Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Parameter estimation and inference with spatial lags and cointegration

Jan Mutl & Leopold Sögner

To cite this article: Jan Mutl & Leopold Sögner (2019) Parameter estimation and inference with spatial lags and cointegration, Econometric Reviews, 38:6, 597-635, DOI: 10.1080/07474938.2017.1382803

To link to this article: https://doi.org/10.1080/07474938.2017.1382803

ECONOMETRIC REVIEWS
2019, VOL. 38, NO. 6, 597–635
https://doi.org/10.1080/07474938.2017.1382803

Parameter estimation and inference with spatial lags and cointegration

Jan Mutl[a] and Leopold Sögner[b]

[a]EBS Real Estate Management Institute, EBS Business School, Germany; [b]Institute for Advanced Studies and Vienna Graduate School of Finance, Austria

ABSTRACT

This article studies dynamic panel data models in which the long run outcome for a particular cross-section is affected by a weighted average of the outcomes in the other cross-sections. We show that imposing such a structure implies a model with several cointegrating relationships that, unlike in the standard case, are nonlinear in the coefficients to be estimated. Assuming that the weights are exogenously given, we extend the dynamic ordinary least squares methodology and provide a dynamic two-stage least squares estimator. We derive the large sample properties of our proposed estimator under a set of low-level assumptions. Then our methodology is applied to US financial market data, which consist of *credit default* swap spreads, as well as firm-specific and industry data. We construct the economic space using a "closeness" measure for firms based on input–output matrices. Our estimates show that this particular form of spatial correlation of credit default swap spreads is substantial and highly significant.

KEYWORDS

Cointegration; credit risk; dynamic ordinary least squares; spatial autocorrelation

JEL CLASSIFICATION

C31; C33; G13

1. Introduction

This article investigates the estimation of nonstandard cointegrating relationships under the presence of regressor endogeneity and serial correlation in the disturbances. Following literature on panel cointegration, we augment the cointegrating vectors by peer or neighborhood effects, which are modeled as spatial lags following Cliff and Ord (1973). In addition to the kind of endogeneity typically dealt with in panel cointegration models (Mark and Sul, 2003; Mark et al., 2005; Baltagi, 2008; Pesaran, 2015), this article shows that the spatial lag results in *further regressor endogeneity* of a different type.

Several approaches have emerged in the literature to estimate cointegrating relationships and to perform statistical inference. One possibility is to use a simple estimation routine, e.g., *ordinary least squares* (*OLS*) and then work out the (sometimes complicated) large sample distribution of the estimated parameters (Phillips and Hansen, 1990; Phillips and Loretan, 1991). Another opportunity is to adjust the estimation routine, such that the large sample distribution is either simpler or free of nuisance parameters. Examples along these lines are the *fully modified least squares* estimator (see, e.g., Phillips and Hansen, 1990; Phillips and Moon, 1999; Pedroni, 2000), the *integrated modified least squares* estimator (Vogelsang and Wagner, 2014, in which integrated modified least squares estimation is linked to fixed-b inference) and the *dynamic least squares* approach. Dynamic least squares estimation includes time-series leads and lags of the first differences of the regressors to control for the serial correlation and regressor endogeneity. This kind of estimator has been proposed by Phillips and Loretan (1991), Saikkonen (1991), and Stock and Watson (1993). It has been applied to panel data, e.g., in

CONTACT Leopold Sögner soegner@ihs.ac.at Institute for Advanced Studies and Vienna Graduate School of Finance, Josefstädter Straße 39, 1080 Vienna, Austria.

Supplemental data for this article can be accessed on the publisher's website.

Kao and Chiang (2000), Mark and Sul (2003) and Mark et al. (2005), in which a seemingly unrelated regression (SUR) type model is considered (for an overview on panel cointegration, see, e.g., Pesaran, 2015, Chapter 31). A further alternative is provided by the autoregressive distributed lag approach, in which cointegrating relationships including lagged variables can be investigated (the reader is referred to Pesaran and Shin, 1995; Binder et al., 2005; Chudik and Pesaran, 2013a,b). Almost recently, spatial correlation and cointegration have been investigated in Yu et al. (2008) and Sögner and Wagner (2017). While Yu et al. (2008) considered maximum-likelihood estimation in a spatial dynamic panel model, Sögner and Wagner (2017) and Sögner et al. (2017) investigate *fully modified* as well as *integrated modified OLS* estimation in a model close to the model presented in this article.

This article develops an econometric tool suitable for investigating situations, in which the long run outcome for a particular cross-section cannot be assumed to be independent of the outcomes of the other cross-sections and, at the same time, autocorrelation of the disturbances and regressor endogeneity are present. To adequately cope with the endogeneity arising from the inclusion of the spatial lags, we propose using a *dynamic two-stage least squares* (*D2SLS*) estimator, which combines *dynamic least squares* (*DOLS*) and *two stage least squares* (*2SLS*) estimation. Section 2 describes the model assumptions and Section 3 obtains the large sample distribution of our estimator and shows how to correctly conduct inference. The finite sample properties are investigated by a Monte Carlo study in Section 4.

Finally, Section 5 applies our methodology to a financial dataset, in which we construct the economic space using a "closeness" measure for firms based on input–output matrices. The weights matrix obtained from input–output data should approximate possible correlation patterns due to technology and demand shocks working their way through the economy. We find that our particular form of cross-sectional spillovers is substantial and highly significant.

2. The model and assumptions

Suppose that the data are generated from[1]

$$y_{it} = \rho \sum_{j=1}^{n} W_{ij} y_{jt} + \boldsymbol{\beta}_I' \mathbf{x}_{Iit} + \boldsymbol{\beta}_C' \mathbf{x}_{Ct} + \boldsymbol{\beta}_L' \mathbf{x}_{Li} + \alpha_i + \lambda_t + u_{it}^{\dagger} = \rho y_{it}^{*} + \boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\beta}_L' \mathbf{x}_{Li} + \alpha_i + \lambda_t + u_{it}^{\dagger}, \quad (1)$$

where y_{it} is the scalar response random variable and $\mathbf{x}_{it}$ is a $k \times 1$ column vector of prediction random variables. The vector $\mathbf{x}_{it}$ is integrated of order one ($I(1)$) and permitted to consist of *individual specific regressors* $\mathbf{x}_{Iit}$ of dimension $k_I \geq 1$, and *cross-sectionally common regressors* $\mathbf{x}_{Ct}$ of dimension $k_C \geq 0$. The regressors $\mathbf{x}_{Lt}$, of dimension $k_L \geq 0$, are time invariant. Hence, $\boldsymbol{\beta} := \left(\boldsymbol{\beta}_I', \boldsymbol{\beta}_C'\right)' \in \mathbb{R}^k$, with $\boldsymbol{\beta}_I \in \mathbb{R}^{k_I}$, $\boldsymbol{\beta}_C \in \mathbb{R}^{k_C}$ and $k = k_I + k_C$. In addition, $\mathbf{x}_{it} := \left(\mathbf{x}_{Iit}', \mathbf{x}_{Ct}'\right)' \in \mathbb{R}^k$. The time index is $t = 1, \ldots, T$, while $i = 1, \ldots, n$ is the cross-sectional index. The individual and time effects, α_i, $i = 1, \ldots, n$, and λ_t, $t = 1, \ldots, T$, are (as in fixed effects model) allowed to be correlated with $\mathbf{x}_{it}$ and $\mathbf{x}_{Li}$. The term $y_{it}^{*} := \sum_{j=1}^{n} W_{ij} y_{jt}$ is referred to as a *spatial lag* and represents the contemporaneous impact of the neighboring observations on y_{it} (see, e.g., Cliff and Ord, 1973; Kelejian and Prucha, 1998, 1999; Kapoor et al., 2007).

[1] In this article, the following notation will be applied: For vectors and matrices we use boldface. If not otherwise stated, the vectors considered are column vectors. Given a $r_M \times c_M$ matrix $\mathbf{M}$, the term $\mathbf{M}_{(r_a:r_b, c_a:c_b)} = [\mathbf{M}]_{(r_a:r_b, c_a:c_b)}$ stands for "from row r_a to row r_b and from column c_a to column c_b of matrix $\mathbf{M}$". The i, j element of $\mathbf{M}$ is $[\mathbf{M}]_{(i,j)}$ or in shorter notation M_{ij}. $\mathbf{0}_{(a \times b)}$ and $\mathbf{1}_{(a \times b)}$ stands for $a \times b$ matrix of zeros and ones; $\mathbf{0}_{(a)}$ and is used to abbreviate $\mathbf{0}_{(a \times 1)}$; $\mathbf{I}_a$ is the $a \times a$ identity matrix, while $\mathbb{I}_{(\cdot)}$ stands for an indicator function. Given a vector $\mathbf{x} \in \mathbb{R}^n$, $diag(\mathbf{x})$ transforms $\mathbf{x}$ into a $n \times n$ diagonal matrix, while for $\mathbf{x}_i \in \mathbb{R}^{\ell \times k}$ $(i = 1, \ldots, n)$, $diag(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ yields a $n\ell \times nk$ block diagonal matrix. $\otimes$ denotes the Kronecker product. 2 E-3 stands for $2 \cdot 10^{-3} = 0.002$. $[Tr]$ denotes the integer part of Tr, $\Rightarrow$ stands for *weak convergence* (see, e.g., Klenke, 2008, Chapter 13), while $\overset{P}{\to}$ and $\overset{a.s.}{\to}$ denote convergence in probability and almost sure convergence (see, e.g., Klenke, 2008, Chapter 6). $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{a}, \mathbf{A})$ denotes that $\boldsymbol{\nu}$ is a normally distributed vector with mean $\mathbf{a}$ and covariance $\mathbf{A}$, while χ_n^2 denotes a χ^2 random variable with n degrees of freedom. Variables where the within-transform described in (9), (12), or (14) is applied, are—if not otherwise stated—denoted by the superscripts $\tilde{}$, $\grave{}$, and $\breve{}$, respectively. $\Delta \mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$ abbreviates the first difference of $\mathbf{x}_t$.

We collect the weights W_{ij} into an $n \times n$ spatial weights matrix $\mathbf{W}$ and follow the spatial econometrics literature by maintaining the following assumption regarding the cross-sectional (or spatial) structure of the model[2]:

Assumption 1 (Spatial Lag). *The spatial weights W_{ij} are nonstochastic and observable with $W_{ii} = 0$ and $\mathbf{W} \neq \mathbf{0}_{(n\times n)}$. The parameter ρ is such that the largest absolute eigenvalue of $\rho\mathbf{W}$ is smaller than one.*

The restriction that $W_{ii} = 0$ is a normalization of the model, which requires that no observation is its own neighbor. The second part of the assumption guarantees that the matrix $(\mathbf{I}_n - \rho\mathbf{W})$ is invertible (see, e.g., Horn and Johnson, 1985, Corollary 5.6.16). That is, the inverse $\mathbf{K} := (\mathbf{I}_n - \rho\mathbf{W})^{-1}$ is well-defined.[3] The invertibility of the matrix $(\mathbf{I}_n - \rho\mathbf{W})$ is needed to provide a unique solution of the model and rule out multiple solutions for y_{it} that would be consistent with the explanatory variables and disturbances.

To obtain the prediction variables, let $\mathbf{x}_{It} := \left(\left(\mathbf{x}'_{Iit}\right)_{i=1,\dots,n}\right)' \in \mathbb{R}^{k_I\cdot n}$ and $\mathbf{x}_t := \left(\mathbf{x}'_{It}, \mathbf{x}'_{Ct}\right)' \in \mathbb{R}^{k_I n + k_C}$. We assume that $\mathbf{x}_t$ follows a vector random walk process, i.e., $\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t$, where $\mathbf{v}_{It} := \Delta\mathbf{x}_{It}$, $\mathbf{v}_{Ct} := \Delta\mathbf{x}_{Ct}$ and $\mathbf{v}_t := \left(\mathbf{v}'_{It}, \mathbf{v}'_{Ct}\right)'$. In addition, we define $\mathbf{v}_{Iit} := \Delta\mathbf{x}_{Iit}$ and $\mathbf{v}_{it} := \left(\mathbf{v}'_{Iit}, \mathbf{v}'_{Ct}\right)' \in \mathbb{R}^{k}$. Together with the noise terms $\mathbf{u}^\dagger_t := \left(u^\dagger_{1t}, \dots, u^\dagger_{nt}\right)' \in \mathbb{R}^n$, we define $\boldsymbol{\eta}^\dagger_{Iit} := \left(u^\dagger_{it}, \mathbf{v}'_{Iit}\right)' \in \mathbb{R}^{k_I+1}$, $\boldsymbol{\eta}^\dagger_{It} := \left(\mathbf{u}^{\dagger\prime}_t, \mathbf{v}'_{It}\right)' \in \mathbb{R}^{(k_I+1)\cdot n}$ and $\boldsymbol{\eta}^\dagger_t := \left(\mathbf{u}^{\dagger\prime}_t, \mathbf{v}'_{It}, \mathbf{v}'_{Ct}\right)' \in \mathbb{R}^{(k_I+1)\cdot n + k_C}$. The (auto-) *covariance* matrices of $\boldsymbol{\eta}^\dagger_t$, abbreviated by $\boldsymbol{\Gamma}^\dagger_\ell \in \mathbb{R}^{((k_I+1)\cdot n + k_C)\times((k_I+1)\cdot n+k_C)}$, for any lag $\ell \in \mathbb{Z}$, are

$$\boldsymbol{\Gamma}^\dagger_\ell := \mathbb{E}\left(\boldsymbol{\eta}^\dagger_{t-\ell}\boldsymbol{\eta}^{\dagger\prime}_t\right) = \begin{pmatrix} \boldsymbol{\Gamma}^\dagger_{\ell,uu} & \boldsymbol{\Gamma}^\dagger_{\ell,uv} \\ \boldsymbol{\Gamma}^\dagger_{\ell,vu} & \boldsymbol{\Gamma}_{\ell,vv} \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(\mathbf{u}^\dagger_{t-\ell}\mathbf{u}^{\dagger\prime}_t\right) & \mathbb{E}\left(\mathbf{u}^\dagger_{t-\ell}\mathbf{v}'_t\right) \\ \mathbb{E}\left(\mathbf{v}_{t-\ell}\mathbf{u}^{\dagger\prime}_t\right) & \mathbb{E}\left(\mathbf{v}_{t-\ell}\mathbf{v}'_t\right) \end{pmatrix}, \quad \text{where}$$
$$\boldsymbol{\Gamma}^\dagger_{\ell,uu} := \mathbb{E}\left(\mathbf{u}^\dagger_{t-\ell}\mathbf{u}^{\dagger\prime}_t\right) = \operatorname{diag}\left(\Gamma^\dagger_{\ell,u_1u_1}, \dots, \Gamma^\dagger_{\ell,u_nu_n}\right) \in \mathbb{R}^{n\times n}. \tag{2}$$

Let $\boldsymbol{\Lambda}^\dagger := \sum_{\ell=1}^{\infty}\mathbb{E}\left(\boldsymbol{\eta}^\dagger_{t-\ell}\boldsymbol{\eta}^{\dagger\prime}_t\right) \in \mathbb{R}^{((k_I+1)\cdot n+k_C)\times((k_I+1)\cdot n+k_C)}$. Then, the *long run covariance matrix* $\boldsymbol{\Omega}^\dagger \in \mathbb{R}^{((k_I+1)\cdot n+k_C)\times((k_I+1)\cdot n+k_C)}$ and the *half-long run covariance matrix* $\boldsymbol{\Delta}^\dagger \in \mathbb{R}^{((k_I+1)\cdot n+k_C)\times((k_I+1)\cdot n+k_C)}$ of $\boldsymbol{\eta}^\dagger_t$ are given by (see, e.g., Phillips and Hansen, 1990)

$$\boldsymbol{\Omega}^\dagger := \sum_{\ell=-\infty}^{\infty}\mathbb{E}\left(\boldsymbol{\eta}^\dagger_{t-\ell}\boldsymbol{\eta}^{\dagger\prime}_t\right) = \boldsymbol{\Gamma}^\dagger_0 + \boldsymbol{\Lambda}^\dagger + \boldsymbol{\Lambda}^{\dagger\prime} \quad\text{and}\quad \boldsymbol{\Delta}^\dagger := \sum_{\ell=0}^{\infty}\mathbb{E}\left(\boldsymbol{\eta}^\dagger_{t-\ell}\boldsymbol{\eta}^{\dagger\prime}_t\right) = \boldsymbol{\Gamma}^\dagger_0 + \boldsymbol{\Lambda}^\dagger. \tag{3}$$

$\boldsymbol{\Omega}^\dagger$ contains the submatrices: $\boldsymbol{\Omega}_{vv} := \sum_{\ell=-\infty}^{\infty}\mathbb{E}\left(\mathbf{v}^\dagger_{t-\ell}\mathbf{v}^{\dagger\prime}_t\right) \in \mathbb{R}^{(k_I\cdot n+k_C)\times(k_I\cdot n+k_C)}$, $\boldsymbol{\Omega}_{v_iv_j} := \sum_{\ell=-\infty}^{\infty}\mathbb{E}\left(\mathbf{v}^\dagger_{i,t-\ell}\mathbf{v}^{\dagger\prime}_{j,t}\right) \in \mathbb{R}^{(k_I+k_C)\times(k_I+k_C)}$, and $\boldsymbol{\Omega}^\dagger_{uu} := \sum_{\ell=-\infty}^{\infty}\mathbb{E}\left(\mathbf{u}^\dagger_{t-\ell}\mathbf{u}^{\dagger\prime}_t\right) = \operatorname{diag}\left(\Omega^\dagger_{u_1u_1}, \dots, \Omega^\dagger_{u_nu_n}\right) \in \mathbb{R}^{n\times n}$. The same notation is applied to $\boldsymbol{\Delta}^\dagger$. In order to fully specify the model, we augment our set of assumptions by

Assumption 2 (Error Dynamics I). *$\boldsymbol{\eta}^\dagger_{Iit}$ and $\boldsymbol{\eta}^\dagger_{Ijt}$ are independent for all $i,j = 1,\dots,n$ with $i \neq j$. The common components $\mathbf{v}_{Ct} \in \mathbb{R}^{k_C}$ are permitted to be correlated with $\boldsymbol{\eta}^\dagger_{Iit}$. The stochastic process $\left(\boldsymbol{\eta}^\dagger_t\right)_{t\in\mathbb{Z}}$ is weakly stationary and obeys a functional central limit theorem. That is,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\boldsymbol{\eta}^\dagger_t \Rightarrow \mathcal{B}^\dagger(r) = \left(\boldsymbol{\Omega}^\dagger\right)^{1/2}\mathcal{W}^\dagger(r), \tag{4}$$

where $r \in [0,1]$ and $\mathcal{B}^\dagger(r)$ is a Brownian motion in $\mathbb{R}^{(k_I+1)\cdot n+k_C}$, while $\mathcal{W}^\dagger(r)$ is a standard Brownian motion in $\mathbb{R}^{(k_I+1)\cdot n+k_C}$. The long run covariance matrix $\boldsymbol{\Omega}^\dagger \in \mathbb{R}^{((k_I+1)\cdot n+k_C)\times((k_I+1)\cdot n+k_C)}$ is finite and of

[2]Throughout the analysis we only consider one spatial lag term. However, the theory considered in this article can also be applied to a model where $y_{it} = \rho_1\sum_{j=1}^{n}W_{1,ij}y_{jt} + \dots + \rho_{k_\rho}\sum_{j=1}^{n}W_{k_\rho,ij}y_{jt} + \beta'\mathbf{x}_{it} + \alpha_i + u^\dagger_{it}$ in a straightforward way. The restriction that only one matrix $\mathbf{W}$ is included is used to keep the notation simple.

[3]The spectral radius is the lower bound for every induced matrix norm (cf. Theorem 5.6.9 in Horn and Johnson, 1985). Our assumption will, for example, be satisfied when the maximum absolute row or column sums of $\rho\mathbf{W}$ are less than one.

full rank; $0 < \mathbf{\Omega}^\dagger < \infty$ *in short form. Moreover,*

$$\frac{1}{T}\sum_{t=2}^{T}\left(\sum_{j=1}^{t-1}\mathbf{v}_j\right)u_t^\dagger \Rightarrow \int_0^1 \mathcal{B}_v(s)d\mathcal{B}_u^\dagger(s)' + \mathbf{\Lambda}_{vu}^\dagger\,. \tag{5}$$

In addition, for $\boldsymbol{\eta}_t^\dagger\boldsymbol{\eta}_t^{\dagger\prime}$ *a weak law of large numbers holds, i.e.* $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t^\dagger\boldsymbol{\eta}_t^{\dagger\prime} \stackrel{P}{\rightarrow} \mathbb{E}\left(\boldsymbol{\eta}_t^\dagger\boldsymbol{\eta}_t^{\dagger\prime}\right)$*. To* $\mathbf{v}_t u_{it}^\dagger$ *a central limit theorem can be applied. That is,* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{v}_t u_{it}^\dagger - \mathbb{E}\left(\mathbf{v}_t u_{it}^\dagger\right)\right) \Rightarrow \boldsymbol{\nu}_{(vu^\dagger)}$*, where* $\boldsymbol{\nu}_{(vu^\dagger)} \sim \mathcal{N}\left(\mathbf{0}_{(k_I n+k_C\times 1)}, \mathbf{D}_{\left(\mathbf{v}_t u_{it}^\dagger\right)}\right)$ *and* $0 < \mathbf{D}_{\left(\mathbf{v}_t u_{it}^\dagger\right)} < \infty$*. Convergence in (4), (5) and* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{v}_t u_{it}^\dagger - \mathbb{E}\left(\mathbf{v}_t u_{it}^\dagger\right)\right)$ *is joint.*

The parameter vector $\boldsymbol{\beta} \neq \mathbf{0}_{(k\times 1)}$ *and* $k \geq k_I \geq 1$*. If* $k_L > 0$ *the time invariant regressors* $\mathbf{x}_{Li} \in \mathbb{R}^{k_L}$ *are independent of* $\mathbf{x}_{js}$ *as well as* $u_{js}^\dagger$ *for all* $i,j = 1,\ldots,n$ *and all* $t = 1,\ldots,T$*.*

The Brownian motion $\mathcal{B}^\dagger(r) \in \mathbb{R}^{(k_I+1)\cdot n+k_c}$ consists of the pairwise independent Brownian motions $\mathcal{B}_{vI_i}(r)$, $\mathcal{B}_{vI_j}(r) \in \mathbb{R}^{k_I}$, for $i,j = 1,\ldots,n$, the Brownian motion $\mathcal{B}_{vC}(r) \in \mathbb{R}^{k_C}$ arising from the common component $\mathbf{v}_{Ct}$, and the Brownian motions $\mathcal{B}_{u_i}^\dagger(r) \in \mathbb{R}$ arising from the common components $u_{it}^\dagger$. Then, $\mathcal{B}_v(r) := \left(\mathcal{B}_{vI_1}(r)',\ldots,\mathcal{B}_{vI_n}(r)',\mathcal{B}_{vC}(r)'\right)' \in \mathbb{R}^{k_I\cdot n+k_C}$. The same notation is applied to $\mathcal{W}^\dagger(r)$. In addition, $\mathcal{B}_{v_i}(r) := \left(\mathcal{B}_{vI_i}(r)',\mathcal{B}_{vC}(r)'\right)' \in \mathbb{R}^k$, where, in general, $\mathcal{B}_{u_i}^\dagger(r)$ and $\mathcal{B}_{v_i}(r)$ as well as $\mathcal{B}_{u_i}^\dagger(r)$ and $\mathcal{B}_v(r)$ are correlated.

Conditions on the stochastic process $(\boldsymbol{\eta}_t^\dagger)_{t\in\mathbb{Z}}$, where a functional central limit theorem holds, are provided in de Jong and Davidson (2000), Davidson (1994), and White (2001), as well as in Mark and Sul (2003), Mark et al. (2005), and Phillips (2014). In the latter articles $(\boldsymbol{\eta}_t^\dagger)$ has a moving average representation $\boldsymbol{\eta}_t^\dagger = \mathbf{\Psi}^\dagger(L)\boldsymbol{\varepsilon}_t^\dagger$, where $\mathbf{\Psi}^\dagger(L)$ is a lag polynomial and $\boldsymbol{\varepsilon}_t^\dagger = \left(\boldsymbol{\varepsilon}_{I1t}^{\dagger\prime},\ldots,\boldsymbol{\varepsilon}_{Int}^{\dagger\prime},\boldsymbol{\varepsilon}_{Ct}'\right)' \in \mathbb{R}^{(k_I+1)n+k_C}$ is a noise term. This also allows for heteroscedastic $\boldsymbol{\varepsilon}_t^\dagger$, which can be important if financial data sets are analyzed (Section 5).

Weak stationary of $(\boldsymbol{\eta}_t^\dagger)$, in particular of the components $u_{it}^\dagger$, is necessary to obtain a cointegrating relationship in (1). To see this, by Assumption 2, $(\mathbf{x}_t)$ follows a vector random walk process and is therefore integrated of order one. y_{it} arises from a weighted sum of $I(1)$ random variables, the fixed effects α_i and λ_t, the (stationary) term $\boldsymbol{\beta}_L'\mathbf{x}_{Li}$ as well as the stationary noise term $u_{it}^\dagger$. To exclude cointegration relationships between the components of $\mathbf{x}_{it}$ and to guarantee that y_{it} is $I(1)$, we imposed the assumption that $\boldsymbol{\beta} \neq \mathbf{0}_{(k)}$ and that the long run covariance matrix is of full rank. Hence, by Assumption 2 the rank of $\mathbf{\Omega}_{vv}$ is $k_I\cdot n+k_C$. The regressors $\mathbf{x}_{Li}$ are assumed to be independent of $\boldsymbol{\eta}_t^\dagger$ and therefore strictly exogenous, such that $u_{jt}^\dagger$ and $\mathbf{x}_{Li}$ are uncorrelated for all pairs $i,j = 1,\ldots,n$ and $t = 1,\ldots,T$.

Remark 1. By $\mathbf{y}_t := (y_{1t},\ldots,y_{nt})'$, $\mathbf{y}_t^* := (y_{1t}^*,\ldots,y_{nt}^*)'$, $\mathbf{u}_t^\dagger := (u_{1t}^\dagger,\ldots,u_{nt}^\dagger)'$, $\mathbf{x}_L := (\mathbf{x}_{L1}',\ldots,\mathbf{x}_{Ln}')' \in \mathbb{R}^{nk_L}$, $\boldsymbol{\lambda} := (\lambda_1,\ldots,\lambda_T)'$ $\boldsymbol{\alpha} := (\alpha_1,\ldots,\alpha_n)'$, $\tilde{\boldsymbol{\beta}} := \mathbf{I}_n \otimes \boldsymbol{\beta}' \in \mathbb{R}^{n\times nk}$, and $\tilde{\boldsymbol{\beta}}_L := \mathbf{I}_n \otimes \boldsymbol{\beta}_L' \in \mathbb{R}^{n\times nk_L}$, we obtain the *structural form* (triangular system)

$$\mathbf{y}_t = \rho\mathbf{y}_t^* + \tilde{\boldsymbol{\beta}}\mathbf{x}_t + \tilde{\boldsymbol{\beta}}_L\mathbf{x}_L + \boldsymbol{\alpha} + \lambda_t\mathbf{1}_{(n\times 1)} + \mathbf{u}_t^\dagger\,,\ \text{where}\ \mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t\,. \tag{6}$$

Assumption 1 guarantees that $\mathbf{I}_n - \rho\mathbf{W}$ has the full rank n. This allows us to obtain the *reduced form*

$$\mathbf{y}_t = (\mathbf{I}_n - \rho\mathbf{W})^{-1}\left(\tilde{\boldsymbol{\beta}}\mathbf{x}_t + \tilde{\boldsymbol{\beta}}_L\mathbf{x}_L + \boldsymbol{\alpha} + \lambda_t\mathbf{1}_{(n\times 1)} + \mathbf{u}_t^\dagger\right)\,,\ \text{where}\ \mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t\,. \tag{7}$$

Observe that the system constitutes n cointegrating equations. The cointegrating relationships do not have the usual linear form in the sense that the solution for y_{it} is a nonlinear function of the parameter ρ. When we consider the data generated by (1) we observe that: (i) $\mathbf{x}_{it}$ and $u_{it}^\dagger$ can be correlated, since

Assumption 2 does not exclude correlation between $\mathbf{v}_{it}$ and $u^{\dagger}_{it}$. (ii) For $\rho \neq 0$, y_{jt} depends on y_{it} and vice versa. (iii) $u^{\dagger}_{it}$ and $u^{\dagger}_{jt}$ are independent by Assumption 2. (iv) Since y_{jt} depends on y_{it} we know that $y^{*}_{it} = W_{ij}y_{jt}$ and $u^{\dagger}_{it}$ are correlated in general. (v) By Assumption 2 there is no correlation between $u^{\dagger}_{jt}$ and $\mathbf{x}_{L}$.

3. Estimation procedure and large sample results

The goal of the following analysis is to construct a dynamic two-stage least squares (*D2SLS*) estimator and to show that it leads to asymptotically unbiased estimates of the parameters ρ, $\boldsymbol{\beta}_I$ and $\boldsymbol{\beta}_C$ or $\boldsymbol{\beta}_L$. In particular, Section 3.1 applies the methodology of projecting on the leads and lags of the (first differenced) dependent variables $\mathbf{x}_{it} \in \mathbb{R}^k$ or $\mathbf{x}_t \in \mathbb{R}^{kn}$ introduced in Saikkonen (1991). We shall observe that these projections eliminate the correlation between $\mathbf{x}_{it}$ and $u^{\dagger}_{it}$ but not the y^{*}_{it} and $u^{\dagger}_{it}$ correlation discussed in Remark 1. To get rid of the latter type of correlation, instrumental variables will be applied. Since $\boldsymbol{\beta}_I$ as well as $\boldsymbol{\beta}_L$ or $\boldsymbol{\alpha}$ as well as $\boldsymbol{\lambda}$ are not separately estimable (see, e.g., Hsiao, 2015, Section 3.6.1), we consider different cases: Section 3.2 considers a model without time effects $\boldsymbol{\lambda}$ and without longitudinally common regressors $\mathbf{x}_{Li}$, i.e., $k_L = 0$. In this case model (1) becomes

$$y_{it} = \rho \sum_{j=1}^{n} W_{ij}y_{jt} + \boldsymbol{\beta}'_I\mathbf{x}_{Iit} + \boldsymbol{\beta}'_C\mathbf{x}_{Ct} + \alpha_i + u^{\dagger}_{it} = \rho y^{*}_{it} + \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + u^{\dagger}_{it}\,, \text{ where}$$

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t,\ \ \mathbf{x}_t = (\mathbf{x}'_{I1t},\ldots,\mathbf{x}'_{Int},\mathbf{x}'_{Ct})' \in \mathbb{R}^{nk_I+k_C},\ \ \boldsymbol{\beta} = \left(\boldsymbol{\beta}'_I,\boldsymbol{\beta}'_C\right)' \in \mathbb{R}^k, \text{ and } k = k_I + k_C. \quad (8)$$

Motivated by the empirical example discussed in Section 5, model (8) will be considered to be the leading case in the following. To simplify the algebra, we apply the within-transformation (see, e.g., Baltagi, 2008, p. 11) and derive the asymptotic distribution of the estimates of the slope coefficients ρ and $\boldsymbol{\beta}$ using within-transformed data. That is, the variables in deviations from their individual means (means taken with respect to the time series dimension) are

$$\widetilde{y}_{it} := y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}\,,\ \widetilde{\mathbf{x}}_{it} := \mathbf{x}_{it} - \frac{1}{T_\star}\sum_{t=1}^{T}\mathbf{x}_{it}\,,\ \widetilde{y}^{*}_{it} := \sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt} \text{ and } \widetilde{u}^{\dagger}_{it} := u^{\dagger}_{it} - \frac{1}{T}\sum_{t=1}^{T} u^{\dagger}_{it}\,, \quad (9)$$

such that (8) after applying the within-transformation (9) reads as follows:

$$\widetilde{y}_{it} = \rho \sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt} + \boldsymbol{\beta}'_I\widetilde{\mathbf{x}}_{Iit} + \boldsymbol{\beta}'_C\widetilde{\mathbf{x}}_{Ct} + \widetilde{u}^{\dagger}_{it_\star} = \rho\widetilde{y}^{*}_{it} + \boldsymbol{\beta}'\widetilde{\mathbf{x}}_{it} + \widetilde{u}^{\dagger}_{it_\star}\,. \quad (10)$$

For this model Section 3.2 obtains the $T \to \infty$-limit distribution for the ordinary least squares (*OLS*), the dynamic ordinary least squares (*DOLS*) and the two stage least squares (*2SLS*) estimator, where second-order asymptotic bias terms show up. Then, the $T \to \infty$-limit distribution of the *D2SLS* estimator is provided, where no second-order bias term shows up and the asymptotic limit distribution is a zero mean Gaussian mixture distribution (for a definition of the Gaussian mixture distribution see, e.g., Johansen, 1995, Chapter 13.1).

In a second step, Section 3.3 will consider the case where "$k_C = 0$, $k_L > 0$, no cross-sectional fixed effects are present ($\boldsymbol{\alpha} = \mathbf{0}_{(n\times 1)}$) but time fixed effects are included (i.e., $\boldsymbol{\lambda} \neq \mathbf{0}_{(T\times 1)}$)." In this case model (1) becomes

$$y_{it} = \rho \sum_{j=1}^{n} W_{ij}y_{jt} + \boldsymbol{\beta}'_I\mathbf{x}_{Iit} + \boldsymbol{\beta}'_L\mathbf{x}_{Li} + \lambda_t + u^{\dagger}_{it} = \rho y^{*}_{it} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\beta}'_L\mathbf{x}_{Li} + \lambda_t + u^{\dagger}_{it}\,,$$

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t\,,\text{ where } \mathbf{x}_t = \mathbf{x}_{It} = (\mathbf{x}'_{I1t},\ldots,\mathbf{x}'_{Int})' \in \mathbb{R}^{nk_I}\,,\mathbf{v}_t = \Delta\mathbf{x}_t\,,k_I = k\,,\text{ and } \boldsymbol{\beta} = \boldsymbol{\beta}_I\,. \quad (11)$$

Another within-transformation can be applied to model (11) to get rid of the time fixed effects $\boldsymbol{\lambda}$ (see, e.g., Baltagi, 2008). That is,

$$\grave{y}_{it} := y_{it} - \frac{1}{n}\sum_{i=1}^{n} y_{it}\,,\ \ \grave{y}^*_{it} := y^*_{it} - \frac{1}{n}\sum_{j=1}^{n} y^*_{jt}\,,\ \ \grave{\mathbf{x}}_{it} := \grave{\mathbf{x}}_{Iit} = \mathbf{x}_{Iit} - \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_{Ijt}\,,$$

$$\grave{\mathbf{x}}_{Li} := \mathbf{x}_{Li} - \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_{Lj} \text{ and } \grave{u}^\dagger_{it} := u_{it} - \frac{1}{n}\sum_{j=1}^{n} u^\dagger_{jt}\,, \text{ such that}$$

$$\grave{y}_{it} = \rho\sum_{j=1}^{n} W_{ij}\grave{y}_{jt} + \boldsymbol{\beta}_I'\grave{\mathbf{x}}_{Iit} + \boldsymbol{\beta}_L'\grave{\mathbf{x}}_{Li} + \grave{u}^\dagger_{it} = \rho\grave{y}^*_{it} + \boldsymbol{\beta}_I'\grave{\mathbf{x}}_{Iit} + \boldsymbol{\beta}_L'\grave{\mathbf{x}}_{Li} + \grave{u}^\dagger_{it}\,. \tag{12}$$

For this within-transformed model a D2SLS estimator can be developed. However, in this article we shall consider model (11) jointly with the case where "$k_L = k_C = 0$ and cross-sectional fixed effects as well as time fixed effects are allowed." In this case model (1) becomes equal to[4]

$$y_{it} = \rho\sum_{j=1}^{n} W_{ij}y_{jt} + \boldsymbol{\beta}_I'\mathbf{x}_{Iit} + \alpha_i + \lambda_t + u^\dagger_{it} = \rho y^*_{it} + \boldsymbol{\beta}_I'\mathbf{x}_{Iit} + \alpha_i + \lambda_t + u^\dagger_{it}\,,$$

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t\,, \text{ where } \mathbf{x}_t = \mathbf{x}_{It} = (\mathbf{x}'_{I1t},\ldots,\mathbf{x}'_{Int})' \in \mathbb{R}^k\,,\ \ \boldsymbol{\beta} = \boldsymbol{\beta}_I\,,\ \ k = k_I, \ \ \text{and } \mathbf{v}_t = \Delta\mathbf{x}_t\,. \tag{13}$$

To model (13) the within-transformation (see, e.g., Baltagi, 2008)

$$\breve{y}_{it} := y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} - \frac{1}{n}\sum_{j=1}^{n} y_{jt} + \frac{1}{Tn}\sum_{j=1}^{n}\sum_{t=1}^{T} y_{jt}\,,\ \ \breve{y}^*_{it} := \sum_{j=1}^{n} W_{ij}\breve{y}_{jt}\,,$$

$$\breve{\mathbf{x}}_{it} := \breve{\mathbf{x}}_{Iit} := \mathbf{x}_{Iit} - \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it} - \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_{jt} + \frac{1}{Tn}\sum_{j=1}^{n}\sum_{t=1}^{T} \mathbf{x}_{jt}\,,$$

$$\breve{u}^\dagger_{it} := u^\dagger_{it} - \frac{1}{T}\sum_{t=1}^{T} u^\dagger_{it} - \frac{1}{n}\sum_{j=1}^{n} u^\dagger_{jt} + \frac{1}{Tn}\sum_{j=1}^{n}\sum_{t=1}^{T} u^\dagger_{jt} \text{ is applied to obtain}$$

$$\breve{y}_{it} = \rho\sum_{j=1}^{n} W_{ij}\breve{y}_{jt} + \boldsymbol{\beta}_I'\breve{\mathbf{x}}_{Iit} + \breve{u}^\dagger_{it} = \rho\breve{y}^*_{it} + \boldsymbol{\beta}_I'\breve{\mathbf{x}}_{Iit} + \breve{u}^\dagger_{it}\,. \tag{14}$$

In addition, since $\breve{\mathbf{x}}_{Li} := \mathbf{x}_{Li} - \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{Li} - \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_{Lj} + \frac{1}{Tn}\sum_{j=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{Lj} = \mathbf{x}_{Li} - \frac{T}{T}\mathbf{x}_{Li} - \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_{Lj} + \frac{T}{Tn}\sum_{j=1}^{n}\mathbf{x}_{Lj} = \mathbf{0}_{(k_L\times 1)}$, applying the within-transform used in (14) to model (11) also results in $\breve{y}_{it} = \rho\breve{y}^*_{it} + \boldsymbol{\beta}_I'\breve{\mathbf{x}}_{Iit} + \breve{u}^\dagger_{it}$. This allows for a joint treatment of the models (11) and (13). For model (14) the limit distribution for the parameters ρ and $\boldsymbol{\beta}_I$ will be obtained when T or (n, T) become large in Section 3.3. For the model (11), given the estimates of ρ and $\boldsymbol{\beta}_I$, an estimate of $\boldsymbol{\beta}_L$ will be derived by means of a further ordinary least squares regression.

