GMM Estimation of short dynamic panel data models with error cross-sectional dependence

Sarafidis, Vasilis

University of Sydney, Monash University

January 2009

GMM Estimation of Short Dynamic Panel Data Models With Error Cross-Sectional Dependence*

Vasilis Sarafidis†

This version: December 2011

Abstract

This paper considers estimation of short dynamic panel data models with error cross-sectional dependence. It is shown that under spatially correlated errors, an additional, generally non-redundant, set of moment conditions becomes available for each i – specifically, instruments with respect to the individual(s) which unit i is spatially correlated with. We demonstrate that these moment conditions remain valid when the error term contains a common factor component, in which situation the standard moment conditions with respect to individual i itself are invalidated, and thereby the standard dynamic panel GMM estimators are inconsistent. The resulting estimators are computationally attractive and do not require estimating the number of unobserved factors. Simulated experiments show that the resulting method of moments estimators perform well in terms of both median bias and root median square error.

Key Words: Dynamic Panel Data, Spatial Dependence, Factor Structure Dependence, Generalised Method of Moments.
JEL Classification: C13; C15; C33.

1 Introduction

In developing the theory of GMM estimation of short dynamic panel data models, it is commonly assumed that the residuals are independently distributed across individuals (see e.g. Anderson and Hsiao, 1981, pg. 598, Arellano and Bond, 1991, pg. 278, Arellano, 1993, pg. 88, Ahn and Schmidt, 1995, pg. 7, Blundell and Bond, 1998, page 118, and others). This assumption is usually made for identification purposes rather than descriptive accuracy with the hope, presumably, that by conditioning on a sufficient number of explanatory variables, what is left over can be treated as a purely idiosyncratic disturbance that is uncorrelated in the cross-sectional dimension. On the other hand, this rather strong assumption is somewhat relaxed in empirical applications involving dynamic panels by allowing for common variations in the dependent variable at any given point in time using a two-way error components disturbance (e.g. Arellano and Bond, 1991, pg. 288, Blundell and Bond, 1998, pg. 137, Bover and Watson, 2005, pg. 1975). In practice, however, this formulation is unlikely to be adequate to remove all correlated behaviour in the residuals and this may invalidate the point estimates of the parameters, as well as inferences; see e.g. Sarafidis and Robertson (2009).

Error cross-sectional dependence may arise for various reasons in practice; for example, it may be due to the presence of spatial correlations specified on the basis of economic and social distance (Conley, 1999) or relative location (Anselin, 1988), as well as due to the presence of unobserved components that give rise to a common factor specification in the disturbances with a fixed number

*I would like to thank Takashi Yamagata and Neville Weber for helpful comments and suggestions.

†Discipline of Operations Management and Econometrics, University of Sydney, and Department of Econometrics and Business Statistics, Monash University, Australia. e-mail: vasilis.sarafidis@sydney.edu.au.

of factors (e.g. Goldberger, 1972, and Jöreskog and Goldberger, 1975). Methods that account for spatial dependence in panel data models have been proposed by Mutl (2006), Kapoor, Kelejian and Prucha (2007), Lee and Yu (2010) among others. Methods that deal with a multi-factor error structure have been proposed by Robertson and Symons (2007), Phillips and Sul (2003), Moon and Perron (2004), Bai (2006), Pesaran (2006), Sarafidis and Yamagata (2010) among others. These methods are theoretically justified in panels where the number of time series observations (T) is large and/or (some of) the covariates are strictly exogenous with respect to the purely idiosyncratic disturbance. Valid methods for fixed T and weakly exogenous, or endogenous regressors have been proposed by Ahn, Lee and Schmidt (2006), Bai (2010), Robertson, Sarafidis and Symons (2010). These methods are non-linear and require estimating the number of unobserved factors as well as the factors themselves. An overview of recent developments in the literature is provided by Sarafidis and Wansbeek (2012).

The present paper investigates the effect of spatial dependence in dynamic panel data models. It is shown that an additional set of moment conditions becomes available $-$ in particular, instruments with respect to the individual(s) which unit i is spatially correlated with. In many practical circumstances these moment conditions are not redundant in the sense that the asymptotic variance of the GMM estimator from the enlarged set of moment conditions is smaller than the GMM estimator that uses the smaller set of moment conditions, i.e. those instruments with respect to individual i only. We develop two GMM estimators. One is based on first-differenced equations and is similar to the Arellano and Bond (1991) GMM estimator. The other one combines equations in first-differences and in levels, yielding a system GMM estimator. Unlike the standard system GMM, however, this estimator remains consistent even if the process is not mean-stationary. This is important because mean-stationarity cannot be theoretically founded in a large number of applications.

Most notably, it is demonstrated that the spatial moment conditions remain valid even when the error term contains a common factor component, in which case the standard moment conditions with respect to lagged values of the endogenous regressor are invalidated. The resulting estimators are computationally attractive since the moment conditions are linear in the parameters, and they do not require estimating the number of unobserved factors or the factors themselves (assuming that theory suggests a particular number of factors to exist) for consistent estimation of the structural parameters. In addition, the set of regressors can be strictly exogenous, or endogenous, while T can be either fixed or large, provided that the number of moment conditions utilised does not grow with T. The main requirement is the specification of a spatial weighting matrix, which is common practice in the spatial literature.

The structure of the paper is as follows. The following section specifies the panel regression model, discusses the basic assumptions employed and derives the consistency and asymptotic normality of the standard first-differenced and system GMM estimators under spatial dependence. Section 3 analyses the properties of the spatial instruments that become available. Section 4 demonstrates that these instruments remain valid even if the error contains a common factor component. The performance of the resulting estimators is investigated in Section 5 using simulated data. A final section concludes.

2 Model Specification and Standard Moment Conditions

This section investigates the effect of spatial dependence on dynamic panel data estimation. Without loss of generality and for easy of exposition we will consider the following panel AR(1) model:

$$
\begin{aligned}
y_{it} &= \alpha y_{it-1} + u_{it},\ |\alpha| < 1,\ i = 1, ..., N,\ t = 1, ..., T, \\
u_{it} &= \eta_i + \varepsilon_{it},\ \varepsilon_{it} = \theta \sum_{j=1}^{N} w_{ij,N} v_{jt} + v_{it},
\end{aligned}
$$

where the initial observation is given by

$$y_{i0} = \delta_0 \eta_i + \delta_1 \varepsilon_{i0},\ \varepsilon_{i0} = \theta \sum_{j=1}^{N} w_{ij,N} \upsilon_{j0} + \upsilon_{i0}. \tag{1}$$

For $\delta_0 = 1/(1-\alpha)$ the process is mean-stationary, and if, in addition, $\delta_1 = \pm\sqrt{1/(1-\alpha^2)}$ the process is covariance-stationary. We do not necessarily want to impose these restrictions at this stage.

Stacking the model over i yields

$$\mathbf{y}_t = \alpha \mathbf{y}_{t-1} + \mathbf{u}_t = \alpha \mathbf{y}_{t-1} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_t = \alpha \mathbf{y}_{t-1} + \boldsymbol{\eta} + P_N \boldsymbol{\upsilon}_t, \tag{2}$$

where $\mathbf{y}_t = (y_{1t}, ..., y_{Nt})'$, $\mathbf{y}_{t-1} = (y_{1t-1}, ..., y_{Nt-1})'$, $\mathbf{u}_t = (u_{1t}, ..., u_{Nt})'$, $\boldsymbol{\eta} = (\eta_1, ..., \eta_N)'$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$, $\boldsymbol{\upsilon}_t = (\upsilon_{1t}, ..., \upsilon_{Nt})'$, W_N is an $N \times N$ matrix and $P_N = I_N + \theta W_N$. $\mathbf{y}_t$ can be written as

$$\mathbf{y}_t = \alpha^t \mathbf{y}_0 + \left(\frac{1-\alpha^t}{1-\alpha} \right) \boldsymbol{\eta} + P_N \sum_{\tau=0}^{t-1} \alpha^\tau \boldsymbol{\upsilon}_{t-\tau}, \tag{3}$$

and from (1) we have $\mathbf{y}_0 = \delta_0 \boldsymbol{\eta} + \delta_1 P_N \boldsymbol{\upsilon}_0$. Therefore, $\mathbf{y}_t$ can be expressed as a linear form of the innovations, $\boldsymbol{\eta}$ and $\boldsymbol{\upsilon}$,

$$\underset{N\times 1}{\mathbf{y}_t} = \underset{N\times N(T+1)}{(\boldsymbol{\beta}_t \otimes P_N)} \underset{N(T+1)\times 1}{\boldsymbol{\upsilon}} + \delta_{0,t}^{*} \underset{N\times 1}{\boldsymbol{\eta}}, \tag{4}$$

where, following a similar approach to Mutl (2006), $\boldsymbol{\beta}_t = \left(\delta_1 \alpha^t, \alpha^{t-1}, ..., \alpha^0, \mathbf{0}_{1\times T-t}\right)$ is a $1 \times (T+1)$ row vector, $\boldsymbol{\upsilon} = (\boldsymbol{\upsilon}_0', ..., \boldsymbol{\upsilon}_T')'$ is a $N(T+1) \times 1$ column vector that contains all the elements of the purely idiosyncratic error component, while $\delta_{0,t}^{*} = \left[\frac{1}{1-\alpha} + \alpha^t \left(\delta_0 - \frac{1}{1-\alpha}\right)\right]$. Observe that $\Delta \boldsymbol{\varepsilon}_t$ can also be expressed as a linear form of $\boldsymbol{\upsilon}$ as follows:

$$\underset{N\times 1}{\Delta \boldsymbol{\varepsilon}_t} = \underset{N\times N(T+1)}{(\boldsymbol{d}_t \otimes P_N)} \underset{N(T+1)\times 1}{\boldsymbol{\upsilon}}, \tag{5}$$

where $\boldsymbol{d}_t$ is a $1 \times (T+1)$ row vector and consists of the $(t-1)^{th}$ row of $(\mathbf{0}_{T-1\times 1}, D)$, while D is the $(T-1) \times T$ matrix first-difference operator (see e.g. Arellano, 2003, pg. 15) defined as

$$D \equiv \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & & 0 & 0 \\ \vdots & & & \ddots & & \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}.$$

Similarly, $\mathbf{u}_t$ can be expressed as

$$\underset{N\times 1}{\mathbf{u}_t} = \underset{N\times N(T+1)}{\left(\boldsymbol{e}_{t+1}' \otimes P_N\right)} \underset{N(T+1)\times 1}{\boldsymbol{\upsilon}} + \underset{N\times 1}{\boldsymbol{\eta}}, \tag{6}$$

where $\boldsymbol{e}_{t+1}$ denotes the elementary $(T+1) \times 1$ vector with 1 in the $(t+1)^{th}$ position.

Taking first-differences in (2) yields

$$\Delta \mathbf{y}_t = \alpha \Delta \mathbf{y}_{t-1} + \Delta \boldsymbol{\varepsilon}_t = \alpha \Delta \mathbf{y}_{t-1} + P_N \Delta \boldsymbol{\upsilon}_t,\ t = 2, ..., T. \tag{7}$$

One can express $\Delta \mathbf{y}_t$ as a linear form of the innovations as follows:

$$\underset{N\times 1}{\Delta \mathbf{y}_t} = \underset{N\times N(T+1)}{(\boldsymbol{\gamma}_t \otimes P_N)} \underset{N(T+1)\times 1}{\boldsymbol{\upsilon}} + \delta_{0,t}^{**} \underset{N\times 1}{\boldsymbol{\eta}}, \tag{8}$$

with $\boldsymbol{\gamma}_{t-1} = \left[\delta_1 (\alpha - 1) \alpha^{t-2}, (\alpha - 1) \alpha^{t-3}, ..., (\alpha - 1) \alpha^0, 1, \mathbf{0}_{1\times T-(t-1)}\right]$ is a $1 \times (T+1)$ row vector, $\delta^{**}_{0,t-1} = \left[\alpha^{t-2} - \delta_0 (1-\alpha) \alpha^{t-2}\right]$, while $\boldsymbol{\upsilon}$, $\boldsymbol{\eta}$ have been defined above. Stacking (2) and (7) over $t = 2, ..., T$ yields

$$\underset{N(T-1)\times 1}{\mathbf{y}} = \alpha \underset{N(T-1)\times 1}{\mathbf{y}_{-1}} + \underset{N(T-1)\times 1}{\mathbf{u}}, \tag{9}$$

and

$$\underset{N(T-1)\times 1}{\Delta\mathbf{y}} = \alpha \underset{N(T-1)\times 1}{\Delta\mathbf{y}_{-1}} + \underset{N(T-1)\times 1}{\Delta\boldsymbol{\varepsilon}}, \tag{10}$$

respectively, where $\mathbf{y} = (\mathbf{y}'_2, ..., \mathbf{y}'_T)'$, $\mathbf{y}_{-1} = \left(\mathbf{y}'_1, ..., \mathbf{y}'_{T-1}\right)'$, $\mathbf{u} = (\mathbf{u}'_2, ..., \mathbf{u}'_T)'$, $\Delta\mathbf{y} = (\Delta\mathbf{y}'_2, ..., \Delta\mathbf{y}'_T)'$, $\Delta\mathbf{y}_{-1} = \left(\Delta\mathbf{y}'_1, ..., \Delta\mathbf{y}'_{T-1}\right)'$, $\Delta\boldsymbol{\varepsilon} = (\Delta\boldsymbol{\varepsilon}'_2, ..., \Delta\boldsymbol{\varepsilon}'_T)'$.

Let $Z_D = diag\left(Y^0, Y^1, ..., Y^{T-2}\right)$ be a $N(T-1) \times T(T-1)/2$ block-diagonal matrix, where a typical block is $Y^s = (\mathbf{y}_{0,}\mathbf{y}_{1,}, ..., \mathbf{y}_s)$, a $N \times (s+1)$ matrix, where $\mathbf{y}_\tau = (y_{1\tau}, y_{2\tau}, ..., y_{N\tau})'$, a $N \times 1$ vector. Also, let $Z_L = diag(\Delta\mathbf{y}_1, \Delta\mathbf{y}_2, ..., \Delta\mathbf{y}_{T-1})$ be a $N(T-1) \times (T-1)$ matrix, where each block is given by $\Delta\mathbf{y}_s = (\Delta y_{1s}, \Delta y_{2s}, ..., \Delta y_{Ns})'$, a $N \times 1$ vector. The following assumptions are maintained:

Assumption 1 (error components): (i) The random variables $\{\upsilon_{it} : 1 \leq i \leq N,\ 0 \leq t \leq T\}$ are independently distributed with zero mean and finite variance σ^2_υ. Furthermore, $\sup_{1\leq i\leq N, 0\leq t\leq T} E\left|\upsilon_{it}\right|^{4+\delta} < \infty$ for some $\delta > 0$. (ii) The random variables $\{\eta_i : 1 \leq i \leq N\}$ are independently distributed with zero mean and finite variance σ^2_η. Furthermore, $\sup_{1\leq i\leq N} E\left|\eta_i\right|^{4+\delta} < \infty$ for some $\delta > 0$. (iii) The processes $\{\upsilon_{it}\}$ and $\{\eta_i\}$ are totally independent.

Assumption 2 (weighting matrix and space of MA parameter): (i) All diagonal elements of W_N equal zero. (ii) The spatial moving average parameter satisfies $\theta \in (-c_{1,\theta}, c_{2,\theta})$ with $0 < c_{1,\theta}, c_{2,\theta} \leq c_\theta < \infty$. (iii) The matrix W_N is non-singular and $P_N = I_N + \theta W_N$ is non-singular for all $\theta \in (-c_{1,\theta}, c_{2,\theta})$. (iv) The row and column sums of W_N and $(I_N + \theta W_N)$ are bounded uniformly in absolute value.

The assumptions above are standard in the spatial literature, see e.g. Kelejian and Prucha (2010). Notice that Assumption 1 permits cross-sectional heteroskedasticity in ε_{it}, through the weighting matrix W_N. Serial independence in the error can be relaxed by allowing ε_{it} to follow a finite MA process. An AR process can be accommodated using further lags of y on the right-hand side of the model. Assumption 2(i) is just a normalisation of the model and implies that no individual is viewed as its own neighbour. Assumptions 2(ii)-(iiii) concern the parameter space of θ and are discussed in detail by Kelejian and Prucha (2010, Section 2.2). Assumption 2(iv) implies that there is no dominant unit in the sample, i.e. an individual unit that is correlated with all remaining individuals. We will study the factor structure case, which violates this scenario, in Section 4. Notice that the assumptions above do not depend on a particular ordering of the data, which can be arbitrary. For reasons of generality the elements of W_N, and by implication of $\mathbf{y}$ with a slight abuse of notation, are permitted to depend on N, that is to form triangular arrays. This is due to the fact that for "boundary" elements the connectedness structure may change as new data points are added. This implies that the asymptotics require the use of a CLT for triangular arrays (see e.g. Davidson, 1994, Ch. 24).

The following proposition shows that the following moment conditions remain valid for the panel autoregressive model with spatially correlated errors.

Proposition 1 *Under Assumptions 1-2, the following $T(T-1)/2$ moment conditions are valid in the first-differenced model* (10)*:*

$$\mathbf{m}_{N,D}(\alpha) = N^{-1} Z'_D \Delta\boldsymbol{\varepsilon} \xrightarrow{p} \mathbf{0}. \tag{11}$$

Furthermore, under mean-stationarity, $\delta_0 = 1/(1-\alpha)$, the following $T-1$ moment conditions are valid in the levels model (9)*:*

$$\mathbf{m}_{N,L}(\alpha) = N^{-1} Z'_L \mathbf{u} \xrightarrow{p} \mathbf{0}. \tag{12}$$

Proof. See Appendix A. ■

The above proposition demonstrates that instruments with respect to lagged values of the endogenous regressor remain valid under spatial dependence. Therefore, under certain regularity conditions it will be shown that Generalised Method of Moment estimators making use of these moment conditions are consistent and asymptotically normal with mean zero. In particular, define

$$Z_S \equiv \begin{bmatrix} Z_D & 0 \\ 0 & Z_L \end{bmatrix}; \mathbf{y}_S \equiv \begin{bmatrix} \Delta\mathbf{y} \\ \mathbf{y} \end{bmatrix}; \mathbf{y}_{-1,S} \equiv \begin{bmatrix} \Delta\mathbf{y}_{-1} \\ \mathbf{y}_{-1} \end{bmatrix}; \mathbf{u}_S \equiv \begin{bmatrix} \Delta\boldsymbol{\varepsilon} \\ \mathbf{u} \end{bmatrix}.$$

Also, let $A_{1,D,N}$ and $A_{1,S,N}$ be sequences of possibly random, non-negative definite matrices of order $\zeta_{1,D} \times \zeta_{1,D}$ and $\zeta_{1,S} \times \zeta_{1,S}$, respectively, where $\zeta_{1,D} = T(T-1)/2$ and $\zeta_{1,S} = T(T-1)/2 + T - 1$. The following assumption is employed for the identification of the autoregressive parameter, α.

Assumption 3(i) (identification of α**)**: $N^{-1}Z_D' Z_D \xrightarrow{p} Q_{Z_D}$, $N^{-1}Z_D'\Delta\mathbf{y}_{-1} \xrightarrow{p} \mathbf{q}_{Z_D\Delta y_{-1}}$, $N^{-1}Z_S'Z_S \xrightarrow{p} Q_{Z_S}$, $N^{-1}Z_S'\mathbf{y}_{-1,S} \xrightarrow{p} \mathbf{q}_{Z_S y_{-1,S}}$, all finite matrices (vectors) with full column rank (non-zero entries). $A_{1,D,N}$ and $A_{1,S,N}$ have full rank and $A_{1,D,N} \xrightarrow{p} A_{1,D}$, $A_{1,S,N} \xrightarrow{p} A_{1,S}$.

The first-differenced (FD) GMM estimator is defined as the minimiser of the following quadratic form:

$$\widehat{\alpha}_{D\left(A_{1,D,N}\right)} \equiv \arg\min_{\alpha} \mathbf{m}_{N,D}(\alpha)' A_{1,D,N}\mathbf{m}_{N,D}(\alpha). \tag{13}$$

Combining (11) and (12) yields

$$\mathbf{m}_{N,S}(\alpha) = N^{-1}Z_S'\mathbf{u}_S \xrightarrow{p} \mathbf{0}. \tag{14}$$

The system (SYS) GMM estimator is defined as the minimiser of the following quadratic form:

$$\widehat{\alpha}_{S\left(A_{1,S,N}\right)} \equiv \arg\min_{\alpha} \mathbf{m}_{N,S}(\alpha) A_{1,S,N}\mathbf{m}_{N,S}(\alpha), \tag{15}$$

Setting the first-order conditions equal to zero and solving for the unknown value of α in (13) and (15) yields

$$\widehat{\alpha}_D = \left[\Delta\mathbf{y}_{-1}' Z_D A_{1,D,N} Z_D' \Delta\mathbf{y}_{-1}\right]^{-1}\left[\Delta\mathbf{y}_{-1}' Z_D A_{1,D,N} Z_D' \Delta\mathbf{y}\right], \tag{16}$$

and

$$\widehat{\alpha}_S = \left[\mathbf{y}_{-1,S}' Z_S A_{1,S,N} Z_S' \mathbf{y}_{-1,S}\right]^{-1}\left[\mathbf{y}_{-1,S}' Z_S A_{1,S,N} Z_S' \mathbf{y}_S\right], \tag{17}$$

respectively. The following theorem establishes the consistency and asymptotic normality of the above estimators.

