



# QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices<sup>☆</sup>



Xi Qu<sup>a,\*</sup>, Lung-fei Lee<sup>b</sup>, Jihai Yu<sup>c</sup>

<sup>a</sup> Shanghai Jiao Tong University, China

<sup>b</sup> Ohio State University, United States

<sup>c</sup> Peking University, China

## ARTICLE INFO

### Article history:

Received 26 August 2015

Received in revised form

20 April 2016

Accepted 14 November 2016

Available online 5 December 2016

### JEL classification:

C13

C23

R15

### Keywords:

Spatial autoregression

Dynamic panels

Fixed effects

Endogenous spatial weights matrix

QMLE

## ABSTRACT

In spatial panel data models, when a spatial weights matrix is constructed from economic or social distance, spatial weights could be endogenous and also time varying. This paper presents model specification and proposes QMLE estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. Asymptotic properties of the proposed QMLE are rigorously established. We extend the notion of spatial near-epoch dependence to allow time dependence. By using spatial-time LLN for near-epoch dependence process and CLT for martingale difference sequence, we establish the consistency and asymptotic normality of QMLE. Monte Carlo experiments show that the proposed estimators have satisfactory finite sample performance.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Spatial econometrics deals with interactions of economic units in space, where the space can be physical or economic in nature. It can be applied to various fields in economics such as tax/expenditure competition in public economics, comovement of commodity price in regional economics, spillover effects in growth theory, and peer effects in labor economics, etc. Recently there

has been much interest in spatial econometrics in panel data settings, because panel data models take into account the dynamic and spatial dependence and also control for unobservable heterogeneity.<sup>1</sup> Some related recent surveys are Anselin et al. (2008), Elhorst (2010), Baltagi (2011), and Lee and Yu (2010a, 2015).

In constructing the spatial weights matrix, when it is based only on geographical information, it is usually exogenous and time invariant. However, when it is also constructed with economic or social characteristics of regions, it could be time varying, and more importantly, endogenous. For example, Case et al. (1993) on state spending have weights based on the difference in the percentage of the population that is black. Brueckner (1998) and Brueckner and Saavedra (2001) inflate weights according to population figures in studies of local governments competitions on economic policies. Baicker (2005) constructs a weights matrix with the degree of

<sup>☆</sup> We would like to thank the co-editor Oliver Linton and three anonymous referees for helpful comments and suggestions. An earlier version of this paper has been presented in Tsinghua International Conference on Econometrics at Tsinghua University, May 14–15, 2015; the 11th International Symposium on Econometric Theory and Applications (SETA), May 30–31, 2015; A Cross-Strait Dialogue (III) in Academia Sinica, April 12, 2016. We thank participants of the conferences for their comments. Qu acknowledges funding from the National Science Foundation of China (Project no. 71601115). Yu acknowledges funding from the National Science Foundation of China (Project no. 71322105 and 71532001) and support from the Center for Statistical Science of Peking University.

\* Corresponding author.

E-mail addresses: [xiqu@sjtu.edu.cn](mailto:xiqu@sjtu.edu.cn) (X. Qu), [lee.1777@osu.edu](mailto:lee.1777@osu.edu) (L.-f. Lee), [jihai.yu@gmail.com](mailto:jihai.yu@gmail.com) (J. Yu).

<http://dx.doi.org/10.1016/j.jeconom.2016.11.004>

0304-4076/© 2016 Elsevier B.V. All rights reserved.

<sup>1</sup> For static spatial panel data models, we have various models depending on fixed/random effects specifications and spatial features, see Anselin (1988), Baltagi et al. (2003, 2013), Kapoor et al. (2007), Fingleton (2008), and Lee and Yu (2010b, 2012a), etc. For dynamic models with spatial interaction, see Elhorst (2005), Mutl (2006), Yu et al. (2008, 2012), Korniotis (2010), Yu and Lee (2010), Lee and Yu (2012b), Su and Yang (2015) for detailed discussions.

population mobility between regions. So does Rincke (2010), who exploits information on commuting flows between locations. These spatial weights matrices are time varying, and could also be endogenous.

The time varying feature of exogenous spatial weights matrices is investigated in Lee and Yu (2012b), which shows that ignoring time varying feature would cause estimation bias and lead to incorrect inferences. The endogenous feature of spatial weights matrix is only considered under the cross-sectional spatial autoregressive (SAR) models in Qu and Lee (2015), where disturbances in the regression equation may be correlated with elements that generate the weights. In a panel data setting, with dynamic features and a time varying spatial weights matrix, it is of practical relevance to investigate an endogenous spatial weights matrix. Therefore, with increasing availability of panel data, it is of interest to investigate spatial dynamic panel data (SDPD) models with endogenous time varying spatial weights matrices.

Kelejian and Piras (2014) propose a 2SLS estimation for a short spatial panel data with endogenous regressors and endogenous weights, utilizing some high-level assumptions on the existence of relevant statistics. In this paper, we will explicitly model sources of endogeneity of a specified spatial weights matrix similar to Qu and Lee (2015), so that we have two estimation equations: one is the spatial panel outcome equation, and the other is entries of a spatial weights matrix. When disturbances in the spatial panel outcome and in the entries of spatial weights matrix are correlated, we have the case of an endogenous spatial weights matrix. By specifying conditional mean independence and homoscedasticity, Qu and Lee (2015) overcome the endogeneity problem in the spatial outcome equation by using a control function from the spatial weights matrix entries' equation. They make use of the law of large numbers (LLN) for spatial near-epoch dependence (NED) process developed in Jenish and Prucha (2012). The concept of NED originates in the time series literature. The spatial NED process in Jenish and Prucha (2012) generalizes that notion to spatial setting. In the current paper, the time series and spatial NED notions are merged to capture both time and space features, and applied to our spatial dynamic panel setting. However, contrary to the cross-sectional case in Qu and Lee (2015), neither the central limit theorem (CLT) for spatial NED nor that for time series can be applied to panel data models with both large  $n$  and  $T$  due to the presence of individual and/or time effects.<sup>2</sup> For asymptotic distribution of quasi-maximum likelihood (QML) estimators, we develop a modified martingale difference CLT, which can be applied to our endogenous spatial dynamic situation.<sup>3</sup>

In addition to providing asymptotic properties of QML estimation for SDPD models with endogenous time varying spatial weights matrices, the current paper offers the following contributions to the literature. First, we extend the NED analysis to the spatial dynamic panel setting, which is a theoretical contribution. To show consistency and asymptotic normality of QML estimators, the classical asymptotic theory for the linear and quadratic forms cannot be directly applied. Although Qu and Lee (2015) applied NED analysis in a cross-sectional data setting, the extension to general double arrays is important to build our spatial-time LLN. Second, the spatial panel model with endogenous weights matrix in

the current paper is important for various empirical applications with time varying weights matrix. When the time varying weights matrices are constructed from economic or social variables, these variables could be endogenous, which invalidates the usual estimation procedure that leads to estimation bias. The method proposed in the current paper provides a robust approach regardless whether the weights matrices are endogenous or not. Third, due to individual and time effects in the dynamic panel data setting, bias correction is needed not only for the estimation of the regression parameters but also for the standard errors of these estimates. These procedures cannot be implied by the cross sectional case in Qu and Lee (2015), so that a separate treatment of the endogenous issue in the panel setting is needed.

The rest of the paper is organized as follows. Section 2 presents the model and a space specification. Section 3 derives the log likelihood function for the model and presents a spatial-time LLN needed for asymptotic analysis, which extends the NED property in Jenish and Prucha (2012) to the SDPD process. A modified martingale difference CLT is also presented. Section 4 investigates consistency and asymptotic distribution of the QMLE. As asymptotic biases may be present, we also propose a bias corrected estimator. Monte Carlo experiments are conducted in Section 5 to study the finite sample performances of various estimators, and an empirical application on cigarette demand is provided. Relevant algebraic expressions and proofs are collected in Appendices.

## 2. The model

### 2.1. The model and endogenous $W_{nt}$

The SDPD model considered is

$$Y_{nt} = \lambda_0 W_{nt} Y_{nt} + \rho_0 W_{n,t-1} Y_{n,t-1} + \gamma_0 Y_{n,t-1} + X_{1nt} \beta_0 + \mathbf{c}_{n10} + \alpha_{t10} l_n + V_{nt}, \quad t = 1, 2, \dots, T, \quad (1)$$

where  $Y_{nt} = (y_{1t}, y_{2t}, \dots, y_{nt})'$  and  $V_{nt} = (v_{1t}, v_{2t}, \dots, v_{nt})'$  are  $n \times 1$  column vectors, and  $v_{it}$ 's are *i.i.d.* across  $i$  and  $t$  with zero mean and variance  $\sigma_0^2$ . The  $X_{1nt}$  is an  $n \times k_1$  matrix of individual and time varying nonstochastic regressors,<sup>4</sup>  $\mathbf{c}_{n10}$  is an  $n \times 1$  column vector of individual fixed effects, and  $\alpha_{t10}$  is the  $t$ th element of the  $T \times 1$  time fixed effect vector  $\boldsymbol{\alpha}_{T10}$ , and  $l_n$  is the  $n \times 1$  vector of ones. The initial values in  $Y_{n0}$  are observable.

Denote  $S_{nt}(\lambda) = I_n - \lambda W_{nt}$  for an arbitrary  $\lambda$  at time period  $t$  and let  $S_{nt} = S_{nt}(\lambda_0)$  be its true value. The reduced form of (1) is

$$Y_{nt} = A_{nt} Y_{n,t-1} + S_{nt}^{-1} (X_{1nt} \beta_0 + \mathbf{c}_{n10} + \alpha_{t10} l_n + V_{nt}),$$

where  $A_{nt} = S_{nt}^{-1} (\gamma_0 I_n + \rho_0 W_{n,t-1})$ . A sufficient condition for the eigenvalues of  $A_{nt}$  being smaller than 1 is  $|\lambda c_w| < 1$  and  $\frac{|\gamma| + |\rho c_w|}{1 - |\lambda c_w|} < 1$  where  $c_w \geq \sup_{n,t} \|W_{nt}\|_\infty$  is a bounded constant.

This is so, because  $\|S_{nt}^{-1}\|_\infty \leq 1 + |\lambda_0 c_w| + |\lambda_0 c_w|^2 + \dots = \frac{1}{1 - |\lambda_0 c_w|}$ , and hence,  $\|A_{nt}\|_\infty \leq \|S_{nt}^{-1}\|_\infty (|\gamma_0| + |\rho_0| \|W_{nt}\|_\infty) \leq \frac{|\gamma_0| + |\rho_0 c_w|}{1 - |\lambda_0 c_w|}$ .

<sup>2</sup> If the data process starts in an initial period (not too far) from observations periods and the sample  $T$  periods are small, then it extends the cross-sectional model in Qu and Lee (2015), but we can use the spatial NED analysis for the asymptotic properties. This case is discussed in the working paper of Qu et al. (2016). However, with the large  $T$  setting, we need to explore the new spatial-time NED property along with  $n$  and  $T$  asymptotics.

<sup>3</sup> This modified martingale difference CLT is not the same one in Yu et al. (2008). The latter, which is developed for linear and quadratic forms with nonstochastic linear coefficients and quadratic matrix, cannot be applied to spatial dynamic panel model with endogenous spatial weights matrices.

<sup>4</sup> When there are additional endogenous regressors in the estimating equation, which are not contained in  $Z_{nt}$ , the QML estimation in the current paper cannot be applied directly. When the endogenous explanatory regressors are different from  $Z_{nt}$  variables, which explains  $W_{nt}$ , generalized method of moments (GMM) estimation can deal with such a situation. Brief discussions of the two stage instrumental variable (2SIV) and GMM estimation of such a model are provided in a supplement file available upon request (see Appendix E). When the endogenous variables are just from  $Z_{nt}$  variables, the ML method in the current paper is still applicable as long as the regressors in the  $Z_{nt}$  equation have relevant distinct variables from  $X_{1nt}$  (such as  $Z_{n,t-1}$ ). See the empirical application in Section 5.2. We would have an under-identification situation if  $Z_{nt}$  is also a (endogenous) regressor in the  $Y_{nt}$  equation along with all relevant explanatory variables,  $Z_{n,t-1}$  and  $X_{2nt}$ . In the later situation, the control variables would be linearly dependent on all explanatory variables of  $Y_{nt}$ .

It follows by the spectral radius theorem (see Theorem 5.6.9 in [Horn and Johnson, 1985](#)) that all the eigenvalues of  $A_{nt}$  are less than one in absolute value.

The spatial weights matrix  $W_{nt}$  generates the dependence of  $y_{it}$ 's across spatial units, and it could be time varying. When it is constructed from economic or social distance, it could be endogenous. We allow  $W_{nt}$  to be time varying and endogenous, by assuming that  $W_{nt}$  can be constructed from  $Z_{nt}$  where  $Z_{nt} = (z_{1t}, z_{2t}, \dots, z_{nt})'$  is an  $n \times p$  matrix generated by<sup>5</sup>

$$Z_{nt} = Z_{n,t-1}\gamma_0 + X_{2nt}\Gamma_0 + \mathbf{c}_{n20} + l_n\alpha'_{t20} + \epsilon_{nt}, \quad (2)$$

where  $\mathbf{c}_{n20}$  is of dimension  $n \times p$ ,  $\alpha_{t20}$  is a  $p \times 1$  vector, and  $\epsilon_{nt}$  is of dimension  $n \times p$ . Specifically, we assume that  $(W_{nt})_{ij} = w_{ij,nt} = h(z_{i,nt}, z_{j,nt})$  where  $h(\cdot)$  is a bounded function. Here,  $Z_{nt}$  is a VAR process with individual effects  $\mathbf{c}_{n20}$  and time effects  $\alpha_{t20}$ , where  $\gamma_0$  is a  $p \times p$  VAR dynamic coefficient matrix<sup>6</sup> and  $\Gamma_0$  is a  $k_2 \times p$  coefficient matrix of individual and time varying regressors of  $X_{2nt}$ .

**Assumption 1.** The error terms  $v_{it}$  and  $\epsilon_{it}$  have the distribution  $(v_{it}, \epsilon'_{it})' \sim \text{i.i.d. } (0, \Sigma_{v\epsilon 0})$  where  $\Sigma_{v\epsilon 0} = \begin{pmatrix} \sigma_{v0}^2 & \sigma'_{v\epsilon 0} \\ \sigma_{v\epsilon 0} & \Sigma_{\epsilon 0} \end{pmatrix}$ ,  $\sigma_{v0}^2$  is a scalar variance of  $v_{it}$ ,  $\Sigma_{\epsilon 0}$  is a  $p \times p$  variance matrix of  $\epsilon_{it} = (\epsilon_{it,1}, \dots, \epsilon_{it,p})'$ , and  $\sigma_{v\epsilon 0}$  is  $p \times 1$  covariance matrix between  $v_{it}$  and  $(\epsilon_{it,1}, \dots, \epsilon_{it,p})'$ . The  $\sup_{i,t} E|v_{it}|^{4+\delta_e}$  and  $\sup_{i,t} E|\epsilon_{it}|^{4+\delta_e}$  are finite for some  $\delta_e > 0$ . Furthermore,  $E(v_{it}|\epsilon_{it}) = \epsilon'_{it}\delta_0$  and  $\text{Var}(v_{it}|\epsilon_{it}) = \sigma_{\xi 0}^2$ .

**Assumption 1** specifies that, as  $W_{nt}$  is a function of  $\epsilon_{nt}$ ,  $W_{nt}$  is endogenous in the estimation equation (1) if  $\epsilon_{nt}$  and  $V_{nt}$  are correlated. Therefore, unless  $\sigma_{v\epsilon} = 0$ , the endogeneity of  $W_{nt}$  will cause estimation bias if we disregard this feature.

For estimation, we need the conditional mean of  $v_{it}$  (conditional on  $\epsilon_{it}$ ). From the joint distribution of  $(v_{it}, \epsilon'_{it})'$ , we have  $E(v_{it}|\epsilon_{it}) = \epsilon'_{it}\delta_0$  where  $\delta_0 = \Sigma_{\epsilon 0}^{-1}\sigma_{v\epsilon 0}$  and  $\text{Var}(v_{it}|\epsilon_{it}) = \sigma_{\xi 0}^2$  where  $\sigma_{\xi 0}^2 = \sigma_{v0}^2 - \sigma'_{v\epsilon 0}\Sigma_{\epsilon 0}^{-1}\sigma_{v\epsilon 0}$ . Denote  $\xi_{nt} = V_{nt} - \epsilon_{nt}\delta_0$ . Then, the expectation of  $\xi_{nt}$  conditional on  $\epsilon_{nt}$  is zero and its conditional variance matrix is  $\sigma_{\xi 0}^2 I_n$ . Thus,  $\xi_{nt}$  is uncorrelated with the terms of  $\epsilon_{nt}$ . Therefore, (1) can be rewritten as

$$\begin{aligned} Y_{nt} = & \lambda_0 W_{nt} Y_{nt} + \rho_0 W_{n,t-1} Y_{n,t-1} + \gamma_0 Y_{n,t-1} + X_{1nt} \beta_0 \\ & + \mathbf{c}_{n10} + \alpha_{t10} l_n \\ & + (Z_{nt} - Z_{n,t-1} \gamma_0 - X_{2nt} \Gamma_0 - \mathbf{c}_{n20} - l_n \alpha'_{t20}) \delta_0 + \xi_{nt}, \end{aligned} \quad (3)$$

where  $E(\xi_{nt}|\epsilon_{nt}) = 0$  and  $\text{Var}(\xi_{nt}|\epsilon_{nt}) = \sigma_{\xi 0}^2 I_n$ , and elements of  $\xi_{nt}$  are i.i.d. across  $i$  and  $t$ . Here,  $(Z_{nt} - Z_{n,t-1} \gamma_0 - X_{2nt} \Gamma_0 - \mathbf{c}_{n20} - l_n \alpha'_{t20})$  are control variables in  $Y_{nt}$  to control the endogeneity of  $W_{nt}$ .

## 2.2. Topological specification

Following [Jenish and Prucha \(2012\)](#), we consider spatial processes located on a possibly unevenly spaced lattice. Denote  $\|Z\|_p = [E|Z|^p]^{1/p}$  when the absolute  $p$ th moment exists.

<sup>5</sup> We notice that the  $Z$  equation is an augmented equation to specify the endogeneity of spatial weights matrix, but not part of a linear simultaneous equation system as in [Beenstock and Felsenstein \(2007\)](#). More generally, the spatial lag term  $M_{nt}Z_{nt}$  could be specified in the  $Y$  or  $Z$  equations as long as  $M_{nt}$  is exogenous. If  $M_{nt}$  is the endogenous  $W_{nt}$ , (1) and (2) becomes a highly nonlinear system and we may have possible multiple solutions. Due to these complicated issues, we do not cover endogenous weights matrix in  $Z$  equation.

<sup>6</sup> We note that the VAR representation includes the multivariate linear model as a special case.

**Definition 1.** Let  $Z = \{Z_{l,L}, l \in D_L, L \geq 1\}$  be a random field with  $\|Z_{l,L}\|_p < \infty$ ,  $p \geq 1$  and  $\epsilon = \{\epsilon_{l,L}, l \in D_L, L \geq 1\}$  be another random field, where  $|D_L| \rightarrow \infty$  as  $L \rightarrow \infty$ , and let  $d = \{d_{l,L}, l \in D_L, L \geq 1\}$  be an array of finite positive constants. Then the random field  $Z$  is said to be  $L_p(d)$ -near-epoch dependent on the random field  $\epsilon$  if

$$\|Z_{l,L} - E(Z_{l,L}|\mathcal{F}_{l,L}(s))\|_p \leq d_{l,L}\psi(s)$$

for some sequence  $\psi(s) \geq 0$  with  $\lim_{s \rightarrow \infty} \psi(s) = 0$ , where  $\mathcal{F}_{l,L}(s) = \sigma(\epsilon_{j,L} : j \in D_L, \rho(l, j) \leq s)$  is the  $\sigma$ -field generated by  $\epsilon_{j,L}$  within distance  $s$  from  $l$ . Here  $l$  is an index as well as a location for simplification.

For the SDPD model (1), while we understand that we are working in general double arrays, it might be convenient to adopt the subscript  $n$  to simplify notation. For such a model, due to non-linearity in  $W_{nt}$ 's, we consider the sequences  $\{y_{it}\}$  and  $\{(v_{it}, \epsilon'_{it})'\}$  as random fields over space and time for analysis. We are going to relate our random variables  $\{y_{it}\}$  to the notion of a random field in [Jenish and Prucha \(2012\)](#) so that we can use their setting and the spatial NED to build our spatial-time LLN. At time  $t = 0$ , the  $n$  units are located in the Euclidean space  $R^d$  of dimension  $d \geq 1$ , which is a subspace (plane) of  $R^{d+1}$ . At time  $t = 1$ , the  $n$  units shift vertically upward to an affine plane parallel to  $R^d$  plane. As time passes, it keeps shifting upward each time by one time unit. We may also accommodate variables in past periods (history) by shifting the plane  $R^d$  at  $t = 0$  downward to an affine plane at  $t = -1$ , etc. So for a sample with  $n$  units over  $T$  periods, we have located individuals in the  $R^{d+1}$  space. Using the maximum metric ([Jenish and Prucha, 2012](#))

$$\|l(i, t) - l(j, \tau)\|_\infty = \max\{|t - \tau|, \|l(i, 0) - l(j, 0)\|_\infty\},$$

each observation indexed by  $(i, t)$  will be located at  $l(i, t) \in R^{d+1}$  where  $l(i, 0)$  is the physical location of an individual at time 0. We shall assume at  $t = 0$ , any two individuals are located in the  $R^d$  plane at least one unit apart. For each spatial unit  $i$ , the location  $l(i, t)$  is always one unit apart from  $l(i, t-1)$  with respect to the time dimension. Denote  $L = nT$  the total number of observations in a spatial panel with  $n$  cross sectional units and  $T$  time periods. We have  $L$  locations in  $R^{d+1}$ . Denote the collection of those locations by  $D_L$ . With an observed  $W_{nt}$  and  $W_{n,t-1}$ , for our spatial and time dynamic specification, individual spatial unit  $i$  at time  $t$  directly links to  $j$  at time  $t$  if  $w_{ij,nt} \neq 0$ ; and it links to itself at time  $t-1$  and other spatial units  $j$  whenever  $w_{ij,n,t-1} \neq 0$ . Due to the time interaction specification, any individual at time  $t$  does not directly link to others including itself at the future  $t + \tau$  for  $\tau \geq 1$ , and not the past  $t - \tau$  with  $\tau \geq 2$ . However, an individual  $i$  at time  $t$  may link to other units at  $t - \tau$  for  $\tau \geq 2$  indirectly. In particular,  $i$  links indirectly to itself in past periods.

So in this setting, we have implicitly defined a mapping of  $(i, t)$  to a location  $l$  in  $D_L$ , i.e.,  $l = l(i, t) \in D_L$  with  $L = nT$ . And we have  $y_l$  corresponds to a  $y_{it}$ . Therefore, we can study the NED with a base  $\sigma$ -field generated by  $\{v_{it}, \epsilon'_{it}\}$ . For our model, at location  $l = l(i, t)$ , may define the base  $\sigma$ -field for spatial NED process by

$$\mathcal{F}_{l,L}(s) = \sigma((v_{j,\tau}, \epsilon'_{j,\tau}) : l(j, \tau) \in D_L, \|l(j, \tau) - l(i, t)\|_\infty \leq s).$$

The  $Y = \{y_{l,t}\}$  is an NED if  $\|y_{l,L} - E(y_{l,L}|\mathcal{F}_{l,L}(s))\|_p \leq d_{l,L}\psi(s)$  for some  $\psi(s)$ , which goes to zero as  $s \rightarrow \infty$ .

From the SDPD model, as  $Y_{nt}$  depends on  $W_{nt}Y_{nt}$ ,  $Y_{n,t-1}$  and  $W_{n,t-1}Y_{n,t-1}$ , we can imagine the following possible “direct” links. At time  $t$ ,

(1) via  $W_{nt}Y_{nt}$ , unit  $(i, t)$  interacts with units  $(j, t)$  where  $w_{ij,nt} \neq 0$ ;

(2) via  $Y_{n,t-1}$ , unit  $(i, t)$  interacts with unit  $(i, t-1)$  due to time lag (state dependence);

(3) via  $W_{n,t-1}Y_{n,t-1}$ , unit  $(i, t)$  interacts with units  $(j, t-1)$  where  $w_{ij,n(t-1)} \neq 0$  (diffusion).

In the cross-sectional case, we have assumed that individual units can directly interact with each other within a limited distance, say  $\rho_c > 1$ . For the extension to a panel setting, we might assume also that at each time  $t$ , the cross-sectional interactions of spatial units are still within the distance  $\rho_c$ , which does not change over time. It is possible to allow  $\rho_c$  to change over time, but we still need that sequence to be bounded; without loss of generality, we may take  $\rho_c$  to be invariant over time. Note that in our geographic setting, we do not allow locations of any individual in physical space to be changed over time. This might be a limitation if the geographic space is defined in terms of some economic characteristics (so characteristic space). Still in the setting, even physical locations do not change, the strength of links in  $w_{ij,nt}$  can change with  $z$ 's over time. Within the distance  $\rho_c$ , even if  $w_{ij,n(t-1)} = 0$ , it is possible to have  $w_{ij,nt} \neq 0$ .

### 2.3. The model in the matrix form

To make use of the space setting in Section 2.2, we may put the model (1) into a big matrix form. Denote

$$Y_L = \begin{pmatrix} Y_{n1} \\ Y_{n2} \\ \vdots \\ Y_{nT} \end{pmatrix}, \mathcal{J}_0(\gamma_0, \rho_0) = \begin{pmatrix} \gamma_0 Y_{n0} + \rho_0 W_{n0} Y_{n0} \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\alpha_{1L} = \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{T1} \end{pmatrix} \otimes I_n, \alpha_{2L} = \begin{pmatrix} \alpha'_{12} \\ \alpha'_{22} \\ \vdots \\ \alpha'_{T2} \end{pmatrix} \otimes I_n,$$

$X_{1L}, \epsilon_L, \xi_L, Z_L$  and  $X_{2L}$  are defined similarly. Also, denote

$$W_L(\eta) = \lambda W_{1L} + \gamma W_{2L} + \rho W_{3L}$$

with

$$W_{1L} = \begin{pmatrix} W_{n1} & & & \\ & W_{n2} & & \\ & & \ddots & \\ & & & W_{nT} \end{pmatrix}, W_{2L} = \begin{pmatrix} 0 & & & \\ I_n & 0 & & \\ & \ddots & \ddots & \\ & & I_n & 0 \end{pmatrix},$$

$$W_{3L} = \begin{pmatrix} 0 & & & \\ W_{n1} & 0 & & \\ & \ddots & \ddots & \\ & & W_{n,T-1} & 0 \end{pmatrix}$$

so that

$$W_L(\eta_0) = \begin{pmatrix} \lambda_0 W_{n1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \gamma_0 I_n + \rho_0 W_{n1} & \lambda_0 W_{n2} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \gamma_0 I_n + \rho_0 W_{n,T-1} & \lambda_0 W_{nT} \end{pmatrix},$$

where  $\eta = (\lambda, \gamma, \rho)$  and  $\eta_0$  is its true value. Therefore, the model (1) with (2) can be written in the matrix form as:

$$Y_L = W_L(\eta_0)Y_L + X_{1L}\beta_0 + \mathcal{J}_0(\gamma_0, \rho_0) + \mathbf{c}_{1L0} + \alpha_{1L0} + \epsilon_L\delta_0 + \xi_L,$$

$$Z_L = Z_{L,-1}\gamma_0 + X_{2L}\Gamma_0 + \mathbf{c}_{2L0} + \alpha_{2L0} + \epsilon_L,$$

where  $\mathbf{c}_{1L} = I_T \otimes \mathbf{c}_{1n}$ ,  $\mathbf{c}_{2L} = I_T \otimes \mathbf{c}_{2n}$ .

We are interested in the  $nT \times nT$  matrix  $W_L(\eta)$  because we can investigate locations of individual units over time in  $\mathbb{R}^{d+1}$  which allows us to investigate spatial-time NED processes. This specification has the following convenience for analysis:

(i) The time index  $t$  and the row block number of  $W_L$  are in a compatible increasing order. At location  $l(i, t)$ , it corresponds to a unit numbered  $l$  among the total  $L = nT$  units in panel as  $l = n(t-1) + i$  for  $i = 1, \dots, n$  and  $t = 1, \dots, T$ .

(ii) One can imagine that at  $t = 1$ , individual units link with each other and form  $W_{n1}$ . At time  $t$ , individual units keep their links of the previous period  $t-1$ , i.e.,  $W_{n,t-1}$  is kept, and attempt to create links with others at time  $t$ , i.e., to form  $W_{nt}$  as new links.

(iii) For future, at time  $t$ , there are no attempt to form  $W_{n,t+s}$  for  $s \geq 1$ .

In the cross-sectional case, there are no cross constraints on the links of  $W_n$  (except for a maximum number of possible links). However, for  $W_L$  in the dynamic panel, there are restrictions across rows of  $W_L$ . For example, at the location  $l(i, t)$ , the links with  $l(i, t-1)$  and  $l(j, t-1)$  for  $j \neq i$  are already given (determined by the past time under the dynamic specification), and new links are possible only with  $l(j, t)$  at time  $t$ .

From Assumption 1, we have  $V_L = \epsilon_L\delta_0 + \xi_L$  where  $\xi_L \sim (0, \sigma_{\xi 0}^2 I_L)$ . Hence,

$$Y_L = W_L(\eta_0)Y_L + X_{1L}\beta_0 + \mathcal{J}_0(\gamma_0, \rho_0) + \mathbf{c}_{1L0} + \alpha_{1L0} \\ + (Z_L - Z_{L,-1}\gamma_0 - X_{2L}\Gamma_0 - \mathbf{c}_{2L0} - \alpha_{2L0})\delta_0 + \xi_L \\ = W_L(\eta_0)Y_L + X_{1L}\beta_0 + \mathcal{J}_0(\gamma_0, \rho_0) + \mathbf{c}_{1L0} + \alpha_{1L0} \\ + (Z_L - Z_{L,-1}\gamma_0 - X_{2L}\Gamma_0)\delta_0 + \xi_L, \quad (4)$$

where  $\mathbf{c}_{1L0} = \mathbf{c}_{1L0} - \mathbf{c}_{2L0}\delta_0$  and  $\alpha_{1L0} = \alpha_{1L0} - \alpha_{2L0}\delta_0$ .

For the  $W_{jL}$ ,  $j = 1, 2, 3$ , matrices in  $W_L(\eta_0)$ , as  $\lambda$ ,  $\gamma$  and  $\rho$  are unknown parameters, when we take derivatives of  $W_L(\eta)$  with respect to  $\eta$ , those  $W_{jL}$  matrices will show up. As  $W_L(\eta) = \lambda W_{1L} + \gamma W_{2L} + \rho W_{3L}$ , by assuming  $\sup_{n,t} \|W_{nt}\|_\infty \leq c_w$ , we have

(1)  $\|W_L(\eta)\|_\infty \leq |\lambda| \|W_{1L}\|_\infty + |\gamma| \|W_{2L}\|_\infty + |\rho| \|W_{3L}\|_\infty \leq c_\theta$ , where  $c_\theta = |\lambda|c_w + |\gamma| + |\rho|c_w$ .

(2) When  $c_\theta$  is assumed to be less than one,  $\sup_\delta \|(I_L - W_L(\eta))^{-1}\|_\infty \leq \sup_\delta (1 + \|W_L(\eta)\|_\infty + \dots + \|W_L^l(\eta)\|_\infty + \dots) \leq 1 + c_\theta + \dots + c_\theta^l + \dots = \frac{1}{1-c_\theta} < \infty$ .

Denote  $S_L(\eta) = I_L - W_L(\eta)$ ,  $S_L = I_L - W_L(\eta_0)$  and correspondingly  $G_{jL}(\eta) = W_{jL}S_L^{-1}(\eta)$  and  $G_{jL} = W_{jL}S_L^{-1}$  for  $j = 1, 2, 3$ . The following properties can be easily checked with the Neumann series expansion of  $S_L^{-1}(\eta)$ .

(3)  $\text{tr}(G_{1L}(\eta)) = \sum_{t=1}^T \text{tr}(G_{nt}(\lambda))$  where  $G_{nt}(\lambda) = W_{nt}S_{nt}^{-1}(\lambda)$ .

(4)  $\text{tr}(G_{2L}(\eta)) = 0$ .

(5)  $\text{tr}(G_{3L}(\eta)) = 0$ .

The item (3) follows because the diagonal blocks of  $G_{1L}(\eta)$  consist of  $G_{n1}(\lambda), \dots, G_{nT}(\lambda)$ . The items (4) and (5) follow because the diagonal blocks of  $G_{2L}(\eta)$  and  $G_{3L}(\eta)$  are all zeros.

### 3. Quasi-maximum likelihood estimation

#### 3.1. The concentrated log likelihood function and some key statistics

For asymptotic analysis, it is useful to have the likelihood function presented in matrix form involving  $W_L(\eta)$ . Denote  $\theta = (\lambda, \phi'_1, \delta', \phi'_2, \alpha', \sigma_\xi^2)'$  where  $\phi_1 = (\gamma, \rho, \beta')'$ ,  $\phi_2 = \text{vec}(\Phi_2)$  with  $\Phi_2 = (\gamma', \Gamma')'$  and  $\alpha$  is the  $J \times 1$  column vector of distinct elements in  $\Sigma_\epsilon$ . Corresponding to  $\phi_1$  and  $\Phi_2$ , we define  $R_{nt} = [Y_{n,t-1}, W_{n,t-1}Y_{n,t-1}, X_{1nt}]$ ,  $\mathcal{K}_{nt} = [Z_{n,t-1}, X_{2nt}]$  and their matrix form  $R_L$  and  $\mathcal{K}_L$ . The current model is recursive as  $Y_L$  depends on  $Z_L$  through  $W_L$ , and  $Z_L$  depends on  $\epsilon_L$  which is correlated with the disturbance term in  $Y_L$ . But  $Z_L$  does not directly depend on  $Y_L$ , so the unconditional pdf of  $Z_L$  is straightforward. Due to the correlation of disturbances in  $Y_L$  and  $Z_L$ , the formation of  $Z_L$  provides useful information to infer parameters in the  $Y_L$  equation. Therefore, to write down the log likelihood function, it is natural to construct the joint Gaussian pdf of  $Y_L$  and  $Z_L$ , obtained through the conditional Gaussian pdf of  $Y_L$  given  $Z_L$  and the marginal Gaussian pdf of  $Z_L$ .



Under this Gaussian pdf assumption, the likelihood would be the exact one. Meanwhile, we can also regard this likelihood function as a quasi-likelihood, by assuming that conditional on  $Z_L$ ,  $Y_L$  has a linear expectation, and the first and second conditional moments satisfy [Assumption 1](#), but does not impose additional assumptions on third or four moments.