[4]To distinguish a model where the econometrician decides to include different kinds for fixed effects (given that $k_C = 0$ or $k_L = 0$), indicator functions of the form $\mathbb{I}_{(\alpha)}$ and $\mathbb{I}_{(\lambda)}$ can be included into model (1). To simplify the notation and to improve the readability we assume that if $k_C = 0$ and $k_L = 0$, then time fixed effects as well as cross-sectionally fixed effects are included.

3.1. Projection facility

Assumption 2 implies that potentially all leads and lags of $\Delta \mathbf{x}_{Iit} = \mathbf{v}_{Iit}$ and $\Delta \mathbf{x}_{Ct} = \mathbf{v}_{Ct}$ are correlated with $u_{it}^{\dagger}$. In the next step, we follow *DOLS* literature and remove the correlation of $u_{it}^{\dagger}$ and $\mathbf{v}_{it}$ by projecting on the leads and lags of the prediction variables. This implies that for each i, we project on $\Delta \mathbf{x}_{it-s} = \left(\Delta \mathbf{x}'_{Iit-s}, \Delta \mathbf{x}'_{Ct-s}\right)' = \left(\mathbf{v}'_{Iit-s}, \mathbf{v}'_{Ct-s}\right)'$ for $s = -p, \ldots, 0, \ldots, p$. The projection of $u_{it}^{\dagger}$ on the p leads and lags of $\Delta \mathbf{x}_{it}$ yields a truncation component $\sum_{s=-p}^{+p} \boldsymbol{\delta}'_{i,s} \Delta \mathbf{x}_{i,t-s}$, vectors of projection coefficients $\boldsymbol{\delta}_{i,s} \in \mathbb{R}^k$ (for $s = -p, \ldots, p$), a truncation error $\mathrm{e}_{p;it} := \sum_{s>p,s<-p} \boldsymbol{\delta}'_{i,s} \Delta \mathbf{x}_{i,t-s}$ plus a *new* disturbance u_{it}, such that

$$u_{it}^{\dagger} = \underline{\boldsymbol{\delta}}'_{p;i} \boldsymbol{\zeta}_{p;it} + \mathrm{e}_{p;it} + u_{it} = \underline{\boldsymbol{\delta}}'_{p;i} \boldsymbol{\zeta}_{p;it} + \underline{u}_{p;it} \text{ and } \underline{u}_{p;it} := \mathrm{e}_{p;it} + u_{it}\,, \text{ where}$$

$$\boldsymbol{\zeta}_{p;it} := \left(\mathbf{v}'_{Ii,t-p}, \mathbf{v}'_{C,t-p}, \ldots, \mathbf{v}'_{Ii,t}, \mathbf{v}'_{C,t}, \ldots, \mathbf{v}'_{Ii,t+p}, \mathbf{v}'_{C,t+p}\right)' \in \mathbb{R}^{(2p+1)k} \text{ and}$$

$$\underline{\boldsymbol{\delta}}_{p;i} := \left(\boldsymbol{\delta}'_{i,-p}, \ldots, \boldsymbol{\delta}'_{i,0}, \ldots, \boldsymbol{\delta}'_{i,+p}\right)' \in \mathbb{R}^{(2p+1)k}\,. \tag{15}$$

The subscript p denotes that the truncation error $\mathrm{e}_{p;it}$ as well as the noise term $\underline{u}_{p;it} = \mathrm{e}_{p;it} + u_{it}$ depend on the number of leads and lags p. $\boldsymbol{\zeta}_{p;it}$ is by construction orthogonal to the new noise term u_{it}, while the term $\underline{u}_{p;it} = \mathrm{e}_{p;it} + u_{it}$ can still be correlated with $\Delta \mathbf{x}_{it}$ for some $p < \infty$.

A further alternative is to follow the system DOLS approach (see, e.g., Park and Ogaki, 1991) and project on the leads and lags of *all* cross-sections. That is, $\boldsymbol{\zeta}_{\sharp p;it} = \boldsymbol{\zeta}_{\sharp p;t} := \left(\mathbf{v}'_{t-p}, \ldots, \mathbf{v}'_{t}, \ldots \mathbf{v}'_{t-p}\right)' \in \mathbb{R}^{(2p+1)(k_I n+k_C)}$, where in Section 2, we already defined $\mathbf{v}_t = \Delta \mathbf{x}_t = \left(\mathbf{v}'_{I1t}, \ldots, \mathbf{v}'_{Int}, \mathbf{v}'_{Ct}\right)' \in \mathbb{R}^{n \cdot k_I + k_C}$. Then,

$$u_{it}^{\dagger} = \underline{\boldsymbol{\delta}}'_{\sharp p;i} \boldsymbol{\zeta}_{\sharp p;t} + \mathrm{e}_{\sharp p;it} + u_{it} = \underline{\boldsymbol{\delta}}'_{\sharp p;i} \boldsymbol{\zeta}_{\sharp p;t} + \underline{u}_{\sharp p;it}\,,$$

$$\boldsymbol{\zeta}_{\sharp p;t} = \left(\mathbf{v}'_{t-p}, \ldots, \mathbf{v}'_{t}, \ldots, \mathbf{v}'_{t+p}\right)' \in \mathbb{R}^{(2p+1)(k_I n+k_C)}\,,$$

$$\underline{u}_{\sharp p;it} := \mathrm{e}_{\sharp p;it} + u_{it}\,, \text{ and } \underline{\boldsymbol{\delta}}_{\sharp p;i} := \left(\boldsymbol{\delta}'_{\sharp i,-p}, \ldots, \boldsymbol{\delta}'_{\sharp i,0}, \ldots, \boldsymbol{\delta}'_{\sharp i,+p}\right)' \in \mathbb{R}^{(2p+1)(k_I n+k_C)}\,. \tag{16}$$

We shall observe that projecting on the own leads and lags (15) is sufficient to eliminate the correlation asymptotically between the regressors and the noise term in model (8), while with time fixed effects or $k_L > 0$, where the within-transform (14) is applied, the projection on all leads and lags (16) is used to get rid of the correlation between $\breve{\mathbf{x}}_{Iit}$ and $\breve{u}_{it}^{\dagger}$.[5] Now we impose an additional restriction on the error dynamics that will guarantee that the truncation error $\mathrm{e}_{p;it}$ (or $\mathrm{e}_{\sharp p;it}$) converges to zero when T becomes large:

Assumption 3 (Error Dynamics II; see Saikkonen (1991), Mark et al. (2005)). *Suppose that $p = p(T)$. Then $p(T)$ has to fulfill $\frac{p(T)^3}{T} \to 0$ and $\sqrt{T} \sum_{|s|>p(T)} \|\boldsymbol{\delta}_{i,s}\|_2 \to 0$ (or $\sqrt{T} \sum_{|s|>p(T)} \|\boldsymbol{\delta}_{\sharp i,s}\|_2 \to 0$ when projecting on the full cross-section of leads and lags) as $T \to \infty$, where $\|.\|_2$ stands for the Euclidean norm.*

Assumption 3 requires that $p(T)$ does not grow too fast, while the second part restricts the dependence between the noise term and the regressors. Based on Assumptions 2–3, for model (10), if T becomes large then—due to the increase in the number of leads and lags $p(T)$—the truncation error $\mathrm{e}_{p;it}$ becomes small. As a result, the difference between $\underline{u}_{p;it}$ and u_{it} becomes small, such that $\underline{u}_{p;it}$ becomes

[5] By contrast, when skipping the assumption of independent cross-sections, Online Appendix A-4 demonstrates that in this case a projection on all leads and lags becomes necessary already for model (8). The Online Appendix can be downloaded from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2256929.

orthogonal to $\boldsymbol{\zeta}_{p;it}$ as $T \to \infty$. For the transformed model (14), $\breve{\mathbf{x}}_{Iit}$ as well as $\breve{u}^{\dagger}_{it}$ contain weighted elements from the other cross-sections $j = 1, \dots, n, j \neq i$. The projections on the full cross-section of leads and lags is applied to the transformed model (14). Then the truncation error $e_{\sharp p;it}$ becomes small such that $\underline{\breve{u}}_{\sharp p,it}$ converges to $\breve{u}_{it}$ and the $\breve{\mathbf{x}}_{Iit}$, $\underline{\breve{u}}_{\sharp p,it}$-correlation gets small for large T due to Assumption 3.

After applying the projections (15) for the transformed model (10) as well as (16) with the transformed model (14), we arrive at the *new* covariance stationary process $(\boldsymbol{\eta}_t)_{t\in\mathbb{Z}}$, where $\boldsymbol{\eta}_t := (\mathbf{u}_t', \mathbf{v}_t')' \in \mathbb{R}^{(k_I+1)n+k_C}$, $\boldsymbol{\eta}_{It} := ((\boldsymbol{\eta}_{it}')_{i=1,\dots,n})' \in \mathbb{R}^{(k_I+1)n}$, $\boldsymbol{\eta}_{it} := (u_{it}, \mathbf{v}_{it}')' \in \mathbb{R}^{k_I+1}$, $\mathbf{v}_t = (\mathbf{v}_{I1t}', \dots, \mathbf{v}_{Int}', \mathbf{v}_{Cit}')' \in \mathbb{R}^{n\cdot k_I+k_C}$, and $\mathbf{u}_t := (u_{1t}, \dots, u_{nt})' \in \mathbb{R}^n$). For any $\ell \in \mathbb{Z}$ we get $\boldsymbol{\Gamma}_\ell := \mathbb{E}\left(\boldsymbol{\eta}_{t-\ell}\boldsymbol{\eta}_t'\right) \in \mathbb{R}^{((k_I+1)\cdot n+k_C)\times((k_I+1)\cdot n+k_C)}$, the new long run covariance matrix $\boldsymbol{\Omega} := \sum_{\ell=-\infty}^{\infty} \mathbb{E}\left(\boldsymbol{\eta}_{t-\ell}\boldsymbol{\eta}_t'\right)$ and new half long covariance $\Delta := \sum_{\ell=0}^{\infty} \mathbb{E}\left(\boldsymbol{\eta}_{t-\ell}\boldsymbol{\eta}_t'\right)$ $u^{\dagger}_{it}$ (16) the projection on all leads and lags is used to get rid of the correlation between $\breve{\mathbf{x}}_{Iit}$ and $u^{\dagger}_{it}$ (16). By applying the notation introduced in (2) and (3), we get $\boldsymbol{\Gamma}_{\ell,vu} = \boldsymbol{\Omega}_{vu} = \boldsymbol{\Delta}_{vu} = \mathbf{0}_{(nk\times n)}$ as well as $\Omega_{u_iu_j} = \Gamma_{\ell,u_iu_j} = 0$, for $i \neq j$ by Assumption 2. By the construction of the noise process u_{it}, the correlation between u_{it} and $\mathbf{v}_{j,t-\ell}$ is zero for all $j = 1, \dots, n$ and $\ell \in \mathbb{Z}$.

In addition, while the time index t goes from 1 to T in (1), after the projection facilities are applied only the observations $p+1, \dots, T-p$ can be used to estimate the model parameters. By defining $t_\star := t - p + 1$ we obtain a new time index which accounts for the projection facility. Then $T_\star = T - 2p$ and $t_\star = 1, \dots, T_\star$.

Given the assumptions on the error dynamics, a *functional central limit theorem* can be applied, such that if $T_\star \to \infty$:

$$\frac{1}{\sqrt{T_\star}} \sum_{t_\star=1}^{[T_\star r]} \boldsymbol{\eta}_{t_\star} \Rightarrow \mathcal{B}(r) = \boldsymbol{\Omega}^{1/2}\mathcal{W}(r)\,. \tag{17}$$

$\mathcal{W}(r)$ is a standard Brownian motion in $\mathbb{R}^{k_I\cdot n+k_C}$, while $\mathcal{B}(r) = \boldsymbol{\Omega}^{1/2}\mathcal{W}(r)$. These Brownian motions contain the components, $\mathcal{B}_{v_i} = \left(\mathcal{B}_{vI_i}(r)', \mathcal{B}_{vC}(r)'\right)'$ and $\mathcal{W}_{v_i} := \left(\mathcal{W}_{vI_i}(r)', \mathcal{W}_{vC}(r)'\right)'$. By Assumption 2 and the construction of $\boldsymbol{\eta}_{t_\star}$, $\mathcal{B}_{u_i}(r)$ and $\mathcal{B}_{v_i}(r)$ as well as $\mathcal{B}_{u_i}(r)$ and $\mathcal{B}_v(r)$ are independent for all $i = 1, \dots, n$. This also yields $\mathcal{B}_{u_i}(r) = \sqrt{\Omega_{u_iu_i}}\mathcal{W}_{u_i}(r)$. By Davidson (1994, Theorem 30.2), $\frac{1}{\sqrt{T_\star}}\tilde{\mathbf{x}}_{i[rT_\star]} \Rightarrow \mathcal{B}_{v_i}(r) - \int_0^1 \mathcal{B}_{v_i}(s)ds$ and $\frac{1}{T_\star^2}\sum_{t_\star=1}^{[rT_\star]} \tilde{\mathbf{x}}_{it_\star}\tilde{\mathbf{x}}_{it_\star}' \Rightarrow \int_0^r \tilde{\mathcal{B}}_{v_i}(s)\tilde{\mathcal{B}}_{v_i}'(s)ds$, for $T_\star \to \infty$, where the *demeaned* Brownian motion $\mathcal{B}_{v_i}(r) - \int_0^1 \mathcal{B}_{v_i}(s)ds$ is abbreviated by $\tilde{\mathcal{B}}_{v_i}(r)$. Assumption 2 and some algebra results in $\frac{1}{T_\star}\sum_{t_\star=1}^{[T_\star]} \tilde{\mathbf{x}}_{it_\star}\tilde{u}_{it_\star} \Rightarrow \sqrt{\Omega_{u_iu_i}}\int_0^1 \tilde{\mathcal{B}}_{v_i}(s)d\mathcal{W}_{u_i}(s) + \boldsymbol{\Delta}_{v_iu_i}$ (see also Davidson, 1994, Theorem 30.13). Since, $\mathbf{v}_{it_\star}$ and $u_{it_\star}$ are uncorrelated, $\boldsymbol{\Delta}_{v_iu_i} = \mathbf{0}_{(k)}$ and $\frac{1}{T_\star}\sum_{t_\star=1}^{[T_\star]} \tilde{\mathbf{x}}_{it_\star}\tilde{u}_{it_\star} \Rightarrow \sqrt{\Omega_{u_iu_i}}\int_0^1 \tilde{\mathcal{B}}_{v_i}(s)d\mathcal{W}_{u_i}(s)$. By contrast, $\frac{1}{T_\star}\sum_{t_\star=1}^{[T_\star]} \tilde{\mathbf{x}}_{it_\star}\tilde{u}^{\dagger}_{it_\star} \Rightarrow \sqrt{\Omega_{u_iu_i}}\int_0^1 \tilde{\mathcal{B}}_{v_i}(s)d\mathcal{W}^{\dagger}_{u_i}(s) + \boldsymbol{\Delta}^{\dagger}_{v_iu_i}$, where - in general - $\boldsymbol{\Delta}^{\dagger}_{v_iu_i} \neq \mathbf{0}_{(k)}$. In addition, we derive $\frac{1}{\sqrt{T_\star}}\grave{\mathbf{x}}_{i[rT_\star]} \Rightarrow \mathcal{B}_{v_i}(r) - \frac{1}{n}\sum_{j=1}^n \mathcal{B}_{v_j}(r) =: \grave{\mathcal{B}}_{v_i}(r)$, where $\grave{\mathcal{B}}_{u_i}(r)$ and $\grave{\mathcal{B}}_v(r)$ are defined in the same way, and $\frac{1}{\sqrt{T_\star}}\breve{\mathbf{x}}_{i[rT_\star]} \Rightarrow \mathcal{B}_{v_i}(r) - \int_0^1 \mathcal{B}_{v_i}(s)ds - \frac{1}{n}\sum_{j=1}^n \mathcal{B}_{v_j}(r) + \frac{1}{n}\sum_{j=1}^n \int_0^1 \mathcal{B}_{v_j}(s)ds =: \breve{\mathcal{B}}_{v_i}(r)$, where $\breve{\mathcal{B}}_{u_i}(r)$ and $\breve{\mathcal{B}}_v(r)$ are defined in an equivalent way (for more details see Online Appendix A-2).

3.2. Large sample properties of some parameter estimators for model (8)

This subsection investigates the model defined in (8), where $k_L = 0$ and no time effects are included. To write down our estimator in a compact way, we define the model in a stacked notation. Define $\widetilde{\mathbf{y}} := (\widetilde{y}_{11}, \dots, \widetilde{y}_{1T_\star}, \dots, \widetilde{y}_{n1}, \dots, \widetilde{y}_{nT_\star})'$, $\mathbf{y}^* := \left(\widetilde{y}^*_{11}, \dots, \widetilde{y}^*_{1T_\star}, \dots, \widetilde{y}^*_{n1}, \dots, \widetilde{y}^*_{nT_\star}\right)'$, $\widetilde{\mathbf{x}} := \left(\widetilde{\mathbf{x}}_{11}, \dots, \widetilde{\mathbf{x}}_{1T_\star}, \dots, \widetilde{\mathbf{x}}_{n1}, \dots, \widetilde{\mathbf{x}}_{nT_\star}\right)'$, and $\widetilde{\underline{\mathbf{u}}}_p := \left(\widetilde{\underline{u}}_{p;11}, \dots, \widetilde{\underline{u}}_{p;1T_\star}, \dots, \widetilde{\underline{u}}_{p;n1}, \dots, \widetilde{\underline{u}}_{p;nT_\star}\right)'$, where $\widetilde{\mathbf{y}}$, $\widetilde{\mathbf{y}}^*$, and $\widetilde{\underline{\mathbf{u}}}_p$ are of dimension $nT_\star \times 1$, while $\widetilde{\mathbf{x}}$ is an $nT_\star \times k$ matrix. Furthermore, we have

$$\widetilde{\boldsymbol{\zeta}}_p := \begin{pmatrix} \widetilde{\boldsymbol{\zeta}}'_{p;11} & \mathbf{0}_{(1\times(2p+1)k)} & \mathbf{0}_{(1\times(2p+1)k)} \\ \vdots & & \\ \widetilde{\boldsymbol{\zeta}}'_{p;1T_\star} & \mathbf{0}_{(1\times(2p+1)k)} & \mathbf{0}_{(1\times(2p+1)k)} \\ \mathbf{0}_{(1\times(2p+1)k)} & \widetilde{\boldsymbol{\zeta}}_{p;21} & \mathbf{0}_{(1\times(2p+1)k)} \\ & & \ddots \\ \mathbf{0}_{(1\times(2p+1)k)} & \mathbf{0}_{(1\times(2p+1)k)} & \widetilde{\boldsymbol{\zeta}}_{p;nT_\star} \end{pmatrix} \in \mathbb{R}^{nT_\star\times(2p+1)k\cdot n} \text{ and}$$

$$\underline{\boldsymbol{\delta}}_p := \begin{pmatrix} \underline{\boldsymbol{\delta}}_{p;1} \\ \vdots \\ \underline{\boldsymbol{\delta}}_{p;n} \end{pmatrix} \in \mathbb{R}^{(2p+1)k\cdot n\times 1} . \tag{18}$$

This provides us with model (10) in stacked form:

$$\widetilde{\mathbf{y}} = \rho\widetilde{\mathbf{y}}^* + \widetilde{\mathbf{x}}\boldsymbol{\beta} + \widetilde{\boldsymbol{\zeta}}_p\underline{\boldsymbol{\delta}}_p + \widetilde{\underline{\mathbf{u}}}_p = \left(\widetilde{\mathbf{y}}^*, \widetilde{\mathbf{x}}\right)\boldsymbol{\gamma} + \widetilde{\boldsymbol{\zeta}}_p\underline{\boldsymbol{\delta}}_p + \widetilde{\underline{\mathbf{u}}}_p = \widetilde{\mathbf{X}}_p\left(\boldsymbol{\gamma}', \underline{\boldsymbol{\delta}}'_p\right)' + \widetilde{\underline{\mathbf{u}}}_p = \widetilde{\mathbf{X}}_p\boldsymbol{\theta}'_p + \widetilde{\underline{\mathbf{u}}}_p , \tag{19}$$

where $\boldsymbol{\gamma} := (\rho, \boldsymbol{\beta}')' \in \mathbb{R}^{1+k}$ and $\boldsymbol{\theta}_p := \left(\boldsymbol{\gamma}', \underline{\boldsymbol{\delta}}'_p\right)' \in \mathbb{R}^{1+k+(2p+1)\cdot kn}$. The right-hand side variables are collected in $\widetilde{\mathbf{X}}_p = \left(\widetilde{\mathbf{y}}^*, \widetilde{\mathbf{x}}, \boldsymbol{\zeta}_p\right) \in \mathbb{R}^{nT_\star\times k+1+(2p+1)\cdot kn}$. The transpose of the rows of the matrix $\widetilde{\mathbf{X}}_p$ are the column vectors $\widetilde{\mathbf{X}}_{p;it_\star} := \left(\tilde{y}^*_{it_\star}, \tilde{\mathbf{x}}'_{it_\star}, \mathbf{0}_{(1\times(2p+1)k\cdot(i-1))}, \widetilde{\boldsymbol{\zeta}}'_{p;it_\star}, \mathbf{0}_{(1\times(2p+1)k\cdot(n-i-1))}\right)' \in \mathbb{R}^{k+1+(2p+1)\cdot kn}$. Including the projection facility (15), model (8) can be written as $\tilde{y}_{it_\star} = \widetilde{\mathbf{X}}'_{p;it_\star}\boldsymbol{\theta}_p + \underline{\tilde{u}}_{p;it_\star}$. In addition, we apply the following notation: Let $\mathbf{W}_i \in \mathbb{R}^{1\times n}$ stand for the ith row of $\mathbf{W}$. The $n \cdot (k_I + k_C) \times nk_I + k_C$ matrix $\mathbf{C}$ transforms $\tilde{\mathbf{x}}_t = \left(\tilde{\mathbf{x}}'_{I1t}, \ldots, \tilde{\mathbf{x}}'_{Int}, \tilde{\mathbf{x}}'_{Ct}\right)' \in \mathbb{R}^{k_In+k_C}$ into $\left(\tilde{\mathbf{x}}'_{I1t}, \tilde{\mathbf{x}}'_{Ct}, \ldots, \tilde{\mathbf{x}}'_{Int}, \tilde{\mathbf{x}}'_{Ct}\right)' \in \mathbb{R}^{(k_I+k_C)\cdot n}$. Note that each row of $\mathbf{C}$ contains exactly one element equal to 1, while the other elements are zero. In addition, $\mathbf{C}$ has full column rank $nk_I + k_C$. For $k_C = 0$, $\mathbf{C} = \mathbf{I}_{nk_I}$. Moreover, let $\underline{\tilde{\mathbf{u}}}_t := \left(\underline{\tilde{u}}_{1t}, \ldots, \underline{\tilde{u}}_{nt}\right)'$ and $\tilde{\boldsymbol{\zeta}}_{p,t} = diag(\tilde{\boldsymbol{\zeta}}'_{p,1t}, \ldots, \tilde{\boldsymbol{\zeta}}'_{p,nt}) \in \mathbb{R}^{n\times(2p+1)\cdot kn}$. Then, $\tilde{y}^*_{it} = \sum_{j=1}^n W_{ij}\sum_{l=1}^n K_{jl}\left(\beta\tilde{\mathbf{x}}_{lt} + \underline{\boldsymbol{\delta}}'_{p;l}\tilde{\boldsymbol{\zeta}}_{p;lt_*} + \underline{\tilde{u}}_{lt}\right)$ can be expressed more compactly by $\mathbf{W}_i\mathbf{K}\left(\tilde{\boldsymbol{\beta}}\mathbf{C}\tilde{\mathbf{x}}_t + \tilde{\boldsymbol{\zeta}}_{p,t}\underline{\boldsymbol{\delta}}_p + \underline{\tilde{\mathbf{u}}}_t\right)$. In the same way as $\left(\tilde{\mathbf{x}}'_{1t}, \ldots, \tilde{\mathbf{x}}'_{nt}\right)' = \mathbf{C}\tilde{\mathbf{x}}_t$ we proceed with $\tilde{\mathbf{v}}$, resulting in $\mathbf{C}\tilde{\mathbf{v}}_t = \mathbf{C}\Delta\tilde{\mathbf{x}}_t$ the half-long covariance matrix $\mathbf{C}\boldsymbol{\Delta}_{vu} \in \mathbb{R}^{kn\times n}$, and the superposition of the components of the demeaned Brownian motion $\mathbf{C}\tilde{\mathcal{B}}_v \in \mathbb{R}^{kn}$.

Let us start with the *OLS*-estimator, where $p = \emptyset$, $\underline{\tilde{u}}_{it} = \tilde{u}^\dagger_{it}$, $T = T_\star$, $t = t_\star$, $\widetilde{\mathbf{X}}^{OLS}_{it} := \left(\tilde{y}^*_{it}, \tilde{\mathbf{x}}'_{it}\right)'$ and

$$\widehat{\boldsymbol{\gamma}}_{OLS} := \left(\hat{\rho}_{OLS}, \hat{\boldsymbol{\beta}}'_{OLS}\right)' = \left(\sum_{i=1}^n\sum_{t=1}^T \widetilde{\mathbf{X}}^{OLS}_{it}\widetilde{\mathbf{X}}^{OLS\prime}_{it}\right)^{-1}\sum_{i=1}^n\sum_{t=1}^T \widetilde{\mathbf{X}}^{OLS}_{it}\tilde{y}_{it} . \tag{20}$$

To obtain the $T \to \infty$-limit distribution, $\widetilde{\mathbf{X}}^{OLS}_{it}$ is scaled by $1/T$. This yields

Proposition 1. *Consider the fixed effects spatial correlation model (8) and the OLS estimator (20) based on the within-transformed model (10). Suppose that the Assumptions 1 and 2 hold.*

Then, for n fixed and $T \to \infty$, it follows that the $T \to \infty$ limits of $\mathbf{M}^{OLS}_{\tilde{X}\tilde{X},Tni} := \frac{1}{T^2}\sum_{t=1}^T \widetilde{\mathbf{X}}^{OLS}_{it}\widetilde{\mathbf{X}}^{OLS\prime}_{it}$ and $\mathbf{M}^{OLS}_{\tilde{X}\tilde{X},Tn} := \sum_{i=1}^n \mathbf{M}^{OLS}_{\tilde{X}\tilde{X},Tni}$ are

$$\mathbf{M}_{\tilde{X}\tilde{X},ni} := \int_0^1 \mathbf{g}_i(r)\mathbf{g}_i(r)'dr , \ \mathbf{g}_i(r) := \begin{pmatrix} \mathbf{W}_i\mathbf{K}\tilde{\boldsymbol{\beta}}\mathbf{C}\tilde{\mathcal{B}}_v(r) \\ \tilde{\mathcal{B}}_{v_i}(r) \end{pmatrix} , \ \textit{and } \mathbf{M}_{\tilde{X}\tilde{X},n} := \sum_{i=1}^n \mathbf{M}_{\tilde{X}\tilde{X},ni} , \tag{21}$$

while the $T \to \infty$ *limits of* $\mathbf{m}^{OLS}_{\tilde{X}\tilde{u}^\dagger,ni} := \frac{1}{T}\sum_{t=1}^{T} \widetilde{\mathbf{X}}^{OLS}_{it}\tilde{u}^\dagger_{it}$ *and* $\mathbf{m}^{OLS}_{\tilde{X}\tilde{u}^\dagger,n} := \sum_{i=1}^{n} \mathbf{m}^{OLS}_{\tilde{X}\tilde{u}^\dagger,ni}$ *are*

$$\mathbf{m}_{\tilde{X}\tilde{u}^\dagger,ni} := \begin{pmatrix} \mathbf{W}_i\mathbf{K}\int_0^1 \tilde{\boldsymbol{\beta}}\mathbf{C}\tilde{\mathcal{B}}_v(r)d\mathcal{B}^\dagger_{u_i}(r) \\ \sqrt{\Omega^\dagger_{u_iu_i}}\int_0^1 \tilde{\mathcal{B}}_{v_i}(r)d\mathcal{W}^\dagger_{u_i}(r) \end{pmatrix} + \begin{pmatrix} \mathbf{W}_i\mathbf{K}\left[\tilde{\boldsymbol{\beta}}\mathbf{C}\boldsymbol{\Delta}^\dagger_{vu_i} + \boldsymbol{\Gamma}^\dagger_{0,uu_i}\right] \\ \boldsymbol{\Delta}^\dagger_{v_iu_i} \end{pmatrix}, \text{ and } \mathbf{m}_{\tilde{X}\tilde{u}^\dagger,n} := \sum_{i=1}^{n} \mathbf{m}_{\tilde{X}\tilde{u}^\dagger,ni}\,.$$

For the centered and scaled ordinary least squares estimator of $\boldsymbol{\gamma}$ *we observe:*

$$T\left(\hat{\boldsymbol{\gamma}}_{OLS} - \boldsymbol{\gamma}\right) \Rightarrow \mathbf{M}^{-1}_{\tilde{X}\tilde{X},n}\mathbf{m}_{\tilde{X}\tilde{u}^\dagger,n}\,. \tag{22}$$

Proof. See Appendix A.

Note that $\mathbf{m}_{\tilde{X}\tilde{u}^\dagger,n}$ contains the "usual" second-order bias term $\sum_{i=1}^{n} \boldsymbol{\Delta}^\dagger_{v_iu_i}$ in the coordinates 2 to $k+1$, while in the first component of $\mathbf{m}_{\tilde{X}\tilde{u}^\dagger,n}$ we observe the second-order bias term $\sum_{i=1}^{n} \mathbf{W}_i\mathbf{K}\left(\tilde{\boldsymbol{\beta}}\mathbf{C}\boldsymbol{\Delta}^\dagger_{vu_i} + \boldsymbol{\Gamma}^\dagger_{0,uu_i}\right)$ arising from a spatial lag. Next, the panel *DOLS* estimator, derived in Mark and Sul (2003) as

$$\hat{\boldsymbol{\theta}}_{DOLS;p} := \left(\widetilde{\mathbf{X}}_p'\widetilde{\mathbf{X}}_p\right)^{-1}\widetilde{\mathbf{X}}_p'\tilde{\mathbf{y}}\,, \tag{23}$$

results in $\hat{\boldsymbol{\theta}}_{DOLS;p} - \boldsymbol{\theta}_p = \left(\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star} \widetilde{\mathbf{X}}_{p;it_\star}\widetilde{\mathbf{X}}'_{p;it_\star}\right)^{-1}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\widetilde{\mathbf{X}}_{p;it_\star}\tilde{\underline{u}}_{p;it_\star}$. To obtain the $T \to \infty$-limit distribution of $\hat{\boldsymbol{\gamma}}_{DOLS;p} := \left(\hat{\rho}_{DOLS;p}, \hat{\boldsymbol{\beta}}'_{DOLS;p}\right)'$, the first $k+1$ components of $\widetilde{\mathbf{X}}_{p;it_\star}$ are scaled by $1/T_\star$, while the remaining components are scaled by $1/\sqrt{T_\star}$, resulting in the scaling matrix $\mathbf{A}_{\tilde{X}p} := diag\left(T_\star^{-1}\cdot\mathbf{I}_{k+1}, T_\star^{-0.5}\cdot\mathbf{I}_{(2p+1)nk}\right) \in \mathbb{R}^{k+1+(2p+1)nk\times k+1+(2p+1)nk}$. For the *DOLS* estimator we obtain:

Proposition 2. *Consider the fixed effects spatial correlation model (8) and the DOLS estimator (23) based on the within-transformed model (10). Suppose that the Assumptions 1 to 3 hold. Let* $T_\star = T - 2p(T)$. *Then, for n fixed and* $T \to \infty$, *it follows that:*

(a) $T_\star(\hat{\boldsymbol{\gamma}}_{DOLS;p} - \boldsymbol{\gamma})$ *and* $\sqrt{T_\star}(\hat{\underline{\boldsymbol{\delta}}}_{DOLS,p} - \underline{\boldsymbol{\delta}}_p)$ *are asymptotically independent.*

(b) $T_\star\left(\hat{\boldsymbol{\gamma}}_{DOLS;p} - \boldsymbol{\gamma}\right) \Rightarrow \mathbf{M}^{-1}_{\tilde{X}\tilde{X},n}\mathbf{m}_{\tilde{X}\tilde{u},n}$, *where* $\mathbf{M}_{\tilde{X}\tilde{X},n}$ *follows from (21),*

$$\mathbf{m}_{\tilde{X}\tilde{u},ni} := \begin{pmatrix} \mathbf{W}_i\mathbf{K}\int_0^1 \tilde{\boldsymbol{\beta}}\mathbf{C}\tilde{\mathcal{B}}_v(r)d\mathcal{B}_{u_i}(r) \\ \sqrt{\Omega_{u_iu_i}}\int_0^1 \tilde{\mathcal{B}}_{v_i}(r)d\mathcal{W}_{u_i}(r) \end{pmatrix} + \begin{pmatrix} \mathbf{W}_i\mathbf{K}\boldsymbol{\Gamma}_{0,uu_i} \\ \mathbf{0}_{(k)} \end{pmatrix}, \text{ and } \mathbf{m}_{\tilde{X}\tilde{u},n} := \sum_{i=1}^{n}\mathbf{m}_{\tilde{X}\tilde{u},ni}\,.$$

$\mathbf{M}_{\tilde{X}\tilde{X},n}$, $\mathbf{m}_{\tilde{X}\tilde{u},ni}$ *and* $\mathbf{m}_{\tilde{X}\tilde{u},n}$ *are the* $T \to \infty$*-limits of* $\left[\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\mathbf{A}_{\tilde{X}p}\widetilde{\mathbf{X}}_{p;it_\star}\widetilde{\mathbf{X}}'_{p;it_\star}\mathbf{A}_{\tilde{X}p}\right]_{(1:k+1,1:k+1)} = \mathbf{M}_{\tilde{X}\tilde{X},nT}$, $\left[\sum_{t_\star=1}^{T_\star}\mathbf{A}_{\tilde{X}p}\widetilde{\mathbf{X}}_{p;it_\star}\tilde{u}_{it_\star}\right]_{(1:k+1,1)} =: \mathbf{m}_{\tilde{X}\tilde{u},nTi}$ *and* $\left[\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\mathbf{A}_{\tilde{X}p}\widetilde{\mathbf{X}}_{p;it_\star}\tilde{u}_{it_\star}\right]_{(1:k+1,1)} =: \mathbf{m}_{\tilde{X}\tilde{u},nT}$.

Proof. See Appendix A.