Theorem 2 *Suppose Assumptions 1-3(i), and* (11) *hold true. Let* $\Omega_{1,D,N} = var\left[\sqrt{N}\mathbf{m}_{N,D}(\alpha)\right]$ *be a sequence of symmetric, non-negative definite matrices with rank greater than or equal to* $\zeta_{1,D}$*, such that* $\lambda_{\min}(\Omega_{1,D,N}) \geq c > 0$*, and* $\Omega_{1,D,N} \xrightarrow{p} \Omega_{1,D} = asy.var\left[\sqrt{N}\mathbf{m}_{N,D}(\alpha)\right]$*. The GMM estimator in* (16) *is consistent and*

$$\sqrt{N}\left(\widehat{\alpha}_{D\left(A_{1,D,N}\right)} - \alpha\right) \xrightarrow{d} N(0, V_D), \tag{18}$$

where

$$V_D = \left[\mathbf{q}_{Z_D\Delta y_{-1}} A_{1,D}' \mathbf{q}_{Z_D\Delta y_{-1}}'\right]^{-1} \mathbf{q}_{Z_D\Delta y_{-1}} A_{1,D}\Omega_{1,D}A_{1,D}\mathbf{q}_{Z_D\Delta y_{-1}}' \left[\mathbf{q}_{Z_D\Delta y_{-1}} A_{1,D}\mathbf{q}_{Z_D\Delta y_{-1}}'\right]^{-1}. \tag{19}$$

In addition to the assumptions above, suppose that (12) *holds true and let* $\Omega_{1,S,N} = var\left[\sqrt{N}\mathbf{m}_{N,S}(\alpha)\right]$ *be a sequence of symmetric, non-negative definite matrices with rank greater than or equal to* $\zeta_{1,S}$*, such that* $\lambda_{\min}(\Omega_{1,S,N}) \geq c > 0$*, and* $\Omega_{1,S,N} \xrightarrow{p} \Omega_{1,S} = asy.var\left[\sqrt{N}\mathbf{m}_{N,S}(\alpha)\right]$*. The GMM estimator in* (17) *is consistent and*

$$\sqrt{N}(\widehat{\alpha}_S - \alpha) \xrightarrow{d} N(0, V_S), \tag{20}$$

where

$$V_S = \left[\mathbf{q}_{Z_S y_{-1,S}} A_{1,S} \mathbf{q}'_{Z_S y_{-1,S}}\right]^{-1} \mathbf{q}_{Z_S y_{-1,S}} A_{1,S} \Omega_{1,S} A_{1,S} \mathbf{q}'_{Z_S y_{-1,S}} \left[\mathbf{q}_{Z_S y_{-1,S}} A_{1,S} \mathbf{q}'_{Z_S y_{-1,S}}\right]^{-1}. \quad (21)$$

Proof. *See Appendix A.* ∎

A first-stage choice for $A_{1,D,N}$ can be such that

$$A_{1,D} = N^{-1} Z'_D \left(DD' \otimes I_N\right) Z_D,$$

which takes into account that the first-differenced operator creates serial correlation in the errors but ignores spatial correlation. Similarly, for $A_{1,S,N}$ one can choose

$$A_{1,S} = N^{-1} Z'_S \begin{bmatrix} (D \otimes I_N)(D \otimes I_N)' & 0 \\ 0 & (I_{T-1} \otimes I_N) \end{bmatrix} Z_S.$$

The optimal GMM estimators are obtained by replacing $A_{1,D,N}$ and $A_{1,S,N}$ by $\Omega^{-1}_{1,D,N}$ and $\Omega^{-1}_{1,S,N}$, respectively, in which case (19) and (21) reduce to

$$V_D = \left[\mathbf{q}_{Z_D \Delta y_{-1}} \Omega^{-1}_{1D} \mathbf{q}'_{Z_D \Delta y_{-1}}\right]^{-1},$$

and

$$V_S = \left[\mathbf{q}_{Z_S y_{-1,S}} \Omega^{-1}_{1,S} \mathbf{q}'_{Z_S y_{-1,S}}\right]^{-1}.$$

The distributional results hold as well if the unobserved $\Omega^{-1}_{1,D,N}$, $\Omega^{-1}_{1,S,N}$ are replaced by consistent estimates. In particular, notice that $\Omega_{1,D,N}$ can be partitioned as follows:

$$\Omega_{1,D,N} = \begin{bmatrix} \Omega_{1,22,D,N} & \cdots & \Omega_{1,2T,D,N} \\ & \ddots & \\ \Omega_{1,T2,D,N} & \cdots & \Omega_{1,TT,D,N} \end{bmatrix},$$

where $\Omega_{1,ts,D,N} = N^{-1} E Y^{t-2\prime} \Delta\boldsymbol{\varepsilon}_t \Delta\boldsymbol{\varepsilon}'_s Y^{s-2}$. Letting the pq^{th} element of $\Omega_{1,ts,D,N}$ be denoted by $\omega_{1,pq,ts,D,N}$, we have

$$\begin{aligned} \omega_{1,pq,ts,D,N} &= N^{-1} E \mathbf{y}'_p \Delta\boldsymbol{\varepsilon}_t \Delta\boldsymbol{\varepsilon}'_s \mathbf{y}_q \\ &= N^{-1} \left[\boldsymbol{v}' \left(\boldsymbol{\beta}'_p \boldsymbol{d}_t \otimes P'_N P_N\right) \boldsymbol{v} + \delta^*_{0,p} \boldsymbol{\eta}' \left(\boldsymbol{d}_t \otimes P_N\right) \boldsymbol{v}\right] \\ &\quad \times \left[\boldsymbol{v}' \left(\boldsymbol{d}'_s \boldsymbol{\beta}_q \otimes P'_N P_N\right) \boldsymbol{v} + \delta^*_{0,q} \boldsymbol{\eta}' \left(\boldsymbol{d}_s \otimes P_N\right) \boldsymbol{v}\right] \\ &= N^{-1} 2tr \left[\left(\boldsymbol{\beta}'_p \boldsymbol{d}_t \otimes P'_N P_N\right) \Sigma_v \left(\boldsymbol{d}'_s \boldsymbol{\beta}_q \otimes P'_N P_N\right) \Sigma_v\right] \\ &\quad + \delta^*_{0,p} \delta^*_{0,q} tr \left[\left(\boldsymbol{d}_t \otimes P_N\right) \Sigma_v \left(\boldsymbol{d}_s \otimes P_N\right) \Sigma_\eta\right], \end{aligned}$$

where $\Sigma_v = \sigma^2_v I_{N(T+1)}$, $\Sigma_\eta = \sigma^2_\eta I_N$ and the remaining variables have been already defined.[1] Therefore, an expectations based operator for $\omega_{1,pq,ts,D,N}$ will replace the true value of the parameters above by their consistent estimates, obtained from the first stage. A consistent estimate for θ, required to compute P_N, can be obtained based on the estimator proposed by Fingleton (2008), applied on the residual vector $\widehat{\mathbf{u}} = \mathbf{y} - \widehat{\alpha}\mathbf{y}_{-1}$; see also Kapoor, Kelejian and Prucha (2007).

An alternative estimator for $\Omega_{1,ts,D,N}$ can be obtained by ignoring the fact that the instruments are stochastic variables, based on

$$\widetilde{\Omega}_{1,D,N} = N^{-1} Z'_D \widehat{\Sigma}_{\Delta_\varepsilon,N} Z_D,$$

[1] This expression easily follows from the expectation of $\psi^2_{3,ts,N}$ in the proof of Proposition 3 in the Appendix with $\widetilde{W}_N$ replaced by I_N.

where $\widehat{\Sigma}_{\Delta_\varepsilon,N}$ is a consistent estimate of

$$E\left[\Delta\boldsymbol{\varepsilon}\Delta\boldsymbol{\varepsilon}'\right] = E\left[\left(D\otimes P_N\right)\boldsymbol{v}\boldsymbol{v}'\left(D'\otimes P_N'\right)\right] = \sigma_v^2\left(D\otimes P_N\right)\left(D'\otimes P_N'\right),$$

with unknown parameters σ_v^2 and θ. This is sub-optimal in the sense that $\widetilde{\Omega}_{1,D,N}$ is not a consistent estimator for $\Omega_{1,D,N}$, however, it is computationally simpler and results in a consistent GMM estimator of α. Similar analysis applies to $\Omega_{1,S,N}$. Block bootstrapping procedures for spatially dependent observations are also available; see e.g. Hall (1985) and Anselin (1990). We will explore this alternative in Section 3.1.

3 Spatial Instruments: Validity, Relevance and Redundancy

It turns out that under spatially correlated errors, an additional set of moment conditions becomes valid and is relevant in the sense that it is correlated with the endogenous regressor. This is demonstrated in the proposition below. In particular, let $\widetilde{Z}_D = diag\left(\widetilde{W}_N Y^0, \widetilde{W}_N Y^1, ..., \widetilde{W}_N Y^{T-2}\right)$ be a $N(T-1)\times T(T-1)/2$ block-diagonal matrix, where $\widetilde{W}_N = W_N + W_N'$ is a symmetric matrix, and Y^s has been defined above. Effectively $\widetilde{W}_N$ is a matrix the i^{th} row of which contains non-zero values at the entries corresponding to the individuals which unit i is spatially correlated with. Also, let $\widetilde{Z}_L = diag\left(\widetilde{W}_N\Delta\mathbf{y}_1, \widetilde{W}_N\Delta\mathbf{y}_2, ..., \widetilde{W}_N\Delta\mathbf{y}_{T-1}\right)$ be a $N(T-1)\times(T-1)$ matrix, where $\Delta\mathbf{y}_s$ has been defined previously.

Proposition 3 *Under Assumptions 1-2, the following $T(T-1)/2$ moment conditions are valid in the first-differenced model* (10)*:*

$$\widetilde{\mathbf{m}}_{N,D}(\alpha) = N^{-1}\widetilde{Z}_D'\Delta\boldsymbol{\varepsilon} \xrightarrow{p} \mathbf{0}, \tag{22}$$

with

$$\widetilde{\mathbf{g}}_{N,D}(\alpha) = N^{-1}\widetilde{Z}_D'\Delta\boldsymbol{y}_{-1} \xrightarrow{p} \mathbf{q}_{\widetilde{Z}_D\Delta y_{-1}}, \tag{23}$$

where $\mathbf{q}_{\widetilde{Z}_D\Delta y_{-1}} = \left(q_{1,D}, ..., q_{T(T-1)/2,D}\right)'$ denotes a $T(T-1)/2\times 1$ column vector with $q_{k,D}\neq 0$, in general. Furthermore, the following $T-1$ moment conditions are valid in the levels model (9)*:*

$$\widetilde{\mathbf{m}}_{N,L}(\alpha) = N^{-1}\widetilde{Z}_L'\boldsymbol{u} \xrightarrow{p} \mathbf{0}, \tag{24}$$

with

$$\widetilde{\mathbf{g}}_{N,L}(\alpha) = N^{-1}\widetilde{Z}_L'\boldsymbol{y}_{-1} \xrightarrow{p} \mathbf{q}_{\widetilde{Z}_L y_{-1}}, \tag{25}$$

where $\mathbf{q}_{\widetilde{Z}_L y_{-1}} = \left(q_{1,L}, ..., q_{T-1,L}\right)'$ denotes a $(T-1)\times 1$ column vector with $q_{k,L}\neq 0$, in general.

Proof. See Appendix A. ■

The above proposition shows that the spatial instruments are valid and relevant as well, so long as $\theta\neq 0$, as shown in the appendix.

Remark 4 *Observe that unlike Proposition 1 we have not imposed mean stationarity for the equations in levels. Intuitively, this is because $E\eta_i\eta_j = 0\ \forall i\neq j$, under the maintained assumptions. Therefore, the spatial moment conditions in the equations in levels are valid in this case even if the standard moment conditions are not. We will investigate the consequences of this result in simulations.*

Define

$$\underset{N(T-1)\times T(T-1)}{Z_{\widetilde{D}}} \equiv \left[Z_D\ ,\ \widetilde{Z}_D\right];\quad \underset{N(T-1)\times 2(T-1)}{Z_{\widetilde{L}}} \equiv \left[Z_L\ ,\ \widetilde{Z}_L\right];\quad \underset{N(T-1)\times[T(T-1)+2(T-1)]}{Z_{\widetilde{S}}} \equiv \begin{bmatrix} Z_{\widetilde{D}} & 0 \\ 0 & Z_{\widetilde{L}} \end{bmatrix},$$

$$\mathbf{m}_{N,\widetilde{D}}(\alpha) \equiv N^{-1} Z'_{\widetilde{D}} \Delta \boldsymbol{\varepsilon},$$

and

$$\mathbf{m}_{N,\widetilde{S}}(\alpha) \equiv N^{-1} Z'_{\widetilde{S}} \Delta \mathbf{u}_S.$$

Let $A_{2,D,N}$ and $A_{2,S,N}$ be sequences of possibly random, non-negative definite matrices of order $\zeta_{2,D} \times \zeta_{2,D}$, and $\zeta_{2,S} \times \zeta_{2,S}$, respectively, where $\zeta_{2,D} = T(T-1)$, $\zeta_{2,S} = T(T-1) + 2(T-1)$. Furthermore, let $\mathbf{q}_{Z_{\widetilde{D}}\Delta y_{-1}} = \left(\mathbf{q}'_{Z_D \Delta y_{-1}}, \mathbf{q}'_{\widetilde{Z}_D \Delta y_{-1}}\right)'$, $\mathbf{q}_{Z_{\widetilde{S}} y_{-1,S}} = \left(\mathbf{q}'_{Z_{\widetilde{D}}\Delta y_{-1}}, \mathbf{q}'_{Z_{\widetilde{L}} y_{-1}}\right)'$ with $\mathbf{q}_{Z_{\widetilde{L}} y_{-1}} = \left(\mathbf{q}'_{Z_L y_{-1}}, \mathbf{q}'_{\widetilde{Z}_L y_{-1}}\right)'$. The following assumption is employed for the identification of α:

Assumption 3(ii) (identification of α): $N^{-1} Z'_{\widetilde{D}} Z_{\widetilde{D}} \xrightarrow{p} Q_{Z_{\widetilde{D}}}$, $N^{-1} Z'_{\widetilde{D}} \Delta \mathbf{y}_{-1} \xrightarrow{p} \mathbf{q}_{Z_{\widetilde{D}}\Delta y_{-1}}$, $N^{-1} Z'_{\widetilde{S}} Z_{\widetilde{S}} \xrightarrow{p} Q_{Z_{\widetilde{S}}}$, , $N^{-1} Z'_{\widetilde{S}} \mathbf{y}_{-1,S} \xrightarrow{p} \mathbf{q}_{Z_{\widetilde{S}} y_{-1,S}}$, all finite matrices (vectors) with full column rank (non-zero entries). $A_{2,D,N}$ and $A_{2,S,N}$ have full rank, such that $A_{2,D,N} \xrightarrow{p} A_{2,D}$, $A_{2,S,N} \xrightarrow{p} A_{2,S}$.

Let

$$\widetilde{\alpha}_{D(A_{2,D,N})} = \left[\Delta \mathbf{y}'_{-1} Z_{\widetilde{D}} A_{2,D,N} Z'_{\widetilde{D}} \Delta \mathbf{y}_{-1}\right]^{-1} \left[\Delta \mathbf{y}'_{-1} Z_{\widetilde{D}} A_{2,D,N} Z'_{\widetilde{D}} \Delta \mathbf{y}\right], \tag{26}$$

and

$$\widetilde{\alpha}_{S(A_{2,S,N})} = \left[\mathbf{y}'_{-1,S} Z_{\widetilde{S}} A_{2,S,N} Z'_{\widetilde{S}} \mathbf{y}_{-1,S}\right]^{-1} \left[\mathbf{y}'_{-1,S} Z_{\widetilde{S}} A_{2,S,N} Z'_{\widetilde{S}} \mathbf{y}_S\right], \tag{27}$$

be the FD and SYS GMM estimators that combine the standard and spatial instruments. The following theorem establishes the consistency and asymptotic normality of these estimators.

Theorem 5 *Suppose Assumptions 1-3(ii), and* (22), (23) *hold true. Let* $\Omega_{2,D,N} = var\left[\sqrt{N}\mathbf{m}_{N,\widetilde{D}}(\alpha)\right]$ *be a sequence of symmetric non-negative definite matrices with rank greater than or equal to* $\zeta_{2,D}$, *such that* $\lambda_{\min}(\Omega_{2,D,N}) \geq c > 0$, *and* $\Omega_{2,D,N} \xrightarrow{p} \Omega_{2,D} = asy.var\left[\sqrt{N}\mathbf{m}_{N,\widetilde{D}}(\alpha)\right]$. *The GMM estimator in* (26) *is consistent and*

$$\sqrt{N}(\widetilde{\alpha}_D - \alpha) \xrightarrow{d} N\left(0, V_{\widetilde{D}}\right), \tag{28}$$

where

$$V_{\widetilde{D}} = \left[\mathbf{q}_{Z_{\widetilde{D}}\Delta y_{-1}} A_{2,D} \mathbf{q}'_{Z_{\widetilde{D}}\Delta y_{-1}}\right]^{-1} \mathbf{q}_{Z_{\widetilde{D}}\Delta y_{-1}} A_{2,D} \Omega_{2,D} A_{2,D} \mathbf{q}'_{Z_{\widetilde{D}}\Delta y_{-1}} \left[\mathbf{q}_{Z_{\widetilde{D}}\Delta y_{-1}} A_{2,D} \mathbf{q}'_{Z_{\widetilde{D}}\Delta y_{-1}}\right]^{-1}.$$

In addition to the assumptions above, suppose that (12), (24)-(25) *hold true and let* $\Omega_{2,S,N} = var\left[\sqrt{N}\mathbf{m}_{N,\widetilde{S}}(\alpha)\right]$ *be a sequence of symmetric, non-negative definite matrices with rank greater than or equal to* $\zeta_{2,S}$, *such that* $\lambda_{\min}(\Omega_{2,S,N}) \geq c > 0$, *and* $\Omega_{2,S,N} \xrightarrow{p} \Omega_{2,S} = asy.var\left[\sqrt{N}\mathbf{m}_{N,\widetilde{S}}(\alpha)\right]$. *The GMM estimator in* (27) *is consistent and*

$$\sqrt{N}(\widetilde{\alpha}_S - \alpha) \xrightarrow{d} N\left(0, V_{\widetilde{S}}\right), \tag{29}$$

where

$$V_{\widetilde{S}} = \left[\mathbf{q}_{Z_{\widetilde{S}} y_{-1,S}} A_{2,S} \mathbf{q}'_{Z_{\widetilde{S}} y_{-1,S}}\right]^{-1} \mathbf{q}_{Z_{\widetilde{S}} y_{-1,S}} A_{2,S} \Omega_{2,S} A_{2,S} \mathbf{q}'_{Z_{\widetilde{S}} y_{-1,S}} \left[\mathbf{q}_{Z_{\widetilde{S}} y_{-1,S}} A_{2,S} \mathbf{q}'_{Z_{\widetilde{S}} y_{-1,S}}\right]^{-1}.$$

Proof. *See Appendix A.* ■

One interesting issue that arises is whether these spatial moment conditions are redundant or not. It is well known that adding more moment conditions will not hurt asymptotically since the asymptotic variance of the GMM estimator that arises from the enlarged set of moment conditions is less than or equal to the asymptotic variance of the GMM estimator from the smaller (nested) set. However, if the enlarged set of moment conditions does not increase the asymptotic efficiency of the GMM estimator, the additional instruments are redundant.[2] To shed some light on this issue we will consider the case where $T = 2$ for the model in first-differences; there is a single equation given by

$$\Delta y_{i2} = \alpha \Delta y_{i1} + \Delta \varepsilon_{i2},$$

and a single instrument with respect to lagged values of the endogenous regressor of individual i, such that the reduced form (instrumental variable regression) equation is

$$\Delta y_{i1} = \gamma y_{i0} + e_i.$$

The FD GMM estimator of α reduces to a simple IV estimator, given by

$$\begin{aligned} \widehat{\alpha} &= \frac{\widehat{\text{cov}}\left(\Delta y_{i2}, \widehat{\gamma} y_{i0}\right)}{\widehat{\text{var}}\left(\widehat{\gamma} y_{i0}\right)} = \frac{\widehat{\text{cov}}\left[\alpha\left(\Delta y_{i1} + \Delta \varepsilon_{i2}\right), \widehat{\gamma} y_{i0}\right]}{\widehat{\text{var}}\left(\widehat{\gamma} y_{i0}\right)} \\ &= \frac{\widehat{\text{cov}}\left[\alpha\left(\widehat{\gamma} y_{i0} + e_i\right) + \Delta \varepsilon_{i2}, \widehat{\gamma} y_{i0}\right]}{\widehat{\text{var}}\left(\widehat{\gamma} y_{i0}\right)} = \alpha + \frac{\widehat{\text{cov}}\left(\Delta \varepsilon_{i2}, y_{i0}\right)}{\widehat{\gamma}\widehat{\text{var}}\left(y_{i0}\right)}, \end{aligned}$$

where the last equality holds because e_i is orthogonal to y_{i0} by construction. Hence, we have

$$\sqrt{N}\left(\widehat{\alpha} - \alpha\right) = \sqrt{N}\frac{\widehat{\text{cov}}\left(\Delta \varepsilon_{i2}, y_{i0}\right)}{\widehat{\gamma}\widehat{\text{var}}\left(y_{i0}\right)} = \frac{1}{\widehat{\gamma}\widehat{\text{var}}\left(y_{i0}\right)} \sum_{i=1}^{N} \Delta \varepsilon_{i2} y_{i0} / \sqrt{N}. \tag{30}$$

Using Proposition 1 and Theorem 2 we have

$$\frac{1}{\widehat{\gamma}\widehat{\text{var}}\left(y_{i0}\right)} \xrightarrow{p} \frac{1}{\gamma \text{var}\left(y_{i0}\right)}, \tag{31}$$

while

$$\frac{\sum_{i=1}^{N} \Delta \varepsilon_{i2} y_{i0}}{\sqrt{N}} \xrightarrow{d} N\left[0, \text{var}\left(\Delta \varepsilon_{i2}\right) \text{var}\left(y_{i0}\right)\right], \tag{32}$$

since

$$\text{var}\left(\frac{\sum_{i=1}^{N} \Delta \varepsilon_{i2} y_{i0}}{\sqrt{N}}\right) = \frac{1}{N}\text{var}\left(\sum_{i=1}^{N} \Delta \varepsilon_{i2} y_{i0}\right) = E\left(\Delta \varepsilon_{i2}^2\right) E\left(y_{i0}^2\right). \tag{33}$$

A direct application of Slutsky theorem yields

$$\sqrt{N}\left(\widehat{\alpha} - \alpha\right) \xrightarrow{d} N\left(0, V_{\widehat{\alpha}}\right), \tag{34}$$

where

$$V_{\widehat{\alpha}} = \frac{\text{var}\left(\Delta \varepsilon_{i2}\right)}{\gamma^2 \text{var}\left(y_{i0}\right)}. \tag{35}$$

In addition, there exist spatial instruments with respect to the individual(s) unit i is spatially correlated with. The instrumental variable equation for the two-stage least-squares (2SLS) estimator that uses the enlarged set of moment conditions is

$$\Delta y_{i1} = \gamma_1 y_{i0} + \gamma_2 \sum_{j=1}^{N} w_{ij} y_{j0} + w_i.$$

The 2SLS estimator of α is given by

$$\widetilde{\alpha} = \frac{\widehat{\text{cov}}\left[\Delta y_{i2}, \widehat{\gamma}_1 y_{i0} + \widehat{\gamma}_2 \sum_{j=1}^{N} w_{ij} y_{j0}\right]}{\widehat{\text{var}}\left[\widehat{\gamma}_1 y_{i0} + \widehat{\gamma}_2 \sum_{j=1}^{N} w_{ij} y_{j0}\right]} = \alpha + \frac{\widehat{\text{cov}}\left(\Delta \varepsilon_{i2}, \widehat{\gamma}_1 y_{i0} + \widehat{\gamma}_2 \sum_{j=1}^{N} w_{ij} y_{j0}\right)}{\widehat{\text{var}}\left(\widehat{\gamma}_1 y_{i0} + \widehat{\gamma}_2 \sum_{j=1}^{N} w_{ij} y_{j0}\right)}, \tag{36}$$

[2] Breusch, Qian, Schmidt and Wyhowski (1999) provide a general treatment of redundancy of moment conditions.

