mathbf{w}}_l$ for any $l \in \{1,..,m_w\}$. Result (102) follows from equation (109) 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 (103) directly follows from results (100)-(102). This completes the proof. ■

Lemma 4 *Let $\mathbf{x}_t$ be generated by model (32), Assumptions 7-12 hold, and $N, T \xrightarrow{j} \infty$ at any rate. Then for any sequence of non-random matrices of weights $\mathbf{W}$ of growing dimension $N \times m_w$ satisfying conditions (33)-(34),*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}' - \mathbf{C}_i \xrightarrow{p} \mathbf{0}, \tag{110}$$

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

Proof. Result (110) directly follows from Lemmas 1, 2 and 3. ■

Lemma 5 *Let $\mathbf{x}_t$ be generated by model (32), Assumptions 7-12 hold, and $N, T \xrightarrow{j} \infty$ at any rate. Then for any sequence of non-random matrices of weights $\mathbf{W}$ of growing dimension $N \times m_w$ satisfying conditions (33)-(34), and for any fixed $p \geq 0$,*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{W}'\boldsymbol{\upsilon}_{t-p}u_{it} \xrightarrow{p} \mathbf{0}, \tag{111}$$

where the process $\boldsymbol{\upsilon}_t$ is defined in equation (40). If in addition $T/N \to \varkappa$, with $0 \leq \varkappa < \infty$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{W}'\boldsymbol{\upsilon}_{t-p}u_{it} \xrightarrow{p} \mathbf{0}. \tag{112}$$

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

$$\boldsymbol{\kappa}_{Nit} = \frac{1}{\sqrt{T_N}}\left\{\mathbf{W}'\boldsymbol{\upsilon}_{t-p}u_{it} - E\left(\mathbf{W}'\boldsymbol{\upsilon}_{t-p}u_{it}\right)\right\}, \tag{113}$$

where the subscript N is used to emphasize the number of cross section units.[33] Let $\{\mathcal{F}_t\}$ denote an increasing sequence of σ-fields ($\mathcal{F}_{t-1} \subset \mathcal{F}_t$) with $\boldsymbol{\kappa}_{Nit}$ measurable with respect of $\mathcal{F}_t$. First it is established that for any *fixed* $i \in \mathbb{N}$, the vector array $\left\{\{\boldsymbol{\kappa}_{Nit}/c_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is uniformly integrable, where $c_{Nt} = \frac{1}{\sqrt{NT_N}}$. For $p > 0$, we can write

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

where $\|\mathbf{W}\|^2 = O\left(N^{-1}\right)$ by condition (33), $\|\boldsymbol{\Sigma}\| = O(1)$ by Assumption 9, and $\sum_{\ell=0}^{\infty} \left\|\boldsymbol{\Phi}^{\ell}\right\|^2 = O(1)$ by Assumption 8. For $p = 0$, we have

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

where as before $\boldsymbol{\Psi}_{ii}$ is $N \times N$ symmetric matrix with the element (n, s) equal to $E\left(u_{it}u_{it}u_{nt}u_{st}\right)$. 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 $\{\boldsymbol{\kappa}_{Nit}/c_{Nt}\}$.

$$E\left|E\left(\boldsymbol{\kappa}_{Nit} \mid \mathcal{F}_{t-n}\right)\right| = \begin{cases} \mathbf{0} & \text{for any } n > 0 \text{ and any fixed } p \geq 0 \\ \boldsymbol{\tau}_{m_w} c_{Nt} O(1) & \text{for } n = 0 \text{ and any fixed } p \geq 0 \end{cases}, \tag{114}$$

and $\left\{\{\boldsymbol{\kappa}_{Nit}, \mathcal{F}_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$.[34] 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

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

Therefore for each fixed $i \in \mathbb{N}$, each of the m_w two-dimensional arrays given by the elements of vector array

[33] 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.

[34] The last equality in equation (114) takes advatage of Liapunov's inequality. $\boldsymbol{\tau}_{m_w}$ is $m_w \times 1$ dimensional vector of ones.

$\left\{\{\boldsymbol{\kappa}_{Nit}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ satisfies conditions of a mixingale weak law[35], 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

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

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 (112).

