CESifo Working Papers

Fixed Effects and Random Effects Estimation of Higher-Order Spatial Autoregressive Models with Spatial Autoregressive and Heteroskedastic Disturbances

Harald Badinger
Peter Egger

Fixed Effects and Random Effects Estimation of Higher-Order Spatial Autoregressive Models with Spatial Autoregressive and Heteroskedastic Disturbances

Abstract

This paper develops a unified framework for fixed and random effects estimation of higher-order spatial autoregressive panel data models with spatial autoregressive disturbances and heteroskedasticity of unknown form in the idiosyncratic error component. We derive the moment conditions and optimal weighting matrix without distributional assumptions for a generalized moments (GM) estimation procedure of the spatial autoregressive parameters of the disturbance process and define both a random effects and a fixed effects spatial generalized two-stage least squares estimator for the regression parameters of the model. We prove consistency of the proposed estimators and derive their joint asymptotic distribution, which is robust to heteroskedasticity of unknown form in the idiosyncratic error component. Finally, we derive a robust Hausman-test of the spatial random against the spatial fixed effects model.

JEL-Code: C130, C210, C230.

Keywords: higher-order spatial dependence, generalized moments estimation, heteroskedasticity, two-stage least squares, asymptotic statistics.

Harald Badinger
Institute for International Economics
Vienna University of Economics and
Business
Augasse 2-6
Austria - 1090 Vienna
harald.badinger@wu.ac.at

Peter Egger
Department of Management, Technology,
and Economics
ETH Zurich
WEH E6, Weinbergstr. 35
Switzerland - 8092 Zurich
egger@kof.ethz.ch

June 2014

I. Introduction

This paper considers the estimation of panel data models with higher-order spatially autocorrelated error components and spatially autocorrelated dependent variables. Spatial interactions in data may originate from various sources such as strategic interaction between jurisdictions (to attract firms or other mobile agents) and firms (in their price, quantity, or quality setting) or general equilibrium effects which disseminate with spatial decay due to their transmission through trade flows, migration, or input-output relationships.[1] Data sets used in empirical studies often share three features: first, they are available in the form of panel data, with a large cross-sectional and a small time series dimension; second, spatial interactions of various kinds co-exist − such as geography-related, trade-related, migration-related interactions − or the decay function of a single spatial interaction is unknown; third, it is unclear whether spatial interactions are local − and affect only immediate neighbors − or global − and affect second third and other neighbors with repercussions. The estimator proposed here addresses the mentioned three features in a unified framework. It allows for panel data with a fixed but arbitrary number of channels or decay segments of spatial interaction in both the error components and the dependent variable, referred to as SARAR(*R*,*S*).

Estimation and testing of both random and fixed effects spatial regressive panel data models with homoskedastic error terms has been considered in the recent literature using a maximum likelihood framework (Baltagi, Song, and Koh, 2003; Lee and Yu, 2010) or a generalized moments approach (Kapoor, Kelejian and Prucha, 2007; Mutl and Pfaffermayr, 2011). The present paper builds on Kapoor et al. (2007). They propose a generalized moments (GM) estimator for the parameters of the spatial regressive error process in a homoskedastic random effects panel data model without endogenous explanatory variables (such as spatial lags of the dependent variable), derive a simplified weighting matrix for the moment conditions under the assumption of normally and identically distributed error components, and prove consistency of the GM estimates. They also establish the asymptotic distribution of the feasible generalized least squares (FGLS) estimates of the parameters of the exogenous regressors.

The present paper extends and generalizes the analysis in Kapoor et al. (2007) in several respects. First, we allow the explanatory variables to be related to the time-invariant error component, i.e., we provide an estimation framework that nests both the fixed and random

---

[1] See Cliff and Ord (1973, 1981), Anselin (1988), and Cressie (1993) for classic references about spatial econometric models in general. Recent theoretical contributions of spatial panel data models include Baltagi, Song, and Koh (2003), Baltagi, Song, Jung, and Koh (2007), Kapoor et al. (2007), Baltagi, Egger, and Pfaffermayr (2008), and Lee and Yu (2008). Recent applications of spatial panel data models include Arbia, Basile, and Piras (2005), Egger, Pfaffermayr, and Winner (2005), Baltagi, Egger, and Pfaffermayr (2007), and Badinger and Egger (2009).

effects setup. Second, we allow for higher-order rather than only first-order spatial regressive processes in both the dependent variable and the error process, enabling a more flexible design and specification tests of the 'spatial' interdependence decay function.[2] Third, we allow for endogenous variables, including spatial lags of the dependent variable in the main equation, which is shown to affect the optimal weighting matrix for the moment conditions as well as the distribution of the GM estimates. Fourth, we do not only prove consistency of the estimates of the model parameters but also derive their joint asymptotic distribution (which is affected by the presence of endogenous variables in a nontrivial way). Fifth, we dispense with the assumption of normally distributed error components, used by Kapoor et al. (2007) to derive a simplified weighting matrix of the moments. In particular, we relax the restrictive assumption that the idiosyncratic errors are identically distributed and allow for heteroskedasticity of arbitrary form over cross-sectional units and time in the idiosyncratic error terms. Under these assumptions, we derive a robust variance-covariance matrix, drawing on recent results by Stock and Watson (2008). We emphasize that the framework of the present paper, the advantage of the GM approach over maximum likelihood (ML) estimation goes beyond that of imposing less restrictive distributional assumptions and computational simplicity, since ML yields inconsistent parameters estimates in the SARAR(R,S) framework with heteroskedasticity of unknown form (see Lee and Yu, 2010). Sixth, we derive a Hausman-test that allows to test the spatial fixed effects against the random effects model in the presence of heteroskedasticity. Seventh and finally, we provide some limited Monte Carlo evidence on the small sample performance of the proposed estimation procedures. In sum this provides a fairly flexible framework for applied work, allowing specification tests, estimation, and inference in random and fixed effects panel data models with potentially higher-order cross-sectional interdependence and heteroskedasticity.

The remainder of the paper is organized as follows. Section II introduces the basic model specification, discusses the fixed versus the random effects model, and provides an overview of the key assumptions of the proposed estimation procedure. Section III proposes GM estimators for the parameters of spatial dependence in the error components. Section IV derives a two-stage least squares (TSLS) and spatial generalized TSLS procedure for estimation of the regression parameters of the model and derives a joint heteroskedasticity-robust asymptotic variance-covariance matrix of the GM and TSLS estimates of the model parameters. Section V derives a consistent estimator of the variance-covariance matrix. Section VI proposes a Hausman-type test of the random versus the fixed effects model. Section VII presents results of a Monte Carlo simulation exercise. Section VIII summarizes our main findings and concludes. The detailed proofs are relegated to a technical appendix.

---

[2] In a cross-sectional framework, estimation of higher order spatial regressive models is considered by Lee and Liu (2010) under homoskedasticity and by Badinger and Egger (2008) under heteroskedasticity.

II. The Basic Model
1. Specification and Key Assumptions

We consider an R-th order spatial regressive panel data model with S-th order spatial regressive error components, referred to as SARAR(R,S) panel data error components model. The basic model comprises $i = 1,...,N$ cross-sectional units and $t = 1,...,T$ time periods. Throughout, subscript N indicates that the variables or parameters are allowed to depend on sample size. For time period t, the model reads

$$\mathbf{y}_{t,N} = \mathbf{X}_{t,N}\boldsymbol{\beta}_N + \sum_{r=1}^{R}\lambda_{r,N}\mathbf{W}_{r,N}\mathbf{y}_{t,N} + \mathbf{u}_{t,N}, \text{ or} \tag{1a}$$

$$\mathbf{y}_{t,N} = \mathbf{Z}_{t,N}\boldsymbol{\delta}_N + \mathbf{u}_{t,N}, \tag{1b}$$

where $\mathbf{y}_{t,N}$ is an $N \times 1$ vector with cross-sectional observations of the dependent variable in year t, $\mathbf{X}_{t,N}$ is an $N \times K$ matrix of observations on K non-stochastic explanatory variables, i.e., $\mathbf{X}_{t,N} = (\mathbf{x}'_{1t,N},...,\mathbf{x}'_{Nt,N})'$, where each of the N vectors $\mathbf{x}_{it,N} = (x_{1,it,N},...,x_{K,it,N})$ is of dimension $1 \times K$, containing the observations on the K explanatory variables for cross-section i and period t. For later reference, define the $T \times K$ matrix $\mathbf{X}_{i,N} = (\mathbf{x}'_{i1,N},...,\mathbf{x}'_{iT,N})'$ as observations on the K explanatory variables for cross-section i and all periods $t = 1,...,T$.

The structure of spatial dependence in $\mathbf{y}_{t,N}$ is determined by the time-invariant $N \times N$ matrices $\mathbf{W}_{r,N}$, $r = 1,...,R$, whose elements $w_{ij,r,N}$ are assumed to be known and will often (but need not) be specified as a decreasing function of geographical distance between the cross-sectional units i and j. The expression $\bar{\mathbf{y}}_{t,r,N} = \mathbf{W}_{r,N}\mathbf{y}_{t,N}$ is referred to as the r-th spatial lag of $\mathbf{y}_N$. The specification of a higher-order process allows the strength of spatial interdependence in the dependent variable (reflected in the spatial autoregressive parameters $\lambda_{r,N}$, $r = 1,...,R$) to vary across a fixed number of R subsets of relations between cross-sectional units.

In equation (1b), the $N \times (K+R)$ design matrix is given by $\mathbf{Z}_{t,N} = (\mathbf{X}_{t,N}, \bar{\mathbf{Y}}_{t,N})$, with $\bar{\mathbf{Y}}_{t,N} = [\bar{\mathbf{y}}_{t,1,N},...,\bar{\mathbf{y}}_{t,R,N}]$, and $\boldsymbol{\delta}_N = (\boldsymbol{\beta}'_N, \boldsymbol{\lambda}'_N)'$, where the $K \times 1$ parameter vector of the exogenous variables is given by $\boldsymbol{\beta}_N = (\beta_{1,N},...,\beta_{K,N})'$ and the $R \times 1$ vector of spatial autoregressive parameters of $\mathbf{y}_N$ is defined as $\boldsymbol{\lambda}_N = (\lambda_{1,N},...,\lambda_{R,N})'$.

The $N \times 1$ vector of error terms $\mathbf{u}_{t,N} = (u_{1t,N},...,u_{Nt,N})'$ is assumed to follow a spatial autoregressive process given by

$$\mathbf{u}_{t,N} = \sum_{s=1}^{S} \rho_{s,N} \mathbf{M}_{s,N} \mathbf{u}_{t,N} + \boldsymbol{\varepsilon}_{t,N}, \tag{1c}$$

$$\boldsymbol{\varepsilon}_{t,N} = \boldsymbol{\mu}_N + \mathbf{v}_{t,N}, \tag{1d}$$

where $\rho_{s,N}$ and $\mathbf{M}_{s,N}$ denote the time-invariant, unknown parameters and the known $N \times N$ matrix of spatial interdependence, respectively. The structure of spatial correlation in the disturbances is determined by the S different, time-invariant $N \times N$ matrices $\mathbf{M}_{s,N}$. As in equation (1a), the specification of a higher-order process allows the strength of spatial interdependence in the disturbances (reflected in the parameters $\rho_{s,N}$, $s = 1,...,S$) to vary across a fixed number of S subsets of relations between cross-sectional units. This enables a more flexible parameterization of the decay of spatial dependence than with a first-order process along two lines: by capturing more than just one channel of interdependence and by allowing for estimation of several parameters $\rho_{s,N}$ for S segments of the decay function (e.g., rings of neighbors or segments of distance). The expression $\bar{\mathbf{u}}_{t,s,N} = \mathbf{M}_{s,N} \mathbf{u}_{t,N}$ is referred to as the s-th spatial lag of $\mathbf{u}_N$. The $S \times 1$ vector of the spatial autoregressive parameters of $\mathbf{u}_{t,N}$ is defined as $\boldsymbol{\rho}_N = (\rho_{1,N},...,\rho_{S,N})'$.

Finally, the $N \times 1$ vector of error terms $\boldsymbol{\varepsilon}_{t,N}$ consists of two error components, a cross-section specific, time-invariant error component $\boldsymbol{\mu}_N$ and an idiosyncratic error component $\mathbf{v}_{t,N}$, which is specific to both the cross-sectional unit and the time period. The typical elements of $\boldsymbol{\varepsilon}_{t,N}$ and $\mathbf{v}_{t,N}$ are the scalars $\varepsilon_{it,N}$ and $v_{it,N}$, respectively, and the $N \times 1$ vector of unit-specific error components is given by $\boldsymbol{\mu}_N = (\mu_{1,N},...,\mu_{N,N})'$.

Stacking observations for all time periods such that t is the slow index and i is the fast index with all vectors and matrices, the model reads

$$\mathbf{y}_N = \mathbf{X}_N \boldsymbol{\beta}_N + \overline{\mathbf{Y}}_N \boldsymbol{\lambda}_N + \mathbf{u}_N, \text{ or} \tag{2a}$$

$$\mathbf{y}_N = \mathbf{Z}_N \boldsymbol{\delta}_N + \mathbf{u}_N, \tag{2b}$$

with the $NT \times K$ regressor matrix $\mathbf{X}_N = (\mathbf{X}'_{1,N},...,\mathbf{X}'_{T,N})'$, and $\overline{\mathbf{Y}}_N = (\overline{\mathbf{y}}_{1,N},...,\overline{\mathbf{y}}_{R,N})$, where $\mathbf{y}_{r,N} = (\mathbf{y}'_{1,r,N},...,\mathbf{y}'_{T,r,N})'$ is the $NT \times 1$ vector of observations on the r-th spatial lag of the dependent variable $\overline{\mathbf{y}}_{r,N}$. The $NT \times 1$ vector of disturbances $\mathbf{u}_N = (\mathbf{u}'_{1,N},...,\mathbf{u}'_{T,N})'$ for the spatial autoregressive process of order S is given by

$$\mathbf{u}_N = \sum_{s=1}^{S} \rho_{s,N} (\mathbf{I}_T \otimes \mathbf{M}_{s,N}) \mathbf{u}_N + \boldsymbol{\varepsilon}_N, \tag{2c}$$

where $\mathbf{I}_T$ is an identity matrix of dimension $T \times T$. The $NT \times 1$ vector $\boldsymbol{\varepsilon}_N = (\boldsymbol{\varepsilon}'_{1,N}, ..., \boldsymbol{\varepsilon}'_{T,N})'$ is specified as

$$\boldsymbol{\varepsilon}_N = (\mathbf{e}_T \otimes \mathbf{I}_N)\boldsymbol{\mu}_N + \mathbf{v}_N = \mathring{\boldsymbol{\mu}}_N + \mathbf{v}_N, \tag{3a}$$

where $\mathbf{e}_T$ is a unit vector of dimension $T \times 1$ and $\mathbf{I}_N$ is an identity matrix of dimension $N \times N$. In light of (2c), the error term can also be written as

$$\boldsymbol{\varepsilon}_N = \mathbf{u}_N - \sum_{s=1}^{S}\rho_{s,N}(\mathbf{I}_T \otimes \mathbf{M}_{s,N})\mathbf{u}_N = \mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{s=1}^{S}\rho_{s,N}\mathbf{M}_{s,N})\mathbf{u}_N. \tag{3b}$$

It follows that

$$\mathbf{u}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{s=1}^{S}\rho_{s,N}\mathbf{M}_{s,N})^{-1}]\boldsymbol{\varepsilon}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{s=1}^{S}\rho_{s,N}\mathbf{M}_{s,N})^{-1}](\mathring{\boldsymbol{\mu}}_N + \mathbf{v}_N), \text{ and} \tag{4a}$$

$$\mathbf{y}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{r=1}^{R}\lambda_{r,N}\mathbf{W}_{r,N})^{-1}]\mathbf{X}_N\boldsymbol{\beta}_N + [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{r=1}^{R}\lambda_{r,N}\mathbf{W}_{r,N})^{-1}]\mathbf{u}_N, \tag{4b}$$

A more general specification of (4a) would allow the spatial regressive parameters (and possibly the weighting matrices) associated with the two error components $\boldsymbol{\mu}_N$ and $\mathbf{v}_N$ to differ as in Baltagi, Egger, and Pfaffermayr (2009). With a higher order process as considered in the present paper, such a specification would be both difficult to identify and computationally involved. Hence we assume the pattern of the spatial regressive disturbance process to be the same for $\boldsymbol{\mu}_N$ and $\mathbf{v}_N$ as in Kapoor et al. (2007).

2. Key Assumptions

As it is standard in the spatial econometric panel data literature, we assume that the explanatory variables collected in $\mathbf{X}_N$ are nonstochastic with elements that are bounded uniformly in absolute value.[3] Without loss of generality we further assume that each explanatory variable changes over time, at least for some cross-section *i*. (Under random effects estimation, this assumption could be relaxed in a straightforward way without invalidating the asymptotic results.) Beyond those, the following assumptions are maintained throughout this paper.

Assumption 1.

Let *T* be a fixed positive integer. (a) For all $1 \leq t \leq T$ and $1 \leq i \leq N, N \geq 1$, the error components $v_{it,N}$ are (mutually) independently distributed with $E(v_{it,N}) = 0$, $E(v_{it,N}^2) = \sigma_{v,it}^2$,

---

[3] See Kapoor, Kelejian, and Prucha (2007, p. 100), Lee and Yu (2008, p.3), Lee and Yu (2010, Assumption 6, p. 6) or Mutl and Pfaffermayr (2011, p.51).

where $0 < \sigma^2_{v,it} < \infty$, and $E|v_{it,N}|^{4+\eta} < \infty$ for some $\eta > 0$. Hence, the idiosyncratic disturbances exhibit heteroskedasticity of unknown form.

(b) For all $1 \le i \le N, N \ge 1$, the unit-specific error components $\mu_{i,N}$ are identically and (mutually) independently distributed with $E(\mu^2_{i,N}) = \sigma^2_\mu$, where $0 < \sigma^2_\mu < b_\mu < \infty$, and $E|\mu_{i,N}|^{4+\eta} < \infty$ for some $\eta > 0$. Following Mundlak (1978), it is assumed that $\mu_{i,N} = \mathbf{x}_{it,N}\boldsymbol{\pi} + w_{it,N}$. Averaging over time periods $t = 1,...,T$ we obtain $\mu_{i,N} = \dot{\mathbf{x}}_{i,N}\boldsymbol{\pi} + \dot{w}_{i,N}$ for $i = 1,...,N$ "between-transformed" observations, where $\dot{\mathbf{x}}_{i,N} = T^{-1}\sum_{t=1}^{T} \mathbf{x}_{it,N}$ and $\dot{w}_{i,N} = T^{-1}\sum_{t=1}^{T} w_{it,N}$ are both $N \times 1$ vectors, and $\dot{w}_{i,N} \sim (0, \sigma^2_{w,N})$. In the random effects model, we have $\boldsymbol{\pi} = \mathbf{0}$, which implies that the time-invariant error component is uncorrelated with the explanatory variables $\mathbf{x}_{it,N}$ in any time period t. In the fixed effects model, we have $\boldsymbol{\pi} \ne \mathbf{0}$, i.e., the explanatory variables are correlated with the time-invariant error component. More precisely, in the random effects specification we have $E(\mu_{i,N}|\mathbf{X}_N) = 0$, whereas in the fixed effects model it holds that $E(\mu_{i,N}|\mathbf{X}_N) = f(\mathbf{X}_N) \ne 0$.[4]

(c) The processes $\{v_{it,N}\}$ and $\{\mu_{i,N}\}$ are independent of each other.[5]

We emphasize that the estimation framework considered here assumes that the spatial regressive structure of the empirical model given by (1a) to (1d) is identical under the fixed effects and the random effects setup, i.e., the time-invariant error component displays the same spatial regressive structure through equations (1b) and (1c), irrespective of the properties of the covariates. This differs from the specification of the spatial regressive fixed effects models in Lee and Yu (2010) as well as Mutl and Pfaffermayr (2011), who exclude the time-invariant error component from the spatial regressive error process.[6]

---

[4] Strictly speaking, with non-stochastic regressors, the two expectations could also be stated unconditionally (see Greene, 2008, p. 18).

[5] Assumption 1 is maintained throughout the paper. For some results, in particular for consistent estimation of the variance-covariance matrix of the GM estimates without distributional assumptions, Assumption 1 will have to be strengthened, assuming that $E|v_{it,N}|^8 < \infty$ for all $1 \le t \le T$ and $1 \le i \le N$.

[6] Lee and Yu (2008) consider maximum likelihood estimation of a homoskedastic spatial regressive fixed effects panel data model; Mutl and Pfaffermayr (2008) consider a Hausman-test for random versus fixed effects first order SARAR(1,1) model with homoskedastic error components. Both partial out the time invariant error component $\boldsymbol{\mu}_N$ from the spatial regressive disturbance process under fixed effects estimation. This choice implies that the time-invariant error component displays the spatial regressive structure of the dependent variable. The difference of Mutl and Pfaffermayr (2008) to our approach is apparent from the

As we will see below, our specification implies that the spatial generalized least squares (GLS) transformed model nests the standard fixed and random effects panel data models. Hence, we regard the nature of the spatial regressive process and the properties of the explanatory variables under random versus fixed effects as two separate sets of assumptions. Our approach allows for cross-sectional interdependence (with a known spatial structure) not only in unobserved variables captured by $\mathbf{v}_N$ but also in unobserved time-invariant variables subsumed in $\mathbf{\mu}_N$ in SARAR(0,S) models (i.e., without a spatially lagged dependent variable). More importantly, this approach allows us to use the same set of four moment conditions both under random effects and fixed effects estimation. Finally, when considering a Hausman test of the random effects versus the fixed effects model in section VI, we wish to consider two model specifications, i.e., the random effects and the fixed effects model, whose assumptions regarding the spatial regressive structure of the error components and the nature of heteroskedasticity are identical and which only differ with regard to whether or not $\mathbf{\pi} = \mathbf{0}$ Assumption (1b).

Assumption 1 implies that

$$E(\varepsilon_{it,N}\varepsilon_{js,N}) = \sigma_\mu^2 + \sigma_{v,it}^2 \text{ for } i = j \text{ and } t = s, \tag{5a}$$

$$E(\varepsilon_{it,N}\varepsilon_{js,N}) = \sigma_\mu^2 \text{ for } i = j \text{ and } t \neq s, \tag{5b}$$

$$E(\varepsilon_{it,N}\varepsilon_{js,N}) = 0, \text{ otherwise.} \tag{5c}$$

As a consequence, the variance-covariance matrix of the stacked error term $\mathbf{\varepsilon}_N$ reads

$$\mathbf{\Omega}_{\varepsilon,N} = E(\mathbf{\varepsilon}_N \mathbf{\varepsilon}_N') = \sigma_\mu^2 (\mathbf{J}_T \otimes \mathbf{I}_N) + \mathbf{\Sigma}_N, \tag{6a}$$

where $\mathbf{J}_T = \mathbf{e}_T \mathbf{e}_T'$ is a $T \times T$ matrix with unitary elements, $\mathbf{I}_{NT}$ is an identity matrix of dimension $NT \times NT$, and

$$\mathbf{\Sigma}_N = E(\mathbf{v}_N \mathbf{v}_N') = diag_{n=1}^{NT} E(v_{n,N}^2) = diag_{n=1}^{NT} E(\varepsilon_{n,N}^2) - \sigma_\mu^2 \mathbf{I}_{NT}. \tag{6b}$$

Note that we use single indexation $n = 1,...,NT$ in equation (6b) to denote elements of the stacked vectors or matrices. We will adopt this convention at several points in the paper in order to simplify notation, when there is no possibility for confusion.

---

specification of the 'Mundlak assumption' in (9). In matrix form and using notation in an obvious way, we assume that $\mathbf{\mu}_N = \dot{\mathbf{X}}_N \mathbf{\pi} + \dot{\mathbf{w}}_N$, whereas Mutl and Pfaffermayr (2008) assume that $\mathbf{\mu}_N = (\mathbf{I}_N - \rho_N \mathbf{W}_N) \dot{\mathbf{X}}_N \mathbf{\pi} + \dot{\mathbf{w}}_N$ (which differs from the specification of their random effects model).

Next, we define two matrices $\mathbf{Q}_{0,N}$ and $\mathbf{Q}_{1,N}$, which are central to the estimation of error component models and the moment conditions of the GM estimator:

$$\mathbf{Q}_{0,N} = (\mathbf{I}_T - \frac{\mathbf{J}_T}{T}) \otimes \mathbf{I}_N \tag{7a}$$

$$\mathbf{Q}_{1,N} = \frac{\mathbf{J}_T}{T} \otimes \mathbf{I}_N . \tag{7b}$$

Pre-multiplying an $NT \times 1$ vector with $\mathbf{Q}_{0,N}$ transforms its elements into deviations from cross-section specific sample means taken over time ("within-transformation"). We will refer to "within-transformed" vectors or matrices with an underbar, e.g., $\mathbf{Q}_{0,N}\mathbf{Z}_N = \underline{\mathbf{Z}}_N$. Pre-multiplying a vector by $\mathbf{Q}_{1,N}$ transforms its elements into cross-section specific sample means ("between-transformation"). Notice that $\mathbf{Q}_{0,N}$ and $\mathbf{Q}_{1,N}$ are both of order $NT \times NT$, symmetric, idempotent, orthogonal to each other, and sum up to $\mathbf{I}_{NT}$.[7]

Assumption 2.

(a) All diagonal elements of the matrices $\mathbf{W}_{r,N}$, $r = 1,...,R$, and $\mathbf{M}_{s,N}$, $s = 1,...,S$, are zero.

(b) The admissible parameter space for the spatial lag of the dependent variable is given by

$$\lambda_{r,N} \in (-a_{L,N}^{\lambda_r}, a_{U,N}^{\lambda_r}), \text{ with } 0 < a_{L,N}^{\lambda_r}, \; a_{U,N}^{\lambda_r} \le a^{\lambda,r} \le a^{\lambda} < \infty, \; r = 1,...,R, \text{ and } \sum_{r=1}^{R} |\lambda_{r,N}| < A_{\lambda} < \infty,$$

where we define a^{λ} such that $a^{\lambda} = \max_{r=1,...,R} (a^{\lambda,r})$ holds.

Analogous assumptions are made for the parameters of the spatial autoregressive error process: $\rho_{s,N} \in (-a_{L,N}^{\rho_s}, a_{U,N}^{\rho_s})$, with $0 < a_{L,N}^{\rho_s}$, $a_{U,N}^{\rho_s} \le a^{\rho,s} \le a^{\rho} < \infty$, $s = 1,...,S$, and

$\sum_{m=1}^{S} |\rho_{m,N}| < A_{\rho} < \infty$, where we define a^{ρ} such that $a^{\rho} = \max_{s=1,...,S} (a^{\rho,s})$ holds.

(c) The matrices $(\mathbf{I}_N - \sum_{r=1}^{R} \lambda_{r,N} \mathbf{W}_{r,N})$ and $(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})$ are nonsingular for $\lambda_r \in (-a_{L,N}^{\lambda_r}, a_{U,N}^{\lambda_r})$ and $\rho_s \in (-a_{L,N}^{\rho_r}, a_{U,N}^{\rho_r})$, respectively.

Part (a) of Assumption 2 is standard. Assumption (2b) requires the spatial regressive parameters to be finite. The admissible value of the scalars A_{λ} (A_{ρ}) will generally depend on the properties of the weights matrices $\mathbf{W}_{r,N}$ ($\mathbf{M}_{s,N}$). For example, with row-normalized matrices $\mathbf{W}_{r,N}$, $r = 1,...,R$, choosing $A_{\lambda} = 1$ ensures that $(\mathbf{I}_N - \sum_{r=1}^{R} \lambda_{r,N} \mathbf{W}_{r,N})$ is invertible, as

---

[7] See Remark A.2 in Appendix A for further properties of $\mathbf{Q}_{0,N}$ and $\mathbf{Q}_{1,N}$.

required in Assumption (2c).[8] Finally, Assumption (2c) ensures that $\mathbf{u}_N$ and $\mathbf{y}_N$ are uniquely identified through equations (4a) and (4b).

Assumption 3.

The row and column sums of the matrices $\mathbf{W}_{r,N}$, $r = 1,...,R$, $\mathbf{M}_{s,N}$, $s = 1,...,S$, $(\mathbf{I}_N - \sum_{r=1}^{R} \lambda_{r,N} \mathbf{W}_{r,N})^{-1}$, and $(\mathbf{I}_N - \sum_{s=1}^{S} \rho_{s,N} \mathbf{M}_{s,N})^{-1}$ are bounded uniformly in absolute value.

In light of Assumptions 1-3 and Remark A.1 in the Appendix, it follows that $E(\mathbf{u}_N) = \mathbf{0}$ and the variance-covariance matrix of $\mathbf{u}_N$ is given by

$$\mathbf{\Omega}_{\mathbf{u},N} = E(\mathbf{u}_N \mathbf{u}_N') = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{s=1}^{S} \rho_{s,N} \mathbf{M}_{s,N})^{-1}]\mathbf{\Omega}_{\boldsymbol{\varepsilon},N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{s=1}^{S} \rho_{s,N} \mathbf{M}_{s,N}')^{-1}], \qquad (8)$$

For the sake generality, all explanatory variables and parameters (except for the variances of the error components $\boldsymbol{\mu}_N$ and $\mathbf{v}_N$) are allowed to depend on sample size *N*. (Of course, all results hold up in the case where parameters do not depend on *N*.) In spatial econometric models this degree of generality is important, given that spatial lags (and disturbance processes) depend on normalized weights matrices. Depending on the weighting scheme, both the spatially weights and the corresponding parameters will change with the size of the cross section dimension, *N*, since a growing *N* (e.g., a growing number of countries or regions) requires renormalizing the weights matrices. Such a specification is consistent, for example, with models where the weights matrices are row-normalized and the number of neighbours of a given cross-sectional unit depends on sample size (see Kapoor et al., 2007, p. 102) or where the strength of interdependence (in terms of the spatial autoregressive parameters) changes with the number of neighbours.

As a result, the model specification in equations (1a)-(1c) is fairly general, allowing for higher-order spatial dependence in the dependent variable, the explanatory variables, and the disturbances, and enabling specification tests to determine to proper structure of cross-sectional interdependence in applied work.

3. Overview of Estimation Procedure
In the following, we outline the estimation procedure proposed in the present paper. Details and proofs of the claims made here are given in the subsequent sections.

---

[8] If the matrices $\mathbf{W}_{r,N}$ are not row-normalized, Assumption (2c) is implied by $A_\lambda = \left(\max_{r=1,...,R} \|\mathbf{W}_{r,N}\| \right)^{-1}$ for some matrix norm $\|\cdot\|$ (see Lee and Liu, 2010, Horn and Johnson, 1985, p. 301).

- In a first step, the regression parameters in model (1a), i.e., $\boldsymbol{\delta}_N$, are consistently estimated by fixed effects two-stage least squares (TSLS), ignoring the spatial regressive structure of $\mathbf{u}_N$ (see Amemiya, 1971, Baltagi, 2005). Under the maintained assumptions, this yields consistent estimates of the disturbances $\hat{\mathbf{u}}_N = \mathbf{y}_N - \mathbf{Z}_N \hat{\boldsymbol{\delta}}_N$. Under stronger assumptions, consistent estimates can also be obtained by pooled two-stage least squares or two-stage least squares with random effects.

- Based on the estimates of the disturbances $\mathbf{u}_N$, a generalized moments (GM) estimator can be used to obtain consistent estimates of the parameters of the spatial regressive disturbance process ($\boldsymbol{\rho}_N$) and the variance of the time-invariant error component (σ_μ^2), denoted as $\tilde{\boldsymbol{\rho}}_N$ and $\tilde{\sigma}_{\mu,N}^2$.

- The joint variance-covariance matrix for the estimates of the regression parameters $\boldsymbol{\delta}_N$ and the spatial regressive parameters $\boldsymbol{\rho}_N$ derived in the present paper, which is robust to both the spatial dependence in $\mathbf{u}_N$ as well as arbitrary heteroskedasticity in the idiosyncratic error term $\mathbf{v}_N$, can be used for specification tests to determine the proper form of the interdependence decay function.[9]

- To improve efficiency (the estimates of) the parameters $\boldsymbol{\rho}_N$ can be used to obtain a (feasible) spatial generalized least squares (GLS) transformed variant of model (1a), which corresponds to a "standard" (fixed or random effects) panel data model without spatial dependence in the disturbances but with heteroskedasticity of unknown form in the idiosyncratic error term $\mathbf{v}_N$. Using this transformed model, feasible spatial generalized two-stage least squares (TSLS) estimates of the regression parameters $\hat{\tilde{\boldsymbol{\delta}}}_N^*$ can be obtained. (The asterisk indicates that the estimates are based on a transformed model; the tilde indicates that the model transformation is based on $\tilde{\boldsymbol{\rho}}_N$, i.e., the GM estimates of $\boldsymbol{\rho}_N$). Again a heteroskedasticity-robust joint variance-covariance of $\hat{\tilde{\boldsymbol{\delta}}}_N^*$ and $\tilde{\boldsymbol{\rho}}_N$ is derived, allowing for joint inference regarding the regression parameters and the spatial regressive parameters of the model.

- The estimation procedure can also be implemented in an iterative way, i.e., the feasible spatial generalized TSLS estimates $\hat{\tilde{\boldsymbol{\delta}}}_N^*$ can be used to obtain iterated

---

[9] The possibility that joint hypotheses about $\boldsymbol{\delta}_N$ and $\boldsymbol{\rho}_N$ may be formulated and tested is an advantage of the proposed two-step approach over the use of (spatial-dependence and heteroskedasticity) robust standard errors. In particular, it allows for specification tests a la Anselin et al. (1996) in a higher order setting and under less restrictive distributional assumptions.

estimates of the disturbances $\mathbf{u}_N$, which can in turn be used to obtain a new set of estimates for $\boldsymbol{\rho}_N$, etc.

- The obtained (feasible or iterated) heteroskedasticity-robust fixed and random effects models can then be tested against each other by a Hausman test which is derived in this paper.

To keep the analysis general, we first consider only the GM estimation of the disturbance process (1c), without assuming a particular form of model (1a) or how consistent estimates of the residuals $\tilde{\mathbf{u}}_N$ of model (1a) are obtained. The advantage of this approach is that the results are potentially applicable to the disturbances of a wider class of regression models, e.g., nonlinear specifications of equation (1a). Then, we consider the estimation of the main equation (1a), using a modular approach with general notation that covers the four estimators considered in the present paper: random effects and fixed effects estimation of both the original and the spatial GLS transformed model.

III. GM Estimation of a SAR(S) Process

In the following, we consider GM estimators for the spatial regressive parameters $\boldsymbol{\rho}_N$ of the disturbance process in equation (1c) and the variance of the time-invariant error component σ_μ^2 and establish their asymptotic joint distribution. In this subsection, we only consider the process in equation (1c) for the disturbances $\mathbf{u}_N$, but not necessarily the one in equation (1a) for $\mathbf{y}_N$. These disturbances $\mathbf{u}_N$ are unknown and thus have to be obtained in a first-step, using consistent estimates of $\boldsymbol{\delta}_N$ in the main equation (1a) (or from some other model), ignoring the spatial regressive error structure in $\mathbf{u}_N$. The assumptions sufficient to establish the asymptotic properties of the GM estimates (consistency and normality) are stated in general terms in Assumptions 4 to 7 in this section and will be made more specific in section IV, where we consider TSLS and spatial generalized TSLS estimation of model (1a). It will also become apparent in this section that the asymptotic distribution of the (second-step) GM estimates of $\boldsymbol{\rho}_N$, which are based on estimated disturbances $\mathbf{u}_N$, is affected in a non-trivial way by the properties of the first-step estimation (fixed versus random effects) and by the presence of endogenous right hand side variables.

1. Moment Conditions

A set of three moment conditions for GM estimation of first-order spatial regressive error processes was introduced in the seminal paper by Kelejian and Prucha (1999) for the case of a single cross-section under homoskedasticity. The extension of this estimator to a random effects panel data error component model by Kapoor et al. (2007) (under homoskedasticity) yields a set of six moment conditions. Heteroskedasticity has so far only been considered in the cross-sectional SARAR(*1,1*) framework by Kelejian and Prucha (2010), who use two of

the three moment conditions in Kelejian and Prucha (1999), and in the SARAR($0,1$) framework by Lin and Lee (2010).

An analogous approach to Kelejian and Prucha (2010) is pursued here in the derivation of the moment conditions under heteroskedasticity, but for (both fixed and random effects) panel data models. For this, we use four of the six moment conditions akin to the ones in Kapoor et al. (2007). Moreover, with an S-th-order rather than a first-order process (SAR(S), with $S > 1$), additional moment conditions are available, associated with each weights matrix $\mathbf{M}_{s,N}$, $s = 1,...,S$, and each pair of weights matrices $\mathbf{M}_{s,N}$, $\mathbf{M}_{s',N}$, $s,=1,...,S; s' > s,...,S$. Define

$$\bar{\boldsymbol{\varepsilon}}_{s,N} = (\mathbf{I}_T \otimes \mathbf{M}_{s,N})\boldsymbol{\varepsilon}_N = (\mathbf{I}_T \otimes \mathbf{M}_{s,N})[\mathbf{u}_N - \sum_{m=1}^{S} \rho_{m,N}(\mathbf{I}_T \otimes \mathbf{M}_{m,N})\mathbf{u}_N]. \tag{10}$$

Under Assumptions 1 to 3, we then have the following set of moment conditions for $T \geq 2$, and $s,=1,...,S; s' > s,...,S$:

$$\mathrm{M}_1^{s,s'}: \quad E[\frac{1}{N(T-1)}\bar{\boldsymbol{\varepsilon}}'_{s,N}\mathbf{Q}_{0,N}\bar{\boldsymbol{\varepsilon}}_{s',N}] = \frac{1}{N(T-1)}Tr[diag_{n=1}^{NT}E(v_{n,N}^2)\mathbf{Q}_{0,N}(\mathbf{I}_T \otimes \mathbf{M}'_{s,N}\mathbf{M}_{s',N})] \tag{11a}$$

$$\mathrm{M}_2^s: \quad E[\frac{1}{N(T-1)}\bar{\boldsymbol{\varepsilon}}'_{s,N}\mathbf{Q}_{0,N}\boldsymbol{\varepsilon}_N] = 0, \text{ or} \tag{11b}$$

$$\mathrm{M}_3^{s,s'}: \quad E[\frac{1}{N}\bar{\boldsymbol{\varepsilon}}'_{s,N}\mathbf{Q}_{1,N}\bar{\boldsymbol{\varepsilon}}_{s',N}] = \frac{T}{N}\sigma_\mu^2 tr(\mathbf{M}'_{s,N}\mathbf{M}_{s',N}) + \frac{1}{N}tr[diag_{n=1}^{NT}E(v_{n,N}^2)\mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{M}'_{s,N}\mathbf{M}_{s',N})] \tag{11c}$$

$$\mathrm{M}_4^s: \quad E[\frac{1}{N}\bar{\boldsymbol{\varepsilon}}'_{s,N}\mathbf{Q}_{1,N}\boldsymbol{\varepsilon}_N] = 0. \tag{11d}$$

Unless part of the weights matrices are orthogonal, there are $4S + S(S-1)$ moment conditions.[10] For the case of a first-order spatial regressive process, i.e., $S = 1$, they nest the moment conditions of the aforementioned GM estimators as special cases. Under homoskedasticity, i.e., $diag_{n=1}^{NT}E(v_{it,N}^2) = \sigma_v^2\mathbf{I}_{NT}$, the corresponding four moment conditions in Kapoor et al. (2007) are then obtained. In the cross-sectional case, i.e., for $T = 1$ (and $\mathbf{Q}_{0,N} = \mathbf{0}$) the moment conditions M_1 and M_2 become uninformative and M_3 and M_4 reduce to the corresponding the two moment conditions in Kelejian and Prucha (2010) under heteroskedasticity with the $N \times N$ matrix $diag_{i=1}^{N}E(v_i^2)$, or the two moment conditions in

---

[10] If some pairs of matrices are orthogonal, $\mathbf{M}'_{s,N}\mathbf{M}_{s',N} = \mathbf{0}$ for some $s \neq s'$, the corresponding moment condition is trivially satisfied for any set of (finite) parameter values. Hence, if all weights matrices were pairwise orthogonal, there would be $4S$ moment conditions.

Kelejian and Prucha (1999) under homoskedasticity with the $N \times N$ matrix $diag_{i=1}^{N} E(v_{i,N}^2) = \mathbf{I}_N \sigma_{v,N}^2$.

Note that the moment conditions can also be written as quadratic forms in the vector $\boldsymbol{\varepsilon}_N$:

$$\text{M}_1^{s,s'}: \quad E[\frac{1}{N(T-1)}\boldsymbol{\varepsilon}_N' \mathbf{A}_{1,N}^{s,s'} \boldsymbol{\varepsilon}_N] = 0 \text{, with} \tag{12a}$$

$$\mathbf{A}_{s,s',N}^1 = \mathbf{Q}_{0,N}\{(\mathbf{I}_T \otimes \mathbf{M}_{s',N}' \mathbf{M}_{s,N}) - diag_{n=1}^{NT}[\mathbf{Q}_{0,N}(\mathbf{I}_T \otimes \mathbf{M}_{s',N}' \mathbf{M}_{s,N})_{nn}]\}\mathbf{Q}_{0,N}.$$

$$\text{M}_2^s: \quad E[\frac{1}{N(T-1)}\boldsymbol{\varepsilon}_N' \mathbf{A}_{2,N}^s \boldsymbol{\varepsilon}_N] = 0 \text{, with } \mathbf{A}_{2,N}^s = \mathbf{Q}_{0,N}(\mathbf{I}_T \otimes \mathbf{M}_{s,N}). \tag{12b}$$

$$\text{M}_3^{s,s'}: \quad E[\frac{1}{N}\boldsymbol{\varepsilon}_N' \mathbf{A}_{3,N}^{s,s'} \boldsymbol{\varepsilon}_N] - \frac{T}{N}\sigma_\mu^2 tr(\mathbf{M}_{s,N}' \mathbf{M}_{s',N}) = 0 \text{, with} \tag{12c}$$

$$\mathbf{A}_{3,N}^{s,s'} = \{\mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{M}_{s',N}' \mathbf{M}_{s,N}) - \mathbf{Q}_{0,N}[diag_{n=1}^{NT}(\mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{M}_{s',N}' \mathbf{M}_{s,N})_{nn}]\mathbf{Q}_{0,N}\}.$$

$$\text{M}_4^s: \quad E[\frac{1}{N}\boldsymbol{\varepsilon}_N' \mathbf{A}_{4,N}^s \boldsymbol{\varepsilon}_N] = 0 \text{, with } \mathbf{A}_{4,N}^s = \mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{M}_{s,N}). \tag{12d}$$

Substituting equations (3a), (3b), (6b), and (10) into the $4S + S(S-1)$ moment conditions (11) yields a $4S + S(S-1)$ equation system in $(\rho_{1,N},...,\rho_{S,N},\sigma_\mu^2)$, which can be written as

$$\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N \mathfrak{b}_N = 0, \tag{13a}$$

where $\mathfrak{b}_N$ is a $[S + S(S-1)/2 + 1] \times 1$ vector given by

$$\mathfrak{b}_N = (\rho_{1,N},...,\rho_{S,N},\rho_{1,N}^2,...,\rho_{S,N}^2,\rho_{1,N}\rho_{2,N},...,\rho_{1,N}\rho_{S,N},...,\rho_{S-1,N}\rho_{S,N},\sigma_\mu^2)',$$

i.e., $\mathfrak{b}_N$ contains S linear terms $\rho_{m,N}$, $m = 1,...,S$, S quadratic terms $\rho_{m,N}^2$, $m = 1,...,S$, $S(S-1)/2$ cross products $\rho_{m,N}\rho_{l,N}$, $m = 1,...,(S-1), l = (m+1),...,S$, as well as σ_μ^2. For later reference, we define the $(S+1) \times 1$ vector of all parameters as $\boldsymbol{\theta}_N = (\boldsymbol{\rho}_N', \sigma_\mu^2)' = (\rho_{1,N},...,\rho_{S,N},\sigma_\mu^2)'$.

$\boldsymbol{\gamma}_N$ is a $[4S + S(S-1)] \times 1$ vector with elements $(\gamma_{i,N})$, $i = 1,...,4S + S(S-1)$, and $\boldsymbol{\Gamma}_N$ is a $[4S + S(S-1)] \times [2S + S(S-1)/2 + 1]$ matrix with elements $(\gamma_{i,j,N})$, $i = 1,...,4S + S(S-1)$, $j = 1,...,2S + S(S-1)/2 + 1$. The elements $\gamma_{i,N}$ and $\gamma_{i,j,N}$ will be defined below.

Throughout the paper, we adopt the following convention with respect to the ordering of the rows in equation system (13). The first four rows are associated with the 4 moment conditions $M_1^{s,s'}$, M_2^s, $M_3^{s,s'}$, and M_4^s with $s = s' = 1$. The next four rows are associated with $M_1^{s,s'}$,

M_2^s, $M_3^{s,s'}$, and M_4^s with $s = s' = 2$, and so forth up to $s = s' = S$. This yields $4S$ rows of the equation system. These moment conditions are always available under Assumptions 1 and 2. Unless part of the weights matrices are orthogonal, there are $S(S-1)$ further moment conditions available, resulting from $M_1^{s,s'}$ and $M_3^{s,s'}$ with $s = 1,...,(S-1)$, $s' = (s+1),...,S$. These are added to the equation system, starting from row $4S+1$, as follows. The next row ($4S+1$) is associated with $M_1^{s,s'}$ and $s = 1$ and $s' = 2$; the next rows with $M_1^{s,s'}$ with $s = 1$ and $s' = 3$, and so forth up to $s = 1$ and $s' = S$; this yields $S(S-1)/2$ rows. We then proceed with $M_3^{s,s'}$ in the same way, yielding another $S(S-1)/2$ rows.

The sample analogue to equation system (13a) is given by

$$\tilde{\boldsymbol{\gamma}}_N - \tilde{\boldsymbol{\Gamma}}_N \mathbf{b}_N = \vartheta_N(\boldsymbol{\theta}_N), \tag{13b}$$

where the elements of $\tilde{\boldsymbol{\gamma}}_N$ and $\tilde{\boldsymbol{\Gamma}}_N$ are equal to those of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ with the expectations operator suppressed and the disturbances $\mathbf{u}_N$ replaced by (consistent) estimates $\tilde{\mathbf{u}}_N$.

GM estimates of the parameters $\rho_{1,N},...,\rho_{S,N}$, σ_μ^2 are then obtained as the solution to

$$\underset{\rho_1,\rho_2,...,\rho_S,\sigma_\mu^2}{\arg\min} \; [(\tilde{\boldsymbol{\gamma}}_N - \tilde{\boldsymbol{\Gamma}}_N \mathbf{b}_N)' \tilde{\boldsymbol{\Theta}}_N (\tilde{\boldsymbol{\gamma}}_N - \tilde{\boldsymbol{\Gamma}}_N \mathbf{b}_N)] = [\vartheta_N(\boldsymbol{\theta}_N)' \tilde{\boldsymbol{\Theta}}_N \vartheta_N(\boldsymbol{\theta}_N)], \tag{14}$$

i.e., the parameter estimates can be obtained from a (weighted) non-linear least squares regression of $\tilde{\boldsymbol{\gamma}}_N$ on the columns of $\tilde{\boldsymbol{\Gamma}}_N$; $\vartheta_N(\boldsymbol{\theta}_N)$ can then be viewed as a vector of regression residuals. The optimal choice of the $[4S + S(S-1)] \times [4S + S(S-1)]$ weighting matrix $\boldsymbol{\Theta}_N$ and its estimation will be discussed below.