Since $\Gamma_{0,u_iu_i} > 0$ projecting on the leads and lags $\Delta\mathbf{x}_{it_\star+s}$, $s = -p(T),\ldots,-1,0,1,\ldots,p(T)$, is not sufficient to obtain convergence to a zero mean Gaussian mixture distribution. Comparing the *OLS* limit $\mathbf{M}^{-1}_{\tilde{X}\tilde{X},n}\mathbf{m}_{\tilde{X}\tilde{u}^\dagger,n}$ to the *DOLS* limit $\mathbf{M}^{-1}_{\tilde{X}\tilde{X},n}\mathbf{m}_{\tilde{X}\tilde{u},n}$, we observe that by projecting on the leads and lags, the second-order bias terms $\mathbf{C}\boldsymbol{\Delta}^\dagger_{vu_i}$ and $\boldsymbol{\Delta}^\dagger_{v_iu_i}$ are removed, but not the correlation term $\mathbf{W}_i\mathbf{K}\boldsymbol{\Gamma}_{0,uu_i}$ arising from the spatial lag.

To obtain convergence to a mean zero Gaussian mixture distribution, we shall estimate the model using instruments for the endogenous variable $\tilde{y}^*_{it} = \sum_{j=1}^{n} W_{ij}\tilde{y}_{jt} = \mathbf{W}_i\tilde{\mathbf{y}}_t$, for which we assume:

Assumption 4 (Valid Instruments; see Kitamura and Phillips (1997)). *The instruments $\tilde{\mathbf{z}}^*_{it} \in \mathbb{R}^{q_\rho}$ fulfill the requirements for instrumental variable estimation as stated, e.g., in Ruud (2000, Chapter 20), Phillips and Hansen (1990) and Kitamura and Phillips (1997). In particular, (i) the number of instruments is larger or equal to the number of parameters (order condition), (ii) $\frac{1}{T_\star^2}\sum_{t_\star=1}^{T_*}(\tilde{y}^*_{it_\star}, \tilde{\mathbf{x}}'_{it_\star})'((\tilde{\mathbf{z}}^*_{it_\star})', \tilde{\mathbf{x}}'_{it_\star})$ weakly converges to a matrix of rank $k+1$ (almost surely) as $T_\star \to \infty$, and (iii) $\frac{1}{T_\star^2}\sum_{t_\star=1}^{T_\star}((\tilde{\mathbf{z}}^*_{it_\star})', \tilde{\mathbf{x}}'_{it_\star})'((\tilde{\mathbf{z}}^*_{it_\star})', \tilde{\mathbf{x}}'_{it_\star})$ weakly converges to a matrix of rank $k + q_\rho$ (almost surely) as $T_\star \to \infty$.*

By following Kelejian and Prucha (1998), we base the instruments on the spatial lags of the explanatory variables. In more detail, our model can be solved as $\tilde{\mathbf{y}} = \left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho\mathbf{W})^{-1}\right]\left(\tilde{\mathbf{x}}\boldsymbol{\beta} + \tilde{\boldsymbol{\zeta}}_p\underline{\boldsymbol{\delta}}_p + \tilde{\underline{\mathbf{u}}}_p\right)$. The matrix $(\mathbf{I}_n - \rho\mathbf{W})^{-1}$ can then be expanded as $(\mathbf{I}_n - \rho\mathbf{W})^{-1} = \sum_{s=0}^{\infty}(\rho\mathbf{W})^s$ (see, e.g., Horn and Johnson, 1985, Corollary 5.6.16). This implies that variables of the form $\sum_{j=1}^n W_{ij}\tilde{x}_{jt_\star\kappa}, \sum_{j=1}^n W_{ij}^2\tilde{x}_{jt_\star\kappa}, \ldots$ are suitable instruments for $\mathbf{W}\tilde{\mathbf{y}}$. $\tilde{x}_{jt_\star\kappa}$ is the coordinate κ of $\tilde{\mathbf{x}}_{jt_\star}$, while W_{ij}^τ stands for $[\mathbf{W}^\tau]_{(i,j)}$ and $\mathbf{W}_i^\tau$ for $[\mathbf{W}^\tau]_{(i,1:n)}$, where $\tau \in \mathbb{N}$. If $\tilde{x}_{jt_\star\kappa}$ is a component specific variable, $\tilde{x}_{jt_\star\kappa}$ and $\tilde{\underline{u}}_{it_\star}$, for $i \neq j$, are independent by Assumption 2, while if $\tilde{x}_{jt_\star\kappa}$ is a common variable, then asymptotic independence will be established for the *D2SLS* estimator when $T \to \infty$. Note that these instruments have an intuitive interpretation: we instrument the W_{ij} weighted sum of the neighbors/peers $\tilde{y}_{jt_\star}$ by the W_{ij} weighted sum of the characteristics of the neighbors (their $\tilde{\mathbf{x}}_{it_\star}$ values). The higher-order spatial lags as instruments then use the characteristics of the neighbors of the neighbors, etc. Hence, we work with the instruments

$$\tilde{z}^*_{it_\star\kappa} = \tilde{x}^*_{it_\star\kappa} := \sum_{j=1}^{n} W_{it_\star}^{\tau_\kappa}\tilde{x}_{jt_\star\kappa}\ , \tag{24}$$

where $\kappa \in \mathbb{K} \subset \{1,\ldots,k\}$, $\mathbb{K}$ is an index set collecting the indices of the instruments used, and $\tau_\kappa \in \mathbb{N}$. Let $\mathbb{K}_{(l)}$ stand for the lth element of the set $\mathbb{K}$. Then, $\tilde{\mathbf{z}}^*_{it_\star} = \left(\tilde{x}^*_{it_\star\mathbb{K}_{(1)}}, \ldots, \tilde{x}^*_{it_\star\mathbb{K}_{(q_\rho)}}\right)' \in \mathbb{R}^{q_\rho}$, the exponents τ_κ are $\tau_{\mathbb{K}_{(l)}}, l = 1,\ldots,q_\rho$. By the $n\times nk$ selector matrices $\mathbf{C}_{(\mathbb{K}_{(j)})}$ (where the coefficients $\left[C_{\mathbb{K}(j)}\right]_{(\iota,(\iota-1)n+n\iota)}$ are equal for all $\iota = 1,\ldots,n$), we get $\tilde{z}^*_{it_\star\mathbb{K}_{(j)}} = \tilde{x}^*_{it_\star\mathbb{K}_{(j)}} = \mathbf{W}_i^{\tau_{\mathbb{K}(j)}}\mathbf{C}_{(\mathbb{K}_{(j)})}\tilde{\mathbf{x}}_{t_\star}$, for $j = 1,\ldots,q_\rho$. To keep the notation simple, we consider - as already stated at the beginning of Section 2 - a model with one spatial lag ($k_\rho = 1$). Hence, with $q_\rho \geq 1$ the order condition is met.

We collect the variables $\tilde{\mathbf{z}}^*_{it_\star} = \left(\mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\tilde{\mathbf{x}}_{t_\star}, \ldots, \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\tilde{\mathbf{x}}_{t_\star}\right)' \in \mathbb{R}^{q_\rho}$ in the $nT_\star \times q_\rho$ matrix $\mathbf{z}^* := \left(\tilde{\mathbf{z}}^*_{1t_\star}, \ldots, \tilde{\mathbf{z}}^*_{nt_\star}\right)'$. The set of our instruments is then $\tilde{\mathbf{Z}}_p := \left(\mathbf{z}^*, \mathbf{x}, \boldsymbol{\zeta}_p\right) \in \mathbb{R}^{T_\star n\times q_\rho+k+(2p+1)k\cdot n}$. The rows of $\tilde{\mathbf{Z}}_p \in \mathbb{R}^{nT\times q_\rho+k+(2p+1)k\cdot n}$ are the transpose of the $q_\rho+k+(2p+1)k\cdot n$-dimensional column vectors $\tilde{\mathbf{Z}}_{p;it_\star} := \left(\tilde{\mathbf{z}}^{*\prime}_{it_\star}, \tilde{\mathbf{x}}'_{it_\star}, \mathbf{0}_{(1\times(2p+1)k\cdot(i-1))}, \tilde{\boldsymbol{\zeta}}'_{p;it_\star}, \mathbf{0}_{(1\times(2p+1)k\cdot(n-i-1))}\right)'$. Next we consider *two-stage least squares* (2*SLS*) estimation. Since no projection on leads and lags is applied with 2*SLS*, we get $p = \emptyset$, $T = T_\star$ and $t = t_\star$ as well as $\tilde{\mathbf{X}}^{2SLS}_{it} := \left(\tilde{y}^*_{it}, \tilde{\mathbf{x}}'_{it}\right)' = \tilde{\mathbf{X}}^{OLS}_{it}$ and $\tilde{\mathbf{Z}}^{2SLS}_{it} := \left(\tilde{\mathbf{z}}^{*\prime}_{it}, \tilde{\mathbf{x}}'_{it}\right)'$. Collecting $\tilde{\mathbf{X}}^{2SLS}_{it}$ and $\tilde{\mathbf{Z}}^{2SLS}_{it}$ results in $\tilde{\mathbf{X}}^{2SLS}$ and $\tilde{\mathbf{Z}}^{2SLS}$. The noise term and the projection operator are given by $\tilde{u}^\dagger_{it}$ and $\mathcal{P}_{2SLS} := \tilde{\mathbf{Z}}^{2SLS}\left(\tilde{\mathbf{Z}}^{2SLS\prime}\tilde{\mathbf{Z}}^{2SLS}\right)^{-1}\tilde{\mathbf{Z}}^{2SLS\prime}$. The term $\tilde{\mathbf{x}}_{it}$ contained in $\tilde{\mathbf{Z}}^{2SLS}_{it}$, is still correlated with $\tilde{u}^\dagger_{it}$. This correlation does not vanish if $T \to \infty$. To see this, consider the two stage least squares estimator

$$\hat{\boldsymbol{\gamma}}_{2SLS} := \left(\tilde{\mathbf{X}}'_{2SLS}\mathcal{P}_{2SLS}\tilde{\mathbf{X}}'_{2SLS}\right)^{-1}\tilde{\mathbf{X}}'_{2SLS}\mathcal{P}_{2SLS}\tilde{\mathbf{y}}\ . \tag{25}$$

By scaling $\tilde{\mathbf{X}}^{2SLS}_{it}$ and $\tilde{\mathbf{Z}}^{2SLS}_{it}$ by $1/T$ we obtain:

Proposition 3. *Consider the fixed effects spatial correlation model (8) and the* 2*SLS estimator (25) based on the within-transformed model (10). Suppose that the Assumptions 1 to 4 hold. Instruments are based on (24).*

Then, for n fixed and $T \to \infty$, it follows that the $T \to \infty$-limits of $\mathbf{M}^{2SLS}_{\tilde{X}\tilde{Z},nTi} := \frac{1}{T^2}\sum_{t=1}^{T}\widetilde{\mathbf{X}}^{2SLS}_{it}\widetilde{\mathbf{Z}}^{2SLS\prime}_{it}$, $\mathbf{M}^{2SLS}_{\tilde{X}\tilde{Z},nT} := \sum_{i=1}^{n}\mathbf{M}^{2SLS}_{\tilde{X}\tilde{Z},ni}$, $\mathbf{M}^{2SLS}_{\tilde{Z}\tilde{Z},nTi} := \frac{1}{T^2}\sum_{t=1}^{T}\widetilde{\mathbf{Z}}^{2SLS}_{it}\widetilde{\mathbf{Z}}^{2SLS\prime}_{it}$, $\mathbf{M}^{2SLS}_{\tilde{Z}\tilde{Z},nT} := \sum_{i=1}^{n}\mathbf{M}^{2SLS}_{\tilde{Z}\tilde{Z},ni}$, $\mathbf{m}^{2SLS}_{\tilde{Z}\tilde{u},nTi} := \frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{Z}}^{2SLS}_{it}\tilde{u}^{\dagger}_{it}$, *and* $\mathbf{m}^{2SLS}_{\tilde{Z}\tilde{u},nT} := \sum_{i=1}^{n}\mathbf{m}^{2SLS}_{\tilde{Z}\tilde{u},nTi}$ *are provided by*

$$\mathbf{M}_{\tilde{X}\tilde{Z},ni} := \int_0^1 \mathbf{g}_i(r)\mathbf{h}_i(r)'dr\,,\ \ \mathbf{M}_{\tilde{X}\tilde{Z},n} := \sum_{i=1}^{n}\mathbf{M}_{\tilde{X}\tilde{Z},ni}\,,\ \ \textit{where}$$

$$\mathbf{h}_i(r) := \left(\mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\tilde{\mathcal{B}}_v(r),\ldots,\mathbf{W}_i^{\tau_{\mathbb{K}(j)}}\mathbf{C}_{(\mathbb{K}_{(j)})}\tilde{\mathcal{B}}_v(r),\ldots,\mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\tilde{\mathcal{B}}_v(r),\tilde{\mathcal{B}}_{v_i}(r)'\right)' \in \mathbb{R}^{q_\rho+k}\,,$$

$$\mathbf{M}_{\tilde{Z}\tilde{Z},ni} := \int_0^1 \mathbf{h}_i(r)\mathbf{h}_i(r)'dr\,,\ \mathbf{M}_{\tilde{Z}\tilde{Z},n} := \sum_{i=1}^{n}\mathbf{M}_{\tilde{Z}\tilde{Z},ni}\ \textit{and}$$

$$\mathbf{m}_{\tilde{Z}\tilde{u}^{\dagger},ni} := \begin{pmatrix} \mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\int_0^1\mathbf{C}_{(\mathbb{K}_{(1)})}\tilde{\mathcal{B}}_v(r)d\mathcal{B}^{\dagger}_{u_i}(r)\\ \vdots \\ \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\int_0^1\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\tilde{\mathcal{B}}_v(r)d\mathcal{B}^{\dagger}_{u_i}(r)\\ \sqrt{\Omega^{\dagger}_{u_iu_i}}\int_0^1\tilde{\mathcal{B}}_{v_i}(r)d\mathcal{W}^{\dagger}_{u_i}(r)\end{pmatrix} + \begin{pmatrix}\mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\mathbf{\Delta}^{\dagger}_{vu_i}\\ \vdots \\ \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\mathbf{\Delta}^{\dagger}_{vu_i}\\ \mathbf{\Delta}^{\dagger}_{v_iu_i}\end{pmatrix},$$

$$\textit{as well as}\ \mathbf{m}_{\tilde{Z}\tilde{u}^{\dagger},n} := \sum_{i=1}^{n}\mathbf{m}_{\tilde{Z}\tilde{u}^{\dagger},ni}\,. \tag{26}$$

The asymptotic limit distributed of the centered and scaled 2SLS estimator of $\boldsymbol{\gamma}$ is provided by:

$$T\left(\hat{\boldsymbol{\gamma}}_{2SLS}-\boldsymbol{\gamma}\right) \Rightarrow \left(\mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{M}'_{\tilde{X}\tilde{Z},n}\right)^{-1}\mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{m}_{\tilde{Z}\tilde{u}^{\dagger},n}\,. \tag{27}$$

Proof. See Appendix A.

Note that 2*SLS* eliminates the correlation term $\sum_{i=1}^{n}\mathbf{W}_i\mathbf{K}\mathbf{\Gamma}^{\dagger}_{0,uu_i}$ in $\mathbf{m}_{\tilde{X}\tilde{u}^{\dagger},ni}$, while the second-order bias arising from serial correlation is still present. We do not attain convergence to a zero mean Gaussian mixture distribution.

We now construct a two stage-least square procedure for our panel setting where leads and lags of $\Delta\tilde{\mathbf{x}}_{it}$ as well as instruments $\tilde{\mathbf{z}}^{*}_{it}$ are included. Let us define the *projection operator* $\mathscr{P}_{Hp}$ projecting on the column space spanned by $\widetilde{\mathbf{Z}}_p$ (see, e.g., Ruud, 2000, Chapter 3). In formal terms $\mathscr{P}_{Hp} := \widetilde{\mathbf{Z}}_p\left(\widetilde{\mathbf{Z}}'_p\widetilde{\mathbf{Z}}_p\right)^{-1}\widetilde{\mathbf{Z}}_p$. Since $\widetilde{\mathbf{Z}}_p$ is a $T_\star n \times q_\rho + k + (2p+1)k\cdot n$ matrix, $\mathscr{P}_{Hp}$ has to be a $T_\star n \times T_\star n$ matrix. The *dynamic two-stage least squares* estimator of $\boldsymbol{\theta}_p = (\rho,\boldsymbol{\beta}',\underline{\boldsymbol{\delta}}'_p)' = (\boldsymbol{\gamma}',\underline{\boldsymbol{\delta}}'_p)'$ is defined as follows:

$$\widehat{\boldsymbol{\theta}}_{D2SLS;p} := \left(\widetilde{\mathbf{X}}'_p\mathscr{P}_{Hp}\widetilde{\mathbf{X}}_p\right)^{-1}\widetilde{\mathbf{X}}'_p\mathscr{P}_{Hp}\tilde{\mathbf{y}} = (\boldsymbol{\gamma}',\underline{\boldsymbol{\delta}}'_p)' + \left(\widetilde{\mathbf{X}}'_p\mathscr{P}_{Hp}\widetilde{\mathbf{X}}_p\right)^{-1}\widetilde{\mathbf{X}}'_p\mathscr{P}_{Hp}\underline{\tilde{\mathbf{u}}}_p\,. \tag{28}$$

Let $\widetilde{\mathbf{X}}_{it_\star,(1:k+1)}$ and $\widetilde{\mathbf{Z}}_{it_\star,(1:k+q_\rho)}$ stand for the first $k+1$ and $k+q_\rho$ elements of $\widetilde{\mathbf{X}}_{p;it_\star}$ and $\widetilde{\mathbf{Z}}_{p;it_\star}$. For the asymptotic analysis we apply the scaling matrix $\mathbf{A}_{\tilde{X}p}$ as well as the scaling matrix $\mathbf{A}_{\tilde{Z}p} := diag\left(T_\star^{-1}\cdot\mathbf{I}_{k+q_\rho},T_\star^{-0.5}\cdot\mathbf{I}_{(2p+1)nk}\right) \in \mathbb{R}^{k+q_\rho+(2p+1)nk\times k+q_\rho+(2p+1)nk}$. The matrix $\mathbf{A}_{\tilde{X}p}$ is diagonal with $1/T_\star$ in the first $k+1$ elements while the matrix $\mathbf{A}_{\tilde{Z}p}$ is diagonal with $1/T_\star$ in the first $k+q_\rho$ elements. The other scaling factors, i.e. all the further elements on the main diagonals of $\mathbf{A}_{\tilde{X}p}$ and $\mathbf{A}_{\tilde{Z}p}$ are $T_\star^{-0.5}$. Let $\mathbf{M}^{\star}_{\tilde{X}\tilde{X},nTi} := \sum_{t_\star=1}^{T_\star}\mathbf{A}_{\tilde{X}p}\widetilde{\mathbf{X}}_{it_\star}\widetilde{\mathbf{X}}'_{it_\star}\mathbf{A}_{\tilde{X}p}$, $\mathbf{M}^{\star}_{\tilde{X}\tilde{X},nT} := \sum_{i=1}^{n}\mathbf{M}^{\star}_{\tilde{X}\tilde{X},nTi}$, $\mathbf{M}^{\star}_{\tilde{X}\tilde{Z},nTi} := \sum_{t_\star=1}^{T_\star}\mathbf{A}_{\tilde{X}p}\widetilde{\mathbf{X}}_{it_\star}\widetilde{\mathbf{Z}}'_{it_\star}\mathbf{A}_{\tilde{Z}p}$, $\mathbf{M}^{\star}_{\tilde{X}\tilde{Z},nT} := \sum_{i=1}^{n}\mathbf{M}^{\star}_{\tilde{X}\tilde{Z},nTi}$, $\mathbf{M}^{\star}_{\tilde{Z}\tilde{Z},nTi} := \sum_{t_\star=1}^{T_\star}\mathbf{A}_{\tilde{Z}p}\widetilde{\mathbf{Z}}_{it_\star,}\widetilde{\mathbf{Z}}'_{it_\star,}\mathbf{A}_{\tilde{Z}p}$, $\mathbf{M}^{\star}_{\tilde{X}\tilde{Z},nT} := \sum_{i=1}^{n}\mathbf{M}^{\star}_{\tilde{X}\tilde{Z},nTi}$, $\mathbf{m}^{\star}_{\tilde{Z}\tilde{u},nTi} :=$

$\sum_{t_\star=1}^{T_\star} \mathbf{A}_{\tilde{Z}p}\widetilde{\mathbf{Z}}_{it_\star,(1:k+1)}\underline{\tilde{\mathbf{u}}}_{it_\star}$, and $\mathbf{m}^\star_{Zu,nT} := \sum_{i=1}^{n} \mathbf{m}^\star_{\tilde{Z}\tilde{u},nTi}$, as well as

$$\mathbf{M}_{\tilde{X}\tilde{X},nTi} := \frac{1}{T_\star^2}\sum_{t_\star=1}^{T_\star}\left(\widetilde{\mathbf{X}}_{it_\star,(1:k+1)}\widetilde{\mathbf{X}}'_{it_\star,(1:k+1)}\right) = \left[\mathbf{M}^\star_{\tilde{X}\tilde{X},nTi}\right]_{(1:k+1,1:k+1)} ,$$

$$\mathbf{M}_{\tilde{X}\tilde{X},nT} := \sum_{i=1}^{n}\mathbf{M}_{\tilde{X}\tilde{X},nTi} \in \mathbb{R}^{k+1\times k+1} ,$$

$$\mathbf{M}_{\tilde{X}\tilde{Z},nTi} := \frac{1}{T_\star^2}\sum_{t_\star=1}^{T_\star}\left(\widetilde{\mathbf{X}}_{it_\star,(1:k+1)}\widetilde{\mathbf{Z}}'_{it_\star,(1:k+q_\rho)}\right) = \left[\mathbf{M}^\star_{\tilde{X}\tilde{Z},nTi}\right]_{(1:k+1,1:k+q_\rho)} ,$$

$$\mathbf{M}_{\tilde{X}\tilde{Z},nT} := \sum_{i=1}^{n}\mathbf{M}_{\tilde{X}\tilde{Z},nTi} \in \mathbb{R}^{k+1\times k+q_\rho} ,$$

$$\mathbf{M}_{\tilde{Z}\tilde{Z},nTi} := \frac{1}{T_\star^2}\sum_{t_\star=1}^{T_\star}\left(\widetilde{\mathbf{Z}}_{it_\star,,(1:k+q_\rho)}\widetilde{\mathbf{Z}}'_{it_\star,,(1:k+q_\rho)}\right) = \left[\mathbf{M}^\star_{\tilde{Z}\tilde{Z},nTi}\right]_{(1:k+q_\rho,1:k+q_\rho)} ,$$

$$\mathbf{M}_{\tilde{Z}\tilde{Z},nT} := \sum_{i=1}^{n}\mathbf{M}_{\tilde{Z}\tilde{Z},nTi} \in \mathbb{R}^{k+q_\rho\times k+q_\rho} ,$$

$$\mathbf{m}_{\tilde{Z}\tilde{u},nTi} := \frac{1}{T_\star}\sum_{t_\star=1}^{T_\star}\left(\widetilde{\mathbf{Z}}_{it_\star,(1:k+1)}\underline{\tilde{\mathbf{u}}}_{it_\star}\right) = \left[\mathbf{m}^\star_{\tilde{Z}\tilde{u},nTi}\right]_{(1:k+q_\rho,1)} \quad \text{and}$$

$$\mathbf{m}_{\tilde{Z}\tilde{u},nT} := \sum_{i=1}^{n}\mathbf{m}_{\tilde{Z}\tilde{u},nTi} = \left[\mathbf{m}^\star_{\tilde{Z}\tilde{u},nT}\right]_{(1:k+q_\rho,1)} \in \mathbb{R}^{k+q_\rho} .$$

Next, we summarize the large sample properties of the *D2SLS* estimator:

Theorem 1 ($T \to \infty$ limits for *D2SLS* Estimation). *Consider the fixed effects spatial correlation model (8) and the D2SLS estimator (28) based on the within-transformed model (10). Suppose that Assumptions 1–4 hold. Let $T_\star = T - 2p(T)$. Then, for n fixed and $T \to \infty$, it follows that*

1. *$T_\star(\hat{\boldsymbol{\gamma}}_{D2SLS;p} - \boldsymbol{\gamma})$ and $\sqrt{T_\star}(\hat{\underline{\boldsymbol{\delta}}}_{D2SLS;p} - \underline{\boldsymbol{\delta}}_p)$ are asymptotically independent.*
2. *$\mathbf{M}_{\tilde{X}\tilde{X},nTi}$, $\mathbf{M}_{\tilde{X}\tilde{X},nT}$, $\mathbf{M}_{\tilde{Z}\tilde{Z},nTi}$, $\mathbf{M}_{\tilde{Z}\tilde{Z},nT}$, $\mathbf{M}_{\tilde{X}\tilde{Z},nT}$, $\mathbf{M}_{\tilde{X}\tilde{X},nT}$, $\mathbf{m}_{\tilde{Z}\tilde{u},nTi}$, and $\mathbf{m}_{\tilde{Z}\tilde{u},nT}$ converge weakly to $\mathbf{M}_{\tilde{X}\tilde{X},ni}$, $\mathbf{M}_{\tilde{X}\tilde{X},n}$, $\mathbf{M}_{\tilde{Z}\tilde{Z},ni}$, $\mathbf{M}_{\tilde{Z}\tilde{Z},n}$, $\mathbf{M}_{\tilde{X}\tilde{Z},n}$, $\mathbf{M}_{\tilde{X}\tilde{X},n}$, $\mathbf{m}_{\tilde{Z}\tilde{u},ni}$, and $\mathbf{m}_{\tilde{Z}\tilde{u},n}$, where*

$$\mathbf{m}_{\tilde{Z}\tilde{u},ni} := \int_0^1 \mathbf{h}_i(r)d\mathcal{B}_{u_i}(r) = \sqrt{\Omega_{u_iu_i}}\int_0^1 \mathbf{h}_i(r)d\mathcal{W}_{u_i}(r) \quad \text{and} \quad \mathbf{m}_{\tilde{Z}\tilde{u},n} := \sum_{i=1}^{n}\mathbf{m}_{\tilde{Z}\tilde{u},ni} . \tag{29}$$

$\mathbf{M}_{\tilde{X}\tilde{X},n}$ and $\mathbf{M}_{\tilde{X}\tilde{X},ni}$ are provided in (21), while $\mathbf{M}_{\tilde{X}\tilde{Z},ni}$, $\mathbf{M}_{\tilde{X}\tilde{Z},n}$, $\mathbf{M}_{\tilde{Z}\tilde{Z},n}$, and $\mathbf{M}_{\tilde{Z}\tilde{Z},ni}$ are provided in (26). In addition, $T_\star\left(\hat{\boldsymbol{\gamma}}_{D2SLS;p} - \boldsymbol{\gamma}\right)$ converges weakly to $\mathbf{M}_n^{-1}\mathbf{m}_n$, where

$$\mathbf{M}_n := \mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{M}'_{\tilde{X}\tilde{Z},n} \quad \text{and} \quad \mathbf{m}_n := \sum_{i=1}^{n}\mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{m}_{\tilde{Z}\tilde{u},ni} = \mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{m}_{\tilde{Z}\tilde{u},n} . \tag{30}$$

3. *Suppose that $\hat{\boldsymbol{\Omega}}_{uu}$ is a consistent estimator of $\boldsymbol{\Omega}_{uu} = diag(\Omega_{u_iu_i})_{i=1,\dots,n}$, then*

$$\begin{aligned}\mathbf{V}_{nT} &:= \left[\mathbf{M}_{\tilde{X}\tilde{Z},nT}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},nT}\mathbf{M}'_{\tilde{X}\tilde{Z},nT}\right]^{-1}\mathbf{D}_{nT}\left[\mathbf{M}_{\tilde{X}\tilde{Z},nT}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},nT}\mathbf{M}'_{\tilde{X}\tilde{Z},nT}\right]^{-1} \\ &\Rightarrow \left[\mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{M}'_{\tilde{X}\tilde{Z},n}\right]^{-1}\mathbf{D}_n\left[\mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{M}'_{\tilde{X}\tilde{Z},n}\right]^{-1} =: \mathbf{V}_n , \quad \textit{where}\end{aligned}$$

$$\mathbf{D}_{nT} := \mathbf{M}_{\tilde{X}\tilde{Z},nT}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},nT}\left(\sum_{i=1}^{n}\hat{\Omega}_{u_iu_i}\mathbf{M}_{\tilde{Z}\tilde{Z},nTi}\right)\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},nT}\mathbf{M}'_{\tilde{X}\tilde{Z},nT}\text{, and}$$

$$\mathbf{D}_{n} := \mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\left(\sum_{i=1}^{n}\Omega_{u_iu_i}\mathbf{M}_{\tilde{Z}\tilde{Z},ni}\right)\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{M}'_{\tilde{X}\tilde{Z},n}\,.$$

Given an $\mathtt{s} \times (k+1)$ *restriction matrix* $\mathbf{R}$, *the Wald type statistic*

$$\mathscr{W}_{\gamma,nT} = \left(T_\star\mathbf{R}\left(\widehat{\boldsymbol{\gamma}}_{D2SLS;p} - \boldsymbol{\gamma}\right)\right)'\left(\mathbf{R}\mathbf{V}_{nT}\mathbf{R}'\right)^{-1}\left(T_\star\mathbf{R}\left(\widehat{\boldsymbol{\gamma}}_{D2SLS;p} - \boldsymbol{\gamma}\right)\right), \tag{31}$$

converges in distribution to a χ^2 *random variable with* $\mathtt{s}$ *degrees of freedom.*

Proof. See Appendix A.

Remark 2. Since the signal to noise ratio goes to infinity, we observe that the *OLS*, the *DOLS*, the 2*SLS*, and the *D*2*SLS* estimator is consistent, when considering $T \to \infty$-limits. Sufficient conditions for consistent estimation of the covariance matrix $\boldsymbol{\Omega}_{uu}$ are discussed in Jansson (2002) and in Online Appendix A-6.

In addition, observe that if $\boldsymbol{\beta} = \mathbf{0}_{(k\times 1)}$ or $k = 0$, the variable y_{it} becomes $I(0)$ (see, e.g., Eqs. (1) and (7)). In this case the signal to noise ratio does not go to infinity and the ordinary least squares estimator is not consistent (for a proof see Sögner and Wagner, 2017).

Remark 3. Projecting on *all* leads and lags as proposed in system-*DOLS* (see Park and Ogaki, 1991) and the *DSUR* approach (see Mark et al., 2005) does not eliminate this bias. Based on Mark et al. (2005), *D2SLS* can be augmented to a richer correlation structure by projecting on the leads and lags of all regressors as described in (16). However, by this projection facility the dimension of the nuisance parameter becomes $(2p+1)\cdot(k_In^2 + k_Cn)$. Due to numerical constraints, this estimator can hardly be implemented when n is large (see Section A-4 in the Online Appendix and Section 4).

Remark 4. The following subsection also derives limits when both n and T become large. For the model (10), where $\tilde{\mathbf{x}}_{Ct}$ and $\tilde{\mathbf{x}}_{Iit}$ are used as regressors, we observe that $\frac{1}{nT_\star^2}\sum_{j=1}^{n}\sum_{t_\star=1}^{T_\star}\tilde{\mathbf{x}}_{Ct}\tilde{\mathbf{x}}'_{Ct} = \frac{1}{T_\star^2}\sum_{t_\star=1}^{T_\star}\tilde{\mathbf{x}}_{Ct}\tilde{\mathbf{x}}'_{Ct}$ such that the $T \to \infty$ limit as well as the $(n,T) \to \infty$ limit remain random variables. By this fact, the $(n,T) \to \infty$ limit of the centered and scaled parameter vector is not a mean zero normal vector if $k_C > 0$.[6]

3.3. Large sample properties of the D2SLS parameter estimator for model (13)

Using the within-transform (14) and the projection facility (16), we get by $\breve{\underline{\boldsymbol{\delta}}}_{\sharp p;i} := \left(\underline{\boldsymbol{\delta}}_{\sharp p;j} - \frac{1}{n}\sum_{j=1}^{n}\underline{\boldsymbol{\delta}}_{\sharp p;j}\right)$ and some algebra (see Online Appendix A-3)[7]

$$\breve{y}_{it} = \rho\breve{y}^*_{it_\star} + \boldsymbol{\beta}'_I\breve{\mathbf{x}}_{Iit_\star} + \breve{\underline{\boldsymbol{\delta}}}'_{\sharp p;i}\tilde{\boldsymbol{\zeta}}_{\sharp p;t_\star} + \breve{\underline{u}}_{p;it_\star} = \mathbf{W}_i\mathbf{K}\left(\tilde{\boldsymbol{\beta}}_I\breve{\mathbf{x}}_{t_\star} + \left(\mathbf{I}_n \otimes \tilde{\boldsymbol{\zeta}}'_{\sharp p;t_\star}\right)\breve{\underline{\boldsymbol{\delta}}}_{\sharp p} + \breve{\underline{\mathbf{u}}}_{t_\star}\right), \tag{32}$$

[6] In a former version we obtained $(n,T) \to \infty$-limits for model with locally common variables, in which case the joint limit is a normally distributed zero mean random vector. These results are available on request.