From (4), the log likelihood function in the matrix form is

$$\begin{aligned} \ln L_L(\theta, \mathbf{c}_{1L}^-, \alpha_{1L}^-, \mathbf{c}_{2L}, \alpha_{2L}) = & -\frac{nT}{2} \ln 2\pi + \ln |I_L - W_L(\eta)| \\ & -\frac{nT}{2} \ln \sigma_\xi^2 - \frac{nT}{2} \ln |\Sigma_\epsilon| \\ & -\frac{1}{2} \text{vec}'(Z_L - \mathcal{K}_L \Phi_2 - \mathbf{c}_{2L} - \alpha_{2L})(\Sigma_\epsilon^{-1} \otimes I_L) \\ & \times \text{vec}(Z_L - \mathcal{K}_L \Phi_2 - \mathbf{c}_{2L} - \alpha_{2L}) \\ & -\frac{1}{2\sigma_\xi^2} [Y_L - W_L(\eta)Y_L - X_{1L}\beta - (Z_L - \mathcal{K}_L \Phi_2)\delta \\ & - \mathcal{I}_0(\gamma, \rho) - \mathbf{c}_{1L}^- - \alpha_{1L}^-]' \\ & \cdot [Y_L - W_L(\eta)Y_L - X_{1L}\beta - (Z_L - \mathcal{K}_L \Phi_2)\delta \\ & - \mathcal{I}_0(\gamma, \rho) - \mathbf{c}_{1L}^- - \alpha_{1L}^-]. \end{aligned}$$

In terms of  $J_n = I_n - \frac{1}{n}I_n I_n'$  and  $J_T = I_T - \frac{1}{T}I_T I_T'$ , which are the two orthogonal projectors for taking time deviation and cross sectional deviation from their means, the concentrated log likelihood is<sup>7</sup>

$$\begin{aligned} \ln L_L^c(\theta) = & -\frac{nT}{2} \ln 2\pi + \ln |I_L - W_L(\eta)| - \frac{nT}{2} \ln \sigma_\xi^2 - \frac{nT}{2} \ln |\Sigma_\epsilon| \\ & -\frac{1}{2} \text{vec}'(Z_L - \mathcal{K}_L \Phi_2)(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(Z_L - \mathcal{K}_L \Phi_2) \\ & -\frac{1}{2\sigma_\xi^2} [Y_L - W_L(\eta)Y_L - X_{1L}\beta_0 - (Z_L - \mathcal{K}_L \Phi_2)\delta - \mathcal{I}_0(\gamma, \rho)]' \\ & \cdot (J_T \otimes J_n)[Y_L - W_L(\eta)Y_L - X_{1L}\beta_0 - (Z_L - \mathcal{K}_L \Phi_2)\delta - \mathcal{I}_0(\gamma, \rho)]. \end{aligned} \quad (5)$$

For the consistency of the QMLE, we need to show that  $\frac{1}{L}[\ln L_L^c(\theta) - E(\ln L_L^c(\theta))] \xrightarrow{p} 0$  uniformly in  $\theta$ . From the reduced form, we have

$$\begin{aligned} \text{vec}'(Z_L - \mathcal{K}_L \Phi_2)(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(Z_L - \mathcal{K}_L \Phi_2) \\ = \text{vec}'[\mathcal{K}_L(\Phi_{20} - \Phi_2) + \epsilon_L](\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \\ \times \text{vec}[\mathcal{K}_L(\Phi_{20} - \Phi_2) + \epsilon_L], \end{aligned}$$

and

$$\begin{aligned} [Y_L - W_L(\eta)Y_L - X_{1L}\beta - (Z_L - \mathcal{K}_L \Phi_2)\delta - \mathcal{I}_0(\gamma, \rho)]'(J_T \otimes J_n) \\ \times [Y_L - W_L(\eta)Y_L - X_{1L}\beta - (Z_L - \mathcal{K}_L \Phi_2)\delta - \mathcal{I}_0(\gamma, \rho)] \\ = \{[W_L(\eta_0) - W_L(\eta)][I_L - W_L(\eta_0)]^{-1}(X_{1L}\beta_0 + \mathcal{I}_0(\gamma_0, \rho_0) \\ + \mathbf{c}_{1L0} + \alpha_{1L0} + \epsilon_L\delta_0 + \xi_L) + X_{1L}(\beta_0 - \beta) \\ + \mathcal{I}_0(\gamma_0, \rho_0) - \mathcal{I}_0(\gamma, \rho) + \epsilon_L(\delta_0 - \delta) + \xi_L \\ - \mathcal{K}_L(\Phi_{20} - \Phi_2)\delta\}' \cdot (J_T \otimes J_n) \\ \cdot \{[W_L(\eta_0) - W_L(\eta)][I_L - W_L(\eta_0)]^{-1}(X_{1L}\beta_0 + \mathcal{I}_0(\gamma_0, \rho_0) \\ + \mathbf{c}_{1L0} + \alpha_{1L0} + \epsilon_L\delta_0 + \xi_L) + X_{1L}(\beta_0 - \beta) \end{aligned}$$

<sup>7</sup> As  $T$  is large, we can ignore the influence of the initial condition. When  $T$  is fixed, we need to specify the initial condition if MLE is used; and we may consider the GMM estimation where lagged dependent variables can be used as IVs. Relevant discussions are provided in the supplement file available upon request (see [Appendix E](#)).

$$+ \mathcal{I}_0(\gamma_0, \rho_0) - \mathcal{I}_0(\gamma, \rho) + \epsilon_L(\delta_0 - \delta) + \xi_L - \mathcal{K}_L(\Phi_{20} - \Phi_2)\delta\}.$$

Furthermore,

$$\ln |I_L - W_L(\eta)| = \ln |I_L - \lambda W_{1L}| = -\sum_{l=1}^L \left[ \sum_{k=1}^{\infty} \frac{\lambda^k}{k} (W_{1L}^k)_{ll} \right],$$

where the first equality holds because the determinant of the block triangular matrix is the product of determinants of diagonal blocks<sup>8</sup> and the second equality is from [Qu and Lee \(2015\)](#). Also, from the first order conditions of (5) the score vector is

$$\begin{aligned} \frac{\partial \ln L_L^c(\theta_0)}{\partial \theta} = & \begin{pmatrix} \frac{\partial \ln L_L^c(\theta_0)}{\partial \lambda} \\ \frac{\partial \ln L_L^c(\theta_0)}{\partial \phi_1} \\ \frac{\partial \ln L_L^c(\theta_0)}{\partial \delta} \\ \frac{\partial \ln L_L^c(\theta_0)}{\partial \phi_2} \\ \frac{\partial \ln L_L^c(\theta_0)}{\partial \sigma_\xi^2} \\ \frac{\partial \ln L_L^c(\theta_0)}{\partial \alpha} \end{pmatrix} \\ = & \begin{pmatrix} \frac{1}{\sigma_{\xi 0}^2} (W_{1L} Y_L)' (J_T \otimes J_n) \xi_L - \text{tr}(G_{1L}) \\ \frac{1}{\sigma_{\xi 0}^2} R_L' (J_T \otimes J_n) \xi_L \\ \frac{1}{\sigma_{\xi 0}^2} \epsilon_L' (J_T \otimes J_n) \xi_L \\ (\Sigma_{\epsilon 0}^{-1} \otimes [\mathcal{K}_L' (J_T \otimes J_n)]) \text{vec}(\epsilon_L) - \frac{1}{\sigma_{\xi 0}^2} \delta_0 \otimes \mathcal{K}_L' (J_T \otimes J_n) \xi_L \\ -\frac{nT}{2\sigma_{\xi 0}^2} + \frac{1}{2\sigma_{\xi 0}^4} \xi_L' (J_T \otimes J_n) \xi_L \\ -\frac{nT}{2} \frac{\partial \ln |\Sigma_{\epsilon 0}|}{\partial \alpha} - \frac{1}{2} \frac{\partial}{\partial \alpha} (\text{tr}[\Sigma_{\epsilon 0}^{-1} \epsilon_L' (J_T \otimes J_n) \epsilon_L]) \end{pmatrix}. \end{aligned} \quad (6)$$

Hence, in general, we need to show<sup>9</sup>

$$\begin{aligned} \frac{1}{L} \zeta_{1L}' (J_T \otimes J_n) W_{j_1 L} W_{L'}^m \zeta_{2L}; \quad \frac{1}{L} (G_{j_1 L} \zeta_{1L})' (J_T \otimes J_n) \zeta_{2L}; \quad (7) \\ \frac{1}{L} (G_{j_1 L} \zeta_{1L})' (J_T \otimes J_n) G_{j_2 L} \zeta_{2L}; \quad \frac{1}{L} \sum_{l=1}^L \left[ \sum_{k=1}^{\infty} \frac{\lambda^k}{k} (W_{1L}^k)_{ll} \right] \end{aligned}$$

converge to their expectations, where  $m$  is a finite integer,  $j_1$  and  $j_2$  can be either 1, 2, 3, corresponding to  $W_{1L}$ ,  $W_{2L}$ ,  $W_{3L}$ ;  $\zeta_{1L}$  and  $\zeta_{2L}$  can be either some nonstochastic regressor vectors,  $\epsilon_L$ ,  $\xi_L$ , or  $Z_{L-1}$ . Also, we need to show the asymptotic distribution of  $A_L' \epsilon_L + B_L' \xi_L + [\xi_L' C_L \xi_L - E(\xi_L' C_L \xi_L | \epsilon_L)] + [\epsilon_L' D_L \epsilon_L - E(\epsilon_L' D_L \epsilon_L)] + \xi_L' E_L \epsilon_L$ , where the components satisfy Assumption CLT in a subsequent section.

<sup>8</sup> As a result, the computation burden is still restricted to  $n \times n$  dimensional matrices but not to  $nT \times nT$  dimensional matrices. This numerical feature is important especially when  $T$  is large. This comment is also applicable to the computation of matrix inverses related to spatial matrices. This becomes apparent in the reduced form of (1) where the needed inverses are simply the  $n \times n$  matrices  $S_{nt}^{-1}$  for all  $t = 1, \dots, T$ .

<sup>9</sup> As the  $Y_L$  and  $Y_{L-1}$  terms can be solved in terms of some nonstochastic regressor vectors,  $\epsilon_L$ ,  $\xi_L$ , and  $Z_{L-1}$ , the convergence of (7) can imply the convergence of (6).

### 3.2. Assumptions and topological structure

To derive the asymptotic properties of the QMLE, we need some LLN and CLT to incorporate the endogenous  $W_{nt}$  under the spatial panel data setting. In [Qu and Lee \(2015\)](#) for the cross sectional case, the distance between individuals  $i$  and  $j$  in  $R^d$  is  $\rho_{ij} = \|l(i, 0) - l(j, 0)\|_\infty$ . With the space–time process, the distance is  $\|l(i, t) - l(j, \tau)\|_\infty$  in  $R^{d+1}$  (but still keep the notation of  $\rho_{ij}$  in  $R^d$ ). We make the following assumptions similar to [Qu and Lee \(2015\)](#).

**Assumption 2.** The lattice  $D \subset R^{d+1}$ ,  $d \geq 1$ , is infinitely countable. The location  $l : \{1, \dots, L\} \rightarrow D_L \subset D$  is a mapping of individual  $i$  at time  $t$  to its location  $l(i, t) \in D \subset R^{d+1}$ . All elements in  $D$  are located at distances of at least  $s_0 > 0$ ; w.l.o.g. we assume that  $s_0 = 1$ .

**Assumption 3.** (i) The spatial weight  $w_{ij,t} \geq 0$ , and  $w_{ii,t} = 0$  for all  $i, j$ , and  $t$ ; (ii) For those nonzero weights between  $i \neq j$ ,  $w_{ij,t} = h(z_{it}, z_{jt}) \cdot I(\rho_{ij} \leq \rho_c)$  or the row-normalized version  $w_{ij,t} = h(z_{it}, z_{jt}) \cdot I(\rho_{ij} \leq \rho_c) / \sum_{\rho_{ik} \leq \rho_c} h_{ik}(z_{it}, z_{kt})$ , where  $h_{ij}(\cdot)$ 's are non-negative, uniformly bounded functions. (iii) For two different periods  $t$  and  $t'$ , the Lipschitz condition holds such that  $|h(z_{it'}, z_{jt'}) - h(z_{it}, z_{jt})| \leq c_0(|z_{it'} - z_{it}| + |z_{jt'} - z_{jt}|)$ .

**Assumption 4.**  $\sup_{n,t} \|W_{nt}\|_\infty \leq c_w$ ,  $\|\gamma_0\|_1 < 1$ , and  $|\lambda_0|c_w + |\rho_0| + |\gamma_0|c_w < 1$ .

**Assumption 5.** Let  $X_{nt}$  denote the collection of distinct columns in  $X_{1nt}$  and  $X_{2nt}$ . Elements of  $X_{nt}$  are nonstochastic and bounded, and  $\lim_{n,T \rightarrow \infty} \frac{1}{nT} \sum_{t=1}^T \tilde{X}_{nt}' J_n \tilde{X}_{nt}$  exists and is nonsingular, where  $\tilde{X}_{nt} = X_{nt} - \frac{1}{T} \sum_{t=1}^T X_{nt}$  and  $J_n = I_n - \frac{1}{n} I_n I_n'$ . Furthermore,  $\lim_{n,T \rightarrow \infty} \frac{1}{nT} \sum_{t=1}^T E \left( \tilde{Z}_{n,t-1}^{(-1)'} J_n \tilde{Z}_{n,t-1}^{(-1)} \right)$  exists and is nonsingular, where  $\tilde{Z}_{n,t-1}^{(-1)} = Z_{n,t-1} - \frac{1}{T} \sum_{t=1}^T Z_{n,t-1}$ .

**Assumption 6.**  $n$  is an increasing function of  $T$ , and  $T$  goes to infinity.

**Assumption 7.** The vector of parameters  $\theta = (\lambda, \gamma, \rho, \beta', \delta, \text{vec}'(\Gamma, \Gamma), \alpha, \sigma_\xi^2)'$  is in a compact set  $\Theta$  and the true parameter vector  $\theta_0$  is in the interior of  $\Theta$ .

[Assumption 2](#) specifies that the asymptotic inference under the space–time NED is based on increasing domain instead of infill domain. Therefore, given each period  $t$ , all spatial units may have their physical distances at least  $s_0 = 1$  apart even if there were a countable number of units. [Assumption 3](#) provides essential features for the time varying spatial weights matrix. From [Assumption 3\(i\)](#), for each unit, their physical positions in each period remain fixed, although their linkages  $w_{ij,t}$  can become weak or strong with time; from [Assumption 3\(ii\)](#), spatial weights (possibly row-normalized) may depend on economic factors, but once the geographic distance exceeds a threshold, two units would be not be spatially connected<sup>10</sup>; [Assumption 3\(iii\)](#) is a smoothing condition on  $h$  so that the time NED property for  $w_{ij,t}$  can be derived from  $z_{it}$  and  $z_{jt}$ . The value of link  $w_{ij,t}$  changes little to  $w_{ij,t'}$  if  $z_{it}$  and  $z_{jt}$  change little, respectively, to  $z_{it'}$  and  $z_{jt'}$ .

Note that while  $w_{ij,n}$  can be a function of endogenous variables  $z_{in}$  and  $z_{jn}$  (one popular example is the economic distance specification as  $1/|z_i - z_j|$ ), it may also include all those exogenous variables such as  $d_{ij}$ . Take the gravity type of specification  $w_{ij} =$

$\gamma_0 P_i^{\gamma_1} P_j^{\gamma_2} / d_{ij}^{\gamma_3}$  as an example. If all  $P_i$ ,  $P_j$ , and  $d_{ij}$  are exogenous, our model can handle this in the current setting, but with the estimation of coefficient  $\delta$  of the control variables as if  $\delta$  were unknown. So it is a model for estimation without imposing the “true” constraint of  $\delta = 0$ . If the constraint  $\delta = 0$  is imposed, the estimation method would correspond to conventional estimation methods of the SDPD model. If  $P_i$  and  $P_j$  are indeed endogenous and  $d_{ij}$  is the pre-determined physical distance, then this is exactly the second case in [Qu and Lee \(2015\)](#) cross sectional paper with  $P$  being our endogenous variable  $Z$ . In the current panel data setting where the  $P_{it}$ ,  $P_{jt}$  and  $d_{ij,t}$  would in general vary with  $t$ , our  $w_{ij,t}$  will allow it to be a function of  $d_{ij,t}$ 's in addition to  $P_{it}$  and  $P_{jt}$ . However, from [Assumption 3\(ii\)](#), we have imposed sparsity in our spatial weights matrices and the spatial weights do not involve unknown parameters. The sparsity assumption is convenient for asymptotic analysis but it would not have destructive limitation in model specifications for empirical studies. Corresponding to the specification  $w_{ij,t} = P_{it}^{\gamma_1} P_{jt}^{\gamma_2} / d_{ij,t}^{\gamma_3}$ , our sparsity assumption will be satisfied by setting  $w_{ij,t} = 0$  if  $d_{ij,t}$  is large enough over a critical value. As long as the critical value is really large enough, this is an adequate approximation because when  $d_{ij,t}$  is really large, the specified  $w_{ij,t}$  without truncation would have rather small values and hence tiny weights.

[Assumption 4](#) imposes restrictions on the stability of the dynamic process. Equivalently, it controls the degree of the magnitude of spatial interaction for the  $W_L(\eta_0)$  matrix where both contemporary and lagged space–time effects are considered. To obtain the NED property for  $y_{it}$ , we need  $|\lambda_0|c_w + |\rho_0| + |\gamma_0|c_w < 1$ . [Assumption 5](#) assumes that the exogenous variables in deviation form from cross section and time means are linearly independent so that there is no issue of multicollinearity. Of course, implicitly all the exogenous variables cannot be either time or cross-sectional invariant variables. If any of them is invariant in any dimension, the corresponding individual and/or time effects would absorb them. Except for the invariant cases, we note that we allow variables in  $X_{1nt}$  and  $X_{2nt}$  to overlap so that there could be common variables. Theoretically and empirically,  $X_{1nt}$  and  $X_{2nt}$  could even be the same as identification is not an issue here. This is so, because the parameters in the equations  $Z_{nt}$  can be identified and consistently estimated solely from (2), and  $X_{1nt}$  and the control variable is not linearly dependent, even though  $X_{1nt}$  and  $X_{2nt}$  might be so. [Assumption 6](#) specifies that the asymptotic analysis is based on large  $n$  and large  $T$  case. We note that if  $T$  were fixed and finite, due to the incidental parameter problem, QMLE would be inconsistent and the generalized method of moments (GMM) estimation with individual effects eliminated would be preferred. However, under this situation, the time–space specification for asymptotic analysis would be similar to the cross-sectional case (see [Qu and Lee, 2015](#)) once the initial values of  $y$ 's are properly dealt with, and the GMM estimation in [Lee and Yu \(2014\)](#) can be applied without much theoretical difficulty.<sup>11</sup> [Assumption 7](#) is a standard assumption in spatial econometrics, so that the parameter space is compact and the true parameter vector is in the interior of the parameter space.

To show the uniform convergence of the log likelihood function (5), the LLN of NED process in [Jenish and Prucha \(2012\)](#) can be used for subsequent analyses. To extend the LLN in [Jenish and Prucha \(2012\)](#) to the SDPD setting, we denote  $\varsigma_l$  as a function of  $\epsilon_l, \xi_l, z_l, X_L, \mathbf{c}_{n10}$  and  $\alpha_{T10} = (\alpha_{1,10}, \dots, \alpha_{T,10})'$  so that

$$\varsigma_l = f(\epsilon_l, \xi_l, z_{l-n}, X_L, \mathbf{c}_{n10}, \alpha_{T10}).$$

For the location  $l = l(i, t)$ , it corresponds to a unit  $(i, t)$  among the total  $L = nT$  units in panel as  $l = n(t-1) + i$ . The above relation

<sup>10</sup> This condition is a technical one. Asymptotic analysis becomes much more complex without it. In empirical studies, such a spatial setting is quite popular.

<sup>11</sup> For the GMM estimation under finite  $T$  case, please see the supplement file available upon request from the authors (see [Appendix E](#)).

means that  $\varsigma_{it}$  is a function of  $\epsilon_{it}$ ,  $\xi_{it}$ ,  $z_{i,t-1}$ ,  $X_{nT}$ ,  $\mathbf{c}_{n10}$  and  $\alpha_{T10}$ . This is because  $z_{it}$  is an AR(1) process for each  $i$ , so  $z_{it}$  is a function of  $z_{i,t-1}$  and  $\epsilon_{it}$ . For the spatial units, their physical distances remain fixed over time so that they are least  $s_0 = 1$  apart. Also, across time but at the same spatial location, we set their distance as one if they are one period apart. Thus, for  $i \neq j$ , as  $\rho_{ij} = \|l(i, t) - l(j, t-1)\|_\infty = \|l(i, t) - l(j, t)\|_\infty$ , we see that the distance between  $(i, t)$  and  $(j, t-1)$  is the same as  $(i, t)$  and  $(j, t)$ . For  $(i, t)$  and  $(i, t-1)$ , their distance is one, but  $(i, t)$  and  $(i, t)$  have zero distance. Therefore,  $\mathcal{F}_{l(i,t),L}(s) = \sigma(\varsigma_{l(j,\tau)}; l(j, \tau) \in D_L : \|l(i, t) - l(j, \tau)\|_\infty \leq s)$ .

Using this space specification, we now can investigate the terms in (7). The derivation of the following propositions can be found in Appendix B.

**Proposition 1 (LLN).** Under Assumptions 1–6, for any finite integer  $m$ ,

$$\begin{aligned} & \frac{1}{L} \varsigma'_{1L} (J_T \otimes J_n) W_{j_1L} W_L^m \varsigma_{2L} - E \left[ \frac{1}{L} \varsigma'_{1L} (J_T \otimes J_n) W_{j_1L} W_L^m \varsigma_{2L} \right] = o_p(1); \\ & \frac{1}{L} (G_{j_1L} \varsigma_{1L})' (J_T \otimes J_n) \varsigma_{2L} - E \left[ \frac{1}{L} (G_{j_1L} \varsigma_{1L})' (J_T \otimes J_n) \varsigma_{2L} \right] = o_p(1); \\ & \frac{1}{L} (G_{j_1L} \varsigma_{1L})' (J_T \otimes J_n) G_{j_2L} \varsigma_{2L} - E \left[ \frac{1}{L} (G_{j_1L} \varsigma_{1L})' (J_T \otimes J_n) G_{j_2L} \varsigma_{2L} \right] \\ & = o_p(1), \end{aligned}$$

where  $j_1$  and  $j_2$  can be either 1, 2, 3, corresponding to  $W_{1L}$ ,  $W_{2L}$ ,  $W_{3L}$ ;  $\varsigma_{1L}$  and  $\varsigma_{2L}$  can be either some nonstochastic regressor vectors,  $\epsilon_L$ ,  $\xi_L$ , or  $Z_{L,-1}$

**Corollary 1 (ULLN).** Under Assumptions 1–7,

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \frac{1}{L} \varsigma'_{1L}(\theta) (J_T \otimes J_n) W_{j_1L} W_L^m \varsigma_{2L}(\theta) - E \left[ \frac{1}{L} \varsigma'_{1L}(\theta) (J_T \otimes J_n) W_{j_1L} W_L^m \varsigma_{2L}(\theta) \right] \right| = o_p(1); \\ & \sup_{\theta \in \Theta} \left| \frac{1}{L} (G_{j_1L} \varsigma_{1L}(\theta))' (J_T \otimes J_n) \varsigma_{2L}(\theta) - E \left[ \frac{1}{L} (G_{j_1L} \varsigma_{1L}(\theta))' (J_T \otimes J_n) \varsigma_{2L}(\theta) \right] \right| = o_p(1); \\ & \sup_{\theta \in \Theta} \left| \frac{1}{L} (G_{j_1L} \varsigma_{1L}(\theta))' (J_T \otimes J_n) G_{j_2L} \varsigma_{2L}(\theta) - E \left[ \frac{1}{L} (G_{j_1L} \varsigma_{1L}(\theta))' (J_T \otimes J_n) G_{j_2L} \varsigma_{2L}(\theta) \right] \right| = o_p(1). \end{aligned}$$

Also, we need to consider the CLT for the form of

$$\begin{aligned} U_L = & \frac{1}{\sqrt{L}} (A'_L \varsigma_L + B'_L \xi_L + \xi'_L C_L \xi_L + \varsigma'_L D_L \varsigma_L \\ & + \xi'_L E_L \varsigma_L - E(\xi'_L C_L \xi_L | \varsigma_L) - E(\varsigma'_L D_L \varsigma_L)), \end{aligned}$$

where  $\varsigma_L$  and  $\xi_L$  are random vectors with each entry  $\varsigma_L$  and  $\xi_L$  having zero means and, respective nonzero variances  $\Sigma_\varsigma$  and  $\Sigma_\xi$ ,  $(\varsigma_L, \xi_L)'$ s are independently distributed across  $L$ ,  $A_L$  and  $D_L$  are  $L$  dimensional non-stochastic vector and matrix, but elements in  $B_L$ ,  $C_L$ , and  $E_L$  may contain  $\varsigma_L$ . We have the conditional mean independence  $E(\xi_L | \varsigma_L) = 0$  and conditional homoscedasticity of  $E(|\xi_L|^r | \varsigma_L)$  for  $r = 2, 3, 4$ .<sup>12</sup>

**Assumption CLT.** (1) For some constant  $\delta_U > 0$ ,  $E(\xi_L^{4+\delta_U})$  and  $E(\varsigma_L^{4+\delta_U})$  exist. Also,  $E(\xi_L | \varsigma_L) = 0$  and  $E(|\xi_L|^r | \varsigma_L)$  are constant across  $l$  for  $r = 2, 3, 4$ .

(2) For the nonstochastic vector  $A_L$ , elements are uniformly bounded; the nonstochastic matrix  $D_L$  is bounded in both row and column sums.

(3) For stochastic vector  $B_L$ , matrices  $C_L$  and  $E_L$ , each can be either  $W_L$ ,  $(I - W_L)^{-1}$ , or  $W_{jL}(I - W_L)^{-1}$ .

**Proposition 2.** Denote  $\sigma_{U_L}^2$  as the variance of  $U_L$ . Under Assumption CLT and assume that  $\sigma_{U_L}^2$  is bounded away from zero,<sup>13</sup> then  $U_L/\sigma_{U_L} \xrightarrow{d} N(0, 1)$ .

## 4. Asymptotic properties of QMLEs

### 4.1. Model identification and consistency

To prove consistency, we will show (i) the  $\frac{1}{L} \ln L_L^c(\theta) - \frac{1}{L} E \ln L_L^c(\theta)$  converges to zero in probability uniformly in  $\theta \in \Theta$ ; (2) the uniform equicontinuity of  $\frac{1}{L} E \ln L_L^c(\theta)$ , and (3) global identification. Using Proposition 1 and Corollary 1, the uniform convergence and uniform equicontinuity can be obtained. For identification under the likelihood theory, the most well-known one is the identification framework of Rothenberg (1971). For the log likelihood function  $\ln L_L^c(\theta)$ , the information inequality is  $E(\ln L_L^c(\theta)) \leq E(\ln L_L^c(\theta_0))$  for any  $\theta$  in its parameter space. The identification of the true parameter vector  $\theta_0$  can then be based on the strict information inequality that  $E(\ln L_L^c(\theta)) < E(\ln L_L^c(\theta_0))$  for all  $\theta \neq \theta_0$ , so that  $\theta_0$  is the unique maximizer of  $E(\ln L_L^c(\theta))$ . The information inequality is valid also for concentrated expected log likelihood, and can be generalized to accommodate some quasi-likelihood functions.

For a spatial panel model with homoscedastic disturbances, the parameter vector of explanatory variables, which are usually in a linear index form, can be easily concentrated out for estimation, and the identification of a model can then be focused on the identification of  $\lambda$  (see Lee and Yu (2016)).  $Q_{1L} = G_{1L}(X_{1L}\beta_0 + \mathcal{J}_0(\gamma_0, \rho_0) + \epsilon_L\delta_0 + \mathbf{c}_{1L0} + \alpha_{1L0})$ . We have the following assumption for parameter identification.

**Assumption 8.** Either (a)  $\lim_{L \rightarrow \infty} \frac{1}{L} E[(Q_{1L}, R_L, \epsilon_L)'(J_T \otimes J_n)(Q_{1L}, R_L, \epsilon_L)]$  exists and is nonsingular; or (b)  $S'_L(\eta)S_L(\eta)$  is not proportional to  $S'_L S_L$  with probability one whenever  $\eta \neq \theta_0$ .

Assumption 8(a) is a rank condition, essentially requiring that the regressors in the estimation equations are linearly independent. If Assumption 8(a) fails, we can still obtain the identification via the stochastic part of the process, which is Assumption 8(b). A sufficient condition is that  $I_L$ ,  $(W_L(\eta_0) + W'_L(\eta_0))$ , and  $W'_L(\eta_0)W_L(\eta_0)$  are linearly independent. With this identification, the uniform convergence of  $\frac{1}{L} [\ln L_L^c(\theta) - E(\ln L_L^c(\theta))]$  and the equicontinuity of  $\frac{1}{L} E(\ln L_L^c(\theta_0))$  imply the consistency of QML estimates.

**Theorem 1.** Under Assumptions 1–8,  $\theta_0$  is globally identified and the QMLE  $\hat{\theta}_{nT}$  that maximizes (5) is consistent such that  $\hat{\theta}_{nT} \xrightarrow{p} \theta_0$ .

<sup>12</sup> Due to the projectors which eliminate individual and time effects, CLT of an NED process cannot be easily used due to time averaging and/or cross sectional averaging over spatial units.

<sup>13</sup> For example, if  $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{i=1}^L \sum_{j \neq i}^L (d_{ij}^2 + d_{ji}^2) > 0$  or  $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{i=1}^L \sum_{j \neq i}^L E(c_{ij}^2 + c_{ji}^2) > 0$ , then  $\sigma_{U_L}^2$  will be bounded away from zero.

#### 4.2. Asymptotic normality

For items in (6), we have  $\frac{\partial \ln L_{1,L}^C(\theta_0)}{\partial \theta} = \frac{\partial \ln L_{1,L}^C(\theta_0)}{\partial \theta} + \Delta_L$  where

$$\frac{\partial \ln L_{1,L}^C(\theta_0)}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma_{\xi_0}^2} (W_{1L} Y_L)' (J_T \otimes J_n) \xi_L - \text{tr}(G_{1L} (J_T \otimes J_n)) \\ \frac{1}{\sigma_{\xi_0}^2} R_L' (J_T \otimes J_n) \xi_L - [\text{tr}(G_{2L} (J_T \otimes J_n)), \text{tr}(G_{3L} (J_T \otimes J_n)), 0]' \\ \frac{1}{\sigma_{\xi_0}^2} \epsilon_L' (J_T \otimes J_n) \xi_L \\ \left[ \begin{aligned} & \left( (\Sigma_{\epsilon_0}^{-1} \otimes \mathcal{K}_L' (J_T \otimes J_n)) \text{vec}(\epsilon_L) \right. \\ & \quad \left. + \text{vec} \left( \frac{(n-1)}{T} \left( \sum_{t=1}^{T-1} \sum_{h=1}^{T-t} \gamma^{(h-1)} \right) \Sigma_{\epsilon_0}^{-1} \right) \right) \\ & \quad - \frac{1}{\sigma_{\xi_0}^2} \delta_0 \otimes \mathcal{K}_L' (J_T \otimes J_n) \xi_L \\ & \quad - \frac{(n-1)(T-1)}{2\sigma_{\xi_0}^2} + \frac{1}{2\sigma_{\xi_0}^4} \xi_L' (J_T \otimes J_n) \xi_L \\ & \quad - \frac{(n-1)(T-1)}{2} \frac{\partial \ln |\Sigma_{\epsilon_0}|}{\partial \alpha} - \frac{1}{2} \frac{\partial}{\partial \alpha} (\text{tr}[\Sigma_{\epsilon_0}^{-1} \epsilon_L' (J_T \otimes J_n) \epsilon_L]) \end{aligned} \right] \end{pmatrix}, \quad (8)$$

and

$$\Delta_L = \begin{pmatrix} \text{tr}(W_{1L} (I_L - W_L)^{-1} (J_T \otimes J_n)) - \text{tr}(G_{1L}) \\ [\text{tr}(G_{2L} (J_T \otimes J_n)), \text{tr}(G_{3L} (J_T \otimes J_n)), 0]' \\ \mathbf{0}_{p \times 1} \\ -\text{vec} \left( \frac{(n-1)}{T} \left( \sum_{t=1}^{T-1} \sum_{h=1}^{T-t} \gamma^{(h-1)} \right) \Sigma_{\epsilon_0}^{-1} \right) \\ -\frac{n+T-1}{2\sigma_{\xi_0}^2} \\ -\frac{n+T-1}{2} \frac{\partial \ln |\Sigma_{\epsilon_0}|}{\partial \alpha} \end{pmatrix}. \quad (9)$$

The first component  $\frac{\partial \ln L_{1,L}^C(\theta_0)}{\partial \theta} = \frac{\partial \ln L_{1,L}^C(\theta_0)}{\partial \theta} - E\left(\frac{\partial \ln L_{1,L}^C(\theta_0)}{\partial \theta}\right)$  has zero mean and is asymptotically normally distributed, and  $\Delta_L = E\left(\frac{\partial \ln L_{1,L}^C(\theta_0)}{\partial \theta}\right)$  is the mean part which may cause asymptotic bias of the QMLE.

We notice that  $\frac{\partial \ln L_{1,L}^C(\theta_0)}{\partial \theta}$  has two components, one is the linear part and the other is the quadratic part. From the reduced form  $Y_L = (I_L - W_L(\eta_0))^{-1} (X_{1L} \beta_0 + \mathbf{1}_0(\gamma_0, \rho_0) + \mathbf{c}_{1L0} + \alpha_{1L0} + \epsilon_L \delta_0 + \xi_L)$ , the  $(W_{jL} Y_L)' (J_T \otimes J_n) \xi_L$  can be decomposed into

$$(W_{jL} Y_L)' (J_T \otimes J_n) \xi_L = (G_{2L} (X_{1L} \beta_0 + \mathbf{1}_0(\gamma_0, \rho_0) + \mathbf{c}_{1L0} + \alpha_{1L0} + \epsilon_L \delta_0))' (J_T \otimes J_n) \xi_L + (G_{jL} \xi_L)' (J_T \otimes J_n) \xi_L.$$

For  $\frac{\partial \ln L_{1,L}^C(\theta_0)}{\partial \theta}$ , we can use the linear and quadratic form of CLT. Although we have endogenous spatial weights matrix, after we use the control function, the CLT in Proposition 2 can be applied with the following assumption.

**Assumption 9.** The  $E(|\xi_{it}|^3 | \epsilon_{it})$  and  $E(|\xi_{it}|^4 | \epsilon_{it})$  are constant across  $i$  and  $t$ .

Thus, under Assumption 9,

$$\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{1,L}^C(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \lim_{n,T \rightarrow \infty} (\mathcal{I}_{nT, \theta_0} + \Omega_{nT, \theta_0}))$$

where  $\mathcal{I}_{nT, \theta_0}$  is in (24) and  $\Omega_{nT, \theta_0}$  is in (25).<sup>14</sup>

For the bias part  $\Delta_L$ , we have

$$\Delta_L = (n-1)a_{1, \theta_0} + Ta_{2, \theta_0},$$

where

$$a_{1, \theta_0} = \begin{pmatrix} -\frac{1}{n-1} \text{tr} \left( G_{1L} \left( \frac{1}{T} I_T I_T' \otimes J_n \right) \right) \\ \frac{1}{n-1} \left[ \text{tr} \left( G_{2L} \left( \frac{1}{T} I_T I_T' \otimes J_n \right) \right), \text{tr} \left( G_{3L} \left( \frac{1}{T} I_T I_T' \otimes J_n \right) \right), 0 \right]' \\ \mathbf{0}_{p \times 1} \\ -\text{vec} \left( \frac{1}{T} \left( \sum_{t=1}^{T-1} \sum_{h=1}^{T-t} \gamma_0^{(h-1)} \right) \right) \\ -\frac{1}{2\sigma_{\xi_0}^2} \\ -\frac{1}{2} \frac{\partial \ln |\Sigma_{\epsilon_0}|}{\partial \alpha} \end{pmatrix} \quad (10)$$

and

$$a_{2, \theta_0} = \begin{pmatrix} -\frac{1}{T} \text{tr} \left( G_{1L} \left( I_T \otimes \frac{1}{n} I_n I_n' \right) \right) \\ \mathbf{0}_{(k_1+2) \times 1} \\ \mathbf{0}_{p \times 1} \\ \mathbf{0}_{(p^2 + pk_2) \times 1} \\ -\frac{1}{2\sigma_{\xi_0}^2} \\ -\frac{1}{2} \frac{\partial \ln |\Sigma_{\epsilon_0}|}{\partial \alpha} \end{pmatrix}. \quad (11)$$

Here, the bias  $a_{1, \theta_0}$  is due to individual effects and the bias  $a_{2, \theta_0}$  is due to time effects.