where, similarly as above, w_i is orthogonal to y_{i0} and $\sum_{j=1}^{N} w_{ij} y_{j0}$ by construction.[3] Therefore, one has

$$\frac{1}{\widehat{\text{var}}\left[\widehat{\gamma}_1 y_{i0} + \widehat{\gamma}_2 \sum_{j=1}^{N} w_{ij} y_{j0}\right]} \xrightarrow{p} \frac{1}{\text{var}\left[\gamma_1 y_{i0} + \gamma_2 \sum_{j=1}^{N} w_{ij} y_{j0}\right]}, \tag{37}$$

and

$$\frac{\sum_{i=1}^{N} \Delta\varepsilon_{i2} \left[\widehat{\gamma}_1 y_{i0} + \widehat{\gamma}_2 \sum_{j=1}^{N} w_{ij} y_{j0}\right]}{\sqrt{N}} \overset{a}{\sim} N\left[0, \text{var}\left(\Delta\varepsilon_{i2}\right) \times \text{var}\left(\gamma_1 y_{i0} + \gamma_2 \sum_{j=1}^{N} w_{ij} y_{j0}\right)\right].$$

The asymptotic distribution of $\sqrt{N}\left(\widetilde{\alpha} - \alpha\right)$ is

$$\sqrt{N}\left(\widetilde{\alpha} - \alpha\right) \xrightarrow{d} N\left(0, V_{\widetilde{\alpha}}\right), \tag{38}$$

where

$$\begin{aligned} V_{\widetilde{\alpha}} &= \frac{\text{var}\left(\Delta\varepsilon_{i2}\right)}{\text{var}\left(\gamma_1 y_{i0} + \gamma_2 \sum_{j=1}^{N} w_{ij} y_{j0}\right)} \\ &= \frac{\text{var}\left(\Delta\varepsilon_{i2}\right)}{\gamma_1^2 \text{var}\left(y_{i0}\right) + \gamma_2^2 \text{var}\left(\sum_{j=1}^{N} w_{ij} y_{j0}\right) + 2\gamma_1\gamma_2 \text{cov}\left(y_{i0}, \sum_{j=1}^{N} w_{ij} y_{j0}\right)}. \end{aligned} \tag{39}$$

It is straightforward to show that the denominator in (39) is larger than in (35) and therefore $\widetilde{\alpha}$ is asymptotically more efficient than $\widehat{\alpha}$. This holds true unless $\gamma_2 = 0$, in which case $\gamma_1 = \gamma$ and $\widetilde{\alpha}$ is asymptotically equivalent to $\widehat{\alpha}$. This is an intuitive result because $\gamma_2 = 0$ implies that, *conditional* on y_{i0}, $\sum_{j=1}^{N} w_{ij} y_{j0}$ is not correlated with the endogenous regressor. We will investigate further the condition $\gamma_2 = 0$ by considering the first-stage coefficient $\widehat{\gamma}_2$, which equals

$$\begin{aligned} \widehat{\gamma}_2 &= \frac{\widehat{\text{cov}}\left(\Delta y_{i1}, \sum_{j=1}^{N} w_{ij} y_{j0}\right) \widehat{\text{var}}\left(y_{i0}\right) - \widehat{\text{cov}}\left(\Delta y_{i1}, y_{i0}\right) \widehat{\text{cov}}\left(y_{i0}, \sum_{j=1}^{N} w_{ij} y_{j0}\right)}{\widehat{\text{var}}\left(y_{i0}\right) \widehat{\text{var}}\left(\sum_{j=1}^{N} w_{ij} y_{j0}\right) - \widehat{\text{cov}}\left(y_{i0}, \sum_{j=1}^{N} w_{ij} y_{j0}\right)^2} \\ \xrightarrow{p} \gamma_2 &= \frac{\text{cov}\left(\Delta y_{i1}, \sum_{j=1}^{N} w_{ij} y_{j0}\right) \text{var}\left(y_{i0}\right) - \text{cov}\left(\Delta y_{i1}, y_{i0}\right) \text{cov}\left(y_{i0}, \sum_{j=1}^{N} w_{ij} y_{j0}\right)}{\text{var}\left(y_{i0}\right) \text{var}\left(\sum_{j=1}^{N} w_{ij} y_{j0}\right) - \text{cov}\left(y_{i0}, \sum_{j=1}^{N} w_{ij} y_{j0}\right)^2}. \end{aligned} \tag{40}$$

[3] Notice that it is also possible to investigate the properties of the GMM estimator that makes use of the optimal weighting matrix. This is asymptotically more efficient than $\widehat{\alpha}$ and $\widetilde{\alpha}$ when the spatial instruments are not redundant. In order to concentrate on the issue of redundancy of the set of additional instruments, however, it suffices to study the properties of the 2SLS estimator.

Without loss of generality suppose that the weighting matrix used is circular[4], such that

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & . & . & . & . & . & 0 \\ 0 & 0 & 1 & 0 & . & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ 0 & . & . & . & . & 0 & 0 & 0 & 1 \\ 1 & . & . & . & . & . & 0 & 0 & 0 \end{bmatrix}. \tag{41}$$

One can show that

$$\begin{aligned} \operatorname{var}(y_{i0}) &= \delta_0^2\sigma_\eta^2 + \delta_1^2\left(1+\theta^2\right)\sigma_v^2; \\ \operatorname{var}\left(\sum_{j=1}^N w_{ij}y_{j0}\right) &= 2\left[\delta_0^2\sigma_\eta^2 + \delta_1^2\left(1+\theta^2\right)\sigma_v^2\right]; \\ \operatorname{cov}(\Delta y_{i1}, y_{i0}) &= (\alpha-1)\left[\delta_0\sigma_\eta^2\left(\delta_0 - \frac{1}{1-\alpha}\right) + \delta_1^2\sigma_u^2\left(1+\theta^2\right)\right]; \\ \operatorname{cov}\left(\Delta y_{i1}, \sum_{j=1}^N w_{ij}y_{j0}\right) &= 2(\alpha-1)\theta\delta_1^2\sigma_v^2; \\ \operatorname{cov}\left(y_{i0}, \sum_{j=1}^N w_{ij}y_{j0}\right) &= 2\theta\delta_1^2\sigma_v^2. \end{aligned} \tag{42}$$

As a result, we have

$$\begin{aligned} \gamma_2 &= \frac{2(\alpha-1)\theta\delta_1^2\sigma_v^2\left[\left[\delta_0^2\sigma_\eta^2+\delta_1^2\left(1+\theta^2\right)\sigma_v^2\right] - \left[\delta_0\sigma_\eta^2\left(\delta_0-\frac{1}{1-\alpha}\right)+\delta_1^2\left(1+\theta^2\right)\sigma_v^2\right]\right]}{2\left[\left[\delta_0^2\sigma_\eta^2+\delta_1^2\left(1+\theta^2\right)\sigma_v^2\right]^2 - 2\left(\theta\delta_1^2\sigma_v^2\right)^2\right]} \\ &= \frac{(\alpha-1)\theta\delta_1^2\sigma_v^2\left[\delta_0^2\sigma_\eta^2 - \delta_0\sigma_\eta^2\left(\delta_0-\frac{1}{1-\alpha}\right)\right]}{\left[\delta_0^2\sigma_\eta^2+\delta_1^2\left(1+\theta^2\right)\sigma_v^2\right]^2 - 2\left(\theta\delta_1^2\sigma_v^2\right)^2} \\ &= -\frac{\theta\delta_0\delta_1^2\sigma_\eta^2\sigma_v^2}{\delta_0^4\sigma_\eta^4 + 2\delta_0^2\delta_1^2\sigma_\eta^2\left(1+\theta^2\right)\sigma_v^2 + \delta_1^4\left(1+\theta^4\right)\sigma_v^4}. \end{aligned} \tag{43}$$

Therefore, we can see that $\gamma_2 = 0$ for either $\delta_0 = 0$, or $\delta_1 = 0$ or $\theta = 0$. The last two zero conditions imply that the covariance between the endogenous regressor and the spatial instruments equals zero, as it is clear from (42). It is worth mentioning that γ_2 does not depend on the value of α so long as δ_0 and δ_1 are not a functions of α. Furthermore, since the denominator in γ_2 is always a non-negative number, $\gamma_2 \leq 0$ for $\theta \geq 0$ and $\delta_0 \geq 0$. The following figure illustrates graphically the value of $-\gamma_2$ for $\theta \in [0,1]$ and $\delta_0 \in [0,4]$, setting $\delta_1^2 = \sigma_\eta^2 = \sigma_v^2 = 1$. Observe that for any $\delta_0 > 0$, the value of the function increases as θ approaches unity. On the other hand, for any given $\theta > 0$ the value of the function initially gets larger as δ_0 increases from zero to a positive value, although it approaches zero as δ_0 increases further to large positive values. As an implication, if the y process is mean-stationary such that $\delta_0 = 1/(1-\alpha)$, the spatial instruments become redundant as $\alpha \to 1$ regardless of the value of θ. Of course, at the same time instruments with respect to lagged values of the endogenous regressor for individual i become weak for $\alpha \to 1$. This in turn implies that the spatial instruments become weak as well, given redundancy. We will investigate further the properties of GMM estimators that make use of spatial instruments using simulated data in Section 3.1.

[4] See e.g. Baltagi, Bresson and Pirotte (2007).

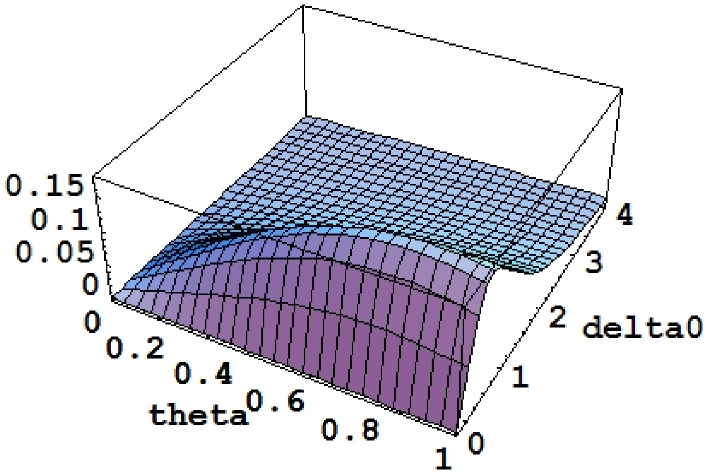

Figure 1

Another issue that arises is dominance; in particular, suppose that there is a dominant cross-sectional unit which is correlated with all remaining individuals. It turns out that instruments with respect to the dominant individual are always redundant. To see this, let the dominat unit be the N^{th} individual. We have

$$\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}y_{Ns}\Delta\varepsilon_{it}=y_{Ns}\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}\Delta\varepsilon_{it}=0, \tag{44}$$

and

$$\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}y_{Ns}\Delta y_{it-1}=y_{Ns}\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}\Delta y_{it-1}=0. \tag{45}$$

Therefore, the instrument with respect to the dominant unit is uncorrelated with Δy_{it-1}. Intuitively, if the same variable is used as instrument for all individuals, then it is fixed in the cross-sectional dimension and therefore it is asymptotically uncorrelated with the endogenous regressor. Notice that the existence of a dominant unit violates the uniform boundedness condition of spatial dependence, and indeed one of the conditions in Theorem 1.

3.1 A Short Monte Carlo Investigation

We will investigate the finite-sample performance of the estimators above using simulated data. The underlying generating process is given by

$$y_{it}=\alpha y_{it-1}+\eta_i+\varepsilon_{it},\ \varepsilon_{it}=\theta\sum_{j=1}^{N}w_{ij,N}\upsilon_{jt}+\upsilon_{it},\ i=1,...,N,\ t=1,...,T, \tag{46}$$

where $w_{ij,N}$ denotes the ij^{th} element of W_N, which is formulated as in (41), $\eta_i\sim i.i.d.N\left(0,\sigma_\eta^2\right)$, $\upsilon_{it}\sim i.i.d.N\left(0,\sigma_\upsilon^2\right)$. The initial value is

$$y_{i0}=\delta_0\eta_i+\delta_1\varepsilon_{i0},\ \varepsilon_{i0}=\theta\sum_{j=1}^{N}w_{ij,N}\upsilon_{j0}+\upsilon_{i0}.$$

We set $\alpha = 0.5$, $\theta = 0.5$, $T = 6$ and we normalise $\sigma_\eta^2 = \sigma_\varepsilon^2 = 1$, while $N \in \{100, 400, 800\}$. The initial conditions are such that $\delta_0 \in \left\{0, 1, (1-\alpha)^{-1}\right\}$ and $\delta_1 = \left(1-\alpha^2\right)^{-1/2}$. For $\delta_0 = (1-\alpha)^{-1}$ the process is mean-stationary and also it is variance-stationary given the chosen value for δ_1.

The results are provided in the table below. FD, FD† and FD* denote the first-differenced GMM estimators that utilise Z_D, $\widetilde{Z}_D$ and $Z_{\widetilde{D}}$, respectively, as defined earlier in the paper, while SYS, SYS† and SYS* denote the system GMM estimators that utilise Z_S, $\widetilde{Z}_S$ and $Z_{\widetilde{S}}$, where $\widetilde{Z}_S$ is defined in (58). Therefore, FD (SYS) makes use of the standard instruments that are available with respect to individual i, FD† (SYS†) makes use of the spatials instruments with respect to the individuals which unit i is spatially correlated with, and FD* (SYS*) combine the two sets of instruments.

Table 1. Performance in terms of mean point estimates and RMSE, $\alpha = 0.5$.

$T=6$	$\delta_0 = 0$			$\delta_0 = 1$			$\delta_0 = 2$		
	FD	FD†	FD*	FD	FD†	FD*	FD	FD†	FD*
N = 100	.450	.405	.429	.395	.380	.377	.440	.370	.407
	(.132)	(.192)	(.133)	(.200)	(.221)	(.197)	(.143)	(.240)	(.159)
N = 400	.486	.472	.480	.472	.467	.466	.486	.463	.477
	(.065)	(.092)	(.063)	(.095)	(.102)	(.090)	(.067)	(.100)	(.071)
N = 800	.493	.486	.490	.484	.482	.481	.492	.480	.487
	(.045)	(.065)	(.043)	(.067)	(.072)	(.063)	(.048)	(.079)	(.051)
	SYS	SYS†	SYS*	SYS	SYS†	SYS*	SYS	SYS†	SYS*
N = 100	.711	.489	.713	.642	.492	.646	.500	.498	.507
	(.224)	(.142)	(.257)	(.167)	(.143)	(.167)	(.099)	(.136)	(.094)
N = 400	.725	.495	.734	.654	.498	.663	.500	.501	.502
	(.229)	(.081)	(.237)	(.161)	(.081)	(.169)	(.053)	(.082)	(.052)
N = 800	.723	.497	.737	.656	.498	.666	.499	.499	.500
	(.223)	(.059)	(.238)	(.160)	(.060)	(.169)	(.034)	(.062)	(.038)

As we can see, the performace of FD and FD* is similar. In most cases FD has slightly less bias and slightly larger RMSE. This is not surprising; it is known that in finite samples and with a fixed value of N, using a larger number of instruments results in a trade-off between bias and efficiency. Of course, asymptotically the GMM estimator with the enlarged set of moment conditions is more efficient, providing that the additional moment conditions are not redundant. FD† is generally dominated by FD and FD* both in terms of bias as well as RMSE. Its performance deteriorates with higher values of δ_0, which, however, is also the case for the remaining first-differenced GMM estimators. Intuitively, this is a weak instruments problem; as δ_0 increases the proportion of the variance of the total disturbance that is due to the variance of the individual-specific effects gets larger. Essentially, abusing the notation, for $\delta_0 \to \infty$ we have $\sigma_\eta^2/\sigma_\varepsilon^2 \to \infty$ and the instruments become weak (see Blundell and Bond, 1998). The same intuition holds for the spatial instruments. Similarly to FD and FD*, the performance of SYS and SYS* is similar under all circumstances. However, both estimators are consistent only under mean-stationarity and they appear to exhibit a large upwards bias otherwise. SYS†, on the other hand, performs well under all situations and largely dominates SYS and SYS*, unless the process is mean-stationarity. Importantly, SYS† appears to dominate FD and FD* as well, unless δ_0 is large. We have experimented also with $\alpha = 0.8$; in this case SYS† uniformly dominates FD and FD* under all circumstances. To save space we do not report these results. In the section below it will be demonstrated that the spatial moment conditions can be used to construct consistent GMM estimators in situations were the standard GMM estimators are not consistent.

4 Spatial and Factor Structure Dependence

In this section we will consider a panel autoregressive model in which the disturbance contains a common factor structure, such that

$$\begin{aligned}\widetilde{y}_{it} &= \alpha\widetilde{y}_{it-1}+\widetilde{u}_{it},\ |\alpha|<1,\ i=1,...,N,\ t=1,...,T,\\ \widetilde{u}_{it} &= \eta_i+\boldsymbol{\lambda}_i'\boldsymbol{\phi}_t+\varepsilon_{it},\ \varepsilon_{it}=\theta\sum_{j=1}^{N}w_{ij,N}\upsilon_{jt}+\upsilon_{it},\end{aligned}$$

where $\boldsymbol{\phi}_t=(\phi_{1t},\phi_{2t},...,\phi_{nt})'$ is a $n\times 1$ vector of factors and $\boldsymbol{\lambda}_i=(\lambda_{1i},\lambda_{2i},...,\lambda_{ni})'$ is a $n\times 1$ vector of factor loadings. A similar structure is also studied by Pesaran and Tosetti (2011) and Chudik, Pesaran and Tosetti (2011). We make the following assumption regarding the factors and their loadings:

Assumption 4 (common factor component): (i) The random variables $\{\lambda_{ri}:1\leq i\leq N,\ 1\leq r\leq n\}$ are independently distributed with zero mean and finite variance $\sigma^2_{\lambda_r}$. Furthermore, $\sup_{1\leq i\leq N,0\leq r\leq n} E\left|\lambda_{ri}\right|^{4+\delta}<\infty$ for some $\delta>0$. (ii) $\boldsymbol{\phi}_t$ is non-stochastic and has uniformly bounded elements, such that $\left\|\boldsymbol{\phi}_t\right\|\leq c<\infty\ \forall\ t$. (iii) The processes $\{\lambda_{ri}\}$, $\{\upsilon_{it}\}$ and $\{\eta_i\}$ are totally independent.

Assumption 4 is standard in factor analysis; see for example, Sarafidis, Yamagata and Robertson (2009) and Sarafidis and Wansbeek (2012). The zero mean assumption on the vector $\boldsymbol{\lambda}_i$ is not restrictive because the model can be expressed in terms of deviations from time-specific averages, which will eliminate the non-zero mean of $\boldsymbol{\lambda}_i$ (e.g. Sarafidis and Robertson, 2009). The vector of factors is treated as fixed and the factor loadings as random variables because the asymptotics apply for large N, T fixed. Observe that $\boldsymbol{\lambda}_i$ is correlated with the lagged dependent variable by construction and cannot be eliminated using the first-difference transformation because $\boldsymbol{\lambda}_i$ is multiplicative with $\boldsymbol{\phi}_t$, which is time-varying. One may think of the loadings in this context as reflecting different sources of unobserved heterogeneity, the impact of which is not constant through time. Rewriting the model in vector form yields

$$\widetilde{\mathbf{y}}_t=\alpha\widetilde{\mathbf{y}}_{t-1}+\widetilde{\mathbf{u}}_t,\ \widetilde{\mathbf{u}}_t=\boldsymbol{\eta}+\Lambda\boldsymbol{\phi}_t+P_N\boldsymbol{\upsilon}_t, \tag{47}$$

where $\Lambda=(\boldsymbol{\lambda}_1,...,\boldsymbol{\lambda}_N)'$ is a $N\times n$ matrix. Observe that $\widetilde{\mathbf{y}}_t$ can be written as

$$\widetilde{\mathbf{y}}_t=\alpha^t\widetilde{\mathbf{y}}_0+\left(\frac{1-\alpha^t}{1-\alpha}\right)\boldsymbol{\eta}+\Lambda\sum_{\tau=0}^{t-1}\alpha^{\tau}\boldsymbol{\phi}_{t-\tau}+P_N\sum_{\tau=0}^{t-1}\alpha^{\tau}\boldsymbol{\upsilon}_{t-\tau}, \tag{48}$$

and the initial observation is now given by $\widetilde{\mathbf{y}}_0=\delta_0\boldsymbol{\eta}+\delta_1P_N\boldsymbol{\upsilon}_0+\delta_2\Lambda\boldsymbol{\phi}_0$. Therefore, $\widetilde{\mathbf{y}}_t$ and $\widetilde{\mathbf{u}}_t$ can be expressed as linear forms of the innovations $\boldsymbol{\eta}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\upsilon}$:

$$\widetilde{\mathbf{y}}_t=\delta^{*}_{0,t}\boldsymbol{\eta}+\left(I_N\otimes\boldsymbol{\psi}_{1,t}\Phi\right)\boldsymbol{\lambda}+\left(\boldsymbol{\beta}_t\otimes P_N\right)\boldsymbol{\upsilon}, \tag{49}$$

and

$$\widetilde{\mathbf{u}}_t=\boldsymbol{\eta}+\left(I_N\otimes\boldsymbol{\phi}_t'\right)\boldsymbol{\lambda}+\left(\boldsymbol{e}_{t+1}'\otimes P_N\right)\boldsymbol{\upsilon}, \tag{50}$$

where $\boldsymbol{\psi}_{1,t}=\left(\delta_2\alpha^t,\alpha^{t-1},...,\alpha^0\right)$ is a $1\times(T+1)$ row vector, $\Phi=(\boldsymbol{\phi}_0,...,\boldsymbol{\phi}_T)'$ is a $(T+1)\times n$ matrix, $\boldsymbol{\lambda}=vec\left(\Lambda'\right)$ is a $nN\times 1$ row vector, while the remaining terms have been defined in Section 2. Stacking (47) over $t=2,...,T$ yields

$$\widetilde{\mathbf{y}}=\alpha\widetilde{\mathbf{y}}_{-1}+\widetilde{\mathbf{u}}, \tag{51}$$

where $\widetilde{\mathbf{y}} = (\widetilde{\mathbf{y}}_2', ..., \widetilde{\mathbf{y}}_T')'$, $\widetilde{\mathbf{y}}_{-1} = \left(\widetilde{\mathbf{y}}_1', ..., \widetilde{\mathbf{y}}_{T-1}'\right)'$, $\widetilde{\mathbf{u}} = (\widetilde{\mathbf{u}}_2', ..., \widetilde{\mathbf{u}}_T{}')'$.

