Spatial Dynamic Models with Intertemporal Optimization

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree
Doctor of Philosophy in the Graduate School of The Ohio State
University

By

Hanbat Jeong, B.A., M.A.

Graduate Program in Economics

The Ohio State University

2019

Dissertation Committee:

Lung-fei Lee, Advisor

Jason Blevins

Robert de Jong

Abstract

My dissertation research introduces new econometric model specifications for a spatial panel data set describing intertemporal strategic interactions of forward-looking economic agents. We assume that each agent in our economy has its fixed geographic location. Estimation methods with their statistical properties for my new econometric models are studied. For each econometric model, I also conduct an empirical application to show how to implement the model. Our models are appropriate to analyze local governments' behaviors such as their expenditures and tax rates.

The first chapter, Spatial dynamic models with intertemporal optimization I: specification and estimation, firstly introduces a dynamic spatial interaction econometric model. There are n forward-looking agents of them each has a parametric linear-quadratic payoff, and interacting with neighbors through a spatial network. Considering a Markov perfect equilibrium (MPE), we derive a unique equilibrium equation and construct a new spatial dynamic panel data (SDPD) model. For estimation, we suggest mainly the quasi-maximum likelihood (QML) method. Asymptotic properties of the QML estimator are investigated. In a Monte Carlo study, we estimate the models parameters and compare the results with those from traditional SDPD models. The model is applied to an empirical study on counties public safety spending in North Carolina. We conduct impulse response and welfare analyses corresponding to changing exogenous characteristics in a region.

The second chapter, Spatial dynamic models with intertemporal optimization II: coevolution of economic activities and networks, introduces a panel data model describing agents' intertemporal optimization decisions with spatial interactions and spatial network evolution. The main purpose is to establish an estimation equation that explains spatial/time dependencies among observed agents' actions and endogenously changing spatial networks. To provide a theoretical foundation of our model, we establish a network interaction model for forward-looking agents. An agent's current action can affect his/her own and neighbors' future marginal payoffs via future spatial network links. Since parameters characterize an agent's payoff, a corresponding parametric econometric model is established. To estimate the model's parameters, we consider a GMM estimation method based on first-order conditions of agents' lifetime problems. Asymptotic properties of the GMM estimator are studied for statistical inferences. For practical uses, we introduce an estimation method for spatial network links using flow variables. Using our model, we study policy interdependence of U.S. states' health expenditures.

The third chapter, Spatial dynamic models with intertemporal optimization III: a dynamic Stackelberg game with spatial interactions, introduces a spatial dynamic panel data (SDPD) model explaining relationships between two types of forward-looking agents: a leader and multiple followers. They empirically represent the central and local governments. Hence, the main purpose of our model is to account for the intertemporal spatial interactions of them: (i) interactions between a leader and followers and (ii) interactions among followers. As an economic foundation of our estimation equation, we establish a dynamic Stackelberg game played on a spatial

network. Derived optimal actions for both types of agents lead to a spatial econometric model. Next, we introduce how to implement the quasi-maximum likelihood (QML) method for recovering parametric payoff functions of the two types of agents. Asymptotic and finite sample properties of the QML estimator are investigated. Last, we employ our model to examine (i) policy interdependence among U.S states' general expenditures and (ii) interrelations between their expenditures and grants from the federal government.

For my family

In particular, this is dedicated to my wife, Hyo Jin.

Acknowledgments

First of all, I would like to greatly thank my advisor, Lung-fei Lee, for his sincere guidance. I have been happy to discuss interesting research issues with him. Meeting with him at the Ohio State University is the greatest luck in my life. Second, I thank Jason Blevins, Stephen Cosslett, and Robert de Jong. As faculty members in econometrics, they have given fruitful suggestions for my works. Also, I thank Bruce Weinberg for having an opportunity to learn empirical works. Last, I appreciate former Lung-fei's students, Xiaoyi Han, Chih-Sheng Hsieh, Tuo Liu, Cheng Wei, Shi Wei, Shin-Yi Wu, Xingbai Xu, Chao Yang, and Kai Yang. I have learned a lot of things during my Ph.D. life.

Vita

2010	B.A. Korean Language and Literature, Sogang University
2013	M.A. Economics, Sogang University
2012-2013	R.A. Korea Productivity Center
2014	M.A. Economics, The Ohio State University
2013-present	Ph.D. Economics, The Ohio State University

Publications

Research Publications

Choi, I, and Jeong, H. " Model selection criteria for factor analysis: some new criteria and performance comparison". *Econometric Reviews*, forthcoming

Fields of Study

Major Field: Economics

Table of Contents

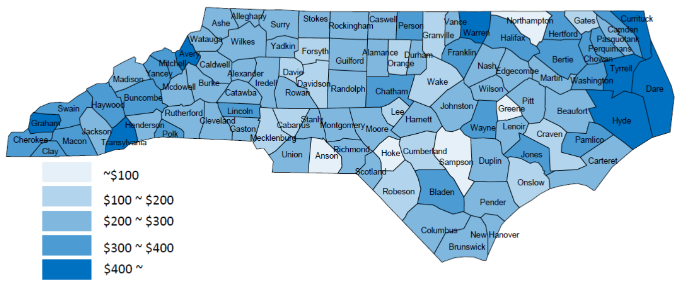
List of Tables

List of Figures

Chapter 1: Introduction

Recently, an economic researcher is facing a data-rich environment. In a large panel data set, we have been observed cross-sectional or time series dependencies in a lot of economic variables. If information in space is available and it involves individuals' characteristics, a spatial econometric model can be a tool that explains the role of space in those dependencies. For example, Case et al. (1993) find that a state government's expenditure is positively correlated with its neighbors' expenditures by using a spatial econometric model. In a cross-sectional setting, a linear spatial autoregressive (SAR) model is popular. Examples are Cliff and Ord (1973), Ord (1975), Anselin (1988), and Lee (2004, 2007). The linear SAR model can be considered as an equilibrium equation if we take a parametric linear-quadratic payoff function (Ballester et al. (2006), Calvo-Armengol, Patacchini and Zenou (2009), Chapter 4 in Jackson and Zenou (2014), and Ushchev and Zenou (2018)). Then, the dependent variables in a SAR model can represent agents continuous type optimal actions (or ourcomes), and their actions can be interrelated with spatial network matrices. Geographic locations of agents usually form spatial networks. If an agent is a local government, its action can be some specific expenditure or tax rates (i.e., local government's action). As an example, Figure shows public safety spending of county governments in North Carolina in 2015. [1]

Figure 1.1: Public safety spending of county governments in North Carolina in 2015



In my dissertation chapters, we extend a SAR model for a large panel data set by considering appropriate economic reasonings. Note that a large panel data set includes a lot of cross-section and time series units. With spatial interactions, hence, dynamic interactions (which mean the interactions of spatial units across different time periods) can be captured. Currently, there are various researches in panel data extensions of SAR models, which are called spatial dynamic panel data (SDPD) models. Examples are Kapoor et al. (2007), Baltagi et al. (2007), Yu et al. (2008), Lee and Yu (2010, 2012, 2014), Shi and Lee (2017), Qu et al. (2018), Han et al. (2019), and LeSage et al. (2019). Those SDPD models can be justified by an extended linear-quadratic payoff function with myopic behaviors. By basic features of a panel data set, a practitioner can capture dynamics of individuals actions (or outcomes). If

[1]If a local government is regarded as an economic agent and its action is a regional policy (e.g., tax rate, expenditure, and so on), the linear-quadratic payoff function represents a representative resident's utility. In this setting, hence, we can interpret that a local government tries to maximize a common person's utility in its region. For more details, refer to Brett and Pinkse (2000), Brueckner (2003) and Revelli (2005).

we consider rational economic agents, observed actions from a panel data set might come from forward-looking behaviors. Hence, we try to construct panel data spatial econometric models based on forward-looking agents' behaviors with the extended parametric linear-quadratic payoff specification. That is, our three model specifications belong to structural econometric models which are products of Lucas critique (1976).

First, we introduce a basic dynamic spatial interaction econometric model. There are n forward-looking agents of them each has a parametric linear-quadratic payoff, and interacting with neighbors through a spatial network. In the first model specification, each economic agent is assumed to have its fixed and innate geographic location. That is, a spatial network is constructed by only geographical arrangements implying that a spatial network is time-invariant and exogenous. Hence, a local government can be a good example of an agent. For each period, for example, a local government makes its fiscal decision by considering neighbors' current and expected future decisions with their demographic characteristics. Considering a Markov perfect equilibrium (MPE), we derive a unique equilibrium equation and construct a new SDPD model. As a refined version of subgame perfect Nash equilibrium, an MPE is popular in structural econometric modeling because an optimal action under the MPE concept is only a (time-invariant) function of state variables. Deriving the MPE equation relies on solving algebraic matrix Riccati equations. For details in MPE, refer to Maskin and Tirole (1988a, 1988b, 2001) and Chapter 7.6 in Ljungqvist and Sargent (2012). Since the derived optimal actions (dependent variables) are linear in state variables, we can fully characterize the correlation structure of dependent variables.

Hence, we can derive the likelihood function for estimation. For estimation, we suggest mainly the quasi-maximum likelihood (QML) method. Asymptotic properties of the QML estimator are investigated. Since the resulted models dependent variables are still linear in disturbances, we can apply the same asymptotic technique in Yu et al. (2008), which is a panel data extension of the martingale difference central limit theorem (MD-CLT) for a linear-quadratic form (Kelejian and Prucha (2001)). Since the model equation includes incidental parameters showing unobserved individual and time characteristics, we need to adjust the asymptotic biases from them. A famous research about the asymptotic bias in dynamic panel models (without interactions) is Hahn and Kuersteiner (2002). As Lee and Yu (2010, 2014), our bias correction is based on evaluating the expected values of scores at the true parameter values. In a Monte Carlo study, we estimate the model's parameters and compare the results with those from traditional SDPD models. The model is applied to an empirical study on counties' public safety spending in North Carolina. We compare our estimation results with those of Yang and Lee (2017).

Second, we introduce a spatial panel data model describing forward-looking agents' decisions with spatial interactions and spatial network evolution. The main purpose is to establish an estimation equation that explains spatial/time dependencies among observed agents' actions and endogenously changing spatial networks. There are various researches in SAR models (or SDPD models) with endogenous spatial networks. Examples are Kelejian and Piras (2014), Qu and Lee (2015), Han and Lee (2016), Hsieh and Lee (2017), Qu et al. (2017), Johnsson and Moon (2017), Kuersteiner and Prucha (2018), and Han, Hsieh, and Ko (2019). However, our second model specification firstly considers a SDPD model specification with time-varying endogenous spatial networks by the forward-looking agent assumption. We establish a network interaction model for forward-looking agents. The forward-looking agent assumption can yield reasonable economic interpretations of having time-varying endogenous spatial networks. An agent's current action can affect his/her own and neighbors' future marginal payoffs via future spatial network links. This feature is similar to habit formation models in macroeconomics (e.g., Fuhrer (2000)). Since parameters characterize an agent's payoff, a corresponding parametric econometric model is established. In contrast to the first model specification, deriving the agents' optimal actions is challenging due to highly nonlinearity. To estimate the model's parameters, hence, we consider a GMM estimation method based on first-order conditions of agents' lifetime problems. This method is motivated by Hansen and Singleton (1982). Consistency and the asymptotic distribution of the GMM estimator are studied. To establish the law of large numbers (LLN) for consistency, we employ the notion of spatial-time near epoch dependence (NED) in Jenish and Prucha (2012) and Qu et al. (2017) since dependent variables of our model (optimal actions) might not be a linear function of disturbances. To study the asymptotic distribution of the GMM estimator, we consider asymptotic properties of the main statistics conditional on unspecified exogenous components stemming from spatial network formation. That is, we utilize a central limit theorem (CLT) for a linear quadratic form of martingale difference arrays with the C-stable convergence concept established in Kuersteiner and Prucha (2013, 2018). This CLT belongs to a CLT with random norming. A basic idea can be seen in Chapter 25.2 in Davidson (1994). To test whether spatial networks evolve exogenously or not, the Wald test can be applied. By Theorem 4 in Kuersteiner and

Prucha (2018), the Wald test statistic asymptotically follows the unconditional chi-square distribution. For practical uses, by considering formulations of gravity models in international trade literature (e.g., Anderson and Wincoop (2003)) and Qu and Lee (2018), we introduce an estimation equation for spatial network links using flow variables. Using our model, we study policy interdependence of U.S. states' health expenditures.

Motivated by Chapter 19 in Ljungqvist and Sargent (2012), third, we introduce a spatial dynamic panel data (SDPD) model explaining the relationships between two types of forward-looking agents: a leader and multiple followers. In a practical application, a leader can represent the central U.S. government while followers can be state governments. The main purpose of the third model is to explain intertemporal spatial interactions of them: (i) interactions between a leader and followers and (ii) interactions among followers. As an economic foundation of our estimation equation, we establish a dynamic Stackelberg game played on a spatial network. As a review of dynamic Stackelberg game models, refer to Li and Sethi (2017). Under the rational expectation equilibrium, derived optimal actions for both types of agents lead to a spatial econometric model. Based on the induced correlation structure of the dependent variables, we derive the log likelihood function and introduce how to apply the quasi-maximum likelihood (QML) method for recovering parametric payoff functions of the two types of agents. Asymptotic and finite sample properties of the QML estimator are investigated.

This thesis proceeds as follows. Chapter 2 introduces estimation and specification methods if a spatial network is time-invariant. Chapter 3 considers the case of time-varying endogenous spatial networks. Chapter 4 deals with intertemporal spatial interactions among two types of economic agents: a leader and multiple followers. For each model specification, a corresponding empirical application will be introduced.

Chapter 2: Spatial dynamic models with intertemporal optimization I: Specification and estimation

2.1 Introduction

Interactions among rational economic agents are characterized by a network (a spatial weights or socio-economic matrix). Since rational agents might be forward-looking instead of myopic, we focus on their behaviors by considering intertemporal optimization. Specification on forward-looking agents' decision-making with network interactions will be introduced. We formulate an econometric model for recovering economic agents' payoff. The econometric model is a new spatial dynamic panel data (SDPD) model, which can be estimated by panel data and it can be regarded as a product of Lucas critique (1976).[2] For the econometric model, identification, estimation, and asymptotic properties of estimators are investigated. Using the new SDPD model, empirical economists can conduct (i) forecasting on future economic activities, (ii) impulse response analyses, and (iii) welfare and counterfactual analyses. As an application of our econometric model, we study counties' public safety spending competition. We recover key parameters describing counties' decision-making and

[2] It means our econometric model is a structural model and its interpretations do not rely on just statistical relationships among economic variables.

compare estimation results with those from traditional models. We give various and fruitful policy implications from this research.

Three contributions will be established in this paper. The first is a theoretical one. We introduce a forward-looking agent's decision-making model with network interactions. There are n economic agents in the economy and their interactions are characterized by an $n \times n$ socio-matrix, which is assumed to be time-invariant and known to agents as well as econometricians. An outcome of an agent's economic activity is assumed to be a continuous one. For example, players select how much time or effort on some economic activity. In order to specify agent's payoff, we take a parametric linear-quadratic payoff function (Ballester et al. (2006) and Calvo-Armengol, Patacchini and Zenou (2009)). The most notable advantages in taking this payoff structure are (i) easily characterizing an equilibrium and (ii) specifying agent's payoff by some key parameters, in addition that a linear-quadratic payoff function might provide a good approximation to an underlying nonlinear function. Chapter 4 in Jackson and Zenou (2014) provides a review for that structure. Based on the payoff function, an agent's choice problem is to maximize his/her discounted lifetime payoff by intertemporally choosing his/her effort. An agent will face future uncertainty and form expectation for it. In addition to future economic shocks, another source of uncertainty is due to unknown future changing exogenous environments of an economy. From that, we describe how an agent forms expectations for series of future decisions and possibly changing exogenous environments.

To derive a complete model, our next step is characterizing an equilibrium under a game setting. An "equilibrium" is a result of rationality of economic agents. Forward-looking decisions on an equilibrium realize the "rationality" of economic agents. For this, we employ a Markov perfect equilibrium (MPE). In the MPE, agents' current decisions depend only on their payoff relevant previous actions, and backward induction can be applied to specify the equilibrium. Under some stability conditions, we have agents' optimizing values, which are results from solving dynamic (differential) games problems, and they are linear-quadratic. In consequence, the vector of agents' equilibrium decisions becomes a unique Nash equilibrium (NE) solution of a linear system. The derived equilibrium equations describe the dynamics of individuals' forward-looking decisions by reflecting series of (discounted) expected future actions and exogenous characteristics in a dynamic NE game setting. As the implied model equations are linear in outcomes, we have a unique NE equilibrium so to obtain a bijective mapping from the model to a likelihood function for estimation.[3]

Second, we deliver an econometric contribution. The popular spatial autoregressive (SAR) model from Cliff and Ord (1973), Ord (1975), Anselin (1988) and Lee (2004, 2007) can be considered as an equilibrium equation of a static quadratic utility model with network interactions. In the literature, panel data can capture the dynamics of individuals' decisions (but mostly without interactions). For spatial interaction issues, there are fruitful studies with spatial dynamic panel data (SDPD) models. Kapoor et al. (2007), Baltagi et al. (2007), Yu et al. (2008), Lee and Yu (2010, 2014) are papers in this area. For the various SDPD models, Lee and Yu (2015) provide a review. Those SDPD models can only be justified by myopic behaviors. In this paper, the designed framework analyzes agents' forward-looking behaviors. With proper panel data, revealed economic activities might be results of dynamic optimization instead of considering only current payoffs. Our derived equilibrium

[3]For this, see Section 8 in Amemiya (1985).

equation provides a new estimable SDPD model. Our SDPD nests traditional SDPD models as special cases if economic agents are myopic.

For estimation, we suggest the quasi-maximum likelihood (QML) method. Identification of the model and asymptotic properties (consistency and asymptotic normality) of the QML estimator are investigated. Because our specification includes individual and time fixed effects, which are infinite incidental parameters and, in consequence, may lead to asymptotic biases in estimates, a bias correction for the QML estimator is studied. Estimating the individual and time dummies relies on residuals, so their asymptotic distributions are affected by convergence rates of the QML estimator of the main parameters. We observe using residuals based on the bias-corrected QML estimator has a mild condition for ratios of n and T relative to using those from the QML estimator without bias-correction. As an alternatively simpler but inefficient estimation, the nonlinear two-stage least squares (NL2S) method is also briefly introduced. Monte Carlo simulations are conducted to evaluate (i) finite sample performance of the QML estimator and its bias correction and (ii) misspecification, when a traditional SDPD specification is taken for estimation as if agents were not forward-looking, i.e., myopic. We find that the QML estimator and its bias correction show reliable performance in small samples. We observe that significant misspecification errors on estimators would appear even for large samples, as the traditional SDPD specification is mistakenly used. When selecting a time-discounting factor, we suggest considering likelihood measures (e.g., sample log-likelihood) if a signal is high with sufficiently many observations. The NL2S estimator shows relatively small biases but does not provide efficient estimates compared to those of the QML estimator.

Finally, we give an empirical study with policy implications on counties' public safety spending. In this application, an economic agent is a local government, and its decision variable is the public safety spending for a county. Yang and Lee (2017) provide a theoretical model for this issue and apply it to cities in North Carolina. They find strong free-riding effects: there are strategic interactions among local governments and, which induce a negative relationship between a city's public safety spending and its neighbors'. In this paper, we revisit this issue with an extended panel data set. We estimate structural parameters using our dynamic interaction model and compare the estimation results with those from the traditional SDPD model. In explaining the spillover effects of local governments' public safety spending, our intertemporal SAR specification turns out to be more statistically favorable than the traditional SDPD model. We find some evidence of persistency of public safety spending, positive diffusion effects from previous neighbors' decisions, positive effects of own total revenue, and negative externalities from neighboring total revenues, but no significant contemporaneous spilled over effects. From the recovered counties' payoff function, we also investigate cumulative effects in the MPE and conduct impulse response analyses corresponding to changing exogenous characteristics in a region. An overshooting impact in the sense of a negative neighboring revenue effect is observed.[4] In the welfare analysis, we observe giving subsidy to the county which has a small number of neighbors turns out to be the most effective policy in the sense of public safety spending.

[4]It means that the contemporaneous negative revenue effect converts to the positive effect after some periods and finally decays.

2.2 A spatial dynamic game with intertemporal optimization

In this section, we give a theoretical economic foundation and suggest a corresponding econometric model. First, we review some motivating literature on the spatial autoregressive model in a cross-sectional setting and then its extension to dynamic panel data model in the econometric literature. From these, we motivate our formulation of a dynamic spatial autoregressive model with agents' decision processes which take into account intertemporal consequences.

2.2.1 Literature review: spatial dynamic panel models and myopic choices

We assume there are n economic agents in an economy and they choose a continuous type economic activity. A tax rate or public spending can be a good example of a continuous economic activity when an agent is a local government. There are interactions among agents' activities through a certain network relationship. Since there are n economic agents, a network is characterized by an $n \times n$ matrix W_n with prespecified non-negative entries (links), which can be formed by social, geographical and/or economic distances. All the diagonal elements of W_n are assumed to be zero to exclude self-influence. From economic reasoning, a way of modeling agents' interactions is to formulate agents' decisions in a game setting. Given existing network connections in W_n, one may specify a linear-quadratic payoff function for each individual (e.g., Ballester et al. (2006) and Calvō-Armengol et al. (2009)) with

$$u_i \left(Y_n, \eta_{it} \right) = \eta_i y_i + \lambda_0 y_i w_{i.} Y_n - \frac{1}{2} y_i^2 \tag{2.1}$$

where $Y_n = (y_1, \cdots, y_n)'$ denotes the vector of agents' decisions (activities, outcomes), η_i is i's exogenous heterogeneity containing his/her exogenous characteristics, $w_{i.}$

denotes the i^{th} row of W_n, and λ_0 determines the strength of strategic interaction among agents while elements of W_n represent relative strength if there are interactions. The first part, $\eta_i y_i$, describes a choice-specific benefit from i's characteristics in his index η_i. Increasing η_i by one unit leads to rising i's marginal payoff $\frac{\partial u_i(Y_n,\eta_{it})}{\partial y_i}$. From i's perspective, decisions by others linked to i will be strategic complements if $\lambda_0 > 0$, strategic substitutes if $\lambda_0 < 0$, and no interactions when $\lambda_0 = 0$. The last quadratic term represents a cost for y_i being taken. Let $\eta_n = (\eta_1, \cdots, \eta_n)'$, $X_n = (x_1, \cdots, x_n)'$ where $x_i = (x_{i1}, \cdots, x_{iK})'$ denotes agent i's observed characteristics, and $\mathcal{E}_n = (\epsilon_1, \cdots, \epsilon_n)'$ be an $n \times 1$ vector of unobservable (for econometrician) components. By specifying η_n as a regression function, $\eta_n = X_n\beta_0 + \mathcal{E}_n$, agents' optimized decisions in a perfect information game give rise to the spatial autoregressive (SAR) model

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + \mathcal{E}_n \tag{2.2}$$

where Y_n is the vector of Nash equilibrium (NE). The system (2.2) can have a unique NE and can be stable under the assumption that $\|\lambda_0 W_n\| < 1$ for some matrix norm $\|\cdot\|$.

The SAR model provides a static model for strategic interactions with a given network. On the other hand, with various panel data sets, one can go beyond the static setting and may track the dynamics of individual's decisions. With panel data, observed decisions of individuals might come from dynamic optimization. Let $\{Y_{nt}, X_{nt}\}$ be a set of panel data where $Y_{nt} = (y_{1t}, \cdots, y_{nt})'$ stands for a vector of individuals' decisions at time t and $X_{nt} = (x_{1t}, \cdots, x_{nt})'$ denotes an $n \times K$ matrix of t^{th}-period observable (for econometricians) exogenous variables. Existing spatial panel data (SDPD) models in the literature (e.g., Kapoor et al. (2007), Baltagi et al. (2007), Yu et al. (2008), Lee and Yu (2010, 2014)) actually take a similar form as the SAR model (2.2) but with additional time lags $Y_{n,t-1}$, diffusion $W_n Y_{n,t-1}$ and individual and time fixed effects:

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + \mathcal{E}_{nt} \tag{2.3}$$

where $\mathbf{c}_{n0}$ is an n-dimensional column vector of individual fixed effects, α_{t0} captures the t^{th}-period time specific effect with l_n being an n-dimensional vector of ones. This equation can be justified by a game framework with agent i's payoff

$$u_i\left(Y_{nt}, Y_{n,t-1}, \eta_{it}\right) = \eta_{it} y_{it} + \lambda_0 y_{it} w_{i.} Y_{nt} + \rho_0 y_{it} w_{i.} Y_{n,t-1} - c\left(y_{it}, y_{i,t-1}\right) \tag{2.4}$$

and $c\left(y_{it}, y_{i,t-1}\right) = \frac{\gamma_0}{2}\left(y_{it} - y_{i,t-1}\right)^2 + \frac{1-\gamma_0}{2} y_{it}^2$ where $0 < \gamma_0 < 1$.[5] The η_{it} denotes the $t^{th}-$period index of heterogeneity of agent i containing those exogenous characteristics, which might evolve over time.[6] The third component, $\rho_0 y_{it} w_{i.} Y_{n,t-1}$, describes agent's learning process. Learning or adopting new technology is a time-consuming process as an agent has to spend some time to understand his/her friends'

[5]In this paper, we use the normalized payoff due to identification easiness. We can consider the following alternative cost specification $\dot{c}\left(y_{it}, y_{i,t-1}\right) = \frac{\gamma_{1,0}}{2}\left(y_{it} - y_{i,t-1}\right)^2 + \frac{\gamma_{2,0}}{2} y_{it}^2$ where $0 < \gamma_{1,0},\ \gamma_{2,0} < 1$. Then, the first order conditions of maximizing the per period payoff can yield $\left(\gamma_{1,0} + \gamma_{2,0}\right) Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_{1,0} Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + \mathcal{E}_{nt}$. It's impossible to identify all the parameters at the same time.

Note that an affine transformation preserves cardinal preferences realized by Von Neumann-Morgenstern utilities. If we consider the payoff normalized by $\frac{1}{\gamma_{1,0}+\gamma_{2,0}}$, we have structural parameters are normalized by $\frac{1}{\gamma_{1,0}+\gamma_{2,0}}$.

[6]In this framework, η_{it} represents i's $t^{th}-$period "overall" characteristic by including (i) agent i's own exogenous characteristics (time-invariant and/or time-variant), (ii) his/her rivals' characteristics combined with elements in W_n showing externalities and (iii) common economic shocks globally affecting all individuals' decision-making.

past decisions and accommodate to the new environment innovated by new technologies.[7] In this setting, individual's learning comes from his/her recent past neighboring decisions.[8] The parameter ρ_0 determines how past neighboring actions affect agent i's current decision. If $\rho_0 > 0$ and agent j (who is an i's friend) increased his/her effort yesterday, agent i may choose a higher level of effort today (because $\frac{\partial^2 u_i(Y_{nt}, Y_{n,t-1}, \eta_{it})}{\partial y_{i,t-1} \partial y_{it}} = \rho_0 w_{ij} \geq 0$). With $\rho_0 < 0$ if agent j increased his/her effort yesterday, agent i tends to select a low level of effort (since $\frac{\partial^2 u_i(Y_{nt}, Y_{n,t-1}, \eta_{it})}{\partial y_{i,t-1} \partial y_{it}} = \rho_0 w_{ij} \leq 0$). The fourth part, $c\left(y_{it}, y_{i,t-1}\right)$, represents a cost of i's decision.[9] In our framework, $c\left(y_{it}, y_{i,t-1}\right)$ consists of two parts: (i) dynamic adjustment cost, $\frac{\gamma_0}{2}\left(y_{it} - y_{i,t-1}\right)^2$, and (ii) agent's cost $\frac{1-\gamma_0}{2} y_{it}^2$ of selecting activity level y_{it}. If there is a large gap between i's current decision y_{it} and his/her recent previous decision $y_{i,t-1}$, the term $\frac{\gamma_0}{2}\left(y_{it} - y_{i,t-1}\right)^2$ may give a high penalty on i's payoff, therefore, it may cause persistency on i's behavior. The parameter γ_0 captures the persistent tendency of agents'

[7] In the case of policy effect analyses, this part also shows policy lags. i.e., affecting neighboring policies on my city's one is time-consuming.

[8] It means that agent's learning follows a Markov process. However, the entire history of past decisions could be relevant to the agents' current choices. In this case, agents' learning process is a Polya process. For the details, refer to Liu et al. (2010). They study peer group effects in laboratory experiments based on Milgrom and Roberts' (1982) entry limit pricing game and use two specifications for agents learning: (i) A Markov model and (ii) a Polya model.

[9] In this paper, we adopt the specification of the quadratic adjustment cost (the famous study about that is Kennan (1979)). Alternatively, Engsted and Haldrup (1994) employ the following quadratic adjustment cost for analyzing the demand for labor,

$$\gamma_0 (l_t - l_t^*)^2 + \left(l_t - l_{t-1}\right)^2 \tag{2.5}$$

where l_t is the $t-$period labor demand, l_t^* denotes the steady-state level of the variable l_t and parameter γ is the relative cost parameter.

However, if we consider $\frac{1-\gamma_0}{2}(y_{it} - y^*)^2$ where y^* denotes a time-invariant social norm showing agents' stereotype, identification of y^* is difficult (in the sense of econometrics). In case of an econometric model based on a static framework, y^* will be absorbed in the intercept. In the case of dynamic one, it will be a part of individual fixed effects.

choices. The term $\frac{1-\gamma_0}{2} y_{it}^2$ is a kind of social cost, which prevents an agent from choosing an extremely high effort.

At time t, agent i maximizes his/her payoff $u_i\left(y_{it}, Y_{-i,t}, Y_{n,t-1}, \eta_{it}\right)$ where $Y_{-i,t} = \left(y_{1t}, \cdots, y_{i-1,t}, y_{i+1,t}, \cdots, y_{nt}\right)'$. It means that agent i knows the optimum choices $Y_{-i,t}$ of others. The first order conditions of such optimization problems give equation (2.3) which characterizes a NE at time t. Since each agent only maximizes his/her per period payoff, this model assumes agents are myopic in their decisions. In this project, we attempt to go beyond myopic behaviors of agents. We consider an agent's intertemporal choice problem and characterize the NE in an infinite horizon in order to derive an estimating equation.[10] Under the linear-quadratic payoff (2.4), this will result in a new spatial dynamic panel data (SDPD) model.

2.2.2 Intertemporal choices

The main feature of our model is that agents are not myopic but rational to expect what would happen in the future based on their available information. An agent considers a series of his/her (expected) future payoffs when he/she makes a current decision based on currently available information, and he/she expects that future realized decisions of all agents will result in an NE. Let $\mathcal{B}_{it}$ be the $t^{th}-$period information set of agent i's perceivable events and it is defined by

$$\mathcal{B}_{it} = \sigma\left(\{y_{js}\}_{j=1}^{n}\big|_{s-\infty}^{t-1}, \{\eta_{js}\}_{j=1}^{n}\big|_{s-\infty}^{t}\right),$$

[10] The derivation can also be done for a finite horizon case if one knows the terminal period.

where $\sigma(\cdot)$ denotes the $\sigma-$field[11] generated by the argument inside. This specification is assumed to be a complete information game from the past to the current period t with uncertainty only for future periods. The η_{it} contains both time-invariant η_i^{iv} and time-varying η_{it}^{v} individual characteristics (some of them might not be observable by econometricians).

To understand the implication of intertemporal choices on spatial interactions, it will be simpler to consider an intertemporal choice problem with two periods. Denote $\eta_{nt} = (\eta_{1t}, \cdots, \eta_{nt})'$ for each t. Given (Y_{n0}, η_{n1}), agent i $(i = 1, \cdots, n)$ is assumed to maximize the expected discounted intertemporal payoff for $t = 1$ and 2: at $t = 1$, $u_i(Y_{n1}, Y_{n0}, \eta_{i1}) + \delta E\left(u_i(Y_{n2}, Y_{n1}, \eta_{i2})|\mathcal{B}_{i1}\right)$; and at $t = 2 : u_i(Y_{n2}, Y_{n1}, \eta_{i2})$, by sequentially selecting y_{it} for $t = 1, 2$. By considering the subgame perfect NE (SPNE) economic activities, the agent i's equilibrium decision at the period 1 is

$$\begin{aligned} y_{i1}^*(Y_{n0}, \eta_{n1}) &= \gamma_0 y_{i0} + \rho_0 w_{i.} Y_{n0} + \lambda_0 w_{i.} Y_{n1}^*(Y_{n0}, \eta_{1n}) \\ &\quad + \delta\left(\Delta_i e_i' A_n^{trad} Y_{n1}^*(Y_{n0}, \eta_{n1}) - \gamma_0 y_{i1}^*(Y_{n0}, \eta_{n1})\right) \\ &\quad + \eta_{i1} + \delta \Delta_i e_i' S_n^{-1} E(\eta_{2n}|\mathcal{B}_{i1}) \end{aligned}$$

where $A_n^{trad} = S_n^{-1}\left(\gamma_0 I_n + \rho_0 W_n\right)$ and $\Delta_i = \frac{\partial e_i' S_n^{-1}(\gamma_0 I_n + \rho_0 W_n) Y_1}{\partial y_{i1}} = e_i' A_n^{trad} e_i$. The quantity Δ_i means a marginal change of the future expected equilibrium decisions of i corresponding to changing y_{i1}.[12] Let $\Delta_n = Diag(A_n^{trad})$. Then, the NE vector at

[11] In a measure theoretical interpretation, the sequence of $\mathcal{B}_{it}$'s is a filteration on $(\Omega, \mathcal{B}_i)$. Ω contains all possible outcomes and $\mathcal{B}_i$ can be defined by

$$\mathcal{B}_i = \sigma\left(\{y_{js}\}_{j=1}^n |_{s-\infty}^{\infty}, \{\eta_{js}\}_{j=1}^n |_{s-\infty}^{\infty}\right).$$

Then, for $t_1 \leq t_2$, $\mathcal{B}_{i,t_1} \subseteq \mathcal{B}_{i,t_2} \subseteq \mathcal{B}_i$, which means agents' knowledge increases over time.

[12] Since there is no additional future period, the expected NE decisions at $t = 2$ are $E\left(Y_{n2}^*(Y_{n1}, \eta_{n2})|\mathcal{B}_{i1}\right) = A_n^{trad} Y_{n1} + S_n^{-1} E\left(\eta_{n2}|\mathcal{B}_{i1}\right)$ for all i.

$t = 1$ can be characterized by a modified SAR equation:

$$Y_{n1}^*(Y_{n0}, \eta_{n1}) = \lambda_0 W_n Y_{n1}^*(Y_{n0}, \eta_{n1}) + \delta \left[\Delta_n A_n^{trad} - \gamma_0 I_n\right] Y_{n1}^*(Y_{n0}, \eta_{n1}) + \left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n0} + \eta_{n1} + \delta \Delta_n S_n^{-1} E_1(\eta_{2n})$$

where $E_t(\cdot)$ denotes the mathematical conditional expectation on $(Y_{n,t-1}, \eta_{nt})$ at $t = 1$ and 2. Let $R_{n1} = (1 + \delta\gamma_0) I_n - \lambda_0 W_n - \delta \Delta_n A_n^{trad}$. By assuming invertibility for R_{n1}, the unique NE can be characterized as

$$Y_{n1}^*(Y_{n0}, \eta_{n1}) = R_{n1}^{-1} \left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n0} + R_{n1}^{-1} \left(\eta_{n1} + \delta \Delta_n S_n^{-1} E_1(\eta_{2n})\right). \tag{2.6}$$

From equation (2.6), we see that taking into account the expected outcomes in the second period, as $\delta > 0$, it brings in the additional spatial influence $\delta \Delta_n A_n^{trad} Y_{n1}^*(Y_{n0}, \eta_{n1})$ and the time influence $\delta \gamma_0 I_n$ due to their effects on possible future outcomes.

Based on recursion, we extend this two-period model to an infinite horizon model. At each time t, given $Y_{n,t-1} = (y_{1,t-1},, \cdots, y_{n,t-1})'$ and $\eta_{nt} = (\eta_{1t}, \cdots, \eta_{nt})'$, each agent, say i, is assumed to maximize the expected discounted intertemporal payoff

$$u_i\left(y_{it}, Y_{-i,t}, Y_{n,t-1}, \eta_{it}\right) + \sum_{s=1}^{\infty} \delta^s E\left(u_i\left(Y_{n,t+s}, Y_{n,t+s-1}, \eta_{i,t+s}\right) | \mathcal{B}_{it}\right) \tag{2.7}$$

by selecting y_{it}. The time-discounting factor $\delta \in [0, 1)$ is introduced to give weights on agent's future choices. The main reason considering an infinite horizon problem is to allow that possibility, and in that case one can get a same functional form (over time periods) of an estimable equation with given information.[13]

[13]From a panel data set, in practice, a researcher might not know initial and terminal periods of agents' decision-making. When we consider a time-invariant equation as an estimating model, utilizing that model is available without concerning specific time period t relative to a finite terminal period.

In perspective of economics, employing an infinite horizon model is prevalent in a lot of theoretical and/or empirical studies. Even though agents actually have a terminal decision-making period, they might keep the same pattern of decision-making at the terminal period because of (i) leaving a bequest, (ii) keeping a nice reputation and so on.

2.2.3 Nash equilibrium characterization

In this subsection, we characterize the NE. In the infinite horizon model, the Markov perfect equilibrium (hereafter, MPE) characterizes the equilibrium strategies of all agents as best responses to one another and helps to yield a unique equilibrium equation. "Markov" means that agent i's t^{th}-period optimal strategy only depends on the state variables $(Y_{n,t-1}, \eta_{nt})$ and does not rely on other earlier parts of its histories (Maskin and Tirole (1988a)). "Perfect" means that the NE constructs an optimizing behavior of each individual for all possible uncertain future states. Hence, an MPE is a refined version of subgame perfect NE. As its old definition is "closed-loop equilibrium", the definition of the MPE involves a dynamic programming equation (the Bellman equation).[14] Since the $t^{th}-$period optimal decisions only depend on $(Y_{n,t-1}, \eta_{nt})$ and, under the Markov assumption other past histories and exogenous characteristics are irrelevant to the current decision-making, $E(\cdot|\mathcal{B}_{it}) = E(\cdot|Y_{n,t-1}, \eta_{nt})$ for all $i = 1, \cdots, n$. Hence, we can simply define the conditional expectation operator $E_t(\cdot)$ by $E_t(\cdot) = E(\cdot|Y_{n,t-1}, \eta_{nt})$. Also, time itself is not payoff-relevant, so we can drop the subscript "t" from agents' optimal policy functions $y^*_{it}(Y_{n,t-1}, \eta_{nt})$ (for $i = 1, \cdots, n$) in the definition of MPE.

Definition 2.2.1 (Markov perfect equilibrium) *A MPE will be a set of value functions $V_i(\cdot)$ $(i = 1, \cdots, n)$ and a set of policy functions $f_i(\cdot)$ $(i = 1, \cdots, n)$ such that*

(i) (Markov strategy) $y^*_{it}(Y_{n,t-1}, \eta_{nt}) = f_i(Y_{n,t-1}, \eta_{nt})$,

[14] For more information in MPE, refer to Maskin and Tirole (1988a, 1988b, 2001) and Chapter 7.6. in Ljungqvist and Sargent (2012).

(ii) given $f_1, \cdots, f_{i-1}, f_{i+1}, \cdots, f_n$, V_i satisfies the Bellman equation

$$V_i(Y_{n,t-1}, \eta_{nt}) = \max_{y_{it}} \left\{ \begin{array}{c} u_i\left(y_{it}, Y^*_{-i,t}(Y_{n,t-1}, \eta_{nt}), Y_{n,t-1}, \eta_{it}\right) \\ +\delta E_t\left(V_i(y_{it}, Y^*_{-i,t}(Y_{n,t-1}, \eta_{nt}), \eta_{n,t+1})\right) \end{array} \right\} \tag{2.8}$$

where

$Y^*_{-i,t}(.,.) = \left(y^*_{1t}(.,.), \cdots, y^*_{i-1,t}(.,.), y^*_{i+1,t}(.,.), \cdots, y^*_{nt}(.,.)\right)'$, *and*

(iii) (principle of optimality) the policy function $f_i(\cdot) = y^*_{it}(\cdot)$ attains the right side of the Bellman equation (2.8).

The principle of optimality characterizes the equivalent relationship between the two solutions to the intertemporal choice problem (2.7) and the functional equation (2.8). In other words, given $(Y_{n,t-1}, \eta_{nt})$,

$$\begin{aligned} V_i(Y_{n,t-1}, \eta_{nt}) &= u_i\left(Y^*_{nt}\left(Y_{n,t-1}, \eta_{nt}\right), Y_{n,t-1}, \eta_{it}\right) + \delta E_t\left(V_i(Y^*_{nt}\left(Y_{n,t-1}, \eta_{nt}\right), \eta_{n,t+1})\right) \\ &= u_i\left(Y^*_{nt}\left(Y_{n,t-1}, \eta_{nt}\right), Y_{n,t-1}, \eta_{it}\right) \\ &\quad + \sum_{s=1}^{\infty} \delta^s E_t\left(u_i\left(Y^*_{n,t+s}\left(Y_{n,t+s-1}, \eta_{n,t+s}\right), Y^*_{n,t+s-1}\left(Y_{n,t+s-2}, \eta_{n,t+s-1}\right), \eta_{i,t+s}\right)\right) \end{aligned}$$

where $Y^*_t\left(Y_{n,t-1}, \eta_{nt}\right) = \left(f_1\left(Y_{n,t-1}, \eta_{nt}\right), \cdots, f_n\left(Y_{n,t-1}, \eta_{nt}\right)\right)'$.

Since payoff (2.4) is linear-quadratic and there is a time-discounting factor δ, the agent i's intertemporal choice problem in an infinite horizon setting belongs to a discounted linear regulator problem. The agent i's value function $V_i(\cdot)$ takes the form

$$V_i\left(Y_{n,t-1}, \eta_{nt}\right) = Y'_{n,t-1} Q_i Y_{n,t-1} + Y'_{n,t-1} L_i \eta_{nt} + \eta'_{nt} G_i \eta_{nt} + c_i \tag{2.9}$$

for some $n \times n$ matrices Q_i, L_i, G_i, and a scalar c_i for each $i = 1, \cdots, n$. Note that Q_i, L_i, G_i and c_i are the unique solutions of the algebraic matrix Riccati equations stemming from a recursive relationship.[15] To have a well-defined Bellman equation

[15]Formation of the algebraic matrix Riccati equations can be found in Appendix A. When we are only interested in agents' optimal policies rather than values, computational advantages are enjoyable since obtaining Q_i and L_i is sufficient for that. This fact is consistent with that Howard's

(a recursive relationship), $V_i(\cdot)$ should be a continuous and bounded function. When we consider a conventional intertemporal choice problem in economics, a choice set is usually limited by a budget or a resource constraint. Due to the existence of a constraint, agent's value will not be explosive, so it becomes continuous and bounded. In our problem, however, while there is no explicit constraint on agents' choices, there are costs which limit choices. The Bellman equation (2.8) can be characterized by using the maximum operator $\mathcal{T}$:

$$\begin{aligned} V_i(Y_{n,t-1},\eta_{nt}) &= \mathcal{T}(V_i)(Y_{n,t-1},\eta_{nt}) \\ &= \max_{y_{it}} \left\{ \begin{array}{c} u_i\left(y_{it}, Y^*_{-i,t}(Y_{n,t-1},\eta_{nt}), Y_{n,t-1}, \eta_{it}\right) \\ +\delta E_t\left(V_i(y_{it}, Y^*_{-i,t}(Y_{n,t-1},\eta_{nt}), \eta_{t+1})\right) \end{array} \right\}, \end{aligned}$$

where the functional solution $V_i(\cdot)$ will be a fixed point of the operator $\mathcal{T}$ in an infinite horizon setting. The existence and uniqueness of the value functions $V_i(\cdot)$'s for all agents can be guaranteed by imposing regularity conditions on $u_i(\cdot)$, W_n, and strength of interactions so that $\mathcal{T}$ is a contraction mapping.[16] For this, define

$$Q^*_n = [(Q_1 + Q'_1)\, e_1, \cdots, (Q_n + Q'_n)\, e_n]' and L^*_n = [L'_1 e_1, \cdots, L'_n e_n]'. \tag{2.10}$$

Assumption 2.2.1 *We assume*

(i) (Process of η^v_{nt}) For each t, $\eta^v_{n,t+1} = \Pi_n \eta^v_{nt} + \xi_{n,t+1}$ where $\|\Pi_n\| < 1$, $\|\cdot\|$ denotes a proper matrix norm, $\eta^v_{nt} = (\eta^v_{1t}, \cdots, \eta^v_{nt})'$, $E_t(\xi_{n,t+1}) = 0$ and $E_t\left(\xi_{n,t+1}\xi'_{n,t+1}\right) = \Omega_\xi$ which is positive definite.

(ii) For each $i = 1, \cdots, n$, all entries of Q_i, L_i, G_i and c_i are bounded.

improvement algorithm (policy function iteration) often converges faster than value function iteration. For more details in the Riccati equation and relevant issues, refer to Chapters 3 and 5 in Ljungqvist and Sargent (2012).

[16] The detailed arguments can be found in Appendix A.

Under Assumption 2.2.1 (i), we have a linear expectation $E_t\left(\eta^v_{n,t+1}\right) = E\left(\eta^v_{n,t+1}|\eta^v_{nt}\right) = \Pi_n \eta^v_{nt}$ and other parts of histories (e.g., $\eta^v_{n,t-1}$, $\eta^v_{n,t-2}$, $\cdots$) are not relevant.[17] Since we assume $\|\Pi_n\| < 1$ and $E_t\left(\xi_{n,t+1}\xi'_{n,t+1}\right) = \Omega_\xi > 0$, it implies $\max_{i=1,\cdots,n} \sup_t E_t\left(|\eta_{i,t+1}|^2\right) < \infty$. If some elements of $\eta_{n,t+1}$ are invariant over time, it would be reasonable to assume them to be known for all agents, then corresponding coefficients in Π_n would be one and $\xi_{n,t+1}$ would be zero. By controlling Q_i, L_i, G_i and c_i, the restrictions of Assumption 2.2.1 (iii) help to avoid agents' extreme decisions so that lifetime values would not be explosive. The restriction on Q_i makes manageable dependence between $Y_{n,t-1}$ and Y_{nt}. The restriction on L_i comes from forward-looking features of our model, but would not appear in a myopic model. By imposing this restriction, expected remote future exogenous effects on the current decisions become negligible.[18]

As $\mathcal{T}$ is a contraction mapping, with an initial guess function $V^{(0)}(\cdot)$, it can iteratively generate a sequence of functions $V^{(j)}(\cdot)$ such that $V^{(j)}(\cdot) = \mathcal{T}\left(V^{(j-1)}\right)(\cdot)$, and the value function V will be the limiting value, i.e., $V_i(\cdot) = \lim_{j\to\infty} \mathcal{T}\left(V_i^{(j-1)}\right)(\cdot)$ for each agent i.[19] The Bellman equation thus characterizes the value function. With an available limiting value $V_i(\cdot)$, the agent i's optimum activity y_{it} can be solved from the maximization problem with

$$y^*_{it}(Y_{n,t-1}, \eta_{nt}) = \arg\max_{y_{it}} \left\{ \begin{array}{c} u_i\left(y_{it}, Y^*_{-i,t}(Y_{n,t-1}, \eta_{nt}), Y_{n,t-1}, \eta_{it}\right) \\ +\delta E_t\left(V_i(y_{it}, Y^*_{-i,t}(Y_{n,t-1}, \eta_{nt}), \eta_{t+1})\right) \end{array} \right\}.$$

[17]The linear conditional expectation would likely be used for practical estimation. Theoretically, it can be generalized to nonlinear functions if needed and desirable. It is convenient in notation here.

[18]Note that G_i and c_i are not relevant to agents' equilibrium decisions. However, controlling them is needed to have bounded V_i's.

[19]This process is called "the method of successive approximations" (Stoket et al. (1989)).

For our model, because the payoff function $u_i(\cdot)$ is a linear-quadratic form in Y_{nt} and $(Y_{n,t-1}, \eta_{nt})$, we would expect that the value function $V_i(\cdot)$ would be a linear-quadratic form. The Bellman equation with a fixed point for $V_i(\cdot)$ would provide the characterization of coefficients of the linear-quadratic form, which in turn, may provide us a system of estimation equations for $y_{it}^*(\cdot)$ for $i = 1, \cdots, n$ at each t. For the system of estimation equations, we shall consider its estimation with methods such as the quasi-maximum likelihood (QML) and a possibly simpler nonlinear two-stage least squares (NL2S).

Whether the value function is indeed in a linear-quadratic form can be revealed by fixed point iterations of the contraction mapping $\mathcal{T}$ and be confirmed by mathematical induction. Indeed, iterations of $\mathcal{T}$ would provide value functions, and then optimized activities of agents can also be derived in a finite horizon setting. For either a finite horizon or infinite horizon setting, one should start with the initial $V_i^{(0)} = 0$ (i.e., a zero initial function) and then have the iterations,

$$V_i^{(j)}(Y_{n,t-1}, \eta_{nt}) = \max_{y_{it}} \left\{ \begin{array}{c} u_i\left(y_{it}, Y_{-i,t}^{*(j)}(Y_{n,t-1}, \eta_{nt}), Y_{n,t-1}, \eta_{it}\right) \\ +\delta E_t\left(V_i^{(j-1)}\left(y_{it}, Y_{-i,t}^{*(j)}(Y_{n,t-1}, \eta_{nt}), \eta_{t+1}\right)\right) \end{array} \right\},$$

for $j = 1, 2, \cdots$. We see that with $V_i^{(0)} = 0$, $V_i^{(1)}(\cdot)$ is the value function of agent i at t being the terminal period; $V_i^{(2)}(\cdot)$ would be the value function at t while $t+1$ were the terminal period, and in general, $V_i^{(J+1)}(\cdot)$ would be the value function at t while $t+J$ were the terminal period. So for a model with a finite horizon of future J periods at time t, the corresponding optimum activity could be derived as

$$y_{it}^{*(J+1)}(Y_{n,t-1}, \eta_{nt}) = \arg\max_{y_{it}} \left\{ \begin{array}{c} u_i\left(y_{it}, Y_{-i,t}^{*(J+1)}(Y_{n,t-1}, \eta_{nt}), Y_{n,t-1}, \eta_{it}\right) \\ +\delta E_t\left(V_i^{(J)}\left(y_{it}, Y_{-i,t}^{*(J+1)}(Y_{n,t-1}, \eta_{nt}), \eta_{n,t+1}\right)\right) \end{array} \right\}$$

and the value function for agent i would be $V_i^{(J+1)}(\cdot)$.