**Theorem 2.** Under Assumptions 1–9, the QMLE  $\hat{\theta}_{nT}$  that maximizes (5) has

$$\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) - \sqrt{\frac{n}{T}} b_{1, \theta_0, nT} - \sqrt{\frac{T}{n}} b_{2, \theta_0, nT} + O_p \left( \max \left( \sqrt{\frac{n}{T^3}}, \sqrt{\frac{T}{n^3}} \right) \right)$$

$$\xrightarrow{d} N(0, \lim_{T \rightarrow \infty} \mathcal{I}_{nT, \theta_0}^{-1} (\mathcal{I}_{nT, \theta_0} + \Omega_{nT, \theta_0}) \mathcal{I}_{nT, \theta_0}^{-1}),$$

where  $b_{1, \theta_0, nT} = \mathcal{I}_{nT, \theta_0}^{-1} a_{1, \theta_0}$  and  $b_{2, \theta_0, nT} = \mathcal{I}_{nT, \theta_0}^{-1} a_{2, \theta_0}$  are  $O(1)$ .

<sup>14</sup> We can rely on  $J_n \Delta \xi_{nt}$  to estimate the variance and fourth moments such that  $\hat{\sigma}_{\xi}^2 = \frac{1}{2(n-1)(T-1)} \sum_{t=2}^T \Delta \xi_{nt}' J_n \Delta \xi_{nt}$  and  $\hat{\mu}_4 = \frac{1}{2(n-2+\frac{1}{n})} \left[ \frac{1}{T-1} \sum_{t=2}^T (\Delta \xi_{nt}' J_n \Delta \xi_{nt})^2 - 2(2n^2 - 3n + 4 - \frac{3}{n}) \hat{\sigma}^4 \right]$ . For the third moment,

we can use  $\hat{\mu}_3 = \frac{\sum_{i=1}^n \sum_{t=1}^T [U_n \hat{\xi}_{nt} \mathbf{1}_i]^3}{(n-3+\frac{2}{n})(T-3+\frac{2}{T})}$  where  $\tilde{\xi}_{nt} = \xi_{nt} - \frac{1}{T} \sum_{t=1}^T \xi_{nt}$ . Details for the derivation are provided in the supplement file to this paper and available upon request (see Appendix E).



### 4.3. Bias correction

We see that  $\hat{\theta}_{nT}$  is consistent, with additive bias terms of  $O(\frac{1}{n})$  and  $O(\frac{1}{T})$ . From Theorem 2, we have three cases for the QMLE  $\hat{\theta}_{nT}$ . When  $\frac{n}{T} \rightarrow k < \infty$ ,  $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) + \sqrt{k}b_{1,\theta_0,nT} + \sqrt{1/k}b_{2,\theta_0,nT} \xrightarrow{d} N(0, \mathcal{I}_{\theta_0}^{-1}(\mathcal{I}_{\theta_0} + \Omega_{\theta_0})\mathcal{I}_{\theta_0}^{-1})$ ; the QML estimator  $\hat{\theta}_{nT}$  has the bias  $\frac{1}{T}b_{1,\theta_0,nT}$  and  $\frac{1}{n}b_{2,\theta_0,nT}$ , and a conventional confidence interval will not be properly centered. When  $\frac{n}{T} \rightarrow 0$ ,  $n(\hat{\theta}_{nT} - \theta_0) + b_{2,\theta_0,nT} \xrightarrow{p} 0$ . When  $\frac{n}{T} \rightarrow \infty$ ,  $T(\hat{\theta}_{nT} - \theta_0) + b_{1,\theta_0,nT} \xrightarrow{p} 0$ . In all these cases, we see that a bias term is of order  $O(\max(\frac{1}{n}, \frac{1}{T}))$ . With bias correction, the asymptotic distribution of the bias-corrected estimate

$$\hat{\theta}_{nT}^1 = \hat{\theta}_{nT} - \frac{\hat{B}_{1,nT}}{T} - \frac{\hat{B}_{2,nT}}{n}$$

can be properly centered, where  $\hat{\theta}_{nT}$  is the QMLE that maximizes (5) and  $\hat{B}_{1,nT} = [\mathcal{J}_{nT,\theta}^{-1} \cdot a_{1,\theta}]|_{\theta=\hat{\theta}_{nT}}$  with  $\mathcal{J}_{nT,\theta_0}^{-1}$  from (24) and  $\hat{B}_{2,nT} = [\mathcal{J}_{nT,\theta}^{-1} \cdot a_{2,\theta}]|_{\theta=\hat{\theta}_{nT}}$ . Similar to Yu et al. (2008), the bias correction can be valid<sup>15</sup> with the following assumption.

**Assumption 10.**  $\frac{\partial a_1(\theta)}{\partial \theta} < \infty$  and  $\frac{\partial a_2(\theta)}{\partial \theta} < \infty$  in the neighborhood of  $\theta_0$ .

**Theorem 3.** Under Assumptions 1–10, if  $\frac{n}{T^3} \rightarrow 0$  and  $\frac{T}{n^3} \rightarrow 0$ ,

$$\sqrt{nT}(\hat{\theta}_{nT}^1 - \theta_0) \xrightarrow{d} N(0, \lim_{T \rightarrow \infty} \mathcal{J}_{nT,\theta_0}^{-1}(\mathcal{I}_{nT,\theta_0} + \Omega_{nT,\theta_0})\mathcal{J}_{nT,\theta_0}^{-1}).$$

### 4.4. A transformation approach

Instead of a direct estimation of time effects, we can eliminate them before estimation, for both (1) and (2). Let  $(F_{n,n-1}, I_n/\sqrt{n})$  be the orthonormal matrix of eigenvectors of  $J_n = I_n - \frac{1}{n}I_n I_n'$  where  $F_{n,n-1}$  corresponds to the eigenvalues of ones and  $I_n/\sqrt{n}$  corresponds to the eigenvalue zero. Denote  $Y_{nt}^* = F_{n,n-1}'Y_{nt}$  and other variables similarly. With  $W_{nt}$  being row normalized, (1) and (2) can be transformed into

$$Y_{nt}^* = \lambda W_{nt}^* Y_{nt}^* + \rho W_{n,t-1}^* Y_{n,t-1}^* + \gamma Y_{n,t-1}^* + X_{1nt}^* \beta_0 + \mathbf{c}_{n1}^* + V_{nt}^*, \quad (12)$$

and

$$Z_{nt}^* = Z_{n,t-1}^* \gamma_0 + X_{2nt}^* \Gamma_0 + \mathbf{c}_{n20}^* + \epsilon_{nt}^*,$$

where  $W_{nt}^* = F_{n,n-1}'W_{nt}F_{n,n-1}$ ,  $X_{1nt}^* = F_{n,n-1}'X_{1nt}$ ,  $\mathbf{c}_{n1}^* = F_{n,n-1}'\mathbf{c}_{n1}$ ,  $V_{nt}^* = F_{n,n-1}'V_{nt}$  and  $V_{nt}^*$  is an  $(n-1)$  dimensional disturbance vector with zero mean and variance matrix  $\sigma_0^2 I_{n-1}$ . This equation is in the format of a typical SAR model in panel data, where the time effects are eliminated and the number of observations is  $T(n-1)$ . Also,  $|I_{n-1} - \lambda W_{nt}^*| = \frac{1}{1-\lambda} |I_n - \lambda W_{nt}|$ ,  $(I_{n-1} - \lambda W_{nt}^*)^{-1} = F_{n,n-1}'(I_n - \lambda W_{nt})^{-1}F_{n,n-1}$  and  $V_{nt}^{*'}(\theta)V_{nt}^*(\theta) = V_{nt}'(\theta)J_n V_{nt}(\theta)$ .

Using  $F_{n,n-1}'F_{n,n-1} = J_n$ , the log (partial) likelihood function is now

$$\ln L_{L^*}^c(\theta) = -\frac{(n-1)T}{2} \ln 2\pi + \ln |I_L - W_L^*(\eta)|$$

$$\begin{aligned} & -\frac{(n-1)T}{2} \ln \sigma_\xi^2 - \frac{(n-1)T}{2} \ln |\Sigma_\epsilon| \\ & -\frac{1}{2} \text{vec}'(Z_L - \mathcal{K}_L \Phi_2)(\Sigma_{\Sigma_\epsilon}^{-1} \otimes J_T \otimes J_n) \text{vec}(Z_L - \mathcal{K}_L \Phi_2) \\ & -\frac{1}{2\sigma_\xi^2} [Y_L - W_L(\eta)Y_L - X_{1L}\beta_0 - (Z_L - \mathcal{K}_L \Phi_2)\delta - \mathcal{I}_0(\gamma, \rho)]' \\ & \cdot (J_T \otimes J_n)[Y_L - W_L(\eta)Y_L - X_{1L}\beta_0 \\ & - (Z_L - \mathcal{K}_L \Phi_2)\delta - \mathcal{I}_0(\gamma, \rho)]. \end{aligned} \quad (13)$$

The difference between the direct approach (5) and the transformation approach (13) is that  $n$  is replaced by  $n-1$  and  $\ln |I_L - W_L(\eta)|$  is replaced by  $\ln |I_L - W_L^*(\eta)|$ . For the first order condition of (13), we see that the last two entries of the corresponding  $a_{2,\theta_0}$  in (11) will be zero due to the degrees of freedom adjustment. Also, we see that the first entry of the corresponding  $a_{2,\theta_0}$  in (11) will be zero because  $\text{tr}(W_{1L}(I_L - W_L(\eta))^{-1}(I_T \otimes \frac{1}{n}I_n I_n')) = \frac{T}{1-\lambda}$  under row-normalization of  $W_{nt}$ , which will be canceled by  $\frac{\partial}{\partial \lambda} T \ln(\frac{1}{1-\lambda}) = \frac{T}{1-\lambda}$  from the relation  $|I_{n-1} - \lambda W_{nt}^*| = \frac{1}{1-\lambda} |I_n - \lambda W_{nt}|$  and the block diagonality of  $W_L(\eta)$ . Therefore, the bias term  $a_{2,\theta_0}$  from the direct estimation of time effects will disappear if we use the transformation approach.

However, the transformation approach requires the row-normalization of  $W_{nt}$ , which makes our model less general. The simulation in Section 5 compares the data transformation approach with the direct approach in the current paper. It turns out that both estimators have satisfactory performances.

## 5. Monte Carlo and empirical application

### 5.1. Monte Carlo

We conduct Monte Carlo experiments to investigate the finite sample performance of QML estimates. The value of  $\delta_0 = \Sigma_{\epsilon 0}^{-1} \sigma_{v \in 0}$  in the DGP can be 0, 0.5 and 0.8, which correspond, respectively, to exogenous, mild endogenous, and strong endogenous  $W_{nt}$ . We estimate the parameters using the QML estimation taking into account endogenous  $W_{nt}$ , and compare them with the estimates using the exogenous  $W_{nt}$  specification in Lee and Yu (2012a,b). We expect to see that (i) for the DGP with endogenous  $W_{nt}$ , the QMLE taking into account endogenous  $W_{nt}$  yields consistent estimates, while the QMLE using exogenous  $W_{nt}$  yields inconsistent estimates; (ii) for the DGP with exogenous  $W_{nt}$ , both QMLEs are consistent but the QMLE using exogenous  $W_{nt}$  will be more efficient.

Data is generated according to (1) and (2) with each of  $X_{1nt}$ ,  $X_{2nt}$  and  $Z_{nt}$  consists of a univariate variable, where

$$\begin{aligned} \theta_0 &= (\lambda_0, \gamma_0, \rho_0, \beta_0, \delta_0, \gamma_0, \Gamma_0, \sigma_{\xi 0}^2, \alpha_0)' \\ &= (0.2, 0.2, 0.2, 1, \delta_0, 0.2, 1, 1 - \delta_0^2, 1), \end{aligned}$$

and  $\delta_0$  takes the values from 0 to 0.8 while other parameters are held constant. The  $X_{1nt}$ ,  $X_{2nt}$ ,  $\mathbf{c}_{n0}$ , and  $\alpha_{t0}$  are generated from independent standard normal distributions.  $V_{nt}$  and  $\epsilon_{nt}$  are generated from a bivariate normal distribution  $\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \delta_0 \\ \delta_0 & 1 \end{pmatrix}$ . Throughout simulation, the initial observations are generated with a similar pattern.<sup>16</sup>

The spatial weights matrix  $W_{nt}$  is generated by  $W_{nt} = W_n^d \circ W_{nt}^e$ , i.e.,  $w_{ij,nt} = w_{ij}^d w_{ij,nt}^e$ , where  $W_n^d$  is a time invariant queen

<sup>15</sup> We note that although the bias correction estimates use biased estimate, the main component of the bias will be eliminated as both  $n$  and  $T$  are large, and the bias corrected terms do not introduce additional asymptotic variance for the estimates.

<sup>16</sup> We generated the spatial panel data with  $20 + T$  periods where the starting value is from  $N(0, I_n)$ , and then take the last  $T$  periods as our sample. By doing so, the initial value in the estimation is close to the steady state.

**Table 1**  
Estimators under endogenous  $W_{nt}$  specification: direct approach.

	$T$	$n$	$\delta$		$\lambda$	$\gamma$	$\rho$	$\beta$	$\delta$	$\gamma$	$\Gamma$	$\sigma_{\xi}^2$	$\alpha$
Before bias correction													
(1)	10	100	0	Bias	−0.015	−0.062	0.002	−0.008	0.000	−0.062	−0.006	−0.116	−0.117
				SD	0.031	0.023	0.033	0.034	0.033	0.024	0.034	0.042	0.042
				RMSE	0.040	0.066	0.044	0.044	0.044	0.067	0.045	0.124	0.125
				CP	0.904	0.218	0.934	0.916	0.934	0.248	0.922	0.214	0.196
(2)	10	100	0.5	Bias	−0.011	−0.054	0.001	−0.008	0.001	−0.055	−0.006	−0.086	−0.116
				SD	0.028	0.020	0.029	0.030	0.029	0.021	0.029	0.032	0.042
				RMSE	0.036	0.058	0.038	0.039	0.038	0.059	0.039	0.092	0.124
				CP	0.914	0.210	0.932	0.906	0.928	0.222	0.922	0.222	0.208
(3)	10	100	0.8	Bias	−0.008	−0.034	−0.001	−0.005	0.001	−0.035	−0.005	−0.041	−0.113
				SD	0.019	0.015	0.020	0.021	0.020	0.015	0.021	0.015	0.041
				RMSE	0.026	0.037	0.026	0.027	0.026	0.039	0.027	0.044	0.121
				CP	0.926	0.336	0.954	0.924	0.930	0.316	0.922	0.230	0.238
(4)	50	100	0	Bias	−0.013	−0.012	0.001	0.001	0.000	−0.012	−0.002	−0.028	−0.033
				SD	0.032	0.010	0.014	0.015	0.015	0.013	0.017	0.024	0.037
				RMSE	0.022	0.017	0.019	0.019	0.02	0.018	0.020	0.039	0.041
				CP	0.846	0.754	0.954	0.948	0.952	0.732	0.920	0.650	0.652
(5)	50	100	0.5	Bias	−0.012	−0.010	0.000	0.001	0.001	−0.010	−0.002	−0.021	−0.033
				SD	0.032	0.009	0.013	0.012	0.013	0.011	0.016	0.019	0.041
				RMSE	0.020	0.015	0.017	0.017	0.017	0.016	0.018	0.029	0.042
				CP	0.866	0.776	0.932	0.952	0.928	0.724	0.922	0.680	0.622
(6)	50	100	0.8	Bias	−0.010	−0.006	−0.001	0.000	0.001	−0.007	−0.002	−0.009	−0.036
				SD	0.029	0.006	0.009	0.008	0.010	0.007	0.012	0.010	0.052
				RMSE	0.016	0.010	0.012	0.011	0.012	0.011	0.012	0.014	0.044
				CP	0.856	0.832	0.914	0.954	0.932	0.764	0.928	0.672	0.620
After bias correction													
(1)	10	100	0	Bias	−0.004	−0.011	−0.001	−0.001	0.000	−0.015	−0.001	−0.018	−0.020
				SD	0.032	0.024	0.034	0.034	0.033	0.024	0.034	0.046	0.046
				RMSE	0.039	0.033	0.046	0.045	0.045	0.035	0.046	0.064	0.063
				CP	0.946	0.912	0.950	0.938	0.950	0.898	0.950	0.912	0.900
(2)	10	100	0.5	Bias	−0.002	−0.010	−0.001	−0.001	0.000	−0.013	−0.001	−0.013	−0.019
				SD	0.028	0.020	0.029	0.030	0.029	0.021	0.030	0.035	0.047
				RMSE	0.036	0.029	0.039	0.039	0.039	0.031	0.040	0.048	0.063
				CP	0.942	0.924	0.948	0.934	0.946	0.910	0.948	0.918	0.912
(3)	10	100	0.8	Bias	−0.002	−0.006	−0.001	−0.001	0.001	−0.008	0.000	−0.005	−0.015
				SD	0.019	0.015	0.020	0.021	0.020	0.015	0.021	0.017	0.046
				RMSE	0.025	0.021	0.027	0.027	0.027	0.022	0.028	0.023	0.063
				CP	0.948	0.930	0.958	0.948	0.942	0.928	0.952	0.922	0.916
(4)	50	100	0	Bias	−0.005	−0.001	−0.002	0.001	0.000	−0.002	−0.001	0.001	−0.004
				SD	0.032	0.011	0.014	0.015	0.015	0.013	0.017	0.025	0.039
				RMSE	0.020	0.014	0.019	0.020	0.020	0.015	0.020	0.029	0.031
				CP	0.920	0.952	0.958	0.952	0.954	0.904	0.924	0.932	0.930
(5)	50	100	0.5	Bias	−0.005	0.000	−0.002	0.001	0.000	−0.002	−0.002	0.001	−0.004
				SD	0.032	0.009	0.013	0.012	0.013	0.011	0.016	0.020	0.042
				RMSE	0.018	0.012	0.017	0.017	0.017	0.013	0.018	0.022	0.032
				CP	0.926	0.956	0.932	0.956	0.938	0.896	0.926	0.918	0.930
(6)	50	100	0.8	Bias	−0.006	0.000	−0.002	0.000	0.001	−0.002	−0.001	0.001	−0.007
				SD	0.029	0.006	0.009	0.008	0.010	0.007	0.012	0.011	0.054
				RMSE	0.015	0.008	0.012	0.012	0.012	0.009	0.013	0.011	0.034
				CP	0.930	0.958	0.932	0.954	0.936	0.908	0.932	0.926	0.928

matrix based on geographic contiguity and  $W_{nt}^e$  is generated by  $w_{ij,nt}^e = 1/|z_{it} - z_{jt}|$  if  $i \neq j$  and  $w_{ij,nt}^e = 0$  otherwise. We use  $T = 10, 50$ , and  $n = 100$ . For each set of generated sample observations, we calculate the QMLE estimator  $\hat{\theta}_{nt}$  and evaluate the bias  $\hat{\theta}_{nt} - \theta_0$ . We do this 500 times to obtain the empirical bias  $\frac{1}{500} \sum_{i=1}^{500} (\hat{\theta}_{nt} - \theta_0)_i$ . Finite sample properties of these estimators are summarized in Tables 1–2, where Table 1 is the QMLE using the correct endogenous specification, and Table 2 is the QMLE using the possibly misspecified exogenous specification which ignores correlation of disturbances. For each case, we report the bias, empirical standard deviation (SD), empirical root mean square error (RMSE), and coverage probability (CP).

From Table 1, the estimates under a correctly specified endogenous  $W_{nt}$  model has small biases for all the estimates after bias correction. Before bias correction, the estimates of  $\gamma$ ,  $\gamma$ ,  $\sigma_{\xi}^2$ , and  $\alpha$  have large bias, which causes the CP to be inaccurate; these bias

become smaller when  $T$  becomes larger, but still remains. After bias correction, these biases are much reduced and the CPs are improved. Also, the performance of QMLEs are satisfactory under different degrees of endogeneity of  $W_{nt}$ . For estimates before or after bias corrections, when  $T$  becomes larger, the bias becomes smaller.

From Table 2, we see that QMLEs under the  $W_{nt}$  ignoring its correlation with disturbances of the SAR equation have significant biases, mainly for the spatial lag coefficient. These biases remain even with the bias correction procedure. Also, the magnitude of the bias will increase with the degree of endogeneity of  $W_{nt}$ . When  $T$  becomes larger, the bias still remains similar.

By comparing the estimates in tables where we have exogenous spatial weights matrix specification ( $\rho = 0$ ), the correctly restricted model has smaller SD and RMSE of estimates for  $\lambda$  than those of the estimation as if  $W_{nt}$  is endogenous while other

**Table 2**Estimators under exogenous  $W_{nt}$  specification: direct approach.

	$T$	$n$	$\delta$		$\lambda$	$\gamma$	$\rho$	$\beta$	$\sigma_v^2$
Before bias correction									
(1)	10	100	0	Bias	−0.014	−0.062	0.002	−0.008	−0.115
				SD	0.029	0.023	0.033	0.034	0.041
				RMSE	0.040	0.066	0.044	0.044	0.123
				CP	0.906	0.208	0.936	0.916	0.208
(2)	10	100	0.5	Bias	0.012	−0.063	0.002	−0.008	−0.118
				SD	0.030	0.022	0.033	0.034	0.041
				RMSE	0.040	0.067	0.044	0.044	0.126
				CP	0.888	0.180	0.938	0.916	0.186
(3)	10	100	0.8	Bias	0.049	−0.064	−0.001	−0.009	−0.126
				SD	0.031	0.022	0.032	0.034	0.041
				RMSE	0.060	0.069	0.043	0.044	0.133
				CP	0.530	0.174	0.950	0.900	0.142
(4)	50	100	0	Bias	−0.010	−0.012	0.000	0.001	−0.029
				SD	0.012	0.010	0.014	0.014	0.021
				RMSE	0.019	0.017	0.019	0.019	0.038
				CP	0.862	0.748	0.954	0.950	0.652
(5)	50	100	0.5	Bias	0.014	−0.013	−0.003	0.000	−0.032
				SD	0.014	0.010	0.014	0.014	0.021
				RMSE	0.022	0.018	0.019	0.019	0.040
				CP	0.738	0.702	0.958	0.954	0.600
(6)	50	100	0.8	Bias	0.052	−0.015	−0.008	−0.001	−0.041
				SD	0.018	0.010	0.014	0.014	0.021
				RMSE	0.053	0.019	0.020	0.019	0.047
				CP	0.066	0.668	0.914	0.952	0.454
After bias correction									
(1)	10	100	0	Bias	−0.003	−0.010	−0.001	−0.001	−0.019
				SD	0.029	0.024	0.034	0.034	0.046
				RMSE	0.039	0.033	0.046	0.045	0.063
				CP	0.950	0.910	0.950	0.936	0.910
(2)	10	100	0.5	Bias	0.023	−0.011	−0.003	−0.001	−0.022
				SD	0.031	0.023	0.034	0.034	0.046
				RMSE	0.045	0.033	0.045	0.045	0.064
				CP	0.840	0.910	0.942	0.942	0.894
(3)	10	100	0.8	Bias	0.061	−0.013	−0.006	−0.002	−0.031
				SD	0.031	0.024	0.033	0.034	0.045
				RMSE	0.070	0.034	0.045	0.045	0.066
				CP	0.396	0.902	0.954	0.942	0.852
(4)	50	100	0	Bias	−0.001	−0.001	−0.002	0.001	−0.001
				SD	0.012	0.011	0.014	0.014	0.022
				RMSE	0.017	0.014	0.019	0.019	0.028
				CP	0.942	0.950	0.956	0.952	0.940
(5)	50	100	0.5	Bias	0.023	−0.002	−0.005	0.000	−0.004
				SD	0.014	0.010	0.014	0.014	0.021
				RMSE	0.028	0.014	0.020	0.019	0.028
				CP	0.510	0.948	0.940	0.956	0.934
(6)	50	100	0.8	Bias	0.061	−0.004	−0.010	−0.001	−0.013
				SD	0.018	0.011	0.014	0.014	0.022
				RMSE	0.062	0.014	0.021	0.019	0.030
				CP	0.020	0.920	0.890	0.952	0.866

estimates still have similar SD and RMSE. This indicates that the correct restricted model will yield more efficient estimates.<sup>17</sup>

We also investigate the performances of estimates under the transformation approach. Under the row-normalization of  $W_{nt}$ , theoretically, the transformation approach is free of the bias  $a_{2,\theta_0}$  due to time effects. From Tables 3–4, we see that estimates of  $\lambda$  have a slightly smaller bias than the direct approach, while other estimates have similar magnitude of bias as the direct approach. This may be due to the fact that  $n = 100$ , so that the magnitude of the  $O(\frac{1}{n})$  bias due to time effects is small and even smaller than the  $O(\frac{1}{T})$  bias due to individual effects.

## 5.2. Empirical application

We consider the dynamic demand for cigarette consumption as our empirical illustration. Baltagi and Levin (1986, 1992) find that there is significant price elasticity but the income elasticity is small. The “bootlegging” effect is found to be significant so that the minimum price of neighboring states influences the cigarette sale in a state. However, this bootlegging specification ignores the possibility that cross border shopping can take place in different neighboring states, but not just the minimum price of neighboring states. For these different neighboring states, only those with a lower price might affect the demand of the home state. Similar to Kelejian and Piras (2014), we propose the following equation for estimation

$$\ln C_{nt} = \lambda W_{nt} \ln C_{nt} + \gamma \ln C_{n,t-1} + \rho W_{n,t-1} \ln C_{n,t-1} + \beta_1 \ln I_{nt} + \beta_2 \ln P_{nt} + \mathbf{c}_n + \alpha_t I_n + V_{nt}.$$

<sup>17</sup> For the Monte Carlo experiments, we also try different settings such as  $n = 49$ , two exogenous variables and two  $Z$  variables. The results are similar, and they are available in the supplement file (see Appendix E).

**Table 3**Estimators under endogenous  $W_{nt}$  specification: transformation approach.

	$T$	$n$	$\delta$		$\lambda$	$\gamma$	$\rho$	$\beta$	$\delta$	$\gamma$	$\Gamma$	$\sigma_{\xi}^2$	$\alpha$
Before bias correction													
(1)	10	100	0	Bias	−0.004	−0.062	0.004	−0.008	0.000	−0.062	−0.006	−0.108	−0.108
				SD	0.029	0.023	0.034	0.034	0.033	0.023	0.034	0.042	0.042
				RMSE	0.038	0.067	0.044	0.044	0.044	0.066	0.045	0.117	0.116
				CP	0.946	0.212	0.936	0.916	0.936	0.252	0.926	0.262	0.288
(2)	10	100	0.5	Bias	−0.003	−0.054	0.002	−0.008	0.000	−0.054	−0.006	−0.081	−0.108
				SD	0.031	0.020	0.029	0.030	0.029	0.021	0.030	0.032	0.044
				RMSE	0.036	0.058	0.038	0.039	0.038	0.059	0.039	0.088	0.116
				CP	0.926	0.206	0.934	0.908	0.936	0.236	0.920	0.288	0.264
(3)	10	100	0.8	Bias	−0.002	−0.034	0.001	−0.005	0.001	−0.035	−0.004	−0.038	−0.104
				SD	0.019	0.015	0.020	0.021	0.020	0.015	0.021	0.015	0.042
				RMSE	0.025	0.038	0.027	0.027	0.027	0.039	0.027	0.042	0.113
				CP	0.932	0.334	0.952	0.924	0.934	0.326	0.928	0.292	0.306
(4)	50	100	0	Bias	−0.001	−0.013	0.001	0.001	0.000	−0.012	−0.001	−0.021	−0.021
				SD	0.024	0.010	0.014	0.014	0.015	0.012	0.015	0.022	0.027
				RMSE	0.018	0.018	0.019	0.019	0.020	0.018	0.020	0.034	0.034
				CP	0.934	0.734	0.954	0.950	0.954	0.738	0.934	0.790	0.800
(5)	50	100	0.5	Bias	−0.003	−0.010	0.001	0.000	0.000	−0.011	−0.001	−0.015	−0.022
				SD	0.030	0.009	0.013	0.012	0.013	0.010	0.015	0.018	0.036
				RMSE	0.017	0.015	0.017	0.017	0.017	0.016	0.017	0.025	0.035
				CP	0.936	0.756	0.936	0.954	0.932	0.734	0.930	0.798	0.822
(6)	50	100	0.8	Bias	−0.004	−0.006	0.000	0.000	0.000	−0.007	−0.001	−0.007	−0.025
				SD	0.025	0.006	0.009	0.008	0.01	0.007	0.011	0.009	0.049
				RMSE	0.014	0.010	0.012	0.012	0.012	0.011	0.012	0.012	0.038
				CP	0.936	0.814	0.914	0.956	0.934	0.760	0.930	0.784	0.792
After bias correction													
(1)	10	100	0	Bias	−0.002	−0.011	−0.001	−0.001	0.000	−0.015	0.000	−0.018	−0.018
				SD	0.029	0.024	0.034	0.034	0.033	0.024	0.034	0.046	0.046
				RMSE	0.039	0.033	0.046	0.045	0.045	0.035	0.046	0.064	0.063
				CP	0.946	0.912	0.952	0.938	0.950	0.900	0.954	0.920	0.904
(2)	10	100	0.5	Bias	−0.001	−0.010	−0.001	−0.001	0.000	−0.013	−0.001	−0.013	−0.018
				SD	0.031	0.020	0.029	0.030	0.029	0.021	0.030	0.036	0.048
				RMSE	0.036	0.029	0.039	0.039	0.039	0.031	0.040	0.048	0.064
				CP	0.944	0.922	0.942	0.932	0.948	0.906	0.948	0.916	0.916
(3)	10	100	0.8	Bias	−0.001	−0.006	0.000	−0.001	0.000	−0.008	0.000	−0.005	−0.014
				SD	0.019	0.015	0.020	0.021	0.020	0.015	0.021	0.017	0.046
				RMSE	0.025	0.021	0.028	0.027	0.027	0.022	0.028	0.023	0.063
				CP	0.948	0.924	0.964	0.948	0.942	0.928	0.950	0.918	0.924
(4)	50	100	0	Bias	−0.001	−0.001	−0.001	0.001	0.000	−0.002	−0.001	−0.001	−0.001
				SD	0.024	0.011	0.014	0.014	0.015	0.012	0.015	0.023	0.027
				RMSE	0.018	0.014	0.019	0.019	0.020	0.014	0.020	0.029	0.029
				CP	0.932	0.946	0.960	0.952	0.954	0.910	0.936	0.938	0.940
(5)	50	100	0.5	Bias	−0.003	−0.001	0.000	0.001	0.000	−0.002	−0.001	0.000	−0.003
				SD	0.030	0.009	0.013	0.012	0.013	0.010	0.015	0.019	0.037
				RMSE	0.018	0.012	0.017	0.017	0.017	0.013	0.018	0.022	0.031
				CP	0.934	0.956	0.940	0.958	0.938	0.900	0.934	0.922	0.934
(6)	50	100	0.8	Bias	−0.004	−0.001	0.000	0.000	0.000	−0.002	−0.001	0.000	−0.005
				SD	0.025	0.006	0.009	0.008	0.010	0.007	0.011	0.010	0.050
				RMSE	0.014	0.008	0.012	0.012	0.012	0.009	0.012	0.011	0.033
				CP	0.938	0.958	0.914	0.956	0.940	0.904	0.938	0.928	0.938

Here,  $C_{nt}$  is the cigarette sales in packs per capita,  $I_{nt}$  is per capita disposable income, and  $P_{nt}$  is price per pack of cigarettes. The sample includes 46 states<sup>18</sup> and 30 years (1983–1992).

We allow the endogeneity of spatial weights matrices by the following equation

$$\ln P_{nt} = \gamma_P \ln P_{n,t-1} + \Gamma_1 \ln I_{nt} + \Gamma_2 \text{Pop16}_{nt} + \mathbf{c}_{n2} + \alpha_{t2} I_n + \epsilon_{nt},$$

where  $\text{Pop16}_{nt}$  is the ratio of population with age over 16, which is the legal age to smoke. For identification purposes, as long as the  $\ln P_{nt}$  includes different exogenous variables such as  $\ln P_{n,t-1}$  or  $\text{Pop16}_{nt}$ ,  $\beta_2$  will be identified.<sup>19</sup>

<sup>18</sup> The omitted states are Alaska, Colorado, Hawaii, North Carolina and Oregon.

<sup>19</sup> In general, if the  $P_{nt}$  variable is not included in the estimation equation of  $Y_{nt}$ , the exogenous variables  $X_{1nt}$  and  $X_{2nt}$  in both equations are allowed to be the same. Here  $P_{nt}$  is the  $Z_{nt}$  in our format. Furthermore  $P_{nt}$  itself is an endogenous regressor which may have a direct impact on sale.

We have two specifications of the spatial weights matrix. The first one specified as

$$w_{ij,t} = \frac{p_{jt}}{p_{it}} d_{ijt}, \quad (14)$$

where  $d_{ijt}$  is a dummy variable specifies the desirability of cross border shopping such that  $d_{ijt} = 1$  if  $i$  and  $j$  are neighbors and  $p_{it} < p_{jt}$ ,  $d_{ijt} = 0$  if  $i$  and  $j$  are not neighbors or  $p_{it} \geq p_{jt}$ . Here, a higher price in a neighboring state  $j$  would cause its residents to cross-border shopping so that the sale in state  $i$  would be higher. A higher value of  $\frac{p_{jt}}{p_{it}}$  indicates a stronger motivation for cross border shopping. The second one is

$$w_{ij,t} = \left( \frac{p_{jt}}{p_{it}} - 1 \right) d_{ijt}, \quad (15)$$

which avoids the abrupt change of  $w_{ij,t}$  at the point of  $\frac{p_{jt}}{p_{it}} = 1$ .