Similarly, taking first-differences in (47) yields

$$\Delta\widetilde{\mathbf{y}}_t = \alpha\Delta\widetilde{\mathbf{y}}_{t-1} + \Lambda\Delta\boldsymbol{\phi}_t + P_N\Delta\boldsymbol{\upsilon}_t. \tag{52}$$

$\Delta\widetilde{\mathbf{y}}_t$ and $\Delta\widetilde{\mathbf{u}}_t$ can be expressed as linear forms of the innovations as follows:

$$\Delta\widetilde{\mathbf{y}}_t = \delta_{0,t}^{**}\boldsymbol{\eta} + \left(I_N \otimes \boldsymbol{\psi}_{2,t}\Phi\right)\boldsymbol{\lambda} + \left(\boldsymbol{\gamma}_t \otimes P_N\right)\boldsymbol{\upsilon}, \tag{53}$$

and

$$\Delta\widetilde{\mathbf{u}}_t = \left(I_N \otimes \Delta\boldsymbol{\phi}_t'\right)\boldsymbol{\lambda} + \left(\boldsymbol{d}_t \otimes P_N\right)\boldsymbol{\upsilon}, \tag{54}$$

where $\boldsymbol{\psi}_{2,t} = \left[\delta_2\left(\alpha-1\right)\alpha^{t-1}, \left(\alpha-1\right)\alpha^{t-2}, ..., \left(\alpha-1\right)\alpha^0, 1\right]$ is a $1\times(t+1)$ row vector. Stacking the column vectors in (52) over $t = 2, ..., T$ yields

$$\Delta\widetilde{\mathbf{y}} = \alpha\Delta\widetilde{\mathbf{y}}_{-1} + \Delta\widetilde{\mathbf{u}}. \tag{55}$$

As shown by Sarafidis and Robertson (2009) for the case where $\theta = 0$, the standard moment conditions that utilise instruments with respect to lagged values of y_{it-1} in the first-differenced equations and Δy_{it-1} in the levels equations are invalidated under a factor structure in the residuals. A similar result applies for $\theta \neq 0$ of course.[5] Therefore both $\widehat{\alpha}_D$ and $\widehat{\alpha}_S$ as defined in (16) and (17) respectively, are not consistent. However, as the following proposition demonstrates, the moment conditions that utilise instruments with respect to the individuals which unit i is spatially correlated with, remain valid. These moment conditions will be used to obtain consistent estimates of the structural parameter α.

Proposition 6 *Under Assumptions 1-4, the following $T(T-1)/2$ moment conditions are valid in the first-differenced model* (55)*:*

$$\widetilde{\mathbf{m}}_{N,\widetilde{D}}\left(\alpha\right) = N^{-1}\widetilde{Z}_D'\Delta\widetilde{\mathbf{u}} \xrightarrow{p} \mathbf{0}. \tag{56}$$

Furthermore, the following $T-1$ moment conditions are valid in the levels model (51)*:*

$$\widetilde{\mathbf{m}}_{N,\widetilde{L}}\left(\alpha\right) = N^{-1}\widetilde{Z}_L'\widetilde{\mathbf{u}} \xrightarrow{p} \mathbf{0}. \tag{57}$$

Proof. See Appendix A. ■

Observe that we have made no assumptions about mean-stationarity of the process. Indeed this assumption is always violated when there exists a common factor component because the mean of the process shifts every time period according to the value of $\boldsymbol{\phi}_t$. Therefore, one requires an estimator that does not rely on this assumption. Define

$$\widetilde{Z}_S \equiv \begin{bmatrix} \widetilde{Z}_D & 0 \\ 0 & \widetilde{Z}_L \end{bmatrix}, \tag{58}$$

and let $A_{3,D,N}$ and $A_{3,S,N}$ be sequences of possibly random, non-negative definite matrices of order $\zeta_{1,D}\times\zeta_{1,D}$, and $\zeta_{1,S}\times\zeta_{1,S}$, respectively. The following assumption is employed for the identification of α:

Assumption 3(iii) (identification of α): $N^{-1}\widetilde{Z}_D'\widetilde{Z}_D \xrightarrow{p} Q_{\widetilde{Z}_D}$, $N^{-1}\widetilde{Z}_D'\Delta\widetilde{\mathbf{y}}_{-1} \xrightarrow{p} \mathbf{q}_{\widetilde{Z}_D\Delta\widetilde{\mathbf{y}}_{-1}}$, $N^{-1}\widetilde{Z}_S'\widetilde{Z}_S \xrightarrow{p} Q_{\widetilde{Z}_S}$, , $N^{-1}\widetilde{Z}_S'\widetilde{\mathbf{y}}_{-1,S} \xrightarrow{p} \mathbf{q}_{\widetilde{Z}_S\widetilde{\mathbf{y}}_{-1,S}}$, all finite matrices (vectors) with full column rank (non-zero entries). $A_{3,D,N}$ and $A_{3,S,N}$ have full rank, such that $A_{3,D,N} \xrightarrow{p} A_{3,D}$, $A_{3,S,N} \xrightarrow{p} A_{3,S}$.

[5] For the case of a degenerate factor structure, which takes the form of a singe individual-invariant time effect, a similar result has been shown by Hsiao and Tahmiscioglu (2008).

Consider the following GMM estimators:

$$\widehat{\alpha}_{\widetilde{D}\left(A_{3,D,N}\right)} = \left[\Delta\widetilde{\mathbf{y}}_{-1}'\widetilde{Z}_D A_{3,D,N}\widetilde{Z}_D'\Delta\widetilde{\mathbf{y}}_{-1}\right]^{-1}\left[\Delta\widetilde{\mathbf{y}}_{-1}'\widetilde{Z}_D A_{3,D,N}\widetilde{Z}_D'\Delta\widetilde{\mathbf{y}}\right], \tag{59}$$

and

$$\widehat{\alpha}_{\widetilde{S}\left(A_{3,S,N}\right)} = \left[\widetilde{\mathbf{y}}_{-1,S}'\widetilde{Z}_S A_{3,S,N}\widetilde{Z}_L'\widetilde{\mathbf{y}}_{-1,S}\right]^{-1}\left[\widetilde{\mathbf{y}}_{-1,S}'\widetilde{Z}_S A_{3,S,N}\widetilde{Z}_S'\widetilde{\mathbf{y}}_S\right]. \tag{60}$$

The following theorem establishes the consistency and asymptotic normality of the above estimators under a factor structure and spatially correlated idiosyncratic components.

Theorem 7 *Suppose Assumptions 1-4, and* (56) *hold true. Let* $\Omega_{3,D,N} = var\left[\sqrt{N}\widetilde{\mathbf{m}}_{N,\widetilde{D}}(\alpha)\right]$ *be a sequence of symmetric non-negative definite matrices with rank greater than or equal to* $\zeta_{1,D}$*, such that* $\lambda_{\min}\left(\Omega_{3,D,N}\right) \geq c > 0$*, and* $\Omega_{3,D,N} \xrightarrow{p} \Omega_{3,D} = asy.var\left[\sqrt{N}\widetilde{\mathbf{m}}_{N,\widetilde{D}}(\alpha)\right]$*. The GMM estimator in* (59) *is consistent and*

$$\sqrt{N}\left(\widehat{\alpha}_{\widetilde{D}} - \alpha\right) \xrightarrow{d} N\left(0, \widetilde{V}_{\widetilde{D}}\right), \tag{61}$$

where

$$\widetilde{V}_{\widetilde{D}} = \left[\mathbf{q}_{\widetilde{Z}_D\Delta\widetilde{y}_{-1}}A_{3,D}\mathbf{q}_{\widetilde{Z}_D\Delta\widetilde{y}_{-1}}'\right]^{-1}\mathbf{q}_{\widetilde{Z}_D\Delta\widetilde{y}_{-1}}A_{3,D}\Omega_{3,D}A_{3,D}\mathbf{q}_{\widetilde{Z}_D\Delta\widetilde{y}_{-1}}'\left[\mathbf{q}_{\widetilde{Z}_D\Delta\widetilde{y}_{-1}}A_{3,D}\mathbf{q}_{\widetilde{Z}_D\Delta\widetilde{y}_{-1}}'\right]^{-1}. \tag{62}$$

In addition, suppose that (57) *holds true and let* $\Omega_{3,S,N} = var\left[\sqrt{N}\widetilde{\mathbf{m}}_{N,\widetilde{L}}(\alpha)\right]$ *be a sequence of symmetric non-negative definite matrices with rank greater than or equal to* $\zeta_{1,S}$*, such that* $\lambda_{\min}\left(\Omega_{3,S,N}\right) \geq c > 0$*, and* $\Omega_{3,S,N} \xrightarrow{p} \Omega_{3,S} = asy.var\left[\sqrt{N}\widetilde{\mathbf{m}}_{N,\widetilde{L}}(\alpha)\right]$*. The GMM estimator in* (60) *is consistent and*

$$\sqrt{N}\left(\widehat{\alpha}_{\widetilde{S}} - \alpha\right) \xrightarrow{d} N\left(0, \widetilde{V}_{\widetilde{S}}\right), \tag{63}$$

where

$$\widetilde{V}_{\widetilde{S}} = \left[\mathbf{q}_{\widetilde{Z}_S\widetilde{y}_{-1}}A_{3,S}\mathbf{q}_{\widetilde{Z}_S\widetilde{y}_{-1}}'\right]^{-1}\mathbf{q}_{\widetilde{Z}_S\widetilde{y}_{-1}}A_{3,S}\Omega_{3,S}A_{3,S}\mathbf{q}_{\widetilde{Z}_S\widetilde{y}_{-1}}'\left[\mathbf{q}_{\widetilde{Z}_S\widetilde{y}_{-1}}A_{3,S}\mathbf{q}_{\widetilde{Z}_S\widetilde{y}_{-1}}'\right]^{-1}.$$

Proof. *See Appendix A.* ■

The optimal GMM estimators are obtained by replacing $A_{3,D,N}$, $A_{3,S,N}$ by consistent estimates of $\Omega_{3,D,N}^{-1}$ and $\Omega_{3,S,N}^{-1}$, respectively. As in Section (2), $\Omega_{3,D,N}$ can be partitioned as

$$\Omega_{3,D,N} = \begin{bmatrix} \Omega_{3,22,D,N} & \cdots & \Omega_{3,2T,D,N} \\ & \ddots & \\ \Omega_{3,T2,D,N} & \cdots & \Omega_{3,TT,D,N} \end{bmatrix},$$

where $\Omega_{3,ts,D,N} = N^{-1}EY^{t-2\prime}\widetilde{W}_N\Delta\widetilde{\mathbf{u}}_t\Delta\widetilde{\mathbf{u}}_s'Y^{s-2}$. The pq^{th} element of $\Omega_{3,ts,D,N}$ is $\omega_{3,pq,ts,D,N} = N^{-1}E\widetilde{\mathbf{y}}_p'\widetilde{W}_N\Delta\widetilde{\mathbf{u}}_t\Delta\widetilde{\mathbf{u}}_s'\widetilde{W}_N\widetilde{\mathbf{y}}_q$. Therefore, from the proof of Proposition 6, equation (81), it is clear that to obtain a consistent estimate of $\Omega_{3,D,N}$, one requires estimating the number of unobserved factors and the factors themselves. For inference purposes, a consistent estimate of $\Omega_{3,D}$ is required even for a sub-optimal GMM estimator, as it is clear from (62). The same issue applies for $\Omega_{3,S,N}$. To avoid this complication, the standard errors for the sub-optimal GMM estimators can be computed using spatial block bootstrapping (e.g. Hall, 1985 and Anselin, 1990). We investigate this approach in simulations.

5 A Simulation Study

We will investigate the performance of the estimators analysed above in finite samples using simulated data. The main focus lies on examining the impact of the relative weight of the unobserved factor component in the total error process, as well as the effect of different values of N and α.

5.1 Design

The underlying generating process is given by

$$y_{it} = \alpha y_{it-1} + u_{it}, \, , \, u_{it} = \lambda_i \phi_t + \theta \sum_{j=1}^{N} w_{ij,N} v_{jt} + v_{it}, \; i = 1, ..., N, \; t = 1, ..., T, \tag{64}$$

where $w_{ij,N}$ is formed as in (41), $\lambda_i \sim i.i.d.U\left[-0.25, 0.25\right]$, $\phi_t \sim i.i.d.N\left(0,1\right)$ and $v_{it} \sim i.i.d.N\left(0,1\right)$.

The performance of the estimators will depend on the proportion of σ_u^2 attributed to the variance of the common factor component $-$ hereafter this proportion is denoted by ξ. Therefore, noticing that

$$\sigma_u^2 = \mu_\lambda^2 \sigma_\phi^2 + \sigma_\lambda^2 \sigma_\phi^2 + \sigma_v^2 \left(1 + \theta^2\right), \tag{65}$$

and normalising $\sigma_\phi^2 = 1$, we have

$$\sigma_v^2 = \frac{(1-\xi)}{\xi} \frac{\sigma_\lambda^2}{\left(1+\theta^2\right)} \tag{66}$$

We set $\theta = 0.5$, which implies that σ_v^2 will change only according to ξ since the value of σ_λ^2 is fixed in the design. As the value of ξ approaches unity, the impact of the factor component in the total error process increases. We choose the following values for ξ:

$$\begin{cases} \text{Low impact of factor structure on } u_{it}: & \xi = 1/3 \\ \text{Medium impact of factor structure on } u_{it}: & \xi = 1/2 \\ \text{Medium-to-high impact of factor structure on } u_{it}: & \xi = 2/3 \\ \text{High impact of factor structure on } u_{it}: & \xi = 3/4 \end{cases}$$

We set $T = 10$, and we experiment between $\alpha \in \{.2, .5, .8\}$ and $N \in \{100, 400, 800\}$. Notice that under a factor structure, mean-stationarity is always violated by construction because the mean shifts every time period according to the value of ϕ_t. To enhance transparency and save space, we simply set the initial value of the process as $y_{i0} = \lambda_i \phi_0 + \varepsilon_{i0}$. All experiments are based on 2,000 replications.

To compute empirical standard errors for the estimators, we use block bootstrapping. The algorithm can be outlined as follows: each individual i is assigned an equal probability of being selected with replacement. If unit i is selected, the complete time series of unit i is sampled to preserve the serial correlation structure of the data. The complete time series of unit i's neighbours, as reflected on W_N, are sampled as well. This process is repeated until the data set equals the original size of N. Once the sampling process is complete, estimates of the autoregressive parameter are obtained using the various methods employed in this simulation experiment. For the estimators that rely on spatial instruments, we make use of the spatial neighbours information, i.e. instruments are utilised with respect to the individuals, unit i is spatially correlated. The same procedure is followed over 200 bootstrapped samples, and the empirical standard errors of the estimators are computed in each replication.

5.1.1 Results

Table A1 in the appendix reports mean bias, root mean square error and empirical size (nominal size is 5%) for the estimators employed in this study. WG is the within-group estimator, FD (SYS) and FD† (SYS†) denote the first-differenced (system) GMM estimators that utilise Z_D (Z_S) and $\widetilde{Z}_D$ ($\widetilde{Z}_S$), respectively, as defined earlier in the paper.[6]

The performance of all estimators depends on the value of ξ, α and the size of N. Specifically, as the value of ξ increases for a given value of α and N, the performance of the estimators deteriorates in terms of bias and RMSE. This is illustrated in Figure 2 for $\alpha = 0.5$, $N = 400$.

[6] For FD† we set $A_{3,D,N} = N^{-1} \widetilde{Z}_D' \left(DD' \otimes I_N\right) \widetilde{Z}_D$, and similarly for FD, expect that $\widetilde{Z}_D$ is replaced by Z_D. For SYS† we set $A_{3,S,N} = N^{-1} \widetilde{Z}_S' diag \left[\left(D \otimes I_N\right)\left(D \otimes I_N\right)', \left(I_{T-1} \otimes I_N\right)\right] \widetilde{Z}_S$, while SYS makes use of Z_S instead.

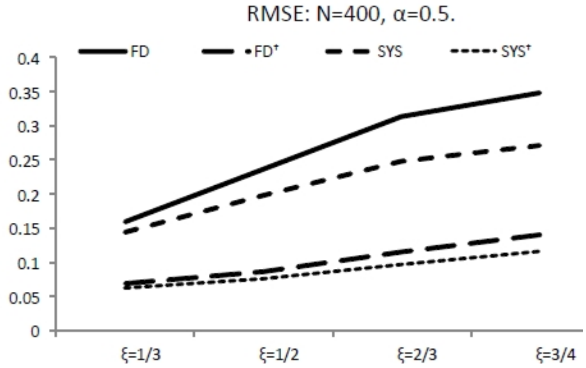

Figure 2

This is expected because higher values of ξ imply that the invalidity of the instruments used with respect to unit i itself (utilised by FD and SYS) is magnified; for the estimators that make use of the spatial instruments, the increase in bias and RMSE is also intuitive because as the value of ξ increases the contribution of the spatial component in the total error process, and thereby the correlation between the endogenous variable and the spatial instruments, decreases.

However, it is important to emphasise two points; firstly, both FD† and SYS† uniformly outperform FD and SYS, respectively, in terms of RMSE. The same holds for bias, unless N is small. Secondly, as the value of N increases, the bias and RMSE of FD† and SYS† decreases considerably, which is natural as these estimators are consistent. This is not the case for the conventional estimators, FD and SYS, the performance of which does not improve with larger values of N. This is illustrated in Figure 3 below for $\xi = 1/3$ and $\xi = 3/4$. Similar graphs apply for the remaining values of ξ, not illustrated here.

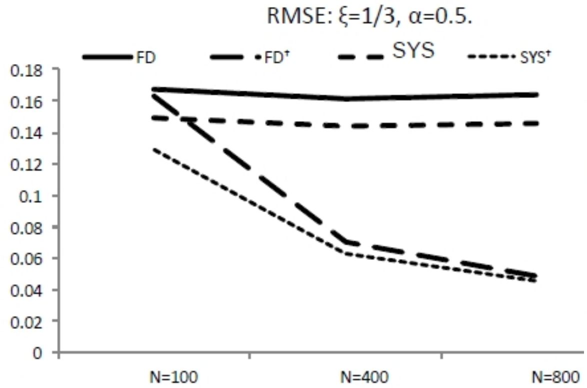

Figure 3a

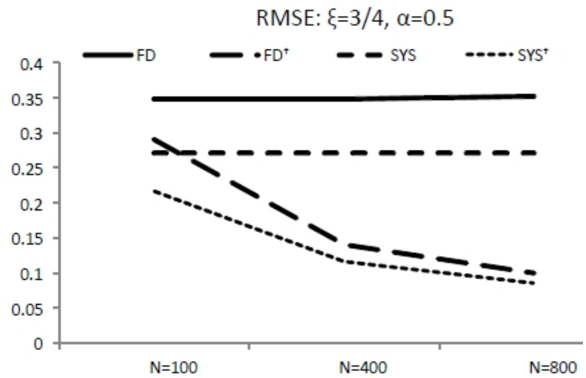

Figure 3b

It is also worth mentioning that SYS† appears to outperform FD† in terms of both bias and RMSE in all circumstances, with the relative difference in performance increasing according to the value of α.

In terms of empirical size, the results indicate that this is largely distorted for the conventional estimators, WG, FD and SYS, which is natural since these estimators exhibit large bias. The same applies to FD† and SYS† when N is small. However, as N increases, size improves considerably and appears to converge to the nominal level, especially for SYS†, which contrary to the conventional system GMM, it does not require mean-stationarity and therefore it is consistent under very mild assumptions on the initial conditions.

6 Concluding Remarks

Error cross-sectional dependence is an increasingly popular research area in the analysis of panel data. This papers considers spatial dependence and factor structure dependence in dynamic panel data models. It is shown that under spatially correlated errors, an additional set of moment conditions arises − in particular, instruments with respect to the individual(s) which unit i is correlated with. We demonstrate that these moment conditions remain valid when the errors contains a common factor component, in which case the standard instruments are invalidated. The resulting estimators are attractive because, aside from specifying a weighting matrix W, they are computationally simple and provide consistent estimates of the structural parameters without requiring estimation of the number of unobserved factors, or the factors themselves. Simulation experiments show that the proposed estimators largely outperform the conventional ones, in terms of both bias and root mean square error. This result is even more pronounced as N becomes larger.

References

[1] Ahn, S. C., Lee, Y.H., Schmidt, P., 2001. Panel data models with multiple time-varying individual effects, Mimeo.

[2] Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. Journal of Econometrics 68, 5-28.

[3] Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error components. Journal of the American Statistical Association 76, 598-606.

[4] Anselin, L., 1990. Some Robust Approaches to Testing and Estimation in Spatial Econometrics. Regional Science and Urban Economics 20, 141-163.

[5] Anselin, L., 2001. Spatial Econometrics. In: Baltagi, B.H. (Eds.), A Companion to Theoretical Econometrics. Blackwell Publishing, pp. 310-330.

[6] Arellano, M., 2003. Panel Data Econometrics. Oxford University Press, Oxford.

[7] Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. Review of Economic Studies 58, 277-297.

[8] Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-Component models. Journal of Econometrics 68, 29-51.

[9] Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

[10] Baltagi, B., Bresson, G., Pirotte, A., 2007. Panel unit root tests and spatial dependence. Journal of Applied Econometrics 22, 339-360.

[11] Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115-143.

[12] Bover, O., Watson, N., 2005. Are there economies of scale in the demand for money by firms? Some panel data estimates. Journal of Monetary Economics 52, 1569-1589.

[13] Breusch, T., Qian, H., Schmidt, P., Wyhowski, D., 1999. Redundancy of moment conditions. Journal of Econometrics 91, 89-111.

[14] Chudik, A., Pesaran, H., Tosetti, E., 2011. Weak and strong cross section dependence and estimation of large panels. The Econometrics Journal 14, c45-c90.

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

[16] Fingleton, B., 2008. A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices. Empirical Economics 34, 35-57.

[17] Goldberger, A.S., 1972. Structural equation methods in the social sciences. Econometrica 40, 979-1001.

[18] Hall, P., 1985. Resampling a coverage pattern. Stochastic Processes and their Applications 20, 231-246.

[19] Hsiao, C., Tahmiscioglu, A.K., 2008. Estimation of dynamic panel data models with both individual and time specific effects. Journal of Statistical Planning and Inference 138, 2698-2721.

[20] Jöreskog, K.G., Goldberger, A.S., 1975. Estimation of a model with multiple indicators and multiple causes of a single latent variable. Journal of the American Statistical Association 70, 631-639.

[21] Kapoor, M., Kelejian, H.H., Prucha I.R., 2007. Panel data models with spatially correlated error components. Journal of Econometrics 140, 97-130.

[22] Kelejian, H., Prucha I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53-67.

[23] Lee, L., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 137, 489-514.

[24] Lee, L. Yu, J., 2010. Efficient GMM estimation of spatial dynamic panel data models with fixed effects, Mimeo.

[25] Moon, H.R., Perron, B., 2004. Efficient estimation of the SUR cointegrating regression model and testing for purchasing power parity. Econometric Reviews 23, 293-323.

[26] Mutl, J., 2006. Dynamic panel data models with spatially correlated disturbances. PhD Thesis, University of Maryland, College Park.

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

[28] Pesaran, H., Tosetti, E., 2011. Large panels with common factors and spatial correlations. Journal of Econometrics 161, 182-202.

[29] Phillips, P., Sul, S., 2003. Dynamic panel estimation and homogeneity testing under cross-sectional dependence. The Econometrics Journal 6, 217-259.

[30] Robertson, D., Symons, J., 2007. Maximum likelihood factor analysis with rank deficient sample covariance matrices. Journal of Multivariate Analysis 98, 813-828.

[31] Sarafidis, V., Yamagata, T., 2010. Instrumental variable estimation of dynamic linear panel data models with defactored regressors under cross-sectional dependence. Mimeo.

[32] Sarafidis, V., Yamagata, T., Robertson, D., 2009. A test of error cross-sectional dependence for a linear dynamic panel model with regressors. Journal of Econometrics 148, 149-161.

[33] Sarafidis, V., Robertson, D., 2009. On the impact of cross-sectional dependence in short dynamic panel estimation. The Econometrics Journal 12, 62-81.