Result (111) 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 (113) 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 $\left\{\left\{\sqrt{T_N}\mathbf{q}_{Nit}/c_{Nt}, \mathcal{F}_t\right\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is uniformly integrable because $\left\{\{\boldsymbol{\kappa}_{Nit}/c_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is uniformly integrable. Furthermore, $\left\{\{\mathbf{q}_{Nit}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ is L_1-mixingale with respect to the constant array $\left\{\frac{1}{\sqrt{T_N}}c_{Nt}\right\}$ since $\left\{\{\boldsymbol{\kappa}_{Nit}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=i}^{\infty}$ 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[36] implies

$$\sum_{t=1}^{T_N}\mathbf{q}_{Nit} \xrightarrow{L_1} 0 \text{ as } N \to \infty. \tag{115}$$

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 (111). ■

Lemma 6 *Let $\mathbf{x}_t$ be generated by model (32), Assumptions 7-12 hold and $N, T \xrightarrow{j} \infty$ such that $T/N \to \varkappa$, with $0 \le \varkappa < \infty$. Then for any sequence of non-random matrices of weights $\mathbf{W}$ of growing dimension $N \times m_w$ satisfying conditions (33)-(34), we have,*

[35] See Davidson (1994, Theorem 19.11).
[36] See Davidson (1994, Theorem 19.11).

a) under Assumption 13,

$$\frac{1}{\sigma_{ii}}\mathbf{C}_i^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{\mathbf{g}}_{it}u_{it}\xrightarrow{D}N\left(0,\mathbf{I}_{k_i}\right), \tag{116}$$

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

b) under Assumption 14,

$$\frac{1}{\sigma_{ii}\sqrt{T}}\boldsymbol{\Omega}_{vi}^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{v}_{i,t-1}u_{it}\xrightarrow{D}N\left(0,\mathbf{I}_{h_i}\right), \tag{117}$$

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 $T_N = T(N)$ 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 $\{\kappa_{Nt}, \mathcal{F}_t\}$ is a stationary martingale difference array.[37] 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

$$\sum_{t=1}^{T_N}\kappa_{Nt}^2\xrightarrow{p}1. \tag{118}$$

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

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

Using Liapunov's theorem (Davidson, 1994, Theorem 23.11), Lindeberg condition[38] holds, which in turn implies

$$\max_{1\leq t\leq T_N}\left|\kappa_{Nt}\right|\xrightarrow{p}0 \text{ as } N\to\infty. \tag{119}$$

Results (118), (119) and the martingale difference array central limit theorem (Davidson, 1994, Theorem 24.3) establish

$$\sum_{t=1}^{T_N}\kappa_{Nt} = \frac{1}{\sqrt{T_N}\sigma_{ii}}\mathbf{a}'\mathbf{C}_i^{-\frac{1}{2}}\sum_{t=1}^{T_N}\widetilde{\mathbf{g}}_{it}u_{it}\xrightarrow{D}N\left(0,\mathbf{I}_{k_i}\right) \tag{120}$$

Since equation (120) holds for any $k_i \times 1$ dimensional vector $\mathbf{a}$ such that $\|\mathbf{a}\| = 1$, result (116) directly follows from equation (120) and Davidson (1994, Theorem 25.6).

Result (117) can be established in the same way as the result (116), but this time we set $\kappa_{Nt} = \frac{1}{\sqrt{T_N}\sigma_{ii}}\mathbf{a}'\boldsymbol{\Omega}_{vi}^{-\frac{1}{2}}\mathbf{v}_{i,t-1}u_{it}$,

[37] As before, let $\{\mathcal{F}_t\}$ denote an increasing sequence of σ-fields ($\mathcal{F}_{t-1} \subset \mathcal{F}_t$) with κ_{Nt} measurable with respect of $\mathcal{F}_t$.

[38] See Davidson (1994, Condition 23.17).

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 (32), and suppose Assumptions 7-12 hold and $N, T \xrightarrow{j} \infty$ at any rate. Then for any arbitrary matrix of weights $\mathbf{W}$ satisfying conditions (33)-(34), and for any $p, q \in \{0, 1\}$:*

$$\frac{1}{T}\sum_{t=1}^{T} \overline{\boldsymbol{v}}_{W,t-p} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{121}$$

$$\frac{1}{T}\sum_{t=1}^{T} \overline{\boldsymbol{v}}_{W,t-p}\mathbf{f}'_{t-q} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{122}$$

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{v}_{i,t-p}\overline{\boldsymbol{v}}'_{W,t-q} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{123}$$

$$\frac{1}{T}\sum_{t=1}^{T} \overline{\boldsymbol{v}}_{W,t-p}\overline{\boldsymbol{v}}'_{W,t-q} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{124}$$

$$\frac{\boldsymbol{\Upsilon}'_i\mathbf{Q}}{T} = o_p(1). \tag{125}$$

Furthermore,

$$\frac{\mathbf{H}'\mathbf{Q}}{T} = \mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{126}$$