**Table 4**Estimators under exogenous  $W_{nt}$  specification: transformation approach.

	$T$	$n$	$\delta$		$\lambda$	$\gamma$	$\rho$	$\beta$	$\sigma_v^2$
Before bias correction									
(1)	10	100	0	Bias	−0.004	−0.062	0.004	−0.008	−0.108
				SD	0.029	0.023	0.034	0.034	0.042
				RMSE	0.038	0.067	0.044	0.044	0.116
				CP	0.944	0.210	0.936	0.914	0.260
(2)	10	100	0.5	Bias	0.022	−0.063	0.003	−0.008	−0.111
				SD	0.030	0.022	0.033	0.034	0.042
				RMSE	0.044	0.068	0.044	0.044	0.119
				CP	0.838	0.178	0.936	0.912	0.242
(3)	10	100	0.8	Bias	0.060	−0.065	0.001	−0.009	−0.119
				SD	0.031	0.022	0.032	0.034	0.041
				RMSE	0.068	0.069	0.044	0.044	0.126
				CP	0.396	0.174	0.948	0.900	0.180
(4)	50	100	0	Bias	0.000	−0.013	0.001	0.001	−0.021
				SD	0.012	0.010	0.014	0.014	0.021
				RMSE	0.017	0.018	0.019	0.019	0.033
				CP	0.942	0.738	0.954	0.952	0.792
(5)	50	100	0.5	Bias	0.024	−0.014	−0.002	0.000	−0.024
				SD	0.015	0.010	0.014	0.014	0.021
				RMSE	0.028	0.018	0.019	0.019	0.035
				CP	0.490	0.684	0.958	0.954	0.736
(6)	50	100	0.8	Bias	0.062	−0.015	−0.007	−0.002	−0.033
				SD	0.018	0.010	0.014	0.014	0.021
				RMSE	0.064	0.020	0.020	0.019	0.041
				CP	0.018	0.642	0.924	0.950	0.578
After bias correction									
(1)	10	100	0	Bias	−0.002	−0.011	−0.001	−0.001	−0.018
				SD	0.029	0.024	0.034	0.034	0.046
				RMSE	0.039	0.033	0.046	0.045	0.063
				CP	0.950	0.910	0.952	0.936	0.912
(2)	10	100	0.5	Bias	0.025	−0.012	−0.002	−0.001	−0.022
				SD	0.031	0.023	0.034	0.034	0.046
				RMSE	0.045	0.033	0.045	0.045	0.064
				CP	0.836	0.910	0.942	0.942	0.894
(3)	10	100	0.8	Bias	0.063	−0.014	−0.006	−0.002	−0.031
				SD	0.031	0.024	0.033	0.034	0.045
				RMSE	0.071	0.034	0.045	0.045	0.066
				CP	0.386	0.902	0.952	0.942	0.862
(4)	50	100	0	Bias	0.000	−0.001	−0.001	0.001	−0.001
				SD	0.012	0.011	0.014	0.014	0.022
				RMSE	0.017	0.014	0.019	0.019	0.028
				CP	0.940	0.948	0.958	0.952	0.938
(5)	50	100	0.5	Bias	0.024	−0.002	−0.004	0.000	−0.005
				SD	0.015	0.010	0.014	0.014	0.022
				RMSE	0.029	0.014	0.020	0.019	0.028
				CP	0.484	0.942	0.946	0.956	0.932
(6)	50	100	0.8	Bias	0.063	−0.004	−0.009	−0.001	−0.014
				SD	0.018	0.011	0.014	0.014	0.022
				RMSE	0.064	0.014	0.021	0.019	0.030
				CP	0.016	0.916	0.900	0.952	0.860

We first estimate the model by using the endogenous specification in Table 5. We use different specifications allowing  $\rho = 0$  and  $\rho \neq 0$  under direct or transformation approaches, using  $\frac{p_{ijt}}{p_{it}} d_{ijt}$  or  $\left(\frac{p_{ijt}}{p_{it}} - 1\right) d_{ijt}$  as weights. From Table 5, we can see that the income elasticity is significant, while price elasticity is insignificant. For all the model specifications, the spatial effect is significant, which implies the existence of a cross border effect. Also, from the estimate of  $\delta$ , we see that it is negative and significant, which implies the endogeneity of the spatial weights matrix. From the model specification, the endogeneity of the spatial weights matrix comes from the fact that the unobservable  $\epsilon_{nt}$  that affect the  $\ln P_{nt}$  equation are correlated with the unobservable  $V_{nt}$  in the  $\ln C_{nt}$  equation. This is understandable if we consider some common factors that will affect the supply of cigarettes. Comparing  $\frac{p_{ijt}}{p_{it}} d_{ijt}$  or  $\left(\frac{p_{ijt}}{p_{it}} - 1\right) d_{ijt}$ , they yield similar

estimates, while the specification of  $\left(\frac{p_{ijt}}{p_{it}} - 1\right) d_{ijt}$  has a slightly higher value of log likelihood. Table 6 is the result for exogenous specification of spatial weights matrices. Compared with the endogenous specification (which is more general), we see that there are small magnitude changes in estimates, i.e., spatial effect estimate is larger, dynamic effect is smaller, income elasticity is smaller, and price elasticity is larger.

## 6. Conclusion

This paper analyzes QMLE estimation of spatial dynamic panel data models, where spatial weights matrices are endogenous and time varying. We generalize the notion of spatial NED in Jenish and Prucha (2012) to incorporate near-epoch for time series to a panel setting. By using spatial-time LLN for near-epoch dependence process and CLT for the martingale difference sequence, we establish

**Table 5**

Cigarette sales equation: endogenous weights matrices.

	$w_{ij,t} = (p_{jt}/p_{it})d_{ij,t}$				$w_{ij,t} = (p_{jt}/p_{it} - 1)d_{ij,t}$			
	direct		transformation		direct		transformation	
	$\rho = 0$	$\rho \neq 0$	$\rho = 0$	$\rho \neq 0$	$\rho = 0$	$\rho \neq 0$	$\rho = 0$	$\rho \neq 0$
$W_{nt} \ln c_{nt}$	0.028 (0.013)	0.033 (0.015)	0.029 (0.013)	0.034 (0.015)	0.029 (0.012)	0.042 (0.015)	0.031 (0.013)	0.044 (0.015)
$\ln c_{n,t-1}$	0.906 (0.013)	0.907 (0.013)	0.906 (0.013)	0.907 (0.013)	0.906 (0.013)	0.909 (0.013)	0.905 (0.013)	0.909 (0.013)
$W_{n,t-1} \ln c_{n,t-1}$		−0.009 (0.015)		−0.010 (0.015)		−0.022 (0.015)		−0.023 (0.015)
$\ln income$	0.086 (0.025)	0.087 (0.025)	0.086 (0.025)	0.087 (0.025)	0.086 (0.024)	0.088 (0.025)	0.086 (0.024)	0.088 (0.025)
$\ln price$	−0.021 (0.023)	−0.020 (0.023)	−0.022 (0.023)	−0.021 (0.023)	−0.024 (0.023)	−0.021 (0.023)	−0.025 (0.023)	−0.022 (0.023)
$\delta$	−0.390 (0.027)	−0.390 (0.027)	−0.390 (0.027)	−0.390 (0.027)	−0.387 (0.027)	−0.390 (0.027)	−0.386 (0.027)	−0.390 (0.027)
$\ln price_{t-1} \ln Z$	0.712 (0.020)	0.712 (0.020)	0.712 (0.020)	0.712 (0.020)	0.712 (0.020)	0.712 (0.020)	0.712 (0.020)	0.712 (0.020)
$\ln income \ln Z$	−0.033 (0.023)	−0.033 (0.023)	−0.033 (0.024)	−0.033 (0.024)	−0.034 (0.023)	−0.033 (0.023)	−0.033 (0.024)	−0.033 (0.024)
$\ln pop16 \ln Z$	−0.046 (0.091)	−0.046 (0.091)	−0.046 (0.091)	−0.047 (0.091)	−0.044 (0.091)	−0.045 (0.091)	−0.045 (0.091)	−0.046 (0.091)
log likelihood ( $10^3$ )	6.4786	6.4786	6.3100	6.3101	6.4789	6.4792	6.3104	6.3110

**Table 6**

Cigarette sales equation: exogenous weights matrices.

	$w_{ij,t} = (p_{jt}/p_{it})d_{ij,t}$				$w_{ij,t} = (p_{jt}/p_{it} - 1)d_{ij,t}$			
	direct		transformation		direct		transformation	
	$\rho = 0$	$\rho \neq 0$	$\rho = 0$	$\rho \neq 0$	$\rho = 0$	$\rho \neq 0$	$\rho = 0$	$\rho \neq 0$
$W_{nt} \ln c_{nt}$	0.036 (0.013)	0.034 (0.015)	0.037 (0.013)	0.035 (0.016)	0.040 (0.013)	0.044 (0.015)	0.041 (0.013)	0.047 (0.016)
$\ln c_{n,t-1}$	0.850 (0.013)	0.850 (0.013)	0.850 (0.013)	0.850 (0.013)	0.850 (0.013)	0.851 (0.013)	0.849 (0.013)	0.850 (0.013)
$W_{n,t-1} \ln c_{n,t-1}$		0.003 (0.015)		0.003 (0.015)		−0.009 (0.015)		−0.010 (0.016)
$\ln income$	0.079 (0.024)	0.079 (0.024)	0.079 (0.024)	0.078 (0.024)	0.078 (0.024)	0.079 (0.024)	0.078 (0.024)	0.078 (0.024)
$\ln price$	−0.267 (0.023)	−0.267 (0.023)	−0.267 (0.023)	−0.268 (0.023)	−0.269 (0.023)	−0.269 (0.023)	−0.269 (0.023)	−0.269 (0.023)
log likelihood ( $10^3$ )	2.6215	2.6218	2.5514	2.5515	2.6226	2.6227	2.5527	2.5527

the consistency and asymptotic normality of QMLE. Bias corrected estimates are also derived. Monte Carlo experiments show that the proposed QMLE have satisfactory finite sample performance.

## Appendix A. Notations

The following list summarizes some frequently used notations in the paper:

$\Sigma_{v \in 0} = \begin{pmatrix} \sigma_{v0}^2 & \sigma'_{v \in 0} \\ \sigma_{v \in 0} & \Sigma_{\epsilon 0} \end{pmatrix}$  is the variance matrix of  $v_{it}$  and  $\epsilon_{it}$ .  
 $\xi_{nt} = V_{nt} - \epsilon_{nt}\delta_0$  where  $\delta_0 = \Sigma_{\epsilon 0}^{-1}\sigma_{v \in 0}$  and  $\sigma_{\xi 0}^2 = \sigma_{v0}^2 - \sigma'_{v \in 0}\Sigma_{\epsilon 0}^{-1}\sigma_{v \in 0}$ .

$\theta = (\lambda, \phi_1', \delta, \phi_2', \alpha, \sigma_{\xi}^2)'$  where  $\phi_1 = (\gamma, \rho, \beta')'$ ,  $\phi_2 = \text{vec}(\Phi_2)$  with  $\Phi_2 = (\gamma', I')'$  and  $\alpha$  is the  $J \times 1$  column vector of distinct elements in  $\Sigma_{\epsilon}$ .

$R_{nt} = [Y_{n,t-1}, W_{n,t-1}Y_{n,t-1}, X_{1nt}]$ ,  $\mathcal{K}_{nt} = [Z_{n,t-1}, X_{2nt}]$  which correspond to  $\phi_1$  and  $\phi_2$ .

$Y_L, \epsilon_L, X_{1L}, Z_L, X_{2L}, R_L$  and  $\mathcal{K}_L$  are vectorized form of corresponding variables with row dimension  $L = nT$ .

$$W_L(\eta_0) = \begin{pmatrix} \lambda_0 W_{n1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \gamma_0 I_n + \rho_0 W_{n1} & \lambda_0 W_{n2} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \gamma_0 I_n + \rho_0 W_{n,T-1} & \lambda_0 W_{nT} \end{pmatrix}$$

where  $\eta = (\lambda, \gamma, \rho)$  and  $\eta_0$  is its true value.

$$\mathcal{I}_0(\gamma_0, \rho_0) = \begin{pmatrix} \gamma Y_{n0} + \rho W_{n0} Y_{n0} \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ is related to the initial condi-}$$

tions for the DGP.

$$W_{1L} = \begin{pmatrix} W_{n1} & & & \\ & W_{n2} & & \\ & & \ddots & \\ & & & W_{nT} \end{pmatrix}, W_{2L} = \begin{pmatrix} 0 & & & \\ I_n & 0 & & \\ & \ddots & \ddots & \\ & & I_n & 0 \end{pmatrix},$$

$$W_{3L} = \begin{pmatrix} 0 & & & \\ W_{n1} & 0 & & \\ & \ddots & \ddots & \\ & & W_{n,T-1} & 0 \end{pmatrix} \text{ so that } W_L(\eta_0) = \lambda_0 W_{1L} + \gamma_0 W_{2L} + \rho_0 W_{3L}.$$

$S_{nt}(\lambda) = I_n - \lambda W_{nt}$  for any possible  $\lambda$  and  $G_{nt}(\lambda) = W_{nt} S_{nt}^{-1}(\lambda)$ .

$S_L(\eta) = I_L - W_L(\eta)$  and  $S_L = I - W_L(\eta_0)$ .

$G_{jL}(\eta) = W_{jL} S_L^{-1}(\eta)$  and  $G_{jL} = W_{jL} S_L^{-1}$  for  $j = 1, 2, 3$ .

$Q_{1L} = G_{1L}(X_{1L}\beta_0 + \mathcal{I}_0(\gamma_0, \rho_0) + \epsilon_L\delta_0 + \mathbf{c}_{1L0} + \alpha_{1L0})$ .

$J_n = I_n - \frac{1}{n} l_n l_n'$  and  $J_T = I_T - \frac{1}{T} l_T l_T'$ , which are the two orthogonal projectors.

$\varsigma_l = f(\epsilon_l, \xi_l, Z_{l-n}, X_L, \mathbf{c}_{n10}, \alpha_{T10})$  and  $\varsigma_L = (\varsigma_{1,L}, \dots, \varsigma_{L,L})'$  is an  $nT$ -dimension column vector.

$\frac{\partial \ln L_L^c(\theta_0)}{\partial \theta} = \frac{\partial \ln L_{1,L}^c(\theta_0)}{\partial \theta} + \Delta_L$  where  $\frac{\partial \ln L_{1,L}^c(\theta_0)}{\partial \theta} = \frac{\partial \ln L_L^c(\theta_0)}{\partial \theta} - E(\frac{\partial \ln L_L^c(\theta_0)}{\partial \theta})$  and  $\Delta_L = E(\frac{\partial \ln L_L^c(\theta_0)}{\partial \theta})$ .

$\mathcal{J}_{nT, \theta_0} = E\left(\frac{1}{L} \frac{\partial^2 \ln L_{1,L}^c(\theta_0)}{\partial \theta \partial \theta'}\right)$  and  $E\left(\frac{1}{L} \frac{\partial \ln L_{1,L}^c(\theta_0)}{\partial \theta} \frac{\partial \ln L_{1,L}^c(\theta_0)}{\partial \theta'}\right) = \mathcal{J}_{nT, \theta_0} + \Omega_{nT, \theta_0} + o_p(1)$  where  $\Omega_{nT, \theta_0}$  is related to the third and fourth moment of  $\xi_L$ .

## Appendix B. Lemmas and proofs

Denote  $\varsigma_{l,L}$  as a function of  $\epsilon_l, \xi_l, Z_{l-n}, X_L, \mathbf{c}_{n10}$  and  $\alpha_{T10} = (\alpha_{1,10}, \dots, \alpha_{T,10})'$ . As an  $nT$ -dimension column vector,  $\varsigma_L = (\varsigma_{1,L}, \dots, \varsigma_{L,L})'$ . In the subsequent lemmas and corollaries,  $\varsigma_{1L}$  and  $\varsigma_{2L}$  are such  $nT$ -dimensional vectors.<sup>20</sup>

By denoting  $B_{n,t-1} = \gamma_0 I_n + \rho_0 W_{n,t-1}$ , we can write  $W_L(\eta_0) = \lambda_0 W_{1L} + \mathcal{B}_L$  where

$$\mathcal{B}_L = \begin{pmatrix} \mathbf{0} & & & \\ B_{n1} & \mathbf{0} & & \\ & B_{n2} & \ddots & \\ & & \ddots & \ddots \\ & & & B_{n,T-1} & \mathbf{0} \end{pmatrix}.$$

Therefore,

$$\begin{aligned} [I_L - W_L(\eta_0)]^{-1} &= \sum_{m=0}^{\infty} W_L^m(\eta_0) = \sum_{m=0}^{\infty} (\lambda_0 W_{1L} + \mathcal{B}_L)^m \\ &= \sum_{m=0}^{\infty} \sum_{i_1, \dots, i_m = (0,1)} \prod_{k=1}^m (\lambda_0 W_{1L})^{i_k} \mathcal{B}_L^{1-i_k}. \end{aligned}$$

For the powers of  $W_L$ , using the block diagonal  $W_{1L}$  and lower diagonal  $\mathcal{B}_L$ , we have (for simplicity, we ignore the subscript  $n$  in relevant matrices), for example, the equation in Box 1.

In general, these terms that need to be analyzed in  $W_L^m$  can be written in a general pattern as

$$W_{1L}^{p_1} \mathcal{B}_L^{q_1} W_{1L}^{p_2} \mathcal{B}_L^{q_2} \dots W_{1L}^{p_m} \mathcal{B}_L^{q_m}, \text{ with } \sum_{k=1}^m (p_k + q_k)$$

$= m$  and  $0 \leq p_k, q_k \leq m$ .

Denote  $C_s^{(l_1, l_2)} = W_{s+1}^{l_1} \prod_{k=1}^{l_2} B_{s+1-k}$  and further denote  $D_{1,m} = \prod_{l=1}^m C_{s_{ml}+t-1}^{(p_l, q_l)}$  with  $s_{ml} = \sum_{d=1}^m q_d$  such that  $D_{1,m} = C_{q_1+1}^{(p_1, q_1)} \dots C_{q_{m-1}+q_m}^{(p_{m-1}, q_{m-1})} C_{q_m}^{(p_m, q_m)}$  and  $D_{T-s_{m1}, m} = C_{T-1}^{(p_1, q_1)} \dots C_{T-1-(q_1+\dots+q_{m-2})}^{(p_{m-1}, q_{m-1})} C_{T-1-(q_1+\dots+q_{m-1})}^{(p_m, q_m)}$ . Then, by recursive mathematical induction, we have

$$W_{1L}^{p_1} \mathcal{B}_L^{q_1} W_{1L}^{p_2} \mathcal{B}_L^{q_2} \dots W_{1L}^{p_m} \mathcal{B}_L^{q_m} = \begin{pmatrix} \mathbf{0}_{ns_{m1} \times n} & & & \\ D_{1,m} & \mathbf{0} & & \\ \mathbf{0} & \ddots & & \\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & \dots & D_{T-s_{m1}, m} & \mathbf{0}_{n \times ns_{m1}} \end{pmatrix}.$$

<sup>20</sup> We note that  $\varsigma_{l,L}$  depends on time lag  $Z_{l-n}$  instead of just independent disturbances. Therefore, even though the statements of Lemmas 1–3 are similar to those in the cross sectional model in Qu and Lee (2015), the proofs in Qu and Lee (2015) do not directly go through.

For example, if all  $p_k$ 's are zeros, then  $q_1 + \dots + q_m = m$ ,  $s_{m1} = m$ ,  $D_{1,m} = \prod_{l=1}^m C_{s_{ml}+t-1}^{(0, q_l)} = C_{m+t-1}^{(0, m)} = \prod_{k=1}^m B_{m+t-k}$ , and

$$\mathcal{B}_L^m = \begin{pmatrix} \mathbf{0}_{mn \times n} & & & \\ B_m B_{m-1} \dots B_1 & \mathbf{0} & & \\ \mathbf{0} & B_{m+1} B_m \dots B_2 & \ddots & \\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & \dots & B_{T-1} B_{T-2} \dots B_{T-m} & \mathbf{0}_{n \times mn} \end{pmatrix};$$

indeed, for this case,  $W_{1L}^{p_1} \mathcal{B}_L^{q_1} \dots W_{1L}^{p_m} \mathcal{B}_L^{q_m} = \mathcal{B}_L^m$ . If all the  $q_k$ 's are zero, then  $p_1 + \dots + p_m = m$ ,  $D_{1,m} = \prod_{l=1}^m C_{t-1}^{(p_l, \cdot)} = C_{t-1}^{(m, 0)} = W_t^m$  and

$$W_{1L}^m = \begin{pmatrix} W_1^m & & & \\ \mathbf{0} & W_2^m & & \\ & \mathbf{0} & \ddots & \\ \vdots & & \mathbf{0} & \ddots \\ \mathbf{0} & \mathbf{0} & \dots & W_T^m \end{pmatrix}.$$

Indeed, for this case,  $W_{1L}^{p_1} \mathcal{B}_L^{q_1} \dots W_{1L}^{p_m} \mathcal{B}_L^{q_m} = W_{1L}^m$ .

**Claim 1.** The  $(i, j)$ th element of  $D_{1,m} = \prod_{l=1}^m C_{s_{ml}+t-1}^{(p_l, q_l)}$  is nonzero only if  $j \in B_i(m, \rho_c)$ , where  $C_s^{(l_1, l_2)} = W_{s+1}^{l_1} \prod_{k=1}^{l_2} B_{s+1-k}$ ,  $s_{ml} = \sum_{d=1}^m q_d$ , and  $B_s = \gamma_0 I_n + \rho_0 W_{ns}$ . For any time period  $s$ , the  $(i, j)$ th element of  $W_s D_{1,m}$  is nonzero only if  $j \in B_i(\rho_c[1+m])$ .

**Lemma 1.** Under Assumptions 1–6, for any finite integers  $a$  and  $b$ ,  $\frac{1}{L} \varsigma_{1L}' W_L^a W_L^b \varsigma_{2L} - E(\frac{1}{L} \varsigma_{1L}' W_L^a W_L^b \varsigma_{2L}) = o_p(1)$ ; and also for any  $j = 1, 2, 3$ ,  $\frac{1}{L} \varsigma_{1L}' W_L^a W_{jL} W_L^b \varsigma_{2L} - E(\frac{1}{L} \varsigma_{1L}' W_L^a W_{jL} W_L^b \varsigma_{2L}) = o_p(1)$ .

**Lemma 2.** Under Assumptions 1–6,  $\frac{1}{L} \varsigma_{1L}' [I_L - W_L(\eta_0)]^{-1} \varsigma_{2L} - E(\frac{1}{L} \varsigma_{1L}' [I_L - W_L(\eta_0)]^{-1} \varsigma_{2L}) = o_p(1)$ . Furthermore, for any  $j = 1, 2, 3$ ,  $\frac{1}{L} \varsigma_{1L}' G_{jL} \varsigma_{2L} - E(\frac{1}{L} \varsigma_{1L}' G_{jL} \varsigma_{2L}) = o_p(1)$ .

**Lemma 3.** Under Assumptions 1–6, for any  $j = 1, 2, 3$ ,  $\frac{1}{L} \varsigma_{1L}' [I_L - W_L(\eta_0)]^{-1} G_{jL} \varsigma_{2L} - E(\frac{1}{L} \varsigma_{1L}' [I_L - W_L(\eta_0)]^{-1} G_{jL} \varsigma_{2L}) = o_p(1)$ .

Now we consider  $\frac{1}{L} \varsigma_{1L}' (J_T \otimes J_n) W_L^m \varsigma_{2L}$  and  $\frac{1}{L} \varsigma_{1L}' (I_L - W_L)^{-1} (J_T \otimes J_n) (I_L - W_L)^{-1} \varsigma_{2L}$ , where

$$J_T \otimes J_n = I_L - \frac{1}{T} l_T l_T' \otimes I_n - I_T \otimes \frac{1}{n} l_n l_n' + \frac{1}{L} l_L l_L'.$$

The terms  $\frac{1}{L} \varsigma_{1L}' W_L^a W_L^b \varsigma_{2L}$  and  $\frac{1}{L} \varsigma_{1L}' (I_L - W_L)^{-1} (I_L - W_L)^{-1} \varsigma_{2L}$  hold from Lemmas 1 and 3. We will separately show the convergence of other terms.

For terms corresponding to  $\frac{1}{T} l_T l_T' \otimes I_n$ ,  $I_T \otimes \frac{1}{n} l_n l_n'$ , and to  $\frac{1}{L} l_L l_L'$ , we have the following Lemmas 4 to 6, respectively.

**Lemma 4.** For any finite integers  $a$  and  $b$ ,

$$\begin{aligned} &\frac{1}{L} \varsigma_{1L}' W_L^a \left( \frac{1}{T} (l_T l_T') \otimes I_n \right) W_L^b \varsigma_{2L} \\ &- E \left[ \frac{1}{L} \varsigma_{1L}' W_L^a \left( \frac{1}{T} (l_T l_T') \otimes I_n \right) W_L^b \varsigma_{2L} \right] = o_p(1); \\ &\frac{1}{L} \varsigma_{1L}' (I_L - W_L)^{-1} \left( \frac{1}{T} (l_T l_T') \otimes I_n \right) (I_L - W_L)^{-1} \varsigma_{2L} \\ &- E \left[ \frac{1}{L} \varsigma_{1L}' (I_L - W_L)^{-1} \left( \frac{1}{T} (l_T l_T') \otimes I_n \right) (I_L - W_L)^{-1} \varsigma_{2L} \right] \\ &= o_p(1). \end{aligned}$$

$$W_{1L}^r \mathcal{B}_L^s W_{1L}^p \mathcal{B}_L^q = \begin{pmatrix} \mathbf{0}_{(s+q)n \times n} & & \\ W_{s+q+1}^r B_{s+q} \cdots B_{q+1} W_{q+1}^p B_q \cdots B_1 & \mathbf{0} & \\ \vdots & \ddots & \\ \mathbf{0} & \cdots & W_1^r B_{T-1} \cdots B_{T-s} W_{T-s}^p B_{T-s-1} \cdots B_{T-s-q} \end{pmatrix} \mathbf{0}_{n \times (s+q)n}.$$

Box I.

The most difficult terms to analyze are terms corresponding to  $I_T \otimes \frac{1}{n} l_n l_n'$ , i.e.,  $\frac{1}{L} \mathcal{S}'_L W_L^a (I_T \otimes \frac{1}{n} l_n l_n') W_L^b \mathcal{S}_{2L}$  and  $\frac{1}{L} \mathcal{S}'_L (I_L - W_L)^{-1} (I_T \otimes \frac{1}{n} l_n l_n') (I_L - W_L)^{-1} \mathcal{S}_{2L}$ . For those terms, the structure in [Assumption 3](#)(iii) on the Lipschitz condition  $|h(z_{it'}, z_{jt'}) - h(z_{it}, z_{jt})| \leq c_0(|z_{it'} - z_{it}| + |z_{jt'} - z_{jt}|)$  plays an important role. Many functions of  $h$  would satisfy this assumption.

**Example 1.** If  $h(z_i, z_j) = \frac{1}{|z_i - z_j|}$  when  $|z_i - z_j| \geq c$  and  $h(z_i, z_j) = \frac{1}{c}$  when  $|z_i - z_j| < c$ , where  $c$  is a positive constant, then we can verify the above Lipschitz condition. There are four different cases based on whether  $|z_i - z_j|$  exceeds the threshold  $c$ . Case (1), both  $|z_{it} - z_{jt}| \geq c$  and  $|z_{it'} - z_{jt'}| \geq c$ . Hence,

$$\begin{aligned} |h(z_{it}, z_{jt}) - h(z_{it'}, z_{jt'})| &= \left| \frac{|z_{it'} - z_{jt'}| - |z_{it} - z_{jt}|}{|z_{it} - z_{jt}| \cdot |z_{it'} - z_{jt'}|} \right| \\ &\leq \frac{1}{c^2} (|z_{it'} - z_{it}| + |z_{jt'} - z_{jt}|). \end{aligned}$$

This inequality holds because the numerator follows from the triangular inequality. Case (2), both  $|z_{it} - z_{jt}| < c$  and  $|z_{it'} - z_{jt'}| < c$ , then  $|w_{ij,t} - w_{ij,t'}| = 0$ . Case (3),  $|z_{it} - z_{jt}| \geq c$  and  $|z_{it'} - z_{jt'}| < c$ .

$$\begin{aligned} |h(z_{it'}, z_{jt'}) - h(z_{it}, z_{jt})| &= \frac{1}{c} - \frac{1}{|z_{it} - z_{jt}|} \leq \frac{|z_{it} - z_{jt}| - c}{c|z_{it} - z_{jt}|} \\ &< \frac{1}{c^2} (|z_{it} - z_{jt}| - |z_{it'} - z_{jt'}|) < \frac{1}{c^2} (|z_{it'} - z_{it}| + |z_{jt'} - z_{jt}|). \end{aligned}$$

By symmetry, Case (4) also holds.

**Example 2.** For the above  $h(z_i, z_j)$ , we consider the row-normalized  $W_{nt}^*$ . Because  $|w_{ij,t}^* - w_{ij,t'}^*| = 0$  if  $j \notin B_i(\rho_c)$ , we only need to consider the case  $j \in B_i(\rho_c)$ . As  $w_{ij,t}^* = \frac{w_{ij,t}}{\sum_{k=1}^n w_{ik,t}} = \frac{w_{ij,t}}{\sum_{k \in B_i(\rho_c)} w_{ik,t}}$ ,

$$\begin{aligned} |w_{ij,t}^* - w_{ij,t'}^*| &= \frac{1}{\left| \sum_{k \in B_i(\rho_c)} w_{ik,t} \cdot \sum_{k \in B_i(\rho_c)} w_{ik,t'} \right|} \\ &\quad \times \left| w_{ij,t} \sum_{k \in B_i(\rho_c)} w_{ik,t'} - w_{ij,t'} \sum_{k \in B_i(\rho_c)} w_{ik,t} \right| \\ &\leq \frac{1}{c_L^2} \left( |w_{ij,t}| \sum_{k \in B_i(\rho_c)} |w_{ik,t'} - w_{ik,t}| \right. \\ &\quad \left. + |w_{ij,t'} - w_{ij,t}| \sum_{k \in B_i(\rho_c)} |w_{ik,t}| \right) \\ &\leq \frac{1}{c_L^2} \frac{c_w}{c^2} \sum_{k \in B_i(\rho_c)} (|z_{kt'} - z_{kt}| + |z_{it'} - z_{it}|) \\ &\quad + (|z_{jt'} - z_{jt}| + |z_{it'} - z_{it}|) \\ &\leq \frac{2}{c_L^2} \frac{c_w}{c^2} \sum_{k \in B_i(\rho_c)} (|z_{kt'} - z_{kt}| + |z_{it'} - z_{it}|), \end{aligned}$$

where  $c_L$  is the lower bound of  $\sum_{k=1}^n w_{ik,t}$ , i.e.,  $\inf_{i,n} \sum_{k=1}^n w_{ik,nt} \geq c_L$ , which means that all individuals have at least one neighbor with positive weights bounded away from zero. It has a similar Lipschitz structure as that in [Assumption 3](#).

**Example 3.** If  $h(z_i, z_j) = z_i^{\gamma_1} z_j^{\gamma_2} g(d_{ij})$ , where  $d_{ij}$  is the time invariant physical distance,  $z_i$  and  $z_j$  are positive bounded values that may change over time, and  $\gamma_1, \gamma_2 > 0$ , then by the mean value theorem,

$$\begin{aligned} |h(z_{it}, z_{jt}) - h(z_{it'}, z_{jt'})| &= |g(d_{ij})| \cdot |z_{it}^{\gamma_1} (z_{jt}^{\gamma_2} - z_{jt'}^{\gamma_2}) + z_{jt'}^{\gamma_2} (z_{it}^{\gamma_1} - z_{it'}^{\gamma_1})| \\ &\leq |g(d_{ij})| \left( |z_{jt} - z_{jt'}| \cdot \gamma_2 \tilde{z}_{jt}^{\gamma_2-1} + |z_{it} - z_{it'}| \cdot \gamma_1 \tilde{z}_{it}^{\gamma_1-1} \right) \sup_{i,t} |z_{it}^{\gamma_1}| \\ &\leq (\gamma_1 + \gamma_2) |g(d_{ij})| (|z_{jt} - z_{jt'}| + |z_{it} - z_{it'}|) \sup_{i,t} |z_{it}^{\gamma_1}| \\ &\quad \times \left( \sup_{i,t} |z_{it}^{\gamma_2-1}| + \sup_{i,t} |z_{it}^{\gamma_1-1}| \right) \\ &\leq c_0 (|z_{jt} - z_{jt'}| + |z_{it} - z_{it'}|). \end{aligned}$$

**Claim 2.** For any finite integer  $m$ ,  $D_{t,m,j}$  satisfies the NED property that  $\sup_{i,j,n,t} \|D_{t,m,j} - E(D_{t,m,j} | \mathcal{F}_{t,T}(S))\|_2 \leq C_m |r_0|^s$ , where  $|r_0| = \|\gamma_0\|_1 < 1$  and  $\mathcal{F}_{t,T}(S)$  is the  $\sigma$ -field generated by  $(\epsilon_{1t'}, v_{1t'}, \dots, \epsilon_{nt'}, v_{nt'}, X_L, \theta_0)$ 's with the time period from  $t' \in [t - s, t + s]$ .

Using this NED property, we can show the convergence of terms involving the deviation from cross sectional mean operator  $I_T \otimes \frac{1}{n} l_n l_n'$ .

**Lemma 5.** Under [Assumptions 1–6](#), for any finite integers  $a$  and  $b$ ,

$$\begin{aligned} \frac{1}{L} \mathcal{S}'_L W_L^a (I_T \otimes \frac{1}{n} l_n l_n') W_L^b \mathcal{S}_{2L} - E \left[ \frac{1}{L} \mathcal{S}'_L W_L^a (I_T \otimes \frac{1}{n} l_n l_n') W_L^b \mathcal{S}_{2L} \right] &= o_p(1); \\ \frac{1}{L} \mathcal{S}'_L (I_L - W_L)^{-1} \left( I_T \otimes \frac{1}{n} l_n l_n' \right) (I_L - W_L)^{-1} \mathcal{S}_{2L} & \\ - E \left[ \frac{1}{L} \mathcal{S}'_L (I_L - W_L)^{-1} \left( I_T \otimes \frac{1}{n} l_n l_n' \right) (I_L - W_L)^{-1} \mathcal{S}_{2L} \right] &= o_p(1). \end{aligned}$$

**Lemma 6.** Under [Assumptions 1–6](#), for any finite integers  $a$  and  $b$ ,

$$\begin{aligned} \frac{1}{L} l'_L W_L^a \mathcal{S}_{1L} \cdot \frac{1}{L} l'_L W_L^b \mathcal{S}_{2L} - E \left( \frac{1}{L} l'_L W_L^a \mathcal{S}_{1L} \cdot \frac{1}{L} l'_L W_L^b \mathcal{S}_{2L} \right) &= o_p(1); \\ \frac{1}{L^2} l'_L (I_L - W_L)^{-1} \mathcal{S}_{1L} l'_L (I_L - W_L)^{-1} \mathcal{S}_{2L} & \\ - E \left[ \frac{1}{L^2} l'_L (I_L - W_L)^{-1} \mathcal{S}_{1L} l'_L (I_L - W_L)^{-1} \mathcal{S}_{2L} \right] &= o_p(1). \end{aligned}$$

**Proof for Claim 1.** The  $(i, j)$ th element of  $D_{t,m} = \prod_{l=1}^m C_{s_{ml}+t-1}^{(p_l, q_l)}$  is nonzero only if  $j \in B_i(\rho_c) \sum_{k=1}^m (p_k + q_k) = B_i(m\rho_c)$ . Furthermore, the  $(i, j)$ th element of  $D_{s,a} D_{t,b}$  is nonzero only if  $j \in B_i(\rho_c(a+b))$ .