[7] To simplify the notation we write $\breve{\underline{\mathbf{u}}}_{it_\star}$ and $\breve{\underline{\mathbf{u}}}_{t_\star}$ when the within-transform (14) is applied to $\underline{\mathbf{u}}_{\sharp it_\star}$ and $\underline{\mathbf{u}}_{\sharp t_\star}$ (i.e., we skip the symbol $\sharp$). In addition, although $\underline{\boldsymbol{\delta}}_{\sharp p;j} - \frac{1}{n}\sum_{j=1}^{n}\underline{\boldsymbol{\delta}}_{\sharp p;j}$ corresponds to the within-transform defined in (12) we use the abbreviation $\breve{\underline{\boldsymbol{\delta}}}_{\sharp p;i}$ to emphasize that this parameter is related to the model based on the within-transform (14). To obtain (32) note that with $k_C = 0$, we have $\boldsymbol{\beta} = \boldsymbol{\beta}_I$, $\tilde{\boldsymbol{\beta}}_I = \tilde{\boldsymbol{\beta}} = \mathbf{I}_n \otimes \boldsymbol{\beta}_I \in \mathbb{R}^{n\times k_I n}$, and $\mathbf{C} = \mathbf{I}_{nk_I}$ such that $\mathbf{W}_i\mathbf{K}\left(\tilde{\boldsymbol{\beta}}\mathbf{C}\breve{\mathbf{x}}_{t_\star} + \left(\mathbf{I}_n \otimes \tilde{\boldsymbol{\zeta}}'_{\sharp p;t_\star}\right) + \breve{\underline{\mathbf{u}}}_{t_\star}\right) = \mathbf{K}\left(\tilde{\boldsymbol{\beta}}_I\breve{\mathbf{x}}_{t_\star} + \left(\mathbf{I}_n \otimes \tilde{\boldsymbol{\zeta}}'_{\sharp p;t_\star}\right) + \breve{\underline{\mathbf{u}}}_{t_\star}\right)$.

where $\breve{\mathbf{x}}_{t_\star} := \breve{\mathbf{x}}_{It_\star} := \left(\breve{\mathbf{x}}_{I1t_\star}, \ldots, \breve{\mathbf{x}}_{Int_\star}\right)' \in \mathbb{R}^{nk_I \times 1}$, $\breve{\underline{\mathbf{u}}}_{t_\star} := \left(\breve{\underline{u}}_{1t_\star}, \ldots, \breve{\underline{u}}_{nt_\star}\right)' \in \mathbb{R}^{n\times 1}$. The vector $\tilde{\underline{\boldsymbol{\delta}}}'_{\sharp p} \in \mathbb{R}^{(2p+1)k_I n^2}$ and the matrix $\mathbf{I}_n \otimes \tilde{\boldsymbol{\zeta}}'_{\sharp p;t_\star} \in \mathbb{R}^{n\times(2p+1)k_I n^2}$ collect the projection parameters $\underline{\boldsymbol{\delta}}_{\sharp p;i}$ and the corresponding leads and lags of $\Delta\tilde{\mathbf{x}}_{Iit_\star}$, while $\mathbf{W}_i$ is the ith row of the $n \times n$ matrix $\mathbf{W}$, describing the impact by all other $n-1$ cross-sections on y_{it} (as well as $\breve{y}_{it_\star}$). In addition, let $\breve{\mathbf{X}}_{p;it_\star} := \left(\breve{y}^*_{it_\star}, \breve{\mathbf{x}}'_{it_\star}, \mathbf{0}_{(1\times(2p+1)k_I\cdot(i-1))}, \tilde{\boldsymbol{\zeta}}'_{\sharp p;it_\star}, \mathbf{0}_{(1\times(2p+1)k_I\cdot(N-i-1))}\right)' \in \mathbb{R}^{1+k_I+(2p+1)k_I n^2}$ and $\breve{\mathbf{Z}}_{p;it_\star} := \left(\breve{\mathbf{z}}^{*\prime}_{it_\star}, \breve{\mathbf{x}}'_{it_\star}, \mathbf{0}_{(1\times(2p+1)k_I\cdot(i-1))}, \tilde{\boldsymbol{\zeta}}'_{\sharp p;it_\star}, \mathbf{0}_{(1\times(2p+1)k_I\cdot(N-i-1))}\right)' \in \mathbb{R}^{q_\rho+k_I+(2p+1)k_I n^2}$, where $\breve{\mathbf{z}}^*_{it_\star} = \left(\breve{z}_{it_\star\mathbb{K}_{(1)}}, \ldots, \breve{z}^*_{it_\star\mathbb{K}_{(q_\rho)}}\right)' \in \mathbb{R}^{q_\rho}$. The suffix $\breve{}$ denotes that the within-transformation – described in (14) – is applied to the instruments $z^*_{it_\star}$ defined in (24). Then, $\breve{y}_{it} = \breve{\mathbf{X}}'_{p;it_\star}\breve{\boldsymbol{\theta}}_p + \breve{\underline{u}}_{p;it_\star}$, where $\breve{\boldsymbol{\theta}}_p := \left(\breve{\gamma}', \breve{\underline{\boldsymbol{\delta}}}'_{\sharp p}\right)'$, $\breve{\boldsymbol{\gamma}} := \left(\rho, \boldsymbol{\beta}'_I\right)'$ and $\breve{\underline{\boldsymbol{\delta}}}_{\sharp p} := \left(\breve{\underline{\boldsymbol{\delta}}}'_{\sharp p;i}, \ldots, \breve{\underline{\boldsymbol{\delta}}}'_{\sharp p;i}, \ldots, \breve{\underline{\boldsymbol{\delta}}}'_{\sharp p;n}\right)' \in \mathbb{R}^{(2p+1)n^2 k_I}$. By collecting $\breve{\mathbf{X}}_{it_\star}$ and $\breve{\mathbf{Z}}_{it_\star}$, we obtain the matrices $\breve{\mathbf{X}} \in \mathbb{R}^{nT_\star\times 1+k_I+(2p+1)n^2k_I}$ and $\breve{\mathbf{Z}} \in \mathbb{R}^{nT_\star\times q_\rho+k_I+(2p+1)n^2k_I}$. Finally, let $\breve{\mathbf{y}} := \left(\breve{y}_{1p+1}, \ldots, \breve{y}_{nT-p}\right)' \in \mathbb{R}^{nT_\star}$, $\breve{\underline{\mathbf{u}}}_p := \left(\breve{\underline{u}}_{1p+1}, \ldots, \breve{\underline{u}}_{nT-p}\right)' \in \mathbb{R}^{nT_\star}$ and $\mathscr{P}_{\breve{H}p} := \breve{\mathbf{Z}}_p\left(\breve{\mathbf{Z}}'_p\breve{\mathbf{Z}}_p\right)^{-1}\breve{\mathbf{Z}}_p$. From (32) we deduce the dynamic two-stage least squares estimator of $\breve{\boldsymbol{\theta}}_p$:

$$\widehat{\breve{\boldsymbol{\theta}}}_{D2SLS;p} = \left(\widehat{\breve{\boldsymbol{\gamma}}}'_{D2SLS;p}, \widehat{\breve{\underline{\boldsymbol{\delta}}}}'_{D2SLS;\sharp p}\right)' := \left(\breve{\mathbf{X}}'_p\mathscr{P}_{\breve{H}p}\breve{\mathbf{X}}_p\right)^{-1}\breve{\mathbf{X}}'_p\mathscr{P}_{\breve{H}p}\breve{\mathbf{y}} = \breve{\boldsymbol{\theta}}_p + \left(\breve{\mathbf{X}}'_p\mathscr{P}_{\breve{H}p}\breve{\mathbf{X}}_p\right)^{-1}\breve{\mathbf{X}}'_p\mathscr{P}_{\breve{H}p}\breve{\underline{\mathbf{u}}}_p\,. \tag{33}$$

To obtain the $T\to\infty$ asymptotic limit distribution of the estimator $\widehat{\breve{\boldsymbol{\theta}}}_{D2SLS;p}$, we apply the scaling matrices $\mathbf{A}_{\breve{X}p} := diag\left(T_\star^{-1}\cdot\mathbf{I}_{k_I+k_C+1}, T_\star^{-0.5}\cdot\mathbf{I}_{(2p+1)(n^2k_I+k_Cn)}\right) \in \mathbb{R}^{k_I+k_C+1+(2p+1)(n^2k_I+nk_C)\times k_I+k_C+1+(2p+1)(n^2k_I+nk_C)}$ and $\breve{\mathbf{A}}_{\breve{Z}p} := diag\left(T_\star^{-1}\cdot\mathbf{I}_{k_I+k_C+q_\rho}, T_\star^{-0.5}\cdot\mathbf{I}_{(2p+1)(n^2k_I+nk_C)}\right) \in \mathbb{R}^{k_I+k_C+q_\rho+(2p+1)(n^2k_I+nk_C)\times k_I+k_C+q_\rho+(2p+1)(n^2k_I+k_Cn)}$; (note that $k_C = 0$ in this subsection). That is, the terms arising from the spatial lag and the I(1) components are scaled by $1/T_\star$, while all projection variables are scaled by $1/\sqrt{T_\star}$. Given these scaling factors, we define $\mathbf{M}^\star_{\breve{X}\breve{Z},nTi} := \sum_{t_\star=1}^{T_\star}\mathbf{A}_{\breve{X}p}\breve{\mathbf{X}}_{p;it_\star}\breve{\mathbf{Z}}'_{p;it_\star}\mathbf{A}_{\breve{Z}p}$, $\mathbf{M}^\star_{\breve{Z}\breve{Z},nTi} := \sum_{t_\star=1}^{T_\star}\mathbf{A}_{\breve{Z}p}\breve{\mathbf{Z}}_{p;it_\star}\breve{\mathbf{Z}}'_{p;it_\star}\mathbf{A}_{\breve{Z}p}$, $\mathbf{M}_{\breve{X}\breve{Z},nTi} := \left[\mathbf{M}^\star_{\breve{X}\breve{Z},nTi}\right]_{(1:1+k_I,1:q_\rho+k_I)}$, $\mathbf{M}_{\breve{Z}\breve{Z},nTi} := \left[\mathbf{M}^\star_{\breve{Z}\breve{Z},nTi}\right]_{(1:q_\rho+k_I,1:q_\rho+k_I)}$, $\mathbf{m}^\star_{\breve{Z}\breve{u},nTi} := \sum_{t_\star=1}^{T_\star}\mathbf{A}_{\breve{Z}p}\breve{\mathbf{Z}}_{p;it_\star}\breve{\underline{u}}_{p;it_\star}$, and $\mathbf{m}_{\breve{Z}\breve{u},nTi} := \left[\mathbf{m}^\star_{\breve{Z}\breve{u},nTi}\right]_{(1:q_\rho+k_I)}$, as well as

$$\breve{\mathbf{g}}_i(r) := \begin{pmatrix}\mathbf{W}_i\mathbf{K}\tilde{\boldsymbol{\beta}}\mathbf{C}\breve{\mathcal{B}}_v(r)\\ \breve{\mathcal{B}}_{v_i}(r)\end{pmatrix} = \begin{pmatrix}\mathbf{W}_i\mathbf{K}\tilde{\boldsymbol{\beta}}_I\breve{\mathcal{B}}_v(r)\\ \breve{\mathcal{B}}_{v_i}(r)\end{pmatrix} \in \mathbb{R}^{k_I+1}\,,$$

$$\breve{\mathbf{h}}_i(r) := \begin{pmatrix}\mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\breve{\mathcal{B}}_v(r)\\ \vdots\\ \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\breve{\mathcal{B}}_v(r)\\ \breve{\mathcal{B}}_{v_i}(r)'\end{pmatrix} \in \mathbb{R}^{q_\rho+k_I}\,,\ \text{and}$$

$$\mathbf{m}_{\breve{Z}\breve{u},ni} := \begin{pmatrix}\int_0^1\mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\left[\int_0^1\breve{\mathcal{B}}_v(r)\left(d\mathcal{B}_{u_i}(r)-\frac{1}{n}\sum_{j=1}^n d\mathcal{B}_{u_j}(r)\right)\right]\\ \vdots\\ \int_0^1\mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\left[\int_0^1\breve{\mathcal{B}}_v(r)\left(d\mathcal{B}_{u_i}(r)-\frac{1}{n}\sum_{j=1}^n d\mathcal{B}_{u_j}(r)\right)\right]\\ \int_0^1\breve{\mathcal{B}}_{v_i}(r)\left(d\mathcal{B}_{u_i}(r)-\frac{1}{n}\sum_{j=1}^n d\mathcal{B}_{u_j}(r)\right)\end{pmatrix}, \tag{34}$$

where $\mathbf{m}_{\breve{Z}\breve{u},ni} \in \mathbb{R}^{q_\rho+k_I}$, $\breve{\mathcal{B}}_v := \left(\breve{\mathcal{B}}'_{v_1}, \ldots, \breve{\mathcal{B}}'_{v_i}, \ldots, \breve{\mathcal{B}}'_{v_n}\right)' \in \mathbb{R}^{nk_I}$, and $\breve{\mathcal{B}}_{v_i} = \mathcal{B}_{v_i} - \int_0^1\mathcal{B}_{v_i}(r)dr - \frac{1}{n}\sum_{j=1}^n\mathcal{B}_{v_j} + \frac{1}{n}\sum_{j=1}^n\int_0^1\mathcal{B}_{v_j}(r)dr$. Since $k_C = 0$, $\mathcal{B}_{v_i} = \mathcal{B}_{vI_i}$, $\breve{\mathcal{B}}_{v_i} = \breve{\mathcal{B}}_{vI_i}$, and $\breve{\mathcal{B}}_v = \breve{\mathcal{B}}_{vI}$. In addition, let $\mathbf{N}_{\breve{h}} := \frac{1}{n}\mathbf{1}_{(1\times n)}\otimes\mathbf{I}_{q_\rho+k_I} \in \mathbb{R}^{(q_\rho+k_I)\times(q_\rho+k_I)n}$, $\breve{\mathbf{h}}(r) := \big(\breve{\mathbf{h}}_1(r)', \ldots, \breve{\mathbf{h}}_i(r)', \ldots,$

$\check{\mathbf{h}}_n(r)')' \in \mathbb{R}^{(q_\rho+k_l)n}$, $\mathbf{M}^{\Sigma}_{\check{Z}\sum\check{Z},nTi} := \sum_{t_\star=1}^{T_\star}\left([\mathbf{A}_{\check{Z}p}\check{\mathbf{Z}}_{p;it_\star}]_{(1:q_\rho+k_k,1)}\right)\left(\frac{1}{n}\sum_{l=1}^{n}[\mathbf{A}_{\check{Z}p}\check{\mathbf{Z}}_{p;lt_\star}]_{(1:q_\rho+k_k,1)}\right)'$, $\mathbf{M}^{\Sigma}_{\sum\check{Z}\check{Z},nTi} := \sum_{t_\star=1}^{T_\star}\left(\frac{1}{n}\sum_{l=1}^{n}[\mathbf{A}_{\check{Z}p}\check{\mathbf{Z}}_{p;lt_\star}]_{(1:q_\rho+k_k,1)}\right)\left([\mathbf{A}_{\check{Z}p}\check{\mathbf{Z}}_{p;it_\star}]_{(1:q_\rho+k_k,1)}\right)'$, and $\mathbf{M}^{\Sigma}_{\sum\check{Z}\sum\check{Z},nTi} := \sum_{t_\star=1}^{T_\star}\left(\frac{1}{n}\sum_{l=1}^{n}[\mathbf{A}_{\check{Z}p}\check{\mathbf{Z}}_{p;lt_\star}]_{(1:q_\rho+k_k,1)}\right)\left(\frac{1}{n}\sum_{j=1}^{n}[\mathbf{A}_{\check{Z}p}\check{\mathbf{Z}}_{p;jt_\star}]_{(1:q_\rho+k_k,1)}\right)'$. Then, for the dynamic two stage least squares estimator $\widehat{\check{\boldsymbol{\theta}}}_{D2SLS;p}$ we obtain:

Theorem 2 ($T \to \infty$ limits for $D2SLS$ Estimation)**.** *Consider the fixed effects spatial correlation models (11) and (13) and the D2SLS estimator (33) based on the within-transformed model (32). Suppose that Assumptions 1–4 hold. Let $T_\star = T - 2p(T)$.*

Then, for n fixed and $T \to \infty$, it follows that

1. *$T_\star(\widehat{\check{\boldsymbol{\gamma}}}_{D2SLS;p} - \check{\boldsymbol{\gamma}})$ and $\sqrt{T_\star}(\widehat{\check{\underline{\boldsymbol{\delta}}}}_{D2SLS;\sharp p} - \check{\underline{\boldsymbol{\delta}}}_{\sharp p})$ are asymptotically independent.*
2. *$T_\star\left(\widehat{\check{\boldsymbol{\gamma}}}_{D2SLS;p} - \check{\boldsymbol{\gamma}}\right)$ converges weakly to $\check{\mathbf{M}}_n^{-1}\check{\mathbf{m}}_n$, where $\mathbf{M}_{\check{X}\check{Z},ni} := lim_{T\to\infty}\mathbf{M}_{\check{X}\check{Z},nTi} = \int_0^1 \check{\mathbf{g}}_i(r)\check{\mathbf{h}}_i(r)'dr$, $\mathbf{M}_{\check{Z}\check{Z},ni} := lim_{T\to\infty}\mathbf{M}_{\check{Z}\check{Z},nTi} = \int_0^1 \check{\mathbf{h}}_i(r)\check{\mathbf{h}}_i(r)'dr$, $\mathbf{M}_{\check{X}\check{Z},n} := lim_{T\to\infty}\mathbf{M}_{\check{X}\check{Z},nT} = \sum_{i=1}^{n}\mathbf{M}_{\check{X}\check{Z},ni}$, $\mathbf{M}_{\check{Z}\check{Z},n} := lim_{T\to\infty}\mathbf{M}_{\check{Z}\check{Z},nT} = \sum_{i=1}^{n}\mathbf{M}_{\check{X}\check{Z},ni}$, $\check{\mathbf{M}}_n := \mathbf{M}_{\check{X}\check{Z},n}\mathbf{M}^{-1}_{\check{Z}\check{Z},n}\mathbf{M}'_{\check{X}\check{Z},n}$, $\mathbf{m}_{\check{Z}\check{u},ni} = lim_{T\to\infty}\mathbf{m}_{\check{Z}\check{u},nTi}$, $\mathbf{m}_{\check{Z}\check{u},n} := \sum_{i=1}^{n}\mathbf{m}_{\check{Z}\check{u},ni}$, and $\check{\mathbf{m}}_n := \mathbf{M}_{\check{X}\check{Z},n}\mathbf{M}^{-1}_{\check{Z}\check{Z},n}\mathbf{m}'_{\check{Z}\check{u},n}$.*
3. *Suppose that $\widehat{\boldsymbol{\Omega}}_{uu}$ is a consistent estimator of $\boldsymbol{\Omega}_{uu} = diag(\Omega_{u_iu_i})_{i=1,\ldots,n}$, then*

$$
\begin{aligned}
\check{\mathbf{V}}_{nT} &:= \left[\mathbf{M}_{\check{X}\check{Z},nT}\mathbf{M}^{-1}_{\check{Z}\check{Z},nT}\mathbf{M}'_{\check{X}\check{Z},nT}\right]^{-1}\check{\mathbf{D}}_{nT}\left[\mathbf{M}_{\check{X}\check{Z},nT}\mathbf{M}^{-1}_{\check{Z}\check{Z},nT}\mathbf{M}'_{\check{X}\check{Z},nT}\right]^{-1} \\
\Rightarrow &\quad \left[\mathbf{M}_{\check{X}\check{Z},n}\mathbf{M}^{-1}_{\check{Z}\check{Z},n}\mathbf{M}'_{\check{X}\check{Z},n}\right]^{-1}\check{\mathbf{D}}_n\left[\mathbf{M}_{\check{X}\check{Z},n}\mathbf{M}^{-1}_{\check{Z}\check{Z},n}\mathbf{M}'_{\check{X}\check{Z},n}\right]^{-1} =: \check{\mathbf{V}}_n\,, \quad \textit{where} \\
&\check{\mathbf{D}}_{nT} := \mathbf{M}_{\check{X}\check{Z},nT}\mathbf{M}^{-1}_{\check{Z}\check{Z},nT}\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},nT}\mathbf{M}^{-1}_{\check{Z}\check{Z},nT}\mathbf{M}'_{\check{X}\check{Z},nT}\,, \\
&\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},nT} = \sum_{i=1}^{n}\hat{\Omega}_{u_iu_i}\left(\mathbf{M}_{\check{Z}\check{Z},nTi} - \mathbf{M}^{\Sigma}_{\check{Z}\sum\check{Z},nTi} - \mathbf{M}^{\Sigma}_{\sum\check{Z}\check{Z},nTi} + \mathbf{M}^{\Sigma}_{\sum\check{Z}\sum\check{Z},nTi}\right)\,, \\
&\check{\mathbf{D}}_n := \mathbf{M}_{\check{X}\check{Z},n}\mathbf{M}^{-1}_{\check{Z}\check{Z},n}\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},n}\mathbf{M}^{-1}_{\check{Z}\check{Z},n}\mathbf{M}'_{\check{X}\check{Z},n}\,,\quad \mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},n} = \sum_{i=1}\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},ni}\,, \textit{ and} \\
&\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},ni} = \Omega_{u_iu_i}\left(\int_0^1\check{\mathbf{h}}_i(r)\check{\mathbf{h}}_i(r)'dr - \int_0^1\check{\mathbf{h}}_i(r)\left(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r)\right)'dr - \int_0^1\left(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r)\right)\check{\mathbf{h}}_i(r)'dr\right. \\
&\left.\quad + \int_0^1\left(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r)\right)\left(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r)\right)'dr\right)\,. \qquad (35)
\end{aligned}
$$

Given an $\mathtt{s} \times (k+1)$ restriction matrix $\mathbf{R}$, the Wald type test statistic

$$
\check{\mathscr{W}}_{\gamma,nT} = \left(T_\star\mathbf{R}\left(\widehat{\check{\boldsymbol{\gamma}}}_{D2SLS;p} - \check{\boldsymbol{\gamma}}\right)\right)'\left(\mathbf{R}\check{\mathbf{V}}_{nT}\mathbf{R}'\right)^{-1}\left(T_\star\mathbf{R}\left(\widehat{\check{\boldsymbol{\gamma}}}_{D2SLS;p} - \check{\boldsymbol{\gamma}}\right)\right)\,, \qquad (36)
$$

converges in distribution to a χ^2 random variable with $\mathtt{s}$ degrees of freedom.

Proof. See Appendix 2.

In the following the joint limit theory developed in Phillips and Moon (1999, 2000) will be applied to obtain joint limits, i.e., $(n, T) \to \infty$-limits.[8] Consider *draws* with double index (i, t), where i, $i = 1, \ldots, n$, stands for the cross-sectional index of the draw and t, $t = 1, \ldots, T$, for the time-series dimension of the draw. To derive the joint $(n, T) \to \infty$-limit distribution we impose

[8]The definitions of sequential convergence ("first $T \to \infty$, then $n \to \infty$") and joint convergence ("$T, n \to \infty$") in probability are provided in Definitions 1 and 2 in Phillips and Moon (2000).

Assumption 5.

(a) *For the stochastic process* $\left(\boldsymbol{\eta}_t\right)_{t\in\mathbb{Z}} = \left(\boldsymbol{\eta}_{It}\right)_{t\in\mathbb{Z}}$ *Assumptions 1–10 stated in Phillips and Moon (1999) hold.*

(b) *If* $k_L > 0$ *a law of large numbers can be applied to* $\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'$. *That is,* $\frac{1}{n}\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'$ *converges in probability to* $\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)$ *for* $n\to\infty$, *where the matrix* $\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)$ *is regular. For* $\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}$ *a joint central limit theorem can be applied. That is,* $\frac{1}{\sqrt{nT_\star}}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger} \Rightarrow \boldsymbol{\nu}_{(\grave{\mathbf{x}}_{Li}\grave{u}_{it}^{\dagger})}$ *for* $(n,T)\to\infty$, *where* $\boldsymbol{\nu}_{(\grave{\mathbf{x}}_{Li}\grave{u}_{it}^{\dagger})} \sim \mathcal{N}\left(\mathbf{0}_{(k_L\times 1)},\mathbf{D}_{(\grave{\mathbf{x}}_{Li}\grave{u}_{it}^{\dagger})}\right)$ *and* $0 < \mathbf{D}_{(\grave{\mathbf{x}}_{Li}\grave{u}_{it}^{\dagger})} < \infty$. *The joint limit* $\boldsymbol{\nu}_{(\grave{\mathbf{x}}_{Li}\grave{u}_{it}^{\dagger})}$ *is equal to the sequential "first T, then* $n\to\infty$*"-limit.*

Part (b) of Assumption 5 will be used to obtain the asymptotic limit distribution of an estimator of $\boldsymbol{\beta}_L$. By the last part of Assumption 2 we already get $\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{u}_{it}^{\dagger}\right) = \mathbf{0}_{(k_L\times 1)}$. Sufficient conditions for a joint central limit theorem to hold, where the joint limit agrees with the sequential limit are provided in Phillips and Moon (1999, pp. 1070–1071).

For a panel cointegration model joint limit theory was, e.g., applied Kao and Chiang (2000), Pedroni (2000), Mark and Sul (2003), Baltagi (2008, Chapter 12.6), and Pesaran (2015, Chapter 31.10). By Assumption 5 Lemma 1(d) of Phillips and Moon (1999) applies, such that for the long run covariances $\boldsymbol{\Omega}_i = \begin{pmatrix}\boldsymbol{\Omega}_{u_iu_i} & \boldsymbol{\Omega}_{u_iv_i}\\ \boldsymbol{\Omega}_{v_iu_i} & \boldsymbol{\Omega}_{v_iv_i}\end{pmatrix}$, $i=1,\ldots,n$, we observe that $\mathbb{E}\left(\boldsymbol{\Omega}_i\right) := \bar{\boldsymbol{\Omega}} := \begin{pmatrix}\bar{\boldsymbol{\Omega}}_{u_iu_i} & \bar{\boldsymbol{\Omega}}_{u_iv_i}\\ \bar{\boldsymbol{\Omega}}_{v_iu_i} & \bar{\boldsymbol{\Omega}}_{v_iv_i}\end{pmatrix}$. Moreover, let $j_1 = k_I+1$, $l_0 = q_\rho+1$, and $l_1 = q_\rho+k_I$, then it follows that

$$\begin{aligned}
&\left[\lim_{(n,T)\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\mathbf{A}_{\breve{X}p}\breve{\mathbf{X}}_{p;it_\star}\breve{\mathbf{Z}}_{p;it_\star}'\mathbf{A}_{\breve{Z}p}\right]_{(2:j_1,l_0:l_1)}\\
&=\left[\lim_{(n,T)\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\mathbf{A}_{\breve{Z}p}\breve{\mathbf{Z}}_{p;it_\star}\breve{\mathbf{Z}}_{p;it_\star}'\mathbf{A}_{\breve{Z}p}\right]_{(l_0:l_1,l_0:l_1)}\\
&=\lim_{(n,T)\to\infty}\frac{1}{nT_\star}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\breve{\mathbf{x}}_{it_\star}\breve{\mathbf{x}}_{it_\star}' = \mathbb{E}\left(\int_0^1\breve{\mathcal{B}}_{v_i}(r)\breve{\mathcal{B}}_{v_i}(r)'dr\right) = \frac{1}{6}\bar{\boldsymbol{\Omega}}_{v_iv_i}\,.
\end{aligned} \tag{37}$$

For more details on these limits see Online Appendix A-2. To obtain $\breve{y}_{it_\star}^{*}$, the spatial weights matrix $\mathbf{W}_{\{n\}}$ satisfies:

Assumption 6 (Spatial Lag II). *The requirements of Assumption 1 continue to hold for the* $n\times n$ *matrix* $\mathbf{W}_{\{n\}}$. *That is, for all* $n\in\mathbb{N}$ *as well as for* $n\to\infty$ *the spatial weights* $\left(W_{\{n\},ij}\right)_{i,j=1,\ldots,n}$ *are nonstochastic and observable with* $W_{\{n\},ii}=0$ *and* $\mathbf{W}_{\{n\}}\neq\mathbf{0}_{(n\times n)}$. *For any* $\rho\in(-1,1)$, *the largest absolute eigenvalue of* $\rho\mathbf{W}_{\{n\}}$ *is smaller than one and the sequences of the largest absolute eigenvalues of* $\rho\mathbf{W}_{\{n\}}$ *is bounded away from one. In addition,* $|W_{\{n\},ij}|\leq\bar{w}$ *and* $|K_{ij}| = |\left[(\mathbf{I}_n-\rho\mathbf{W}_{\{n\}})^{-1}\right]_{(ij)}|\leq\bar{\omega}$ *for all* $i,j=1,\ldots,n$ *and* $n\in\mathbb{N}\cup\infty$.

The inverse $\mathbf{K}_{\{n\}} := \left(\mathbf{I}_n-\rho\mathbf{W}_{\{n\}}\right)^{-1}$ exists by the assumption on the largest eigenvalue of $\rho\mathbf{W}_{\{n\}}$ for each finite n as well as for $n\to\infty$ (see, e.g., Heuser, 1992, Theorem 12.4 on the Neumann Series). By $|W_{\{n\},ij}|\leq\bar{w}$ and $|K_{\{n\},ij}|\leq\bar{w}$, the elements of $\mathbf{W}_{\{n\}}$ and $\mathbf{K}_{\{n\}}$ are bounded. Since an eigenvalue of $\rho\mathbf{W}_{\{n\}}$ can be kept small by decreasing ρ when n becomes large, we postulate $\rho\in(-1,1)$.[9]

To obtain the $(n,T)\to\infty$ asymptotic limit distribution of the estimator *D2SLS* estimator (33), we apply the scaling matrices $\frac{1}{\sqrt{n}}\mathbf{A}_{\breve{X}p}$ and $\frac{1}{\sqrt{n}}\mathbf{A}_{\breve{Z}p}$. That is, the terms arising from the spatial lag and the $I(1)$

[9]On spatial weights $\mathbf{W}_{\{n\}}$ for a large cross-sectional dimension see, e.g., Kapoor et al. (2007), Kelejian and Prucha (2008), and Drukker et al. (2013). The subscript $\{n\}$ is added to express the dependence on n.

components are scaled by $1/(\sqrt{n}T_\star)$, while all projection variables are scaled by $1/\sqrt{nT_\star}$. Given these scaling factors, we define $\mathbf{Q}^\star_{\breve{X}\breve{Z},nT} := \frac{1}{n}\sum_{i=1}^n \mathbf{M}^\star_{\breve{X}\breve{Z},nTi} = \frac{1}{n}\sum_{i=1}^n \sum_{t_\star=1}^{T_\star} \mathbf{A}_{\breve{X}p}\breve{\mathbf{X}}_{p;it_\star}\breve{\mathbf{Z}}'_{p;it_\star}\mathbf{A}_{\breve{Z}p}$, $\mathbf{Q}^\star_{\breve{Z}\breve{Z},nT} := \frac{1}{n}\sum_{i=1}^n \mathbf{M}^\star_{\breve{Z}\breve{Z},nTi} = \frac{1}{n}\sum_{i=1}^n \sum_{t_\star=1}^{T_\star} \mathbf{A}_{\breve{Z}p}\breve{\mathbf{Z}}_{p;it_\star}\breve{\mathbf{Z}}'_{p;it_\star}\mathbf{A}_{\breve{Z}p}$, $\mathbf{Q}_{\breve{X}\breve{Z},nT} := \left[\mathbf{Q}^\star_{\breve{X}\breve{Z},nT}\right]_{(1:1+k_I,1:q_\rho+k_I)}$, $\mathbf{Q}_{\breve{Z}\breve{Z},nT} := \left[\mathbf{Q}^\star_{\breve{Z}\breve{Z},nT}\right]_{(1:q_\rho+k_I,1:q_\rho+k_I)}$, $\mathbf{q}^\star_{\breve{Z}\breve{u},nT} := \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{m}^\star_{\breve{Z}\breve{u},nTi} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_{t_\star=1}^{T_\star} \mathbf{A}_{\breve{Z}p}\breve{\mathbf{Z}}_{p;it_\star}\underline{\breve{u}}_{p;it_\star}$, and $\mathbf{q}_{\breve{Z}\breve{u},nT} := \left[\mathbf{q}^\star_{\breve{Z}\breve{u},nT}\right]_{(1:q_\rho+k_I,1)}$. If the joint limits of these terms exist, the joint limits will be abbreviated by $\mathbf{Q}^\star_{\breve{X}\breve{Z}},\dots,\mathbf{q}_{\breve{Z}\breve{u}}$.

In addition, under the premise that all expectations containing a term arising from a spatial lag exist, we get the expectations $\mathbb{E}\left(\mathbf{M}^\star_{\breve{X}\breve{Z}nTi}\right)$ and $\mathbb{E}\left(\mathbf{M}^\star_{\breve{Z}\breve{Z}nTi}\right)$, where—by the congruence of the joint and the sequential limits (implied by our model assumptions)—we observe that $\mathbb{E}\left(\mathbf{M}^\star_{\breve{X}\breve{Z}nTi}\right)$ and $\mathbb{E}\left(\mathbf{M}^\star_{\breve{Z}\breve{Z}nTi}\right)$ are block diagonal with $\mathbb{E}\left(\mathbf{M}_{\breve{X}\breve{Z}nTi}\right)$ and $\mathbb{E}\left(\mathbf{M}_{\breve{Z}\breve{Z}nTi}\right)$ in the north-west. The south-east blocks of both matrices are equal and contain expectations of the autocovariance matrices $\boldsymbol{\Gamma}_{\ell,v_iv_i}$. Observe that

$$
\mathbb{E}\left(\mathbf{M}_{\breve{X}\breve{Z}nTi}\right) = \begin{pmatrix} \mathbb{E}\left(\left[\mathbf{M}_{\breve{X}\breve{Z},nTi}\right]_{(1,1)}\right) & \mathbb{E}\left(\left[\mathbf{M}_{\breve{X}\breve{Z},nTi}\right]_{(1,2:k+q_\rho)}\right) \\ \mathbb{E}\left(\left[\mathbf{M}_{\breve{X}\breve{Z},nTi}\right]_{(2:k+1,1)}\right) & \mathbb{E}\left(\int_0^1 \breve{\mathcal{B}}_{v_i}(r)\breve{\mathcal{B}}_{v_i}(r)'dr\right) \end{pmatrix}
$$
$$
= \begin{pmatrix} \mathbb{E}\left(\int_0^1 \sum_{j=1}^n\sum_{\ell=1}^n W_{\{n\},ij}K_{\{n\},j\ell}\boldsymbol{\beta}_I'\breve{\mathcal{B}}_{v_\ell}(r)\begin{bmatrix} \mathbf{W}^{\tau_{\mathbb{K}(1)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(1)}\breve{\mathcal{B}}_v(r) \\ \vdots \\ \mathbf{W}^{\tau_{\mathbb{K}(q_\rho)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(q_\rho)}\breve{\mathcal{B}}_v(r)\end{bmatrix}' dr\right) & \mathbb{E}\left(\sum_{j=1}^n\sum_{\ell=1}^n W_{\{n\},ij}K_{\{n\},j\ell}\boldsymbol{\beta}_I'\int_0^1 \breve{\mathcal{B}}_{v_\ell}(r)\breve{\mathcal{B}}_{v_i}(r)'dr\right) \\ \mathbb{E}\left(\int_0^1 \breve{\mathcal{B}}_{v_i}(r)\begin{bmatrix} \mathbf{W}^{\tau_{\mathbb{K}(1)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(1)}\breve{\mathcal{B}}_v(r) \\ \vdots \\ \mathbf{W}^{\tau_{\mathbb{K}(q_\rho)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(q_\rho)}\breve{\mathcal{B}}_v(r)\end{bmatrix}' dr\right)' & \frac{1}{6}\bar{\boldsymbol{\Omega}}_{v_iv_i} \end{pmatrix}, \text{ and} \tag{38}
$$

$$
\mathbb{E}\left(\mathbf{M}_{\breve{Z}\breve{Z}nTi}\right) := \begin{pmatrix} \mathbb{E}\left(\left[\mathbf{M}_{\breve{Z}\breve{Z},nTi}\right]_{(1:q_\rho,1:q_\rho)}\right) & \mathbb{E}\left(\left[\mathbf{M}_{\breve{Z}\breve{Z},nTi}\right]_{(1:q_\rho,2:k+q_\rho)}\right) \\ \mathbb{E}\left(\left[\mathbf{M}_{\breve{X}\breve{Z},nTi}\right]_{(2:k+q_\rho,1:q_\rho)}\right) & \mathbb{E}\left(\int_0^1 \breve{\mathcal{B}}_{v_i}(r)\breve{\mathcal{B}}_{v_i}(r)'dr\right) \end{pmatrix}
$$
$$
= \begin{pmatrix} \mathbb{E}\left(\int_0^1 \begin{bmatrix} \mathbf{W}^{\tau_{\mathbb{K}(1)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(1)}\breve{\mathcal{B}}_v(r) \\ \vdots \\ \mathbf{W}^{\tau_{\mathbb{K}(q_\rho)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(q_\rho)}\breve{\mathcal{B}}_v(r)\end{bmatrix}\begin{bmatrix} \mathbf{W}^{\tau_{\mathbb{K}(1)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(1)}\breve{\mathcal{B}}_v(r) \\ \vdots \\ \mathbf{W}^{\tau_{\mathbb{K}(q_\rho)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(q_\rho)}\breve{\mathcal{B}}_v(r)\end{bmatrix}' dr\right) & \mathbb{E}\left(\int_0^1 \begin{bmatrix} \mathbf{W}^{\tau_{\mathbb{K}(1)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(1)}\breve{\mathcal{B}}_v(r) \\ \vdots \\ \mathbf{W}^{\tau_{\mathbb{K}(q_\rho)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(q_\rho)}\breve{\mathcal{B}}_v(r)\end{bmatrix}\breve{\mathcal{B}}_{v_i}(r)'dr\right) \\ \mathbb{E}\left(\int_0^1 \breve{\mathcal{B}}_{v_i}(r)\begin{bmatrix} \mathbf{W}^{\tau_{\mathbb{K}(1)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(1)}\breve{\mathcal{B}}_v(r) \\ \vdots \\ \mathbf{W}^{\tau_{\mathbb{K}(q_\rho)}}_{\{n\},i}\mathbf{C}_{\mathbb{K}(q_\rho)}\breve{\mathcal{B}}_v(r)\end{bmatrix}' dr\right) & \frac{1}{6}\bar{\boldsymbol{\Omega}}_{v_iv_i} \end{pmatrix}. \tag{39}
$$

Given Assumption 5, we observe from (37) that the expectations as well as the joint limits for the terms in the south-east are equal to $\frac{1}{6}\mathbb{E}\left(\boldsymbol{\Omega}_{v_iv_i}\right) = \frac{1}{6}\bar{\boldsymbol{\Omega}}_{v_iv_i}$. In addition, the other terms in (38) and (39) contain sums arising from a spatial lag. To guarantee the existence of the expectations in (38) and (39) and to make a joint weak law of large numbers applicable, we impose:

Assumption 7.