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

$$\frac{\mathbf{H}'\mathbf{H}}{T} = \mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{128}$$

$$\frac{\mathbf{H}'\mathbf{u}_{i\circ}}{T} = \mathbf{A}'\frac{\mathbf{Q}'\mathbf{u}_{i\circ}}{T} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{129}$$

where

$$\underset{T\times h_i}{\boldsymbol{\Upsilon}_i} = (\mathbf{v}_{i0}, \mathbf{v}_{i1}, ..., \mathbf{v}_{i,T-1})', \tag{130}$$

$\mathbf{v}_{it} = \mathbf{S}'_i \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}^{\ell}\mathbf{u}_{t-\ell}$, matrices $\mathbf{H}$ and $\mathbf{Z}_i$ are defined below equation (53), and matrices $\mathbf{Q}$, $\mathbf{F}$ and $\mathbf{A}$ are defined in equations (54)-(55).

Proof. Result (121) follows directly from equation (87) 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 (122) follows from result (121) by noting that $\mathbf{f}_t$ is independently distributed of $\overline{\boldsymbol{v}}_{W,t}$ 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 (125) follows. Results (123) and (124) follows directly from equation (88) of Lemma 1 by noting that

$$\sqrt{N}E\left(\boldsymbol{v}_{i,t-p}\overline{\boldsymbol{v}}'_{W,t-q}\right) = O\left(\frac{1}{\sqrt{N}}\right) \tag{131}$$

as well as[39]

$$\sqrt{N}E\left(\overline{\boldsymbol{v}}_{W,t-p}\overline{\boldsymbol{v}}'_{W,t-q}\right) = O\left(\frac{1}{\sqrt{N}}\right). \tag{132}$$

In order to prove equations (126)-(129), first note that the row t of the matrix $\mathbf{H} - \mathbf{QA}$ is $\left(0, \overline{\boldsymbol{v}}'_{Wt}, \overline{\boldsymbol{v}}'_{W,t-1}\right)$.

[39] Results (131) and (132) are straightforward to establish by taking the row norm and by noting that the granularity conditions (33)-(34) imply $\|\mathbf{W}\|_r = O\left(N^{-1}\right)$.

Using results (121)-(124), we have

$$\frac{(\mathbf{H}-\mathbf{QA})'\mathbf{Q}}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix} 0 \\ \overline{\boldsymbol{v}}_{Wt} \\ \overline{\boldsymbol{v}}_{W,t-1} \end{pmatrix}\begin{pmatrix} 1, & \mathbf{f}_t', & \mathbf{f}_{t-1}' \end{pmatrix}\right] = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{133}$$

$$\frac{\mathbf{Z}_i'(\mathbf{H}-\mathbf{QA})}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\boldsymbol{\xi}_{i,t-1}\begin{pmatrix} 0 \\ \overline{\boldsymbol{v}}_{Wt} \\ \overline{\boldsymbol{v}}_{W,t-1} \end{pmatrix}'\right] = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{134}$$

$$\frac{\mathbf{H}'(\mathbf{H}-\mathbf{QA})}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix} \overline{\mathbf{x}}_{Wt} \\ \overline{\mathbf{x}}_{W,t-1} \end{pmatrix}\begin{pmatrix} 0 \\ \overline{\boldsymbol{v}}_{Wt} \\ \overline{\boldsymbol{v}}_{W,t-1} \end{pmatrix}'\right] = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{135}$$

$$\frac{(\mathbf{H}-\mathbf{QA})'(\mathbf{H}-\mathbf{QA})}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix} 0 \\ \overline{\boldsymbol{v}}_{Wt} \\ \overline{\boldsymbol{v}}_{W,t-1} \end{pmatrix}\begin{pmatrix} 0 \\ \overline{\boldsymbol{v}}_{Wt} \\ \overline{\boldsymbol{v}}_{W,t-1} \end{pmatrix}'\right] = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{136}$$

Equations (133)-(134) establish results (126) and (127). Note that

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

where the last equality uses equations (133) and (135). This completes the proof of result (128).