[34] Sarafidis, V., Wansbeek, T., 2012. Cross-sectional dependence in panel data analysis. Econometric Reviews, forthcoming.

Appendix A

The $(t-1)^{th}$ block of $N^{-1}\widetilde{Z}_D'\Delta\boldsymbol{\varepsilon}$, for $t=2,...,T$, where $\widetilde{Z}_D = diag\left(\widetilde{W}_N Y^0, \widetilde{W}_N Y^1, ..., \widetilde{W}_N Y^{T-2}\right)$ and $\widetilde{W}_N = W_N + W_N'$, with $\widetilde{W}_N' = (W_N + W_N')' = W_N' + W_N = \widetilde{W}_N$, is given by

$$\begin{aligned}
N^{-1}\left(\widetilde{W}_N Y^{t-2}\right)'\Delta\boldsymbol{\varepsilon}_t &= N^{-1}\begin{bmatrix} \mathbf{y}_0'\widetilde{W}_N\Delta\boldsymbol{\varepsilon}_t \\ \mathbf{y}_1'\widetilde{W}_N\Delta\boldsymbol{\varepsilon}_t \\ \vdots \\ \mathbf{y}_{t-2}'\widetilde{W}_N\Delta\boldsymbol{\varepsilon}_t \end{bmatrix} \\
&= N^{-1}\begin{bmatrix} \boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_0'\boldsymbol{d}_t \otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon} + \delta_{0,s}^{*}\boldsymbol{\eta}'\left(\boldsymbol{d}_t \otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon} \\ \boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_1'\boldsymbol{d}_t \otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon} + \delta_{0,s}^{*}\boldsymbol{\eta}'\left(\boldsymbol{d}_t \otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon} \\ \vdots \\ \boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_{t-2}'\boldsymbol{d}_t \otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon} + \delta_{0,s}^{*}\boldsymbol{\eta}'\left(\boldsymbol{d}_t \otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon} \end{bmatrix},
\end{aligned}$$

since, using (4) and (5), we have

$$\begin{aligned}
\mathbf{y}_s'\widetilde{W}_N\Delta\boldsymbol{\varepsilon}_t &= \left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s' \otimes P_N'\right) + \delta_{0,s}^{*}\boldsymbol{\eta}'\right]\widetilde{W}_N\left[\left(\boldsymbol{d}_t \otimes P_N\right)\boldsymbol{\upsilon}\right] \\
&= \boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s' \otimes P_N'\right)\widetilde{W}_N\left(\boldsymbol{d}_t \otimes P_N\right) + \delta_{0,s}^{*}\boldsymbol{\upsilon} + \delta_{0,s}^{*}\boldsymbol{\eta}'\widetilde{W}_N\left(\boldsymbol{d}_t \otimes P_N\right)\boldsymbol{\upsilon} \\
&= \boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s' \otimes P_N'\right)\left(1 \otimes \widetilde{W}_N\right)\left(\boldsymbol{d}_t \otimes P_N\right)\boldsymbol{\upsilon} + \delta_{0,s}^{*}\boldsymbol{\eta}'\left(1 \otimes \widetilde{W}_N\right)\left(\boldsymbol{d}_t \otimes P_N\right)\boldsymbol{\upsilon} \\
&= \boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t \otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon} + \delta_{0,s}^{*}\boldsymbol{\eta}'\left(\boldsymbol{d}_t \otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}. \qquad (67)
\end{aligned}$$

The $(t-1)^{th}$ block of $N^{-1}Z_D'\Delta\boldsymbol{\varepsilon}$ is identical except that $\widetilde{W}_N$ is replaced by I_N, the $N \times N$ identity matrix.

Similarly, using (6) and (8), the $(t-1)^{th}$ block of $\widetilde{Z}_l'\boldsymbol{u}$ can be written as

$$\begin{aligned}
\Delta\mathbf{y}_{t-1}'\widetilde{W}_N\mathbf{u}_t &= \left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}' \otimes P_N'\right) + \delta_{0,t-1}^{**}\boldsymbol{\eta}'\right]\widetilde{W}_N\left[\left(\boldsymbol{e}_{t+1}' \otimes P_N\right)\boldsymbol{\upsilon} + \boldsymbol{\eta}\right] \\
&= \boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}' \otimes P_N'\right)\widetilde{W}_N\left(\boldsymbol{e}_{t+1}' \otimes P_N\right)\boldsymbol{\upsilon} + \boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}' \otimes P_N'\right)\widetilde{W}_N\boldsymbol{\eta} \\
&\quad + \delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\left(\boldsymbol{e}_{t+1}' \otimes P_N\right)\boldsymbol{\upsilon} + \delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta} \\
&= \boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}' \otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon} + \boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}' \otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta} \\
&\quad + \delta_{0,t-1}^{**}\boldsymbol{\eta}'\left(\boldsymbol{e}_{t+1}' \otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon} + \delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}. \qquad (68)
\end{aligned}$$

The $(t-1)^{th}$ block of $Z_L'\boldsymbol{u}$ is identical except that $\widetilde{W}_N$ is replaced by I_N.

Furthermore, the $(t-1)^{th}$ block of $N^{-1}\widetilde{Z}_D'\Delta\boldsymbol{y}_{-1}$, for $t=2,...,T$, is

$$N^{-1}\left(\widetilde{W}_N Y^{t-2}\right)'\Delta\boldsymbol{y}_{-1} = N^{-1}\begin{bmatrix} \mathbf{y}_0'\widetilde{W}_N\Delta\boldsymbol{y}_{t-1} \\ \mathbf{y}_1'\widetilde{W}_N\Delta\boldsymbol{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-2}'\widetilde{W}_N\Delta\boldsymbol{y}_{t-1} \end{bmatrix},$$

where

$$\begin{aligned}
\mathbf{y}_s'\widetilde{W}_N\Delta\mathbf{y}_{t-1} &= \left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s' \otimes P_N'\right) + \delta_{0,s}^{*}\boldsymbol{\eta}'\right]\widetilde{W}_N\left[\left(\boldsymbol{\gamma}_{t-1} \otimes P_N\right)\boldsymbol{\upsilon} + \delta_{0,t-1}^{**}\boldsymbol{\eta}\right] \\
&= \boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1} \otimes P_n'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon} + \delta_{0,t-1}^{**}\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s' \otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta} \\
&\quad + \delta_{0,s}^{*}\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1} \otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon} + \delta_{0,s}^{*}\delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}, \qquad (69)
\end{aligned}$$

while using (4) and (8), the $(t-1)^{th}$ block of $N^{-1}\widetilde{Z}_L'\mathbf{y}_{-1}$ can be written as

$$\begin{aligned} N^{-1}\Delta\mathbf{y}_{t-1}'\widetilde{W}_N\mathbf{y}_{t-1} &= \left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\right)+\delta_{0,t-1}^{**}\boldsymbol{\eta}'\right]\widetilde{W}_N\left[\left(\boldsymbol{\beta}_{t-1}\otimes P_N\right)\boldsymbol{\upsilon}+\delta_{0,t-1}^{*}\boldsymbol{\eta}\right] \\ &= \boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{\beta}_{t-1}\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{\upsilon}+\delta_{0,t-1}^{**}\boldsymbol{\eta}'\left(\boldsymbol{\beta}_{t-1}\otimes\widetilde{W}_NP_N\right)\boldsymbol{\upsilon} \\ &\quad +\delta_{0,t-1}^{*}\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}+\delta_{0,t-1}^{*}\delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}. \qquad (70) \end{aligned}$$

We define the following terms:

$$\begin{aligned} \psi_{1,st,N} &\equiv N^{-1}\mathbf{y}_s'\Delta\boldsymbol{\varepsilon}_t; \\ \psi_{2,t,N} &\equiv N^{-1}\Delta\mathbf{y}_{t-1}'\mathbf{u}_t; \\ \psi_{3,st,N} &\equiv N^{-1}\mathbf{y}_s'\widetilde{W}_N\Delta\boldsymbol{\varepsilon}_t; \\ \psi_{4,t,N} &\equiv N^{-1}\Delta\mathbf{y}_{t-1}'\widetilde{W}_N\mathbf{u}_t; \\ \psi_{5,st,N} &\equiv N^{-1}\mathbf{y}_s'\widetilde{W}_N\Delta\mathbf{y}_{t-1}; \\ \psi_{6,t,N} &\equiv N^{-1}\Delta\mathbf{y}_{t-1}'\widetilde{W}_N\mathbf{y}_{t-1}. \end{aligned}$$

PROOF OF PROPOSITION 1

We need to show that (i) $E\psi_{1,st,N}=0$, $E\psi_{1,st,N}^2\to 0$ as $N\to\infty$ for $t=2,...,T$, $s\le t-2$, and (ii) $E\psi_{2,t,N}=0$, $E\psi_{2,t,N}^2\to 0$ as $N\to\infty$ for $t=2,...,T$. This is entirely straightforward from the proof in Proposition 3 by replacing $\widetilde{W}_N$ by I_N and using the mean-stationarity assumption for $\psi_{2,t,N}$, which implies that $\delta_{0,t-1}^{**}=0$ $\forall t$. The claims in Proposition 1 then follow from Chebychev's inequality. QED

PROOF OF PROPOSITION 3

Firstly, we will show that (i) $E\psi_{3,st,N}=0$, $E\psi_{3,st,N}^2\to 0$ as $N\to\infty$ for $t=2,...,T$, $s\le t-2$, and (ii) $E\psi_{4,t,N}=0$, $E\psi_{4,t,N}^2\to 0$ as $N\to\infty$ for $t=2,...,T$.

We have

$$\begin{aligned} E\psi_{3,st,N} &= N^{-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{\upsilon}+\delta_{0,s}^{*}\boldsymbol{\eta}'\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)\boldsymbol{\upsilon}\right] \\ &= N^{-1}tr\left[\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_NP_N\right)E\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\right]+\delta_{0,s}^{*}tr\left[\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)E\boldsymbol{\upsilon}\boldsymbol{\eta}'\right] \\ &= N^{-1}tr\left[\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_NP_N\right)\Sigma_\upsilon\right], \qquad (71) \end{aligned}$$

since $E\boldsymbol{\upsilon}\boldsymbol{\eta}'=\mathbf{0}$ under the maintained assumptions. Observe that $\boldsymbol{\beta}_s'\boldsymbol{d}_t$ is a $(T+1)\times(T+1)$ matrix that contains zeros on the main diagonal $s\le t-2$. Therefore, $\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_n'\widetilde{W}_NP_n\right)$ is a $N(T+1)\times N(T+1)$ matrix with zeros on the main diagonal and by Lemma 9 it has uniformly bounded row and column sums, setting $\boldsymbol{\beta}_s'\boldsymbol{d}_t=H$ and $P_N'\widetilde{W}_NP_N=C_{1,N}$. In addition, Σ_υ is a $N(T+1)\times N(T+1)$ diagonal matrix under the maintained assumptions with uniformly bounded elements. Hence, $tr\left[\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'P_N\right)\Sigma_\upsilon\right]=0$ by Lemma 10(i), setting $\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_NP_N=C_{1,\ell N}^0$ and $\Sigma_\upsilon=D_{1,\ell N}$, with $\ell=T+1$. Therefore, $E\psi_{3,st,N}=0$.

The variance of $\psi_{3,st,N}$ equals

$$\begin{aligned} E\psi_{3,st,N}^2 &= N^{-2}E\{\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{\upsilon}+\delta_{0,s}^{*}\boldsymbol{\eta}'\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)\right] \\ &\quad\times\left[\boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\boldsymbol{\beta}_s\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{\upsilon}+\delta_{0,s}^{*}\boldsymbol{\upsilon}'\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)\boldsymbol{\eta}\right]\} \\ &= N^{-2}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\mathbf{d}_t'\boldsymbol{\beta}_s\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{\upsilon}\right] \\ &\quad +N^{-2}\delta_{0,s}^{*2}E\left[\boldsymbol{\eta}'\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right] \\ &\quad +N^{-2}2\delta_{0,s}^{*}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right]. \qquad (72) \end{aligned}$$

The first term on the right-hand side of the last equality above equals

$$N^{-2}E\left[\boldsymbol{v}'\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}\boldsymbol{v}'\left(\mathbf{d}_s'\boldsymbol{\beta}_t\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}\right]=N^{-2}2tr\left(C_{1,\ell N}^0\Sigma_vC_{1,\ell N}^{0\prime}\Sigma_v\right),$$

with $\ell=T+1$. This follows from Lemma 11(i) and the fact that $c_{ii,1,\ell N}^0=0$ for $s\leq t-2$. Given this property and since Σ_v is diagonal, it follows from Lemma 10(iii) that $N^{-2}2tr\left(C_{1,\ell N}^0\Sigma_vC_{1,\ell N}^{0\prime}\Sigma_v\right)=o(1)$. The second term is

$$\begin{aligned}
& N^{-2}E\left[\boldsymbol{\eta}'\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)\boldsymbol{v}\boldsymbol{v}'\left(\boldsymbol{d}_t'\otimes\widetilde{W}_NP_N'\right)\boldsymbol{\eta}\right]\\
={}& N^{-2}trE\left[\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)\boldsymbol{v}\boldsymbol{v}'\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\boldsymbol{\eta}'\right]\\
={}& N^{-2}tr\left[\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)\Sigma_v\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\right)\Sigma_\eta\right]\\
={}& N^{-2}tr\left[\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)\Sigma_v\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\right)(1\otimes\Sigma_\eta)\right]\\
={}& N^{-2}tr\left[\left(\boldsymbol{d}_t\otimes P_N\widetilde{W}_N\right)\Sigma_v\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\Sigma_\eta\right)\right]. \qquad (73)
\end{aligned}$$

By Lemma 9 the row and column sums of $\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)$ and $\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\Sigma_\eta\right)$ are uniformly bounded. Furthermore, Σ_v is a diagonal matrix with uniformly bounded diagonal entries. As a result, $N^{-2}tr\left[\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)\Sigma_v\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\Sigma_\eta\right)\right]=o(1)$ by Lemma 10(ii), setting $\left(\boldsymbol{d}_t\otimes\widetilde{W}_NP_N\right)=C_{2\ell N}^0$, $\Sigma_v=D_{1,\ell N}$ and $\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\Sigma_\eta\right)=C_{3\ell N}^0$ The third term equals zero by Lemma 11(iv). It follows from Chebychev's inequality that $N^{-1}\mathbf{y}_s'\widetilde{W}_N\Delta\boldsymbol{\varepsilon}_t\xrightarrow{p}0$.

For $\psi_{4,st,N}$ we have, using (6) and (8),

$$\begin{aligned}
E\psi_{4,t,N} ={}& N^{-1}E\left[\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}'\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}\right]+N^{-1}E\left[\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right]\\
&+\delta_{0,t-1}^{**}N^{-1}E\left[\boldsymbol{\eta}'\left(\boldsymbol{e}_{t+1}'\otimes\widetilde{W}_NP_N\right)\boldsymbol{v}\right]+\delta_{0,t-1}^{**}N^{-1}E\left[\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right]\\
={}& N^{-1}tr\left[\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}'\otimes P_N'\widetilde{W}_NP_N\right)E\boldsymbol{v}\boldsymbol{v}'\right]+N^{-1}tr\left[\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)E\boldsymbol{\eta}\boldsymbol{v}'\right]\\
&+\delta_{0,t-1}^{**}N^{-1}tr\left[\left(\boldsymbol{e}_{t+1}'\otimes\widetilde{W}_NP_N\right)E\boldsymbol{v}\boldsymbol{\eta}'\right]+\delta_{0,t-1}^{**}N^{-1}tr\left[\widetilde{W}_NE\boldsymbol{\eta}\boldsymbol{\eta}'\right]. \qquad (74)
\end{aligned}$$

Under the maintained assumptions $E\boldsymbol{\eta}\boldsymbol{v}'=0$. In addition, both $\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}'\otimes P_N'\widetilde{W}_NP_N\right)$ and $\widetilde{W}_N$ have uniformly bounded row and column sums and contain zeros on the main diagonal. Therefore, by Lemma 10(i) $E\psi_{4,t,N}=0$. Notice that the expression for $N^{-1}\Delta\mathbf{y}_{t-1}'\mathbf{u}_t$ is obtained by replacing $\widetilde{W}_N$ by the identity matrix. In this case the last term is zero only if $\delta_{0,t-1}^{**}=0$, which is satisfied under mean-stationarity of the process.

The variance of $\psi_{4,st,N}$ equals

$$\begin{aligned}
E\psi_{4,t,N}^2 ={}& N^{-2}E\{[\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}'\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}+\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\\
&+\delta_{0,t-1}^{**}\boldsymbol{\eta}'\left(\boldsymbol{e}_{t+1}'\otimes\widetilde{W}_NP_N\right)\boldsymbol{v}+\delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}]\\
&\times[\boldsymbol{v}'\left(\boldsymbol{e}_{t+1}\boldsymbol{\gamma}_{t-1}\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}+\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1}\otimes\widetilde{W}_NP_N\right)\boldsymbol{v}\\
&+\delta_{0,t-1}^{**}\boldsymbol{v}'\left(\boldsymbol{e}_{t+1}\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}+\delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}]\}=
\end{aligned}$$

$$
\begin{aligned}
&= N^{-2}E\left[\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}'\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}\boldsymbol{e}_{t+1}\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}\right]\\
&\quad +N^{-2}2E\left[\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}'\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1}\otimes \widetilde{W}_NP_N\right)\boldsymbol{v}\right]\\
&\quad +N^{-2}2\delta_{0,t-1}^{**}E\left[\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}'\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}\boldsymbol{v}'\left(\boldsymbol{e}_{t+1}\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right]\\
&\quad +N^{-2}2\delta_{0,t-1}^{**}E\left[\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}'\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right]\\
&\quad +N^{-2}E\left[\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1}\otimes \widetilde{W}_NP_N\right)\boldsymbol{v}\right]\\
&\quad +N^{-2}2\delta_{0,t-1}^{**}E\left[\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\boldsymbol{v}'\left(\boldsymbol{e}_{t+1}\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right]\\
&\quad +N^{-2}2\delta_{0,t-1}^{**}E\left[\boldsymbol{v}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right]\\
&\quad +N^{-2}\delta_{0,t-1}^{**2}E\left[\boldsymbol{\eta}'\left(\boldsymbol{e}_{t+1}'\otimes \widetilde{W}_NP_N\right)\boldsymbol{v}\boldsymbol{v}'\left(\boldsymbol{e}_{t+1}\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right]\\
&\quad +N^{-2}2\delta_{0,t-1}^{**2}E\left[\boldsymbol{\eta}'\left(\boldsymbol{e}_{t+1}'\otimes \widetilde{W}_NP_N\right)\boldsymbol{v}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right]\\
&\quad +N^{-2}\delta_{0,t-1}^{**2}E\left[\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right]\\
&= N^{-2}2tr\left[C_{4,\ell N}^0\Sigma_vC_{4,\ell N}^{0\prime}\Sigma_v\right]+N^{-2}2\delta_{0,t-1}^{**}tr\left[C_{4,\ell N}^0\Sigma_\eta\widetilde{W}_N\Sigma_v\right]\\
&\quad +N^{-2}tr\left[G_{1,\ell N}\Sigma_\eta G_{1,\ell N}'\Sigma_v\right]+N^{-2}2\delta_{0,t-1}^{**}tr\left[G_{1,n}\Sigma_\eta C_{5,\ell N}^0\Sigma_v\right]\\
&\quad +N^{-2}2\delta_{0,t-1}^{**2}tr\left[C_{5,\ell N}^{0\prime}\Sigma_\eta C_{5,\ell N}^0\Sigma_v\right]+N^{-2}2\delta_{0,t-1}^{**2}tr\left[\widetilde{W}_N\Sigma_\eta\widetilde{W}_N\Sigma_\eta\right],
\end{aligned}
$$

where $C_{4,\ell N}^0=\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}'\otimes P_N'\widetilde{W}_NP_N\right)$, $G_{1,n}=\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)$, $C_{5,\ell N}^0=\left(\boldsymbol{e}_{t+1}\otimes P_N'\widetilde{W}_N\right)$, using Lemma 11 repeatedly. Finally, by Lemma 10(iii)-(iv) it follows immediately that $E\psi_{4,t,N}^2=o(1)$ and so $N^{-1}\boldsymbol{\Delta}\mathbf{y}_{t-1}'\widetilde{W}_N\mathbf{u}_t\xrightarrow{p}0$ by Chebychev's inequality.

Next, we will show that (i) $E\psi_{5,st,N}=q_{st,d}\neq 0$, in general, $E\psi_{5,st,N}^2\to 0$ as $N\to\infty$ for $t=2,...,T$, $s\leq t-2$, and (ii) $E\psi_{6,t,N}=q_{t,s}\neq 0$, in general, $E\psi_{6,t,N}^2\to 0$ as $N\to\infty$ for $t=2,...,T$.

The expected value of $E\psi_{5,st,N}$, using (4) and (8), is given by

$$
\begin{aligned}
E\psi_{5,st,N} &= N^{-1}E\left[\boldsymbol{v}'\left(\boldsymbol{\beta}_s'\otimes P_N'\right)+\delta_{0,t}^*\boldsymbol{\eta}'\right]\widetilde{W}_N\left[\left(\boldsymbol{\gamma}_{t-1}\otimes P_N\right)\boldsymbol{v}+\delta_{0,t-1}^{**}\boldsymbol{\eta}\right]\\
&= N^{-1}E\left[\boldsymbol{v}'\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1}\otimes P_N'\widetilde{W}_NP_N\right)\boldsymbol{v}\right]+\delta_{0,s}^*\delta_{0,t-1}^{**}N^{-1}E\left[\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right]\\
&\quad +\delta_{0,t-1}^{**}N^{-1}E\left[\boldsymbol{v}'\left(\boldsymbol{\beta}_s'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right]+\delta_{0,s}^*N^{-1}E\left[\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1}\otimes \widetilde{W}_NP_N\right)\boldsymbol{v}\right]\\
&= N^{-1}tr\left[\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1}\otimes P_N'\widetilde{W}_NP_N\right)E\boldsymbol{v}\boldsymbol{v}'\right]+\delta_{0,s}^*\delta_{0,t-1}^{**}N^{-1}tr\left[\widetilde{W}_NE\boldsymbol{\eta}\boldsymbol{\eta}'\right]\\
&\quad +N^{-1}\delta_{0,t-1}^{**}tr\left[\left(\boldsymbol{\beta}_s'\otimes P_N'\widetilde{W}_N\right)E\boldsymbol{\eta}\boldsymbol{v}'\right]+N^{-1}\delta_{0,s}^*tr\left[\left(\boldsymbol{\gamma}_{t-1}\otimes \widetilde{W}_NP_N\right)E\boldsymbol{v}\boldsymbol{\eta}'\right]\\
&= N^{-1}tr\left[\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1}\otimes P_N'\widetilde{W}_NP_N\right)\Sigma_v\right]=q_{st,D},
\end{aligned}
\tag{75}
$$

since $E\boldsymbol{v}\boldsymbol{\eta}'=E\boldsymbol{\eta}\boldsymbol{v}'=\mathbf{0}$ under the maintained assumptions. We have also used the fact that $tr\left[\widetilde{W}_NE\boldsymbol{\eta}\boldsymbol{\eta}'\right]=tr\left[\widetilde{W}_N\Sigma_\eta\right]=0$ from Lemma 10(i), given that $\widetilde{W}_N$ contains zeros on the main diagonal and Σ_η is diagonal. Observe that $q_{st,D}\neq 0$ in general, unless $\theta=0$, in which case $P_N=I_N\Rightarrow P_N'\widetilde{W}_NP_N=\widetilde{W}_N$. As a result, the kronecker product matrix contains zeros on the main diagonal and so $q_{st,D}=0$. Therefore, the spatial instruments are not correlated with the endogenous regressor − an intuitive result.