For the situation with infinite horizon, the iterations continue to infinity and the stable system of NE is

$$Y_{nt}^{*}\left(Y_{n,t-1},\eta_{nt}\right)=\left(\lambda_{0}W_{n}+\delta Q_{n}^{*}\right)Y_{nt}^{*}\left(Y_{n,t-1},\eta_{nt}\right)+\left(\gamma_{0}I_{n}+\rho_{0}W_{n}\right)Y_{n,t-1}+\left(I_{n}+\delta L_{n}^{*}\Pi_{n}\right)\eta_{nt}, \tag{2.11}$$

which captures the contemporaneous spatial spillover effect through $\lambda_0 W_n Y_{nt}^{*}\left(Y_{n,t-1},\eta_{nt}\right)$, dynamic effect $\gamma_0 Y_{n,t-1}$, spatial- past time effect or diffusion $\rho_0 W_n Y_{n,t-1}$, and additional expected spatial- future time effect $\delta Q_n^{*} Y_{nt}^{*}\left(Y_{n,t-1},\eta_{nt}\right)$. The additional term $\delta L_n^{*}\Pi_n\eta_{nt}$ is due to expected future unknown explanatory factors and disturbances, as η_{nt} may contain time-varying and invariant explanatory variables and disturbances. The spatial-time filter of our model is defined by

$$R_n = S_n - \delta Q_n^{*}, where S_n = I_n - \lambda_0 W_n. \tag{2.12}$$

So the NE activity vector at time t is

$$Y_{nt}^{*}\left(Y_{n,t-1},\eta_{nt}\right) = A_n Y_{n,t-1} + B_n \eta_{nt} \tag{2.13}$$

where $A_n = R_n^{-1}\left(\gamma_0 I_n + \rho_0 W_n\right)$ and $B_n = R_n^{-1}\left(I_n + \delta L_n^{*}\Pi_n\right)$. Note that the transformation R_n characterizes the interrelation among agents' decisions. Due to the forward-looking feature of our model, direct influences (i.e., first-order spatial effects) can come from all spatial units even for a sparse W_n.[20] In the view of SAR models, R_n would reduce to the conventional $S_n = I_n - \lambda_0 W_n$ when $\delta = 0$, i.e., with completely discount of future values, or equivalently with myopic behavior. The transformation

[20]For illustrative purposes, suppose there is no isolated spatial unit. Then, all elements in Q_n^{*} are nonzero. In our system equation (2.11), note that the direct influences can be composed by two parts: (i) $\lambda_0 W_n Y_{nt}^{*}$ and (ii) $\delta Q_n^{*} Y_{nt}^{*}$. If $w_{ij} = 0$, there is no direct contemporaneous spill over effect (i.e., $\lambda_0 w_{ij} y_{jt} = 0$ if $w_{ij} = 0$). Even for $w_{ij} = 0$, $\delta\left[Q_n^{*}\right]_{ij} y_{jt} \neq 0$ since agent i has in mind j's expected future indirect influences (i.e., future NE) in his/her current decision-making.

L_n^* can be represented by

$$L_n^* = \sum_{m=1}^{\infty} \delta^{m-1} D_{n,m} \Pi_n^{m-1} \tag{2.14}$$

where $D_{n,m}$ $(m = 1, 2, \cdots)$ denote some $n \times n$ matrices, which only rely on λ_0, γ_0, ρ_0, and δ with W_n.[21] In estimating parameters, both the structural and nuisance parameters (related to Π_n) are included in the linear term L_n^*, but the parts of structural parameters and nuisance one can be distinguished. Using $D_{n,1}$, moreover, we find the relationship between Q_n^* and L_n^*:

$$Q_n^* = D_{n,1} \left(\gamma_0 I_n + \rho_0 W_n\right) - \gamma_0 I_n, \tag{2.15}$$

which implies

$$\begin{aligned} Y_{nt}^* \left(Y_{n,t-1}, \eta_{nt}\right) &= \left(\lambda_0 W_n + \delta D_{n,1} \left(\gamma_0 I_n + \rho_0 W_n\right) - \delta\gamma_0 I_n\right) Y_{nt}^* \left(Y_{n,t-1}, \eta_{nt}\right) \\ &\quad + \left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n,t-1} + \left(I_n + \sum_{m=1}^{\infty} \delta^m D_{n,m} \Pi_n^m\right) \eta_{nt} \end{aligned} \tag{2.16}$$

and

$$R_n = \left(1 + \delta\gamma_0\right) I_n - \lambda_0 W_n - \delta D_{n,1} \left(\gamma_0 I_n + \rho_0 W_n\right) \tag{2.17}$$

Equation (2.16) describes a role of future relevant components combined with δ. The additional components $\delta\gamma_0 I_n$ and $-\delta D_{n,1}\left(\gamma_0 I_n + \rho_0 W_n\right)$ in R_n are due to agents' forward-looking decision-making and they are respectively counterparts of the time influence $\delta\gamma_0 I_n$ and the additional spatial influence $\delta\Delta_n A_n^{trad}$ in the two-period model. Note that $e_i' Q_n^* = e_i'\left(Q_i + Q_i'\right)$ and $e_i' L_n^* = e_i' L_i$ for all $i = 1, \cdots, n$. To explain equation (2.16), consider the first-order condition of agent i's arbitrary t period problem:

$$y_{it}^*(Y_{n,t-1}, \eta_{nt}) \quad = \quad \eta_{it} + \gamma_0 y_{i,t-1} + \rho_0 w_{i.} Y_{n,t-1} + \lambda_0 w_{i.} Y_{nt}^*(Y_{n,t-1}, \eta_{nt})$$

[21] Detailed forms and their derivations can be found in our Appendix A.

$$
\begin{aligned}
&+\delta\left(e_i Q_n^* Y_{nt}^*(Y_{n,t-1},\eta_t) + \sum_{m=1}^{\infty}\delta^{m-1}e_i' D_{n,m}\Pi_n^m \eta_{nt}\right) \\
= \quad & \eta_{it} + \gamma_0 y_{i,t-1} + \rho_0 w_{i.} Y_{n,t-1} + \lambda_0 w_{i.} Y_{nt}^*(Y_{n,t-1},\eta_{nt}) - \delta\gamma_0 y_{it}^*(Y_{n,t-1},\eta_{nt}) \\
&+\delta e_i' D_{n,1}\left(\left(\gamma_0 I_n + \rho_0 W_n\right) Y_{nt}^*(Y_{n,t-1},\eta_{nt}) + \Pi_n \eta_{nt}\right) + \sum_{m=2}^{\infty}\delta^m e_i' D_{n,m}\Pi_n^m \eta_{nt}.
\end{aligned}
$$

Hence, we can observe $\delta D_{n,1}\left(\left(\gamma_0 I_n + \rho_0 W_n\right) Y_{nt}^*(Y_{n,t-1},\eta_{nt}) + \Pi_n \eta_{nt}\right)$ plays a similar role to the additional terms in the two-period model except the additional exogenous influences $\sum_{m=2}^{\infty}\delta^m e_i' D_{n,m}\Pi_n^m \eta_{nt}$. The reason why only $D_{n,1}$ appears in R_n and $Y_{nt}^*\left(Y_{n,t-1},\eta_{nt}\right)$ just relies on the payoff relevant history $Y_{n,t-1}$ are due to the Markov property of agents' decision-making.

2.3 The econometric model

In this section, we construct an econometric model and suggest estimation methods for this model with a panel data set. Assume a researcher has observed $\left(\{Y_{nt}, X_{nt}\}_{t=0}^{T}\right)$ and W_n from a panel data set, where Y_{nt} is an $n \times 1$ vector of dependent variables and $X_{nt} = (X_{nt,1}\cdots, X_{nt,K})$ with $X_{nt,k} = (x_{1t,k},\cdots, x_{nt,k})'$ for $k = 1,\cdots,K$ is an $n \times K$ matrix of (exogenous) explanatory variables.[22] Each Y_{nt} is supposed to be realized as an equilibrium, (i.e., $Y_{nt} = Y_{nt}^*\left(Y_{n,t-1},\eta_{nt}\right)$). For estimation, we assume some structures on η_{nt}. First, η_{nt} contains time-varying explanatory variables (X_{nt}) with coefficients $\beta_0 = (\beta_{1,0},\cdots,\beta_{K,0})'$ and disturbances. In addition, fixed individual and time effects can be introduced as components of η_{nt}. It is of interest to note for the infinite horizon case, the modified dynamic SAR equation can allow the specification of additive individual effect $c_{i,0}^*$ and time effect $\alpha_{t,0}$. With all individual effects in a vector $\mathbf{c}_{n0}^* = \left(c_{1,0}^*,\cdots,c_{n,0}^*\right)'$ which is invariant over time, the corresponding Π_n

[22] After the subsection, we add the subscript n (or T) to point out that it is constructed by n (or T) sample points.

would be an identity matrix, thus individual effects would be reparameterized into $\mathbf{c}_{n0} = (I_n + \delta L_n^*)\,\mathbf{c}_{n0}^*$. For a time effect $\alpha_{t,0} l_n$, if $\alpha_{t,0}$'s are random shocks which might influence every agent, then its corresponding Π_n is zero, so the time effect $\alpha_{t,0} l_n$ can be additive.

Hence, we have the model specification

$$Y_{nt} = (\lambda_0 W_n + \delta Q_n^*)\, Y_{nt} + (\gamma_0 I_n + \rho_0 W_n)\, Y_{n,t-1} + (I_n + \delta L_n^* \Pi_n)\, X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t,0} l_n + \mathcal{E}_{nt} \tag{2.18}$$

for $t = 1, \cdots, T$, where $\mathcal{E}_{nt} = (\epsilon_{1t}, \cdots, \epsilon_{nt})'$ is an n-dimensional vector of i.i.d. disturbances with mean zero and variance $\sigma_{\epsilon,0}^2 > 0$. The main parameters are λ_0, γ_0, ρ_0, β_0 and $\sigma_{\epsilon,0}^2$. The time-discounting factor δ is considered as a primitive parameter and the incidental parameters in Π_n are assumed to be covered by the process of X_{nt}'s already. We shall explore the estimation approach in the situation of both n and T being large. In this situation, it is appropriate to consider the estimation of the structural parameter vector $\theta_0 = \left(\lambda_0, \gamma_0, \rho_0, \beta_0', \sigma_{\epsilon,0}^2\right)'$ together with the fixed individual and time effects $\mathbf{c}_{n0}$ and α_{T0}, where $\alpha_{T0} = (\alpha_{1,0}, \cdots, \alpha_{T,0})'$ is the vector of time effects.

As special cases of model specification (2.18), we consider two cases because they have distinct features. First, consider $\lambda_0 = \rho_0 = 0$, which means no spatial interactions but not myopic due to individual own time lag effect. In this case, $R_n = z I_n$ such that $z = 1 + \delta\gamma_0 + \frac{-\delta\gamma_0^2}{1+\delta\gamma_0+\frac{-\delta\gamma_0^2}{1+\delta\gamma_0+\frac{-\delta\gamma_0^2}{\cdots}}}$. Using the formula of infinite continued fractions[23], we have

$$R_n = \frac{1}{2}\left(1 + \delta\gamma_0 + \sqrt{1 + 2\delta\gamma_0 - \delta\gamma_0^2\,(4 - \delta)}\right) I_n. \tag{2.19}$$

[23]This is, $\sqrt{x^2 + y} = x + \frac{y}{2x+\frac{y}{2x+\frac{y}{\cdots}}}$.

To obtain validity of (2.19), $1+2\delta\gamma_0 - \delta\gamma_0^2(4-\delta) > 0$ is required. The second case is $\lambda_0 = 0$, which means no direct contemporaneous spatial interaction. In conventional SDPD models, there is no contemporaneous spatial interaction if $\lambda_0 = 0$. In our case, however, the forward-looking spatial filter R_n becomes $I_n - \delta Q_n^*$ where the i^{th}-row of Q_n^* is $e_i' A_n' \left[-e_i e_i' + \delta\left(Q_i + Q_i'\right)\right] A_n + \gamma_0 e_i' \left[A_n' e_i e_i' + (A_n - I_n)\right]$. It implies that (i) $Q_n^* \neq \mathbf{0}_{n\times n}$ even for $\lambda_0 = 0$ since agents' consider the expected future diffusion effects, and (ii) Q_n^* would be simpler than that of $\lambda_0 = 0$ case.

The reduced form of equation (2.18) is

$$Y_{nt} = A_n Y_{n,t-1} + R_n^{-1}\left[\left(I_n + \delta L_n^* \Pi_n\right) X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t,0} l_n + \mathcal{E}_{nt}\right] \tag{2.20}$$

where $A_n = R_n^{-1}\left(\gamma_0 I_n + \rho_0 W_n\right)$ with $R_n = I_n - \left(\lambda_0 W_n + \delta Q_n^*\right)$. Stability of system (2.18) means the spatial-time dependence should be manageable. Note that $Q_n^* = D_{n,1}\left(\gamma_0 I_n + \rho_0 W_n\right) - \gamma_0 I_n$, $L_n^* = \sum_{m=1}^{\infty} \delta^{m-1} D_{n,m} \Pi_n^{m-1}$ and $D_{n,m}$ $(m = 2, 3, \cdots)$ are generated by $D_{n,1}$. Then, assuming uniform boundedness of $D_{n,1}$ yields well-definedness and uniformly boundedness of L_n^*. Hence, the current and expected future exogenous effects $I_n + \delta L_n^* \Pi_n$ become manageable.[24] When absolute summability for $\sum_{j=1}^{\infty} A_n^j$ and its uniform boundedness in row and column sums hold, we have the infinite summation representation

$$Y_{nt} = \sum_{j=0}^{\infty} A_n^j R_n^{-1}\left[\left(I_n + \delta L_n^* \Pi_n\right) X_{n,t-j}\beta_0 + \mathbf{c}_{n0} + \alpha_{t-j,0} l_n + \mathcal{E}_{n,t-j}\right]. \tag{2.21}$$

As n increases, $\|A_n\| < 1$ and uniform boundedness of R_n^{-1} guarantees the variance of each y_{it} is not explosive and remains to be bounded.

[24]If $\mathbf{c}_{n0}^*$ is a vector of uniformly bounded constants, $\mathbf{c}_{n0} = \left(I_n + \delta L_n^*\right)\mathbf{c}_{n0}^*$ is also uniformly bounded if $\|D_{n,1}\| < c_D$.

2.4 Estimation

2.4.1 Quasi-maximum likelihood estimation

To estimate equation (2.18), we firstly suggest the quasi-maximum likelihood estimation (QML) method, which gives a fundamental background in parameter estimation. Asymptotic results for the QML estimator are based on the increasing-domain asymptotic.[25] Let $\theta = (\lambda, \gamma, \rho, \beta', \sigma_\epsilon^2)'$ be the set of structural parameters for estimation, where θ_0 is the true value of θ. The dimension of the parameters is $4+K$. To distinguish the individual- or time-specific effects for estimation, we denote $\mathbf{c}_n = (c_1, \cdots, c_n)'$ and $\alpha_T = (\alpha_1, \cdots, \alpha_T)'$. Let $\theta_{1,0}$ be the true $\theta_1 = (\lambda, \gamma, \rho)'$, which consists of parameters involved in L_n^* and Q_n^*. For each θ_1, we define $Q_n^*(\theta_1)$ and $L_n^*(\theta_1)$ with $R_n(\theta_1) = I_n - \lambda W_n - \delta Q_n^*(\theta_1)$ and $A_n(\theta_1) = R_n^{-1}(\theta_1)(\gamma I_n + \rho W_n)$. The log-likelihood function with a panel with nT observations will be

$$\begin{aligned} \ln L_{nT}(\theta, \mathbf{c}_n, \alpha_T) &= -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma_\epsilon^2 + T\ln|R_n(\theta_1)| \\ &\quad -\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}\mathcal{E}'_{nt}(\theta, \mathbf{c}_n, \alpha_T)\,\mathcal{E}_{nt}(\theta, \mathbf{c}_n, \alpha_T) \end{aligned} \tag{2.22}$$

where $\mathcal{E}_{nt}(\theta, \mathbf{c}_n, \alpha_T) = R_n(\theta_1) Y_{nt} - (\gamma I_n + \rho W_n) Y_{n,t-1} - (I_n + \delta L_n^*(\theta_1)\Pi_n) X_{nt}\beta - \mathbf{c}_n - \alpha_t l_n$.

The computation of this model will be more complicated than that of the conventional SDPD model. Note that the conventional SDPD model is linear in parameters except $\sigma_{\epsilon,0}^2$. But for the equation from the intertemporal dynamic spatial model, the implied matrices Q_n^* and L_n^* are both functions of the parameters λ_0, γ_0, ρ_0 and the time-discounting factor δ. Hence, we need to numerically evaluate $Q_n^*(\theta_1)$ and $L_n^*(\theta_1)$

[25] It means that sample observations are from a growing observation region (spatial domain). In case of the fixed-domain asymptotic, a spatial domain (a region) is fixed and bounded and the number of observations in that spatial domain increases.

for each θ_1 (i.e., inner loop). As the total number of individual and time fixed effects in $\mathbf{c}_{n0}$ and α_{T0} is $n+T$, it is desirable to focus on the use of the concentrated log-likelihood function with the fixed effects $\mathbf{c}_{n0}$ and α_{T0} concentrated out. In consequence, the optimization of the concentrated log-likelihood function is on a fixed number of structural parameters. As the fixed effects are linear in the generalized SAR equation, they can be estimated as regression coefficients when other structural parameters in the equation are given.

Let $\bar{Y}_{nT} = \frac{1}{T}\sum_{s=1}^{T} Y_{ns}$, $\bar{Y}_{nT,-1} = \frac{1}{T}\sum_{s=0}^{T-1} Y_{ns}$ and $\bar{X}_{nT} = \frac{1}{T}\sum_{s=1}^{T} X_{ns}$. With fixed individual and time effects concentrated out, the concentrated log-likelihood with parameter subvector θ is

$$\ln L_{nT,c}(\theta) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma_\epsilon^2 + T\ln|R_n(\theta_1)| - \frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}\tilde{\mathcal{E}}_{nt}'(\theta) J_n \tilde{\mathcal{E}}_{nt}(\theta) \quad (2.23)$$

where $\tilde{\mathcal{E}}_{nt}(\theta) = R_n(\theta_1)\tilde{Y}_{nt} - (\gamma I_n + \rho W_n)\tilde{Y}_{n,t-1}^{(-)} - (I_n + \delta L_n^*(\theta_1)\Pi_n)\tilde{X}_{nt}\beta$ with $\tilde{Y}_{nt} = Y_{nt} - \bar{Y}_{nT}$, $\tilde{Y}_{n,t-1}^{(-)} = Y_{n,t-1} - \bar{Y}_{nT,-1}$, and $\tilde{X}_{nt} = X_{nt} - \bar{X}_{nT}$ in deviation from time mean, and $J_n = I_n - \frac{1}{n}l_n l_n'$ being the deviation from group mean operator.[26] From (2.23), we obtain the maximum likelihood estimators, $\hat{\theta}_{ml,nT} = \arg\max_{\theta\in\Theta}\ln L_{nT,c}(\theta)$, where Θ denotes the parameter space of θ. For computation, in particular, with a large size sample, we shall put more attention on the evaluation of the determinant $|R_n(\theta_1)|$ and its inverse $R_n^{-1}(\theta_1)$. In the spatial literature, the suggestion by Lesage and Pace (2009) on a Taylor series analytic expansion of the determinant $|I_n - \lambda W_n|$ in λ may be useful. For the inverse of $R_n(\theta_1)$, one might also consider the Neumann series expansion. That Neumann series expansion can be justified by the stability of our spatial dynamic process.

[26]Note that we cannot eliminate the time fixed effects by introducing a traditional orthonormal transformation like Lee and Yu (2010) and derive a partial likelihood for estimation because the spatial filter matrix R_n does not have a row-normalization property.

Define $R_{n\lambda}(\theta_1) = \frac{\partial R_n(\theta_1)}{\partial \lambda}$, $R_{n\gamma}(\theta_1) = \frac{\partial R_n(\theta_1)}{\partial \gamma}$, $R_{n\rho}(\theta_1) = \frac{\partial R_n(\theta_1)}{\partial \rho}$, $L_{n\lambda}^*(\theta_1) = \frac{\partial L_n^*(\theta_1)}{\partial \lambda}$, $L_{n\gamma}^*(\theta_1) = \frac{\partial L_n^*(\theta_1)}{\partial \gamma}$, and $L_{n\rho}^*(\theta_1) = \frac{\partial L_n^*(\theta_1)}{\partial \rho}$. Note that $R_{n\lambda}$, $R_{n\gamma}$, $R_{n\rho}$, $L_{n\lambda}^*$, $L_{n\gamma}^*$, and $L_{n\rho}^*$ denote those quantities at $\theta = \theta_0$. Here are assumptions for asymptotic properties of $\hat{\theta}_{ml,nT}$. Subsequent asymptotic analysis of the QMLE extends properly that in Yu et al. (2008).

Assumption 2.4.1 *(i) The diagonal elements of W_n are zero.*

(ii) W_n is strictly exogenous and uniformly bounded in row and column sums in absolute value.

Assumption 2.4.2 *For all i and t, ϵ_{it} i.i.d. $\left(0, \sigma_{\epsilon,0}^2\right)$, and $E\left|\epsilon_{it}\right|^{4+\eta} < \infty$ for some $\eta > 0$.*

Assumption 2.4.3 *The parameter space Θ of θ is compact. The true parameter θ_0 is in int(Θ).*

Assumption 2.4.4 *$\{X_{nt}\}_{t=1}^T$, $\{\alpha_{t0}\}_{t=1}^T$ and $\mathbf{c}_{n0}$ are conditional upon nonstochastic values with*

$\sup_{n,T} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left|x_{it,k}\right|^{2+\eta} < \infty$ *for all k,* $\sup_T \frac{1}{T} \sum_{t=1}^T \left|\alpha_{t0}\right|^{2+\eta} < \infty$ *and* $\sup_n \frac{1}{n} \sum_{i=1}^n \left|c_{i,0}\right|^{2+\eta} < \infty$ *for some $\eta > 0$.*

Assumption 2.4.5 *Let Θ_1 be the compact parameter space for θ_1.*

(i) $R_n(\theta_1)$ is invertible for $\theta_1 \in \Theta_1$. $Q_n^*(\theta_1)$ and $L_n^*(\theta_1)$ uniformly bounded in both row and column norms, uniformly in $\theta_1 \in \Theta_1$.

(ii) At any $\theta \in$ int(Θ), the first, second and third derivatives of $R_n(\theta_1)$ and $L_n^*(\theta_1)$ with respect to θ_1 exist and are uniformly bounded in both row and column sum norms, uniformly in $\theta_1 \in \Theta_1$.

(iii) $\sum_{h=1}^{\infty} abs\left(A_n^h\right)$ *is uniformly bounded in both row and column sum norms, where* $[abs(A_n)]_{ij} = \left|[A_n]_{ij}\right|$.

(iv) $\|\delta D_{n,1}\Pi_n\| < 1$ *where* $\|\cdot\|$ *is a proper matrix norm.*

Assumption 2.4.6 *We assume that* T *goes to infinity and* n *is an increasing function of* T.

Assumption 2.4.1 is a standard assumption in spatial econometrics. By assuming uniform boundedness of W_n, spatial dependence becomes not too large and manageable (spatial stability condition). Assumption 2.4.2 (i) assumes *i.i.d.* disturbances across i and t for simplicity. Assuming a compact parameter space (Assumption 2.4.3) is for theoretical analyses (for details, refer to Chapter 4 in Amemiya (1985)). Assumption 2.4.4 means the conditioning argument and is for simplicity of asymptotic analyses for the QMLE. In our economic environment, X_{nt} and α_{t0} are stochastic, so agents can make predictions about their future values. For estimation of the implied structural equation (2.18), X_{nt}, $\mathbf{c}_{n0}$ and α_{t0} are conditional upon as constants and we introduce the higher than the second empirical moment restrictions for X_{nt}, α_{t0} and $\mathbf{c}_{n0}$.[27] Assumption 2.4.5 is for well-definedness of our model. Invertibility of $R_n(\theta_1)$ for $\theta_1 \in \Theta_1$ guarantees for existence and uniqueness of the equilibrium system (2.18) for any $\theta_1 \in \Theta_1$ (Assumption 2.4.5 (i)). Uniform boundedness assumption for $R_n(\theta_1)$ for $\theta_1 \in \Theta_1$ means spatial dependence of dependent varaibles from our model is manageable (stable spatial process). Assumption 2.4.5 (ii) is a trivial requirement. Existence and uniformly boundedness of the first and second derivatives of $R_n(\theta_1)$ and $L_n^*(\theta_1)$ should be required so that $\frac{\partial \ln L_{nT,c}(\theta)}{\partial \theta}$ and $\frac{\partial^2 \ln L_{nT,c}(\theta)}{\partial \theta \partial \theta'}$ for $\theta \in \Theta$

[27]By Kelejian and Prucha (2001), these higher than the second moment restrictions (with the higher than the fourth-moment restriction for ϵ_{it}) are required to apply a central limit theorem for a linear quadratic form.

are well-defined. The reason for having the third derivatives of $R_n(\theta_1)$ and $L_n^*(\theta_1)$ is for the uniform convergence of the second order derivatives of the log-likelihood function. Assumption 2.4.5 (iii) plays a crucial role to study the asymptotic properties of $\hat{\theta}_{ml,nT}$ by restricting dependence between time series and between cross sectional units so that the process is stable in both the space and time dimensions. Under Assumption 2.4.5 (iii) and large T, the initial value Y_{n0} does not affect asymptotic properties of $\hat{\theta}_{ml,nT}$. A sufficient condition for absolute summability is $\|A_n\|_\infty < 1$, so the infinite sum $\sum_{h=0}^{\infty} A_n^h$ exists and is $(I_n - A_n)^{-1}$. If we have Assumption 2.4.5 (iv), $\sum_{h=1}^{\infty} \delta^{h-1} D_{n,1}^h \Pi_n^{h-1} = D_{n,1}(1 - \delta D_{n,1}\Pi_n)^{-1}$.[28] It means expected future exogenous effects become manageable, so the remote (expected) future exogenous effects on Y_{nt} are small to be asymptotically ignorable. Assumption 2.4.6 is needed to consistently estimate the individual and time dummies. Large T is for consistent estimation of $\mathbf{c}_{n0}$ and large n is required for consistent estimation of α_{t0}.

For asymptotic analysis of $\hat{\theta}_{ml,nT}$, note that $\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \theta}$ takes the following linear-quadratic form[29]:

$$\frac{1}{\sqrt{nT}} \sum_{t=1}^{T} \left[B_{y,n} \tilde{Y}_{n,t-1}^{(-)} + D_{nt} \right]' J_n \mathcal{E}_{nt} + \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} \left[\mathcal{E}_{nt}' B_{q,n}' J_n \mathcal{E}_{nt} - \sigma_{\epsilon,0}^2 tr\left(B_{q,n}\right) \right] \quad (2.24)$$

where $B_{y,n}$ and $B_{q,n}$ are some $n \times n$ uniformly bounded (in n) matrices and D_{nt} denotes some time-varying nonstochastic component. By (2.24), $\hat{\theta}_{ml,nT}$ can be asymptotically biased because $\bar{Y}_{nT,-1}$ and $\bar{\mathcal{E}}_{nT}$ are correlated even for large n and T due to many incidential paramaters of individual and time effects. To derive the asymptotic distribution of $\hat{\theta}_{ml,nT}$ and adjust its asymptotic bias, we can decompose $\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \theta}$ into an uncorrelated part and a correlated part. For this, consider the decomposition

[28]Since $D_{n,h}$'s $(h = 2, 3, \cdots)$ are generated by $D_{n,1}$, $L_n^* = \sum_{h=1}^{\infty} \delta^{h-1} D_{n,h} \Pi_n^{h-1}$ is uniformly bounded in n.

[29]The formulas of $\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \theta}$ can be found in Appendix A.

$J_n\tilde{Y}_{n,t-1}^{(-)} = J_n\tilde{Y}_{n,t-1}^{(-)(u)} - J_n\bar{U}_{nT,-1}$ where

$$\begin{aligned} J_n\tilde{Y}_{n,t-1}^{(-)(u)} &= J_n\left[\sum_{h=0}^{\infty} A_n^h R_n^{-1}\left[\left(I_n + \delta L_n^*\Pi_n\right)\tilde{X}_{n,t-j-1}\beta_0 + \tilde{\alpha}_{t-h-1,0}l_n\right]\right] \\ &\quad + J_n\left[\sum_{h=0}^{\infty} A_n^h R_n^{-1}\mathcal{E}_{n,t-h-1}\right] \end{aligned}$$

and $\bar{U}_{nT,-1} = \frac{1}{T}\sum_{t=0}^{T-1}\sum_{h=0}^{\infty} A_n^h R_n^{-1}\mathcal{E}_{n,t-h}$.

Using the decomposition, we have $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{nT}}\frac{\partial \ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta} - \Delta_{1,nT} - \Delta_{2,nT}$. Note that

$$\begin{aligned} \frac{1}{\sqrt{nT}}\frac{\partial \ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta} &= \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left[B_{y,n}\tilde{Y}_{n,t-1}^{(-)(u)} + D_{nt}\right]' J_n\mathcal{E}_{nt} \\ &\quad + \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left[\mathcal{E}_{nt}' B_{q,n}' J_n\mathcal{E}_{nt} - \sigma_{\epsilon,0}^2 tr\left(J_n B_{q,n}\right)\right], \end{aligned} \tag{2.25}$$

which determines the asymptotic distribution of $\hat{\theta}_{ml,nT}$. The terms $\Delta_{1,nT}$ and $\Delta_{2,nT}$ characterize asymptotic biases. Note that $\Delta_{1,nT}$ and $\Delta_{2,nT}$ are respectively $\sqrt{\frac{T}{n}}\left[\left(B_{y,n}\bar{U}_{nT,-1}\right)' J_n\bar{\mathcal{E}}_{nT} + \bar{\mathcal{E}}_{nT}' B_{q,n}' J_n\bar{\mathcal{E}}_{nT}\right]$ and $\sqrt{\frac{T}{n}}\left[\sigma_{\epsilon,0}^2\left(tr\left(B_{q,n}\right) - tr\left(J_n B_{q,n}\right)\right)\right]$ where the detailed forms of $\Delta_{1,nT}$ and $\Delta_{2,nT}$ can be found in Appendix B. $\Delta_{1,nT}$ comes from estimating $\mathbf{c}_{n0}$ while $\Delta_{2,nT}$ is generated from estimating $\{\alpha_{t0}\}_{t=1}^{T}$. The main stochastic components of $\Delta_{1,nT}$ are $\bar{U}_{nT,-1}' B_n\bar{\mathcal{E}}_{nT}$, and $\bar{\mathcal{E}}_{nT}' B_n\bar{\mathcal{E}}_{nT}$ where B_n denotes some uniformly bounded (in n) matrix in row and column sum norms. However, $\Delta_{2,nT}$ is determined by non-stochastic components, $tr\left(-R_{n\lambda}R_n^{-1}\right) - tr\left(J_n(-R_{n\lambda}R_n^{-1})\right)$, $tr\left(-R_{n\gamma}R_n^{-1}\right) - tr\left(J_n(-R_{n\gamma}R_n^{-1})\right)$, $tr\left(-R_{n\rho}R_n^{-1}\right) - tr\left(J_n(-R_{n\rho}R_n^{-1})\right)$, and $\frac{1}{2\sigma_{\epsilon,0}^2}$. By Lemmas 2.1 and 2.2 in our supplementary file, $\Delta_{1,nT} = \sqrt{\frac{n}{T}}a_{n,1}(\theta_0) + O\left(\sqrt{\frac{n}{T^3}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$, where $a_{n,1}(\theta_0) = O(1)$, and, $\Delta_{2,nT} = \sqrt{\frac{T}{n}}a_{n,2}(\theta_0)$, where $a_{n,2}(\theta_0)$ are $O(1)$. The formulas of $a_{n,1}(\theta_0)$ and $a_{n,2}(\theta_0)$ can be found in Appendix A.

Consistency and asymptotic normality

First, consider consistency of $\hat{\theta}_{ml,nT}$. For each $\theta \in \Theta$, define

$$\begin{aligned} Q_{nT}(\theta) &= \frac{1}{nT} E \ln L_{nT,c}(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma_\epsilon^2 + \frac{1}{n}\ln |R_n(\theta_1)| \\ &\quad -\frac{1}{2\sigma_\epsilon^2}\frac{1}{nT} E\left(\sum_{t=1}^{T} \overset{\heartsuit}{\mathcal{E}}'_{nt}(\theta) J_n \overset{\heartsuit}{\mathcal{E}}_{nt}(\theta)\right) \end{aligned}$$

To show consistency, the first step is verifying uniform convergence of sample average of the log-likelihood function, $\sup_{\theta\in\Theta}\left|\frac{1}{nT}\ln L_{nT,c}(\theta) - Q_{nT}(\theta)\right| \to_p 0$ as $n, T \to \infty$. After this, we show $Q_{nT}(\theta)$ is well-behaved at any point θ in Θ by verifying uniform equicontinuity of $Q_{nT}(\theta)$ Obtaining the identification uniqueness completes the proof of consistency. The assumption below describes the identification uniqueness conditions.

Assumption 2.4.7 (Identification) *To identify θ_0, we assume*

(i) $\lim_{n,T\to\infty}\left[\frac{1}{n}\ln\left|\sigma_{\epsilon,0}^2 R_n^{-1\prime} R_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_{\epsilon,nT}^2(\theta_1) R_n^{-1\prime}(\theta_1) R_n^{-1}(\theta_1)\right|\right] \neq 0$ *for* $\theta_1 \neq \theta_{1,0}$ *where*

$$\begin{aligned} \sigma_{\epsilon,nT}^2(\theta_1) &= \frac{1}{nT}\sum_{t=1}^{T} E\left(\begin{array}{c} \tilde{Z}_{nt}(\theta_1) - \tilde{\mathbf{X}}_{nt}(\theta_1)\left[\sum_{s=1}^{T}\tilde{\mathbf{X}}'_{ns}(\theta_1) J_n \tilde{\mathbf{X}}_{ns}(\theta_1)\right]^{-1} \\ \times \sum_{s=1}^{T}\tilde{\mathbf{X}}'_{ns}(\theta_1) J_n \tilde{Z}_{ns}(\theta_1) \end{array}\right)' \\ &\times J_n\left(\tilde{Z}_{nt}(\theta_1) - \tilde{\mathbf{X}}_{nt}(\theta_1)\left[\sum_{s=1}^{T}\tilde{\mathbf{X}}'_{ns}(\theta_1) J_n \tilde{\mathbf{X}}_{ns}(\theta_1)\right]^{-1}\sum_{s=1}^{T}\tilde{\mathbf{X}}'_{ns}(\theta_1) J_n \tilde{Z}_{ns}(\theta_1)\right) \\ &+\frac{\sigma_{\epsilon,0}^2}{n-1} tr\left(R_n^{-1\prime} R'_n(\theta_1) J_n R_n(\theta_1) R_n^{-1}\right) + o(1), \end{aligned}$$

$\tilde{Z}_{nt}(\theta_1) = \left[R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)\right]\tilde{Y}_{n,t-1}^{(-)} + R_n(\theta_1) R_n^{-1}\left[\tilde{\mathbf{X}}_{nt}\beta_0 + \tilde{\alpha}_{t,0} l_n\right]$, *and* $\tilde{\mathbf{X}}_{nt}(\theta_1) = (I_n + \delta L_n^*(\theta_1)\Pi_n)\tilde{X}_{nt}$ *with* $\tilde{\mathbf{X}}_{nt} = \tilde{\mathbf{X}}_{nt}(\theta_{1,0})$.

(ii) $\lim_{T\to\infty}\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbf{X}}'_{nt} J_n \tilde{\mathbf{X}}_{nt}$ *exists and is nonsingular.*

Let $Q_{nT,c}(\theta_1) = Q_{nT}\left(\theta_1, \beta_{nT}(\theta_1), \sigma^2_{\epsilon,nT}(\theta_1, \beta_{nT}(\theta_1))\right)$ where $\sigma^2_{\epsilon,nT}(\theta_1, \beta) = \arg\max_{\sigma^2_\epsilon} Q_{nT}(\theta_1, \beta, \sigma^2_\epsilon)$ and $\beta_{nT}(\theta_1) = \arg\max_\beta Q_{nT}(\theta_1, \beta, \sigma^2_\epsilon)$. Assumption 2.4.7 (i) comes from the information inequality for the concentrated expected log-likelihood function $Q_{nT,c}(\theta_1)$. Note that $\sigma^2_{\epsilon,nT}(\theta_1) = \frac{1}{nT} E\left(\sum_{t=1}^{T} \tilde{\mathcal{E}}'_{nt}(\theta_1, \beta_{nT}(\theta_1)) J_n \tilde{\mathcal{E}}_{nt}(\theta_1, \beta_{nT}(\theta_1))\right)$ and this expectation does not depend on a normal distribution, but it comes from the correctly specified first two moments. Also, we observe $\sigma^2_{\epsilon,nT}(\theta_1) = \sigma^2_{\epsilon,nT,1}(\theta_1) + \sigma^2_{\epsilon,nT,2}(\theta_1) + o(1)$ where

$$\sigma^2_{\epsilon,nT,1}(\theta_1) = \frac{1}{nT} \sum_{t=1}^{T} E\left(\tilde{Z}_{nt}(\theta_1) - \tilde{\mathbf{X}}_{nt}(\theta_1)\beta_{nT}(\theta_1)\right)' J_n \left(\tilde{Z}_{nt}(\theta_1) - \tilde{\mathbf{X}}_{nt}(\theta_1)\beta_{nT}(\theta_1)\right) \tag{2.26}$$

and $\sigma^2_{\epsilon,nT,2}(\theta_1) = \frac{\sigma^2_{\epsilon,0}}{n-1} tr\left(R_n^{-1\prime} R'_n(\theta_1) J_n R_n(\theta_1) R_n^{-1}\right)$. Note that $J_n \tilde{Z}_{nt}(\theta_1)$ is an approximation function for $J_n \tilde{\mathbf{X}}_{nt}\beta_0$ since $J_n \tilde{Z}_{nt}(\theta_{1,0}) = J_n \tilde{\mathbf{X}}_{nt}\beta_0$. Hence, the first term, $\sigma^2_{\epsilon,nT,1}(\theta_1)$, is a quadratic function of the difference between the two approximation functions for $J_n \tilde{\mathbf{X}}_{nt}\beta_0$ while $\sigma^2_{\epsilon,nT,2}(\theta_1) = E\left(\tilde{\mathcal{E}}'_{nt} R_n^{-1\prime} R'_n(\theta_1) J_n R_n(\theta_1) R_n^{-1} \tilde{\mathcal{E}}_{nt}\right)$, which is strictly positve. When θ_1 approaches to $\theta_{1,0}$, $\sigma^2_{\epsilon,nT,1}(\theta_1)$ is close to zero. Hence, $\sigma^2_{\epsilon,nT,2}(\theta_1)$ will play a main role in identifying $\theta_{1,0}$ if θ_1 is around $\theta_{1,0}$. Identifying β_0 is done by Assumption 2.4.7 (ii), which is analogous to identification of β_0 in a standard linear regression once $\theta_{1,0}$ is identified. When replacing $\tilde{\mathbf{X}}_{nt}$ by $\tilde{X}_{nt}$, we can observe this feature and Assumption 2.4.7 (ii) becomes equivalent to the identification condition of β_0 in conventional SDPD models. These conditions (i) and (ii) validate the strict information inequality (in the limit at least) so that θ_0 is globally identifiable.

Here is the theorem showing consistency of $\hat{\theta}_{ml,nT}$.

Theorem 2.4.1 *Suppose Assumptions 2.4.1 - 2.4.7 hold. Then,* $\hat{\theta}_{ml,nT} \to_p \theta_0$ *as* $T \to \infty$.

Next, we will derive the asymptotic distribution of $\hat{\theta}_{ml,nT}$. Denote $\Sigma_{\theta_0,nT} = -E\left(\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}(\theta_0)}{\partial\theta\partial\theta'}\right)$ and $\Omega_{\theta_0,nT} = E\left(\frac{1}{nT}\frac{\partial \ln L^{(u)}_{nT,c}(\theta_0)}{\partial\theta}\frac{\partial \ln L^{(u)}_{nT,c}(\theta_0)}{\partial\theta'}\right)$. For that, we introduce the following assumption.

Assumption 2.4.8 $\liminf_{n,T\to\infty} \phi_{\min}\left(\Omega_{\theta_0,nT}\right) > 0$ *and* $\liminf_{n,T\to\infty} \phi_{\min}\left(\Sigma_{\theta_0,nT}\right) > 0$ *where* $\phi_{\min}(\cdot)$ *denotes the smallest eigenvalue.*

Due to Assumption 2.4.5 (ii), we have continuity of $\Sigma_{\theta,nT} = -E\left(\frac{1}{nT}\frac{\partial \ln L_{nT,c}(\theta)}{\partial\theta\partial\theta'}\right)$ in $\theta \in \mathcal{N}(\theta_0)$ where $\mathcal{N}(\theta_0)$ denotes some neighborhood of θ_0. Hence, assuming $\inf_{n,T}\phi_{\min}\left(\Sigma_{\theta_0,nT}\right) > 0$ implies that $\Sigma_{\theta,nT}$ is also nonsingular for any $\theta \in \mathcal{N}(\theta_0)$. The derivation of the asymptotic normality of $\hat{\theta}_{ml,nT}$ will be based on the mean value theorem, and the central limit theorem for martingale difference arrays to $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{(u)}_{nT,c}(\theta_0)}{\partial\theta}$. The theorem below gives the asymptotic distribution of $\hat{\theta}_{ml,nT}$.

Theorem 2.4.2 *Suppose Assumptions 2.4.1 - 2.4.8 hold. Then,*

$$\begin{aligned}&\sqrt{nT}\left(\hat{\theta}_{ml,nT} - \theta_0\right) + \sqrt{\frac{n}{T}}\Sigma^{-1}_{\theta_0,nT}a_{n,1}(\theta_0) + \sqrt{\frac{T}{n}}\Sigma^{-1}_{\theta_0,nT}a_{n,2}(\theta_0)\\&+O_p\left(\max\left(\sqrt{\frac{n}{T^3}},\sqrt{\frac{T}{n^3}},\sqrt{\frac{1}{T}}\right)\right)\\&\to_d N\left(0,\Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}\right),\end{aligned}$$

where $\Omega_{\theta_0} = \lim_{T\to\infty}\Omega_{\theta_0,nT}$ *and* $\Sigma_{\theta_0} = \lim_{T\to\infty}\Sigma_{\theta_0,nT}$.

By Theorem 2.4.2, we have the results: (i) if $\frac{n}{T}\to 0$, $n\left(\hat{\theta}_{ml,nT} - \theta_0\right) + \Sigma^{-1}_{\theta_0,nT}a_{n,2}(\theta_0) \to_p 0$, (ii) if $\frac{n}{T}\to c \in (0,\infty)$, $\sqrt{nT}\left(\hat{\theta}_{ml,nT} - \theta_0\right) + \sqrt{c}\Sigma^{-1}_{\theta_0,nT}a_{n,1}(\theta_0) + \sqrt{\frac{1}{c}}\Sigma^{-1}_{\theta_0,nT}a_{n,2}(\theta_0) \to_d N\left(0,\Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}\right)$, and (iii) if $\frac{n}{T}\to\infty$, $T\left(\hat{\theta}_{ml,nT} - \theta_0\right) + \Sigma^{-1}_{\theta_0,nT}a_{n,1}(\theta_0) \to_p 0$. $\hat{\theta}_{ml,nT}$ has an asymptotic bias of order $O\left(\max\left\{\frac{1}{n},\frac{1}{T}\right\}\right)$ due to $-\frac{1}{T}\Sigma^{-1}_{\theta_0,nT}a_{n,1}(\theta_0) - \frac{1}{n}\Sigma^{-1}_{\theta_0,nT}a_{n,2}(\theta_0)$. Hence, the confidence interval for $\hat{\theta}_{ml,nT}$ is not properly centered at θ_0 even if n and

T have the same order (that is, $\frac{n}{T} \to c \in (0, \infty)$). If n and T do not have the same order, $\hat{\theta}_{ml,nT}$ will be degenerated. Hence, a bias corrected estimator constructed by

$$\hat{\theta}^{c}_{ml,nT} = \hat{\theta}_{ml,nT} - \frac{1}{T}\left[-\Sigma^{-1}_{\theta_0,nT} a_{n,1}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \frac{1}{n}\left[-\Sigma^{-1}_{\theta_0,nT} a_{n,2}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}}, \quad (2.27)$$

can be valuable. The assumption below is introduced for $\hat{\theta}^{c}_{ml,nT}$.

Assumption 2.4.9 *$\sum_{h=0}^{\infty} A_n^h(\theta_1)$ and $\sum_{h=1}^{\infty} hA_n^{h-1}(\theta_1)$ are uniformly bounded in either row or column sums uniformly in a neighborhood of θ_0.*

Under Assumption 2.4.9, we have

$$\sqrt{\frac{n}{T}}\left(\left[\Sigma^{-1}_{\theta,nT} a_{n,1}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \Sigma^{-1}_{\theta_0,nT} a_{n,1}(\theta_0)\right) \to_p 0, and$$
$$\sqrt{\frac{T}{n}}\left(\left[\Sigma^{-1}_{\theta,nT} a_{n,2}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \Sigma^{-1}_{\theta_0,nT} a_{n,2}(\theta_0)\right) \to_p 0$$

when $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$. Hence, we can apply the asymptotic equivalence.[30]

Corollary 2.4.3 *Under the additional Assumption 2.4.9, $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$, then*

$$\sqrt{nT}\left(\hat{\theta}^{c}_{ml,nT} - \theta_0\right) \to_d N\left(0, \Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}\right). \quad (2.28)$$

For the bias-adjusted estimator $\hat{\theta}^{c}_{ml,nT}$, if n and T are not too much large relative to each other, it can have a nondegenerate distribution and its confidence interval can properly be centered. For finite samples performance, results from Monte Carlo simulations are in Section 2.5.

[30]That is, if (i)

$$\sqrt{nT}\left(\hat{\theta}^{c}_{ml,nT} - \theta_0\right) - \sqrt{nT}\left(\hat{\theta}_{ml,nT} - \frac{1}{T}\left[-\Sigma^{-1}_{\theta_0,nT} a_{n,1}(\theta_0)\right] - \frac{1}{n}\left[-\Sigma^{-1}_{\theta_0,nT} a_{n,2}(\theta_0)\right] - \theta_0\right) \to_p 0$$

and (ii) $\sqrt{nT}\left(\hat{\theta}_{ml,nT} - \frac{1}{T}\left[-\left(\Sigma_{\theta_0,nT}\right)^{-1} a_{n,1}(\theta_0)\right] - \frac{1}{n}\left[-\Sigma^{-1}_{\theta_0,nT} a_{n,2}(\theta_0)\right] - \theta_0\right) \to_d N(0, *)$ where $*$ denotes the asymptotic variance derived in Corollary 2.4.3, we also have $\sqrt{nT}\left(\hat{\theta}^{c}_{ml,nT} - \theta_0\right) \to_d N(0, *)$.

Next, consider asymptotic properties of $\hat{\mathbf{c}}_{n,ml}(\hat{\theta}_{ml,nT})$ and $\hat{\alpha}_{t,ml}(\hat{\theta}_{ml,nT})$ for $t = 1, \cdots, T$. Recovering $\mathbf{c}_{n0}$ and α_{t0}'s is meaningful because they are employed to obtain welfare measures. To identify $\mathbf{c}_{n0}$ and α_{t0}'s, we impose the normalization restriction $\sum_{t=1}^{T} \alpha_{t0} = 0$ because $c_{i,0} + \alpha_{t0} = (c_{i,0} + x) + (\alpha_{t0} - x)$ for any x. Since T goes to infinity and n is an increasing function of T, consistently estimating $\mathbf{c}_{n0}$ and α_{t0}'s is feasible. For each θ, define $\hat{r}_{nt}(\theta) = R_n(\theta_1) Y_{nt} - (\gamma I_n + \rho W_n) Y_{n,t-1} - (I_n + \delta L_n^*(\theta_1) \Pi_n) X_{nt}\beta$. Because we impose $\sum_{t=1}^{T} \alpha_{t0} = 0$, $\hat{\mathbf{c}}_{n,ml}(\theta) = \frac{1}{T} \sum_{t=1}^{T} \hat{r}_{nt}(\theta)$ and $\hat{\alpha}_{t,ml}(\theta) = \frac{1}{n} l_n' [\hat{r}_{nt}(\theta) - \hat{\mathbf{c}}_{n,ml}(\theta)]$. Two estimates for $\mathbf{c}_{n0} + \alpha_{t0} l_n + \mathcal{E}_{nt}$ can be considered: (i) $\hat{r}_{nt}\left(\hat{\theta}_{ml,nT}\right)$, and (ii) $\hat{r}_{nt}\left(\hat{\theta}^c_{ml,nT}\right)$. The theorem below shows their asymptotic properties.