Equation (115) (see proof or Lemma 5) implies

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

as $N,T \xrightarrow{j} \infty$ at any rate. Result (129) follows by noting that $\sqrt{N}E\left(\overline{\boldsymbol{v}}_{W,t-p}u_{it}\right) = O\left(N^{-\frac{1}{2}}\right)$. This completes the proof. ■

Lemma 8 *Let $\mathbf{x}_t$ be generated by model (32), suppose Assumptions 7-12, 14 hold, and $N,T \xrightarrow{j} \infty$ at any rate. Then for any arbitrary matrix of weights $\mathbf{W}$ satisfying conditions (33)-(34) and Assumption 14, we have*

$$\frac{\mathbf{Q}'\mathbf{Q}}{T} \xrightarrow{p} \boldsymbol{\Omega}_Q, \tag{137}$$

$\boldsymbol{\Omega}_{QQ}$ is nonsingular, and

$$\frac{\boldsymbol{\Upsilon}_i'\boldsymbol{\Upsilon}_i}{T} - \boldsymbol{\Omega}_{vi} \xrightarrow{p} 0, \tag{138}$$

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 (54), and matrix* $\boldsymbol{\Upsilon}_i = (\mathbf{v}_{i0}, \mathbf{v}_{i1}, ..., \mathbf{v}_{i,T-1})'$.

Proof. Assumption 11 implies matrix $\boldsymbol{\Omega}_Q$ is nonsingular. Result (137) 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 $\frac{1}{T}\sum_{t=1}^{T}\mathbf{S}'_i\boldsymbol{\upsilon}_t\boldsymbol{\upsilon}'_t\mathbf{S}_i - E(\mathbf{S}'_i\boldsymbol{\upsilon}_t\boldsymbol{\upsilon}'_t\mathbf{S}_i) \xrightarrow{p} 0$.[40] This completes the proof. ■

Lemma 9 *Let* $\mathbf{x}_t$ *be generated by model (32), suppose Assumptions 7-12 and 14 hold, and* $N, T \xrightarrow{j} \infty$ *at any rate. Then for any arbitrary matrix of weights* $\mathbf{W}$ *satisfying conditions (33)-(34) and Assumption 14, we have*

$$\frac{\mathbf{Z}'_i\mathbf{M}_H\mathbf{Z}_i}{T} = \frac{\mathbf{Z}'_i\mathbf{M}_Q\mathbf{Z}_i}{T} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{139}$$

$$\frac{\mathbf{Z}'_i\mathbf{M}_Q\mathbf{Z}_i}{T} - \boldsymbol{\Omega}_{vi} \xrightarrow{p} 0, \tag{140}$$

$$\frac{\mathbf{Z}'_i\mathbf{M}_H\mathbf{Q}}{\sqrt{T}} = o_p\left(\sqrt{\frac{T}{N}}\right), \tag{141}$$

$$\frac{\mathbf{Z}'_i\mathbf{M}_H\mathbf{u}_{i\circ}}{\sqrt{T}} = \frac{\boldsymbol{\Upsilon}'_i\mathbf{M}_Q\mathbf{u}_{i\circ}}{\sqrt{T}} + o_p\left(\sqrt{\frac{T}{N}}\right), \tag{142}$$

where $\boldsymbol{\Omega}_{vi}$ *is defined in Assumption 14,* $\mathbf{M}_H = \mathbf{I}_T - \mathbf{H}(\mathbf{H}'\mathbf{H})^+\mathbf{H}'$, *matrices* $\mathbf{H}$ *and* $\mathbf{Z}_i$ *are defined below equation (53), matrices* $\mathbf{Q}$ *and* $\mathbf{F}$ *are defined in equation (54), and matrix* $\boldsymbol{\Upsilon}_i = (\mathbf{v}_{i0}, \mathbf{v}_{i1}, ..., \mathbf{v}_{i,T-1})'$.

Proof.

$$\frac{\mathbf{Z}'_i\mathbf{M}_H\mathbf{Z}_i}{T} = \frac{\mathbf{Z}'_i\mathbf{Z}_i}{T} - \frac{\mathbf{Z}'_i\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^+\frac{\mathbf{H}'\mathbf{Z}_i}{T}. \tag{143}$$

Results (127)-(128) of Lemma 7 imply

$$\frac{\mathbf{Z}'_i\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^+\frac{\mathbf{H}'\mathbf{Z}_i}{T} = \frac{\mathbf{Z}'_i\mathbf{Q}}{T}\mathbf{A}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^+\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Z}_i}{T} + o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{144}$$

Using definition of the Moore-Penrose inverse, it follows

$$\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^+\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right) = \left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right). \tag{145}$$

Multiply equation (145) 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[41]

$$\mathbf{A}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^+\mathbf{A}' = \left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}. \tag{146}$$

[40] $\|\mathbf{S}_i\|_c = O(1)$ by Assumption 7.