Furthermore, we have

$$
\begin{aligned}
E\psi_{5,st,N}^{2} &= N^{-2}E\{[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1}\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}+\delta_{0,t-1}^{**}\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta} \\
&\quad +\delta_{0,s}^{*}\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1}\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}+\delta_{0,s}^{*}\delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}] \\
&\quad \times[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{\beta}_s\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}+\delta_{0,t-1}^{**}\boldsymbol{\eta}'\left(\boldsymbol{\beta}_s\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon} \\
&\quad +\delta_{0,s}^{*}\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}+\delta_{0,s}^{*}\delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}]\} \\
&= N^{-2}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1}\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{\beta}_s\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\right] \\
&\quad +N^{-2}2\delta_{0,t-1}^{**}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1}\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\eta}'\left(\boldsymbol{\beta}_s\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\right] \\
&\quad +N^{-2}2\delta_{0,s}^{*}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1}\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right] \\
&\quad +N^{-2}2\delta_{0,s}^{*}\delta_{0,t-1}^{**}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1}\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right] \\
&\quad +N^{-2}\delta_{0,t-1}^{**2}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\boldsymbol{\eta}'\left(\boldsymbol{\beta}_s\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\right] \\
&\quad +N^{-2}2\delta_{0,s}^{*}\delta_{0,t-1}^{**}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right] \\
&\quad +N^{-2}2\delta_{0,s}^{*}\delta_{0,t-1}^{**2}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right] \\
&\quad +N^{-2}\delta_{0,s}^{*2}E\left[\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1}\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right] \\
&\quad +N^{-2}2\delta_{0,s}^{*2}\delta_{0,t-1}^{**}E\left[\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1}\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right] \\
&\quad +N^{-2}\delta_{0,s}^{*2}\delta_{0,t-1}^{**2}E\left[\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right] \\
&= N^{-2}\left[2tr\left[C_{1,\ell N}\Sigma_\upsilon C_{1,\ell N}'\Sigma_\upsilon\right]+\sum_{i=1}^{\ell N}\alpha_{ii,k,\ell N}^{2}\left(\mu_{\upsilon_i}^{4}-3\sigma_{\upsilon_i}^{4}\right)\right] \\
&\quad +N^{-2}2\delta_{0,s}^{*}\delta_{0,t-1}^{**}tr\left[C_{1,\ell N}\Sigma_\upsilon\right]tr\left[\widetilde{W}_N\Sigma_\eta\right] \\
&\quad +N^{-2}\delta_{0,t-1}^{**2}tr\left[G_{2,n}\Sigma_\eta G_{2,n}\Sigma_\upsilon\right]+tr\left[G_{2,n}'\Sigma_{\xi_1}G_{3,n}\Sigma_{\xi_2}\right] \\
&\quad +N^{-2}\delta_{0,s}^{*2}tr\left[G_{3,n}\Sigma_\eta G_{3,n}'\Sigma_\upsilon\right]+N^{-2}\delta_{0,s}^{*2}\delta_{0,t-1}^{**2}tr\left[\widetilde{W}_N\Sigma_\eta\widetilde{W}_N\Sigma_\eta\right],
\end{aligned}
$$

where $C_{1,\ell N}=\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1}\otimes P_N'\widetilde{W}_N P_N\right)$, $G_{2,n}=\left(\boldsymbol{\beta}_s'\otimes P_N'\widetilde{W}_N\right)$, $G_{3,n}=\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)$, using Lemma 11 repeatedly. Using, by Lemma 10(iii)-(iv) it is straightforward to show that $E\psi_{5,t,N}^{2}=o(1)$.

Finally, for $\psi_{6,t,N}$ we have

$$
\begin{aligned}
E\psi_{6,t,N} &= N^{-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\right)+\delta_{0,t-1}^{*}\boldsymbol{\eta}'\right]\widetilde{W}_N\left[\left(\boldsymbol{\beta}_{t-1}\otimes P_N\right)\boldsymbol{\upsilon}+\delta_{0,t-1}^{**}\boldsymbol{\eta}\right] \\
&= N^{-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{\beta}_{t-1}\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\right]+\delta_{0,t-1}^{*}E\left[\boldsymbol{\upsilon}^{0}\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right] \\
&\quad +N^{-1}\delta_{0,t-1}^{**}E\left[\boldsymbol{\eta}'\left(\boldsymbol{\beta}_{t-1}\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\right]+N^{-1}\delta_{0,t-1}^{*}\delta_{0,t-1}^{**}E\left[\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right] \\
&= N^{-1}tr\left[\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{\beta}_{t-1}\otimes P_N'\widetilde{W}_N P_N\right)E\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\right]+\delta_{0,t-1}^{*}\delta_{0,t-1}^{**}N^{-1}tr\left[\widetilde{W}E\boldsymbol{\eta}\boldsymbol{\eta}'\right] \\
&\quad +N^{-1}\delta_{0,t-1}^{*}tr\left[\left(\boldsymbol{\gamma}_{t-1}'\otimes P_N'\widetilde{W}_N\right)E\boldsymbol{\eta}\boldsymbol{\upsilon}'\right]+N^{-1}\delta_{0,t-1}^{**}tr\left[\left(\boldsymbol{\beta}_{t-1}\otimes \widetilde{W}_N P_N\right)E\boldsymbol{\upsilon}\boldsymbol{\eta}'\right] \\
&= N^{-1}tr\left[\left(\boldsymbol{\beta}_s'\boldsymbol{\gamma}_{t-1}\otimes P_n'\widetilde{W}_N P_N\right)\Sigma_\upsilon\right]=q_{t,L}, \qquad (76)
\end{aligned}
$$

since $E\boldsymbol{\upsilon}\boldsymbol{\eta}'=E\boldsymbol{\eta}\boldsymbol{\upsilon}'=\mathbf{0}$ under the maintained assumptions, and recalling that $tr\left[\widetilde{W}_N E\boldsymbol{\eta}\boldsymbol{\eta}'\right]=tr\left[\widetilde{W}_N\Sigma_\eta\right]=0$. Notice, as before, $q_{t,L}=0$ for $\theta=0$.

The variance of $\psi_{6,t,N}$ equals

$$
\begin{aligned}
E\psi^2_{6,st,N} &= N^{-2}E\{[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}'_{t-1}\boldsymbol{\beta}_{t-1}\otimes P'_N\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}+\delta^*_{0,t-1}\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}'_{t-1}\otimes P'_N\widetilde{W}_N\right)\boldsymbol{\eta} \\
&\quad +\delta^{**}_{0,t-1}\boldsymbol{\eta}'\left(\boldsymbol{\beta}_{t-1}\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}+\delta^*_{0,t-1}\delta^{**}_{0,t-1}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}] \\
&\quad \times[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}'_{t-1}\boldsymbol{\gamma}_{t-1}\otimes P'_N\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}+\delta^*_{0,t-1}\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1}\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon} \\
&\quad +\delta^{**}_{0,t-1}\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}'_{t-1}\otimes P'_N\widetilde{W}_N\right)\boldsymbol{\eta}+\delta^*_{0,t-1}\delta^{**}_{0,t-1}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}]\} \\
&= N^{-2}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}'_{t-1}\boldsymbol{\beta}_{t-1}\otimes P'_N\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}'_{t-1}\boldsymbol{\gamma}_{t-1}\otimes P'_N\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\right] \\
&\quad +N^{-2}2\delta^*_{0,t-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}'_{t-1}\boldsymbol{\beta}_{t-1}\otimes P'_N\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1}\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\right]- \\
&\quad +N^{-2}2\delta^{**}_{0,t-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}'_{t-1}\boldsymbol{\beta}_{t-1}\otimes P'_N\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}'_{t-1}\otimes P'_N\widetilde{W}_N\right)\boldsymbol{\eta}\right]- \\
&\quad +N^{-2}2\delta^*_{0,t-1}\delta^{**}_{0,t-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}'_{t-1}\boldsymbol{\beta}_{t-1}\otimes P'_N\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right] \\
&\quad +N^{-2}\delta^{*2}_{0,t-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}'_{t-1}\otimes P'_N\widetilde{W}_N\right)\boldsymbol{\eta}\boldsymbol{\eta}'\left(\boldsymbol{\gamma}_{t-1}\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\right] \\
&\quad +N^{-2}2\delta^*_{0,t-1}\delta^{**}_{0,t-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}'_{t-1}\otimes P'_N\widetilde{W}_N\right)\boldsymbol{\eta}\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}'_{t-1}\otimes P'_N\widetilde{W}_N\right)\boldsymbol{\eta}\right] \\
&\quad +N^{-2}2\delta^{*2}_{0,t-1}\delta^{**}_{0,t-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}'_{t-1}\otimes P'_N\widetilde{W}_N\right)\boldsymbol{\eta}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right]- \\
&\quad +N^{-2}\delta^{**2}_{0,t-1}E\left[\boldsymbol{\eta}'\left(\boldsymbol{\beta}_{t-1}\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}'_{t-1}\otimes P'_N\widetilde{W}_N\right)\boldsymbol{\eta}\right] \\
&\quad +N^{-2}2\delta^*_{0,t-1}\delta^{**2}_{0,t-1}E\left[\boldsymbol{\eta}'\left(\boldsymbol{\beta}_{t-1}\otimes \widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\eta}'\widetilde{\mathbf{W}}_{\mathbf{N}}\boldsymbol{\eta}\right]- \\
&\quad +N^{-2}\delta^{*2}_{0,t-1}\delta^{**2}_{0,t-1}E\left[\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right] \\
&= \left[2tr\left[C_{1,\ell N}\Sigma_\upsilon C'_{1,\ell N}\Sigma_\upsilon\right]+\right] \\
&\quad N^{-2}\left[2tr\left[C_{2,\ell N}\Sigma_\upsilon C'_{2,\ell N}\Sigma_\upsilon\right]+\sum_{i=1}^{\ell N}\alpha^2_{ii,2,\ell N}\left(\mu^4_{\upsilon_i}-3\sigma^4_{\upsilon_i}\right)\right] \\
&\quad +N^{-2}2\delta^*_{0,t-1}\delta^{**}_{0,t-1}tr\left[C_{2,\ell N}\Sigma_\upsilon\right]tr\left[\widetilde{W}_N\Sigma_\eta\right] \\
&\quad +N^{-2}2\delta^{**}_{0,t-1}tr\left[G_{3,n}\Sigma_\eta G'_{3,n}\Sigma_\upsilon\right]+N^{-2}2\delta^*_{0,t-1}\delta^{**}_{0,t-1}tr\left[G'_{3,n}\Sigma_\eta C_{3,\ell N}\Sigma_\upsilon\right] \\
&\quad +N^{-2}\delta^{*2}_{0,t-1}tr\left[C_{3,\ell N}\Sigma_\eta C'_{3,\ell N}\Sigma_\upsilon\right]+N^{-2}\delta^{*2}_{0,t-1}\delta^{**2}_{0,t-1}tr\left[\widetilde{W}_N\Sigma_\eta\widetilde{W}_N\Sigma_\eta\right],
\end{aligned}
\tag{77}
$$

where $C_{2,\ell N}=\left(\boldsymbol{\gamma}'_{t-1}\boldsymbol{\beta}_{t-1}\otimes P'_N\widetilde{W}_N P_N\right)$, $G_{3,n}=\left(\boldsymbol{\gamma}'_{t-1}\otimes P'_N\widetilde{W}_N\right)$, $C_{3,\ell N}=\left(\boldsymbol{\beta}_{t-1}\otimes\widetilde{W}_N P_N\right)$, using Lemma 11 repeatedly. Finally, by Lemma 10(iii)-(iv) it follows that $E\psi^2_{6,t,N}=o(1)$. QED

PROOF OF THEOREM 2

Firstly, notice that using the expression for $\widehat{\alpha}_D$ in (16), we can write

$$
\sqrt{N}\left(\widehat{\alpha}_D-\alpha\right)=\left[\frac{\Delta\mathbf{y}'_{-1}Z_D}{N}A_{1,D,N}\frac{Z'_D\Delta\mathbf{y}_{-1}}{N}\right]^{-1}\left[\frac{\Delta\mathbf{y}'_{-1}Z_D}{N}A_{1,D,N}\frac{Z'_D\Delta\boldsymbol{\varepsilon}}{\sqrt{N}}\right]. \tag{78}
$$

By the maintained assumptions we know that $N^{-1}Z'_D\Delta\mathbf{y}_{-1}\xrightarrow{p}\mathbf{q}_{Z_D\Delta y_{-1}}$, $N^{-1}Z'_DZ_D\xrightarrow{p}Q_{Z_D}$, and $A_{1,D,N}\xrightarrow{p}A_{1,D}$, where all limiting matrices are finite with full column rank. In addition, as shown in (67) with $\widetilde{W}_N$ replaced by I_N, the elements of $Z'_D\Delta\boldsymbol{\varepsilon}$ are quadratic forms of the innovations, $\boldsymbol{\upsilon}$ and $\boldsymbol{\eta}$. By Assumption 1 these random variables satisfy Assumptions B.1, B.3 in Appendix B. Furthermore, Assumption 2 ensures that Assumption B.2. is also met. As a result, we have

$\Omega_{1,D,N}^{-1/2} Z_D' \Delta \boldsymbol{\varepsilon} \xrightarrow{d} N\left(\mathbf{0}, I_{\zeta_{1,D}}\right)$, with $\Omega_{1,D,N} = var\left[\sqrt{N}\mathbf{m}_{N,D}(\alpha)\right]$, by virtue of the central limit theorem provided by Lemma 13. Hence,

$$\sqrt{N} Z_D' \Delta \boldsymbol{\varepsilon} = \sqrt{N}\Omega_{1,D,N}^{1/2}\Omega_{1,D,N}^{-1/2} Z_D' \Delta \boldsymbol{\varepsilon} \xrightarrow{d} N(\mathbf{0}, \Omega_{1,D}).$$

The desired result follows from the generalised Slutsky theorem. Of course, this also implies that $\widehat{\alpha}_D$ is consistent. The proof of (20) is similar and therefore it will be omitted. QED

PROOF OF THEOREM 5

$\widetilde{\alpha}_D$ can be written as

$$\sqrt{N}(\widetilde{\alpha}_D - \alpha) = \left[\frac{\Delta \mathbf{y}_{-1}' Z_{\widetilde{D}}}{N} A_{2,D,N} \frac{Z_{\widetilde{D}}' \Delta \mathbf{y}_{-1}}{N}\right]^{-1} \left[\frac{\Delta \mathbf{y}_{-1}' Z_{\widetilde{D}}}{N} A_{2,D,N} \frac{Z_{\widetilde{D}}' \Delta \boldsymbol{\varepsilon}}{\sqrt{N}}\right]. \tag{79}$$

Combining Assumption 3 and Proposition 3, namely $N^{-1} Z_D' \Delta \mathbf{y}_{-1} \xrightarrow{p} \mathbf{q}_{Z_D \Delta y_{-1}}$ and $N^{-1} \widetilde{Z}_D' \Delta \boldsymbol{y}_{-1} \xrightarrow{p} \mathbf{q}_{\widetilde{Z}_D \Delta y_{-1}}$, we have that $N^{-1} \Delta \mathbf{y}_{-1}' Z_{\widetilde{D}} \xrightarrow{p} \mathbf{q}_{Z_{\widetilde{D}} \Delta y_{-1}}'$, such that all entries are different from zero. Furthermore, under the maintained assumptions in Section 3 we have $N^{-1} Z_{\widetilde{D}}' Z_{\widetilde{D}} \xrightarrow{p} Q_{Z_{\widetilde{D}}}$, and $A_{2,D,N} \xrightarrow{p} A_{2,D}$, where all limiting matrices are finite with full column rank. It is straightforward to show that the elements of $Z_{\widetilde{D}}' \Delta \boldsymbol{\varepsilon}$ are stacked random variables consisting of $\mathbf{y}_s' \Delta \boldsymbol{\varepsilon}_t$, and $\mathbf{y}_s' \widetilde{W}_N \Delta \boldsymbol{\varepsilon}_t$. As shown in (67) these are quadratic forms of the innovations, $\boldsymbol{\upsilon}$ and $\boldsymbol{\eta}$, and satisfy Assumptions B.1, B.3 in Appendix B, given Assumption 1. Moreover, since $\widetilde{W}_N = W_N + W_N'$, it has uniformly bounded row and column sums by Lemma 9. This implies that Assumption B.2. is satisfied. Combining Propositions 1 and 3 it follows that

$$\mathbf{m}_{N,\widetilde{D}}(\alpha) \xrightarrow{p} \mathbf{0}. \tag{80}$$

As a result, $\Omega_{2,\widetilde{D},N}^{-1/2} Z_{\widetilde{D}}' \Delta \boldsymbol{\varepsilon} \xrightarrow{d} N\left(\mathbf{0}, I_{\zeta_{2,D}}\right)$, with $\Omega_{2,\widetilde{D},N} = var\left[\sqrt{N}\mathbf{m}_{N,\widetilde{D}}(\alpha)\right]$, by virtue of the central limit theorem provided by Lemma 13. Hence,

$$\sqrt{N} Z_{\widetilde{D}}' \Delta \boldsymbol{\varepsilon} = \sqrt{N}\Omega_{2,\widetilde{D},N}^{1/2}\Omega_{2,\widetilde{D},N}^{-1/2} Z_{\widetilde{D}}' \Delta \boldsymbol{\varepsilon} \xrightarrow{d} N\left(\mathbf{0}, \Omega_{2,\widetilde{D}}\right).$$

The result follows by the generalised Slutsky theorem. Of course, this also implies that $\widetilde{\alpha}_D$ is consistent. The proof of the second part of the theorem is similar and it will be omitted. QED

PROOF OF PROPOSITION 6

Define the following terms:

$$\begin{aligned}
\psi_{7,st,N} &\equiv N^{-1} \widetilde{\mathbf{y}}_s' \widetilde{W}_N \Delta \widetilde{\mathbf{u}}_t; \\
\psi_{8,t,N} &\equiv N^{-1} \Delta \widetilde{\mathbf{y}}_{t-1}' \widetilde{W}_N \widetilde{\mathbf{u}}_t.
\end{aligned}$$

The aim is to show that $E\psi_{k,st,N} = 0$, $E\psi_{k,st,N}^2 \to 0$ as $N \to \infty$ for $k = 7, 8$. Using (49) and (54) we have

$$\begin{aligned}
E\psi_{7,st,N} &= N^{-1} E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s' \otimes P_N'\right) + \boldsymbol{\lambda}'\left(I_N \otimes \Phi' \boldsymbol{\psi}_{1,s}'\right) + \delta_{0,s}^* \boldsymbol{\eta}'\right] \widetilde{W}_N \left[\left(I_N \otimes \Delta \boldsymbol{\phi}_t'\right) \boldsymbol{\lambda} + (\boldsymbol{d}_t \otimes P_N) \boldsymbol{\upsilon}\right] \\
&= N^{-1} E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s' \otimes P_N'\right) \widetilde{W}_N \left(I_N \otimes \Delta \boldsymbol{\phi}_t'\right) \boldsymbol{\lambda}\right] + N^{-1} E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s' \otimes P_N'\right) \widetilde{W}_N (\boldsymbol{d}_t \otimes P_N) \boldsymbol{\upsilon}\right] \\
&\quad + N^{-1} E\left[\boldsymbol{\lambda}'\left(I_N \otimes \Phi' \boldsymbol{\psi}_{1,s}'\right) \widetilde{W}_N \left(I_N \otimes \Delta \boldsymbol{\phi}_t'\right) \boldsymbol{\lambda}\right] \\
&\quad + N^{-1} E\left[\boldsymbol{\lambda}'\left(I_N \otimes \Phi' \boldsymbol{\psi}_{1,s}'\right) \widetilde{W}_N (\boldsymbol{d}_t \otimes P_N) \boldsymbol{\upsilon}\right] \\
&\quad + N^{-1} E\left[\delta_{0,s}^* \boldsymbol{\eta}' \widetilde{W}_N \left(I_N \otimes \Delta \boldsymbol{\phi}_t'\right) \boldsymbol{\lambda}\right] + N^{-1} E\left[\delta_{0,s}^* \boldsymbol{\eta}' \widetilde{W}_N (\boldsymbol{d}_t \otimes P_N) \boldsymbol{\upsilon}\right] \\
&= N^{-1} tr\left[\left(\widetilde{W}_N \otimes \Phi' \boldsymbol{\psi}_{1,s}' \Delta \boldsymbol{\phi}_t'\right) \Sigma_\lambda\right],
\end{aligned}$$

where $\Sigma_\lambda = E\boldsymbol{\lambda}\boldsymbol{\lambda}'$, which follows from (i) using repeatedly the property $\mathbf{b}_1 A \mathbf{b}_2 = tr\left[A\mathbf{b}_2\mathbf{b}_1'\right]$ for any vectors $\mathbf{b}_1$ and $\mathbf{b}_2$, (ii) the fact that $E\boldsymbol{\lambda}\boldsymbol{\upsilon}' = E\boldsymbol{\lambda}\boldsymbol{\eta}' = E\boldsymbol{\upsilon}\boldsymbol{\eta}' = \mathbf{0}$ by assumption, and (iii) the result $N^{-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\otimes P_N'\right)\widetilde{W}_N\left(\boldsymbol{d}_t\otimes P_N\right)\boldsymbol{\upsilon}\right] = 0$ from the proof of Proposition 3. Since $\widetilde{W}_N$ contains zeros on the main diagonal, the kronecker product matrix $\left(\widetilde{W}_N\otimes\Phi'\boldsymbol{\psi}_{1,s}'\Delta\boldsymbol{\phi}_t'\right)$ preserves this property. On the other hand, Σ_λ is a diagonal matrix. Therefore, $tr\left[\left(\widetilde{W}_N\otimes\Phi'\boldsymbol{\psi}_{1,s}'\Delta\boldsymbol{\phi}_t'\right)\Sigma_\lambda\right] = 0$ by Lemma 10(i).