In the following, we define the elements of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$, grouped by the corresponding moment conditions. Thereby, we use the following notation:

$$\bar{\mathbf{u}}_{s,N} = (\mathbf{I}_T \otimes \mathbf{M}_{s,N})\mathbf{u}_N, \; s = 1,...,S, \text{ and} \tag{15a}$$

$$\bar{\bar{\mathbf{u}}}_{sm,N} = (\mathbf{I}_T \otimes \mathbf{M}_{s,N})(\mathbf{I}_T \otimes \mathbf{M}_{m,N})\mathbf{u}_N = (\mathbf{I}_T \otimes \mathbf{M}_{s,N}\mathbf{M}_{m,N})\mathbf{u}_N, \; s = 1,...,S, \; m = 1,...,S. \tag{15b}$$

In the derivation of the elements of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$, we also make use of the fact that

$$diag_{n=1}^{NT}(v_{n,N}^2) = diag_{n=1}^{NT}(u_{n,N}^2) - 2\sum_{m=1}^{S} \rho_m diag_{n=1}^{NT}(\bar{u}_{m,n,N}u_{n,N})$$

$$+\sum_{m=1}^{S}\sum_{l=1}^{S}\rho_m\rho_l diag_{n=1}^{NT}(\bar{u}_{m,n,N}\bar{u}_{l,n,N})-\sigma_\mu^2\mathbf{I}_{NT}\,, \tag{16}$$

where $\bar{u}_{m,n,N}$ denotes the n-th element of the vector $\bar{\mathbf{u}}_{m,N}$.

Moment Condition $\mathrm{M}_1^{s,s'}$

Due to the adopted convention regarding the ordering of the rows in equation system (13), the row index for moment condition $\mathrm{M}_1^{s,s'}$ – denoted as row($\mathrm{M}_1^{s,s'}$) – is given by $4(s-1)+1$ for $s=s'$ and $4S+(s-1)S-s(s-1)/2+s'-s$ for $s<s'$. Hence, moment condition $\mathrm{M}_{s,s'}^1$ delivers S rows of equation system (in row 1, 5, …,$4S-3$)) for $s=s'$ and $S(S-1)/2$ rows of equation system (in rows $4S+1,....,4S+S(S-1)/2$) for $s<s'$. The corresponding elements of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ are defined as follows:[11]

$$\gamma_{\mathrm{row}(\mathrm{M}_1^{s,s'})}=\frac{1}{N(T-1)}E\{\bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{0,N}\bar{\mathbf{u}}_{s',N}-Tr[\mathbf{Q}_{0,N}diag_{n=1}^{NT}(u_{n,N}^2)(\mathbf{I}_T\otimes\mathbf{M}'_{s,N}\mathbf{M}_{s',N})]\} \tag{17a}$$

$$=\frac{1}{N(T-1)}E(\mathbf{u}'_N\mathfrak{A}_{1,N}^{s,s'}\mathbf{u}_N)\,,$$

where $\mathfrak{A}_{1,N}^{s,s'}=\{\mathbf{Q}_{0,N}(\mathbf{I}_T\otimes\mathbf{M}'_{s',N}\mathbf{M}_{s,N})-diag_{n=1}^{NT}[\mathbf{Q}_{0,N}(\mathbf{I}_T\otimes(\mathbf{M}'_{s',N}\mathbf{M}_{s,N})]_{nn}\}$.

$$\gamma_{\mathrm{row}(\mathrm{M}_1^{s,s'}),m}=\frac{2}{N(T-1)}E\{\bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{0,N}\bar{\bar{\mathbf{u}}}_{s'm,N}-Tr[\mathbf{Q}_{0,N}diag_{n=1}^{NT}(\bar{u}_{m,n,N}u_{n,N})(\mathbf{I}_T\otimes\mathbf{M}'_{s,N}\mathbf{M}_{s',N})]\}$$

$$=\frac{2}{N(T-1)}E[\mathbf{u}'_N(\mathbf{I}_T\otimes\mathbf{M}'_{m,N})\mathfrak{A}_{1,N}^{s,s'}\mathbf{u}_N]\,,\text{ associated with }\rho_m\,,\ m=1,...,S\,.$$

$$\gamma_{\mathrm{row}(\mathrm{M}_1^{s,s'}),S+m}=-\frac{1}{N(T-1)}E\{\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{0,N}\bar{\bar{\mathbf{u}}}_{s'm,N}-Tr[\mathbf{Q}_{0,N}diag_{n=1}^{NT}(\bar{u}_{m,n,N}^2)(\mathbf{I}_T\otimes\mathbf{M}'_{s,N}\mathbf{M}_{s',N})]\}$$

$$=-\frac{1}{N(T-1)}E[\mathbf{u}'_N(\mathbf{I}_T\otimes\mathbf{M}'_{m,N})\mathfrak{A}_{1,N}^{s,s'}(\mathbf{I}_T\otimes\mathbf{M}_{m,N})\mathbf{u}_N]\,,\text{ associated with }\rho_m^2\,,\ m=1,...,S\,.$$

$$\gamma_{\mathrm{row}(\mathrm{M}_1^{s,s'}),S(m+1)-m(m-1)/2+l-m}=-\frac{2}{N(T-1)}E\{\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{0,N}\bar{\bar{\mathbf{u}}}_{s'l,N}-Tr[\mathbf{Q}_{0,N}diag_{n=1}^{NT}(\bar{u}_{m,n,N}\bar{u}_{l,n,N})(\mathbf{I}_T\otimes\mathbf{M}'_{s,N}\mathbf{M}_{s',N})]\}$$

$$=-\frac{2}{N(T-1)}E[\mathbf{u}'_N(\mathbf{I}_T\otimes\mathbf{M}'_{l,N})\mathfrak{A}_{1,N}^{s,s'}(\mathbf{I}_T\otimes\mathbf{M}_{m,N})\mathbf{u}_N]\,,$$

associated with $\rho_m\rho_l$, $m=1,...,(S-1)$; $l=(m+1),...,S$.

$$\gamma_{\mathrm{row}(\mathrm{M}_1^{s,s'}),2S+S(S-1)/2+1}=-\frac{1}{N}tr(\mathbf{M}'_{s,N}\mathbf{M}_{s',N})\,.$$

Moment Condition M_s^2

Due to the adopted convention with respect to the ordering of the rows in equation system (13a), the row index for moment condition $M_2^{s,s'}$ is given by $4(s-1)+2$. (For $M_2^{s,s'}$ we always have $s=s'$ such that we use only a single subscript.) Hence, moment condition M_2^s delivers S rows of the equation system (in rows 2, 6, …, $4S-2$). The corresponding elements of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ are defined as follows:

$$\gamma_{4(s-1)+2} = \frac{1}{N(T-1)} E(\bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{0,N}\mathbf{u}_N) \tag{17b}$$

$$= \frac{1}{N(T-1)} E(\mathbf{u}'_N\mathfrak{A}^s_{2,N}\mathbf{u}_N), \text{ where } \mathfrak{A}^s_{2,N} = \mathbf{A}^s_{2,N}.$$

$$\gamma_{4(s-1)+2,m} = \frac{1}{N(T-1)} E(\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{0,N}\mathbf{u}_N + \bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{0,N}\bar{\mathbf{u}}_{m,N})$$

$$= \frac{1}{N(T-1)} E[\mathbf{u}'_N(\mathbf{I}_T \otimes \mathbf{M}'_{m,N})(\mathfrak{A}^{s'}_{2,N} + \mathfrak{A}^s_{2,N})\mathbf{u}_N],$$

$$\gamma_{4(s-1)+2,S+m} = -\frac{1}{N(T-1)} E(\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{0,N}\bar{\mathbf{u}}_{m,N})$$

$$= -\frac{1}{N(T-1)} E[\mathbf{u}'_N(\mathbf{I}_T \otimes \mathbf{M}'_{m,N})\mathfrak{A}^s_{2,N}(\mathbf{I}_T \otimes \mathbf{M}_{m,N})\mathbf{u}_N],$$

$$\gamma_{4(s-1)+2,S(m+1)-m(m-1)/2+l-m} = -\frac{1}{N(T-1)} E[\bar{\bar{\mathbf{u}}}'_{sl,N}\mathbf{Q}_{0,N}\bar{\mathbf{u}}_{m,N} + \bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{0,N}\bar{\mathbf{u}}_{l,N}]$$

$$= -\frac{1}{N(T-1)} E[\mathbf{u}'_N(\mathbf{I}_T \otimes \mathbf{M}'_{l,N})(\mathfrak{A}^{s'}_{2,N} + \mathfrak{A}^s_{2,N})(\mathbf{I}_T \otimes \mathbf{M}_{m,N})\mathbf{u}_N],$$

$$\gamma_{4(s-1)+2,2S+S(S-1)/2+1} = 0.$$

Moment Condition $M_3^{s,s'}$

Due to the adopted convention regarding the ordering of the rows in equation system (13), the row index for moment condition $M_3^{s,s'}$ – denoted as row($M_3^{s,s'}$) – is given by $4(s-1)+3$ for $s=s'$ and $4S+S(S-1)/2+(s-1)S-s(s-1)/2+s'-s$ for $s<s'$. Hence, moment condition $M_3^{s,s'}$ delivers S rows of the equation system (in rows 3, 6, …, $4S-1$) for $s=s'$ and $S(S-1)/2$ rows of the equation system (in rows $4S+S(S-1)/2+1,....,4S+S(S-1)$) for $s<s'$.

$$\gamma_{\text{row}(M_3^{s,s'})} = \frac{1}{N} E\{\bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{1,N}\bar{\mathbf{u}}_{s',N} - Tr[\mathbf{Q}_{1,N}diag^{NT}_{n=1}(u^2_{n,N})(\mathbf{I}_T \otimes \mathbf{M}'_{s,N}\mathbf{M}_{s',N})]\} \tag{17c}$$

$$= \frac{1}{N} E(\mathbf{u}'_N\mathfrak{A}^{s,s'}_{3,N}\mathbf{u}_N),$$

where $\mathfrak{A}^{s,s'}_{3,N} = [\mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{M}'_{s',N}\mathbf{M}_{s,N}) - diag^{NT}_{n=1}[\mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{M}'_{s',N}\mathbf{M}_{s,N})]_{nn}]$.

$$\gamma_{\mathrm{row}(\mathbf{M}_3^{s,s'}),m} = \frac{2}{N} E\{\overline{\mathbf{u}}'_{s,N}\mathbf{Q}_{1,N}\overline{\overline{\mathbf{u}}}_{s'm,N} - Tr[\mathbf{Q}_{1,N}diag_{n=1}^{NT}(\overline{u}_{m,n,N}u_{n,N})(\mathbf{I}_T \otimes \mathbf{M}'_{s,N}\mathbf{M}_{s',N})]\}$$

$$= \frac{2}{N} E[\mathbf{u}'_N(\mathbf{I}_T \otimes \mathbf{M}'_{m,N})\mathfrak{A}_{3,N}^{s,s'}\mathbf{u}_N],$$

$$\gamma_{\mathrm{row}(\mathbf{M}_3^{s,s'}),S+m} = -\frac{1}{N} E\{\overline{\overline{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{1,N}\overline{\overline{\mathbf{u}}}_{s'm,N} - Tr[\mathbf{Q}_{1,N}diag_{n=1}^{NT}(\overline{u}_{m,n,N}^2)(\mathbf{I}_T \otimes \mathbf{M}'_{s,N}\mathbf{M}_{s',N})]\}$$

$$= -\frac{1}{N} E[\mathbf{u}'_N(\mathbf{I}_T \otimes \mathbf{M}'_{m,N})\mathfrak{A}_{3,N}^{s,s'}(\mathbf{I}_T \otimes \mathbf{M}_{m,N})\mathbf{u}_N],$$

$$\gamma_{\mathrm{row}(\mathbf{M}_3^{s,s'}),S(m+1)-m(m-1)/2+l-m} = -\frac{2}{N} E\{\overline{\overline{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{1,N}\overline{\overline{\mathbf{u}}}_{s'l,N} - Tr[\mathbf{Q}_{1,N}diag_{n=1}^{NT}(\overline{u}_{m,n,N}\overline{u}_{l,n,N})(\mathbf{I}_T \otimes \mathbf{M}'_{s,N}\mathbf{M}_{s',N})]\}$$

$$= -\frac{2}{N} E[\mathbf{u}'_N(\mathbf{I}_T \otimes \mathbf{M}'_{l,N})\mathfrak{A}_{3,N}^{s,s'}(\mathbf{I}_T \otimes \mathbf{M}_{m,N})\mathbf{u}_N],$$

$$\gamma_{\mathrm{row}(\mathbf{M}_3^{s,s'}),2S+S(S-1)/2+1} = \frac{(T-1)}{N}tr(\mathbf{M}'_{s,N}\mathbf{M}_{s',N}), \text{ associated with } \sigma_{\mu,N}^2.$$

Moment Condition $\mathbf{M}_4^{s,s'}$

The row index for moment condition $\mathbf{M}_4^{s,s'}$ is given by $4(s-1)+4$, i.e., moment condition $\mathbf{M}_4^s$ delivers S rows of the equation system (in rows 4, 8, …, $4S$). Moment condition $\mathbf{M}_2$ delivers S rows of the equation system (in rows 2, 6, …, $4S-2$). The corresponding elements of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ are defined as follows:

$$\gamma_{4(s-1)+4} = \frac{1}{N} E(\overline{\mathbf{u}}'_{s,N}\mathbf{Q}_{1,N}\mathbf{u}_N) \tag{17d}$$

$$= \frac{1}{N} E(\mathbf{u}'_N\mathfrak{A}_{4,N}^s\mathbf{u}_N), \text{ where } \quad \mathfrak{A}_{4,N}^s = \mathbf{A}_{4,N}^s.$$

$$\gamma_{4(s-1)+4,m} = \frac{1}{N} E(\overline{\overline{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{1,N}\mathbf{u}_N + \overline{\mathbf{u}}'_{s,N}\mathbf{Q}_{1,N}\overline{\mathbf{u}}_{m,N})$$

$$= \frac{1}{N} E[\mathbf{u}'_N(\mathbf{I}_T \otimes \mathbf{M}'_{m,N})(\mathfrak{A}_{4,N}^{s\,'} + \mathfrak{A}_{4,N}^s)\mathbf{u}_N],$$

$$\gamma_{4(s-1)+4,S+m} = -\frac{1}{N} E(\overline{\overline{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{1,N}\overline{\mathbf{u}}_{m,N})$$

$$= -\frac{1}{N} E[\mathbf{u}'_N(\mathbf{I}_T \otimes \mathbf{M}'_{m,N})\mathfrak{A}_{4,N}^s(\mathbf{I}_T \otimes \mathbf{M}_{m,N})\mathbf{u}_N],$$

$$\gamma_{4(s-1)+4,S(m+1)-m(m-1)/2+l-m} = -\frac{1}{N} E[\overline{\overline{\mathbf{u}}}'_{sl,N}\mathbf{Q}_{1,N}\overline{\mathbf{u}}_{m,N} + \overline{\overline{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{1,N}\overline{\mathbf{u}}_{l,N}]$$

$$= -\frac{1}{N} E[\mathbf{u}'_N(\mathbf{I}_T \otimes \mathbf{M}'_{l,N})(\mathfrak{A}_{4,N}^{s\,'} + \mathfrak{A}_{4,N}^s)(\mathbf{I}_T \otimes \mathbf{M}_{m,N})\mathbf{u}_N],$$

$$\gamma_{4(s-1)+4,2S+S(S-1)/2+1} = 0.$$

This completes the specification of the elements of the matrices γ_N and Γ_N. The similarity between the structure of the expressions resulting from moment conditions $M_1^{s,s'}$ and M_2^s on the one hand and $M_3^{s,s'}$ and M_4^s on the other hand is apparent. Apart from a slight discrepancy in the definition of the element corresponding to σ_μ^2 between $M_1^{s,s'}$ and $M_3^{s,s'}$, the other elements differ only by the normalization factor and the corresponding matrix of quadratic forms, $Q_{0,N}$ and $Q_{1,N}$, respectively.

2. Definition of GM Estimator

It is a well known result from the literature on generalized method of moments estimation that, for weighting matrix Θ_N in (14), it is optimal to use the inverse of the (properly normalized) variance-covariance matrix of the sample moments, evaluated at the true parameter values. Denote the optimal weighting matrix, which will be derived in Subsection 3.2, by Ψ_N^{-1} and its estimate by $\tilde{\Psi}_N^{-1}$. The optimally weighted GM estimator uses $\tilde{\Theta}_N = \tilde{\Psi}_N^{-1}$ and is defined as

$$(\tilde{\rho}_{1,N},...,\tilde{\rho}_{S,N},\tilde{\sigma}_{\mu,N}^2,) = \arg\min \{\vartheta_N(\underline{\theta})'\tilde{\Theta}_N\vartheta_N(\underline{\theta}), -a^\rho \leq \underline{\underline{\rho}}_s \leq a^\rho, s = 1,..,S, \underline{\underline{\sigma}}_\mu^2 \in [0,b_\mu]\},$$

$$\text{with } \vartheta_N(\underline{\theta}) = \vartheta_N(\underline{\rho},\underline{\underline{\sigma}}_\mu^2,) = (\tilde{\gamma}_N - \tilde{\Gamma}_N\underline{b}). \tag{18}$$

In a first step, we will assume that $\mu_{i,N}$ and $v_{it,N}$ are normally distributed in the derivation of the optimal weighting matrix Ψ_N^{-1} as in Kapoor et al. (2007). In the Appendix, the optimal weighting matrix Ψ_N^{-1} will be derived without distributional assumptions (apart from the ones in Assumption 1). It is worth emphasizing that the use of estimated disturbances together with the presence of endogenous variables in (1a) introduces a difference between the optimal weighting matrix Ψ_N^{-1} and the inverse of the variance-covariance matrix of the sample moments. Under fixed effects, this is also true, even if there are no endogenous variables in the main equation (1a). This will become apparent in section 3.2., where the optimal weighting matrix Ψ_N^{-1} and an estimate $\tilde{\Psi}_N^{-1}$ are derived.

3. Asymptotic Properties of the GM Estimator for θ_N

3.1 Consistency

In order to prove consistency of the estimator $\tilde{\theta}_N$, the following additional assumptions are introduced:

Assumption 4.

Assume that $\tilde{\mathbf{u}}_N - \mathbf{u}_N = \mathbf{D}_N \boldsymbol{\Delta}_N$, i.e., $\tilde{u}_{n,N} - u_{n,N} = \mathbf{d}_{n.,N} \boldsymbol{\Delta}_N$, for $n = 1,...,NT$, where $\mathbf{D}_N$ is an $NT \times P$ matrix, the $1 \times P$ vector $\mathbf{d}_{n.,N}$ denotes the n-th row of $\mathbf{D}_N$ and $\boldsymbol{\Delta}_N$ is a $P \times 1$ vector. Let $d_{nj,N}$ be the j-th element of $\mathbf{d}_{n.,N}$. We assume that $E|d_{nj,N}(t)|^{2+\delta} \leq c_d < \infty$ for some $\delta > 0$, where c_d does not depend on N, and that $N^{1/2} \|\boldsymbol{\Delta}_N\| = O_p(1)$.

Assumption 4 will be fulfilled in many settings, e.g., if model (1a) contains endogenous variables (such as spatial lags of $\mathbf{y}_N$) and is estimated by fixed or random effects two-stage least squares. In that case, $\boldsymbol{\Delta}_N$ denotes the difference between the parameter estimates and the true parameter values and $\mathbf{d}_{n.,N}$ is the (negative of the) n-th row of the design matrix $\mathbf{Z}_N$ under random effects or of the within transformed design matrix $\underline{\mathbf{Z}}_N = \mathbf{Q}_{0,N} \mathbf{Z}_N$ under fixed effects (see subsection 2 of Section IV). Under certain conditions, Assumption 4 will also be satisfied if model (1a) involves a non-linear specification (see Kelejian and Prucha, 2010). Finally, Assumption 4 implies that $(NT)^{-1} \sum_{n=1}^{NT} \|\mathbf{d}_{n.,N}\|^{2+\delta}$ is $O_p(1)$.

Assumption 5.
(a) The smallest eigenvalues of $\boldsymbol{\Gamma}_N' \boldsymbol{\Gamma}_N$ are bounded uniformly away from zero, i.e., $\lambda_{\min}(\boldsymbol{\Gamma}_N' \boldsymbol{\Gamma}_N) \geq \lambda_* > 0$. (b) $\tilde{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta}_N = o_p(1)$, where $\boldsymbol{\Theta}_N$ are $[4S + S(S-1)] \times [4S + S(S-1)]$ non-stochastic, symmetric, positive definite matrices. (c) The largest eigenvalues of $\boldsymbol{\Theta}_N$ are bounded uniformly from above, i.e., $\lambda_{\max}(\boldsymbol{\Theta}_N) \leq \bar{\lambda}_* < \infty$ and the smallest eigenvalues of $\boldsymbol{\Theta}_N$ are bounded uniformly away from zero, i.e., $\lambda_{\min}(\boldsymbol{\Theta}_N) \geq \underline{\lambda}_* > 0$.

Assumption 5 implies that the smallest eigenvalues of $\boldsymbol{\Gamma}_N' \boldsymbol{\Theta}_N \boldsymbol{\Gamma}_N$ are bounded uniformly away from zero, ensuring that the true parameter vector $\boldsymbol{\theta}_N$ is identifiable unique. Moreover, by the equivalence of matrix norms, it follows from Assumption 5 that $\boldsymbol{\Theta}_N$ and $\boldsymbol{\Theta}_N^{-1}$ are $O(1)$.

Assumptions 1-5 ensure consistency of the GM estimators for $\boldsymbol{\theta}_N = (\boldsymbol{\rho}_N', \sigma_\mu^2)$. We summarize these results in the following theorem, which is proven in Appendix B.

Theorem 1. Consistency of Weighted GM Estimator $\tilde{\boldsymbol{\theta}}_N$
Suppose Assumptions 1-5 hold. Then, provided the optimization space contains the parameter space, the weighted GM estimators $\tilde{\boldsymbol{\theta}}_N(\tilde{\boldsymbol{\Theta}}_N) = [\tilde{\rho}_{1,N}(\tilde{\boldsymbol{\Theta}}_N),...,\tilde{\rho}_{S,N}(\tilde{\boldsymbol{\Theta}}_N), \tilde{\sigma}_{\mu,N}^2(\tilde{\boldsymbol{\Theta}}_N)]'$ defined by (18) are consistent for $\rho_{1,N},...,\rho_{S,N}$ and σ_μ^2, i.e.,

$$\tilde{\rho}_{s,N}(\tilde{\boldsymbol{\Theta}}_N) - \rho_{s,N} \xrightarrow{p} 0, \ s = 1,...,S, \text{ and } \tilde{\sigma}_{\mu,N}^2(\tilde{\boldsymbol{\Theta}}_N) - \sigma_\mu^2 \xrightarrow{p} 0 \text{ as } N \to \infty.$$

This result holds for an arbitrary weighting matrix (that satisfies Assumption 5). Hence, it applies to both the optimally weighted GM estimator defined by (18) with $\tilde{\boldsymbol{\Theta}}_N = \tilde{\boldsymbol{\Psi}}_N^{-1}$ or the initial unweighted GM estimator with $\tilde{\boldsymbol{\Theta}}_N = \mathbf{I}_N$.

3.2 Asymptotic Distribution of GM Estimator for $\boldsymbol{\theta}_N$

In the following we consider the asymptotic distribution of the optimally weighted GM estimator $\tilde{\boldsymbol{\theta}}_N$. To establish asymptotic normality of $\tilde{\boldsymbol{\theta}}_N = (\tilde{\boldsymbol{\rho}}_N, \tilde{\sigma}^2_{\mu,N})$, we need some additional assumptions.

Assumption 6.

Let $\mathbf{D}_N$ be defined as in Assumption 4, such that $\tilde{\mathbf{u}}_N - \mathbf{u}_N = \mathbf{D}_N \boldsymbol{\Delta}_N$. For any real $NT \times NT$ matrix $\mathbf{A}_N$, whose row and column sums are bounded uniformly in absolute value, it holds that $N^{-1} \mathbf{D}_N' \mathbf{A}_N \mathbf{u}_N - N^{-1} E(\mathbf{D}_N' \mathbf{A}_N \mathbf{u}_N) = o_p(1)$.

A sufficient condition for Assumption 6 is, e.g., that the columns of $\mathbf{D}_N$ are of the form $\boldsymbol{\pi}_N + \boldsymbol{\Pi}_N \boldsymbol{\varepsilon}_N$, where the elements of $\boldsymbol{\pi}_N$ are bounded uniformly in absolute value and the row and column sums of $\boldsymbol{\Pi}_N$ are bounded uniformly in absolute value (see Remark A.1 in the Appendix). This will be the case in many applications, e.g., for model (1a), when $\mathbf{D}_N$ equals (the negative of) the design matrix $\mathbf{Z}_N$ or the within-transformed design matrix $\underline{\mathbf{Z}}_N$ (compare subsection 2 of Section IV).

Assumption 7.

Let $\boldsymbol{\Delta}_N$ be defined as in Assumption 4. Then,

$(NT)^{1/2} \boldsymbol{\Delta}_N = (NT)^{-1/2} \mathbf{T}_N' \boldsymbol{\xi}_N + o_p(1)$, with $\mathbf{T}_N = (\mathbf{T}_{\mathbf{v},N}', \mathbf{T}_{\boldsymbol{\mu},N}')'$, $\boldsymbol{\xi}_N = (\mathbf{v}_N', \boldsymbol{\mu}_N')'$, i.e.,

$(NT)^{1/2} \boldsymbol{\Delta}_N = (NT)^{-1/2} \mathbf{T}_{\mathbf{v},N}' \mathbf{v}_N + (NT)^{-1/2} \mathbf{T}_{\boldsymbol{\mu},N}' \boldsymbol{\mu}_N + o_p(1)$,

where $\mathbf{T}_N$ is an $(NT + N) \times P$-dimensional real non-stochastic matrix whose elements are bounded uniformly in absolute value; its submatrices $\mathbf{T}_{\mathbf{v},N}$ and $\mathbf{T}_{\boldsymbol{\mu},N}$ are of dimension $(NT \times P)$ and $(N \times P)$, respectively. As remarked above, $\boldsymbol{\Delta}_N$ typically denotes the difference between the parameter estimates and the true parameter values. Assumption 7 is kept general and will be satisfied by many estimators, which differ in the definition of $\mathbf{T}_N$. In Section IV, we verify that it holds if the model in equation (1a) is estimated by (random or fixed effects) two-stage least squares (TSLS) or feasible spatial generalized TSLS.

In Appendix B, the limiting distribution of the GM estimator of $\boldsymbol{\theta}_N$ is shown to depend on (the inverse of) the matrix $\mathbf{J}_N' \boldsymbol{\Theta}_N \mathbf{J}_N$ and the variance-covariance matrix of a vector of

quadratic forms in $\mathbf{v}_N$ and $\boldsymbol{\mu}_N$, denoted as $\mathfrak{q}_N$. We consider each of these expressions in the following.

The $[4S + S(S-1)] \times (S+1)$ matrix $\mathbf{J}_N$ of derivatives of the $[4S + S(S-1)] \times 1$ vector of moment conditions in (11) is given by

$$\mathbf{J}_N(\boldsymbol{\theta}_N) = \frac{\partial(\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N \mathbf{b}_N)}{\partial \boldsymbol{\theta}'} = (j_{i,1,N}, ..., j_{i,S,N}, j_{i,S+1,N}), \text{ with} \tag{19a}$$

$$j_{i,s,N} = \frac{\partial(\boldsymbol{\gamma}_{i.,N} - \boldsymbol{\Gamma}_{i.,N} \mathbf{b}_N)}{\partial \rho_s}, \ i = 1, ..., 4S + S(S-1), \ s = 1, ..., S,$$

$$j_{i,S+1,N} = \frac{\partial(\boldsymbol{\gamma}_{i.,N} - \boldsymbol{\Gamma}_{i.,N} \mathbf{b}_N)}{\partial \sigma_\mu}, \ i = 1, ..., 4S + S(S-1),$$

where $\boldsymbol{\gamma}_{i.,N}$ and $\boldsymbol{\Gamma}_{i.,N}$ denote the i-th row of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ respectively.

Using $\frac{\partial \boldsymbol{\gamma}_N}{\partial \boldsymbol{\theta}'} = \mathbf{0}$ and ignoring the negative sign, we have

$$\mathbf{J}_N(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}'} \boldsymbol{\Gamma}_N \mathbf{b}_N = \boldsymbol{\Gamma}_N \mathfrak{B}_N, \tag{19b}$$

where $\boldsymbol{\Gamma}_N$ is defined above and of dimension $[4S + S(S-1)] \times [2S + S(S-1)/2 + 1]$ and $\mathfrak{B}_N$ is a $[2S + S(S-1)/2 + 1] \times (S+1)$ matrix of the form

$$\mathfrak{B}_N = (\mathfrak{B}_1, \mathfrak{B}'_{2,N}, \mathfrak{B}'_{3,N}, \mathfrak{B}'_{4,N})', \tag{20a}$$

with $\mathfrak{B}_1 = (\mathbf{I}_S, \mathbf{0}_{S\times1})$ and $\mathfrak{B}_{2,N} = [diag_{s=1}^{S}(2\rho_{s,N}), \mathbf{0}_{S\times1})]$. The $S(S-1)/2 \times (S+1)$ matrix $\mathfrak{B}_{3,N} = [(\mathfrak{B}'_{3,1,N}, ..., \mathfrak{B}'_{3,S-1,N})', \mathbf{0}_{S(S-1)/2\times1}]$ consists of $(S-1)$ vertically arranged blocks $\mathfrak{B}_{3,m,N}$, $m = 1, ..., (S-1)$, which have the following structure:

$$\mathfrak{B}_{3,m,N} = (\mathfrak{C}_{m,N}, \mathfrak{d}_{m,N}, \mathfrak{E}_{m,N}), \tag{20b}$$

where $\mathfrak{C}_{m,N}$ is a $(S-m)\times(m-1)$ matrix of zeros,[12] $\mathfrak{d}_{m,N}$ is a $(S-m)\times1$ vector, defined as $\mathfrak{d}_{m,N} = (\rho_{m+1,N}, ..., \rho_{S,N})'$, and $\mathfrak{E}_{m,N} = \rho_{m,N}\mathbf{I}_{S-m}$. Finally, $\mathfrak{B}_{4,N}$ is a $1\times(S+1)$ vector, defined as

---

[12] I.e., there is no block $\mathfrak{C}_{1,N}$ in $\mathfrak{B}_{3,1,N}$.

$$\mathfrak{B}_{4,N} = (\mathbf{0}_{1 \times S}, 1). \tag{20c}$$

For later reference, note that $\mathfrak{B}_N$ has full column rank $(S+1)$; as a consequence, the $(S+1) \times (S+1)$ matrix $\mathfrak{B}'_N \mathfrak{B}_N$ is positive definite (see, e.g., Greene, 2003, p. 835).

We next consider the vector $\mathbf{q}_N$ and its limiting distribution. First, define $\mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N)$ as the $[4S + S(S-1)] \times 1$ vector of sample moments with the expectation operator suppressed, evaluated at the true parameter values, and ignoring the deterministic constants. It is made up of the following quadratic forms in $\tilde{\mathbf{u}}_N$:

$$\mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N) = N^{-1}(\tilde{\mathbf{u}}'_N \mathbf{C}^{s,s'}_{c,N} \tilde{\mathbf{u}}_N) \text{ for } c = 1,...,4 \text{ and } s, s' = 1,...,S. \tag{21}$$

Hence, each element of this vector corresponds to a particular moment condition, indexed by c, each of which is associated with a particular weights matrix $\mathbf{M}_{s,N}$ through (12b) and (12d) for moment conditions M^s_2 and M^s_4, or through (12a) and (12c) with a pair of weights matrices $\mathbf{M}_{s,N}$ and $\mathbf{M}_{s',N}$ for moment conditions $\mathrm{M}^{s,s'}_1$ and $\mathrm{M}^{s,s'}_3$. The arrangement of the elements is the same as in equation system (13).

In light of (12), the matrices $\mathbf{C}^{s,s'}_{c,N}$, $c = 1,...,4$, and $s, s' = 1,...,S$, are defined as follows:

$$\mathbf{C}^{s,s'}_{1,N} = \frac{1}{2(T-1)} \mathbf{R}'_N [\mathbf{A}^{s,s'}_{1,N} + (\mathbf{A}^{s,s'}_{1,N})'] \mathbf{R}_N, \tag{22}$$

$$\mathbf{C}^{s}_{2,N} = \frac{1}{2(T-1)} \mathbf{R}'_N [\mathbf{A}^{s}_{2,N} + (\mathbf{A}^{s}_{2,N})'] \mathbf{R}_N,$$

$$\mathbf{C}^{s,s'}_{3,N} = \frac{1}{2} \mathbf{R}'_N [\mathbf{A}^{s,s'}_{3,N} + (\mathbf{A}^{s,s'}_{3,N})'] \mathbf{R}_N,$$

$$\mathbf{C}^{s}_{4,N} = \frac{1}{2} \mathbf{R}'_N [\mathbf{A}^{s}_{4,N} + (\mathbf{A}^{s}_{4,N})'] \mathbf{R}_N,$$

where we have used the definition $\mathbf{R}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})]$.

By Assumption 3 and Remark A.1 in Appendix A, the row and column sums of the symmetric $NT \times NT$ matrices $\mathbf{C}^{s,s}_{c,N}$, $c = 1,...,4$, and $s, s' = 1,...,S$, are bounded uniformly in absolute value. Using equation (21) and invoking Lemma B.1 (see Appendix B), the elements of $N^{1/2} \mathbf{q}_N(\boldsymbol{\rho}_N, \boldsymbol{\Delta}_N)$ can be expressed as

$$N^{-1/2}(\tilde{\mathbf{u}}'_N \mathbf{C}^{s,s'}_{c,N} \tilde{\mathbf{u}}_N) = N^{-1/2}(\mathbf{u}'_N \mathbf{C}^{s,s'}_{c,N} \mathbf{u}_N) + (\boldsymbol{\alpha}^{s,s'}_{c,N})' N^{1/2} \boldsymbol{\Delta}_N + o_p(1) \tag{23}$$

with $\alpha_{c,N}^{s,s'} = N^{-1}E[\mathbf{D}_N'(\mathbf{C}_{c,N}^{s,s'} + \mathbf{C}_{c,N}^{s,s'}{}')\mathbf{u}_N] = 2N^{-1}E(\mathbf{D}_N'\mathbf{C}_{c,N}^{s,s'}\mathbf{u}_N)$ since $\mathbf{C}_{c,N}^{s,s'}$ is symmetric. By Lemma B.1 the elements of the $P \times 1$ vectors $\alpha_{c,N}^{s,s'}$, $c = 1,...,4$, and $s,s' = 1,...,S$, are bounded uniformly in absolute value. As evident from (23), $\alpha_{c,N}^{s,s'} = \mathbf{0}$ when $E(\mathbf{u}_N|\mathbf{D}_N) = \mathbf{0}$, which is the case under random effects estimation if there are no endogenous variables.

Note that $\mathbf{Q}_{0,N}\boldsymbol{\varepsilon}_N = \mathbf{Q}_{0,N}\mathbf{v}_N$ and that for symmetric $N \times N$ matrices $\mathbf{A}_N$, we have $\boldsymbol{\varepsilon}_N'\mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{A}_N)\mathbf{Q}_{1,N}\boldsymbol{\varepsilon}_N = T\boldsymbol{\mu}_N'\mathbf{A}_N\boldsymbol{\mu}_N + \mathbf{v}_N'\mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{A}_N)\mathbf{Q}_{1,N}\mathbf{v}_N + 2\mathbf{v}_N'(\mathbf{e}_T \otimes \mathbf{A}_N)\boldsymbol{\mu}_N$. Using (22), (23), and Assumption 7 we can rewrite the vector of sample moments as

$$N^{1/2}\mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N) = N^{-1/2}\overset{*}{\mathbf{q}}_N + o_p(1) = \mathbf{q}_N + o_p(1), \tag{24}$$

where each element of the $[4S + S(S-1)] \times 1$ vector $\overset{*}{\mathbf{q}}_{c,N} = (\overset{*}{q}_{c,N}^{s,s'})$ can be written as linear quadratic form of the $(NT + N) \times 1$ vector $\boldsymbol{\xi}_N = (\mathbf{v}_N', \boldsymbol{\mu}_N')'$:

$$\begin{aligned}
\overset{*}{q}_{c,N}^{s,s'} &= [\boldsymbol{\xi}_N'\overline{\mathbf{A}}_{c,N}^{s,s'}\boldsymbol{\xi}_N + (\overline{\mathbf{a}}_{c,N}^{s,s'})'\boldsymbol{\xi}_N] + o_p(1) \\
&= [\boldsymbol{\xi}_N'\overline{\mathbf{A}}_{c,N}^{s,s'}\boldsymbol{\xi}_N + (\mathbf{a}_{c,\mathbf{v},N}^{s,s'})'\mathbf{v}_N + (\mathbf{a}_{c,\boldsymbol{\mu},N}^{s,s'})'\boldsymbol{\mu}_N] + o_p(1),
\end{aligned} \tag{25}$$

where

$$\overline{\mathbf{A}}_{c,N}^{s,s'} = \begin{bmatrix} \mathbf{A}_{c,\mathbf{v},N}^{s,s'} & \mathbf{A}_{c,\mathbf{v},\boldsymbol{\mu},N}^{s,s'} \\ (\mathbf{A}_{c,\mathbf{v},\boldsymbol{\mu},N}^{s,s'})' & \mathbf{A}_{c,\boldsymbol{\mu},N}^{s,s'} \end{bmatrix},$$

$\overline{\mathbf{a}}_{c,N}^{s,s'} = T^{-1}\mathbf{T}_N\boldsymbol{\alpha}_{c,N}^{s,s'}$, $c = 1,...,4$, $s,s' = 1,...,S$, or

$\overline{\mathbf{a}}_{c,N}^{s,s'} = [(\mathbf{a}_{c,\mathbf{v},N}^{s,s'})', (\mathbf{a}_{c,\boldsymbol{\mu},N}^{s,s'})']' = T^{-1}[(\mathbf{T}_{\mathbf{v},N}\boldsymbol{\alpha}_{c,N}^{s,s'})', (\mathbf{T}_{\boldsymbol{\mu},N}\boldsymbol{\alpha}_{c,N}^{s,s'})']'$, for $c = 1,...,4$, and $s,s' = 1,...,S$.

Observe that the elements of $\overline{\mathbf{a}}_{c,N}^{s,s'}$, $c = 1,...,4$, and $s,s' = 1,...,S$, are bounded uniformly in absolute value by Assumption 7 and Lemma B.1. The symmetric matrices $\overline{\mathbf{A}}_{c,N}^{s,s'}$, $\mathbf{A}_{c,\mathbf{v},N}^{s,s'}$, $\mathbf{A}_{c,\mathbf{v},\boldsymbol{\mu},N}^{s,s'}$, and $\mathbf{A}_{c,\boldsymbol{\mu},N}^{s,s'}$ are of dimension $(NT + N) \times (NT + N)$, $NT \times NT$, $NT \times N$, and $N \times N$, respectively, and defined as follows.

For moment condition $M_1^{s,s'}$, we have

$$\mathbf{A}_{1,\mathbf{v},N}^{s,s'} = \frac{1}{2(T-1)}[\mathbf{A}_{1,N}^{s,s'} + (\mathbf{A}_{1,N}^{s,s'})'], \quad \mathbf{A}_{1,\mathbf{v},\boldsymbol{\mu},N}^{s,s'} = \mathbf{0}_{NT \times N}, \text{ and } \mathbf{A}_{1,\boldsymbol{\mu},N}^{s,s'} = \mathbf{0}_{N \times N}. \tag{26a}$$

For moment condition M_2^s we have

$$\mathbf{A}_{2,\mathbf{v},N}^{s} = \frac{1}{2(T-1)}[\mathbf{A}_{2,N}^{s} + (\mathbf{A}_{2,N}^{s})'], \ \mathbf{A}_{2,\mathbf{v},\mathbf{\mu},N}^{s} = \mathbf{0}_{NT \times N}, \text{ and } \mathbf{A}_{2,\mathbf{\mu},N}^{s} = \mathbf{0}_{N \times N}. \tag{26b}$$

For moment condition $\mathrm{M}_3^{s,s'}$ we have

$$\mathbf{A}_{3,\mathbf{v},N}^{s,s'} = \frac{1}{2}[\mathbf{A}_{3,N}^{s} + (\mathbf{A}_{3,N}^{s})'], \ \mathbf{A}_{3,\mathbf{v},\mathbf{\mu},N}^{s,s'} = \frac{1}{2}[\mathbf{e}_{T} \otimes (\mathbf{M}_{s',N}'\mathbf{M}_{s,N} + \mathbf{M}_{s,N}'\mathbf{M}_{s',N})], \text{ and}$$

$$\mathbf{A}_{3,\mathbf{\mu},N}^{s,s'} = \frac{T}{2}(\mathbf{M}_{s',N}'\mathbf{M}_{s,N} + \mathbf{M}_{s,N}'\mathbf{M}_{s',N}). \tag{26c}$$

For moment condition M_4^{s}, we have

$$\mathbf{A}_{4,\mathbf{v},N}^{s} = \frac{1}{2}[\mathbf{A}_{4,N}^{s} + (\mathbf{A}_{4,N}^{s})'], \ \mathbf{A}_{4,\mathbf{v},\mathbf{\mu},N}^{s} = \frac{1}{2}[\mathbf{e}_{T} \otimes (\mathbf{M}_{s,N} + \mathbf{M}_{s,N}')], \text{ and}$$

$$\mathbf{A}_{4,\mathbf{\mu},N}^{s,s'} = \frac{T}{2}(\mathbf{M}_{s,N} + \mathbf{M}_{s,N}'). \tag{26d}$$

Note that the row and column sums of the symmetric matrices $\mathbf{A}_{c,N}^{s,s'}$, $\mathbf{A}_{c,\mathbf{v},N}^{s,s'}$, $\mathbf{A}_{c,\mathbf{v},\mathbf{\mu},N}^{s,s'}$, and $\mathbf{A}_{c,\mathbf{\mu},N}^{s,s'}$ are bounded uniformly in absolute value by Assumption 3 and Remark A.1 in the Appendix. Moreover, the elements of the $\mathbf{\xi}_{N} = (\mathbf{v}_{N}', \mathbf{\mu}_{N}')'$ are independently distributed by Assumption 1, and the variance-covariance matrix of $\mathbf{\xi}_{N}$ is

$$\mathbf{\Omega}_{\xi,N} = \begin{bmatrix} \mathbf{\Sigma}_{N} & \mathbf{0}_{NT \times N} \\ \mathbf{0}_{N \times NT} & \sigma_{\mu}^{2}\mathbf{I}_{N} \end{bmatrix}. \tag{27}$$

In order to calculate the variance-covariance matrix of $\mathbf{q}_{N}$, given by the $[4S + S(S-1)] \times [4S + S(S-1)]$ matrix $\mathbf{\Psi}_{N} = N^{-1}E(\mathbf{\mathring{q}}_{N}\mathbf{\mathring{q}}_{N}')$, we invoke Lemma A.1 in Kelejian and Prucha (2010). For the moment, assume that the error components $\mathbf{\mu}_{N}$ and $\mathbf{v}_{N}$ are normally distributed.[13] The distribution of the GM estimates without distributional assumptions (apart from Assumption 1) is considered in the Appendix. Under normality, the covariance between two elements of the vector $\mathbf{q}_{N}$ is given by:

$$\mathfrak{C}_{c,c',N}^{s,s';t,t'} = N^{-1}Cov(\mathring{q}_{c,N}^{s,s'}, \mathring{q}_{c',N}^{t,t'}) \tag{28a}$$

$$= N^{-1}Cov[\mathbf{\xi}_{N}'\overline{\mathbf{A}}_{c,N}^{s,s'}\mathbf{\xi}_{N} + (\overline{\mathbf{a}}_{c,N}^{s,s'})'\mathbf{\xi}_{N}, \mathbf{\xi}_{N}'\overline{\mathbf{A}}_{c',N}^{t,t'}\mathbf{\xi}_{N} + (\overline{\mathbf{a}}_{c',N}^{t,t'})'\mathbf{\xi}_{N}]$$

$$= N^{-1}Cov[\mathbf{\xi}_{N}'\overline{\mathbf{A}}_{c,N}^{s,s'}\mathbf{\xi}_{N} + (\mathbf{a}_{c,\mathbf{v},N}^{s,s'})'\mathbf{v}_{N} + (\mathbf{a}_{c,\mathbf{\mu},N}^{s,s'})'\mathbf{\mu}_{N}, \mathbf{\xi}_{N}'\overline{\mathbf{A}}_{c',N}^{t,t'}\mathbf{\xi}_{N} + (\mathbf{a}_{c',\mathbf{v},N}^{t,t'})'\mathbf{v}_{N} + (\mathbf{a}_{c',\mathbf{\mu},N}^{t,t'})'\mathbf{\mu}_{N}]$$

---

[13] In that case, in Assumption 1, the requirement of finite $4 + \eta$-th moments of the error components can be relaxed to the requirement of finite variances.

$$= 2N^{-1}Tr(\mathbf{A}^{s,s'}_{c,\mathbf{v},N}\boldsymbol{\Sigma}_N\mathbf{A}^{t,t'}_{c',\mathbf{v},N}\boldsymbol{\Sigma}_N + 2\sigma^2_\mu(\mathbf{A}^{s,s'}_{c,\mathbf{v},\boldsymbol{\mu},N})'\boldsymbol{\Sigma}_N\mathbf{A}^{t,t'}_{c',\mathbf{v},\boldsymbol{\mu},N} + \sigma^4_\mu\mathbf{A}^{s,s'}_{c,\boldsymbol{\mu},N}\mathbf{A}^{t,t'}_{c',\boldsymbol{\mu},N}) \; .$$

$$+ N^{-1}[(\mathbf{a}^{s,s'}_{c,\mathbf{v},N})'\boldsymbol{\Sigma}_N\mathbf{a}^{t,t'}_{c',\mathbf{v},N} + \sigma^2_\mu(\mathbf{a}^{s,s'}_{c,\boldsymbol{\mu},N})'\mathbf{a}^{t,t'}_{c',\boldsymbol{\mu},N}],$$

with $c,c' = 1,...,4$, $s,t = 1,...,S$ for $s = s'$ and $t = t'$, and $s = 1,...,S-1$, $s' > s$. Note that the each combination of indices c, s, s' (and also c', t, t') is associated with a particular row of $\mathbf{q}_N$. Hence, $\mathfrak{E}^{s,s';t,t'}_{c,c',N}$ is the covariance between the element of $\mathbf{q}_N$ associated with moment condition $\mathrm{M}^{s,s'}_c$ and the element of $\mathbf{q}_N$ associated with moment condition $\mathrm{M}^{t,t'}_{c'}$. (For the second and fourth moment condition we always have $s = s'$ and $t = t'$).

In equation (28), $a^{s,s'}_{c,\mathbf{v},nn,N}$ and $a^{s,s'}_{c,\boldsymbol{\mu},ii,N}$ denote the n-th and i-th main diagonal element of the matrices $\mathbf{A}^{s,s'}_{c,\mathbf{v},N}$ and $\mathbf{A}^{s,s'}_{c,\boldsymbol{\mu},N}$, respectively, and $a^{s,s'}_{c,\mathbf{v},n,N}$ and $a^{s,s'}_{c,\boldsymbol{\mu},i,N}$ denote the n-th and i-th element of the vectors $\mathbf{a}^{s,s'}_{c,\mathbf{v},N}$ and $\mathbf{a}^{s,s'}_{c,\boldsymbol{\mu},N}$ respectively.

The arrangement of the elements $\boldsymbol{\Psi}_N = (\psi_{i,j,N})$, $i = 1,...,[4S + S(S-1)]$, $j = 1,...,[4S + S(S-1)]$ is straightforward and follows naturally from the ordering of the elements in the vector $\mathbf{q}_N$, though it is notationally burdensome to state in the general case.