(a) *For any n, T as well as $(n,T)\to\infty$, the expectations (38) and (39) exist. For any $T, n \in \mathbb{N}$ as well as for $n\to\infty$ also the second moments of $\left[\mathbf{M}_{\breve{X}\breve{Z},nTi}\right]_{(1,1)}$, $\left[\mathbf{M}_{\breve{X}\breve{Z},nTi}\right]_{(1,2:k+q_\rho)}$, $\left[\mathbf{M}_{\breve{X}\breve{Z},nTi}\right]_{(2:k+1,1:q_\rho)}$,*

$\left[\mathbf{M}_{\check{Z}\check{Z},nTi}\right]_{(1:q_\rho,1:q_\rho)}$, $\left[\mathbf{M}_{\check{Z}\check{Z},nTi}\right]_{(1:q_\rho,q_\rho+1:k+q_\rho)} = \left[\mathbf{M}_{\check{Z}\check{Z},nTi}\right]'_{(q_\rho+1:k+q_\rho,1:q_\rho)}$, $\mathbf{M}^{\Sigma}_{\sum\check{Z}\check{Z},nTi} = \mathbf{M}^{\Sigma\prime}_{\check{Z}\sum\check{Z},nTi}$, $\mathbf{M}^{\Sigma}_{\sum\check{Z}\sum\check{Z},nTi}$ *and* $\left[\mathbf{m}_{\check{Z}\check{u},nTi}\right]_{(1:q_\rho,1)}$ *exist.*

(b) *If the* $(T,n)\to\infty$*-limits exist, the rank condition and the order condition are met.*

By Assumption 5 the limit theory developed in Phillips and Moon (1999) can be applied immediately to all components of $\mathbf{M}_{\cdot\cdot,nTi}$ and $\mathbf{m}_{\check{Z}\check{u},nTi}$ not containing terms with a spatial lag. Due to Assumption 7(a) the expected values $\mathbb{E}(\mathbf{M}^{\star}_{\check{X}\check{Z},nTi})$ as well as $\mathbb{E}(\mathbf{M}^{\star}_{\check{Z}\check{Z},nTi})$ exist, and the joint limits $\left[\lim_{(n,T)\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\mathbf{A}_{\check{X}p}\check{\mathbf{X}}_{p;it_\star}\check{\mathbf{Z}}'_{p;it_\star}\mathbf{A}_{\check{Z}p}\right]_{(1:j_1,1:l_1)} =: \mathbf{Q}_{\check{X}\check{Z}}$ and $\left[\lim_{(n,T)\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\mathbf{A}_{\check{Z}p}\check{\mathbf{Z}}_{p;it_\star}\check{\mathbf{Z}}'_{p;it_\star}\mathbf{A}_{\check{Z}p}\right]_{(1:l_1,1:l_1)} =: \mathbf{Q}_{\check{Z}\check{Z}}$ satisfy $\mathbf{Q}_{\check{X}\check{Z}} = \mathbb{E}(\mathbf{M}_{\check{X}\check{Z},nTi}) = \mathbb{E}(\mathbf{M}_{\check{X}\check{Z},ni})$ as well as $\mathbf{Q}_{\check{Z}\check{Z}} = \mathbb{E}(\mathbf{M}^{\star}_{\check{Z}\check{Z},nTi}) = \mathbb{E}(\mathbf{M}^{\star}_{\check{Z}\check{Z},ni})$. By the assumption on the moments of $\mathbf{M}^{\Sigma}_{\sum\check{Z}\check{Z},nTi} = \mathbf{M}^{\Sigma\prime}_{\check{Z}\sum\check{Z},nTi}$ and $\mathbf{M}^{\Sigma}_{\sum\check{Z}\sum\check{Z},nTi}$ the joint limit of $\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},nT}$ exists and is equal to the expectation of $\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},ni}$.

Note that the existence of the expectations of the nonsouth-east terms in (38) and (39) is nontrivial. To see this, the joint limits should be equal (given the existence of the expectations and regularity conditions following from Phillips and Moon, 1999) to the expectations of the $T\to\infty$ limits of $\mathbf{M}_{\check{X}\check{Z},nTi}$ and $\mathbf{M}_{\check{Z}\check{Z},nTi}$ obtained in Theorem 2. By considering, e.g., the $(1,1)$ element of $\mathbb{E}(\mathbf{M}_{\check{X}\check{Z},ni})$ we observe $|[\mathbb{E}(\mathbf{M}_{\check{X}\check{Z},ni})]_{(1,1)}| = |\mathbb{E}(\int_0^1\sum_{j=1}^{n}\sum_{\ell=1}^{n}W_{\{n\},ij}K_{\{n\},j\ell}\boldsymbol{\beta}'\breve{\mathcal{B}}_{v_\ell}(r)\sum_{\iota=1}^{n}W_{\{n\},i\iota}[\mathbf{C}_{\mathbb{K}(1)}]_{(\iota,(\iota-1)n+1\iota)}\breve{\mathcal{B}}_{v_\iota}(r)dr)| \le \bar{\omega}^3 n\cdot|\sum_{\ell=1}^{n}\sum_{\iota=1}^{n}\mathbb{E}(\int_0^1|\boldsymbol{\beta}'\breve{\mathcal{B}}_{v_\ell}(r)[\mathbf{C}_{\mathbb{K}(1)}]_{(\iota,(\iota-1)n+1\iota)}\breve{\mathcal{B}}_{v_\iota}(r)dr)|$, where $|W_{\{n\},ij}|$ and $|K_{\{n\},ij}|\le\bar{\omega}$ by Assumption 6 and $\mathbb{E}(\int_0^1\boldsymbol{\beta}'\breve{\mathcal{B}}_{v_j}(r)[\mathbf{C}_{\mathbb{K}(1)}]_{(\iota,(\iota-1)n+1\iota)}\breve{\mathcal{B}}_{v_\iota}(r)) = \mathbf{0}_{(k_I\times 1)}$ for $j\neq\iota$, while for $j=\iota$ it is a $\boldsymbol{\beta}$, $[\mathbf{C}_{\mathbb{K}(1)}]_{(\iota,(\iota-1)n+1\iota)}$ weighted sum of $\frac{1}{6}\bar{\boldsymbol{\Omega}}_{v_iv_i}$ (note that the coefficients $[\mathbf{C}_{\mathbb{K}(1)}]_{(\iota,(\iota-1)n+1\iota)}$ are equal for all $\iota=1,\ldots,n$). In addition, suppose that $W_{\{n\},ij}\ge\underline{\omega}>0$ for all $i\neq j$ and $K_{\{n\},ij}\ge\underline{\omega}>0$ for all i,j, then $|[\mathbb{E}(\mathbf{M}_{\check{X}\check{Z},ni})]_{(1,1)}| = |\mathbb{E}(\int_0^1\sum_{j=1}^{n}\sum_{\ell=1}^{n}W_{\{n\},ij}K_{\{n\},j\ell}\boldsymbol{\beta}'\breve{\mathcal{B}}_{v_\ell}(r)\sum_{\iota=1}^{n}W_{\{n\},i\iota}[\mathbf{C}_{\mathbb{K}(1)}]_{(\iota,(\iota-1)n+1\iota)}\breve{\mathcal{B}}_{v_\iota}(r)dr)| \ge \underline{\omega}^3 n\cdot|\mathbb{E}(\boldsymbol{\beta}'\breve{\mathcal{B}}_{v_\iota}(r)[\mathbf{C}_{\mathbb{K}(1)}]_{(\iota,(\iota-1)n+1\iota)}\breve{\mathcal{B}}_{v_\iota}(r)dr)|$, which becomes large if n becomes large. Hence, in general the $(1,1)$ element of $\mathbb{E}(\mathbf{M}_{\check{X}\check{Z},ni})$ need not be finite. Similar calculations can be performed with the north-east and the south-west element as well as for $\mathbb{E}(\mathbf{M}_{\check{Z}\check{Z},ni})$. Intuitively, either only a finite subset of spatial weights is nonzero or the weights decay sufficiently fast such that these expected values exist.

The existence of the second moments of $\mathbf{M}_{\check{X}\check{Z},nTi}$, $\mathbf{M}_{\check{Z}\check{Z},nTi}$, and $\mathbf{m}_{\check{Z}\check{u},nTi}$, will be used to apply a joint central limit theory to $\mathbf{q}_{\check{Z}\check{u},nT}$. The variance will be provided by $\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u}} := \mathbb{E}\big(\Omega_{u_iu_i}\big(\int_0^1\check{\mathbf{h}}_i(r)\check{\mathbf{h}}_i(r)'dr - \int_0^1\check{\mathbf{h}}_i(r)(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r))'dr - \int_0^1(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r))\check{\mathbf{h}}_i(r)'dr + \int_0^1(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r))(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r))'dr\big)\big) = \bar{\boldsymbol{\Omega}}_{u_iu_i}\mathbb{E}\big(\int_0^1\check{\mathbf{h}}_i(r)\check{\mathbf{h}}_i(r)'dr - \int_0^1\check{\mathbf{h}}_i(r)(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r))'dr - \int_0^1(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r))\check{\mathbf{h}}_i(r)'dr + \int_0^1(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r))(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r))'dr\big)$ (see Online Appendix A-2 and Theorems 1 and 2 of Phillips and Moon, 1999). This yields:

Theorem 3 (Joint $(n,T)\to\infty$-Limits for *D2SLS* Estimation). *Consider the fixed effects spatial correlation model (10) and the D2SLS estimator (33) based on the within-transformed model (14). Suppose that the Assumptions 1 to 7 hold. Let* $T_\star = T-2p(T)$*, then for* $(n,T)\to\infty$*, where* $n^6/T\to 0$*, it follows that:*

I. $\mathbf{Q}_{\check{Z}\check{Z},nT}\xrightarrow{P}\mathbf{Q}_{\check{Z}\check{Z}}$, $\mathbf{Q}_{\check{X}\check{Z},nT}\xrightarrow{P}\mathbf{Q}_{\check{X}\check{Z}}$. *The asymptotic distribution of* $\sqrt{n}T_\star(\check{\boldsymbol{\gamma}}_{D2SLS;p}-\boldsymbol{\gamma})$ *is a normal distribution with mean vector* $\mathbf{0}_{(k+1)}$ *and covariance matrix*

$$\check{\mathbf{V}}_{\check{Q}} := \left[\mathbf{Q}_{\check{X}\check{Z}}\mathbf{Q}^{-1}_{\check{Z}\check{Z}}\mathbf{Q}'_{\check{X}\check{Z}}\right]^{-1}\check{\mathbf{D}}_{\check{Q}}\left[\mathbf{Q}_{\check{X}\check{Z}}\mathbf{Q}^{-1}_{\check{Z}\check{Z}}\mathbf{Q}'_{\check{X}\check{Z}}\right]^{-1},\quad \text{where } \check{\mathbf{D}}_{\check{Q}} := \mathbf{Q}_{\check{X}\check{Z}}\mathbf{Q}^{-1}_{\check{Z}\check{Z}}\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u}}\mathbf{Q}^{-1}_{\check{Z}\check{Z}}\mathbf{Q}'_{\check{X}\check{Z}}. \quad (40)$$

II. *Suppose that a consistent estimator of* $\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u}}$*, denoted by* $\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},nT}$ *is available, such that* $\check{\mathbf{D}}_{\check{Q},nT} := \mathbf{Q}_{\check{X}\check{Z},nT}\mathbf{Q}^{-1}_{\check{Z}\check{Z},,nT}\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},,nT}\mathbf{Q}^{-1}_{\check{Z}\check{Z},,nT}\mathbf{Q}'_{\check{X}\check{Z},,nT}$ *and* $\check{\mathbf{V}}_{\check{Q},nT} := \left[\mathbf{Q}_{\check{X}\check{Z},nT}\mathbf{Q}^{-1}_{\check{Z}\check{Z},nT}\mathbf{Q}'_{\check{X}\check{Z},nT}\right]^{-1}$

$\breve{\mathbf{D}}_{\breve{Q},nT}\left[\mathbf{Q}_{\breve{X}\breve{Z},nT}\mathbf{Q}_{\breve{Z}\breve{Z},nT}^{-1}\mathbf{Q}_{\breve{X}\breve{Z},nT}'\right]^{-1}$ *are estimators of* $\breve{\mathbf{D}}_{\breve{Q}}$ *and* $\breve{\mathbf{V}}_{\breve{Q}}$. *Then,* $\breve{\mathbf{D}}_{\breve{Q},nT}$ *and* $\breve{\mathbf{V}}_{\breve{Q},nT}$ *converge in probability to* $\breve{\mathbf{D}}_{\breve{Q}}$ *as well as* $\breve{\mathbf{V}}_{\breve{Q}}$.

Proof. See Section B.

From Theorem 3 it follows that the Wald type test defined in Theorem 1 can still be applied in a setting where T and n are large. The limit distribution of the Wald statistic is still a χ^2-distribution with $\mathfrak{s}$ degrees of freedom. In addition, we observe that the *D2SLS* estimator is $\sqrt{nT}$ consistent, with $n^6/T \to 0$. Although the *OLS*, the *DOLS*, the *2SLS*, and the *D2SLS* estimators are consistent for n fixed and $T \to \infty$, the *OLS*, the *DOLS*, and the *2SLS* are in general (due to the second-order bias terms) not $\sqrt{nT}$ consistent.

In a final step we investigate model (11), where—in addition to ρ and $\boldsymbol{\beta}_I$—the parameter $\boldsymbol{\beta}_L$ has to be estimated. From (12) we derive the regression model

$$\grave{y}_{it_\star} - \rho\sum_{j=1}^{n} W_{ij}\grave{y}_{jt_\star} - \boldsymbol{\beta}_I'\grave{\mathbf{x}}_{Iit_\star} = \boldsymbol{\beta}_L'\grave{\mathbf{x}}_{Li} + \grave{u}_{it_\star}^{\dagger}\,, \tag{41}$$

and the infeasible estimator

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_L &= \left(\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\left(\grave{y}_{it_\star} - \rho\sum_{j=1}^{n} W_{ij}\grave{y}_{jt_\star} - \boldsymbol{\beta}_I'\grave{\mathbf{x}}_{Iit_\star}\right)\\ &= \left(T_\star\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\sum_{t_\star=1}^{T_\star}\left(\boldsymbol{\beta}_L'\grave{\mathbf{x}}_{Li} + \grave{u}_{it_\star}^{\dagger}\right)\,.\end{aligned} \tag{42}$$

To obtain a feasible estimator we plug in $\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS}$ into (42), resulting in

$$\widehat{\boldsymbol{\beta}}_L = \left(\sum_{i=1}^{n}\sum_{t=1}^{T_\star}\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\left(\grave{y}_{it_\star} - \widehat{\rho}\sum_{j=1}^{n} W_{ij}\grave{y}_{jt_\star} - \widehat{\boldsymbol{\beta}}_I'\grave{\mathbf{x}}_{Iit_\star}\right)\,,\ \text{such that}$$

$$\widehat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}_L == \left(\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\frac{1}{T_\star}\sum_{\star=1}^{T_\star}\left[\grave{u}_{it_\star}^{\dagger} + \left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS} - \breve{\boldsymbol{\gamma}}\right)'\left(\grave{y}_{it_\star}^{*}, \grave{\mathbf{x}}_{Iit_\star}'\right)'\right]. \tag{43}$$

For the feasible estimator (43) Appendix C shows that the $(n, T) \to \infty$-asymptotic limit distribution[10] of $\sqrt{nT_\star}\left(\hat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}_L\right)$ is a normal distribution with mean vector $\mathbf{0}_{(k_L\times 1)}$ and covariance matrix $\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}\mathbf{D}_{\left(\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}\right)}\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1} = \mathbb{E}\left(\Omega_{u_iu_i}^{\dagger}\right)\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}$. In more detail, by Assumption 5.(b) the term $\frac{1}{n}\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'$ converges in probability to $\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)$ by a weak law of large numbers. In addition, by Assumption 5(b) a joint central limit theorem applies to $\frac{1}{\sqrt{nT_\star}}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\grave{u}_{it}^{\dagger}$, such that $\frac{1}{\sqrt{nT_\star}}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\grave{u}_{it}^{\dagger}$ converges to a normally distributed vector $\boldsymbol{\nu}_{(\grave{x}_{Li}\grave{u}_{ti}^{\dagger})}$ where $\boldsymbol{\nu}_{(\grave{x}_{Li}\grave{u}_{ti}^{\dagger})} \sim \mathcal{N}\left(\mathbf{0}_{(k_L\times 1)}, \mathbf{D}_{\left(\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}\right)}\right)$ and $\mathbf{D}_{\left(\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}\right)} = \mathbb{E}\left(\Omega_{u_iu_i}^{\dagger}\right)\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)$. Hence, we observe that $\lim_{n,T\to\infty}\sqrt{nT_\star}\left(\widehat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}_L\right) = \mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}\boldsymbol{\nu}_{(\grave{x}_{Li}\grave{u}_{ti}^{\dagger})}$. Suppose that $\hat{\bar{\Omega}}_{u_iu_i}^{\dagger} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_{u_iu_i}^{\dagger}$ consistently estimates

[10] For $(n, T) \to \infty$ and $n^6/T \to 0$.

$\bar{\Omega}^{\dagger}_{u_iu_i} := \mathbb{E}\left(\Omega^{\dagger}_{u_iu_i}\right)$ for $(T,n) \to \infty$, then the covariance matrix $\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}'_{Li}\right)^{-1}\mathbf{D}_{\left(\grave{\mathbf{x}}_{Li}\grave{u}^{\dagger}_{it_\star}\right)}\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}'_{Li}\right)^{-1} = \mathbb{E}\left(\Omega^{\dagger}_{u_iu_i}\right)\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}'_{Li}\right)^{-1}$ can be estimated consistently using the the finite sample analogs $\hat{\bar{\Omega}}^{\dagger}_{u_iu_i}$ as well as $\frac{1}{n}\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}'_{Li}$.

4. Monte Carlo simulations

This section investigates the small sample properties of the *D2SLS* estimator as well as the *size* and *power* of the Wald type test obtained in the Theorems 1 and 2. We generate the data based on an error process that follows from Assumption 2. Regarding the error dynamics we consider $\boldsymbol{\eta}^{\dagger}_{Iit}$ and $\mathbf{v}_{Ct}$ generated from

$$\boldsymbol{\eta}^{\dagger}_{Iit} = \boldsymbol{\Psi}^{\dagger}_{Ii}(L)\,\boldsymbol{\varepsilon}^{\dagger}_{Iit}\,,\ \text{ for } i = 1,\dots,n, \text{ and, } \ \mathbf{v}_{Ct} = \sum_{i=1}^{n}\boldsymbol{\Psi}_{CIi}\boldsymbol{\eta}^{\dagger}_{Iit} + \boldsymbol{\Psi}_{C}(L)\,\boldsymbol{\varepsilon}_{Ct}\,, \tag{44}$$

where $\boldsymbol{\Psi}_{CIi}$, $i = 1,\dots,n$, are $k_C \times k_I + 1$ matrices. To operationalize this, we need to specify the lag polynomials $\boldsymbol{\Psi}^{\dagger}_{Ii}(L)$ and $\boldsymbol{\Psi}_{C}(L)$. In particular, we have to specify the error dynamics of the vectors $\boldsymbol{\eta}^{\dagger}_{Iit}$ and $\mathbf{v}_{Ct}$, where we assume the same error dynamics for all cross-sections $i = 1,\dots,n$ as well as for the common regressors. For model (8) we use four explanatory variables, where $k_I = k_C = 2$ and $\boldsymbol{\beta}_I = \boldsymbol{\beta}_C = (1,1)'$. Hence, $k = 4$ and $\boldsymbol{\beta} = (1,1,1,1)'$. In this case the number of instruments is $q_\rho = 2$. The individual variables $\mathbf{x}_{Iit}$ are used to construct the instruments $\tilde{\mathbf{z}}^{*}_{it}$. To model the correlation between $\boldsymbol{\eta}^{\dagger}_{Iit}$ and $\mathbf{v}_{Ct}$ we use (44), where the $k_C \times k_I$ matrix $\boldsymbol{\Psi}_{CIi} = \frac{0.1}{n}\cdot\mathbf{1}_{(2\times3)}$ for $i = 1,\dots,n$. For the model (13) the individual variables are used to construct the instruments as well, such that $\breve{\mathbf{z}}^{*}_{it} = \breve{\mathbf{x}}^{*}_{it}$, $k_I = k = 2$, and $q_\rho = 2$. In both cases the exponents τ used to construct these instruments in (24) are set to one.

Regarding the error dynamics we use stationary designs close to Binder et al. (2005) to generate the data for the vectors $\boldsymbol{\eta}^{\dagger}_{Iit}$, for $i = 1,\dots,n$, and $\mathbf{v}_{ct}$. The innovations $\boldsymbol{\varepsilon}^{\dagger}_{Iit}$ are generated as independent draws from $\boldsymbol{\varepsilon}^{\dagger}_{Iit} \sim \mathcal{N}\left(\mathbf{0}_{(k_I+1)}, \boldsymbol{\Sigma}_{I\boldsymbol{\varepsilon}}\right)$, where $\mathcal{N}(.,.)$ stands for a normal distribution. To obtain $\mathbf{v}_{Ct}$, the innovations $\boldsymbol{\varepsilon}_{Ct}$ are *iid* normal, where $\boldsymbol{\varepsilon}_{Ct} \sim \mathcal{N}\left(\mathbf{0}_{(k_C)}, \boldsymbol{\Sigma}_{C\boldsymbol{\varepsilon}}\right)$. In the following Monte Carlo experiments $[\boldsymbol{\Sigma}_{I\boldsymbol{\varepsilon}}]_{(i,i)} = [\boldsymbol{\Sigma}_{C\boldsymbol{\varepsilon}}]_{(i,i)} = 1$ for all diagonal terms and $[\boldsymbol{\Sigma}_{I\boldsymbol{\varepsilon}}]_{(i,j)} = [\boldsymbol{\Sigma}_{C\boldsymbol{\varepsilon}}]_{(i,j)} = 0.8$ for all off-diagonal elements.

In the first three designs we generate $\boldsymbol{\eta}^{\dagger}_{Iit}$ and $\mathbf{v}_{Ct}$ by the first-order vector autoregressive system $(VAR(1))$ $\boldsymbol{\eta}^{\dagger}_{Iit} = \boldsymbol{\Phi}^{\dagger}_{Ii}\boldsymbol{\eta}^{\dagger}_{Ii,t-1} + \boldsymbol{\varepsilon}^{\dagger}_{Iit}$ and $\mathbf{v}_{Ct} = \boldsymbol{\Phi}_{C}\mathbf{v}_{C,t-1} + \boldsymbol{\varepsilon}_{Ct}$, where the 3×3 matrix $\boldsymbol{\Phi}^{\dagger}_{Ii}$ and the 2×2 matrix $\boldsymbol{\Phi}_{C}$ come from one of the following designs: Design $DGP = 1$, stands for stationary $VAR(1)$ with maximum eigenvalue of 0.6, where $\left[\boldsymbol{\Phi}^{\dagger}_{Ii}\right]_{(i,i)} = \left[\boldsymbol{\Phi}_{C}\right]_{(i,i)} = 0.4$ and $\left[\boldsymbol{\Phi}^{\dagger}_{Ii}\right]_{(i,j)} = \left[\boldsymbol{\Phi}_{C}\right]_{(i,j)} = 0.1$ (for $i \neq j$). In design $DGP = 2$ we consider a stationary $VAR(1)$ with maximum eigenvalue of 0.8, where $\left[\boldsymbol{\Phi}_{\cdot}\right]_{(i,i)} = 0.6$ and $\left[\boldsymbol{\Phi}_{\cdot}\right]_{(i,j)} = 0.1$, while with design $DGP = 3$, $\left[\boldsymbol{\Phi}_{\cdot}\right]_{(i,i)} = 0.75$, and $\left[\boldsymbol{\Phi}_{\cdot}\right]_{(i,j)} = 0.1$ yield a largest eigenvalue of 0.95. In addition, we consider a finite-order vector moving average (MA) processes of the form $\boldsymbol{\eta}^{\dagger}_{Iit} = \boldsymbol{\varepsilon}^{\dagger}_{Iit} + \sum_{l=1}^{q}\boldsymbol{\Psi}^{\dagger}_{Iil}\boldsymbol{\varepsilon}^{\dagger}_{Ii,t-l}$ and $\mathbf{v}_{Ct} = \boldsymbol{\varepsilon}_{Ct} + \sum_{l=1}^{q}\boldsymbol{\Psi}_{Cl}\boldsymbol{\varepsilon}_{C,t-l}$ where we choose: Design $DGP = 4$, which is a first-order MA process with parameter $\boldsymbol{\Psi}^{\dagger}_{Ii1}$ (presented in (45)) and $\boldsymbol{\Psi}^{\dagger}_{C1}$, while with $DGP = 5$ we use an MA(2) model with parameters

$$\boldsymbol{\Psi}^{\dagger}_{Ii1} = \begin{pmatrix} 0.6 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.6 \end{pmatrix},\quad \boldsymbol{\Psi}^{\dagger}_{Ii2} = \begin{pmatrix} 0.4 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.4 \end{pmatrix},$$
$$\boldsymbol{\Psi}_{C1} = \begin{pmatrix} 0.6 & 0.1 \\ 0.1 & 0.6 \end{pmatrix},\ \text{ and }\ \boldsymbol{\Psi}_{C2} = \begin{pmatrix} 0.4 & 0.1 \\ 0.1 & 0.4 \end{pmatrix}.$$

Recall that the disturbance in the equation for y_{it} is given by the first element of the vector $\boldsymbol{\eta}^{\dagger}_{Iit}$, while its remaining elements contain $\Delta\mathbf{x}_{Iit}$. The maximum number of leads and lags of the explanatory variables that are conditionally correlated with the disturbances is equal to *one* in the Designs 1-3, while for the Designs 4 and 5 all lags of the explanatory variables are conditionally correlated with the disturbances.

The remaining parameters of the model are chosen as follows: We generate the individual effects α_i from $\alpha_i \sim \mathcal{N}(0,1)$, while if time fixed effects are included then $\lambda_t \sim \mathcal{N}(0,1)$. For a model where $k_L > 0$, we work with $\boldsymbol{\beta}_L = (1,1)'$ and $\mathbf{x}_{Li} \sim \mathcal{N}\left(\mathbf{0}_{(2\times1)}, \mathbf{I}_2\right)$ for $i = 1,\ldots,n$.

The spatial correlation parameter ρ is chosen from the set $\{-0.95, -0.5, -0.1, 0, 0.1, 0.5, 0.95\}$. The choice of $\mathbf{W}$ is based on Kapoor et al. (2007). In more details we consider: (i) A "one step ahead-one step behind circular world" with corresponding entries $1/2$, where the last element in each row is subject to some noise. That is $W_{i,i+1} = 0.5 - \zeta_i$ and $W_{i+1,i} = 0.5$ for $i = 1,\ldots,n-1$. $W_{1,n} = 0.5 - \zeta_1$ and $W_{n,1} = 0.5$, the other entries are zeros. ζ_i, $i = 1,\ldots,n$, are *iid* uniformly distributed noise terms on the interval $[0, 0.2]$. (ii) A "three step ahead-three step behind circular world" with corresponding entries $1/6$. (iii) A "five step ahead-five step behind circular world" with corresponding entries $1/10$. (iv) A "one step ahead-one step behind Rook constellation" with corresponding entries $1/2$. This design is noncircular. Here $W_{i,i+1} = 0.5 - \zeta_i$ and $W_{i+1,i} = 0.5$ for $i = 1,\ldots,n-1$; the other entries are zero. (v) A "two step ahead-two step behind Queen constellation". In this noncircular design $W_{i,i+1} = 0.3$, $W_{i,i+2} = 0.2 - \zeta_i$, $W_{i+1,i} = 0.3$, and $W_{i+2,i} = 0.2$ for $i = 1,\ldots,n-2$; the other entries are zero. Thus, we have in total 175 different data generating processes (5, 7, 5 different settings for the autoregressive structure of $\boldsymbol{\eta}^{\dagger}_{it}$, the spatial correlation parameter ρ and the spatial correlation matrix $\mathbf{W}$, respectively). If the noise terms ζ_i are zero, we observed that some of the matrices become difficult to invert for model (13) when the weights W_{ij} are proportional to $1/n$ for all $i \neq j$. In this case also instruments based on $\tilde{\mathbf{x}}^*_{it_\star}$ can be used, where we did not observe that problem. The asymptotic limit distribution with these instruments is obtained in Online Appendix A-5.

Estimates of $\Omega_{u_iu_i}$ are obtained by $\widehat{\Omega}_{\underline{u}_i\underline{u}_i} = \frac{1}{T_\star}\sum_{t_\star=1}^{T_\star}\sum_{s_\star=1}^{T_\star} \mathtt{k}\left(\frac{|t_\star - s_\star|}{b_T}\right)\widehat{\tilde{u}}_{it_\star}\widehat{\tilde{u}}_{is_\star}$, where $\mathtt{k}(.)$ is a kernel function with bandwidth b_T and $\widehat{\tilde{u}}_{it_\star}$ are the residuals.[11] For the estimation of the long run covariance $\Omega_{u_iu_i}$ we applied the Bartlett and the truncated kernel.[12] The truncated kernel exhibits a better performance than the Bartlett kernel. For the truncated kernel $\mathtt{k}\left(\frac{|t_\star - s_\star|}{b_T}\right) = 1$ for $\frac{|t_\star - s_\star|}{b_T} \leq 1$ and 0 for $\frac{|t_\star - s_\star|}{b_T} > 1$. Hence, only $\widehat{\tilde{u}}_{it_\star}\widehat{\tilde{u}}_{is_\star}$ where $|t_\star - s_\star| \leq b_T$ are used to estimate $\Omega_{u_iu_i}$. In our simulation runs we use a $b_T \leq 15$, where b_T depends on the serial correlation of the residuals.

When implementing the *D2SLS* estimators (28) and (33), the number of leads and lags p included in the regression has to be chosen. Recent literature proposed to choose p by information criteria (see, e.g., Kejriwal and Perron, 2008; Kurozumi and Tuvaandorj, 2010). With small T and n the implementation of such criteria is straightforward. However, since a dataset with (relatively) large n and T is going to be considered, working with small p becomes necessary due to computational restrictions. In particular, we set $p = 2$ for all components $i = 1,\ldots,n$. For all designs working with $p = 2$ performed better than working with $p = 1$.

Last but not least $\mathtt{M}$ is the number of Monte Carlo steps and $\mathtt{m}$ is the index of the corresponding iteration. For $n = 5$ and 10, $\mathtt{M} = 2{,}000$ while for $n = 50$ and $n = 100$ due to the higher computational requirements $\mathtt{M} = 1{,}000$. Although, we know that for *OLS*, 2*SLS*, and *DOLS* we did not obtain weak convergence to a zero mean Gaussian mixture distribution, the Wald statistic is also calculated for these estimation methods. For the model (13) we only projected on the own leads an lags $\breve{\boldsymbol{\zeta}}_{it_\star}$ with the *DOLS* and the *D2SLS* estimator. For OLS and 2SLS we do not project on any leads and lags, e.g., $p = \emptyset$, while $\breve{\mathbf{Z}}_{p;it_\star}$ is replaced by $\breve{\mathbf{X}}_{p;it_\star}$ with the OLS and the DOLS estimator of $\breve{\gamma}$.

We tried to consider the cases where $T = 200$ and $n = 5, 10, 50$ and 100. With these n we investigated the *size* of the Wald statistic and obtained the percentages of the simulation runs where the true null hypothesis $\rho = 0$ was rejected at $\alpha_c = \{0.01, 0.05, 0.1\}$ significance levels. To obtain the *power* of the Wald type test we choose $\rho = \{-0.95, -0.5, -0.1, 0.1, 0.5, 0.95\}$ and investigate how often the *false* null-hypothesis of $\rho = 0$ is rejected.[13]

[11] For model (13) we work the the residuals $\widehat{\tilde{u}}_{it_\star}$.
[12] For more details in consistent estimation of the covariance see Section A-6.
[13] Tables A-1 to A-9 in the Online Appendix present summary statistics from these simulation runs for the leading case given by (8). For model (13) further results are provided in the Tables A-11 to A-13.

As a first result we observe that projecting on all leads and lags can only be performed for small n due to numerical constraints. We already observe that for $n = 5$ the performance of the *SD2SLS* estimators is poor (i.e. (A-24) discussed Online Appendix A-4 for model (8) and (33) for model (13)). Although the projection on all leads and lags is not necessary to obtain a mean zero Gaussian mixture limit distribution (Theorem1) for model (8), for model (13) where the estimator (33) is applied, theoretically correct estimates of the parameter $\check{\boldsymbol{\gamma}}$ can only be obtained when the cross-sectional dimension is small. For larger n, i.e. already for $n = 20$ and larger, the software package was hardly able/was not able to invert the matrices contained in the estimator (33). Hence, we were not able to perform a Monte Carlo study for model (13) with $n > 10$. In addition, for $n = 5$ and $n = 10$, Table 2 shows substantial oversizing for the estimator (33). The columns where (33) is applied are abbreviated by *SD2SLS*. Regarding the size of the Wald type test, the (second-order biased) estimators 2*SLS*, *DOLS*, and *D2SLS* (and still *OLS*) perform better than the estimator (33). For $\rho \neq 0$, and $n = 5$ or $n = 10$, the estimators reject the wrong null-hypothesis $\rho = 0$ in more than 88% of our simulation runs. That is, the power is acceptable. Online Appendix A-7 demonstrates that the bias and the root mean squared error of the estimator (33) are high. In addition, given a small n and $k_L > 0$, we observed very poor results for the estimator (43). Therefore, parameter estimation for $\boldsymbol{\beta}_L$ based on first using (33) to estimate $\boldsymbol{\beta} = \boldsymbol{\beta}_I$ and ρ and then applying the estimator (43) does not work in practice. We claim that this effect is caused by the properties of the estimator (33) and the fact that a small cross-sectional dimension n is available to estimate $\boldsymbol{\beta}_L$.

Hence, the remaining part of this section investigates the small sample properties of model (8) and the estimator (28). Regarding power, for $\rho = \{-0.95, -0.5, -0.1, 0.1, 0.5, 0.95\}$ the *false* null-hypothesis of $\rho = 0$ has been rejected in almost all of the simulation runs for the above simulation designs. When considering the $5 \times 6 \times 5 = 150$ different designs, where $\rho \neq 0$, we observe that even with $n = 5$, in almost 99% of all simulation runs the false null hypothesis was rejected at a 5% significance level (see the last rows of Table 1). The smallest rejection rates are observed with $\rho = \pm 0.1$ and the moving average designs $DGP = 4$ and $DGP = 5$. For $n = 10$, $n = 50$, and $n = 100$ we observed that the false null hypothesis has been rejected in almost all cases. To analyze the *size* of the Wald type test, the rejection rates of the Wald type test for the true null hypothesis $\rho = 0$ are investigated. The comparison of *D2SLS* to *DOLS* is of special interest. With $n = 5$, the oversizing remains modest for *DOLS* and *D2SLS*. The rejection rates observed are very similar, although *D2SLS* uses the instrumental variables where the numerical complexity is increased. With the moving average process stronger oversizing effects are obtained. The performance of *DOLS* is very close to the performance of our *D2SLS* estimator; here in some settings undersizing is observed. With 2*SLS* the oversizing observed is large, while substantial oversizing can be observed when *OLS* is applied. If the correlation of $\mathbf{v}_{it}$ and $u_{it}^{\dagger}$ is decreased (e.g., by choosing a diagonal $\boldsymbol{\Sigma}_{.\boldsymbol{\varepsilon}}$ or a *VAR* model with smaller eigenvalues), the oversizing behavior of *OLS* and 2*SLS* decreases. With small correlations, the performance of *OLS* and 2*SLS* is comparable to the performance of *DOLS* and *D2SLS*. There also exist data generating processes where the performance of the Wald type test for *DOLS* is much worse than for *D2SLS*. This takes place if $\Gamma_{0,u_i u_i}$ is large compared to the variance of $\mathbf{v}_{it}$. This effect can be expected by looking at the asymptotic bias term arising for *DOLS* (Proposition 2).

Remark 5. The question also arises whether the oversizing effect observed with the *D2SLS* estimator can be attributed to instrumental variable estimation, the choice of the instruments or the inclusion of common variables. Note that for $\rho = 0$ (and $W_{ii} = 0$, for $i = 1, \ldots, n$), the asymptotic bias of the *DOLS* estimator is zero.[14] By comparing the rejection rates of the *DOLS* and the *D2SLS* estimator, oversizing with *D2SLS* is approximately equal to—and some cases even smaller—than oversizing with *DOLS* (these effects are present with or without common variables).

[14]To see this, by Proposition 2 the asymptotic bias term is given by $\mathbf{W}_i \mathbf{K} \boldsymbol{\Gamma}_{0,uu_i}$ for $i = 1, \ldots, n$. For $\rho = 0$ we get $\mathbf{K} = (\mathbf{I}_n - 0\mathbf{W})^{-1} = \mathbf{I}_n$ and $\mathbf{W}_i \mathbf{K} = \mathbf{W}_i \mathbf{I}_n = \mathbf{W}_i$. By Assumption 1 $W_{ii} = 0$, while by Assumption 2 $\Gamma_{0,u_j u_i} = 0$ for $i \neq j$. This yields $\mathbf{W}_i \mathbf{K} \boldsymbol{\Gamma}_{0,uu_i} = \sum_{j=1}^{n} W_{ij} \Gamma_{0,u_j u_i} = 0$. The assumption that $\Gamma_{0,u_j u_i} \neq 0$, for $i \neq j$, is important to obtain an asymptotically unbiased DOLS estimator for $\rho = 0$.

Table 1. Model (8): Size and Power for the Parameter ρ, Simulation designs $DGP = 1 - 5$ and $W = (i) - (v)$: First four blocks: Rejections rates of the Wald type test for the *true* null hypothesis $\rho = 0$ in percentage terms, given the significance levels $\alpha_c = \{0.01, 0.05, 0.1\}$. $Q(\alpha)$ stands for the α quantile of the corresponding Monte Carlo samples. Last block: Rejections rates of the Wald type test for the *false* null hypothesis $\rho = 0$ in percentage terms, given the significance levels $\alpha_c = \{0.01, 0.05, 0.1\}$.