Remark 8 *Observe that if $\widetilde{W}_N$ is replaced by I_N, the trace of the matrix in the square brackets is $O(N)$ and so $N^{-1}tr\left[\left(\widetilde{W}_N\otimes\Phi'\boldsymbol{\psi}_{1,s}'\Delta\boldsymbol{\phi}_t'\right)\Sigma_\lambda\right] = O(1)$; the moment conditions are not valid in this case.*

For $E\psi_{7,st,N}^2$ we have

$$
\begin{aligned}
E\psi_{7,st,N}^2 &= N^{-2}E\widetilde{\mathbf{y}}_s'\widetilde{W}_N\Delta\widetilde{\mathbf{u}}_t\Delta\widetilde{\mathbf{u}}_t'\widetilde{W}_N\widetilde{\mathbf{y}}_s \\
&\quad N^{-1}E\{\left[\delta_{0,s}^*\boldsymbol{\eta}' + \boldsymbol{\lambda}'\left(I_N\otimes\Phi'\boldsymbol{\psi}_{1,s}'\right) + \boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\otimes P_N'\right)\right]\widetilde{W}_N\left[\left(I_N\otimes\Delta\boldsymbol{\phi}_t'\right)\boldsymbol{\lambda} + \left(\boldsymbol{d}_t\otimes P_N\right)\boldsymbol{\upsilon}\right] \\
&\quad \times\left[\boldsymbol{\lambda}'\left(I_N\otimes\Delta\boldsymbol{\phi}_t\right) + \boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\otimes P_N'\right)\right]\widetilde{W}_N\left[\delta_{0,s}^*\boldsymbol{\eta} + \left(I_N\otimes\boldsymbol{\psi}_{1,s}\Phi\right)\boldsymbol{\lambda} + \left(\boldsymbol{\beta}_s\otimes P_N\right)\boldsymbol{\upsilon}\right]\} \\
&= N^{-2}E\{[\delta_{0,s}^*\boldsymbol{\eta}'\left(\widetilde{W}_N\otimes\Delta\boldsymbol{\phi}_t'\right)\boldsymbol{\lambda} + \delta_{0,s}^*\boldsymbol{\eta}'\left(\boldsymbol{d}_t\otimes\widetilde{W}_N P_N\right)\boldsymbol{\upsilon} \\
&\quad + \boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Phi'\boldsymbol{\psi}_{1,s}'\Delta\boldsymbol{\phi}_t'\right)\boldsymbol{\lambda} + \boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Phi'\boldsymbol{\psi}_{1,s}'\right)\left(\boldsymbol{d}_t\otimes P_N\right)\boldsymbol{\upsilon} \\
&\quad + \boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\otimes P_N'\right)\left(\widetilde{W}_N\otimes\Delta\boldsymbol{\phi}_t'\right)\boldsymbol{\lambda} + \boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}] \\
&\quad \times[\delta_{0,s}^*\boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Delta\boldsymbol{\phi}_t\right)\boldsymbol{\eta} + \boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Delta\boldsymbol{\phi}_t\boldsymbol{\psi}_{1,s}\Phi\right)\boldsymbol{\lambda} + \boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Delta\boldsymbol{\phi}_t\right)\left(\boldsymbol{\beta}_s\otimes P_N\right)\boldsymbol{\upsilon} \\
&\quad + \delta_{0,s}^*\boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta} + \boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\otimes P_N'\right)\left(\widetilde{W}_N\otimes\boldsymbol{\psi}_{1,s}\Phi\right)\boldsymbol{\lambda} + \boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\boldsymbol{\beta}_s\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}]\} \\
&= N^{-2}E\left[\delta_{0,s}^{*2}\boldsymbol{\eta}'\left(\widetilde{W}_N\otimes\Delta\boldsymbol{\phi}_t'\right)\boldsymbol{\lambda}\boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Delta\boldsymbol{\phi}_t\right)\boldsymbol{\eta}\right] \\
&\quad + N^{-2}E\left[\delta_{0,s}^{*2}\boldsymbol{\eta}'\left(\boldsymbol{d}_t\otimes\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\otimes P_N'\widetilde{W}_N\right)\boldsymbol{\eta}\right] \\
&\quad + N^{-2}E\left[\boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Phi'\boldsymbol{\psi}_{1,s}'\Delta\boldsymbol{\phi}_t'\right)\boldsymbol{\lambda}\boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Delta\boldsymbol{\phi}_t\boldsymbol{\psi}_{1,s}\Phi\right)\boldsymbol{\lambda}\right] \\
&\quad + N^{-2}E\left[\boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Phi'\boldsymbol{\psi}_{1,s}'\Delta\boldsymbol{\phi}_t'\right)\boldsymbol{\lambda}\boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\boldsymbol{\beta}_s\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\right] \\
&\quad + N^{-2}E\left[\boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Phi'\boldsymbol{\psi}_{1,s}'\right)\left(\boldsymbol{d}_t\otimes P_N\right)\boldsymbol{\upsilon}\boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Delta\boldsymbol{\phi}_t\right)\left(\boldsymbol{\beta}_s\otimes P_N\right)\boldsymbol{\upsilon}\right] \\
&\quad + N^{-2}2E\left[\boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Phi'\boldsymbol{\psi}_{1,s}'\right)\left(\boldsymbol{d}_t\otimes P_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\otimes P_N'\right)\left(\widetilde{W}_N\otimes\boldsymbol{\psi}_{1,s}\Phi\right)\boldsymbol{\lambda}\right] \\
&\quad + N^{-2}2E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\lambda}'\left(\widetilde{W}_N\otimes\Delta\boldsymbol{\phi}_t\boldsymbol{\psi}_{1,s}\Phi\right)\boldsymbol{\lambda}\right] \\
&\quad + N^{-2}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\left(\boldsymbol{d}_t'\boldsymbol{\beta}_s\otimes P_N'\widetilde{W}_N P_N\right)\boldsymbol{\upsilon}\right] =
\end{aligned}
$$

$$
\begin{aligned}
= \; & N^{-2}\delta_{0,s}^{*2} tr\left[\left(\widetilde{W}_N \otimes \Delta\boldsymbol{\phi}_t'\right)\Sigma_\lambda\left(\widetilde{W}_N \otimes \Delta\boldsymbol{\phi}_t\right)\Sigma_\eta\right] \\
& +N^{-2}\delta_{0,s}^{*2} tr\left[\left(\boldsymbol{d}_t \otimes \widetilde{W}_N P_N\right)\Sigma_\upsilon\left(\boldsymbol{d}_t' \otimes P_N'\widetilde{W}_N\right)\Sigma_\eta\right] \\
& +N^{-2} 2tr\left[\left(\widetilde{W}_N \otimes \Phi'\boldsymbol{\psi}_{1,s}'\Delta\boldsymbol{\phi}_t'\right)\Sigma_\lambda\left(\widetilde{W}_N \otimes \Delta\boldsymbol{\phi}_t\boldsymbol{\psi}_{1,s}\Phi\right)\Sigma_\lambda\right] \\
& +N^{-2} tr\left[\left(\widetilde{W}_N \otimes \Phi'\boldsymbol{\psi}_{1,s}'\Delta\boldsymbol{\phi}_t'\right)\Sigma_\lambda\right] tr\left[\left(\boldsymbol{d}_t'\boldsymbol{\beta}_s \otimes P_N'\widetilde{W}_N P_N\right)\Sigma_\upsilon\right] \\
& +N^{-2} tr\left[\left(\widetilde{W}_N \otimes \Phi'\boldsymbol{\psi}_{1,s}'\right)\left(\boldsymbol{d}_t \otimes P_N\right)\Sigma_\upsilon\left(\boldsymbol{\beta}_s' \otimes P_N'\right)\left(\widetilde{W}_N \otimes \Delta\boldsymbol{\phi}_t'\right)\Sigma_\lambda\right] \\
& +N^{-2} 2tr\left[\left(\widetilde{W}_N \otimes \Phi'\boldsymbol{\psi}_{1,s}'\right)\left(\boldsymbol{d}_t \otimes P_N\right)\Sigma_\upsilon\left(\boldsymbol{d}_t' \otimes P_N'\right)\left(\widetilde{W}_N \otimes \boldsymbol{\psi}_{1,s}\Phi\right)\Sigma_\lambda\right] \\
& +N^{-2} 2tr\left[\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t \otimes P_N'\widetilde{W}_N P_N\right)\Sigma_\upsilon\right] tr\left[\left(\widetilde{W}_N \otimes \Delta\boldsymbol{\phi}_t\boldsymbol{\psi}_{1,s}\Phi\right)\Sigma_\lambda\right] \\
& +N^{-2} 2tr\left[\left(\boldsymbol{\beta}_s'\boldsymbol{d}_t \otimes P_N'\widetilde{W}_N P_N\right)\Sigma_\upsilon\left(\boldsymbol{d}_t'\boldsymbol{\beta}_s \otimes P_N'\widetilde{W}_N P_N\right)\Sigma_\upsilon\right], \qquad (81)
\end{aligned}
$$

using Lemma 11(iii) for the first, second and sixth terms of the last equality, Lemma 11(ii) for the third and last terms, Lemma 11(v) for the fourth seventh terms and Lemma 11(vi) for the remaining term. It is easily seen now using Lemma 10 that $E\psi_{7,st,N}^2 = o(1)$.

Using (53) and (50) we have for $\psi_{8,t,N}$

$$
\begin{aligned}
E\psi_{8,t,N} = \; & N^{-1}E\{\left[\boldsymbol{\upsilon}\left(\boldsymbol{\gamma}_{t-1}' \otimes P_N'\right) + \boldsymbol{\lambda}'\left(I_N \otimes \Phi'\boldsymbol{\psi}_{2,t-1}'\right) + \delta_{0,t-1}^{**}\boldsymbol{\eta}'\right] \\
& \times\widetilde{W}_N\left[\boldsymbol{\eta} + \left(I_N \otimes \boldsymbol{\phi}_t'\right)\boldsymbol{\lambda} + \left(\boldsymbol{e}_{t+1}' \otimes P_N\right)\boldsymbol{\upsilon}\right]\} \\
= \; & N^{-1}E\left[\boldsymbol{\upsilon}\left(\boldsymbol{\gamma}_{t-1}' \otimes P_N'\right)\widetilde{W}_N\boldsymbol{\eta}\right] + N^{-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}' \otimes P_N'\right)\widetilde{W}_N\left(I_N \otimes \boldsymbol{\phi}_t'\right)\boldsymbol{\lambda}\right] \\
& +N^{-1}E\left[\boldsymbol{\upsilon}'\left(\boldsymbol{\gamma}_{t-1}' \otimes P_N'\right)\widetilde{W}_N\left(\boldsymbol{e}_{t+1}' \otimes P_N\right)\boldsymbol{\upsilon}\right] + N^{-1}E\left[\boldsymbol{\lambda}'\left(I_N \otimes \Phi'\boldsymbol{\psi}_{2,t-1}'\right)\widetilde{W}_N\boldsymbol{\eta}\right] \\
& +N^{-1}E\left[\boldsymbol{\lambda}'\left(I_N \otimes \Phi'\boldsymbol{\psi}_{2,t-1}'\right)\widetilde{W}_N\left(I_N \otimes \boldsymbol{\phi}_t'\right)\boldsymbol{\lambda}\right] \\
& +N^{-1}E\left[\boldsymbol{\lambda}'\left(I_N \otimes \Phi'\boldsymbol{\psi}_{2,t-1}'\right)\widetilde{W}_N\left(\boldsymbol{e}_{t+1}' \otimes P_N\right)\boldsymbol{\upsilon}\right] \\
& +N^{-1}E\left[\delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\boldsymbol{\eta}\right] + N^{-1}E\left[\delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\left(I_N \otimes \boldsymbol{\phi}_t'\right)\boldsymbol{\lambda}\right] \\
& +N^{-1}E\left[\delta_{0,t-1}^{**}\boldsymbol{\eta}'\widetilde{W}_N\left(\boldsymbol{e}_{t+1}' \otimes P_N\right)\boldsymbol{\upsilon}\right] \\
= \; & N^{-1}tr\left[\left(\boldsymbol{\gamma}_{t-1}'\boldsymbol{e}_{t+1}' \otimes P_N'\widetilde{W}_N P_N\right)\Sigma_\upsilon\right] + N^{-1}\delta_{0,t-1}^{**} tr\left[\widetilde{W}_N\Sigma_\eta\right] \\
& +N^{-1}tr\left[\left(\widetilde{W}_N \otimes \Phi'\boldsymbol{\psi}_{2,t-1}'\boldsymbol{\phi}_t'\right)\Sigma_\lambda\right].
\end{aligned}
$$

Observe that the first two terms in the last equality are zero from the proof of Proposit ion 3. The last term is also zero by Lemma 10(i). Finally, it is entirely straightforward (but again tedious) to show that $E\psi_{8,t,N}^2 = o(1)$. QED

PROOF OF THEOREM 7

We have

$$
\sqrt{N}\left(\widehat{\alpha}_{\widetilde{D}} - \alpha\right) = \left[\frac{\Delta\widetilde{\mathbf{y}}_{-1}'\widetilde{Z}_D}{N} A_{3,D,N}\frac{\widetilde{Z}_D'\Delta\widetilde{\mathbf{y}}_{-1}}{N}\right]^{-1}\left[\frac{\Delta\widetilde{\mathbf{y}}_{-1}'\widetilde{Z}_D}{N} A_{3,D,N}\frac{\widetilde{Z}_D'\Delta\widetilde{\mathbf{u}}}{\sqrt{N}}\right]. \qquad (82)
$$

Under the maintained assumptions $N^{-1}\widetilde{Z}_D'\Delta\widetilde{\boldsymbol{y}}_{-1} \xrightarrow{p} \mathbf{q}_{\widetilde{Z}_D\Delta y_{-1}}$, a vector with non-zero entries, and $A_{2,D,N} \xrightarrow{p} A_{2,D}$, where the limiting matrix is finite with full column rank. It is easy to see that the elements of $\widetilde{Z}_D'\Delta\widetilde{\mathbf{u}}$ are stacked random variables consisting of $\widetilde{\mathbf{y}}_s'\widetilde{W}_N\Delta\widetilde{\mathbf{u}}_t$. As shown in Proposition

6 these are quadratic forms of the innovations, $\boldsymbol{\upsilon}$, $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ that satisfy Assumptions B.1, B.3 in Appendix B, given Assumption 1 and Assumption 4. Moreover, $\widetilde{W}_N$ has uniformly bounded row and column sums by Lemma 9 and satisfies Assumption B.2. As a result, $\Omega_{3,D,N}^{-1/2}\widetilde{Z}_D'\Delta\widetilde{\mathbf{u}} \xrightarrow{d} N\left(\mathbf{0}, I_{\zeta_{1,D}}\right)$, with $\Omega_{3,D,N} = var\left[\sqrt{N}\widetilde{\mathbf{m}}_{N,\widetilde{D}}(\alpha)\right]$, by virtue of the central limit theorem provided by Lemma 13. Hence,

$$\sqrt{N}\widetilde{Z}_D'\Delta\widetilde{\mathbf{u}} = \sqrt{N}\Omega_{3,D,N}^{1/2}\Omega_{3,D,N}^{-1/2}\widetilde{Z}_D'\Delta\widetilde{\mathbf{u}} \xrightarrow{d} N\left(\mathbf{0}, \Omega_{3,D}\right).$$

The result follows by the generalised Slutsky theorem. Of course, this also implies that $\widehat{\alpha}_{\widetilde{D}}$ is consistent. The proof of the second part of the theorem is similar and it will be omitted. QED

Appendix B

Lemma 9 *Let $C_{k,N}$ be a (sequence of) $N \times N$ matrices, for $k = 1,...,$ whose row and column sums are uniformly bounded in absolute value by finite constants, c_k, respectively. Then the row and column sums of the product of a finite number of $C_{k,N}$, e.g. $C_{1,N}C_{2,N}C_{3,N}$, and of the sum of a finite number of $C_{k,N}$, e.g. $C_{1,N}+C_{2,N}+C_{3,N}$ are uniformly bounded in absolute value by $\prod_{k=1} c_k$ and $\sum_k c_k$, respectively. Furthermore, consider a finite $l_1 \times l_2$ matrix H, whose rows and columns sums are uniformly bounded in absolute value by a constant c_h. Then $H \otimes C_{k,N}$ has uniformly bounded row and column sums.*

Proof. *These are easily verified. See, for example, Kapoor (2003) and Kelejian Prucha (1999).* ■

Lemma 10 *Let $C_{k,\ell N}^0 = \left[c_{ij,k,\ell N}^0\right]$ and $C_{k,\ell N} = [c_{ij,k,\ell N}]$ be (sequences of) $\ell N \times \ell N$ matrices, with $l \geq 1$ fixed and some k, whose row and columns sums are uniformly bounded in absolute value by $c_{k,C}^0$, $c_{k,C}$, respectively, with $c_{ii,k,\ell N}^0 = 0\ \forall i$. Let $G_{k,\ell N} = [g_{ij,k,\ell N}]$ be a (sequence of) $\ell N \times N$ matrices, whose row and columns sums are uniformly bounded in absolute value by $c_{k,G}$. Furthermore, let $D_{k,\ell N} = [d_{ij,k,\ell N}]$ and $\Delta_{k,N} = [\delta_{ij,k,N}]$ be $\ell N \times \ell N$ and $N \times N$ diagonal matrices respectively, for some k, whose diagonal entries are uniformly bounded in absolute value by $c_{k,D}$ and $c_{k,\Delta}$. Then*

$$\begin{aligned}
(i) &\ :\ tr\left[C_{k,\ell N}^0 D_{k,\ell N}\right] = 0;\\
(ii) &\ :\ N^{-1}tr\left[C_{k\ell N}D_{\ell N}C_{k'\ell N}\right] = O(1);\\
(iii) &\ :\ N^{-1}tr\left[C_{k,\ell N}D_{k,\ell N}C_{k',\ell N}D_{k',\ell N}\right] = O(1);\\
(iv) &\ :\ N^{-1}tr\left[G_{k,\ell N}\Delta_{k,N}G_{k',\ell N}'D_{k,\ell N}\right] = O(1).
\end{aligned}$$

Proof: For part (i) of the lemma we have

$$N^{-1}tr\left[C_{k,\ell N}^0 D_{k,\ell N}\right] = N^{-1}\sum_{i=1}^{\ell N}\sum_{j=1}^{\ell N}\left[c_{ij,k,\ell N}^0 d_{ji,k,\ell N}\right] = 0, \tag{83}$$

since $c_{ij,k,\ell N}^0 = 0$ for $j = i$, while $d_{ji,k,\ell N} = 0\ \forall j \neq i$ (all k). For part (ii) we have

$$\begin{aligned}
& N^{-1}tr\left[C_{k,\ell N}D_{\ell N}C_{k',\ell N}\right] = N^{-1}\sum_{i=1}^{\ell N}\sum_{j=1}^{\ell N} c_{ij,k,\ell N}c_{ji,k',\ell N}d_{jj,k,\ell N}\\
\leq\ & c_{k,D}N^{-1}\sum_{i=1}^{\ell N}\sum_{j=1}^{\ell N}\left|c_{ij,k,\ell N}\right|\left|c_{ji,k',\ell N}\right| \leq c_{k,D}N^{-1}\sum_{i=1}^{\ell N}\sum_{j=1}^{\ell N}\left|c_{ij,k,\ell N}\right|\sum_{j=1}^{\ell N}\left|c_{ji,k',\ell N}\right|\\
\leq\ & c_{k,D}N^{-1}\sum_{i=1}^{\ell N} c_{k,C}c_{k',C} = O(1). \qquad (84)
\end{aligned}$$

For part (iii) we have

$$
\begin{aligned}
& N^{-1} tr\left[C_{k,\ell N} D_{k,\ell N} C_{k'\ell N} D_{k',\ell N}\right] = N^{-1} \sum_{i=1}^{\ell N} \sum_{j=1}^{\ell N} c_{ij,k,\ell N} c_{ji,k',\ell N} d_{jj,k,\ell N} d_{ii,k',\ell N} \\
\leq \; & N^{-1} \sum_{i=1}^{\ell N} d_{ii,k',\ell N} \sum_{j=1}^{\ell N} \left|c_{ij,k,\ell N}\right| \left|c_{ji,k',\ell N}\right| d_{jj,k,\ell N} \\
\leq \; & c_{k,D} N^{-1} \sum_{i=1}^{\ell N} d_{ii,k',\ell N} \sum_{j=1}^{\ell N} \left|c_{ij,k,\ell N}\right| \left|c_{ji,k',\ell N}\right| \\
\leq \; & c_{k,D} N^{-1} \sum_{i=1}^{\ell N} d_{ii,k',\ell N} \sum_{j=1}^{\ell N} \left|c_{ij,k,\ell N}\right| \sum_{j=1}^{\ell N} \left|c_{ji,k',\ell N}\right| \\
\leq \; & c_{k,D} c_{k,C} c_{k',C} N^{-1} \sum_{i=1}^{\ell N} d_{ii,k',\ell N} \leq c_{k,D} c_{k',D} c_{k,C} c_{k',C} = O(1). \qquad (85)
\end{aligned}
$$

Finally, part (iv) equals

$$
\begin{aligned}
& N^{-1} tr\left[G_{k,\ell N} \Delta_{k,N} G'_{k',\ell N} D_{k,\ell N}\right] = N^{-1} \sum_{i=1}^{\ell N} \sum_{j=1}^{N} g_{ij,k,\ell N} g_{ij,k',\ell N} d_{ii,k,\ell N} \delta_{jj,k,N} \\
\leq \; & N^{-1} \sum_{i=1}^{\ell N} d_{ii,k,\ell N} \sum_{j=1}^{N} \left|g_{ij,k,\ell N}\right| \left|g_{ij,k',\ell N}\right| \delta_{jj,k,N} \\
\leq \; & c_{k,\Delta} N^{-1} \sum_{i=1}^{\ell N} d_{ii,k,\ell N} \sum_{j=1}^{N} \left|g_{ij,k,\ell N}\right| \left|g_{ij,k',\ell N}\right| \\
\leq \; & c_{k,\Delta} N^{-1} \sum_{i=1}^{\ell N} d_{ii,k',\ell N} \sum_{j=1}^{N} \left|g_{ij,k,\ell N}\right| \sum_{j=1}^{N} \left|g_{ij,k',\ell N}\right| \\
\leq \; & c_{k,\Delta} c_{k,G} c_{k',G} N^{-1} \sum_{i=1}^{\ell N} d_{ii,k',\ell N} \leq c_{k,\Delta} c_{k,G} c_{k',G} c_{k',D} = O(1). \qquad (86)
\end{aligned}
$$

The following lemma concerns the variance and covariance of various quadratic forms.