The expression in (28) holds generally. Part of the elements of $\boldsymbol{\Psi}_N$ can be stated in simpler terms: in particular, the submatrices $\mathbf{A}^{s,s'}_{c,\boldsymbol{\mu},N}$, are zero for $c = 1$ and $c = 2$ such that $\mathfrak{E}^*_{\boldsymbol{\mu},N}$ drops out for the respective elements. If both sub-matrices associated with $\mu_{it,N}$ are zero ($c = 1$ or $c = 2$ and $c' = 1$ or $c' = 2$), $\mathfrak{E}^{**}_{\boldsymbol{\mu},N}$ drops out as well. Under fixed effects estimation, the terms $\mathfrak{E}^{**}_{\boldsymbol{\mu},N}$ (the expressions involving $\mathbf{a}^{s,s'}_{c,\boldsymbol{\mu},N}$) are equal to zero. Finally, since the main diagonal elements of the matrices $\mathbf{A}^s_{2,N}$ and $\mathbf{A}^s_{4,N}$ are zero, the term $\mathfrak{E}^{*,\bullet}_{\mathbf{v},N}$ does not show up for elements where $c = 2$ or $c = 4$ (or where $c' = 2$ or $c' = 4$).

To derive the asymptotic distribution of $\mathbf{q}_N$ and $\tilde{\boldsymbol{\theta}}_N$ we invoke the central limit theorem for vectors of linear quadratic forms given by Kelejian and Prucha (2010, Theorem A.1) and Corollary F4 in Pötscher and Prucha (1997). We summarize the results regarding the asymptotic distribution of $\tilde{\boldsymbol{\theta}}_N$ in the following Theorem, which is proven in Appendix B.

Theorem 2. Asymptotic Normality of $\tilde{\boldsymbol{\theta}}_N$

Let $\tilde{\boldsymbol{\theta}}_N$ be the GM estimator defined by (18). Suppose Assumptions 1-7 hold and, furthermore, that $\lambda_{\min}(\boldsymbol{\Psi}_N) \geq c^*_{\boldsymbol{\Psi}} > 0$. Then, provided the optimization space contains the parameter space, we have

$$N^{1/2}(\widetilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N) = (\mathbf{J}_N'\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}_N'\boldsymbol{\Theta}_N\boldsymbol{\Psi}_N^{1/2}\boldsymbol{\xi}_N + o_p(1)\,, \text{ with}$$

$$\mathbf{J}_N = \frac{\partial}{\partial\boldsymbol{\theta}'}\boldsymbol{\Gamma}_N\mathbf{b}_N = \boldsymbol{\Gamma}_N\boldsymbol{\mathfrak{B}}_N\,, \text{ and}$$

$$\boldsymbol{\xi}_N = \boldsymbol{\Psi}_N^{-1/2}\mathbf{q}_N \xrightarrow{d} N(0,\mathbf{I}_{4S+S(S-1)})\,,$$

where $\boldsymbol{\Psi}_N = E(\mathbf{q}_N\mathbf{q}_N')$ and $\boldsymbol{\Psi}_N = (\boldsymbol{\Psi}_N^{1/2})(\boldsymbol{\Psi}_N^{1/2})'$.

Furthermore $N^{1/2}(\widetilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N) = O_p(1)$ and

$$\boldsymbol{\Omega}_{\widetilde{\boldsymbol{\theta}}_N}(\boldsymbol{\Theta}_N) = (\mathbf{J}_N'\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}_N'\boldsymbol{\Theta}_N\boldsymbol{\Psi}_N\boldsymbol{\Theta}_N\mathbf{J}_N(\mathbf{J}_N'\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1}\,,$$

where $\boldsymbol{\Omega}_{\widetilde{\boldsymbol{\theta}}_N}$ is positive definite.

Theorem 2 implies that the difference between the cumulative distribution function of $N^{1/2}(\widetilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$ and that of $N(0,\boldsymbol{\Omega}_{\widetilde{\boldsymbol{\theta}}_N})$ converges pointwise to zero, which justifies the use of the latter as an approximation of the former.[14] Theorem 2 holds both under normality and non-normality of the error components, the difference being only the definition of the elements of $\boldsymbol{\Psi}_N$ (and the requirement regarding the finiteness of the moments of the error components in Assumption 1).

Note that $\boldsymbol{\Omega}_{\widetilde{\boldsymbol{\theta}}_N}(\boldsymbol{\Psi}_N^{-1}) = (\mathbf{J}_N'\boldsymbol{\Psi}_N^{-1}\mathbf{J}_N)^{-1}$ and that $\boldsymbol{\Omega}_{\widetilde{\boldsymbol{\theta}}_N}(\boldsymbol{\Theta}_N) - \boldsymbol{\Omega}_{\widetilde{\boldsymbol{\theta}}_N}(\boldsymbol{\Psi}_N^{-1})$ is positive semidefinite. Thus, using a consistent estimator of $\boldsymbol{\Psi}_N^{-1}$ (which will be derived below) as weighting matrix $\boldsymbol{\Theta}_N$ leads to the efficient GM estimator. We add that $\boldsymbol{\Psi}_N$ is not exactly equal to the variance-covariance matrix of the moments, if there is an endogenous right-hand side variable in equation (1), since the GM estimates are based on estimated rather than the true disturbances. (See also the discussion surrounding equation (23)).

IV. Estimation of Regression Parameters $\boldsymbol{\delta}_N$ and Joint Asymptotic Distribution

In the following, we consider estimators for the regression parameters $\boldsymbol{\delta}_N$ in model (1a) and establish their joint asymptotic distribution with the GM estimates $\widetilde{\boldsymbol{\theta}}_N$ derived in section III. We keep the analysis general first, allowing us to state our results in a succinct way that nests both random and fixed effects estimation of the original model as well as the spatial GLS transformed model. We will then be more specific about the properties and the respective expressions for the TSLS and spatial generalized TSLS estimation of model (1a).

---

[14] Compare Corollary F4 in Pötscher and Prucha (1997).

1. General Statement of Estimator and Joint Asymptotic Distribution

Key to establishing the asymptotic properties of the GM estimates $\tilde{\boldsymbol{\theta}}_N$, which are based on the estimated disturbances of model (1a), is Assumption 7, which holds that the (properly normalized) difference between the true parameters and the estimates ($\boldsymbol{\Delta}_N$) is linear in the stacked vector of error terms, i.e., $(NT)^{1/2}\boldsymbol{\Delta}_N = (NT)^{-1/2}\mathbf{T}_N'\boldsymbol{\xi}_N + o_p(1)$.

For all estimators of $\boldsymbol{\delta}_N$ in model (1a) considered in the present paper, the matrix $\mathbf{T}_N$ has the following structure:

$$\mathbf{T}_N = \mathbf{F}_N\mathbf{P}_N \quad \text{with } \mathbf{F}_N = (\mathbf{F}_{\mathbf{v},N}', \mathbf{F}_{\boldsymbol{\mu},N}')', \tag{29a}$$

which can also be written as

$$\mathbf{T}_N = (\mathbf{T}_{\mathbf{v},N}', \mathbf{T}_{\boldsymbol{\mu},N}')' \quad \text{with } \mathbf{T}_{\mathbf{v},N} = \mathbf{F}_{\mathbf{v},N}\mathbf{P}_N, \ \mathbf{T}_{\boldsymbol{\mu},N} = \mathbf{F}_{\boldsymbol{\mu},N}\mathbf{P}_N, \tag{29b}$$

where $\mathbf{F}_{\mathbf{v},N}$ is a real non-stochastic $NT \times P_*$ matrix, $\mathbf{F}_{\boldsymbol{\mu},N}$ is a real non-stochastic $N \times P_*$ matrix, and $\mathbf{P}_N$ is a real non-stochastic $P_* \times P$ matrix, with P as in Assumption 7. The definition of $\mathbf{P}_N$, $\mathbf{F}_{\mathbf{v},N}$, $\mathbf{F}_{\boldsymbol{\mu},N}$ will be seen to depend on the estimated model (original versus spatial GLS transformed model) and the estimation approach (random versus fixed effects). In general, $\mathbf{P}_N$ is a function of the original or within-transformed design matrix $\mathbf{Z}_N$ and a real non-stochastic $NT \times P_*$ matrix of instruments $\mathbf{H}_N$, (or spatial GLS transformed variants thereof); $\mathbf{F}_{\mathbf{v},N}$ and $\mathbf{F}_{\boldsymbol{\mu},N}$ depend on the original or within-transformed instruments $\mathbf{H}_N$ (or spatial GLS transformed variants thereof), and – in the untransformed model – on the matrix $[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N}')^{-1}]$.

Since both $N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$ and $(NT)^{1/2}\boldsymbol{\Delta}_N$, and thus also $N^{1/2}\boldsymbol{\Delta}_N$ are asymptotically linear in $\boldsymbol{\xi}_N$, the joint distribution of the vector $[N^{1/2}\boldsymbol{\Delta}_N', N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)']'$ can be derived invoking the central limit theorem for vectors of quadratic forms by Kelejian and Prucha (2010).

Consider the $[(P_* + 4S + S(S-1)] \times 1$ vector of linear and linear quadratic forms in $\boldsymbol{\xi}_N$:

$$\mathbf{w}_N = \begin{bmatrix} (NT)^{-1/2}\mathbf{F}_N'\boldsymbol{\xi}_N \\ \mathbf{q}_N \end{bmatrix}. \tag{30}$$

Its variance-covariance matrix is of dimension $[P_* + 4S + S(S-1)] \times [(P_* + 4S + S(S-1)]$ and given by:

$$Var(\mathbf{w}_N) = \mathbf{\Psi}_{\mathbf{w},N} = E \begin{bmatrix} (NT)^{-1}\mathbf{F}'_N\boldsymbol{\xi}_N\boldsymbol{\xi}'_N\mathbf{F}_N & (NT)^{-1/2}\mathbf{F}'_N\boldsymbol{\xi}_N\mathbf{q}'_N \\ (NT)^{-1/2}\mathbf{q}_N\boldsymbol{\xi}'_N\mathbf{F}_N & \mathbf{q}_N\mathbf{q}'_N \end{bmatrix} = \begin{bmatrix} \mathbf{\Psi}_{\Delta\Delta,N} & \mathbf{\Psi}_{\Delta\theta,N} \\ \mathbf{\Psi}'_{\Delta\theta,N} & \mathbf{\Psi}_N \end{bmatrix}, \qquad (31)$$

where the $[4S + S(S-1)] \times [(4S + S(S-1)]$ matrix $\mathbf{\Psi}_N$ is defined above in (28).

The $P_* \times P_*$ matrix $\mathbf{\Psi}_{\Delta\Delta,N}$ is defined as

$$\mathbf{\Psi}_{\Delta\Delta,N} = \mathbf{\Psi}^{\mathbf{v}}_{\Delta\Delta,N} + \mathbf{\Psi}^{\boldsymbol{\mu}}_{\Delta\Delta,N}, \text{ with} \qquad (32a)$$

$$\mathbf{\Psi}^{\mathbf{v}}_{\Delta\Delta,N} = (NT)^{-1}(\mathbf{F}'_{\mathbf{v},N}\boldsymbol{\Sigma}_N\mathbf{F}_{\mathbf{v},N}) \text{ and } \mathbf{\Psi}^{\boldsymbol{\mu}}_{\Delta\Delta,N} = (NT)^{-1}\sigma^2_{\mu}\mathbf{F}'_{\boldsymbol{\mu},N}\mathbf{F}_{\boldsymbol{\mu},N}.$$

The $P_* \times [4S + S(S-1)]$ matrix $\mathbf{\Psi}_{\Delta\theta,N}$ is given by

$$\mathbf{\Psi}_{\Delta\theta,N} = E[(NT)^{-1/2}\mathbf{F}'_N\boldsymbol{\xi}_N\mathbf{q}'_N], \qquad (32b)$$

which is made up by $[4S + S(S-1)]$ columns of dimension $P_* \times 1$, each of them associated with a set of indices c, s, and s' and thus a particular moment condition. Under normality of $\boldsymbol{\mu}_N$ and $\mathbf{v}_N$, the columns are defined as

$$\boldsymbol{\psi}_{\Delta\theta,.,(c,s,s'),N} = \boldsymbol{\psi}^{\mathbf{v}}_{\Delta\theta,.,(c,s,s'),N} + \boldsymbol{\psi}^{\boldsymbol{\mu}}_{\Delta\theta,.,(c,s,s'),N}, \ c = 1,...,4, \ s,s' = 1,...,S, \text{ with} \qquad (32c)$$

$$\boldsymbol{\psi}^{\mathbf{v}}_{\Delta\theta,.,(c,s,s'),N} = \frac{1}{N}\frac{1}{T^{1/2}}(\mathbf{F}'_{\mathbf{v},N}\boldsymbol{\Sigma}_N\mathbf{a}^{s,s'}_{c,\mathbf{v},N}), \text{ and}$$

$$\boldsymbol{\psi}^{\boldsymbol{\mu}}_{\Delta\theta,.,(c,s,s'),N} = \frac{1}{N}\frac{1}{T^{1/2}}\sigma^2_{\mu}(\mathbf{F}'_{\boldsymbol{\mu},N}\mathbf{a}^{s,s'}_{c,\boldsymbol{\mu},N}).$$

In Appendix 1.2, $\mathbf{\Psi}_{\Delta\theta,N}$ is defined for the general case without distributional assumptions (apart from Assumption 1).

Regarding the joint limiting distribution of $N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$ and $(NT)^{1/2}\boldsymbol{\Delta}_N$, we now have the following result, which is proven in Appendix B.

Theorem 3. Joint Distribution of $\tilde{\boldsymbol{\theta}}_N$ and Regression Parameters

Suppose that Assumptions 1-7 hold. Moreover, assume also that $\mathbf{H}_N = O(1)$ (see Assumption 9 below) and that $\mathbf{F}_N = O(1)$; the latter assumption will be verified, once we have defined the

matrix $\mathbf{F}_N$ for the particular estimators used. Moreover, assume that $\lambda_{\min}(\boldsymbol{\Psi}_{\mathbf{w},N}) \geq c^*_{\boldsymbol{\Psi}_{\mathbf{w}}} > 0$. Then,

$$\begin{bmatrix} N^{1/2}\boldsymbol{\Delta}_N \\ N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N) \end{bmatrix} = \begin{bmatrix} T^{-1/2}\mathbf{P}'_N & \mathbf{0} \\ \mathbf{0} & (\mathbf{J}'_N\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}'_N\boldsymbol{\Theta}_N \end{bmatrix} \boldsymbol{\Psi}^{1/2}_{\mathbf{w},N}\boldsymbol{\xi}_{o,N} + o_p(1), \text{ with}$$

$$\boldsymbol{\xi}_{o,N} = \boldsymbol{\Psi}^{-1/2}_{\mathbf{w},N}\mathbf{w}_N = \boldsymbol{\Psi}^{-1/2}_{\mathbf{w},N}[(NT)^{-1/2}\boldsymbol{\xi}'_N\mathbf{F}_N, \mathbf{q}'_N]' \overset{d}{\to} N(\mathbf{0}, \mathbf{I}_{P^*+4S+S(S-1)}), \text{ and}$$

$$\boldsymbol{\Omega}_{\mathbf{w},N} = \begin{bmatrix} T^{-1/2}\mathbf{P}'_N & \mathbf{0} \\ \mathbf{0} & (\mathbf{J}'_N\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}'_N\boldsymbol{\Theta}_N \end{bmatrix} \boldsymbol{\Psi}_{\mathbf{w},N} \begin{bmatrix} T^{-1/2}\mathbf{P}_N & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta}_N\mathbf{J}_N(\mathbf{J}'_N\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1} \end{bmatrix}.$$

Theorem 3 implies that the difference between the joint cumulative distribution function of $[N^{1/2}\boldsymbol{\Delta}'_N, N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)']'$ and that of $N(\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{w},N})$ converges pointwise to zero, which justifies the use of the latter distribution as an approximation of the former.

Remark 2.
Theorem 3 holds under both normality and non-normality of the error components, the difference being the definition of the elements of $\boldsymbol{\Psi}_{\mathbf{w},N}$, in particular those of $\boldsymbol{\Psi}_N$ and $\boldsymbol{\Psi}_{\boldsymbol{\Delta\theta},N}$. Obviously, Theorem 3 can also be used to obtain the joint distribution of $N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$ and some other estimator $N^{1/2}\boldsymbol{\Delta}^\times_N$, where $(NT)^{1/2}\boldsymbol{\Delta}^\times_N = (NT)^{-1/2}\mathbf{T}^{\times\prime}_N\boldsymbol{\xi}_N + o_p(1)$, $\mathbf{T}^\times_N = \mathbf{F}^\times_N\mathbf{P}^\times_N$, assuming that analogous assumptions are maintained for this estimator. In particular, the results remain valid, but with $\mathbf{F}_N$ and $\mathbf{P}_N$ replaced by $\mathbf{F}^\times_N$ and $\mathbf{P}^\times_N$ in the definitions of $\boldsymbol{\Psi}_{\boldsymbol{\Delta\Delta},N}$ as well as $\boldsymbol{\Psi}_{\boldsymbol{\Delta\theta},N}$.

2. Two-Stage Least Squares (TSLS) and Spatial Generalized TSLS Estimation of $\boldsymbol{\delta}_N$

Obviously $E(\bar{\mathbf{Y}}_N\mathbf{u}'_N) \neq \mathbf{0}$ in model (1a). In the following we consider four TSLS estimators for $\boldsymbol{\delta}_N$. First, depending on whether $\boldsymbol{\pi} = \mathbf{0}$ or not in equation (9), we consider random effects or fixed effects estimation. Second, we consider (both fixed and random effects) estimation of the original model (1a) as well as of the spatial generalized LS transformed model, which is obtained by premultiplying model (1a) with the transformation matrix $\mathbf{R}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})]$. Regarding notation, we use an underbar to refer to within-transformed variables, e.g., $\underline{\mathbf{Z}}_N = \mathbf{Q}_{0,N}\mathbf{Z}_N$. Spatial generalized LS transformed variables are indicated by an asterix, e.g., $\mathbf{Z}^*_N = \mathbf{R}_N\mathbf{Z}_N$. Matrices and vectors that are both within- and spatial GLS transformed variables are indicated, accordingly, e.g., $\underline{\mathbf{Z}}^*_N = \mathbf{Q}_{0,N}\mathbf{Z}^*_N = \mathbf{Q}_{0,N}\mathbf{R}_N\mathbf{Z}_N$. By the properties of $\mathbf{Q}_{0,N}$, an equivalent way writing this is

$\mathbf{\underline{Z}}_N^* = \mathbf{R}_N \mathbf{\underline{Z}}_N = \mathbf{Q}_{0,N} \mathbf{R}_N \mathbf{Z}_N$, i.e., the order, in which the transformations are performed is immaterial.

2.1 Assumptions

Some properties of the regressor matrix $\mathbf{X}_N$ have already been discussed in subsection 3 of section II. The following further assumptions are maintained.

Assumption 8.

The non-stochastic instrument matrix $\mathbf{H}_N$ has full column rank $P_* \geq K + R$ (for N large enough). Furthermore, the elements of $\mathbf{H}_N$ are bounded uniformly in absolute value. Under fixed effects estimation, we also assume that each instrument changes over time (at least for some cross-section i). Moreover, it hold that $\mathbf{Q_{HH}} = \lim_{N \to \infty}[(NT)^{-1}\mathbf{H}_N'\mathbf{H}_N]$ and $\mathbf{Q_{HZ}} = \text{plim}_{N \to \infty}[(NT)^{-1}\mathbf{H}_N'\mathbf{Z}_N]$ are finite and non-singular.

Regarding the choice of instruments, note that

$$E(\sum_{r=1}^{R}\mathbf{W}_{r,N}\mathbf{y}_N) = \sum_{r=1}^{R}\mathbf{W}_{r,N}E(\mathbf{y}_N) = \sum_{r=1}^{R}\mathbf{W}_{r,N}E\{[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{r=1}^{R}\lambda_{r,N}\mathbf{W}_{r',N})^{-1}]\mathbf{X}_N\boldsymbol{\beta}_N\}$$
$$= \sum_{r=1}^{R}\mathbf{W}_{r,N}\{\mathbf{I}_T \otimes [\mathbf{I}_N + \sum_{i=1}^{\infty}(\sum_{r'=1}^{R}\lambda_{r',N}\mathbf{W}_{r',N})^i]\}\mathbf{X}_N\boldsymbol{\beta}_N ,$$

provided that $\left\|\sum_{r'=1}^{R}\lambda_{r',N}\mathbf{W}_{r',N}\right\| < 1$ for some matrix norm $\|\cdot\|$ (compare Horn and Johnson, 1985, p. 301). The instrument matrix $\mathbf{H}_N$ is used to instrument $\mathbf{Z}_N = (\mathbf{X}_N, \overline{\mathbf{Y}}_N)$ in a least squares regression of $\mathbf{Z}_N$ on $\mathbf{H}_N$, obtaining $\hat{\mathbf{Z}}_N = \mathbf{P}_{\mathbf{H}_N}\mathbf{Z}_N$, where $\mathbf{P}_{\mathbf{H}_N} = \mathbf{H}_N(\mathbf{H}_N'\mathbf{H}_N)^{-1}\mathbf{H}_N'$. It is thus reasonable to select $\mathbf{H}_N$ to include $\mathbf{X}_N$ and a subset of the linearly independent columns of terms of the sum $[\mathbf{I}_T \otimes \sum_{i=1}^{Q}(\sum_{r'=1}^{R}\mathbf{W}_{r',N})^i]\mathbf{X}_N$, where Q is some predefined constant.[15] Note that such a choice of $\mathbf{H}_N$ implies that the second part of Assumption 9 will be fulfilled (by Assumptions 3 and 8) and that $\mathbf{X}_N$ is projected on itself.

Analogous assumptions are maintained for the within-transformed regressor and instrument matrices $\mathbf{\underline{X}}_N$ and $\mathbf{\underline{H}}_N$. Assumption 8 then also hold for the spatial GLS transformed variables

---

[15] Kelejian, Prucha, and Yuzefovich (2004) consider the results using alternative sets of instruments in the estimation of a cross-section SARAR(1,1) model. Their Monte Carlo simulation results suggest that choosing $Q = 2$ will be sufficient in many applications.

$\mathbf{X}_N^*$ and $\mathbf{H}_N^*$ (under random effects estimation) or $\underline{\mathbf{X}}_N^*$ and $\underline{\mathbf{H}}_N^*$ (under fixed effects estimation).

2.2 Definition of TSLS Estimator and Asymptotic Results

2.2.1 Random Effects Estimation
The random effects TSLS estimator of model (1a) is defined as

$$\tilde{\boldsymbol{\delta}}_N = (\hat{\mathbf{Z}}_N' \mathbf{Z}_N)^{-1} \hat{\mathbf{Z}}_N' \mathbf{y}_N, \text{ where} \tag{33}$$

$$\hat{\mathbf{Z}}_N = \mathbf{P}_{\mathbf{H}_N} \mathbf{Z}_N = (\mathbf{X}_N, \hat{\overline{\mathbf{Y}}}_N), \text{ and}$$

$$\hat{\overline{\mathbf{Y}}}_N = \mathbf{P}_{\mathbf{H}_N} \overline{\mathbf{Y}}_N \text{ with } \mathbf{P}_{\mathbf{H}_N} = \mathbf{H}_N (\mathbf{H}_N' \mathbf{H}_N)^{-1} \mathbf{H}_N'.$$

As already mentioned, under random effects estimation, the **Z**-matrix typically includes a constant. The following lemma shows that the various assumptions maintained in Section III are automatically satisfied by the random effects TSLS estimator $\tilde{\boldsymbol{\delta}}_N$ and the corresponding residuals $\tilde{\mathbf{u}}_N = \mathbf{y}_n - \mathbf{Z}_N \tilde{\boldsymbol{\delta}}_N$, which are used in the GM estimation of the parameters $\rho_{s,N}$, $s = 1,...,S$, and σ_μ^2. A proof of Lemma 1 is given in Appendix B.

Lemma 1
Suppose that Assumptions 1-3 and 8 hold, and that $\sup_N \|\boldsymbol{\beta}_N\| \le b < \infty$. Let $\mathbf{D}_N = -\mathbf{Z}_N$, then, the fourth moments of the elements of $\mathbf{D}_N$ are bounded uniformly in absolute value, Assumption 6 holds, and

(a) $(NT)^{1/2} (\tilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N) = (NT)^{-1/2} \mathbf{T}_{\mathbf{v},N}' \mathbf{v}_N + (NT)^{-1/2} \mathbf{T}_{\boldsymbol{\mu},N}' \boldsymbol{\mu}_N + o_p(1)$, where

$$\mathbf{T}_{\mathbf{v},N} = \mathbf{F}_{\mathbf{v},N} \mathbf{P}_N, \quad \mathbf{T}_{\boldsymbol{\mu},N} = \mathbf{F}_{\boldsymbol{\mu},N} \mathbf{P}_N,$$

$$\mathbf{P}_N = \mathbf{Q}_{\mathbf{HH}}^{-1} \mathbf{Q}_{\mathbf{HZ}} (\mathbf{Q}_{\mathbf{HZ}}' \mathbf{Q}_{\mathbf{HH}}^{-1} \mathbf{Q}_{\mathbf{HZ}})^{-1},$$

$$\mathbf{F}_{\mathbf{v},N} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^S \rho_{m,N} \mathbf{M}_{m,N}')^{-1}] \mathbf{H}_N, \text{ and}$$

$$\mathbf{F}_{\boldsymbol{\mu},N} = (\mathbf{e}_T' \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^S \rho_{m,N} \mathbf{M}_{m,N}')^{-1}] \mathbf{H}_N.$$

(b) $(NT)^{-1/2} \mathbf{T}_{\mathbf{v},N}' \mathbf{v}_N + (NT)^{-1/2} \mathbf{T}_{\boldsymbol{\mu},N}' \boldsymbol{\mu}_N = O_p(1)$;

(c) $\mathbf{P}_N = O_p(1)$ and $\tilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$, with

$$\tilde{\mathbf{P}}_N = [(NT)^{-1} \mathbf{H}_N' \mathbf{H}_N]^{-1} [(NT)^{-1} \mathbf{H}_N' \mathbf{Z}_N] \{ [(NT)^{-1} \mathbf{Z}_N' \mathbf{H}_N] [(NT)^{-1} \mathbf{H}_N' \mathbf{H}_N]^{-1} [(NT)^{-1} \mathbf{H}_N' \mathbf{Z}_N] \}^{-1}.$$

Note that (a) and (b) together imply that $\tilde{\boldsymbol{\delta}}_N$ is a $N^{1/2}$-consistent estimator of $\boldsymbol{\delta}_N$. Regarding Assumption 4, we now have $\tilde{\mathbf{u}}_N - \mathbf{u}_N = \mathbf{D}_N \boldsymbol{\Delta}_N$, where $\mathbf{D}_N = -\mathbf{Z}_N$ and $\boldsymbol{\Delta}_N = \tilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N$. Lemma

1 shows that under Assumptions 1-3 and 8 the TSLS residuals automatically satisfy the conditions postulated in Assumptions 4, 6, and 7 with respect to $\mathbf{D}_N$, $\mathbf{\Lambda}_N$, and $\mathbf{T}_N$. Hence, Theorems 1 and 2 apply to the GM estimator $\tilde{\mathbf{\theta}}_N$, which is based on the TSLS residuals. The lemma also establishes that the elements of $\mathbf{D}_N = -\mathbf{Z}_N$ are bounded uniformly in absolute value, gives explicit expressions for $\mathbf{P}_N$ and $\tilde{\mathbf{P}}_N$, and verifies that the conditions concerning these matrices made in Theorem 3 are fulfilled. Hence, Theorem 3 covers the GM estimator $\tilde{\mathbf{\theta}}_N$ and the TSLS estimator $\tilde{\mathbf{\delta}}_N$, and gives the joint limiting distribution of $N^{1/2}(\tilde{\mathbf{\theta}}_N - \mathbf{\theta}_N)$ and $N^{1/2}(\tilde{\mathbf{\delta}}_N - \mathbf{\delta}_N)$, where the matrices $\mathbf{P}_N, \tilde{\mathbf{P}}_N$, $\mathbf{F}_{\mathbf{v},N}$, $\mathbf{F}_{\mathbf{\mu},N}$ are as in Lemma 1.

2.2.2 Fixed Effects Estimation
The fixed effects TSLS estimator of model (1a) is defined as

$$\tilde{\underline{\mathbf{\delta}}}_N = (\hat{\underline{\mathbf{Z}}}'_N \underline{\mathbf{Z}}_N)^{-1}\hat{\underline{\mathbf{Z}}}'_N \underline{\mathbf{y}}_N, \text{ where} \tag{34}$$

$$\hat{\underline{\mathbf{Z}}}_N = \mathbf{P}_{\underline{\mathbf{H}}_N}\underline{\mathbf{Z}}_N = \mathbf{P}_{\underline{\mathbf{H}}_N}\mathbf{Z}_N \text{ with } \mathbf{P}_{\underline{\mathbf{H}}_N} = \underline{\mathbf{H}}_N(\underline{\mathbf{H}}'_N\underline{\mathbf{H}}_N)^{-1}\underline{\mathbf{H}}'_N\underline{\mathbf{Z}}_N.$$

The fixed effects estimates $\tilde{\underline{\mathbf{\delta}}}_N$ can then be used to obtain consistent estimates of the disturbances, given by $\tilde{\mathbf{u}}_N = \mathbf{y}_n - \mathbf{Z}_N\tilde{\underline{\mathbf{\delta}}}_N$, which are then used for the GM estimation of the parameters $\rho_{s,N}$, $s = 1,...,S$, and σ_μ^2. These should not be confused with the fixed effects residuals $\tilde{\underline{\mathbf{u}}}_N = \underline{\mathbf{y}}_n - \underline{\mathbf{Z}}_N\tilde{\underline{\mathbf{\delta}}}_N$, which are an estimate of $\mathbf{Q}_{0,N}\mathbf{u}_N$.

The results for the fixed effects estimation are exactly as in Lemma 1, with $\mathbf{T}_N$, $\mathbf{P}_N$, $\mathbf{H}_N$ replaced with their within-transformed counterparts $\underline{\mathbf{T}}_N$, $\underline{\mathbf{P}}_N$, $\underline{\mathbf{H}}_N$, and with

$$\underline{\mathbf{T}}_{\mathbf{\mu},N} = \mathbf{0}, \; \underline{\mathbf{F}}_{\mathbf{\mu},N} = \mathbf{0}, \text{ and}$$

$$\underline{\mathbf{F}}_{\mathbf{v},N} = \mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}'_{m,N})^{-1}]\underline{\mathbf{H}}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}'_{m,N})^{-1}]\underline{\mathbf{H}}_N.^{16}$$

3. Definition of Spatial Generalized Two-Stage Least Squares (GTSLS) Estimator and Asymptotic Results

---

[16] By the idempotency of the within-transformation matrix $\mathbf{Q}_{0,N}$, one could equivalently use the fixed effects residuals $\underline{\mathbf{v}}_N = \mathbf{Q}_{0,N}\mathbf{v}_N$ in the expression $(NT)^{1/2}\mathbf{\Delta}_N = (NT)^{-1/2}\underline{\mathbf{T}}'_{\mathbf{v},N}\underline{\mathbf{v}}_N + o_p(1)$. However, since the derivation of the heteroskedasticity-robust variance-covariance matrix relies on the use of the original residuals, we also define the fixed effects estimator as a linear form in the original residuals $\mathbf{v}_N$.

3.1. Random Effects Estimation

The spatial GLS transformed version of model (1b) is given by

$$\mathbf{y}_N^* = \mathbf{Z}_N^* \boldsymbol{\delta}_N + \mathbf{u}_N^*, \tag{34}$$

where $\mathbf{y}_N^* = \mathbf{R}_N \mathbf{y}_N$, $\mathbf{Z}_N^* = \mathbf{R}_N \mathbf{Z}_N$, and $\mathbf{u}_N^* = \mathbf{R}_N \mathbf{u}_N = \boldsymbol{\varepsilon}_N$ and the transformation matrix $\mathbf{R}_N$ is given by $\mathbf{R}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})]$.

The random effects spatial GTSLS estimator, denoted as $\hat{\boldsymbol{\delta}}_N^*$, is then obtained as a TSLS estimator applied to the transformed model (37), using the transformed instruments $\mathbf{H}_N^* = \mathbf{R}_N \mathbf{H}_N$, i.e.,

$$\hat{\boldsymbol{\delta}}_N^* = (\hat{\mathbf{Z}}_N^{*\prime} \mathbf{Z}_N^*)^{-1} \hat{\mathbf{Z}}_N^{*\prime} \mathbf{y}_N^*, \tag{35a}$$

with $\hat{\mathbf{Z}}_N^* = \mathbf{P}_{\mathbf{H}_N^*} \mathbf{Z}_N^*$ and $\mathbf{P}_{\mathbf{H}_N} = \mathbf{H}_N^* (\mathbf{H}_N^{*\prime} \mathbf{H}_N^*)^{-1} \mathbf{H}_N^{*\prime}$.

The feasible random effects spatial GTSLS estimator, denoted as $\hat{\tilde{\boldsymbol{\delta}}}_N^*$, is defined analogously, replacing the transformation matrix $\mathbf{R}_N$ by its estimate $\tilde{\mathbf{R}}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \tilde{\rho}_{m,N} \mathbf{M}_{m,N})]$, i.e.,

$$\hat{\tilde{\boldsymbol{\delta}}}_N^* = (\hat{\tilde{\mathbf{Z}}}_N^{*\prime} \tilde{\mathbf{Z}}_N^*)^{-1} \hat{\tilde{\mathbf{Z}}}_N^{*\prime} \tilde{\mathbf{y}}_N^*, \tag{35b}$$

where the tilde indicates that the transformation is based on the estimate of $\mathbf{R}_N$.

The following lemma shows that the various assumptions maintained in Section III are automatically satisfied by the (feasible) random effects spatial GTSLS estimator $\hat{\tilde{\boldsymbol{\delta}}}_N^*$ and the corresponding residuals $\mathbf{u}_N(\hat{\tilde{\boldsymbol{\delta}}}_N^*) = \mathbf{y}_N - \mathbf{Z}_N \hat{\tilde{\boldsymbol{\delta}}}_N^*$. The proof is given in Appendix B.

Lemma 2.

Suppose the Assumptions of Lemma 1 hold, and let $\hat{\tilde{\boldsymbol{\delta}}}_N^*$ be defined as in (39), where $\breve{\boldsymbol{\theta}}_N$ is any $N^{1/2}$-consistent estimator of $\boldsymbol{\theta}_N$ (such as the GM estimator $\tilde{\boldsymbol{\theta}}_N$ based on the TSLS residuals). Then

(a) $(NT)^{1/2} \Delta_N^* = (NT)^{-1/2} \mathbf{T}_{\mathbf{v},N}^{*\prime} \mathbf{v}_N + (NT)^{-1/2} \mathbf{T}_{\boldsymbol{\mu},N}^{*\prime} \boldsymbol{\mu}_N + o_p(1)$, where

$$\mathbf{T}_{\mathbf{v},N}^* = \mathbf{F}_{\mathbf{v},N}^* \mathbf{P}_N^*, \ \mathbf{T}_{\boldsymbol{\mu},N}^* = \mathbf{F}_{\boldsymbol{\mu},N}^* \mathbf{P}_N^*,$$

$$\mathbf{P}_N^* = \mathbf{Q}_{\mathbf{H}^*\mathbf{H}^*}^{-1} \mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*} (\mathbf{Q}'_{\mathbf{H}^*\mathbf{Z}^*} \mathbf{Q}_{\mathbf{H}^*\mathbf{H}^*}^{-1} \mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*})^{-1},$$

$$\mathbf{F}_{\mathbf{v},N}^* = \mathbf{H}_N^*,$$

$$\mathbf{F}_{\boldsymbol{\mu},N}^* = (\mathbf{e}'_T \otimes \mathbf{I}_N)\mathbf{H}_N^* .$$

(b) $(NT)^{-1/2}\mathbf{T}_{\mathbf{v},N}^{*\prime}\mathbf{v}_N + (NT)^{-1/2}\mathbf{T}_{\boldsymbol{\mu},N}^{*\prime}\boldsymbol{\mu}_N = O_p(1)$.

(c) $\mathbf{P}_N^* = O(1)$ and $\breve{\mathbf{P}}_N^* - \mathbf{P}_N^* = o_p(1)$ for

$$\breve{\mathbf{P}}_N^* = [(NT)^{-1}\breve{\mathbf{H}}_N^{*\prime}\breve{\mathbf{H}}_N^*]^{-1}[(NT)^{-1}\breve{\mathbf{H}}_N^{*\prime}\breve{\mathbf{Z}}_N^*] \times \{[(NT)^{-1}\breve{\mathbf{Z}}_N^{*\prime}\breve{\mathbf{H}}_N^*][(NT)^{-1}\breve{\mathbf{H}}_N^{*\prime}\breve{\mathbf{H}}_N^*]^{-1}[(NT)^{-1}\breve{\mathbf{H}}_N^{*\prime}\breve{\mathbf{Z}}_N^*]\}^{-1}.$$

In light of Lemmata 1 and 2 the joint limiting distribution of the (feasible) spatial GTSLS estimator $\hat{\breve{\boldsymbol{\delta}}}_N^*$ and the GM estimator $\breve{\boldsymbol{\theta}}_N$ follows from Theorem 3 and the discussion thereafter, with $\boldsymbol{\Delta}_N^* = \hat{\breve{\boldsymbol{\delta}}}_N^* - \boldsymbol{\delta}_N$.

Note that in light of Lemma 2 the residuals $\hat{\breve{\mathbf{u}}}_N^*(\hat{\breve{\boldsymbol{\delta}}}_N^*) = \mathbf{y}_N - \mathbf{Z}_N\hat{\breve{\boldsymbol{\delta}}}_N^* = \mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N^*$ can be used to estimate $\boldsymbol{\theta}_N$ by the GM estimator defined by (18), where the discussion surrounding Lemma 1 applies analogously here. Taking this argument one step further, $\boldsymbol{\theta}_N$ and $\boldsymbol{\delta}_N$ can also be estimated by an iterative procedure.

3.2. Fixed Effects Estimation

The fixed effects spatial GTSLS estimator, denoted as $\hat{\underline{\boldsymbol{\delta}}}_N^*$, is defined as

$$\hat{\underline{\boldsymbol{\delta}}}_N^* = (\hat{\underline{\mathbf{Z}}}_N^{*\prime}\,\underline{\mathbf{Z}}_N^*)^{-1}\hat{\underline{\mathbf{Z}}}_N^{*\prime}\,\underline{\mathbf{y}}_N^*, \tag{36a}$$

with $\hat{\underline{\mathbf{Z}}}_N^* = \mathbf{P}_{\underline{\mathbf{H}}_N^*}\underline{\mathbf{Z}}_N^*$ and $\mathbf{P}_{\underline{\mathbf{H}}_N} = \underline{\mathbf{H}}_N^*(\underline{\mathbf{H}}_N^{*\prime}\,\underline{\mathbf{H}}_N^*)^{-1}\underline{\mathbf{H}}_N^{*\prime}$.

The feasible fixed effects spatial GTSLS estimator, denoted as $\hat{\underline{\tilde{\boldsymbol{\delta}}}}_N^*$, is defined analogously, using the estimate of the transformation matrix $\tilde{\mathbf{R}}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \tilde{\rho}_{m,N}\mathbf{M}_{m,N})]$, i.e.,

$$\hat{\underline{\tilde{\boldsymbol{\delta}}}}_N^* = (\hat{\underline{\tilde{\mathbf{Z}}}}_N^{*\prime}\,\underline{\tilde{\mathbf{Z}}}_N^*)^{-1}\hat{\underline{\tilde{\mathbf{Z}}}}_N^{*\prime}\,\underline{\tilde{\mathbf{y}}}_N^*. \tag{36b}$$

The results for the fixed effects estimation are exactly as in Lemma 1, with $\mathbf{T}_N^*$, $\mathbf{P}_N^*$, $\mathbf{H}_N^*$ replaced with their within-transformed counterparts $\underline{\mathbf{T}}_N^*$, $\underline{\mathbf{P}}_N^*$, $\underline{\mathbf{H}}_N^*$, and with

$$\underline{\mathbf{T}}_{\boldsymbol{\mu},N}^{*} = \mathbf{0} \text{ and } \underline{\mathbf{F}}_{\boldsymbol{\mu},N}^{*} = \mathbf{0}, \text{ and}$$

$$\underline{\mathbf{F}}_{\mathbf{v},N}^{*} = \underline{\mathbf{H}}_{N}^{*}.$$

Again notice that it is not the fixed effects residuals but the estimated disturbances $\tilde{\mathbf{u}}_N(\hat{\tilde{\boldsymbol{\delta}}}_N^{*}) = \mathbf{y}_N - \mathbf{Z}_N\hat{\tilde{\boldsymbol{\delta}}}_N^{*}$, which can be used in the GM estimation of $\boldsymbol{\theta}_N$.

V. Variance-Covariance Matrix Estimation

As evident from Theorem 3, the matrix $\boldsymbol{\Omega}_{\mathbf{w},N}$ is of sandwich form. Both under random and fixed effects estimation, the "sandwiched" middle term, i.e., $\boldsymbol{\Psi}_{\mathbf{w},N}$, is seen to depend (among others) on the idiosyncratic error terms $\mathbf{v}_N$. A complication in deriving a consistent estimator for $\boldsymbol{\Psi}_{\mathbf{w},N}$ arises from the well-known fact that one can only obtain consistent estimates of the vector of fixed effects residuals $\underline{\mathbf{v}}_N = (\underline{v}_{it,N})$, i.e., the within-transformed residuals, but not of the original idiosyncratic errors $\mathbf{v}_N$ – a manifestation of the so-called incidental parameter problem (Lancaster, 2000).

This point was prominently made in a recent paper by Stock and Watson (2008), who suggest a heteroskedasticity-robust bias-corrected variance-covariance matrix estimator for nonspatial fixed effects panel data models. A closely related issue arises in the estimation of the variance-covariance matrix of the GM estimates $\boldsymbol{\theta}_N$ given by (28). In the following, we will derive bias-corrected estimators for the joint asymptotic variance-covariance matrix of all model parameters under both fixed and random effects estimation, pursuing an approach analogous to that in Stock and Watson (2008).

1. Estimation of $\boldsymbol{\Psi}_{\mathbf{w},N}$

In the following, we derive estimators for the each block of $\boldsymbol{\Psi}_{\mathbf{w},N}$. We start by defining an estimator for $\boldsymbol{\Psi}_{\boldsymbol{\Delta\Delta},N}$, required for inference with respect to the parameters $\boldsymbol{\delta}_N$ of the main equation (1a). In a next step we turn to the estimation of the (inverse) of the optimal weighting matrix for the GM estimation $\boldsymbol{\Psi}_N$, which is also a key element in the estimation of the variance-covariance matrix of the GM estimates of $\boldsymbol{\theta}_N$. Finally, we turn to the estimation of $\boldsymbol{\Psi}_{\boldsymbol{\theta\Delta}N}$, required for joint tests regarding $\boldsymbol{\theta}_N$ and $\boldsymbol{\delta}_N$.

1.1 Estimation of $\boldsymbol{\Psi}_{\boldsymbol{\Delta\Delta},N}$

Consider

$$\boldsymbol{\Psi}_{\Delta\Delta,N} = \boldsymbol{\Psi}_{\Delta\Delta,N}^{\mathbf{v}} + \boldsymbol{\Psi}_{\Delta\Delta,N}^{\mu} \text{, where}$$

$$\boldsymbol{\Psi}_{\Delta\Delta,N}^{\mathbf{v}} = (NT)^{-1}(\mathbf{F}_{\mathbf{v},N}'\boldsymbol{\Sigma}_N\mathbf{F}_{\mathbf{v},N}) \text{ and } \boldsymbol{\Psi}_{\Delta\Delta,N}^{\mu} = (NT)^{-1}\sigma_{\mu}^2\mathbf{F}_{\mu,N}'\mathbf{F}_{\mu,N} \, .$$

Under random effects estimation, the estimators for $\mathbf{F}_N$ (original model) and $\mathbf{F}_N^*$ (spatial GLS transformed model) are defined as

$$\tilde{\mathbf{F}}_{\mathbf{v},N} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \tilde{\rho}_{m,N}\mathbf{M}_{m,N}')^{+}]\mathbf{H}_N \, , \tag{37a}$$

$$\tilde{\mathbf{F}}_{\mu,N} = (\mathbf{e}_T' \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \tilde{\rho}_{m,N}\mathbf{M}_{m,N}')^{+}]\mathbf{H}_N \, ,$$

and

$$\tilde{\mathbf{F}}_{\mathbf{v},N}^* = \mathbf{H}_N^* = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \tilde{\rho}_{m,N}\mathbf{M}_{m,N})]\mathbf{H}_N \, , \tag{37b}$$

$$\tilde{\mathbf{F}}_{\mu,N}^* = (\mathbf{e}_T' \otimes \mathbf{I}_N)\mathbf{H}_N^* = (\mathbf{e}_T' \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \tilde{\rho}_{m,N}\mathbf{M}_{m,N})]\mathbf{H}_N \, ,$$

Under fixed effects estimation, the estimators for $\mathbf{F}_N$ and $\mathbf{F}_N^*$ are defined as

$$\underline{\tilde{\mathbf{F}}}_{\mathbf{v},N} = \mathbf{Q}_{0,N}\tilde{\mathbf{F}}_{\mathbf{v},N} \, , \quad \underline{\tilde{\mathbf{F}}}_{\mu,N} = \mathbf{0} \text{, and} \tag{37c}$$

$$\underline{\tilde{\mathbf{F}}}_{\mathbf{v},N}^* = \mathbf{Q}_{0,N}\tilde{\mathbf{F}}_{\mathbf{v},N}^* \, , \quad \underline{\tilde{\mathbf{F}}}_{\mu,N}^* = \mathbf{0} \, . \tag{37d}$$

Hence – under random effects estimation of the untransformed model – the estimator for $\boldsymbol{\Psi}_{\Delta\Delta,N}^{\mu}$ is given by

$$\tilde{\boldsymbol{\Psi}}_{\Delta\Delta,N}^{\mu} = \frac{1}{NT}\tilde{\sigma}_{\mu}^2\tilde{\mathbf{F}}_{\mu,N}'\tilde{\mathbf{F}}_{\mu,N} \, , \tag{38}$$

where $\tilde{\sigma}_{\mu}^2$ is the GM estimate of σ_{μ}^2 (based on the residuals generated using the random effects estimator $\tilde{\mathbf{u}}_N(\tilde{\boldsymbol{\delta}}_N) = \mathbf{y}_N - \mathbf{Z}_N\tilde{\boldsymbol{\delta}}_N$). For the other estimators considered, $\tilde{\boldsymbol{\Psi}}_{\Delta\Delta,N}^{\mu}$ is defined in the same way, properly replacing the $\mathbf{F}$-matrices and the estimates of the disturbances $\tilde{\mathbf{u}}_N$.

As already mentioned above, due to the heteroskedasticity of $\mathbf{v}_N$ and the fact that the variance covariance matrix depends on the idiosyncratic error terms in levels $\mathbf{v}_N$ rather than the fixed effects residuals $\underline{\mathbf{v}}_N$, a bias correction is required. As shown in Lemma C.2 of the

Appendix, adopting an approach analogous to that in Stock and Watson (2008) in the present framework yields the following bias-corrected estimator for $\Psi^{\mathbf{v}}_{\Lambda\Lambda,N}$:[17]

$$\tilde{\Psi}^{\mathbf{v}}_{\Lambda\Lambda,N} = \frac{1}{N(T-2)}(\tilde{\mathbf{F}}'_{\mathbf{v},N}\tilde{\Sigma}^{HR}_{N}\tilde{\mathbf{F}}_{\mathbf{v},N}),\tag{39}$$

where $\tilde{\Sigma}^{HR}_{N} = diag^{NT}_{n=1}[(\tilde{v}^{HR}_{n,N})^2]$ with $(\tilde{v}^{HR}_{it,N})^2 = \tilde{v}^2_{it,N} - \frac{1}{T(T-1)}\sum_{r=1}^{T}\tilde{v}^2_{ir,N}$. The estimates of the fixed effects residual are given by $\tilde{\underline{\mathbf{v}}}_N = (\tilde{\underline{v}}_{it,N}) = \mathbf{Q}_{0,N}\tilde{\boldsymbol{\varepsilon}}_N = \mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})\tilde{\mathbf{u}}_N$. Again the modification modifications of (39) for other estimators are straightforward, replacing $\tilde{\mathbf{F}}_{\mathbf{v},N}$ properly.