Theorem 2.4.4 *Suppose Assumptions 2.4.1 - 2.4.8 hold. Additionally, assume* $\sum_{t=1}^{T} \alpha_{t0} = 0$. *Then,*

(i) for each i, if $\frac{\sqrt{T}}{n} \to 0$, $\sqrt{T}(\hat{c}_{i,ml} - c_{i,0}) \to_d N(0, \sigma^2_{\epsilon,0})$ where $\hat{c}_{i,ml} = \hat{c}_{i,ml}(\hat{\theta}_{ml,nT})$ and they are asymptotically independent with each other.

(ii) For each t, if $\frac{\sqrt{n}}{T} \to 0$, $\sqrt{n}(\hat{\alpha}_{t,ml} - \alpha_{t0}) \to_d N(0, \sigma^2_{\epsilon,0})$ where $\hat{\alpha}_{t,ml} = \hat{\alpha}_{t,ml}(\hat{\theta}_{ml,nT})$ and they are asymptotically independent with each other.

(iii) Assume Assumption 2.4.9, $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$. For each i, $\sqrt{T}\left(\hat{c}^c_{i,ml} - c_{i,0}\right) \to_d N(0, \sigma^2_{\epsilon,0})$ where $\hat{c}^c_{i,ml} = \hat{c}_{i,ml}(\hat{\theta}^c_{ml,nT})$. For each t, $\sqrt{n}\left(\hat{\alpha}^c_{t,ml} - \alpha_{t0}\right) \to_d N(0, \sigma^2_{\epsilon,0})$ where $\hat{\alpha}^c_{t,ml} = \hat{\alpha}_{t,ml}(\hat{\theta}^c_{ml,nT})$. Asymptotic independence holds like (i) and (ii).

Parts (i) and (ii) show that the conditions are symmetric for the other effects. By Theorem 2.4.2, we have the convergence rate of $\hat{\theta}_{ml,nT}$
(i.e., $\hat{\theta}_{ml,nT} - \theta_0 = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \frac{1}{n}\right)\right)$). Then, $\hat{c}_{i,ml} - c_{i,0} = \frac{1}{T}\sum_{t=1}^{T} \epsilon_{it} + O_p(1) \cdot \left\|\hat{\theta}_{ml,nT} - \theta_0\right\|$ and $\hat{\alpha}_{t,ml} - \alpha_{t0} = \frac{1}{n}\sum_{i=1}^{n} \epsilon_{it} + O_p(1) \cdot \left\|\hat{\theta}_{ml,nT} - \theta_0\right\|$. Hence, the conditions

$\frac{\sqrt{T}}{n} = o(1)$ for $\hat{c}_{i,ml}$ and $\frac{\sqrt{n}}{T} = o(1)$ for $\hat{\alpha}_{t,ml}$ come respectively from[31] $\sqrt{T}\left(\hat{c}_{i,ml} - c_{i,0}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{it} + O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}, \frac{\sqrt{T}}{n}\right)\right)$,

and $\sqrt{n}\left(\hat{\alpha}_{t,ml} - \alpha_{t0}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{it} + O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}, \frac{\sqrt{n}}{T}\right)\right)$. Note that the residuals $\hat{r}_{nt}\left(\hat{\theta}_{ml,nT}\right)$ contain the individual- and time-dummy as an additive way. If T is large with small n, there exists a $O\left(\frac{1}{n}\right)$ bias for the regression coefficients since there are only n observations for each time dummy. For the estimate of individual effects, $\hat{c}_{i,ml}$, so $\frac{\sqrt{T}}{n} \to 0$ would appear in its asymptotic distribution normalized by $\frac{1}{\sqrt{T}}$. The symmetric argument can be applied to $\hat{\alpha}_{t,ml}$.

Part (iii) means the ratio conditions of n and T can be relaxed when we employ the residuals based on $\hat{\theta}^{c}_{ml,nT}$. Corollary 2.4.3 implies $\hat{\theta}^{c}_{ml,nT} - \theta_0 = O_p\left(\frac{1}{\sqrt{nT}}\right)$ if $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$. Then, $\sqrt{T}\left(\hat{c}^{c}_{i,ml} - c_{i,0}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{it} + O_p\left(\frac{1}{\sqrt{n}}\right)$, and $\sqrt{n}\left(\hat{\alpha}^{c}_{t,ml} - \alpha_{t0}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{it} + O_p\left(\frac{1}{\sqrt{T}}\right)$. Since $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$ are milder conditions than $\frac{\sqrt{T}}{n} \to 0$ and $\frac{\sqrt{n}}{T} \to 0$, estimating both $\mathbf{c}_{n0}$ and $\alpha_{T,0}$ via (ii) $\hat{r}_{nt}\left(\hat{\theta}^{c}_{ml,nT}\right)$ would be beneficial compared to employing $\hat{r}_{nt}\left(\hat{\theta}_{ml,nT}\right)$.

2.4.2 Nonlinear two-stage least squares (NL2S) estimation

In practical applications, we may like to have a robust estimator to unknown heteroskedasticity and/or unknown serial/cross-sectional correlations. Under a limited information setting, the NL2S method can be a reasonable estimation approach. In addition to possible robustness, it might have computational advantage relative to the ML or QML methods by avoiding evaluating $\ln|R_n(\theta_1)|$. In this subsection, we briefly discuss the implementation of this method.

[31] In conventional SDPD literature (e.g., Yu et al. (2008), and Lee and Yu (2012)), the convergence rate of the QMLE is $O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}\right)\right)$. In this case, the condition $\frac{\sqrt{T}}{n} = o(1)$ for $\hat{c}_{i,ml}$ is not required. Since we adopt the direct estimation approach of estimating $\mathbf{c}_{n0}$ and $\alpha_{T,0}$, we have the different convergence rate of the QMLE.

For each t, let Z_{nt} be the $n \times q$ IV matrix where $q \geq 4+K$ means the order condition of identifiability. By observing the form of additional endogenous component $Q_n^* Y_{nt}$, we can consider $[Y_{n,t-1}, X_{nt}]$ and its transformations by $[I_n, W_n, W_n', W_n' W_n, W_n^2, \cdots]$ as IVs. Define the sample moment function $g_{nT}^{\mathbf{L}}(\theta, \mathbf{c}_n, \alpha_T) = \frac{1}{nT}\sum_{t=1}^T Z_{nt}' \mathcal{E}_{nt}(\theta, \mathbf{c}_n, \alpha_t)$ and observe $E\left(g_{nT}^{\mathbf{L}}(\theta_0, \mathbf{c}_{n0}, \alpha_{T,0})\right) = \mathbf{0}_{q\times 1}$. Then, the NL2S estimator (NL2SE) can be obtained by minimizing the objective function:
$g_{nT}^{\mathbf{L}\prime}(\theta, \mathbf{c}_n, \alpha_T)\left(\frac{1}{nT}\sum_{t=1}^T Z_{nt}' Z_{nt}\right)^{-1} g_{nT}^{\mathbf{L}}(\theta, \mathbf{c}_n, \alpha_T)$.[32] For regularity conditions about IV Z_{nt}, we need to assume existence of $\text{plim}_{n,T\to\infty}\frac{1}{nT}\sum_{t=1}^T Z_{nt}' Z_{nt}$ and nonsingularity of it. Remaining conditions for consistency and asymptotic normality can be achieved by our suggested assumptions for the QML method.[33] In next section, we compare estimation results by the QML and NL2S methods to investigate whether the NL2S estimation method could work well.

2.5 Simulations

In this section, we report Monte Carlo simulation results on small sample performance of the QMLE. For $t = 1, \cdots, T$, the DGP for our simulation is

$$R_n Y_{nt} = \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + \sum_{k=1}^K (I_n + \delta L_n^* \Pi_n)(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n) X_{nt,k} + \mathbf{c}_{n0} + \alpha_{t0} l_n + \mathcal{E}_{nt} \tag{2.29}$$

and the expectation function Π_n is specified based on

$$X_{nt,k} = A_{k,n} X_{n,t-1,k} + \mathbf{c}_{n,k,0} + \alpha_{t,k,0} l_n + V_{nt,k} \tag{2.30}$$

[32]Since the incidental parameters c_{n0} and $\alpha_{T,0}$ are linear in $\mathcal{E}_{nt}(\theta, \mathbf{c}_n, \alpha_t)$, the concentrated statistical objected function will be $g_{nT,c}^{\mathbf{L}\prime}(\theta)\left(\frac{1}{nT}\sum_{t=1}^T Z_{nt}' Z_{nt}\right)^{-1} g_{nT,c}^{\mathbf{L}}(\theta)$ where $g_{nT,c}^{\mathbf{L}}(\theta) = \frac{1}{nT}\sum_{t=1}^T Z_{nt}' J_n \mathcal{E}_{nt}(\theta)$.

[33]For basic discussions on the NL2SE, refer to Theorems 8.1.1 and 8.1.2 in Amemiya (1985).

for $k = 1, \cdots, K$ where $A_{k,n} = \gamma_{k,0} I_n + \rho_{k,0} W_n$. We consider the joint estimation for the main parameter vector θ_0 and the nuisance parameters $\left\{\gamma_{k,0}, \rho_{k,0}, \sigma^2_{V,k,0}\right\}_{k=1}^{K}$ where $\sigma^2_{V,k,0} I_n$ is the variance of $V_{nt,k}$.[34]

For sample sizes, we consider the combinations of $n = 49$, 81 and $T = 10$, 30. We generate our data with $30 + T$ periods where the starting value is drawn from $N\left(\mathbf{0}_{n\times 1}, I_n\right)$, but employ the last T periods as our sample. This design makes the initial value Y_{n0} close to be in steady state. We experiment two cases with the primitive δ, (i) $\delta = 0.5$ (large discounted for the future) and (ii) $\delta = 0.95$ (small discounted for the future). The $\mathbf{c}_{n0}$, $\mathbf{c}_{n,k,0}$, α_{t0}, $\alpha_{t,k,0}$, $\mathcal{E}_{nt}$, and $V_{nt,k}$'s $(k = 1, \cdots, K)$ are independently drawn from the standard normal distribution. For W_n, a row-normalized rook matrix as for a chess board is utilized. We consider $K = 1$, and fix $\gamma_0 = 0.4$, $\beta_{1,1,0} = 0.4$, $\beta_{2,1,0} = 0.4$, $\sigma^2_{\epsilon,0} = 1$, $\gamma_{1,0} = 0.4$, $\rho_{1,0} = 0.1$ and $\sigma^2_{V,1,0} = 1$ throughout the experiment. For (λ_0, ρ_0), we consider four scenarios: (i) $(\lambda_0, \rho_0) = (0.2, 0.2)$, (ii) $(\lambda_0, \rho_0) = (0.2, -0.2)$, (iii) $(\lambda_0, \rho_0) = (-0.2, 0.2)$ and (iv) $(\lambda_0, \rho_0) = (-0.2, -0.2)$. The tolerance level of the inner loop is 0.0001 (evaluated by $\|\cdot\|_\infty$).[35] We compare performance of four estimators, (i) the QMLE $\hat{\theta}_{ml,nT}$ (ii) the bias corrected QMLE $\hat{\theta}^c_{ml,nT}$, (iii) QMLE as if $\delta = 0$ (denoted by $\hat{\theta}^S_{ml,nT}$) and (iv) the bias corrected QMLE as if $\delta = 0$ (denoted by $\hat{\theta}^{S,c}_{ml,nT}$). That is, $\hat{\theta}^S_{ml,nT}$ and $\hat{\theta}^{S,c}_{ml,nT}$ are the QMLEs based on Lee and Yu's (2010). In order to evaluate performance of estimators, we consider four criteria: (i) empirical bias, (ii) standard deviation (SD), (iii) empirical root

[34] As a simpler alternative, we can consider a two-step estimation instead of the joint estimation. In the first step, the nuisance parameters are estimated and generated regressors from the first step are used in the second step to estimate the structural parameters θ_0. However, it sometimes might yield a bad statistical inference without taking into account the asymptotic influence of the first step estimate through the generated regressors. See, e.g., Pagan (1984) and Murphy and Topel (1985). For the empirical analyses, we also take the joint estimation.

[35] This level is also applied to our empirical analysis.

mean square error (RMSE) and (iv) 95% coverage probability (CP).[36] The number of sample repetitions I is 400. The obtained MC results reported in Table 2.1 with $\delta = 0.95$ are summarized in Subsections 5.1 and 5.2.

Table 2.1: Performance of the QML estimators when $\delta = 0.95$
$(n, T) = (49, 10)$ and $(\lambda, \rho) = (0.2, 0.2)$

		λ	γ	ρ	β_1	β_2	σ_ϵ^2	γ_1	ρ_1	$\sigma_{V,1}^2$
		0.2	0.4	0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	−0.02	−0.15	0.05	0.04	0.06	−0.21	−0.15	−0.01	−0.14
	SD	0.07	0.06	0.08	0.04	0.08	0.07	0.04	0.09	0.06
	$RMSE$	0.07	0.16	0.10	0.06	0.10	0.23	0.16	0.09	0.15
	CP	0.93	0.22	0.86	0.84	0.88	0.14	0.06	0.94	0.31
$\hat{\theta}^c_{ml,nT}$	$Bias$	0.00	−0.03	0.00	0.02	0.03	−0.05	−0.03	−0.00	−0.04
	SD	0.07	0.07	0.09	0.04	0.08	0.09	0.05	0.09	0.07
	$RMSE$	0.07	0.07	0.09	0.05	0.09	0.10	0.06	0.09	0.08
	CP	0.93	0.84	0.90	0.92	0.93	0.79	0.87	0.91	0.81
$\hat{\theta}^S_{ml,nT}$	$Bias$	0.02	−0.19	0.02	0.00	0.03	−0.43			
	SD	0.06	0.04	0.07	0.04	0.07	0.04			
	$RMSE$	0.06	0.19	0.07	0.04	0.08	0.44			
	CP	0.92	0.00	0.95	0.94	0.92	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	0.05	−0.10	−0.04	−0.01	0.01	−0.37			
	SD	0.06	0.05	0.07	0.04	0.07	0.04			
	$RMSE$	0.08	0.11	0.08	0.04	0.07	0.37			
	CP	0.81	0.32	0.89	0.94	0.93	0.00			

2.5.1 The overall results

(i) The empirical biases of $\hat{\theta}_{ml,nT}$ and $\hat{\theta}^c_{ml,nT}$ tend to decrease when n and T are large. In particular, we have biases for $\hat{\gamma}_{ml,nT}$ $(\hat{\gamma}^c_{ml,nT})$, $\hat{\sigma}^2_{ml,nT}$ $(\hat{\sigma}^{2,c}_{ml,nT})$, $\hat{\gamma}_{1,ml,nT}$

[36] The 95% coverage probability is defined by

$$\frac{1}{I}\#_I\left\{[\theta_0]_l \in \left[\left[\hat{\theta}\right]_l - \frac{1.96}{\sqrt{nT}}\left[\Sigma_{\theta_0}^{-1}\widehat{\Omega_{\theta_0}}\Sigma_{\theta_0}^{-1}\right]_{ll}^{\frac{1}{2}}, \left[\hat{\theta}\right]_l + \frac{1.96}{\sqrt{nT}}\left[\Sigma_{\theta_0}^{-1}\widehat{\Omega_{\theta_0}}\Sigma_{\theta_0}^{-1}\right]_{ll}^{\frac{1}{2}}\right]\right\}$$

for $l = 1, \cdots, 4 + 5K$, I is the total number of sample repetitions,, $\#_I\{\cdot\}$ denotes the number of counts of coverage, where, $\hat{\theta}$ is an estimate of θ_0 and $\widehat{\Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1}}$ denotes a consistent estimate of $\Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1}$. We employ $\left[\Sigma_{\theta}^{-1}\Omega_{\theta}\Sigma_{\theta}^{-1}\right]_{\theta=\hat{\theta}_{ml,nT}}$ for $\Sigma_{\theta_0}^{-1}\widehat{\Omega_{\theta_0}}\Sigma_{\theta_0}^{-1}$.

Continued Table 2.1: $(n, T) = (49, 10)$ and $(\lambda, \rho) = (0.2, -0.2)$

		λ	γ	ρ	β_1	β_2	σ_ϵ^2	γ_1	ρ_1	$\sigma_{V,1}^2$
		0.2	0.4	−0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	*Bias*	−0.08	−0.16	0.10	0.03	0.04	−0.24	−0.15	−0.02	−0.14
	SD	0.07	0.06	0.09	0.04	0.08	0.07	0.04	0.09	0.06
	RMSE	0.11	0.17	0.13	0.05	0.09	0.26	0.16	0.09	0.15
	CP	0.78	0.16	0.77	0.89	0.91	0.07	0.06	0.94	0.31
$\hat{\theta}^c_{ml,nT}$	*Bias*	−0.02	−0.04	0.02	0.02	0.02	−0.06	−0.03	−0.01	−0.04
	SD	0.08	0.07	0.10	0.04	0.08	0.09	0.05	0.09	0.07
	RMSE	0.08	0.08	0.10	0.05	0.08	0.11	0.06	0.10	0.08
	CP	0.89	0.82	0.88	0.92	0.93	0.71	0.86	0.91	0.81
$\hat{\theta}^S_{ml,nT}$	*Bias*	−0.11	−0.20	0.12	−0.02	0.00	−0.46			
	SD	0.06	0.04	0.07	0.04	0.07	0.04			
	RMSE	0.13	0.20	0.14	0.04	0.07	0.46			
	CP	0.51	0.00	0.61	0.93	0.94	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	*Bias*	−0.08	−0.11	0.06	−0.03	−0.02	−0.39			
	SD	0.06	0.05	0.08	0.04	0.07	0.04			
	RMSE	0.10	0.12	0.10	0.05	0.07	0.40			
	CP	0.69	0.24	0.85	0.88	0.94	0.00			

Continued Table 2.1: $(n, T) = (49, 10)$ and $(\lambda, \rho) = (-0.2, 0.2)$

		λ	γ	ρ	β_1	β_2	σ_ϵ^2	γ_1	ρ_1	$\sigma_{V,1}^2$
		−0.2	0.4	0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	*Bias*	−0.02	−0.15	0.01	0.03	0.03	−0.23	−0.15	−0.02	−0.14
	SD	0.07	0.06	0.08	0.04	0.08	0.07	0.04	0.09	0.06
	RMSE	0.07	0.16	0.08	0.05	0.08	0.25	0.16	0.09	0.15
	CP	0.94	0.19	0.97	0.90	0.92	0.10	0.06	0.94	0.31
$\hat{\theta}^c_{ml,nT}$	*Bias*	−0.00	−0.03	−0.00	0.02	0.01	−0.06	−0.03	−0.01	−0.04
	SD	0.08	0.06	0.10	0.04	0.08	0.09	0.05	0.10	0.07
	RMSE	0.08	0.07	0.10	0.04	0.08	0.10	0.06	0.10	0.08
	CP	0.92	0.83	0.91	0.92	0.93	0.74	0.86	0.91	0.81
$\hat{\theta}^S_{ml,nT}$	*Bias*	0.05	−0.19	−0.02	−0.02	−0.02	−0.46			
	SD	0.06	0.04	0.07	0.04	0.07	0.04			
	RMSE	0.08	0.20	0.07	0.04	0.07	0.46			
	CP	0.87	0.00	0.94	0.90	0.93	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	*Bias*	0.07	−0.10	−0.04	−0.03	−0.03	−0.39			
	SD	0.06	0.04	0.08	0.04	0.07	0.04			
	RMSE	0.10	0.11	0.09	0.05	0.08	0.40			
	CP	0.72	0.23	0.87	0.86	0.93	0.00			

($\hat{\gamma}^c_{1,ml,nT}$), $\hat{\rho}_{1,ml,nT}$ ($\hat{\rho}^c_{1,ml,nT}$) and $\hat{\sigma}^2_{V,1,ml,nT}$ ($\hat{\sigma}^{2,c}_{V,1,ml,nT}$), which are reduced substantially as sample sizes become larger. While the empirical biases diminish when n and T

Continued Table 2.1: $(n,T) = (49,10)$ and $(\lambda,\rho) = (-0.2,-0.2)$

		λ	γ	ρ	β_1	β_2	σ^2_ϵ	γ_1	ρ_1	$\sigma^2_{V,1}$
		-0.2	0.4	-0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.07	-0.16	0.07	0.03	0.03	-0.23	-0.15	-0.02	-0.14
	SD	0.07	0.06	0.08	0.04	0.08	0.07	0.04	0.09	0.06
	$RMSE$	0.10	0.17	0.11	0.05	0.08	0.24	0.16	0.09	0.15
	CP	0.78	0.14	0.86	0.88	0.92	0.10	0.06	0.94	0.31
$\hat{\theta}^c_{ml,nT}$	$Bias$	-0.02	-0.04	0.03	0.02	0.01	-0.06	-0.03	-0.01	-0.04
	SD	0.07	0.06	0.09	0.04	0.08	0.09	0.05	0.09	0.07
	$RMSE$	0.08	0.07	0.10	0.05	0.08	0.10	0.06	0.10	0.08
	CP	0.90	0.79	0.89	0.91	0.93	0.74	0.86	0.91	0.81
$\hat{\theta}^S_{ml,nT}$	$Bias$	-0.07	-0.19	0.09	-0.01	-0.01	-0.44			
	SD	0.06	0.04	0.07	0.04	0.07	0.04			
	$RMSE$	0.09	0.20	0.11	0.04	0.07	0.44			
	CP	0.69	0.00	0.74	0.94	0.94	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	-0.05	-0.10	0.06	-0.02	-0.02	-0.37			
	SD	0.06	0.05	0.08	0.04	0.07	0.04			
	$RMSE$	0.08	0.11	0.10	0.04	0.08	0.38			
	CP	0.81	0.25	0.84	0.90	0.92	0.00			

Continued Table 2.1: $(n,T) = (49,30)$ and $(\lambda,\rho) = (0.2,0.2)$

		λ	γ	ρ	β_1	β_2	σ^2_ϵ	γ_1	ρ_1	$\sigma^2_{V,1}$
		0.2	0.4	0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.04	-0.04	0.04	0.02	0.04	-0.08	-0.05	0.00	-0.06
	SD	0.04	0.03	0.05	0.02	0.04	0.04	0.03	0.05	0.04
	$RMSE$	0.06	0.06	0.06	0.03	0.06	0.09	0.06	0.05	0.07
	CP	0.84	0.70	0.84	0.82	0.84	0.53	0.49	0.93	0.60
$\hat{\theta}^c_{ml,nT}$	$Bias$	-0.00	-0.00	-0.00	0.00	0.01	-0.01	-0.00	0.00	-0.01
	SD	0.04	0.03	0.05	0.02	0.04	0.05	0.03	0.05	0.04
	$RMSE$	0.04	0.03	0.05	0.02	0.04	0.05	0.03	0.05	0.04
	CP	0.96	0.93	0.96	0.96	0.96	0.90	0.91	0.93	0.91
$\hat{\theta}^S_{ml,nT}$	$Bias$	0.02	-0.11	-0.00	-0.01	0.03	-0.38			
	SD	0.03	0.02	0.04	0.02	0.04	0.03			
	$RMSE$	0.04	0.11	0.04	0.02	0.05	0.38			
	CP	0.87	0.00	0.96	0.94	0.87	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	0.06	-0.08	-0.04	-0.02	0.01	-0.35			
	SD	0.03	0.02	0.04	0.02	0.04	0.03			
	$RMSE$	0.07	0.09	0.05	0.03	0.04	0.36			
	CP	0.55	0.06	0.83	0.85	0.95	0.00			

increase, contribution of large T for reducing biases is relatively larger compared to that of large n.

Continued Table 2.1: $(n, T) = (49, 30)$ and $(\lambda, \rho) = (0.2, -0.2)$

		λ 0.2	γ 0.4	ρ -0.2	β_1 0.4	β_2 0.4	σ_ϵ^2 1	γ_1 0.4	ρ_1 0.1	$\sigma_{V,1}^2$ 1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.06	-0.05	0.04	0.02	0.03	-0.09	-0.05	-0.00	-0.06
	SD	0.04	0.03	0.05	0.02	0.04	0.04	0.03	0.05	0.04
	$RMSE$	0.07	0.06	0.06	0.03	0.05	0.10	0.06	0.05	0.07
	CP	0.75	0.59	0.90	0.87	0.89	0.36	0.49	0.93	0.60
$\hat{\theta}^c_{ml,nT}$	$Bias$	-0.00	-0.01	-0.00	0.00	0.00	-0.01	-0.00	0.00	-0.01
	SD	0.04	0.03	0.05	0.02	0.04	0.05	0.03	0.05	0.04
	$RMSE$	0.04	0.03	0.05	0.02	0.04	0.05	0.03	0.05	0.04
	CP	0.96	0.93	0.95	0.95	0.95	0.88	0.91	0.93	0.91
$\hat{\theta}^S_{ml,nT}$	$Bias$	-0.11	-0.12	0.07	-0.02	-0.00	-0.41			
	SD	0.03	0.02	0.04	0.02	0.04	0.02			
	$RMSE$	0.12	0.12	0.08	0.03	0.04	0.41			
	CP	0.08	0.00	0.62	0.78	0.95	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	-0.08	-0.09	0.04	-0.03	-0.02	-0.38			
	SD	0.03	0.02	0.04	0.02	0.04	0.03			
	$RMSE$	0.09	0.09	0.06	0.04	0.04	0.38			
	CP	0.31	0.04	0.81	0.65	0.95	0.00			

Continued Table 2.1: $(n, T) = (49, 30)$ and $(\lambda, \rho) = (-0.2, 0.2)$

		λ -0.2	γ 0.4	ρ 0.2	β_1 0.4	β_2 0.4	σ_ϵ^2 1	γ_1 0.4	ρ_1 0.1	$\sigma_{V,1}^2$ 1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.03	-0.05	0.01	0.02	0.02	-0.09	-0.05	-0.00	-0.06
	SD	0.04	0.03	0.05	0.02	0.04	0.04	0.03	0.05	0.04
	$RMSE$	0.05	0.06	0.05	0.03	0.05	0.10	0.06	0.05	0.07
	CP	0.88	0.65	0.95	0.88	0.90	0.39	0.49	0.93	0.60
$\hat{\theta}^c_{ml,nT}$	$Bias$	-0.00	-0.01	-0.01	0.00	0.00	-0.01	-0.00	0.00	-0.01
	SD	0.04	0.03	0.05	0.02	0.04	0.05	0.03	0.05	0.04
	$RMSE$	0.04	0.03	0.05	0.02	0.04	0.05	0.03	0.05	0.04
	CP	0.95	0.92	0.94	0.95	0.96	0.88	0.91	0.93	0.91
$\hat{\theta}^S_{ml,nT}$	$Bias$	0.05	-0.12	-0.03	-0.03	-0.02	-0.41			
	SD	0.03	0.02	0.04	0.02	0.04	0.02			
	$RMSE$	0.06	0.12	0.05	0.04	0.04	0.41			
	CP	0.72	0.00	0.89	0.67	0.95	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	0.08	-0.09	-0.05	-0.04	-0.03	-0.38			
	SD	0.03	0.02	0.04	0.02	0.04	0.03			
	$RMSE$	0.08	0.09	0.06	0.04	0.05	0.38			
	CP	0.35	0.03	0.79	0.55	0.87	0.00			

(ii) $\hat{\theta}^c_{ml,nT}$ performs better with smaller empirical biases and RMSE compared to those of $\hat{\theta}_{ml,nT}$. The biases observed in $\hat{\gamma}_{ml,nT}$, $\hat{\rho}_{ml,nT}$, $\hat{\sigma}^2_{ml,nT}$, $\hat{\gamma}_{1,ml,nT}$, $\hat{\rho}_{1,ml,nT}$ and $\hat{\sigma}^2_{V,1,ml,nT}$ can be corrected by the bias correction procedure.

Continued Table 2.1: $(n,T) = (49,30)$ and $(\lambda,\rho) = (-0.2,-0.2)$

		λ	γ	ρ	β_1	β_2	σ_ϵ^2	γ_1	ρ_1	$\sigma_{V,1}^2$
		-0.2	0.4	-0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.05	-0.05	0.03	0.02	0.02	-0.09	-0.05	-0.00	-0.06
	SD	0.04	0.03	0.05	0.02	0.04	0.04	0.03	0.05	0.04
	$RMSE$	0.06	0.06	0.05	0.03	0.05	0.10	0.06	0.05	0.07
	CP	0.78	0.52	0.93	0.88	0.92	0.36	0.49	0.93	0.60
$\hat{\theta}^c_{ml,nT}$	$Bias$	-0.00	-0.01	0.00	0.00	0.00	-0.01	-0.00	0.00	-0.01
	SD	0.04	0.03	0.05	0.02	0.04	0.05	0.03	0.05	0.04
	$RMSE$	0.04	0.03	0.05	0.02	0.04	0.05	0.03	0.05	0.04
	CP	0.96	0.93	0.95	0.95	0.95	0.89	0.91	0.93	0.91
$\hat{\theta}^S_{ml,nT}$	$Bias$	-0.08	-0.12	0.05	-0.02	-0.01	-0.39			
	SD	0.03	0.02	0.04	0.02	0.04	0.03			
	$RMSE$	0.09	0.12	0.07	0.03	0.04	0.39			
	CP	0.25	0.00	0.72	0.82	0.95	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	-0.06	-0.08	0.04	-0.03	-0.02	-0.35			
	SD	0.03	0.02	0.04	0.02	0.04	0.03			
	$RMSE$	0.06	0.09	0.05	0.03	0.04	0.36			
	CP	0.58	0.05	0.85	0.74	0.93	0.00			

Continued Table 2.1: $(n,T) = (81,10)$ and $(\lambda,\rho) = (0.2,0.2)$

		λ	γ	ρ	β_1	β_2	σ_ϵ^2	γ_1	ρ_1	$\sigma_{V,1}^2$
		0.2	0.4	0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	0.00	-0.15	0.05	0.03	0.05	-0.21	-0.15	-0.01	-0.13
	SD	0.05	0.04	0.06	0.04	0.07	0.06	0.03	0.07	0.04
	$RMSE$	0.05	0.16	0.08	0.05	0.08	0.22	0.15	0.07	0.14
	CP	0.93	0.06	0.82	0.83	0.88	0.03	0.01	0.94	0.16
$\hat{\theta}^c_{ml,nT}$	$Bias$	0.01	-0.03	0.00	0.02	0.02	-0.04	-0.02	-0.00	-0.04
	SD	0.06	0.05	0.07	0.04	0.07	0.07	0.04	0.08	0.05
	$RMSE$	0.06	0.06	0.07	0.04	0.07	0.08	0.05	0.08	0.06
	CP	0.91	0.80	0.90	0.90	0.93	0.78	0.86	0.90	0.80
$\hat{\theta}^S_{ml,nT}$	$Bias$	0.04	-0.19	0.02	-0.01	0.02	-0.43			
	SD	0.05	0.03	0.06	0.03	0.06	0.03			
	$RMSE$	0.06	0.19	0.06	0.03	0.06	0.43			
	CP	0.80	0.00	0.92	0.91	0.90	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	0.06	-0.10	-0.04	-0.01	0.01	-0.37			
	SD	0.05	0.04	0.06	0.03	0.06	0.03			
	$RMSE$	0.08	0.10	0.07	0.03	0.06	0.37			
	CP	0.69	0.14	0.86	0.89	0.93	0.00			

(iii) In the case of $\hat{\theta}_{ml,nT}$, the coverage probabilities increase for all cases and approach to 0.95. The coverage probabilities of $\hat{\theta}_{ml,nT}$ also increase and are close to

Continued Table 2.1: $(n, T) = (81, 10)$ and $(\lambda, \rho) = (0.2, -0.2)$

		λ	γ	ρ	β_1	β_2	σ_ϵ^2	γ_1	ρ_1	$\sigma_{V,1}^2$
		0.2	0.4	-0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.06	-0.16	0.09	0.02	0.03	-0.24	-0.15	-0.01	-0.13
	SD	0.06	0.04	0.07	0.04	0.07	0.06	0.03	0.07	0.04
	$RMSE$	0.08	0.17	0.12	0.04	0.07	0.24	0.15	0.07	0.14
	CP	0.79	0.04	0.67	0.86	0.89	0.01	0.01	0.93	0.16
$\hat{\theta}_{ml,nT}^c$	$Bias$	-0.01	-0.04	0.02	0.02	0.01	-0.06	-0.02	-0.00	-0.04
	SD	0.06	0.05	0.08	0.04	0.07	0.07	0.04	0.08	0.05
	$RMSE$	0.06	0.06	0.08	0.04	0.07	0.09	0.05	0.08	0.06
	CP	0.90	0.78	0.88	0.90	0.92	0.74	0.86	0.91	0.80
$\hat{\theta}_{ml,nT}^S$	$Bias$	-0.10	-0.20	0.11	-0.02	-0.01	-0.45			
	SD	0.05	0.03	0.06	0.03	0.06	0.03			
	$RMSE$	0.11	0.20	0.12	0.04	0.06	0.45			
	CP	0.43	0.00	0.51	0.85	0.93	0.00			
$\hat{\theta}_{ml,nT}^{S,c}$	$Bias$	-0.08	-0.10	0.06	-0.03	-0.02	-0.39			
	SD	0.05	0.04	0.06	0.03	0.06	0.03			
	$RMSE$	0.09	0.11	0.08	0.04	0.06	0.39			
	CP	0.59	0.09	0.79	0.78	0.92	0.00			

Continued Table 2.1: $(n, T) = (81, 10)$ and $(\lambda, \rho) = (-0.2, 0.2)$

		λ	γ	ρ	β_1	β_2	σ_ϵ^2	γ_1	ρ_1	$\sigma_{V,1}^2$
		-0.2	0.4	0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	0.00	-0.15	0.01	0.02	0.02	-0.23	-0.15	-0.01	-0.13
	SD	0.06	0.04	0.07	0.04	0.07	0.05	0.03	0.07	0.04
	$RMSE$	0.06	0.16	0.07	0.04	0.07	0.23	0.15	0.07	0.14
	CP	0.95	0.05	0.94	0.88	0.90	0.01	0.01	0.94	0.16
$\hat{\theta}_{ml,nT}^c$	$Bias$	0.00	-0.03	-0.00	0.01	0.01	-0.05	-0.02	-0.00	-0.04
	SD	0.06	0.05	0.08	0.03	0.07	0.06	0.04	0.08	0.05
	$RMSE$	0.06	0.06	0.08	0.04	0.07	0.08	0.05	0.08	0.06
	CP	0.93	0.80	0.92	0.91	0.92	0.75	0.86	0.91	0.80
$\hat{\theta}_{ml,nT}^S$	$Bias$	0.06	-0.19	-0.03	-0.03	-0.02	-0.45			
	SD	0.05	0.03	0.06	0.03	0.06	0.03			
	$RMSE$	0.08	0.19	0.06	0.04	0.06	0.45			
	CP	0.72	0.00	0.91	0.82	0.92	0.00			
$\hat{\theta}_{ml,nT}^{S,c}$	$Bias$	0.08	-0.10	-0.05	-0.03	-0.04	-0.39			
	SD	0.05	0.03	0.06	0.03	0.06	0.03			
	$RMSE$	0.09	0.11	0.08	0.05	0.07	0.39			
	CP	0.61	0.09	0.85	0.73	0.89	0.00			

0.95 when we increase n and T. Overall, the results (i), (ii) and (iii) also hold for $\delta = 0.5$.

Continued Table 2.1: $(n, T) = (81, 10)$ and $(\lambda, \rho) = (-0.2, -0.2)$

		λ	γ	ρ	β_1	β_2	σ^2_ϵ	γ_1	ρ_1	$\sigma^2_{V,1}$
		-0.2	0.4	-0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.06	-0.16	0.06	0.03	0.02	-0.22	-0.15	-0.01	-0.13
	SD	0.05	0.04	0.07	0.03	0.07	0.05	0.03	0.07	0.04
	$RMSE$	0.08	0.16	0.09	0.04	0.07	0.23	0.15	0.07	0.14
	CP	0.78	0.03	0.82	0.86	0.91	0.01	0.01	0.94	0.16
$\hat{\theta}^c_{ml,nT}$	$Bias$	-0.02	-0.04	0.02	0.02	0.01	-0.05	-0.02	-0.00	-0.04
	SD	0.06	0.05	0.08	0.03	0.07	0.06	0.04	0.08	0.05
	$RMSE$	0.06	0.06	0.08	0.04	0.07	0.08	0.05	0.08	0.06
	CP	0.90	0.78	0.88	0.90	0.92	0.78	0.86	0.91	0.80
$\hat{\theta}^S_{ml,nT}$	$Bias$	-0.06	-0.19	0.08	-0.02	-0.01	-0.44			
	SD	0.05	0.03	0.06	0.03	0.06	0.03			
	$RMSE$	0.08	0.20	0.10	0.04	0.06	0.44			
	CP	0.66	0.00	0.67	0.87	0.92	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	-0.05	-0.10	0.05	-0.03	-0.03	-0.37			
	SD	0.05	0.03	0.06	0.03	0.06	0.03			
	$RMSE$	0.07	0.11	0.08	0.04	0.06	0.37			
	CP	0.76	0.09	0.79	0.83	0.91	0.00			

Continued Table 2.1: $(n, T) = (81, 30)$ and $(\lambda, \rho) = (0.2, 0.2)$

		λ	γ	ρ	β_1	β_2	σ^2_ϵ	γ_1	ρ_1	$\sigma^2_{V,1}$
		0.2	0.4	0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.02	-0.05	0.03	0.02	0.04	-0.07	-0.05	-0.00	-0.05
	SD	0.03	0.02	0.04	0.02	0.04	0.03	0.02	0.04	0.03
	$RMSE$	0.04	0.05	0.05	0.03	0.05	0.08	0.05	0.04	0.06
	CP	0.86	0.57	0.85	0.81	0.80	0.41	0.31	0.94	0.54
$\hat{\theta}^c_{ml,nT}$	$Bias$	-0.00	-0.00	0.00	0.00	0.01	-0.01	-0.00	0.00	-0.01
	SD	0.03	0.03	0.04	0.02	0.04	0.03	0.02	0.04	0.03
	$RMSE$	0.03	0.03	0.04	0.02	0.04	0.04	0.02	0.04	0.03
	CP	0.92	0.93	0.93	0.93	0.94	0.92	0.94	0.94	0.93
$\hat{\theta}^S_{ml,nT}$	$Bias$	0.04	-0.11	-0.00	-0.01	0.03	-0.38			
	SD	0.03	0.02	0.03	0.02	0.03	0.02			
	$RMSE$	0.05	0.11	0.03	0.02	0.04	0.38			
	CP	0.62	0.00	0.95	0.88	0.84	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	0.06	-0.08	-0.03	-0.02	0.01	-0.35			
	SD	0.03	0.02	0.03	0.02	0.03	0.02			
	$RMSE$	0.07	0.08	0.05	0.02	0.04	0.35			
	CP	0.31	0.01	0.80	0.80	0.92	0.00			

(iv) For $\hat{\theta}^S_{ml,nT}$ and $\hat{\theta}^{S,c}_{ml,nT}$, they do not have a good pattern of performance. The RMSEs and the coverage probabilities of $\hat{\theta}^S_{ml,nT}$ and $\hat{\theta}^{S,c}_{ml,nT}$ even tend to increase after the bias correction. Also, this tendency does not disappear for large n and T. For

Continued Table 2.1: $(n, T) = (81, 30)$ and $(\lambda, \rho) = (0.2, -0.2)$

		λ	γ	ρ	β_1	β_2	σ_ϵ^2	γ_1	ρ_1	$\sigma_{V,1}^2$
		0.2	0.4	-0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.04	-0.05	0.04	0.01	0.03	-0.08	-0.05	-0.00	-0.05
	SD	0.04	0.02	0.04	0.02	0.04	0.03	0.02	0.04	0.03
	$RMSE$	0.05	0.06	0.05	0.02	0.05	0.09	0.05	0.04	0.06
	CP	0.76	0.44	0.85	0.87	0.86	0.26	0.30	0.94	0.54
$\hat{\theta}^c_{ml,nT}$	$Bias$	-0.00	-0.00	0.00	0.00	0.00	-0.01	-0.00	0.00	-0.01
	SD	0.04	0.03	0.04	0.02	0.04	0.03	0.02	0.04	0.03
	$RMSE$	0.04	0.03	0.04	0.02	0.04	0.04	0.02	0.04	0.03
	CP	0.94	0.94	0.92	0.93	0.94	0.92	0.94	0.94	0.93
$\hat{\theta}^S_{ml,nT}$	$Bias$	-0.10	-0.12	0.07	-0.03	-0.00	-0.40			
	SD	0.03	0.02	0.03	0.02	0.03	0.02			
	$RMSE$	0.10	0.12	0.08	0.03	0.03	0.40			
	CP	0.04	0.00	0.38	0.64	0.93	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	-0.08	-0.09	0.04	-0.03	-0.02	-0.37			
	SD	0.03	0.02	0.03	0.02	0.03	0.02			
	$RMSE$	0.08	0.09	0.05	0.04	0.04	0.37			
	CP	0.16	0.01	0.73	0.50	0.90	0.00			

Continued Table 2.1: $(n, T) = (81, 30)$ and $(\lambda, \rho) = (-0.2, 0.2)$

		λ	γ	ρ	β_1	β_2	σ_ϵ^2	γ_1	ρ_1	$\sigma_{V,1}^2$
		-0.2	0.4	0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.02	-0.05	0.01	0.01	0.02	-0.08	-0.05	-0.00	-0.05
	SD	0.04	0.02	0.04	0.02	0.04	0.03	0.02	0.04	0.03
	$RMSE$	0.04	0.05	0.04	0.02	0.04	0.08	0.05	0.04	0.06
	CP	0.88	0.51	0.94	0.87	0.87	0.31	0.30	0.94	0.54
$\hat{\theta}^c_{ml,nT}$	$Bias$	-0.00	-0.00	-0.00	0.00	0.00	-0.01	-0.00	0.00	-0.01
	SD	0.04	0.03	0.04	0.02	0.04	0.03	0.02	0.04	0.03
	$RMSE$	0.04	0.03	0.04	0.02	0.04	0.03	0.02	0.04	0.03
	CP	0.93	0.94	0.93	0.93	0.94	0.91	0.94	0.94	0.93
$\hat{\theta}^S_{ml,nT}$	$Bias$	0.06	-0.12	-0.03	-0.03	-0.02	-0.40			
	SD	0.03	0.02	0.03	0.02	0.03	0.02			
	$RMSE$	0.07	0.12	0.04	0.03	0.04	0.40			
	CP	0.38	0.00	0.87	0.52	0.88	0.00			
$\hat{\theta}^{S,c}_{ml,nT}$	$Bias$	0.08	-0.09	-0.04	-0.04	-0.03	-0.37			
	SD	0.03	0.02	0.03	0.02	0.03	0.02			
	$RMSE$	0.08	0.09	0.05	0.04	0.05	0.37			
	CP	0.19	0.01	0.75	0.37	0.79	0.00			

all cases, $\hat{\theta}^S_{ml,nT}$ and $\hat{\theta}^{S,c}_{ml,nT}$ do not seem to work well due to crucial misspecification errors.

Continued Table 2.1: $(n, T) = (81, 30)$ and $(\lambda, \rho) = (-0.2, -0.2)$

		λ	γ	ρ	β_1	β_2	σ_ϵ^2	γ_1	ρ_1	$\sigma_{V,1}^2$
		-0.2	0.4	-0.2	0.4	0.4	1	0.4	0.1	1
$\hat{\theta}_{ml,nT}$	$Bias$	-0.04	-0.05	0.03	0.01	0.02	-0.08	-0.05	-0.00	-0.05
	SD	0.03	0.02	0.04	0.02	0.04	0.03	0.02	0.04	0.03
	$RMSE$	0.05	0.06	0.04	0.02	0.04	0.08	0.05	0.04	0.06
	CP	0.78	0.39	0.88	0.86	0.88	0.32	0.30	0.94	0.54
$\hat{\theta}_{ml,nT}^c$	$Bias$	-0.00	-0.01	0.00	0.00	0.00	-0.01	-0.00	0.00	-0.01
	SD	0.03	0.02	0.04	0.02	0.04	0.03	0.02	0.04	0.03
	$RMSE$	0.03	0.03	0.04	0.02	0.04	0.03	0.02	0.04	0.03
	CP	0.94	0.93	0.95	0.93	0.93	0.92	0.94	0.94	0.93
$\hat{\theta}_{ml,nT}^S$	$Bias$	-0.07	-0.11	0.05	-0.02	-0.01	-0.38			
	SD	0.03	0.02	0.03	0.02	0.03	0.02			
	$RMSE$	0.07	0.12	0.06	0.03	0.04	0.38			
	CP	0.22	0.00	0.56	0.73	0.90	0.00			
$\hat{\theta}_{ml,nT}^{S,c}$	$Bias$	-0.05	-0.08	0.04	-0.03	-0.02	-0.35			
	SD	0.03	0.02	0.03	0.02	0.03	0.02			
	$RMSE$	0.06	0.08	0.05	0.03	0.04	0.35			
	CP	0.42	0.01	0.73	0.60	0.88	0.00			

2.5.2 The results for specific parameters

(λ_0) In terms of empirical biases and coverage probabilities, $\hat{\lambda}_{ml,nT}^c$ works relatively better than $\hat{\lambda}_{ml,nT}$. For most cases, downward biases are observed. When $\rho_0 < 0$, it seems that $\hat{\lambda}_{ml,nT}$ and $\hat{\lambda}_{ml,nT}^c$ have relatively low coverage probabilities.

Based on $\hat{\lambda}_{ml,nT}^{S,c}$ when $(n, T) = (49, 10)$, the signs of misspecification biases are positive if $\rho_0 > 0$, but are negative if $\rho_0 < 0$. From these results, the sign of ρ_0 determines the sign of the misspecification bias of $\hat{\lambda}_{ml,nT}^{S,c}$ while the sign of λ_0 would not be so.

(γ_0) Under small T, $\hat{\gamma}_{ml,nT}$ has significant downward biases for all cases. When T increases, the absolute values of biases decrease. This result is consistent with those of Hahn and Kuersteiner (2002) for dynamic panels (with neither spatial nor intertemporal effects). The bias corrected $\hat{\gamma}_{ml,nT}^c$ reduces the bias.

Focusing on $\hat{\gamma}^{S,c}_{ml,nT}$ when $(n, T) = (49, 10)$, we observe misspecification biases in estimating γ_0 are negative and their degree of bias might be affected by values of λ_0 and ρ_0.

(ρ_0) For ρ_0, the magnitude of biases is smaller than that of γ_0. For all cases, we observe upward biases in $\hat{\rho}_{ml,nT}$. If $\lambda_0 > 0$ and $\rho_0 < 0$, substantial upward biases in $\hat{\rho}_{ml,nT}$ are observed. On the other hand, we detect relatively small upward biases in $\hat{\rho}_{ml,nT}$ if $\lambda_0 < 0$ and $\rho_0 > 0$. By introducing the bias correction to $\hat{\rho}_{ml,nT}$ or increasing n or T, the amount of bias decreases and coverage probabilities become better.

Consider the misspecification bias by focusing on $\hat{\rho}^{S,c}_{ml,nT}$. Based on $\hat{\rho}^{S,c}_{ml,nT}$ when $(n, T) = (49, 10)$, misspecification biases turn to be upward if $\rho_0 < 0$, but are downward if $\rho_0 > 0$. It seems that the sign of misspecification bias takes the opposite sign of ρ_0 but can be irrelevant to signs of λ_0.

($\beta_{1,1,0}$) Performances of $\hat{\beta}_{1,1,ml,nT}$ and $\hat{\beta}^{c}_{1,1,ml,nT}$ are reasonable in biases and coverage probabilities. For all cases, upward biases in $\hat{\beta}_{1,1,ml,nT}$ are detected but they diminish after correcting biases or increasing n or T.

To analyze the misspecification bias, consider $\hat{\beta}^{S,c}_{1,1,ml,nT}$ when $(n, T) = (49, 10)$. We observe downward biases and those biases increase when δ increases in absolute values.

($\beta_{2,1,0}$) Like the case of $\beta_{1,1,0}$, we detect upward biases in $\hat{\beta}_{2,1,ml,nT}$ but they decrease and coverage probabilities become better after correcting the biases or increasing n or T.

To study misspecification errors, focus on $\hat{\beta}^{S,c}_{2,1,ml,nT}$ with $(n, T) = (49, 10)$. When both λ_0 and $\rho_0 > 0$, there are upward misspecification biases in $\hat{\beta}^{S,c}_{2,1,ml,nT}$. For other cases, however, downward misspecification biases in $\hat{\beta}^{S,c}_{2,1,ml,nT}$ are observed.

$\left(\sigma^2_{\epsilon,0}\right)$ When n and T are small, biases of $\hat{\sigma}^2_{\epsilon,ml,nT}$ are downward and the bias correction is needed.

For all cases of $\hat{\sigma}^{2,S}_{\epsilon,ml,nT}$ and $\hat{\sigma}^{2,S,c}_{\epsilon,ml,nT}$, there are downward biases.

$(\gamma_{1,0})$ Properties of $\hat{\gamma}_{1,ml,nT}$ of X processes are very similar to $\hat{\gamma}_{ml,nT}$. That is, large downward biases in $\hat{\gamma}_{1,ml,nT}$ are observed but the bias can be reduced and the coverage probability can become more adequate from the bias correction.

$(\rho_{1,0})$ In case of $\rho_{1,0}$, $\hat{\rho}_{1,ml,nT}$ and $\hat{\rho}^c_{1,ml,nT}$ perform well with small biases and adequate coverage probabilities even for small samples.