Lemma 11 *Let* $\boldsymbol{\xi}_1 = \left(\xi_{1,1}, ..., \xi_{\ell N,1}\right)' \sim \left(\mathbf{0}, \Sigma_{\xi_1}\right)$, *and* $\boldsymbol{\xi}_2 = \left(\xi_{1,2}, ..., \xi_{N,2}\right)' \sim \left(\mathbf{0}, \Sigma_{\xi_2}\right)$, *where* Σ_{ξ_1}, Σ_{ξ_2} *are positive-definite* $\ell N \times \ell N$ *and* $N \times N$ *matrices, respectively, for some fixed* ℓ. *Let* $A_{k,\ell N} = [\alpha_{ij,k,\ell N}]$ *and* $B_{k,N} = [\alpha_{ij,k,N}]$ *be sequences of* $\ell N \times \ell N$ *and* $N \times N$ *non-stochastic symmetric matrices, respectively, for some* k, *while* $G_{k,N} = [g_{ij,k,N}]$ *be a sequence of non-stochastic matrices of order* $\ell N \times N$. *Consider the decomposition* $\Sigma_{\xi_1} = S_1 S_1'$, *let* $A^*_{k,\ell N} = \left[\alpha^*_{ij,k,\ell N}\right] = S_1' A_{k,\ell N} S_1$, $B^*_{k,N} = \left[b^*_{ij,k,N}\right] = S_2' B_{k,N} S_2$, *and* $G^*_{k,N} = S_1' G_{k,N} S_2$. *Furthermore, let* $\boldsymbol{\eta}_1 = \left(\eta_{1,1}, ..., \eta_{\ell N,1}\right)' = S_1^{-1} \boldsymbol{\xi}_1$, $\boldsymbol{\eta}_2 = \left(\eta_{1,2}, ..., \eta_{\ell N,2}\right)' = S_2^{-1} \boldsymbol{\xi}_2$. *Then assuming that the elements of* $\boldsymbol{\eta}_1$ *and* $\boldsymbol{\eta}_2$ *are (mutually) independently distributed with zero mean, unit variance and finite third and fourth order moments*

$E\eta_{1,i}^3 = \mu_{\eta_{1,i}}^3$, $E\eta_{1,i}^4 = \mu_{\eta 1,i}^4$, $E\eta_{2,i}^3 = \mu_{\eta_{2,i}}^3$, $E\eta_{2,i}^4 = \mu_{\eta_{2,i}}^4$, *we have*

$$\begin{aligned}
(i) &: E\left(\boldsymbol{\xi}_1' A_{k,\ell N}\boldsymbol{\xi}_1\right) = tr\left(A_{k,\ell N}^*\right) = tr\left(A_{k,\ell N}\Sigma_{\xi_1}\right);\\
(ii) &: Cov\left(\boldsymbol{\xi}_1' A_{k,\ell N}\boldsymbol{\xi}_1, \boldsymbol{\xi}_1' A_{k',\ell N}\boldsymbol{\xi}_1\right) = 2tr\left(A_{k,\ell N}\Sigma_{\xi_1}A_{k',\ell N}\Sigma_{\xi_1}\right)\\
&\quad +\sum_{i=1}^{\ell N}\alpha_{ii,k,\ell N}\alpha_{ii,k',\ell N}\left(\mu_{\xi_{1,i}}^4 - 3\sigma_{\xi_1}^4\right);\\
(iii) &: Cov\left(\boldsymbol{\xi}_1' G_{k,\ell N}\boldsymbol{\xi}_2, \boldsymbol{\xi}_2' G_{k',\ell N}'\boldsymbol{\xi}_1\right) = tr\left[G_{k,n}\Sigma_{\xi_2}G_{k',\ell N}'\Sigma_{\xi_1}\right];\\
(iv) &: Cov\ \left(\boldsymbol{\xi}_1' A_{k,\ell N}\boldsymbol{\xi}_1, \boldsymbol{\xi}_1' G_{k,\ell N}\boldsymbol{\xi}_2\right) = 0;\\
(v) &: Cov\left(\boldsymbol{\xi}_1' A_{k,\ell N}\boldsymbol{\xi}_1, \boldsymbol{\xi}_2' B_{k,N}\boldsymbol{\xi}_2\right) = tr\left[A_{k,\ell N}\Sigma_{\xi_1}\right]tr\left[B_{k,N}\Sigma_{\xi_1}\right];\\
(vi) &: Cov\left(\boldsymbol{\xi}_1' G_{k,\ell N}\boldsymbol{\xi}_2, \boldsymbol{\xi}_1' G_{k',\ell N}\boldsymbol{\xi}_2\right) = tr\left[G_{k,n}\Sigma_{\xi_2}G_{k',n}'\Sigma_{\xi_1}\right].
\end{aligned}$$

Proof. *The expression for the expectation follows from the assumed independence of* $\boldsymbol{\eta}_1$ *because*

$$\begin{aligned}
& E\left(\boldsymbol{\xi}_1' A_{k,\ell N}\boldsymbol{\xi}_1\right) = E\left(\boldsymbol{\xi}_1' S_1^{-1\prime}S_1' A_{k,\ell N}S_1 S_1^{-1}\boldsymbol{\xi}_1\right) = E\left(\boldsymbol{\eta}_1' A_{k,\ell N}^*\boldsymbol{\eta}_1\right)\\
= & \ tr\left(A_{k,\ell N}^* E\boldsymbol{\eta}_1\boldsymbol{\eta}_1'\right) = tr\left(A_{k,\ell N}^*\right) = tr\left(S_1' A_{k,\ell N}S_1\right) = tr\left(S_1 S_1' A_{k,\ell N}\right)\\
= & \ tr\left(A_{k,\ell N}\Sigma_{\xi_1}\right).
\end{aligned} \tag{87}$$

The derivation of the expression of part (ii) follows from the proof of Theorem 1 in Kelejian and Prucha (2001, pg. 242). Observe that the last term drops out if the diagonal elements of $A_{k,n}^*$ *and/or* $A_{k',n}^*$ *equal zero, i.e.* $\alpha_{ii,k,\ell N} = \alpha_{ii,k',\ell N} = 0$. *In this case* $E\left(\boldsymbol{\xi}_1' A_{k,\ell N}\boldsymbol{\xi}_1\right) = tr\left(A_{k,\ell N}\Sigma_{\xi_1}\right) = 0$ *from Lemma 10(i). This term also drops out under normality of* $\boldsymbol{\xi}_1$, *in which case the expression for the covariance is provided in Magnus and Neudecker (1979, Corrolary 4.1(ii)). For parts (iii) and (iv) we have*

$$\begin{aligned}
& Cov\left(\boldsymbol{\xi}_1' G_{k,N}\boldsymbol{\xi}_2, \boldsymbol{\xi}_2' G_{k',N}'\boldsymbol{\xi}_1\right) = E\boldsymbol{\xi}_1' G_{k,N}\boldsymbol{\xi}_2\boldsymbol{\xi}_2' G_{k',N}'\boldsymbol{\xi}_1\\
= & \ E\boldsymbol{\xi}_1' S_1^{-1\prime}S_1' G_{k,N}S_2 S_2^{-1}\boldsymbol{\xi}_2\boldsymbol{\xi}_2' S_2^{-1\prime}S_2' G_{k',N}' S_1 S_1^{-1}\boldsymbol{\xi}_1 = E\boldsymbol{\eta}_1' G_{k,N}^*\boldsymbol{\eta}_2\boldsymbol{\eta}_2' G_{k',N}^{*\prime}\boldsymbol{\eta}_1\\
= & \ E\sum_{i=1}^{\ell N}\sum_{p=1}^{\ell N}\sum_{j=1}^{N}\sum_{q=1}^{N} g_{ij,k,N}^* g_{pq,k',N}\eta_{i,1}\eta_{p,1}\eta_{j,2}\eta_{q,2} = \sum_{i=1}^{\ell N}\sum_{j=1}^{N} g_{ij,k,N}^* g_{pq,k',N}^* E\left(\eta_{i,1}^2\right)E\left(\eta_{j,2}^2\right)\\
= & \ tr\left[G_{k,N}^{\prime *}G_{k',N}^*\right] = tr\left[S_2' G_{k,N}' S_1 S_1' G_{k',N}S_2\right] = tr\left[S_2 S_2' G_{k,N}' S_1 S_1' G_{k',N}\right]\\
= & \ tr\left[G_{k',N}\Sigma_{\xi_2}G_{k,N}'\Sigma_{\xi_1}\right];
\end{aligned} \tag{88}$$

and

$$\begin{aligned}
& Cov\left(\boldsymbol{\xi}_1' A_{k,\ell N}\boldsymbol{\xi}_1, \boldsymbol{\xi}_1' G_{k,\ell N}\boldsymbol{\xi}_2\right) = E\boldsymbol{\xi}_1'' A_{k,\ell N}\boldsymbol{\xi}_1\boldsymbol{\xi}_1' G_{k,\ell N}\boldsymbol{\xi}_2\\
= & \ E\boldsymbol{\xi}_1'' S_1^{-1\prime}S_1' A_{k,\ell N}S_1 S_1^{-1}\boldsymbol{\xi}_1\boldsymbol{\xi}_1' S_1^{-1\prime}S_1' G_{k,\ell N}S_2 S_2^{-1}\boldsymbol{\xi}_2 = E\boldsymbol{\eta}_1' A_{k,\ell N}^*\boldsymbol{\eta}_1\boldsymbol{\eta}_1' G_{k,\ell N}^*\boldsymbol{\eta}_2\\
= & \ E\sum_{i=1}^{\ell N}\sum_{j=1}^{\ell N}\sum_{p=1}^{\ell N}\sum_{q=1}^{N}\alpha_{ij,k,\ell N}^* g_{pq,k,\ell N}^*\eta_{i,1}\eta_{j,1}\eta_{p,1}\eta_{q,2}\\
= & \ \sum_{i=1}^{\ell N}\sum_{j=1}^{\ell N}\sum_{p=1}^{\ell N}\sum_{q=1}^{N}\alpha_{ij,k,\ell N}^* g_{pq,k,\ell N}^* E\left(\eta_{i,1}\eta_{j,1}\eta_{p,1}\right)E\left(\eta_{q,2}\right) = 0.
\end{aligned} \tag{89}$$

For parts (v) and (vi) we have

$$\begin{aligned}
& Cov\left(\boldsymbol{\xi}_1' A_{k,\ell N}\boldsymbol{\xi}_1, \boldsymbol{\xi}_2' B_{k,N}\boldsymbol{\xi}_2\right) = E\boldsymbol{\eta}_1' A_{k,\ell N}^*\boldsymbol{\eta}_1\boldsymbol{\eta}_2' B_{k,N}^*\boldsymbol{\eta}_2\\
= & \ E\sum_{i=1}^{\ell N}\sum_{j=1}^{\ell N}\alpha_{ij,k,\ell N}^*\eta_{j,1}\eta_{i,1}\sum_{p=1}^{N}\sum_{q=1}^{N} b_{pq,k,N}^*\eta_{p,2}\eta_{q,2}\\
= & \ \sum_{i=1}^{\ell N}\alpha_{ii,k,\ell N}^* E\left(\eta_{i,1}^2\right)\sum_{p=1}^{N} b_{pp,k,N}^* E\left(\eta_{p,2}^2\right)\\
= & \ tr\left[A_{k,\ell N}^*\right]tr\left[B_{k,N}^*\right] = tr\left[A_{k,\ell N}\Sigma_{\xi_1}\right]tr\left[B_{k,N}\Sigma_{\xi_1}\right],
\end{aligned} \tag{90}$$

and finally

$$
\begin{aligned}
& Cov\left(\boldsymbol{\xi}_1' G_{k,\ell N}\boldsymbol{\xi}_2, \boldsymbol{\xi}_1' G_{k',\ell N}\boldsymbol{\xi}_2\right) = E\boldsymbol{\eta}_1' G^*_{k,\ell N}\boldsymbol{\eta}_2\boldsymbol{\eta}_1' G^*_{k',\ell N}\boldsymbol{\eta}_2 \\
= \; & E\sum_{i=1}^{\ell N}\sum_{j=1}^{\ell N}\sum_{p=1}^{N}\sum_{q=1}^{N} g^*_{ip,k,\ell N} g^*_{jq,k',\ell N}\eta_{i,1}\eta_{j,1}\eta_{p,2}\eta_{q,2} \\
= \; & \sum_{i=1}^{\ell N}\sum_{p=1}^{N} g^*_{ip,k,\ell N} g^*_{pi,k',\ell N} E\left(\eta^2_{i,1}\right) E\left(\eta^2_{p,2}\right) \\
= \; & tr\left[G^{*\prime}_{k,\ell N} G^*_{k',\ell N}\right] = tr\left[G'_{k,\ell N}\Sigma_{\xi_1} G_{k',\ell N}\Sigma_{\xi_2}\right]. \qquad (91)
\end{aligned}
$$

■

Remark 12 *The assumption of symmetry for $A_{k,\ell N}$ and $B_{k,N}$ is without loss of generality because for any square matrix A, $\mathbf{x}' A\mathbf{x} = \mathbf{x}'\left(\frac{A+A}{2}\right)\mathbf{x}$.*

Consider the following $r \times 1$ vectors of linear quadratic forms

$$
\mathbf{q}_{1,N} = \begin{pmatrix} q_{1,1,N} \\ \vdots \\ q_{r,1,N} \end{pmatrix} = \begin{pmatrix} (\ell N)^{-1/2}\,\boldsymbol{\xi}_1' A_{1,\ell N}\boldsymbol{\xi}_1 \\ \vdots \\ (\ell N)^{-1/2}\,\boldsymbol{\xi}_1' A_{r,\ell N}\boldsymbol{\xi}_1 \end{pmatrix},
$$

$$
\mathbf{q}_{2,N} = \begin{pmatrix} q_{1,2,N} \\ \vdots \\ q_{r,2,N} \end{pmatrix} = \begin{pmatrix} (\ell N)^{-1/2}\,\boldsymbol{\xi}_1' G_{1,N}\boldsymbol{\xi}_2 \\ \vdots \\ (\ell N)^{-1/2}\,\boldsymbol{\xi}_1' G_{r,N}\boldsymbol{\xi}_2 \end{pmatrix},
$$

and

$$
\mathbf{q}_N = \begin{pmatrix} q_{1,N} \\ \vdots \\ q_{r,N} \end{pmatrix} = \begin{pmatrix} (\ell N)^{-1/2}\left[\boldsymbol{\xi}_1' A_{1,\ell N}\boldsymbol{\xi}_1 + \boldsymbol{\xi}_1' G_{1,N}\boldsymbol{\xi}_2\right] \\ \vdots \\ (\ell N)^{-1/2}\left[\boldsymbol{\xi}_1' A_{r,\ell N}\boldsymbol{\xi}_1 + \boldsymbol{\xi}_1' G_{r,N}\boldsymbol{\xi}_2\right] \end{pmatrix},
$$

where $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are ℓN and N vectors of random variables and $A_{k,\ell N}$ and $G_{k,N}$ denote $\ell N \times \ell N$ and $\ell N \times N$ non-stochastic matrices. Let $\boldsymbol{\mu}_{q_N} = E\mathbf{q}_N = E\left[\mathbf{q}_{1,N} + \mathbf{q}_{2,N}\right]$ and $\Sigma_{q_N} = var\left(\mathbf{q}_N\right) = var\left(\mathbf{q}_{1,N}\right) + var\left(\mathbf{q}_{2,N}\right)$, since by Lemma 11 cov$(\mathbf{q}_{1,N} + \mathbf{q}_{2,N}) = 0$.

The following assumptions are maintained:

Assumption B.1. The random variables of the arrays $\left\{\boldsymbol{\xi}_{1,i} : 1 \le i, j \le \ell N\right\}$ and $\left\{\boldsymbol{\xi}_{2,i} : 1 \le i \le N\right\}$ have zero expectation and are totally independent.

Assumption B.2. The elements of the array of real numbers $\{\alpha_{ij,k} : 1 \le i \le \ell N\}$ satisfy $\alpha_{ij,k} = \alpha_{ji,k}$ and $\sup_{j\le \ell N}\sum_{i=1}^{\ell N}|\alpha_{ij,k}| < \infty$, $\sup_{j\le N}\sum_{i=1}^{\ell N}|g_{ij,k}| < \infty$.

Assumption B.3. For all k, the following conditions hold true: $\sup_{i\le \ell N} E\left|\boldsymbol{\xi}_{1,i}\right|^{2+\delta} < \infty$ for some $\delta > 0$ and $\alpha_{ii,k} = 0$, $\sup_{i\le \ell N} E\left|\boldsymbol{\xi}_{1,i}\right|^{4+\delta} < \infty$ for some $\delta > 0$ and $\alpha_{ii,k} \ne 0$, $\sup_{i\le N} E\left|\boldsymbol{\xi}_{2,i}\right|^{2+\delta} < \infty$ for some $\delta > 0$ and $g_{ii,k} = 0$, $\sup_{i\le N} E\left|\boldsymbol{\xi}_{2,i}\right|^{4+\delta} < \infty$ for some $\delta > 0$ and $g_{ii,k} \ne 0$.

The following Lemma provides a central limit theorem for vectors of linear quadratic forms, due to Kelejian and Prucha (2010).

Lemma 13 *Suppose Assumptions B.1 – B.3 hold true and $\lambda_{\min}\left(\Sigma_q\right) \ge c$ for some c, where $\lambda_{\min}\left(\Sigma_q\right)$ denotes the smallest eigenvalue of Σ_q. Letting $\Sigma_q = \left(\Sigma_q^{1/2}\right)\left(\Sigma_q^{1/2}\right)'$, then*

$$
\Sigma_{q_N}^{-1/2}\left(\mathbf{q}_N - \boldsymbol{\mu}_{q_N}\right) \xrightarrow{d} N\left(\mathbf{0}, I_r\right).
$$

Proof. *This follows directly from Kelejian and Prucha (2010), Theorem A.1.* ■

Table A1. Performance of estimators in terms of mean point estimates, Root Mean Squared Error and size

$T=10$	$\alpha=0.2$					$\alpha=0.5$					$\alpha=0.8$				
	WG	FD	FD†	SYS	SYS†	FE	FD	FD†	SYS	SYS†	FE	FD	FD†	SYS	SYS†
$N=100$															
$\xi=1/3$	.071	.150	.134	.173	.161	.322	.423	.404	.457	.445	.543	.625	.618	.732	.729
	(.164)	(.141)	(.133)	(.137)	(.117)	(.203)	(.167)	(.163)	(.149)	(.129)	(.272)	(.269)	(.253)	(.156)	(.137)
	[.992]	[.548]	[.581]	[.405]	[.437]	[.991]	[.878]	[.522]	[.729]	[.234]	[.993]	[.892]	[.384]	[.414]	[.276]
$\xi=1/2$	.068	.116	.113	.158	.148	.314	.369	.372	.431	.427	.532	.525	.566	.693	.708
	(.196)	(.199)	(.161)	(.188)	(.136)	(.233)	(.240)	(.202)	(.202)	(.154)	(.298)	(.380)	(.312)	(.210)	(.165)
	[.991]	[.743]	[.651]	[.489]	[.311]	[.982]	[.838]	[.581]	[.745]	[.783]	[.994]	[.926]	[.796]	[.499]	[.372]
$\xi=2/3$	.063	.078	.082	.140	.128	.302	.307	.327	.401	.401	.515	.444	.500	.656	.677
	(.235)	(.257)	(.199)	(.235)	(.166)	(.272)	(.313)	(.256)	(.250)	(.189)	(.334)	(.461)	(.387)	(.258)	(.205)
	[.984]	[.639]	[.584]	[.317]	[.369]	[.992]	[.872]	[.609]	[.895]	[.528]	[.951]	[.885]	[.734]	[.496]	[.347]
$\xi=3/4$	.060	.058	.061	.131	.116	.294	.277	.295	.387	.381	.504	.414	.451	.639	.654
	(.256)	(.285)	(.226)	(.257)	(.188)	(.294)	(.348)	(.291)	(.272)	(.216)	(.357)	(.490)	(.434)	(.280)	(.234)
	[.955]	[.792]	[.629]	[.348]	[.520]	[.972]	[.901]	[.893]	[.825]	[.506]	[.963]	[.952]	[.919]	[.463]	[.417]
$N=400$															
$\xi=1/3$	.073	.159	.184	.179	.190	.324	.433	.476	.464	.486	.545	.642	.749	.739	.780
	(.160)	(.139)	(.060)	(.134)	(.058)	(.200)	(.161)	(.070)	(.144)	(.063)	(.269)	(.256)	(.103)	(.151)	(.069)
	[.983]	[.436]	[.061]	[.204]	[.061]	[.987]	[.813]	[.067]	[.682]	[.057]	[.973]	[.917]	[.146]	[.333]	[.068]
$\xi=1/2$	.070	.124	.176	.135	.163	.316	.376	.464	.436	.479	.533	.538	.725	.697	.771
	(.193)	(.200)	(.072)	(.199)	(.068)	(.230)	(.238)	(.087)	(.199)	(.076)	(.295)	(.373)	(.134)	(.208)	(.084)
	[.927]	[.421]	[.056]	[.234]	[.077]	[.981]	[.801]	[.069]	[.648]	[.072]	[.917]	[.984]	[.163]	[.289]	[.071]
$\xi=2/3$	.065	.084	.163	.145	.177	.303	.312	.443	.404	.466	.515	.451	.679	.658	.755
	(.233)	(.259)	(.093)	(.235)	(.087)	(.269)	(.314)	(.116)	(.248)	(.098)	(.332)	(.456)	(.189)	(.260)	(.100)
	[.888]	[.438]	[.082]	[.450]	[.084]	[.984]	[.837]	[.096]	[.776]	[.081]	[.947]	[.873]	[.211]	[.370]	[.094]
$\xi=3/4$	.062	.063	.151	.135	.169	.295	.281	.424	.388	.454	.503	.419	.641	.640	.736
	(.255)	(.287)	(.111)	(.257)	(.103)	(.292)	(.349)	(.141)	(.271)	(.117)	(.356)	(.486)	(.232)	(.283)	(.131)
	[.948]	[.570]	[.113]	[.307]	[.094]	[.963]	[.876]	[.147]	[.794]	[.092]	[.949]	[.937]	[.289]	[.446]	[.097]
$N=800$															
$\xi=1/3$	.070	.140	.192	.177	.196	.321	.428	.487	.461	.493	.543	.638	.771	.736	.789
	(.163)	(.130)	(.042)	(.134)	(.041)	(.203)	(.164)	(.049)	(.146)	(.046)	(.271)	(.259)	(.073)	(.153)	(.050)
	[.971]	[.399]	[.048]	[.158]	[.057]	[.982]	[.783]	[.056]	[.632]	[.053]	[.966]	[.941]	[.124]	[.288]	[.059]
$\xi=1/2$	.065	.118	.188	.159	.193	.310	.370	.481	.431	.490	.529	.531	.758	.694	.784
	(.198)	(.201)	(.051)	(.187)	(.050)	(.235)	(.242)	(.061)	(.202)	(.055)	(.299)	(.379)	(.095)	(.214)	(.061)
	[.924]	[.453]	[.051]	[.210]	[.060]	[.987]	[.797]	[.057]	[.651]	[.058]	[.982]	[.976]	[.131]	[.291]	[.062]
$\xi=2/3$	.055	.077	.180	.139	.188	.300	.306	.469	.399	.482	.510	.443	.730	.655	.773
	(.238)	(.259)	(.067)	(.233)	(.064)	(.276)	(.318)	(.083)	(.250)	(.072)	(.338)	(.466)	(.134)	(.260)	(.080)
	[.934]	[.461]	[.071]	[.453]	[.079]	[.991]	[.810]	[.081]	[.784]	[.071]	[.934]	[.917]	[.189]	[.371]	[.081]
$\xi=3/4$	.054	.055	.174	.128	.183	.287	.274	.457	.383	.475	.500	.411	.703	.637	.769
	(.261)	(.288)	(.080)	(.254)	(.077)	(.300)	(.353)	(.100)	(.272)	(.086)	(.362)	(.496)	(.165)	(.283)	(.097)
	[.962]	[.592]	[.081]	[.312]	[.082]	[.982]	[.893]	[.111]	[.812]	[.079]	[.962]	[.943]	[.227]	[.457]	[.090]