We summarize the consistency result of the estimators given by (38) and (39) with the following theorem.

Theorem 4a. Consistency of $\tilde{\Psi}_{\Lambda\Lambda,N}$

Let $\tilde{\Psi}_{\Lambda\Lambda,N} = \tilde{\Psi}^{\mu}_{\Lambda\Lambda,N} + \tilde{\Psi}^{\mathbf{v}}_{\Lambda\Lambda,N}$ with $\tilde{\Psi}^{\mu}_{\Lambda\Lambda,N}$ and $\tilde{\Psi}^{\mathbf{v}}_{\Lambda\Lambda,N}$ defined in (38) and (39). Suppose that the Assumptions of Theorem 3, apart from Assumptions 5 and 7, hold and that additionally all of the fourth moments of the elements of $\mathbf{D}_N$ are bounded uniformly. Suppose furthermore (a) $\sup_N \sum_{s=1}^{S}|\rho_{s,N}| < 1$ and that the row and column sums of $\mathbf{M}_N$ are bounded uniformly in absolute value by one and some finite constant respectively, and (b) $\tilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$ with $\mathbf{P}_N = O(1)$. Then, $\tilde{\Psi}_{\Lambda\Lambda,N} - \Psi_{\Lambda\Lambda,N} = o_p(1)$ and $\tilde{\Psi}^{+}_{\Lambda\Lambda,N} - \Psi^{-1}_{\Lambda\Lambda,N} = o_p(1)$.

Proof. Theorem 4a follows from Lemmata C.2 and C.3 in Appendix C.[18]

Remark 3: Under estimation of the spatial GLS transformed model (where the inverse of $\mathbf{R}_N$ cancels out), condition (a) can be dropped. Under TSLS (or spatial GTSLS estimation), condition (b) in Theorem 4a is automatically fulfilled (see Lemmata 1 and 2).

1.2 Estimation of Ψ_N

---

[17] The result in Stock and Watson (2008) is obtained as a special case for $\rho_N = \mathbf{0}$ and if there are no endogenous right-hand side variables, i.e., $\underline{\mathbf{F}}_N = \underline{\mathbf{X}}_N$.

[18] Note that Lemma C.2 uses a slightly different definition of $\tilde{\Sigma}^{HR}_{N}$, factoring out $T/(T-2)$, for notational convenience of the proof.

Consider the elements of $\boldsymbol{\Psi}_N$ as defined in (28). For estimation, it will turn out convenient to rewrite the part of the elements of $\boldsymbol{\Psi}_N$ as given by (28a) with the main diagonal elements of the matrices $\mathbf{A}_{c,\mathbf{v},N}^{s,s'}$ set to zero in the first expression of the trace in the first line. Furthermore, to simplify the exposition we drop the indices c, s, and s' in the following derivation and to adopt the following notational convention. We refer to the matrix $\mathbf{A}_{c,\mathbf{v},N}^{s,s'}$, associated with the set of indices c, s, s' as $\mathbf{A}_{\mathbf{v},N} = (a_{\mathbf{v},nn,N}) = (a_{\mathbf{v},it,js,N})$, and to the matrix $\mathbf{A}_{c,\mathbf{v},N}^{s,s'}$ with its main diagonal elements set to zero as $\mathbf{A}_{\mathbf{v},N}^{\circ} = \mathbf{A}_{\mathbf{v},N} - diag_{n=1}^{NT}(a_{\mathbf{v},nn,N})$. Analogously, the matrix $\mathbf{A}_{c',\mathbf{v},N}^{t,t'}$, associated with the set of indices c', t, t' is denoted as $\mathbf{B}_{\mathbf{v},N} = (b_{\mathbf{v},nn,N}) = (b_{\mathbf{v},it,js,N})$, and $\mathbf{B}_{\mathbf{v},N}^{\circ} = \mathbf{B}_{\mathbf{v},N} - diag_{n=1}^{NT}(b_{\mathbf{v},nn,N})$. We adopt the same convention for the matrices $\mathbf{A}_{c,\boldsymbol{\mu},N}^{s,s'}$, henceforth denoted as $\mathbf{A}_{\boldsymbol{\mu},N} = (a_{\boldsymbol{\mu},nn,N}) = (a_{\boldsymbol{\mu},it,js,N})$, as well as $\mathbf{A}_{c',\boldsymbol{\mu},N}^{t,t'}$, henceforth denoted as $\mathbf{B}_{\boldsymbol{\mu},N}$, and also for the vectors $\mathbf{a}_{c,\mathbf{v},N}^{s,s'}$ and $\mathbf{a}_{c',\mathbf{v},N}^{t,t'}$, henceforth denoted $\mathbf{a}_{\mathbf{v},N} = (a_{\mathbf{v},n,N}) = (a_{\mathbf{v},it,N})$ and $\mathbf{b}_{\mathbf{v},N} = (b_{\mathbf{v},n,N}) = (b_{\mathbf{v},it,N})$, respectively. Finally, we refer to products of equally indexed elements of $\mathbf{A}_{\mathbf{v},N}$ and $\mathbf{B}_{\mathbf{v},N}$ as $c_{\mathbf{v},it,N} = a_{\mathbf{v},it,js,N} b_{\mathbf{v},it,js,N}$ (or $c_{n,N} = a_{\mathbf{v},nn,N} b_{\mathbf{v},nn,N}$), and we define $d_{\mathbf{v},n,N} = (a_{\mathbf{v},n,N} b_{\mathbf{v},nn,N} + a_{\mathbf{v},nn,N} b_{\mathbf{v},n,N})$ and $d_{\boldsymbol{\mu},n,N} = (a_{\boldsymbol{\mu},n,N} b_{\boldsymbol{\mu},nn,N} + a_{\boldsymbol{\mu},nn,N} b_{\boldsymbol{\mu},n,N})$.

In that case, equation (28a) can be written, for given a given pair of index sets c, s, s' and c', t, t', as

$$\mathfrak{E}_N = \mathfrak{E}_{\mathbf{v},N}^{*,\circ} + \mathfrak{E}_{\mathbf{v},N}^{*} + \mathfrak{E}_{\boldsymbol{\mu},N}^{*} + \mathfrak{E}_{\mathbf{v}\boldsymbol{\mu},N}^{*} + \mathfrak{E}_{\mathbf{v},N}^{**} + \mathfrak{E}_{\boldsymbol{\mu},N}^{**},\tag{40}$$

where

$$\mathfrak{E}_{\mathbf{v},N}^{*,\circ} = 2N^{-1} Tr(\mathbf{A}_{\mathbf{v},N}^{\circ} \boldsymbol{\Sigma}_N \mathbf{B}_{\mathbf{v},N}^{\circ} \boldsymbol{\Sigma}_N) = 2N^{-1} E \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\substack{j=1\\js\neq it}}^{N}\sum_{s=1}^{T} c_{\mathbf{v},it,js,N} v_{it,N}^2 v_{js,N}^2,$$

$$\mathfrak{E}_{\mathbf{v},N}^{*,\bullet} = 2N^{-1} \sum_{n=1}^{NT} c_{\mathbf{v},n,N} \sigma_{v,n,N}^4 = \frac{2}{3} N^{-1} \sum_{n=1}^{NT} c_{\mathbf{v},n,N} \sigma_{v,n,N}^{(4)},$$

$$\mathfrak{E}_{\mathbf{v}\boldsymbol{\mu},N}^{*} = 4\sigma_{\mu}^2 N^{-1} Tr(\mathbf{A}_{\mathbf{v},\boldsymbol{\mu},N}' \boldsymbol{\Sigma}_N \mathbf{B}_{\mathbf{v},\boldsymbol{\mu},N}),$$

$$\mathfrak{E}_{\boldsymbol{\mu},N}^{*} = 2\sigma_{\mu}^4 N^{-1} Tr(\mathbf{A}_{\boldsymbol{\mu},N} \mathbf{B}_{\boldsymbol{\mu},N}),$$

$$\mathfrak{E}_{\mathbf{v},N}^{**} = N^{-1} \mathbf{a}_{\mathbf{v},N}' \boldsymbol{\Sigma}_N \mathbf{b}_{\mathbf{v},N},$$

$$\mathfrak{E}_{\boldsymbol{\mu},N}^{**} = N^{-1} \sigma_{\mu}^2 \mathbf{a}_{\boldsymbol{\mu},N}' \mathbf{b}_{\boldsymbol{\mu},N},$$

Notice that the terms $c_{\mathbf{v},n,N} \sigma_{v,n,N}^4$, $n = 1,..., NT$, associated with the main diagonal elements of $\mathbf{A}_{\mathbf{v},N}$ and $\mathbf{B}_{\mathbf{v},N}$, in the expression $\mathbf{A}_{\mathbf{v},N} \boldsymbol{\Sigma}_N \mathbf{B}_{\mathbf{v},N} \boldsymbol{\Sigma}_N$, are not included in $\mathfrak{E}_{\mathbf{v},N}^{*,\circ}$. To rewrite $\mathfrak{E}_{\mathbf{v},N}^{*,\bullet}$,

we have used the fact that $\sigma_{v,n,N}^4 = \sigma_{v,n,N}^{(4)}/3$ under normality, where $\sigma_{v,n,N}^{(4)}$ is the fourth moment of $\mathbf{v}_N$.

We next define the estimates for $\mathbf{a}_{\mathbf{v},N}$ and $\mathbf{a}_{\boldsymbol{\mu},N}$:

$$\tilde{\mathbf{a}}_{\mathbf{v},N} = (\tilde{a}_{\mathbf{v},n,N}) = \tilde{\mathbf{T}}_{\mathbf{v},N}\tilde{\boldsymbol{\alpha}}_N, \ \ \tilde{\mathbf{a}}_{\boldsymbol{\mu},N} = (\tilde{a}_{\boldsymbol{\mu},n,N}) = \tilde{\mathbf{T}}_{\boldsymbol{\mu},N}\tilde{\boldsymbol{\alpha}}_N, \text{ with} \tag{41}$$

$\tilde{\mathbf{T}}_{\mathbf{v},N} = \tilde{\mathbf{F}}_{\mathbf{v},N}\tilde{\mathbf{P}}_N$, $\tilde{\mathbf{T}}_{\boldsymbol{\mu},N} = \tilde{\mathbf{F}}_{\boldsymbol{\mu},N}\tilde{\mathbf{P}}_N$, and $\tilde{\boldsymbol{\alpha}}_N = 2N^{-1}(\mathbf{D}_N'\tilde{\mathbf{C}}_N\tilde{\mathbf{u}}_N)$. The (properly indexed) matrices $\tilde{\mathbf{C}}_N$, i.e., $\mathbf{C}_{c,N}^{s,s'}$, are given by (22) with $\boldsymbol{\rho}_N$ replaced by $\tilde{\boldsymbol{\rho}}_N$ and the estimates of the disturbances are given by $\tilde{\mathbf{u}}_N(\tilde{\boldsymbol{\delta}}_N) = \mathbf{y}_N - \mathbf{Z}_N\tilde{\boldsymbol{\delta}}_N$. Expression (41) as written holds for random effects estimation of the original model; the modifications for the other estimators are obvious, appropriately replacing $\mathbf{D}_N = -\mathbf{Z}_N$, $\tilde{\mathbf{T}}_N$, and $\tilde{\mathbf{u}}_N$. Of course, analogous definitions apply to $\mathbf{b}_{\mathbf{v},N}$ and $\mathbf{b}_{\boldsymbol{\mu},N}$.

We next define estimators for the terms in (40), starting with the "homoskedastic" terms, involving only the time-invariant error component $\mu_{i,N}$.

1.2.1. Estimation of "homoskedastic" terms
Consistent estimators of the expressions in (40), associated (only) with the homoskedastic, time-invariance error component $\mu_{i,N}$, are given by

$$\tilde{\mathfrak{E}}_{\boldsymbol{\mu},N}^* = 2\tilde{\sigma}_\mu^4 N^{-1} Tr(\mathbf{A}_{\boldsymbol{\mu},N}\mathbf{B}_{\boldsymbol{\mu},N}), \tag{42a}$$

$$\tilde{\mathfrak{E}}_{\boldsymbol{\mu},N}^{**} = N^{-1}\tilde{\sigma}_\mu^2 \tilde{\mathbf{a}}_{\boldsymbol{\mu},N}'\tilde{\mathbf{b}}_{\boldsymbol{\mu},N}, \tag{42b}$$

The consistency proofs for the estimators defined in (42) are easily seen to be special cases of those for the heteroskedastic terms considered in the next section and thus omitted for the sake of brevity.

1.2.2. Estimation of "heteroskedastic" terms
Consider first $\mathfrak{E}_{\mathbf{v},N}^{*,\circ}$ as defined in (40). Its estimation is simplified by the fact that the matrices $\mathbf{A}_{\mathbf{v},N}^\circ$ and $\mathbf{B}_{\mathbf{v},N}^\circ$ and thus the elements $c_{it,js,N}$ are time-invariant, i.e., $c_{it,js,N} = c_{i,j,N}$. As shown in Lemma C.5 in Appendix C, a consistent estimate of $\mathfrak{E}_{\mathbf{v},N}^{*,\circ}$ is given by

$$\tilde{\mathfrak{E}}_{\mathbf{v},N}^{*,\circ} = \frac{T^2}{(T-1)^2}\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N c_{i,j,N}\sum_{t=1}^T\sum_{s=1}^T \tilde{\underline{v}}_{it,N}^2 \tilde{\underline{v}}_{js,N}^2. \tag{43a}$$

Next, consider $\mathfrak{E}^{*,\bullet}_{\mathbf{v},N} = 2N^{-1}\sum_{n=1}^{NT} c_{\mathbf{v},n,N}\sigma^4_{v,n,N}$. Under normality, and noting that elements $c_{it,N}$ are time-invariant, this can also be written as weighted sum of fourth moments as $\mathfrak{E}^{*,\bullet}_{\mathbf{v},N} = \frac{2}{3}N^{-1}\sum_{i=1}^{N} c_{\mathbf{v},i,N}\sum_{t=1}^{T}\sigma^{(4)}_{v,it,N}$, which can be estimated consistently using

$$\widetilde{\mathfrak{E}}^{*,\bullet}_{\mathbf{v},N} = \frac{2}{3}[\frac{k_0}{(1-k_1 m_1)}\widetilde{\underline{\varpi}}_N + \frac{k_1 m_0}{(1-k_1 m_1)}\widetilde{\underline{\mathfrak{a}}}_N], \text{ where} \tag{43b}$$

$$\widetilde{\underline{\varpi}}_N = \frac{1}{N}\sum_{i=1}^{N} c_{i,N}\sum_{t=1}^{T}\widetilde{\underline{v}}^4_{it,N},$$

$$\widetilde{\underline{\mathfrak{a}}}_N = \frac{1}{N}\sum_{i=1}^{N} c_{i,N}\sum_{t=1}^{T}\widetilde{\underline{v}}^2_{it,N}\frac{1}{T-1}\sum_{\substack{r=1\\r\neq t}}^{T}\widetilde{\underline{v}}^2_{ir,N},$$

$$m_0 = \frac{T^3}{T^3 - 2T^2 - 3T + 9} , \quad m_1 = -\frac{(2T-3)}{T^3 - 2T^2 - 3T + 9},$$

$$k_0 = \frac{T^2(T-1)}{T^3 - 4T^2 + 6T - 3} , \quad k_1 = -\frac{(T-1)(6T-9)}{T^3 - 4T^2 + 6T - 3}.$$

The derivation, using a bias correction in the spirit of Stock and Watson (2008), and the proof of consistency is given in Lemma C.6a in Appendix C.

In light of the previous results, estimation of $\mathfrak{E}^*_{\mathbf{v}\boldsymbol{\mu},N}$ is straightforward; exploiting the fact that the weights matrices are time-invariant, a consistent estimate is given by

$$\widetilde{\mathfrak{E}}^*_{\mathbf{v}\boldsymbol{\mu},N} = 4\widetilde{\sigma}^2_\mu \frac{1}{N}\frac{T}{T-1}Tr(\mathbf{A}'_{\mathbf{v},\boldsymbol{\mu},N}\widetilde{\underline{\boldsymbol{\Sigma}}}_N\mathbf{B}_{\mathbf{v},\boldsymbol{\mu},N}), \tag{44a}$$

where $\widetilde{\underline{\boldsymbol{\Sigma}}}_N = diag^{NT}_{n=1}[(\widetilde{\underline{v}}_{n,N})^2]$.

Finally, an estimate of $\mathfrak{E}^{**}_{\mathbf{v},N}$ is given by

$$\widetilde{\mathfrak{E}}^{**}_{\mathbf{v},N} = \frac{T}{N(T-2)}\widetilde{\mathbf{a}}'_{\mathbf{v},N}\widetilde{\boldsymbol{\Sigma}}^{HR}_N\widetilde{\mathbf{b}}_{\mathbf{v},N}. \tag{44b}$$

That $\widetilde{\mathfrak{E}}^{**}_{\mathbf{v},N} - \mathfrak{E}^{**}_{\mathbf{v},N} = o_p(1)$ follows from Lemmata C.2 and C.3 and Remark C.1 thereafter in Appendix C.

We summarize the results of section 1.2 with the following theorem.

Theorem 4b. Consistency of $\widetilde{\boldsymbol{\Psi}}_N$

Suppose all of the assumptions of Theorem 4a and Assumption 7 holds and that $\mathbf{v}_N$ and $\boldsymbol{\mu}_N$ are normally distributed. Let the elements of $\tilde{\boldsymbol{\Psi}}_N$ be defined as above (from (39) to (44)). Then, $\tilde{\boldsymbol{\Psi}}_N - \boldsymbol{\Psi}_N = o_p(1)$ and $\tilde{\boldsymbol{\Psi}}_N^+ - \boldsymbol{\Psi}_N^{-1} = o_p(1)$.

Remark 4: Under non-normality, Theorem 4.b holds under additional assumptions regarding the moments of $\mathbf{v}_N$ and $\boldsymbol{\mu}_N$ and with augmented definitions of the elements of $\boldsymbol{\Psi}_N$ and $\tilde{\boldsymbol{\Psi}}_N$; details are given in the Appendix.

1.3 Estimation of $\boldsymbol{\Psi}_{\boldsymbol{\theta}\Delta,N}$

It remains to provide an estimate of $\boldsymbol{\Psi}_{\boldsymbol{\theta}\Delta,N}$, which is required for tests of joint hypotheses concerning the regression parameters $\boldsymbol{\delta}_N$ and the parameters associated with the spatial regressive disturbance process $\boldsymbol{\theta}_N$.

As evident from the results in section 1.2, the assumptions maintained in Theorem 4b are sufficient to prove that the following expressions consistently estimate the columns of $\boldsymbol{\Psi}_{\boldsymbol{\theta}\Delta,N}$ as defined in light of (32c), provided that $\mathbf{v}_N$ and $\boldsymbol{\mu}_N$ are normally distributed:

$$\tilde{\psi}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N} = \tilde{\psi}^{\mathbf{v}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N} + \tilde{\psi}^{\boldsymbol{\mu}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N}, \ c = 1,...,4, \ s,s' = 1,...,S, \text{ with} \tag{45}$$

$$\tilde{\psi}^{\mathbf{v}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N} = \frac{T^{1/2}}{N(T-2)} (\mathbf{F}'_{\mathbf{v},N} \tilde{\boldsymbol{\Sigma}}^{HR}_N \tilde{\mathbf{a}}^{s,s'}_{c,\mathbf{v},N}), \text{ and}$$

$$\tilde{\psi}^{\boldsymbol{\mu}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N} = \frac{1}{N} \frac{1}{T^{1/2}} \tilde{\sigma}^2_{\mu} (\mathbf{F}'_{\boldsymbol{\mu},N} \tilde{\mathbf{a}}^{s,s'}_{c,\boldsymbol{\mu},N}).$$

Theorem 4c. Consistency of $\tilde{\boldsymbol{\Psi}}_{\Delta\boldsymbol{\theta},N}$

Suppose the assumptions of Theorem 4b hold and let (the columns of) $\tilde{\boldsymbol{\Psi}}_{\Delta\boldsymbol{\theta},N}$ be defined by (45). Then, we have $\boldsymbol{\Psi}_{\boldsymbol{\theta}\Delta,N} = O(1)$, $\tilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}\Delta,N} - \boldsymbol{\Psi}_{\boldsymbol{\theta}\Delta,N} = o_p(1)$, and $\tilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}\Delta,N} = O_p(1)$.

Remark 5: Under non-normality, Theorem 4c holds under additional assumptions and with augmented definitions of the columns of $\boldsymbol{\Psi}_{\Delta\boldsymbol{\theta},N}$ and $\tilde{\boldsymbol{\Psi}}_{\Delta\boldsymbol{\theta},N}$; details are given in the appendix.

2. Estimation of $\boldsymbol{\Omega}_{\mathbf{w},N}$

The estimate of $\mathbf{J}_N$ is given by

$$\tilde{\mathbf{J}}_N = \tilde{\boldsymbol{\Gamma}}_N \tilde{\mathfrak{B}}_N. \tag{46}$$

The elements of $\tilde{\Gamma}_N$ are defined in (17) with the expectations operator suppressed and the disturbances $\mathbf{u}_N$ replaced by their estimated counterparts. For simplicity of notation, the estimated disturbances are denoted as $\tilde{\mathbf{u}}_N$ throughout, though it should be clear that they are generated by the respective estimators $\tilde{\boldsymbol{\delta}}_N$, $\underline{\tilde{\boldsymbol{\delta}}}_N$, $\hat{\tilde{\boldsymbol{\delta}}}_N^*$, or $\hat{\underline{\tilde{\boldsymbol{\delta}}}}_N^*$ defined above. For example, under fixed effects (feasible) spatial generalized LS estimation, we have $\tilde{\mathbf{u}}_N = \mathbf{y}_N - \mathbf{Z}_N \hat{\underline{\tilde{\boldsymbol{\delta}}}}_N^*$. The matrix $\tilde{\mathfrak{B}}_N$ is given by (20) with $\rho_{s,N}$ replaced by the GM estimates $\tilde{\rho}_{s,N}$, $s = 1, ..., S$.

Theorem 5. Consistency of $\tilde{\Omega}_{\mathbf{w}_N}$

Suppose that Assumptions 1-7 hold. Let $\tilde{\Psi}_{\mathbf{w},N}$ be defined as above (from (39) to (45)). Define

$$\tilde{\Omega}_{\mathbf{w},N} = \begin{bmatrix} T^{-1/2}\tilde{\mathbf{P}}_N' & \mathbf{0} \\ \mathbf{0} & (\tilde{\mathbf{J}}_N'\tilde{\Theta}_N\tilde{\mathbf{J}}_N)^{-1}\tilde{\mathbf{J}}_N'\tilde{\Theta}_N \end{bmatrix} \tilde{\Psi}_{\mathbf{w},N} \begin{bmatrix} T^{-1/2}\tilde{\mathbf{P}}_N & \mathbf{0} \\ \mathbf{0} & \tilde{\Theta}_N\tilde{\mathbf{J}}_N(\tilde{\mathbf{J}}_N'\tilde{\Theta}_N\tilde{\mathbf{J}}_N)^{-1} \end{bmatrix}.$$

It follows that $\tilde{\Omega}_{\mathbf{w}_N} - \Omega_{\mathbf{w}_N} = o_p(1)$, $\Omega_{\mathbf{w}_N} = O(1)$, and $\tilde{\Omega}_{\mathbf{w}_N} = O_p(1)$.

Proof.

Above we showed that $\tilde{\Psi}_{\mathbf{w},N} - \Psi_{\mathbf{w},N} = o_p(1)$. By assumption, $\tilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$, $\mathbf{P}_N = O(1)$, and $\tilde{\mathbf{P}}_N = O_p(1)$ as well as $\tilde{\Theta}_N - \Theta_N = o_p(1)$, $\Theta_N = O(1)$ and $\tilde{\Theta}_N = O_p(1)$. In the proof of Theorem 2 it was shown that $\tilde{\mathbf{J}}_N - \mathbf{J}_N = o_p(1)$, $\mathbf{J}_N = O(1)$, and $\tilde{\mathbf{J}}_N = O_p(1)$, and that $(\tilde{\mathbf{J}}_N'\tilde{\Theta}_N\tilde{\mathbf{J}}_N)^+ - (\mathbf{J}_N'\Theta_N\mathbf{J}_N)^{-1} = o_p(1)$, $(\mathbf{J}_N'\Theta_N\mathbf{J}_N)^{-1} = O(1)$, and $(\tilde{\mathbf{J}}_N'\tilde{\Theta}_N\tilde{\mathbf{J}}_N)^+ = O_p(1)$. It now follows that $\tilde{\Omega}_{\mathbf{w},N} - \Omega_{\mathbf{w},N} = o_p(1)$ and $\Omega_{\mathbf{w},N} = O(1)$ and thus $\tilde{\Omega}_{\mathbf{w},N} = O_p(1)$.

Remark 5: Under non-normality, Theorem 5 holds under additional assumptions and with augmented definitions of $\Psi_{\mathbf{w},N}$ and $\tilde{\Psi}_{\mathbf{w},N}$; details are given in the appendix.

VI. Random vs. Fixed Effects. A Heteroskedasticity-Robust Hausman Test

In the following we derive a Hausman-type test of the spatial random effects versus the spatial fixed effects model under heteroskedasticity of unknown form. Both estimators considered are based on the spatial GLS transformed model (which removes the cross-sectional interdependence) and use a heteroskedasticity-robust robust variance-covariance matrix for inference. In general, neither of these two estimators will be efficient, such that we use a generalized Hausman-test for inference (see, Weesie, 1999; Creel, 2004).

Consider the stacked vector of random and fixed effects estimates of the regression parameters, which is given by

$$\overline{\mathbf{d}}_N = \begin{bmatrix} N^{1/2} \mathbf{\Delta}_N^* \\ N^{1/2} \underline{\mathbf{\Delta}}_N^* \end{bmatrix} = \begin{bmatrix} N^{1/2} (\hat{\mathbf{\delta}}_N^* - \mathbf{\delta}_N) \\ N^{1/2} (\hat{\underline{\mathbf{\delta}}}_N^* - \mathbf{\delta}_N) \end{bmatrix}. \tag{47}$$

By Theorem A.1 in Kelejian and Prucha (2010)

$$\overline{\mathbf{d}}_N \overset{d}{\to} \overline{\mathbf{\Omega}}_N = \begin{bmatrix} \mathbf{\Omega}_{\hat{\mathbf{\delta}}_N^*} & \mathbf{\Omega}_{\hat{\mathbf{\delta}}_N^* \underline{\hat{\mathbf{\delta}}}_N^*} \\ \mathbf{\Omega}'_{\hat{\mathbf{\delta}}_N^* \underline{\hat{\mathbf{\delta}}}_N^*} & \mathbf{\Omega}_{\underline{\hat{\mathbf{\delta}}}_N^*} \end{bmatrix}. \tag{48}$$

As evident from (48), $\mathbf{\Omega}_{\hat{\mathbf{\delta}}_N^*} = T^{-1} \mathbf{P}_N^{*\prime} \mathbf{\Psi}_{\mathbf{\Delta\Delta},N} \mathbf{P}_N^*$ and $\mathbf{\Omega}_{\underline{\hat{\mathbf{\delta}}}_N^*} = T^{-1} \underline{\mathbf{P}}_N^{*\prime} \mathbf{\Psi}_{\mathbf{\Delta\Delta},N} \underline{\mathbf{P}}_N^*$. The off-diagonal block of $\overline{\mathbf{\Omega}}_N$ is given by

$$\mathbf{\Omega}_{\hat{\mathbf{\delta}}_N^* \underline{\hat{\mathbf{\delta}}}_N^*} = \frac{1}{T} \mathbf{P}_N^{*\prime} [\frac{1}{NT} \mathbf{F}_{\mathbf{v},N}^{*\prime} \mathbf{\Sigma}_N \underline{\mathbf{F}}_{\mathbf{v},N}^*] \underline{\mathbf{P}}_N^*, \tag{49a}$$

which can be estimated consistently – by the same logic as $\mathbf{\Omega}_{\hat{\mathbf{\delta}}_N^*}$ and $\mathbf{\Omega}_{\underline{\hat{\mathbf{\delta}}}_N^*}$ – using

$$\tilde{\mathbf{\Omega}}_{\hat{\mathbf{\delta}}_N^* \underline{\hat{\mathbf{\delta}}}_N^*} = \frac{1}{T} \tilde{\mathbf{P}}_N^{*\prime} [\frac{1}{N(T-2)} \tilde{\mathbf{F}}_{\mathbf{v},N}^{*\prime} \tilde{\mathbf{\Sigma}}_N^{HR} \underline{\tilde{\mathbf{F}}}_{\mathbf{v},N}^*] \underline{\tilde{\mathbf{P}}}_N^*. \tag{49b}$$

The Hausman test, which is derived under the null hypothesis that the random effects model as specified in section II is the true model, takes the form of a Wald-type test of the restriction that $\hat{\mathbf{\delta}}_N^* = \underline{\hat{\mathbf{\delta}}}_N^*$. Define the discrepancy vector $\hat{\mathfrak{m}}_N = \mathfrak{R}_N \hat{\mathbf{q}}_N$, where $\hat{\mathbf{q}}_N = (\hat{\mathbf{\delta}}_N^{*\prime}, \underline{\hat{\mathbf{\delta}}}_N^{*\prime})'$. Note that typically, the dimension of the parameter vector under random effects exceeds the parameter vector under fixed effects by 1 due to the inclusion of a constant. Hence, for comparison of the two estimators, we focus on a joint test regarding the slope parameters, i.e., we test

$$H_0 : \mathfrak{R} \mathbf{q}_N = \mathbf{0} \text{ against } H_1 : \mathfrak{R} \mathbf{q}_N \neq \mathbf{0}, \tag{50}$$

where $\mathfrak{R} = (\mathbf{0}_{P \times 1}, \mathbf{I}_P, -\mathbf{I}_P)$, assuming that the constant appears in the first row of the random effects estimator $\hat{\mathbf{\delta}}_N^*$. We use a generalized Wald-type test (e.g., Greene, 2003, pp. 95, 487), which takes the form[19]

$$\mathfrak{m}_N' (\mathfrak{R} \tilde{\mathbf{\Omega}}_N \mathfrak{R}')^+ \mathfrak{m}_N \sim \chi^2(P), \tag{51}$$

---

[19] If one of the estimators is efficient, the off-diagonal blocks are equal to zero and equation (51) reduces to the standard Hausman test.

where $\tilde{\mathfrak{Q}}_N = N^{-1}\tilde{\mathbf{\Omega}}_{\hat{\mathbf{\delta}}_N^* \underline{\hat{\mathbf{\delta}}}_N^*}$ and P is the number of restrictions, which is equal to the number of slope parameters in the present case.[20]

VII. Some Monte Carlo Evidence

In the following we provide some limited Monte Carlo evidence on the performance of the estimation procedure suggested in the present paper. A comprehensive assessment, using a broad range of parameter constellations, alternative distributional assumptions, and alternative specifications of the weights matrices is beyond the scope of the present paper and left for future research. We consider a SARAR(2,2) specification with two explanatory variables, assuming that $\mathbf{W}_N = \mathbf{M}_N$:[21]

$$\mathbf{y} = \mathbf{x}_1 \beta_1 + \mathbf{x}_2 \beta_2 + \sum_{r=1}^{2} \lambda_r (\mathbf{I}_T \otimes \mathbf{W}_r) \mathbf{y} + \mathbf{u}, \tag{52a}$$

$$\mathbf{u} = \sum_{s=1}^{2} \rho_s (\mathbf{I}_T \otimes \mathbf{W}_s) \mathbf{u} + \mathbf{\varepsilon}. \tag{52b}$$

We consider three sample sizes: $N = 50$, $N = 100$, and $N = 250$ and assume $T = 5$ throughout. For each Monte Carlo experiment, we consider 1000 draws. The explanatory variables $\mathbf{x}_1$ and $\mathbf{x}_2$ are generated as random draws from a standard normal distribution, scaled with a factor of five, and treated as fixed in repeated samples. The parameters are as follows: $\beta_1 = \beta_2 = 1$, $\lambda_1 = 0.5$, $\lambda_2 = 0.25$, $\rho_1 = 0.4$, and $\rho_2 = 0.2$.

The unnormalized $N \times N$ matrix $\mathbf{W}^0$ consists of two $N \times N$ matrices $\mathbf{W}_1^0$ and $\mathbf{W}_2^0$, where $\mathbf{W}_1^0 + \mathbf{W}_2^0 = \mathbf{W}^0$. The matrices $\mathbf{W}_1^0$ and $\mathbf{W}_2^0$ are specified such that they contain the elements of $\mathbf{W}^0$ for a different band of neighbours each. Otherwise, they have zero elements. In line with Kelejian and Prucha (2010), we choose a design, where $\mathbf{W}_1^0$ corresponds to an 'up to 3 ahead and up to 3 behind' specification and $\mathbf{W}_2^0$ corresponds to a '4 to 6 ahead and 4 to 6 behind' specification. The final weights matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are obtained by individually row-normalizing $\mathbf{W}_1^0$ and $\mathbf{W}_2^0$. As already mentioned, we have $\mathbf{M}_1 = \mathbf{W}_1$ and $\mathbf{M}_2 = \mathbf{W}_2$.

---

[20] The theory underlying Hausman tests with not fully efficient estimators is derived in White (1982, 1994). In a non-spatial context, such a generalized Hausman test is considered, e.g., in Weesie (1999) or Creel (2004). Sufficient assumptions to ensure well-behaved asymptotic properties in generalized Wald tests are derived and discussed in Andrews (1987) and Vuong (1987).

[21] For simplicity of notation, the subscript N is suppressed in the following.

Regarding the choice of instruments, we include linearly independent terms of up to second order spatial lags of the exogenous variables. In particular, the matrix of untransformed instruments $\mathbf{H}$ contains 12 columns and is given by

$$\mathbf{H} = [\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_1)\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_2)\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_1^2)\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_2^2)\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_1\mathbf{W}_2)\mathbf{X}]. \tag{53}$$

The elements of the error term $\boldsymbol{\varepsilon}$ are specified as $\varepsilon_{it} = \mu_i + v_{it}$, where the idiosyncratic error is given by $v_{it} = 0.5\zeta_{it}\sqrt{(0.5 + 0.1x_{1,it}^2 + 0.1x_{2,it}^2)} + \sqrt{\kappa_{it}}\vartheta_{it}$. Thereby ζ_{it} and ϑ_{it} are draws from a standard normal distribution and κ_{it} is a draw from a uniform distribution with support $[0.5, 1.5]$, which is treated as fixed in repeated samples. Hence, v_{it} exhibits both conditional and unconditional heteroskedasticity.

The individual effect is specified as $\mu_i = \pi_1\bar{x}_{1,i,} + \pi_2\bar{x}_{2,i} + \bar{w}_i$, where $\bar{w}_{i,N}$ is a draw from normal distribution with variance 0.5. We consider two specifications: in the random effects model we have $\pi_1 = \pi_2 = 0$ (and, hence, $Var(\mu_{i,N}) = Var(\bar{w}_{i,N})$); in the fixed effects model we have $\pi_1 = \pi_2 = 0.25$ (and, hence, $Var(\mu_{i,N}) = Var(\bar{\mathbf{x}}_{i,})\boldsymbol{\pi}^2 + Var(\bar{w}_i)$).

Results for the estimates of ρ_1 and ρ_2 are obtained by the GM estimator defined in equation (18), using the optimal weighting matrix under normality $(\tilde{\boldsymbol{\Psi}}_N^\circ)^{-1}$. The estimates reported for the regression parameters are FGTSLS estimates as defined in (35) and (36) using the transformed set of instruments $\tilde{\mathbf{H}}^{**}$. For each single coefficient, we report the average bias and root mean squared error for each parameter constellation and the rejection rates for the test that the coefficient is equal to the true parameter value. For the random effects models, we also show the results for the Hausman test.

< Table 1 >

Table 1 reports the results of the Monte Carlo analysis for the three different sample sizes considered, both under the random and fixed effects specification. Given that the natural habitat of GM estimation is large samples, the performance in the smallest sample with $N = 50$ is acceptable. In the random effects (fixed effects) specification, the average bias and RMSE amount to 0.0008385 and 0.0246475 (0.001719 and 0.027935) for the estimates of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$ and -0.0096335 and 0.2563835 (-0.0106385 and 1.050696) for the estimates of $\boldsymbol{\rho} = (\rho_1, \rho_2)'$. With an average rejection rate of 0.0685 and 0.139 (0.0650 and 0.1225), the performance of the single hypotheses tests referring to $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ is not too bad as well. The Hausman test is oversized with a rejection rate of 0.1060.

For moderately sized samples with $N = 250$, the bias has virtually disappeared: in relative terms it amounts to 0.01560 (0.0102) percent for estimates of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$ and to -0.1647

(-0.280) percent for the estimates of $\boldsymbol{\rho} = (\rho_1, \rho_2)'$ under random effects (fixed effects). The average RMSE of the estimates of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$ shrinks to 0.011376 (0.011466), that of the estimates of $\boldsymbol{\rho} = (\rho_1, \rho_2)'$ shrinks to 0.213485 (0.800393) under random effects (fixed effects). The size of the tests improves, but it approaches nominal size of 5 percent relatively slowly. The reason for the latter partly accrues to the fact that the data for $\mathbf{x}_1$ and $\mathbf{x}_2$ are generated as random draws. A second reason relates to the specific 'ahead-behind' design of the spatial weights matrices, which – together with the properties of the explanatory variables – results in a fairly high correlation between spatial lags of different orders. With explanatory variables as in many empirical applications and less artificial spatial weights matrices, there will be less correlation between the spatial lags of the explanatory variables and spatial lags of different orders and the size of tests can be expected to approach the nominal size faster than in the chosen design. Regarding the GM estimates of $\boldsymbol{\rho}$, the average size amounts to 0.139 (0.123), that for the FGTSLS estimates of $\boldsymbol{\lambda}$ to 0.0555 (0.139). The performance of the Hausman test is worth mentioning, which has already approached its nominal size with a rejection rate of 0.056.

The final column in Table 1 considers the case with N = 250 and where the sum of the parameters of the spatial lag of the dependent variable is closer to 1, i.e., with $\lambda_1 = 0.6$ and $\lambda_2 = 0.35$. As can be seen from the results, the performance in terms of bias and size is comparable with the parameter constellation where the sum of λ_1 and λ_2 is smaller in magnitude.

Overall, the Monte Carlo experiments illustrate that the proposed estimators work reasonably well in terms of bias and RMSE, even in very small samples. Regarding the estimates of the variance-covariance matrix of the parameter estimates, in particular those relating to the disturbance process, some care is warranted in the interpretation of the results in small samples, though the tests appear to be conservative in the sense that they under-reject the null and the *p*-values converge from above for reasons mentioned in the previous paragraph. It should also be emphasized that the results here are based on a correctly specified model with a high signal to noise ratio. Hence, apart from a comprehensive Monte Carlos study using alternative distributional assumptions and 'real world' explanatory variables and weights matrices, an interesting extension for future research would be to explore small sample corrections or re-sampling methods for the GM estimators considered in the present paper in order to improve the performance in small samples or in empirical models with poor fit.

VIII. Conclusions

This paper derived a two-step estimation procedure for spatial regressive panel data models with spatial regressive disturbances of the SARAR(R,S) type under both random and fixed effects assumptions and allowing for heteroskedasticity of arbitrary form in the idiosyncratic

error terms. The regression model is estimated by two-stage least squares (TSLS) to obtain consistent estimates of the disturbances, which are then used in the second step to obtain generalized moments (GM) estimates of the parameters of the spatial regressive disturbance process.

We provide a detailed study of the asymptotic properties of the proposed two-step TSLS and GM estimators of the model parameters, prove their consistency and establish asymptotic normality. Both for the original model and the spatial generalized least squares (GLS) transformed model, we derive the joint and asymptotic variance-covariance matrix, which is robust to (cross-sectional interdependence and) heteroskedasticity of unknown form. This enables robust tests of the general SARAR(R,S) model against restricted alternatives such as SARAR(0,S) and SARAR(R,0) or SARAR(1,1) with random and fixed effects panel data models under heteroskedasticity. We also propose a generalized Hausman-type test of the spatial random versus the spatial fixed effects model.

The framework suggested in the present paper provides a flexible tool for applied econometric researchers for empirical models with cross-sectional interdependence and allows to study the strength and pattern of spatial interdependence more flexibly and under less restrictive assumptions than existing SARAR(1,1) models assuming homoskedasticity. Allowing for alternative modes of interdependence and determining the proper pattern of the interdependence decay function is not only of interest in itself but also a prerequisite for a correct model specification and valid inference.

Table 1. Monte Carlo Results, 1000 draws

	$N = 50$		$N = 100$		$N = 250$		$N = 250$	
	RE	FE	RE	FE	RE	FE	RE	FE
$\lambda_1 = 0.5$							$\lambda_1 = 0.6$	
Bias	0.001082	0.002029	-0.000113	0.000359	0.000237	-0.0000468	0.000161	0.000587
RMSE	0.02395	0.026601	0.018157	0.016493	0.010799	0.011269	0.009209	0.009012
Rej. Rate	0.074	0.068	0.045	0.049	0.046	0.042	0.056	0.048
$\lambda_2 = 0.25$							$\lambda_2 = 0.35$	
Bias	0.000595	0.001409	0.000259	0.000791	-0.000120	-0.0000301	-0.000073	-0.000654
RMSE	0.025345	0.029269	0.019095	0.017378	0.011953	0.011663	0.010053	0.009981
Rej. Rate	0.063	0.062	0.05	0.054	0.065	0.051	0.047	0.052
$\beta_1 = 1$							$\beta_1 = 1$	
Bias	-0.000313	0.000429	-0.000564	-0.000313	0.000187	-0.000447	0.000286	0.00000197
RMSE	0.017017	0.01953	0.013611	0.012649	0.008170	0.008024	0.008164	0.007508
Rej. Rate	0.049	0.057	0.064	0.064	0.049	0.054	0.061	0.042
$\beta_2 = 1$							$\beta_2 = 1$	
Bias	-0.000125	-0.000613	0.000103	-0.000262	0.000005	-0.0000729	-0.000561	0.000248
RMSE	0.018706	0.019158	0.012461	0.011945	0.008016	0.007815	0.00777	0.008334
Rej. Rate	0.057	0.087	0.054	0.047	0.047	0.053	0.048000	0.052000
$\rho_1 = 0.4$							$\rho_1 = 0.4$	
Bias	-0.002348	0.014094	0.000859	-0.00071	0.000162	0.003757	-0.000614	0.004045
RMSE	0.184249	0.954723	0.146618	0.670901	0.121458	0.701775	0.126152	0.747754
Rej. Rate	0.147	0.122	0.126	0.115	0.146	0.131	0.125000	0.133000
$\rho_2 = 0.2$							$\rho_2 = 0.2$	
Bias	-0.016919	-0.035371	-0.004592	-0.010192	-0.000656	-0.005854	-0.005464	-0.011615
RMSE	0.328518	1.146669	0.314663	0.864678	0.305513	0.899012	0.312235	0.945605
Rej. Rate	0.131	0.123	0.131	0.116	0.132	0.115	0.118000	0.128000
Hausman-test								
Rej. Rate	0.106		0.058		0.056		0.054	

References

Amemiya, T. (1971). The estimation of the variances in a variance-components model. *International Economic Review*, 12, 1-13.

Andrews, D. (1987). Asymptotic results of generalized wald tests. *Econometric Theory*, 3, 348-358.

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

Anselin, L., Bera, A.K., Florax, R. and Yoon, M.J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26, 77-104.

Arbia, G., Basile, R., and Piras, G. (2005), Using spatial panel data in modeling regional growth and convergence. ISAE Working Paper no. 55, Rome.

Arraiz, I., Drukker, D.M., Kelejian, H., and Prucha, I. (2010). A spatial Cliff-Ord-type model with heteroskedastic innovations: Small and large sample results. Regional Science and Urban Economics, 50(2), 592-614.

Audretsch, D.B. and Feldmann, M.P. (1996). R&D spillovers and the geography of innovation and production. *American Economic Review*, 86, 630-640.

Badinger, H. and Egger, P. (2008). GM estimation of higher-order spatial autoregressive processes in cross-section models with heteroskedastic disturbances. CESifo Working Paper no. 2356, Munich.

Badinger, H. and Egger, P. (2009). Horizontal versus vertical interdependence in multinational activity. CESifo Working Paper no. 2327, Munich.

Baltagi, B.H. (2005). *Econometric Analysis of Panel Data*, third edition. Chichester: Wiley.

Baltagi, B.H. (2006). Random effects and spatial autocorrelation with equal weights. *Econometric Theory*, 22(5), 973-84.

Baltagi, B.H., Egger, P., and Pfaffermayr, M. (2007). Estimating models of complex FDI: Are there third-country effects? *Journal of Econometrics*, 140(1), 260-281.

Baltagi, B.H., Egger, P., and Pfaffermayr, M. (2009). A generalized spatial panel data model with random effects. Working Paper No. 113, Center for Policy Research, University of Syracuse.

Baltagi, B.H. and Li, D. (2001). LM test for functional form and spatial error correlation. *International Regional Science Review*, 24, 194-225.

Baltagi, B.H., Song, S.H., and Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, 117, 123-150.

Baltagi, B.H., Song, S.H., Jung, B.C., and Koh, W. (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics*, 140, 5-51.

Bell, K.P. and Bockstael, N.E. (2000). Applying the generalized-moments estimation approach to spatial problems involving microlevel data. *The Review of Economics and Statistics,* 82(1), 72–82.

Besley, T. and Case, A. (1995). Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *American Economic Review*, 85, 25-45.

Case, A., Hines Jr., J. and Rosen, H. (1993). Budget spillovers and fiscal policy independence: Evidence from the States. *Journal of Public Economics*, 52, 285-307.

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

Cliff, A. and Ord, J. (1981). *Spatial Processes, Models and Applications*. London: Pion, 1981.

Cohen, J.P. and Morrison Paul, C. (2007). The impacts of transportation infrastructure on property values: A higher-order spatial econometrics approach. *Journal of Regional Science*, 47(3), 457-478.

Cohen, J.P. and Morrison Paul, C.J. (2004). Public infrastructure investment, interstate spatial spillovers, and manufacturing costs. *The Review of Economics and Statistics,* 86(2), 551-560.

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

Creel, M. (2004). Modified Hausman tests for inefficient estimators. *Applied Economics*, 36(21), 2373-2376.

Egger, P., Pfaffermayr, M. and Winner, H. (2005). An unbalanced spatial panel data approach to US state tax competition. *Economics Letters*, 88, 329-335.

Gilbert, S. (2002). Testing the distribution of error components in panel data models. *Economics Letters*, 77, 47-53.

Greene, W.H. (2003). *Econometric Analysis*, fifth edition. Pearson, Upper Saddle River, New Jersey.

Holtz-Eakin, D. (1994). Public sector capital and the productivity puzzle. *Review of Economics and Statistics*, 76, 12-21.

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

Kapoor, M., Kelejian, H.H., and Prucha, I.R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140, 97-130.

Kelejian, H. and Robinson, D. (1992). Spatial autocorrelation: A new computationally simple test with an application to per capita county police expenditures. *Regional Science and Urban Economics*, 22, 317–331.

Kelejian, H.H. and Prucha, I.R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17, 99-121.

Kelejian, H.H. and Prucha, I.R. (1999). A generalized moments estimator for the autoregresssive parameter in a spatial model. *International Economic Review*, 40, 509-533.