	$\alpha_c = 0.01$					$\alpha_c = 0.05$					$\alpha_c = 0.10$				
	OLS	*2SLS*	*DOLS*	*D2SLS*	*SD2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*SD2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*SD2SLS*
$n = 5, T = 200, \rho = 0$															
min	0.6	0.7	1.7	1.6	13.4	2.5	2.6	6.5	5.6	25.0	5.4	5.7	11.6	10.7	33.4
max	29.3	5.3	5.7	3.8	49.9	48.3	15.0	13.9	9.0	59.6	59.5	24.8	21.3	15.1	65.1
$Q(99)$	29.2	5.3	5.3	3.7	49.5	48.2	14.9	13.5	8.9	59.5	59.3	24.7	20.5	15.1	65.0
$Q(95)$	24.2	5.1	4.1	3.4	48.1	41.8	14.7	11.7	8.6	59.2	52.1	24.5	17.7	15.0	64.6
$Q(90)$	5.2	5.1	3.6	3.2	48.0	15.6	14.5	9.5	8.6	59.0	25.2	24.4	15.4	14.8	64.6
$n = 10, T = 200, \rho = 0$															
min	0.6	0.5	1.5	1.5		2.3	2.3	5.4	5.2		4.1	4.3	9.9	9.7	
max	3.1	3.3	2.4	2.4		10.1	11.6	7.9	7.7		18.0	19.3	13.8	13.7	
$Q(99)$	3.0	3.3	2.4	2.4		10.1	11.5	7.8	7.7		17.8	19.2	13.7	13.5	
$Q(95)$	3.0	3.3	2.2	2.2		9.9	10.9	7.7	7.7		17.0	18.5	13.4	12.9	
$Q(90)$	2.8	3.1	2.1	2.1		9.6	9.0	7.7	7.6		16.1	15.6	12.8	12.7	
$n = 50, T = 200, \rho = 0$															
min	0.2	0.2	0.2	0.2		3.6	3.8	2.0	2.0		9.2	8.6	5.4	5.4	
max	29.0	27.0	5.2	5.2		58.8	56.8	13.4	13.0		76.8	75.0	19.6	19.2	
$Q(99)$	29.0	27.0	5.2	5.1		58.6	56.8	13.4	12.9		76.7	75.0	19.5	19.2	
$Q(95)$	28.3	26.8	5.0	4.8		57.8	56.8	12.9	12.3		76.1	74.8	18.8	18.6	
$Q(90)$	26.2	26.4	4.8	4.5		57.5	56.6	11.7	11.8		75.4	74.0	18.1	17.2	
$n = 100, T = 200, \rho = 0$															
min	0.4	0.2	0.6	0.6		4.2	4.4	3.2	3.2		9.0	9.0	8.0	7.8	
max	56.2	53.0	4.6	4.6		79.2	77.6	13.6	13.2		89.0	88.8	19.2	20.0	
$Q(99)$	56.2	53.0	4.6	4.6		79.1	77.6	13.4	13.2		88.9	88.8	19.2	19.8	
$Q(95)$	56.0	52.8	4.3	4.4		78.7	77.2	12.6	13.1		88.6	88.4	19.0	19.1	
$Q(90)$	55.8	52.5	3.8	4.0		78.4	76.3	12.2	12.4		88.4	87.6	18.2	18.8	
$n = 5, T = 200, \rho \neq 0, \rho \in \{-0.9, -0.5, -0.1, 0.1, 0.5, 0.9\}$															
min	100.0	99.8	100.0	100.0	97.2	100.0	99.8	100.0	100.0	97.6	100.0	99.8	100.0	100.0	98.2
max	100.0	99.8	100.0	100.0	96.8	100.0	99.8	100.0	100.0	97.0	100.0	100.0	100.0	100.0	97.0
$Q(1)$	100.0	100.0	100.0	100.0	99.2	100.0	100.0	100.0	100.0	99.2	100.0	100.0	100.0	100.0	99.2
$Q(5)$	100.0	100.0	100.0	99.8	99.4	100.0	100.0	100.0	99.8	99.4	100.0	100.0	100.0	100.0	99.4
$Q(10)$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 2. Model (13): Size and Power for the Parameter ρ, Simulation designs $DGP = 1 - 5$ and $W = (i) - (v)$: First four blocks: Rejections rates of the Wald type test for the *true* null hypothesis $\rho = 0$ in percentage terms, given the significance levels $\alpha_c = \{0.01, 0.05, 0.1\}$. $Q(\alpha)$ stands for the α quantile of the corresponding Monte Carlo samples. Last two blocks: Rejections rates of the Wald type test for the *false* null hypothesis $\rho = 0$ in percentage terms, given the significance levels $\alpha_c = \{0.01, 0.05, 0.1\}$.

	$\alpha_c = 0.01$					$\alpha_c = 0.05$					$\alpha_c = 0.10$				
	OLS	*2SLS*	*DOLS*	*D2SLS*	*SD2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*SD2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*SD2SLS*
$n = 5, T = 200, \rho = 0$															
min	1.7	1.6	0.7	0.3	13.7	5.7	5.5	3.7	2.0	25.4	10.3	9.9	7.6	4.4	33.9
max	18.8	9.0	21.5	16.3	41.3	31.3	18.1	35.0	27.4	52.3	39.6	26.1	42.5	34.9	60.1
$Q(99)$	1.7	1.6	0.7	0.4	13.8	5.7	5.5	3.7	2.2	25.6	10.4	9.9	7.6	4.7	34.2
$Q(95)$	1.9	1.7	0.9	0.6	14.4	5.9	5.5	3.9	2.7	26.6	10.9	10.1	7.8	5.9	35.2
$Q(90)$	2.1	1.9	0.9	0.8	15.1	6.4	5.8	4.3	3.3	26.9	11.7	10.7	8.1	6.8	35.5
$n = 5, T = 200, \rho \neq 0, \rho \in \{-0.9, -0.5, -0.1, 0.1, 0.5, 0.9\}$,															
min	94.3	94.8	98.9	74.3	82.4	98.0	98.0	99.4	81.2	86.5	98.8	99.0	99.5	84.5	88.8
max	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$Q(1)$	95.4	95.3	99.0	75.9	83.7	98.2	98.5	99.4	82.1	88.0	99.0	99.1	99.6	85.0	89.8
$Q(5)$	98.1	98.6	100.0	79.1	89.6	98.8	99.3	100.0	85.7	92.4	99.4	99.6	100.0	88.8	93.7
$Q(10)$	100.0	100.0	100.0	84.7	94.5	100.0	100.0	100.0	89.1	96.0	100.0	100.0	100.0	90.6	96.5
$n = 10, T = 200, \rho = 0$															
min	1.2	1.0	0.4	0.1	30.8	4.1	4.3	1.8	0.7	42.4	9.8	9.0	5.0	3.1	48.6
max	15.6	9.0	21.0	7.2	87.0	28.4	17.0	37.0	15.0	89.6	37.4	26.4	43.8	20.6	92.2
$Q(99)$	1.2	1.0	0.4	0.1	31.0	4.4	4.3	2.1	0.9	43.0	10.1	9.0	5.2	3.1	49.4
$Q(95)$	1.4	1.0	0.6	0.2	31.9	5.5	4.6	3.2	1.4	44.8	11.3	9.4	6.2	3.2	51.8
$Q(90)$	1.8	1.5	0.7	0.2	32.6	7.2	5.9	3.7	1.4	45.2	12.5	11.7	7.2	3.3	52.6
$n = 10, T = 200, \rho \neq 0, \rho \in \{-0.9, -0.5, -0.1, 0.1, 0.5, 0.9\}$															
min	100.0	100.0	100.0	95.0	97.6	100.0	100.0	100.0	96.0	98.6	100.0	100.0	100.0	96.8	98.6
max	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$Q(1)$	100.0	100.0	100.0	95.8	98.4	100.0	100.0	100.0	96.7	98.7	100.0	100.0	100.0	97.2	98.8
$Q(5)$	100.0	100.0	100.0	98.8	99.6	100.0	100.0	100.0	99.2	99.6	100.0	100.0	100.0	99.5	99.6
$Q(10)$	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Summing up, we observe that the estimator (28) exhibits (in most cases) some oversizing behavior as already observed in the literature where dynamic least squares estimation has been applied (see, e.g., Mark and Sul, 2003). However, even with the true null-hypothesis $\rho = 0$, where no spatial endogenetiy is present, the *D2SLS* estimator in most cases outperforms the *DOLS*, the *OLS* as well as the 2*SLS* estimator.

5. Empirical illustration

In this section, we apply the tools developed in the former sections to credit risk data. To model corporate default swap (CDS) spreads we follow Berndt et al. (2008) and use the distance to default, the debt to value ratio, interest rates and the VIX volatility index as explanatory variables. By the matrix $\mathbf{W}$ we model a specific form of default risk correlation, where $\mathbf{W}$ will be derived from input–output data obtained the *Bureau of Labor Statistics* (BLS). The CDS dataset already used in Schneider et al. (2010), comprises CDS spreads of 278 firms obtained from the *Markit Group*. We focus on the five year maturities which are typically the most liquid ones. The observation period is January 2, 2001 to May 30, 2008. In line with a bulk of quantitative finance literature we stick to weekly data, such that $T = 230$. The CDS data are matched with firm specific characteristics obtained from *Thomson Datastream* and *Compustat* data. We construct the KMV distance to default, DD_{it}, from firm specific data by following Crosbie and Bohn (2003). Moreover, we calculate the debt to value ratio, DVR_{it}. DVR_{it} is measured in percentage terms. We also include the *VIX* volatility from the *Chicago Board Options Exchange* (http://www.cboe.com/micro/VIX/vixintro.aspx) as an explanatory variable. Additionally, we include a the year interest rate, denoted by r_{2t} and measured in percentage terms, from the *Federal Reserve* (http://federalreserve.gov/releases/h15/data.htm). After matching the firm specific data with the CDS data and excluding observations where data problems are observed, with work with a cross-section of $n = 148$, $y_{it} = \ln CDS_{it}$, the common variables are $\mathbf{x}_{Ct} = (r_{2t}, VIX_t)'$, while $\mathbf{x}_{Iit} = (DD_{it}, \ln DVR_{it})'$.[15]

Using our data set, we apply model (8) and estimate the parameter vector $\boldsymbol{\gamma}$ by two-stage least squares, *DOLS*, *OLS*, and *D2SLS*. The results are presented in Table 3. Based on our theoretical results, only the *D2SLS* estimator should be used. The results from the other estimation methods are included only for comparison. When instrumental variables are used in the estimation, the logarithm of the distance to default and the debt-to-value ratio are used in (24), i.e., $q_\rho = 2$. All the p-values presented in Table 3 are obtained by a Wald type test as described in Theorem 1.

For the distance to default and the debt to value ratio the parameters are highly significant and have the signs expected from finance literature. The impact of the short term interest rate r_{2t} is significant as well. When the short term interest rate r_{2t} increases, the logarithm of the CDS spread decreases. The *VIX* volatility index is not significant when *D2SLS* estimation is performed and default significance levels (1%, 5%, 10%) are applied. With the dynamic two stage least squares estimator the spatial correlation parameter ρ is positive as expected and highly significant.

Table 3. Parameter Estimates for model (8) applied to CDS data. The response variable y_{it} is the natural logarithm of the CDS spread on a firm level. The explanatory variables are the distance to default, DD_{it}, the logarithm of the debt to value ratio, $\ln DVR_{it}$, a two year bond yield r_{2t} and the VIX volatility index VIX_t. $T = 230$, $n = 148$, $p = 2$ leads and lags are used; the number of instruments is $q_\rho = 2$.

$\hat{\gamma}$	*OLS*		*2SLS*		*DOLS*		*D2SLS*	
ρ	0.8023	< 0.001	0.5413	< 0.001	0.7861	< 0.001	0.5021	< 0.001
β_{DD}	−0.0434	0.0059	−0.0516	0.0017	−0.0751	< 0.001	−0.0893	< 0.001
$\beta_{\ln DVR}$	0.4233	< 0.001	0.4519	< 0.001	0.4220	< 0.001	0.4509	< 0.001
β_{r2}	−0.1354	< 0.001	−0.1616	< 0.001	−0.1209	< 0.001	−0.1468	< 0.001
β_{VIX}	0.0007	0.3225	0.0009	0.2018	−0.0006	0.2720	−0.0006	0.2726

[15]For more details on the data see Section A-8 in the Online Appendix.

6. Conclusions

In this paper, we studied panel data models with a cointegration relationship including a spatial lag. Due to this spatial lag, standard estimation techniques do not provide us with appropriate tools to estimate the parameters and to perform inference. Based on this problem we stick to the usual assumptions used in the dynamic least squares estimation and develop a dynamic two stage least squares estimator. We show that the parameter vector of interest is asymptotically independent of the nuisance parameters. Moreover, we derive the asymptotic distribution of the parameters when the time-series dimension becomes large. Convergence to a zero mean Gaussian mixture is attained, which also allows the application of a Wald type test. In addition, a limit result, where the time-series and the cross-sectional dimension become large is obtained. The limit distribution is a Gaussian distribution.

Our estimation methodology is applied to simulated data to investigate the small sample properties, and to financial data to test for the impact of spatial correlation on credit default swap spreads. Given this financial data set and a spatial correlation matrix obtained from input–output data, our analysis shows that spatial correlation is highly significant.

A. Proof of the Propositions 1–3 as well as Theorems 1 and 2

The estimators considered in Section 3 can be expressed by means of

$$\widehat{(\boldsymbol{\gamma}',\underline{\boldsymbol{\delta}}_p')}'^{(\mathrm{m})} - (\boldsymbol{\gamma}',\underline{\boldsymbol{\delta}}_p')' = \left(\widetilde{\mathbf{X}}_p'^{(\mathrm{m})}\mathscr{P}_{Hp}\widetilde{\mathbf{X}}_p^{(\mathrm{m})}\right)^{-1}\widetilde{\mathbf{X}}_p'^{(\mathrm{m})}\mathscr{P}_{Hp}^{(\mathrm{m})}\underline{\mathbf{u}}_p^{(\mathrm{m})}\ ,\quad \mathscr{P}_{Hp} = \widetilde{\mathbf{Z}}_p^{(\mathrm{m})}\left(\widetilde{\mathbf{Z}}_p'^{(\mathrm{m})}\widetilde{\mathbf{Z}}_p^{(\mathrm{m})}\right)^{-1}\widetilde{\mathbf{Z}}_p'^{(\mathrm{m})}\ ,\ \text{and}$$

$$\left(\widehat{\breve{\boldsymbol{\gamma}}}',\widehat{\breve{\underline{\boldsymbol{\delta}}}}_{\sharp p}'\right)' - \left(\breve{\boldsymbol{\gamma}}',\breve{\underline{\boldsymbol{\delta}}}_{\sharp p}'\right)' = \left(\breve{\mathbf{X}}_p'\mathscr{P}_{\breve{H}p}\breve{\mathbf{X}}_p\right)^{-1}\breve{\mathbf{X}}_p'\mathscr{P}_{\breve{H}p}\underline{\breve{\mathbf{u}}}_p\ ,\quad \mathscr{P}_{\breve{H}p} = \breve{\mathbf{Z}}_p\left(\breve{\mathbf{Z}}_p'\breve{\mathbf{Z}}_p\right)^{-1}\breve{\mathbf{Z}}_p'\ , \tag{45}$$

where $\mathrm{m} = OLS, \ldots, D2SLS$. For OLS and $2SLS$, $p = \emptyset$, while for OLS and $DOLS$ we have $\widetilde{\mathbf{Z}}_p^{(\mathrm{m})} = \widetilde{\mathbf{X}}_p^{(\mathrm{m})}$, such that $\widetilde{\mathbf{Z}}_p^{(\mathrm{m})}$ is equal to the identity matrix. $\widetilde{\mathbf{X}}_p^{(\mathrm{m})}$ is a $T_\star n \times 1 + k + (2p+1)k \cdot n$ matrix while $\widetilde{\mathbf{Z}}_p^{(\mathrm{m})}$ is of dimension $T_\star n \times q_\rho + k + (2p+1)k \cdot n$, where $\widetilde{\mathbf{X}}_{it_\star;p}^{(\mathrm{m})}$ and $\widetilde{\mathbf{Z}}_{it_\star;p}^{(\mathrm{m})}$ are the transpose of the $it_\star$ elements of $\widetilde{\mathbf{X}}_p^{(\mathrm{m})}$ and $\widetilde{\mathbf{Z}}_p^{(\mathrm{m})}$. The variables containing the suffix $\breve{}$ were defined in Section 3.3. Then, with $r \in [0,1]$, we obtain for $DOLS$ and $D2SLS$

$$T_\star^{0.5}\cdot\mathbf{A}_{\tilde{X}p}\widetilde{\mathbf{X}}_{i[rT_\star];p}^{(\mathrm{m})} = \begin{pmatrix} T_\star^{0.5}\cdot\frac{1}{T_\star}\cdot\left(\mathbf{W}_i\mathbf{K}\left(\tilde{\boldsymbol{\beta}}\mathbf{C}\tilde{\mathbf{x}}_{[rT_\star]} + \tilde{\boldsymbol{\zeta}}_{p;i[rT_\star]}\underline{\boldsymbol{\delta}}_p + \underline{\tilde{\mathbf{u}}}_{[rT_\star]}^{(\mathrm{m})}\right)\right) \\ T_\star^{0.5}\cdot\frac{1}{T_\star}\cdot\tilde{\mathbf{x}}_{i[rT_\star]} \\ \mathbf{0}_{((2p+1)k\cdot(i-1)\times 1)} \\ \tilde{\boldsymbol{\zeta}}_{p;i[rT_\star]} \\ \mathbf{0}_{((2p+1)k\cdot(n-i-1)\times 1)} \end{pmatrix} \quad \text{and}$$

$$T_\star^{0.5}\cdot\mathbf{A}_{\breve{X}p}\breve{\mathbf{X}}_{i[rT_\star];p} = \begin{pmatrix} T_\star^{0.5}\cdot\frac{1}{T_\star}\cdot\left(\mathbf{W}_i\mathbf{K}\left((\mathbf{I}_n\otimes\breve{\boldsymbol{\beta}})\breve{\mathbf{x}}_{I[rT_\star]} + \breve{\boldsymbol{\zeta}}_{\sharp p;[rT_\star]}\breve{\underline{\boldsymbol{\delta}}}_p + \underline{\breve{\mathbf{u}}}_{[rT_\star]}\right)\right) \\ T_\star^{0.5}\cdot\frac{1}{T_\star}\cdot\breve{\mathbf{x}}_{Ii[rT_\star]} \\ \mathbf{0}_{((2p+1)kn_I\cdot(i-1)\times 1)} \\ \breve{\boldsymbol{\zeta}}_{\sharp p;[rT_\star]} \\ \mathbf{0}_{((2p+1)k_In\cdot(n-i-1)\times 1)} \end{pmatrix}. \tag{46}$$

By Assumption 2 the components 1 to $k+1$ provided in (46) weakly converge to $\mathbf{g}_i(r)$ and $\breve{\mathbf{g}}_i(r)$ as $T \to \infty$ (defined in (21) as well as (34) in the main text). For the OLS and the $2SLS$ estimator, (46) implies that $\left[T_\star^{0.5}\cdot\mathbf{A}_{\tilde{Z}p}\widetilde{\mathbf{X}}_{i[rT_\star];p=}^{(\mathrm{m})}\right]_{(1:k+1,1)}$ converges weakly to $\mathbf{g}_i(r)$. In addition, for $D2SLS$ we get $T_\star^{0.5}\cdot\mathbf{A}_{\tilde{Z}p}\widetilde{\mathbf{Z}}_{i[rT_\star];p}^{\mathrm{m}} = T_\star^{0.5}\cdot\mathbf{A}_{\tilde{Z}p}\left(T_\star^{-0.5}\tilde{\mathbf{z}}_{i[rT_\star]}^{*}, T_\star^{-0.5}\tilde{\mathbf{x}}_{i[rT_\star]}', \mathbf{0}_{((2p+1)k(i-1)\times 1)}', \tilde{\boldsymbol{\zeta}}_{p;i[rT_\star]}', \mathbf{0}_{((2p+1)k\cdot(n-i-1)\times 1)}'\right)'$, where the first $k+q_\rho$ components weakly converge to $\mathbf{h}_i(r)$ as $T \to \infty$ (Proposition 3). Also for the $2SLS$ estimator we observe that the $T \to \infty$ limit of $T_\star^{-0.5}\widetilde{\mathbf{Z}}_{i[rT_\star];p}^{(2SLS)}$ is $\mathbf{h}_i(r)$. For the first $q_\rho + k_I$ coordinates of $T_\star^{0.5} \cdot \mathbf{A}_{\breve{Z}p}\breve{\mathbf{Z}}_{i[rT_\star];p}^{\mathrm{m}}$ we derive weak convergence to $\breve{\mathbf{h}}_i(r)$ defined in (34). Note that for $k_C = 0$, $\boldsymbol{\beta} = \boldsymbol{\beta}_I$, $k = k_I$, and $\mathbf{C} = \mathbf{I}_{nk_I}$.

Step 1 (*Asymptotic Limit Distribution*): First we consider:

$$\mathbf{M}^\star_{\tilde{X}\tilde{Z},nTi} := \sum_{t_\star=1}^{T_\star} \mathbf{A}_{\tilde{X}p}\tilde{\mathbf{X}}_{it_\star;p}\tilde{\mathbf{Z}}'_{it_\star;p}\mathbf{A}_{\tilde{Z}p} = \frac{1}{T_\star}\sum_{t_\star=1}^{T_\star} \begin{pmatrix} T_\star^{-0.5}\tilde{y}^*_{it_\star} \\ T_\star^{-0.5}\tilde{\mathbf{x}}_{it_\star} \\ \mathbf{0}_{((2p+1)k\cdot(i-1)\times 1)} \\ \tilde{\boldsymbol{\zeta}}_{p;it_\star} \\ \mathbf{0}_{((2p+1)k\cdot(n-i-1)\times 1)} \end{pmatrix} \begin{pmatrix} T_\star^{-0.5}\tilde{\mathbf{z}}^*_{it_\star} \\ T_\star^{-0.5}\tilde{\mathbf{x}}_{it_\star} \\ \mathbf{0}_{((2p+1)k\cdot(i-1)\times 1)} \\ \tilde{\boldsymbol{\zeta}}_{p;it_\star} \\ \mathbf{0}_{((2p+1)k\cdot(n-i-1)\times 1)} \end{pmatrix}' . \tag{47}$$

By Theorem 30.2 in Davidson (1994) $\left[\lim_{T\to\infty} \mathbf{M}^\star_{\tilde{X}\tilde{Z},nTi}\right]_{(1:k+1,1:k+q_\rho)} = \int_0^1 \mathbf{g}_i(r)\mathbf{h}_i(r)'dr = \mathbf{M}_{\tilde{X}\tilde{Z},ni}$. In addition, by a law of large numbers (see, e.g., White, 2001, Chapter 3.2), the terms in the south-east converge in probability to elements of $\boldsymbol{\Gamma}_{\ell,vv}$. Last but not least the terms in the south-west and the terms in the north-east convergence weakly when scaled by $1/T$, that is $\frac{1}{T_\star}\sum_{t_\star=1}^{T_\star} \tilde{\mathbf{x}}_{it_\star}\underline{\tilde{u}}_{it_\star} \Rightarrow \int_0^1 \tilde{\mathcal{B}}_{v_i}(r)d\mathcal{B}^\dagger_{u_j}(r) + \boldsymbol{\Delta}^\dagger_{v_iu_i}$ such that $T_\star^{-3/2}\sum_{t_\star=1}^{T_\star} \tilde{\mathbf{x}}_{it_\star}\underline{\tilde{u}}^\dagger_{it_\star}$ converges to zero in probability. The same result is observed for $T_\star^{-3/2}\sum_{t_\star=1}^{T_\star} \tilde{\mathbf{x}}_{it_\star}\underline{\tilde{u}}_{it_\star}$. Hence, $\left[\mathbf{M}^\star_{\tilde{X}\tilde{Z},ni}\right]_{(1:k+1,k+q_\rho+1:k+q_\rho+(2p+1)nk)} = \left[\mathbf{M}^\star_{\tilde{Z}\tilde{Z},ni}\right]'_{(k+q_\rho+(2p+1)nk,1:k+1)} = \mathbf{0}_{(k+1\times(2p+1)nk)}$. In the same way $\mathbf{M}^\star_{\tilde{Z}\tilde{Z},nTi}$ converges to $\left[\mathbf{M}^\star_{\tilde{Z}\tilde{Z},nTi}\right]_{(1:k+1,1:k+q_\rho)} = \int_0^1 \mathbf{h}_i(r)\mathbf{h}_i(r)'dr = \mathbf{M}_{\tilde{Z}\tilde{Z},ni}$, the south-eastern block contains elements of $\boldsymbol{\Gamma}_{\ell,vv}$, while the remaining blocks are zero. $\mathbf{M}^\star_{\tilde{X}\tilde{X},nTi}$ converges to $\left[\mathbf{M}^\star_{\tilde{X}\tilde{X},nTi}\right]_{(1:k+1,1:k+1)} = \int_0^1 \mathbf{g}_i(r)\mathbf{g}_i(r)'dr = \mathbf{M}_{\tilde{X}\tilde{X},ni}$, the south-eastern block contains elements of $\boldsymbol{\Gamma}_{\ell,vv}$, while the remain blocks are zero. The same steps also apply to $\mathbf{M}^\star_{\breve{X}\breve{Z},nTi}$ and $\mathbf{M}^\star_{\breve{Z}\breve{Z},nTi}$.

In the second step we obtain the limit of the terms containing $\underline{\tilde{u}}_{it}^{(\mathrm{m})}$ or $\underline{\breve{u}}_{it}$: For the OLS estimator we obtain the limit of $\frac{1}{T}\sum_{t=1}^{T} \tilde{\mathbf{X}}_{it}^{OLS}\tilde{u}^\dagger_{it}$ by applying Assumptions 1 and 2 as follows:

$$\begin{aligned} \mathbf{m}_{\tilde{X}\tilde{u}^\dagger,nTi} &= \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} \mathbf{W}_i\mathbf{K}\left(\tilde{\boldsymbol{\beta}}\mathbf{C}\tilde{\mathbf{x}}_t + \tilde{\boldsymbol{\zeta}}_{p,t}\underline{\boldsymbol{\delta}}_p + \underline{\tilde{\mathbf{u}}}_t\right) \\ \tilde{\mathbf{x}}_{it} \end{pmatrix} \tilde{u}^\dagger_{it} \\ &\Rightarrow \begin{pmatrix} \mathbf{W}_i\mathbf{K}\left[\int_0^1 \tilde{\boldsymbol{\beta}}\left[\mathbf{C}\tilde{\mathcal{B}}_v(r)d\mathcal{B}^\dagger_{u_i}(r) + \mathbf{C}\boldsymbol{\Delta}^\dagger_{vu_i}\right] + \boldsymbol{\Gamma}^\dagger_{0,uu_i}\right] \\ \sqrt{\Omega^\dagger_{u_iu_i}}\int_0^1 \tilde{\mathcal{B}}_{v_i}(r)d\mathcal{W}^\dagger_{u_i}(r) + \boldsymbol{\Delta}^\dagger_{v_iu_i} \end{pmatrix} = \mathbf{m}_{\tilde{X}\tilde{u}^\dagger,ni}\,. \end{aligned}$$

Then, $\mathbf{m}_{\tilde{X}\tilde{u}^\dagger,n} = \sum_{i=1}^{n} \mathbf{m}_{\tilde{X}\tilde{u}^\dagger,ni}$ and $T\left(\widehat{\boldsymbol{\gamma}}_{OLS} - \boldsymbol{\gamma}\right) \Rightarrow \mathbf{M}^{-1}_{\tilde{X}\tilde{X},n}\mathbf{m}_{\tilde{X}\tilde{u}^\dagger,n}$, by the continuous mapping theorem. This proves Proposition 1(b). For the *DOLS* estimator we have to obtain the limit of

$$\mathbf{m}^\star_{\tilde{X}\tilde{u},nTi} := \frac{1}{T_\star^{0.5}}\sum_{t_\star=1}^{T_\star} \begin{pmatrix} T_\star^{-0.5}\tilde{y}^*_{it_\star} \\ T_\star^{-0.5}\tilde{\mathbf{x}}_{it_\star} \\ \mathbf{0}_{((2p+1)k\cdot(i-1)\times 1)} \\ \tilde{\boldsymbol{\zeta}}_{p;it_\star} \\ \mathbf{0}_{((2p+1)k\cdot(n-i-1)\times 1)} \end{pmatrix} \underline{\tilde{u}}_{it} = \sum_{t_\star=1}^{T_\star} \begin{pmatrix} T_\star^{-1}\mathbf{W}_i\mathbf{K}\left(\tilde{\boldsymbol{\beta}}\mathbf{C}\tilde{\mathbf{x}}_{t_\star} + \tilde{\boldsymbol{\zeta}}_{p,t_\star}\underline{\boldsymbol{\delta}}_p + \underline{\tilde{\mathbf{u}}}_{t_\star}\right) \\ T_\star^{-1}\tilde{\mathbf{x}}_{it_\star} \\ \mathbf{0}_{((2p+1)k\cdot(i-1)\times 1)} \\ T_\star^{-0.5}\tilde{\boldsymbol{\zeta}}_{p;it_\star} \\ \mathbf{0}_{((2p+1)k\cdot(n-i-1)\times 1)} \end{pmatrix} \underline{\tilde{u}}_{it}\,. \tag{48}$$

By Assumptions 1–3, the $T \to \infty$-limit of $\frac{1}{T_\star}\sum_{t_\star=1}^{T_\star} \tilde{\boldsymbol{\zeta}}_{p,t_\star}\tilde{u}_{it} = \mathbf{0}_{((2p+1)nk\times 1)}$. Moreover, $T_\star^{-1}\sum_{t_\star=1}^{T_\star} \underline{\tilde{\mathbf{u}}}_{t_\star}\tilde{u}_{it}$ converges to $\boldsymbol{\Gamma}_{0,uu_i}$ by a law of large numbers and the fact that $|\underline{\tilde{\mathbf{u}}}_{t_\star} - \tilde{\mathbf{u}}_{t_\star}|$ goes to zero sufficiently fast by

Assumption 3. Hence, we observe that

$$\left[\mathbf{m}^{\star}_{\tilde{X}\tilde{u},nTi}\right]_{(1:k+1)} \Rightarrow \left[\lim_{T\to\infty} \mathbf{m}^{\star}_{\tilde{X}\tilde{u},nTi}\right]_{(1:k+1)} = \mathbf{m}_{\tilde{X}\tilde{u},ni} = \begin{pmatrix} \mathbf{W}_i\mathbf{K}\left(\int_0^1 \tilde{\boldsymbol{\beta}}\mathbf{C}\tilde{\mathcal{B}}_v(r)d\mathcal{B}_{u_i}(r) + \boldsymbol{\Gamma}_{0,uu_i}\right) \\ \sqrt{\Omega_{u_iu_i}}\int_0^1 \tilde{\mathcal{B}}_{v_i}(r)d\mathcal{W}_{u_i}(r) \end{pmatrix}.$$

To obtain the limit of $\left[\lim_{T\to\infty} \mathbf{m}^{\star}_{\tilde{X}\tilde{u},nTi}\right]_{(k+2:(2p+1)nk)}$ a central limit theorem was assumed to hold in Assumption 2. From (49) we obtain $\mathbf{m}_{\tilde{X}\tilde{u},n} = \sum_{i=1}^n \mathbf{m}_{\tilde{X}\tilde{u},ni}$. Since $\mathbf{M}_{\tilde{X}\tilde{X},n}$ is a regular matrix, we observe that $T_\star\left(\widehat{\boldsymbol{\gamma}}_{DOLS;p} - \boldsymbol{\gamma}\right) \Rightarrow \mathbf{M}^{-1}_{\tilde{X}\tilde{X},n}\mathbf{m}_{\tilde{X}\tilde{u},n}$. This proves Proposition 2.(b).

For 2*SLS* we consider the limit of $\mathbf{m}_{\tilde{Z}\tilde{u}^\dagger,nTi} := \sum_{i=1}^n \frac{1}{T}\sum_{t=1}^T \widetilde{\mathbf{Z}}^{2SLS}_{it}\tilde{u}^\dagger_{it}$. We observe that

$$\frac{1}{T}\sum_{t=1}^T \begin{pmatrix} \mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\tilde{\mathbf{x}}_t \\ \vdots \\ \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\tilde{\mathbf{x}}_t \\ \tilde{\mathbf{x}}_{it} \end{pmatrix}\tilde{u}^\dagger_{it} \Rightarrow \begin{pmatrix} \mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\left(\int_0^1 \tilde{\mathcal{B}}_v(r)dB^\dagger_{u_i}(r) + \boldsymbol{\Delta}^\dagger_{vu_i}\right) \\ \vdots \\ \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\left(\int_0^1 \tilde{\mathcal{B}}_v(r)dB^\dagger_{u_i}(r) + \boldsymbol{\Delta}_{vu_i}\right) \\ \int_0^1 \tilde{\mathcal{B}}_{v_i}(r)dB^\dagger_{u_i}(r) + \boldsymbol{\Delta}^\dagger_{vu_i} \end{pmatrix}$$

$$= \int_0^1 \mathbf{h}_i(r)dB^\dagger_{u_i}(r) + \begin{pmatrix} \mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\boldsymbol{\Delta}^\dagger_{vu_i} \\ \vdots \\ \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\boldsymbol{\Delta}^\dagger_{vu_i} \\ \boldsymbol{\Delta}^\dagger_{v_iu_i} \end{pmatrix} = \mathbf{m}_{\tilde{Z}\tilde{u}^\dagger,ni}. \quad (49)$$

$\mathbf{m}_{\tilde{Z}\tilde{u}^\dagger,n} = \sum_{i=1}^n \mathbf{m}_{\tilde{Z}\tilde{u}^\dagger,ni}$ Then, the continuous mapping theorem results in $T\left(\widehat{\boldsymbol{\gamma}}_{2SLS} - \boldsymbol{\gamma}\right) \Rightarrow \left(\mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{M}'_{\tilde{X}\tilde{Z},n}\right)^{-1}\mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{m}_{\tilde{Z}\tilde{u}^\dagger,n}$, which proves Proposition 3(b).

Finally, for the *D2SLS* estimator (28), we derive the $T\to\infty$-limit of $\mathbf{m}^{\star}_{\tilde{Z}\tilde{u},nT} = \sum_{i=1}^n \mathbf{m}^{\star}_{\tilde{Z}\tilde{u},nTi}$, where

$$\mathbf{m}^{\star}_{\tilde{Z}\tilde{u},nTi} = \sum_{t=1}^{T_\star}\mathbf{A}_{\tilde{Z}p}\begin{pmatrix} \mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\tilde{\mathbf{x}}_{t_\star} \\ \vdots \\ \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\tilde{\mathbf{x}}_{t_\star} \\ \tilde{\mathbf{x}}_{it_\star} \\ \mathbf{0}_{((2p+1)k\cdot(i-1)\times 1)} \\ \tilde{\boldsymbol{\zeta}}_{p;it_\star} \\ \mathbf{0}_{((2p+1)k\cdot(n-i-1)\times 1)} \end{pmatrix} u_{it_\star} \Rightarrow \begin{pmatrix} \mathbf{W}_i^{\tau_{\mathbb{K}(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\int_0^1 \tilde{\mathcal{B}}_v(r)dB_{u_i}(r) \\ \vdots \\ \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}}\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\int_0^1 \tilde{\mathcal{B}}_v(r)dB_{u_i}(r) \\ \int_0^1 \tilde{\mathcal{B}}_{v_i}(r)dB_{u_i}(r) \\ \mathbf{0}_{((2p+1)k\cdot(i-1)\times 1)} \\ \boldsymbol{\nu}_{(\tilde{\zeta}_{p;it_\star}\tilde{u}_{it})} \\ \mathbf{0}_{((2p+1)k\cdot(n-i-1)\times 1)} \end{pmatrix}. \quad (50)$$

Since $\mathbf{v}_{i_\star}$ and $u_{i_\star}$ are uncorrelated, $\boldsymbol{\Delta}_{v_iu_i} = \mathbf{0}_{(k)}$ and therefore no correlation terms show up in (50). To obtain $\lim_{T\to\infty} T_\star^{-0.5}\sum \tilde{\boldsymbol{\zeta}}_{p;it_\star}\tilde{u}_{it_\star}$ we assumed that a central limit theorem can be applied and convergence to a normally distribution vector $\boldsymbol{\nu}_{(\tilde{\zeta}_{p;it_\star}\tilde{u}_{it})}$ with mean zero takes place. For the first $k+q_\rho$ components we observe that $\left[\mathbf{m}^{\star}_{\tilde{Z}\tilde{u},ni}\right]_{(1:k+q_\rho)} = \int_0^1 \mathbf{h}_i(r)d\mathcal{B}_{u_i}(r) = \mathbf{m}_{\tilde{Z}\tilde{u},ni}$.