**Proof for Lemma 1.** If  $s_{a1} \leq s_{b1}$ , i.e.,  $q'_1 + \dots + q'_a \leq q_1 + \dots + q_b$ , then

$$\left( W_{1L}^{p'_1} \mathcal{B}_L^{q'_1} \cdots W_{1L}^{p'_a} \mathcal{B}_L^{q'_a} \right) \left( W_{1L}^{p_1} \mathcal{B}_L^{q_1} \cdots W_{1L}^{p_b} \mathcal{B}_L^{q_b} \right)$$



$$\begin{aligned}
&= \begin{pmatrix} \mathbf{0}_{n \times ns_{a1}} & D'_{1,a} & \mathbf{0} & \cdots & \mathbf{0} \\ & \mathbf{0} & \ddots & & \mathbf{0} \\ & & & \ddots & \vdots \\ & & & & D'_{T-s_{a1},a} \\ & & & & \mathbf{0}_{ns_{a1} \times n} \end{pmatrix} \\
&\times \begin{pmatrix} \mathbf{0}_{ns_{b1} \times n} & & & & \\ D_{1,b} & \mathbf{0} & & & \\ \mathbf{0} & \ddots & & & \\ \vdots & & \ddots & & \\ \mathbf{0} & \mathbf{0} & \cdots & D_{T-s_{b1},b} & \mathbf{0}_{n \times ns_{b1}} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{0}_{n(s_{b1}-s_{a1}) \times n} & & & & \\ D'_{s_{b1}-s_{a1}+1,a} D_{1,b} & \mathbf{0} & & & \\ \mathbf{0} & D'_{s_{b1}-s_{a1}+2,a} D_{2,b} & \mathbf{0} & & \\ \vdots & & \ddots & & \\ \mathbf{0} & \cdots & & D'_{T-s_{a1},a} D_{T-s_{b1},b} & \mathbf{0}_{n \times ns_{b1}} \end{pmatrix}.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\frac{1}{L} \varsigma'_{1L} \left( W_{1L}^{p'_1} \mathcal{B}_L^{q'_1} \cdots W_{1L}^{p'_a} \mathcal{B}_L^{q'_a} \right) \left( W_{1L}^{p_1} \mathcal{B}_L^{q_1} \cdots W_{1L}^{p_b} \mathcal{B}_L^{q_b} \right) \varsigma_{2L} \\
&= \frac{1}{T} \sum_{t=1}^{T-s_{b1}} \varsigma'_{1,s_{b1}-s_{a1}+t} D'_{s_{b1}-s_{a1}+t,a} D_{t,b} \varsigma_{2,t} \\
&= \frac{1}{n} \sum_{i=1}^n T_{i,n}, \text{ where } T_{i,n} = \frac{1}{T} \sum_{t=1}^{T-s_{b1}} \sum_{j \in B_i((a+b)\rho_c)} \varsigma_{1i,s_{b1}-s_{a1}+t} \\
&\quad \times (D'_{s_{b1}-s_{a1}+t,a} D_{t,b})_{ij} \varsigma_{2j,t}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\text{Var} \left[ \frac{1}{L} \varsigma'_{1L} \left( W_{1L}^{p'_1} \mathcal{B}_L^{q'_1} \cdots W_{1L}^{p'_a} \mathcal{B}_L^{q'_a} \right) \left( W_{1L}^{p_1} \mathcal{B}_L^{q_1} \cdots W_{1L}^{p_b} \mathcal{B}_L^{q_b} \right) \varsigma_{2L} \right] \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(T_{i,n}, T_{j,n}) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j \in B_i(2(a+b)\rho_c)} \text{Cov}(T_{i,n}, T_{j,n}) \\
&\leq \frac{1}{n^2} \sum_{i=1}^n \sum_{j \in B_i(2(a+b)\rho_c)} \sup_i \mathbb{E} \left| \frac{1}{T} \sum_{t=1}^{T-s_{b1}} \sum_{j \in B_i((a+b)\rho_c)} \varsigma_{1i,s_{b1}-s_{a1}+t} \right. \\
&\quad \left. \times (D'_{s_{b1}-s_{a1}+t,a} D_{t,b})_{ij} \varsigma_{2j,t} \right|^2 \leq \frac{1}{n} c((a+b)\rho_c)^{3d}.
\end{aligned}$$

The second “=” holds because all individuals involved in  $T_{i,n}$  are within  $B_i((a+b)\rho_c)$ , so when locations  $i$  and  $j$  are more than  $2(a+b)\rho_c$  apart,  $T_{i,n}$  and  $T_{j,n}$  has no common elements and also  $\varsigma_{1i,t}$  and  $\varsigma_{2j,s}$  are independent whenever  $i \neq j$ , for all  $t$  and  $s$ . The last “ $\leq$ ” holds because  $\sup_{i,j,t_1,t_2} \mathbb{E} |\varsigma_{1i,t_1} (D'_{t_1,a} D_{t_2,b})_{ij} \varsigma_{2j,t_2}|^2 < \infty$ .

If  $s_{a1} \geq s_{b1}$ , the analysis would be similar to the case of  $s_{a1} \leq s_{b1}$  by taking a transpose.

For any finite integers  $a$  and  $b$ ,  $\frac{1}{L} \varsigma'_{1L} W_L^a W_L^b \varsigma_{2L}$  can be written as a summation of  $2^{(a+b)}$  terms of  $\frac{1}{L} \varsigma'_{1L} \left( W_{1L}^{p'_1} \mathcal{B}_L^{q'_1} \cdots W_{1L}^{p'_a} \mathcal{B}_L^{q'_a} \right) \left( W_{1L}^{p_1} \mathcal{B}_L^{q_1} \cdots W_{1L}^{p_b} \mathcal{B}_L^{q_b} \right) \varsigma_{2L}$ . The same arguments can be applied to  $\frac{1}{L} \varsigma'_{1L} W_L^a W_{jL} W_L^b \varsigma_{2L}$  for  $j = 1, 2$ , and  $3$ . Therefore,

$$\begin{aligned}
&\text{Var} \left[ \frac{1}{L} \varsigma'_{1L} W_L^a W_{jL} W_L^b \varsigma_{2L} \right] = O \left( \frac{1}{n} \right) \text{ and } \frac{1}{L} \left[ \varsigma'_{1L} W_L^a W_{jL} W_L^b \varsigma_{2L} \right. \\
&\quad \left. - \mathbb{E}(\varsigma'_{1L} W_L^a W_{jL} W_L^b \varsigma_{2L}) \right] = O_p \left( \frac{1}{\sqrt{n}} \right). \quad (16)
\end{aligned}$$

**Proof for Lemma 2.** We decompose  $\frac{1}{L} \varsigma'_{1L} [I_L - W_L(\eta_0)]^{-1} \varsigma_{2L}$  into two parts and show each part separately:

$$\begin{aligned}
\frac{1}{L} \varsigma'_{1L} [I_L - W_L(\eta_0)]^{-1} \varsigma_{2L} &= \frac{1}{L} \sum_{m=0}^{\infty} \varsigma'_{1L} W_L^m \varsigma_{2L} = \frac{1}{L} \sum_{m=0}^M \varsigma'_{1L} W_L^m \varsigma_{2L} \\
&\quad + \frac{1}{L} \sum_{m=M+1}^{\infty} \varsigma'_{1L} W_L^m \varsigma_{2L}.
\end{aligned}$$

Here,

$$W_L^m = \sum_{r=0}^m \lambda_0^{m-r} \sum_{\substack{p_1+p_2+\cdots+p_m=m-r \\ q_1+q_2+\cdots+q_m=r}} W_{1L}^{p_1} \mathcal{B}_L^{q_1} W_{1L}^{p_2} \mathcal{B}_L^{q_2} \cdots W_{1L}^{p_m} \mathcal{B}_L^{q_m},$$

and

$$\begin{aligned}
&\frac{1}{L} \varsigma'_{1L} \left( W_{1L}^{p_1} \mathcal{B}_L^{q_1} W_{1L}^{p_2} \mathcal{B}_L^{q_2} \cdots W_{1L}^{p_m} \mathcal{B}_L^{q_m} \right) \varsigma_{2L} \\
&= \frac{1}{nT} \sum_{t=1}^{T-s_{m1}} \varsigma'_{1,s_{m1}+t} D_{t,m} \varsigma_{2,t}.
\end{aligned}$$

As  $D_{t,m} = \prod_{l=1}^m C_{s_{ml}+t-1}^{(p_l, q_l)}$  with  $C_s^{(l_1, l_2)} = W_{s+1}^{l_1} \prod_{k=1}^{l_2} B_{s+1-k}$  and  $s_{ml} = \sum_{d=1}^m q_d$ ,

$$\begin{aligned}
\mathbb{E} |\varsigma'_{1,s_{m1}+t} D_{t,m} \varsigma_{2,t}| &\leq \sum_{i=1}^n \sum_{j \in B_i(m\rho_c)} \mathbb{E} |D_{t,m,ij} \varsigma_{1i,s_{m1}+t} \varsigma_{2j,t}| \\
&\leq \sum_{i=1}^n \sum_{j \in B_i(m\rho_c)} \sup_t |D_{t,m,ij}| \sqrt{\sup_{i,t} \mathbb{E}(\varsigma_{1i,t}^2) \sup_{i,t} \mathbb{E}(\varsigma_{2i,t}^2)} \\
&\leq n(m\rho_c)^d \sup_{n,t} \|W_{nt}\|_{\infty}^{m-r} \cdot \sup_t \|B_t\|_{\infty}^r C_{2v} \\
&= nm^d c_w^{m-r} (|\rho_0| + |\gamma_0| c_w)^r C, \text{ for some constants } C_{2v} \text{ and } C.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{1}{L} \mathbb{E} \left| \varsigma'_{1L} \left( W_{1L}^{p_1} \mathcal{B}_L^{q_1} W_{1L}^{p_2} \mathcal{B}_L^{q_2} \cdots W_{1L}^{p_m} \mathcal{B}_L^{q_m} \right) \varsigma_{2L} \right| \\
&\leq \frac{1}{nT} \sum_{t=1}^{T-s_{m1}} nm^d c_w^{m-r} (|\rho_0| + |\gamma_0| c_w)^r C \\
&\leq m^d C c_w^{m-r} (|\rho_0| + |\gamma_0| c_w)^r; \\
&\frac{1}{L} \mathbb{E} |\varsigma'_{1L} W_L^m \varsigma_{2L}| \leq m^d C \sum_{r=0}^m \binom{m}{r} |\lambda_0 c_w|^{m-r} \cdot (|\rho_0| + |\gamma_0| c_w)^r \\
&= m^d C (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^m; \quad (17)
\end{aligned}$$

and

$$\frac{1}{L} \sum_{m=M+1}^{\infty} \mathbb{E} |\varsigma'_{1L} W_L^m \varsigma_{2L}| \leq C_2 M^d (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^M.$$

For any  $1 > \delta > 0$ , there exists an  $M_\delta$  such that  $M_\delta^d (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^{M_\delta} < \delta^2 / (4C_2)$ . We can choose  $M$  big enough ( $M \geq M_\delta$ ), but independent of  $n$  and  $T$ , such that

$$\begin{aligned}
&\frac{1}{L} \sum_{m=M+1}^{\infty} \mathbb{E} |\varsigma'_{1L} W_L^m \varsigma_{2L}| < \frac{\delta^2}{4}, \text{ and hence,} \\
&P \left( \left| \frac{1}{L} \sum_{m=M+1}^{\infty} \varsigma'_{1L} W_L^m \varsigma_{2L} \right| > \frac{\delta}{4} \right) \leq \frac{\frac{1}{L} \sum_{m=M+1}^{\infty} \mathbb{E} |\varsigma'_{1L} W_L^m \varsigma_{2L}|}{\delta/4} < \delta.
\end{aligned}$$

Denote

$$\begin{aligned}\mathcal{R}_L &= \frac{1}{L} \varsigma'_{1L} (I_L - W_L)^{-1} \varsigma_{2L} - \mathbb{E} \left[ \frac{1}{L} \varsigma'_{1L} (I_L - W_L)^{-1} \varsigma_{2L} \right] \\ &= \mathcal{R}_{1L} + \mathcal{R}_{2L}, \text{ where} \\ \mathcal{R}_{1L} &= \frac{1}{L} \sum_{m=0}^M \varsigma'_{1L} W_L^m \varsigma_{2L} - \mathbb{E} \left( \frac{1}{L} \sum_{m=0}^M \varsigma'_{1L} W_L^m \varsigma_{2L} \right) \text{ and } \mathcal{R}_{2L} \\ &= \frac{1}{L} \sum_{m=M+1}^{\infty} \varsigma'_{1L} W_L^m \varsigma_{2L} - \mathbb{E} \left( \frac{1}{L} \sum_{m=M+1}^{\infty} \varsigma'_{1L} W_L^m \varsigma_{2L} \right).\end{aligned}$$

We want to show that for any  $1 > \delta > 0$ , there exists an  $L_\delta$  such that when  $L \geq L_\delta$ ,  $P_L = P(|\mathcal{R}_L| > \delta) < 2\delta$ . As

$$\begin{aligned}P_L &= P(|\mathcal{R}_{1L} + \mathcal{R}_{2L}| > \delta) \leq P(|\mathcal{R}_{1L}| > \delta/2) + P(|\mathcal{R}_{2L}| > \delta/2) \\ &\leq P(|\mathcal{R}_{1L}| > \delta/2) + P \left( \left| \frac{1}{L} \sum_{m=M+1}^{\infty} \varsigma'_{1L} W_L^m \varsigma_{2L} \right| > \delta/4 \right) \\ &\quad + P \left( \mathbb{E} \left| \frac{1}{L} \sum_{m=M+1}^{\infty} \varsigma'_{1L} W_L^m \varsigma_{2L} \right| > \delta/4 \right),\end{aligned}$$

when we choose  $M = M_\delta$ ,  $P_L \leq \text{Var} \left[ \frac{1}{L} \varsigma'_{1L} W_L^M \varsigma_{2L} \right] / (\delta/2)^2 + \delta = O(1/n) + \delta$  by Lemma 1 and preceding analysis. The same arguments can be applied to  $\frac{1}{L} \varsigma'_{1L} W_{jL} (I_L - W_L)^{-1} \varsigma_{2L}$  for  $j = 1, 2$ , and 3. Therefore,

$$\frac{1}{L} \varsigma'_{1L} G_{jL} \varsigma_{2L} - \left[ \frac{1}{L} \varsigma'_{1L} G_{jL} \varsigma_{2L} \right] = o_p(1).$$

**Proof for Lemma 3.** We decompose it into two parts.

$$\begin{aligned}[I_L - W'_L(\eta_0)]^{-1} [I_L - W_L(\eta_0)]^{-1} &= \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} W'^a_L W^b_L \\ &= \sum_{m=0}^{\infty} \sum_{a=0}^m W'^a_L W^{m-a}_L = \sum_{m=0}^M \sum_{a=0}^m W'^a_L W^{m-a}_L \\ &\quad + \sum_{m=M+1}^{\infty} \sum_{a=0}^m W'^a_L W^{m-a}_L.\end{aligned}$$

For any finite  $M$ , the convergence of  $\frac{1}{L} \varsigma'_{1L} \sum_{m=0}^M \sum_{a=0}^m W'^a_L W^{m-a}_L \varsigma_{2L}$  can be shown from Lemma 1. Now we need to consider  $\frac{1}{L} \varsigma'_{1L} \sum_{m=M+1}^{\infty} \sum_{a=0}^m W'^a_L W^{m-a}_L \varsigma_{2L}$ . As

$$\begin{aligned}\frac{1}{L} \varsigma'_{1L} \left( W'^{p'_1}_{1L} \mathcal{B}^{q'_1}_{1L} \cdots W'^{p'_{a'}}_{1L} \mathcal{B}^{q'_{a'}}_{1L} \right) \left( W^{p_1}_{1L} \mathcal{B}^{q_1}_{1L} \cdots W^{p_b}_{1L} \mathcal{B}^{q_b}_{1L} \right) \varsigma_{2L} \\ = \frac{1}{T} \sum_{t=1}^{T-s_{b1}} \varsigma'_{1, s_{b1}-s_{a1}+t} D'_{s_{b1}-s_{a1}+t, a} D_{t, b} \varsigma_{2, t},\end{aligned}$$

similar to the proof of Lemma 2, we can show

$$\begin{aligned}\mathbb{E} \left| \varsigma'_{1, s_{b1}-s_{a1}+t} D'_{s_{b1}-s_{a1}+t, a} D_{t, b} \varsigma_{2, t} \right| \\ \leq n(a+b)^d c_w^{a+b-r_1-r_2} (|\rho_0| + |\gamma_0| c_w)^{r_1+r_2} C,\end{aligned}$$

where  $C$  is a constant independent of  $T$  or  $n$ , expression in  $W_L^a$  has products of  $B_t$ 's with power of  $r_1$  terms and  $W_{nt}$ 's with  $(a-r_1)$  terms, i.e.,  $q'_1 + \cdots + q'_{a'} = r_1$ ,  $p'_1 + \cdots + p'_{a'} = a-r_1$ ,  $q_1 + \cdots + q_b = r_2$ , and  $p_1 + \cdots + p_b = b-r_2$ . Hence,

$$\begin{aligned}\frac{1}{L} \mathbb{E} \left| \varsigma'_{1L} W'^a_L W^b_L \varsigma_{2L} \right| &\leq (a+b)^d C \sum_{r_1=0}^a \binom{a}{r_1} \\ &\quad \times \sum_{r_2=0}^b \binom{b}{r_2} |\lambda_0 c_w|^{a+b-r_1-r_2} \cdot (|\rho_0| + |\gamma_0| c_w)^{r_1+r_2}\end{aligned}$$

$$= (a+b)^d C (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^{a+b}; \quad (18)$$

$$\frac{1}{L} \mathbb{E} \left| \varsigma'_{1L} \sum_{a=0}^m W'^a_L W^{m-a}_L \varsigma_{2L} \right| \leq m^{d+1} C (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^m;$$

$$\begin{aligned}\frac{1}{L} \sum_{m=M+1}^{\infty} \mathbb{E} \left| \varsigma'_{1L} \sum_{a=0}^m W'^a_L W^{m-a}_L \varsigma_{2L} \right| \\ \leq C_2 M^{d+1} (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^M.\end{aligned}$$

Using the similar argument in the proof of Lemma 2, for any  $j = 1, 2, 3$ , we have

$$\begin{aligned}\frac{1}{L} \varsigma'_{1L} [I_L - W_L(\eta_0)']^{-1} W_{jL} [I_L - W_L(\eta_0)]^{-1} \varsigma_{2L} \\ - \mathbb{E} \left( \frac{1}{L} \varsigma'_{1L} [I_L - W_L(\eta_0)']^{-1} W_{jL} [I_L - W_L(\eta_0)]^{-1} \varsigma_{2L} \right) = o_p(1).\end{aligned}$$

**Proof for Lemma 4.**

$$\begin{aligned}\frac{1}{L} \varsigma'_{1L} W'^a_L \left( \frac{1}{T} (I_T I'_T) \otimes I_n \right) W^b_L \varsigma_{2L} &= \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{T} (I'_T \otimes e'_{i,n}) W^a_L \varsigma_{1L} \right) \\ &\quad \times \left( \frac{1}{T} (I'_T \otimes e'_{i,n}) W^b_L \varsigma_{2L} \right) = \frac{1}{n} \sum_{i=1}^n a_i b_i.\end{aligned}$$

Similar to the proof of Lemma 1,  $a_i$  can be expressed as  $2^a$  terms of

$$\begin{aligned}\frac{1}{T} (I'_T \otimes e'_{i,n}) W^{p_1}_{1L} \mathcal{B}^{q_1}_{1L} \cdots W^{p_a}_{1L} \mathcal{B}^{q_a}_{1L} \varsigma_{1L} \\ = \frac{1}{T} \sum_{t=1}^{T-s_{a1}} \sum_{j_1 \in B_i(a\rho_c)} D_{t, a, ij_1} \varsigma_{1j_1, t} = \frac{1}{T} \sum_{t=1}^{T-s_{a1}} \sum_{j \in B_i(a\rho_c)} D_{t, a, ij} \varsigma_{1j, t}\end{aligned}$$

and  $a_i b_i$  can be expressed as  $2^{a+b}$  terms of

$$H_i = \frac{1}{T^2} \sum_{t_1=1}^{T-s_{a1}} \sum_{j_1 \in B_i(a\rho_c)} D_{t_1, a, ij_1} \varsigma_{1j_1, t_1} \sum_{t_2=1}^{T-s_{b1}} \sum_{j_2 \in B_i(b\rho_c)} D_{t_2, b, ij_2} \varsigma_{2j_2, t_2}.$$

From this, we can see  $H_i$  only involves  $(\epsilon_{j_1}, \xi_{j_1}, \dots, \epsilon_{j_T}, \xi_{j_T})$ 's within the ball  $B_i(\max(a, b)\rho_c)$ . When  $k \notin B_i(2\max(a, b)\rho_c)$ ,  $H_i$  and  $H_k$  have no common variables and hence are independent.

$$\frac{1}{L} \varsigma'_{1L} W'^a_L \left( \frac{1}{T} (I_T I'_T) \otimes I_n \right) W^b_L \varsigma_{2L} = \frac{1}{n} \sum_{i=1}^n a_i b_i = \sum_{m=1}^{2^{a+b}} \frac{1}{n} \sum_{i=1}^n H_i^{(m)}.$$

For each  $H_i^{(m)}$ ,  $m = 1, \dots, 2^{a+b}$ , using similar arguments in the proof of Lemma 1,

$$\begin{aligned}\text{Var} \left( \frac{1}{n} \sum_{i=1}^n H_i^{(m)} \right) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \text{Cov}(H_i^{(m)}, H_k^{(m)}) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{k \in B_i(2\max(a, b)\rho_c)} \text{Cov}(H_i^{(m)}, H_k^{(m)}) \\ &\leq \frac{1}{n} C[(a+b)\rho_c]^{5d} \sqrt{\sup_{i, t} E(\varsigma_{1i, t}^4) \sup_{i, t} E(\varsigma_{2i, t}^4)} = o\left(\frac{1}{n}\right).\end{aligned}$$

From this,  $\frac{1}{n} \sum_{i=1}^n H_i^{(m)} - E\left(\frac{1}{n} \sum_{i=1}^n H_i^{(m)}\right) = O_p(1/\sqrt{n})$  holds for any  $m = 1, \dots, 2^{a+b}$ . Hence,

$$\begin{aligned}\frac{1}{L} \varsigma'_{1L} W'^a_L \left( \frac{1}{T} (I_T I'_T) \otimes I_n \right) W^b_L \varsigma_{2L} \\ - E \left[ \frac{1}{L} \varsigma'_{1L} W'^a_L \left( \frac{1}{T} (I_T I'_T) \otimes I_n \right) W^b_L \varsigma_{2L} \right] = o_p\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

We decompose  $H_L = \frac{1}{L} \zeta'_{1L} (I_L - W'_L)^{-1} \left( \frac{1}{T} (l_T l'_T) \otimes I_n \right) (I_L - W_L)^{-1} \zeta_{2L}$  into three terms:

$$\begin{aligned} H_L &= \frac{1}{L} \zeta'_{1L} \sum_{a=0}^{\infty} W_L'^a \left( \frac{1}{T} (l_T l'_T) \otimes I_n \right) \sum_{b=0}^{\infty} W_L^b \zeta_{2L} \\ &= \frac{1}{L} \zeta'_{1L} \sum_{a=0}^A W_L'^a \left( \frac{1}{T} (l_T l'_T) \otimes I_n \right) \sum_{b=0}^B W_L^b \zeta_{2L} \\ &\quad + \frac{1}{L} \sum_{a=A+1}^{\infty} \zeta'_{1L} W_L'^a \left( \frac{1}{T} (l_T l'_T) \otimes I_n \right) \sum_{b=0}^B W_L^b \zeta_{2L} \\ &\quad + \frac{1}{L} \zeta'_{1L} \sum_{a=0}^{\infty} W_L'^a \left( \frac{1}{T} (l_T l'_T) \otimes I_n \right) \sum_{b=B+1}^{\infty} W_L^b \zeta_{2L} \\ &= H_{2L} + H_{3L} + H_{4L}. \end{aligned}$$

For any finite  $A$  and  $B$ , and any  $\delta > 0$ , we have shown  $P(|H_{2L} - E(H_{2L})| > \delta/3) \leq O(n^{-1/2})$ . Now we consider  $E|H_{3L}|$  and  $E|H_{4L}|$ . As

$$\begin{aligned} E \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t_1=1}^{T-s_{a1}} \sum_{j_1 \in B_i(a\rho_c)} D_{t_1, a, ij_1} \zeta_{1j_1, t_1} \sum_{t_2=1}^{T-s_{b1}} \sum_{j_2 \in B_i(b\rho_c)} D_{t_2, b, ij_2} \zeta_{2j_2, t_2} \right| \\ \leq a^d b^d \rho_c^{2d} \sup_{n, t} \|W_{nt}\|_{\infty}^{a+b-r_1-r_2} \\ \cdot \sup_t \|B_t\|_{\infty}^{r_1+r_2} \sup_{i, t} \sqrt{E(\zeta_{1i, t}^2) E(\zeta_{2i, t}^2)} \\ \leq (ab)^d c_w^{a+b-r_1-r_2} (|\rho_0| + |\gamma_0| c_w)^{r_1+r_2} C_1, \end{aligned}$$

where  $C_1$  is a constant independent of  $T$  or  $n$ ,  $r_1$  and  $r_2$  are the same as the proof of Lemma 3. Hence,

$$\begin{aligned} \frac{1}{L} E \left| \zeta'_{1L} W_L'^a \left( \frac{1}{T} (l_T l'_T) \otimes I_n \right) W_L^b \zeta_{2L} \right| \\ \leq (ab)^d C_1 \sum_{r_1=0}^a \binom{a}{r_1} \sum_{r_2=0}^b \binom{b}{r_2} |\lambda_0 c_w|^{a+b-r_1-r_2} (|\rho_0| + |\gamma_0| c_w)^{r_1+r_2} \\ = C_1 (ab)^d (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^{a+b}, \end{aligned}$$

and

$$\begin{aligned} E|H_{3L}| &\leq \sum_{a=A+1}^{\infty} \sum_{b=0}^B \frac{1}{L} E \left| \zeta'_{1L} W_L'^a \left( \frac{1}{T} (l_T l'_T) \otimes I_n \right) W_L^b \zeta_{2L} \right| \\ &\leq A^d C (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^A; \\ E|H_{4L}| &\leq \sum_{a=0}^{\infty} \sum_{b=B+1}^{\infty} \frac{1}{L} E \left| \zeta'_{1L} W_L'^a \left( \frac{1}{T} (l_T l'_T) \otimes I_n \right) W_L^b \zeta_{2L} \right| \\ &\leq B^d C (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^B \end{aligned}$$

for some constant  $C$  and  $C_1$ . For any  $\delta > 0$ , there exists an  $M_\delta$  big enough such that  $M_\delta^d (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^{M_\delta} < \delta^2/(6C)$ . We can choose  $A$  and  $B$  big enough ( $A, B \geq M_\delta$ ), but independent of  $n$  and  $T$ , such that

$$\begin{aligned} E|H_{3L}| &< \frac{\delta^2}{6}, E|H_{4L}| < \frac{\delta^2}{6}, P(|H_{3L}| > \frac{\delta}{6}) \leq \frac{E|H_{3L}|}{\delta/6} < \delta, \text{ and} \\ P(|H_{4L}| > \frac{\delta}{6}) &< \delta. \end{aligned}$$

We want to show that for any  $1 > \delta > 0$ , there exists an  $L_\delta$  such that when  $L \geq L_\delta$ ,  $P_L = P(|H_L - E(H_L)| > \delta) \rightarrow 0$ . As

$$\begin{aligned} P_L &= P(|H_{2L} - E(H_{2L})| + |H_{3L} - E(H_{3L})| + |H_{4L} - E(H_{4L})| > \delta) \\ &\leq P(|H_{2L} - E(H_{2L})| > \delta/3) + P(|H_{3L}| > \delta/6) \\ &\quad + P(E|H_{3L}| > \delta/6) + P(|H_{4L}| > \delta/6) + P(E|H_{4L}| > \delta/6), \end{aligned}$$

when we choose  $A = M_\delta$  and  $B = M_\delta$ ,

$$\begin{aligned} P_L &\leq P(|H_{2L} - E(H_{2L})| > \delta/3) + P(|H_{3L}| > \delta/6) \\ &\quad + P(|H_{4L}| > \delta/6) \leq O(n^{-1/2}) + 2\delta. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{1}{L} \zeta'_{1L} (I_L - W'_L)^{-1} \left( \frac{1}{T} (l_T l'_T) \otimes I_n \right) (I_L - W_L)^{-1} \zeta_{2L} \\ - E \left[ \frac{1}{L} \zeta'_{1L} (I_L - W'_L)^{-1} \left( \frac{1}{T} (l_T l'_T) \otimes I_n \right) (I_L - W_L)^{-1} \zeta_{2L} \right] \\ = o_p(1). \end{aligned}$$

**Proof for Claim 2.** First we show the time NED property of  $z_{it}$ . As

$$\begin{aligned} z_{it} &= \gamma'_0 z_{i, t-1} + \Gamma'_0 x_{2i, t} + c_{i, 20} + \alpha_{t20} + \epsilon_{i, t} \\ &= \lim_{L \rightarrow \infty} \sum_{q=0}^{t+L} \gamma_0'^q (\Gamma_0' x_{2i, t-q} + c_{i, 20} + \alpha_{t-q, 20} + \epsilon_{i, t-q}), \end{aligned}$$

we have

$$\begin{aligned} \|z_{it} - E(z_{it} | \mathcal{T}_{t, T}(s))\|_p &= \left\| \lim_{L \rightarrow \infty} \sum_{q=s+1}^{t+L} \gamma_0'^q \epsilon_{i, t-q} \right\|_p \\ &\leq \lim_{L \rightarrow \infty} \sum_{q=s+1}^{t+L} \|\gamma_0'^q\|_{\infty} \|\epsilon_{i, t-q}\|_p \leq c_{\epsilon p} \frac{|r_0|^s}{1 - |r_0|}, \end{aligned}$$

where  $|r_0| = \|\gamma_0\|_1 < 1$ . Then we show the time  $L_2$ -NED of  $w_{ij, t}$ . If  $j \notin B_i(\rho_c)$ , then  $w_{ij, t} = 0$ , so we only need to consider the case with  $j \in B_i(\rho_c)$ . By the Lipschitz condition in Assumption 3(iii),

$$\begin{aligned} \|w_{ij, t} - E(w_{ij, t} | \mathcal{T}_{t, T}(s))\|_2 &= \|h(z_{it}, z_{jt}) - E(h(z_{it}, z_{jt}) | \mathcal{T}_{t, T}(s))\|_2 \\ &\leq \|h(z_{it}, z_{jt}) - h(E(z_{it} | \mathcal{T}_{t, T}(s)), E(z_{jt} | \mathcal{T}_{t, T}(s)))\|_2 \\ &\leq c_0 [\|z_{it} - E(z_{it} | \mathcal{T}_{t, T}(s))\|_2 + \|z_{jt} - E(z_{jt} | \mathcal{T}_{t, T}(s))\|_2] \leq |r_0|^s C. \end{aligned}$$

Thus,

$$\begin{aligned} \|W_{nt, ij}^2 - E(W_{nt, ij}^2 | \mathcal{T}_{t, T}(s))\|_2 &\leq \sum_{k \in B_i(\rho_c)} \|w_{ik, t} w_{kj, t} \\ &\quad - E(w_{kj, t} | \mathcal{T}_{t, T}(s)) E(w_{ik, t} | \mathcal{T}_{t, T}(s))\|_2 \\ &\leq \sum_{k \in B_i(\rho_c)} (c_w \|w_{ik, t} - E(w_{ik, t} | \mathcal{T}_{t, T}(s))\|_2 + c_w \|w_{kj, t} \\ &\quad - E(w_{kj, t} | \mathcal{T}_{t, T}(s))\|_2) = 2c_w (\rho_c)^d |r_0|^s C_1. \end{aligned}$$

Similarly, we have

$$\begin{aligned} A_m &\equiv \sup_{i, j, n, t} \|W_{nt, ij}^m - E(W_{nt, ij}^m | \mathcal{T}_{t, T}(s))\|_2 \\ &\leq \sup_{i, j, n, t} \sum_{k \in B_i(m\rho_c)} \|W_{nt, ik}^{m-1} w_{nt, kj} \\ &\quad - E(W_{nt, ik}^{m-1} | \mathcal{T}_{t, T}(s)) E(w_{nt, kj} | \mathcal{T}_{t, T}(s))\|_2 \\ &\leq \sup_{i, j, n, t} \sum_{k \in B_i(m\rho_c)} (c_w^{m-1} \|w_{nt, kj} - E(w_{nt, kj} | \mathcal{T}_{t, T}(s))\|_2 + \|W_{nt, ik}^{m-1} \\ &\quad - E(W_{nt, ik}^{m-1} | \mathcal{T}_{t, T}(s))\|_2 \cdot c_w) \\ &\leq (m\rho_c)^d c_w^{m-1} |r_0|^s C_2 + (m\rho_c)^d c_w \sup_{i, k, n, t} \|W_{nt, ik}^{m-1} \\ &\quad - E(W_{nt, ik}^{m-1} | \mathcal{T}_{t, T}(s))\|_2 \leq m^d c_w^{m-1} |r_0|^s C + m^d C A_{m-1}. \end{aligned}$$

By deduction, for any finite  $m$ , we have

$$A_m \equiv \sup_{i, j, n, t} \|W_{nt, ij}^m - E(W_{nt, ij}^m | \mathcal{T}_{t, T}(s))\|_2 \leq C_m |r_0|^s.$$

Now we show the  $L_2$ -NED of  $D_{t, m, ij}$ . As  $D_{t, m} = \prod_{l=1}^m C_{s_{ml}+t-1}^{(p_l, q_l)}$  with  $\sum_{k=1}^m (p_k + q_k) = m$ ,  $D_{t, m}$  is a product of  $m$  terms of  $W_{nt}$ 's

and  $B_{nt}$ 's. Since  $B_{nt}$ 's have the same sparse structure as  $W_{nt}$ 's,  $D_{t,m}$  has the same NED property as  $W_{nt}^m$ , i.e.,  $\sup_{i,j,n,t} \|D_{t,m,ij} - E(D_{t,m,ij}|\mathcal{T}_{t,T}(s))\|_2 \leq C_m |r_0|^s$ .

**Proof for Lemma 5.** Denote

$$\frac{1}{L} \zeta'_{1L} W_L'^a \left( I_T \otimes \frac{1}{n} (l_n l_n') \right) W_L^b \zeta_{2L} = \frac{1}{T} \sum_{t=1}^T R_t, \text{ where}$$

$$R_t = \left[ \frac{1}{n} (e'_{t,T} \otimes l_n') W_L^a \zeta_{1L} \right]' \left[ \frac{1}{n} (e'_{t,T} \otimes l_n') W_L^b \zeta_{2L} \right].$$

The  $R_t$  can be expressed as  $2^{a+b}$  terms of

$$\frac{1}{n^2} \sum_{i_1=1}^n \sum_{j_1 \in B_{i_1}(a\rho_c)} D_{t-s_{a1}, a, i_1 j_1} \zeta_{1j_1, t-s_{a1}} \\ \times \sum_{i_2=1}^n \sum_{j_2 \in B_{i_2}(b\rho_c)} D_{t-s_{b1}, b, i_2 j_2} \zeta_{2j_2, t-s_{b1}}.$$

For  $\zeta_{1j,t}$ , it has the property  $\sup_{j,t} \|\zeta_{1j,t}\|_4 < \infty$  and the time NED as  $\sup_{j,t} \|\zeta_{1j,t} - E(\zeta_{1j,t}|\mathcal{T}_{t,T}(s))\|_4 \leq C|r_0|^s$ . The same for  $\zeta_{2j,t}$ . As the  $\sigma$ -field  $\mathcal{T}_{t,T}(s + s_{a1} + s_{b1})$  covers both  $\mathcal{T}_{t-s_{a1},T}(s)$  and  $\mathcal{T}_{t-s_{b1},T}(s)$ , we have

$$\begin{aligned} & \|\zeta_{1j_1, t-s_{a1}} \zeta_{2j_2, t-s_{b1}} - E(\zeta_{1j_1, t-s_{a1}} \zeta_{2j_2, t-s_{b1}} | \mathcal{T}_{t,T}(s + s_{a1} + s_{b1}))\|_2 \\ & \leq \|\zeta_{1j_1, t-s_{a1}} \zeta_{2j_2, t-s_{b1}} - E(\zeta_{1j_1, t-s_{a1}} | \mathcal{T}_{t-s_{a1},T}(s))\|_2 \\ & \quad \times E(\zeta_{2j_2, t-s_{b1}} | \mathcal{T}_{t-s_{b1},T}(s))\|_2 \\ & \leq \sup_{j,t} \|\zeta_{1j,t}\|_4 \cdot \sup_{j,t} \|\zeta_{2j,t} - E(\zeta_{2j,t} | \mathcal{T}_{t,T}(s))\|_4 \\ & \quad + \sup_{j,t} \|E(\zeta_{2j,t} | \mathcal{T}_{t,T}(s))\|_4 \cdot \sup_{j,t} \|\zeta_{1j,t} - E(\zeta_{1j,t} | \mathcal{T}_{t,T}(s))\|_4 \\ & \leq C_1 |r_0|^s. \end{aligned}$$

Since  $0 \leq s_{a1} \leq a$ ,  $0 \leq s_{b1} \leq b$ , where  $a$  and  $b$  are finite integers, we have

$$\begin{aligned} & \|\zeta_{1j_1, t-s_{a1}} \zeta_{2j_2, t-s_{b1}} - E(\zeta_{1j_1, t-s_{a1}} \zeta_{2j_2, t-s_{b1}} | \mathcal{T}_{t,T}(s))\|_2 \\ & \leq C_1 |r_0|^{s-s_{a1}-s_{b1}} \leq C_2 |r_0|^s. \end{aligned}$$

From Lemma 4 in Xu and Lee (2014), if  $\sup_{i,n} \|Y_{i,n}\|_{2r} \leq \Delta < \infty$  and  $\sup_{i,n} \|Z_{i,n}\|_{2r} \leq \Delta < \infty$  for some  $r > 2$ ,  $\|Y_{i,n} - E(Y_{i,n} | \mathcal{F}_{i,n}(s))\|_2 \leq d_{i,Y_n} \rho^s$  and  $\|Z_{i,n} - E(Z_{i,n} | \mathcal{F}_{i,n}(s))\|_2 \leq d_{i,Z_n} \rho^s$ , then the product  $Y_{i,n} Z_{i,n}$  has the NED property

$$\|Y_{i,n} Z_{i,n} - E(Y_{i,n} Z_{i,n} | \mathcal{F}_{i,n}(s))\|_2 \leq d_{i,n} \tilde{\rho}^s,$$

where  $d_{i,n} = 2^{(3r-2)/(r-1)} (d_{i,Y_n} + d_{i,Z_n})^{(r-2)/(2r-2)} \Delta^{(3r-2)/(2r-2)}$  and  $\tilde{\rho} = \rho^{(r-2)/(2r-2)}$ .