Kelejian, H.H. and Prucha, I.R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics*, 118, 27-50.

Kelejian, H.H. and Prucha, I.R. (2007). HAC Estimation in a Spatial Framework. *Journal of Econometrics*, 140(1), 131-154.

Kelejian, H.H. and Prucha, I.R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1), 53-67.

Kelejian, H.H., Prucha, I.R. and Yuzefovich, E. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. In: LeSage, J. and Pace, K. (eds.), *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*. Elsevier, New York, 163-198.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2), 391–413.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6), 1899-1925.

Lee, L.F. and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154(2), 165-185.

Lee, L.F. and Liu, X. (2010). Efficient GMM estimation of high order spatial autoregressive models. *Econometric Theory*, 26, 187-230.

Lin, X. and L.F. Lee (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*, 157(1), 34-52.

Mittelhammer, R.C. (1996). *Mathematical Statistics for Economics and Business*. New York: Springer.

Mundlak, Y. (1978). On Pooling Time Series and Cross Section Data, *Econometrica*, 46(1), 69-85.

Mutl, J. and Pfaffermayr, M. (2011). The Hausman Test in a Cliff and Ord Panel Model. *Econometrics Journal*. 14, 48-76.

Pinkse, J. and Slade, M.E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, 85, 125-154.

Pinkse, J., Slade, M.E., and Brett, C. (2002). Spatial price competition: A semiparametric approach. *Econometrica*, 70, 1111-1153.

Pötscher, B.M. and Prucha, I.R. (1997). *Dynamic Nonlinear Econometric Models, Asymptotic Theory*. New York: Springer.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition. New York: Wiley.

Resnik, S. (1999). *A Probability Path*. Boston: Birkhäuser.

Shroder, M. (1995). Games the States don't play: Welfare benefits and the theory of fiscal federalism. *Review of Economics and Statistics*, 77, 183-191.

Stock and Watson (2008). Heterokedasticity-robust standard errors for fixed effects panel data regression. *Econometrica*, 76(1), 155-174.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *Review of Economic Studies*, 68, 261-295.

Van der Vaart, H.R. and Yen, H.E. (1968). Weak sufficient conditions for Fatou's Lemma and Lebesgue's Dominated Convergence Theorem. *Mathematics Magazine*, 41(3), 109-117.

Vuong, Q.H. (1987). Generalized inverses and asymptotic properties of Wald test. *Economics Letters*, 24, 343-347.

Weesie, J. (1999) Seemingly unrelated estimation and the cluster-adjusted sandwich estimator, *Stata Technical Bulletin* STB-52.

APPENDIX. Variance-Covariance Matrix Under Non-Normality of Error Components

As already mention in the main text, Theorems 4b and 4c as well as Theorem 5 also hold under non-normality with different definitions of Ψ_N and $\Psi_{\theta\Delta,N}$ respectively. In the following, we provide the definitions of the respective elements under non-normality and define consistent estimates for them.

1.1 Distribution of GM Estimates under Non-Normality (Definition of Ψ_N)

If we drop the assumption that μ_N and v_N are normally distributed, equation (28b) becomes

$$\mathfrak{E}_{c,c',N}^{s,s';t,t'} = N^{-1}Cov(\overset{*}{q}_{c,N}^{s,s'}, \overset{*}{q}_{c',N}^{t,t'}) \tag{A.1a}$$

$$= N^{-1}Cov[\xi_N' \overline{\mathbf{A}}_{c,N}^{s,s'}\xi_N + (\overline{\mathbf{a}}_{c,N}^{s,s'})'\xi_N, \xi_N' \overline{\mathbf{A}}_{c',N}^{t,t'}\xi_N + (\overline{\mathbf{a}}_{c',N}^{t,t'})'\xi_N]$$

$$= N^{-1}Cov[\xi_N' \overline{\mathbf{A}}_{c,N}^{s,s'}\xi_N + (\mathbf{a}_{c,v,N}^{s,s'})'\mathbf{v}_N + (\mathbf{a}_{c,\mu,N}^{s,s'})'\mu_N, \xi_N' \overline{\mathbf{A}}_{c',N}^{t,t'}\xi_N + (\mathbf{a}_{c',v,N}^{t,t'})'\mathbf{v}_N + (\mathbf{a}_{c',\mu,N}^{t,t'})'\mu_N]$$

$$= 2N^{-1}Tr(\mathbf{A}_{c,v,N}^{s,s'}\boldsymbol{\Sigma}_N \mathbf{A}_{c',v,N}^{t,t'}\boldsymbol{\Sigma}_N + 2\sigma_\mu^2(\mathbf{A}_{c,v,\mu,N}^{s,s'})'\boldsymbol{\Sigma}_N \mathbf{A}_{c',v,\mu,N}^{t,t'} + \sigma_\mu^4 \mathbf{A}_{c,\mu,N}^{s,s'}\mathbf{A}_{c',\mu,N}^{t,t'}).$$

$$+ N^{-1}[(\mathbf{a}_{c,v,N}^{s,s'})'\boldsymbol{\Sigma}_N \mathbf{a}_{c',v,N}^{t,t'} + \sigma_\mu^2(\mathbf{a}_{c,\mu,N}^{s,s'})'\mathbf{a}_{c',\mu,N}^{t,t'}]$$

$$+ N^{-1}\sum_{n=1}^{NT} a_{c,v,nn,N}^{s,s'} a_{c',v,nn,N}^{t,t'}(\sigma_{v,n,N}^{(4)} - 3\sigma_{v,n,N}^4) + (\sigma_\mu^{(4)} - 3\sigma_\mu^4)N^{-1}\sum_{i=1}^{N} a_{c,\mu,ii,N}^{s,s'} a_{c',\mu,ii,N}^{t,t'}$$

$$+ N^{-1}\sum_{n=1}^{NT}(a_{c,v,N}^{s,s'}a_{c',v,nn,N}^{t,t'} + a_{c,v,nn,N}^{s,s'}a_{c',v,N}^{t,t'})\sigma_{v,n,N}^{(3)} + \sigma_\mu^{(3)}N^{-1}\sum_{i=1}^{N}(a_{c,\mu,i,N}^{s,s'}a_{c',\mu,ii,N}^{t,t'} + a_{c,\mu,ii,N}^{s,s'}a_{c',\mu,i,N}^{t,t'}).$$

Adopting the notational convention introduced in section V, subsection 1.2, (A.1a) can be written as

$$\mathfrak{E}_N = \mathfrak{E}_{v,N}^{*,\circ} + \mathfrak{E}_{\mu,N}^{*} + \mathfrak{E}_{v\mu,N}^{*} + \mathfrak{E}_{v,N}^{**} + \mathfrak{E}_{\mu,N}^{**} \tag{A.1b}$$

$$+ \mathfrak{E}_{v,N}^{***} + \mathfrak{E}_{\mu,N}^{***} + \mathfrak{E}_{v,N}^{****} + \mathfrak{E}_{\mu,N}^{****},$$

i.e., $\mathfrak{E}_{v,N}^{*} = 0$, and the additional terms, appearing in the second row of (A.1b), are defined as:

$$\mathfrak{E}_{v,N}^{***} = N^{-1}\sum_{n=1}^{NT} c_{v,n,N}(\sigma_{v,n,N}^{(4)} - \sigma_{v,n,N}^4) \tag{A.2a}$$

$$\mathfrak{E}_{\mu,N}^{***} = (\sigma_\mu^{(4)} - 3\sigma_\mu^4)N^{-1}\sum_{i=1}^{N} c_{\mu,ii,N} \tag{A.2b}$$

$$\mathfrak{E}_{v,N}^{****} = N^{-1}\sum_{n=1}^{NT} d_{v,n,N}\sigma_{v,n,N}^{(3)} \tag{A.2c}$$

$$\mathfrak{E}_{\mu,N}^{****} = \sigma_\mu^{(3)}N^{-1}\sum_{i=1}^{N} d_{\mu,i,N}, \tag{A.2d}$$

where $\sigma_\mu^{(3)}$ and $\sigma_\mu^{(4)}$ ($\sigma_v^{(3)}$ and $\sigma_v^{(4)}$) denote the third and fourth moments of $\boldsymbol{\mu}_N$ ($\mathbf{v}_N$), respectively.[22]

As shown in Lemma C.4 of Appendix C, the third and fourth moments of $\mu_{i,N}$, denoted as $\sigma_\mu^{(3)}$ and $\sigma_\mu^{(4)}$, can be estimated consistently using

$$\tilde{\sigma}_{\mu,N}^{(3)} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}_{it}^2 \text{, and} \tag{A.3a}$$

$$\tilde{\sigma}_{\mu,N}^{(4)} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}_{it,N}^3 \tag{A.3b}$$

$$- \frac{3}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}_{it,N} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\varepsilon}_{it,N}^2 - \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}_{it,N} \right),$$

where $\tilde{\boldsymbol{\varepsilon}}_N = \mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \tilde{\rho}_{m,N} \mathbf{M}_{m,N}) \tilde{\mathbf{u}}_N$.

Hence, consistent estimators of the expressions in (A1), associated (only) with the homoskedastic, time-invariance error component $\mu_{i,N}$, are given by

$$\tilde{\mathfrak{E}}_{\boldsymbol{\mu},N}^{***} = (\tilde{\sigma}_\mu^{(4)} - 3\tilde{\sigma}_\mu^4) N^{-1} \sum_{i=1}^{N} c_{\boldsymbol{\mu},ii,N}, \tag{A.4a}$$

$$\tilde{\mathfrak{E}}_{\boldsymbol{\mu},N}^{****} = \tilde{\sigma}_\mu^{(3)} N^{-1} \sum_{i=1}^{N} \tilde{d}_{\boldsymbol{\mu},i,N} \text{ with } \tilde{d}_{\boldsymbol{\mu},n,N} = (\tilde{a}_{\boldsymbol{\mu},n,N} b_{\boldsymbol{\mu},nn,N} + a_{\boldsymbol{\mu},nn,N} \tilde{b}_{\boldsymbol{\mu},n,N}). \tag{A.4b}$$

Next turn to $\mathfrak{E}_{\mathbf{v},N}^{***}$, which we rewrite as

$$\mathfrak{E}_{\mathbf{v},N}^{***} = \mathfrak{E}_{\mathbf{v}1,N}^{***} - \mathfrak{E}_{\mathbf{v}2,N}^{***} \text{ with } \mathfrak{E}_{\mathbf{v}1,N}^{***} = N^{-1} \sum_{n=1}^{NT} c_{\mathbf{v},n,N} \sigma_{v,n,N}^{(4)} \text{ and } \mathfrak{E}_{\mathbf{v}2,N}^{***} = N^{-1} \sum_{n=1}^{NT} c_{\mathbf{v},n,N} \sigma_{v,n,N}^4. \tag{A.5a}$$

We first consider $\mathfrak{E}_{\mathbf{v}1,N}^{***}$ and note that the elements $c_{\mathbf{v},n,N}$ are time-invariant. By Lemma C.6a, a consistent estimator of $\mathfrak{E}_{\mathbf{v}1,N}^{***}$ is given by

---

[22] For the elements, where both $c = 2$ (or $c = 4$) and $c' = 2$ (or $c' = 4$), the terms involving the third and fourth moments drop out.

$$\widetilde{\mathfrak{E}}^{***}_{\mathbf{v}1,N} = \frac{k_0}{1-k_1 m_1} \frac{1}{NT} \sum_{i=1}^{N} c_{\mathbf{v},i,N} \sum_{t=1}^{T} \widetilde{\underline{v}}^4_{it,N} + \frac{k_1 m_0}{1-k_1 m_1} \frac{1}{N} \sum_{i=1}^{N} c_{\mathbf{v},i,N} \frac{1}{T} \sum_{t=1}^{T} \widetilde{\underline{v}}^2_{it,N} \frac{1}{T-1} \sum_{\substack{r=1 \\ r \neq t}}^{T} \widetilde{\underline{v}}^2_{ir,N} \;, \quad \text{(A.5a)}$$

where

$$m_0 = \frac{T^3}{T^3 - 2T^2 - 3T + 9} \;,\; m_1 = -\frac{2T-3}{T^3 - 2T^2 - 3T + 9},$$
$$k_0 = \frac{T^2(T-1)}{(T^3 - 4T^2 + 6T - 3)} \;,\; k_1 = -\frac{(T-1)(6T-9)}{(T^3 - 4T^2 + 6T - 3)} \;.$$

Next consider $\mathfrak{E}^{***}_{\mathbf{v}2,N}$, which involves a weighted sum of the squared variance. Without distributional assumptions and unknown heteroskedasticity over both cross-sections and time, it is not possible to obtain an estimates of (a weighted sum of) the squared variances. (Using the fourth power of the residuals estimates a weighted sum of the fourth moments.) Hence, an approximation is required, assuming that the idiosyncratic error components are heteroskedastic only over cross-sections, but not over time, i.e., $v_{it,N} \sim i.d.(0,\sigma_i^2)$. Under that assumption, the following expression consistently estimates $\mathfrak{E}^{***}_{\mathbf{v}2,N}$ as shown in Lemma C.6b in Appendix C:

$$\widetilde{\mathfrak{E}}^{***}_{\mathbf{v}2,N} = \frac{m_0}{1-m_1 k_1} \frac{1}{N} \sum_{i=1}^{N} c_{\mathbf{v},i,N} \frac{1}{T} \sum_{t=1}^{T} \widetilde{\underline{v}}^2_{it,N} \frac{1}{T-1} \sum_{\substack{r=1 \\ r \neq t}}^{T} \widetilde{\underline{v}}^2_{ir,N} + \frac{m_1 k_0}{1-m_1 k_1} \frac{1}{NT} \sum_{i=1}^{N} c_{\mathbf{v},i,N} \sum_{t=1}^{T} \widetilde{\underline{v}}^4_{it,N} \;, \quad \text{(A.5b)}$$

where m_0, m_1, k_0, k_1 are defined as above.

Finally, a consistent estimate of $\mathfrak{E}^{****}_{\mathbf{v},N}$ is given by

$$\widetilde{\mathfrak{E}}^{****}_{\mathbf{v},N} = \frac{T}{N(T-1)} \sum_{n=1}^{NT} \widetilde{d}_{\mathbf{v},n,N} (\widetilde{v}^{HR}_{n,N})^3 \quad \text{with} \quad \widetilde{d}_{\mathbf{v},n,N} = (\widetilde{a}_{\mathbf{v},n,N} b_{\mathbf{v},nn,N} + a_{\mathbf{v},nn,N} \widetilde{b}_{\mathbf{v},n,N}) \,, \quad \text{(A.6)}$$

where $(\widetilde{v}^{HR}_{it,N})^3 = \frac{T(T-1)}{T^2 - 3T + 3} [\widetilde{\underline{v}}^3_{it,N} + \frac{1}{(T-1)^3} \sum_{r=1}^{T} \widetilde{\underline{v}}^3_{ir,N}]$. The consistency of $\widetilde{\mathfrak{E}}^{****}_{\mathbf{v},N}$ follows from Lemma C.7 and Remark C.3 thereafter in Appendix C.

1.2 Joint Distribution of Regression Parameters and GM Estimates under Non-normality (Definition of $\Psi_{\theta\Delta,N}$)

Under non-normality, equation (32c) becomes augmented by terms involving the third moments of the error components as follows:

$$\Psi_{\Delta\theta,..,(c,s,s'),N} = \Psi^{\mathbf{v}}_{\Delta\theta,..,(c,s,s'),N} + \Psi^{\mu}_{\Delta\theta,..,(c,s,s'),N} \,, \quad c = 1,...,4, \; s,s' = 1,...,S, \; \text{with} \quad \text{(A.7a)}$$

$$\boldsymbol{\psi}^{\mathbf{v}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N} = \frac{1}{N}\frac{1}{T^{1/2}}[\mathbf{F}'_{\mathbf{v},N}(\boldsymbol{\Sigma}^{(3)}_N\boldsymbol{\kappa}_{\mathbf{A}^{s,s'}_{c,\mathbf{v},N}} + \boldsymbol{\Sigma}_N\mathbf{a}^{s,s'}_{c,\mathbf{v},N})]\text{, and}$$

$$\boldsymbol{\psi}^{\boldsymbol{\mu}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N} = \frac{1}{N}\frac{1}{T^{1/2}}[\mathbf{F}'_{\boldsymbol{\mu},N}(\sigma^{(3)}_\mu\boldsymbol{\kappa}_{\mathbf{A}^{s,s'}_{c,\boldsymbol{\mu},N}} + \sigma^2_\mu\mathbf{a}^{s,s'}_{c,\boldsymbol{\mu},N})\text{,}$$

where $\boldsymbol{\Sigma}^{(3)}_N = diag(\sigma^{(3)}_{v,n})$ is an $NT \times NT$ diagonal matrix with third moments $\sigma^{(3)}_{v,n}$, $n = 1,...,NT$, $\boldsymbol{\kappa}_{\mathbf{A}^{s,s'}_{c,\mathbf{v},N}}$ is an $NT \times 1$ vector with the main diagonal elements of $\mathbf{A}^{s,s'}_{c,\mathbf{v},N}$, and $\boldsymbol{\kappa}_{\mathbf{A}^{s,s'}_{c,\boldsymbol{\mu},N}}$ is an $N \times 1$ vector with the main diagonal elements of $\mathbf{A}^{s,s'}_{c,\boldsymbol{\mu},N}$.

In light (32c) and the results of section 1.2, the assumptions maintained in Theorem 4b are sufficient to prove that the following expressions consistently estimate the elements of $\boldsymbol{\Psi}_{\boldsymbol{\theta}\Delta,N}$:

$$\widetilde{\boldsymbol{\psi}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N} = \widetilde{\boldsymbol{\psi}}^{\mathbf{v}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N} + \widetilde{\boldsymbol{\psi}}^{\boldsymbol{\mu}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N}\text{, } c = 1,...,4\text{, } s,s' = 1,...,S\text{, with} \qquad (\text{A.7b})$$

$$\boldsymbol{\psi}^{\mathbf{v}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N} = \frac{1}{N}\frac{T^{1/2}}{(T-1)}[\mathbf{F}'_{\mathbf{v},N}(\widetilde{\boldsymbol{\Sigma}}^{HR,(3)}_N\boldsymbol{\kappa}_{\mathbf{A}^{s,s'}_{c,\mathbf{v},N}} + \widetilde{\boldsymbol{\Sigma}}^{HR}_N\widetilde{\mathbf{a}}^{s,s'}_{c,\mathbf{v},N})]\text{, and}$$

$$\widetilde{\boldsymbol{\psi}}^{\boldsymbol{\mu}}_{\Delta\boldsymbol{\theta},..,(c,s,s'),N} = \frac{1}{N}\frac{1}{T^{1/2}}[\mathbf{F}'_{\boldsymbol{\mu},N}(\widetilde{\sigma}^{(3)}_\mu\boldsymbol{\kappa}_{\mathbf{A}^{s,s'}_{c,\boldsymbol{\mu},N}} + \widetilde{\sigma}^2_\mu\widetilde{\mathbf{a}}^{s,s'}_{c,\boldsymbol{\mu},N})\text{.}$$

TECHNICAL APPENDIX

APPENDIX A

Notation

We adopt the standard convention to refer to matrices and vectors with acronyms in boldface. Let $\mathbf{A}_N$ denote some matrix. Its elements are referred to as $a_{ij,N}$; $\mathbf{a}_{i.,N}$ and $\mathbf{a}_{.i,N}$ denote the i-th row and the i-th column of $\mathbf{A}_N$ respectively. If $\mathbf{A}_N$ is a square matrix, $\mathbf{A}_N^{-1}$ denotes its inverse; if $\mathbf{A}_N$ is singular, $\mathbf{A}_N^{+}$ denotes its generalized inverse. The (submultiplicative) matrix norm $\|\cdot\|$ is defined as $\|\mathbf{A}_N\| = [Tr(\mathbf{A}_N'\mathbf{A}_N)]^{1/2}$. In several places, we use single indexation, e.g., $n = 1,...,NT$, to denote elements of the vectors or matrices that are stacked over time periods.[23]

Remark A.1

i) Definition of row and column sum boundedness (Kapoor, Kelejian, and Prucha, 2007, p. 99): Let $\mathbf{A}_N, N \geq 1$, be some sequence of $NT \times NT$ matrices with T some fixed positive integer. We will then say that the row and column sums of the (sequence of) matrices $\mathbf{A}_N$ are bounded uniformly in absolute value, if there exists a constant $c < \infty$, which does not depend on N, such that

$$\max_{1 \leq n \leq NT} \sum_{j=1}^{NT} |a_{nj,N}| \leq c \text{ and } \max_{1 \leq j \leq NT} \sum_{n=1}^{NT} |a_{nj,N}| \leq c \text{ for all } N \geq 1.$$

ii) Let $\mathbf{A}_N$ be a (sequence of) $N \times N$ matrices whose row and column sums are bounded uniformly in absolute value, and let $\mathbf{S}$ be some $T \times T$ matrix (with $T \geq 1$ fixed). Then the row and column sums of the matrix $\mathbf{S} \otimes \mathbf{A}_N$ are bounded uniformly in absolute value (compare Kapoor, Kelejian, and Prucha, 2007, p. 118).

iii) If $\mathbf{A}_N$ and $\mathbf{B}_N$ are (sequences of) $NT \times NT$ matrices (with $T \geq 1$ fixed), whose row and column sums are bounded uniformly in absolute value (by c_A and c_B), then so are the row and column sums of $\mathbf{A}_N\mathbf{B}_N$ and $\mathbf{A}_N + \mathbf{B}_N$ (by $c_A c_B$ and $c_A + c_B$). If $\mathbf{Z}_N$ is a (sequence of) $NT \times P$ matrices whose elements are bounded uniformly in absolute value, then so are the elements of $\mathbf{A}_N\mathbf{Z}_N$ and $(NT)^{-1}\mathbf{Z}_N'\mathbf{A}_N\mathbf{Z}_N$. Of course, this also covers the case $(NT)^{-1}\mathbf{Z}_N'\mathbf{Z}_N$ for $\mathbf{A}_N = \mathbf{I}_{NT}$ (compare Kapoor, Kelejian, and Prucha, 2007, p. 119).

---

[23] Take the vector $\mathbf{v}_N = (\mathbf{v}_{1,N}',...,\mathbf{v}_{T,N}')'$, for example. Using indexation $n = 1,...,NT$, the elements $v_{n,N}, n = 1,...,N$, refer to period $t = 1$, elements $u_{n,N}, n = N+1,...,2N$ refer to $t = 2$, etc., and elements $u_{n,N}, n = (T-1)N+1,...,NT$ refer to period $t = T$.

iv) Suppose that the row and columns sums of the $NT \times NT$ matrices $\mathbf{A}_N = (a_{ij,N})$ are bounded uniformly in absolute value by some finite constant c_A; then $\sum_{n=1}^{NT} |a_{nj,N}|^q \le c_A^q$ for $q > 1$ (see Kelejian and Prucha, 2009, Remark C.1).

v) Let $\boldsymbol{\xi}_N$ and $\boldsymbol{\eta}_N$ be $NT \times 1$ random vectors (with $T \ge 1$ fixed), where, for each N, the elements are independently distributed with zero mean and finite variances. Then the elements of $(NT)^{-1/2} \mathbf{Z}'_N \boldsymbol{\xi}_N$ are $O_p(1)$ and $(NT)^{-1} \boldsymbol{\xi}'_N \mathbf{A}_N \boldsymbol{\eta}_N$ is $O_p(1)$.[24]

vi) Let $\boldsymbol{\zeta}_N$ be a $NT \times 1$ random vector (with $T \ge 1$ fixed), where, for each N, the elements are distributed with zero mean and finite fourth moments. Let $\boldsymbol{\pi}_N$ be some nonstochastic $NT \times 1$ vector, whose elements are bounded uniformly in absolute value and let $\boldsymbol{\Pi}_N$ be a $NT \times NT$ nonstochastic matrix whose row and column sums are bounded uniformly in absolute value. Define the column vector $\mathbf{d}_N = \boldsymbol{\pi}_N + \boldsymbol{\Pi}_N \boldsymbol{\zeta}_N$. It follows that the elements of $\mathbf{d}_N$ have finite fourth moments.[25]

Remark A.2

The matrices $\mathbf{Q}_{0,N}$ and $\mathbf{Q}_{1,N}$ have the following properties (see Kapoor, Kelejian, and Prucha, 2007, p. 101):

$$tr(\mathbf{Q}_{0,N}) = N(T-1), \ tr(\mathbf{Q}_{1,N}) = N, \ \mathbf{Q}_{0,N}(\mathbf{e}_T \otimes \mathbf{I}_N) = \mathbf{0}, \ \mathbf{Q}_{1,N}(\mathbf{e}_T \otimes \mathbf{I}_N) = (\mathbf{e}_T \otimes \mathbf{I}_N),$$

$$\mathbf{Q}_{0,N} \boldsymbol{\varepsilon}_N = \mathbf{Q}_{0,N} \mathbf{v}_N, \ \mathbf{Q}_{1,N} \boldsymbol{\varepsilon}_N = (\mathbf{e}_T \otimes \mathbf{I}_N)\boldsymbol{\mu}_N + \mathbf{Q}_{1,N} \mathbf{v}_N, \ (\mathbf{I}_T \otimes \mathbf{D}_N)\mathbf{Q}_{0,N} = \mathbf{Q}_{0,N}(\mathbf{I}_T \otimes \mathbf{D}_N),$$

$$(\mathbf{I}_T \otimes \mathbf{D}_N)\mathbf{Q}_{1,N} = \mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{D}_N), \ tr[(\mathbf{I}_T \otimes \mathbf{D}_N)\mathbf{Q}_{0,N}] = (T-1)tr(\mathbf{D}_N),$$

$$tr[(\mathbf{I}_T \otimes \mathbf{D}_N)\mathbf{Q}_{1,N}] = tr(\mathbf{D}_N),$$

---

[24] Kelejian and Prucha (2004) consider the case $T = 1$ and identically distributed elements of $\boldsymbol{\xi}_N$ and $\boldsymbol{\eta}_N$. Results hold up for (fixed) $T \ge 1$ and under heteroskedasticity, as long as the variances of the elements of $\boldsymbol{\xi}_N$ and $\boldsymbol{\eta}_N$ are bounded uniformly in absolute value.

[25] Kelejian and Prucha (2009, Lemma C.2) give a proof for $T = 1$ and independent elements of $\boldsymbol{\zeta}_N$. The extension to (fixed) $T \ge 1$ is obvious. Independence of the elements of $\boldsymbol{\zeta}_N$ is not required for the result to hold. The fourth moments of the elements of $\mathbf{d}_N = \boldsymbol{\pi}_N + \boldsymbol{\Pi}_N \boldsymbol{\zeta}_N$ are given by $E(\pi_{i,N} + \sum_{j=1}^{NT} \pi_{ij,N} \zeta_{j,N})^4 \le 2^4 E[\pi_{i,N}^4 + (\sum_{j=1}^{NT} \pi_{ij,N} \zeta_{j,N})^4]$

$\le 2^4 [\pi_{i,N}^4 + \sum_{j=1}^{NT} \sum_{k=1}^{NT} \sum_{l=1}^{NT} \sum_{m=1}^{NT} |\pi_{ij,N}| \|\pi_{ik,N}\| \|\pi_{il,N}\| |\pi_{im,N}| E|\zeta_{j,N}| \|\zeta_{k,N}\| \|\zeta_{l,N}\| |\zeta_{m,N}|] \le K < \infty$, by Hölder's inequality as long as the fourth moments of the elements of $\boldsymbol{\zeta}_N$ are bounded uniformly.

where $\mathbf{D}_N$ is an arbitrary $N \times N$ matrix. Obviously, the row and column sums of $\mathbf{Q}_{0,N}$ and $\mathbf{Q}_{1,N}$ are bounded uniformly in absolute value.

APPENDIX B

Lemma B.1[26]

Let $\mathbf{A}_N$ be some nonstochastic $NT \times NT$ matrix (with T fixed), whose row and column sums are bounded uniformly in absolute value. Let $\mathbf{u}_N$ be defined by (2c) and $\tilde{\mathbf{u}}_N$ be a predictor for $\mathbf{u}_N$. Suppose that Assumptions 1 to 4 hold. Then

(a) $N^{-1}E|\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N| = O(1)$, $Var(N^{-1}\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N) = o(1)$, and $N^{-1}(\tilde{\mathbf{u}}'_N\mathbf{A}_N\tilde{\mathbf{u}}_N) - N^{-1}E(\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N) = o_p(1)$.

(b) $N^{-1}E|\mathbf{d}'_{.j,N}\mathbf{A}_N\mathbf{u}_N| = O(1)$, $j = 1,...,P$, where $\mathbf{d}_{.j,N}$ is the j-th column of the $NT \times P$ matrix $\mathbf{D}_N$, and $N^{-1}\mathbf{D}'_N\mathbf{A}_N\tilde{\mathbf{u}}_N - N^{-1}E(\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N) = o_p(1)$.

(c) If furthermore Assumption 6 holds, then
$N^{-1/2}\tilde{\mathbf{u}}'_N\mathbf{A}_N\tilde{\mathbf{u}}_N = N^{-1/2}\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N + \boldsymbol{\alpha}'_N N^{1/2}\boldsymbol{\Delta}_N + o_p(1)$ with $\boldsymbol{\alpha}_N = N^{-1}E[\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N]$.
In light of (b), we have $\boldsymbol{\alpha}_N = O(1)$ and $N^{-1}\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\tilde{\mathbf{u}}_N - \boldsymbol{\alpha}_N = o_p(1)$.

Proof of part (a)
Let

$$\vartheta_N = N^{-1}\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N \text{ and } \tilde{\vartheta}_N = N^{-1}\tilde{\mathbf{u}}'_N\mathbf{A}_N\tilde{\mathbf{u}}_N. \tag{B.1}$$

Given (4a), we have $\vartheta_N = N^{-1}\boldsymbol{\varepsilon}'_N\mathfrak{S}_N\boldsymbol{\varepsilon}_N$, with the symmetric $NT \times NT$ matrix $\mathfrak{S}_N$ defined as

$$\mathfrak{S}_N = (1/2)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}'_{m,N})^{-1}](\mathbf{A}_N + \mathbf{A}'_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}] \tag{B.2}$$

By Assumptions 1-3 and Remark A.1 in Appendix A, the row and column sums of the matrices $\mathfrak{S}_N$ are bounded uniformly in absolute value. Let $\boldsymbol{\Omega}_{\varepsilon,N} = \sigma_\mu^2(\mathbf{J}_T \otimes \mathbf{I}_N) + \boldsymbol{\Sigma}_N$, then given Assumption 2, the row and column sums of the matrices $\mathfrak{S}_N\boldsymbol{\Omega}_{\varepsilon,N}\mathfrak{S}_N\boldsymbol{\Omega}_{\varepsilon,N}$ are bounded uniformly in absolute value.

In the following let $K < \infty$ be a common bound for the row and column sums of the absolute value of the elements of $\mathfrak{S}_N, \boldsymbol{\Omega}_{\varepsilon,N}$, and $\mathfrak{S}_N\boldsymbol{\Omega}_{\varepsilon,N}\mathfrak{S}_N\boldsymbol{\Omega}_{\varepsilon,N}$ and of the absolute value of their respective elements. Then

$$E|\vartheta_N| = E\left|N^{-1}\sum_{n=1}^{NT}\sum_{j=1}^{NT}b_{nj,N}\varepsilon_{n,N}\varepsilon_{j,N}\right| \tag{B.3}$$

---

[26] Compare Lemma C.1 in Kelejian and Prucha (2009) for the case of a cross-sectional SARAR(1,1) model and Lemma C.1 in Badinger and Egger (2008b) for the case of a cross-sectional SARAR(R,S) model.

$$\leq N^{-1}\sum_{n=1}^{NT}\sum_{j=1}^{NT}\left|b_{nj,N}\right|E\left|\varepsilon_{n,N}\right|\left|\varepsilon_{j,N}\right|$$

$$\leq N^{-1}\sum_{n=1}^{NT}\sum_{j=1}^{NT}\left|b_{nj,N}\right|\sigma_{\varepsilon,n}\sigma_{\varepsilon,j}\leq TK^{3},$$

where we used Hölder's inequality in the last step. This proves that $E\left|\vartheta_{N}\right|$ is O(1).

Now consider $Var(\vartheta_{N})$, rewriting ϑ_{N} as quadratic form in $\xi_{N}=(\mathbf{v}'_{N},\boldsymbol{\mu}'_{N})'$ and invoking Lemma A.1 in Kelejian and Prucha (2009):

$$Var(\vartheta_{N})=Cov(N^{-1}\boldsymbol{\varepsilon}'_{N}\mathfrak{S}_{N}\boldsymbol{\varepsilon}_{N},N^{-1}\boldsymbol{\varepsilon}'_{N}\mathfrak{S}_{N}\boldsymbol{\varepsilon}_{N}) \qquad (\text{B.4})$$

$$=N^{-2}Cov(\xi'_{N}\overline{\mathfrak{S}}_{N}\xi_{N},\xi'_{N}\overline{\mathfrak{S}}_{N}\xi_{N})$$

$$=2N^{-2}Tr(\overline{\mathfrak{S}}_{N}\boldsymbol{\Omega}_{\xi_{N}}\overline{\mathfrak{S}}_{N}\boldsymbol{\Omega}_{\xi_{N}})+N^{-2}\sum_{n=1}^{NT}\bar{s}_{nn,*}^{2}[E(\eta_{n,N}^{4})-3],$$

$$=2N^{-2}Tr(\overline{\mathfrak{S}}_{N}\boldsymbol{\Omega}_{\xi_{N}}\overline{\mathfrak{S}}_{N}\boldsymbol{\Omega}_{\xi_{N}})+N^{-2}Tr\{diag_{n=1,...,NT}(\bar{s}_{nn*,N}^{2})diag_{n=1,...,NT}[E(\eta_{n,N}^{4})-3]\},$$

where $\overline{\mathfrak{S}}_{N}$ is a $N(T+1)\times N(T+1)$ matrix, whose elements and row and column sums are bounded uniformly in absolute value by some constant K^{*}. Next, $\bar{s}_{nn*,N}$ is the n-th diagonal element of $\overline{\mathfrak{S}}_{N}^{*}=(s_{nn,*,N})=\mathbf{S}'_{N}\overline{\mathfrak{S}}'_{N}\mathbf{S}_{N}$, with $\mathbf{S}_{N}\mathbf{S}'_{N}=\boldsymbol{\Omega}_{\xi_{N}}$, where $\boldsymbol{\Omega}_{\xi_{N}}$ is the variance-covariance matrix of ξ_{N}, which is diagonal with elements $\sigma_{v,n}^{2}$ for $n=1,...,NT$ and elements σ_{μ}^{2} for $n=NT+1,...,N(T+1)$. Finally, the vector $\boldsymbol{\eta}_{N}=\mathbf{S}_{N}^{-1}\xi_{N}$. In light of Assumption 1, the row and column sums (and the elements) of $\mathbf{S}_{N}$ are bounded uniformly in absolute value by some finite constant, say K^{**}. Moreover, the row and column sums (and the elements) of $\mathbf{S}_{N}^{-1}$ are also bounded uniformly in absolute value by some constant K^{***}. Finally, in light of Remark A.1 and Assumption 1 it follows that the elements of $\boldsymbol{\eta}_{N}=\mathbf{S}^{-1}\xi_{N}$ have finite fourth moments. Denote their bound by K^{****}. Without loss of generality we assume that the bound K used above is chosen such that $K^{*}\leq K$, $K^{**}\leq K$, $K^{***}\leq K$, and $K^{****}\leq K$. Hence, we have

$$Var(\vartheta_{N})\leq 2N^{-2}Tr(K\mathbf{I}_{N(T+1)})+N^{-2}Tr[diag_{n=1,...,N(T+1)}(K^{2}K)]=N^{-1}(T+1)(2K+K^{3})=o(1).$$

The claim in part (a) of Lemma B.1 that $N^{-1}(\mathbf{u}'_{N}\mathbf{A}_{N}\mathbf{u}_{N})-N^{-1}E(\mathbf{u}'_{N}\mathbf{A}_{N}\mathbf{u}_{N})=o_{p}(1)$ now follows from Chebychev's inequality (see, for example, White, 2001, p. 35).

We now prove the second part of (a), i.e., $N^{-1}(\tilde{\mathbf{u}}'_{N}\mathbf{A}_{N}\tilde{\mathbf{u}}_{N})-N^{-1}E(\mathbf{u}'_{N}\mathbf{A}_{N}\mathbf{u}_{N})=o_{p}(1)$. Since $\vartheta_{N}-E(\vartheta_{N})=o_{p}(1)$, it suffices to show that $\tilde{\vartheta}_{N}-\vartheta_{N}=o_{p}(1)$. By Assumption 4, we have

$\tilde{\mathbf{u}}_N - \mathbf{u}_N = \mathbf{D}_N \Delta_N$, where $\mathbf{D}_N = (\mathbf{d}'_{1.,N},...,\mathbf{d}'_{NT.,N})'$. Substituting $\tilde{\mathbf{u}}_N = \mathbf{u}_N + \mathbf{D}_N \Delta_N$ into the expression for $\tilde{\vartheta}_N$ in (B.1), we obtain

$$\tilde{\vartheta}_N - \vartheta_N = N^{-1}(\mathbf{u}'_N + \Delta'_N \mathbf{D}'_N)\mathbf{A}_N(\mathbf{u}_N + \mathbf{D}_N \Delta_N) - N^{-1}\mathbf{u}'_N \mathbf{A}_N \mathbf{u}_N \qquad (\text{B.5})$$
$$= N^{-1}[\Delta'_N \mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N + \Delta'_N \mathbf{D}'_N \mathbf{A}_N \mathbf{D}_N \Delta_N]$$
$$= \phi_N + \psi_N,$$

where

$$\phi_N = N^{-1}[\Delta'_N \mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N], \qquad (\text{B.6})$$
$$= N^{-1}\{\Delta'_N \mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]\boldsymbol{\varepsilon}_N\}$$
$$= N^{-1}(\Delta'_N \mathbf{D}'_N \mathbf{C}_N \boldsymbol{\varepsilon}_N),$$

with $\mathbf{C}_N = (\mathbf{A}_N + \mathbf{A}'_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}] = (\mathbf{c}'_{1.,N},..,\mathbf{c}'_{NT.,N})'$, and

$$\psi_N = N^{-1}\Delta'_N \mathbf{D}'_N \mathbf{A}_N \mathbf{D}_N \Delta_N. \qquad (\text{B.7})$$

By Assumption 3 and Remark A.1, the row and column sums of $\mathbf{C}_N$ are bounded uniformly in absolute value. We next prove that $\phi_N = o_p(1)$ and $\psi_N = o_p(1)$.

Proof that $\phi_N = o_p(1)$:

$$|\phi_N| = N^{-1}|\Delta'_N \mathbf{D}'_N \mathbf{C}_N \boldsymbol{\varepsilon}_N| \qquad (\text{B.8})$$
$$= N^{-1}\left\|\sum_{n=1}^{NT}\Delta'_N \mathbf{d}'_{n.,N}\mathbf{c}_{n.,N}\boldsymbol{\varepsilon}_N\right\|$$
$$\leq N^{-1}\sum_{n=1}^{NT}\|\Delta'_N\|\|\mathbf{d}'_{n.,N}\|\|\mathbf{c}_{n.,N}\boldsymbol{\varepsilon}_N\|$$
$$= N^{-1}\|\Delta'_N\|\sum_{n=1}^{NT}\|\mathbf{d}'_{n.,N}\|\left\|\sum_{j=1}^{NT}c_{nj,N}\varepsilon_{j,N}\right\|$$
$$\leq N^{-1}\|\Delta'_N\|\sum_{n=1}^{NT}\|\mathbf{d}'_{n.,N}\|\sum_{j=1}^{NT}\|c_{nj,N}\varepsilon_{j,N}\|$$
$$= N^{-1}\|\Delta'_N\|\sum_{n=1}^{NT}\|\mathbf{d}'_{n.,N}\|\sum_{j=1}^{NT}|c_{nj,N}||\varepsilon_{j,N}|$$
$$= N^{-1}\|\Delta'_N\|\sum_{j=1}^{NT}|\varepsilon_{j,N}|\sum_{n=1}^{NT}\|\mathbf{d}'_{n.,N}\||c_{nj,N}|$$
$$\leq N^{-1}\|\Delta'_N\|\sum_{j=1}^{NT}|\varepsilon_{j,N}|\left(\sum_{n=1}^{NT}\|\mathbf{d}'_{n.,N}\|^{p}\right)^{1/p}\left(\sum_{n=1}^{NT}|c_{nj,N}|^{q}\right)^{1/q}$$

$$= N^{1/p-1/2}\left(N^{1/2}\|\mathbf{\Delta}_N\|\right)\left(N^{-1}\sum_{j=1}^{NT}\left|\varepsilon_{j,N}\right|\right)\left(N^{-1}\sum_{n=1}^{NT}\left\|\mathbf{d}'_{n.,N}\right\|^p\right)^{1/p}\left(\sum_{n=1}^{NT}\left|c_{nj,N}\right|^q\right)^{1/q}.$$

Note that $\left(\sum_{n=1}^{NT}\left|c_{nj,N}\right|\right)\leq\overline{K}<\infty$ by Assumption. In the following we denote by $\overline{K}$ the uniform bound for the row and column sums of the absolute value of the elements of $\mathbf{A}_N$ and $\mathbf{C}_N$. From Remark A.1 in Appendix A, it follows that $\left(\sum_{n=1}^{NT}\left|c_{nj,N}\right|^q\right)\leq\overline{K}^q$ and thus $\left(\sum_{n=1}^{NT}\left|c_{nj,N}\right|^q\right)^{1/q}$ $\leq\overline{K}$. Factoring $\overline{K}$ out of the sum yields

$$\left|\phi_N\right|\leq\overline{K}N^{1/p-1/2}\left(N^{1/2}\|\mathbf{\Delta}_N\|\right)\left(N^{-1}\sum_{j=1}^{NT}\left|\varepsilon_{j,N}\right|\right)\left(T(NT)^{-1}\sum_{n=1}^{NT}\left\|\mathbf{d}'_{n.,N}\right\|^p\right)^{1/p}.$$

This holds for $p=2+\delta$ for some $\delta>0$ as in Assumption 4 and $1/p+1/q=1$. By Assumption 4, $N^{1/2}\|\mathbf{\Delta}_N\|=O_p(1)$. Assumption 4 also implies that $\left((NT)^{-1}\sum_{n=1}^{NT}\left\|\mathbf{d}'_{n.,N}\right\|^p\right)^{1/p}=O_p(1)$ for $p=2+\delta$ and some $\delta>0$.

Moreover, $E\left|\varepsilon_{j,N}\right|\leq K<\infty$, which implies that $N^{-1}\sum_{n=1}^{NT}\left|\varepsilon_{n,N}\right|=O_p(1)$. Since $N^{1/p-1/2}\to 0$ as $N\to\infty$ it follows that $\left|\phi_N\right|=o_p(1)$. For later reference, note that $\left|\phi_N\right|=\overline{K}o_p(1)=o_p(1)$, where we can choose $\overline{K}=2c_Ac_P$, where c_A and c_P are the bounds for the row and column sums of the absolute values of the elements of $\mathbf{A}_N$ and $[\mathbf{I}_T\otimes(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]$, respectively. (Compare (B.6) and Remark A.1).

Next consider

$$\left|\psi_N\right|=N^{-1}\left|\mathbf{\Delta}'_N\mathbf{D}'_N\mathbf{A}_N\mathbf{D}_N\mathbf{\Delta}_N\right|=N^{-1}\left|\sum_{n=1}^{NT}\sum_{j=1}^{NT}\mathbf{\Delta}'_N\mathbf{d}'_{n.,N}a_{nj,N}\mathbf{d}_{j.,N}\mathbf{\Delta}_N\right|\tag{B.9}$$

$$\leq N^{-1}\|\mathbf{\Delta}_N\|^2\sum_{n=1}^{NT}\left\|\mathbf{d}'_{n.,N}\right\|\sum_{j=1}^{NT}\left\|\mathbf{d}_{j.,N}\right\|\left|a_{nj,N}\right|$$

$$\leq N^{-1}\|\mathbf{\Delta}_N\|^2\sum_{n=1}^{NT}\left\|\mathbf{d}'_{n.,N}\right\|\left(\sum_{j=1}^{NT}\left\|\mathbf{d}_{j.,N}\right\|^p\right)^{1/p}\left(\sum_{j=1}^{NT}\left|a_{nj,N}\right|^q\right)^{1/q}$$

$$\leq N^{1/p}\overline{K}\|\mathbf{\Delta}_N\|^2\left(N^{-1}\sum_{n=1}^{NT}\left\|\mathbf{d}'_{n.,N}\right\|\right)\left(N^{-1}\sum_{j=1}^{NT}\left\|\mathbf{d}_{j.,N}\right\|^p\right)^{1/p}$$

$$=N^{1/p-1/2}N^{-1/2}\overline{K}\left(N^{1/2}\|\mathbf{\Delta}_N\|\right)^2\left(N^{-1}\sum_{j=1}^{NT}\left\|\mathbf{d}_{j.,N}\right\|^p\right)^{2/p}=o_p(1).$$

From the last inequality we can also see that $N^{1/2}\psi_N = o_p(1)$. Note that $|\psi_N| = \overline{K}o_p(1) = o_p(1)$, where we can choose $\overline{K} = 2c_A c_P$. Summing up, we have proved that $\psi_N = o_p(1)$.

Proof of part (b)

Denote by $\vartheta^*_{s,N}$ the s-th element of $N^{-1}\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N$. By Assumptions 3 and 4 and Remark A.1 in Appendix A there exists a constant $\overline{\overline{K}} < \infty$ such that $E(u^2_{i,N}) \le \overline{\overline{K}}$ and $E|d_{ij,N}|^p \le \overline{\overline{K}}$ with $p = 2 + \delta$ for some $\delta > 0$. Without loss of generality we assume that the row and column sums of the matrices $\mathbf{A}_N$ are bounded uniformly by $\overline{\overline{K}} < \infty$. Notice first that

$$
\begin{aligned}
E|u_{n,N}\|d_{js,N}| &\le \left(Eu^2_{n,N}\right)^{1/2}\left(Ed^2_{js,N}\right)^{1/2} \\
&\le \left(Eu^2_{n,N}\right)^{1/2}\left(E|d_{js,N}|^p\right)^{1/p} \\
&\le \overline{\overline{K}}^{1/2}\overline{\overline{K}}^{1/p} = \overline{\overline{K}}^{1/2+1/p} \text{ with } p \text{ as before.}
\end{aligned}
$$

It follows that

$$
E|\vartheta^*_{s,N}| = N^{-1}\sum_{n=1}^{NT}\sum_{j=1}^{NT}|a_{nj,N}|\,E|u_{n,N}\|d_{js,N}| \tag{B.10}
$$

$$
\le \overline{\overline{K}}^{1/2+1/p}N^{-1}\sum_{n=1}^{NT}\sum_{j=1}^{NT}|a_{nj,N}| \le \overline{\overline{K}}^{1/2+1/p}N^{-1}NT\overline{\overline{K}} = T\overline{\overline{K}}^{3/2+1/p} < \infty,
$$

which shows that $E|N^{-1}\mathbf{d}'_{.s,N}\mathbf{A}_N\mathbf{u}_N| = O(1)$, and also that $\boldsymbol{\alpha}_N = N^{-1}E[\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N] = O(1)$. It is readily verified that $Var(\vartheta^*_s) = o(1)$, such that we have $\vartheta^*_s - E(\vartheta^*_s) = o_p(1)$. Next observe that

$$
N^{-1}\mathbf{D}'_N\mathbf{A}_N\tilde{\mathbf{u}}_N = N^{-1}\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N + \phi^*_N, \tag{B.11}
$$

where $\phi^*_N = N^{-1}\mathbf{D}'_N\mathbf{A}_N\mathbf{D}_N\boldsymbol{\Delta}_N$. By arguments analogous to the proof that $\phi_N = N^{-1}[\boldsymbol{\Delta}'_N\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N] = o_p(1)$, it follows that $\phi^*_N = o_p(1)$. Hence $\tilde{\vartheta}^*_s - \vartheta^*_s = o_p(1)$, and thus $\tilde{\vartheta}^*_s - E(\vartheta^*_s) = o_p(1)$, which also shows that $N^{-1}\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\tilde{\mathbf{u}}_N - \boldsymbol{\alpha}_N = o_p(1)$.