Hence, by the block diagonal structure of $\mathbf{M}^{\star}_{\tilde{X}\tilde{Z},n}$ and $\mathbf{M}^{\star}_{\tilde{Z}\tilde{Z},nT}$, the invertability of $\mathbf{M}_n = \mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},nT}\mathbf{M}'_{\tilde{X}\tilde{Z},n}$ and the continuous mapping theorem, we observe that $T_\star\left(\widehat{\boldsymbol{\gamma}}_{D2SLS;p} - \boldsymbol{\gamma}\right)$ converges weakly to $\mathbf{M}_n^{-1}\mathbf{m}_n$, where $\mathbf{m}_n = \sum_{i=1}^n \mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{m}_{\tilde{Z}\tilde{u},ni} = \mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{m}_{\tilde{Z}\tilde{u},n}$. In addition, $\mathcal{B}_v$ and $\mathcal{B}_{u_i}$ are uncorrelated and therefore independent, no second-order bias terms show up. Hence, we observe converge to zero mean Gaussian mixture distribution. This proves Theorem 1(b).

In a similar way, for the *D2SLS* estimator (33) we derive the $T \to \infty$-limit of $\mathbf{m}^{\star}_{\breve{Z}\breve{u},nT} = \sum_{i=1}^{n} \mathbf{m}^{\star}_{\breve{Z}\breve{u},nTi}$. Using (A-8) obtained in Online Appendix A-2, we derive

$$\mathbf{m}^{\star}_{\breve{Z}\breve{u},nTi} = \sum_{t=1}^{T_\star} \mathbf{A}_{\breve{Z}p} \begin{pmatrix} \mathbf{W}_i^{\tau_{\mathbb{K}(1)}} \mathbf{C}_{(\mathbb{K}_{(1)})} \breve{\mathbf{x}}_{t_\star} \\ \vdots \\ \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}} \mathbf{C}_{(\mathbb{K}_{(q_\rho)})} \breve{\mathbf{x}}_{t_\star} \\ \breve{\mathbf{x}}_{it_\star} \\ \mathbf{0}_{((2p+1)kn\cdot(i-1)\times 1)} \\ \breve{\boldsymbol{\zeta}}_{\sharp p;t_\star} \\ \mathbf{0}_{((2p+1)kn\cdot(n-i-1)\times 1)} \end{pmatrix} u_{it_\star} \Rightarrow \begin{pmatrix} \mathbf{m}_{\breve{Z}\breve{u},ni} \\ \mathbf{0}_{((2p+1)kn\cdot(i-1)\times 1)} \\ \boldsymbol{\nu}_{(\breve{\zeta}_{\sharp p;t_\star} \breve{u}_{it})} \\ \mathbf{0}_{((2p+1)kn\cdot(n-i-1)\times 1)} \end{pmatrix}, \quad \text{where}$$

$$\mathbf{m}_{\breve{Z}\breve{u},ni} := \begin{pmatrix} \int_0^1 \mathbf{W}_i^{\tau_{\mathbb{K}(1)}} \mathbf{C}_{(\mathbb{K}_{(1)})} \left[\breve{\mathcal{B}}_v(r) \left(d\mathcal{B}_{u_i}(r) - \frac{1}{n}\sum_{j=1}^n d\mathcal{B}_{u_j}(r)\right)\right] \\ \vdots \\ \int_0^1 \mathbf{W}_i^{\tau_{\mathbb{K}(q_\rho)}} \mathbf{C}_{(\mathbb{K}_{(q_\rho)})} \left[\breve{\mathcal{B}}_v(r) \left(d\mathcal{B}_{u_i}(r) - \frac{1}{n}\sum_{j=1}^n d\mathcal{B}_{u_j}(r)\right)\right] \\ \int_0^1 \breve{\mathcal{B}}_{v_i}(r) \left(d\mathcal{B}_{u_i}(r) - \frac{1}{n}\sum_{j=1}^n d\mathcal{B}_{u_j}(r)\right) \end{pmatrix}. \tag{51}$$

The term $\boldsymbol{\nu}_{(\breve{\zeta}_{\sharp p;t_\star} \breve{u}_{it})}$ is normally distributed with mean zero. Therefore, by the block diagonal structure of $\mathbf{M}^{\star}_{\breve{X}\breve{Z},n}$ and $\mathbf{M}^{\star}_{\breve{Z}\breve{Z},nT}$, the invertability of $\breve{\mathbf{M}}_n = \mathbf{M}_{\breve{X}\breve{Z},n}\mathbf{M}^{-1}_{\breve{Z}\breve{Z},nT}\mathbf{M}'_{\breve{X}\breve{Z},n}$ and the continuous mapping theorem, we observe that $T_\star\left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS;p} - \breve{\boldsymbol{\gamma}}\right)$ converges weakly to $\breve{\mathbf{M}}_n^{-1}\breve{\mathbf{m}}_n$, where $\breve{\mathbf{m}}_n = \sum_{i=1}^n \mathbf{M}_{\breve{X}\breve{Z},n}\mathbf{M}^{-1}_{\breve{Z}\breve{Z},n}\mathbf{m}_{\breve{Z}\breve{u},ni}$.

`Step 2` (*Asymptotic independence*): By this block diagonal structure of the limits of $\mathbf{M}^{\star}_{\breve{Z}\breve{Z},nT}$ and $\mathbf{M}^{\star}_{\breve{X}\breve{Z},nT}$, we observe that $T_\star(\widehat{\boldsymbol{\gamma}}_p^{(\mathtt{m})} - \boldsymbol{\gamma})$ and $\sqrt{T_\star}(\widehat{\underline{\boldsymbol{\delta}}}_p^{(\mathtt{m})} - \underline{\boldsymbol{\delta}})$ are uncorrelated for $\mathtt{m} = OLS$, *DOLS*, *2SLS* and *D2SLS*. Due to the properties of Brownian motion and normal random variables these terms are independent. In same way we observe that $\mathbf{M}^{\star}_{\breve{X}\breve{Z},nTi}$ and $\mathbf{M}^{\star}_{\breve{Z}\breve{Z},nTi}$ converge to a block diagonal matrices, where the block in the south east contains covariance matrices of the leads and lags of $\breve{\mathbf{v}}_{t_\star}$ and $\breve{\mathbf{v}}_{t_\star}$. By the continuous mapping theorem (see, e.g., Klenke, 2008, p. 257), for the elements in the north-west we obtain $\left[\lim_{T\to\infty}\mathbf{M}^{\star}_{\breve{X}\breve{Z},nTi}\right]_{(1:k+1,1:k+q_\rho)} = \int_0^1 \breve{\mathbf{g}}_i(r)\breve{\mathbf{h}}_i(r)'dr = \mathbf{M}_{\breve{X}\breve{Z},ni}$ as well as $\left[\lim_{T\to\infty}\mathbf{M}^{\star}_{\breve{Z}\breve{Z},nTi}\right]_{(1:k+1,1:k+q_\rho)} = \int_0^1 \breve{\mathbf{h}}_i(r)\breve{\mathbf{h}}_i(r)'dr = \mathbf{M}_{\breve{Z}\breve{Z},ni}$.

The matrices $\mathbf{M}_{\tilde{X}\tilde{X},n}$, $\mathbf{M}_{\tilde{Z}\tilde{Z},n}$ and $\mathbf{M}_{\tilde{X}\tilde{Z},n}$ are matrices of full rank by Assumption 4. In addition, Section A-1 in the Online Appendix provides sufficient conditions where $\mathbf{M}_{\tilde{X}\tilde{X},n}$, $\mathbf{M}_{\breve{X}\breve{X},n}$, $\mathbf{M}_{\tilde{Z}\tilde{Z},n}$ and $\mathbf{M}_{\breve{Z}\breve{Z},n}$ are full rank matrices.

`Step 3` (*Wald statistic $\mathscr{W}_{\boldsymbol{\gamma},n}$*): We follow Phillips and Hansen (1990), Johansen (1995), and Park and Phillips (1988) to derive the so called *observed Wald-statistic* $\mathscr{W}_{\boldsymbol{\gamma},nT}$ and its limit $\mathscr{W}_{\boldsymbol{\gamma},n}$. Consider the $s \times k+1$ restriction matrix $\mathbf{R}$.

For the *D2SLS* estimator (28), the $T \to \infty$-limit of $\mathbf{m}^{\star}_{\tilde{Z}\tilde{u},nT} = \sum_{i=1}^n \mathbf{m}^{\star}_{\tilde{Z}\tilde{u},nTi}$ was derived in (50). Conditional on $\tilde{\mathcal{B}}_v(r)$, $r \in [0,1]$, we observe that $\mathbb{V}\left(\mathbf{m}_{\tilde{Z}\tilde{u},n}\mathbf{m}'_{\tilde{Z}\tilde{u},n}|\tilde{\mathcal{B}}_v(r),\ r\in[0,1]\right) = \mathbb{E}\left(\mathbf{m}_{\tilde{Z}\tilde{u},n}\mathbf{m}'_{\tilde{Z}\tilde{u},n}|\tilde{\mathcal{B}}_v(r),\ r\in[0,1]\right) = \sum_{i=1}^n \Omega_{u_iu_i}\mathbf{M}_{\tilde{Z}\tilde{Z},ni}$. This follows from the cross-sectional independence imposed in Assumption 2 and the result that $\int_0^1 \tilde{\mathcal{B}}_v(r)d\mathcal{B}_{u_i}(r)$ is a mean zero Gaussian mixture distribution where $\tilde{\mathcal{B}}_v(r)$ and $\mathcal{B}_{u_i}(r)$ are independent (see (A-14) in the Online Appendix).

Endowed with a consistent estimator, $\hat{\boldsymbol{\Omega}}_{uu}$, of the long run covariance matrix $\boldsymbol{\Omega}_{uu} = diag(\Omega_{u_iu_i})_{i=1,\dots,n}$ we obtain by the above convergence results and the continuous mapping theorem $\mathbf{V}_{nT} = \left[\mathbf{M}_{\tilde{X}\tilde{Z},nT}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},nT}\mathbf{M}'_{\tilde{X}\tilde{Z},nT}\right]^{-1}\mathbf{D}_{nT}\left[\mathbf{M}_{\tilde{X}\tilde{Z},nT}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},nT}\mathbf{M}'_{\tilde{X}\tilde{Z},nT}\right]^{-1} \Rightarrow \left[\mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n} - \mathbf{M}'_{\tilde{X}\tilde{Z},n}\right]^{-1}$

$\mathbf{D}_n\left[\mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{M}'_{\tilde{X}\tilde{Z},n}\right]^{-1} := \mathbf{V}_n$, where $\mathbf{D}_{nT} = \mathbf{M}_{\tilde{X}\tilde{Z},nT}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},nT}\left(\sum_{i=1}^n \hat{\Omega}_{u_iu_i}\mathbf{M}_{\tilde{Z}\tilde{Z},nTi}\right)\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},nT}\mathbf{M}'_{\tilde{X}\tilde{Z},nT}$ and $\mathbf{D}_n = \mathbf{M}_{\tilde{X}\tilde{Z},n}\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\left(\sum_{i=1}^n \Omega_{u_iu_i}\mathbf{M}_{\tilde{Z}\tilde{Z},ni}\right)\mathbf{M}^{-1}_{\tilde{Z}\tilde{Z},n}\mathbf{M}'_{\tilde{X}\tilde{Z},n}$. Then the Wald statistic (31) satisfies

$$\mathscr{W}_{\boldsymbol{\gamma},nT} \Rightarrow \mathscr{W}_{\boldsymbol{\gamma},n} = \left(T_\star\mathbf{R}\left(\widehat{\boldsymbol{\gamma}}_{D2SLS;p}-\boldsymbol{\gamma}\right)\right)'\left(\mathbf{R}\mathbf{V}_n\mathbf{R}'\right)^{-1}\left(T_\star\mathbf{R}\left(\widehat{\boldsymbol{\gamma}}_{D2SLS;p}-\boldsymbol{\gamma}\right)\right). \tag{52}$$

Due to the fact that we have derived a zero mean normal mixture distribution, under the null hypothesis the Wald statistic $\mathscr{W}_{\boldsymbol{\gamma},n}$ follows a χ^2 distribution with $\mathfrak{s}$ degrees of freedom. This completes the proof of Theorem 1.

In a further step we consider the Wald type test obtained in Theorem 2. Assume that a consistent estimator of $\boldsymbol{\Omega}_{uu} = diag(\Omega_{u_iu_i})_{i=1,\dots,n}$ is available. From (A-17) obtained in the Online Appendix, we observe that $\mathbb{V}\left(\mathbf{m}_{\breve{Z}\breve{u},n}\mathbf{m}'_{\breve{Z}\breve{u},n}|\tilde{\mathcal{B}}_v(r),\ r\in[0,1]\right) = \mathbb{E}\left(\check{\mathbf{m}}_{\breve{Z}\breve{u},n}\mathbf{m}'_{\breve{Z}\breve{u},n}|\tilde{\mathcal{B}}_v(r),\ r\in[0,1]\right)$ and

$$\begin{aligned}
&\mathbb{E}\left(\mathbf{m}_{\breve{Z}\breve{u},n}\mathbf{m}'_{\breve{Z}\breve{u},n}|\tilde{\mathcal{B}}_v(r),\ r\in[0,1]\right)\\
&= \sum_{i=1}^n\sum_{l=1}^n\left(\Omega_{u_iu_l}-\frac{1}{n}\Omega_{u_iu_i}-\frac{1}{n}\Omega_{u_lu_l}+\frac{1}{n^2}\sum_{j=1}^n\Omega_{u_ju_j}\right)\int_0^1\breve{\mathbf{h}}_i(r)\breve{\mathbf{h}}_l(r)'dr\\
&\quad\text{or using the more compact notation used in (A-18)}\\
&= \sum_{i=1}^n\Omega_{u_iu_i}\left(\int_0^1\breve{\mathbf{h}}_i(r)\breve{\mathbf{h}}_i(r)'dr-\int_0^1\breve{\mathbf{h}}_i(r)\left(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r)\right)'dr-\int_0^1\left(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r)\right)\breve{\mathbf{h}}_i(r)'dr\right.\\
&\quad\left.+\int_0^1\left(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r)\right)\left(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r))\right)'dr\right).
\end{aligned} \tag{53}$$

Note that by Assumption 2, $\Omega_{u_iu_l} = 0$ for any $i \neq l$. By the assumption that $\hat{\Omega}_{u_iu_i}$ consistently estimates $\Omega_{u_iu_i}$, Eq. (53) and the continuous mapping theorem we get $\breve{\mathbf{V}}_{nT} = \left[\mathbf{M}_{\breve{X}\breve{Z},nT}\mathbf{M}^{-1}_{\breve{Z}\breve{Z},nT}\mathbf{M}'_{\breve{X}\breve{Z},nT}\right]^{-1}\breve{\mathbf{D}}_{nT}\left[\mathbf{M}_{\breve{X}\breve{Z},nT}\mathbf{M}^{-1}_{\breve{Z}\breve{Z},nT}\mathbf{M}'_{\breve{X}\breve{Z},nT}\right]^{-1} \Rightarrow \left[\mathbf{M}_{\breve{X}\breve{Z},n}\mathbf{M}^{-1}_{\breve{Z}\breve{Z},n}-\mathbf{M}'_{\breve{X}\breve{Z},n}\right]^{-1}$ $\breve{\mathbf{D}}_n\left[\mathbf{M}_{\breve{X}\breve{Z},n}\mathbf{M}^{-1}_{\breve{Z}\breve{Z},n}\mathbf{M}'_{\breve{X}\breve{Z},n}\right]^{-1} := \breve{\mathbf{V}}_n$, where

$$\begin{aligned}
\breve{\mathbf{D}}_{nT} &= \mathbf{M}_{\breve{X}\breve{Z},nT}\mathbf{M}^{-1}_{\breve{Z}\breve{Z},nT}\mathbf{P}_{\breve{Z}\breve{u}\breve{Z}\breve{u},nT}\mathbf{M}^{-1}_{\breve{Z}\breve{Z},nT}\mathbf{M}'_{\breve{X}\breve{Z},nT}, \quad\text{with}\\
\mathbf{P}_{\breve{Z}\breve{u}\breve{Z}\breve{u},nT} &:= \sum_{i=1}^n\sum_{l=1}^n\left(\hat{\Omega}_{u_iu_l}-\frac{1}{n}\hat{\Omega}_{u_iu_i}-\frac{1}{n}\hat{\Omega}_{u_lu_l}+\frac{1}{n^2}\sum_{j=1}^n\hat{\Omega}_{u_ju_j}\right)\\
&\quad\times\left[\sum_{t_\star=1}^{T_\star}\mathbf{A}_{\breve{Z}p}\breve{\mathbf{Z}}_{p;it_\star}\breve{\mathbf{Z}}'_{p;lt_\star}\mathbf{A}_{\breve{Z}p}\right]_{(1:q_\rho+k_I:1:q_\rho+k_I)}\quad\text{and}\\
\breve{\mathbf{D}}_n &= \mathbf{M}_{\breve{X}\breve{Z},n}\mathbf{M}^{-1}_{\breve{Z}\breve{Z},n}\mathbf{P}_{\breve{Z}\breve{u}\breve{Z}\breve{u},n}\mathbf{M}^{-1}_{\breve{Z}\breve{Z},n}\mathbf{M}'_{\breve{X}\breve{Z},n}, \quad\text{with}\\
\mathbf{P}_{\breve{Z}\breve{u}\breve{Z}\breve{u},n} &= \sum_{i=1}^n\mathbf{P}_{\breve{Z}\breve{u}\breve{Z}\breve{u},ni}, \quad\text{where by (53) and (A-18)}\\
\mathbf{P}_{\breve{Z}\breve{u}\breve{Z}\breve{u},ni} &= \Omega_{u_iu_i}\left(\int_0^1\breve{\mathbf{h}}_i(r)\breve{\mathbf{h}}_i(r)'dr-\int_0^1\breve{\mathbf{h}}_i(r)\left(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r)\right)'dr-\int_0^1\left(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r)\right)\breve{\mathbf{h}}_i(r)'dr\right.\\
&\quad\left.+\int_0^1\left(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r)\right)\left(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r))\right)'dr\right).
\end{aligned}$$

Then the Wald statistic (31) satisfies

$$\breve{\mathscr{W}}_{\gamma,nT} = \left(T_\star\mathbf{R}\left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS;p}-\breve{\boldsymbol{\gamma}}\right)\right)'\left(\mathbf{R}\breve{\mathbf{V}}_{nT}\mathbf{R}'\right)^{-1}\left(T_\star\mathbf{R}\left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS;p}-\breve{\boldsymbol{\gamma}}\right)\right), \tag{54}$$

and converges to a χ^2-distributed random variable with $\mathfrak{s}$ degrees of freedom.

B. Joint limits: Proof of Theorem 3

By Assumption 5, the process $\left(\boldsymbol{\eta}_{it}^{\dagger}\right)_{t\in\mathbb{Z}}$, $\boldsymbol{\eta}_{it}\in\mathbb{R}^{k_I}$, allows to apply the joint asymptotic limit theory developed in Phillips and Moon (1999, 2000). Phillips and Moon (1999, Lemma 2) show that for $\left(\boldsymbol{\eta}_{it}^{\dagger}\right)_{t\in\mathbb{Z}}$, a panel Beveridge-Nelson decomposition exists (see also Johansen, 1995, p. 36), given that their Assumptions 1 and 2 are met. Hence, we consider a linear process $\left(\boldsymbol{\eta}_{it}^{\dagger}\right)_{t\in\mathbb{Z}}$, with Wold representation $\boldsymbol{\eta}_{it}^{\dagger}=\sum_{s=0}^{\infty}\mathrm{C}_{s;\eta_i}^{\dagger}\boldsymbol{\varepsilon}_{it-s}^{\dagger}$, where $\mathbf{C}_{s;\eta_i}^{\dagger}\in\mathbb{R}^{(k_I+1)\times(k_I+1)}$ for $i=1,\ldots,n$ and $\boldsymbol{\eta}_{t}^{\dagger}=\sum_{s=1}^{\infty}\mathrm{C}_{s;\eta}^{\dagger}\boldsymbol{\varepsilon}_{t-s}^{\dagger}$, where $\mathrm{C}_{s;\eta}^{\dagger}=diag\left(\mathrm{C}_{s;\eta_1}^{\dagger},\ldots,\mathrm{C}_{s;\eta_n}^{\dagger}\right)$ and $\boldsymbol{\varepsilon}_{t-s}^{\dagger}:=\left(\boldsymbol{\varepsilon}_{1t-s}^{\dagger\prime},\ldots,\boldsymbol{\varepsilon}_{nt-s}^{\dagger\prime}\right)'$. Since u_{it} can be written as $u_{it}=u_{it}^{\dagger}-\boldsymbol{\Omega}_{u_i v}\boldsymbol{\Omega}_{vv}^{-1}\mathbf{v}_t=u_{it}^{\dagger}-\boldsymbol{\Omega}_{u_i v_i}\boldsymbol{\Omega}_{v_i v_i}^{-1}\mathbf{v}_{it}$ (the last equality follows from independence across i), also for $\left(\boldsymbol{\eta}_{it}\right)_{t\in\mathbb{Z}}$, a Wold representation $\boldsymbol{\eta}_{it_\star}=\sum_{s=0}^{\infty}\mathrm{C}_{s;\eta_i}\boldsymbol{\varepsilon}_{it_\star-s}$ as well as a panel Beveridge-Nelson decomposition exists. After applying the within-transform defined in (32), we obtain

$$\breve{\boldsymbol{\eta}}_{it_\star}=\mathrm{C}_{\eta_i}(1)\breve{\boldsymbol{\varepsilon}}_{it_\star}-\mathring{\boldsymbol{\eta}}_{it_\star}+\mathring{\boldsymbol{\eta}}_{it_\star-1},\ \ \mathring{\boldsymbol{\eta}}_{it_\star}:=\sum_{s=0}^{\infty}\mathring{\mathrm{C}}_{s;\eta_i}\breve{\boldsymbol{\varepsilon}}_{it_\star-s_\star},\ \ \mathring{\mathrm{C}}_{s;\eta_i}:=\sum_{\kappa=s+1}^{\infty}\mathrm{C}_{\kappa;\eta_i},\ \ \mathrm{C}_{\eta_i}(1):=\sum_{s=0}^{\infty}\mathrm{C}_{s,\eta_i},$$

$$\breve{\mathbf{v}}_{it_\star}=\mathrm{C}_{v_i}(1)\breve{\boldsymbol{\varepsilon}}_{it_\star}-\mathring{\mathbf{v}}_{it_\star}+\mathring{\mathbf{v}}_{it_\star-1}\ ,\ \ \mathring{\mathbf{v}}_{it_\star}:=\sum_{s=0}^{\infty}\mathring{\mathrm{C}}_{s;v_i}\breve{\boldsymbol{\varepsilon}}_{it_\star-s_\star}\ ,\ \ \mathring{\mathrm{C}}_{s;v_i}:=\sum_{\kappa=s+1}^{\infty}\mathrm{C}_{\kappa;v_i}\ \text{and}$$

$$\breve{u}_{it_\star}=\mathrm{C}_{u_i}(1)\breve{\boldsymbol{\varepsilon}}_{u_i t_\star}-\mathring{u}_{it_\star}+\mathring{u}_{i_\star-1}\ ,\ \mathring{u}_{it_\star}:=\sum_{s=0}^{\infty}\mathring{\mathrm{C}}_{s;u_i}\breve{\boldsymbol{\varepsilon}}_{t_\star-s_\star}\ ,\ \ \mathring{\mathrm{C}}_{s;u_i}:=\sum_{\kappa=s+1}^{\infty}\mathrm{C}_{\kappa;u_i}\ ,$$

such that for partial sums we get

$$\sum_{t_\star=1}^{[rT_\star]}\breve{\boldsymbol{\eta}}_{it_\star}=\mathrm{C}_{\eta_i}(1)\sum_{t_\star=1}^{[rT_\star]}\breve{\boldsymbol{\varepsilon}}_{it_\star}+\mathring{\boldsymbol{\eta}}_{i0_\star}-\mathring{\boldsymbol{\eta}}_{i[rT_\star]}\ \text{and}\ \breve{\mathbf{x}}_{i[rT_\star]}=\breve{\mathbf{x}}_{i0_\star}+\left[\sum_{t_\star=1}^{[rT_\star]}\breve{\boldsymbol{\eta}}_{it_\star}\right]_{(2:k_I+1)}=\breve{\mathbf{x}}_{i0_\star}+\sum_{t_\star=1}^{[rT_\star]}\breve{\mathbf{v}}_{it_\star}\ . \tag{55}$$

The term $\mathring{\boldsymbol{\eta}}_{i0_\star}$ denotes $\mathring{\boldsymbol{\eta}}_{it_\star}$ at $t_\star=0$. $\mathrm{C}_{s;\eta_i}$ is a matrix polynomial satisfying the conditions of Phillips and Moon (1999, Assumption 1), while $\mathrm{C}_{s;\eta_i}(1):=\sum_{s=0}^{\infty}\mathrm{C}_{s;\eta_i}$ as well as $\mathrm{C}_{s;v_i}$ and $\mathrm{C}_{s;u_i}$ are submatrix polynomials of $\mathrm{C}_{s;\eta_i}$. Note that u_{it} and $\mathbf{v}_t$ as well as $\mathbf{v}_{it}$ are uncorrelated by construction. Hence, $\mathrm{C}_{\eta_i}(1)$ is block diagonal. $\mathrm{C}_{u_i}(1)$ stands for the corresponding row of $\mathrm{C}_{\eta_i}(1)$ to obtain u_{it}. The same notation is applied to obtain $\mathbf{v}_{it}$. Using the Beveridge–Nelson decomposition (55) and $t_\star=t-p$, we obtain

$$\mathbf{Q}_{\check{Z}\check{Z}nT}:=\frac{1}{n}\sum_{i=1}^{n}\mathbf{M}_{\check{Z}\check{Z},nTi}=\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{M}}_{\check{Z}\check{Z},nTi}+\frac{1}{n}\sum_{i=1}^{n}\mathbf{R}_{\check{Z}\check{Z},nTi}\ ,\ \ \text{where}\ \ \bar{\mathbf{M}}_{\check{Z}\check{Z},nTi}:=\frac{1}{T_\star^2}\sum_{t_\star=1}^{T_\star}\bar{\mathscr{M}}_{\check{Z},it_\star}\bar{\mathscr{M}}_{\check{Z},it_\star}'\ ,$$

$$\bar{\mathscr{M}}_{\check{Z},it_\star}:=\begin{pmatrix}\sum_{j=1}^{n}W_{\{n\},ij}^{\tau\mathbb{K}_{(1)}}\left[\mathbf{C}_{(\mathbb{K}_{(1)})}\right]_{(j,(j-1)n+1:jn)}\left(\mathrm{C}_{v_j}(1)\sum_{s_\star=1}^{t_\star}\breve{\boldsymbol{\varepsilon}}_{js_\star}\right)\\ \vdots\\ \sum_{j=1}^{n}W_{\{n\},ij}^{\tau\mathbb{K}_{(q_\rho)}}\left[\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\right]_{(j,(j-1)n+1:jn)}\left(\mathrm{C}_{v_j}(1)\sum_{s_\star=1}^{t_\star}\breve{\boldsymbol{\varepsilon}}_{j_\star}\right)\\ \mathrm{C}_{v_i}(1)\sum_{s_\star=1}^{t_\star}\breve{\boldsymbol{\varepsilon}}_{is_\star}\end{pmatrix}\in\mathbb{R}^{q_\rho+k_I}\ ,\ \text{and}$$

$$\mathbf{R}_{\check{Z}\check{Z},nTi}:=\mathbf{M}_{\check{Z}\check{Z},nTi}-\bar{\mathbf{M}}_{\check{Z}\check{Z},nTi}\in\mathbb{R}^{q_\rho+k\times q_\rho+k}\ . \tag{56}$$

With $K_{\{n\},j\ell}:=\left[\mathbf{K}_{\{n\}}\right]_{(j,\ell)}$ we get

$$\mathbf{Q}_{\check{X}\check{Z},nT}:=\frac{1}{n}\sum_{i=1}^{i}\mathbf{M}_{\check{X}\check{Z}nTi}=\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{M}}_{\check{X}\check{Z}nTi}+\frac{1}{i}\sum_{i=1}^{n}\mathbf{R}_{\check{X}\check{Z}nT}\ ,\ \ \text{where}\ \ \bar{\mathbf{M}}_{\check{X}\check{Z}nTi}:=\frac{1}{T_\star^2}\sum_{t_\star=1}^{T_\star}\bar{\mathscr{M}}_{\check{X},it_\star}\bar{\mathscr{M}}_{\check{Z},it_\star}'\ ,$$

$$\bar{\mathscr{M}}_{\breve{X},it_\star} := \begin{pmatrix} \sum_{j=1}^{n}\sum_{\ell=1}^{n} W_{\{n\},ij}K_{\{n\},j\ell}\breve{\boldsymbol{\beta}}'\left(\mathsf{C}_{v_\ell}(1)\sum_{s_\star=1}^{t_\star}\boldsymbol{\varepsilon}_{\ell s_\star}\right) \\ \mathsf{C}_{v_i}(1)\sum_{s_\star=1}^{t_\star}\boldsymbol{\varepsilon}_{is_\star} \end{pmatrix} \in \mathbb{R}^{1+k_I},$$

$$\mathbf{R}_{\breve{X}\breve{Z},nTi} := \mathbf{M}_{\breve{X}\breve{Z},nTi} - \bar{\mathbf{M}}_{\breve{X}\breve{Z},nTi} \in \mathbb{R}^{1+k\times q_\rho+k}, \tag{57}$$

and

$$\mathbf{q}_{\breve{Z}\breve{u},nT} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{m}_{\breve{Z}\breve{u},nTi} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\mathbf{m}}_{\breve{Z}\breve{u},nTi} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{R}_{\breve{Z}\breve{u},nTi}, \quad \text{where}$$

$$\bar{\mathbf{m}}_{Zu,nTi} := \frac{1}{T_\star}\sum_{t_\star=1}^{T_\star}\bar{\mathscr{M}}_{\breve{Z},it_\star}\bar{\mathscr{M}}_{\breve{u},it_\star}, \quad \bar{\mathscr{M}}_{\breve{u},it_\star} := \mathsf{C}_{u_i}(1)\breve{\boldsymbol{\varepsilon}}_{it_\star} \in \mathbb{R}, \quad \mathbf{R}_{\breve{Z}\breve{u},nTi} := \mathbf{m}_{\breve{Z}\breve{u},nTi} - \bar{\mathbf{m}}_{\breve{Z}\breve{u},nTi} \in \mathbb{R}^{q_\rho+k}. \tag{58}$$

`Step 1,` $(T, n) \to \infty$*-Limits of* $\mathbf{Q}_{\breve{X}\breve{Z},nT}$, $\mathbf{Q}_{\breve{Z}\breve{Z},nT}$ *and* $\mathbf{q}_{\breve{Z}\breve{u},nT}$: In the following steps we adapt Phillips and Moon (1999, Lemmata 13 and 16) to the requirements of our model. First we show:

Lemma 1. *Suppose that the expectations* $\mathbf{Q}_{\breve{Z}\breve{Z}}$ *and* $\mathbf{Q}_{\breve{X}\breve{Z}}$ *exist, Assumptions 1–6 hold, and* $\mathbf{W}_{\{n\}}$ *as well as the error structure are such that* $\mathbb{E}\left(\left[\mathbf{M}_{\breve{X}\breve{Z},nTi}\right]^2_{(l_r l_c)}\right) < \infty$, $l_r = 1, \ldots, k+1$ *for* $l_c = 1$ *and* $l_c = 1, \ldots, k+q_\rho$ *for* $l_r = 1, \ldots, q_\rho$, $\mathbb{E}\left(\left[\mathbf{M}_{\breve{Z}\breve{Z},nTi}\right]^2_{(l_r,l_c)}\right) < \infty$, $l_r = 1, \ldots, k+1$ *for* $l_c = 1, \ldots, q_\rho$ *and* $l_c = 1, \ldots, k+q_\rho$ *for* $l_r = 1, \ldots, q_\rho$ *for all* $T \in \mathbb{N}$ *and* $n \in \mathbb{N} \cup \{\infty\}$.

Then $\mathbf{Q}_{\breve{X}\breve{Z},nT}$ *and* $\mathbf{Q}_{\breve{Z}\breve{Z},nT}$ *converge in probability to* $\mathbf{Q}_{\breve{X}\breve{Z}}$ *and* $\mathbf{Q}_{\breve{Z}\breve{Z}}$ *as* $(n, T) \to \infty$ *with* $\frac{n^6}{T} \to 0$.

Proof. For $\left[\mathbf{Q}_{\breve{X}\breve{Z},nT}\right]_{(2:k+1,q_\rho:q_\rho+k)}$ and $\left[\mathbf{Q}_{\breve{Z}\breve{Z},nT}\right]_{(q_\rho:q_\rho+k,q_\rho:q_\rho+k)}$ joint convergence in probability results already from Phillips and Moon (1999, Lemma 13(a)). That is, $\left[\mathbf{Q}_{\breve{X}\breve{Z}}\right]_{(2:k+1,q_\rho:q_\rho+k)} = \left[\mathbf{Q}_{\breve{Z}\breve{Z}}\right]_{(q_\rho:q_\rho+k,q_\rho:q_\rho+k)} = \frac{1}{6}\bar{\boldsymbol{\Omega}}_{v_iv_i}$. For the remaining components by (56) and a panel Beveridge-Nelson (55) decomposition we obtain:

$$\mathbf{Q}_{\breve{Z}\breve{Z},nT} = \frac{1}{n}\sum_{i=1}\mathbf{M}_{\breve{Z}\breve{Z},nTi} = \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{M}}_{\breve{Z}\breve{Z},nTi} + \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{R}}_{\breve{Z}\breve{Z},nTi}, \quad \text{where } \mathbf{M}_{\breve{Z}\breve{Z},nTi} = \frac{1}{T_\star^2}\sum_{t_\star=1}^{T_\star}\mathscr{M}_{\breve{Z},it_\star}\mathscr{M}'_{\breve{Z},it_\star},$$

$$\mathscr{M}_{\breve{Z},it_\star} := \bar{\mathscr{M}}_{\breve{Z},it_\star} + \begin{pmatrix} \sum_{j=1}^{n} W^{\tau_{\mathbb{K}(1)}}_{\{n\},ij}\left[\mathbf{C}_{(\mathbb{K}_{(1)})}\right]_{(j,(j-1)n+1:jn)}\left(\mathring{\mathbf{x}}_{j0_\star} - \mathring{\mathbf{v}}_{jt_\star} + \mathring{\mathbf{v}}_{j0_\star}\right) \\ \vdots \\ \sum_{j=1}^{n} W^{\tau_{\mathbb{K}(q_\rho)}}_{\{n\},ij}\left[\mathbf{C}_{(\mathbb{K}_{(q_\rho)})}\right]_{(j,(j-1)n+1:jn)}\left(\mathring{\mathbf{x}}_{j0_\star} - \mathring{\mathbf{v}}_{jt_\star} + \mathring{\mathbf{v}}_{j0_\star}\right) \\ \left(\mathring{\mathbf{x}}_{i0_\star} - \mathring{\mathbf{v}}_{it_\star} + \mathring{\mathbf{v}}_{i0_\star}\right) \end{pmatrix} \quad \text{and}$$

$$\bar{\mathbf{R}}_{\breve{Z}\breve{Z},nTi} := \mathbf{R}_{\breve{Z}\breve{Z},nTi} = \mathbf{M}_{\breve{Z}\breve{Z},nTi} - \bar{\mathbf{M}}_{\breve{Z}\breve{Z},nTi}. \tag{59}$$

For $\mathbf{Q}_{\breve{X}\breve{Z},nT}$, we use (57) to obtain:

$$\mathbf{Q}_{\breve{X}\breve{Z},nT} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}_{\breve{X}\breve{Z},nTi} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{M}}_{\breve{X}\breve{Z},nTi} + \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{R}}_{\breve{X}\breve{Z},nTi}, \quad \text{where}$$

$$\mathbf{M}_{\breve{X}\breve{Z},nTi} = \frac{1}{T_\star^2}\sum_{t_\star=1}^{T_\star} \mathscr{M}_{\breve{X},it_\star}\mathscr{M}'_{\breve{Z},it_\star}, \quad \check{\mathscr{M}}_{\breve{X},it_\star} := \bar{\mathscr{M}}_{\breve{X},it_\star} + \begin{pmatrix} \sum_{j=1}^{n}\sum_{\ell=1}^{n} \mathbf{W}_{\{n\},ij}K_{\{n\},j\ell}\breve{u}^{\dagger}_{\ell t_\star} \\ \mathbf{0}_{(k)} \end{pmatrix},$$

$$\breve{\mathbf{u}}^{\dagger}_{it_\star} = \mathsf{C}_{u_i}(1)^{\dagger}\breve{\boldsymbol{\varepsilon}}_{it_\star} + \mathring{\mathbf{u}}^{\dagger}_{it_\star} - \mathring{\mathbf{u}}^{\dagger}_{it_\star-1}, \quad \mathring{\mathbf{u}}^{\dagger}_{it_\star} = \sum_{s=0}^{\infty}\mathsf{C}^{\dagger}_{s;u_i}\breve{\boldsymbol{\varepsilon}}^{\dagger}_{it-s},$$

$$\mathscr{M}_{\breve{X},it_\star} := \check{\mathscr{M}}_{\breve{X},it_\star} + \begin{pmatrix} \sum_{j=1}^{n}\sum_{\ell=1}^{n} \mathbf{W}_{\{n\},ij}\,[\mathbf{K}]_{(j,\ell)}\,\breve{\boldsymbol{\beta}}'\left(\mathring{\mathbf{x}}_{j0_\star} - \mathring{\mathbf{v}}_{jt_\star} + \mathring{\mathbf{v}}_{j0_\star}\right) \\ \left(\mathring{\mathbf{x}}_{i0_\star} - \mathring{\mathbf{v}}_{it_\star} + \mathring{\mathbf{v}}_{i0_\star}\right) \end{pmatrix},$$

$$\bar{\mathbf{R}}_{\breve{X}\breve{Z},nTi} = \mathbf{M}_{\breve{X}\breve{Z},nTi} - \check{\mathbf{M}}_{\breve{X}\breve{Z},nTi} \text{ and}$$
$$\mathbf{R}_{\breve{X}\breve{Z},nTi} = \mathbf{M}_{\breve{X}\breve{Z},nTi} - \bar{\mathbf{M}}_{\breve{X}\breve{Z},nTi} = \bar{\mathbf{R}}_{\breve{X}\breve{Z},nTi} + \tilde{\mathbf{M}}_{\breve{X}\breve{Z},nTi} - \bar{\mathbf{M}}_{\breve{X}\breve{Z},nTi}. \tag{60}$$

By following Phillips and Moon (1999, p. 1100), expression (59) decomposes $\mathbf{Q}_{\breve{Z}\breve{Z},nT}$ into the term $\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{M}}_{\breve{Z}\breve{Z},nTi}$, consisting of weighted sums containing $\mathsf{C}_{v_i}(1)\sum_{s=1}^{t_\star+p}\breve{\boldsymbol{\varepsilon}}_{is}$ and $\breve{u}^{\dagger}_{it}$, and the residual term $\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{R}}_{\breve{Z}\breve{Z},nTi}$. A similar decomposition is obtained for $\mathbf{Q}_{\breve{X}\breve{Z},nT}$ in (60).