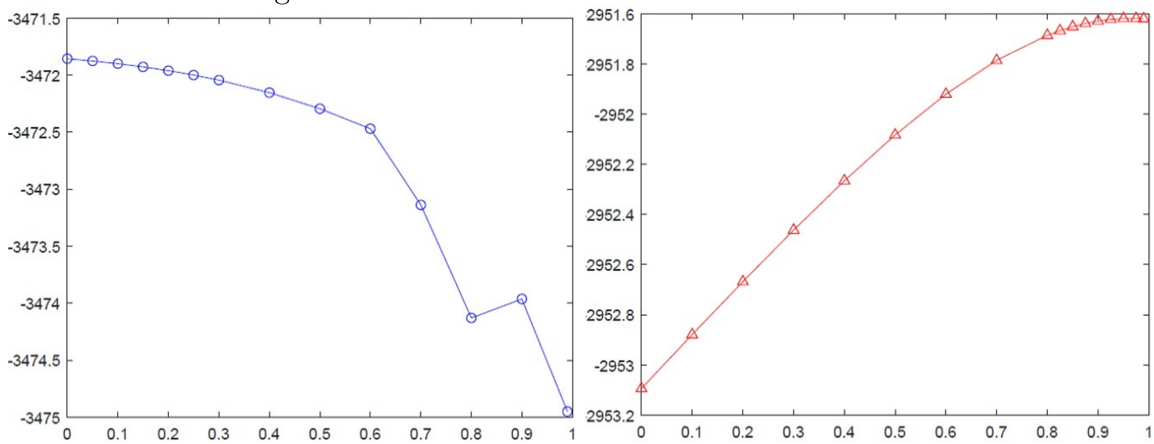
$\left(\sigma^2_{V,1,0}\right)$ Lastly, consider $\hat{\sigma}^2_{V,1,ml,nT}$ and $\hat{\sigma}^{2,c}_{V,1,ml,nT}$. Similar to $\sigma^2_{\epsilon,0}$, we detect a substantial downward bias for small $T = 10$ cases. By introducing the bias correction or increasing sample size T, biases are reduced and coverage probabilities are improved.

2.5.3 Identification of δ and effects of misspecified δ on estimation

In nonlinear structural econometric analyses, identifying the true time-discounting factor (δ_0) is a challenging issue since the statistical objective function is very flat around δ_0.[37] Hence, we conduct an additional experiment on identifying δ_0 the true time-discounting factor. To identify the true δ_0, we suggest using the log-likelihood measures such as the sample log-likelihood function, Akaike information criterion (AIC), and Bayesian information criterion (BIC). Employing those likelihood measures can be justified by the information inequality in likelihood theory. Via Figures 2.1 and 2.2, we report the sample likelihood functions across various δ's and the misspecification errors of estimating λ_0, γ_0, and ρ_0 in terms of the RMSE for the two representative cases: (i) $\delta_0 = 0$ and (ii) $\delta_0 = 0.95$ with a large finite sample

[37] Komarova et al. (2017) discuss this issue in a framework of dynamic discrete choice models.

Figure 2.1: Selection of δ via likelihood measures

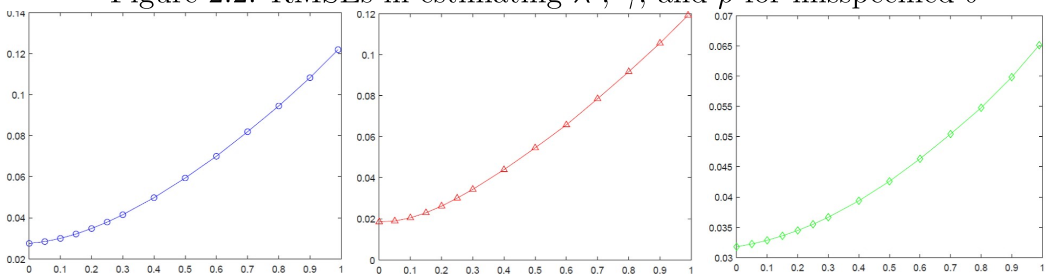


Note: We show two representative cases: (i) Myopic: $\delta = 0$, $(n, T) = (81, 30)$, and $K = 2$ (ii) Forward-looking: $\delta = 0.95$, $(n, T) = (81, 30)$, and $K = 2$ We set $\lambda = 0.2$, $\gamma = 0.4$, and $\rho = 0.2$, and other circumstances are the same as the main simulation. The x-axis shows δs while the y-axis reports the sample log-likelihood.

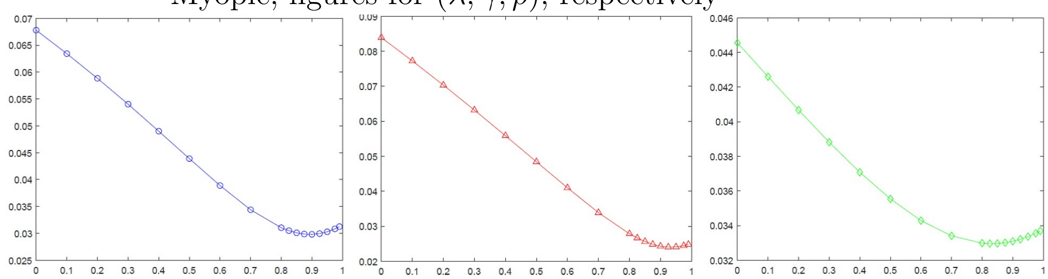
and rich exogenous variables. Additional results and discussions can be found in the supplementary file of Jeong and Lee (2018).

Throughout all cases, three observations can be summarized. First, having sufficiently large observations is needed to identify the true δ_0. If we do not have sufficient observations, we may not distinguish the true model via the likelihood measures. Second, the number of significant exogenous variables also affects identifying δ_0. Under same circumstance, including additional exogenous variables means a (relatively) high signal-to-noise ratio. If a portion of the explainable part is large, we can distinguish the myopic and forward-looking models by the likelihood measures and estimation results are less affected by misspecified δ's. Third, it is easier to identify δ_0 if the true model is a myopic one. It seems that the myopic model's complexity is much simpler, so less information might be required to identify δ_0, which is zero.

Figure 2.2: RMSEs in estimating λ , γ, and ρ for misspecified δ



Myopic, figures for (λ, γ, ρ), respectively



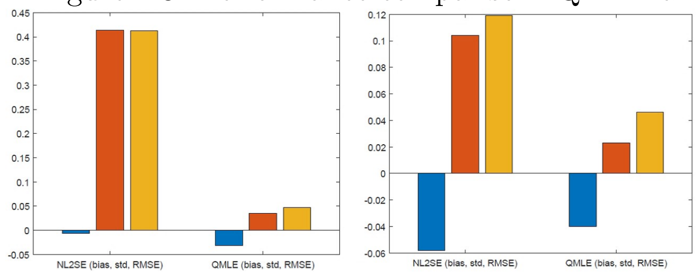
Forward-looking, figures for (λ, γ, ρ), respectively

Note: We show two representative cases: (i) Myopic: $\delta = 0$, $(n, T) = (81, 30)$, and $K = 2$ (ii) Forward-looking: $\delta = 0.95$, $(n, T) = (81, 30)$, and $K = 2$ We set $\lambda = 0.2$, $\gamma = 0.4$, and $\rho = 0.2$, and other circumstances are the same as the main simulation. The x-axis shows δs while the y-axis reports the sample log-likelihood.

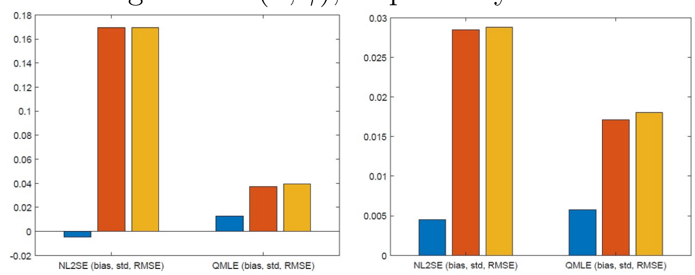
2.5.4 Performance comparison: QML and NL2S methods

In this subsection, we compare estimation performance of the QML and NL2S estimators. For this experiment, we set $(n, T) = (81, 30)$, $\delta = 0.95$, $\lambda_0 = 0.2$, $\gamma_0 = 0.4$, $\rho_0 = 0$, $\beta_{1,1,0} = \beta_{1,2,0} = 0.4$, $\beta_{2,1,0} = \beta_{2,2,0} = 0$, and other circumstances are the same as in the main simulation. This design means no spatial time lag as well as no Durbin regressor for simplicity. As IVs, we employ $[Y_{n,t-1}, X_{nt}]$ and its transformations by $[I_n, W_n, W_n', W_n'W_n, W_n^2]$. Under this circumstance, $W_n\,[Y_{n,t-1}, X_{nt}]$ can play an important role in identifying θ_0.

Figure 2.3: Performance comparison: QMLE and NL2SE



Figures for (λ, γ), respectively



Figures for (ρ, β), respectively

Note: We set $(n, T) = (81, 30)$, $\delta = 0.95$, $\lambda_0 = 0.2$, $\gamma_0 = 0.4$, $\rho_0 = 0$, $\beta_{1,1,0} = \beta_{1,2,0} = 0.4$, no Durbin regressor, and other circumstances are the same as the main simulation. As IVs for the NL2SE, we consider $[Y_{n,t-1}, X_{nt}]$, and its transformations by $[I_n, W_n, W_n', W_nW_n', W_n^2]$.

For each estimation method and parameter value, we report empirical bias, standard deviation, and RMSE as bar graphs (Figure 2.3).[38] Except for ρ_0, two methods show the same signs of empirical biases (negative for λ_0 and γ_0, and positive for $\beta_{1,1,0}$). The NL2SE tends to yield smaller magnitude of empirical biases than that of the QMLE (except for γ_0). In terms of standard deviation and RMSE, however, the NL2SE is worse than the QMLE. This implies the NL2SE is not efficient, so we may need to include more IVs or consider quadratic moment conditions to improve efficiency. If we include many moment conditions, however, it leads to additional

[38]We do not report results for $\beta_{1,2,0}$, which are similar to those of $\beta_{1,1,0}$.

biases (Lee and Yu (2014)). Compared to the main structural parameters λ_0, γ_0, and ρ_0, there is the relatively small gap in efficiency in estimating $\beta_{1,1,0}$.

In the aspect of computation costs, it seems using the NL2S method does not reduce computation time. In the inner loop, solutions of algebraic matrix Riccati equation $Q_n^*(\theta)$ and $L_n^*(\theta)$ are obtained for given θ, so $\overset{\heartsuit}{\mathcal{E}}_{nt}(\theta)$'s are calculated. Note that this procedure is required for both estimation methods. In the outer loop, however, parameter searching on Θ is conducted by optimizing different statistical objective functions. We expect reduced computation time in the outer loop by avoiding calculating $\ln|R_n(\theta)|$ if we use the NL2S method. Hence, the main computation costs might be originated from the inner loop. If we have very large n, calculating $\ln|R_n(\theta)|$ can be also demanding. For this situation, using approximation methods for $\ln|R_n(\theta)|$ will be helpful.

2.6 Application

In this section, we consider an application of our model. Since our model is based on strategic interactions stemming from fixed locations, we consider analyzing spillover effects of local governments' welfare spending. Two sources of strategic interactions can be considered in making local policies. First, welfare recipients can move in from or out to nearby cities to enjoy more beneficial policies. Second, the "yardstick competition" is considered. It means that a decision-maker of a local government has an incentive to make an efficient fiscal decision by comparing its decision with those of neighboring local governments. Since there exists "vote" to evaluate the performance of a local government by residents, this type of competitions arises. To econometrically investigate these strategic interactions, SAR and/or SDPD models describe optimal reaction functions of local governments when they play a simultaneous move game at each period. With payoff specification (2.4), conventional SDPD models present the vector of myopic best response functions while the intertemporal spatial dynamic model shows the forward-looking best responses.

In this paper, we consider public safety spending competitions among counties in North Carolina. Both myopic and forward-looking policy reaction functions are considered.[39] In the case of the public safety spending competition, a decision maker shall consider specific policy externalities. Those policy externalities arise since criminals can commit crimes with moving to neighboring cities and they are punished in every city. On one hand, a local government has an incentive to decrease its safety spending to enjoy "free-riding" effects when its neighbor spends more on public safety (substitution effect). On the other hand, a local government can increase its effort (public safety spending) to reduce overall criminal activities corresponding to a substantial safety spending in a neighboring city (similar to income effect in consumer theory). Yang and Lee (2017) consider a criminal's payoff function describing an incentive to commit a crime. Under certain conditions of payoff, they show the substitution effect will dominate. In both complete and incomplete information settings, they establish a SAR equation as a policy reaction function and find significant estimated substitution effects in cities' public safety spending. However, their framework is based on a static game, so a cross-sectional data set is employed.

We revisit this issue with a panel data set and two kinds of econometric specifications: (i) conventional SDPD model, and (ii) our intertemporal SAR model. From

[39]Reasons for considering our forward-looking model are that (i) a policymaker can be assumed to be benevolent (for the regional economic growth) and (ii) he/she has an incentive to make a forward-looking decision to keep his/her political reputation.

Table 2.2: Descriptive statistics: counties in North Carolina

Variables	Mean	Standard dev.	Min	Max
Public safety spending (10^6)	20.55	26.34	0.00	237.37
Total revenue (10^6)	126.23	216.43	0.00	1786.45
Proportion on total expenditure	0.19	0.06	0.00	0.45
Population (10^3)	94.43	140.55	4.14	1035.61
Land area (km^2)	1259.18	497.48	446.70	2457.92
Population density ($/km^2$)	74.76	99.90	3.40	763.93
Median ages	40.08	4.58	23.90	51.30
Median household income (10^4)	4.14	0.77	2.51	7.06
Distance (km)	248.15	147.84	12.26	751.90
No. of observations	1200	-	-	-

Note: Sample is 100 counties in North Carolina from 2005 to 2016. Dollar amounts are real values adjusted by the GDP deflator with base year 2009.

the North Carolina Department of State Treasurer's website, we obtain the government finance data. The data on counties' demographic and economic characteristics are from the United States Census Bureau. We have samples of 100 counties in North Carolina from 2005 to 2016 (total 1,200 observations). We construct a panel data set, so it might capture the dynamics of local governments' decision-making and their demographic/economic characteristics.[40] Table 2.2 summarizes the sample statistics. All dollar amounts are real values adjusted by the GDP deflator with the base year 2009. We observe that counties have distinct characteristics in financial status as well as economic/demographic characteristics. There are substantial differences among county governments' revenues, amounts of public safety spending, and proportion of expenditures on public safety. The maximal public safety spending is

[40]For some demographic and economic variables (Median ages and Median household income), there are some missing observations from 2005 to 2008 (164 observations among 1,200 observations). To get a balanced panel data set, we conduct the extrapolation scheme.

237.365 million dollars, and the minimal one is zero. The number of observations taking zero is 31 among a total of 1,200 observations (2.58%).[41] In the proportion of expenditures on public safety, the average is 19.3%, and the standard deviation is 0.06%. The largest portion is 44.8% while the smallest one is 0%. County governments in North Carolina also differ in demographic/economic status. The smallest population is 4,127 in 2016 (Tyrrell county) while two big counties are: Mecklenburg county (1,035,605 in 2016) and Wake county (1,007,631 in 2016). The population density is calculated by $\frac{Population}{Landarea(km^2)}$, where the minimum and maximum areas are respectively 446.701 km^2 and 2457.924 km^2. The average median age of counties is 40.08, and the median household income is 41,410 dollars.

For construction of a network W_n, we employ a concept of "neighbors" such that $w_{ij} = \frac{\tilde{w}_{ij}}{\sum_{k=1}^{n} \tilde{w}_{ik}}$ where $\tilde{w}_{ij} = 1$ if i and j are "neighbors"; $\tilde{w}_{ij} = 0$ otherwise. To define "neighbors", geographic distances among counties are considered. The kilometer-base geographic distance between two counties i and j (denoted by d_{ij}) is evaluated by the Haversine formula:

$$d_{ij} = 2r_E \arcsin\left(\sin^2\left(\frac{\varphi_j - \varphi_i}{2}\right) + \cos(\varphi_j)\cos(\varphi_i)\sin^2\left(\frac{\tau_j - \tau_i}{2}\right)\right) \quad (2.31)$$

where $r_E = 6356.752$ km denotes the Earth radius, φ_i and φ_j are latitudes, and τ_i and τ_j are longitudes in radians.[42] If $d_{ij} < d_c$ where d_c is a specified cutoff value, i and j are "neighbors". We consider four sets of model pairs (myopic model v.s forward-looking model) by choosing four different cutoff values, $d_c = 50$, 65, 80, and 95. On average, a county has 4.34 neighbors if $d_c = 50$; 7.34 neighbors if $d_c = 65$; 10.54 neighbors if $d_c = 80$; and 14.76 neighbors if $d_c = 95$.

[41] Because the zero proportion is small, so we do not build a Tobit model for this application.

[42] That is, county i's location is characterized by a pair (φ_i, τ_i).

This application studies the main structural parameters. λ_0, γ_0, and ρ_0 under two different assumptions for agents. i.e., myopic v.s forward-looking agents. Instead of directly estimating the time-discounting factor δ, we consider and compare two values of δ: (i) $\delta = 0$ (myopic agents) and (ii) $\delta = 0.9704$ (forward-looking agents). The value $\delta = 0.9704$ is set by $\frac{1}{1+\bar{\imath}_r}$ where $\bar{\imath}_r = 0.0305$ is the average annual 10-year Treasury Constant Maturity Rate from 2005 to 2016.[43] To achieve a stable process of a decision variable, we consider counties' public safety spending per capita as a dependent variable. Since a local government's public safety spending is based on its budget, the annual revenue (per capita) of a county is considered as an explanatory variable. Since the population size and residents' wealth level might affect the scale of criminal activities, a decision of a local government reflects those features. To control them, the population density and the median household income are included in a set of explanatory variables. We also include the median age of residents of a county. Lastly, Durbin regressors ($W_n X_{nt}$) of all explanatory variables are also considered so that they describe the externalities of explanatory variables affecting decisions. For estimation of the structural and nuisance parameters, we consider the joint estimation of the equations (2.29) and (2.30).[44]

The estimation results are summarized in four tables of Table 2.3: They are respectively for various neighboring systems with $d_c = 50$, 65, 80, and 95. For both $\delta = 0$ and 0.9704, and all cutoff values, county government's public safety spending (per

[43]In macroeconomic literature, δ is calibrated with targeting to the first moment of capital to output ratio (about 3) or is set to be a reciprocal of the gross long-run (risk-free) interest rate. They usually take a value from 0.95 to 0.99 if an annual data set is considered. We select the latter approach, which implies $\delta(1 + \bar{\imath}_r) = 1$. In a conventional intertemporal consumption-saving model, $\delta(1 + \bar{\imath}_r) = 1$ means completely smoothed consumption. For the detailed discussion, refer to Chapter 1.3 in Ljungqvist and Sargent (2012).

[44]Derivation and statistical properties (including asymptotic properties) of the joint QML method can be found in the supplement file.

Table 2.3: Model estimation

	Myopic	Forward-looking
Total revenue per capita	0.1008*** (0.0054)	0.1226*** (0.0066)
Population density	0.0002 (0.0003)	0.0002 (0.0003)
Median ages	0.0035 (0.0022)	0.003 (0.0022)
Median Household income	0.0011 (0.0011)	0.001 (0.0011)
Neighboring total revenue per capita	-0.0295*** (0.0096)	-0.0379*** (0.0117)
Neighboring population density	-0.0001 (0.0006)	0 (0.0005)
Neighboring median ages	0.0011 (0.0041)	0.0008 (0.004)
Neighboring median household income	-0.0018 (0.0021)	-0.0017 (0.0022)
λ	-0.0309 (0.043)	-0.0623 (0.0561)
γ	0.384*** (0.0252)	0.5099*** (0.069)
ρ	0.0582 (0.0515)	0.1154* (0.0662)
σ_ϵ^2	0.003*** (0.0001)	0.0051*** (0.0003)
Conditional log likelihood	-2713.5	-2713.3
AIC	4935	4934.6
BIC	5610.4	5610
No. of Obs	1200	1200
No. of neighbors	4.3400 (1.4229)	4.3400 (1.4229)
Cutoff distance	50	50

Note: The conditional log likelihood is the sample log likelihood for Y_{nt} given X_{nt}. AIC and BIC are the values of information criteria. Theoretical standard deviations are in parenthesis. Estimates that are significant at the 10 percent, 5 percent, and 1 percent levels are respectively marked by *, **, and ***.

	Myopic	Forward-looking
Total revenue per capita	0.1012*** (0.0053)	0.1226*** (0.0066)
Population density	0.0002 (0.0003)	0.0002 (0.0002)
Median ages	0.0032 (0.0022)	0.0027 (0.0021)
Median Household income	0.0011 (0.0011)	0.001 (0.0011)
Neighboring total revenue per capita	-0.0394*** (0.0129)	-0.053*** (0.0157)
Neighboring population density	-0.0001 (0.0006)	0 (0.0005)
Neighboring median ages	-0.001 (0.0054)	-0.001 (0.0053)
Neighboring median household income	-0.0027 (0.0027)	-0.0026 (0.0028)
λ	-0.0308 (0.0559)	-0.0321 (0.072)
γ	0.3796*** (0.0251)	0.5228*** (0.0656)
ρ	0.0747 (0.0657)	0.1486* (0.0833)
σ_ϵ^2	0.003*** (0.0001)	0.0051*** (0.0003)
Conditional log likelihood	-2712.9	-2712.5
AIC	4932.2	4931.3
BIC	5607.6	5606.7
No. of Obs	1,200	1,200
No. of neighbors	7.3400 (2.1937)	7.3400 (2.1937)
Cutoff distance	65	65

Note: The conditional log likelihood is the sample log likelihood for Y_{nt} given X_{nt}. AIC and BIC are the values of information criteria. Theoretical standard deviations are in parenthesis. Estimates that are significant at the 10 percent, 5 percent, and 1 percent levels are respectively marked by *, **, and ***.

	Myopic	Forward-looking
Total revenue per capita	0.1023*** (0.0054)	0.1239*** (0.0066)
Population density	0.0002 (0.0003)	0.0002 (0.0002)
Median ages	0.0032 (0.0022)	0.0028 (0.0021)
Median Household income	0.0011 (0.0011)	0.001 (0.0011)
Neighboring total revenue per capita	-0.052*** (0.0158)	-0.0667*** (0.0191)
Neighboring population density	-0.0003 (0.0007)	-0.0002 (0.0006)
Neighboring median ages	-0.0028 (0.0074)	-0.0031 (0.0072)
Neighboring median household income	-0.0041 (0.0034)	-0.0036 (0.0036)
λ	0.0142 (0.0657)	0.0058 (0.0845)
γ	0.3739*** (0.0251)	0.5081*** (0.065)
ρ	0.0705 (0.0784)	0.1726* (0.0984)
σ_ϵ^2	0.003*** (0.0001)	0.0051*** (0.0003)
Conditional log likelihood	-2712.9	-2712.5
AIC	4927.8	4927.1
BIC	5603.3	5602.5
No. of Obs	1,200	1,200
No. of neighbors	10.5400 (3.0465)	10.5400 (3.0465)
Cutoff distance	80	80

Note: The conditional log likelihood is the sample log likelihood for Y_{nt} given X_{nt}. AIC and BIC are the values of information criteria. Theoretical standard deviations are in parenthesis. Estimates that are significant at the 10 percent, 5 percent, and 1 percent levels are respectively marked by *, **, and ***.

	Myopic	Forward-looking
Total revenue per capita	0.1031*** (0.0054)	0.1237*** (0.0066)
Population density	0.0003 (0.0003)	0.0002 (0.0002)
Median ages	0.0033 (0.0022)	0.0028 (0.0021)
Median Household income	0.0012 (0.0011)	0.0011 (0.0011)
Neighboring total revenue per capita	-0.0673*** (0.0187)	-0.082*** (0.0226)
Neighboring population density	-0.0006 (0.0008)	-0.0004 (0.0008)
Neighboring median ages	-0.0044 (0.0088)	-0.0041 (0.0086)
Neighboring median household income	-0.0028 (0.0042)	-0.0024 (0.0044)
λ	0.0434 (0.0805)	0.027 (0.1049)
γ	0.3616*** (0.0252)	0.506*** (0.0661)
ρ	0.1607 (0.1002)	0.1696 (0.1255)
σ_ϵ^2	0.003*** (0.0001)	0.0051*** (0.0003)
Conditional log likelihood	-2713.1	-2713.2
AIC	4927.1	4927.4
BIC	5602.5	5602.8
No. of Obs	1,200	1,200
No. of neighbors	14.7600 (4.1709)	14.7600 (4.1709)
Cutoff distance	95	95

Note: The conditional log likelihood is the sample log likelihood for Y_{nt} given X_{nt}. AIC and BIC are the values of information criteria. Theoretical standard deviations are in parenthesis. Estimates that are significant at the 10 percent, 5 percent, and 1 percent levels are respectively marked by *, **, and ***.

capita) is persistent itself, the total revenue is significantly positive, but the neighboring total revenue is significantly negative. The current competition parameter λ_0 is negative for $d_c = 50$ and 65 while it is positive for $d_c = 80$ and 95. However, those are not significant. For the learning and/or diffusion parameter ρ_0, the sign is positive for all cases, but it is significant only for the forward-looking agent model (except $d_c = 95$) at the 10% significance level. Thus, for the forward-looking agent model, this result indicates that the learning and diffusion effects diminish when d_c characterizing "neighbors" becomes 95 kilometers. The population density, median age, median household income and their Durbin regressors do not have significant effects. To evaluate the model's performance, we consider three likelihood measures: sample conditional log-likelihood values[45], values of Akaike information criterion (AIC) and Bayesian information criterion (BIC). In choosing a spatial weight matrix, Chapter 2 in Lee (2008) suggests using the goodness-of-fit measures (e.g., adjusted R^2 or log-likelihood). Via Section 2.5, we provide evidence for using likelihood measures in selecting δ. Based on those likelihood measures, hence, the forward-looking agent model with cutoff value $d_c = 80$ is the best one among the 8 model specifications. For each cutoff value d_c, the forward-looking agent model is more favorable than the myopic model except $d_c = 95$. For both myopic and forward-looking models, $d_c = 80$ is selected in general as preferred.[46]

Here we provide economic interpretations based on the forward-looking agent model with $d_c = 80$. We can recover the cost function: $c\left(y_{it}, y_{i,t-1}\right) = 0.2541\left(y_{it} - y_{i,t-1}\right)^2 +$

[45]It means the log-likelihood function conditional on exogenous variables.

[46]However, AIC selects $d_c = 95$ in case of the myopic model.

$0.2459y_{it}^2$. The marginal direct effect of increasing previous own public safety spending (per capita) by one thousand dollars on the current one is 0.508 thousand dollars. The marginal direct effect of increasing previous neighbors' public safety spending (per capita) by one thousand dollars is $\rho_0 \sum_{j=1}^{n} w_{ij} = \rho_0 = 0.1726$ thousand dollars.[47] Consider the direct marginal effects of own and neighbor's revenues on the public safety spending. When the current revenue (per capita) of a county increases by one thousand dollars, it induces an increment of 0.124 thousand dollars directly on its public safety spending (per capita). On the other hand, the direct effect of neighbors' revenues (per capita) by increasing one thousand dollars will decrease the public safety spending (per capita) by 0.067 thousand dollars. It provides evidence of the negative externalities of revenues on the public safety spending.

Since our intertemporal SAR equation describes an equilibrium system, the cumulative marginal effects of an increase in the total revenue can be evaluated. The formula of the cumulative marginal effects from j's k^{th}-exogenous characteristic on i's decision is

$$\frac{\partial y_{it}}{\partial x_{jt,k}} = \left[R_n^{-1} \left(I_n + \delta \mathbf{D}_{n,k} A_{k,n}\right) \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n\right)\right]_{ij} \tag{2.32}$$

where $\mathbf{D}_{n,k} = \sum_{l=1}^{\infty} \delta^{l-1} D_{n,l} A_{k,n}^{l-1}$ for each $k = 1, \cdots, K$. Correspondingly, the cumulative own marginal effects are $[R_n^{-1} (I_n + \delta \mathbf{D}_{n,k} A_{k,n}) (\beta_{1,k,0} I_n + \beta_{2,k,0} W_n)]_{ii}$. On the other hand, the direct neighboring marginal effect is $\beta_{2,k,0} w_{ij}$ while the direct own marginal effect is $\beta_{1,k,0}$. Equation (2.32) says the cumulative marginal effects differ across spatial units and heterogeneity of these comes from the network W_n. To investigate the cumulative effect, we select two specific counties based on the number of neighbors. Based on $d_c = 80$, Iredell county has the largest number of neighbors

[47] For specific j's effect on i's decision, it will be $\rho_0 w_{ij} = \frac{\rho_0}{Number of i's neighbors}$.

Table 2.4: The direct and cumulative effect of increasing the total revenue (per capita) by one thousand dollars

		Iredell county	Dare county
Direct	Own effect	0.1239	0.1239
	Neighboring effect	-0.0039	-0.0222
Cumulative	Own effect	0.1046	0.1045
	Neighboring effect	-0.003	-0.0167
No. of neighbors		17	3

(17 neighbors) while Dare county has the smallest number of neighbors (3 neighbors). Figure 2.4 describes neighbors of the two counties.

Figure 2.4: Neighbors of the two counties (based on $d_c = 80$)

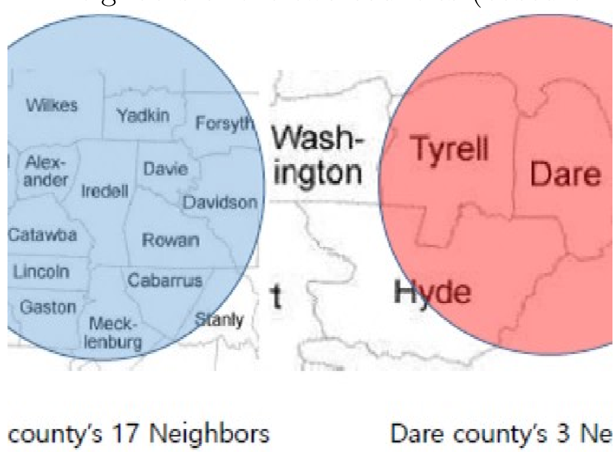


Table 2.4 shows direct own/neighboring effects and cumulative own/neighboring effects for the two counties. First, magnitudes of neighboring effects (both direct and cumulative) are bigger for the isolated county. Second, the negative direct neighboring

effects are smaller than the negative neighboring cumulative effects. For Dare county, that negative effect is weakened by 29.28% while 23.07% of the effect is alleviated for Iredell county in the equilibrium. Third, the positive direct effects are also weakened in the equilibrium. For Dare county, the positive own effect is alleviated by 15.66% and 15.58% of the positive effect is weakened for Iredell county. These results might be affected by a structure of W_n and structural parameters θ_0.[48]

A notable advantage of using dynamic models is doing impulse response analyses. The effect of changing j's t^{th}-period k^{th}-exogenous characteristic $x_{jt,k}$ on i's $(t+h)^{th}$-period economic activity $y_{i,t+h}$ $(h = 1, 2, \cdots)$ is characterized by the impulse response function:

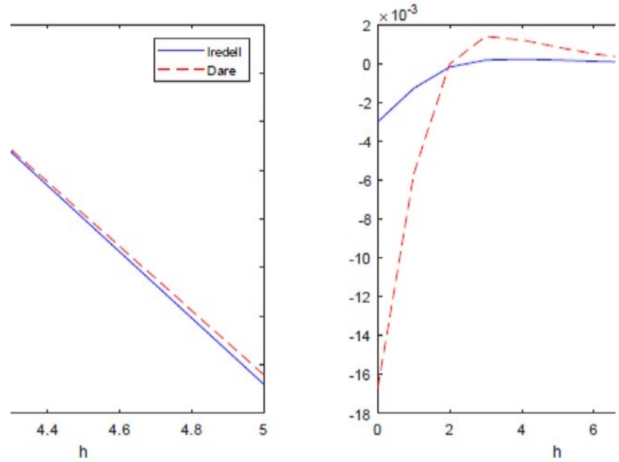
$$\frac{\partial \left[E_t\left(Y_{n,t+h}\right)\right]_i}{\partial x_{jt,k}} = \left[\sum_{g=0}^{h} A_n^{h-g} R_n^{-1} \left(I_n + \delta \mathbf{D}_{n,k} A_{k,n}\right) \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n\right) A_{k,n}^{g}\right]_{ij}. \tag{2.33}$$

Using formula (2.33), we plot the impulse response functions of own effects $\frac{\partial \left[E_t\left(Y_{n,t+h}\right)\right]_i}{\partial x_{it,k}}$ and neighboring effects $\frac{\partial \left[E_t\left(Y_{n,t+h}\right)\right]_i}{\partial x_{jt,k}}$ (j is a neighbor of i) for the two counties.

First, observe the impulse response functions of own effects. Note that Iredell county's own cumulative effect (impulse response function at $h = 0$) is slightly larger than that of Dare county (see Table 2.4). However, there is a crossover at $h = 4$. Since two impulse responses are so close in this case, we only plot the impulse response functions of the two counties between $h = 4$ and 5 to show the intersecting point. It means Dare county's own effects will be larger than that of Iredell county after $h = 4$. Second, we capture the overshooting effects for both counties. The negative neighboring effects are alleviated by $h = 2$. After $h = 3$, the neighboring effects become positive and they are diminishing when h increases. In case of Dare county,

[48] Additional comments for this issues can be found in the supplement file of Jeong and Lee (2018).

Figure 2.5: Impulse response functions: own effects (left) and neighboring effects (right)



that overshooting effect is more distinct relative to that of Iredell county. It seems that the negative neighboring effects diminish over time combined with other positive effects: self-reinforcing effects, positive diffusion effects, and positive own revenue effects. Since we consider a row-normalized W_n, nonzero elements in the row of W_n for Dare county are much larger than those of Iredell county. This fact may be a primary reason for distinct overshooting effects in case of Dare county.

Last, we want to deliver policy implications by conducting welfare analyses. We consider a situation that the North Carolina state government gives some amount of subsidy (per capita) to a county in 2016. So, the initial period is set to be 2016 in this analysis. Let Δ_x denote the amount of subsidy and $k = 1$ for the index of a county's total revenue. Then, we generate a new regressor $X_{nT,1}$ (denoted by $\ddot{X}_{nT,1}$)

$$\ddot{X}_{nT,1} = [x_{1,T,1}, \cdots, x_{j,T,1} + \Delta_x, \cdots, x_{n,T,1}]' \tag{2.34}$$

Table 2.5: Changes of social welfare if a countys total revenue (per capita) increases by one thousand dollars

	Case 1	Case 2	Case 3	Case 4
Welfare change $\hat{\Delta}_W$	-0.0013	0.0097	0.0121	0.0918

Note: We select four specific counties: (Case 1) Mecklenburg county (richest and the most populated county), (Case 2) Tyrrell county (poorest and the least populated county), (Case 3) Iredell county (the largest number of neighbors (17 neighbors)), and (Case 4) Dare county (the most isolated one (3 neighbors)).

describing a changed economic environment, where j denotes a subsidy recipient. Note that the realized pair $\{Y_{nT}, X_{nT,1}\}$ and the generated one $\left\{Y_{nT}, \ddot{X}_{nT,1}\right\}$ yield distinct dynamics, so they have different expected lifetime values as well as social welfare. Using the bias corrected QMLE ($\hat{\theta}^c_{ml,nT}$), we can compute a change of welfare

$$\hat{\Delta}_{\mathcal{W}} = \widehat{\mathcal{W}}^F \left(\left\{Y_{nT}, \ddot{X}_{nT,1}\right\}; \hat{\theta}^c_{ml,nT}\right) - \widehat{\mathcal{W}}^F \left(\{Y_{nT}, X_{nT,1}\}; \hat{\theta}^c_{ml,nT}\right) \tag{2.35}$$

where $\widehat{\mathcal{W}}^F (\{Y_{nT}, X_{nT,1}\}; \theta)$ stands for the welfare measure defined by the summation of counties' (expected) lifetime payoffs with the initial value $\{Y_{nT}, X_{nT,1}\}$ and parameter θ. $\widehat{\mathcal{W}}^F \left(\left\{Y_{nT}, \tilde{X}_{nT,1}\right\}; \hat{\theta}^c_{ml,nT}\right)$ captures social welfare when a county receives some subsidy while $\widehat{\mathcal{W}}^F \left(\{Y_{nT}, X_{nT,1}\}; \hat{\theta}^c_{ml,nT}\right)$ evaluates social welfare in a given realized economic environment. The difference between $\widehat{\mathcal{W}}^F \left(\left\{Y_{nT}, \tilde{X}_{nT,1}\right\}; \hat{\theta}^c_{ml,nT}\right)$ and $\widehat{\mathcal{W}}^F \left(\{Y_{nT}, X_{nT,1}\}; \hat{\theta}^c_{ml,nT}\right)$ will capture a welfare change corresponding to the change of policy.[49]

For convenience of analysis, we only select four specific counties: (Case 1) Mecklenburg county (richest and the most populated county), (Case 2) Tyrrell county

[49] The detailed derivation and specification can be found in the supplement file of Jeong and Lee (2018).

(poorest and the least populated county), (Case 3) Iredell county (has the largest number of neighbors (17 neighbors)), and (Case 4) Dare county (has the smallest number of neighbors (3 neighbors)). The amount of subsidy (per capita) from the state government is set to be one thousand dollars (i.e., $\Delta_x = \$1,000$). Table 2.5 reports $\hat{\Delta}_{\mathcal{W}}$'s for Cases 1 - 4. First, we observe that the number of neighbors affects social welfare more than population and/or level of revenues in our framework. When the state government increases Mecklenburg county's revenue (per capita) by $1,000, social welfare decreases by 0.0013 welfare measure. This negative welfare effect might come from the negative externalities of revenues on the public safety spending. Welfare increases for each of the other three cases. By comparing Cases 3 and 4, giving subsidy to the county whose number of neighbors is small increases social welfare more in the sense of public safety spending.

2.7 Conclusion

In this paper, we consider the specification and estimation of a spatial intertemporal competition model in a dynamic (differential) game setting. Agents are linked in a given spatial network. To characterize agent's payoff function, a linear-quadratic one is considered. By the MPE with a unique NE equation, we build an econometric model and consider model identification and estimation. In particular, we investigate the QML estimator. We obtain consistency and asymptotic normality of the QML estimator under some regularity conditions. Due to the presence of many nuisance parameters, bias correction of the QML estimator is needed. To fortify those results and investigate finite sample performance of the estimator, we conduct Monte Carlo simulations. From the simulations, the QML estimator and its bias-correction reveal reliable performance. In particular, for small T, the bias corrected QML estimator is recommended. For a misspecified conventional SDPD model, which ignores the intertemporal decision, significant empirical biases of estimates and low coverage probabilities are detected. Using the established model, we analyze strategic spillover effects of counties' public safety spending in North Carolina. We estimate structural parameters and compare the estimation results with those from the conventional SDPD model. First, our intertemporal SAR specification turns out to be more statistically favorable than the corresponding traditional SDPD model. Second, we find some evidence of persistency of public safety spending, positive learning and/or diffusion effects from previous neighbors' decisions, positive effects of own total revenue, and negative externalities from neighboring total revenues. An overshooting effect is captured for the case of negative neighboring revenue effect. In the welfare analysis, we observe giving subsidy to counties whose number of neighbors is small can be effective in the sense of public safety spending.

Chapter 3: Spatial dynamic models with intertemporal optimization II: Coevolution of economic activities and networks

3.1 Introduction

For many economic variables, cross-sectional or time series dependencies have been observed. For example, Case et al. (1993) find that a state government's expenditure is positively correlated with its neighbors' expenditures. In this paper, by focusing on interactions among rational agents (characterized by spatial units), we establish a new spatial dynamic panel data (SDPD) model. The purpose of our model is to explain both dependencies with time-evolving endogenous spatial networks by economic reasonings. That is, it accounts for the co-dynamics of (i) interrelated actions of rational agents and (ii) their network relationships. To construct a model with a rigorous economic foundation, two aspects of rationality are considered. Because agents may live multi-periods when we focus on a panel data set, their economic decisions do not arise once but for every period. Hence, we assume that agents are forward-looking instead of myopic. We formulate their interactions via a spatial network, which leads to a game played on a spatial network. The network game can characterize optimal actions by the Nash equilibrium (NE) concept and our estimation equation will specify their optimal actions. Since our target is a long panel data set showing multiple agents (n denotes the number of agents) with long decision-making periods, (i) strength of connections of a spatial network represented by an $n \times n$ matrix W_{nt} might be changing over time t[50], and (ii) it can be affected by economic actions. Corresponding research questions are verifying sources of endogenous spatial network evolution and relating this issue to network interactions. We formulate an estimation equation by considering both problems at the same time, namely, (i) network interactions, and (ii) a network formation. In consequence, our econometric model gives a tool to explain spatial/time dependencies among agents' actions as well as time-varying spatial networks affected by agents' actions.

As an economic foundation for our econometric model, first, we introduce a theoretical model that establishes a connection between agents' optimal actions and evolution of spatial networks by the forward-looking agent assumption.[51] A motivating example is studying interdependencies of local governments' expenditures. For example, an agent is a state government, and it selects an amount of health expenditure for each period by considering neighboring states' current and expected future health expenditures with their demographic characteristics. We formulate a network interaction model by internalizing agents' decisions on network links. In

[50]Empirical evidence for time-varying relationships among agents can be found in Tables 1 and 2 in Goldsmith-Pinkham and Imbens (2013). Using the Add Health survey, they report 534 students' friendship links at two points in time (Wave I and Wave II). In their Table 2, we observe that 1.17% of friendship links are changed.

[51]A recent study in the coevolution of agents' actions and networks is Han, Hsieh, and Ko (2019). There are two different features of our model. First, our key mechanism generating coevolution is the forward-looking agent assumption. The mechanism of Han, Hsieh, and Ko (2019) is the dynamic feature of latent variables affecting both network interactions and formation. Second, we focus on the case of spatial networks rather than friendship networks.

detail, agents' current actions (automatically) construct their future network links. For this, we assume that a spatial network W_{nt} is composed of nonnegative values exhibiting (relative) intensities of interactions and differentiable with respect to agents' actions. Fixed geographic locations of agents basically form a spatial network W_{nt}. However, the policy interdependence dynamically arises via W_{nt} driven by agents' economic similarities[52] in addition to geographic similarities. If spatial networks evolve with time-varying economic indicators, which are affected by optimal actions of forward-looking agents, this gives a recursive structure and we might face an issue on time-varying endogenous spatial networks.

To describe endogenously changing spatial networks, we establish a differential network game model. For identification of parameters about agents' preferences, we consider a parametric payoff function, which is a dynamic extension of Ballester et al. (2006). The designed payoff function contains a time-varying spatial network W_{nt}, and rivals' actions can affect an agent's payoff via W_{nt}. The theoretical model for our econometric specification is based on the payoff function with the forward-looking agent assumption. For each period, state variables consist of recent past actions and currently realized exogenous characteristics. We assume complete information up to the current period for agents' information to form expectations of future exogenous characteristics. Based on the payoff specification and the information set, a lifetime payoff is defined by a weighted sum of expected per period payoffs with a time-discounting factor. Each agent's current action can affect his/her future economic

[52]Economic indicators showing the economic status of agents construct economic similarities. In the case of regional expenditure decisions, a regional income level can be an economic indicator. As an example of economic similarities, hence, a reciprocal value of an absolute difference of two different regional income levels can be considered.

indicators forming future spatial network links. Compared to Jeong and Lee (2018)[53], even though the payoff function is still linear-quadratic (LQ), our specification has an additional nonlinear channel with that agents' current actions can affect their future payoffs via future spatial network links by evolution of economic similarities (indicators). In consequence, feedback effects can arise from that (i) agents' current actions help to form future spatial network links, and (ii) (expected) future spatial network links also influence their current actions due to the forward-looking agent assumption. Therefore, our model provides a connection between optimal actions of forward-looking agents and their time-evolving spatial networks.

Second, we provide a new econometric model for a spatial panel data set describing the theoretical specification. Since, in this paper, networks (W_{nt}) are allowed to evolve with actions, our model does not belong to a LQ dynamic programming. As a result, it is difficult to derive optimal actions in an explicit functional form and our econometric model belongs to a nonlinear SDPD model with endogenous networks. Instead of deriving the optimal action vector, we characterize optimal actions via Euler equations for an estimation purpose (Hansen and Singleton (1982)). Since Euler equations can show the relationship among past, current and expected future optimal actions, we can study the marginal effect of changing a current action on future payoffs through changing future spatial network links. By estimating some structural parameters in Euler equations, we can detect the existence of co-dynamics between agents' current actions and their future spatial network links. The derived Euler equations involve infinite expected future actions since an agent's current action

[53]Under an intertemporal optimization specification based on a linear-quadratic (LQ) pyoff function, but with an exogenous spatial network, they derive a linear optimal decision vector in state variables via a LQ programming setting.

nonlinearly affects his/her own and rivals' future period payoffs. For a practical use, we approximate some (expected) future components in an Euler equation by the LQ perturbation method[54] and obtain a tractable measure for estimation. With linear moment orthogonality conditions, the nonlinear two-stage least squares estimation (NL2S) method can be considered. To improve estimation efficiency, in addition to instrumental variables (IV) moments for orthogonality conditions, we also consider quadratic moment conditions, which capture spatial correlation. By those moment conditions, we study the GMM estimation method for our model.

With both large numbers of spatial units and time series periods, under the increasing domain asymptotic framework, consistency and the asymptotic distribution of the GMM estimator (GMME) are studied. As dependent variables of our model (agents' optimal actions) can be serially and spatially correlated, but might not be a linear function of disturbances, we employ the notion of spatial-time near epoch dependence (NED) in Jenish and Prucha (2012) to establish the law of large numbers (LLN) for the GMME. In order to derive the asymptotic distribution of the GMME, we utilize a central limit theorem (CLT) for a LQ form of martingale difference arrays based on the $\mathcal{C}$-stable convergence concept. In Kuersteiner and Prucha (2013, 2018), this notion is a joint convergence concept of main statistics and a $\mathcal{C}$-measurable random variable. Then, the asymptotic distribution of the GMME would be normal conditional on $\mathcal{C}$. An advantage of taking this concept is to analyze asymptotic properties of main statistics conditional on unspecified exogenous components stemming from spatial network formation. From the derived asymptotic distribution, we observe the existence of asymptotic biases due to incidental parameters due to individual and

[54]Since the large number of agents (n) is targeted, we need to consider a computation method which is free to a curse of dimensionality.

time effects and, so we suggest a bias correction method. For testing whether spatial networks evolve exogenously or not, the Wald test can be applied. The Wald test statistic follows the asymptotic unconditional chi-square distribution.

Lastly, we apply our econometric model to empirically investigate policy interdependence of U.S. states' health expenditures. The state's health expenditure yields its human capital accumulation potentially improving regional future economic status. To implement our model, we should know (or prespecify) a formation function of spatial networks to evaluate the marginal effect of changing a current action on future spatial network links. For practical reasons, hence, we suggest an estimation method to specify formation of spatial networks by using flow variables (e.g., state-to-state migration flows).[55] Coefficients of the spatial (for geographic and economic distances) network formation model mean elasticities of intensities of interactions. Under this specification, 1% change in states' economic distances yields -0.1313% change in intensities of interactions. A positive spatial spillover effect is captured in states' health expenditures, but its estimated coefficient is not significant. The Wald statistic shows that the coevolution of states' health expenditures and their spatial networks is not significant. We observe that health expenditures of U.S. states are persistent. A positive effect of federal grants on the state government's marginal payoff is detected.

3.2 Model specification

There are two main components in our analysis. The first is a spatial econometric model specification.[56] In this section, we have two issues to rigorously formulate a

[55]Such formulations are many as pointed out in Qu and Lee (2018), which considers flows with multiple fixed effects for estimation.

[56]Another aspect is statistical theories for spatial econometric models (e.g., large sample theory) due to the dependencies generated by spatial networks. This issue will be discussed in Section 3.4.

spatial econometric model. The first issue is to reveal how spatial networks shape forward-looking agents' actions. The second issue is how agents' actions affects spatial network links in the aspect of network interaction models.

At first, we introduce a spatial dynamic panel data (SDPD) model and a corresponding payoff function specification justifying it. Second, the agent's lifetime problem will be designed based on the payoff function. That is, we build a theoretical foundation of an estimation equation based on the payoff function. Third, we will characterize optimum actions by Euler equations to construct the estimation equation.