Proof of part (c)

In light of the proof of part (a)

$$
N^{-1/2}\tilde{\mathbf{u}}'_N\mathbf{A}_N\tilde{\mathbf{u}}_N = N^{-1/2}\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N + [N^{-1}\mathbf{u}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{D}_N]N^{1/2}\boldsymbol{\Delta}_N + N^{1/2}\psi_N, \tag{B.12}
$$

where $N^{1/2}\psi_N = o_p(1)$ as shown above, and in light of (b) and since $N^{1/2}\boldsymbol{\Delta}_N = O_p(1)$ by Assumption 4, we have

$$
N^{-1/2}\tilde{\mathbf{u}}'_N\mathbf{A}_N\tilde{\mathbf{u}}_N = N^{-1/2}\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N + N^{1/2}\boldsymbol{\alpha}'_N\boldsymbol{\Delta}_N + o_p(1). \tag{B.13}
$$

Proof of Theorem 1. Consistency of the Weighted GM Estimator

We first show that Assumption 5 also implies that the smallest eigenvalue of $\boldsymbol{\Gamma}'_N\boldsymbol{\Theta}_N\boldsymbol{\Gamma}_N$ is bounded away from zero, i.e., that $\lambda_{\min}(\boldsymbol{\Gamma}'_N\boldsymbol{\Theta}_N\boldsymbol{\Gamma}_N) \geq \lambda_0$ for some $\lambda_0 > 0$. By Assumption 5 and in light of Rao (1973, p. 62),

$$\lambda_{\min}(\boldsymbol{\Gamma}'_N\boldsymbol{\Gamma}_N) = \inf_x \frac{\mathbf{x}'\boldsymbol{\Gamma}'_N\boldsymbol{\Gamma}_N\mathbf{x}}{\mathbf{x}'\mathbf{x}} \geq \lambda_* > 0 . \tag{B.14}$$

Using Mittelhammer (1996, p. 254) we have

$$\lambda_{\min}(\boldsymbol{\Gamma}'_N\boldsymbol{\Theta}_N\boldsymbol{\Gamma}_N) = \inf_x \frac{\mathbf{x}'\boldsymbol{\Gamma}'_N\boldsymbol{\Theta}_N\boldsymbol{\Gamma}_N x}{\mathbf{x}'\mathbf{x}} \geq \lambda_{\min}(\Xi_N^{-1})\inf_x \frac{\mathbf{x}'\boldsymbol{\Gamma}'_N\boldsymbol{\Gamma}_N\mathbf{x}}{\mathbf{x}'\mathbf{x}}$$

$$= \lambda_{\min}(\boldsymbol{\Theta}_N)\lambda_{\min}(\boldsymbol{\Gamma}'_N\boldsymbol{\Gamma}_N) \geq \lambda_0 > 0 , \tag{B.15}$$

with $\lambda_0 = \underline{\lambda}_*\lambda_*$ since $= \lambda_{\min}(\boldsymbol{\Theta}_N) \geq \underline{\lambda}_* > 0$ by Assumption 5.

The objective function of the weighted GM estimator and its nonstochastic counterpart are given by

$$R_N(\underline{\boldsymbol{\theta}}) = (\tilde{\boldsymbol{\gamma}}_N - \tilde{\boldsymbol{\Gamma}}_N\underline{\mathbf{b}})'\boldsymbol{\Theta}_N(\tilde{\boldsymbol{\gamma}}_N - \tilde{\boldsymbol{\Gamma}}_N\underline{\mathbf{b}}) \text{ and} \tag{B.16a}$$

$$\overline{R}_N(\underline{\boldsymbol{\theta}}) = (\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N\underline{\mathbf{b}})'\boldsymbol{\Theta}_N(\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N\underline{\mathbf{b}}) \tag{B.16b}$$

Since $\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N\mathbf{b}_N = 0$, we have $\overline{R}_N(\boldsymbol{\theta}_N) = 0$, i.e., $\overline{R}_N(\underline{\boldsymbol{\theta}}) = 0$ at the true parameter vector $\boldsymbol{\theta}_N = (\rho_{1N},...,\rho_{S,N},\sigma_\mu)'$. Hence,

$$\overline{R}_N(\underline{\boldsymbol{\theta}}) - \overline{R}_N(\boldsymbol{\theta}_N) = (\underline{\mathbf{b}} - \mathbf{b}_N)'\boldsymbol{\Gamma}'_N\boldsymbol{\Theta}_N\boldsymbol{\Gamma}_N(\underline{\mathbf{b}} - \mathbf{b}_N) . \tag{B.17a}$$

In light of Rao (1973, p. 62) and Assumption 5, it follows that:

$$\overline{R}_N(\underline{\boldsymbol{\theta}}) - \overline{R}_N(\boldsymbol{\theta}_N) \geq \lambda_{\min}(\boldsymbol{\Gamma}'_N\boldsymbol{\Theta}_N\boldsymbol{\Gamma}_N)(\underline{\mathbf{b}} - \mathbf{b}_N)'(\underline{\mathbf{b}} - \mathbf{b}_N) \text{ and} \tag{B.17b}$$

$$\overline{R}_N(\underline{\boldsymbol{\theta}}) - \overline{R}_N(\boldsymbol{\theta}_N) \geq \lambda_0(\underline{\mathbf{b}} - \mathbf{b}_N)'(\underline{\mathbf{b}} - \mathbf{b}_N) .$$

By the properties of the norm $\|\mathbf{A}\| = [tr(\mathbf{A}\mathbf{A})]^{1/2}$, we have $\|\underline{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \leq (\underline{\mathbf{b}} - \mathbf{b}_N)'(\underline{\mathbf{b}} - \mathbf{b}_N)$ such that $\overline{R}_N(\underline{\boldsymbol{\theta}}) - \overline{R}_N(\boldsymbol{\theta}_N) \geq \lambda_0\|\underline{\boldsymbol{\theta}} - \boldsymbol{\theta}_N\|^2$. Hence, for every $\varepsilon > 0$

$$\lim_{N \to \infty} \inf_{\{\boldsymbol{\theta}: \|\underline{\boldsymbol{\theta}} - \boldsymbol{\theta}_N\| \geq \varepsilon\}}[\overline{R}_N(\underline{\boldsymbol{\theta}}) - \overline{R}_N(\boldsymbol{\theta}_N)] \geq \inf_{\{\boldsymbol{\theta}: \|\underline{\boldsymbol{\theta}} - \boldsymbol{\theta}_N\| \geq \varepsilon\}} \lambda_0\|\underline{\boldsymbol{\theta}} - \boldsymbol{\theta}_N\|^2 = \lambda_0\varepsilon^2 > 0 , \tag{B.18}$$

which proves that the true parameter vector $\boldsymbol{\theta}_N = (\rho_{1,N},...,\rho_{S,N},\sigma_\mu^2)'$ is identifiable unique (compare Lemma 4.1 in Pötscher and Prucha, 1997).

Moreover, let $\mathfrak{F}_N = (\tilde{\boldsymbol{\gamma}}_N,-\tilde{\boldsymbol{\Gamma}}_N)$ and $\boldsymbol{\Phi}_N = (\boldsymbol{\gamma}_N,-\boldsymbol{\Gamma}_N)$. Then, the difference between the objective function and its nonstochastic counterpart can then be written as

$$R_N(\omega,\underline{\boldsymbol{\theta}}) = (1,\underline{\mathbf{b}}')\mathfrak{F}'_N\tilde{\boldsymbol{\Theta}}_N\mathfrak{F}_N(1,\underline{\mathbf{b}}')' \text{ and} \tag{B.19a}$$

$$\overline{R}_N(\underline{\boldsymbol{\theta}}) = (1, \underline{\mathbf{b}}') \boldsymbol{\Phi}_N' \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N (1, \underline{\mathbf{b}}')' , \tag{B.19b}$$

such that

$$\left| R_N(\omega, \underline{\boldsymbol{\theta}}) - \overline{R}_N(\underline{\boldsymbol{\theta}}) \right| = \left| (1, \underline{\mathbf{b}}') (\mathfrak{F}_N' \widetilde{\boldsymbol{\Theta}}_N \mathfrak{F}_N - \boldsymbol{\Phi}_N' \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N)(1, \underline{\mathbf{b}}')' \right|$$

$$\leq \left\| \mathfrak{F}_N' \widetilde{\boldsymbol{\Theta}}_N \mathfrak{F}_N - \boldsymbol{\Phi}_N' \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N \right\| \left\| (1, \underline{\mathbf{b}}') \right\|^2$$

$$\leq \left\| \mathfrak{F}_N' \widetilde{\boldsymbol{\Theta}}_N \mathfrak{F}_N - \boldsymbol{\Phi}_N' \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N \right\| [1 + S(a^\rho)^2 + \frac{2S + S(S-1)}{2}(a^\rho)^4 + b_\mu^2].$$

As evident from (17), the elements of the matrices $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ are all of the form $\mathbf{u}_N' \mathfrak{A}_N \mathbf{u}_N$, where $\mathfrak{A}_N$ are nonstochastic $NT \times NT$ matrices, whose row and column sums are bounded uniformly in absolute value. In light of Lemma B.1, the elements of $\boldsymbol{\Phi}_N$ are $O(1)$ and it follows that $\left\| \mathfrak{F}_N - \boldsymbol{\Phi}_N \right\| \overset{p}{\to} 0$ and $\left\| \mathfrak{F}_N' \widetilde{\boldsymbol{\Theta}}_N \mathfrak{F}_N - \boldsymbol{\Phi}_N' \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N \right\| \overset{p}{\to} 0$ as $N \to \infty$. As a consequence, we have (for finite S)

$$\sup_{-a^\rho \leq \underline{\rho}_s \leq a^\rho, s=1,\dots S, \underline{\sigma}_\mu^2 \in [0, b_\mu]} \left| R_N(\omega, \underline{\boldsymbol{\theta}}) - \overline{R}_N(\underline{\boldsymbol{\theta}}) \right| \leq \left\| \mathfrak{F}_N' \mathfrak{F}_N - \boldsymbol{\Phi}_N' \boldsymbol{\Phi}_N \right\| [1 + S(a^\rho)^2 + \frac{S(S-1)}{2}(a^\rho)^4 + b^2] \overset{p}{\to} 0 \text{ as } N \to \infty.$$

$$\tag{B.20}$$

Together with identifiable uniqueness, the consistency of $\widetilde{\boldsymbol{\theta}}_N = (\widetilde{\rho}_{1,N}, \dots, \widetilde{\rho}_{S,N}, \widetilde{\sigma}_{\mu,N}^2)$ now follows directly from Lemma 3.1 in Pötscher and Prucha (1997).

Proof of Theorem 2. Asymptotic Normality of $\widetilde{\boldsymbol{\theta}}_N$

To derive the asymptotic distribution of the vector $\mathbf{q}_N$, defined in (30) we invoke the central limit theorem for vectors of linear quadratic forms given by Kelejian and Prucha (2009, Theorem A.1). The vector of quadratic forms in the present context, to which the Theorem is applied is $\overset{*}{\mathbf{q}}_N = N^{1/2} \mathbf{q}_N$; its variance-covariance matrix is given by $\overset{*}{\boldsymbol{\Psi}}_N = N \boldsymbol{\Psi}_N$ and $(\overset{*}{\boldsymbol{\Psi}}_N)^{-1/2} = N^{-1/2} \boldsymbol{\Psi}_N^{-1/2}$.

Note that in light of Assumptions 1, 2 and 7 (and Lemma B.1), the stacked innovations $\boldsymbol{\xi}_N$, the matrices $\mathbf{A}_{c,\mathbf{v},N}^{s,s'}$, $\mathbf{A}_{c,\boldsymbol{\mu},N}^{s,s'}$, and the vectors $\mathbf{a}_{c,\mathbf{v},N}^{s,s'}$ and $\mathbf{a}_{c,\boldsymbol{\mu},N}^{s,s'}$, $c = 1,\dots,4$, $s, s' = 1,\dots,S$, satisfy the assumptions of central limit theorem by Kelejian and Prucha (2009, Theorem A.1).

It follows that

$$-(\overset{*}{\boldsymbol{\Psi}}_N)^{-1/2} \overset{*}{\mathbf{q}}_N = -N^{-1/2} \boldsymbol{\Psi}_N^{-1/2} \overset{*}{\mathbf{q}}_N = -\boldsymbol{\Psi}_N^{-1/2} \mathbf{q}_N \overset{d}{\to} (\mathbf{0}, \mathbf{I}_{4S+2}) , \tag{B.21}$$

since $N^{-1} \lambda_{\min}(\overset{*}{\boldsymbol{\Psi}}_N) = \lambda_{\min}(\boldsymbol{\Psi}_N) > 0$ by assumption as required in Theorem A.1.

Since the row and column sums of the matrices $\overline{\mathbf{A}}_{c,N}^{s,s'}$, the elements of the vectors $\overline{\mathbf{a}}_{c,N}^{s,s'}$, $c = 1,...,4$, and $s,s' = 1,...,S$, and the $(4+\eta)-th$ moments of $\mathbf{v}_N$ and $\boldsymbol{\mu}_N$ are bounded uniformly in absolute value, it follows in light of (28) that the elements of $\boldsymbol{\Psi}_N$ and also those of $\boldsymbol{\Psi}_N^{1/2}$ are bounded uniformly in absolute value.

We next turn to the derivation of the limiting distribution of the GM estimator $\tilde{\boldsymbol{\theta}}_N$. In Theorem 1 we showed that the GM estimator $\tilde{\boldsymbol{\theta}}_N$ defined by (18) is consistent. It follows that – apart from a set of the sample space whose probability tends to zero – the estimator satisfies the following first order condition:

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{q}_N(\tilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)' \tilde{\boldsymbol{\Theta}}_N \mathbf{q}_N(\tilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N) = \frac{\partial \mathbf{q}_N(\tilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}'} \tilde{\boldsymbol{\Theta}}_N \mathbf{q}_N(\tilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N) = \mathbf{0}, \qquad \text{(B.22)}$$

which is a $(S+1) \times 1$ vector, the rows corresponding the partial derivatives of the criterion function with respect to $\rho_{s,N}$, $s = 1,...,S$, and σ_μ^2.

Substituting the mean value theorem expression

$$\mathbf{q}_N(\tilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N) = \mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N) + \frac{\partial \mathbf{q}_N(\overline{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}'} (\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N), \qquad \text{(B.23)}$$

where $\overline{\boldsymbol{\theta}}_N$ is some between value, into the first-order condition yields

$$\frac{\partial \mathbf{q}_N(\tilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}} \tilde{\boldsymbol{\Theta}}_N \frac{\partial \mathbf{q}_N(\overline{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}'} N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N) = -\frac{\partial \mathbf{q}_N(\tilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}} \tilde{\boldsymbol{\Theta}}_N N^{1/2} \mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N). \text{ (B.24)}$$

Observe that $\dfrac{\partial \mathbf{q}_N(\boldsymbol{\theta}, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}'} = \tilde{\boldsymbol{\Gamma}}_N \boldsymbol{\mathfrak{B}}_N$ and consider the two $(S+1) \times (S+1)$ matrices

$$\tilde{\boldsymbol{\Xi}}_N = \frac{\partial \mathbf{q}_N(\tilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}} \tilde{\boldsymbol{\Theta}}_N \frac{\partial \mathbf{q}_N(\overline{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}'} = \tilde{\boldsymbol{\mathfrak{B}}}_N' \tilde{\boldsymbol{\Gamma}}_N' \tilde{\boldsymbol{\Theta}}_N \tilde{\boldsymbol{\Gamma}}_N \overline{\boldsymbol{\mathfrak{B}}}_N, \qquad \text{(B.25)}$$

$$\boldsymbol{\Xi}_N = \boldsymbol{\mathfrak{B}}_N' \boldsymbol{\Gamma}_N' \boldsymbol{\Theta}_N \boldsymbol{\Gamma}_N \boldsymbol{\mathfrak{B}}_N, \qquad \text{(B.26)}$$

where $\tilde{\boldsymbol{\mathfrak{B}}}_N$ and $\overline{\boldsymbol{\mathfrak{B}}}_N$ correspond to $\boldsymbol{\mathfrak{B}}_N$ as defined above with $\tilde{\boldsymbol{\theta}}_N$ and $\overline{\boldsymbol{\theta}}_N$ substituted for $\boldsymbol{\theta}_N$. Notice that $\boldsymbol{\Xi}_N$ is positive definite, since $\boldsymbol{\Gamma}_N$ and $\boldsymbol{\Theta}_N$ are positive definite by assumption and the $[2S + S(S-1)/2+1] \times (S+1)$ matrix $\boldsymbol{\mathfrak{B}}_N$ has full column rank.

In the proof of Theorem 1 (and Lemma B.1) we have demonstrated that $\tilde{\boldsymbol{\Gamma}}_N - \boldsymbol{\Gamma}_N \xrightarrow{p} \mathbf{0}$ and that the elements of $\boldsymbol{\Gamma}_N$ and $\tilde{\boldsymbol{\Gamma}}_N$ are $O(1)$ and $O_p(1)$, respectively. By Assumption 5, $\tilde{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta}_N = o_p(1)$, $\boldsymbol{\Theta}_N = O(1)$ and $\tilde{\boldsymbol{\Theta}}_N = O_p(1)$. Since $\tilde{\boldsymbol{\rho}}_N$ and $\overline{\boldsymbol{\rho}}_N$ (and thus also $\tilde{\boldsymbol{\mathfrak{B}}}_N$ and

$\overline{\mathfrak{B}}_N$) are consistent and bounded uniformly in probability, if follows that $\widetilde{\widetilde{\Xi}}_N - \Xi_N = o_p(1)$, $\widetilde{\widetilde{\Xi}}_N = O_p(1)$, and $\Xi_N = O(1)$. Moreover, Ξ_N is positive definite and thus invertible, and its inverse Ξ_N^{-1} is also $O(1)$.

Denote $\widetilde{\widetilde{\Xi}}_N^+$ as the generalized inverse of $\widetilde{\widetilde{\Xi}}_N$. It then follows as a special case of Lemma F1 in Pötscher and Prucha (1997) that $\widetilde{\widetilde{\Xi}}_N$ is non-singular with probability approaching 1 as $N \to \infty$, that $\widetilde{\widetilde{\Xi}}_N^+$ is $O_p(1)$, and that $\widetilde{\widetilde{\Xi}}_N^+ - \Xi_N^{-1} = o_p(1)$.

Pre-multiplying (B.24) with $\widetilde{\widetilde{\Xi}}_N^+$ we obtain, after rearranging terms,

$$N^{1/2}(\widetilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N) = (\mathbf{I}_{S+1} - \widetilde{\widetilde{\Xi}}_N^+ \widetilde{\widetilde{\Xi}}_N)N^{1/2}(\widetilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N) - N^{1/2}\widetilde{\widetilde{\Xi}}_N^+ \frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}} \widetilde{\boldsymbol{\Theta}}_N \mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N) . \text{(B.27)}$$

In light of the discussion above, the first term on the right-hand side is zero on ω-sets of probability approaching 1 (compare Pötscher and Prucha, 1997, pp. 228). This yields

$$N^{1/2}(\widetilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N) = -\widetilde{\widetilde{\Xi}}_N^+ \frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}} \widetilde{\boldsymbol{\Theta}}_N N^{1/2}\mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N) + o_p(1) . \tag{B.28}$$

Next observe that

$$\widetilde{\widetilde{\Xi}}_N^+ \frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}} \widetilde{\boldsymbol{\Theta}}_N - \Xi_N^{-1}\mathfrak{B}_N' \boldsymbol{\Gamma}_N' \boldsymbol{\Theta}_N = o_p(1) , \tag{B.29}$$

since $\widetilde{\widetilde{\Xi}}_N^+ - \Xi_N^{-1} = o_p(1)$ and $\frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\theta}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\theta}} - \mathfrak{B}_N' \boldsymbol{\Gamma}_N' = o_p(1)$.

As we showed in section III, the elements of $N^{1/2}\mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N)$ can be expressed as

$$N^{1/2}\mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N) = N^{-1/2}\mathbf{q}_N^* + o_p(1) = \mathbf{q}_N + o_p(1) . \tag{B.30}$$

where $\mathbf{q}_N^*$ is defined in (24), and that

$$-(\boldsymbol{\Psi}_N^*)^{-1/2}\mathbf{q}_N^* = -N^{-1/2}\boldsymbol{\Psi}_N^{-1/2}\mathbf{q}_N^* = -\boldsymbol{\Psi}_N^{-1/2}\mathbf{q}_N \xrightarrow{d} (\mathbf{0}, \mathbf{I}_{4S+2}) . \tag{B.31}$$

It now follows from (B.28), (B.29), and (B.30) that

$$N^{1/2}(\widetilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N) = \Xi_N^{-1}\mathbf{J}_N' \boldsymbol{\Theta}_N \boldsymbol{\Psi}_N^{1/2}(-\boldsymbol{\Psi}_N^{-1/2}\mathbf{q}_N) + o_p(1) . \tag{B.32}$$

Since all nonstochastic terms on the right hand side from (B.32) are $O(1)$ it follows that $N^{1/2}(\widetilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$ is $O_p(1)$. To derive the asymptotic distribution of $N^{1/2}(\widetilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$, we invoke Corollary F4 in Pötscher and Prucha (1997). In the present context, we have

$$\zeta_N = -\Psi_N^{-1/2}\mathbf{q}_N \xrightarrow{d} \zeta \sim N(\mathbf{0}, \mathbf{I}_{4S+S(S-1)}),$$

$$N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N) = \mathfrak{X}_N\zeta_N + o_p(1), \text{ with}$$

$$\mathfrak{X}_N = \Xi_N^{-1}\mathbf{J}_N'\Theta_N\Psi_N^{1/2}.$$

Furthermore, $N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N) = O_p(1)$ and its variance-covariance matrix is

$$\Omega_{\tilde{\boldsymbol{\theta}}_N}(\Theta_N) = (\mathbf{J}_N'\Theta_N\mathbf{J}_N)^{-1}\mathbf{J}_N'\Theta_N\Psi_N\Theta_N'\mathbf{J}_N(\mathbf{J}_N'\Theta_N\mathbf{J}_N)^{-1},$$

where $\Omega_{\tilde{\boldsymbol{\theta}}_N}$ is positive definite.

As a final point it has to be shown that $\liminf_{N\to\infty}\lambda_{\min}(\mathfrak{X}_N\mathfrak{X}_N') > 0$ as required in Corollary F4 in Pötscher and Prucha (1997). Observe that

$$\lambda_{\min}(\mathfrak{X}_N\mathfrak{X}_N') = \lambda_{\min}(\Xi_N^{-1}\mathbf{J}_N'\Theta_N\Psi_N\Theta_N'\mathbf{J}_N\Xi_N^{-1}) \tag{B.33}$$

$$\geq \lambda_{\min}(\Psi_N)\lambda_{\min}(\Theta_N\Theta_N')\lambda_{\min}(\Xi_N^{-1}\Xi_N^{-1})\lambda_{\min}(\Gamma_N'\Gamma_N)\lambda_{\min}(\mathfrak{B}_N'\mathfrak{B}_N) > 0,$$

since the matrices involved are all positive definite.

Proof of Theorem 3. Joint Distribution of $\tilde{\boldsymbol{\rho}}_N$ and Other Model Parameters

The first line in Theorem 3 holds in light of Assumption 7 (for $N^{1/2}\Delta_N$), bearing in mind that $\mathbf{T}_N = \mathbf{F}_N\mathbf{P}_N$, and Theorem 2 (for $N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$).

We next prove that $\boldsymbol{\xi}_{o,N} = \Psi_{\mathbf{w},N}^{-1/2}[(NT)^{-1/2}\boldsymbol{\xi}_N'\mathbf{F}_N, \mathbf{q}_N']' \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{P^*+4S+S(S-1)})$ by verifying that the assumptions of the central limit theorem A.1 by Kelejian and Prucha (2009) are fulfilled. Note that $\lambda_{\min}(\Psi_{\mathbf{w},N}) \geq c_{\Psi_{\mathbf{w}}}^* > 0$ by assumption. In Theorem 2, we verified that the stacked innovations $\boldsymbol{\xi}_N$, the matrices $\mathbf{A}_{c,\mathbf{v},N}^{s,s'}$, $\mathbf{A}_{c,\boldsymbol{\mu},N}^{s,s'}$, and the vectors $\mathbf{a}_{c,\mathbf{v},N}^{s,s'}$ and $\mathbf{a}_{c,\boldsymbol{\mu},N}^{s,s'}$, $c=1,...,4$, $s,s'=1,...,S$, satisfy the assumptions of central limit theorem by Kelejian and Prucha (2009, Theorem A.1).

For the estimators considered in the present paper, the elements of the matrix $\mathbf{F}_N = (\mathbf{F}_{\mathbf{v},N}', \mathbf{F}_{\boldsymbol{\mu},N}')'$ are bounded uniformly in absolute value, provided that the elements of the matrix $\mathbf{H}_N$ are bounded uniformly in absolute value (see Lemmata 1 and 2). Hence, the linear form $\mathbf{F}_N'\boldsymbol{\xi}_N = \mathbf{F}_{\mathbf{v},N}'\mathbf{v}_N + \mathbf{F}_{\boldsymbol{\mu},N}'\boldsymbol{\mu}_N$ fulfils the assumptions of Theorem A.1; as a consequence,

$$\boldsymbol{\xi}_{o,N} \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{P^*+4S+S(S-1)}).$$

Proof of Lemma 1.

Consider the case of random effects estimation first. In light of equations (4a) and (4b), Assumptions 3 and 8, as well as $\sup_N \|\boldsymbol{\beta}_N\| \leq b < \infty$, it follows that all columns of $\mathbf{Z}_N = (\mathbf{X}_N, \overline{\mathbf{Y}}_N)$ are of the form $\vartheta_N = \boldsymbol{\pi}_N + \boldsymbol{\Pi}_N \boldsymbol{\varepsilon}_N$, where the elements of the vector $\boldsymbol{\pi}_N$ and the row and column sums of the matrix $\boldsymbol{\Pi}_N$ are bounded uniformly in absolute value. It follows that the fourth moments of the elements of the matrix $\mathbf{D}_N = -\mathbf{Z}_N$ are bounded uniformly by some finite constant and that Assumption 6 holds (see Remark A.1 in Appendix A).

Next, note that

$$(NT)^{1/2}(\tilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N) = \tilde{\mathbf{P}}_N'(NT)^{-1/2}\mathbf{F}_{\mathbf{v},N}'\mathbf{v}_N + \tilde{\mathbf{P}}_N'(NT)^{-1/2}\mathbf{F}_{\boldsymbol{\mu},N}'\boldsymbol{\mu}_N,$$

where $\tilde{\mathbf{P}}_N$ is defined in the Lemma, and

$$\mathbf{F}_{\mathbf{v},N} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N}')^{-1}]\mathbf{H}_N, \text{ and}$$

$$\mathbf{F}_{\boldsymbol{\mu},N} = (\mathbf{e}_T' \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N}')^{-1}]\mathbf{H}_N.$$

In light of Assumption 8, $\tilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$ and $\mathbf{P}_N = O(1)$, with $\mathbf{P}_N$ as defined in the Lemma. By Assumptions 2, 3 and 8, the elements of $\mathbf{F}_{\mathbf{v},N}$ and $\mathbf{F}_{\boldsymbol{\mu},N}$ are bounded uniformly in absolute value. By Assumption 1, $E(\mathbf{v}_N) = \mathbf{0}$, $E(\boldsymbol{\mu}_N) = \mathbf{0}$, and the diagonal variance-covariance matrices of $\mathbf{v}_N$ and $\boldsymbol{\mu}_N$ have uniformly bounded elements. Thus, $E[(NT)^{-1/2}\mathbf{F}_{\mathbf{v},N}'\mathbf{v}_N] = \mathbf{0}$ and the elements of the variance-covariance matrix of $N^{-1/2}\mathbf{F}_{\mathbf{v},N}'\mathbf{v}_N$, i.e., $(NT)^{-1}\mathbf{F}_{\mathbf{v},N}'\boldsymbol{\Sigma}_N\mathbf{F}_{\mathbf{v},N}$, are bounded uniformly in absolute value. Moreover, $E[(NT)^{-1/2}\mathbf{F}_{\boldsymbol{\mu},N}'\boldsymbol{\mu}_N] = \mathbf{0}$, and the elements of the variance-covariance matrix of $N^{-1/2}\mathbf{F}_{\boldsymbol{\mu},N}'\boldsymbol{\mu}_N$, i.e., $(NT)^{-1}\sigma_{\mu}^2\mathbf{F}_{\boldsymbol{\mu},N}'\mathbf{F}_{\boldsymbol{\mu},N}$, are bounded uniformly in absolute value (see Remark A.1 in Appendix A). It follows from Chebychev's inequality that $(NT)^{-1/2}\mathbf{F}_{\mathbf{v},N}'\mathbf{v}_N = O_p(1)$, $(NT)^{-1/2}\mathbf{F}_{\boldsymbol{\mu},N}'\boldsymbol{\mu}_N = O_p(1)$, and consequently $(NT)^{1/2}(\tilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N) = \mathbf{P}_N(NT)^{-1/2}\mathbf{F}_{\mathbf{v},N}'\mathbf{v}_N + \mathbf{P}_N(NT)^{-1/2}\mathbf{F}_{\boldsymbol{\mu},N}'\boldsymbol{\mu}_N + o_p(1)$ and that $\mathbf{P}_N(NT)^{-1/2}\mathbf{F}_{\mathbf{v},N}'\mathbf{v}_N + \mathbf{P}_N(NT)^{-1/2}\mathbf{F}_{\boldsymbol{\mu},N}'\boldsymbol{\mu}_N = O_p(1)$. This completes the proof, recalling that $\mathbf{T}_N = (\mathbf{T}_{\mathbf{v},N}', \mathbf{T}_{\boldsymbol{\mu},N}')' = (\mathbf{P}_N'\mathbf{F}_{\mathbf{v},N}', \mathbf{P}_N'\mathbf{F}_{\boldsymbol{\mu},N}')'$. Obviously, the same proof applies under fixed effects estimation, using the within-transformed matrices $\underline{\mathbf{T}}_{\mathbf{v},N}$, $\underline{\mathbf{T}}_{\boldsymbol{\mu},N}$, $\underline{\mathbf{P}}_N$, $\underline{\mathbf{H}}_N$, $\underline{\mathbf{F}}_{\mathbf{v},N}$, and $\underline{\mathbf{F}}_{\boldsymbol{\mu},N} = \mathbf{0}$, provided that Assumption 8 is maintained accordingly for $\underline{\mathbf{X}}_N$ and $\underline{\mathbf{H}}_N$.

Proof of Lemma 2.

The random effects spatial generalized TSLS estimator is given by

$$\hat{\breve{\boldsymbol{\delta}}}_N = (\hat{\breve{\mathbf{Z}}}_N^{*\,\prime}\breve{\mathbf{Z}}_N^*)^{-1}\hat{\breve{\mathbf{Z}}}_N^{*\,\prime}\breve{\mathbf{y}}_N^*, \text{ where } \breve{\mathbf{y}}_N^* = \breve{\mathbf{Z}}_N^*\boldsymbol{\delta}_N + \breve{\mathbf{u}}_N^* \text{ with}$$

$$\breve{\mathbf{u}}_N^* = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\breve{\rho}_{m,N}\mathbf{M}_{m,N})]\mathbf{u}_N .$$

Substituting $\hat{\breve{\mathbf{Z}}}_N^* = \mathbf{P}_{\breve{\mathbf{H}}_N^*}\breve{\mathbf{Z}}_N^* = \breve{\mathbf{H}}_N^*(\breve{\mathbf{H}}_N^{*\,\prime}\breve{\mathbf{H}}_N^{**})^{-1}\breve{\mathbf{H}}_N^{*\,\prime}\breve{\mathbf{Z}}_N^*$, we obtain

$$(NT)^{1/2}(\hat{\breve{\boldsymbol{\delta}}}_N - \boldsymbol{\delta}_N) = (NT)^{1/2}\boldsymbol{\Delta}_N^* = (NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\,\prime}\breve{\mathbf{H}}_N^{*\,\prime}\breve{\mathbf{u}}_N^*, \text{ with}$$

$$\widetilde{\mathbf{P}}_N^{*\,\prime} = \{[(NT)^{-1}\breve{\mathbf{Z}}_N^{*\,\prime}\breve{\mathbf{H}}_N^*][(NT)^{-1}\breve{\mathbf{H}}_N^{*\,\prime}\breve{\mathbf{H}}_N^*]^{-1}[(NT)^{-1}\breve{\mathbf{H}}_N^{*\,\prime}\breve{\mathbf{Z}}_N^*]\}^{-1}[(NT)^{-1}\breve{\mathbf{Z}}_N^{*\,\prime}\breve{\mathbf{H}}_N^*][(NT)^{-1}\breve{\mathbf{H}}_N^{*\,\prime}\breve{\mathbf{H}}_N^*]^{-1}.$$

Next note that

$$\breve{\mathbf{u}}_N^* = \mathbf{u}_N^* - \sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N)\mathbf{u}_N ,$$

where $\overline{\mathbf{M}}_N$ is a matrix, whose row and columns sums are bounded uniformly in absolute value, satisfying

$$\sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})\mathbf{M}_{m,N} = \sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})\overline{\mathbf{M}}_N .$$

Substituting for $\breve{\mathbf{u}}_N^*$, we obtain

$$(NT)^{1/2}(\hat{\breve{\boldsymbol{\delta}}}_N - \boldsymbol{\delta}_N) = (NT)^{1/2}\boldsymbol{\Delta}_N^* = \eth_{1,N} - \eth_{2,N}, \text{ where}$$

$$\eth_{1,N} = (NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\,\prime}\breve{\mathbf{H}}_N^{*\,\prime}\boldsymbol{\varepsilon}_N ,$$

$$\eth_{2,N} = (NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\,\prime}\breve{\mathbf{H}}_N^{*\,\prime}\sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N)\mathbf{u}_N$$

$$= (NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\,\prime}\breve{\mathbf{H}}_N^{*\,\prime}\sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]\boldsymbol{\varepsilon}_N ,$$

Note that the feasible generalized TSLS estimator uses generated (transformed) instruments $\breve{\mathbf{H}}_N^*$, based on the estimate $\breve{\boldsymbol{\theta}}_N$. Using

$$\breve{\mathbf{H}}_N^* = \mathbf{H}_N^*(\boldsymbol{\theta}_N) - \sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N)\mathbf{H}_N$$

we obtain $\eth_{i,N} = \sum_{j=1}^{2}\eth_{ij,N}$ such that $(NT)^{1/2}\boldsymbol{\Delta}_N^* = \sum_{i=1}^{2}\sum_{j=1}^{2}\eth_{ij,N}$.

Considering $\mathfrak{d}_{1,N}$, we have

$$\mathfrak{d}_{11,N} = (NT)^{-1/2}\breve{\mathbf{P}}_N^{*\prime}\breve{\mathbf{H}}_N^{*\prime}\boldsymbol{\varepsilon}_N$$

$$= (NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\prime}\breve{\mathbf{H}}_N^{*\prime}[\mathbf{v}_N + (\mathbf{e}_T \otimes \mathbf{I}_N)\boldsymbol{\mu}_N]$$

$$= (NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\prime}\mathbf{F}_{\mathbf{v},N}^{*\prime}\mathbf{v}_N + (NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\prime}\mathbf{F}_{\boldsymbol{\mu},N}^{*\prime}\boldsymbol{\mu}_N,$$

with $\mathbf{F}_{\mathbf{v},N}^{*} = \mathbf{H}_N^{*}$, and $\mathbf{F}_{\boldsymbol{\mu},N}^{*} = (\mathbf{e}_T' \otimes \mathbf{I}_N)\mathbf{H}_N^{*}$.

$$\mathfrak{d}_{12,N} = -(NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\prime}\sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})\mathbf{H}_N'(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N')\boldsymbol{\varepsilon}_N.$$

Regarding $\mathfrak{d}_{2,N}$ we have

$$\mathfrak{d}_{21,N} = (NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\prime}\mathbf{H}_N^{*\prime}\sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]\boldsymbol{\varepsilon}_N$$

$$\mathfrak{d}_{22,N} = -(NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\prime}\sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})\mathbf{H}_N'(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N')$$

$$\times \sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]\boldsymbol{\varepsilon}_N$$

$$= -(NT)^{-1/2}\widetilde{\mathbf{P}}_N^{*\prime}[\sum_{m=1}^{S}(\breve{\rho}_{m,N} - \rho_{m,N})]^2\mathbf{H}_N'(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N'\overline{\mathbf{M}}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]\boldsymbol{\varepsilon}_N,$$

Next note that, in light of Assumption 8 and since $\breve{\boldsymbol{\theta}}_N$ is $N^{1/2}$-consistent, it follows that

$$(NT)^{-1}\hat{\breve{\mathbf{Z}}}_N^{*\prime}\breve{\mathbf{Z}}_N^{*} - \mathbf{Q}_{\mathbf{H^*Z^*}}'\mathbf{Q}_{\mathbf{H^*H^*}}^{-1}\mathbf{Q}_{\mathbf{H^*Z^*}} = o_p(1).$$

By Assumption 8 we also have $\mathbf{Q}_{\mathbf{H^*Z^*}}'\mathbf{Q}_{\mathbf{H^*H^*}}^{-1}\mathbf{Q}_{\mathbf{H^*Z^*}} = O(1)$ and thus $(\mathbf{Q}_{\mathbf{H^*Z^*}}'\mathbf{Q}_{\mathbf{H^*H^*}}^{-1}\mathbf{Q}_{\mathbf{H^*Z^*}})^{-1} = O(1)$. It follows as a special case of Pötscher and Prucha (1997, Lemma F1) that

$$[(NT)^{-1}\hat{\breve{\mathbf{Z}}}_N^{*\prime}\breve{\mathbf{Z}}_N^{*}]^{-1} - (\mathbf{Q}_{\mathbf{H^*Z^*}}'\mathbf{Q}_{\mathbf{H^*H^*}}^{-1}\mathbf{Q}_{\mathbf{H^*Z^*}})^{-1} = o_p(1).$$

It follows further that $\widetilde{\mathbf{P}}_N^{*} - \mathbf{P}_N^{*} = o_p(1)$ and $\mathbf{P}_N^{*} = O(1)$ with $\mathbf{P}_N^{*}$ defined in the Lemma.

Next observe that $(\breve{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = o_p(1)$. Note further that all terms $\mathfrak{d}_{ij,N}$ except for $\mathfrak{d}_{11,N}$ are of the form $o_p(1)\widetilde{\mathbf{P}}_N^{*\prime}(NT)^{-1/2}\mathfrak{D}_N'\boldsymbol{\varepsilon}_N$, where $\mathfrak{D}_N$ are $NT \times P_*$ matrices involving products of $(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N)$, $[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]$, and $\mathbf{H}_N$. By the maintained assumptions regarding these matrices it follows that the elements of $\mathfrak{D}_N$ are bounded uniformly in absolute value.

As a consequence, $E[(NT)^{-1/2}\mathfrak{D}'_N\boldsymbol{\varepsilon}_N]=\mathbf{0}$ and the elements of the variance-covariance matrix of $(NT)^{-1/2}\mathfrak{D}'_N\boldsymbol{\varepsilon}_N$, i.e., $(NT)^{-1}\mathfrak{D}'_N\boldsymbol{\Omega}_{\varepsilon,N}\mathfrak{D}_N$, are bounded uniformly in absolute value (see Remark A.1 in Appendix A). It follows from Chebychev's inequality that $(NT)^{-1/2}\mathfrak{D}'_N\boldsymbol{\varepsilon}_N=O_p(1)$. As a consequence, all terms $\mathfrak{d}_{ij,N}$ except for $\mathfrak{d}_{11,N}$ are $o_p(1)$, and $\mathfrak{d}_{11,N}=O_p(1)$. Finally, observe that $\mathfrak{d}_{11,N}=(NT)^{-1/2}\tilde{\mathbf{P}}_N^{*\prime}\mathbf{F}_{v,N}^{*\ \prime}\mathbf{v}_N+(NT)^{-1/2}\tilde{\mathbf{P}}_N^{*\prime}\mathbf{F}_{\mu,N}^{*\ \prime}\boldsymbol{\mu}_N$, with $\mathbf{F}_{v,N}^{*}=\mathbf{H}_N^{*}$ and $\mathbf{F}_{\mu,N}^{*\ \prime}=\mathbf{H}_N^{*\prime}(\mathbf{e}_T\otimes\mathbf{I}_N)$, recalling that $\mathbf{T}_N^{*}=\mathbf{F}_N^{*}\mathbf{P}_N^{*}$.

APPENDIX C

Lemma C.1

Define the $NT\times 1$ vectors $\tilde{\boldsymbol{\varepsilon}}_N=[\mathbf{I}_T\otimes(\mathbf{I}_N-\sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})]\tilde{\mathbf{u}}_N$ with elements $\tilde{\varepsilon}_{it,N}$ and the vector of fixed effects residuals $\tilde{\underline{\mathbf{v}}}_N=\mathbf{Q}_{0,N}\tilde{\boldsymbol{\varepsilon}}_N$. Suppose that Assumptions 1-4 hold and that the elements of $\mathbf{D}_N=(\mathbf{d}'_{1.,N},...,\mathbf{d}'_{N.,N})'$ have bounded fourth moments. Then $\tilde{\boldsymbol{\varepsilon}}_N=\boldsymbol{\varepsilon}_N+\boldsymbol{\eta}_N$ and $\tilde{\underline{\mathbf{v}}}_N=\underline{\mathbf{v}}_N+\underline{\boldsymbol{\eta}}_N$, with $\left|\eta_{it,N}\right|\le\alpha_N\beta_{it,N}$, $\left|\underline{\eta}_{it,N}\right|\le\alpha_N\underline{\beta}_{it,N}$, and where $N^{1/2}\alpha_N=O_p(1)$, $E\beta_{it,N}^4\le K$, and $E\underline{\beta}_{it,N}^4\le K$ for some $K<\infty$. As a direct consequence,
$$\left|\tilde{\varepsilon}_{it,N}^2-\varepsilon_{it,N}^2\right|\le 2\alpha_N\beta_{it,N}\left|\varepsilon_{it,N}\right|+\alpha_N^2\beta_{it,N}^2 \text{ and } \left|\tilde{\underline{v}}_{it,N}^2-\underline{v}_{it,N}^2\right|\le 2\alpha_N\underline{\beta}_{it,N}\left|\underline{v}_{it,N}\right|+\alpha_N^2\underline{\beta}_{it,N}^2.$$

Proof.
Note first that

$$\tilde{\boldsymbol{\varepsilon}}_N=\boldsymbol{\varepsilon}_N+\boldsymbol{\eta}_N \text{, where} \tag{C.1}$$

$$\boldsymbol{\eta}_N=[\sum_{m=1}^{S}(\rho_{m,N}-\tilde{\rho}_{m,N})(\mathbf{I}_T\otimes\mathbf{M}_{m,N})][\mathbf{I}_T\otimes(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]\boldsymbol{\varepsilon}_N$$

$$+[\mathbf{I}_T\otimes(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})]\mathbf{D}_N\boldsymbol{\Delta}_N+[\sum_{m=1}^{S}(\rho_{m,N}-\tilde{\rho}_{m,N})(\mathbf{I}_T\otimes\mathbf{M}_{m,N})]\mathbf{D}_N\boldsymbol{\Delta}_N.$$

This can also be written as

$$\boldsymbol{\eta}_N=\mathfrak{R}_N\mathbf{g}_N, \tag{C.2}$$

where $\mathfrak{R}_N=(\mathfrak{R}_{1,N},\mathfrak{R}_{2,N},\mathfrak{R}_{3,N})$ with

$$\mathfrak{R}_{1,N}=\mathbf{I}_T\otimes(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})\mathbf{D}_N,$$

$$\mathfrak{R}_{2,N}=\{(\mathbf{I}_T\otimes\mathbf{M}_{1,N})[\mathbf{I}_T\otimes(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]\boldsymbol{\varepsilon}_N,...,(\mathbf{I}_T\otimes\mathbf{M}_{S,N})[\mathbf{I}_T\otimes(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]\boldsymbol{\varepsilon}_N\},$$

$$\mathfrak{R}_{3,N}=[(\mathbf{I}_T\otimes\mathbf{M}_{1,N})\mathbf{D}_N,...,(\mathbf{I}_T\otimes\mathbf{M}_{S,N})\mathbf{D}_N], \text{ and}$$

$$\mathfrak{g}_N = [\Delta'_N, (\boldsymbol{\rho}_N - \tilde{\boldsymbol{\rho}}_N)', (\boldsymbol{\rho}_N - \tilde{\boldsymbol{\rho}}_N)' \otimes \Delta'_N]'.$$

In light of Assumption 3 and since the elements of $\mathbf{D}_N = (\mathbf{d}'_{1.,N}, ..., \mathbf{d}'_{N.,N})'$ have bounded fourth moments, each column of the matrix $\mathfrak{R}_N$ is of the form $\boldsymbol{\pi}_N + \boldsymbol{\Pi}_N \boldsymbol{\zeta}_N$, where the elements of the $NT \times 1$ vector $\boldsymbol{\pi}_N$ are bounded uniformly in absolute value by some finite constant, the row and column sums of the $NT \times NT$ matrix $\boldsymbol{\Pi}_N$ are bounded uniformly in absolute value by some finite constant, and the fourth moments of the elements of $\boldsymbol{\zeta}_N$ are also bounded by some finite constant. It follows that the fourth moments of the elements of $\mathfrak{R}_N$ are also bounded by some finite constant (see Remark A.1 in Appendix A).

As a consequence, $\|\boldsymbol{\eta}_N\| \leq \|\mathfrak{R}_N\| \|\mathfrak{g}_N\|$, or for the n-th element of the $NT \times 1$ vector $\boldsymbol{\eta}_N$,

$$|\eta_{n,N}| \leq \|\mathfrak{g}_N\| \|\mathfrak{r}_{n,N}\| = \alpha_N \beta_{n,N}, \tag{C.3}$$

where $\alpha_N = \|\mathfrak{g}_N\|$, $\mathfrak{r}_{n.,N}$ denotes the n-th row of $\mathfrak{R}_N$, and $\beta_{n,N} = \|\mathfrak{r}_{n.,N}\|$ with $E|\beta_{n,N}|^4 \leq K_\beta < \infty$. Without loss of generality we can select K_β such that $E(\beta_{n,N}^\gamma) \leq K_\beta$ for $\gamma \leq 4$. By Assumption 1 there is also some K_ε such that $E|\varepsilon_{n,N}|^\gamma \leq K_\varepsilon < \infty$ for $\gamma \leq 4$. In the following we use K to denote the larger bound, i.e., $K = \max(K_\varepsilon, K_\beta)$. Also note that $N^{1/2}\alpha_N = O_p(1)$. Replacing index n with index it, we have, from (C.1) and (C.3), that

$$\begin{aligned}
|\tilde{\varepsilon}_{it,N}^2 - \varepsilon_{it,N}^2| &= |(\varepsilon_{it,N} + \eta_{it,N})^2 - \varepsilon_{it,N}^2| = |(\varepsilon_{it,N}^2 + 2\varepsilon_{it,N}\eta_{it,N} + \eta_{it,N}^2) - \varepsilon_{it,N}^2| \\
&= |2\varepsilon_{it,N}\eta_{it,N} + \eta_{it,N}^2| \\
&\leq |2\varepsilon_{it,N}\eta_{it,N}| + |\eta_{it,N}^2| \\
&\leq 2\alpha_N \beta_{i,N}|\varepsilon_{it,N}| + \alpha_N^2 \beta_{it,N}^2 .
\end{aligned}$$

By the same reasoning we have

$$\tilde{\underline{\mathbf{v}}}_N = \mathbf{Q}_{0,N}\tilde{\boldsymbol{\varepsilon}}_N = \mathbf{Q}_{0,N}\boldsymbol{\varepsilon}_N + \mathbf{Q}_{0,N}\boldsymbol{\eta}_N \tag{C.4}$$

$$= \mathbf{Q}_{0,N}\mathbf{v}_N + \underline{\boldsymbol{\eta}}_N$$

$$= \underline{\mathbf{v}}_N + \underline{\boldsymbol{\eta}}_N ,$$

with $\underline{\boldsymbol{\eta}}_N = \underline{\mathfrak{R}}_N \mathfrak{g}_N$, where $\underline{\mathfrak{R}}_N = \mathbf{Q}_{0,N}\mathfrak{R}_N$. Obviously, the elements of the columns of $\underline{\mathfrak{R}}_N$ and their fourth moments remain bounded uniformly after pre-multiplication with $\mathbf{Q}_{0,N}$, such that we have $|\underline{\eta}_{it,N}| \leq \alpha_N \underline{\beta}_{it,N}$ with α_N defined as above and $E|\underline{\beta}_{it,N}|^4 \leq K_{\underline{\beta}} < \infty$. Finally, we

have $\left| \tilde{\underline{v}}_{it,N}^2 - \underline{v}_{it,N}^2 \right| \le 2\alpha_N \beta_{it,N} \left| \underline{v}_{it,N} \right| + \alpha_N^2 \underline{\beta}_{it,N}^2$. Without loss of generality, we choose the bound K in the lemma such that $K \ge K_{\underline{\beta}}$ and $K \ge K_{\beta}$.