[41] Note that $plim_{T\to\infty}\frac{1}{T}\mathbf{Q}'\mathbf{Q}$ is nonsingular by Lemma 8, equation (137). $\mathbf{A}\mathbf{A}'$ is nonsingular, since matrix $\mathbf{A}$ has full row-rank by Assumption 14.

Equations (146) and (144) imply

$$\frac{\mathbf{Z}_i'\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{+}\frac{\mathbf{H}'\mathbf{Z}_i}{T}=\frac{\mathbf{Z}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{Z}_i}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{147}$$

Result (139) follows from equations (147) and (143).

System (32) implies

$$\mathbf{Z}_i=\boldsymbol{\tau}\boldsymbol{\alpha}_i'\mathbf{S}_i+\mathbf{F}\left(-1\right)\boldsymbol{\Gamma}_i'\mathbf{S}_i+\boldsymbol{\Upsilon}_i. \tag{148}$$

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

$$\frac{\mathbf{Z}_i'\mathbf{M}_Q\mathbf{Z}_i}{T}=\frac{\boldsymbol{\Upsilon}_i'\mathbf{M}_Q\boldsymbol{\Upsilon}_i}{T}=\frac{\boldsymbol{\Upsilon}_i'\boldsymbol{\Upsilon}_i}{T}+\frac{\boldsymbol{\Upsilon}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\boldsymbol{\Upsilon}_i}{T}. \tag{149}$$

Using equations (125), (137) and (138), result (140) follows directly from (149).

Results (126)-(128) of Lemma 7 imply

$$\frac{\mathbf{Z}_i'\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{+}\frac{\mathbf{H}'\mathbf{Q}}{T}=\frac{\mathbf{Z}_i'\mathbf{Q}}{T}\mathbf{A}\left(\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}\mathbf{A}\right)^{+}\mathbf{A}'\frac{\mathbf{Q}'\mathbf{Q}}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{150}$$

Substituting equation (146), it follows

$$\frac{\mathbf{Z}_i'\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{+}\frac{\mathbf{H}'\mathbf{Q}}{T}=\frac{\mathbf{Z}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{Q}}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{151}$$

Equation (151) 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 (141).

Results (127)-(129) 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 (146), it follows

$$\frac{\mathbf{Z}_i'\mathbf{H}}{T}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{+}\frac{\mathbf{H}'\mathbf{Q}}{T}=\frac{\mathbf{Z}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{u}_{i\circ}}{T}+o_p\left(\tfrac{1}{\sqrt{N}}\right). \tag{152}$$

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 $\mathbf{Q}=\left[\boldsymbol{\tau},\mathbf{F},\mathbf{F}\left(-1\right)\right]$, equations (152) and (148) 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 (32), and suppose Assumptions 7-12 and 14 hold, and $N,T\xrightarrow{j}\infty$ at any* *rate. Then for any arbitrary matrix of weights* $\mathbf{W}$ *satisfying conditions (33)-(34) and Assumption 14, we have*

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\boldsymbol{\zeta}_i(-1)}{T} = o_p\left(\frac{1}{\sqrt{N}}\right), \tag{153}$$

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\mathbf{u}_{i\circ}}{\sqrt{T}} = \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{i\circ}}{\sqrt{T}} + o_p\left(\sqrt{\frac{T}{N}}\right) + o_p(1), \tag{154}$$

where matrices $\mathbf{M}_H$ *,and* $\mathbf{Z}_i$ *are defined below equation (53),* $\boldsymbol{\Upsilon}_i = (\mathbf{v}_{i0}, \mathbf{v}_{i1}, ..., \mathbf{v}_{i,T-1})$ *and vector* $\boldsymbol{\zeta}_i(-1) = (\zeta_{i,0}, ..., \zeta_{i,T-1})'$.

Proof.