3.2.1 Related literature and payoff specification

Suppose we have a set of panel data $\{Y_{nt}, X_{nt}\}$ where $Y_{nt} = (y_{1t}, \cdots, y_{nt})'$ denotes an $n \times 1$ vector of (continuous type) dependent variables at time t and $X_{nt} = (x_{1t}, \cdots, x_{nt})'$ with $x_{it} = (x_{it,1}, \cdots, x_{it,K})'$ stands for an $n \times K$ matrix of explanatory variables. Using a panel data set, we may capture dynamics of individual units' actions (or outcomes). As there are interactions among individual units and these interactions can be specified by spatial networks, SDPD models formulate dependence in individuals' actions (or outcomes) across individuals and over time periods.[57] Under this framework, each unit i is a spatial unit having its innate and fixed geographic location, e.g., a local government. In the conventional regional science literature, spatial network links are formed by spatial units' physical distances, which indicate that a

[57]Examples are Kapoor et al. (2007), Baltagi et al. (2007), Yu et al. (2008), Lee and Yu (2010, 2012, 2014), and Qu et al. (2017).

spatial weights matrix is fixed over time. However, if intensity of network interactions is affected by economic consequences, spatial networks might vary over time.[58]

If there are n spatial units, an $n \times n$ matrix W_{nt} captures the t^{th}-period socio-economic relationships among units. Each non-diagonal component of W_{nt} shows a (socio-economic) relationship between two spatial units at time t. In the conventional spatial econometrics literature, one assumes all entries in W_{nt} are nonnegative. Elements of W_{nt} only show relative magnitudes of interactions. By assuming all diagonal entries in W_{nt} (for any t) to be zero, we exclude the self-influence. A specification of SDPD models is[59]

$$Y_{nt} = \lambda_0 W_{nt} Y_{nt} + \gamma_0 Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{e}_{nt}, t = 1, \cdots, T \tag{3.1}$$

where $(\lambda_0, \gamma_0, \beta_0')'$ denotes a set of parameters, and $\mathbf{e}_{nt}$ is an n-dimensional vector of unobservables.[60] The time-varying W_{nt} can potentially come from time-evolving economic indicators, which might be affected by individual units' actions (or outcomes). Let $Z_{nt} = (Z_{nt,1}, \cdots, Z_{nt,P})$ (where $Z_{nt,p} = (z_{1t,p}, \cdots, z_{nt,p})'$ for $p = 1, \cdots, P$ and $z_{it} = (z_{it,1}, \cdots, z_{it,P})$ for $i = 1, \cdots, n$) be an $n \times P$ matrix of economic indicators at time t. We allow that some strictly exogenous indicators in Z_{nt} can be components of forming W_{nt}. All indicators in Z_{nt} are assumed to be continuous type variables for

[58]Examples of employing economic variables in the empirical regional network formation are plenty, e.g., Case et al. (1993), and Figlio et al. (1999).

[59]For SDPD models and their statistical properties, refer to Yu et al. (2008), and Lee and Yu (2010, 2012, 2014). Lee and Yu (2015) provide a review of this model structure. The diffusion effect $\rho_0 W_{n,t-1} Y_{n,t-1}$ might be included in equation (3.1). In this paper, we consider the case of $\rho_0 = 0$ for model's simplicity.

[60]$\mathbf{e}_{nt}$ can contain individual/time fixed/random effects and idiosyncratic disturbances.

technical simplification.[61] For the formation of Z_{nt}, we consider the linear specification based on Han and Lee (2016):[62]

$$Z_{nt} = Y_{n,t-1}\psi_0 + V_{nt} \tag{3.2}$$

where $\psi_0 = (\psi_{1,0}, \cdots, \psi_{P,0})$ is an $1 \times P$ vector of parameters showing dependencies upon the past actions $Y_{n,t-1}$, and V_{nt} denotes an $n \times P$ matrix of exogenous components, which may consist of observed as well as unobserved variables.[63] Specification (3.2) means that W_{nt} is generated by Z_{nt}, which is driven by $Y_{n,t-1}$.[64] If $\psi_{p,0} = 0$ for some p, $Z_{nt,p}$ would be a strictly exogenous indicator forming W_{nt}. To illustrate ramifications from $Y_{n,t-1}$, we represent $W_{nt}\left[Y_{n,t-1}\right]$ as the t^{th}-period spatial network and $w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right]$, $w_{t,i.}\left[Y_{n,t-1}\right]$ and $w_{t,.i}\left[Y_{n,t-1}\right]$ denote, respectively, the $(i,j)^{th}$-element, i^{th}-row, and i^{th}-column of $W_{nt}\left[Y_{n,t-1}\right]$.[65]

A properly modified equation (3.1) can be justified by a myopic dynamic game with n agents who have the LQ payoff: the agent i's t^{th}-period payoff (denoted by u_{it} later on) is

$$u_i(y_{it}, Y_{-i,t}, Y_{n,t-1}, \eta_{it}) = \left(\eta_{it} + \lambda_0 w_{t,i.}\left[Y_{n,t-1}\right] Y_{nt}\right) y_{it} - c(y_{it}, y_{i,t-1}) \tag{3.3}$$

[61] In our framework, the network link $w_{t+1,ij}\left[y_{it}, y_{jt}\right]$ obtains differentiability by considering a continuous type economic indicators $z_{i,t+1}$ and $z_{j,t+1}$. However, some indicators (potentially constructing agents' social distances) might be dichotomous. For example, we can consider the discrete political metric: $d_p(i,j) = 0$ if i and j support the same political party; $d_p(i,j) = 1$ otherwise. We leave this issue for future research.

[62] Nowadays, we see some empirical applications which consider flow variables $z_{ij,t}$ affecting formation of W_{nt} (Qu and Lee (2018)). For example, $z_{ij,t}$ can be a linear function of y_{it}, y_{jt}, and other variables. We leave this issue for future works.

[63] We focus on the relationship between Z_{nt} and $Y_{n,t-1}$ and do not have a detailed specification on V_{nt}. The V_{nt} can contain agents' observable/unobservable exogenous characteristics (both time-invariant and variant ones) as well as common economic shocks (time factors). In the estimation part, we introduce how to relate $\{V_{nt}\}$ to the GMM estimator's asymptotic properties.

[64] Note that $z_{i,t+1}$ is only affected by own t^{th}-period action y_{it} (no spatial interaction effects from neighbors' activities y_{jt}'s).

[65] The argument $Y_{n,t-1}$ in those functions emphasizes its important role in the network formation.

where $Y_{-i,t} = (y_{1t}, \cdots, y_{i-1,t}, y_{i+1,t}, \cdots, y_{nt})'$ and $c(y_{it}, y_{i,t-1}) = \frac{\gamma_0}{2}(y_{it} - y_{i,t-1})^2 + \frac{1-\gamma_0}{2} y_{it}^2$ with $0 < \gamma_0 < 1$. At time t, agent i selects his/her action y_{it} to maximize u_{it}. This is a myopic dynamic extension of the LQ payoff discussed by Ballester et al. (2006), and Calvo-Armengol et al. (2009), which concern about static models.[66] In the LQ network game, equation (3.1) shows linear best replies. The η_{it} presents i's t^{th}-period exogenous heterogeneity and, for all n agents, its vector version is denoted by $\eta_{nt} = (\eta_{1t}, \cdots, \eta_{nt})'$ at time t.[67] The η_{it} contains time invariant (η_i^{iv}) and time variant (η_{it}^{v}) individual characteristics affecting decision-making. They are public to all agents (if they belong to the information set) but some of them might not be observable by econometricians. The parameter λ_0 describes strategic interactions among agents' current economic activities. $c(\cdot, \cdot)$ denotes a cost function in agents' actions and consists of two components: (i) a dynamic adjustment cost and (ii) an agent's cost of selecting activity level y_{it}. The coefficient γ_0 captures a relative weight for the adjustment cost, so high γ_0 yields persistency of agent's action. In consequence, equation (3.1) would characterize NE activities if agents were assumed to completely discount future payoffs, i.e., myopic behavior Focusing on studying regional policies (see Section 3.5), payoff (3.3) describes a local government's objective function (refer to Brueckner (2003) and Revelli (2005)).

[66]Payoff (3.3) shows the local-aggregate model (i.e., $y_{it} \sum_{j=1}^{n} w_{t,ij}[y_{i,t-1}, y_{j,t-1}] y_{jt}$) Recently, Ushchev and Zenou (2018) introduce a local average model and its interpretations as a theoretical foundation of the SAR model. The local average means a "social norm", and this specification imposes a penalty when an agent takes a large deviation from the social norm. In a static game setting, the two theoretical specifications yield the same econometric model specification. In our forward-looking agent framework, however, those might generate different econometric models. In this paper, we take a notion of the local aggregate model since there is a limitation in interpreting the parameter λ_0 ($\lambda_0 > 0$) under the notion of the local average model.

[67]In particular, equation (3.1) comes from supposing $\eta_{nt} = X_{nt}\beta_0 + \mathbf{e}_{nt}$ for each t.

3.2.2 Agent's intertemporal choice problem

In this subsection, we set up the forward-looking agent's lifetime problem based on periodic payoff (3.3). We introduce a time-discounting factor $\delta \in [0,1)$ to make different weights on future economic choices to distinguish them from the current one. An agent's lifetime payoff (3.6) below is defined by the discounted summation (by δ) of per period payoffs (3.3). There are four different features compared to a conventional macroeconomic dynamic model: (i) the existence of network interactions, (ii) network heterogeneities stemming from spatial location's heterogeneities, (iii) linear (via adjustment costs) and nonlinear (via future network links) effects of the current action on (expected) future marginal payoffs[68], and (iv) a large number of spatial units (large n).

With going beyond a forward-looking spatial interaction model, an agent knows his/her current action (y_{it}) will influence the next future network links via ($z_{i,t+1}$) by equation (3.2). As there are uncertainty in future exogenous characteristics ($\eta_{n,t+1}, \eta_{n,t+2}, \cdots$), each agent would form expectations on future events based on their currently available information. To specify expectation for uncertainty, let $\mathcal{B}_{it}$ be the t^{th}-period information set of agent i's perceivable events. Below are definitions of an agent's intertemporal choice problem (ICP) and relevant concepts taking into account on interactions with other agents. $\sigma(\cdot)$ denotes the sigma-field generated by arguments inside. The superscript "$*$" denotes agents' optimal actions.

[68]This payoff belongs non-separable preferences. Among them, our model specification is similar to habit formation models (e.g., Fuhrer (2000)). In macroeconomic dynamic models, a payoff function is usually specified by a time-separable one (e.g., Hansen and Singleton (1982)).

Definition 3.2.1 (Intertemporal choice problem (ICP)) *(i) (Information set) For each i at time t, $\mathcal{B}_{it}$ is specified by*

$$\mathcal{B}_{it} = \sigma\left(\{y_{js}\}_{j=1}^{n}|_{s=-\infty}^{t-1}, \left\{\eta_j^{iv}\right\}_{j=1}^{n}, \left\{\eta_{js}^{v}\right\}_{j=1}^{n}|_{s=-\infty}^{t}\right). \tag{3.4}$$

(ii) (Process of η_{nt}^v) $\eta_{nt}^v = (\eta_{1t}^v, \cdots, \eta_{nt}^v)'$ follows a linear Markov process:

$$\eta_{nt}^v = \rho_{\eta,0}\eta_{n,t-1}^v + \xi_{nt} \tag{3.5}$$

where $|\rho_{\eta,0}| < 1$, and $\xi_{nt} \sim i.i.d. N\left(\mathbf{0}_{n\times 1}, \Omega_\xi\right)$ with $\Omega_\xi > 0$.[69]

(iii) Given $(Y_{n,t-1}, \eta_{nt})$, agent i $(i = 1, \cdots, n)$ maximizes his/her expected lifetime payoff by selecting y_{it} for each t: that is,

$$\{y_{is}^*\}_{s=t}^{\infty} = \arg\max_{\{y_{is}\}_{s=t}^{\infty}} \left\{ \begin{array}{c} u_i(y_{it}, Y_{-i,t}, Y_{n,t-1}, \eta_{it}) \\ +\sum_{s=1}^{\infty}\delta^s E\left(u_i(y_{i,t+s}, Y_{-i,t+s}, Y_{n,t+s-1}, \eta_{i,t+s})|\mathcal{B}_{it}\right) \end{array} \right\}. \tag{3.6}$$

Definition 3.2.1 means that $\mathcal{B}_{it}$ contains all previous actions and exogenous heterogeneities up to t. This setting means complete information up to the current period t. In Definition 3.2.1 (ii), the linearity assumption with parameter $\rho_{\eta,0}$ in η_{nt}^v is for simplicity.[70] For agents' innate (time-invariant) characteristics $\eta_n^{iv} = (\eta_1^{iv}, \cdots, \eta_n^{iv})'$, they are known to all agents. Key parts of Definition 3.2.1 (ii) are that (a) η_{nt}^v only depends on $\eta_{n,t-1}^v$ and ξ_{nt} (not other part of history) and (b) ξ_{nt} and $\eta_{n,t-1}^v$ are independent. Since (a) uncertainty only arises by future exogenous heterogeneities $\eta_{n,t+1}^v$ and (b) η_{nt}^v evolves by equation (3.5), $E\left(\eta_{n,t+1}^v|\mathcal{B}_{it}\right) = E\left(\eta_{n,t+1}^v|\eta_{nt}^v\right) = \rho_{\eta,0}\eta_{nt}^v$ for all i and t. Given $(Y_{n,t-1}, \eta_{nt})$, all agents expect that future actions will be realized as

[69] Since the agent's lifetime value would not be linear-quadratic in state variables, we assume normality of ξ_{nt} to have a linear conditional expectation $E\left(\eta_{n,t+1}^v|\mathcal{B}_{it}\right)$. This assumption will be relaxed in the estimation part. For example, the likelihood function can be formed for estimation but would be quasi-likelihood if ξ_{nt} were not really normally distributed.

[70] Or, we can generalize specification (3.5) to a case of finite numbers of parameters governing the process of η_{nt}^v.

NEs, in that all agents form same expectations on future actions. Hence, we have $E(\cdot|\mathcal{B}_{it}) = E(\cdot|Y_{n,t-1}, \eta_{nt})$ for all i and t, where $(Y_{n,t-1}, \eta_{nt})$ represents the initial condition at time t. As agents' rational expectations can be characterized by the mathematical expectation, we define the conditional expectation operator at time t by $E_t(\cdot) = E(\cdot|Y_{n,t-1}, \eta_{nt})$ for simplicity.

By Definition 3.2.1 (iii), we can specify the agent i's lifetime value $V_i^*(\cdot)$: given $(Y_{n,t-1}, \eta_{nt})$

$$V_i^*(Y_{n,t-1}, \eta_{nt}) = u_i(Y_{nt}^*, Y_{n,t-1}, \eta_{it}) + \sum_{s=1}^{\infty} \delta^s E_t\left(u_i(Y_{n,t+s}^*, Y_{n,t+s-1}^*, \eta_{i,t+s})\right) \qquad (3.7)$$

where $Y_{nt}^* = (y_{1t}^*, \cdots y_{nt}^*)'$ at time t. By observing equation (3.7), agent i at time t chooses y_{it} with considering (a) the adjustment cost $\frac{\gamma_0}{2}(y_{i,t+1} - y_{it})^2$ and (b) the future network links $w_{t+1,ij}[y_{it}, y_{jt}]$, which are included in the $(t+1)^{th}$-period payoff. Hence, there exist "feedback effects" due to the forward-looking assumption in decision makers: the optimal actions Y_{nt}^* of all agents influence the future network by equation (3.2) and in turn are affected by the (expected) future network $W_{n,t+1}[Y_{nt}^*]$. If an agent completely discounts future payoffs, feedback effects would not arise and the resulted model would be the myopic one. That is, the relationship between Y_{nt}^* and $W_{n,t+1}[Y_{nt}^*]$ will be simultaneous if agents are forward-looking while evolution of $W_{n,t+1}[Y_{nt}^*]$ becomes adaptive if agents are myopic. Another distinguished feature of Definition 3.2.1 (iii) is that the agent's decision-making problem will not be a LQ programming (his/her payoff function is not LQ in state variables $(Y_{n,t-1}, \eta_{nt})$) unless $\psi_0 = 0$ in equation (3.2), because his/her action would nonlinearly influence his/her own and rivals' future marginal payoffs via future network links.[71] It implies that

[71] When actions take limited dependent variables, we might have other examples of a nonlinear equation of exogenous variables under the linear-quadratic payoff specification. Examples are Xu and Lee (2015, 2016).

agents' optimal actions which will be a function of $(Y_{n,t-1}, \eta_{nt})$ might not be a linear one. Therefore, for the current model, it is difficult to directly derive an optimal decision rule like that in Jeong and Lee (2018).[72] This difficulty motivates us to specify agents' optimal actions based on a set of first-order conditions of (3.7), i.e., Euler equations.

To specify Y_{nt}^* using the Euler equations, we introduce a set of technical assumptions for the payoff function and network links in addition to Z_{nt}, taking into account that Y_{nt} influence the future network links. First of all, the Euler equations should be sufficient to characterize the optimum of the ICP (3.7). For this, consider the potential nonlinear part $w_{t+1,ij}[y_{it}, y_{jt}]$ $(j \neq i)$ of $u_{i,t+1}$. First, $w_{t+1,ij}[y_{it}, y_{jt}]$ $(j \neq i)$ are required to be differentiable with respect to y_{it}. There are additional components affecting $W_{n,t+1}[Y_{nt}]$. We assume that each agent's physical location is given and does not change over time. For relevant components of fixed physical locations, let d_{ij} be a distance between i and j and let d_c denote a finite threshold distance. We assume that when $d_{ij} \leq d_c$, agents i and j make a network link, but on the other hand, if agents i and j are sufficiently far (i.e., $d_{ij} > d_c$), they would not make a link. Then, d_c determines the maximum number of neighbors of agent i, which will be denoted as

[72] To see this feature, consider a simple two-period model. Given $(Y_{n,t-1}, \eta_{nt})$, agent i $(i = 1, \cdots, n)$ maximizes his/her discounted lifetime payoffs: at $t = 1$, $u_{i1} + \delta E_1(u_{i2})$; and at $t = 2$, u_{i2}, by sequentially selecting y_{it} for $t = 1, 2$. Since there were no additional future periods after $t = 2$, the NE activity vector at time $t = 2$ is given by the conventional SAR equation $Y_{n2}^* = \lambda_0 W_{n2}[Y_{n1}] Y_{n2}^* + \gamma_0 Y_{n1} + \eta_{n2}$. Since agents' choice problem at $t = 2$ is considered as a subgame, agent i's SPNE activity at $t = 1$ is derived by maximizing $u_{i1} + \delta E_1(u_{i2}^*)$ where $u_{i2}^* = \frac{1}{2}(y_{i2}^*)^2 - \frac{\gamma_0}{2} y_{i1}^2$ denotes the i's second-period payoff evaluated at Y_{n2}^*. The first-order conditions are

$$0 = \eta_{i1} + \gamma_0 y_{i0} + \lambda_0 w_{1,i.}[Y_{n0}] Y_{n1}^* - (1 + \delta\gamma_0) y_{i1}^* + \delta E_1 \left(\frac{\partial y_{i,2}^*}{\partial y_{i,1}} y_{i,2}^* \right) |_{Y_{n1}^*}, fori = 1, \cdots, n. \quad (3.8)$$

The SPNE vector at $t = 1$ (denoted by Y_{n1}^*) satisfies equations (3.8). Even though we consider the linear-quadratic payoff function, deriving the explicit form of Y_{n1}^* is a challenging issue since it is a highly nonlinear function of (Y_{n0}, η_{n1}).

$d_c(i)$. If d_c is large, the number of i's neighbors $d_c(i)$ might be large but that number remains as n tends to infinity.While there might be some different specifications on W_{nt}, this specification seems the most popular one for sparse spatial networks.

Contrast to conventional friendship network formation models with observable dichotomous links, our spatial network links $w_{t,ij}$'s can have (relative) intensities of interactions. It implies that they take nonnegative real numbers. As agents know that their current actions (y_{it}) can affect strength of future network links ($w_{t+1,ij}[y_{it}, y_{jt}]$ and $w_{t+1,ji}[y_{jt}, y_{it}]$ for $j = 1, \cdots, n$; but $j \neq i$), they optimize their actions by taking into account future strength of network links. That is, we formulate the network interaction model by internalizing individual decisions on network links.

Here is a formal assumption in order to have a tractable network formation model.

Assumption 3.2.1 *At each time t, we assume $w_{t,ij}[y_{it}, y_{jt}] = h(z_{i,t+1}, z_{j,t+1}) \cdot h_d(d_{ij}, d_c)$. For each pair (i,j), $h(z_{i,t+1}, z_{j,t+1})$ is infinitely differentiable a.e*[73]*. The first-order derivatives of $h(\cdot)$, $\frac{\partial h(z_{i,t+1}, z_{j,t+1})}{\partial z_{i,t+1}}$ and $\frac{\partial h(z_{i,t+1}, z_{j,t+1})}{\partial z_{j,t+1}}$, are bounded.*[74]

First, we need to have a smoothness condition for the agent's payoff function since deriving the Euler equations relies on the envelope theorem. By Assumption 3.2.1, the agent' payoff $u_i(Y_{nt}, Y_{n,t-1}, \eta_{it})$ is infinitely differentiable a.e.[75] Then, we can represent

[73]In our model setting, a functional form of $h(\cdot)$ is time-invariant since we seek to a stable economic environment. For $h(z_{i,t+1}, z_{j,t+1})$, e.g., we might take the distance-based specification $h(z_{i,t+1}, z_{j,t+1}) = |z_{i,t+1} - z_{j,t+1}|^{-\alpha_e}$ where $\alpha_e > 0$, which is infinitely differentiable a.e.

[74]Assumption 3.2.1 implies that the network link $w_{t,ij}$ is only affected by y_{it} and y_{jt}. However, this assumption can be generalized to a row-normalized W_{nt}. For example, $w_{t,ij}[Y_{nt}] = \frac{h(z_{i,t+1}, z_{j,t+1}) \cdot h_d(d_{ij}, d_c)}{\sum_{d_{ik} \leq d_c} h(z_{i,t+1}, z_{k,t+1}) \cdot h_d(d_{ik}, d_c)}$. Then, the network link $w_{t,ij}[Y_{nt}]$ can be affected by y_{it}, y_{jt} as well as y_{kt} such that $k \neq i$, $k \neq j$ with $d_{ik} \leq d_c$.

[75]In detail, this requirement is a key for obtaining differentiable agents' lifetime values and optimal decision functions (with respect to $(Y_{n,t-1}, \eta_{nt})$). Also, we can check the second-order condition to obtain the sufficiency of considering the Euler equations for optimality. For this issue, refer to the supplement file. Related results can be found in Theorem 4.11 in Stokey et al. (1989) and Santos (1991).

the agent's optimal action (y_{it}^*) as the solution to the first order condition. Also, the agents' optimal lifetime values and actions are continuously differentiable functions in $(Y_{n,t-1}, \eta_{nt})$. Second, differentiability of $w_{t+1,ij}[y_{it}, y_{jt}]$ gives a formulation of the marginal effects of y_{it} and y_{jt} on $w_{t+1,ij}[y_{it}, y_{jt}]$ in the $(t+1)^{th}$-period payoffs, e.g., $\frac{\partial w_{t+1,ij}[y_{it},y_{jt}]}{\partial y_{it}} = \sum_{p=1}^{P} \psi_{p,0} \frac{\partial h(z_{i,t+1},z_{j,t+1})}{\partial z_{i,t+1,p}} \cdot h_d(d_{ij}, d_c)$ where a bounded function $h_d(\cdot)$ plays a role in controlling the intensity of interactions stemming from d_{ij} (or d_c).[76] For each $p = 1, \cdots, P$, (i) $\frac{\partial z_{i,t+1,p}}{\partial y_{it}} = \psi_{p,0}$ is the effect of changing y_{it} on its p^{th} future economic indicator ($z_{i,t+1,p}$) and (ii) $\frac{\partial h(z_{i,t+1},z_{j,t+1})}{\partial z_{i,t+1,p}} \cdot h_d(d_{ij}, d_c)$ represents the changed network link via $z_{i,t+1,p}$. To implement our model, a functional form of $h(z_{i,t+1}, z_{j,t+1})$ is required to be known for evaluation of $\frac{\partial h(z_{i,t+1},z_{j,t+1})}{\partial z_{i,t+1,p}}$. Third, in order to obtain a tractable measure for estimation, we will employ a perturbation method to approximate $V_i(\cdot)$ using $u_i(Y_{nt}, Y_{n,t-1}, \eta_{it})$ around the population averages.

The following is an illustrative example.

Example 3.2.2 *Consider Han and Lee's (2016) Medicaid related spending competiton model. Then, y_{it} is the state i's t^{th}-period Medicaid related spending. In their specification, z_{it} is the state i's t^{th}-period income per capita, and it is significantly positively affected by $y_{i,t-1}$ ($\psi_0 > 0$). Simply, assume $w_{t+1,ij}[y_{it}, y_{jt}] = \frac{1}{|z_{i,t+1}-z_{j,t+1}|}$ implying $\frac{\partial h(z_{i,t+1},z_{j,t+1})}{\partial z_{i,t+1,p}} = \frac{-sgn(z_{i,t+1}-z_{j,t+1})}{|z_{i,t+1}-z_{j,t+1}|^2}$ Then, the effect of changing y_{it} on $w_{t+1,ij}[y_{it}, y_{jt}]$ is governed by $-sgn(z_{i,t+1} - z_{j,t+1})$ since $|z_{i,t+1} - z_{j,t+1}|^2 > 0$ a.e. First, consider the case that $z_{i,t+1} > z_{j,t+1}$. When y_{it} marginally increases, $z_{i,t+1}$ increases and $w_{t+1,ij}[y_{it}, y_{jt}]$ decreases. It means weakened intensity of interactions at time $t+1$ since the economic distance between i and j becomes far by increasing y_{it}. Second,*

[76]e.g., $h_d(d_{ij}, d_c)$ can be specified by $\frac{1}{d_{ij}}\mathbf{1}\{d_{ij} \leq d_c\} + \frac{1}{c_d}\mathbf{1}\{0 \leq d_{ij} \leq c_d\}$ for some small c_d such that $0 < c_d < d_c$.

consider the case that $z_{i,t+1} < z_{j,t+1}$. *By marginally increasing* y_{it}, $z_{i,t+1}$ *increases, which it yields that* $w_{t+1,ij}[y_{it}, y_{jt}]$ *increases. Increasing* y_{it} *makes that* i *and* j *be economically close, so intensity of interactions becomes strengthened.*

3.2.3 Characterizing the NE actions by the Euler equations

The purpose of this subsection is characterizing agents' NE actions to formulate an equation for estimation. We consider a stable economic environment based on the infinite horizon problem. Then, the agent's decision problem will be the same at each period conditional on its immediate information generated by $(Y_{n,t-1}, \eta_{nt})$. As a result, the ICP (3.6) can be represented by a functional equation (FE). A corresponding equilibrium concept is a Markov perfect equilibrium (MPE).[77] Here is the definition of MPE.

Definition 3.2.3 (Markov perfect equilibrium) *A MPE is a set of value functions* $V_i(\cdot)$ $(i = 1, \cdots, n)$ *and a set of policy functions* $f_i(\cdot)$ $(i = 1, \cdots, n)$ *such that*

(i) (Markov strategy) $y_{it}^* = f_i(Y_{n,t-1}, \eta_{nt})$ *(let* $Y_{nt}^* = f(Y_{n,t-1}, \eta_{nt})$*),*

(ii) given $f_1, \cdots, f_{i-1}, f_{i+1}, \cdots, f_n$, V_i *satisfies the Bellman equation*

$$V_i(Y_{n,t-1}, \eta_{nt}) = \max_{y_{it}} \left\{ u_i(y_{it}, Y_{-i,t}^*, Y_{n,t-1}, \eta_{it}) + \delta E_t \left(V_i(y_{it}, Y_{-i,t}^*, \eta_{n,t+1}) \right) \right\} \tag{3.9}$$

where $Y_{-i,t}^* = \left(y_{1t}^*, \cdots, y_{i-1,t}^*, y_{i+1,t}^*, \cdots, y_{nt}^* \right)'$, *and*

(iii) the policy function $f_i(\cdot)$ *attains the right side of the Bellman equation (3.9).*

[77] The MPE is a refined version of a subgame perfect Nash equilibrium (SPNE). It characterizes the equilibrium strategies of all agents as best responses to one another. Further, each agent's optimal strategy only depends on the state variables $(Y_{n,t-1}, \eta_{nt})$ and does not rely on other parts of histories.

By Definition 3.2.3 (i), the optimal action y_{it}^* is a function of the previous actions of all agents $(Y_{n,t-1})$ and the currently realized exogenous characteristics (η_{nt}). Definition 3.2.3 (iii) is the principle of optimality, which says the equivalence between the solutions of ICP (3.6) and FE (3.9). That is, $V_i(\cdot) = V_i^*(\cdot)$ for all $i = 1, \cdots, n$. In consequence, the optimal activities y_{it}^*'s and the state variables $(Y_{n,t-1}, \eta_{nt})$ have a time-invariant relationship $f(\cdot)$.[78] Since each network link $w_{t,ij}[y_{i,t-1}, y_{j,t-1}]$ is not a LQ function of $y_{i,t-1}$ and $y_{j,t-1}$, we do not generally have a closed form of $V_i(\cdot)$[79] and it is infeasible to directly construct an estimating equation by $f_i(\cdot)$'s.

To build an econometric model based on agents' optimal actions, we utilize the first-order conditions of agents' lifetime problems. The first-order condition of the agent i's t^{th}-period optimal decision is[80]

$$0 = \eta_{it} + \gamma_0 y_{i,t-1} + \lambda_0 w_{t,i.}[Y_{n,t-1}] Y_{nt}^* - y_{it}^* + \delta \frac{\partial}{\partial y_{it}} E_t(V_i(Y_{nt}, \eta_{n,t+1}))|_{Y_{nt}^*} fori = 1, \cdots, n. \tag{3.10}$$

The main issue is to represent the expectation function $\frac{\partial}{\partial y_{it}} E_t(V_i(Y_{nt}, \eta_{n,t+1}))$ in the current and future state variables. For this, we need the smoothness condition for $u_i(\cdot)$ by Assumption 3.2.1 and exchangeability of the differential operator $\frac{\partial}{\partial y_{it}}$ to the conditional expectation $E_t(\cdot)$ by Definition 3.2.1 (ii).[81]

[78]For details in the principle of optimality, refer to the supplement file.

[79]If a network does not evolve, the agent i's lifetime value would be linear-quadratic in state variables and we can derive a closed form expression for $V_i(\cdot)$ and $f(\cdot)$. For that case, refer to Jeong and Lee (2018).

[80] $\frac{\partial F(x)}{\partial x}|_{x^*}$ denotes the derivative of $F(\cdot)$ with respect to x evaluated at $x = x^*$.

[81]Then, by the Lebesgue dominant convergence theorem,

$$\frac{\partial}{\partial y_{it}} \int_{\xi_{n,t+1}} V_i(Y_{nt}, \eta_{n,t+1}) d\mu_{\xi_{n,t+1}} = \int_{\xi_{n,t+1}} \frac{\partial V_i(Y_{nt}, \eta_{n,t+1})}{\partial y_{it}} d\mu_{\xi_{n,t+1}} at Y_{nt}^*$$

where $\mu_{\xi_{n,t+1}}$ is the measure for $\xi_{n,t+1}$.

By the law of iterated expectations and substitutions, we have the Euler equation for agent i

$$\begin{aligned} 0 &= \eta_{it} + \gamma_0 y_{i,t-1} + \lambda_0 w_{t,i.}\left[Y_{n,t-1}\right] Y_{nt}^* - y_{it}^* \\ &\quad -\delta\gamma_0 y_{it}^* + \delta E_t\left[\left(\gamma_0 + \lambda_0 \frac{\partial w_{t+1,i.}\left[Y_{nt}^*\right]}{\partial y_{it}} Y_{n,t+1}^*\right) y_{i,t+1}^*\right] \\ &\quad +\delta E_t \sum_{j\neq i}^{n} \frac{\partial y_{j,t+1}}{\partial y_{it}} \lambda_0 w_{t+1,ij}\left[Y_{nt}^*\right] y_{i,t+1}^* \\ &\quad + \sum_{k=1}^{\infty} \delta^{k+1} E_t \left(\begin{array}{c} \lambda_0 \sum_{j_1,\cdots,j_k\neq i}^{n} \frac{\partial y_{j_1,t+1}}{\partial y_{it}} \cdots \frac{\partial y_{j_k,t+k}}{\partial y_{j_{k-1},t+k-1}} \frac{\partial w_{t+k+1,i.}\left[Y_{n,t+k}^*\right]}{\partial y_{j_k,t+k}} Y_{n,t+k+1}^* y_{i,t+k+1}^* \\ +\lambda_0 \sum_{j_1,\cdots j_{k+1}\neq i}^{n} \frac{\partial y_{j_1,t+1}}{\partial y_{it}} \cdots \frac{\partial y_{j_{k+1},t+k+1}}{\partial y_{j_k,t+k}} w_{t+k+1,i,j_{k+1}} \left[Y_{n,t+k}^*\right] y_{i,t+k+1}^* \end{array} \right) \end{aligned} \tag{3.11}$$

since (i) $\frac{\partial u_i(Y_{n,t+1},Y_{nt},\eta_{i,t+1})}{\partial y_{it}} = \left(\gamma_0 + \lambda_0 \frac{\partial w_{t+1,i.}[Y_{nt}]}{\partial y_{it}} Y_{n,t+1}\right) y_{i,t+1} - \gamma_0 y_{it}$, (ii) $\frac{\partial u_i(Y_{n,t+1},Y_{nt},\eta_{i,t+1})}{\partial y_{jt}} = \lambda_0 \frac{\partial w_{t+1,i.}[Y_{nt}]}{\partial y_{jt}} Y_{n,t+1} y_{i,t+1}$ for $j \neq i$, and (iii) $\frac{\partial u_i(Y_{n,t+1},Y_{nt},\eta_{i,t+1})}{\partial y_{j,t+1}} = \lambda_0 w_{t+1,ij}\left[Y_{nt}\right] y_{i,t+1}$ for $j \neq i$.

Equation (3.10) equates the negative marginal change in u_{it} with the expected future marginal payoffs when there is a small change in y_{it}. By the Euler equations, we can capture the marginal effect of changing y_{it} on future network links (contained in the future marginal payoffs), i.e., the existence of coevolution of agents' actions and spatial network links. However, a difficulty arises on utilizing the Euler equation approach since the agents' current actions can (nonlinearly) affect future marginal payoffs through future network links. Utilizing equation (3.9) and the envelope theorem, substitution yields

$$\begin{aligned} &\frac{\partial V_i\left(Y_{nt}, \eta_{n,t+1}\right)}{\partial y_{it}} \\ &= \frac{\partial u_i(Y_{n,t+1}, Y_{nt}, \eta_{i,t+1})}{\partial y_{it}}\Big|_{Y_{n,t+1}^*} + \sum_{j\neq i}^{n} \frac{\partial y_{j,t+1}}{\partial y_{it}} \left(\begin{array}{c} \frac{\partial u_i(Y_{n,t+1},Y_{nt},\eta_{i,t+1})}{\partial y_{j,t+1}} \\ +\delta E_{t+1}\left(\frac{\partial V_i(Y_{n,t+1},\eta_{n,t+2})}{\partial y_{j,t+1}}\right) \end{array} \right) \Big|_{Y_{n,t+1}^*} \\ &= \frac{\partial u_i(Y_{n,t+1}, Y_{nt}, \eta_{i,t+1})}{\partial y_{it}}\Big|_{Y_{n,t+1}^*} + \sum_{j\neq i}^{n} \frac{\partial y_{j,t+1}}{\partial y_{it}} \frac{\partial u_i(Y_{n,t+1}, Y_{nt}, \eta_{i,t+1})}{\partial y_{j,t+1}}\Big|_{Y_{n,t+1}^*} \end{aligned}$$

$$+\sum_{k=1}^{\infty}\delta^k E_{t+1}\left(\begin{array}{c}\sum_{j_1,\cdots,j_k\neq i}^{n}\frac{\partial y_{j_1,t+1}}{\partial y_{it}}\cdots\frac{\partial y_{j_k,t+k}}{\partial y_{j_{k-1},t+k-1}}\frac{\partial u_i(Y_{n,t+k+1},Y_{n,t+k},\eta_{i,t+k+1})}{\partial y_{j_k,t+k}}\\ +\sum_{j_1,\cdots j_{k+1}\neq i}^{n}\frac{\partial y_{j_1,t+1}}{\partial y_{it}}\cdots\frac{\partial y_{j_{k+1},t+k+1}}{\partial y_{j_k,t+k}}\frac{\partial u_i(Y_{n,t+k+1},Y_{n,t+k},\eta_{i,t+k+1})}{\partial y_{j_{k+1},t+k+1}}\end{array}\right)|_{Y^*_{n,t+k+1}}$$

since the first-order condition (3.10) leads to the envelope theorem $\frac{\partial y^*_{i,t+1}}{\partial y_{it}}\left(\frac{\partial u_i(Y_{n,t+1},Y_{nt},\eta_{i,t+1})}{\partial y_{i,t+1}}+\delta E_t\left(\frac{\partial V_i(Y_{n,t+1},\eta_{n,t+2})}{\partial y_{i,t+1}}\right)\right)|_{Y^*_{n,t+1}}=0$. Hence, we do not need to compute $\frac{\partial y^*_{i,t+1}}{\partial y_{it}}$ for all $i=1,\cdots,n$.[82] Due to the agents' non-cooperative behaviors, however, we observe that $\frac{\partial u_i(Y_{nt},Y_{n,t-1},\eta_{it})}{\partial y_{jt}}|_{Y^*_{nt}}+\delta E_t\left(\frac{\partial V_i(Y_{nt},\eta_{n,t+1})}{\partial y_{jt}}\right)|_{Y^*_{nt}}$ might not be zero for $j\neq i$. From this, we see that a change of i's current action y_{it} generates changes of his/her entire expected future payoffs through changing his/her rivals' next period actions (indirect effect). In our dynamic network game framework, hence, the envelope theorem cannot eliminate $\frac{\partial y_{j,t+1}}{\partial y_{it}}$ for $j\neq i$ due to the non-cooperative feature and hence future terms involving $\frac{\partial y_{j,t+1}}{\partial y_{it}}$ for $j\neq i$ in the above equation, i.e., the Benveniste-Scheinkman formula cannot be applied.

We observe that the coevolution of agents' actions and spatial network links can be summarized by the terms $\frac{\partial w_{t+1,i.}[Y^*_{nt}]}{\partial y_{it}}$ and $\frac{\partial w_{t+k+1,i.}[Y^*_{n,t+k}]}{\partial y_{j,t+k}}$, and the two parameters (λ_0 and ψ_0) control these effects. It implies that the coevolution comes from two channels: (i) evolution of economic indicators driven by agents' actions (ψ_0, see equation (3.2)) and (ii) spatial interaction effects (λ_0). If $\psi_0=0$, the variables forming the economic similarities (e.g., $1/|z_{it}-z_{jt}|$) are not affected by their previous actions, and network links might only come from geographic arrangements. Also, even though $\psi_0\neq 0$, if $\lambda_0=0$, spatial networks do not play a role in agents' actions because every agent solves his/her optimization problem without considering other agents' actions. In this

[82] Basically, the envelope theorem means that a marginal change of the optimizer does not contribute to the change in the optimal value. Then, the marginal effects of state variables on the control variable need not be computed in order to evaluate the marginal change of the objective function (Judd (1998), p453). For example, the Benveniste-Scheinkman formula (well utilized in macroeconomics) says that the marginal change of the value can be characterized by only the marginal change of the per period payoff at the optimum. In our model, however, the envelope theorem gives that no player can improve his/her lifetime payoff by unilaterally marginally changing his/her strategy.

case, the agent i's Euler equation becomes

$$0 = \eta_{it} + \gamma_0 y_{i,t-1} - y_{it}^* + \delta\gamma_0 \left(E_t \left(y_{i,t+1}^* \right) - y_{it}^* \right), \tag{3.12}$$

which implies that the economic system is autarky.

Since equation (B.5) involves a lot of future terms, it is infeasible for estimation. Hence, our next step is to obtain a feasible estimating equation based on (B.5). Note that $\frac{\partial w_{t+1,i.}[Y_{nt}^*]}{\partial y_{it}}$ mainly describes an impact of agent's action on future network links. With highlighting out the marginal effect of y_{it} on his/her $(t+1)^{th}$-period payoff, we represent the other future terms as $\frac{\partial V_i(Y_{n,t+1},\eta_{n,t+2})}{\partial y_{j,t+1}}|_{Y_{n,t+1}^*}$ for $j \neq i$ and try to approximate them using a numerical approximation method. For each $p = 1, \cdots, P$, let $m_{t+1,p,ij}\left[y_{it}^*, y_{jt}^*\right] = \frac{\partial h(z_{i,t+1}, z_{j,t+1})}{\partial z_{i,t+1,p}} h_d(d_{ij}, d_c)$, and $m_{t+1,p,i.}\left[Y_{nt}^*\right] = \left(m_{t+1,p,i1}\left[y_{it}^*, y_{1t}^*\right], \cdots, m_{t+1,p,in}\left[y_{it}^*, y_{nt}^*\right]\right)$. The $m_{t+1,p,ij}\left[y_{it}^*, y_{jt}^*\right]$ describes the marginal impact of the p^{th} economic indicator $z_{i,t+1,p}$ on the $(t+1)^{th}$-period network link $w_{t+1,ij}\left[y_{it}^*, y_{jt}^*\right]$. Then, $\lambda_0 \frac{\partial w_{t+1,i.}[Y_{nt}^*]}{\partial y_{it}} = \sum_{p=1}^{P} \lambda_0 \psi_{p,0} m_{t+1,p,i.}\left[Y_{nt}^*\right]$. By (B.5), the systemized stochastic Euler equation is

$$\begin{aligned} \mathbf{0}_{n\times 1} &= \eta_{nt} + \gamma_0 Y_{n,t-1} + \lambda_0 W_{nt}\left[Y_{n,t-1}\right] Y_{nt}^* - (1+\delta\gamma_0) Y_{nt}^* \\ &+\delta E_t \left\{ \begin{array}{c} \left[\begin{array}{c} \gamma_0 I_n + diag_{i=1}^n \left(\sum_{p=1}^{P} \lambda_0 \psi_{p,0} m_{t+1,p,i.}\left[Y_{nt}^*\right] Y_{n,t+1}^* \right) \\ +\lambda_0 diag_{i=1}^n \left(\left(e_{ni}' \Delta_{n,t+1}^* \circ \tilde{e}_{ni}' \right) w_{t+1,i.}'\left[Y_{nt}^*\right] \right) \end{array} \right] Y_{n,t+1}^* \\ +\delta \nabla V_{n,t+2} \end{array} \right\} \end{aligned} \tag{3.13}$$

where $\Delta_{n,t+1}^*$ is an $n \times n$ matrix having $\frac{\partial y_{j,t+1}}{\partial y_{it}}$ as the (i,j)-element, $\nabla V_{n,t+2}$ is an n-dimensional column vector defined with its i^{th}-element being $\left(e_{ni}' \Delta_{n,t+1}^* \circ \tilde{e}_{ni}'\right)\left(\tilde{e}_{ni} \circ \frac{\partial V_i(Y_{n,t+1},\eta_{n,t+2})}{\partial Y_{n,t+1}}|_{Y_{n,t+1}^*}\right)$, $\circ$ denotes the Hadamard product, and $\tilde{e}_{ni} = l_n - e_{ni}$ with l_n an $n \times 1$ vector of ones. The system equation (3.13) can be simply represented for Y_{nt} at Y_{nt}^* by[83]

$$\mathbf{0}_{n\times 1} = \mathcal{J}\left(Y_{n,t-1}, \eta_{nt}, Y_{nt}\right)|_{Y_{nt}^*}. \tag{3.14}$$

That is, values of $\mathcal{J}\left(Y_{n,t-1}, \eta_{nt}, Y_{nt}\right)$ take zero at the optimum Y_{nt}^* given $(Y_{n,t-1}, \eta_{nt})$. The particular complication in evaluating this $\mathcal{J}(\cdot)$ value is that $V_i\left(Y_{n,t+1}, \eta_{n,t+2}\right)$ captures all the future optimized values after $t+2$ via $\nabla V_{n,t+2}$. To implement equation (3.14) using a spatial panel data set, one may consider a (direct) numerical method (e.g., value/policy function iteration method). However, numerically evaluating those terms will lead to a curse of dimensionality since we consider the large spatial units (n) yielding the corresponding large state space.[84]

Hence, we are motivated to use an alternative approximation method. That is, we try to reduce model's complexity by a behavioral assumption (bounded rationality). By an approximation technique, this concept is useful when we have a difficulty in deriving a solution under full rationality. For the related issues, refer to Chapter 7.4 in Rubinstein (1998). Under the concept of bounded rationality, agents have limitations in their forecasting abilities (limited foresight).

[83]Note that $Y_{n,t+1}^*$ in the conditional expectation $E_t(\cdot)$ is the function of Y_{nt}^* and $\eta_{n,t+1}$, while Y_{nt}^* is a function of $Y_{n,t-1}$ and η_{nt}, $\eta_{n,t+1}$ is a function of η_{nt}.

[84]Due to the existence of W_{nt}, y_{it} might be affected by all previous actions $y_{1,t-1}, \cdots, y_{n,t-1}$ and all realized characteristics $\eta_{1t}, \cdots, \eta_{nt}$. It yields a large dimensional state space. If an agent is a state government and we consider the 48 contiguous states in the U.S, the dimension of state variables is 96. Moreover, the state variables in this case belong to a continuous type. Then, the computation of evaluating $V_i(\cdot)$ by the grid method is extremely challenging.

3.3 Construction of the econometric model

In this section, we build an econometric model by approximating the system Euler equation (3.13) under bounded rationality.[85] Our issue is to estimate structural parameters using a panel data set: $\{Y_{nt}, X_{nt}, Z_{nt}\}$ defined on a common probability space $(\Omega, \mathcal{F}, P)$. From this section, the superscript "*" denoting optimality and parenthesis $[Y_{nt}^*]$ in $w_{t+1,ij}[Y_{nt}^*]$ are dropped. First, we need to achieve uniqueness of the optimal policy function since the unique representations of $\Delta_{n,t+1}^*$ and $\frac{\partial V_i(Y_{n,t+1},\eta_{n,t+2})}{\partial y_{j,t+1}}$ (for $j \neq i$) are required. For this, we introduce the following assumptions.

Assumption 3.3.1 *For each i, $\bar{y}_i^\circ$ and $\bar{\eta}_i^\circ$ denote respectively the population means of $\{y_{it}\}_t$ and $\{\eta_{it}\}_t$. Let $\bar{Y}_n^\circ = (\bar{y}_1^\circ, \cdots, \bar{y}_n^\circ)'$ and $\bar{\eta}_n^\circ = (\bar{\eta}_1^\circ, \cdots, \bar{\eta}_n^\circ)'$. We assume $\sup_{n,t} \phi_{\max}\left(\frac{\partial \mathcal{J}(Y_{n,t-1},\eta_{nt},Y_{nt})}{\partial Y_{nt}'}\right) < 0$ at $(Y_{n,t-1}, \eta_{nt}, Y_{nt}) = \left(\bar{Y}_n^\circ, \bar{\eta}_n^\circ, \bar{Y}_n^\circ\right)$ where $\phi_{\max}(\cdot)$ denotes the largest eigenvalue.*

The main purpose of Assumption 3.3.1 is to have a unique relationship between $(Y_{n,t-1}, \eta_{nt})$ (state variables) and Y_{nt} (decision variables) nearby $\left(\bar{Y}_n^\circ, \bar{\eta}_n^\circ\right)$. Note that the realized $\{Y_{nt}\}$ comes from one of the possible $f(\cdot)$ satisfying equation (3.14).[86]

[85] To have an analytic form of the first order conditions in estimation, Li et al. (2007) approximate the first order condition of the concentrated log-likelihood function based on the Taylor approximation. Even though the objective functions of ours and their cases are different, the purpose of the approximation is the same (approximating the first order condition to a tractable form for estimation).

[86] The system of Euler equations is sufficient to represent the optimum by Assumption 3.3.1. However, there are potentially many data generating processes satisfying the system of Euler equations (3.14) (i.e., multiple equilibria) due to highly nonlinearity of equation (3.14). Then, there are multiple mapplings from the state variables to the decision variables. i.e., We cannot guarantee for a unique form of $\Delta_{n,t+1}^*$. Around the population averages, however, we can have their unique representations by the implicit function theorem.

If the original NEs were multiple, economic agents select the NE near $\left(\bar{Y}_n^\circ, \bar{\eta}_n^\circ, \bar{Y}_n^\circ\right)$, which are known to all agents. It means the equilibrium selection mechanism into the structure. For this, we need to assume that all agents know $\left(\bar{Y}_n^\circ, \bar{\eta}_n^\circ\right)$. For the econometric analysis of multiple equilibria issues, refer to de Paula (2013).

The population mean $\bar{Y}_n^\circ$ in Assumption 3.3.1 comes from the data generating process of the actually realized decision variables and is the baseline of applying the implicit function theorem.[87] The existence of $\bar{Y}_n^\circ$ and $\bar{\eta}_n^\circ$ is implied by the stable economic environment.[88] Assumption 3.3.1 is a key condition of the implicit function theorem. $\sup_{n,t} \phi_{\max}\left(\frac{\partial \mathcal{J}(Y_{n,t-1},\eta_{nt},Y_{nt})}{\partial Y_{nt}'}\right)|_{\left(\bar{Y}_n^\circ,\bar{\eta}_n^\circ,\bar{Y}_n^\circ\right)} < 0$ means that the $n \times n$ Hessian matrix $\frac{\partial \mathcal{J}(Y_{n,t-1},\eta_{nt},Y_{nt})}{\partial Y_{nt}'}$ is negative definite. This is the second-order condition justfying the optimum at the population average values. Also, the maximum eigenvalue of the Hessian should be still negative even for large n. By applying the implicit function theorem, there exists a unique function $f(\cdot)$ such that $(Y_{n,t-1}, \eta_{nt}, Y_{nt})$ around the population averages $\left(\bar{Y}_n^\circ, \bar{\eta}_n^\circ, \bar{Y}_n^\circ\right)$ and $f(\cdot)$ is infinitely differentiable because our $\mathcal{J}(\cdot)$ is infinitely differentiable.[89] Note that this uniqueness is due to the implicit function theorem at the local neighborhood of $\left(\bar{Y}_n^\circ, \bar{\eta}_n^\circ, \bar{Y}_n^\circ\right)$.