Proof of Theorem 4a. Consistency of $\tilde{\boldsymbol{\Psi}}_{\Delta\Delta,N}$

In the following we provide two Lemmata that establish the consistency of $\tilde{\boldsymbol{\Psi}}_{\Delta\Delta,N}^{v}$.[27] As evident from the proof, this also covers the simpler case of $\tilde{\boldsymbol{\Psi}}_{\Delta\Delta,N}^{\mu}$.

Lemma C.2

Suppose Assumptions 1-4 hold and let

$$\underline{\mathbf{v}}_N = \mathbf{Q}_{0,N}\boldsymbol{\varepsilon}_N = \mathbf{Q}_{0,N}\mathbf{v}_N = \mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})]\mathbf{u}_N, \text{ and}$$

$$\underline{\tilde{\mathbf{v}}}_N = \mathbf{Q}_{0,N}\tilde{\boldsymbol{\varepsilon}}_N = \mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})]\tilde{\mathbf{u}}_N,$$

with $\tilde{\mathbf{u}}_N = \mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N$ and $\mathbf{D}_N = (\mathbf{d}_{1.,N}', ..., \mathbf{d}_{NT.,N}')'$, and where the $S \times 1$ vector $\tilde{\boldsymbol{\rho}}_N$ can be any estimator that satisfies $(\tilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = o_p(1)$. Let $\mathbf{a}_N$ and $\mathbf{b}_N$ be $NT \times 1$ vectors, whose elements are bounded uniformly in absolute value by some constant c, and let $\boldsymbol{\Sigma}_N = diag_{n=1}^{NT}(Ev_{n,N}^2) = diag_{n=1}^{NT}(\sigma_{v,n,N}^2)$. Define $\tilde{\boldsymbol{\Sigma}}_N^{HR} = diag_{n=1}^{NT}[(\tilde{v}_{n,N}^{HR})^2] = diag_{it=1}^{NT}[(\tilde{v}_{it,N}^{HR})^2]$ with

$$(\tilde{v}_{it,N}^{HR})^2 = \frac{T}{T-2}\tilde{\underline{v}}_{it,N}^2 - \frac{1}{(T-1)(T-2)}\sum_{r=1}^{T}\tilde{\underline{v}}_{ir,N}^2. \text{ Then}$$

(a) $\frac{1}{NT}\mathbf{a}_N'\tilde{\boldsymbol{\Sigma}}_N^{HR}\mathbf{b}_N - \frac{1}{NT}\mathbf{a}_N'\boldsymbol{\Sigma}_N\mathbf{b}_N = o_p(1)$ and $\frac{1}{NT}\mathbf{a}_N'\boldsymbol{\Sigma}_N\mathbf{b}_N = O(1)$.

(b) There exist random variables ς_N that do not depend on $\mathbf{a}_N$ and $\mathbf{b}_N$ such that

$$\left| \frac{1}{NT}\mathbf{a}_N'\tilde{\boldsymbol{\Sigma}}_N^{HR}\mathbf{b}_N - \frac{1}{NT}\mathbf{a}_N'\boldsymbol{\Sigma}_N\mathbf{b}_N \right| \le K(c)(1+\varsigma_N), \text{ with } \varsigma_N = o_p(1) \text{ and where } K(c) < \infty \text{ is a}$$

constant that depends monotonically on c (as well as on some other bounds maintained in the assumptions).

Proof.

A complication in the estimation of $(NT)^{-1}\mathbf{a}_N'\boldsymbol{\Sigma}_N\mathbf{b}_N$ arises from the fact that $\boldsymbol{\Sigma}_N = diag_{n=1}^{NT}(Ev_{n,N}^2)$ is based on the idiosyncratic error components in levels ($v_{it,N}$), whereas the estimator has to be based on the (demeaned) fixed effects residuals $\underline{v}_{it,N}$. The problem at hand is similar in its structure to that in Stock and Watson (2008), who consider the estimation of a heteroskedasticity-robust variance-covariance matrix in fixed effects panel data models (without spatial correlation). They suggest an asymptotic bias correction that is

---

[27] Related results for the cross-sectional case are obtained by Kelejian and Prucha (2009).

based on an expression, where the error components $v_{it,N}$ are clustered over cross-section units (averaged over time), and which can be estimated consistently with the fixed effects residuals ($\tilde{\underline{v}}^2_{it,N}$). In the following, we adopt the approach by Stock and Watson (2008) to derive bias-corrected estimators in the present framework.

Define

$$\tau_N = \frac{1}{NT} E(\mathbf{a}'_N \Sigma_N \mathbf{b}_N), \tag{C.5a}$$

$$\underline{\tau}_N = \frac{1}{N(T-1)} \mathbf{a}'_N \underline{\Sigma}_N \mathbf{b}_N \text{ with } \underline{\Sigma}_N = diag^{NT}_{n=1}(\underline{v}^2_{n,N}), \text{ and} \tag{C.5b}$$

$$\tilde{\underline{\tau}}_N = \frac{1}{N(T-1)} \mathbf{a}'_N \tilde{\underline{\Sigma}}_N \mathbf{b}_N \text{ with } \tilde{\underline{\Sigma}}_N = diag^{NT}_{n=1}(\tilde{\underline{v}}^2_{n,N}). \tag{C.5c}$$

The bias is derived using the expectation of the infeasible estimate $E\underline{\tau}_N$, which assumes that the true parameters $\boldsymbol{\rho}_N$ and $\boldsymbol{\delta}_N$ are known and omits the degrees of freedom correction for the P regressors. For simplicity of notation, define $c_{it,N} = a_{it,N} b_{it,N}$; without loss of generality, the bound c in the Lemma is chosen such that $c_{it,N} \leq c < \infty$.

Recognizing that $E\underline{\tau}_N = \frac{1}{N(T-1)} E(\mathbf{a}'_N \underline{\Sigma}_N \mathbf{b}_N) = \frac{1}{N} E\sum_{i=1}^{N} \underline{\tau}_{i,N}$ we have, for each i,

$$
\begin{aligned}
E\underline{\tau}_{i,N} &= \frac{1}{T-1} E\sum_{t=1}^{T} c_{it,N} \underline{v}^2_{it,N} \\
&= \frac{1}{T-1} E\sum_{t=1}^{T} c_{it,N} (v^2_{it,N} - 2v_{it,N}\bar{v}_{i,N} + \bar{v}^2_{i,N}) \\
&= \frac{1}{T-1} E\sum_{t=1}^{T} c_{it,N} v^2_{it,N} - \frac{1}{T-1} E\sum_{t=1}^{T} 2c_{it,N} v_{it,N} \frac{1}{T}\sum_{s=1}^{T} v_{is,N} + \frac{1}{T-1} E\sum_{t=1}^{T} \frac{1}{T^2} c_{it,N} \sum_{r=1}^{T}\sum_{s=1}^{T} v_{ir,N} v_{is,N} \\
&= \frac{T-2}{T-1} \tau_{i,N} + \frac{1}{(T-1)} \kappa_{i,N},
\end{aligned}
\tag{C.6}
$$

using $\tau_{i,N} = \frac{1}{T} E\sum_{t=1}^{T} c_{it,N} v^2_{it,N}$ and $\kappa_{i,N} = \frac{1}{T} E\sum_{t=1}^{T} c_{it,N} \frac{1}{T}\sum_{r=1}^{T} v^2_{ir,N}$.

Rearranging terms and averaging over N yields the following bias corrected estimator for τ_N:

$$\tilde{\tau}^{HR}_N = \frac{T-1}{T-2}[\tilde{\underline{\tau}}_N - \frac{1}{T-1}\tilde{\underline{\kappa}}_N], \tag{C.7}$$

where $\tilde{\underline{\tau}}_N = \dfrac{1}{N(T-1)}\displaystyle\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,N}\tilde{\underline{v}}_{it,N}^2$ and $\tilde{\underline{\kappa}}_N = \dfrac{1}{NT}\displaystyle\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,N}\dfrac{1}{T-1}\sum_{r=1}^{T}\tilde{\underline{v}}_{ir,N}^2$.[28] Finally, note

that (C.7) can also be written as $\tilde{\tau}_N^{HR} = \dfrac{1}{NT}\mathbf{a}_N'\tilde{\Sigma}_N^{HR}\mathbf{b}_N$, where $\tilde{\Sigma}_N^{HR}$ is a diagonal matrix with

elements $(\tilde{v}_{it,N}^{HR})^2 = \dfrac{T}{T-2}\tilde{v}_{it,N}^2 - \dfrac{1}{(T-1)(T-2)}\displaystyle\sum_{r=1}^{T}\tilde{v}_{ir,N}^2$.

We next prove that $\tilde{\tau}_N^{HR} - \tau_N = o_p(1)$, considering

$$\tilde{\tau}_N^{HR} - \tau_N = \frac{T-1}{T-2}(\tilde{\underline{\tau}}_N - E\underline{\tau}_N) - \frac{1}{T-2}(\tilde{\underline{\kappa}}_N - \kappa_N) \tag{C.8}$$

and showing that both $(\tilde{\underline{\tau}}_N - E\underline{\tau}_N)$ and $(\tilde{\underline{\kappa}} - \kappa_N)$ are $o_p(1)$ for fixed T as $N \to \infty$.

Consider first $\tilde{\underline{\tau}}_N - E\underline{\tau}_N$. It follows from the triangle inequality that

$$\left|\tilde{\underline{\tau}}_N - E\underline{\tau}_N\right| \le \left|\tilde{\underline{\tau}}_N - \underline{\tau}_N\right| + \left|\underline{\tau}_N - E\underline{\tau}_N\right|. \tag{C.9}$$

By the weak law of large numbers for i.d. variables (e.g., White, 2001, p. 35), we have

$$\underline{\tau}_N - E\underline{\tau}_N = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,N}(\underline{v}_{it,N}^2 - E\underline{v}_{it,N}^2) = o_p(1), \tag{C.10}$$

observing that the fourth moments of $v_{it,N}$ (and $\underline{v}_{it,N}$) are bounded uniformly by Assumption 1. We thus also have $\left|\underline{\tau}_N - E\underline{\tau}_N\right| = o_p(1)$.

Moreover, repeatedly using the triangle inequality, it follows that

$$\left|\underline{\tau}_N - E\underline{\tau}_N\right| = \frac{1}{N(T-1)}\left|\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,N}(\underline{v}_{it,N}^2 - E\underline{v}_{it,N}^2)\right| \tag{C.11}$$

$$\le c\zeta_{1,N}^{\tau} + 2c_* c^2,$$

where $\zeta_{1,N}^{\tau} = \dfrac{1}{N(T-1)}\displaystyle\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\left|\underline{v}_{it,N}^2 - E\underline{v}_{it,N}^2\right| - E\left|\underline{v}_{it,N}^2 - E\underline{v}_{it,N}^2\right|\right|$; the constant c_* is chosen such

that $Ev_{it,N}^2 \le c_*$ and $E\underline{v}_{it,N}^2 \le c_*$. Note that $\zeta_{1,N}^{\tau} = o_p(1)$ by the weak law of large numbers.

Next rewrite

$$\tilde{\underline{\tau}}_N = \frac{1}{N(T-1)}\mathbf{a}_N'\tilde{\underline{\Sigma}}_N\mathbf{b}_N = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,N}\tilde{\underline{v}}_{it,N}^2 = \frac{1}{N(T-1)}\tilde{\underline{\mathbf{v}}}_N'\mathbf{C}_N\tilde{\underline{\mathbf{v}}}_N , \text{ and}$$

---

[28] Note that $\kappa_{i,N} = E\underline{\kappa}_{i,N}$, where $\underline{\kappa}_{i,N} = \dfrac{1}{T}E\displaystyle\sum_{t=1}^{T} c_{it,N}\dfrac{1}{T-1}\sum_{r=1}^{T} v_{ir,N}^2$.

$$\underline{\tau}_N = \frac{1}{N(T-1)} \mathbf{v}'_N \mathbf{C}_N \mathbf{v}_N,$$

where $\mathbf{C}_N = diag_{n=1}^{NT}(c_{n,N}) = diag_{it=1}^{NT}(c_{it,N})$. Hence,

$$\tilde{\underline{\tau}}_N = \frac{1}{N(T-1)} \tilde{\mathbf{v}}'_N \underline{\mathbf{C}}_N \tilde{\mathbf{v}}_N \tag{C.12a}$$

$$= \frac{1}{N(T-1)} \tilde{\mathbf{u}}'_N \underline{\mathbf{C}}_N \tilde{\mathbf{u}}_N - \frac{2}{N(T-1)} \tilde{\mathbf{u}}'_N [\mathbf{I}_T \otimes (\sum_{m=1}^{S} \tilde{\rho}_{m,N} \mathbf{M}'_{m,N})] \underline{\mathbf{C}}_N \tilde{\mathbf{u}}_N$$

$$+ \frac{1}{N(T-1)} \tilde{\mathbf{u}}'_N [\mathbf{I}_T \otimes (\sum_{m=1}^{S} \tilde{\rho}_{m,N} \mathbf{M}'_{m,N})] \underline{\mathbf{C}}_N [\mathbf{I}_T \otimes \sum_{m=1}^{S} \tilde{\rho}_{m,N} \mathbf{M}_{m,N})] \tilde{\mathbf{u}}_N,$$

where $\underline{\mathbf{C}}_N = \mathbf{Q}_{0,N} \mathbf{C}_N \mathbf{Q}_{0,N}$, and

$$\underline{\tau}_N = \frac{1}{N(T-1)} \mathbf{v}'_N \underline{\mathbf{C}}_N \mathbf{v}_N \tag{C.12b}$$

$$= \frac{1}{N(T-1)} \mathbf{u}'_N \underline{\mathbf{C}}_N \mathbf{u}_N - 2(NT)^{-1} \mathbf{u}'_N [\mathbf{I}_T \otimes (\sum_{m=1}^{S} \rho_{m,N} \mathbf{M}'_{m,N})] \underline{\mathbf{C}}_N \mathbf{u}_N$$

$$+ \frac{1}{N(T-1)} \mathbf{u}'_N [\mathbf{I}_T \otimes (\sum_{m=1}^{S} \rho_{m,N} \mathbf{M}'_{m,N})] \underline{\mathbf{C}}_N [\mathbf{I}_T \otimes \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})] \mathbf{u}_N.$$

By the properties of the matrices $\mathbf{Q}_{0,N}$, $\mathbf{C}_N$, and $\mathbf{M}_{m,N}$, $m = 1,...,S$, and in light of Remark A.1, the expressions in (C.12b) are all quadratic forms in matrices whose row and column sums are bounded uniformly in absolute value by some constants that depend monotonically on c as well as on other bounds maintained in the assumptions.

Repeatedly using the triangle inequality, Lemma B.1 in Appendix B, and factoring out the $o_p(1)$ terms it follows that

$$\left| \tilde{\underline{\tau}}_N - \underline{\tau}_N \right| \leq k_\tau(c) \zeta_{2,N}^\tau, \tag{C.13}$$

where $\zeta_{2,N}^\tau = o_p(1)$ and does not depend on $\mathbf{a}_N$ and $\mathbf{b}_N$ and the constant $k_\tau(c)$ depends monotonically on c and other bounds maintained in the assumptions. Obviously, it follows that $\tilde{\underline{\tau}}_N - E\underline{\tau}_N = o_p(1)$. Moreover, we have

$$\left| E\underline{\tau}_N \right| \leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| c_{it,N} \right| \left| E \left| v_{it,N}^2 \right| \right| \leq cc_* < \infty. \tag{C.14}$$

It follows from (C.9), (C.11), and (C.13) that

$$\left| \tilde{\underline{\tau}}_N - \tau_N \right| \leq [c + 2cc_* + k_\tau(c)](\varsigma_N^\tau), \tag{C.15}$$

where $\varsigma_N^\tau = (1 + \zeta_N^\tau)$.

Next consider $(\tilde{\underline{\kappa}} - \kappa_N)$. By the triangle inequality, $\left|\tilde{\underline{\kappa}} - \kappa_N\right| \leq \left|\tilde{\underline{\kappa}} - \underline{\kappa}_N\right| + \left|\underline{\kappa}_N - \kappa_N\right|$ and by the weak law of large numbers

$$\underline{\kappa}_N - \kappa_N = o_p(1), \tag{C.16}$$

observing that the fourth moments of $v_{it,N}$ are bounded uniformly by Assumption 2. We thus also have $\left|\underline{\kappa}_N - \kappa_N\right| = o_p(1)$. Next, rewrite $\left|\underline{\kappa}_N - \kappa_N\right|$ as

$$\left|\underline{\kappa}_N - \kappa_N\right| = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}c_{it,N}(\underline{\bar{v}}_{i,N}^2 - E\bar{v}_{i,N}^2) \tag{C.17a}$$

where $\underline{\bar{v}}_{i,N}^2 = \frac{1}{T-1}\sum_{r=1}^{T}v_{ir,N}^2$ and $\bar{v}_{i,N}^2 = \frac{1}{T}\sum_{r=1}^{T}v_{ir,N}^2$. Moreover, repeatedly using the triangle inequality, it follows that

$$\left|\underline{\kappa}_N - \kappa_N\right| = \frac{1}{NT}\left|\sum_{i=1}^{N}\sum_{t=1}^{T}c_{it,N}(\underline{\bar{v}}_{i,N}^2 - E\bar{v}_{i,N}^2)\right| \tag{C.17b}$$

$$\leq c\zeta_{1,N}^{\kappa} + 2c^*c,$$

where the last step uses $\zeta_{1,N}^{\kappa} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\left|\bar{v}_{i,N}^2 - E\bar{v}_{i,N}^2\right| - E\left|\underline{\bar{v}}_{i,N}^2 - E\bar{v}_{i,N}^2\right|\right|$ and c_* is defined as above. Note that $\zeta_{1,N}^{\kappa} = o_p(1)$ by the weak law of large numbers.

From Lemma C.1 in Appendix C, it follows that $\left|\underline{\tilde{v}}_{it,N}^2 - \underline{v}_{it,N}^2\right| \leq 2\alpha_N \underline{\beta}_{it,N}\left|v_{it,N}\right| + \alpha_N^2\underline{\beta}_{it,N}^2$, where $E\underline{\beta}_{it,N}^4 \leq K_{\underline{\beta}} < \infty$ and $N^{1/2}\alpha_N = O_p(1)$. Using the triangle and Hölder inequality, we have

$$\left|\tilde{\underline{\kappa}}_N - \underline{\kappa}_N\right| = \frac{1}{NT}\left|\sum_{i=1}^{N}\sum_{t=1}^{T}c_{it,N}\frac{1}{T-1}\sum_{r=1}^{T}\underline{\tilde{v}}_{ir,N}^2 - \sum_{i=1}^{N}\sum_{t=1}^{T}c_{it,N}\frac{1}{T-1}\sum_{r=1}^{T}v_{ir,N}^2\right|$$

$$\leq k_{\kappa}(c)\zeta_{2,N}^{\kappa}, \text{ where } \zeta_{2,N}^{\kappa} = o_p(1). \tag{C.18}$$

Obviously, $\left|\tilde{\underline{\kappa}}_N - \underline{\kappa}_N\right| = o_p(1)$, which – together with (C.16) implies that $\tilde{\underline{\kappa}} - \kappa_N = o_p(1)$. It also holds that $\left|\kappa_N\right| < \infty$.

From (C.17) with (C.18) it follows that

$$\left|\tilde{\underline{\kappa}}_N - \kappa_N\right| \leq [c + 2cc_* + k_{\tau}(c)]\varsigma_N^{\kappa}, \tag{C.19}$$

where $\varsigma_N^{\kappa} = (1 + \zeta_N^{\kappa})$. Combining our results that $\tilde{\underline{\tau}}_N - E\underline{\tau}_N = o_p(1)$ and $\tilde{\underline{\kappa}} - \kappa_N = o_p(1)$, result (a) in Lemma C.2 follows in light of (C.8), noting that

$|\tau_N| \le \dfrac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|c_{it,N}|\,E|\sigma_{v,it,N}^2| \le cc_* < \infty$. Result (b) in Lemma follows from (C.15) and (C.19), which yields

$$\left|\tilde{\tau}_N^{HR} - \tau_N\right| \le 2\{[c + 2cc_* + k_\tau(c)]\varsigma_N^\tau + [c + 2cc_* + k_\tau(c)]\varsigma_N^\kappa\}$$

$$\le 2[c + 2cc_* + k(c)]\varsigma_N\,,$$

(C.20)

where $k(c) = \max[k_\tau(c), k_\kappa(c)]$ and $\varsigma_N = \varsigma_N^\tau + \varsigma_N^\kappa = (2 + \zeta_N^\tau + \zeta_N^\kappa)$.

Lemma C.3

Suppose Assumptions 1-4 hold. Furthermore, assume that $\sup_N \sum_{m=1}^{S}|\rho_{m,N}| < 1$, and that the row and column sums of $\mathbf{M}_{m,N}$, $m = 1,\dots,S$ are uniformly bounded in absolute value by 1 and some finite constant respectively. Let $\underline{\mathbf{v}}_N = \mathbf{Q}_{0,N}\boldsymbol{\varepsilon}_N = \mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})]\mathbf{u}_N$, and

let $\quad \underline{\tilde{\mathbf{v}}}_N = \mathbf{Q}_{0,N}\tilde{\boldsymbol{\varepsilon}}_N = \mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})\tilde{\mathbf{u}}_N \quad$ with $\quad \tilde{\mathbf{u}}_N = \mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N \quad$ and

$\mathbf{D}_N = (\mathbf{d}'_{1.,N},\dots,\mathbf{d}'_{NT.,N})'$, and where the $S \times 1$ vector $\tilde{\boldsymbol{\rho}}_N$ can be any estimator that satisfies $(\tilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = o_p(1)$.

Let $\quad \mathbf{F}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]\mathbf{H}_N \quad$ and $\quad \tilde{\mathbf{F}}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})^{+}]\mathbf{H}_N$, where $\mathbf{H}_N$ is an $N \times P_*$ matrix whose elements are uniformly bounded in absolute value by some constant $c < \infty$, and let $\boldsymbol{\Sigma}_N$ and $\tilde{\boldsymbol{\Sigma}}_N^{HR}$ be defined as in Lemma C.2. Then,

$\dfrac{1}{NT}\tilde{\mathbf{F}}_N'\tilde{\boldsymbol{\Sigma}}_N^{HR}\tilde{\mathbf{F}}_N - \dfrac{1}{NT}\mathbf{F}_N'\boldsymbol{\Sigma}_N\mathbf{F}_N = o_p(1)$ and $\dfrac{1}{NT}\mathbf{F}_N'\boldsymbol{\Sigma}_N\mathbf{F}_N = O(1)$.

Proof.

The subsequent proof will focus on the case, where $\mathbf{F}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}]\mathbf{H}_N$ and

$\tilde{\mathbf{F}}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})^{+}]\mathbf{H}_N$; this corresponds to the random effects estimation of the untransformed model (see Lemma 1); it is readily observed from the proof that this covers also the case where $\mathbf{F}_N = \underline{\mathbf{F}}_N$ (fixed effects estimation of untransformed model), $\mathbf{F}_N = \mathbf{F}_N^*$ (random effects estimation of transformed model) as well as $\mathbf{F}_N = \underline{\mathbf{F}}_N^*$ (fixed effects estimation of transformed model).

Under the maintained assumptions there exists a ρ_* with $\sup \sum_{m=1}^{S} |\rho_{m,N}| < \rho_* < 1$. By the properties of the matrices $\mathbf{M}_{m,N}$ the row and column sums of $\rho_* \mathbf{M}_{m,N}$, $m = 1, ..., S$ are uniformly bounded in absolute value by 1 and some finite constant respectively. For later reference, also note that the elements of the vector $\rho_*^k \mathbf{M}_N^k \mathbf{h}_{.s,N}$ are also uniformly bounded in absolute value by c.

In the following, we ignore the division by (the fixed constant) T without consequences for the proof. Denote the (r,s)-th element of the difference $\dfrac{1}{NT} \widetilde{\mathbf{F}}_N' \widetilde{\mathbf{\Sigma}}_N^{HR} \widetilde{\mathbf{F}}_N - \dfrac{1}{NT} \mathbf{F}_N \mathbf{\Sigma}_N \mathbf{F}_N$ as v_N. It is given by

$$v_N = N^{-1}(\widetilde{\mathbf{f}}_{.r,N}' \widetilde{\mathbf{\Sigma}}_N^{HR} \widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.r,N}' \mathbf{\Sigma}_N \mathbf{f}_{.s,N}), \quad r, s = 1, ..., P_*, \tag{C.21}$$

which can be written as $v_N = \sum_{i=1}^{7} v_{i,N}$, where

$$v_{1,N} = N^{-1}(\widetilde{\mathbf{f}}_{.r,N} - \mathbf{f}_{.r,N})'(\widetilde{\mathbf{\Sigma}}_N^{HR} - \mathbf{\Sigma}_N)(\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N}) \tag{C.22}$$

$$v_{2,N} = N^{-1}(\widetilde{\mathbf{f}}_{.r,N} - \mathbf{f}_{.r,N})'(\widetilde{\mathbf{\Sigma}}_N^{HR} - \mathbf{\Sigma}_N)\mathbf{f}_{.s,N}$$

$$v_{3,N} = N^{-1}\mathbf{f}_{.r,N}'(\widetilde{\mathbf{\Sigma}}_N^{HR} - \mathbf{\Sigma}_N)(\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N})$$

$$v_{4,N} = N^{-1}\mathbf{f}_{.r,N}'(\widetilde{\mathbf{\Sigma}}_N^{HR} - \mathbf{\Sigma}_N)\mathbf{f}_{.s,N}$$

$$v_{5,N} = N^{-1}(\widetilde{\mathbf{f}}_{.r,N} - \mathbf{f}_{.r,N})'\mathbf{\Sigma}_N(\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N})$$

$$v_{6,N} = N^{-1}(\widetilde{\mathbf{f}}_{.r,N} - \mathbf{f}_{.r,N})'\mathbf{\Sigma}_N\mathbf{f}_{.s,N}$$

$$v_{7,N} = N^{-1}\mathbf{f}_{.r,N}'\mathbf{\Sigma}_N(\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N}).$$

Next note that $\mathbf{f}_{.s,N} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})^{-1}]\mathbf{h}_{.s,N}$ and thus

$$\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N} = \{\mathbf{I}_T \otimes [(\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N}\mathbf{M}_{m,N})^+ - (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})^{-1}]\}\mathbf{h}_{.s,N} \tag{C.23}$$

We next demonstrate that $\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N} = o_p(1)$ by showing that each summand $v_{i,N} = o_p(1)$, $i = 1, ..., 7$, invoking the following theorem (see, e.g., Resnik, 1999, p. 171): Let $(X, X_N, N \geq 1)$ be real valued random variables. Then, $X_N \xrightarrow{p} X$ if and only if each subsequence X_{N_a} contains a further subsequence $X_{N'_a}$ that converges almost surely to X.

As we show below we will be confronted with terms of the form:

$$\aleph_N^{(k,l)} = N^{-1}p_*^{l+k}\mathbf{h}'_{.r,N}(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N^{\prime l})\,\widetilde{\boldsymbol{\Sigma}}_N^{HR}(\mathbf{I}_T \otimes \mathbf{M}_N^k)\mathbf{h}_{.s,N} - N^{-1}p_*^{l+k}\mathbf{h}'_{.r,N}((\mathbf{I}_T \otimes \overline{\mathbf{M}}_N^{\prime l})\boldsymbol{\Sigma}_N((\mathbf{I}_T \otimes \overline{\mathbf{M}}_N^k)\mathbf{h}_{.s,N}\ .\text{(C.24)}$$

where $\overline{\mathbf{M}}_N$ is a matrix, whose row and column sums are uniformly bounded in absolute value by some constant $c_{\overline{M}}$ It follows that the absolute values of the elements of the vector $(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N^k)\mathbf{h}_{.s,N}$ (and also that of $\rho_*^k(\mathbf{I}_T \otimes \overline{\mathbf{M}}_N^k)\mathbf{h}_{.s,N}$) are uniformly bounded in absolute value by some finite constant $c_* = \overline{c}c$ (and $c_{**} = \rho_*^k \overline{c}c < c_*$). (See Remark A.1 in Appendix A.) Without loss of generality c_* is chosen such that $Ev_{it,N}^2 \leq c_*$ and $E\underline{v}_{it,N}^2 \leq c_*$ holds.

Hence, Lemma C.2 applies and it follows that $\aleph_N^{(k,l)} = o_p(1)$ and that there exist random variables $\varsigma_N = o_p(1)$ such that $\left|\aleph_N^{(k,l)}\right| \leq K(c_*)(1+\varsigma_N)$.

Now, let the index N_a denote some subsequence. In light of the aforementioned equivalence, there exists a subsequence of this subsequence ($N_{a'}$) such that for events $\omega \in A$, with $P(A^C) = 0$, it holds that

$$\left|\aleph_{N_a'}^{(k,l)}(\omega)\right| \to 0,\ \left|\varsigma_{N_a'}(\omega)\right| \to 0,\ \left|\widetilde{\rho}_{m,N_a'}(\omega) - \rho_{m,N_a'}\right| \to 0,\ m=1,...,S \tag{C.25}$$

and that for some $N_a' \geq \overline{N}_\omega$, $\left|\varsigma_{N_a'}(\omega)\right| \leq 1$ and thus

$$\left|\aleph_{N_a'}^{(k,l)}(\omega)\right| \leq K(c_*)(1+\left|\varsigma_{N_a'}(\omega)\right|) \leq 2K(c_*)\,, \tag{C.26}$$

and finally

$$\sum_{m=1}^S \left|\widetilde{\rho}_{m,N_a'}(\omega)\right| \leq p_{**}\,,\ \text{where}\ p_{**} = \frac{\sup_N \sum_{m=1}^S \left|\rho_{m,N}\right| + p_*}{2} < 1\,. \tag{C.27}$$

In the following, assume that $N_a' \geq \overline{N}_\omega$. Since $\sum_{m=1}^S \left|\widetilde{\rho}_{m,N_a'}(\omega)\right| < 1$, it follows from Horn and Johnson (1985, p. 301) that $(\mathbf{I}_N - \sum_{m=1}^S \widetilde{\rho}_{m,N_a'}(\omega)\mathbf{M}_{m,N_a'})^+$ is invertible and that

$$\widetilde{\mathbf{f}}_{.s,N_a'}(\omega) - \mathbf{f}_{.s,N_a'} = \mathbf{I}_T \otimes \{[(\mathbf{I}_N - \sum_{m=1}^S \widetilde{\rho}_{m,N_a'}(\omega)\mathbf{M}_{m,N_a'})^{-1} - \mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^S \rho_{m,N_a'}\mathbf{M}_{m,N_a'})^{-1}]\}\mathbf{h}_{.s,N_a'} \tag{C.28}$$

$$= \mathbf{I}_T \otimes \sum_{l=1}^\infty [(\sum_{m=1}^S \widetilde{\rho}_{m,N_a'}(\omega)\mathbf{M}_{m,N_a'})^l - (\sum_{m=1}^S \rho_{m,N_a'}\mathbf{M}_{m,N_a'})^l]\mathbf{h}_{.s,N_a'}\,.$$

Substituting into the expression for $v_{1,N_a'}$ given by (C.22) yields

$$v_{1,N_a'} = N_a'^{-1}(\widetilde{\mathbf{f}}_{.r,N_a'}(\omega) - \mathbf{f}_{.r,N_a'})'(\widetilde{\boldsymbol{\Sigma}}_{N_a'}^{HR} - \boldsymbol{\Sigma}_{N_a'})(\widetilde{\mathbf{f}}_{.s,N_a'}(\omega) - \mathbf{f}_{.s,N_a'}) \tag{C.29}$$

$$= N_a'^{-1} \mathbf{h}'_{.r,N_a'} \{\mathbf{I}_T \otimes \sum_{l=1}^{\infty} [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}_{m,N_a'})^l - (\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}_{m,N_a'})^l]'\}$$

$$\times (\tilde{\boldsymbol{\Sigma}}_{N_a'} - \boldsymbol{\Sigma}_{N_a'})\{\mathbf{I}_T \otimes \sum_{k=1}^{\infty} [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}_{m,N_a'})^k - (\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}_{m,N_a'})^k]\} \mathbf{h}_{.s,N_a'}$$

$$= N_a'^{-1} \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \mathbf{h}'_{.r,N_a'} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}'_{m,N_a'})^l - (\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}'_{m,N_a'})^l]\} \tilde{\boldsymbol{\Sigma}}_{N_a'}^{HR} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}_{m,N_a'})^k - (\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}_{m,N_a'})^k]\} \mathbf{h}_{.s,N_a'}$$

$$- N_a'^{-1} \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \mathbf{h}'_{.r,N_a'} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}'_{m,N_a'})^l - (\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}'_{m,N_a'})^l]\} \boldsymbol{\Sigma}_{N_a'} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}_{m,N_a'})^k - (\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}_{m,N_a'})^k]\} \mathbf{h}_{.s,N_a'} .$$

A single element with index (k,l) of this infinite double sum over k and l is given by

$$= N_a'^{-1} \mathbf{h}'_{.r,N_a'} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}'_{m,N_a'})^l - (\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}'_{m,N_a'})^l]\} \tilde{\boldsymbol{\Sigma}}_{N_a'}^{HR} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}_{m,N_a'})^k - (\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}_{m,N_a'})^k]\} \mathbf{h}_{.s,N_a'}$$

$$- N_a'^{-1} \mathbf{h}'_{.r,N_a'} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}'_{m,N_a'})^l - (\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}'_{m,N_a'})^l]\} \boldsymbol{\Sigma}_{N_a'} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}_{m,N_a'})^k - (\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}_{m,N_a'})^k]\} \mathbf{h}_{.s,N_a'} .$$

$$(C.30)$$

Next note that for any value of $\boldsymbol{\rho}_{N_a'}$ and any $\tilde{\boldsymbol{\rho}}_{N_a'}(\omega)$ there exist matrices $\overline{\overline{\mathbf{M}}}_{N_a'}$ and $\overline{\overline{\overline{\mathbf{M}}}}_{N_a'}$, whose row and column sums are uniformly bounded in absolute value, such that:

$$\sum_{m=1}^{S} \rho_{m,N_a'} \mathbf{M}_{m,N_a'} = (\sum_{m=1}^{S} \rho_{m,N_a'}) \overline{\overline{\mathbf{M}}}_{N_a'} \text{ and } \sum_{m=1}^{S} \tilde{\rho}_{m,N_a'} \mathbf{M}_{m,N_a'} = (\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'}) \overline{\overline{\overline{\mathbf{M}}}}_{N_a'} . \qquad (C.31)$$

$\overline{\overline{\mathbf{M}}}_{N_a'}$ and $\overline{\overline{\overline{\mathbf{M}}}}_{N_a'}$ can thus be factored out of the sum, yielding

$$(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'}(\omega))^l \overline{\overline{\overline{\mathbf{M}}}}_{N_a'}^l - (\sum_{m=1}^{S} \rho_{m,N_a'})^l \overline{\overline{\mathbf{M}}}_{N_a'}^l . \qquad (C.32)$$

By the same reasoning, for any values of $(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'}(\omega))$ and $(\sum_{m=1}^{S} \rho_{m,N_a'})$, there exists a matrix $\overline{\mathbf{M}}_{N_a'}$, whose row and column sums are uniformly bounded in absolute value, such that:

$$(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'}(\omega))^l \overline{\mathbf{M}}_{N_a'} - (\sum_{m=1}^{S} \rho_{m,N_a'})^l \overline{\mathbf{M}}_{N_a'} = [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'}(\omega))^l - (\sum_{m=1}^{S} \rho_{m,N_a'})^l] \overline{\mathbf{M}}_{N_a'} . \qquad (C.33)$$

Substituting $\overline{\mathbf{M}}_{N_a'}$ into the expression for $\nu_{1,N_a'}$, we obtain

$$(C.127)$$

$$\nu_{1,N_a'} = N_a'^{-1} \mathbf{h}'_{.r,N_a'} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'}(\omega))^l - (\sum_{m=1}^{S} \rho_{m,N_a'})^l] \overline{\mathbf{M}}'^l_{N_a'}\} \tilde{\boldsymbol{\Sigma}}_{N_a'}^{HR} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'}(\omega))^k - (\sum_{m=1}^{S} \rho_{m,N_a'})^k] \overline{\mathbf{M}}'^k_{N_a'}\} \mathbf{h}_{.s,N_a'}$$

$$- N_a'^{-1} \mathbf{h}'_{.r,N_a'} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'}(\omega))^l - (\sum_{m=1}^{S} \rho_{m,N_a'})^l] \overline{\mathbf{M}}'^l_{N_a'}\} \boldsymbol{\Sigma}_{N_a'} \{\mathbf{I}_T \otimes [(\sum_{m=1}^{S} \tilde{\rho}_{m,N_a'}(\omega))^k - (\sum_{m=1}^{S} \rho_{m,N_a'})^k] \overline{\mathbf{M}}'^k_{N_a'}\} \mathbf{h}_{.s,N_a'} .$$

Hence, we can then write

$$v_{1,N'_a}(\omega) = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} X_{N'_a}^{(k,l)}(\omega) \, , \tag{C.34}$$

where $X_{N'_a}^{(k,l)}(\omega) = a_{N'_a}^{(k,l)}(\omega)\aleph_{N'_a}^{(k,l)}(\omega)$ with

$$a_{N'_a}^{(k,l)}(\omega) = \frac{[(\sum_{m=1}^{S}\tilde{\rho}_{m,N'_a}(\omega))^l - (\sum_{m=1}^{S}\rho_{m,N'_a})^l]}{\rho_*^l} \frac{[(\sum_{m=1}^{S}\tilde{\rho}_{m,N'_a})^l - (\sum_{m=1}^{S}\rho_{m,N'_a})^k]}{\rho_*^k} \tag{C.35}$$

and

$$\aleph_{N'_a}^{(k,l)}(\omega) = N'^{-1}p_*^{l+k}\mathbf{h}'_{.r,N'_a}(\mathbf{I}_T\otimes\overline{\mathbf{M}}_{N'_a}'^l)\,\tilde{\mathbf{\Sigma}}_{N'_a}^R(\omega)(\mathbf{I}_T\otimes\overline{\mathbf{M}}_{N'_a}^k)\mathbf{h}_{.s,N'} - N^{-1}p_*^{l+k}\mathbf{h}'_{.r,N'_a}(\mathbf{I}_T\otimes\overline{\mathbf{M}}_{N'_a}'^l)\mathbf{\Sigma}_{N'_a}(\mathbf{I}_T\otimes\overline{\mathbf{M}}_N^k)\mathbf{h}_{.s,N'_a} \tag{C.36}$$

Note that $\left|a_{N'_a}^{(k,l)}\right|\to 0$ as $N'_a\to\infty$ in light of the aforementioned results and thus $\left|X_{N'_a}^{(k,l)}(\omega)\right|\to 0$ since $\left|\aleph_{N'_a}^{(k,l)}(\omega)\right|\le 2K(c_*)$ for N'_a large enough. Moreover,

$$a_{N'_a}^{(k,l)}(\omega) = \frac{[(\sum_{m=1}^{S}\tilde{\rho}_{m,N'_a}(\omega))^l - (\sum_{m=1}^{S}\rho_{m,N'_a})^l]}{\rho_*^l} \frac{[(\sum_{m=1}^{S}\tilde{\rho}_{m,N'_a})^l - (\sum_{m=1}^{S}\rho_{m,N'_a})^k]}{\rho_*^k} \, . \tag{C.37}$$

Hence,

$$\left|a_{N'_a}\right| \le 2\left(\frac{\rho_{**}}{\rho_*}\right)^l 2\left(\frac{\rho_{**}}{\rho_*}\right)^k = 4\left(\frac{\rho_{**}}{\rho_*}\right)^{l+k} \, .$$

For $N\ge\overline{N}_\omega$, $\left|\aleph_{N'_a}^{(k,l)}(\omega)\right|\le 2K(c_*)$, such that

$$\left|X_{N'_a}^{(k,l)}(\omega)\right| \le B^{(l,k)} = 2K(c)4\left(\frac{\rho_{**}}{\rho_*}\right)^{l+k} = 8K(c)\left(\frac{\rho_{**}}{\rho_*}\right)^{l+k} \, . \tag{C.38}$$

Hence, there exists a dominating function $B^{(l,k)}$ for all values of k,l. Moreover, since $\rho_{**}/\rho_* <1$ by construction, the dominating function is integrable (summable), i.e.,

$$\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\left|B^{(l,k)}\right| = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}B^{(l,k)} < \infty \, . \tag{C.39}$$

Hence the assumptions for application of Lebesgue's Dominated Convergence Theorem are fulfilled (see, e.g., Van der Vaart and Yen, 1968), such that

$$\lim_{N'_a\to\infty} v_{1,N'_a}(\omega) = 0 \, . \tag{C.40}$$

The same holds for v_{i,N'_a}, $i=2,...,7$. It follows that $v_{i,N'_a}\to 0$ as $N'_a\to\infty$ and in light of Resnik (1999) it follows that $v_N\to o_p(1)$.

Thus, $N^{-1}\widetilde{\mathbf{F}}'_N\widetilde{\mathbf{\Sigma}}^{HR}_N\widetilde{\mathbf{F}}_N - N^{-1}\mathbf{F}'_N\mathbf{\Sigma}_N\mathbf{F}_N = o_p(1)$. That $N^{-1}\mathbf{F}'_N\mathbf{\Sigma}_N\mathbf{F}_N = O(1)$ follows from the properties maintained for the row and column sums of $(\mathbf{I}_N - \sum_{m=1}^{S}\rho_m\mathbf{M}_{m,N})^{-1}$ and the elements of $\mathbf{\Sigma}_N$ and $\mathbf{H}_N$.

Remark C.1

Regarding $\widetilde{\mathfrak{E}}^{**}_{\mathbf{v},N}$, note that $\mathbf{a}_{\mathbf{v},N} = \mathbf{F}_{\mathbf{v},N}\mathbf{P}_N\boldsymbol{\alpha}_N$ and $\widetilde{\mathbf{a}}_{\mathbf{v},N} = \widetilde{\mathbf{F}}_{\mathbf{v},N}\widetilde{\mathbf{P}}_N\widetilde{\boldsymbol{\alpha}}_N$ (obviously suppressing the indexation of $\widetilde{\boldsymbol{\alpha}}_N$), and accordingly for $\mathbf{b}_{\mathbf{v},N}$. By assumption $\widetilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$, $\mathbf{P}_N = O(1)$ and thus $\widetilde{\mathbf{P}}_N = O_p(1)$, where the dimension of $\mathbf{P}_N$ is $P_* \times P$. Moreover, $\widetilde{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N = o_p(1)$, $\boldsymbol{\alpha}_N = O(1)$ and thus $\widetilde{\boldsymbol{\alpha}}_N = O_p(1)$, where the dimension of $\boldsymbol{\alpha}_N$ is $P \times 1$. By Lemma C.3, we have $\qquad N^{-1}\widetilde{\mathbf{F}}'_N\widetilde{\mathbf{\Sigma}}^{HR}_N\widetilde{\mathbf{F}}_N - N^{-1}\mathbf{F}'_N\mathbf{\Sigma}^{HR}_N\mathbf{F}_N = o_p(1)$. $\qquad$ It $\qquad$ follows $\qquad$ that $\widetilde{\mathfrak{E}}^{**}_{\mathbf{v},N} - \mathfrak{E}^{**}_{\mathbf{v},N} = N^{-1}\widetilde{\boldsymbol{\alpha}}'_N\widetilde{\mathbf{P}}'_N\widetilde{\mathbf{F}}'_N\widetilde{\mathbf{\Sigma}}^{HR}_N\widetilde{\mathbf{F}}_N\widetilde{\mathbf{P}}_N\widetilde{\boldsymbol{\alpha}}_N - N^{-1}\boldsymbol{\alpha}'_N\mathbf{P}'_N\mathbf{F}'_N\mathbf{\Sigma}_N\mathbf{F}_N\mathbf{P}_N\boldsymbol{\alpha}_N = o_p(1)$.

Lemma C.4

Let $\widetilde{\sigma}^{(3)}_{\mu,N}$ and $\widetilde{\sigma}^{(4)}_{\mu,N}$ be defined as in (A.3a) and (A.3b). Suppose that Assumptions 1-4 hold and that the elements of $\mathbf{D}_N = (\mathbf{d}'_{1.,N},...,\mathbf{d}'_{N.,N})'$ have bounded fourth moments. It follows that $\widetilde{\sigma}^{(3)}_{\mu,N} - \sigma^{(3)}_\mu = o_p(1)$, $\sigma^{(3)}_\mu = O(1)$, and thus $\widetilde{\sigma}^{(3)}_{\mu,N} = O_p(1)$, and that $\widetilde{\sigma}^{(4)}_{\mu,N} - \sigma^{(4)}_\mu = o_p(1)$, $\sigma^{(4)}_\mu = O(1)$ and thus $\widetilde{\sigma}^{(4)}_{\mu,N} = O_p(1)$.

Proof.

The subsequent proof builds on Gilbert (2002), who considers the estimation of third and fourth moments in homoskedastic error component models without spatial lags of the dependent variable (or other endogenous variables) and without spatial regressive disturbances.