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

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

$\left\|\boldsymbol{\phi}_{ib}\right\|_r = O\left(N^{-1}\right)$ by Assumption 7, therefore result (153) directly follows from equations (134) and (135).

$$\begin{aligned}\frac{\boldsymbol{\Upsilon}_i'\mathbf{M}_Q\mathbf{u}_{i\circ}}{\sqrt{T}} &= \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{i\circ}}{\sqrt{T}} + \frac{\boldsymbol{\Upsilon}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\mathbf{u}_{i\circ}}{\sqrt{T}},\\ &= \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{i\circ}}{\sqrt{T}} + o_p(1),\end{aligned} \tag{155}$$

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 nonsingular by Lemma 8, and $\frac{\boldsymbol{\Upsilon}_i'\mathbf{Q}}{T} = o_p(1)$ by Lemma 7, equation (125). Substituting (155) into equation (142) implies result (154). This completes the proof. ■

Proof of Theorem 1.

a) Substituting for x_{it} in equation (49) yields

$$\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it} + \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}u_{it}\right). \tag{156}$$

With $N, T \xrightarrow{j} \infty$ in any order, Lemma 5 yields[42]

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}u_{it} \xrightarrow{p} \mathbf{0}. \tag{157}$$

Also using Lemmas 3 and 4 we have

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it} \xrightarrow{p} \mathbf{0}, \tag{158}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}' - \mathbf{C}_{iN} \xrightarrow{p} \mathbf{0}, \tag{159}$$

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

respectively. Assumption 13 postulates that the matrix $\mathbf{C}_{iN}$ is invertible for any $N \geq N_0$ and $\left\|\mathbf{C}_{iN}^{-1}\right\|$ is bounded in N. It follows from equation (159) that

$$\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1} - \mathbf{C}_{iN}^{-1} \xrightarrow{p} \mathbf{0}. \tag{160}$$

Result $\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i \xrightarrow{p} 0$ directly follows from equations (157), (158) and (160).

b) Multiplying equation (156) by $\sqrt{T}$ yields

$$\sqrt{T}\left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right) = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}u_{it}\right). \tag{161}$$

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

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it} \xrightarrow{p} \mathbf{0}. \tag{162}$$

Since $\left\|\mathbf{C}_{iN}^{-1}\right\| = O(1)$, equations (160) and (162) now yield

$$\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}q_{it} \xrightarrow{p} \mathbf{0}. \tag{163}$$

Lemma 5 establishes

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overline{\boldsymbol{v}}_{\mathbf{W},t-p}u_{it} \xrightarrow{p} \mathbf{0} \text{ for } p \in \{0,1\}. \tag{164}$$

It follows from equation (164) that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{g}_{it} - \widetilde{\mathbf{g}}_{it}\right)u_{it} \xrightarrow{p} 0, \tag{165}$$

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

$$\frac{1}{\sigma_{ii,N}}\mathbf{C}_{iN}^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{\mathbf{g}}_{it}u_{it} \xrightarrow{D} N\left(\mathbf{0}, \mathbf{I}_{k_i}\right), \tag{166}$$

Equations (160), (163), (165) and (166) imply result (50).

c) Lemma 4 establishes $\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}' - \mathbf{C}_{iN} \xrightarrow{p} \mathbf{0}$. The estimated residuals from auxiliary regression (48) are equal to $\widehat{u}_{it} = u_{it} - \mathbf{g}_{it}'\left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right)$, which implies

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{it}^2 = \frac{1}{T}\sum_{t=1}^{T}u_{it}^2 - 2\left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right)'\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}u_{it} + \left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right)'\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}'\right)\left(\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right), \tag{167}$$

where $\frac{1}{T}\sum_{t=1}^{T}u_{it}^2 - \sigma_{ii,N}^2 \xrightarrow{p} 0$, $\widehat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i \xrightarrow{p} 0$ is established in part (a) of this proof, $\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}\mathbf{g}_{it}' - \mathbf{C}_{iN} \xrightarrow{p} \mathbf{0}$ 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 (157). This completes the proof.

■

Proof of Theorem 2. Vector $\mathbf{x}_{i\circ}$ can be written, using system (32), as

$$\mathbf{x}_{i\circ} = \boldsymbol{\tau}\left(\alpha_i - \boldsymbol{\delta}_i'\mathbf{S}_i'\boldsymbol{\alpha}\right) + \mathbf{Z}_i\boldsymbol{\delta}_i + \mathbf{F}\gamma_i - \mathbf{F}\left(-1\right)\boldsymbol{\Gamma}'\mathbf{S}_i\boldsymbol{\delta}_i + \boldsymbol{\zeta}_i\left(-1\right) + \mathbf{u}_{i\circ}, \tag{168}$$

where $\boldsymbol{\zeta}_i\left(-1\right) = \left(\zeta_{i0}, ..., \zeta_{i,T-1}\right)'$. Substituting equation (168) into the partition least squares formula (53) and noting that by Lemma 9,