The assumptions stated in Lemma 1 on the moments of $\mathbf{M}_{\breve{Z}\breve{Z},nTi}$ and $\mathbf{M}_{\breve{X}\breve{Z},nTi}$ are sufficient for $\check{\mathbf{M}}_{\breve{X}\breve{Z},nTi}$ and $\bar{\mathbf{M}}_{\breve{Z}\breve{Z},nTi}$ to be uniformly integrable in $T_\star$ (see, e.g., Klenke, 2008, Theorem 6.25); (the assumption of second moments can be made weaker by demanding for moments $1+\epsilon$, $\epsilon > 0$ to exist and then applying Billingsley (1986, p. 348)). For the south-eastern parts of $\bar{\mathbf{M}}_{\breve{X}\breve{Z},nTi}$ and $\bar{\mathbf{M}}_{\breve{Z}\breve{Z},nTi}$ uniform integrability already follows from Phillips and Moon (1999, Assumptions 1 and 2)). Then by Billingsley (1986, Theorem 25.12), we observe that $\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{M}}_{\breve{Z}\breve{Z},nTi} \xrightarrow{P} \mathbb{E}\left(\bar{\mathbf{M}}_{\breve{Z}\breve{Z},nTi}\right) = \mathbf{Q}_{\breve{Z}\breve{Z}}$ and $\frac{1}{n}\sum_{i=1}^{i}\tilde{\mathbf{M}}_{\breve{X}\breve{Z},nTi} \xrightarrow{P} \mathbb{E}\left(\bar{\mathbf{M}}_{\breve{X}\breve{Z},nTi}\right) = \mathbf{Q}_{\breve{X}\breve{Z}}$ as $(T,n)\to\infty$. Note that $\mathbb{E}\left(\bar{\mathbf{M}}_{\breve{X}\breve{Z},nTi}\right) = \mathbf{Q}_{\breve{X}\breve{Z}} = \mathbb{E}\left(\mathbf{M}_{\breve{X}\breve{Z},nTi}\right)$ and $\mathbb{E}\left(\bar{\mathbf{M}}_{\breve{Z}\breve{Z},nTi}\right) = \mathbf{Q}_{\breve{Z}\breve{Z}} = \mathbb{E}\left(\mathbf{M}_{\breve{Z}\breve{Z},nTi}\right)$ as defined in (38) as well as (39).

The residual terms $\bar{\mathbf{R}}_{\breve{X}\breve{Z},nTi}$ and $\bar{\mathbf{R}}_{\breve{Z}\breve{Z},nTi}$ decompose into sums to of the remaining terms obtained in Phillips and Moon (1999, p. 1101, "$R_{ki,T}$" in their notation). These remaining terms in Phillips and Moon (1999) have expectations of the order $\frac{1}{\sqrt{T}}O(1)$ and $\frac{1}{T}O(1)$. In addition the condition $n/T \to 0$ has to be met such that weighted sums of these residual terms become small. Since n, n^2, and n^3 summands of this structure show up in the nonsouth-east terms of $\bar{\mathbf{R}}_{\breve{Z}\breve{Z},nTi}$ and $\bar{\mathbf{R}}_{\breve{X}\breve{Z},nTi}$, and the weights are bounded by our Assumption 6 (i.e., $|W_{\{n\},ij}| \leq \bar{w}$ and $|\mathbf{K}_{\{n\},ij}| \leq \bar{w}$), we obtain the requirement that $\frac{n^3}{\sqrt{T}} \to 0 \Leftrightarrow \frac{n^6}{T} \to 0$ to make the impact of these remaining terms sufficiently small. Then, by Phillips and Moon (1999, Corollary 1) $\lim_{(T,n)\to\infty}\mathbf{Q}_{\breve{X}\breve{Z},nT} = \mathbf{Q}_{\breve{X}\breve{Z}} = \mathbb{E}\left(\mathbf{M}_{\breve{X}\breve{Z},nTi}\right)$ and $\lim_{(T,n)\to\infty}\mathbf{Q}_{\breve{Z}\breve{Z},nT} = \mathbf{Q}_{\breve{Z}\breve{Z}} = \mathbb{E}\left(\mathbf{M}_{\breve{Z}\breve{Z},nTi}\right)$, given that $\frac{n^6}{T} \to 0$.

Next we consider the limit of $\frac{1}{\sqrt{nT_\star}}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\breve{\mathbf{Z}}_{it_\star(1:k+q_\rho)}\underline{\breve{u}}_{it_\star}$:

Lemma 2. *Suppose that the Assumptions of Lemma 1 hold. In addition,* $\mathbf{W}_{\{n\}}$ *and* $\left(\boldsymbol{\eta}_{it}\right)_{t\in\mathbb{Z}}$ *are such that* $\mathbb{E}\left(\left[\mathbf{m}_{\breve{Z}\breve{u},nTi}\right]^2_{(l_r,1)}\right) < \infty$, $l_r = 1,\ldots,q_\rho$ *for all* $i, T \in \mathbb{N}$ *as well as for all* $n \in \mathbb{N}$ *and* $n \to \infty$. *Let* $(n,T)\to\infty$ *and* $n^6/T \to 0$.

Then $\mathbf{q}_{\breve{Z}\breve{u},nT} = \frac{1}{\sqrt{nT_\star}}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\breve{\mathbf{Z}}_{it_\star(1:k+q_\rho)}\underline{\breve{u}}_{it} \Rightarrow \mathbf{q}_{\breve{Z}\breve{u}} \sim \mathcal{N}(\mathbf{0}_{(k+q_\rho,1)}, \boldsymbol{\Xi})$, *where* $\boldsymbol{\Xi} = \mathbf{P}_{\breve{Z}\breve{u}\breve{Z}\breve{u}}$ *and* $\mathbf{P}_{\breve{Z}\breve{u}\breve{Z}\breve{u}} = \bar{\boldsymbol{\Omega}}_{u_iu_i}\mathbb{E}\big(\int_0^1 \breve{\mathbf{h}}_i(r)\breve{\mathbf{h}}_i(r)'dr - \int_0^1 \breve{\mathbf{h}}_i(r)\big(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r)\big)'dr - \int_0^1 \big(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r)\big)\breve{\mathbf{h}}_i(r)'dr + \int_0^1 \big(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r)\big)\big(\mathbf{N}_{\breve{h}}\breve{\mathbf{h}}(r)\big)'dr\big)$.

Proof. Equation (58) results in:

$$\mathbf{q}_{\check{Z}\check{u},nT} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{m}_{\check{Z}\check{u},nTi} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\mathbf{m}}_{\check{Z}\check{u},nTi} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\mathbf{R}}_{\check{Z}\check{u},nTi}$$

$$\mathbf{m}_{\check{Z}\check{u},nTi} = \frac{1}{T_\star}\sum_{t_\star=1}^{T_\star}\mathscr{M}_{\check{Z},it_\star}\left[\mathscr{M}_{\check{u},it_\star} + \breve{e}_{p;it_\star}\right],$$

$$\mathscr{M}_{\check{u},it_\star} := \bar{\mathscr{M}}_{\check{u},it_\star} + \left(-\mathring{u}_{it_\star} + \mathring{u}_{it_\star-1}\right) \quad \text{and} \quad ,$$

$$\bar{\mathbf{R}}_{\check{Z}\check{u},nTi} := \mathbf{m}_{\check{Z}\check{u},nTi} - \bar{\mathbf{m}}_{\check{Z}\check{u},nTi}\,, \quad \mathbf{R}_{\check{Z}\check{u},nTi} = \bar{\mathbf{R}}_{\check{Z}\check{u},nTi} + \frac{1}{T_\star}\sum_{t_\star=1}^{T_\star}\mathscr{M}_{\check{Z},it_\star}\breve{e}_{p;it_\star}\,. \tag{61}$$

Since $\mathscr{M}_{\check{Z},it_\star}$ does not depend on $\check{u}_{it_\star}$ the difference between $\mathbf{R}_{\check{Z}\check{u},nTi}$ and $\bar{\mathbf{R}}_{\check{Z}\check{u},nTi}$ can be obtained in a straightforward way. Equation (61) has two main differences compared to the corresponding term considered at the beginning of the proof of Lemma 16 in Phillips and Moon (1999):

First, $\breve{e}_{\sharp p;it_\star}(=\breve{e}_{p;it_\star}$ in shorter notation) described in (16) shows up, and second, a sum over n terms is included by $\mathbf{W}_{\{n\}i}^{\tau\mathbb{K}_{(1)}}\mathbf{C}_{(\mathbb{K}_{(1)})}\cdots$ due to the instruments used in our article. Note that the projection error terms $e_{\sharp p;it_\star}$ are of the order $o_p(1)$ for $T_\star \to \infty$ and each i. The term $\mathscr{M}_{\check{Z},it_\star}$ is bounded in probability by Assumption 7. By this $\mathbf{R}_{\check{Z}\check{u},nTi} - \bar{\mathbf{R}}_{\check{Z}\check{u},nTi}$ converges to a vector of zeros in probability for $T \to \infty$ for each $n \in \mathbb{N} \cup \{\infty\}$.

Second, we consider the effect caused by the spatial lag: Observe that the first components of $\mathbf{q}_{\check{Z}\check{u},nT}$ can be considered to be n-fold sums over terms considered in Phillips and Moon (1999). By Klenke (2008, Theorem 6.25) uniform integrability of $\left[\bar{\mathbf{m}}_{\check{Z}\check{u},nTi}\right]_{(1:q_\rho,1)}$ in T (as well as in $T_\star$) follows from the fact that the first two moments of $\mathbf{m}_{\check{Z}\check{u},nTi}$ are finite for all $T_\star$ as stated in the assumptions of Lemma 2. For the remaining coordinates of $\mathbf{m}_{\check{Z}\check{u},nTi}$ uniform integrability already follows from the Assumptions in Phillips and Moon (1999, see proof of Lemma 13). This allows us to apply a joint central limit theorem, in particular Phillips and Moon (1999, Theorem 3), where $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\mathbf{m}}_{\check{Z}\check{u},nTi}$ converges to a normal random variable $\mathbf{q}_{\check{Z}\check{u}}$ with a $k_I + q_\rho$-dimensional mean zero and covariance $\boldsymbol{\Xi} = \mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u}}$, where (see also Online Appendix A-2)

$$\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u}} = \bar{\boldsymbol{\Omega}}_{u_iu_i}\mathbb{E}\left(\int_0^1 \check{\mathbf{h}}_i(r)\check{\mathbf{h}}_i(r)'dr - \int_0^1 \check{\mathbf{h}}_i(r)\left(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r)\right)'dr - \int_0^1\left(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r)\right)\check{\mathbf{h}}_i(r)'dr + \int_0^1\left(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r)\right)\left(\mathbf{N}_{\check{h}}\check{\mathbf{h}}(r)\right)'dr\right). \tag{62}$$

In addition, the components $q_\rho+1$ to $q_\rho+k$ of $\bar{\mathbf{R}}_{\check{Z}\check{u},nTi}$, jointly converge to zero if $n/T \to 0$. This follows already from Phillips and Moon (1999, Lemma 13 or 16; for q_ρ to $q_\rho + k$, $\bar{\mathbf{R}}_{\check{Z}\check{u},nTi}$ can be decomposed into "$R_{i,Tk}$" in their notation), where the authors show that the residual terms "$R_{k,iT}$" can be decomposed into components with an expected norm of order $\mathcal{O}(1/T)$, $\mathcal{O}(\sqrt{n/T}) = \sqrt{n/T}\mathcal{O}(1)$ and $\mathcal{O}(\sqrt{n}/T)$ (by this fact Phillips and Moon (1999) added the requirement $n/T \to 0$, which is sufficient to obtain $(n,T) \to \infty$ convergence for these coordinates). In our analysis the first q_ρ components of $\bar{\mathbf{R}}_{\check{Z}\check{u},nTi}$ contain n-fold sums of the terms investigated in Phillips and Moon (1999, p. 1108). Since n terms of this structure show up due to a spatial lag, and the elements of $\mathbf{W}_{\{n\}}$ and $\mathbf{K}_{\{n\}}$ are bounded by $\bar{w}$, the residual term $\bar{\mathbf{R}}_{\check{Z}\check{u},nTi}$ contains terms of the form $n \cdot O(1/T)$, $n \cdot \sqrt{n/T}O(1)$ and $n \cdot O(\sqrt{n}/T)$. These residual terms demand for $n^3/T \to 0$. Hence, $n^6/T \to 0$ is sufficient to meet this requirement and all the terms contained in $\bar{\mathbf{R}}_{\check{Z}\check{u},nTi}$ converge to zero in probability, such that for $(T,n) \to \infty$, weak convergence of $\mathbf{q}_{\check{Z}\check{u},nT}$ to a normal random vector $\mathbf{q}_{\check{Z}\check{u}}$ with mean zero and covariance $\boldsymbol{\Xi} = \mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u}}$ obtained in (62) follows (see, e.g., the central limit theorem provided in Phillips and Moon, 1999, Theorem 3). By our model assumptions $\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u}}$ is equal to the expectation of $\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u},ni}$.

Given that the assumptions stated in Theorem 3 hold, we observe joint convergence (in probability) of $\mathbf{Q}_{\check{X}\check{Z},nT}$, $\mathbf{Q}_{\check{Z}\check{Z},nT}$ to $\mathbf{Q}_{\check{X}\check{Z}}$, $\mathbf{Q}_{\check{Z}\check{Z}}$ for $(T,n)\to\infty$, where $n^6/T\to 0$. In addition, under the conditions stated in Theorem 3, $\mathbf{q}_{\check{Z}\check{u},nT}$ weakly converges to $\mathbf{q}_{\check{Z}\check{u}}$. $\mathbf{q}_{\check{Z}\check{u}}$ is multivariate normal with mean vector $\mathbf{0}_{(k+q_\rho\times 1)}$ and covariance matrix $\boldsymbol{\Xi}=\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u}}$.

`Step 2,` $(T,n)\to\infty$*-Limits of* $\sqrt{n}T_\star\left(\widehat{\check{\boldsymbol{\gamma}}}_{D2SLS;p}-\check{\boldsymbol{\gamma}}\right)$: Consider the south-east blocks of $\mathbf{Q}^\star_{\check{Z}\check{Z},nT}$ and $\mathbf{Q}^\star_{\check{X}\check{Z},nT}$. In both matrices we consider sums of elements $\check{\mathbf{v}}_{it_\star}\check{\mathbf{v}}'_{it_\star+\ell}$ scaled by $\frac{1}{nT_\star}$. By a joint law of large numbers convergence (in probability) to elements of $\boldsymbol{\Gamma}_{\ell,v_iv_i}$ is obtained. The south-west and the north-eastern blocks of these matrices converge to zero in probability. To see this, in these blocks of $\mathbf{Q}^\star_{\check{Z}\check{Z},nT}$ and $\mathbf{Q}^\star_{\check{X}\check{Z},nT}$ we meet terms similar to $\mathbf{q}_{\check{Z}\check{u},nT}$ scaled by a higher rate. In particular, these terms contain $\frac{1}{n}\sum_{i=1}^n \mathbf{m}_{\check{Z}\check{v},nTi}:=\frac{1}{\sqrt{nT_\star^3}}\sum_{i=1}^n \mathscr{M}_{\check{Z},it_\star}\check{\mathbf{v}}'_{is}$ and $\frac{1}{n}\sum_{i=1}^n \mathbf{m}_{\check{Z}\check{v},nTi}:=\frac{1}{\sqrt{nT_\star^3}}\sum_{i=1}^n \mathscr{M}_{\check{Z}\check{u},nt_\star i}\check{\mathbf{v}}'_{is}$. Given Assumptions 1 and 2 of Phillips and Moon (1999), $\mathbf{v}_{it}$ and $\check{\mathbf{v}}_{it}$ are of the same stochastic order as $u^\dagger_{it}$ and $\check{u}_{it}$. Hence, for $\frac{1}{\sqrt{nT_\star^2}}\sum_{i=1}^n \mathscr{M}_{\check{Z},it_\star}\check{\mathbf{v}}'_{is}$ and $\frac{1}{\sqrt{nT_\star^2}}\sum_{i=1}^n \mathscr{M}_{\check{X},it_\star}\check{\mathbf{v}}'_{is}$ a joint central limit theorem holds. Based on these results, the terms $\frac{1}{n}\sum_{i=1}^n \mathbf{m}_{\check{Z}\check{v},nTi}$ and $\frac{1}{n}\sum_{i=1}^n \mathbf{m}_{\check{X}\check{v},nTi}$ converge to zero in probability.

By Lemma 1 we observe convergence in probability of the matrices $\mathbf{Q}_{\check{X}\check{Z},nT}$ and $\mathbf{Q}_{\check{Z}\check{Z},nT}$ to $\mathbf{Q}_{\check{X}\check{Z}}$ and $\mathbf{Q}_{\check{Z}\check{Z}}$. Lemma 2 shows weak convergence to a normal distribution. The requirements of these Lemmata are met by the assumptions stated in Theorem 3. By a mapping theorem for random variables converging in probability (see, e.g., White, 2001, Theorem 2.27), we observe that the $(n,T)\to\infty$-asymptotic distribution of $\sqrt{n}T_\star(\check{\boldsymbol{\gamma}}_{D2SLS;p}-\check{\boldsymbol{\gamma}})$ is a normal distribution with mean vector $\mathbf{0}_{(k+1,1)}$ and a covariance matrix $\mathbf{V}_{\check{Q}}$. Since we have assumed that $\mathbf{P}_{\check{Z}\check{u}\check{Z}\check{u}}$ can be estimated consistently, $\check{\mathbf{D}}_{\check{Q},nT}$ and $\check{\mathbf{V}}_{\check{Q},nT}$ converge in probability to $\check{\mathbf{D}}_{\check{Q}}$ and $\check{\mathbf{V}}_{\check{Q}}$, respectively.

C. Estimation of β_L

This section obtains the limit distributions for the estimator (43). Note that conditional on $\grave{\mathbf{x}}_{Li}$:

$$\begin{aligned}
&\lim_{T\to\infty}\sqrt{T_\star}\left(\widehat{\boldsymbol{\beta}}_L-\boldsymbol{\beta}_L\right)\\
&=\lim_{T\to\infty}\sqrt{T_\star}\left(\sum_{i=1}^n \grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}'_{Li}\right)^{-1}\frac{1}{T_\star}\sum_{i=1}^n \grave{\mathbf{x}}_{Li}\left[\sum_{t_\star=1}^{T_\star}\grave{u}^\dagger_{it_\star}+\left(\widehat{\check{\boldsymbol{\gamma}}}_{D2SLS}-\check{\boldsymbol{\gamma}}\right)'\sum_{t_\star=1}^{T_\star}\left(\grave{y}^*_{it},\grave{\mathbf{x}}'_{Iit}\right)'\right]\\
&=\lim_{T\to\infty}\sqrt{T_\star}\left(\sum_{i=1}^n \grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}'_{Li}\right)^{-1}\sum_{i=1}^n \grave{\mathbf{x}}_{Li}\left[\frac{1}{T_\star}\sum_{t_\star=1}^{T_\star}\grave{u}^\dagger_{it_\star}+\frac{T_\star}{T_\star^2}\left(\widehat{\check{\boldsymbol{\gamma}}}_{D2SLS}-\check{\boldsymbol{\gamma}}\right)'\sum_{t_\star=1}^{T_\star}\left(\grave{y}^*_{it},\grave{\mathbf{x}}'_{Iit}\right)'\right]\\
&=\lim_{T\to\infty}\left(\sum_{i=1}^n \grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}'_{Li}\right)^{-1}\sum_{i=1}^n \grave{\mathbf{x}}_{Li}\left[T_\star^{-1/2}\sum_{t_\star=1}^{T_\star}\grave{u}^\dagger_{it_\star}+T_\star\left(\widehat{\check{\boldsymbol{\gamma}}}_{D2SLS}-\check{\boldsymbol{\gamma}}\right)'T_\star^{-3/2}\sum_{t_\star=1}^{T_\star}\left(\grave{y}^*_{it},\grave{\mathbf{x}}'_{Iit}\right)'\right]\\
&=\left(\sum_{i=1}^n \grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}'_{Li}\right)^{-1}\sum_{i=1}^n \grave{\mathbf{x}}_{Li}\left[\int_0^1 d\mathcal{B}_{\grave{u}_i(r)}+\left[\lim_{T\to\infty}T_\star\left(\widehat{\check{\boldsymbol{\gamma}}}_{D2SLS}-\check{\boldsymbol{\gamma}}\right)\right]'\int_0^1\grave{\mathbf{g}}_i(r)dr\right]\\
&\quad\text{where }\grave{\mathbf{g}}_i(r):=\begin{pmatrix}\mathbf{W}_i\mathbf{KC}\grave{\mathcal{B}}_v(r)\\ \grave{\mathcal{B}}_{v_i}(r)\end{pmatrix}=\begin{pmatrix}\mathbf{W}_i\mathbf{K}\grave{\mathcal{B}}_v(r)\\ \grave{\mathcal{B}}_{v_i}(r)\end{pmatrix}\in\mathbb{R}^{k_I+1}.
\end{aligned}\tag{63}$$

Hence, we observe that conditional on $\grave{\mathbf{x}}_{Li}$ the $T\to\infty$ limit distribution of the centered and scaled estimator of $\grave{\boldsymbol{\beta}}_L$ converges to a nonstandard limit containing the limit of the centered parameter $\check{\boldsymbol{\gamma}}$. $\mathbf{C}=\mathbf{I}_{nk_I}$ since $k_C=0$. The terms in (63) are $\mathcal{O}_p(1)$, such that $\sqrt{T_\star}\left(\widehat{\boldsymbol{\beta}}_L-\boldsymbol{\beta}_L\right)$ is $\mathcal{O}_p(1)$. Since the distribution of $\grave{\mathbf{x}}_{Li}$ has not been specified, the ("unconditional") asymptotic limit distribution of $\sqrt{T_\star}\left(\widehat{\boldsymbol{\beta}}_L-\boldsymbol{\beta}_L\right)$ cannot be obtained.

In the next step, we use the second part of Assumption 5 and apply a joint central limit theorem to $\frac{1}{\sqrt{nT_\star}}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\grave{u}_{it}^{\dagger}$.[16] Then for the joint limit we derive

$$\begin{aligned}
&\lim_{(n,T)\to\infty}\sqrt{nT_\star}\left(\widehat{\boldsymbol{\beta}}_L-\boldsymbol{\beta}_L\right)\\
&=\lim_{(n,T)\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}\cdot\Bigg[n^{-1/2}T_\star^{-1/2}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}\\
&\qquad+n^{1/2}T_\star\left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS}-\breve{\boldsymbol{\gamma}}\right)'n^{-1/2}n^{-1/2}T_\star^{-3/2}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'\Bigg]\\
&=\left(\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)\right)^{-1}\boldsymbol{\nu}_{(\grave{x}_{Li}\grave{u}_{ti})}\ \text{ where }\boldsymbol{\nu}_{(\grave{x}_{Li}\grave{u}_{ti})}\sim\mathcal{N}\left(\mathbf{0}_{(k_L\times 1)},\mathbf{D}_{\left(\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}\right)}\right)\ \text{ and}\\
&\mathbf{D}_{\left(\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}\right)}=\mathbb{E}\left(\Omega_{u_iu_i}^{\dagger}\right)\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)\ .
\end{aligned}\tag{64}$$

Note that by the independence of $\mathbf{x}_{Li}$ and $u_{it}^{\dagger}$ postulated in Assumption 2, $\mathbf{D}_{\left(\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}\right)}=\mathbb{E}\left(\left(\grave{u}_{it_\star}^{\dagger}\right)^2\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)+\sum_{s_\star=-\infty,\,s_\star\neq t_\star}^{\infty}\mathbb{E}\left(\left(\grave{u}_{is_\star}^{\dagger}\grave{u}_{it_\star}^{\dagger}\right)\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)=\mathbb{E}\left(\Omega_{u_iu_i}^{\dagger}\right)\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)$. To obtain (64) we have to show that the term $n^{-1/2}\left(n^{1/2}T_\star\left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS}-\breve{\boldsymbol{\gamma}}\right)'n^{-1/2}T_\star^{-3/2}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'\right)$ converges in probability to zero. To derive this result we show that $\left(n^{1/2}T_\star\left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS}-\breve{\boldsymbol{\gamma}}\right)'n^{-1/2}T_\star^{-3/2}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'\right)$ is $O_p(1)$. Note that $\grave{\mathbf{x}}_{Li}$ is stationary with an expectation equal to zero, $\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'$ is $I(1)$ and independent of $\grave{\mathbf{x}}_{Li}$ and $\sqrt{n}T_\star\left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS}-\breve{\boldsymbol{\gamma}}\right)'=O_p(1)$ [for $(n,T)\to\infty$ and $n^6/T\to 0$]. Then,

$$\begin{aligned}
&\lim_{(n,T)\to\infty}\frac{1}{T_\star}\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS}-\breve{\boldsymbol{\gamma}}\right)'\sum_{t_\star=1}^{T_\star}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'\\
&=\lim_{(n,T)\to\infty}T^{-1}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)\frac{\sqrt{n}T_\star}{\sqrt{n}T_\star}\left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS}-\breve{\boldsymbol{\gamma}}\right)\\
&=\lim_{(n,T)\to\infty}\left(n^{-1/2}T^{-3/2}\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\sum_{t_\star=1}^{T_\star}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'\right)O_p(1)\ .
\end{aligned}\tag{65}$$

For the term $n^{-1/2}T^{-3/2}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'$ we already observed in (63) that $\lim_{T\to\infty}T^{-3/2}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'=\grave{\mathbf{x}}_{Li}\int_0^1\grave{\mathbf{g}}_i(r)dr$. Then, the sequential "first T, then n"-limit of $n^{-1/2}T^{-3/2}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'$ is a normally distributed random vector with a k_L-dimensional mean vector of zeros and and finite covariance. Assumptions 5–7 are sufficient that also the joint limit of the term $n^{-1/2}T^{-3/2}\sum_{i=1}^{n}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'$ is a normal random variable. By this result, the fact that $\sqrt{n}T_\star\left(\widehat{\breve{\boldsymbol{\gamma}}}_{D2SLS}-\breve{\boldsymbol{\gamma}}\right)$ converges to a normal random variable and the continuous mapping theorem,

[16]To apply Phillips and Moon (1999, Theorem 3) let $\mathbf{m}_{\grave{\mathbf{X}}\grave{\mathbf{u}},Tni}:=\frac{1}{\sqrt{T_\star}}\sum_{t_\star=1}^{T_\star}\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}$. If the limes inferior of the smallest eigenvalue of the covariance of $\mathbf{m}_{\grave{X}\grave{u},Tni}$ is larger than zero and $||\mathbf{m}_{\grave{X}\grave{u},Tni}||_2$ is uniformly integrable in $T_\star$, this theorem can be applied. In the notation of Phillips and Moon (1999), $Y_{i,T_\star}=\mathbf{m}_{\grave{\mathbf{X}}\grave{\mathbf{u}},Tni}$ where we set $C_i=\mathbf{I}_{k_L}$ such that $Y_{i,T_\star}=C_iQ_{i,T_\star}=Q_{i,T_\star}$. The joint and the sequential limits are equal if $\frac{1}{n}\sum_{i=1}^{n}\Omega_{u_iu_i}\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'$ converges in probability to $\mathbf{D}_{\left(\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}\right)}$.

we observe that $n^{-1/2}\left(n^{1/2}T_\star\left(\widehat{\check{\boldsymbol{\gamma}}}_{D2SLS}-\check{\boldsymbol{\gamma}}\right)'n^{-1/2}T^{-3/2}\sum_{i=1}^{n}\grave{\mathbf{x}}_{Li}\sum_{t_\star=1}^{T_\star}\left(\grave{y}_{it}^{*},\grave{\mathbf{x}}_{Iit}'\right)'\right)$ converges to zero in probability. Hence, by (64) $\sqrt{nT_\star}\left(\widehat{\boldsymbol{\beta}}_L-\boldsymbol{\beta}_L\right)$ converges to a normally distributed random variable with mean zero and variance $\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}\mathbf{D}_{\left(\grave{\mathbf{x}}_{Li}\grave{u}_{it_\star}^{\dagger}\right)}\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}=\mathbb{E}\left(\Omega_{u_iu_i}^{\dagger}\right)\mathbb{E}\left(\grave{\mathbf{x}}_{Li}\grave{\mathbf{x}}_{Li}'\right)^{-1}$ [for $(n,T)\to\infty$ and $n^6/T\to 0$].

Funding

The authors appreciate helpful comments from the editor Esfandiar Maasoumi and two anonymous referees as well as from Manfred Deistler, Jaroslava Hlouskova, Justinas Pelenis, Julia Reynolds, M. Hashem Pesaran, Benedikt Pötscher, Klaus Pötzelberger, Stefan Schneeberger, Martin Wagner, Derya Uysal, the participants of the *CFE 2012 Conference*, the *3rd Humboldt-Copenhagen Conference*, the *Econometrics of Social Interaction Symposium* at the University of York 2013, the *Conference on Cross-sectional Dependence in Panel Data Models* in Cambridge 2013, the *SSES 2013 Conference*, the *ESEM 2013 Conference*, as well as research seminar participants at the *IHS*, the *European Business School*, the *University of Duisburg-Essen*, *Universitat Pompeu Fabra*, *NYU Abu Dhabi* and *Vienna University of Economics and Business* for interesting discussions and comments. Financial support from the Austrian Central Bank under Grant Nr. 13607 is gratefully acknowledged.

ORCID

Leopold Sögner http://orcid.org/0000-0001-5388-0601

References

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data*. 4th ed. Chichester, UK: Whiley.
Berndt, A., Douglas, R., Duffie, D., Ferguson, M., Schranz, D. (2008). Measuring default risk premia form default swap rates and EDFs. Working paper, Stanford University.
Billingsley, P. (1986). *Probability and Measure*. 2nd ed. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.
Binder, M., Hsiao, C., Pesaran, M. H. (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21(04):795–837.
Chudik, A., Pesaran, M. H. (2013a). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Globalization and Monetary Policy Institute Working Paper 146, Federal Reserve Bank of Dallas.
Chudik, A., Pesaran, M. H. (2013b). Large panel data models with cross-sectional dependence: A survey. Globalization and Monetary Policy Institute Working Paper 153, Federal Reserve Bank of Dallas.
Cliff, A., Ord, J. (1973). *Spatial Autocorrelation*. London: Pion.
Crosbie, P. and Bohn, J. (2003). Modeling default risk - Modeling methodology. New York: *KMV Corporation*.
Davidson, J. (1994). *Stochastic Limit Theory - An Introduction for Econometricians*. New York: Oxford University Press.
de Jong, R. M., Davidson, J. (2000). The functional central limit theorem and weak convergence to stochastic integrals I. *Econometric Theory* 16(05):621–642.
Drukker, D. M., Egger, P., Prucha, I. R. (2013). On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econometric Reviews* 32(5–6):686–733.
Heuser, H. (1992). *Funktionalanalysis (Theorie und Anwendung)*. 3rd ed. Wiesbaden: Teubner.
Horn, R. A., Johnson, C. R. (1985). *Matrix Analysis*. New York: Cambridge University Press.
Hsiao, C. (2015). *Analysis of Panel Data*. 3 ed. Econometric Society Monographs, Vol. 54. Cambridge: Cambridge University Press.
Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18(06):1449–1459.
Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
Kao, C., Chiang, M. (2000). On the estimation and inference of a cointegrated regression in panel data. In: Baltagi, B., ed. *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*. Kao/Chiang: Elsevier, New York.
Kapoor, M., Kelejian, H. H., Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics* 140:97–130.
Kejriwal, M., Perron, P. (2008). Data dependent rules for selection of the number of leads and lags in the dynamic ols cointegrating regression. *Econometric Theory* 24(05):1425–1441.

Kelejian, H. H., Prucha, I. R. (1998). A generalized spatial two-stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17:99–121.
Kelejian, H. H., Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40:509–533.
Kelejian, H. H., Prucha, I. R. (2008). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. CES Working paper 2448.
Kitamura, Y., Phillips, P. C. (1997). Fully modified iv, give and gmm estimation with possibly non-stationary regressions and instruments. *Journal of Econometrics* 80:85–123.
Klenke, A. (2008). *Probability Theory - A Comprehensive Course*. Klenke: Springer, Berlin.
Kurozumi, E., Tuvaandorj, P. (2010). Model selection criteria in multivariate models with multiple structural changes. Global COE Hi-Stat Discussion Paper Series gd10-144, Institute of Economic Research, Hitotsubashi University.
Mark, N. C., Ogaki, M., Sul, D. (2005). Dynamic seemingly unrelated cointegrating regressions. *Review of Economic Studies* 72(3):797–820.
Mark, N. C., Sul, D. (2003). Cointegration vector estimation by panel dols and long-run money demand. *Oxford Bulletin of Economics and Statistics* 65(5):655–680.
Park, J., Ogaki, M. (1991). Seemingly Unrelated Canonical Cointegrating Regressions. RCER Working Papers 280, University of Rochester - Center for Economic Research (RCER).
Park, J. Y., Phillips, P. C. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4(03):468–497.
Pedroni, P. (2000). Fully modified OLS for heterogeneous cointegrated panels. In: Baltagi, B., ed. *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*. Pedroni: Elsevier, New York.
Pesaran, M. (2015). *Time Series and Panel Data Econometrics*. Pesaran: Oxford University Press, Oxford.
Pesaran, M., Shin, Y. (1995). An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis. Cambridge Working Papers in Economics 9514, Faculty of Economics, University of Cambridge.
Phillips, P., Moon, H. (2000). Nonstationary panel data analysis: An overview of some recent developments. *Econometric Reviews* 19(3):263–286.
Phillips, P. C. (2014). Optimal estimation of cointegrated systems with irrelevant instruments. *Journal of Econometrics* 178(Part 2):210–224.
Phillips, P. C. B., Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57(1):99–125.
Phillips, P. C. B., Loretan, M. (1991). Estimating long run economic equilibria. *Review of Economic Studies* 58:407–436.
Phillips, P. C. B., Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica* 67(5): 1057–1112.
Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory*. New York: Oxford University Press.
Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7(1):1–21.
Schneider, P., Sögner, L., Veža, T. (2010). The economic role of jumps and recovery rates in the market for corporate default risk. *Journal of Financial and Quantitative Analysis* 45:1517–1547.
Sögner, L., Vogelsang, T. J., Wagner, M. (2017). Integrated Modified OLS Estimation of Spatially Correlated Cointegrated Systems. Technical report, Institute for Advanced Studies, Vienna Michigan State University and TU Dortmund.
Sögner, L., Wagner, M. (2017). Fully Modified *OLS* Estimation of Spatially Correlated Cointegrated Relationships. Technical report, Institute for Advanced Studies, Vienna and TU Dortmund.
Stock, J. H., Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61(4):783–820.
Vogelsang, T. J., Wagner, M. (2014). Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics* 178(2):741–760.
White, H. (2001). *Asymptotic Theory For Econometricians*. revised ed. Bingley, UK: Emerald Group Publishing.
Yu, J., de Jong, R., Lee, L.-F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. *Journal of Econometrics* 146(1):118–134.