Consider the third moment of $\mu_{i,N}$ and its estimate:

$$\sigma^{(3)}_\mu = E(\varepsilon_{is,N}\varepsilon^2_{it,N}) \text{ for any given } i \text{ and } s \neq t, \text{ and} \qquad (C.41a)$$

$$\widetilde{\sigma}^{(3)}_{\mu,N} = \frac{1}{NT(T-1)}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{\substack{t=1 \\ t \neq s}}^{T}\widetilde{\varepsilon}_{is,N}\widetilde{\varepsilon}^2_{it,N} . \qquad (C.41b)$$

By Assumption 1, $\sigma^{(3)}_\mu$ is invariant to the choice of i, s and t. Using (C.1), we have

$$\widetilde{\sigma}^{(3)}_{\mu,N} = \frac{1}{NT(T-1)}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{\substack{t=1 \\ t \neq s}}^{T}(\varepsilon_{is,N} + \eta_{is,N})(\varepsilon_{it,N} + \eta_{it,N})^2 \qquad (C.42)$$

$$= \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} (\varepsilon_{is,N} \varepsilon_{it,N}^2 + 2\eta_{it,N} \varepsilon_{is,N} + \eta_{it,N}^2 \varepsilon_{is,N} + \eta_{is,N} \varepsilon_{it,N}^2 + 2\eta_{is,N} \eta_{it,N} + \eta_{is,N} \eta_{it,N}^2)$$

$$= \phi_{1,N} + \phi_{2,N} + \phi_{3,N} + \phi_{4,N} + \phi_{5,N} + \phi_{6,N} .$$

Consider

$$\phi_{1,N} = \sum_{i=1}^{N} \sum_{s}^{T} \sum_{t \neq s}^{T} \varepsilon_{is,N} \varepsilon_{it,N}^2 = \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} (\mu_{is,N} + v_{is,N})(\mu_{it,N} + v_{it,N})^2 \qquad \text{(C.43a)}$$

$$= \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} (\mu_{i,N}^3 + 2\mu_{i,N}^2 v_{it,N} + \mu_{i,N} v_{it,N}^2 + v_{is,N} \mu_{i,N}^2 + 2\mu_{i,N} v_{it,N} v_{is,N} + v_{it,N}^2 v_{is,N})$$

$$= \phi_{11,N} + \phi_{12,N} + \phi_{13,N} + \phi_{14,N} + \phi_{15,N} + \phi_{16,N} .$$

By the weak law of large numbers $\phi_{11,N}$ converges in probability to $\sigma_{\mu}^{(3)}$. Notice further that, by the properties of $v_{it,N}$ and $\mu_{it,N}$ (see Assumption 1), $\phi_{12,N}, \phi_{13,N}, \phi_{14,N}, \phi_{15,N}$, and $\phi_{16,N}$ are all $o_p(1)$. As a consequence, $\phi_{1,N}$ converges in probability to $\sigma_{\mu}^{(3)}$.

Next observe that

$$\phi_{2,N} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} 2\eta_{it,N} \varepsilon_{is,N} \qquad \text{(C.43b)}$$

$$\leq \frac{2}{NT(T-1)} \alpha_N \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \varepsilon_{is,N} \beta_{it,N}$$

$$\leq 2N^{1/2} \alpha_N N^{-1/2} \left([NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} |\varepsilon_{is,N}|^2 \right)^{1/2} \left([NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \beta_{it,N}^2 \right)^{1/2} = o_p(1),$$

$$\phi_{3,N} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \eta_{it,N}^2 \varepsilon_{is,N} \qquad \text{(C.43c)}$$

$$\leq (N^{1/2} \alpha_N)^2 N^{-1} \left([NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} |\varepsilon_{is,N}|^2 \right)^{1/2} \left([NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \beta_{it,N}^4 \right)^{1/2} = o_p(1),$$

$$\phi_{4,N} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \eta_{is,N} \varepsilon_{it,N}^2 \qquad \text{(C.43d)}$$

$$\leq N^{1/2} \alpha_N N^{-1/2} \left([NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{s}^{T} \sum_{t \neq s}^{T} |\varepsilon_{is,N}|^4 \right)^{1/2} \left([NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{s}^{T} \sum_{t \neq s}^{T} \beta_{it,N}^2 \right)^{1/2} = o_p(1),$$

$$\phi_{5,N} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s}^{T} \sum_{t \neq s}^{T} 2\eta_{is,N}\eta_{it,N} \tag{C.43e}$$

$$\leq \frac{2}{NT(T-1)} (N^{1/2}\alpha_N)^2 N^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \beta_{is,N}\beta_{it,N}$$

$$\leq 2(N^{1/2}\alpha_N)^2 N^{-1} \left([NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \beta_{is,N}^2 \right)^{1/2} \left([NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \beta_{it,N}^2 \right)^{1/2} = o_p(1),$$

$$\phi_{6,N} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \eta_{is,N}\eta_{it,N}^2 \tag{C.43f}$$

$$\leq 2(N^{1/2}\alpha_N)^3 N^{-3/2} \left([NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \beta_{is,N}^2 \right)^{1/2} \left([NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \beta_{it,N}^4 \right)^{1/2} = o_p(1),$$

because $N^{1/2}\alpha_N$ is $O_p(1)$ and the terms in brackets expressions are all $O_p(1)$, since $E|\varepsilon_{is,N}|^\gamma < K < \infty$ and $E|\beta_{it,N}|^\gamma < K < \infty$ for $\gamma \leq 4$ and all N. It follows that $\tilde{\sigma}_{\mu,N}^{(3)} - \sigma_\mu^{(3)} = o_p(1)$, $\sigma_\mu^{(3)} = O(1)$ by Assumption 1, and that $\tilde{\sigma}_{\mu,N}^{(3)} = O_p(1)$. Obviously, we then also have that $(\tilde{\sigma}_{\varepsilon,N}^{(3)} - \tilde{\sigma}_{\mu,N}^{(3)}) - \sigma_v^{(3)} = \tilde{\sigma}_{v,N}^{(3)} - \sigma_v^{(3)} = o_p(1)$.

Consider next the fourth moment of $\mu_{i,N}$ and its estimate:

$$\sigma_\mu^4 = E(\varepsilon_{is}\varepsilon_{it}^3) - 3E(\varepsilon_{is}\varepsilon_{it})[E(\varepsilon_{it}^2) - E(\varepsilon_{is}\varepsilon_{it})] \text{ for any given } i \text{ and } s \neq t, \tag{C.44a}$$

$$\tilde{\sigma}_{\mu,N}^{(4)} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N}\tilde{\varepsilon}_{it,N}^3 \tag{C.44b}$$

$$- \frac{3}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N}\tilde{\varepsilon}_{it,N} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\varepsilon}_{it,N}^2 - \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N}\tilde{\varepsilon}_{it,N} \right)$$

$$= \delta_{1,N} - \delta_{2,N}(\tilde{\sigma}_{\varepsilon,N}^2 - \delta_{2,N}).$$

Observe that

$$\delta_{1,N} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} (\varepsilon_{is,N} + \eta_{is,N})(\varepsilon_{it,N} + \eta_{it,N})^3 \tag{C.45}$$

$$= \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} (\varepsilon_{is,N}\varepsilon_{it,N}^3 + 3\varepsilon_{is,N}\varepsilon_{it,N}^2\eta_{it,N} + 3\eta_{it,N}^2\varepsilon_{is,N}\varepsilon_{it,N} + \varepsilon_{is,N}\eta_{it,N}^3$$

$$+ \eta_{is,N}\varepsilon_{it,N}^3 + 3\varepsilon_{it,N}^2\eta_{it,N}\eta_{is,N} + 3\eta_{it,N}^2\eta_{is,N}\varepsilon_{it,N} + \eta_{it,N}^3\eta_{is,N})$$

$$= \delta_{11,N} + \delta_{12,N} + \delta_{13,N} + \delta_{14,N} + \delta_{15,N} + \delta_{16,N} + \delta_{17,N} + \delta_{18,N}.$$

The first term $\delta_{11,N}$ can also be written as

$$\delta_{11,N} = \varepsilon_{is,N} \varepsilon_{it,N}^3 = (\mu_{is,N} + v_{is,N})(\mu_{it,N} + v_{it,N})^3 = \tag{C.46}$$

$$= (\mu_{is,N} + v_{is,N})(\mu_{it,N}^3 + 3\mu_{it,N}^2 v_{it,N} + 3\mu_{it,N} v_{it,N}^2 + v_{it,N}^3)$$

$$= (\mu_{is,N}\mu_{it,N}^3 + 3\mu_{is,N}\mu_{it,N}^2 v_{it,N} + 3\mu_{is,N}\mu_{it,N} v_{it,N}^2 + \mu_{is,N} v_{it,N}^3$$

$$+ v_{is,N}\mu_{it,N}^3 + 3\mu_{it,N}^2 v_{is,N} v_{it,N} + 3\mu_{it,N} v_{is,N} v_{it,N}^2 + v_{is,N} v_{it,N}^3)$$

$$= (\mu_{i,N}^4 + 3\mu_{i,N}^3 v_{it,N} + 3\mu_{i,N}^2 v_{it,N}^2 + \mu_{is,N} v_{it,N}^3 + v_{is,N}\mu_{it,N}^3 + 3\mu_{it,N}^2 v_{is,N} v_{it,N} + 3\mu_{it,N} v_{is,N} v_{it,N}^2 + v_{is,N} v_{it,N}^3).$$

By the properties of $v_{it,N}$ and $\mu_{it,N}$ (see Assumption 1), the difference between $\delta_{11,N}$ and

$\sigma_\mu^{(4)} + \sigma_\mu^2 E \dfrac{3}{N}\sum_{i=1}^{N}\dfrac{1}{T}\sum_{t=1}^{T} v_{it,N}^2$ converges in probability to zero by the weak law of large numbers

for i.d. random variables (White, 2001, p. 37, Corollary 3.9).

Moreover, it follows from the properties of $\mathbf{v}_N$ and $\boldsymbol{\mu}_N$ (see Assumption 1), that the terms

$\delta_{12,N}, \delta_{13,N}, \delta_{14,N}, \delta_{15,N}, \delta_{16,N}, \delta_{17,N}, \delta_{18,N}$ are all $o_p(1)$. It follows that the difference between $\delta_{1,N}$

and $\sigma_\mu^{(4)} + \sigma_\mu^2 \dfrac{3}{N}\sum_{i=1}^{N}\dfrac{1}{T}\sum_{t=1}^{T}\sigma_{v,it,N}^2$ converges in probability to zero.

Next consider

$$\delta_{2,N} = \frac{3}{NT(T-1)}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t\neq s}^{T}(\varepsilon_{is,N} + \eta_{it,N})(\varepsilon_{it,N} + \eta_{it,N}) \tag{C.47}$$

$$= \frac{3}{NT(T-1)}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t\neq s}^{T}(\varepsilon_{is,N} + \eta_{is,N})(\varepsilon_{it,N} + \eta_{it,N})$$

$$= \frac{3}{NT(T-1)}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t\neq s}^{T}(\varepsilon_{is,N}\varepsilon_{it,N} + \eta_{is,N}\varepsilon_{it,N} + \varepsilon_{is,N}\eta_{it,N} + \eta_{is,N}\eta_{it,N}),$$

which converges to $3\sigma_\mu^2$ by the weak law of large numbers, since

$E[\dfrac{1}{T}\sum_{s=1}^{T}\varepsilon_{is,N}\dfrac{1}{(T-1)}\sum_{t\neq s}^{T}\varepsilon_{it,N}] = \sigma_\mu^2$ for $s\neq t$ by the properties of $v_{it,N}$ and $\mu_{it,N}$ and the sum

over the remainder terms appearing in $\delta_{2,N}$ are $o_p(1)$ by arguments analogous to those for

$\phi_{2,N}$ and $\phi_{5,N}$ (see (C.43e) and (C.43f)). Finally, the difference between $\tilde{\sigma}_{\varepsilon,N}^2 =$

$\dfrac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\varepsilon}_{it,N}^2$ and $\sigma_\mu^2 + E\dfrac{1}{N}\sum_{i=1}^{N}\dfrac{1}{T}\sum_{t=1}^{T}\sigma_{v,it,N}^2$ converges in probability to zero. As a

consequence, $\tilde{\sigma}_{\mu,N}^{(4)} - \sigma_\mu^{(4)} = o_p(1)$, $\sigma_\mu^{(4)} = O(1)$ by Assumption 1, and $\tilde{\sigma}_{\mu,N}^{(4)} = O_p(1)$.

Lemma C.5

Suppose Assumptions 1-4 hold. Let $\Lambda_N = \dfrac{1}{NT}(\boldsymbol{\sigma}_{v,N}^2)'\mathbf{A}_N\boldsymbol{\sigma}_{v,N}^2$ with $\boldsymbol{\sigma}_N^2 = (\sigma_{v,1,N}^2,...,\sigma_{v,NT,N}^2)'$,

with $\mathbf{A}_N = (a_{it,js,N})$ real, nonstochastic, and symmetric $NT \times NT$ matrices, whose elements are time-invariant ($a_{it,js,N} = a_{i,j,N}$), whose diagonal elements are zero ($a_{i,j,N} = 0$ for $a_{i,j,N} = 0$), and whose row and column sums are bounded uniformly in absolute value. Let

$$\widetilde{\underline{\mathbf{v}}}_N = \mathbf{Q}_{0,N}\widetilde{\boldsymbol{\varepsilon}}_N = \mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})\tilde{\mathbf{u}}_N \qquad \text{with} \qquad \tilde{\mathbf{u}}_N = \mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N \qquad \text{and}$$

$\mathbf{D}_N = (\mathbf{d}'_{1,N},...,\mathbf{d}'_{NT,N})'$, and where the $S \times 1$ vector $\tilde{\boldsymbol{\rho}}_N$ can be any estimator that satisfies $(\tilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = o_p(1)$. Finally, define

$$\widetilde{\Lambda}_N^{HR} = \frac{T^2}{(T-1)^2}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}a_{i,j,N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{\underline{v}}_{it,N}^2\widetilde{\underline{v}}_{js,N}^2 .$$

Then, we have $\widetilde{\Lambda}_N^{HR} - \Lambda_N = o_p(1)$, $\Lambda_N = O(1)$, and $\widetilde{\Lambda}_N^{HR} = O_p(1)$.

Proof.

Note that

$$\Lambda_N = \frac{1}{N}E\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_{ij,N} , \text{ where} \tag{C.48a}$$

$$\Lambda_{ij,N} = a_{i,j,N}E\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it,N}^2 v_{js,N}^2 = a_{i,j,N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sigma_{v,it}^2\sigma_{v,js}^2 ,$$

since $a_{i,j,N} = 0$ for $i = j$ and $t = s$. The corresponding expression based on the fixed effects residuals is given by

$$\underline{\Lambda}_N = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\underline{\Lambda}_{ij,N} \text{ where } \underline{\Lambda}_{ij,N} = a_{i,j,N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{v}_{it,N}^2\underline{v}_{js,N}^2 . \tag{C.48b}$$

Since $\underline{v}_{it,N}$ and $\underline{v}_{js,N}$ are independent for all t,s,

$$E\underline{\Lambda}_{ij,N} = a_{i,j,N}\frac{1}{T}E\sum_{t=1}^{T}\underline{v}_{it,N}^2 E\sum_{s=1}^{T}\underline{v}_{js,N}^2 \tag{C.49}$$

$$= a_{i,j,N}\frac{1}{T}\frac{(T-1)^2}{T^2}E\sum_{t=1}^{T}v_{it,N}^2 E\sum_{s=1}^{T}v_{js,N}^2$$

$$= \frac{(T-1)^2}{T^2}\Lambda_{ij,N} ,$$

which suggests the following bias-corrected estimator:

$$\widetilde{\Lambda}_N^{HR} = \frac{T^2}{(T-1)^2}\widetilde{\underline{\Lambda}}_N , \text{ where} \tag{C.50}$$

$$\underline{\widetilde{\Lambda}}_N = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\underline{\widetilde{\Lambda}}_{ij,N} \text{ with } \underline{\widetilde{\Lambda}}_{ij,N} = a_{i,j,N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{\underline{v}}_{it,N}^2\widetilde{\underline{v}}_{js,N}^2 .$$

To show that $\left|\widetilde{\Lambda}_N^{HR} - \Lambda_N\right| = o_p(1)$, we next demonstrate that $\left|\dfrac{T^2}{(T-1)^2}\underline{\Lambda}_N - \Lambda_N\right| = o_p(1)$ and

$\left|\widetilde{\Lambda}_N^{HR} - \dfrac{T^2}{(T-1)^2}\underline{\Lambda}_N\right| = o_p(1)$.

Consider

$$\underline{\Lambda}_N = \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}a_{i,j,N}\sum_{t=1}^{T}\sum_{s=1}^{T}\underline{v}_{it,N}^2\underline{v}_{js,N}^2 \tag{C.51}$$

$$= \frac{1}{NT}(\underline{\mathbf{v}}_N^2)'\mathbf{A}_N\underline{\mathbf{v}}_N^2 ,$$

using $\underline{\mathbf{v}}_N^2 = (\underline{v}_{n,N}^2) = (\underline{v}_{it,N}^2)$ and $\mathbf{A}_N = (a_{n,n',N}) = (a_{it,js,N})$. Note that $a_{it,js,N} = a_{i,j,N}$ and that $a_{it,js,N} = 0$ for $it = js$. Next, define the $NT \times 1$ vector $\boldsymbol{\zeta}_N = (\underline{v}_{it,N}^2 - \sigma_{v,it,N}^2)$. By Assumption 1, $\sigma_{v,it,N}^2 = E(\underline{v}_{it,N}^2) \le K < \infty$ and the row and column sums of the variance-covariance matrix $\boldsymbol{\Xi}_N = E(\boldsymbol{\zeta}_N\boldsymbol{\zeta}_N')$ are bounded uniformly in absolute value. Next rewrite $\underline{\Lambda}_N = \dfrac{1}{NT}(\boldsymbol{\zeta}_N)'\mathbf{A}_N\boldsymbol{\zeta}_N + E\underline{\Lambda}_N + \dfrac{2}{NT}(\boldsymbol{\sigma}_{\underline{v}}^2)'\mathbf{A}_N\boldsymbol{\zeta}_N$, and note that

$$Var(\underline{\Lambda}_N) \le \{\frac{1}{(NT)^2}Var[(\boldsymbol{\zeta}_N)'\mathbf{A}_N\boldsymbol{\zeta}_N] + \frac{4}{(NT)^2}Var[(\boldsymbol{\sigma}_{\underline{v}}^2)'\mathbf{A}_N\boldsymbol{\zeta}_N]\}^{0.5} \tag{C.52}$$

$$= [\frac{2}{(NT)^2}Tr(\mathbf{A}_N\boldsymbol{\Xi}_N\mathbf{A}_N\boldsymbol{\Xi}_N) + \frac{4}{(NT)^2}(\boldsymbol{\sigma}_{\underline{v}}^2)'(\mathbf{A}_N\boldsymbol{\Xi}_N\mathbf{A}_N)\boldsymbol{\sigma}_{\underline{v}}^2]^{0.5} = o(1) ,$$

such that we have $\Lambda_N - \dfrac{T^2}{(T-1)^2}E\underline{\Lambda}_N = o_p(1)$ by Chebychev's inequality.

Next note that $\left|\widetilde{\Lambda}_N^{HR} - \dfrac{T^2}{(T-1)^2}\underline{\Lambda}_N\right| = \dfrac{T^2}{(T-1)^2}\left|\underline{\widetilde{\Lambda}}_N - \underline{\Lambda}_N\right|$ and consider

$$\left|\underline{\widetilde{\Lambda}}_N - \underline{\Lambda}_N\right| = \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}a_{i,j,N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{\underline{v}}_{it,N}^2\widetilde{\underline{v}}_{js,N}^2 - \underline{v}_{it,N}^2\underline{v}_{js,N}^2)\right| \tag{C.53}$$

$$\le |\vartheta_1| + |\vartheta_2| + |\vartheta_3| ,$$

with

$$|\vartheta_1| = \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}a_{i,j,N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{\underline{v}}_{it,N}^2 - \underline{v}_{it,N}^2)\underline{v}_{js,N}^2 ,$$

$$|\vartheta_2| = \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}a_{i,j,N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{\underline{v}}_{js,N}^2 - \underline{v}_{js,N}^2) ,$$

$$|\vartheta_3| = \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}a_{i,j,N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{\underline{v}}_{it,N}^2 - \underline{v}_{it,N}^2)(\widetilde{\underline{v}}_{js,N}^2 - \underline{v}_{js,N}^2) .$$

By Lemma C.1 we have $\left|\widetilde{v}_{it,N}^2 - \underline{v}_{it,N}^2\right| \le 2\alpha_N \underline{\beta}_{it,N}\left|\underline{v}_{it,N}\right| + \alpha_N^2 \underline{\beta}_{it,N}^2$. In light of the maintained assumptions regarding the properties of $\underline{v}_{it,N}^2$ and $\mathbf{A}_N$, it follows that $\left|\mathcal{G}_1\right| = o_p(1)$, $\left|\mathcal{G}_2\right| = o_p(1)$, $\left|\mathcal{G}_3\right| = o_p(1)$, and thus $\left|\widetilde{\underline{\Lambda}}_N - \underline{\Lambda}_N\right| = o_p(1)$.

Summing up $\left|\widetilde{\Lambda}_N^{HR} - \Lambda_N\right| = o_p(1)$. Finally, $\Lambda_N = \dfrac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T a_{it,js,N}\sigma_{v,it}^2\sigma_{v,js}^2 = O(1)$, such that $\widetilde{\Lambda}_N^{HR} = O_p(1)$, which completes the proof.

Lemma C.6a

Suppose Assumptions 1-4 hold; in addition, assume that $Ev_{it,N}^8 < \infty$ for all $1 \le t \le T$ and

$1 \le i \le N, N \ge 1$. Let $\varpi_N = \dfrac{1}{NT}E\sum_{i=1}^N c_{i,N}\sum_{t=1}^T v_{it,N}^4$, where the nonstochastic, time-invariant

scalars $c_{it,N} = c_{i,N}$ are bounded uniformly in absolute value. Let

$\widetilde{\underline{\mathbf{v}}}_N = \mathbf{Q}_{0,N}\widetilde{\underline{\boldsymbol{\varepsilon}}}_N = \mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^S \widetilde{\rho}_{m,N}\mathbf{M}_{m,N})\widetilde{\mathbf{u}}_N$ with $\widetilde{\mathbf{u}}_N = \mathbf{u}_N + \mathbf{D}_N\Delta_N$ and

$\mathbf{D}_N = (\mathbf{d}'_{1.,N},...,\mathbf{d}'_{NT.,N})'$, and where the $S\times 1$ vector $\widetilde{\boldsymbol{\rho}}_N$ can be any estimator that satisfies $(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = o_p(1)$. Finally, define

$$\widetilde{\varpi}_N^{HR} = \frac{k_0}{1 - k_1 m_1}\widetilde{\underline{\varpi}}_N + \frac{k_1 m_0}{1 - k_1 m_1}\widetilde{\underline{\mathbf{a}}}_N, \text{ where}$$

$$\widetilde{\underline{\varpi}}_N = \frac{1}{NT}\sum_{i=1}^N c_{i,N}\sum_{t=1}^T \widetilde{v}_{it,N}^4,$$

$$\widetilde{\underline{\mathbf{a}}}_N = \frac{1}{N}\sum_{i=1}^N c_{i,N}\frac{1}{T}\sum_{t=1}^T \widetilde{v}_{it,N}^2\frac{1}{T-1}\sum_{\substack{r=1\\r\neq t}}^T \widetilde{v}_{ir,N}^2,$$

$$m_0 = \frac{T^3}{T^3 - 2T^2 - 3T + 9}, \quad m_1 = -\frac{(2T-3)}{T^3 - 2T^2 - 3T + 9},$$

$$k_0 = \frac{T^2(T-1)}{T^3 - 4T^2 + 6T - 3}, \quad k_1 = -\frac{(T-1)(6T-9)}{T^3 - 4T^2 + 6T - 3}.$$

Then, we have $\widetilde{\varpi}_N^{HR} - \varpi_N = o_p(1)$, $\varpi_N = O(1)$, and $\widetilde{\varpi}_N^{HR} = O_p(1)$.

Proof.
Consider

$$\varpi_N = \frac{1}{N}E\sum_{i=1}^N \varpi_{i,N} \text{ with } \varpi_{i,N} = \frac{1}{T}c_{i,N}\sum_{t=1}^T v_{it,N}^4. \tag{C.54a}$$

The corresponding expression based on the fixed effects residuals is given by

$$\underline{\varpi}_N = \frac{1}{N} E \sum_{i=1}^{N} \underline{\varpi}_{i,N} \quad \text{with} \quad \underline{\varpi}_{i,N} = \frac{1}{T-1} c_{i,N} \sum_{t=1}^{T} \underline{v}_{it,N}^4 . \tag{C.54b}$$

Substituting for $\underline{v}_{it,N}^4$, simplifying (exploiting the independence of $v_{it,N}$ and $v_{js,N}$ for $i \neq j$ or $s \neq t$), and collecting terms, we obtain – for each i – that

$$\varpi_{i,N} = k_0 E \underline{\varpi}_{i,N} + k_1 \mathfrak{a}_{i,N} , \quad \text{where} \tag{C.55}$$

$$\mathfrak{a}_{i,N} = c_{i,N} E \frac{1}{T} \sum_{t=1}^{T} v_{it,N}^2 \frac{1}{T-1} \sum_{\substack{r=1 \\ r \neq t}}^{T} v_{ir,N}^2 ,$$

$$k_0 = \frac{T^2(T-1)}{T^3 - 4T^2 + 6T - 3} \quad \text{and} \quad k_1 = -\frac{(T-1)(6T-9)}{T^3 - 4T^2 + 6T - 3} .$$

Since the correction term $\mathfrak{a}_{i,N}$ is also based on original rather than demeaned residuals, another bias correction for is required. Analogous derivations yield the result that

$$\mathfrak{a}_{i,N} = m_0 E \underline{\mathfrak{a}}_{i,N} + m_1 \omega_{i,N} , \quad \text{with} \tag{C.56}$$

$$\underline{\mathfrak{a}}_{i,N} = c_{i,N} E \frac{1}{T} \sum_{t=1}^{T} \underline{v}_{it,N}^2 \frac{1}{T-1} \sum_{\substack{r=1 \\ r \neq t}}^{T} \underline{v}_{ir,N}^2 ,$$

$$m_0 = \frac{T^3}{T^3 - 2T^2 - 3T + 9} \quad \text{and} \quad m_1 = -\frac{(2T-3)}{T^3 - 2T^2 - 3T + 9} .$$

Substituting (C.56) into (C.55), averaging over $i = 1,...,N$ and solving for ϖ_N yields the following bias corrected estimator for ϖ_N:

$$\widetilde{\varpi}_N^{HR} = g_0 \widetilde{\underline{\varpi}}_N + g_1 \underline{\widetilde{\mathfrak{a}}}_N , \quad \text{where} \tag{C.57}$$

$$g_0 = \frac{k_0}{1 - k_1 m_1} \quad \text{and} \quad g_1 = \frac{k_1 m_0}{1 - k_1 m_1} .$$

We next show that $\widetilde{\varpi}_N^{HR} - \varpi_N = o_p(1)$, considering each summand in (C.57). By the weak law of large numbers,

$$\underline{\varpi}_N - E \underline{\varpi}_N = o_p(1) , \tag{C.58a}$$

given that $E|\underline{\varpi}_N|^2 < K < \infty$, since the 8-th moments of $v_{it,N}$ (and thus also those of $\underline{v}_{it,N}$) are finite. Using the triangle inequality and the results in Lemma C.1, we have

$$\left| \widetilde{\underline{\varpi}}_N - \underline{\varpi}_N \right| \leq \frac{1}{N} \sum_{i=1}^{N} \left| \widetilde{\underline{\varpi}}_{i,N} - \underline{\varpi}_{i,N} \right| \tag{C.58b}$$

$$\leq \frac{1}{NT} E \sum_{i=1}^{N} |c_{i,N}| \left| \sum_{t=1}^{T} \left| 4 \underline{v}_{it,N}^3 \underline{\eta}_{it,N} + 6 \underline{v}_{it,N}^2 \underline{\eta}_{it,N}^2 + 4 \underline{v}_{it,N} \underline{\eta}_{it,N}^3 + \underline{\eta}_{it,N}^4 \right| \right|$$

$$\leq \frac{1}{NT} \sum_{i=1}^{N} |c_{i,N}| \left(\sum_{t=1}^{T} (4\alpha_N |\underline{v}_{it,N}|^3 |\underline{\beta}_{it,N}| + 6\alpha_N^2 |\underline{v}_{it,N}|^2 \underline{\beta}_{it,N}^2 + 4\alpha_N^3 |\underline{v}_{it,N}| |\underline{\beta}_{it,N}|^3 + \alpha_N^4 \underline{\beta}_{it,N}^4 \right)$$

$$\leq \frac{1}{NT} \sum_{i=1}^{N} |c_{i,N}| \left(\sum_{k=1}^{4} \delta_{k,it} \right) ,$$

with $\quad \delta_{1,it} = 4\alpha_N \sum_{t=1}^{T}\left|\underline{v}_{it,N}\right|^3\left|\underline{\beta}_{it,N}\right|, \quad \delta_{2,it} = 6\alpha_N^2 \sum_{t=1}^{T}\underline{v}_{it,N}^2\underline{\beta}_{it,N}^2, \quad \delta_{3,it} = 4\alpha_N^3 \sum_{t=1}^{T}\left|\underline{v}_{it,N}\right|\left\|\underline{\beta}_{it,N}\right\|^3, \quad$ and

$\delta_{4,it} = \alpha_N^4 \sum_{r=1}^{T}\underline{\beta}_{it,N}^4$. It is readily verified that $\dfrac{1}{NT}\sum_{i=1}^{N}\left|c_{i,N}\right|\left|\sum_{t=1}^{T}\delta_{k,it}\right| = o_p(1)$ for $k = 1,\dots,4$ under

the maintained assumptions. As an example, consider the case $k = 1$. Using $\left|c_{i,N}\right| \le K < \infty$ for some $K < \infty$, the triangle inequality, and Hölder's inequality, we have

$$\frac{1}{NT^2}\sum_{i=1}^{N}\left|c_{i,N}\right|\left|\sum_{t=1}^{T}\delta_{1,it}\right| \le \frac{4K}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\underline{v}_{it,N}\right|^3\alpha_N\left|\underline{\beta}_{it,N}\right| \tag{C.58c}$$

$$\le 4KN^{-1/2}(N^{1/2}\alpha_N)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\underline{v}_{it,N}\right|^4\right)^{3/4}\left(\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\left|\underline{\beta}_{js,N}\right|^4\right)^{1/4} = o_p(1),$$

since $N^{1/2}\alpha_N = O_p(1)$, $\dfrac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\underline{v}_{it,N}\right|^4 = O_p(1)$, $\dfrac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\left|\underline{\beta}_{js,N}\right|^4 = O_p(1)$, and $N^{-1/2} = o(1)$.

It follows that $\left|\widetilde{\underline{\omega}}_N - \underline{\omega}_N\right| = o_p(1)$ and thus $\left|\widetilde{\underline{\omega}}_N - E\underline{\omega}_N\right| = o_p(1)$.

Next consider $\left|\widetilde{\underline{a}}_N - E\underline{a}_N\right|$. Again, under the maintained assumptions,

$$\underline{a}_N - E\underline{a}_N = o_p(1), \tag{C.59a}$$

and thus $\left|\underline{a}_N - E\underline{a}_N\right| = o_p(1)$ by the weak law of large numbers.

Using the triangle inequality and the results in Lemma C.1, we have

$$\left|\widetilde{\underline{a}}_N - \underline{a}_N\right| \le \frac{1}{N}\sum_{i=1}^{N}\left|\widetilde{\underline{a}}_{i,N} - \underline{a}_{i,N}\right| \tag{C.59b}$$

$$\le \frac{1}{NT(T-1)}\sum_{i=1}^{N}\left|c_{i,N}\right|\left|\sum_{t=1}^{T}\sum_{k=1}^{8}\left|\delta_{k,it}\right|\right|,$$

with $\quad \delta_{1,it} = 2\underline{v}_{it,N}\underline{\eta}_{it,N}\sum_{\substack{r=1\\r\ne t}}^{T}\underline{v}_{ir,N}^2, \qquad \delta_{2,it} = \underline{\eta}_{it,N}^2\sum_{\substack{r=1\\r\ne t}}^{T}\underline{v}_{ir,N}^2, \qquad \delta_{3,it} = \frac{1}{T-1}\underline{v}_{it,N}^2\sum_{\substack{r=1\\r\ne t}}^{T}2\underline{v}_{ir,N}\underline{\eta}_{ir,N},$

$\delta_{4,it} = \dfrac{2}{T-1}\underline{v}_{it,N}\underline{\eta}_{it,N}\sum_{\substack{r=1\\r\ne t}}^{T}\underline{v}_{ir,N}\underline{\eta}_{ir,N}, \; \delta_{5,it} = \dfrac{1}{T-1}\underline{\eta}_{it,N}^2\sum_{\substack{r=1\\r\ne t}}^{T}2\underline{v}_{ir,N}\underline{\eta}_{ir,N}, \; \delta_{6,it} = \dfrac{1}{T-1}\underline{v}_{it,N}^2\sum_{\substack{r=1\\r\ne t}}^{T}\underline{\eta}_{ir,N}^2,$

$\delta_{7,it} = \dfrac{2}{T-1}\underline{v}_{it,N}\underline{\eta}_{it,N}\sum_{\substack{r=1\\r\ne t}}^{T}\underline{\eta}_{ir,N}^2, \; \delta_{8,it} = \dfrac{1}{T-1}\underline{\eta}_{it,N}^2\sum_{\substack{r=1\\r\ne t}}^{T}\underline{\eta}_{ir,N}^2.$

Consider $\delta_{1,it}$. Substituting for $\underline{\eta}_{it,N} = \alpha_N\underline{\beta}_{it,N}$, using the triangle inequality and the generalized Hölder inequality, we obtain – for each of the $(T-1)$ terms with $r \ne t$

$$\frac{1}{NT(T-1)}\sum_{i=1}^{N}\left|c_{i,N}\right|\left|\sum_{t=1}^{T}\left|\delta_{1,it}\right|\right| \le \frac{2K}{N}\frac{1}{T}\frac{1}{(T-1)}\alpha_N\sum_{i=1}^{N}\left|\sum_{t=1}^{T}\underline{v}_{it,N}\underline{\beta}_{it,N}\underline{v}_{ir,N}^2\right| \tag{C.59c}$$

$$\leq \frac{2K}{(T-1)} N^{-1/2}(N^{1/2}\alpha_N)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\underline{v}_{it,N}\right|^2\right)^{1/2}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\left|\underline{\beta}_{is,N}\right|^4\right)^{1/4}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\underline{v}_{ir,N}\right|^4\right)^{1/4}=o_p(1),$$

since $\left|c_{i,N}\right|\leq K$ with $K=O(1)$, $N^{1/2}\alpha_N=O_p(1)$, $\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\underline{v}_{it,N}\right|^2\right)^{1/2}=O_p(1)$,

$\left(\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\left|\underline{\beta}_{js,N}\right|^4\right)^{1/4}=O_p(1)$, $\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\underline{v}_{it,N}\right|^4\right)^{1/4}=O_p(1)$, and $N^{-1/2}=o(1)$.

By analogous arguments, the other terms involving $\delta_{2,it}$ to $\delta_{8,it}$ can be shown to be $o_p(1)$ under the maintained assumptions. It follows that $\left|\tilde{\mathfrak{a}}_{1,N}-\mathfrak{a}_{1,N}\right|=o_p(1)$, and thus $\left|\tilde{\mathfrak{a}}_N-E\underline{\mathfrak{a}}_N\right|=o_p(1)$.

This completes the proof, recognizing that $\left|\varpi_N\right|=O(1)$ under the maintained assumptions.

Lemma C.6b

Suppose Assumptions 1-4 hold; assume further that $v_{it,N}\sim i.d.(0,\sigma_{v,i}^2)$, i.e., there is cross-sectional heteroskedasticity only in $v_{it,N}$ (but no heteroskedasticity over time). Let

$$\chi_N=\frac{1}{NT}\sum_{i=1}^{N}c_{i,N}\sum_{t=1}^{T}\sigma_i^4 \text{ and define}$$

$$\tilde{\chi}_N^{HR}=\frac{m_0}{1-m_1k_1}\tilde{\mathfrak{a}}_N+\frac{m_1k_0}{1-m_1k_1}\tilde{\varpi}_N,$$

where m_0,m_1,k_0,k_1 as well as $\tilde{\mathfrak{a}}_N$ and $\tilde{\varpi}_N$ are as in Lemma C.6a.

Then, we have $\tilde{\chi}_N^{HR}-\chi_N=o_p(1)$, $\chi_N=O(1)$, and $\tilde{\chi}_N^{HR}=O_p(1)$.

Proof.

Notice first that

$$\chi_N=\frac{1}{NT}\sum_{i=1}^{N}c_{i,N}\sum_{t=1}^{T}(Ev_{it,N}^2)^2=\frac{1}{N}\sum_{i=1}^{N}c_{i,N}\sigma_i^4. \tag{C.60b}$$

Under the maintained assumptions, this can be written equivalently in the following (estimable) expression:

$$\chi_N=\frac{1}{N}\sum_{i=1}^{N}\chi_{i,N}, \tag{C.60b}$$

where $\omega_{i,N}=E(\frac{c_i}{T(T-1)}\sum_{t=1}^{T}\sum_{\substack{s=1\\s\neq t}}^{T}v_{it,N}^2v_{is,N}^2)$.

Next, observe that χ_N is equal to $\mathfrak{a}_N$ as defined in the proof of Lemma C.6a. Substituting (C.55) into (C.56), solving for $\mathfrak{a}_N$, and averaging over $i=1,...,N$ the bias corrected estimator

given in Lemma C.6b is obtained. That $\left|\widetilde{\underline{\omega}}_N - E\underline{\omega}_N\right| = o_p(1)$ and $\left|\widetilde{\underline{a}}_N - E\underline{a}_N\right| = o_p(1)$ was already shown in the proof of Lemma C.6a.

Remark C.2

If $v_{it,N}$ is in fact heteroskedastic over both cross-sections and time, the error made by the approximation in Lemma C.6b is given by

$$\Delta = \frac{1}{N}\sum_{i=1}^{N}c_{i,N}\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\substack{s=1\\s\neq t}}^{T}\sigma_{it,N}^{2}\sigma_{is,N}^{2} - \sum_{i=1}^{N}c_{i,N}\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\substack{s=1\\s\neq t}}^{T}\sigma_{it,N}^{2}\sigma_{it,N}^{2}$$

$$= \frac{1}{N}\sum_{i=1}^{N}c_{i,N}\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\substack{s=1\\s\neq t}}^{T}\sigma_{it,N}^{2}(\sigma_{is,N}^{2}-\sigma_{it,N}^{2})\,.$$

Hence, Δ can be assumed to be small for small T and when heteroskedasticity is mainly of the cross-section type (or random over time).

Lemma C.7

Suppose the assumptions of Lemma C.6a hold. Let $\vartheta_{N} = \frac{1}{NT}E\sum_{i=1}^{N}\sum_{t=1}^{T}c_{it,N}v_{it,N}^{3}$, where the nonstochastic scalars $c_{it,N}$ are bounded uniformly in absolute value. Define

$$\widetilde{\vartheta}_{N}^{HR} = f_{0}\widetilde{\underline{\varpi}}_{N} + f_{1}\widetilde{\underline{b}}_{N}\,, \text{ where}$$

$$\widetilde{\underline{\vartheta}}_{N} = \frac{1}{N(T-1)}E\sum_{i=1}^{N}\sum_{t=1}^{T}c_{it,N}\underline{\widetilde{v}}_{it,N}^{3}\,,$$

$$\widetilde{\underline{b}}_{N} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}c_{it,N}\frac{1}{T-1}\sum_{r=1}^{T}\underline{\widetilde{v}}_{ir,N}^{3}\,,$$

$$f_{0} = \frac{T(T-1)}{T^{2}-3T+3} \text{ and } f_{1} = \frac{1}{(T^{2}-3T+3)}\frac{T^{2}}{(T-1)^{2}}\,.$$

Then, we have $\widetilde{\vartheta}_{N}^{HR} - \vartheta_{N} = o_{p}(1)$, $\vartheta_{N} = O(1)$, and $\widetilde{\vartheta}_{N}^{HR} = O_{p}(1)$.

Proof.
Consider

$$\vartheta_{N} = \frac{1}{N}E\sum_{i=1}^{N}\varpi_{i,N} \text{ with } \vartheta_{i,N} = \frac{1}{T}\sum_{t=1}^{T}c_{it,N}v_{it,N}^{3}\,, \tag{C.61a}$$

The corresponding expression based on the fixed effects residuals is given by

$$\underline{\vartheta}_{N} = \frac{1}{N}E\sum_{i=1}^{N}\underline{\vartheta}_{i,N} \text{ with } \underline{\vartheta}_{i,N} = \frac{1}{T-1}\sum_{t=1}^{T}c_{it,N}\underline{v}_{it,N}^{3}\,. \tag{C.61b}$$

Substituting for $\underline{v}_{it,N}^{3}$, simplifying (exploiting the independence of $v_{it,N}$ and $v_{js,N}$ for $i \neq j$ or $s \neq t$), and rearranging terms, we obtain that – for each i

$$\vartheta_{i,N} = f_{0}E\underline{\vartheta}_{i,N} + \dot{f}_{1}b_{i,N}\,, \text{ where} \tag{C.62}$$

$$f_{0} = \frac{T(T-1)}{T^{2}-3T+3} \text{ and } \dot{f}_{1} = \frac{1}{T^{2}-3T+3}\,.$$

Since the correction term $\mathfrak{b}_{i,N}$ is also based on original rather than demeaned residuals, another bias correction is required as well. Analogous derivations yield the result that

$$\mathfrak{b}_{i,N} = \frac{T^2}{(T-1)^2} E\underline{\mathfrak{b}}_{i,N}, \tag{C.63}$$

such that

$$\vartheta_{i,N} = f_0 E\underline{\vartheta}_{i,N} + f_1 E\underline{\mathfrak{b}}_{i,N}, \text{ where}$$

$$f_0 = \frac{1}{f_0'} = \frac{T(T-1)}{T^2 - 3T + 3},$$

$$f_1 = \dot{f}_1 \frac{T^2}{(T-1)^2} = \frac{1}{(T^2 - 3T + 3)} \frac{T^2}{(T-1)^2}$$

$$= \frac{T(T-1)}{T^2 - 3T + 3} \frac{T}{(T-1)^3} = f_0 \frac{T}{(T-1)^3}.$$

Averaging over $i = 1,...,N$, we obtain the following bias corrected estimator for ϖ_N:

$$\tilde{\underline{\vartheta}}_N^{HR} = f_0 \tilde{\underline{\vartheta}}_N + f_1 \tilde{\underline{\mathfrak{b}}}_N, \text{ where} \tag{C.64}$$

$$\tilde{\underline{\vartheta}}_N = \frac{1}{N(T-1)} E \sum_{i=1}^{N} \sum_{t=1}^{T} c_{it,N} \tilde{\underline{v}}_{it,N}^3 = \frac{1}{N} E \sum_{i=1}^{N} \tilde{\underline{\vartheta}}_{i,N}, \text{ and}$$

$$\tilde{\underline{\mathfrak{b}}}_N = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} c_{it,N} \frac{1}{T-1} \sum_{r=1}^{T} \tilde{\underline{v}}_{ir,N}^3.$$

The proof that $\tilde{\varpi}_N^{HR} - \varpi_N = o_p(1)$ is very similar to that in Lemma C.6a and is thus omitted for the sake of brevity. Finally, suppose that ω_N can be written as quadratic form $\vartheta_N = \frac{1}{NT} \mathbf{a}_N' \mathbf{\Sigma}_N^{(3)} \mathbf{b}_N$ with $\mathbf{\Sigma}_N^{(3)} = diag_{n=1}^{NT} E(v_{n,N}^3) = diag_{it=1}^{NT} E(v_{it,N}^3)$ and $c_{it,N} = a_{it,N} b_{it,N}$; then

$$\tilde{\vartheta}_N^{HR} = \frac{1}{N(T-1)} \mathbf{a}_N' \tilde{\mathbf{\Sigma}}_N^{(3),HR} \mathbf{b}_N, \quad \text{and} \quad \tilde{\mathbf{\Sigma}}_N^{(3),HR} = diag_{n=1}^{NT} [(\tilde{v}_{n,N}^{HR})^3] \quad \text{with}$$

$$(\tilde{v}_{it,N}^{HR})^3 = f_0 [\tilde{\underline{v}}_{it,N}^3 + \frac{1}{(T-1)^3} \sum_{r=1}^{T} \tilde{\underline{v}}_{ir,N}^3].$$

Remark C.3

Note that $\mathfrak{E}_{\mathbf{v},N}^{****} = N^{-1} \sum_{n=1}^{NT} d_{\mathbf{v},n,N} \sigma_{v,n,N}^{(3)}$ and $\tilde{\mathfrak{E}}_{\mathbf{v},N}^{****} = N^{-1} \sum_{n=1}^{NT} \tilde{d}_{\mathbf{v},n,N} \sigma_{v,n,N}^{(3)}$. Accounting for the definition of $d_{\mathbf{v},n,N}$, $\mathfrak{E}_{\mathbf{v},N}^{****}$ can be written as sum of the two expressions $\mathbf{a}_{\mathbf{v},N}' \mathbf{\Sigma}_N^{(3)} \mathbf{\kappa}_{\mathbf{B}_{\mathbf{v},N}}$ and $\mathbf{\kappa}_{\mathbf{A}_{\mathbf{v},N}}' \mathbf{\Sigma}_N^{(3)} \mathbf{b}_{\mathbf{v},N}$, where $\mathbf{\kappa}_{\mathbf{B}_{\mathbf{v},N}}$ ($\mathbf{\kappa}_{\mathbf{A}_{\mathbf{v},N}}$) is an $NT \times 1$ vector made up of the main diagonal elements of the matrix $\mathbf{B}_{\mathbf{v},N}$ ($\mathbf{A}_{\mathbf{v},N}$). Next, observe that $\mathbf{a}_{\mathbf{v},N} = \mathbf{F}_{\mathbf{v},N} \mathbf{P}_N \mathbf{\alpha}_N$ and $\tilde{\mathbf{a}}_{\mathbf{v},N} = \tilde{\mathbf{F}}_{\mathbf{v},N} \tilde{\mathbf{P}}_N \tilde{\mathbf{\alpha}}_N$ (obviously suppressing the indexation of $\tilde{\mathbf{\alpha}}_N$). By assumption $\tilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$, $\mathbf{P}_N = O(1)$ and thus $\tilde{\mathbf{P}}_N = O_p(1)$, where the dimension of $\mathbf{P}_N$ is $P_* \times P$. Moreover, $\tilde{\mathbf{\alpha}}_N - \mathbf{\alpha}_N = o_p(1)$, $\mathbf{\alpha}_N = O(1)$ and thus $\tilde{\mathbf{\alpha}}_N = O_p(1)$, where the dimension of $\mathbf{\alpha}_N$ is $P \times 1$. By arguments, analogous to that in Lemma C.3, we have $N^{-1} \tilde{\mathbf{F}}_N' \tilde{\mathbf{\Sigma}}_N^{(3),HR} \mathbf{\kappa}_{\mathbf{B}_{\mathbf{v},N}} - N^{-1} \mathbf{F}_N' \mathbf{\Sigma}_N^{(3),HR} \mathbf{\kappa}_{\mathbf{B}_{\mathbf{v},N}} = o_p(1)$. It

follows that $N^{-1}\widetilde{\boldsymbol{\alpha}}'_N \widetilde{\mathbf{P}}'_N \widetilde{\mathbf{F}}'_N \widetilde{\boldsymbol{\Sigma}}_N^{(3),HR} \boldsymbol{\kappa}_{\mathbf{B}_{\mathbf{v},N}} - N^{-1} \boldsymbol{\alpha}'_{p,r,N} \mathbf{P}'_N \mathbf{F}'_N \boldsymbol{\Sigma}_N^{(3)} \boldsymbol{\kappa}_{\mathbf{B}_{\mathbf{v},N}} = o_p(1)$. By the same reasoning, $\boldsymbol{\kappa}'_{\mathbf{A}_{\mathbf{v},N}} \widetilde{\boldsymbol{\Sigma}}_N^{(3)} \widetilde{\mathbf{b}}_{\mathbf{v},N} - \boldsymbol{\kappa}'_{\mathbf{A}_{\mathbf{v},N}} \boldsymbol{\Sigma}_N^{(3)} \mathbf{b}_{\mathbf{v},N} = o_p(1)$, from which it follows that $\widetilde{\mathfrak{E}}_{\mathbf{v},N}^{****} - \mathfrak{E}_{\mathbf{v},N}^{****} = o_p(1)$.