As  $\sup_{t,n,i,j} |D_{t,a,ij}| < \infty$ , we have

$$\begin{aligned} & \|D_{t-s_{a1}, a, i_1 j_1} D_{t-s_{b1}, b, i_2 j_2} - E(D_{t-s_{a1}, a, i_1 j_1} D_{t-s_{b1}, b, i_2 j_2} \\ & \quad \times |\mathcal{T}_{t,T}(s + s_{a1} + s_{b1}))\|_2 \leq C_{1ab} |r_0|^{1/2} |r_0|^s \end{aligned}$$

and hence,

$$\begin{aligned} & \|D_{t-s_{a1}, a, i_1 j_1} D_{t-s_{b1}, b, i_2 j_2} - E(D_{t-s_{a1}, a, i_1 j_1} D_{t-s_{b1}, b, i_2 j_2} | \mathcal{T}_{t,T}(s))\|_2 \\ & \leq C_{2ab} |r_0|^{1/2} |r_0|^s. \end{aligned}$$

Together with  $\max(\|\zeta_{1i,t}\|_{4+\delta}, \|\zeta_{2i,t}\|_{4+\delta}) \leq \Delta < \infty$ , we have

$$\begin{aligned} & \sup_{i_1, i_2, j_1, j_2, t} \|D_{t-s_{a1}, a, i_1 j_1} D_{t-s_{b1}, b, i_2 j_2} \zeta_{1j_1, t-s_{a1}} \zeta_{2j_2, t-s_{b1}} \\ & \quad - E(D_{t-s_{a1}, a, i_1 j_1} D_{t-s_{b1}, b, i_2 j_2} \zeta_{1j_1, t-s_{a1}} \zeta_{2j_2, t-s_{b1}} | \mathcal{T}_{t,T}(s))\|_2 \\ & \leq C_{3ab} \tilde{r}_0^s, \end{aligned}$$

where  $\tilde{r}_0 = r_0^{\delta/(4+2\delta)}$ . Hence,

$$\left\| \frac{1}{n^2} \sum_{i_1=1}^n \sum_{j_1 \in B_{i_1}(a\rho_c)} D_{t-s_{a1}, a, i_1 j_1} \zeta_{1j_1, t-s_{a1}} \sum_{i_2=1}^n \sum_{j_2 \in B_{i_2}(b\rho_c)} D_{t-s_{b1}, b, i_2 j_2} \zeta_{2j_2, t-s_{b1}} \right\|_2$$

$$\begin{aligned} & \times \sum_{j_2 \in B_{i_2}(b\rho_c)} D_{t-s_{b1}, b, i_2 j_2} \zeta_{2j_2, t-s_{b1}} \\ & - E \left( \frac{1}{n^2} \sum_{i_1=1}^n \sum_{j_1 \in B_{i_1}(a\rho_c)} D_{t-s_{a1}, a, i_1 j_1} \zeta_{1j_1, t-s_{a1}} \sum_{i_2=1}^n \sum_{j_2 \in B_{i_2}(b\rho_c)} D_{t-s_{b1}, b, i_2 j_2} \zeta_{2j_2, t-s_{b1}} | \mathcal{T}_{t,T}(s) \right) \Big\|_2 \\ & \leq C |r_0|^{\delta/(4+2\delta)} |r_0|^s. \end{aligned}$$

To show  $\frac{1}{T} \sum_{t=1}^T R_t - \frac{1}{T} \sum_{t=1}^T E(R_t) = o_p(1)$ , we will show (i)  $\frac{1}{T} \sum_{t=1}^T R_t - \frac{1}{T} \sum_{t=1}^T E(R_t | \mathcal{T}_{t,T}(s)) = o_p(1)$  and (ii)  $\frac{1}{T} \sum_{t=1}^T E(R_t | \mathcal{T}_{t,T}(s)) - \frac{1}{T} \sum_{t=1}^T E(R_t) = o_p(1)$ . As we have already shown, the NED property holds for  $R_t^{(m)}$ , for  $m = 1, \dots, 2^{a+b}$ ,

$$\left\| \frac{1}{T} \sum_{t=1}^T R_t - \frac{1}{T} \sum_{t=1}^T E(R_t | \mathcal{T}_{t,T}(s)) \right\|_2 \leq C_{ab} |r_0|^{\delta/(4+2\delta)} |r_0|^s.$$

Denote  $\tilde{R}_t = E(R_t | \mathcal{T}_{t,T}(s))$ . Then  $\tilde{R}_{t_1}$  and  $\tilde{R}_{t_2}$  are independent when  $|t_1 - t_2| > 2s$ . Hence,

$$\begin{aligned} & E \left| \frac{1}{T} \sum_{t=1}^T [E(R_t | \mathcal{T}_{t,T}(s)) - \frac{1}{T} \sum_{t=1}^T E(R_t)] \right|^2 \\ & = \frac{1}{T^2} \sum_{t_1=1}^T \sum_{t_2=1}^T \text{Cov}(\tilde{R}_{t_1}, \tilde{R}_{t_2}) \leq \frac{1}{T^2} \sum_{t_1=1}^T \sum_{|t_1-t_2| \leq 2s} \text{Cov}(\tilde{R}_{t_1}, \tilde{R}_{t_2}) \\ & \leq \frac{1}{T^2} \sum_{t_1=1}^T \sum_{|t_1-t_2| \leq 2s} \sup_t E |\tilde{R}_t|^2 \leq \frac{4s}{T} \sup_t E(R_t^2) \leq \frac{s}{T} C. \end{aligned}$$

The last inequality holds because  $R_t$  can be expressed as  $2^{a+b}$  terms of

$$\begin{aligned} & \frac{1}{n^2} \sum_{i_1=1}^n \sum_{j_1 \in B_{i_1}(a\rho_c)} D_{t-s_{a1}, a, i_1 j_1} \zeta_{1j_1, t-s_{a1}} \sum_{i_2=1}^n \sum_{j_2 \in B_{i_2}(b\rho_c)} D_{t-s_{b1}, b, i_2 j_2} \zeta_{2j_2, t-s_{b1}} \end{aligned}$$

and

$$\begin{aligned} & \sup_{t_1, t_2, n} E \left| \frac{1}{n^2} \sum_{i_1=1}^n \sum_{j_1 \in B_{i_1}(a\rho_c)} D_{t_1, a, i_1 j_1} \zeta_{1j_1, t_1} \sum_{i_2=1}^n \sum_{j_2 \in B_{i_2}(b\rho_c)} D_{t_2, b, i_2 j_2} \zeta_{2j_2, t_2} \right|^2 \\ & \leq (a\rho_c)^{2d} (b\rho_c)^{2d} \sup_{i,j,t} |D_{t,a,ij} D_{t,b,ij}|^2 \sqrt{\sup_{i,t} E |\zeta_{1i,t}|^4 \cdot \sup_{i,t} E |\zeta_{2i,t}|^4} \\ & \leq C_0, \text{ where } C_0 \text{ does not depend on } T \text{ or } n. \end{aligned}$$

By Chebyshev's inequality,

$$\left| \frac{1}{T} \sum_{t=1}^T \left[ E(R_t | \mathcal{T}_{t,T}(s)) - \frac{1}{T} \sum_{t=1}^T E(R_t) \right] \right| = O_p \left( \sqrt{\frac{s}{T}} \right).$$

Together, we have

$$\begin{aligned} & \left| \frac{1}{T} \sum_{t=1}^T (R_t - E(R_t)) \right| \leq \left| \frac{1}{T} \sum_{t=1}^T (R_t - E(R_t | \mathcal{T}_{t,T}(s))) \right| \\ & \quad + \left| \frac{1}{T} \sum_{t=1}^T [E(R_t | \mathcal{T}_{t,T}(s)) - E(R_t)] \right| = O_p \left( |r_0|^{\delta/(4+2\delta)} |r_0|^s + \sqrt{s/T} \right). \end{aligned}$$

If we choose  $s = T^\alpha$  with  $0 < \alpha < 1$ , say  $\alpha = 1/2$ , then  $\left| \frac{1}{T} \sum_{t=1}^T [R_t - E(R_t)] \right| = O_p(T^{-1/4})$ . Therefore, for any finite  $a$



and  $b$ ,

$$\begin{aligned} & \frac{1}{L} \zeta'_{1L} W_L'^a \left( I_T \otimes \frac{1}{n} (I_n I_n') \right) W_L^b \zeta_{2L} \\ & - E \left[ \frac{1}{L} \zeta'_{1L} W_L'^a \left( I_T \otimes \frac{1}{n} (I_n I_n') \right) W_L^b \zeta_{2L} \right] = o_p(1). \end{aligned}$$

Next we consider

$$\begin{aligned} & \frac{1}{L} \zeta'_{1L} (I_L - W_L')^{-1} \left( I_T \otimes \frac{1}{n} (I_n I_n') \right) (I_L - W_L)^{-1} \zeta_{2L} \\ & = \frac{1}{L} \zeta'_{1L} \left( \sum_{a=0}^A W_L'^a + \sum_{a=A+1}^{\infty} W_L'^a \right) \left( \frac{1}{n} I_T \otimes (I_n I_n') \right) \\ & \quad \times \left( \sum_{b=0}^B W_L^b + \sum_{b=B+1}^{\infty} W_L^b \right) \zeta_{2L} \\ & = \frac{1}{L} \zeta'_{1L} \sum_{a=0}^A W_L'^a \left( \frac{1}{n} I_T \otimes (I_n I_n') \right) \sum_{b=0}^B W_L^b \zeta_{2L} \\ & \quad + \frac{1}{L} \sum_{a=A+1}^{\infty} \zeta'_{1L} W_L'^a \left( \frac{1}{n} I_T \otimes (I_n I_n') \right) \sum_{b=0}^B W_L^b \zeta_{2L} \\ & \quad + \frac{1}{L} \zeta'_{1L} \sum_{a=0}^{\infty} W_L'^a \left( \frac{1}{n} I_T \otimes (I_n I_n') \right) \sum_{b=B+1}^{\infty} W_L^b \zeta_{2L} \\ & = H_{1L}^* + H_{2L}^* + H_{3L}^*. \end{aligned}$$

We have shown that, for any finite  $A$  and  $B$ , and any  $\delta > 0$ ,  $P(|H_{1L}^* - E(H_{1L}^*)| > \frac{\delta}{3}) = O(T^{-1/4})$ . As

$$\begin{aligned} & E \left| \frac{1}{T} \sum_{t=1}^T \frac{1}{n^2} \sum_{i_1=1}^n \sum_{j_1 \in B_{i_1}(a\rho_c)} D_{t-s_{a1}, a, i_1 j_1} \zeta_{1j_1, t-s_{a1}} \right. \\ & \quad \times \left. \sum_{i_2=1}^n \sum_{j_2 \in B_{i_2}(b\rho_c)} D_{t-s_{b1}, b, i_2 j_2} \zeta_{2j_2, t-s_{b1}} \right| \\ & \leq a^d b^d \rho_c^{2d} \sup_{n,t} \|W_{nt}\|_{\infty}^{a+b-r_1-r_2} \cdot \sup_t \|B_t\|_{\infty}^{r_1+r_2} \\ & \quad \times \sqrt{\sup_{i,t} E|\zeta_{1i,t}^2| \cdot \sup_{i,t} E|\zeta_{2i,t}^2|} \\ & \leq (ab)^d c_w^{a+b-r_1-r_2} (|\rho_0| + |\gamma_0| c_w)^{r_1+r_2} C_1, \end{aligned}$$

where  $C_1$  is a constant independent of  $T$  or  $n$ ,  $r_1$  and  $r_2$  are the same as the proof of Lemma 3. Hence,

$$\begin{aligned} & E \left| \frac{1}{L} \zeta'_{1L} W_L'^a \left( \frac{1}{n} I_T \otimes (I_n I_n') \right) W_L^b \zeta_{2L} \right| \\ & \leq (ab)^d C_1 \sum_{r_1=0}^a \binom{a}{r_1} \sum_{r_2=0}^b \binom{b}{r_2} |\lambda_0 c_w|^{a+b-r_1-r_2} (|\rho_0| + |\gamma_0| c_w)^{r_1+r_2} \\ & = C_1 (ab)^d (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^{a+b}. \end{aligned}$$

Using similar arguments in the Proof of Lemma 4, we can show that, for any  $1 > \delta > 0$ ,

$$\begin{aligned} & P(|H_{1L}^* - E(H_{1L}^*) + H_{2L}^* - E(H_{2L}^*) + H_{3L}^* - E(H_{3L}^*)| > \delta) \\ & \leq P \left( |H_{1L}^* - E(H_{1L}^*)| > \frac{\delta}{3} \right) + P \left( |H_{2L}^*| > \frac{\delta}{6} \right) \\ & \quad + P \left( |H_{2L}^*| > \frac{\delta}{6} \right) + P \left( |H_{3L}^*| > \frac{\delta}{6} \right) \\ & \quad + P \left( |H_{3L}^*| > \frac{\delta}{6} \right) \leq O(T^{-1/4}) + 2\delta. \end{aligned}$$

Hence,

$$\begin{aligned} & \frac{1}{L} \zeta'_{1L} (I_L - W_L')^{-1} \left( \frac{1}{T} (I_T I_T') \otimes I_n \right) (I_L - W_L)^{-1} \zeta_{2L} \\ & - E \left[ \frac{1}{L} \zeta'_{1L} \left( \frac{1}{T} (I_T I_T') \otimes I_n \right) (I_L - W_L)^{-1} \zeta_{2L} \right] = o_p(1). \end{aligned}$$

**Proof for Lemma 6.** From earlier results in Eq. (16),

$$\begin{aligned} & \frac{1}{L} l_L' W_L^m \zeta_{1L} - E \left( \frac{1}{L} l_L' W_L^m \zeta_{1L} \right) = O_p(1/\sqrt{n}); \\ & E \left( \frac{1}{L} l_L' W_L^a \zeta_{1L} \cdot \frac{1}{L} l_L' W_L^b \zeta_{2L} \right) - E \left( \frac{1}{L} l_L' W_L^a \zeta_{1L} \right) E \left( \frac{1}{L} l_L' W_L^b \zeta_{2L} \right) \\ & = \text{Cov} \left( \frac{1}{L} l_L' W_L^a \zeta_{1L}, \frac{1}{L} l_L' W_L^b \zeta_{2L} \right) = O \left( \frac{1}{n} \right); \end{aligned}$$

and result in Eq. (17),  $E \left| \frac{1}{L} l_L' W_L^a \zeta_{1L} \right| = O(1)$  and  $\frac{1}{L} l_L' W_L^b \zeta_{2L} = O_p(1)$ . Therefore, for any finite  $a$  and  $b$ , not depending on  $n$  and  $T$ , we have

$$\begin{aligned} & \frac{1}{L} l_L' W_L^a \zeta_{1L} \cdot \frac{1}{L} l_L' W_L^b \zeta_{2L} - E \left( \frac{1}{L} l_L' W_L^a \zeta_{1L} \cdot \frac{1}{L} l_L' W_L^b \zeta_{2L} \right) \\ & = \frac{1}{L} l_L' W_L^a \zeta_{1L} \cdot \frac{1}{L} l_L' W_L^b \zeta_{2L} - E \left( \frac{1}{L} l_L' W_L^a \zeta_{1L} \right) E \left( \frac{1}{L} l_L' W_L^b \zeta_{2L} \right) \\ & \quad + O \left( \frac{1}{n} \right) \\ & = \left( \frac{1}{L} l_L' W_L^a \zeta_{1L} - E \left( \frac{1}{L} l_L' W_L^a \zeta_{1L} \right) \right) \frac{1}{L} l_L' W_L^b \zeta_{2L} \\ & \quad + \left( \frac{1}{L} l_L' W_L^b \zeta_{2L} - E \left( \frac{1}{L} l_L' W_L^b \zeta_{2L} \right) \right) E \left( \frac{1}{L} l_L' W_L^a \zeta_{1L} \right) + O \left( \frac{1}{n} \right) \\ & = O_p \left( \frac{1}{\sqrt{n}} \right). \end{aligned}$$

For any integers  $a$  and  $b$ , similar to Eq. (18),

$$\begin{aligned} & E \left| \frac{1}{L} l_L' W_L^a \zeta_{1L} \cdot \frac{1}{L} l_L' W_L^b \zeta_{2L} \right| \leq \frac{1}{n^2 T^2} \sum_{t=1}^{T-s_{a1}} \sum_{t'=1}^{T-s_{b1}} \sum_{i'=1}^n \sum_{j' \in B_{i'}(a\rho_c)} \sum_{i=1}^n \\ & \quad \times \sum_{j \in B_i(b\rho_c)} E |D_{t,a,ij} \zeta_{1i,t} \zeta_{2j,t'} D_{t',b,i'j'}| \\ & \leq \sum_{j' \in B_{i'}(a\rho_c)} \sum_{j \in B_i(b\rho_c)} \sup_{t,i,j} |D_{t,a,ij}| \sup_{t,i,j} |D_{t,b,ij}| \\ & \quad \times \sqrt{\sup_{i,t} E(\zeta_{1i,t}^2) \sup_{i,t} E(\zeta_{2i,t}^2)} \\ & \leq C(ab)^d (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^{a+b}. \end{aligned}$$

Hence,

$$\begin{aligned} & E \left| \sum_{a=M_1+1}^{\infty} \sum_{b=M_2+1}^{\infty} \frac{1}{L} l_L' W_L^a \zeta_{1L} \cdot \frac{1}{L} l_L' W_L^b \zeta_{2L} \right| \\ & \leq C_2 M_1^d M_2^d (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^{M_1+M_2} \end{aligned}$$

and

$$\begin{aligned} & E \left| \sum_{a=M_1+1}^{\infty} \frac{1}{L} l_L' W_L^a \zeta_{1L} \cdot \sum_{b=0}^{M_2} \frac{1}{L} l_L' W_L^b \zeta_{2L} \right| \\ & \leq C_2 M_1^d M_2^d (|\lambda_0| c_w + |\rho_0| + |\gamma_0| c_w)^{M_1}. \end{aligned}$$

Thus, using similar arguments in the proof of Lemma 4, we can decompose  $\frac{1}{L^2} l_L' (I_L - W_L)^{-1} \zeta_{1L} l_L' (I_L - W_L)^{-1} \zeta_{2L}$  into three terms and show that

$$\frac{1}{L^2} l_L' (I_L - W_L)^{-1} \zeta_{1L} l_L' (I_L - W_L)^{-1} \zeta_{2L}$$

$$-E\left[\frac{1}{L^2}l'_l(I_L - W_L)^{-1} \varsigma_{1L}l'_l(I_L - W_L)^{-1} \varsigma_{2L}\right] = o_p(1).$$

This finishes the proof.

**Proof for Proposition 1.** See Lemmas 3–6.

**Proof for Corollary 1.** From Proposition 1 and compactness of  $\Theta$ , similar to Qu and Lee (2015), we obtain the results.

**Proof for Proposition 2.** We can rewrite  $U_L$  in the following way<sup>21</sup>:

$$\begin{aligned} U_L &= \frac{1}{\sqrt{L}} \sum_{l=1}^L \left( a_l \varsigma_l + b_l \xi_l + \xi_l \sum_{j=1}^L c_{lj} \xi_j - E\left(\xi_l \sum_{j=1}^L c_{lj} \xi_j | \varsigma_L\right) \right. \\ &\quad \left. + \varsigma_l \sum_{j=1}^L d_{lj} \varsigma_j - E\left(\varsigma_l \sum_{j=1}^L d_{lj} \varsigma_j\right) + \xi_l \sum_{j=1}^L e_{lj} \varsigma_j \right) \\ &= \frac{1}{\sqrt{L}} \sum_{l=1}^L \left( a_l \varsigma_l + b_l \xi_l + 2\xi_l \sum_{j=1}^{l-1} c_{lj} \xi_j + c_{ll}(\xi_l^2 - \sigma_\xi^2) \right. \\ &\quad \left. + 2\varsigma_l \sum_{j=1}^{l-1} d_{lj} \varsigma_j + d_{ll}(\varsigma_l^2 - \sigma_\varsigma^2) + \xi_l \sum_{j=1}^L e_{lj} \varsigma_j \right). \end{aligned}$$

Let  $U_L = \frac{1}{\sqrt{L}} \sum_{l=1}^{2L} U_{L,l}$ , then we can define  $U_{L,l}$  and the corresponding  $\sigma$ -field  $\mathcal{F}_{L,l}$  with following expressions:

$$\begin{aligned} U_{L,1} &= \frac{1}{\sqrt{L}} [a_1 \varsigma_1 + d_{11}(\varsigma_1^2 - \sigma_\varsigma^2)], \quad \mathcal{F}_{L,1} = \langle \varsigma_1 \rangle; \\ U_{L,2} &= \frac{1}{\sqrt{L}} [a_2 \varsigma_2 + 2\varsigma_2 d_{21} \varsigma_1 + d_{22}(\varsigma_2^2 - \sigma_\varsigma^2)], \quad \mathcal{F}_{L,2} = \langle \varsigma_1, \varsigma_2 \rangle; \\ &\vdots \\ U_{L,L} &= \frac{1}{\sqrt{L}} \left[ a_L \varsigma_L + 2\varsigma_L \sum_{j=1}^{L-1} d_{Lj} \varsigma_j + d_{LL}(\varsigma_L^2 - \sigma_\varsigma^2) \right], \\ \mathcal{F}_{L,L} &= \langle \varsigma_1, \dots, \varsigma_L \rangle; \\ U_{L,L+1} &= \frac{1}{\sqrt{L}} \left[ b_1 \xi_1 + \xi_1 \sum_{j=1}^L e_{1j} \varsigma_j + c_{11}(\xi_1^2 - \sigma_\xi^2) \right], \\ \mathcal{F}_{L,L+1} &= \langle \varsigma_L, \xi_1 \rangle; \\ U_{L,L+2} &= \frac{1}{\sqrt{L}} \left[ b_2 \xi_2 + \xi_2 \sum_{j=1}^L e_{2j} \varsigma_j + 2\xi_2 c_{21} \xi_1 + c_{22}(\xi_2^2 - \sigma_\xi^2) \right], \\ \mathcal{F}_{L,L+2} &= \langle \varsigma_L, \xi_1, \xi_2 \rangle; \\ &\vdots \\ U_{L,2L} &= \frac{1}{\sqrt{L}} \left[ b_L \xi_L + \xi_L \sum_{j=1}^L e_{Lj} \varsigma_j + 2\xi_L \sum_{j=1}^{L-1} c_{Lj} \xi_j + c_{LL}(\xi_L^2 - \sigma_\xi^2) \right], \\ \mathcal{F}_{L,2L} &= \langle \varsigma_L, \xi_1, \dots, \xi_L \rangle. \end{aligned}$$

Because  $(\varsigma_l, \xi_l)$ 's are independent and homoscedasticity across  $L$ , and  $\xi_l$ 's are mean independent and homoscedasticity conditional on  $\varsigma_l$ 's, we have  $E(U_{L,l} | \mathcal{F}_{L,l-1}) = 0$ , for  $l = 2, \dots, 2L$ . Then  $\{(U_{n,l}, \mathcal{F}_{n,l}) | 1 \leq l \leq 2L\}$  forms a martingale difference double array. We have  $\text{Var}(\sum_{l=1}^{2L} U_{L,l}) = \sum_{l=1}^{2L} E U_{L,l}^2$  as  $U_{L,l}$ 's are martingale

differences. Define this variance as  $\sigma_{U_L}^2$  and the normalized variables  $U_{L,l}^* = U_{L,l}/\sigma_{U_L}$ . In order for the martingale central limit theorem to be applicable, we shall show that there exists a  $\delta_U > 0$  such that

$$\sum_{l=1}^{2L} E|U_{L,l}^*|^{2+\delta_U} \rightarrow 0 \text{ as } L \rightarrow \infty \text{ and } \sum_{l=1}^{2L} E(U_{L,l}^{*2} | \mathcal{F}_{L,l-1}) - 1 \xrightarrow{p} 0.$$

Denote  $\sigma_{U_{1,L}}^2$  and  $\sigma_{U_{2,L}}^2$  as the variance of  $U_{d1,L}$  and  $U_{d2,L}$ , where

$$\begin{aligned} U_{d1,L} &= \frac{1}{\sqrt{L}} A'_L \varsigma_L + \varsigma'_L D_L \varsigma_L - E(\varsigma'_L D_L \varsigma_L) \text{ and} \\ U_{d2,L} &= \frac{1}{\sqrt{L}} B'_L \xi_L + \xi'_L C_L \xi_L + \xi'_L E_L \varsigma_L - E(\xi'_L C_L \xi_L | \varsigma_L), \end{aligned}$$

then  $U_L = U_{d1,L} + U_{d2,L}$  and  $\sigma_{U_L}^2 = \sigma_{U_{1,L}}^2 + \sigma_{U_{2,L}}^2$ . The expressions are

$$\begin{aligned} \sigma_{U_{1,L}}^2 &= \frac{\sigma_\varsigma^2}{L} A'_L A_L + \frac{\sigma_\varsigma^4}{L} \text{tr}(D_L^2 + D_L D'_L) + [E(\varsigma_L^4) - 3\sigma_\varsigma^4] \frac{1}{L} \sum_{l=1}^L d_{ll}^2 \\ &\quad + \frac{2}{L} E(\varsigma_L^3) \sum_{l=1}^L a_l d_{ll} \end{aligned}$$

and

$$\begin{aligned} \sigma_{U_{2,L}}^2 &= \frac{\sigma_\xi^2}{L} E(B'_L B_L) + \frac{\sigma_\xi^4}{L} \text{tr}(C_L^2 + C_L C'_L) + \frac{\sigma_\xi^2}{L} E(\varsigma'_L E'_L E_L \varsigma_L) \\ &\quad + \frac{2\sigma_\xi^2}{L} E(B'_L E_L \varsigma_L) + \frac{1}{L} \sum_{l=1}^L E[(c_{ll}^2 \xi_l^4) - 3c_{ll}^2 \sigma_\xi^4] \\ &\quad + \frac{2}{L} \sum_{l=1}^L E(\xi_l^3 b_l c_{ll}) + \frac{2}{L} \sum_{l=1}^L \sum_{j=1}^L E(\xi_l^3 c_{lj} e_{lj} \varsigma_j). \end{aligned}$$

As  $\sigma_\varsigma^2 \sum_{l=1}^L a_l^2 + \text{Var}(\varsigma_L^2) \sum_{l=1}^L d_{ll}^2 \geq 2\sqrt{\sigma_\varsigma^2 \text{Var}(\varsigma_L^2)} \sum_{l=1}^L |a_l d_{ll}| \geq 2|E(\varsigma_L^3)| \sum_{l=1}^L |a_l d_{ll}|$ , we have

$$\begin{aligned} \sigma_{U_{1,L}}^2 &= \frac{\sigma_\varsigma^2}{L} \sum_{l=1}^L a_l^2 + \frac{\sigma_\varsigma^4}{L} \sum_l \sum_{j \neq l} (d_{lj}^2 + d_{jl}^2) + \frac{\text{Var}(\varsigma_L^2)}{L} \sum_{l=1}^L d_{ll}^2 \\ &\quad + \frac{2}{L} E(\varsigma_L^3) \sum_{l=1}^L a_l d_{ll} \geq \frac{\sigma_\varsigma^4}{L} \sum_{l=1}^L \sum_{j \neq l} (d_{lj}^2 + d_{jl}^2). \end{aligned}$$

So a sufficient condition for  $\sigma_{U_{1,L}}^2$  to be bounded away from zero is that  $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L \sum_{j \neq l} (d_{lj}^2 + d_{jl}^2) > 0$ . Similarly, a sufficient condition for  $\sigma_{U_{2,L}}^2$  to be bounded away from zero is that  $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L \sum_{j \neq l} E(c_{lj}^2 + c_{jl}^2) > 0$ . Therefore, a sufficient condition for  $\sigma_{U_L}^2 > 0$  is that  $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L \sum_{j \neq l} (d_{lj}^2 + d_{jl}^2) > 0$  or  $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L \sum_{j \neq l} E(c_{lj}^2 + c_{jl}^2) > 0$ .

We know that  $E(B'_L B_L)$ ,  $E(C_L^2 + C_L C'_L)$ ,  $E(\varsigma'_L E'_L E_L \varsigma_L)$ , and  $E(B'_L E_L \varsigma_L)$  are  $O(L)$ . Therefore,  $\sigma_{U_{1,L}}^2$ ,  $\sigma_{U_{2,L}}^2$ , and  $\sigma_{U_L}^2$  are  $O(1)$ . For  $l = 1, \dots, L$ ,  $U_{L,l} = \frac{1}{\sqrt{L}} (a_l \varsigma_l + 2\varsigma_l \sum_{j=1}^{l-1} d_{lj} \varsigma_j + d_{ll}(\varsigma_l^2 - \sigma_\varsigma^2))$ . Since  $U_{L,l}/\sigma_{U_{1,L}}$  has the conventional linear quadratic form, the two conditions for the summation of first  $L$  terms are easy to check. We have  $\sum_{l=1}^L E|U_{L,l}^*|^{2+\delta_U} \rightarrow 0$  as  $L \rightarrow \infty$  and  $\sum_{l=1}^L E(U_{L,l}^2 | \mathcal{F}_{L,l-1})/\sigma_{U_L}^2 - 1 \xrightarrow{p} 0$ .

<sup>21</sup> Here we treat  $C_L$  and  $D_L$  as symmetric. If asymmetric, we can always replace  $C_L$  with  $(C_L + C'_L)/2$ . The proof of this CLT originated in an earlier working version of Qu and Lee (2015). The subsequent CLT used in the published version of Qu and Lee (2015) uses a CLT of an NED process. However, the latter CLT cannot be applied to the SDPD model with endogenous  $W_{it}$  due to the presence of the projectors  $J_n$  and  $J_T$  in certain statistics.