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\mathbf{Q}}{\sqrt{T}} = o_p\left(\sqrt{\frac{T}{N}}\right), \tag{169}$$

it follows

$$\sqrt{T}\left(\widehat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i\right) = \left(\frac{\mathbf{Z}_i'\mathbf{M}_H\mathbf{Z}_i}{T}\right)^{-1}\left[\frac{\mathbf{Z}_i'\mathbf{M}_H\left(\mathbf{u}_{i\circ} + \boldsymbol{\zeta}_i\left(-1\right)\right)}{\sqrt{T}} + o_p\left(\sqrt{\frac{T}{N}}\right)\right]. \tag{170}$$

Lemma 9 also establishes that

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\mathbf{Z}_i}{T} - \boldsymbol{\Omega}_{vi} \xrightarrow{p} \mathbf{0}, \text{ as } N, T \xrightarrow{j} \infty \text{ at any rate,} \tag{171}$$

where $\boldsymbol{\Omega}_{vi} = E\left(\mathbf{v}_{it}\mathbf{v}_{it}'\right)$ is nonsingular by Assumption 14.

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

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\boldsymbol{\zeta}_i\left(-1\right)}{\sqrt{T}} \xrightarrow{p} 0, \tag{172}$$

and

$$\frac{\mathbf{Z}_i'\mathbf{M}_H\mathbf{u}_{i\circ}}{\sqrt{T}} = \frac{\boldsymbol{\Upsilon}_i'\mathbf{u}_{i\circ}}{\sqrt{T}} + o_p\left(\sqrt{\frac{T}{N}}\right) + o_p\left(1\right), \tag{173}$$

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

$$\frac{1}{\sigma_{ii}\sqrt{T}}\boldsymbol{\Omega}_{vi}^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{v}_{i,t-1}u_{it} \xrightarrow{D} N\left(\mathbf{0}, \mathbf{I}_{h_i}\right). \tag{174}$$

The desired result (56) now follows from (170)-(174). ■

European Central Bank Working Paper Series

For a complete list of Working Papers published by the ECB, please visit the ECB's website (http://www.ecb.europa.eu).

944 The New Area-Wide Model of the euro area: a micro-founded open-economy model for forecasting and policy analysis" by K. Christoffel, G. Coenen and A. Warne, October 2008.

945 "Wage and price dynamics in Portugal" by C. Robalo Marques, October 2008.

946 "Macroeconomic adjustment to monetary union" by G. Fagan and V. Gaspar, October 2008.

947 "Foreign-currency bonds: currency choice and the role of uncovered and covered interest parity" by M. M. Habib and M. Joy, October 2008.

948 "Clustering techniques applied to outlier detection of financial market series using a moving window filtering algorithm" by J. M. Puigvert Gutiérrez and J. Fortiana Gregori, October 2008.

949 "Short-term forecasts of euro area GDP growth" by E. Angelini, G. Camba-Méndez, D. Giannone, L. Reichlin and G. Rünstler, October 2008.

950 "Is forecasting with large models informative? Assessing the role of judgement in macroeconomic forecasts" by R. Mestre and P. McAdam, October 2008.

951 "Exchange rate pass-through in the global economy: the role of emerging market economies" by M. Bussière and T. Peltonen, October 2008.

952 "How successful is the G7 in managing exchange rates?" by M. Fratzscher, October 2008.

953 "Estimating and forecasting the euro area monthly national accounts from a dynamic factor model" by E. Angelini, M. Bańbura and G. Rünstler, October 2008.

954 "Fiscal policy responsiveness, persistence and discretion" by A. Afonso, L. Agnello and D. Furceri, October 2008.

955 "Monetary policy and stock market boom-bust cycles" by L. Christiano, C. Ilut, R. Motto and M. Rostagno, October 2008.

956 "The political economy under monetary union: has the euro made a difference?" by M. Fratzscher and L. Stracca, November 2008.

957 "Modeling autoregressive conditional skewness and kurtosis with multi-quantile CAViaR" by H. White, T.-H. Kim, and S. Manganelli, November 2008.

958 "Oil exporters: in search of an external anchor" by M. M. Habib and J. Stráský, November 2008.

959 "What drives U.S. current account fluctuations?" by A. Barnett and R. Straub, November 2008.

960 "On implications of micro price data for macro models" by B. Maćkowiak and F. Smets, November 2008.

961 "Budgetary and external imbalances relationship: a panel data diagnostic" by A. Afonso and C. Rault, November 2008.

962 "Optimal monetary policy and the transmission of oil-supply shocks to the euro area under rational expectations" by S. Adjemian and M. Darracq Pariès, November 2008.