Then, the optimal policy function $Y_{nt} = f(Y_{n,t-1}, \eta_{nt})$ can be represented by the first-order Taylor polynomial (a mean value theorem) around $\left(\bar{Y}_n^\circ, \bar{\eta}_n^\circ\right)$,

$$f(Y_{n,t-1}, \eta_{nt}) = f\left(\bar{Y}_n^\circ, \bar{\eta}_n^\circ\right) + \frac{\partial f(Y_{n,t-1}, \eta_{nt})}{\partial Y_{n,t-1}'}|_{\left(Y_{n,t-1}^+, \eta_{nt}^+\right)} \left(Y_{n,t-1} - \bar{Y}_n^\circ\right) \quad (3.15)$$

$$+\frac{\partial f(Y_{n,t-1}, \eta_{nt})}{\partial \eta_{nt}'}|_{\left(Y_{n,t-1}^+, \eta_{nt}^+\right)} \left(\eta_{nt} - \bar{\eta}_n^\circ\right)$$

where $\left(Y_{n,t-1}^+, \eta_{nt}^+\right)$ lies between $(Y_{n,t-1}, \eta_{nt})$ and $\left(\bar{Y}_n^\circ, \bar{\eta}_n^\circ\right)$. That is, we pursue the local unique solution near the population averages. Using equation (3.15), the linear

[87]Note that evolution of η_{nt} consisting of η_{nt}^v and η_n^{iv} is not affected by that of Y_{nt} and η_{nt}. Since $|\rho_{\eta,0}| < 1$ and $E_t(\xi_{n,t+1}) = \mathbf{0}_{n\times 1}$, $\bar{\eta}_i^\circ = \eta_i^{iv}$ for all i.

[88]For each i, a way of generating the data $(f_i(\cdot))$ has a time-invariant functional form.

[89]The theorem is stated in Section 3.4 in Judd (1996). Indeed, if $\mathcal{J}(\cdot)$ is a C^k-function, then $f(\cdot)$ is a C^{k-1}-function. This is reason why Santos (1992) introduces a condition of C^2 on the per period payoff function for differentiability of the optimal policy function.

approximation of $f(\cdot)$ (denoted by $f^e(\cdot)$) is

$$f^e\left(Y_{n,t-1},\eta_{nt}\right)=f\left(\bar{Y}_n^{\circ},\bar{\eta}_n^{\circ}\right)+\frac{\partial f\left(Y_{n,t-1},\eta_{nt}\right)}{\partial Y'_{n,t-1}}|_{\left(\bar{Y}_n^{\circ},\bar{\eta}_n^{\circ}\right)}\left(Y_{n,t-1}-\bar{Y}_n^{\circ}\right) \tag{3.16}$$

$$+\frac{\partial f\left(Y_{n,t-1},\eta_{nt}\right)}{\partial \eta'_{nt}}|_{\left(\bar{Y}_n^{\circ},\bar{\eta}_n^{\circ}\right)}\left(\eta_{nt}-\bar{\eta}_n^{\circ}\right).$$

Since Z_{nt} and $Y_{n,t+1}$ are observable and the formation function $h(\cdot)$ (or $h_d(\cdot)$) is (pre-) specified by an econometrician, they give specifications on $\{w_{t,ij}\}$ and $\{m_{t+1,p,ij}\}$ in equation (3.13).[90]

Now we approximate the Euler equation system (3.13) to a tractable form. To get approximations of $\Delta^*_{n,t+1}$ and $\frac{\partial V_i(Y_{n,t+1},\eta_{n,t+2})}{\partial y_{j,t+1}}$ (for $j\neq i$), the main task is calculating $f^e(\cdot)$ and the corresponding approximated value functions $V_i^e(\cdot)$'s. A source of difficulty is that $u_i\left(Y_{nt},Y_{n,t-1},\eta_{it}\right)$ is not a LQ function of $Y_{n,t-1}$ (in particular, network links $w_{t,ij}\left[y_{i,t-1},y_{j,t-1}\right]$'s). To avoid the curse of dimensionality, we employ a regular LQ perturbation method[91] for approximating $u_i(\cdot)$ around the steady state[92] $\left(\bar{Y}_n^{\circ},\bar{\eta}_n^{\circ}\right)$ (approximated $u_i(\cdot)$ is denoted by $u_i^e(\cdot)$) and get $V_i^e(\cdot)$ using the $u_i^e(\cdot)$.[93] Note that we only need to approximate the smooth functions $\{w_{t,ij}\left[y_{i,t-1},y_{j,t-1}\right]y_{it}y_{jt}\}_{i\neq j}$ in

[90] In contrast to conventional spatial econometric literature, a researcher needs to know the formation functions $h_d(\cdot)$ and $h(\cdot)$ to evaluate $\frac{\partial h(z_{i,t+1},z_{j,t+1})}{\partial z_{i,t+1}}$. One way to get $h_d(\cdot)$ and $h(\cdot)$ is to prespecify them. Another way is to use an estimation method by assuming a parametric functional form of $h(\cdot)$. For this, we suggest an estimation method to specify $h(\cdot)$ and $h_d(\cdot)$ in Appendix C.

[91] The regular perturbation means that a small change in the problem induces a small change in the solution.

[92] From systems theory, a process Y_{nt} is in a steady state if $Y_{nt}=Y_{n,t-1}$ for all t. Steady state values what we select here are the population averages, $\left(\bar{Y}_n^{\circ},\bar{\eta}_n^{\circ}\right)$.

[93] The LQ approximation method is a local approximation method and explains how the dynamic system evolves around the steady state. Famous examples are Magill (1977) and Kydland and Prescott (1982). Benigno and Woodford (2012) provide theoretical discussions in this method. For a review, refer to Chapters 3 and 4 of Judd (1996) and Chapters 13 and 14 of Judd (1998).

$u_i(\cdot)$. For the details about the approximation procedure, refer to Appendix A.[94] In practice, for each i $\bar{y}_i^{\circ}$ and $\bar{\eta}_i^{\circ}$ can be computed by the time averages $\bar{\eta}_{i,T} = \frac{1}{T}\sum_{t=1}^{T} \eta_{it}^{obs}$ and $\bar{y}_{i,T} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ where η_{it}^{obs} denotes the observable part of η_{it}. After computing $V_i^e(\cdot)$'s by solving the algebraic matrix Riccati equations based on the LQ functions $u_i^e(\cdot)$'s, the marginal effect matrix $\Delta_{n,t+1}^*$ can be obtained. Because Δ_{nt}^* is time invariant, it will be denoted by Δ_n^*.[95] By replacing $\Delta_{n,t+1}^*$ and $\frac{\partial V_i(Y_{n,t+1},\eta_{n,t+2})}{\partial y_{j,t+1}}$ (in equation (3.14)) with Δ_n^* and $\frac{\partial V_i^e(Y_{n,t+1},\eta_{n,t+2})}{\partial y_{j,t+1}}$ using $V_i^e(\cdot)$, we have the system of Euler equations under bounded rationality,

$$\mathbf{0}_{n\times 1} = \mathcal{J}^e\left(Y_{n,t-1}, \eta_{nt}, Y_{nt}\right). \tag{3.17}$$

For each t, the observable part of η_{nt}^v ($\eta_{nt}^{obs} = \left(\eta_{1t}^{obs}, \cdots, \eta_{nt}^{obs}\right)'$) is specified by $X_{nt}\beta_0$ where $\beta_0 = (\beta_{1,0}, \cdots, \beta_{K,0})'$ stands for a K-dimensional vector of coefficients. By studying $\lambda_0\psi_{1,0}, \cdots, \lambda_0\psi_{P,0}$, we can capture (i) whether the agents' actions and spatial networks coevolve, and (ii) which economic indicators work to the coevolution. Then, the parameters of our interests are the true values λ_0, γ_0, ψ_0, and β_0, which are summarized as $\theta_0 = (\lambda_0, \gamma_0, \psi_0, \beta_0')'$. The time-discounting factor δ is considered as a primitive parameter[96] and the incidental parameters in $\rho_{\eta,0}$ are supposed to be already

[94] One can analyze the approximated linear optimal policy induced by $V_i^e(\cdot)$'s. To investigate coevolution of economic activities and networks, however, we need to consider the simultaneous relation between the (approximated) outcome equation and the entry equation (3.2). Since the simultaneous effects of network links and optimal actions are highly nonlinear, it is difficult to capture the simultaneous relationship. Since the system Euler equation can represent (i) evolution of Y_{nt} and (ii) effects of evolving spatial networks from Y_{nt} by $diag_{i=1}^n\left(\sum_{p=1}^P \lambda_0\psi_{p,0} m_{t+1,p,i\cdot} Y_{n,t+1}\right)$, we take the Euler equation approach.

[95] Since the approximated $u_i(Y_{nt}, Y_{n,t-1}, \eta_{it})$ becomes a LQ function of its arguments, the $V_i^e(Y_{n,t-1}, \eta_{nt})$ is linear-quadratic in $(Y_{n,t-1}, \eta_{nt})$. Then, the resulted Δ_n^* from $V_i^e(Y_{n,t-1}, \eta_{nt})$'s is a time-invariant matrix.

[96] In structural econometrics, we usually avoid estimating δ due to difficulties in identification. In practice, δ is selected by economic reasonings. For example, $\delta = \frac{1}{1+\bar{r}}$ where $\bar{r}$ is a long-term interest rate.

revealed based on estimating process of X_{nt}.[97] One of the advantages of using panel data is robustly controlling unobserved cross-section and time heterogeneities if n and T are large. To simply control them, we consider the additive individual and time dummies as incidental parameters[98]: $\mathbf{c}_{n0} = (c_{1,0}, \cdots, c_{n,0})'$ and $\alpha_{T,0} = (\alpha_{1,0}, \cdots, \alpha_{T,0})'$. Let $\mathcal{E}_{nt} = (\epsilon_{1t}, \cdots, \epsilon_{nt})'$ denote the idiosyncratic disturbances in implementing the approximated Euler equation system, (3.17).[99] For each pair (i, t), hence, the error term is specified by a linear function, $c_{i,0} + \alpha_{t,0} + \epsilon_{it}$. To achieve robust estimation results, we try to directly estimate $\mathbf{c}_{n0}$ and $\alpha_{T,0}$ instead of considering them as random components.

From equation (3.17), we set up the moment conditions for estimation. For each t, for example, let

$$\ell_t = \sigma\left(\{Y_{ns}\}_{s=0}^{t}, \{X_{ns}\}_{s=0}^{T}, \{Z_{ns}^*\}_{s=0}^{T}, \mathbf{c}_{n0}, \alpha_{T,0}\right) \tag{3.18}$$

where Z_{nt}^* denotes the strictly exogenous part of Z_{nt}. As Z_{nt}^*, agents' geographic and their time varying exogenous characteristics can be considered. The t^{th}-period disturbances, $\epsilon_{1t}, \cdots, \epsilon_{nt}$, are orthogonal to ℓ_{t-1}. Then, equation (3.17) at time t can be econometrically specified by

$$\left[\begin{array}{c} \left[S_n\left(W_{nt}\right)Y_{nt} - \gamma_0 Y_{n,t-1} - X_{nt}\beta_0\right] + \delta\gamma_0\left(Y_{nt} - Y_{n,t+1}^e\right) \\ -\delta\left\{\left[\sum_{p=1}^{P}\lambda_0\psi_{p,0}M_{n,t+1,p} + \lambda_0 N_{n,t+1}\right]Y_{n,t+1}^e\right\} \\ -\delta^2\nabla V_{n,t+2}^e \end{array}\right] = \mathbf{c}_{n0} + \alpha_{t,0}l_n + \mathcal{E}_{nt} \tag{3.19}$$

[97]We can reveal the process of X_{nt} without considering the outcome process Y_{nt} since X_{nt} comes from η_{nt}^v, and it is supposed to follow an exogenous linear Markov process.

[98]As a statistical extension, we can also consider multiplicative (interactive) fixed effects (i.e., factor structure). In practice, however, we should set aside relatively large T for using the specification of interactive fixed effects. In a regional data set (our model's main target), T is not quite large.

[99]The error term ϵ_{it} can be interpreted as the agent i's t^{th}-period expectational error by using the $(t-1)^{th}$-period information set (to predict the t^{th}-period forward-looking structural system).

or

$$E\left[\begin{array}{c}[S_n(W_{nt})Y_{nt}-\gamma_0 Y_{n,t-1}-X_{nt}\beta_0]+\delta\gamma_0\left(Y_{nt}-Y^e_{n,t+1}\right)\\ -\delta\left\{\left[\sum_{p=1}^{P}\lambda_0\psi_{p,0}M_{n,t+1,p}+\lambda_0 N_{n,t+1}\right]Y^e_{n,t+1}\right\}\\ -\delta^2\nabla V^e_{n,t+2}-\mathbf{c}_{n0}-\alpha_{t,0}l_n\end{array}|\ell_{t-1}\right]=\mathbf{0}_{n\times 1} \qquad (3.20)$$

for $t=1,\cdots,T$, where $S_n(W_{nt})=I_n-\lambda_0 W_{nt}$ denotes the spatial filter at time t, $Y^e_{n,t+1}=\left(y^e_{1,t+1},\cdots,y^e_{n,t+1}\right)'$ is the expected $Y_{n,t+1}$ from the LQ method, $M_{n,t+1,p}=diag^n_{i=1}\left(m_{t+1,p,i.}Y^e_{n,t+1}\right)$ for $p=1,\cdots,P$, $N_{n,t+1}=diag^n_{i=1}\left(\left(e'_{ni}\Delta^*_n\circ\tilde{e}'_{ni}\right)w'_{t+1,i.}\right)$, and $\nabla V^e_{n,t+2}$ denotes the approximated $\nabla V_{n,t+2}$. Then, $Y^e_{n,t+1}=A^e_n Y_{nt}+\rho_{\eta,0}B^e_n X_{nt}\beta_0+C^e_n$ and $\Delta^*_n=A^{e\prime}_n$ for some $n\times n$ matrices A^e_n and B^e_n and some $n\times 1$ vector C^e_n. The explicit forms of A^e_n, B^e_n, C^e_n and components in $\nabla V^e_{n,t+2}$ can be found in Appendix A. We observe that equation (3.19) represents a conventional SDPD model if $\delta=0$; Y_{nt} is a nonlinear function of $\mathcal{E}_{nt}$; $M_{n,t+1,p}$ for $p=1,\cdots,P$ show coevolution of actions and spatial networks; $N_{n,t+1}$ represents the indirect changes of the agent i's next period payoff via changes of rivals' next period actions by changing i's current period action; $\nabla V^e_{n,t+2}$ describes the indirect changes of the i's remaining future payoffs via changes of rivals' next period actions by changing i's current period action. Since $M_{n,t+1,p}$ and $\nabla V^e_{n,t+2}$ are functions of Y_{nt}, and $N_{n,t+1}$ and $\nabla V^e_{n,t+2}$ rely on the time average $\bar{Y}_{nT}=(\bar{y}_{1,T},\cdots,\bar{y}_{n,T})'$, they are potentially endogenous.[100] For asymptotic analysis, we will impose some restrictions on those weights matrices in the next section.

Our theoretical model is considered by given n, but our estimation framework is for samples with both large n and T. Hence, here is the model assumption about

[100] If spatial networks and their relevant components are endogenous, estimates of main parameters in a network interaction model can yield incorrect statistical inferences. For this issue, there are recent studies concerning time-evolving socio-economic networks and/or endogeneity of them. Examples are Lee and Yu (2012), Goldsmith-Pinkham and Imbens (2013), Kelejian and Piras (2014), Qu and Lee (2015), Han and Lee (2016), Hsieh and Lee (2017), Qu et al. (2017), Johnsson and Moon (2017), Kuersteiner and Prucha (2018), and Han, Hsieh and Ko (2019).

the population moment conditions and stability of Y_{nt} under large n. For detailed regularity conditions for $\mathcal{E}_{nt}$, we state those in Section 3.4.

Assumption 3.3.2 *(i) At θ_0, $\mathbf{c}_{n0}$, and $\alpha_{t,0}$, equation (B.6) holds.*

(ii) $\sup_{n,t} \rho_{\max}\left(\frac{\partial f^e(Y_{n,t-1},\eta_{nt})}{\partial Y'_{n,t-1}}\right) \leq \tau < 1$ for some $0 \leq \tau < 1$ and $\sup_{n,t} \rho_{\max}\left(\frac{\partial f^e(Y_{n,t-1},\eta_{nt})}{\partial \eta'_{nt}}\right) \leq c_\eta$ for some $c_\eta < \infty$ a.e where $\rho_{\max}(A_n)$ stands for the spectral radius of A_n.

Hence, Assumption B.13 (i) yields the moment conditions for the GMM estimation, which is based on the approximated Euler equation system (3.17). For stability of the system under large n, we impose Assumption B.13 (ii) as sufficient conditions. In particular, $\sup_{n,t} \rho_{\max}\left(\frac{\partial f^e(Y_{n,t-1},\eta_{nt})}{\partial Y'_{n,t-1}}\right) < 1$ and $\sup_{n,t} \rho_{\max}\left(\frac{\partial f^e(Y_{n,t-1},\eta_{nt})}{\partial \eta'_{nt}}\right) \leq c_\eta$ give sufficient conditions to avoid an explosive Euler equation even for large n. This guarantees for bounded lifetime values given initial conditions.[101] Also, they are key devices of showing asymptotic properties of suggested estimators.[102] First, $\sup_{n,t} \rho_{\max}\left(\frac{\partial f^e(Y_{n,t-1},\eta_{nt})}{\partial Y'_{n,t-1}}\right) < 1$ means that $f^e(\cdot)$ is a contraction mapping of $Y_{n,t-1}$. This property gives a background of using a sample mean $\bar{Y}_{nT} = (\bar{y}_{1,T}, \cdots, \bar{y}_{n,T})'$ as an estimate of $\bar{Y}_n^\circ$. Second, $\sup_{n,t} \rho_{\max}\left(\frac{\partial f^e(Y_{n,t-1},\eta_{nt})}{\partial \eta'_{nt}}\right) \leq c_\eta$ implies that $f^e(\cdot)$ is a Lipschitz continuous function of exogenous characteristics given $Y_{n,t-1}$. This device helps to restrict the degree of dependencies in the spatial dimension.

[101] Hence, this assumption implies the principle of optimality: given initial conditions, lifetime values and policy functions are bounded. For this issue, see the supplement file.

[102] In case of myopic spatial-dynamic models, we can characterize the stability conditions using parameter values and properties of a spatial network matrix (see Lee and Yu (2015)). For the forward-looking spatial-dynamic models, it is difficult to capture the relationship between potential causes via (i) the parameter values (λ_0, γ_0), (ii) time-discounting factor δ, and (iii) magnitude and denseness of W_{nt} and stability conditions. Hence, we introduce a high level assumption, Assumption B.13 (ii). Intuitively, we might need to have that (i) λ_0, γ_0 and δ are small, (ii) elements in W_{nt} are small, and (iii) W_{nt}'s are sparse.

3.4 Parameter estimation

This section implements the GMM estimation method based on the derived moment conditions (B.6). Second, we will study statistical properties of the GMM estimator. For this, we characterize a topological space where spatial-time units are located. On the defined space, the set of realized data (e.g., $\{Y_{nt}, X_{nt}\}$ and $\{\mathcal{E}_{nt}\}$) are (weakly dependent) random fields. Distances among spatial-time units on the topological space will specify intensities of spatial-time dependencies. Based on this specification, we formulate sample moment functions (for notational convenience) and investigate asymptotic properties of the GMM estimator.

3.4.1 Topological specification

Each data generating process (DGP) can be indexed by a spatial-time unit on a specific topological space. The purpose of this setting is to characterize (weak) dependencies among spatial-time units based on their distances and network links. It is essential that, if two spatial-time units are sufficiently far, those two spatial-time processes would be nearly uncorrelated. Hence, the issue here is to specify "farness" of spatial-time units. Relevant concepts are borrowed from Chapter 17 in Davidson (1994), Jenish and Prucha (2012) and Qu et al. (2017).

Note that we have n spatial units over T periods as a data set. We assume that a spatial unit i has its fixed location in a subset of $\mathbf{R}^d$ $(d \geq 1)$ for all t. When a state government is considered as an individual (or spatial) unit, it is located on Earth, so its latitude and longitude can specify its physical location. Since a pair of latitude and longitude can be one-to-one transformed to $\mathbf{R}^2$, a state government can be placed on an unevenly spaced lattice in $\mathbf{R}^2$. By the same logic, in addition to physical distance, there might be economic characteristics which can be regarded as other dimension of locations. So, in general, i's t^{th}-period location is characterized in a subset of $\mathbf{R}^{d+1}$ (i.e., nT spatial-time processes on $\mathbf{R}^{d+1}$). Based on this idea, define a location function

$$l : \{1, \cdots, n\} \times \{1, \cdots, T\} \to D_L \subset \mathbf{R}^{d+1} \tag{3.21}$$

by $l\,(i,t) = (l_1(i), \cdots, l_d(i), t)$. Since we have $L = nT$ observations, it means there are L observed locations, and they are specified by the unevenly spaced lattice D_L.[103] For $i = 1, \cdots, n$, $l(i,0) = (l_1^*(i), \cdots, l_d^*(i), 0)$ denotes i's physical location. At time $t = 1$, $l(i,1) = (l_1^*(i), \cdots, l_d^*(i), 1)$. It means that all spatial units $i = 1, \cdots, n$ move vertically and parallelly upward from $t = 0$ to $t = 1$. By the same logic, $l(i,-1) = (l_1^*(i), \cdots, l_d^*(i), -1)$ at $t = -1$. Here is the formal setting.

Assumption 3.4.1 *A possible set of locations is described by a lattice D, which is an infinitely countable subset of $\mathbf{R}^{d+1}$ $(d \geq 1)$. There exists a mapping $l\,(\cdot)$ from $\{1, \cdots, L\}$ to $D_L \subset D$. The minimum distance between two different elements in D is 1.*

Assumption 3.4.1 characterizes weak dependence of the spatial-time processes for our asymptotic inference. The set D_L stands for the collection of locations in $\mathbf{R}^{d+1}$ corresponding to an (available) spatial panel data set. Since there are n cross-section units with T time periods, the number of spatial-time units' locations in D_L is L. The set D is a set of potential locations which can accomodate all n spatial units, where n can tend to infinity, so it contains infinitely countable locations (i.e., $|D| = |\mathbf{N}|$ where $\mathbf{N}$ denotes the set of natural numbers). When a bounded region in D is taken,

[103] Then, $|D_L| = L$ where $|A|$ denotes the cardinality of a set A.

Assumption 3.4.1 implies that there exist at most a finite spatial-time units which can be located there. It implies that our asymptotic inference will be based on the increasing domain asymptotics.[104]

From Jenish and Prucha's (2012) and Qu et al. (2017), we consider the maximum metric to evaluate the distance between two spatial-time units $l(i,t)$ and $l(j,t')$:

$$\|l(i,t) - l(j,t')\|_\infty = \max\{|t-t'|, \|l(i,0) - l(j,0)\|_\infty\}. \tag{3.22}$$

Then, $l(i,t)$ and $l(j,t')$ are neighbors if i and j are physical neighbors and time t' is a near epoch to t. Then, two spatial-time processes indexed by $l(i,t)$ and $l(j,t')$ are nearly uncorrelated if either $|t-t'|$ or $\|l(i,0) - l(j,0)\|_\infty$ is large. This maximum metric (3.22) is employed to define a base σ-field,

$$\mathcal{F}_{l(i,t),L}(s) = \sigma\left(\varsigma_{l(j,t')} : l(j,t') \in D_L, \|l(i,t) - l(j,t')\|_\infty \leq s\right) \tag{3.23}$$

where s denotes a threshold distance, and the random field $\left\{\varsigma_{l(i,t)}\right\}$ denotes a set of baseline processes. Based on the base σ-field $\mathcal{F}_{l(i,t),L}(s)$, a spatial-time process located at $l(i,t)$ will be approximated by $\varsigma_{l(j,t')}$'s such that $\|l(i,t) - l(j,t')\|_\infty \leq s$. Some regularity conditions will be introduced to obtain a controllable dependencies. We will rely on this setting for proving consistency of GMM estimators.

3.4.2 Estimation: nonlinear two-stage least squares (NL2S) and generalized method of moments (GMM) estimation methods

To estimate θ_0, we consider the generalized method of moments (GMM) estimation method based on the approximated Euler equations. We have only a partial

[104] An opposite concept is the infill asymptotic (fixed domain asymptotic) where the number of sampling spatial-time units increases even though a bounded region is selected. In the limit, then, the intensity of interactions could not be controlled (strong dependence). Since our model's primary applications are studying local government's behavior, considering the increasing domain asymptotic would be appropriate.

specification[105] of the model by the system of stochastic Euler equations. Under this limited information setting, considering GMM estimation is a frequently chosen option (e.g., Hansen and Singleton (1982)). By Assumption B.13, we select instrumental variables (IVs) from ℓ_{t-1}. Let $\mathbf{q}_{nt} = (q_{1t}', \cdots, q_{nt}')'$ be an $n \times q$ matrix of IVs and $q_{it} = (q_{it,1}, \cdots, q_{it,q})$ for $i = 1, \cdots, n$. The order condition for identification says that $q \geq 2 + K + P$ is required. Let Θ denote a $(2 + K + P)$-dimensional parameter space. Let $\theta = (\lambda, \gamma, \psi', \beta')'$, $\mathbf{c}_n = (c_1, \cdots, c_n)'$, and α_t denote parameter values. For each $\theta \in \Theta$, define $S_n(W_{nt}, \lambda)$, $A_n^e(\theta)$, $B_n^e(\theta)$, and $C_n^e(\theta)$, $M_{n,t+1,p}(\theta)$, $N_{n,t+1}(\theta)$, and $\nabla V_{n,t+2}^e(\theta)$ to being $S_n(W_{nt}) = S_n(W_{nt}, \lambda_0)$, $A_n^e = A_n^e(\theta_0)$, $B_n^e = B_n^e(\theta_0)$, $C_n^e = C_n^e(\theta_0)$, $M_{n,t+1,p} = M_{n,t+1,p}(\theta_0)$, $N_{n,t+1} = N_{n,t+1}(\theta_0)$, and $\nabla V_{n,t+2}^e = \nabla V_{n,t+2}^e(\theta_0)$. For each $\theta \in \Theta$, hence, $Y_{n,t+1}^e(\theta) = A_n^e(\theta) Y_{nt} + \rho_{\eta,0} B_n^e(\theta) X_{nt}\beta + C_n^e(\theta)$. Then, we have

$$
\begin{aligned}
\mathcal{E}_{nt}(\theta, \mathbf{c}_n, \alpha_t) \;=\; & [S_n(W_{nt}, \lambda) Y_{nt} - \gamma Y_{n,t-1} - X_{nt}\beta] + \delta\gamma \left(Y_{nt} - Y_{n,t+1}^e(\theta)\right) \\
& -\delta \left[\sum_{p=1}^{P} \lambda\psi_p M_{n,t+1,p}(\theta) + \lambda N_{n,t+1}(\theta)\right] Y_{n,t+1}^e(\theta) \\
& -\delta^2 \nabla V_{n,t+2}^e(\theta) - \mathbf{c}_n - \alpha_t l_n
\end{aligned}
$$

by equation (3.19). The i^{th}-element of $\mathcal{E}_{nt}(\theta, \mathbf{c}_n, \alpha_t)$ is denoted by $\epsilon_{it}(\theta, c_i, \alpha_t)$.

For each $l(i,t) \in D_L$ and $(\theta, \mathbf{c}_n, \alpha_t)$, we have $q_{l(i,t),L} = q_{it}$ and $\epsilon_{l(i,t),L}(\theta, c_i, \alpha_t) = \epsilon_{it}(\theta, c_i, \alpha_t)$, and define the linear IV moment function to practically implement (B.6)

$$
g_{l(i,t),L}^{\mathbf{L}}(\theta, c_i, \alpha_t) = q_{l(i,t),L}\epsilon_{l(i,t),L}(\theta, c_i, \alpha_t), \tag{3.24}
$$

so $E\left(g_{l(i,t),L}^{\mathbf{L}}(\theta, c_i, \alpha_t)\right) = \mathbf{0}_{q\times 1}$ at $(\theta, c_i, \alpha_t) = (\theta_0, c_{i,0}, \alpha_{t,0})$ by Assumption B.13. To help identification of ψ_0, we can add the linear moment conditions originated from

[105] Note that we do not fully recover the whole economic structure (including network interactions and formation). We focus on estimating only the network interaction part.

specification (3.2).[106] If we only use the IV moments, our estimation methodology will be the nonlinear two-stage least squares (NL2S) estimation method. Note that the linear IV moment conditions are originated from the population orthogonality condition (B.6), so they become the main source of identification.[107]

We can also consider a quadratic moment introduced in conventional spatial econometrics literature.[108] For each time t, let $R_{nt,1}, \cdots, R_{nt,m}$ be $n \times n$ matrices whose all components are strictly exogenous and all diagonal elements are zero.[109] Under the i.i.d disturbance assumption, for each time t and $l = 1, \cdots, m$,

$$E\left(\mathcal{E}_{nt}' R_{nt,l} \mathcal{E}_{nt}\right) = \sigma_0^2 tr\left(R_{nt,l}\right) = 0. \tag{3.25}$$

The quadratic moment function for each $l(i,t) \in D_L$ can be defined by

$$g_{l(i,t),L}^{\mathbf{Q}}\left(\theta, \mathbf{c}_n, \alpha_t\right) = \begin{bmatrix} \sum_{l(j,t)} \left[R_{nt,1}\right]_{ij} \epsilon_{l(i,t),L}\left(\theta, c_i, \alpha_t\right) \epsilon_{l(j,t),L}\left(\theta, c_j, \alpha_t\right) \\ \vdots \\ \sum_{l(j,t)} \left[R_{nt,m}\right]_{ij} \epsilon_{l(i,t),L}\left(\theta, c_i, \alpha_t\right) \epsilon_{l(j,t),L}\left(\theta, c_j, \alpha_t\right) \end{bmatrix}. \tag{3.26}$$

In employing the quadratic moment condition, the i.i.d. disturbances assumption, strictly exogenous components and zero diagonal entries of $R_{nt,l}$ $(l = 1, \cdots, m)$ are essential.

Hence, we have the $(m+q)$-dimensional moment vector $g_{l(i,t),L}\left(\theta, \mathbf{c}_n, \alpha_t\right) = \begin{bmatrix} g_{l(i,t),L}^{\mathbf{Q}}\left(\theta, \mathbf{c}_n, \alpha_t\right) \\ g_{l(i,t),L}^{\mathbf{L}}\left(\theta, c_i, \alpha_t\right) \end{bmatrix}$ and $E\left(g_{l(i,t),L}\left(\theta_0, \mathbf{c}_{n0}, \alpha_{t0}\right)\right) = \mathbf{0}_{(q+m)\times 1}$ for

[106] For a simple example, $q_{it} v_{it,p}\left(\psi_{p,0}, c_i, \alpha_t\right)$ for $p = 1, \cdots, P$ where $v_{it,p}\left(\psi_{p,0}, c_i, \alpha_t\right) = z_{it,p} - \psi_{p,0} y_{i,t-1} - c_i - \alpha_t$.

[107] We only rely on $E_{t-1}\left(\epsilon_{it}\right) = 0$ for the linear IV moment conditions. It implies that the NL2S estimation method can be robust to unknown heteroskedasticity and unknown serial/spatial correlations since this method does not rely on additional stochastic properties (e.g., heteroskedasticity and correlations) of ϵ_{it}.

[108] Refer to Lee (2007) and Lee and Yu (2014).

[109] A broader class of quadratic moment matrices can be also considered: e.g., $R_{nt,l}$ satisfying $tr\left(R_{nt,l}\right) = 0$. However, this class of quadratic moment matrices would not be valid under unknown heteroskedasticity. Since we do not want to highlight the i.i.d. property of $\left\{\epsilon_{l(i,t),L}\right\}$, we take a narrower class of quadratic moment matrices for robust estimation.

each $l(i,t) \in D_L$.[110] To give different weights for the $(q+m)$-moment conditions, let a_L be a $K_a \times (m+q)$ random matrix with a full row rank greater than or equal to the number of unknown parameters $2+P+K$. Assume $a_0 = \text{plim}_{L\to\infty} a_L$ is of full row rank. Then, the GMM estimator is defined by

$$\left(\hat{\theta}_L, \hat{\mathbf{c}}_{n,L}, \hat{\alpha}_{T,L}\right) = \arg\min_{\theta\in\Theta,\mathbf{c}_n,\alpha_T} \bar{g}'_L(\theta,\mathbf{c}_n,\alpha_T)\, a'_L a_L \bar{g}_L(\theta,\mathbf{c}_n,\alpha_T) \tag{3.27}$$

where $\bar{g}_L(\theta,\mathbf{c}_n,\alpha_T) = \frac{1}{L}\sum_{l(i,t)\in D_L} g_{l(i,t),L}(\theta,\mathbf{c}_n,\alpha_t)$.

Let $\mathcal{E}_L(\theta,\mathbf{c}_n,\alpha_T) = \left(\mathcal{E}'_{n1}(\theta,\mathbf{c}_n,\alpha_t),\cdots,\mathcal{E}'_{nT}(\theta,\mathbf{c}_n,\alpha_t)\right)'$ with $\mathcal{E}_{nt}(\theta,c_i,\alpha_t) = \left(\epsilon_{1t}(\theta,c_i,\alpha_t),\cdots,\epsilon_{nt}(\theta,c_i,\alpha_t)\right)'$ for each t and $\theta\in\Theta$, and $\mathbf{q}_L = \left(\mathbf{q}'_{n1},\cdots,\mathbf{q}'_{nT}\right)'$. Then, we have king vector representations of $\bar{g}_L(\theta,\mathbf{c}_n,\alpha_T)$, which are useful in computation:

$$\bar{g}^{\mathbf{L}}_L(\theta,\mathbf{c}_n,\alpha_T) = \frac{1}{L}\sum_{l(i,t)\in D_L} g^{\mathbf{L}}_{l(i,t),L}(\theta,c_i,\alpha_t) = \frac{1}{L}\mathbf{q}'_L\mathcal{E}_L(\theta,\mathbf{c}_n,\alpha_T) \tag{3.28}$$

and $\bar{g}^{\mathbf{Q}}_L(\theta,\mathbf{c}_n,\alpha_T) = \frac{1}{L}\sum_{l(i,t)\in D_L} g^{\mathbf{Q}}_{l(i,t),L}(\theta,\mathbf{c}_n,\alpha_t) = \frac{1}{L}\begin{bmatrix} \mathcal{E}'_L(\theta,\mathbf{c}_n,\alpha_T)\,\mathbf{R}_{nT,1}\mathcal{E}_L(\theta,\mathbf{c}_n,\alpha_T) \\ \vdots \\ \mathcal{E}'_L(\theta,\mathbf{c}_n,\alpha_T)\,\mathbf{R}_{nT,m}\mathcal{E}_L(\theta,\mathbf{c}_n,\alpha_T) \end{bmatrix}$ where $\mathbf{R}_{nT,l} = diag(R_{n1,l},\cdots,R_{nT,l})$ is a $L\times L$ block diagonal matrix for $l = 1,\cdots,m$. Let $S_L(\theta,\mathbf{c}_n,\alpha_T) = \bar{g}'_L(\theta,\mathbf{c}_n,\alpha_T)\,a'_L a_L \bar{g}_L(\theta,\mathbf{c}_n,\alpha_T)$ for each $(\theta,\mathbf{c}_n,\alpha_T)$ where $\bar{g}_L(\theta,\mathbf{c}_n,\alpha_T) = \begin{bmatrix} \bar{g}^{\mathbf{Q}}_L(\theta,\mathbf{c}_n,\alpha_T) \\ \bar{g}^{\mathbf{L}}_L(\theta,\mathbf{c}_n,\alpha_T) \end{bmatrix}$.

Our next issue is to deal with the incidental parameters $\mathbf{c}_{n0}$ and $\alpha_{T,0}$. We consider the direct estimation approach using the concentrated statistical objective function. Observe that $\epsilon_{l(i,t),L}(\theta,c_i,\alpha_t)$ is nonlinear in θ while it is linear in c_i and α_t. To concentrate out $\mathbf{c}_{n0}$ and $\alpha_{T,0}$ from $S_L(\theta,\mathbf{c}_n,\alpha_T)$, let $J_n = I_n - \frac{1}{n}l_n l'_n$ and $J_T =$

[110] In conventional spatial econometrics models (e.g., linear SAR models), we can observe model's reduced form and derive its (best) moment conditions (functions of exogenous variables and unknown parameters). Since we do not exactly derive a reduced form of our model, it is hard to obtain the best moment conditions. The ideal instruments might be a highly nonlinear function of exogenous variables in ℓ_{t-1} and unknown structural parameters.

$I_T - \frac{1}{T}l_T l_T'$ be the demeaning (orthogonal) operators. Note that J_n eliminates the time fixed effects $\alpha_{T,0}$ while J_T involves in deleting the individual specific effects $\mathbf{c}_{n0}$. As a result, the optimization of the GMM objective function is on the fixed number of parameters.

For each $l(i,t) \in D_L$ and $\theta \in \Theta$, hence, we define

$$\begin{aligned}
\epsilon_{l(i,t),L}(\theta) &= [(e_{ni}' - \lambda w_{t,i.}) Y_{nt} - \gamma y_{i,t-1} - x_{it}\beta] + \delta\gamma \left(y_{it} - y_{i,t+1}^e(\theta)\right) \\
&\quad -\delta \left[\left(\sum_{p=1}^{P} \lambda \psi_p \left[M_{n,t+1,p}(\theta)\right]_{ii} + \lambda \left[N_{n,t+1}(\theta)\right]_{ii} \right) y_{i,t+1}^e(\theta) \right] \\
&\quad -\delta^2 \left[\nabla V_{n,t+2}^e(\theta)\right]_i ,
\end{aligned}$$

and $\mathcal{E}_L(\theta) = (\mathcal{E}_{n1}'(\theta), \cdots, \mathcal{E}_{nT}'(\theta))'$ with $\mathcal{E}_{nt}(\theta) = (\epsilon_{1t}(\theta), \cdots, \epsilon_{nt}(\theta))'$ for each $t = 1, \cdots, T$. The direct estimation approach is to use the demeaned data by the demeaning operators J_T and J_n. That is, we rely on that $(J_T \otimes J_n)\mathcal{E}_L(\theta, \mathbf{c}_n, \alpha_T) = (J_T \otimes J_n)\mathcal{E}_L(\theta)$ for each $\theta \in \Theta$.

Then, the concentrated GMM objective function is obtained using J_n and J_T:

$$S_L^c(\theta) = \bar{g}_L^{c\prime}(\theta) a_L' a_L \bar{g}_L^c(\theta) \, where \bar{g}_L^c(\theta) = \begin{bmatrix} \bar{g}_L^{\mathbf{Q},c}(\theta) \\ \bar{g}_L^{\mathbf{L},c}(\theta) \end{bmatrix}, \tag{3.29}$$

$\bar{g}_L^{\mathbf{Q},c}(\theta) = \frac{1}{L}\begin{bmatrix} \mathcal{E}_L'(\theta)(J_T \otimes J_n)\mathbf{R}_{nT,1}(J_T \otimes J_n)\mathcal{E}_L(\theta) \\ \vdots \\ \mathcal{E}_L'(\theta)(J_T \otimes J_n)\mathbf{R}_{nT,m}(J_T \otimes J_n)\mathcal{E}_L(\theta) \end{bmatrix}$, and
$\bar{g}_L^{\mathbf{L},c}(\theta) = \frac{1}{L}\mathbf{q}_L'(J_T \otimes J_n)\mathcal{E}_L(\theta)$. Hence,

$$\hat{\theta}_L = \arg\min_{\theta \in \Theta} S_L^c(\theta). \tag{3.30}$$

In practice, we conduct the iterative estimation procedure (nested fixed-point algorithm). As individual and time dummies are linear parameters, they can be identified like regression coefficients after the main structural parameters θ_0 are recovered. If the model is over-identified, we can check whether the moment conditions match the data set by using the J test statistic. The J test statistic can capture (i) whether the Euler equation system fits the data well and (ii) whether the conditions of the quadratic moments are valid or not.

Note that the GMM estimator $\hat{\theta}_L$ satisfies $\left[\frac{\partial \bar{g}_L^c(\hat{\theta}_L)}{\partial \theta'}\right]' a_L' a_L \bar{g}_L^c\left(\hat{\theta}_L\right) = \mathbf{0}_{(2+K+P)\times 1}$, so we observe $\frac{\partial \bar{g}_L^c(\theta)}{\partial \theta'}$ and $\bar{g}_L^c(\theta)$ are the key statistics for asymptotic analyses. To specify components of $\frac{\partial \bar{g}_L^c(\theta)}{\partial \theta'}$, define $A_{n\lambda}(\theta) = \frac{\partial A_n(\theta)}{\partial \lambda}$, $A_{n\gamma}(\theta) = \frac{\partial A_n(\theta)}{\partial \gamma}$, $A_{n\psi_p}(\theta) = \frac{\partial A_n(\theta)}{\partial \psi_p}$ for $p = 1, \cdots, P$, $A_{n\beta_k}(\theta) = \frac{\partial A_n(\theta)}{\partial \beta_k}$ for $k = 1, \cdots, K$ where $A_n(\theta)$ is a vector or matrix relying on θ. And, note that $A_{n\lambda}$, $A_{n\gamma}$, $A_{n\psi_p}$ and $A_{n\beta_k}$ represent respectively $A_{n\lambda} = \frac{\partial A_n(\theta)}{\partial \lambda}|_{\theta=\theta_0}$, $A_{n\gamma} = \frac{\partial A_n(\theta)}{\partial \gamma}|_{\theta=\theta_0}$, $A_{n\psi_p} = \frac{\partial A_n(\theta)}{\partial \psi_p}|_{\theta=\theta_0}$ for $p = 1, \cdots, P$ and $A_{n\beta_k} = \frac{\partial A_n(\theta)}{\partial \beta_k}|_{\theta=\theta_0}$ for $k = 1, \cdots, K$.

3.4.3 Asymtotic analysis

In this subsection, we establish consistency and derive the asymptotic distribution of the GMM estimator. Note that the dependent variables of our model are serially and spatially correlated, but those variables may not be a linear function of disturbances. Following Jenish and Prucha (2012), consistency of $\hat{\theta}_L$ will be established based on the near-epoch dependence (NED) of spatial-time processes by controlling dependencies among them. Here is the definition of the spatial-time NED based on $\mathcal{F}_{l(i,t),L}(s)$.

Definition 3.4.1 (L_p near-epoch dependence) *Consider two random fields,* $Y = \left\{y_{l(i,t),L} : l(i,t) \in D_L, L \geq 1\right\}$ *with* $\left\|y_{l(i,t),L}\right\|_{L_p} < \infty$*,* $p \geq 1$ *and* $\varsigma = \left\{\varsigma_{l(i,t),L} : l(i,t) \in D, L \geq 1\right\}$*, and an array of finite positive constants,* $d = \left\{d_{l(i,t),L} : l(i,t) \in D_L, L \geq 1\right\}$*. Assume* $|D_L| \to \infty$ *as* $L \to \infty$ *where* $|D_L|$ *denotes* *a cardinality of* D_L*. Then,* Y *is* L_p*-NED on* ς *if*

$$\left\| y_{l(i,t),L} - E\left(y_{l(i,t),L} | \mathcal{F}_{l(i,t),L}(s) \right) \right\|_{L_p} \leq d_{l(i,t),L}\psi(s) \tag{3.31}$$

for a sequence $\psi(s) \geq 0$ *such that* $\psi(s) \to 0$ *as* $s \to \infty$*. If* $\sup_{L,l(i,t)\in D_L} d_{l(i,t),L} < \infty$*,* Y *is uniformly* L_p*-NED on* ς*.*

Note that the NED is a property of a mapping from ς to Y instead of a stochastic feature of Y itself. Choose $l(i,t)$ for some $i \in \{1, \cdots, n\}$ and $t \in \{1, \cdots, T\}$ for interpretations. First, $E\left(y_{l(i,t),L} | \mathcal{F}_{l(i,t),L}(s) \right)$ means the approximated $y_{l(i,t),L}$ using "some near random elements" $\varsigma_{l(j,t'),L}$'s such that $\left\| l(i,t) - l(j,t') \right\|_{\infty} \leq s$. The random field ς contains "input processes", and it consists of all $\varsigma_{l(j,t'),L}$'s indexed by D. Note that the NED concept relates two random fields, as an example, Y is approximately mixing if the input process ς is mixing. Second, this approximation error is bounded by some constant $d_{l(i,t),L}$ and a decreasing function $\psi(s)$ in s. The first term $d_{l(i,t),L}$ depends on a specific unit $l(i,t)$, which it means that this term controls heterogeneity of unit $l(i,t)$. By the Minkowski and the (conditional) Jensen's inequalities,

$$\left\| y_{l(i,t),L} - E\left(y_{l(i,t),L} | \mathcal{F}_{l(i,t),L}(s) \right) \right\|_{L_p} \leq \left\| y_{l(i,t),L} \right\|_{L_p} + \left\| E\left(y_{l(i,t),L} | \mathcal{F}_{l(i,t),L}(s) \right) \right\|_{L_p} \leq 2 \left\| y_{l(i,t),L} \right\|_{L_p}. \tag{3.32}$$

Then, we can choose $d_{l(i,t),L} \leq 2\left\| y_{l(i,t),L} \right\|_{L_p}$, which it leads to $0 \leq \psi(s) \leq 1$. By introducing some regularity conditions restricting $\left\| y_{l(i,t),L} \right\|_{L_p}$, hence, we focus on the uniform $L_p - NED$ concept. The second term $\psi(s)$ controls weak dependencies on adjacent units, $l(j,t')$'s. This device describes the intensity of dependence, so it should be negligible if two units $l(i,t)$ and $l(j,t')$ are sufficiently far. By the Lyapunov inequality, if Y is L_p-NED on ς, then it is also $L_q - NED$ on ς with the same scaling factor $d_{l(i,t),L}$ and coefficient $\psi(s)$ for $q \leq p$. Xu and Lee (2018) provide a review of related asymptotic techniques.

To derive the asymptotic properties of the GMM estimator, we establish some laws of large numbers (LLN) and central limit theorem (CLT). For this, we produce some regularity assumptions.

Assumption 3.4.2 *For all i and t, $\epsilon_{it} \sim i.i.d.\,(0, \sigma_0^2)$, and $\sup_{i,t} E\left|\epsilon_{it}\right|^{4+\eta_\epsilon} < \infty$ for some $\eta_\epsilon > 0$.*

Assumption 3.4.3 *The parameter space Θ of θ is compact. $\theta_0 \in int\,(\Theta)$.*

Assumption 3.4.4 *$\{X_{nt}\}_{t=0}^T$, $\{\alpha_{t0}\}_{t=1}^T$ and $\mathbf{c}_{n0}$ are conditional upon nonstochastic values.*
$\sup_{n,T} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left|x_{it,k}\right|^{2+\eta} < \infty$ *for all k,* $\sup_T \frac{1}{T} \sum_{t=1}^T \left|\alpha_{t0}\right|^{2+\eta} < \infty$ *and*
$\sup_n \frac{1}{n} \sum_{i=1}^n \left|c_{i,0}\right|^{2+\eta} < \infty$ *for some $\eta > 0$.*

Assumption 3.4.5 *For each t and $i \neq j$, $w_{t,ij} = h(z_{it}, z_{jt}) \cdot h_d\,(d_{ij}) \cdot \mathbf{1}\{d_{ij} \leq d_c\}$ where $h\,(\cdot)$ and $h_d\,(\cdot)$ are uniformly bounded nonnegative functions and $0 < d_c < \infty$.*
$\max\left\{\sup_{n,t} \left\|W_{nt}\right\|_\infty, \sup_{n,t} \left\|W_{nt}\right\|_1\right\} \leq c_w$ *for some $c_w > 0$.*

Assumption 3.4.6 *(i) For $\theta \in \Theta$, $A_n^e\,(\theta)$, $B_n^e\,(\theta)$ and $M_{n,t+1}\,(\theta)$ are uniformly bounded in both row and column sum norms a.e., uniformly in $\theta \in \Theta$. All elements of $C_n^e\,(\theta)$ are uniformly bounded a.e., uniformly in $\theta \in \Theta$. A matrix norm of $A_n^e\,(\theta)$ is bounded by $\tau \in [0, 1)$ a.e.*

(ii) For $\theta \in int\,(\Theta)$, the first, and second derivatives of $A_n^e\,(\theta)$ and $B_n^e\,(\theta)$ with respect to θ are uniformly bounded in both row and column sum norms a.e., uniformly in $\theta \in \Theta$. All elements of the first, and second derivatives of $C_n^e\,(\theta)$ for $\theta \in int\,(\Theta)$ are uniformly bounded a.e., uniformly in $\theta \in \Theta$.

(iii) Let Θ_1 be the parameter space for λ. For any t, $S_n(W_{nt}, \lambda)$ is nonsingular and uniformly bounded in both row and column sum norms, uniformly in $\lambda \in int(\Theta_1)$.

Assumption 3.4.7 *(i) For any t and $i \neq j$,*

$$|h(z_{it}, z_{jt}) - h(\hat{z}_{it}, \hat{z}_{jt})| \leq c_h (|z_{it} - \hat{z}_{it}| + |z_{jt} - \hat{z}_{jt}|) \quad (3.33)$$

for two pairs (z_{it}, z_{jt}) and $(\hat{z}_{it}, \hat{z}_{jt})$ where $c_h > 0$.

(ii) For each $i \neq j$ and p, note that $m_{t,p,ij} = \frac{\partial h(z_{it}, z_{jt})}{\partial z_{it,p}} \cdot h_d(d_{ij}) \cdot \mathbf{1}\{d_{ij} \leq d_c\}$.