For  $l = L + 1, \dots, 2L$ ,  $U_{L,l} = \frac{1}{\sqrt{L}}[b_{l'}\xi_{l'} + 2\xi_{l'} \sum_{j=1}^{l'} c_{l'j}\xi_j + c_{l'l'}(\xi_{l'}^2 - \sigma_\xi^2) + \xi_{l'} \sum_{j=1}^L e_{l'j}\zeta_j]$  with  $l' = l - L$ , so

$$\begin{aligned} E(U_{L,l}^2 | \mathcal{F}_{L,l-1}) &= \frac{1}{L} E \left[ b_{l'}^2 \xi_{l'}^2 + 4\xi_{l'}^2 \sum_{j=1}^{l'-1} c_{l'j}\xi_j \sum_{k=1}^{l'-1} c_{l'k}\xi_k + c_{l'l'}^2 (\xi_{l'}^2 - \sigma_\xi^2)^2 \right. \\ &\quad + \xi_{l'}^2 \sum_{j=1}^L e_{l'j}\zeta_j \sum_{k=1}^L e_{l'k}\zeta_k + 2b_{l'}c_{l'l'}(\xi_{l'}^3 - \xi_{l'}\sigma_\xi^2) \\ &\quad + 4b_{l'}\xi_{l'}^2 \sum_{j=1}^{l'-1} c_{l'j}\xi_j + 2b_{l'}\xi_{l'}^2 \sum_{j=1}^L e_{l'j}\zeta_j + 4c_{l'l'}(\xi_{l'}^3 - \xi_{l'}\sigma_\xi^2) \sum_{j=1}^{l'-1} c_{l'j}\xi_j \\ &\quad \left. + 4\xi_{l'}^2 \sum_{j=1}^{l'-1} c_{l'j}\xi_j \sum_{k=1}^L e_{l'k}\zeta_k + 2c_{l'l'}(\xi_{l'}^3 - \xi_{l'}\sigma_\xi^2) \sum_{j=1}^L e_{l'j}\zeta_j \right] \\ &= \frac{1}{L} \left[ b_{l'}^2 \sigma_\xi^2 + 4\sigma_\xi^2 \sum_{j=1}^{l'-1} c_{l'j}\xi_j \sum_{k=1}^{l'-1} c_{l'k}\xi_k + c_{l'l'}^2 \text{Var}(\xi_{l'}^2) \right. \\ &\quad + \sigma_\xi^2 \sum_{j=1}^L e_{l'j}\zeta_j \sum_{k=1}^L e_{l'k}\zeta_k + 4b_{l'}\sigma_\xi^2 \sum_{j=1}^{l'-1} c_{l'j}\xi_j \\ &\quad + 2b_{l'}c_{l'l'}E(\xi_{l'}^3) + 2b_{l'}\sigma_\xi^2 \sum_{j=1}^L e_{l'j}\zeta_j + 4c_{l'l'}E(\xi_{l'}^3) \sum_{j=1}^{l'-1} c_{l'j}\xi_j \\ &\quad \left. + 4\sigma_\xi^2 \sum_{j=1}^{l'-1} c_{l'j}\xi_j \sum_{k=1}^L e_{l'k}\zeta_k + 2c_{l'l'}E(\xi_{l'}^3) \sum_{j=1}^L e_{l'j}\zeta_j \right]; \\ E(U_{L,l}^2) &= \frac{1}{L} \left[ E(b_{l'}^2)\sigma_\xi^2 + 4\sigma_\xi^4 \sum_{j=1}^{l'-1} E(c_{l'j}^2) + E(c_{l'l'}^2)\text{Var}(\xi_{l'}^2) \right. \\ &\quad + \sigma_\xi^2 \sum_{j=1}^L \sum_{k=1}^L E(e_{l'j}e_{l'k}\zeta_j\zeta_k) + 2E(b_{l'}c_{l'l'})E(\xi_{l'}^3) \\ &\quad \left. + 2\sigma_\xi^2 \sum_{j=1}^L E(b_{l'}e_{l'j}\zeta_j) + 2E(\xi_{l'}^3) \sum_{j=1}^L E(c_{l'l'}e_{l'j}\zeta_j) \right]. \end{aligned}$$

Therefore, for  $l = L + 1, \dots, 2L$  and  $l' = l - L$ ,

$$\begin{aligned} E(U_{L,l}^2 | \mathcal{F}_{L,l-1}) - E(U_{L,l}^2) &= \frac{1}{L} [\sigma_\xi^2(b_{l'}^2) - E(b_{l'}^2) \\ &\quad + 4\sigma_\xi^2 \sum_{j=1}^{l'-1} \left( c_{l'j}\xi_j \sum_{k=1}^{l'-1} c_{l'k}\xi_k - \sigma_\xi^2 E(c_{l'j}^2) \right) \\ &\quad + (c_{l'l'}^2 - E(c_{l'l'}^2))\text{Var}(\xi_{l'}^2) \\ &\quad + \sigma_\xi^2 \sum_{j=1}^L \sum_{k=1}^L (e_{l'j}\zeta_j e_{l'k}\zeta_k - E(e_{l'j}\zeta_j e_{l'k}\zeta_k)) \\ &\quad + 2E(\xi_{l'}^3)(b_{l'}c_{l'l'} - E(b_{l'}c_{l'l'})) + 2\sigma_\xi^2 \sum_{j=1}^L (b_{l'}e_{l'j}\zeta_j - E(b_{l'}e_{l'j}\zeta_j)) \\ &\quad + 4b_{l'}\sigma_\xi^2 \sum_{j=1}^{l'-1} c_{l'j}\xi_j + 4c_{l'l'}E(\xi_{l'}^3) \sum_{j=1}^{l'-1} c_{l'j}\xi_j \\ &\quad + 4\sigma_\xi^2 \sum_{j=1}^{l'-1} c_{l'j}\xi_j \sum_{k=1}^L e_{l'k}\zeta_k + 2E(\xi_{l'}^3) \sum_{j=1}^L (c_{l'l'}e_{l'j}\zeta_j - E(c_{l'l'}e_{l'j}\zeta_j))] \\ &= \frac{1}{L} \{ \sigma_\xi^2 [b_{l'}^2 - E(b_{l'}^2)] + \text{Var}(\xi_{l'}^2) [c_{l'l'}^2 - E(c_{l'l'}^2)] \\ &\quad + 2E(\xi_{l'}^3) [b_{l'}c_{l'l'} - E(b_{l'}c_{l'l'})] \\ &\quad + \frac{4}{L} \sum_{j=1}^{l'-1} \left[ \sigma_\xi^2 \left( c_{l'j}\xi_j \sum_{k=1}^{l'-1} c_{l'k}\xi_k - \sigma_\xi^2 E(c_{l'j}^2) \right) + b_{l'}\sigma_\xi^2 c_{l'j}\xi_j \right. \end{aligned}$$

$$\begin{aligned} &\quad \left. + c_{l'l'}E(\xi_{l'}^3)c_{l'j}\xi_j + \sigma_\xi^2 c_{l'j}\xi_j \sum_{k=1}^L e_{l'k}\zeta_k \right] \\ &\quad + \frac{1}{L} \sum_{j=1}^L \left[ \sigma_\xi^2 \sum_{k=1}^L (e_{l'j}\zeta_j e_{l'k}\zeta_k - E(e_{l'j}\zeta_j e_{l'k}\zeta_k)) + 2\sigma_\xi^2 (b_{l'}e_{l'j}\zeta_j \right. \\ &\quad \left. - E(b_{l'}e_{l'j}\zeta_j)) + 2E(\xi_{l'}^3)(c_{l'l'}e_{l'j}\zeta_j - E(c_{l'l'}e_{l'j}\zeta_j)) \right], \end{aligned}$$

and hence,

$$\begin{aligned} \sum_{l=L+1}^{2L} [E(U_{L,l}^2 | \mathcal{F}_{L,l-1}) - E(U_{L,l}^2)] &= H_{1L} + H_{2L} + H_{3L} + H_{4L} + H_{5L} \\ &\quad + H_{6L} + H_{7L} + H_{8L} + H_{9L} + H_{10L}, \end{aligned}$$

where

$$\begin{aligned} H_{1L} &= \frac{\sigma_\xi^2}{L} \sum_{l'=1}^L [b_{l'}^2 - E(b_{l'}^2)]; \quad H_{2L} = \frac{\text{Var}(\xi_{l'}^2)}{L} \sum_{l'=1}^L [c_{l'l'}^2 - E(c_{l'l'}^2)]; \\ H_{3L} &= \frac{2E(\xi_{l'}^3)}{L} \sum_{l'=1}^L [b_{l'}c_{l'l'} - E(b_{l'}c_{l'l'})]; \\ H_{4L} &= \frac{4\sigma_\xi^2}{L} \sum_{l'=1}^L \sum_{j=1}^{l'-1} \left( c_{l'j}\xi_j \sum_{k=1}^{l'-1} c_{l'k}\xi_k - \sigma_\xi^2 E(c_{l'j}^2) \right); \\ H_{5L} &= \frac{4\sigma_\xi^2}{L} \sum_{l'=1}^L \sum_{j=1}^{l'-1} b_{l'}c_{l'j}\xi_j; \\ H_{6L} &= \frac{4E(\xi_{l'}^3)}{L} \sum_{l'=1}^L \sum_{j=1}^{l'-1} c_{l'l'}c_{l'j}\xi_j; \\ H_{7L} &= \frac{4\sigma_\xi^2}{L} \sum_{l'=1}^L \sum_{j=1}^{l'-1} c_{l'j}\xi_j \sum_{k=1}^L e_{l'k}\zeta_k; \\ H_{8L} &= \frac{2\sigma_\xi^2}{L} \sum_{l'=1}^L \sum_{j=1}^L (b_{l'}e_{l'j}\zeta_j - E(b_{l'}e_{l'j}\zeta_j)); \\ H_{9L} &= \frac{2E(\xi_{l'}^3)}{L} \sum_{l'=1}^L \sum_{j=1}^L (c_{l'l'}e_{l'j}\zeta_j - E(c_{l'l'}e_{l'j}\zeta_j)); \\ H_{10L} &= \frac{\sigma_\xi^2}{L} \sum_{l'=1}^L \sum_{j=1}^L \sum_{k=1}^L (e_{l'j}\zeta_j e_{l'k}\zeta_k - E(e_{l'j}\zeta_j e_{l'k}\zeta_k)). \\ E(U_{L,l}^2) &= \frac{1}{L} \left[ E(b_{l'}^2)\sigma_\xi^2 + 4\sigma_\xi^4 \sum_{j=1}^{l'-1} E(c_{l'j}^2) + E(c_{l'l'}^2)\text{Var}(\xi_{l'}^2) \right. \\ &\quad + \sigma_\xi^2 \sum_{j=1}^L \sum_{k=1}^L E(e_{l'j}e_{l'k}\zeta_j\zeta_k) + 2E(b_{l'}c_{l'l'})E(\xi_{l'}^3) \\ &\quad \left. + 2\sigma_\xi^2 \sum_{j=1}^L E(b_{l'}e_{l'j}\zeta_j) + 2E(\xi_{l'}^3) \sum_{j=1}^L E(c_{l'l'}e_{l'j}\zeta_j) \right]. \end{aligned}$$

We need to show  $H_{jL} = o_p(1)$  for all  $j = 1, \dots, 10$ . As  $H_{1L} = \frac{\sigma_\xi^2}{L} \sum_{l'=1}^L [b_{l'}^2 - E(b_{l'}^2)] = \frac{\sigma_\xi^2}{L} (B_L' B_L - E(B_L' B_L))$  and  $H_{10L} = \frac{\sigma_\xi^2}{L} \sum_{l'=1}^L \sum_{j=1}^L \sum_{k=1}^L (e_{l'j}\zeta_j e_{l'k}\zeta_k - E(e_{l'j}\zeta_j e_{l'k}\zeta_k)) = \frac{\sigma_\xi^2}{L} [S_L' E_L' E_L S_L - E(S_L' E_L' E_L S_L)]$ , if  $B_L$ ,  $C_L$ , and  $E_L$  satisfy conditions in Assumption CLT, we have  $H_{10L} = o_p(1)$ . Similar arguments can be applied to  $H_{jL}$  for  $j = 2, \dots, 9$ . These imply that  $\sum_{l=L+1}^{2L} E(U_{L,l}^2 | \mathcal{F}_{L,l-1}) / \sigma_{U_{2,L}}^2 - 1 \xrightarrow{p} 0$ .

Together with  $\sum_{l=1}^L E(U_{L,l}^2 | \mathcal{F}_{L,l-1}) / \sigma_{U_{L,l}}^2 - 1 \xrightarrow{p} 0$ , we have

$$\sum_{l=1}^{2L} E(U_{L,l}^2 | \mathcal{F}_{L,l-1}) - (\sigma_{U_{L,l}}^2 + \sigma_{U_{2L,l}}^2) \xrightarrow{p} 0, \text{ i.e., } \sum_{l=1}^{2L} E(U_{L,l}^{*2} | \mathcal{F}_{L,l-1}) - 1 \xrightarrow{p} 0.$$

The next step is to show  $\sum_{l=L+1}^{2L} E|U_{L,l}|^{2+\delta_U} \rightarrow 0$  as  $L \rightarrow \infty$ . For any positive constants  $p$  and  $q$  such that  $1/p + 1/q = 1$ , we have

$$\begin{aligned} |U_{L,l}| &= \frac{1}{\sqrt{L}} |b_{l'} \xi_{l'} + 2\xi_{l'} \sum_{j=1}^{l'-1} c_{l'j} \xi_j + c_{l'l'} (\xi_{l'}^2 - \sigma_\xi^2) + \xi_{l'} \sum_j^L e_{l'j} \varsigma_j| \\ &\leq \frac{1}{\sqrt{L}} \left( |b_{l'}| |\xi_{l'}| + 2|\xi_{l'}| \sum_{j=1}^{l'-1} |c_{l'j}| |\xi_j| + |c_{l'l'}| |\xi_{l'}^2 - \sigma_\xi^2| \right. \\ &\quad \left. + |\xi_{l'}| \sum_j^L |e_{l'j}| |\varsigma_j| \right) \\ &= \frac{1}{\sqrt{L}} \left( |b_{l'}|^{\frac{1}{p}} |b_{l'}|^{\frac{1}{q}} |\xi_{l'}| + 2|\xi_{l'}| \sum_{j=1}^{l'-1} |c_{l'j}|^{\frac{1}{p}} |c_{l'j}|^{\frac{1}{q}} |\xi_j| \right. \\ &\quad \left. + |c_{l'l'}|^{\frac{1}{p}} |c_{l'l'}|^{\frac{1}{q}} |\xi_{l'}^2 - \sigma_\xi^2| + |\xi_{l'}| \sum_j^L |e_{l'j}|^{\frac{1}{p}} |e_{l'j}|^{\frac{1}{q}} |\varsigma_j| \right). \end{aligned}$$

By Holder's inequality, we get

$$\begin{aligned} |U_{L,l}|^q &\leq L^{-q/2} \left[ \left( (|b_{l'}|^{\frac{1}{p}})^p + \sum_{j=1}^{l'-1} (|c_{l'j}|^{\frac{1}{p}})^p + 1 + \sum_j^L (|e_{l'j}|^{\frac{1}{p}})^p \right)^{\frac{1}{p}} \right. \\ &\quad \times \left( (|b_{l'}|^{\frac{1}{q}} |\xi_{l'}|)^q + \sum_{j=1}^{l'-1} (2|\xi_{l'}| |c_{l'j}|^{\frac{1}{q}} |\xi_j|)^q \right. \\ &\quad \left. \left. + (|c_{l'l'}|^{\frac{1}{q}} |\xi_{l'}^2 - \sigma_\xi^2|)^q + \sum_j^L (|\xi_{l'}| |e_{l'j}|^{\frac{1}{q}} |\varsigma_j|)^q \right)^{\frac{1}{q}} \right] \\ &= L^{-\frac{q}{2}} \left( |b_{l'}| + \sum_{j=1}^{l'-1} |c_{l'j}| + 1 + \sum_j^L |e_{l'j}| \right)^{\frac{q}{p}} \left( |b_{l'}| |\xi_{l'}^q| \right. \\ &\quad \left. + 2^q \sum_{j=1}^{l'-1} |c_{l'j}| |\xi_{l'}^q| |\xi_j^q| + |c_{l'l'}| |\xi_{l'}^2 - \sigma_\xi^2|^q + |\xi_{l'}^q| \sum_j^L |e_{l'j}| |\varsigma_j^q| \right). \end{aligned}$$

As  $c_L$  and  $e_L$  are uniformly bounded in both row and column sums and elements of  $B_L$  are uniformly bounded, there exists a constant  $c_U$  such that  $|b_{l'}| \leq c_U/3$ ,  $\sum_{j=1}^{l'-1} |c_{l'j}| \leq c_U/3$ , and  $\sum_j^L |e_{l'j}| \leq c_U/3$  for all  $l'$  and  $L$ . Hence,

$$\begin{aligned} |U_{L,l}|^q &\leq L^{-\frac{q}{2}} c_U^{q/p} \left( |b_{l'}| |\xi_{l'}^q| + 2^q \sum_{j=1}^{l'-1} |c_{l'j}| |\xi_{l'}^q| |\xi_j^q| \right. \\ &\quad \left. + |c_{l'l'}| (|\xi_{l'}^2 - \sigma_\xi^2|)^q + \sum_j^L |e_{l'j}| |\xi_{l'}^q| |\varsigma_j^q| \right). \end{aligned}$$

There exists a finite constant  $c_q > 1$  such that  $E|\varsigma_{l'}^q| \leq c_q$  and  $E(|\xi_{l'}^2 - \sigma_\xi^2|)^q \leq c_q$  for all  $l'$  and  $L$ . Taking  $q = 2 + \delta_U/2$ , as  $E|\xi_{l'}^{2(2+\delta_U/2)}|$  exists and  $\sum_j^L (|e_{l'j}| |\xi_j^q|) \leq c_G |\xi_{l'}^q|$  where  $c_G = \|E_L\|_\infty < \infty$ , it follows that

$$\sum_{l=L+1}^{2L} E|U_{L,l}|^{2+\delta_U/2} \leq 2^{2+\delta_U/2} L^{-\frac{2+\delta_U/2}{2}} c_U^{(2+\delta_U/2)/p} c_q^2$$

$$\begin{aligned} &\times \left( |b_{l'}| + \sum_{j=1}^{l'} E|c_{l'j}| + |c_{l'l'}| + \sum_j^L E(|e_{l'j}| |\varsigma_{l'}^{2+\delta_U/2}|) \right) \\ &= O\left(L^{-\frac{\delta_U}{4}}\right). \end{aligned}$$

As  $\sigma_{U_L}^2 = O(1)$  and is bounded away from zero,

$$\begin{aligned} \sum_{l=1}^{2L} E|U_{L,l}|^{2+\delta_U/2} &= \frac{1}{\sigma_{U_L}^{2+\delta_U}} \sum_{l=1}^{2L} E|U_{L,l}|^{2+\delta_U/2} \\ &= O(L^{-\frac{\delta_U}{4}}), \end{aligned}$$

which goes to zero as  $L$  tends to infinity. Conditions for the martingale CLT satisfy, so  $U_L/\sigma_{U_L} \xrightarrow{d} N(0, 1)$ .

## Appendix C. Relevant statistics

### C.1. Expected Value of $\ln L_L^c(\theta)$

To investigate the uniform convergence property of the concentrated log likelihood function  $\ln L_L^c(\theta)$  in (5), we need to derive the  $E \ln L_L^c(\theta)$ .

From (4), as  $Y_L = (I_L - W_L(\eta_0))^{-1}(X_{1L}\beta_0 + \mathcal{I}_0(\gamma_0, \rho_0) + \mathbf{c}_{1L0} + \alpha_{1L0} + \epsilon_L\delta_0 + \xi_L)$ , we have

$$W_{jL} Y_L = W_{jL} S_L^{-1}(\eta)(X_{1L}\beta_0 + \mathcal{I}_0(\gamma_0, \rho_0) + \mathbf{c}_{1L0} + \alpha_{1L0} + \epsilon_L\delta_0 + \xi_L).$$

By denoting  $\mathcal{G}_L = \begin{pmatrix} G_{1L} \\ G_{2L} \\ G_{3L} \end{pmatrix} = \begin{pmatrix} W_{1L} S_L^{-1} \\ W_{2L} S_L^{-1} \\ W_{3L} S_L^{-1} \end{pmatrix} = \mathcal{W}_L S_L^{-1}$  where  $\mathcal{W}_L =$

$\begin{pmatrix} W_{1L} \\ W_{2L} \\ W_{3L} \end{pmatrix}$ ,  $S_L(\eta) S_L^{-1} = I_L - ((\eta - \eta_0)' \otimes I_L) \mathcal{G}_L$ . Since  $\xi_L(\theta, \mathbf{c}_{1L}^-, \alpha_{1L}^-, \mathbf{c}_{2L}, \alpha_{2L}) = (I_L - W_L(\eta)) Y_L - X_{1L}\beta - \mathcal{I}_0(\gamma, \rho) - \mathbf{c}_{1L} - \alpha_{1L} - \epsilon_L(\phi_2, \mathbf{c}_{2L}, \alpha_{1L})\delta$  where  $\epsilon_L(\phi_2, \mathbf{c}_{2L}, \alpha_{2L}) = Z_L - \mathcal{K}_L \Phi_2 - \mathbf{c}_{2L} - \alpha_{2L}$ , we have

$$\begin{aligned} \xi_L(\theta, \mathbf{c}_{1L}^-, \alpha_{1L}^-, \mathbf{c}_{2L}, \alpha_{2L}) &= S_L(\eta) S_L^{-1}(X_{1L}\beta_0 + \mathcal{I}_0(\gamma_0, \rho_0) \\ &\quad + \mathbf{c}_{1L0} + \alpha_{1L0} + \epsilon_L\delta_0 + \xi_L) - X_{1L}\beta - \mathcal{I}_0(\gamma, \rho) - \mathbf{c}_{1L} \\ &\quad - \alpha_{1L} - \epsilon_L(\phi_2, \mathbf{c}_{2L}, \alpha_{2L})\delta \\ &= X_{1L}(\beta_0 - \beta) + (\mathcal{I}_0(\gamma_0, \rho_0) - \mathcal{I}_0(\gamma, \rho)) \\ &\quad + \epsilon_L\delta_0 - \epsilon_L(\phi_2, \mathbf{c}_{2L}, \alpha_{2L})\delta \\ &\quad - ((\eta - \eta_0)' \otimes I_L) \mathcal{G}_L(X_{1L}\beta_0 + \mathcal{I}_0(\gamma_0, \rho_0) + \epsilon_L\delta_0) \\ &\quad + (\mathbf{c}_{1L0} - \mathbf{c}_{1L}) + (\alpha_{1L0} - \alpha_{1L}) \\ &\quad - ((\eta - \eta_0)' \otimes I_L) \mathcal{G}_L(\mathbf{c}_{1L0} + \alpha_{1L0}) + S_L(\eta) S_L^{-1} \xi_L. \end{aligned}$$

As  $\epsilon_L\delta_0 - \epsilon_L(\phi_2, \mathbf{c}_{2L}, \alpha_{2L})\delta = \epsilon_L(\delta_0 - \delta) + [\epsilon_L - \epsilon_L(\phi_2, \mathbf{c}_{2L}, \alpha_{2L})]\delta$  where

$$\begin{aligned} \epsilon_L - \epsilon_L(\phi_2, \mathbf{c}_{2L}, \alpha_{2L}) &= \mathcal{K}_L(\Phi_2 - \Phi_{20}) + (\mathbf{c}_{2L} - \mathbf{c}_{2L0}) \\ &\quad + (\alpha_{2L} - \alpha_{2L0}), \end{aligned} \quad (19)$$

we have

$$\begin{aligned} \epsilon_L\delta_0 - \epsilon_L(\phi_2, \mathbf{c}_{2L}, \alpha_{2L})\delta &= \epsilon_L(\delta_0 - \delta) + \mathcal{K}_L(\Phi_2 - \Phi_{20})\delta \\ &\quad + (\mathbf{c}_{2L} - \mathbf{c}_{2L0})\delta + (\alpha_{2L} - \alpha_{2L0})\delta. \end{aligned} \quad (20)$$

Therefore,

$$\begin{aligned} \xi_L(\theta, \mathbf{c}_{1L}^-, \alpha_{1L}^-, \mathbf{c}_{2L}, \alpha_{2L}) &= X_{1L}(\beta_0 - \beta) + (\mathcal{I}_0(\gamma_0, \rho_0) - \mathcal{I}_0(\gamma, \rho)) \\ &\quad + \epsilon_L(\delta_0 - \delta) + \mathcal{K}_L(\Phi_2 - \Phi_{20})\delta \\ &\quad - ((\eta - \eta_0)' \otimes I_L) \mathcal{G}_L(X_{1L}\beta_0 + \mathcal{I}_0(\gamma_0, \rho_0) + \epsilon_L\delta_0) \\ &\quad + (\alpha_{2L} - \alpha_{2L0})\delta + (\mathbf{c}_{1L0} - \mathbf{c}_{1L}) \\ &\quad + (\alpha_{1L0} - \alpha_{1L}) - ((\eta - \eta_0)' \otimes I_L) \mathcal{G}_L(\mathbf{c}_{1L0} + \alpha_{1L0}) S_L(\eta) S_L^{-1} \xi_L. \end{aligned} \quad (21)$$

With the  $J_T \otimes J_n$  transformation, time and individual effects are eliminated. Therefore, by denoting  $\xi_L(\theta) = Y_L - W_L(\eta)Y_L - X_{1L}\beta_0 - (Z_L - \mathcal{K}_L\Phi_2)\delta - \mathbf{I}_0(\gamma, \rho)$  and  $\epsilon_L(\phi_2) = Z_L - \mathcal{K}_L\Phi_2$ , we have

$$\begin{aligned} (J_T \otimes J_n)\xi_L(\theta) &= (J_T \otimes J_n)[X_{1L}(\beta_0 - \beta) + \mathbf{I}_0(\gamma_0, \rho_0) - \mathbf{I}_0(\gamma, \rho)] \\ &\quad + \epsilon_L(\delta_0 - \delta) + \mathcal{K}_L(\Phi_2 - \Phi_{20})\delta \\ &\quad + (J_T \otimes J_n)[((\eta_0 - \eta)' \otimes I_L)\mathcal{G}_L(X_{1L}\beta_0 + \mathbf{I}_0(\gamma_0, \rho_0) + \epsilon_L\delta_0 \\ &\quad + \mathbf{c}_{1L0} + \alpha_{1L0})] + (J_T \otimes J_n)[S_L(\eta)S_L^{-1}\xi_L], \end{aligned} \quad (22)$$

and

$$(J_T \otimes J_n)\epsilon_L(\phi_2) = (J_T \otimes J_n)(Z_L - \mathcal{K}_L\Phi_2).$$

For the expected value of  $\ln L_L^c(\theta)$ , it is

$$\begin{aligned} E \ln L_L^c(\theta) &= -\frac{nT}{2} \ln 2\pi + E \ln |I_L - W_L(\eta)| - \frac{nT}{2} \ln \sigma_\xi^2 \\ &\quad - \frac{nT}{2} \ln |\Sigma_\epsilon| - \frac{1}{2} E [\text{vec}'(\epsilon_L(\phi_2))(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(\epsilon_L(\phi_2))] \\ &\quad - \frac{1}{2\sigma_\xi^2} E [\xi_L'(\theta)(J_T \otimes J_n)\xi_L(\theta)]. \end{aligned}$$

We need to solve  $E[\text{vec}'(\epsilon_L(\phi_2))(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(\epsilon_L(\phi_2))]$  and  $E[\xi_L'(\theta)(J_T \otimes J_n)\xi_L(\theta)]$ . From (19),

$$\begin{aligned} E[\text{vec}'(\epsilon_L(\phi_2))(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(\epsilon_L(\phi_2))] &= E[\text{vec}'(\epsilon_L)(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(\epsilon_L)] \\ &\quad + E[\text{vec}'(\mathcal{K}_L(\Phi_{20} - \Phi_2))(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(\mathcal{K}_L(\Phi_{20} - \Phi_2))] \\ &\quad + 2E[\text{vec}'(\mathcal{K}_L(\Phi_{20} - \Phi_2))(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(\epsilon_L)]. \end{aligned}$$

We have  $E[\text{vec}'(\epsilon_L)(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(\epsilon_L)] = \text{tr}[(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n)E(\text{vec}(\epsilon_L)\text{vec}'(\epsilon_L))] = \text{tr}[(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n)(\Sigma_{\epsilon 0} \otimes I_T \otimes I_n)] = (n-1)(T-1)\text{tr}(\Sigma_\epsilon^{-1}\Sigma_{\epsilon 0})$ . Also, using uncorrelatedness of  $X_{2L}$  and  $\epsilon_L$ , and  $\frac{1}{T}E(Z_{L-1}'(J_T \otimes J_n)\epsilon_L) = O(\frac{n}{T})$  from Lemma 16 in Yu et al. (2008), we have

$$\begin{aligned} \frac{1}{nT} E[\text{vec}'(\epsilon_L(\phi_2))(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(\epsilon_L(\phi_2))] &= \frac{(n-1)(T-1)}{nT} \text{tr}(\Sigma_\epsilon^{-1}\Sigma_{\epsilon 0}) + \frac{1}{nT} (\phi_{20} - \phi_2)' \\ &\quad \times [E\mathcal{K}_L'(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n)\mathcal{K}_L] (\phi_{20} - \phi_2) + O\left(\frac{1}{T}\right). \end{aligned}$$

From (22), by denoting  $Q_{jL} = G_{jL}(X_{1L}\beta_0 + \mathbf{I}_0(\gamma_0, \rho_0) + \epsilon_L\delta_0 + \mathbf{c}_{1L0} + \alpha_{1L0})$  for  $j = 1, 2, 3$ , we see that

$$[Q_{2L}, Q_{3L}, X_{1L}] + [\mathbf{I}_0(\gamma_0, \rho_0), \mathbf{0}_{L \times k_1}] = R_L.$$

Therefore, (22) can be written as

$$\begin{aligned} (J_T \otimes J_n)\xi_L(\theta) &= (J_T \otimes J_n)[Q_{1L}(\lambda_0 - \lambda) + R_L(\phi_{10} - \phi) \\ &\quad + \mathcal{K}_L(\Phi_2 - \Phi_{20})\delta + \epsilon_L(\delta_0 - \delta) + S_L(\eta)S_L^{-1}\xi_L] \end{aligned}$$

so that we get the equation given in Box II. For the second term, we have  $\frac{1}{nT} E[\text{tr}((S_L(\eta)S_L^{-1}\xi_L)' \cdot (J_T \otimes J_n) \cdot (S_L(\eta)S_L^{-1}\xi_L))] = \frac{\sigma_{\xi 0}^2}{nT} E[\text{tr}((S_L(\eta)S_L^{-1})' \cdot (J_T \otimes J_n) \cdot (S_L(\eta)S_L^{-1}))]$ . For the third term, we have  $\frac{1}{nT} E(Q_{1L}, R_L, \mathcal{K}_L, \epsilon_L)' (J_T \otimes J_n) S_L(\eta) S_L^{-1} \xi_L = 0$ . Therefore, by denoting

$$\mathcal{H}_{1,nT} = \frac{1}{nT} (Q_{1L}, R_L, \mathcal{K}_L, \epsilon_L)' (J_T \otimes J_n) (Q_{1L}, R_L, \mathcal{K}_L, \epsilon_L),$$

we have

$$\begin{aligned} \frac{1}{nT} E[\xi_L'(\theta)(J_T \otimes J_n)\xi_L(\theta)] &= E \begin{bmatrix} ((\lambda_0 - \lambda), (\phi_{10} - \phi)', ((\Phi_2 - \Phi_{20})\delta)', (\delta_0 - \delta)') \\ \cdot \mathcal{H}_{1,nT} \cdot ((\lambda_0 - \lambda), (\phi_{10} - \phi)', ((\Phi_2 - \Phi_{20})\delta)', (\delta_0 - \delta)') \end{bmatrix} \end{aligned}$$

$$+ \frac{\sigma_\xi^2}{nT} E[\text{tr}(S_L(\eta)S_L^{-1})' \cdot (J_T \otimes J_n) \cdot (S_L(\eta)S_L^{-1})].$$

By denoting

$$\mathcal{H}_{2,nT}(\Sigma_\epsilon) = \frac{1}{nT} E[\mathcal{K}_L'(\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n)\mathcal{K}_L]$$

we have

$$\begin{aligned} E \frac{1}{nT} \ln L_L^c(\theta) &= -\frac{1}{2} \ln 2\pi + \frac{1}{nT} E \ln |S_L(\lambda)| - \frac{1}{2} \ln |\sigma_\xi^2| \\ &\quad - \frac{1}{2} \ln |\Sigma_\epsilon| - \frac{1}{2} \frac{(n-1)(T-1)}{nT} \text{tr}(\Sigma_\epsilon^{-1}\Sigma_{\epsilon 0}) \\ &\quad - \frac{1}{2} (\phi_{20} - \phi_2)' \mathcal{H}_{2,nT}(\Sigma_\epsilon) (\phi_{20} - \phi_2) \\ &\quad - \frac{1}{2\sigma_\xi^2} \frac{1}{nT} ((\lambda_0 - \lambda), (\phi_{10} - \phi)', ((\Phi_2 - \Phi_{20})\delta)', (\delta_0 - \delta)') \\ &\quad \cdot \mathcal{H}_{1,nT} \cdot ((\lambda_0 - \lambda), (\phi_{10} - \phi)', ((\Phi_2 - \Phi_{20})\delta)', (\delta_0 - \delta)') \\ &\quad - \frac{1}{2nT} \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} E \text{tr}(S_L(\eta)S_L^{-1})' \cdot (J_T \otimes J_n) \cdot (S_L(\eta)S_L^{-1}) + O\left(\frac{1}{T}\right). \end{aligned}$$

By denoting  $\Phi_1 = (\lambda, \phi_1, (\Phi_2\delta)', \delta')'$ ,  $\Phi_{10}(\delta) = (\lambda_0, \phi_{10}, (\Phi_{20}\delta)', \delta_0)'$ , we have

$$\begin{aligned} E \frac{1}{nT} \ln L_L^c(\theta) &= -\frac{1}{2} \ln 2\pi + \frac{1}{nT} E \ln |S_L(\lambda)| - \frac{1}{2} \ln |\sigma_\xi^2| \\ &\quad - \frac{1}{2} \ln |\Sigma_\epsilon| - \frac{1}{2} \text{tr}(\Sigma_\epsilon^{-1}\Sigma_{\epsilon 0}) \\ &\quad - \frac{1}{2} (\phi_{20} - \phi_2)' E \mathcal{H}_{2,nT}(\Sigma_\epsilon) (\phi_{20} - \phi_2) \\ &\quad - \frac{1}{2nT} \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} E \text{tr}(S_L(\eta)S_L^{-1})' \cdot (J_T \otimes J_n) \cdot (S_L(\eta)S_L^{-1}) \\ &\quad - \frac{1}{2\sigma_\xi^2} (\Phi_{10}(\delta) - \Phi_1)' E \mathcal{H}_{1,nT} (\Phi_{10}(\delta) - \Phi_1) \\ &\quad + O\left(\frac{1}{n}\right) + O\left(\frac{1}{T}\right). \end{aligned} \quad (23)$$

## C.2. First and second order derivatives

By denoting  $\xi_L(\theta) = (I_L - W_L(\eta))Y_L - X_{1L}\beta - (Z_L - \mathcal{K}_L\Phi_2)\delta - \mathbf{I}_0(\gamma, \rho)$  and  $\epsilon_L(\theta) = Z_L - \mathcal{K}_L\Phi_2$ , the first and second order conditions of (5) can be derived.<sup>22</sup>

For the expectation part of the score in (9), to derive the  $E \frac{\partial \ln L_L^c(\theta)}{\partial \phi_2}$  component, we can rely on the iterative substitution of  $Z_{nt} = Z_{n,t-1}\gamma_0 + X_{2nt}\Gamma_0 + \mathbf{c}_{n20} + l_n\alpha'_{t20} + \epsilon_{nt}$ . We can show that  $E(\bar{Z}_{n,T-1}' J_n \bar{\epsilon}_{nT}) = \frac{n-1}{T^2} \sum_{t=1}^{T-1} \sum_{h=1}^{T-t} \gamma_0^h \Sigma_{\epsilon 0}$ . Therefore,

$$E[(\Sigma_{\epsilon 0}^{-1} \otimes \bar{Z}_{n,T-1}' J_n) \text{vec}(\bar{\epsilon}_{nT})] = \frac{n-1}{T^2} \text{vec} \left( \sum_{t=1}^{T-1} \sum_{h=1}^{T-t} \gamma_0^h \right).$$

Alternatively, in terms of the big matrix form, as  $\text{vec}(Z_{L-1}\gamma_0) = (\gamma_0' \otimes W_{2L}) \text{vec}(Z_L)$ , by denoting  $\mathcal{X}_L$  as the vectorized stack of  $X_{2nt}\Gamma_0 + \mathbf{c}_{n20} + l_n\alpha'_{t20}$ , we have  $\text{vec}(Z_L) = (I_L - \gamma_0' \otimes W_{2L})^{-1} (\mathcal{X}_L + \epsilon_L)$  and  $\text{vec}(Z_{L-1}\gamma_0) = (\gamma_0' \otimes W_{2L})(I_L - \gamma_0' \otimes W_{2L})^{-1} (\mathcal{X}_L + \epsilon_L)$ . Denote  $e_j = (0, \dots, 1, \dots, 0)'$  as a  $p \times 1$  vector with 1 in its  $j$ th element and zero otherwise. Then, by denoting  $G_{2L,\gamma} = (\gamma_0' \otimes W_{2L})(I_L - \gamma_0' \otimes W_{2L})^{-1}$ , we have  $E[Z_{L-1,j}' (J_T \otimes J_n) \epsilon_{L,l}] =$

<sup>22</sup> Details for the first and second order derivatives can be found in the supplement file and available upon request (see Appendix E).



$$\begin{aligned}
& E[\xi_L'(\theta)(J_T \otimes J_n)\xi_L(\theta)] \\
&= E \left[ \begin{aligned} & ((\lambda_0 - \lambda), (\phi_{10} - \phi)', ((\Phi_2 - \Phi_{20})\delta)', (\delta_0 - \delta)')(Q_{1L}, R_L, \mathcal{K}_L, \epsilon_L)' \cdot \\ & (J_T \otimes J_n) \cdot (Q_{1L}, R_L, \mathcal{K}_L, \epsilon_L)((\lambda_0 - \lambda)', (\phi_{10} - \phi)', ((\Phi_2 - \Phi_{20})\delta)', (\delta_0 - \delta)')' \end{aligned} \right] \\
&+ Etr[(S_L(\eta)S_L^{-1}\xi_L)' \cdot (J_T \otimes J_n) \cdot (S_L(\eta)S_L^{-1}\xi_L)] \\
&+ E[(\lambda_0 - \lambda), (\phi_{10} - \phi)', ((\Phi_2 - \Phi_{20})\delta)', (\delta_0 - \delta)')(Q_{1L}, R_L, \mathcal{K}_L, \epsilon_L)' \cdot (J_T \otimes J_n) \cdot S_L(\eta)S_L^{-1}\xi_L].
\end{aligned}$$

**Box II.**

$$E \left( -\frac{\partial^2 \ln L_L^c(\theta_0)}{\partial \theta \partial \theta'} \right) = \frac{1}{\sigma_{\xi 0}^2} \cdot \begin{pmatrix} I_{\lambda\lambda} & * & * & * & * & \mathbf{0}_{1 \times J} \\ ER_L^{(u)} J_L Q_{1L} & ER_L' J_L R_L & \mathbf{0}_{(k_1+2) \times p} & * & \mathbf{0}_{(k_1+2) \times 1} & \mathbf{0}_{(k_1+2) \times J} \\ E[\epsilon_L' J_L Q_{1L}] + & \mathbf{0}_{p \times (k_1+2)} & T \cdot E\epsilon_L' J_L \epsilon_L & \mathbf{0}_{p \times k_{\phi_2}} & \mathbf{0}_{p \times 1} & \mathbf{0}_{p \times J} \\ E[\epsilon_L' J_L (G_{1L} \epsilon_L \delta_0)] & -\delta_0 \otimes E\mathcal{K}_L' J_L Q_{1L} & -\delta_0 \otimes E\mathcal{K}_L' J_L R_L & \mathbf{0}_{k_{\phi_2} \times p} & \mathbf{0}_{k_{\phi_2} \times 1} & \mathbf{0}_{k_{\phi_2} \times J} \\ -\delta_0 \otimes E\mathcal{K}_L' J_L Q_{1L} & -\delta_0 \otimes E\mathcal{K}_L' J_L R_L & \mathbf{0}_{k_{\phi_2} \times p} & I_{\phi_2 \phi_2} & \mathbf{0}_{k_{\phi_2} \times 1} & \mathbf{0}_{k_{\phi_2} \times J} \\ Etr(G_L J_L) & \mathbf{0}_{1 \times (k_1+2)} & \mathbf{0}_{1 \times p} & \mathbf{0}_{1 \times k_{\phi_2}} & \frac{1}{\sigma_{\xi 0}^2} \left( \frac{nT}{2} - T \right) & \mathbf{0}_{1 \times J} \\ \mathbf{0}_{J \times 1} & \mathbf{0}_{J \times (k_1+2)} & \mathbf{0}_{J \times p} & \mathbf{0}_{J \times k_{\phi_2}} & \mathbf{0}_{J \times 1} & I_{\alpha\alpha} \end{pmatrix} + O(n), \quad (24)$$

where  $I_{\lambda\lambda} = EQ_L'(J_T \otimes J_n)Q_{1L} + \sigma_{\xi 0}^2 tr(G_{1L}^2 + G_{1L}'(J_T \otimes J_n)G_{1L})$ ,  $I_{\phi_2 \phi_2} = (\sigma_{\xi 0}^2 \Sigma_{\epsilon 0}^{-1} + \delta_0 \delta_0') \otimes E\mathcal{K}_L'(J_T \otimes J_n)\mathcal{K}_L$ , and  $I_{\alpha\alpha}$  is a  $J \times J$  matrix with it  $(k, j)$  element being  $\frac{nT}{2} \sigma_{\xi 0}^2 tr(\Sigma_{\epsilon 0}^{-1} \frac{\partial \Sigma_{\epsilon 0}}{\partial \alpha_k} \Sigma_{\epsilon 0}^{-1} \frac{\partial \Sigma_{\epsilon 0}}{\partial \alpha_j})$ .