963 "Public and private sector wages: co-movement and causality" by A. Lamo, J. J. Pérez and L. Schuknecht, November 2008.

964 "Do firms provide wage insurance against shocks? Evidence from Hungary" by G. Kátay, November 2008.

965 "IMF lending and geopolitics" by J. Reynaud and J. Vauday, November 2008.

966 "Large Bayesian VARs" by M. Bańbura, D. Giannone and L. Reichlin, November 2008.

967 "Central bank misperceptions and the role of money in interest rate rules" by V. Wieland and G. W. Beck, November 2008.

968 "A value at risk analysis of credit default swaps" by B. Raunig and M. Scheicher, November 2008.

969 "Comparing and evaluating Bayesian predictive distributions of asset returns" by J. Geweke and G. Amisano, November 2008.

970 "Responses to monetary policy shocks in the east and west of Europe" by M. Jarociński, November 2008.

971 "Interactions between private and public sector wages" by A. Afonso and P. Gomes, November 2008.

972 "Monetary policy and housing prices in an estimated DSGE for the US and the euro area" by M. Darracq Pariès and A. Notarpietro, November 2008.

973 "Do China and oil exporters influence major currency configurations?" by M. Fratzscher and A. Mehl, December 2008.

974 "Institutional features of wage bargaining in 23 European countries, the US and Japan" by P. Du Caju, E. Gautier, D. Momferatou and M. Ward-Warmedinger, December 2008.

975 "Early estimates of euro area real GDP growth: a bottom up approach from the production side" by E. Hahn and F. Skudelny, December 2008.

976 "The term structure of interest rates across frequencies" by K. Assenmacher-Wesche and S. Gerlach, December 2008.

977 "Predictions of short-term rates and the expectations hypothesis of the term structure of interest rates" by M. Guidolin and D. L. Thornton, December 2008.

978 "Measuring monetary policy expectations from financial market instruments" by M. Joyce, J. Relleen and S. Sorensen, December 2008.

979 "Futures contract rates as monetary policy forecasts" by G. Ferrero and A. Nobili, December 2008.

980 "Extracting market expectations from yield curves augmented by money market interest rates: the case of Japan" by T. Nagano and N. Baba, December 2008.

981 "Why the effective price for money exceeds the policy rate in the ECB tenders?" by T. Välimäki, December 2008.

982 "Modelling short-term interest rate spreads in the euro money market" by N. Cassola and C. Morana, December 2008.

983 "What explains the spread between the euro overnight rate and the ECB's policy rate?" by T. Linzert and S. Schmidt, December 2008.

984 “The daily and policy-relevant liquidity effects” by D. L. Thornton, December 2008.

985 “Portuguese banks in the euro area market for daily funds” by L. Farinha and V. Gaspar, December 2008.

986 “The topology of the federal funds market” by M. L. Bech and E. Atalay, December 2008.

987 “Probability of informed trading on the euro overnight market rate: an update” by J. Idier and S. Nardelli, December 2008.

988 “The interday and intraday patterns of the overnight market: evidence from an electronic platform” by R. Beaupain and A. Durré, December 2008.

989 “Modelling loans to non-financial corporations in the euro area” by C. Kok Sørensen, D. Marqués Ibáñez and C. Rossi, January 2009

990 “Fiscal policy, housing and stock prices” by A. Afonso and R. M. Sousa, January 2009.

991 “The macroeconomic effects of fiscal policy” by A. Afonso and R. M. Sousa, January 2009.

992 “FDI and productivity convergence in central and eastern Europe: an industry-level investigation” by M. Bijsterbosch and M. Kolasa, January 2009.

993 “Has emerging Asia decoupled? An analysis of production and trade linkages using the Asian international input-output table” by G. Pula and T. A. Peltonen, January 2009.

994 “Fiscal sustainability and policy implications for the euro area” by F. Balassone, J. Cunha, G. Langenus, B. Manzke, J. Pavot, D. Prammer and P. Tommasino, January 2009.

995 “Current account benchmarks for central and eastern Europe: a desperate search?” by M. Ca’ Zorzi, A. Chudik and A. Dieppe, January 2009.

996 “What drives euro area break-even inflation rates?” by M. Ciccarelli and J. A. García, January 2009.

997 “Financing obstacles and growth: an analysis for euro area non-financial corporations” by C. Coluzzi, A. Ferrando and C. Martinez-Carrascal, January 2009.

998 “Infinite-dimensional VARs and factor models” by A. Chudik and M. H. Pesaran, January 2009.