Let $\tilde{m}_p(z_{it}, z_{jt}) = m_{t,p,ij}$. Assume $\bar{m} = \max_{p=1,\cdots,P} \sup_{i,j,n,t} |\tilde{m}_p(z_{it}, z_{jt})| < \infty$ and for any t, p and $i \neq j$,

$$|\tilde{m}_p(z_{it}, z_{jt}) - \tilde{m}_p(\hat{z}_{it}, \hat{z}_{jt})| \leq c_m (|z_{it} - \hat{z}_{it}| + |z_{jt} - \hat{z}_{jt}|) \quad (3.34)$$

for two pairs (z_{it}, z_{jt}) and $(\hat{z}_{it}, \hat{z}_{jt})$ where $c_m > 0$.

Assumption 3.4.8 *T goes to infinity and n is an increasing function of T.*

Assumption 3.4.2 states regularity assumptions for ϵ_{it}. We consider *i.i.d.* disturbances across i and t for simplicity. Based on Assumption 3.4.2 with assuming that the IV $\mathbf{q}_{nt}$ is uniformly L_4-bounded, the moment functions $g^c_{l(i,t),L}(\theta)$ would be p-dominated on Θ for $p = 2$. This helps to establish the uniform law of large numbers (ULLN).[111] Also, the higher than the fourth moment assumption for ϵ_{it} is needed to apply a central limit theorem for a LQ form.[112] Compact parameter space in Assumption 3.4.3 is for a well-defined nonlinear extremum estimator (Chapter 4 in Amemiya

[111]This result yields that $\epsilon_{l(i,t),L}(\theta)$ becomes $L_{4+\eta_\epsilon}$-bounded uniformly in i, t, L and uniformly $\theta \in \Theta$. It means that the residual evaluated at $\theta \in \Theta$ ($\epsilon_{l(i,t),L}(\theta)$) has the same stochastic boundedness condition as that of $\epsilon_{l(i,t),L}$.

[112]Related articles are Kelejian and Prucha (2001), Yu et al. (2008), Qu et al. (2017), and Kuersteiner and Prucha (2018). The detailed discussion about the LLN and CLT can be found in Appendix B.

(1985)). To achieve simplicity of asymptotic analyses, we employ the conditioning argument (Assumption 3.4.4) and the boundedness assumption for empirical moments of explanatory and dummy variables. Assumption 3.4.5 is a standard assumption on W_{nt} in the spatial econometric literature.[113] By Assumption 3.4.6 (i), we impose manageable dependence between $Y_{n,t-1}$ and Y_{nt} for each $\theta \in \Theta$ since $A_n^{e\prime}(\theta)$ formulates $\frac{\partial Y_{nt}}{\partial Y_{n,t-1}^{\prime}}$ at $\theta \in \Theta$. For asymptotic analyses, Assumption 3.4.6 (i) supposes that $A_n^e(\theta)$, $B_n^e(\theta)$, $C_n^e(\theta)$ and $M_{n,t+1}(\theta)$ transforming Y_{nt} make manageable changes at $\theta \in \Theta$.[114] This device makes the same orders of moments of $\epsilon_{l(i,t),L}$ and $\epsilon_{l(i,t),L}(\theta)$ for each $\theta \in \Theta$ (i.e., $\epsilon_{l(i,t),L}$ and $\epsilon_{l(i,t),L}(\theta)$ are both $L_{4+\eta_\epsilon}$-bounded). Assumption 3.4.6 (ii) is for having an identification condition and applying the LLN to $\frac{\partial \bar{g}_L^c(\theta)}{\partial \theta^{\prime}}$. Assumption 3.4.6 (iii) characterizes the magnitude of contemporaneous spatial influences. This assumption helps to establish that $y_{l(i,t),L}$ is a spatial-NED process given $Y_{n,t-1}$.

Assumption 3.4.7 (i) is the Lipschitz condition on the formation function $h(\cdot)$ so that the NED property for $w_{t,ij}$ can be obtained from z_{it} and z_{jt}. Assumption 3.4.7 (ii) also imposes the boundedness and Lipschitz conditions on the marginal effect of $y_{i,t-1}$ on $w_{ij,t}$ via $z_{it,p}$. By Assumption 3.4.7, that is, values of $h(z_{it}, z_{jt})$ and $\tilde{m}_p(z_{it}, z_{jt})$ have manageable changes when their arguments, (z_{it}, z_{jt}), are changed. Assumption 3.4.8 means that large n and T cases would be considered for our asymptotic analysis to deal with the incidental parameter problem. Large n is for estimating the time

[113]In detail, we require that the formation functions $h(\cdot)$ and $h_d(\cdot)$ are uniformly bounded. Uniform boundedness of $h(\cdot)$ can be achieved by limiting a maximum intensity of influences: for example, $h(z_{it}, z_{jt}) = \frac{1}{\|z_{it}-z_{jt}\|}$ if $\|z_{it} - z_{jt}\| > c_z$ and $h(z_{it}, z_{jt}) = \frac{1}{c_z}$ otherwise for some $c_z > 0$.

[114]For example, note that $[M_{n,t+1,p}(\theta)]_{ii} = m_{t+1,p,i.}Y_{n,t+1}^e(\theta)$. Note that $Y_{n,t+1}^e(\theta)$ can be an unbounded stochastic component. Hence, Assumption 3.4.6 (i) implies the following specification:

$$[M_{n,t+1,p}(\theta)]_{ii} = \begin{cases} sgn\left(m_{t+1,p,i.}Y_{n,t+1}^e(\theta)\right) c_M if \left|m_{t+1,p,i.}Y_{n,t+1}^e(\theta)\right| > c_M \\ m_{t+1,p,i.}Y_{n,t+1}^e(\theta)\, otherwise \end{cases} \tag{3.35}$$

for some constant $c_M > 0$, which can be selected by a researcher.

dummies $\alpha_{T,0}$ while estimating the individual specific parameters $\mathbf{c}_{n0}$ is required to have large T. Also, the effect of initial values on estimation becomes ignorable under large T.

Under the provided regularity conditions, we establish the large sample properties of the GMM estimator $\hat{\theta}_L$. For consistency of $\hat{\theta}_L$, we require (i) the ULLN, (ii) stochastic equicontinuity, and (iii) identification uniqueness as sufficient conditions. For consistency of $\hat{\theta}_L$, the first issue is to establish the LLN for $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$, which is the main stochastic component of $g^c_{l(i,t),L}(\theta)$. Let $\varsigma_{l(i,t),L}$ be a measurable function of $\epsilon_{l(i,t),L}$, Z^*_L, X_L, $\mathbf{c}_{n0}$, and $\alpha_{T,0}$ where $Z^*_L = (Z^{*\prime}_{n1}, \cdots, Z^{*\prime}_{nT})$, and $X_L = (X'_{n1}, \cdots, X'_{nT})'$, i.e, $\varsigma_{l(i,t),L} = \varsigma\left(\epsilon_{l(i,t),L}, Z^*_L, X_L, \mathbf{c}_{n0}, \alpha_{T,0}\right)$ for each $l(i,t) \in D$. Then, we define the baseline σ-field for approximation:

$$\mathcal{F}_{l(i,t),L}(s) = \sigma\left(\varsigma_{l(j,t')} : l(j,t') \in D, \|l(i,t) - l(j,t')\|_\infty \leq s\right). \tag{3.36}$$

Hence, the sequence $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ forms a random field with respect to another random field $\left\{\varsigma_{l(i,t),L}\right\}_{l(i,t)\in D}$. So, we discuss the NED properties of $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ based on $\mathcal{F}_{l(i,t),L}(s)$. Proposition B.2.5 says that $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ is uniformly $L_{4+\eta_\epsilon}$-bounded and is uniformly L_2-NED on $\left\{\varsigma_{l(i,t),L}\right\}_{l(i,t)\in D}$. Then, we can apply the LLN to $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ in the sense of L_1-norm (Proposition B.2.6). The idea of showing this is to verify two steps: (i) the approximated $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ (which is a function of $\mathcal{F}_{l(i,t),L}(s)$) satisfies the LLN, and then, (ii) the distance between the actual sequence and the approximated one can be closed to zero if distance s grows. A main part of (ii) is to have a nonlinear moving average representation of $y_{l(i,t),L}$ using $\left\{\varsigma_{l(i,t),L}\right\}_{l(i,t)\in D}$. This LLN helps to establish convergence of the sample moment function $\bar{g}^c_L(\theta) = \frac{1}{L}\sum_{l(i,t)\in D_L} g^c_{l(i,t),L}(\theta)$ to its expected value for each $\theta \in \Theta$. To hold $\frac{1}{L}\left[\bar{g}^c_L(\theta) - E\left(\bar{g}^c_L(\theta)\right)\right] \to_p 0$ for each $\theta \in \Theta$, Lemmas B.2.12 and B.2.13 verify respectively $\sup_{L,l\in D_L} E\left|\bar{g}_L^c(\theta)\right|^2 < \infty$ and $\bar{g}_L^c(\theta)$ is uniformly L_1-NED on ς for each $\theta \in \Theta$. This is a pointwise LLN, so for arbitrary finite points θ's in Θ $\bar{g}_L^c(\theta)$ converges its expected value by the same reason of Proposition B.2.6. Since Θ is compact, it can be covered by finite subcovers. To extend the pointwise LLN to the ULLN, $\bar{g}_L^c(\theta)$ satisfies the Lipschitz continuity condition in parameter, which gives enough smoothness of $\bar{g}_L^c(\theta)$ (in θ). In Appendix B, Lemma B.2.14 verifies this feature and gives (i) the uniform convergence $\sup_{\theta\in\Theta}\left[\bar{g}_L^c(\theta) - E\left(\bar{g}_L^c(\theta)\right)\right] \to_p 0$, and (ii) uniformly equicontinuity of $E\left(\bar{g}_L^c(\theta)\right)$.

To obtain consistency of $\hat{\theta}_L$, the last requirement is the identification uniqueness of θ_0. We only consider the linear IV moments as the source of identification since (i) they come from the underlying model assumption (i.e., population orthogonality condition) and (ii) the derivation of identification conditions is relatively simple. To have the identified system, we need to have a unique solution so that for all $l(i,t) \in D_L$ $E\left[g^{\mathbf{L}}_{l(i,t),L}(\theta, c_i, \alpha_t)\right] = \mathbf{0}_{q\times 1}$ if and only if $(\theta, c_i, \alpha_t) = (\theta_0, c_{i,0}, \alpha_{t,0})$. Since we are interested in the structural parameter θ_0, the equation $\text{plim}_{L\to\infty}\bar{g}_L^{\mathbf{L},c}(\theta) = \mathbf{0}_{q\times 1}$ is required to have a unique solution θ_0. For each $l(i,t) \in D_L$ and $\theta \in \Theta$, define the following generated regressors: $L_{\lambda,it}(\theta) = \frac{\partial\epsilon_{l(i,t),L}(\theta)}{\partial\lambda}$, $L_{\gamma,it}(\theta) = \frac{\partial\epsilon_{l(i,t),L}(\theta)}{\partial\gamma}$, $L_{\psi_p,it}(\theta) = \frac{\partial\epsilon_{l(i,t),L}(\theta)}{\partial\psi_p}$ for $p = 1, \cdots, P$, and $L_{\beta_k,it}(\theta) = \frac{\partial\epsilon_{l(i,t),L}(\theta)}{\partial\beta_k}$ for $k = 1, \cdots, K$. For each $\theta \in \Theta$, let $L_{\psi,it}(\theta) = \left[L_{\psi_1,it}(\theta), \cdots, L_{\psi_P,it}(\theta)\right]'$ ($P \times 1$ vector), $L_{\beta,it}(\theta) = \left[L_{\beta_1,it}(\theta), \cdots, L_{\beta_K,it}(\theta)\right]'$ ($K\times 1$ vector), $L_{it}(\theta) = \left[L_{\lambda,it}(\theta), L_{\lambda,it}(\theta), L'_{\psi,it}(\theta), L'_{\beta,it}(\theta)\right]$ for $\theta \in \Theta$ ($(2+P+K)$-dimensional row vector), $\mathbf{L}_{nt}(\theta) = \left(L'_{1t}(\theta), \cdots, L'_{nt}(\theta)\right)'$, and $\mathbf{L}_L(\theta) = \left(\mathbf{L}'_{n1}(\theta), \cdots, \mathbf{L}'_{nT}(\theta)\right)'$. Since $\text{plim}_{L\to\infty}\frac{1}{L}\mathbf{q}'_L(J_T \otimes J_n)\mathcal{E}_L = \mathbf{0}_{q\times 1}$ by Lemma B.2.16, the unique solution of $\text{plim}_{L\to\infty}\bar{g}_L^{\mathbf{L},c}(\theta) = \mathbf{0}_{q\times 1}$ at θ_0 needs that θ_0 is a unique solution of

$$plim_{L\to\infty}\frac{1}{L}\mathbf{q}_L'\left(J_T\otimes J_n\right)\mathbf{L}_L\left(\bar{\theta}\right)\cdot\left(\theta_0-\theta\right)=0 \tag{3.37}$$

where $\bar{\theta}$ lies between θ_0 and θ. Then, if $\text{plim}_{L\to\infty}\frac{1}{L}\mathbf{q}_L'\left(J_T\otimes J_n\right)\mathbf{L}_L\left(\theta\right)$ has the full column rank $2+P+K$ around θ_0, we have a sufficient condition for identification. The two assumptions below state regularity conditions for linear and quadratic moments.

Assumption 3.4.9 *(i) $\mathbf{q}_{nt}$ is predetermined at time t (i.e., $E\left(\mathbf{q}_{nt}|\ell_{t-1}\right)=\mathbf{q}_{nt}$). And, its column dimension of $\mathbf{q}_{nt}$ is fixed ($=q\geq 2+P+K$) for all n and t.*

(ii) For some $\eta_q>0$, $\sup_{i,t,L}E\left|q_{l(i,t),L}\right|^{4+\eta_q}<\infty$ and $\sup_{i,t,n,T}\sum_{j=1}^{n}\sum_{t'=1}^{T}E\left|q_{l(i,t),L}q_{l(j,t'),L}'\right|<\infty$.

(iii) $\left\{q_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ is uniformly L_2-NED on ς.

(iv) $plim_{L\to\infty}\frac{1}{L}\mathbf{q}_L'\left(J_T\otimes J_n\right)\mathbf{L}_L\left(\theta\right)$ is of full column rank $2+P+K$ for $\theta\in\mathcal{N}\left(\theta_0\right)$ where $\mathcal{N}\left(\theta_0\right)$ denotes some neighborhood of θ_0.

Assumption 3.4.10 *(i) For each t and l, all diagonal elements of $R_{nt,l}$ are zero so that $E\left(\frac{1}{L}\mathcal{E}_L'\mathbf{R}_{L,l}\mathcal{E}_L\right)=0$ and all entries in $R_{nt,l}$ are measurable functions of $\left\{\varsigma_{l(i,t),L}^*\right\}_{l(i,t)\in D}$ where $\varsigma_{l(i,t),L}^*=\varsigma^*\left(z_{l(i,t),L}^*,X_L,\mathbf{c}_{n0},\alpha_{T,0}\right)$.*

(ii) The matrices $R_{nt,l}$'s are uniformly bounded in both row and column sums in absolute values.

(iii) For any i, t and l, $\left[R_{nt,l}\right]_{ij}=0$ if $\left\|l\left(i,0\right)-l\left(j,0\right)\right\|_\infty>ad_c$ for some $a\in\mathbf{N}$.

Assumption 3.4.9 (i) implies that for each t $E\left(\mathbf{q}_{nt}'\mathcal{E}_{nt}\right)=E\left(\mathbf{q}_{nt}'E\left(\mathcal{E}_{nt}|\ell_{t-1}\right)\right)=\mathbf{0}_{q\times 1}$ by the law of iterated expectations. Assuming the existence of the (higher than) fourth moment for $q_{l(i,t),L}$ is for the dominance condition of $g_{l(i,t),L}^{\mathbf{L},c}\left(\theta\right)$ for each $\theta\in\Theta$. Via Assumption 3.4.9 (ii), we impose the spatial-time stability condition for IVs. By Assumption 3.4.9 (iii), we achieve that the sequence of sample moment functions

$\left\{g^c_{l(i,t),L}(\theta)\right\}_{l(i,t)\in D_L}$ is uniformly L_1-NED on ς, so the pointwise LLN can be applied to $\frac{1}{L}\sum_{l(i,t)\in D_L} g^c_{l(i,t),L}(\theta)$. Assumption 3.4.9 (iv) is a sufficient condition that θ_0 is a unique solution to equation (3.37).

Assumption 3.4.10 (i) means that $R_{nt,1},\cdots,R_{nt,m}$ are not relevant to $\left\{\epsilon_{l(i,t),L}\right\}_{l(i,t)\in D_L}$, but can be affected by the exogenous part of Z_{nt}, X_{nt}, $\mathbf{c}_{n0}$, and $\alpha_{t,0}$. This implies that for each $l=1,\cdots,m$ $E\left(\mathcal{E}'_L\left(J_T\otimes J_n\right)\mathbf{R}_{L,l}\left(J_T\otimes J_n\right)\mathcal{E}_L\right)=\sigma_0^2 tr\left(E\left(\mathbf{R}_{L,l}\right)\left(J_T\otimes J_n\right)\right)$. As a quadratic moment matrix, we can use the spatial network matrix W_n (if available), and approximated functions for $W_{n,t-1}$ based on the part of strictly exogenous components $\left\{\varsigma^*_{l(i,t),L}\right\}_{l(i,t)\in D_L}$. For example, if $\check{W}_{n,t-1}$ is a projected (approximated) $W_{n,t-1}$ on the space generated by $\left\{\varsigma^*_{l(i,t),L}\right\}_{l(i,t)\in D_L}$, we can use $\check{W}_{n,t-1}$, $\check{W}'_{n,t-1}$, and $\check{W}'_{n,t-1}\check{W}_{n,t-1}-tr\left(\check{W}'_{n,t-1}\check{W}_{n,t-1}\right)$ as a quadratic moment matrix. For the relevant issue, refer to Kelejian and Piras (2014). We do not consider that a quadratic moment matrix depends on unknown parameters.[115] Assumption 3.4.10 (ii) restricts the magnitude of the quadratic moment matrices. By imposing Assumption 3.4.10 (iii), the number of nonzero elements in $R_{nt,l}$ is finite even for large n (i.e, sparse $R_{nt,l}$'s).

Now we obtain consistency of $\hat{\theta}_L$, which is implied by (i) uniform convergence of $\left[S^c_L(\theta)-E\left(S^c_L(\theta)\right)\right]$ in $\theta\in\Theta$, (ii) uniform equicontinuity of $\left\{E\left(S^c_L(\theta)\right)\right\}$ on Θ, and (iii) identification uniqueness. Note that the NL2SE can also achieve consistency since that approach shares those conditions.

Theorem 3.4.2 *Assume Assumptions 3.4.1 - 3.4.10 hold. Then,* $\hat{\theta}_L\to_p\theta_0$ *as* $L\to\infty$.

For statistical inferences, the next step is to derive the asymptotic distribution of $\hat{\theta}_L$. Since $\text{plim}_{L\to\infty}a_L=a_0$, the random function $\sqrt{L}\cdot a_0\bar{g}^c_L(\theta_0)$ mainly characterizes

[115] Since we do not derive the feasible reduced form, it is difficult to analyze the asymptotic impacts of the estimator obtained by the first step on the quadratic moment matrices.

the asymptotic distribution of $\hat{\theta}_L$ with $\frac{\partial \bar{g}_L^c(\hat{\theta})}{\partial \theta'} = \frac{1}{L}\sum_{l(i,t)\in D_L} \frac{\partial g_{l(i,t),L}^c(\hat{\theta})}{\partial \theta'}$. By Theorem 3.4.2 and the continuous mapping theorem, we have $\frac{\partial \bar{g}_L^c(\hat{\theta})}{\partial \theta'} - \frac{\partial \bar{g}_L^c(\theta_0)}{\partial \theta'} = o_p(1)$, which implies that $\frac{\partial \bar{g}_L^c(\theta_0)}{\partial \theta'}$ can be considered instead of $\frac{\partial \bar{g}_L^c(\hat{\theta})}{\partial \theta'}$ in the asymptotic analysis. First, by Propositions B.2.9, B.2.10 and B.2.11 in Appendix B, the random field $\left\{\frac{\partial g_{l(i,t),L}^c(\theta_0)}{\partial \theta'}\right\}_{l(i,t)\in D_L}$ is uniformly L_1-NED on ς. Then, we can apply the LLN to $\frac{\partial \bar{g}_L^c(\theta_0)}{\partial \theta'}$. We denote $G_L = E\left(\frac{\partial \bar{g}_L^c(\theta_0)}{\partial \theta'}\right)$ and G_0 as its limit.

Second, consider $\sqrt{L}\cdot a_0 \bar{g}_L^c(\theta_0)$, which is a linear combination of linear and quadratic moment conditions. The linear moment part $\frac{1}{\sqrt{L}}\mathbf{q}_L'(J_T \otimes J_n)\mathcal{E}_L$ can be composed by the mean zero part $\frac{1}{\sqrt{L}}\mathbf{q}_L'(I_T \otimes J_n)\mathcal{E}_L$ and the asymptotic bias part $-\frac{1}{T\sqrt{L}}\mathbf{q}_L'(l_T l_T' \otimes J_n)\mathcal{E}_L$. By Lemma B.2.16, we have $-\frac{1}{T\sqrt{L}}E(\mathbf{q}_L'(l_T l_T' \otimes J_n)\mathcal{E}_L) = O\left(\sqrt{\frac{n}{T}}\right)$ and
$\frac{1}{T\sqrt{L}}(\mathbf{q}_L'(l_T l_T' \otimes J_n)\mathcal{E}_L - E(\mathbf{q}_L'(l_T l_T' \otimes J_n)\mathcal{E}_L)) = O_p\left(\frac{1}{\sqrt{T}}\right)$. Then, define
$b_L^{\mathbf{L}}(\theta_0, \sigma_0^2) = \frac{1}{n}\sum_{t=1}^T E(\bar{\mathbf{q}}_{nT}' J_n \mathcal{E}_{nt}) = O(1)$. Hence,

$$\frac{1}{\sqrt{L}}\mathbf{q}_L'(J_T \otimes J_n)\mathcal{E}_L = \frac{1}{\sqrt{L}}\mathbf{q}_L'(I_T \otimes J_n)\mathcal{E}_L - \sqrt{\frac{n}{T}}b_L^{\mathbf{L}}\left(\theta_0, \sigma_0^2\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \quad (3.38)$$

Observe the quadratic moment part: for $l = 1, \cdots, m$

$$\begin{aligned} & \frac{1}{\sqrt{L}}\mathcal{E}_L'(J_T \otimes J_n)\mathbf{R}_{L,l}(J_T \otimes J_n)\mathcal{E}_L \\ = & \frac{1}{\sqrt{L}}\left[\mathcal{E}_L'(J_T \otimes J_n)\mathbf{R}_{L,l}(J_T \otimes J_n)\mathcal{E}_L - \sigma_0^2 tr(\mathbf{R}_{L,l}(J_T \otimes J_n))\right] \\ & + \frac{\sigma_0^2}{\sqrt{L}} tr(\mathbf{R}_{L,l}(J_T \otimes J_n)). \end{aligned}$$

By Lemma B.2.17, we have $\frac{\sigma_0^2}{\sqrt{L}} tr(\mathbf{R}_{L,l}(J_T \otimes J_n)) = -\sqrt{\frac{T}{n}}b_{L,l}^{\mathbf{Q}}(\sigma_0^2) + O_p\left(\frac{1}{\sqrt{L}}\right)$ where $b_{L,l}^{\mathbf{Q}}(\sigma_0^2) = \frac{\sigma_0^2}{L}\sum_{t=1}^T\sum_{i=1}^n\sum_{j=1}^n [R_{nt,l}]_{ij}$ for $l = 1, \cdots, m$ and $b_{L,l}^{\mathbf{Q}}(\sigma_0^2) = O_p(1)$ by Assumption 3.4.10 (ii). Let $a_L = \left(a_L^{(1)}, \cdots, a_L^{(m)}, a_L^{(q)}\right)$ and a_0 =plim$_{L\to\infty} a_L$ where $a_0 = \left(a_0^{(1)}, \cdots, a_0^{(m)}, a_0^{(q)}\right)$. By applying the asymptotic equivalence, we can consider the $k_a^{th}-$element of $\sqrt{L} \cdot a_0 \bar{g}_L^c(\theta_0)$ (for $k_a = 1, \cdots, K_a$):

$$\begin{aligned}
& \frac{1}{\sqrt{L}} \mathcal{E}_L' \left(J_T \otimes J_n\right) \left(\sum_{l=1}^m a_{0,k_a}^{(l)} \mathbf{R}_{L,l}\right) \left(J_T \otimes J_n\right) \mathcal{E}_L \\
& + \frac{1}{\sqrt{L}} a_{0,k_a}^{(q)} \mathbf{q}_L' \left(J_T \otimes J_n\right) \mathcal{E}_L \\
= \ & \frac{1}{\sqrt{L}} \left[\begin{array}{c} \mathcal{E}_L' \left(J_T \otimes J_n\right) \left(\sum_{l=1}^m a_{0,k_a}^{(l)} \mathbf{R}_{L,l}\right) \left(J_T \otimes J_n\right) \mathcal{E}_L \\ -\sigma_0^2 tr \left(\sum_{l=1}^m a_{0,k_a}^{(l)} \mathbf{R}_{L,l} \left(J_T \otimes J_n\right)\right) \end{array} \right] \\
& + \frac{1}{\sqrt{L}} a_{0,k_a}^{(q)} \mathbf{q}_L' \left(I_T \otimes J_n\right) \mathcal{E}_L - \sqrt{\frac{n}{T}} a_{0,k_a}^{(q)} b_L^{\mathbf{L}} \left(\theta_0, \sigma_0^2\right) \\
& - \sqrt{\frac{T}{n}} \sum_{l=1}^m a_{0,k_a}^{(l)} b_{L,l}^{\mathbf{Q}} \left(\sigma_0^2\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{L}}\right)
\end{aligned} \tag{3.39}$$

where $a_{0,k_a}^{(l)}$ is the k_a^{th}-element of $a_0^{(l)}$, and $a_{0,k_a}^{(q)}$ is the k_a^{th}-row of $a_0^{(q)}$. Then, the mean zero term $\frac{1}{\sqrt{L}} \tilde{\mathbf{q}}_{L,k_a}' \mathcal{E}_L + \frac{1}{\sqrt{L}} \left[\mathcal{E}_L' \tilde{\mathbf{R}}_{L,k_a} \mathcal{E}_L - \sigma_0^2 tr\left(\tilde{\mathbf{R}}_{L,k_a}\right)\right]$ where $\tilde{\mathbf{R}}_{L,k_a} = \left(J_T \otimes J_n\right) \left(\sum_{l=1}^m a_{0,k_a}^{(l)} \mathbf{R}_{L,l}\right) \left(J_T \otimes J_n\right)$ and $\tilde{\mathbf{q}}_{L,k_a}' = a_{0,k_a}^{(q)} \mathbf{q}_L' \left(I_T \otimes J_n\right)$ characterizes the asymptotic distribution of the GMME. Hence, we can represent $\sqrt{L} \cdot a_L \bar{g}_L^c(\theta_0)$ as $\sqrt{L} \cdot a_L \bar{g}_L^c(\theta_0) = \sqrt{L} \cdot a_L \bar{g}_L^{c,(u)}(\theta_0) - \sqrt{\frac{n}{T}} a_L^{(q)} b_L^{\mathbf{L}}(\theta_0, \sigma_0^2) - \sqrt{\frac{T}{n}} \sum_{l=1}^m a_L^{(l)} b_{L,l}^{\mathbf{Q}}(\sigma_0^2) + o_p(1)$ where $\bar{g}_L^{c,(u)}(\theta_0)$ denotes the mean zero part of $\bar{g}_L^c(\theta_0)$.

For the asymptotic distribution of $\sqrt{L} \cdot a_0 \bar{g}_L^{c,(u)}(\theta_0)$, we need to find an appropriate limit theorem. We take ideas from Kuersteiner and Prucha (2013, 2018) and Qu et al. (2017). Note that $\{Z_{nt}, d_{ij}, d_c\}$ can affect network formation. Except for a channel of affecting $Y_{n,t-1}$ on Z_{nt}, remaining components of Z_{nt} are unspecified. For each $l(i,t)$ the (unspecified) exogenous network formation component $\varsigma_{l(i,t),L}^*$ can be represented by sub-σ-field $\mathcal{C}_l (\subset \mathcal{F})$. Elements in V_{nt} in equation (3.2) can be $\varsigma_{l(i,t),L}^*$. Hence, we specify $\left\{\varsigma_{l(i,t),L}^*\right\}_{l(i,t) \in D_L}$ as $\mathcal{C} = \vee_{l=1}^L \mathcal{C}_l$ and $E\left(\epsilon_{l(i,t),L} | \mathcal{C}\right) = 0$ for all $l(i,t) \in D_L$ where $\vee$ is the notation for the sigma field generated by the union of two sigma fields. Then, the $\mathcal{C}$-stable convergence concept can be established, which it is joint convergence of main statistics and a $\mathcal{C}$-measurable random variable. An advantage of taking the notion is that strictly exogenous components from network formation $\left\{\varsigma^*_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ can be represented by $\mathcal{C}$. If we treat $\left\{\varsigma^*_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ as constants, a conditional limiting distribution will follow a normal distribution meaning a conventional CLT framework. Observe that the main statistics $\sqrt{L}\cdot a_0\bar{g}_L^{c,(u)}(\theta_0)$ takes a LQ form of disturbances conditional on $\mathcal{C}$ if a_0 is $\mathcal{C}$-measurable. Lemma B.2.19 shows that the LQ form converges $\mathcal{C}$-stably. With considering randomness of a $\mathcal{C}$-measurable random variable, $\sqrt{L}\cdot a_0\bar{g}_L^{c,(u)}(\theta_0)$ may follow mixed normal due to random norming by the continuous mapping theorem.[116]

The assumption below states conditions for having the well-defined asymptotic variance of $\hat{\theta}_L$.

Assumption 3.4.11 *G_0 and a_0 are $\mathcal{C}$-measurable.*

$\liminf_{L\to\infty}\phi_{\min}(G_L'a_L'a_LG_L) > 0$ and $\liminf_{L\to\infty}\phi_{\min}(\Sigma_L) > 0$ where $\phi_{\min}(\cdot)$ denotes the smallest eigenvalue and $\Sigma_L = \frac{1}{L}Var\left(\sum_{l(i,t)\in D_L} g^c_{l(i,t),L}(\theta_0)\right)$.

In Hansen's (1982) GMM setting, the optimal GMM (OGMM) estimation can be considered by the approximated Σ_L. Let

$$\omega_{L,m} = \left[vec_D\left((J_T\otimes J_n)\mathbf{R}_{L,1}(J_T\otimes J_n)\right),\cdots,vec_D\left((J_T\otimes J_n)\mathbf{R}_{L,m}(J_T\otimes J_n)\right)\right] \tag{3.40}$$

and

$$\begin{aligned}\kappa_{L,m} &= [vec\left((J_T\otimes J_n)\mathbf{R}'_{L,1}(J_T\otimes J_n)\right),\cdots,vec\left((J_T\otimes J_n)\mathbf{R}'_{L,m}(J_T\otimes J_n)\right)]' \\ &\quad\times[vec\left((J_T\otimes J_n)\mathbf{R}^s_{L,1}(J_T\otimes J_n)\right),\cdots,vec\left((J_T\otimes J_n)\mathbf{R}^s_{L,m}(J_T\otimes J_n)\right)].\end{aligned}$$

[116] Or, we can argue that $\sqrt{L}\cdot a_0\bar{g}_L^{c,(u)}(\theta_0)$ follows asymptotically normal conditional on $\mathcal{C}$. A basic idea of the CLT with random norming, refer to Chapter 25.2 in Davidson (1994).

Then, the Σ_L can be approximated by

$$\tilde{\Sigma}_L = \frac{1}{L}\begin{bmatrix} (\mu_4 - 3\sigma_0^4)\,\omega'_{L,m}\omega_{L,m} + \sigma_0^4\kappa_{L,m} & \mu_3\omega'_{L,m}(I_T \otimes J_n)\,\mathbf{q}_L \\ \mu_3\mathbf{q}'_L(I_T \otimes J_n)\,\omega_{L,m} & \sigma_0^2\mathbf{q}'_L(I_T \otimes J_n)\,\mathbf{q}_L \end{bmatrix} \tag{3.41}$$

where $\mu_3 = E(\epsilon_{it}^3)$ and $\mu_4 = E(\epsilon_{it}^4)$, and Σ_L and $\tilde{\Sigma}_L$ can achieve the same probability limit, i.e., $\Sigma_0 = \text{plim}_{L\to\infty}\Sigma_L (= \text{plim}_{L\to\infty}\tilde{\Sigma}_L)$. Since $\tilde{\Sigma}_L$ depends on nuisance parameters σ_0^2, μ_3 and μ_4, consistent estimates for them are required for practical uses. The Theorem below shows the asymptotic distribution of $\hat{\theta}_L$.

Theorem 3.4.3 *If Assumptions 3.4.1 - 3.4.11 hold,*

$$\sqrt{L}\left(\hat{\theta}_L - \theta_0\right) + \left[G'_L a'_L a_L G_L\right]^{-1}\left\{\begin{array}{c} \sqrt{\frac{n}{T}}G'_L a'_L a_L^{(q)} b_L^{\mathbf{L}}(\theta_0, \sigma_0^2) \\ +\sqrt{\frac{T}{n}}G'_L a'_L \sum_{l=1}^m a_L^{(l)} b_{L,l}^{\mathbf{Q}}(\sigma_0^2) \end{array}\right\} + o_p(1) \tag{3.42}$$
$$\to_d plim_{L\to\infty}\Omega_L(a'_L a_L)^{\frac{1}{2}} \cdot \xi_*$$

where $\Omega_L(a'_L a_L) = [G'_L a'_L a_L G_L]^{-1} G'_L a'_L a_L \Sigma_L a'_L a_L G_L [G'_L a'_L a_L G_L]^{-1}$, *and* $\xi_* \sim N\left(\mathbf{0}_{(2+P+K)\times 1}, I_{2+P+K}\right)$ *is independent of* $plim_{L\to\infty}\Omega_L(a'_L a_L)$ *(because they are* $\mathcal{C}$*-measurable).*

Let $\xi_{*,L} = \sqrt{L}\left(\hat{\theta}_L - \theta_0\right) + [G'_L a'_L a_L G_L]^{-1}$
$\left\{\sqrt{\frac{n}{T}}G'_L a'_L a_L^{(q)} b_L^{\mathbf{L}}(\theta_0, \sigma_0^2) + \sqrt{\frac{T}{n}}G'_L a'_L \sum_{l=1}^m a_L^{(l)} b_{L,l}^{\mathbf{Q}}(\sigma_0^2)\right\}$ *and* $\Omega_0(a'_0 a_0) = plim_{L\to\infty}\Omega_L(a'_L a_L)$ *for further explanations.*

First, Theorem 3.4.3 says the existence of asymptotic biases (they are of $O\left(\max\left\{\frac{1}{n}, \frac{1}{T}\right\}\right)$) due to direct estimation of incidental parameters. And, $\hat{\theta}_L - \theta_0 = O_p\left(\max\left\{\frac{1}{\sqrt{L}}, \frac{1}{n}, \frac{1}{T}\right\}\right)$. If $\frac{n}{T} \to 0$ or $\frac{T}{n} \to 0$, the GMME's asymptotic distribution will be degenerated: (i) if $\frac{n}{T} \to 0$,
$n\left(\hat{\theta}_L - \theta_0\right) + [G'_L a'_L a_L G_L]^{-1} G'_L a'_L \sum_{l=1}^m a_L^{(l)} b_{L,l}^{\mathbf{Q}}(\sigma_0^2) \to_p 0$, and (ii) if $\frac{T}{n} \to 0$,
$T\left(\hat{\theta}_L - \theta_0\right) + [G'_L a'_L a_L G_L]^{-1} G'_L a'_L a_L^{(q)} b_L^{\mathbf{L}}(\theta_0, \sigma_0^2) \to_p 0$. To avoid degenerate distributions, hence, an appropriate ratio of n and T is required. If $\frac{n}{T} \to c \in (0, \infty)$,

$\sqrt{L}\left(\hat{\theta}_L - \theta_0\right) + \sqrt{c}\left[G_L' a_L' a_L G_L\right]^{-1} G_L' a_L' a_L^{(q)} b_L^{\mathbf{L}}\left(\theta_0, \sigma_0^2\right)$
$+ \sqrt{\frac{1}{c}}\left[G_L' a_L' a_L G_L\right]^{-1} G_L' a_L' \sum_{l=1}^{m} a_L^{(l)} b_{L,l}^{\mathbf{Q}}\left(\sigma_0^2\right) \rightarrow_d \Omega_0\left(a_0' a_0\right)^{\frac{1}{2}} \cdot \xi_*.$

Second, consider the case of $\frac{n}{T} \rightarrow c \in (0, \infty)$ for illustrative purposes. Then, Theorem 3.4.3 means that convergence of the characteristic function conditional on $\mathcal{C}$, $E\left(\exp\left(\mathbf{i}\varpi' \xi_{*,L}\right) | \mathcal{C}\right) \rightarrow \exp\left(-\varpi' \Omega_0\left(a_0' a_0\right) \varpi / 2\right)$ where $\mathbf{i} = \sqrt{-1}$, and ϖ is a linear combination. It implies that $E\left(\exp\left(\mathbf{i}\varpi' \xi_{*,L}\right)\right) \rightarrow E\left(\exp\left(-\varpi' \Omega_0\left(a_0' a_0\right) \varpi / 2\right)\right)$, which means that the unconditional asymptotic distribution of $\varpi' \xi_{*,L}$ is a mixed normal. Third, observe that the asymptotic distribution of $\sqrt{L}\left(\hat{\theta}_L - \theta_0\right)$ is not centered at $\mathbf{0}_{(2+P+K)\times 1}$ conditional on $\mathcal{C}$ if $\frac{n}{T}$ has a moderate ratio. Hence, a bias correction for $\hat{\theta}_L$ is needed. Fourth, the optimal GMM weight matrix (conditional on $\mathcal{C}$) can be specified by $a_L' a_L = \Sigma_L^{-1}$ by the Hansen's (1982) GMM setting, i.e., $\Omega_L\left(\Sigma_L^{-1}\right) = \left[G_L' \Sigma_L^{-1} G_L\right]^{-1}$. Then, its probability limit is $\left[G_0' \Sigma_0^{-1} G_0\right]^{-1}$, which is $\mathcal{C}$-measurable. Since Σ_L^{-1} involves unknown nuisance parameters including the second, third and fourth moments of ϵ_{it}, the optimal GMM estimator (OGMME) (denoted by $\hat{\theta}_{L,o}$) is infeasible. After getting consistent estimates of σ_0^2, μ_3, and μ_4, we can employ an estimated one (denoted by $\widehat{\tilde{\Sigma}_L}$) and get the feasible OGMME (FOGMME), $\hat{\theta}_{L,f}$. So long as $\widehat{\tilde{\Sigma}}_L - \Sigma_L = o_p(1)$, the FOGMME and OGMME are asymptotically equivalent.[117]

[117]Due to the model's complexity, many IVs might be needed to approximate endogenous variables in the model or to increase GMME's efficiency. However, we do not recommend using many moment conditions since it can yield an additional asymptotic bias (see Lee and Yu (2014)). That asymptotic bias is of an order proportional to the ratio of the number of moments and the total observation numbers. By observing (3.42), we can also expect that using many quadratic moments leads to large asymptotic biases (since it involves the sum $\sum_{l=1}^{m} a_L^{(l)} b_{L,l}^{\mathbf{Q}}\left(\sigma_0^2\right)$ and it may increase corresponding to m). In practice, therefore, we suggest selecting the number of IVs as well as quadratic moments carefully with considering feasible sample observations.

Bias correction

By observing (3.42), the two kinds of asymptotic biases exist. The first source of bias is due to estimating the incidental parameters $\alpha_{T,0}$ and employing the quadratic moments, $b^{\mathbf{Q}}_{L,l}\left(\sigma_0^2\right) = \frac{\sigma_0^2}{L}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[R_{nt,l}\right]_{ij}$. Since we select $R_{nt,l}$'s, $b^{\mathbf{Q}}_{L,l}\left(\sigma_0^2\right)$ can be evaluated after getting a consistent estimate of σ_0^2. The second source of bias comes from the incidental parameters $\mathbf{c}_{n0}$ and using the linear moments, $b^{\mathbf{L}}_{L}\left(\theta_0,\sigma_0^2\right) = \frac{1}{n}\sum_{t=1}^{T}E\left(\bar{\mathbf{q}}'_{nT}J_n\mathcal{E}_{nt}\right)$. Hence, a key part of bias correction is to evaluate $E\left(\bar{\mathbf{q}}'_{nT}J_n\mathcal{E}_{nt}\right)$. Since $\mathbf{q}_{nt}$ is predetermined, $b^{\mathbf{L}}_{L}\left(\theta_0,\sigma_0^2\right) = \frac{1}{L}\sum_{t=1}^{T}\sum_{t'>t}E\left(\mathbf{q}'_{nt'}J_n\mathcal{E}_{nt}\right)$. Note that possible choices for $\mathbf{q}_{nt}$ are lagged dependent variable $Y_{n,t-1}$, (lagged) exogenous variables $X_{n,t-1}$, X_{nt}, and their affine transformations. Our difficulty is to formulate $E\left(\mathbf{q}'_{nt'}J_n\mathcal{E}_{nt}\right)$ for $t'>t$ if $\mathbf{q}_{nt}$ consists of $Y_{n,t-1}$ and its linear transformations[118] since we do not derive the reduced form of $Y_{n,t-1}$. One way to reduce the bias is to use similar $Y_{n,t-1}$ (e.g., $Y^{e}_{n,t-1}$) which has a feasible reduced form.[119]

Then, we have a bias corrected estimator

$$\hat{\theta}^{c}_{L} = \hat{\theta}_{L} - \frac{1}{T}\left[C'_LC_L\right]^{-1}C'_La^{(q)}_{L}\widehat{b^{\mathbf{L}}_{L}\left(\theta_0,\sigma_0^2\right)} - \frac{1}{n}\left[C'_LC_L\right]^{-1}C'_L\sum_{l=1}^{m}a^{(l)}_{L}\widehat{b^{\mathbf{Q}}_{L,l}\left(\sigma_0^2\right)} \tag{3.43}$$

where $C_L = a_LG_L$, and $\hat{\cdot}$ denotes an estimate of $\cdot$. In case of $\hat{\theta}_{L,o}$ and $\hat{\theta}_{L,f}$, we can have the Cholesky factorization of Σ_L^{-1} (or $\tilde{\Sigma}_L^{-1}$), i.e., $\Sigma_L^{-1} = a'_{\Sigma_L,L}a_{\Sigma_L,L}$ where $a_{\Sigma_L,L} = \left(a^{(1)}_{\Sigma_L,L},\cdots,a^{(m)}_{\Sigma_L,L},a^{(q)}_{\Sigma_L,L}\right)$ is an $(m+q)\times(m+q)$ square matrix and $a_{\Sigma_L,L}$ is positive definite for sufficiently large L by Assumption 3.4.11. Then, $\hat{\theta}^{c}_{L,o}$ and $\hat{\theta}^{c}_{L,f}$ can

[118]When considering a linear combination with $Y_{n,t-1}$ as $\mathbf{q}_{nt}$ (i.e., $B_nY_{n,t-1}$ where B_n is some $n\times n$ matrix), a linear combination matrix should be strictly exogenous. If we use $W_{n,t-1}$ as a linear combination matrix, we need to additionally adjust nonlinearity of $W_{n,t-1}$ since $W_{n,t-1}$ might be a nonlinear function of $Y_{n,t-2}$.

[119]For example, we can choose $Y^{e}_{n,t-1}$ which comes from the LQ perturbation method and formulate $E\left(Y^{e\prime}_{nt'}J_n\mathcal{E}_{nt}\right)$ for $t'\geq t$. If a spatial network matrix W_n is available, we can also use a linear optimal decision vector based on the LQ value functions (Jeong and Lee (2018)).

be also defined. Indeed, the second source of bias $b_L^{\mathbf{L}}(\theta_0, \sigma_0^2)$ might not be perfectly measured even for the large samples since the functional form of actual Y_{nt} is difficult to be achieved. Hence, we firstly consider the ideal situation (i.e., we can successfully evaluate $E(\mathbf{q}_{nt'}' J_n \mathcal{E}_{nt})$ for $t' > t$). For example, (i) all IVs are strictly exogenous; (ii) $\lim_{L\to\infty} b_L^{\mathbf{L}}(\theta_0, \sigma_0^2)$ is obtainable. After that, we investigate the asymptotic impact originated from a misspecification of $b_L^{\mathbf{L}}(\theta_0, \sigma_0^2)$. Here is the ideal regularity condition for $\hat{\theta}_L^c$.

Assumption 3.4.12 *Assume*
$\sqrt{\frac{n}{T}}\left[[C_L'C_L]^{-1} C_L' a_L^{(q)} b_L^{\mathbf{L}}\widehat{(\theta_0, \sigma_0^2)} - [C_L'C_L]^{-1} C_L' a_L^{(q)} b_L^{\mathbf{L}}(\theta_0, \sigma_0^2)\right] \to_p 0,$ *and*
$\sqrt{\frac{T}{n}}\left[[C_L'C_L]^{-1} C_L' \sum_{l=1}^m a_L^{(l)} b_{L,l}^{\mathbf{Q}}\widehat{(\sigma_0^2)} - [C_L'C_L]^{-1} C_L' \sum_{l=1}^m a_L^{(l)} b_{L,l}^{\mathbf{Q}}(\sigma_0^2)\right] \to_p 0.$

Assumption 3.4.12 gives the asymptotic equivalence.[120] Since $G_L \to_p G_0$ and $a_L \to_p a_0$, we need to have $b_L^{\mathbf{L}}\widehat{(\theta_0, \sigma_0^2)} - b_L^{\mathbf{L}}(\theta_0, \sigma_0^2) \to_p \mathbf{0}_{q\times 1}$ and $b_{L,l}^{\mathbf{Q}}\widehat{(\sigma_0^2)} - b_{L,l}^{\mathbf{Q}}(\sigma_0^2) \to_p 0$ for $l = 1, \cdots, m$ to satisfy Assumption 3.4.12. As long as getting a consistent estimate of σ_0^2, we can achieve $b_{L,l}^{\mathbf{Q}}\widehat{(\sigma_0^2)} - b_{L,l}^{\mathbf{Q}}(\sigma_0^2) \to_p 0$. Now we have the following corollary.

Corollary 3.4.4 *Under the same assumptions of Theorem 3.4.3 with Assumption 3.4.12, we have*

$$\sqrt{L}\left(\hat{\theta}_L^c - \theta_0\right) \to_d \Omega_0 \left(a_0' a_0\right)^{\frac{1}{2}} \cdot \xi_*. \tag{3.44}$$

That is, we have

$$\sqrt{L}\left(\hat{\theta}_L^c - \theta_0\right) - \sqrt{L}\begin{pmatrix} \hat{\theta}_L - \frac{1}{T}[C_L'C_L]^{-1} C_L' a_L^{(q)} b_L^{\mathbf{L}}(\theta_0, \sigma_0^2) \\ -\frac{1}{n}[C_L'C_L]^{-1} C_L' \sum_{l=1}^m a_L^{(l)} b_{L,l}^{\mathbf{Q}}(\sigma_0^2) - \theta_0 \end{pmatrix} \to_p 0,$$

[120] If we take Y_{nt}^e for the bias correction and Y_{nt}^e is correctly specified, we can assume that (i) $\sum_{h=0}^{\infty}[A_n^e(\theta)]^h$ and $\sum_{h=1}^{\infty} h [A_n^e(\theta)]^h$ are uniformly bounded in either row or column sums uniformly in a neighborhood of θ_0, and (ii) $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$ for the asymptotic equivalence. They come from Corollary 4.3 in Jeong and Lee (2018).

implying

$$
\begin{aligned}
& \sqrt{L}\left(\hat{\theta}_L^c-\theta_0\right) \\
= \; & \sqrt{L}\left(\begin{array}{c}
\hat{\theta}_L-\frac{1}{T}\left[C_L^{\prime} C_L\right]^{-1} C_L^{\prime} a_L^{(q)} b_L^{\mathbf{L}}\left(\theta_0, \sigma_0^2\right) \\
-\frac{1}{n}\left[C_L^{\prime} C_L\right]^{-1} C_L^{\prime} \sum_{l=1}^m a_L^{(l)} b_{L, l}^{\mathbf{Q}}\left(\sigma_0^2\right)-\theta_0
\end{array}\right)+o_p(1) \\
\rightarrow \; & { }_d \Omega_0\left(a_0^{\prime} a_0\right)^{\frac{1}{2}} \cdot \xi_* .
\end{aligned}
$$

Hence, if we have a moderate ratio of n and T and good approximations of $b_L^{\mathbf{L}}\left(\theta_0, \sigma_0^2\right)$ and $b_{L,l}^{\mathbf{Q}}\left(\sigma_0^2\right)$'s, the asymptotic distribution of $\sqrt{L}\left(\hat{\theta}_L^c-\theta_0\right)$ will be properly centered conditional on $\mathcal{C}$.