**Box III.**

$E[(\epsilon_L' \otimes I_L)G_{2L} \text{vec}(\epsilon_L)]' \cdot (J_T \otimes J_n) \cdot (\epsilon_L' \otimes I_L) \text{vec}(\epsilon_L)$ . Thus, similar results can be obtained by using the explicit form of  $G_{2L}$ .

From the second order derivatives, by denoting  $k_{\phi_2} = p^2 + p(k_2 + 1)$  and  $J_L = J_T \otimes J_n$ , we can show the information matrix as in **Box III**.

The structure of  $E(-\frac{\partial^2 \ln L_L^c(\theta_0)}{\partial \theta \partial \theta'})$  is similar to **Qu and Lee (2015)** with similar conditions for the non-singularity of the information matrix.

**C.3. Variance of the score**

We can show that  $E(\frac{1}{L} \frac{\partial \ln L_L^c(\theta_0)}{\partial \theta} \frac{\partial \ln L_L^c(\theta_0)}{\partial \theta'}) = \mathcal{I}_{nT, \theta_0} + \mathcal{Q}_{nT, \theta_0} + o_p(1)$ , where  $\mathcal{I}_{nT, \theta_0}$  is in (24), and  $\mathcal{Q}_{nT, \theta_0}$  is related to the third and fourth moment of  $\xi_L$  which are zeros under normal distribution. In general, by using  $E(D_L' \xi_L \cdot \xi_L' B_L \xi_L) = \mu_{\xi}^3 \sum_{l=1}^L E(D_L')_l (B_L)_l$  and  $E(\xi_L' A_L \xi_L \cdot \xi_L' B_L \xi_L) = (\mu_{\xi}^4 - 3\sigma_{\xi 0}^4) \sum_{l=1}^L E(A_L)_l (B_L)_l + \sigma_{\xi 0}^4 E[tr(A_L B_L) + tr(A_L B_L') + tr(A_L)tr(B_L)]$ , we have

$$\mathcal{Q}_{nT, \theta_0} = \frac{1}{L} \frac{1}{\sigma_{\xi 0}^4} \begin{pmatrix} R_{\lambda\lambda} & * & * & * & * & \mathbf{0} \\ R_{\lambda\phi_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & * & \mathbf{0} \\ R_{\lambda\delta} & \mathbf{0} & \mathbf{0} & \mathbf{0} & * & \mathbf{0} \\ R_{\lambda\phi_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & * & \mathbf{0} \\ R_{\lambda\sigma_{\xi}} & R_{\phi_1 \sigma_{\xi}} & R_{\delta \sigma_{\xi}} & R_{\phi_2 \sigma_{\xi}} & R_{\sigma_{\xi} \sigma_{\xi}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & R_{\phi_2 \alpha} & \mathbf{0} & R_{\alpha\alpha} \end{pmatrix}, \quad (25)$$

where

$$\begin{aligned}
R_{\lambda\lambda} &= (\mu_{\xi}^4 - 3\sigma_{\xi 0}^4) \sum_{l=1}^L E[(G_{1L}(J_T \otimes J_n))_{ll}]^2 + 2\mu_{\xi}^3 \sum_{l=1}^L E[(Q_{1L}')_l (G_{1L}(J_T \otimes J_n))_{ll}], \\
R_{\lambda\phi_1} &= \mu_{\xi}^3 \sum_{l=1}^L E[(R_L^{(u)})_l (G_{1L}(J_T \otimes J_n))_{ll}], \\
R_{\lambda\delta} &= \mu_{\xi}^3 \sum_{l=1}^L E[(\epsilon_L' (J_T \otimes J_n))_l (G_{1L}(J_T \otimes J_n))_{ll}], \\
R_{\lambda\phi_2} &= R_{\lambda\Gamma} = -\mu_{\xi}^3 \delta_0 \otimes \sum_{l=1}^L E[(\mathcal{K}_L)_l (G_{1L}(J_T \otimes J_n))_{ll}], \\
R_{\lambda\sigma_{\xi}} &= \frac{1}{2\sigma_{\xi 0}^2} (\mu_{\xi}^4 - 3\sigma_{\xi 0}^4) \sum_{l=1}^L E[(G_{1L}(J_T \otimes J_n))_{ll}] + \frac{\mu_{\xi}^3}{2\sigma_{\xi 0}^2} \sum_{l=1}^L E[(Q_{1L}')_l], \\
R_{\phi_1 \sigma_{\xi}} &= \frac{1}{2\sigma_{\xi 0}^2} \mu_{\xi}^3 \sum_{l=1}^L E[(R^{(u)})_l], \\
R_{\delta \sigma_{\xi}} &= \frac{nT \mu_{\xi}^3}{2\sigma_{\xi 0}^2},
\end{aligned}$$

$$R_{\phi_2 \sigma_{\xi}} = -\frac{\mu_{\xi}^3}{2\sigma_{\xi 0}^2} \delta_0 \otimes \sum_{l=1}^L E[(J_T \otimes J_n) \mathcal{K}_L']_l,$$

$$R_{\phi_2 \alpha} = -\frac{\sigma_{\xi 0}^4}{2} E(\Sigma_{\epsilon 0}^{-1} \otimes \mathcal{K}_L'(J_T \otimes J_n)) \text{vec}(\epsilon_L) \cdot \frac{\partial}{\partial \alpha} (tr[\Sigma_{\epsilon 0}^{-1} \epsilon_L' (J_T \otimes J_n) \epsilon_L]),$$

$$R_{\sigma_{\xi} \sigma_{\xi}} = \frac{nT}{4\sigma_{\xi 0}^4} (\mu_{\xi}^4 - 3\sigma_{\xi 0}^4),$$

$$R_{\alpha\alpha} = \frac{\sigma_{\xi 0}^2}{4} E \frac{\partial}{\partial \alpha} (tr[\epsilon_L' (J_T \otimes J_n) \epsilon_L \Sigma_{\epsilon 0}^{-1}]) \cdot \frac{\partial}{\partial \alpha'} (tr[\epsilon_L' (J_T \otimes J_n) \epsilon_L \Sigma_{\epsilon 0}^{-1}]).$$

**Appendix D. Proofs for theorems****D.1. Proof for Theorem 1**

We will show (i) the uniform convergence of  $\frac{1}{L} \ln L_L^c(\theta) - \frac{1}{L} E \ln L_L^c(\theta)$  to zero in probability uniformly in  $\theta$  in  $\Theta$ ; (2) the uniform equicontinuity of  $\frac{1}{L} E \ln L_L^c(\theta)$ , and (3) global identification.

(1) To prove  $\frac{1}{L} \ln L_L^c(\theta) - \frac{1}{L} E \ln L_L^c(\theta) \xrightarrow{p} 0$  uniformly in  $\theta$  in any compact parameter space  $\Theta$ :

From (5), the log likelihood function is

$$\begin{aligned}
\ln L_L^c(\theta) &= -\frac{nT}{2} \ln 2\pi + \ln |S_L(\eta)| - \frac{nT}{2} \ln \sigma_{\xi}^2 - \frac{nT}{2} \ln |\Sigma_{\epsilon}| \\
&- \frac{1}{2\sigma_{\xi}^2} \xi_L'(\theta)(J_T \otimes J_n)\xi_L(\theta), \\
&- \frac{1}{2} \text{vec}'(\epsilon_L(\phi_2))(\Sigma_{\epsilon}^{-1} \otimes J_T \otimes J_n) \text{vec}(\epsilon_L(\phi_2)),
\end{aligned}$$

where  $\xi_L(\theta) = Y_L - W_L(\eta)Y_L - X_{1L}\beta_0 - (Z_L - \mathcal{K}_L\Phi_2)\delta - \mathcal{I}_0(\gamma, \rho)$  and  $\epsilon_L(\phi_2) = Z_L - \mathcal{K}_L\Phi_2$ . Here, with the  $J_T \otimes J_n$  transformation, the time and individual effects in  $(J_T \otimes J_n)\xi_L(\theta, \mathbf{c}_{1L}, \alpha_{1L})$  and  $(J_T \otimes J_n)\epsilon_L(\phi_2, \mathbf{c}_{2L}, \alpha_{2L})$  are eliminated. From (23),

$$\begin{aligned}
E \frac{1}{nT} \ln L_L^c(\theta) &= -\frac{1}{2} \ln 2\pi + \frac{1}{nT} E \ln |S_L(\eta)| - \frac{1}{2} \ln |\sigma_{\xi}^2| \\
&- \frac{1}{2} \ln |\Sigma_{\epsilon}| - \frac{1}{2} tr(\Sigma_{\epsilon}^{-1} \Sigma_{\epsilon 0}) - \frac{1}{2\sigma_{\xi}^2} (\Phi_{10} - \Phi_1)' E\mathcal{H}_{1, nT} (\Phi_{10} - \Phi_1) \\
&- \frac{1}{2} (\phi_{20} - \phi_2)' E\mathcal{H}_{2, nT} (\Sigma_{\epsilon}) (\phi_{20} - \phi_2) \\
&- \frac{1}{2nT} \frac{\sigma_{\xi 0}^2}{\sigma_{\xi}^2} Etr[(S_L(\eta)S_L^{-1})' \cdot (J_T \otimes J_n) \cdot (S_L(\eta)S_L^{-1})]
\end{aligned}$$

$$+ o\left(\frac{1}{n}\right) + o\left(\frac{1}{T}\right),$$

where  $\Phi_1 = (\lambda, \phi_1', (\Phi_2\delta)', \delta')'$ ,  $\Phi_{10}(\delta) = (\lambda_0, \phi_{10}', (\Phi_{20}\delta)', \delta_0')'$ . Therefore,

$$\begin{aligned} \frac{1}{nT} \ln L_L^c(\theta) - \frac{1}{nT} \mathbb{E} \ln L_L^c(\theta) &= -\frac{1}{2} \frac{1}{nT} \text{vec}'(\epsilon_L(\phi_2)) \\ &\quad \times (\Sigma_\epsilon^{-1} \otimes J_T \otimes J_n) \text{vec}(\epsilon_L(\phi_2)) \\ &+ \frac{1}{2} \text{tr}(\Sigma_\epsilon^{-1} \Sigma_{\epsilon 0}) + \frac{1}{2} (\phi_{20} - \phi_2)' \mathbb{E} \mathcal{H}_{2,nT}(\Sigma_\epsilon) (\phi_{20} - \phi_2) \\ &- \frac{1}{2\sigma_\xi^2} \frac{1}{nT} \xi_L'(\theta) (J_T \otimes J_n) \xi_L(\theta) \\ &+ \frac{1}{2\sigma_\xi^2} (\Phi_{10} - \Phi_1)' \mathbb{E} \mathcal{H}_{1,nT}(\Phi_{10} - \Phi_1) \\ &+ \frac{1}{2nT} \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} \mathbb{E} \left[ \text{tr}(S_L(\eta) S_L^{-1})' \cdot (J_T \otimes J_n) \cdot (S_L(\eta) S_L^{-1}) \right] \\ &+ \frac{1}{nT} \ln |S_L(\eta)| - \frac{1}{nT} \mathbb{E} \ln |S_L(\eta)| + o\left(\frac{1}{n}\right) + o\left(\frac{1}{T}\right). \end{aligned}$$

For  $\frac{1}{nT} \ln |S_L(\eta)| - \frac{1}{nT} \mathbb{E} \ln |S_L(\eta)|$ , we have

$$\begin{aligned} \sup_{\eta \in \eta} \frac{1}{L} [\ln |L - W_L(\eta)| - \mathbb{E} (\ln |L - W_L(\eta)|)] \\ = -\sup_{\lambda \in \Lambda} \frac{1}{L} \sum_{l=1}^L \left[ \sum_{k=1}^{\infty} \frac{\lambda^k}{k} (W_{1L}^k)_{ll} - \sum_{k=1}^{\infty} \frac{\lambda^k}{k} \mathbb{E} (W_{1L}^k)_{ll} \right] \xrightarrow{p} 0, \end{aligned}$$

whose proof can be found in the proof of Theorem 3 in [Qu and Lee \(2015\)](#). By [Proposition 1](#) and [Corollary 1](#), other terms converge uniformly. Thus, we have the uniform convergence of the log likelihood function such that  $\frac{1}{L} [\ln L_L^c(\theta) - \mathbb{E} (\ln L_L^c(\theta))] \xrightarrow{p} 0$  uniformly in  $\theta$ .

(2) To prove  $\frac{1}{(n-1)T} \mathbb{E} \ln L_{n,T}(\theta)$  is uniformly equicontinuous in  $\theta$  in any compact parameter space  $\Theta$ :

From (23), these terms can be expressed by the general terms in [Proposition 1](#), so that the pointwise convergence can be established. Also, as all the parameters are bounded and enter the log likelihood function polynomially except for the term  $\frac{1}{nT} \mathbb{E} \ln |S_L(\eta)|$ , we need only show the stochastic equicontinuity of  $\frac{1}{nT} \mathbb{E} \ln |S_L(\eta)|$ . We see that as  $G_{jL} = W_{jL} S_L^{-1}$  with  $\sup_{\lambda, t} \|G_{jL}(\eta)\|_\infty < \infty$ ,

$$\left| \frac{1}{nT} \ln |S_L(\eta_1)| - \frac{1}{nT} \ln |S_L(\eta_2)| \right| = \left| \left( \frac{1}{nT} \text{tr} G_{1L}(\bar{\eta}), \frac{1}{nT} \text{tr} G_{2L}(\bar{\eta}), \frac{1}{nT} \text{tr} G_{3L}(\bar{\eta}) \right) (\eta_2 - \eta_1) \right| \leq C \cdot \|\eta_2 - \eta_1\|_\infty$$

where  $C$  is a constant. Thus, we have the stochastic equicontinuity of  $\frac{1}{nT} \ln |S_L(\eta)|$ .

Also, as all the variance parameters are bounded away from zero in compact parameter spaces, and relevant terms are bounded, we have the result that  $\mathbb{E} \frac{1}{nT} \ln L_L^c(\theta)$  is uniformly equicontinuous.

(3) To prove global identification:

From the expected log likelihood function in (23), we have

$$\begin{aligned} \mathbb{E} \frac{1}{nT} \ln L_L^c(\theta_0) &= -\frac{1}{2} \ln 2\pi + \frac{1}{nT} \mathbb{E} \ln |S_L(\eta_0)| - \frac{1}{2} \ln |\sigma_{\xi 0}^2| \\ &- \frac{1}{2} \ln |\Sigma_{\epsilon 0}| - \frac{1}{2} \text{tr}(I_p) - \frac{1}{2} + o\left(\frac{1}{n}\right) + o\left(\frac{1}{T}\right), \end{aligned}$$

which implies that

$$\mathbb{E} \frac{1}{nT} \ln L_L^c(\theta) - \mathbb{E} \frac{1}{nT} \ln L_L^c(\theta_0)$$

$$\begin{aligned} &= \frac{1}{nT} \mathbb{E} \ln |S_L(\eta)| - \frac{1}{nT} \mathbb{E} \ln |S_L| - \frac{1}{2} \ln |\sigma_\xi^2| + \frac{1}{2} \ln |\sigma_{\xi 0}^2| \\ &- \frac{1}{2} \ln |\Sigma_\epsilon| + \frac{1}{2} \ln |\Sigma_{\epsilon 0}| - \frac{1}{2} \text{tr}(\Sigma_\epsilon^{-1} \Sigma_{\epsilon 0}) + \frac{1}{2} \text{tr}(I_p) \\ &- \frac{1}{2} (\phi_{20} - \phi_2)' \mathbb{E} \mathcal{H}_{2,nT}(\Sigma_\epsilon) (\phi_{20} - \phi_2) \\ &- \frac{1}{2nT} \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} \mathbb{E} \text{tr}(S_L(\eta) S_L^{-1})' \cdot (J_T \otimes J_n) \cdot (S_L(\eta) S_L^{-1}) \\ &+ \frac{1}{2} - \frac{1}{2\sigma_\xi^2} (\Phi_{10} - \Phi_1)' \mathbb{E} \mathcal{H}_{1,nT}(\Phi_{10} - \Phi_1) \\ &+ o\left(\frac{1}{n}\right) + o\left(\frac{1}{T}\right) \\ &= -\frac{1}{2} \left[ \text{tr}(\Sigma_{\epsilon 0}^{1/2} \Sigma_\epsilon^{-1} \Sigma_{\epsilon 0}^{1/2}) - \ln |\Sigma_{\epsilon 0}^{1/2} \Sigma_\epsilon^{-1} \Sigma_{\epsilon 0}^{1/2}| - p \right] \\ &- \frac{1}{2} (\phi_{20} - \phi_2)' \mathbb{E} \mathcal{H}_{2,nT}(\Sigma_\epsilon) (\phi_{20} - \phi_2) \\ &- \frac{1}{2nT} \mathbb{E} \left[ \text{tr} \left( \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} (S_L(\eta) S_L^{-1})' (S_L(\eta) S_L^{-1}) \right) \right. \\ &\quad \left. - \ln \left| \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} (S_L(\eta) S_L^{-1})' (S_L(\eta) S_L^{-1}) \right| - nT \right] \\ &- \frac{1}{2\sigma_\xi^2} (\Phi_{10} - \Phi_1)' \mathbb{E} \mathcal{H}_{1,nT}(\Phi_{10} - \Phi_1) + o\left(\frac{1}{n}\right) \\ &+ o\left(\frac{1}{T}\right), \end{aligned} \tag{26}$$

$$\text{using } \frac{1}{nT} \text{tr} \left( \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} (S_L(\eta) S_L^{-1})' (S_L(\eta) S_L^{-1}) \right) - \frac{1}{nT} \text{tr} \left( \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} (S_L(\eta) S_L^{-1})' (J_T \otimes J_n) (S_L(\eta) S_L^{-1}) \right) = O\left(\frac{1}{n}\right) + O\left(\frac{1}{T}\right).$$

For the scalar case, the function  $f(x) = x - \ln x - 1 \geq 0$  for any  $x > 0$  and it is minimized at  $x = 1$ . For a positive definite real matrix  $M$ ,  $f(M) = \text{tr}(M) - \ln |M| - m = \sum_{i=1}^m (\varphi_i - \ln \varphi_i - 1) \geq 0$  and is minimized at  $M = I_m$ , where  $m$  is the dimension of  $M$  and  $\varphi_i$  is an eigenvalue of  $M$ . Therefore,  $\lim_{n,T \rightarrow \infty} (\mathbb{E} \frac{1}{nT} \ln L_L^c(\theta) - \mathbb{E} \frac{1}{nT} \ln L_L^c(\theta_0)) \leq 0$ . Next, we need to show that  $\lim_{n,T \rightarrow \infty} (\mathbb{E} \frac{1}{nT} \ln L_L^c(\theta) - \mathbb{E} \frac{1}{nT} \ln L_L^c(\theta_0)) = 0$  implies  $\theta = \theta_0$ . As all the four terms in (26) are nonnegative,  $\lim_{n,T \rightarrow \infty} (\mathbb{E} \frac{1}{nT} \ln L_L^c(\theta) - \mathbb{E} \frac{1}{nT} \ln L_L^c(\theta_0)) = 0$  implies that all the four terms have limits being zero. First,  $f(\Sigma_{\epsilon 0}^{1/2} \Sigma_\epsilon^{-1} \Sigma_{\epsilon 0}^{1/2}) = 0$  implies that  $\Sigma_\epsilon = \Sigma_{\epsilon 0}$ . Second, as  $\mathbb{E} \mathcal{H}_{2,nT}(\Sigma_\epsilon)$  is positive definite,  $\phi_2 = \phi_{20}$ . Third, as  $\mathbb{E} \mathcal{H}_{1,nT}$  is positive definite,  $\Phi_1 = \Phi_{10}$ . With  $\lambda = \lambda_0$  implied by  $\Phi_1 = \Phi_{10}$ ,  $\text{tr} \left( \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} (S_L(\eta) S_L^{-1})' (S_L(\eta) S_L^{-1}) \right) - \ln \left| \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} (S_L(\eta) S_L^{-1})' (S_L(\eta) S_L^{-1}) \right| - nT = 0$  implies that  $\frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} - \ln \left| \frac{\sigma_{\xi 0}^2}{\sigma_\xi^2} \right| - 1 = 0$ , which implies that  $\sigma_\xi^2 = \sigma_{\xi 0}^2$ .

With above arguments, the consistency follows.

## D.2. Proof for Theorem 2

To prove asymptotic normality, we have the Taylor expansion:  $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) = (-\frac{1}{nT} \frac{\partial^2 \ln L_L^c(\bar{\theta}_{nT})}{\partial \theta \partial \theta'})^{-1} \cdot (\frac{1}{\sqrt{nT}} \frac{\partial \ln L_L^c(\bar{\theta}_{nT})}{\partial \theta} + \frac{1}{\sqrt{nT}} \Delta_L)$  where  $\frac{1}{\sqrt{nT}} \frac{\partial \ln L_L^c(\bar{\theta}_{nT})}{\partial \theta} \xrightarrow{d} N(0, \lim_{n,T \rightarrow \infty} (\mathcal{J}_{nT, \theta_0} + \Omega_{nT, \theta_0}))$  and  $\bar{\theta}_{nT}$  lies between  $\theta_0$  and  $\hat{\theta}_{nT}$ .

First, we have  $-\frac{1}{nT} \frac{\partial^2 \ln L_L^c(\bar{\theta}_{nT})}{\partial \theta \partial \theta'} = (-\frac{1}{nT} \frac{\partial^2 \ln L_L^c(\bar{\theta}_{nT})}{\partial \theta \partial \theta'} - (-\frac{1}{nT} \frac{\partial^2 \ln L_L^c(\theta_0)}{\partial \theta \partial \theta'})) + (-\frac{1}{nT} \frac{\partial^2 \ln L_L^c(\theta_0)}{\partial \theta \partial \theta'} - \mathcal{J}_{nT, \theta_0}) + \mathcal{J}_{nT, \theta_0}$ . The first term

is  $\|\hat{\theta}_{nT} - \theta_0\| \cdot O_p(1)$  and the second term is  $o_p(1)$ . Therefore,  $-\frac{1}{nT} \frac{\partial^2 \ln L_{n,T}(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} = \|\hat{\theta}_{nT} - \theta_0\| \cdot O_p(1) + o_p(1) + \mathcal{J}_{nT, \theta_0}$ . Because  $\|\hat{\theta}_{nT} - \theta_0\| = o_p(1)$  and  $\Sigma_{\theta_0, nT}$  is nonsingular in the limit,  $-\frac{1}{nT} \frac{\partial^2 \ln L_{n,T}^c(\hat{\theta}_{nT})}{\partial \theta \partial \theta'}$  is invertible for large  $n$  and  $T$  and  $(-\frac{1}{nT} \frac{\partial^2 \ln L_{n,T}^c(\hat{\theta}_{nT})}{\partial \theta \partial \theta'})^{-1}$  is  $O_p(1)$ . Then,  $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) = O_p(1) \cdot (O_p(1) + O(\sqrt{\frac{n}{T}}) + O(\sqrt{\frac{T}{n}}))$ , which implies that

$$\hat{\theta}_{nT} - \theta_0 = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \frac{1}{n}\right)\right) = O_p\left(\max\left(\frac{1}{n}, \frac{1}{T}\right)\right). \quad (27)$$

Hence,

$$\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) = \left(\mathcal{J}_{nT, \theta_0} + O_p\left(\max\left(\frac{1}{n}, \frac{1}{T}\right)\right)\right)^{-1} \cdot \left(\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{1,L}^c(\theta_0)}{\partial \theta} + \frac{1}{\sqrt{nT}} \Delta_L\right). \quad (28)$$

By using the CLT for linear and quadratic forms, we see that the score is asymptotically distributed as  $\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{1,L}^c(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \lim_{n,T \rightarrow \infty} (\mathcal{J}_{nT, \theta_0} + \Omega_{nT, \theta_0}))$ . Also, as  $\frac{1}{\sqrt{nT}} \Delta_L = \sqrt{\frac{n}{T}} a_{1, \theta_0} + \sqrt{\frac{T}{n}} a_{2, \theta_0}$ , we have  $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) + \sqrt{\frac{n}{T}} b_{1, \theta_0, nT} + \sqrt{\frac{T}{n}} b_{2, \theta_0, nT} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{T}{n^3}}, \frac{1}{n}, \frac{1}{T}\right)\right) \xrightarrow{d} N(0, \lim_{n,T \rightarrow \infty} \mathcal{J}_{nT, \theta_0}^{-1} (\mathcal{J}_{nT, \theta_0} + \Omega_{nT, \theta_0}) \mathcal{J}_{nT, \theta_0}^{-1})$ , where  $b_{1, \theta_0, nT} = \mathcal{J}_{nT, \theta_0}^{-1} a_{1, \theta_0}$  and  $b_{2, \theta_0, nT} = \mathcal{J}_{nT, \theta_0}^{-1} a_{2, \theta_0}$  are  $O(1)$ . Because a term of the order  $O_p(\max(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{T}{n^3}}, \frac{1}{n}, \frac{1}{T}))$  is also of the order  $O_p(\max(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{T}{n^3}}))$ , we have the results.

### D.3. Proof for Theorem 3

Theorem 2 states that  $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) + \sqrt{\frac{n}{T}} \mathcal{J}_{nT, \theta_0}^{-1} a_{1, \theta_0} + \sqrt{\frac{T}{n}} \mathcal{J}_{nT, \theta_0}^{-1} a_{2, \theta_0} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{T}{n^3}}\right)\right)$  are asymptotically centered normal. As  $\hat{\theta}_{nT}^1 = \hat{\theta}_{nT} + \frac{1}{T}(-\frac{1}{nT} E \frac{\partial^2 \ln L_{n,T}^c(\hat{\theta}_{nT})}{\partial \theta \partial \theta'})^{-1} a_1(\hat{\theta}_{nT}) + \frac{1}{n}(-\frac{1}{nT} E \frac{\partial^2 \ln L_{n,T}^c(\hat{\theta}_{nT})}{\partial \theta \partial \theta'})^{-1} a_2(\hat{\theta}_{nT})$ , under  $\frac{n}{T^3} \rightarrow 0$  and  $\frac{T}{n^3} \rightarrow 0$ , we will have  $\sqrt{nT}(\hat{\theta}_{nT}^1 - \theta_0) \xrightarrow{d} N(0, \lim_{n,T \rightarrow \infty} \mathcal{J}_{nT, \theta_0}^{-1} (\mathcal{J}_{nT, \theta_0} + \Omega_{nT, \theta_0}) \mathcal{J}_{nT, \theta_0}^{-1})$  if

$$\sqrt{\frac{n}{T}} \left( \left( -\frac{1}{nT} E \frac{\partial^2 \ln L_{n,T}^c(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1} a_1(\hat{\theta}_{nT}) - \mathcal{J}_{nT, \theta_0}^{-1} a_1(\theta_0) \right) \xrightarrow{p} 0, \quad (29)$$

and

$$\sqrt{\frac{T}{n}} \left( \left( -\frac{1}{nT} E \frac{\partial^2 \ln L_{n,T}^c(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1} a_2(\hat{\theta}_{nT}) - \mathcal{J}_{nT, \theta_0}^{-1} a_2(\theta_0) \right) \xrightarrow{p} 0. \quad (30)$$

Assuming that  $\frac{n}{T^3} \rightarrow 0$  and  $\frac{T}{n^3} \rightarrow 0$ , we will prove (29) and (30).

From (27),  $-\frac{1}{nT} E \frac{\partial^2 \ln L_{n,T}^c(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} = \mathcal{J}_{nT, \theta_0}^{-1} + O_p\left(\max\left(\frac{1}{n}, \frac{1}{T}\right)\right)$ . Hence,

$$\begin{aligned} & \sqrt{\frac{n}{T}} \left( \left( -\frac{1}{nT} E \frac{\partial^2 \ln L_{n,T}^c(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1} a_1(\hat{\theta}_{nT}) - \mathcal{J}_{nT, \theta_0}^{-1} a_1(\theta_0) \right) \\ &= \sqrt{\frac{n}{T}} \left( \mathcal{J}_{nT, \theta_0}^{-1} (a_1(\hat{\theta}_{nT}) - a_1(\theta_0)) \right) + \sqrt{\frac{n}{T}} a_1(\hat{\theta}_{nT}) \end{aligned}$$

$$\times O_p\left(\max\left(\frac{1}{n}, \frac{1}{T}\right)\right).$$

As  $\hat{\theta}_{nT} - \theta_0 = O_p\left(\max\left(\frac{1}{T}, \frac{1}{n}\right)\right)$  and  $a_1(\theta_0)$  is  $O(1)$ , according to the Taylor expansion of  $a_1(\hat{\theta}_{nT})$  in (10) around  $a_1(\theta_0)$ , to prove (29) is reduced to proving that elements of  $\frac{\partial a_1(\hat{\theta}_{nT})}{\partial \theta'} < \infty$  where  $\hat{\theta}_{nT}$  lies between  $\hat{\theta}_{nT}$  and  $\theta_0$ . Similarly, (30) is also satisfied under  $\frac{\partial a_2(\hat{\theta}_{nT})}{\partial \theta'} < \infty$ . With explicit forms of  $a_1(\theta)$  and  $a_2(\theta)$ , Assumption 10 would be reasonable.

## Appendix E. Supplementary data

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.jeconom.2016.11.004>.

## References

- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic, The Netherlands.
- Anselin, L., Le Gallo, J., Jayet, H., 2008. Spatial panel econometrics. In: Matyas, L., Sevestre, P. (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Springer, Berlin, pp. 625–660. Chapter 19.
- Baicker, K., 2005. The spillover effects of state spending. *J. Public Econ.* 89, 529–544.
- Baltagi, B.H., 2011. Spatial panels. In: Ullah, A., Giles, D.E.A. (Eds.), *Handbook on Empirical Economics and Finance*. Taylor & Francis Group, New York, pp. 397–434.
- Baltagi, B.H., Egger, P., Pfaffermayr, M., 2013. A generalized spatial panel data model with random effects. *Econometric Rev.* 32, 650–685.
- Baltagi, B., Levin, D., 1986. Estimating dynamic demand for cigarettes using panel data: the effects of bootlegging, taxation and advertising reconsidered. *Rev. Econ. Stat.* 48, 148–155.
- Baltagi, B., Levin, D., 1992. Cigarette taxation: raising revenues and reducing consumptions. *Struct. Change Econ. Dyn.* 3, 321–335.
- Baltagi, B.H., Song, S.H., Koh, W., 2003. Testing panel data regression models with spatial error correlation. *J. Econometrics* 117, 123–150.
- Beenstock, M., Felsenstein, D., 2007. Spatial vector autoregressions. *Spat. Econ. Anal.* 2, 167–196.
- Brueckner, J.K., 1998. Testing for strategic interaction among local governments: the case of growth controls. *J. Urban Econ.* 44, 438–467.
- Brueckner, J.K., Saavedra, L.A., 2001. Do local governments engage in strategic property tax competition? *Natl. Tax J.* 54, 203–229.
- Case, A., Hines, J.R., Rosen, H.S., 1993. Budget spillovers and fiscal policy interdependence: evidence from the states. *J. Public Econ.* 52, 285–307.
- Elhorst, J., 2005. Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis* 37, 85–106.
- Elhorst, J., 2010. Spatial panel data models. In: Fischer, M.M., Getis, A. (Eds.), *Handbook of Applied Spatial Analysis: Software tools, Methods and Applications*. Springer, Heidelberg, Dordrecht, London, New York, pp. 377–407. Chapter C.2.
- Fingleton, B., 2008. A generalized method of moments estimators for a spatial panel model with an endogenous spatial lag and spatial moving average errors. *Spat. Econ. Anal.* 3, 27–44.
- Horn, R., Johnson, C., 1985. *Matrix Algebra*. Cambridge University Press.
- Jenish, N., Prucha, I., 2012. On spatial processes and asymptotic inference under near-epoch dependence. *J. Econometrics* 170, 178–190.
- Kapoor, N.M., Kelejian, H.H., Prucha, I.R., 2007. Panel data models with spatially correlated error components. *J. Econometrics* 140, 97–130.
- Kelejian, H., Piras, G., 2014. Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Reg. Sci. Urban Econ.* 140–149.
- Korniotis, G.M., 2010. Estimating panel models with internal and external habit formation. *J. Bus. Econom. Statist.* 28, 145–158.
- Lee, L.F., Yu, J., 2010a. Some recent developments in spatial panel data models. *Reg. Sci. Urban Econ.* 40, 255–271.
- Lee, L.F., Yu, J., 2010b. Estimation of spatial autoregressive panel data models with fixed effects. *J. Econometrics* 154, 165–185. 2010.
- Lee, L.F., Yu, J., 2012a. Spatial panels: random components vs. fixed effects. *Internat. Econom. Rev.* 53, 1369–1412.
- Lee, L.F., Yu, J., 2012b. QML estimation of spatial dynamic panel data models with time varying spatial weights matrices. *Spat. Econ. Anal.* 7, 31–74.
- Lee, L.F., Yu, J., 2014. Efficient gmm estimation of spatial dynamic panel data models. *J. Econometrics* 180, 174–197.
- Lee, L.F., Yu, J., 2015. Spatial panel data models. In: Baltagi, B. (Ed.), *Oxford Handbook of Panel Data*. Oxford University Press.
- Lee, L.F., Yu, J., 2016. Identification of spatial panel Durbin models. *J. Appl. Econometrics* 31, 133–162.
- Mutl, J., 2006. *Dynamic Panel Data Models with Spatially Correlated Disturbances*. (Ph.D. thesis), University of Maryland, College Park.
- Qu, X., Lee, L.F., 2015. Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *J. Econometrics* 184, 209–232.
- Qu, X., Wang, X., Lee, L.F., 2016. IV estimation of a spatial dynamic panel model with endogenous spatial weights when T is small. Working paper.

- Rincke, J., 2010. A commuting-based refinement of the contiguity matrix for spatial models, and an application to local police expenditures. *Reg. Sci. Urban Econ.* 40, 324–330.
- Rothenberg, T.J., 1971. Identification in parametric models. *Econometrica* 39, 577–591.
- Su, L., Yang, Z., 2015. QML estimation of dynamic panel data models with spatial errors. *J. Econometrics* 185, 230–258.
- 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. *J. Econometrics* 146, 118–134.
- Yu, J., de Jong, R., Lee, L.F., 2012. Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration. *J. Econometrics* 167, 16–37.
- Yu, J., Lee, L.F., 2010. Estimation of unit root spatial dynamic data models. *Econometric Theory* 26, 1332–1362.