However, we might not have $b_L^{\mathbf{L}}\left(\widehat{\theta_0, \sigma_0^2}\right)-b_L^{\mathbf{L}}\left(\theta_0, \sigma_0^2\right) \rightarrow_p \mathbf{0}_{q \times 1}$ since we approximate $E\left(\overline{\mathbf{q}}_{nT}^{\prime} J_n \mathcal{E}_{nt}\right)$ if $\mathbf{q}_{nt}$ includes $Y_{n,t-1}$. Let $b_*^{\mathbf{L}}$ =plim$_{L\rightarrow\infty} b_L^{\mathbf{L}}\left(\widehat{\theta_0, \sigma_0^2}\right)$ (=probability limit of the approximated one) and $b_0^{\mathbf{L}}=\lim_{L\rightarrow\infty} b_L^{\mathbf{L}}\left(\theta_0, \sigma_0^2\right)$ (=probability limit of the actual bias term). Suppose $b_*^{\mathbf{L}} \neq b_0^{\mathbf{L}}$. Then, the first part of Assumption 3.4.12 fails. Observe that

$$
\begin{aligned}
& \sqrt{\frac{n}{T}}\left[\left[C_L^{\prime} C_L\right]^{-1} C_L^{\prime} a_L^{(q)} b_L^{\mathbf{L}}\left(\widehat{\theta_0, \sigma_0^2}\right)-\left[C_L^{\prime} C_L\right]^{-1} C_L^{\prime} a_L^{(q)} b_L^{\mathbf{L}}\left(\theta_0, \sigma_0^2\right)\right] \\
= \; & \sqrt{\frac{n}{T}}\left[C_L^{\prime} C_L\right]^{-1} C_L^{\prime} a_L^{(q)}\left[\left(b_L^{\mathbf{L}}\left(\widehat{\theta_0, \sigma_0^2}\right)-b_*^{\mathbf{L}}\right)-\left(b_L^{\mathbf{L}}\left(\theta_0, \sigma_0^2\right)-b_0^{\mathbf{L}}\right)+\left(b_*^{\mathbf{L}}-b_0^{\mathbf{L}}\right)\right] .
\end{aligned}
$$

If $\frac{n}{T} \rightarrow c \in (0,\infty)$, we have

$$
\sqrt{L}\left(\hat{\theta}_L^c-\theta_0\right) \rightarrow_d \Omega_0\left(a_0^{\prime} a_0\right)^{\frac{1}{2}} \cdot \xi_*+\sqrt{c} plim_{L \rightarrow \infty}\left[C_L^{\prime} C_L\right]^{-1} C_L^{\prime} a_L^{(q)}\left(b_*^{\mathbf{L}}-b_0^{\mathbf{L}}\right) \tag{3.45}
$$

Hence, if we expect large errors in using this bias-correction method, having large T (relative to n) can alleviate the problem.

By using $\hat{\theta}_L^c$, the statistical inferences might yield more accurate results. If we consider the linear constraints[121] R, the Wald statistic

[121] We can also consider a non-linear constraint by applying the delta method.

$T_L = \left\| \left(R\Omega_L \left(a_L' a_L\right) R'\right)^{\frac{1}{2}} \sqrt{L} \left(R\hat{\theta}_L^c - r\right) \right\|^2$ can be used to test $H_0 : R\theta_0 = r$ (against $H_1 : R\theta_0 \neq r$) where r is a $\dim(r)$-dimensional vector and R is a $\dim(r) \times (2 + K + P)$ full row rank matrix. By applying Lemma B.2.20, T_L will follow the asymptotically chi-square distribution: under $H_0 : R\theta_0 = r$, $T_L \rightarrow_d \chi^2_{\dim(r)}$ implying $P\left(T_L > \chi^2_{\dim(r),1-\alpha}\right) \rightarrow \alpha$ where $\chi^2_{\dim(r),1-\alpha}$ denotes the $(1-\alpha)$ quantile of the chi-square distribution with $\dim(r)$ degrees of freedom. Even for the existence of $\mathcal{C}$-measurable random components $\text{plim}_{L\to\infty} \Omega_L \left(a_L' a_L\right)^{\frac{1}{2}}$ in the asymptotic distribution of $\hat{\theta}_L^c$, the Wald statistic asymptotically follows the unconditional chi-square distribution. This result has the same implication as Theorem 4 in Kuersteiner and Prucha (2018).

3.4.4 Finite sample properties

In this subsection, we conduct Monte Carlo experiment with data generated from equation (3.19). There are two purposes of the simulations: (i) performance evaluation of the NL2S and GMM estimators, and (ii) the analysis of misspecification errors when we estimate the model with ignoring the forward-looking feature. The simulation design aligns with the empirical application in Section 3.5 in terms of the sample size $(n, T) = (48, 24)$, the spatial networks (W_{nt}), and the averages of dependent variables $(\bar{Y}_n^\circ)$.[122] We set $K = 1$, $P = 1$, and $\delta = 0.95$. We draw X_{nt}, $\mathbf{c}_{n0}$, $\alpha_{T,0}$, and $\mathcal{E}_{nt}$ mutually independently from the standard normal distributions. We generate the data with $20 + T$ periods, where the starting vector values are $\bar{Y}_n^\circ$, but

[122] That is, we employ the same spatial networks and relevant terms as the application part to have a unique reduced form of equation (3.19). We generate $Y_{n,t+1}^e$ and $\nabla V_{n,t+2}^e$ by the LQ perturbation method around $\bar{Y}_n^\circ$. In the supplementary file, we conduct additional simulations for the case of a time-invariant spatial network matrix. Since this case can have a well-defined correlation structure of Y_{nt}, we can do the quasi maximum likelihood (QML) approach. We compare estimation performance of the NL2S and GMM estimators with that of the QML estimator.

Table 3.1: Simulation results

Case 1: $(\lambda_0, \gamma_0, \psi_0, \beta_0) = (0.1, 0.2, 0.5, 1)$

	NL2S				GMM			
	λ_0	γ_0	ψ_0	β_0	λ_0	γ_0	ψ_0	β_0
Bias	0	-0.0309	-0.0013	-0.0202	-0.0117	-0.0309	-0.0014	-0.0181
SD	0.0208	0.0268	0.0237	0.0346	0.0116	0.0269	0.0239	0.0345
T-SD	0.0201	0.0263	0.0242	0.0346	0.0108	0.0261	0.0242	0.0344
RMSE	0.0208	0.0409	0.0237	0.0401	0.0165	0.041	0.024	0.0389
	GMMC							
	λ_0	γ_0	ψ_0	β_0				
Bias	-0.0011	-0.0037	0.0238	-0.0026				
SD	0.0115	0.0279	0.024	0.0345				
T-SD	0.0108	0.0261	0.0242	0.0344				
RMSE	0.0116	0.0282	0.0338	0.0346				

utilize the last T periods as our sample. For the true parameter values, we consider two sets of θ_0, $(0.1, 0.2, 0.5, 1)'$ and $(-0.1, 0.2, -0.5, -1)'$. We select small amounts of θ_0 to obtain fast convergence in solving algebraic matrix Riccati equations. Linear IVs are $\mathbf{q}_{nt} = [Y_{n,t-1}, W_nY_{n,t-1}, W_n^2Y_{n,t-1}, X_{nt}, W_nX_{nt}, W_n^2X_{nt}]$ where W_n is the row normalized adjacency matrix of the U.S. states. The quadratic moment matrix is formed with $\mathbf{R}_{L,1} = I_T \otimes W_n$.

First, we compare the three estimation methods: (i) NL2S, (ii) GMM, and (iii) GMM with bias correction (denoted by GMMC). By comparing performance of (i) NL2SE and (ii) GMME, we can investigate whether using the quadratic moment condition plays a role in efficiency. To evaluate performance of estimators, we report four criteria: (i) bias, (ii) standard deviation (SD), (iii) theoretical standard deviation

Misspecification errors

	NL2S				GMM			
	λ_0	γ_0	ψ_0	β_0	λ_0	γ_0	ψ_0	β_0
Bias	-0.0122	-0.0525	0.0006	-0.1346	-0.0214	-0.0514	0.0006	-0.1331
	GMMC							
	λ_0	γ_0	ψ_0	β_0				
Bias	-0.0123	-0.0525	0.0302	-0.1344				

Case 2: $(\lambda_0, \gamma_0, \psi_0, \beta_0) = (-0.1, 0.2, -0.5, -1)$

	NL2S				GMM			
	λ_0	γ_0	ψ_0	β_0	λ_0	γ_0	ψ_0	β_0
Bias	0.0029	-0.0309	-0.0021	0.0186	-0.0099	-0.034	-0.0021	0.0241
SD	0.0229	0.0261	0.0228	0.0369	0.0134	0.0259	0.0231	0.0366
T-SD	0.025	0.0263	0.0239	0.0354	0.0135	0.0261	0.0239	0.0345
RMSE	0.0231	0.0404	0.0229	0.0414	0.0166	0.0427	0.0232	0.0438
	GMMC							
	λ_0	γ_0	ψ_0	β_0				
Bias	0.0005	-0.0049	0.0227	0.0025				
SD	0.0136	0.0272	0.0231	0.0367				
T-SD	0.0135	0.0261	0.0239	0.0345				
RMSE	0.0136	0.0277	0.0323	0.0368				

Misspecification errors

	NL2S				GMM			
	λ_0	γ_0	ψ_0	β_0	λ_0	γ_0	ψ_0	β_0
Bias	0.0156	-0.0534	-0.0004	0.1368	0.0028	-0.0547	-0.0004	0.1401
	GMMC							
	λ_0	γ_0	ψ_0	β_0				
Bias	0.0139	-0.0545	0.0292	0.1372				

(T-SD), and (iv) root mean square error (RMSE). To obtain those measures, we conduct 400 repetitions. In Table 3.1, we obtain several findings:

(1) First, consider estimation of λ_0. NL2SE has smaller bias relative to that of GMME in absolute values. There exists downward biases in GMME. After the bias correction, the magnitudes of bias become smaller. In terms of efficiency, GMME is significantly better than NL2SE.

(2) Second, consider estimating γ_0. In NL2SE and GMME, there are downward biases. After correcting the bias, the magnitude of bias decreases. In estimating γ_0, additionally considering $\mathbf{R}_{L,1}$ does not improve efficiency.

(3) Third, consider estimation of ψ_0. In NL2SE and GMME, there are downward biases for both cases, $\lambda_0 > 0$ and $\lambda_0 < 0$. In estimating ψ_0, our bias correction increases magnitudes of biases (in absolute values). It seems that our bias correction method overevaluates the bias for estimating ψ_0.

(4) Fourth, we report estimation results of β_0. For the NL2S and GMM methods, the signs of biases are opposite to those of the true β_0. After applying the bias correction method, we can reduce the magnitude of bias. It seems that additionally using $\mathbf{R}_{L,1}$ does not affect efficiency.

In sum, (i) using the quadratic moment condition is only beneficial in estimating λ_0; (ii) our bias correction works well in estimating λ_0, γ_0 and β_0. For accurate estimation for λ_0, using quadratic moment conditions can be additionally considered (to support NL2SE).

Second, we study misspecification errors when we estimate equation (3.19) with $0 < \delta < 1$ using the conventional SDPD model (equation (3.1)). If $\theta_0 = (0.1, 0.2, 0.5, 1)'$, we detect significant downward biases except for the estimates of ψ_0 when we disregard the forward-looking feature. If $\theta_0 = (-0.1, 0.2, -0.5, -1)'$, we observe upward misspecification biases in estimating λ_0 and ψ_0. That is, estimated coefficients λ_0, γ_0 and β_0 will be underestimated (in absolute values) without the forward-looking behaviors.

3.5 Application

In this section, we introduce an empirical application to show how to implement our model. Note that our model assumptions are appropriate for fixed physical locations of agents with their time-varying intensities of interactions by the economic indicators. Hence, we consider policy interdependencies among the U.S. state's health expenditures.[123] Each state government becomes an economic agent i, and its action at each period is state health expenditure per capita. A government officer makes its health expenditure decision by considering two potential factors (Brueckner (2003)): (i) welfare recipients can move to neighboring states to enjoy favorable welfare environment (i.e., welfare motivated move), and (ii) they can expel the officer when the state government makes an inefficient expenditure decision by comparing that of similar states (i.e., yardstick competition).[124] In consequence, a state decision maker

[123] Since the state's health expenditure is used for health-related research, immunization programs, regulation of air and water quality, etc., it can yield the policy spillover effects to neighboring states.

[124] Brueckner (2003) reviews the justifications of payoff function (3.3) in the framework of local government strategic interactions. The yardstick competition can be explained by the policy spillover model. Then, the payoff specification (3.3) for a local government describes the utility of a representative resident of region i.

The welfare motivated move is justified by the resource-flow model. In consequence, a jurisdiction's payoff would be the same as (3.3) and show a representative resident's utility. In this case, however, jurisdiction i is indirectly affected by other jurisdictions $Y_{-i,t}$ via resources such as population. A resident in region i can be a labor force and can move to other regions to enjoy a favorable economic policy. Hence, jurisdiction i has a motivation of considering neighboring policies.

Table 3.2: Descriptive statistics: 48 contiguous states in U.S.

Variables	Mean	Standard dev.	Min	Max
Health expenditure	0.1425	0.0805	0.0271	0.5855
Proportion on total expenditure	0.0247	0.0124	0.0047	0.0707
Total expenditure	5.7992	1.2951	3.4631	11.6327
Total revenue	5.8646	1.2749	1.5816	14.8746
Grants from the Federal government	1.4947	0.4185	0.5757	4.1586
Population (millions)	6.0305	6.5155	0.4663	39.25
Population density	73.7276	100.0283	469.5999	1.8541
Personal income	35.7279	5.6464	25.7151	59.1866
Unemployment rate	0.0559	0.0187	0.023	0.1361
Proportion of the population aged 0-17	0.2475	0.0216	0.1898	0.3532
Proportion of the population aged 18-65	0.6198	0.0151	0.559	0.6511

Note: Sample is 48 contiguous states from 1992 to 2016. Dollar amounts are in thousands and real per capita values adjusted by the GDP deflator with base year 2012.

considers neighboring actions. Note that the state's health expenditure can positively affect its future economic status by human capital accumulations (Bloom et al. (2004)). In consequence, we take the time-varying W_{nt} since evolution of W_{nt} is driven by the states' economic status. As Han and Lee (2016), economic proximities of states are captured by their personal income per capita (i.e., variable Z_{nt}).

We estimate the underlying incentive structure of U.S. states' actions on health expenditure by assuming forward-looking agents and network evolution. For data, we choose 48 contiguous states in the U.S. (excluding Alaska, Hawaii, and Washington D.C.) and time periods are from 1992 to 2016 (total 1,200 observations for each variable). From the United States Census Bureau, we obtain the states' demographic/economic and finance data.[125] For the additional macroeconomic variables

[125]For the government finance data, two periods of observations (years 2001 and 2003) are not available. Hence, we generate the government finance data for the two periods by the interpolation.

(e.g., GDP deflator, interest rates, states' unemployment rates), we use the website of the Federal reserve bank of S.t. Louis. All dollar amounts are in thousands and real per capita values adjusted by the GDP deflator with the base year 2012. Table 3.2 shows the descriptive statistics for collected variables.

The main issue is verifying whether a local government decision affects their network links by the economic proximities. [126] In our model, the coevolution of economic activities and networks arises if $\psi_0 \neq 0$. By using Z_{nt} and $\{d_{ij}\}$, we need to specify socio-economic networks W_{nt}.[127] We consider W_{nt} based on the following specification

$$w_{t,ij} = h\left(z_{it}, z_{jt}\right) \cdot h_d\left(d_{ij}\right) \cdot \mathbf{1}\left\{(i,j)\, nbd\right\} \tag{3.47}$$

for $i \neq j$ to represent the intensity of interaction between i and j at time t. To measure the economic distance, we consider $E_{t,ij} = |z_{it} - z_{jt}|$.[128] By introducing $\mathbf{1}\left\{(i,j)\, nbd\right\}$, we can have sparse W_{nt}.

To give reasonable backgrounds of selecting spatial networks, we consider the Cobb-Douglas function specification of $w_{t,ij}$:

$$w_{t,ij} = E_{t,ij}^{-\alpha_e} d_{ij}^{-\alpha_d} \left(\frac{y_{j,t-1}}{y_{i,t-1}}\right)^{\alpha_w} \cdot \mathbf{1}\left\{(i,j)\, nbd\right\} \tag{3.48}$$

[126] A famous discussion related to this issue is Barro (1990).

[127] Each state government's physical location can be characterized by the pair of latitude and longitude for its capital (denoted by (φ_i, υ_i) in radians). By the Haversine formula, we compute the physical distance between two states denoted by d_{ij}: Let (φ_i, υ_i) and (φ_j, υ_j) be respectively geographic locations of i and j. Then,

$$d_{ij} = 2r_E \arcsin\left(\sin^2\left(\frac{\varphi_j - \varphi_i}{2}\right) + \cos\left(\varphi_j\right)\cos\left(\varphi_i\right)\sin^2\left(\frac{\upsilon_j - \upsilon_i}{2}\right)\right) \tag{3.46}$$

where $r_E = 6356.752$km denotes the Earth radius.

[128] For a case of $|z_{it} - z_{jt}| < \$200$, we set $E_{t,ij}$ to be \$200 to exclude extremely strong intensity of economic interactions. Selection of the minimum $E_{t,ij}$ does not significantly affect our estimation results.

Table 3.3: Estimated coefficients (elasticities) of spatial-economic network formation models

Coefficients	Specification (1)	Specification (2)
α_e	0.3017*** [0.0266]	0.1313*** [0.0242]
α_d	1.4575*** [0.0272]	2.0421*** [0.0213]
α_w	0.0388* [0.0231]	0.0808* [0.0459]
Fixed effects	No	Yes

Note: Estimates that are significant at the 1 percent, 5 percent, and 1 percent levels are respectively marked by *, **, and ***.

for $i \neq j$ where α_d, α_e, and α_w are coefficients. The estimation procedure is introduced in Appendix B. Table 3.3 shows the estimated coefficients in specification (3.48). The estimated α_e and α_d are significant under the 1% significance level, and estimate of α_w is significant under the 10% significance level.

We estimate model's main parameters based on Assumption B.13 via the NL2SE and GMME. For the GMME, we consider the FOGMME. For the time-discounting factor, we employ the average of 10-year Treasury constant maturity rates during the sample periods: i.e., $\delta = 0.956 \simeq \frac{1}{1+\bar{\imath}_r}$ where $\bar{\imath}_r = 0.0457$ denotes the average interest rates. As explanatory variables, we utilize (i) total revenue, (ii) federal grants, (iii) time differenced population density[129], (iv) unemployment rate, (v) proportion of the population aged 0~17, and (vi) proportion of the population aged 18-65. This specification is guided by Case et al. (1993). For the linear IV moment conditions, we

[129]We find that the U.S. population densities are quite persistent. In our model framework, all exogenous (time-varying) characteristics (η^v_{nt} in the theoretical model) should be stationary. To avoid nonstationary variable issues, we use the time differenced population densities.

Table 3.4: Model estimation

Case 1: $\delta = 0$

Estimation method Variables	NL2S	GMM	GMMC
Grants from the Fed.	0.0173** [0.0071]	0.0179*** [0.0062]	0.0176*** [0.0062]
λ_0	0.0052 [0.0161]	0.0046 [0.0110]	0.0100 [0.0110]
γ_0	0.8963*** [0.0228]	0.8932*** [0.0224]	0.8940*** [0.0224]
ψ_0	0.3627** [0.1859]	0.3613* [0.1859]	0.4050** [0.1859]
No. of Obs.	1152	1152	1152
Sample objective function	3.59E-06	0.0175	0.0192

Note: Estimates that are significant at the 10 percent, 5 percent, and 1 percent levels are respectively marked by *, **, and ***.

consider $[Y_{n,t-1}, X_{nt}]$ and its transformation by W_n and W_n^2.[130] We set $\mathbf{R}_{L,1} = I_T \otimes W_n$. Bias correction of the GMME is conducted based on the optimal policies from the LQ perturbation method.

Table 3.4 shows the main estimation results for two models: (i) $\delta = 0$ (myopic), and (ii) $\delta = 0.956$. The reported standard errors are based on the conditional asymptotic normality on the $\mathcal{C}$-measurable random variables. We also report the Wald test statistics T_L for $H_0 : \lambda_0\psi_0 = 0$ based on the bias corrected GMME. Here is the summary of estimation results. First, significantly positive γ_0 is captured under the 1% significance level. It implies that a state government is hard to extremely change its health expenditure over time due to the high level of adjustment costs. The spatial interaction coefficients (λ_0) are positive, which it means that a state

[130]The spatial network W_n is specified by a row-normalized one $w_{ij} = \frac{\tilde{w}_{ij}}{\sum_{k=1}^{n} \tilde{w}_{ij}}$ where

$$\tilde{w}_{ij} = \mathbf{1}\left\{(i,j)\, nbd\right\} \tag{3.49}$$

if $j \neq i$. Then, W_n is strict exogenous.

Case 2: $\delta = 0.956$

Estimation method Variables	NL2S	GMM	GMMC
Grants from the Fed.	0.0200** [0.0083]	0.0207*** [0.0074]	0.0191** [0.0073]
λ_0	0.0032 [0.0096]	0.0029 [0.0069]	0.0028 [0.0069]
γ_0	0.9839*** [0.0060]	0.9840*** [0.0059]	0.9954*** [0.0059]
ψ_0	0.3627** [0.1859]	0.3612* [0.1859]	0.4095 [0.1859]
T_L for $H_0 : \lambda_0\psi_0 = 0$			0.1566
No. of Obs.	1152	1152	1152
Sample objective function	4.31E-06	0.0176	0.0236

Note: Estimates that are significant at the 10 percent, 5 percent, and 1 percent levels are respectively marked by *, **, and ***.

government decision on health expenditure is reinforced by rival states' decisions. However, the estimated coefficients are not significant. For ψ_0, we detect the positive effect of state's health expenditure on its economic status. The Wald test statistic $T_L = 0.1566$ does not reject $H_0 : \lambda_0\psi_0 = 0$ since $\chi^2_{1,0.95} = 3.82$. It seems that the U.S. health expenditure decisions and the spatial networks do not coevolved. We observe the significantly positive effect of the federal grants on the state's marginal payoff under the 5% significance level. For other exogenous characteristics, there is no significant effect. In comparing the two models, the estimated coefficients for γ_0 and β_0 from the conventional model are lower than those of the forward-looking model.

3.6 Conclusion

In this paper, we introduce a new spatial econometric model describing optimal actions of forward-looking agents and spatial network evolution. Due to the forward-looking agent assumption, the feedback effect arises: agents are not only affected by spatial networks, but their actions can also affect the spatial networks. Since the agent's payoff function is characterized by parameters, a corresponding parametric econometric model is established. To estimate structural parameters, we consider a GMM estimation method based on a set of Euler equations. Asymptotic properties of the GMM estimator are studied for statistical inferences. Using the Wald test, we can test whether spatial networks are exogenous evolve or not. As an application, we explore policy interdependencies among the U.S. states' decisions on health expenditure.

Chapter 4: Spatial dynamic models with intertemporal optimization III: A dynamic Stackelberg game with spatial interactions

4.1 Introduction

This paper suggests a new spatial econometric model for a panel data set describing spatial interactions among two types of forward-looking agents: a leader and multiple followers. In an application of public economics, a leader represents the federal government while followers are for the state governments. Using our estimation equation, we want to empirically investigate two types of policy interdependence at the same time: (i) interactions between the U.S. federal and state governments, and (ii) interactions among the state governments. In traditional literature, regional policy interdependence can be explained by a spatial econometric model (Case et al. (1993)), which can describe strategic interactions among spatial units. For example, a linear spatial autoregressive (SAR) model describes a vector of best responses of a game played on a spatial network with a parametric linear-quadratic (LQ) utility function. However, conventional spatial econometric models are designed for studying policy interrelations among the same level of local governments (e.g., interactions among state governments). To formulate the two types of spatial interactions, we establish a dynamic Stackelberg game with spatial interactions motivated by Chapter 19 in Ljungqvist and Sargent (2012). LQ parametric payoff functions characterize preferences of the two types of agents. Focusing on the linear rational expectations equilibrium, we derive the parametric econometric model. The resulted model is a new spatial dynamic panel data (SDPD) model showing the different levels of interactions.

As the first contribution, we establish a spatial econometric model based on a dynamic Stackelberg game played by the leader and the multiple followers. As a review of dynamic Stackelberg game models, refer to Li and Sethi (2017). Our model specification is useful to study spatial interactions of the central (one leader) and the local governments (multiple followers). Two points are considered in our model specification. The first point is basic features of a spatial panel data set. Since a policy decision of a (central/local) government arises at each fiscal year, observed policies by them are reported in a panel data set. In formulating our theoretical model, we consider strategic interactions among rational economic agents. Hence, we assume that the revealed actions come from agents' intertemporal optimization (i.e., forward-looking agent assumption). At each period, both types of agents rationally expect uncertain future actions and exogenous characteristics given their available information. As the second point of view, we assume that there is a hierarchy in decision-making of the two types of agents. For each period, two stages of decision-making exist. For each period, the leader firstly chooses its continuous type actions (grants) to support n followers (i.e., n decision variables). After observing the leader's actions, the n followers simultaneously choose their own (continuous type) actions (e.g., state's expenditure) by considering (i) neighboring followers' current and expected future actions with their exogenous characteristics and (ii) expected future leader's actions. At each period, the influences from the leader and the followers are unilateral and they are formulated as some parameters. Like a conventional spatial econometric model, interrelations among the followers are characterized by a spatial network (formulated by an $n \times n$ matrix W_n).

We derive the rational expectation equilibrium equations with assuming a stable economic environment. We seek to estimate parameters characterizing agents' payoff functions. Each follower has a LQ payoff function in his/her action.[131] This LQ payoff specification justifies a conventional SDPD model (e.g., Lee and Yu (2010)). Based on the proposed LQ payoff specification, we build the agents' lifetime optimization problems. A follower's payoff can be influenced by rival followers' actions through a spatial network W_n. Leader's payoff is defined by the summation of followers' payoff functions, which shows social welfare. Hence, a follower's payoff can be also affected by the leader's actions because there is a hierarchy in decision-making. Since both agents' payoff functions are LQ in actions as well as state variables, the vector of optimal actions for both types of agents will be linear in state variables. Due to the Stackelberg game structure, for each time period the followers' optimal actions depend on the optimal actions made by the leader. It means that we also need to verify the vector of leader's optimal actions to implement the econometric model.

By giving a specific structure on the followers' exogenous characteristics, we finish specifying the estimation equation. In the econometrician's perspective, they contain (i) observable exogenous variables with sensitivity parameters, (ii) unobserved

[131] This LQ payoff function is a dynamic extension of static network interaction models discussed by Ballester et al. (2006), Calvo-Armengol et al. (2009), and Bramoullé et al. (2014).

individuals' (followers') innate characteristics, (iii) an unobserved time specific shock, and (iv) i.i.d. disturbances for estimation. The individual and time effects are treated as parameters, so we can allow arbitrary correlations between those unobserved effects and the exogenous variables. By the reduced form, we derive the log-likelihood function. For the incidental parameters showing individual- and time dummies, we employ the direct estimation approach by the concentrated log-likelihood function. Hence, estimation of structural parameters can come from optimization under a finite parameter space and the quasi maximum likelihood (QML) estimator is obtained by maximizing the concentrated log-likelihood function.

Large sample properties, consistency and asymptotic normality, are studied based on (i) the conventional arguments for nonlinear extremum estimators and (ii) the statistical theories for a LQ form of martingale differences (Yu et al. (2008)). Asymptotic analyses in this paper are conducted based on (i) large n and T and (ii) increasing domain asymptotic frameworks. From the derived asymptotic distribution of the QML estimator, we observe the existence of asymptotic biases due to the incidental parameters. Achieving a moderate ratio between n and T is crucial to have the nondegenerate distribution of the QML estimator. To have the asymptotically centered confidence intervals, we suggest a bias correction method. We conduct Monte Carlo simulations for finite sample properties of the QML estimator and its bias corrected version. As conventional SDPD model's estimation, the bias correction method is based on the scores' expected values at the true parameters. In many cases, we observe that the QML estimator underestimate the true parameter values in absolute values. After the bias correction, we find that the magnitude of biases tends to diminish.

Last, we employ our model to examine (i) policy interdependence among U.S states' general expenditures and (ii) interrelations between their expenditures and grants from the federal government. Among the U.S. state governments, we find that there is a positive spatial spillover effect in their expenditure. In our estimation results, it seems that there is no effect of the federal grants on the states' expenditures. Also, we observe that the state expenditures are dynamically persistent. We observe that there exists a significant positive effect of the state's total revenue. Also, a significant negative effect of the state's unemployment rate is detected.

4.2 Model specification

In this section, we introduce a dynamic Stackelberg game consisting of one leader and multiple followers. A basic payoff specification will follow a (dynamic version) LQ payoff function (Jeong and Lee (2018)). This payoff function can justify a conventional time space dynamic model for a spatial panel data set. Reversely, we attempt to make a theoretical model using this LQ payoff function with (i) the forward-looking agent assumption and (ii) the existence of the two types of agents with hierarchical decision-making. The derived rational expectation equilibrium equation will lead to an estimation equation.

4.2.1 Spatial network interaction model with a dynamic Stackelberg game

We assume that two types of agents exist: (i) a leader, and (ii) followers. There is only one leader while multiple followers exist (the number of followers is denoted by n). For an empirical example, the U.S. federal government is a leader and followers are state governments. We assume that (i) decision-making periods are infinite[132] for a stable economic environment, and (ii) agents are forward-looking instead of myopic (i.e., they maximize their lifetime payoffs rather than their per period payoffs). Each agent's lifetime payoff (for both types of agents) is defined by a weighted summation of his/her per period payoff. For this, we introduce a time-discounting factor $\delta \in [0, 1)$ to distinguish the future payoffs from the current one.[133] We assume that δ is common to the leader and the followers.

To characterize relations among agents, we specify geographic locations of them. This is because we assume that interactions among agents arise due to geographic arrangements. The leader is indexed by 0 and indexes $i = 1, \cdots, n$ represent the followers. We assume that each agent has its fixed location on the (subset of) Euclidean space $\mathbf{R}^d$ ($d \geq 1$). Hence, each endowed index contains information in the agent's geographic location. Based on those specified locations, interactions among followers are characterized by an $n \times n$ spatial network matrix W_n. Each element in W_n takes a nonnegative real number and represents a (relative) intensity of interaction. To exclude the self-influence, we assume that all diagonal entries in W_n are zero. For notational convenience, $w_{i.}$ and w_{ij} denote respectively the i^{th}-row of W_n and the (i, j)-element of W_n. Interactions between the leader and the followers will be characterized later.

Three types of variables exist in an economy. First, let $Y_{nt} = (y_{1t}, \cdots, y_{nt})'$ be an $n \times 1$ vector of followers' continuous type actions. Second, let $\mathbf{b}_{nt} = (b_{1t}, \cdots, b_{nt})'$

[132]We can also consider a finite horizon problem if an econometrician knows initial and terminal periods of decision-making.

[133]Even for $\delta = 0$ (myopic agents), we will also have a new spatial econometric model specification describing hierarchical decision-making of the two types of agents. After introducing a general case, we will introduce this special case.

be n-dimensional vector of leader's actions for followers at time t. For example, y_{it} is a local government i's action (e.g., expenditure or tax rates) while b_{it} is a level of grants for a local government i. For the followers' behaviors, we will derive the optimal actions Y_{nt}^*:

$$Y_{nt}^*\left(Y_{n,t-1}, \mathbf{b}_{nt}, \eta_{nt}\right) = A_n Y_{n,t-1} + B_n \mathbf{b}_{nt} + C_n \eta_{nt} \tag{4.1}$$

where A_n, B_n, and C_n are some $n \times n$ matrices.[134] That is, the followers' current optimal actions (Y_{nt}^*) rely on their previous ones ($Y_{n,t-1}$), and realized exogenous characteristics (η_{nt}). Since the followers choose their actions after observing the leader's actions ($\mathbf{b}_{nt}$), Y_{nt}^* rely on $\mathbf{b}_{nt}$ (i.e., hierarchical decision-making). Last, let $\eta_{nt} = (\eta_{1t}, \cdots, \eta_{nt})'$ be an $n \times 1$ vector of exogenous characteristics of followers controlled by nature. It can compose of time-invariant ($\eta_n^{iv} = (\eta_1^{iv}, \cdots, \eta_n^{iv})'$) and time-varying ($\eta_{nt}^{v} = (\eta_{1t}^{v}, \cdots, \eta_{nt}^{v})'$) characteristics. Considering the U.S. state governments as an example for followers, the following components in η_{nt} can be included.

Example 4.2.1 *Let's pick an arbitrary state government i. First, η_i^{iv} might include state i's geographic characteristics such as (i) the number of bordering states, (ii) land/water areas, etc. For time varying characteristics, second, state i's time-varying demographic and economic characteristics can be included: by Case et al. (1993), (i) total revenue, (ii) grants from the federal government, (iii) population density, (iv) unemployment rate, and (v) compositions of state's population. In our model specification, (ii) the federal grants are the leader's choice variables.*

[134]For example, equation (4.1) can take a reduced form of a SDPD model specification. We will rigorously characterize A_n, B_n, and C_n in the next step. Since we pursue a stable economic environment, the solutions A_n, B_n, and C_n would be time-invariant.

We assume that there is uncertainty in future exogenous characteristics $\eta_{n,t+1}, \eta_{n,t+2}, \cdots$, so both types of agents rationally expect them in their decision-making. For this, suppose that η_{nt} follows a first-order linear Markov process: $\eta_{nt} = \pi_0 \eta_{n,t-1} + \xi_{nt}$ where $\xi_{nt} \sim i.i.d.\left(0_{n\times 1}, \Omega_{\xi}\right)$. For the part of η_{nt}^{v}, we assume that $|\pi_0| < 1$ for stationary and Ω_{ξ} is positive definite. For the part of η_{n}^{iv}, $\pi_0 = 1$ and $\Omega_{\xi} = 0$.

A justification for equation (4.1) comes from the following linear-quadratic (LQ) payoff[135]: for $i = 1, \cdots, n$ and for each t,

$$\begin{aligned} & u_i\left(y_{it}, Y_{-i,t}, ; Y_{n,t-1}, b_{it}, \eta_{it}\right) \\ = \; & \left(\eta_{it} + \phi_{i,0} b_{it} + \rho_0 w_{i.} Y_{n,t-1} + \lambda_0 w_{i.} Y_{nt}\right) y_{it} - c\left(y_{it}, y_{i,t-1}\right) \end{aligned} \tag{4.2}$$

where $c\left(y_{it}, y_{i,t-1}\right) = \frac{\gamma_0}{2}\left(y_{it} - y_{i,t-1}\right)^2 + \frac{1-\gamma_0}{2} y_{it}^2$ with $0 < \gamma_0 < 1$, and $Y_{-i,t} = \left(y_{1t}, \cdots, y_{i-1,t}, y_{i+1,t}, \cdots, y_{nt}\right)'$. Parameter $\phi_{i,0}$ shows a direct effect of the grants on the follower i's marginal payoff. If a follower is myopic, the first order condition of the follower i's maximization problem yields his/her best response,

$$y_{it}^{*}\left(Y_{n,t-1}, b_{it}, \eta_{it}\right) = \lambda_0 w_{i.} Y_{nt} + \gamma_0 y_{i,t-1} + \rho_0 w_{i.} Y_{n,t-1} + \phi_{i,0} b_{it} + \eta_{it}. \tag{4.3}$$

Without $\phi_{i,0} b_{it}$, equation (4.3) can represent a conventional spatial dynamic panel data (SDPD) model (e.g., Yu et al. (2008) and Lee and Yu (2010)). A new term in this paper is $\phi_{i,0} b_{it}$, and the parameter $\phi_{i,0}$ can represent follower i's (financial) dependency upon the leader. We assume that $\phi_{i,0} = \phi_0 \Lambda_{d,i}$ where $\Lambda_{d,i}$ is a function of geographic distance between i and 0. We will introduce more specification on $\phi_{i,0}$ later. Through $w_{i.}$, other followers' previous and current actions can affect the follower i's t^{th}-period marginal payoff. Directions of influences are governed by two

[135] We consider a (dynamic extension of) parametric linear-quadratic payoff function introduced by Ballester et al. (2006), and Calvo-Armengol et al. (2009).

parameters, λ_0 and ρ_0. If $\lambda_0 > 0$, a follower's action is reinforced by neighboring followers' current actions. On the other hand, neighboring followers' current actions offset each other if $\lambda_0 < 0$. Similar interpretations can be done for ρ_0. To avoid having an extreme option, a cost function specification $(c(y_{it}, y_{i,t-1}))$ exists. It consists of two parts and is governed by parameter γ_0. The first part is a dynamic adjustment cost showing persistency of actions, and the second part shows a cost of selecting a level of activity. The parameter γ_0 captures a (relative) weight for the dynamic adjustment cost. If γ_0 is large, the followers' actions are persistent.

Based on payoffs (4.2), we consider the leader's payoff function. The leader's per period payoff is defined by the summation of followers' payoffs and costs of selecting $b_{1t}, \cdots, b_{nt}$: for each t,

$$\begin{aligned} & \mathcal{W}_{0,t}\left(\mathbf{b}_{nt}; Y_{nt}, Y_{n,t-1}, \eta_{nt}, \tau_{nt}\right) \\ = \; & \sum_{k=1}^{n} u_k\left(Y_{nt}, ; Y_{n,t-1}, b_{kt}, \eta_{kt}\right) - \frac{1}{2}\sum_{k=1}^{n} b_{kt}^2 + \sum_{k=1}^{n} \tau_{kt} b_{kt} \end{aligned} \tag{4.4}$$

where $\tau_{nt} = (\tau_{1t}, \cdots, \tau_{nt})'$, τ_{kt} is the autonomous payment for the follower k at period t. We assume that τ_{kt}s are not relevant to actions of both types of agents. The first part of $\mathcal{W}_{0,t}(\cdot)$ represents social benefits defined by the summation of the followers' payoffs. The second part describes the cost of selecting $b_{1t}, \cdots, b_{nt}$. The third part, $\sum_{k=1}^{n} \tau_{kt} b_{kt}$, represents an incentive of giving the autonomous payments. When $\phi_0 = 0$ without the third component, the optimal grant level for every follower will be zero since taking $b_{kt} > 0$ in this case only raises the quadratic cost $-\frac{1}{2}b_{kt}^2$. The amount of autonomous payment τ_{kt} might depend on aggregate economic shocks and k's innate characteristics. Hence, τ_{kt} is a function of η_k^{iv} and η_{kt}^{v}. In Section 4.3, we will introduce a specification of τ_{kt}.

Note that the leader's choice variables at time t are $\mathbf{b}_{nt}$ given the initial conditions $(Y_{n,t-1}, \eta_{nt})$. For each t, the leader knows that the followers take the strategy (4.1). By observing the leader's payoff (4.4), we can see that there are two channels of b_{it} affecting $\mathcal{W}_{0,t}(\cdot)$. First, selecting b_{it} directly affects (i) $u_i(\cdot)$ via $\phi_{i,0} b_{it} y_{it}$, (ii) $-\frac{1}{2} b_{it}^2$, and (iii) $\tau_{it} b_{it}$. Second, b_{it} can affect $\mathcal{W}_{0,t}(\cdot)$ through changing followers' actions. Note that $\frac{\partial y_{jt}}{\partial b_{it}} = B_{n,ji}$ where $B_{n,ji}$ denotes the (j,i)-element of B_n. Hence, the diagonal element of B_n, $B_{n,ii}$, shows the direct influence of b_{it} on y_{it} in the rational expectation equilibrium. Since the leader's payoff is also LQ in its actions as well as state variables, the leader also adopts the linear optimal policy function of $(Y_{n,t-1}, \eta_{nt})$:[136] for each t

$$\mathbf{b}_{nt}^*(Y_{n,t-1}, \eta_{nt}, \tau_{nt}) = D_n Y_{n,t-1} + E_n \eta_{nt} + F_n \tau_{nt}$$

where D_n, E_n and F_n are $n \times n$ matrices specified in the upcoming subsection.

4.2.2 Intertemporal decision-making

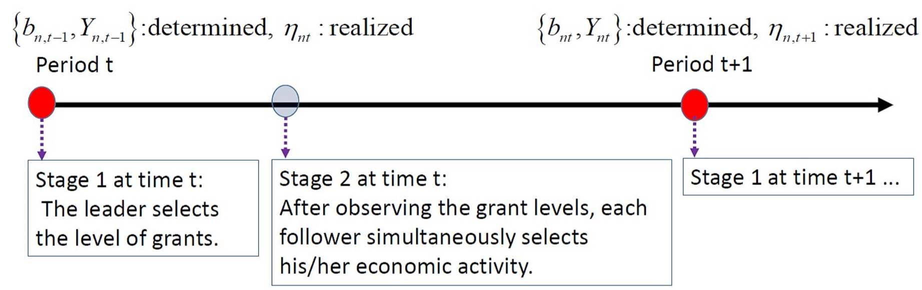
In this subsection, we define the intertemporal choice problems of the two-type agents. The matrices A_n, B_n, C_n, D_n, E_n and F_n will be verified as the solutions to the leader's and followers' intertemporal optimization problems.

At time t, we assume that all agents know all past actions $\{\mathbf{b}_{ns}, Y_{ns}\}_{s=-\infty}^{s=t-1}$ and all realized exogenous characteristics $\{\eta_{ns}\}_{s=-\infty}^{s=t}$ (perfect recall and complete information up to the current period). Based on this information setting, the conditional expectation operator $\mathbf{E}_t(\cdot)$ is defined. In the economy, uncertainty arises due to the future exogenous characteristics $\eta_{n,t+1}$, $\eta_{n,t+2}$, $\cdots$. Expectations for the future exogenous characteristics are rationally (i.e., mathematically) formed through $\mathbf{E}_t(\cdot)$

[136] Note that τ_{nt} is a function of η_{nt}. In order to highlight a role of τ_{nt}, we include τ_{nt} as an argument for $\mathbf{b}_{nt}^*(\cdot)$.

(e.g., $\mathbf{E}_t(\eta_{n,t+1}) = \pi_0\eta_{nt}$). We suppose that $\mathbf{E}_t(\tau_{n,t+1}) = \tilde{\pi}_0\tau_{nt}$ with $|\tilde{\pi}_0| < 1$ and $Var_t(\tau_{n,t+1}) = \Omega_\tau > 0$ where $\tau_n = (\tau_1, \cdots, \tau_n)'$. That is, for each t $\tau_{n,t+1} = \tilde{\pi}_0\tau_{nt} + \xi^\tau_{nt}$ with $\mathbf{E}_t(\xi^\tau_{nt}) = \mathbf{0}_{n\times 1}$. Under this rational expectation framework, the timeline of intertemporal decision-making can be described by the following figure:

Figure 4.1: The timeline of intertemporal decision-making



For each t, $\{\mathbf{b}_{n,t-1}, Y_{n,t-1}\}$ and η_{nt} are determined or realized, and there are two stages of decision-making. In the first stage, the leader chooses the optimal level of grants ($\mathbf{b}^*_{nt}$) for the followers. In the second stage, after observing $\mathbf{b}^*_{nt}$, the followers simultaneously choose their optimal actions Y^*_{nt}. And then, $\{\mathbf{b}_{nt}, Y_{nt}, \eta_{n,t+1}\}$ are determined/realized ($\mathbf{b}_{nt} = \mathbf{b}^*_{nt}$ and $Y_{nt} = Y^*_{nt}$), and the first stage of the $(t+1)^{th}$-period will be open.

Now we formally define the lifetime optimization problems of the two-type agents. Given $(Y_{n,t-1}, \eta_{nt})$, for each t the leader chooses $\{\mathbf{b}_{ns}\}_{s=t}^{\infty}$ by maximizing

$$\begin{aligned} & \mathcal{W}_{0,t}\left(\mathbf{b}_{nt}; Y_{nt}, Y_{n,t-1}, \eta_{nt}, \tau_{nt}\right) \\ & + \mathbf{E}_t \sum_{s=1}^{\infty} \delta^s \mathcal{W}_{0,t+s}\left(\mathbf{b}_{n,t+s}; Y_{n,t+s}, Y_{n,t+s-1}, \eta_{n,t+s}, \tau_{n,t+s}\right) \end{aligned} \tag{4.5}$$

with knowing $Y_{nt}^* = A_n Y_{n,t-1} + B_n \mathbf{b}_{nt} + C_n \eta_{nt}$.[137] Since we consider a stable economic environment, the maximization problem (4.5) can be represented by the recursive relationship (Bellman equation): given $(Y_{n,t-1}, \eta_{nt})$

$$V^L\left(Y_{n,t-1}, \eta_{nt}, \tau_{nt}\right) = \max_{\mathbf{b}_{nt}} \left\{ \begin{array}{c} \mathcal{W}_{0,t}\left(\mathbf{b}_{nt}; Y_{nt}^*, Y_{n,t-1}, \eta_{nt}, \tau_{nt}\right) \\ + \delta \mathbf{E}_t V^L\left(Y_{nt}^*, \eta_{n,t+1}, \tau_{n,t+1}\right) \end{array} \right\} \tag{4.6}$$

subject to $Y_{nt}^* = A_n Y_{n,t-1} + B_n \mathbf{b}_{nt} + C_n \eta_{nt}$. Suppose that A_n, B_n, and C_n are given[138], and they are going to be revealed by solving the followers' problems. Since (i) $\mathcal{W}_{0,t}(\cdot)$ is linear-quadratic in $\mathbf{b}_{nt}$ and (ii) $\mathbf{b}_{nt}^*$ is an affine function of $(Y_{n,t-1}, \eta_{nt}, \tau_{nt})$, the value $V^L(Y_{n,t-1}, \eta_{nt}, \tau_{nt})$ would take a LQ form of $(Y_{n,t-1}, \eta_{nt}, \tau_{nt})$:

$$\begin{aligned} V^L\left(Y_{n,t-1}, \eta_{nt}, \tau_{nt}\right) &= Y_{n,t-1}' Q_n^L Y_{n,t-1} + Y_{n,t-1}' L_n^L \eta_{nt} + Y_{n,t-1}' L_n^{L,\tau} \tau_{nt} \\ &\quad + \eta_{nt}' Q_n^{L,\eta} \eta_{nt} + \tau_{nt}' Q_n^{L,\tau} \tau_{nt} + \tau_{nt}' L_n^{L,\tau,\eta} \eta_{nt} + c_n^L \end{aligned}$$

where $n \times n$ matrices Q_n^L, L_n^L, $L_n^{L,\tau}$, $Q_n^{L,\eta}$, $Q_n^{L,\tau}$, and $L_n^{L,\tau,\eta}$ and a scalar c_n^L are the solutions to the algebraic Riccati equations. Verifying and computing the forms of them can be found in Appendix C.

[137] That is, the leader knows (rationally expects) the followers' optimal actions at the same decision-making period. Then, $Y_{nt}^* = A_n Y_{n,t-1} + B_n \mathbf{b}_{nt} + C_n \eta_{nt}$ plays a role as a linear constraint.

[138] In practice, by applying backward induction, A_n, B_n, and C_n are firstly verified.

Now we can reveal D_n, E_n and F_n from equation (4.6). The first order condition[139] is

$$\begin{aligned}\mathbf{0}_{n\times 1} &= \Phi_{n,0}Y_{nt}^{*}-\mathbf{b}_{nt}^{*}+\tau_{nt} \\ &\quad +B_n'\left[\eta_{nt}+\Phi_{n,0}\mathbf{b}_{nt}^{*}+\left(\gamma_0 I_n+\rho_0 W_n\right)Y_{n,t-1}-S_n^L Y_{nt}^{*}\right] \\ &\quad +\delta B_n'\left[Q_n^{L,*}Y_{nt}^{*}+\pi_0 L_n^L\eta_{nt}+\tilde{\pi}_0 L_n^{L,\tau}\tau_{nt}\right]\end{aligned} \tag{4.7}$$

where $\Phi_{n,0}=diag\left(\phi_{1,0},\cdots,\phi_{n,0}\right)$, $S_n^L=I_n-\lambda_0\left(W_n+W_n'\right)$, and $Q_n^{L,*}=Q_n^L+Q_n^{L\prime}$. By (4.7) and the form of Y_{nt}^{*}, we have

$$\begin{aligned}&\left(I_n-B_n'\Phi_{n,0}\right)\mathbf{b}_{nt}^{*} \\ =&\left(\Phi_{n,0}-B_n'\left(S_n^L-\delta Q_n^{L,*}\right)\right)Y_{nt}^{*}+B_n'\left(\gamma_0 I_n+\rho_0 W_n\right)Y_{n,t-1} \\ &+B_n'\left(I_n+\delta\pi_0 L_n^L\right)\eta_{nt}+\left(I_n+\delta\tilde{\pi}_0 B_n' L_n^{L,\tau}\right)\tau_{nt}\end{aligned}$$

implying

$$\begin{aligned}&\left[I_n-B_n'\Phi_{n,0}-\left(\Phi_{n,0}-B_n'\left(S_n^L-\delta Q_n^{L,*}\right)\right)B_n\right]\mathbf{b}_{nt}^{*} \\ =&\left[\left(\Phi_{n,0}-B_n'\left(S_n^L-\delta Q_n^{L,*}\right)\right)A_n+B_n'\left(\gamma_0 I_n+\rho_0 W_n\right)\right]Y_{n,t-1} \\ &+\left[B_n'\left(I_n+\delta\pi_0 L_n^L\right)+\left(\Phi_{n,0}-B_n'\left(S_n^L-\delta Q_n^{L,*}\right)\right)C_n\right]\eta_{nt} \\ &+\left(I_n+\delta\tilde{\pi}_0 B_n' L_n^{L,\tau}\right)\tau_{nt}.\end{aligned} \tag{4.8}$$

Hence, we obtain

$$\begin{aligned}D_n &= \left(R_n^L\right)^{-1}\left[\left(\Phi_{n,0}-B_n'\left(S_n^L-\delta Q_n^{L,*}\right)\right)A_n+B_n'\left(\gamma_0 I_n+\rho_0 W_n\right)\right], \\ E_n &= \left(R_n^L\right)^{-1}\left[B_n'\left(I_n+\delta\pi_0 L_n^L\right)+\left(\Phi_{n,0}-B_n'\left(S_n^L-\delta Q_n^{L,*}\right)\right)C_n\right],\end{aligned}$$

[139]This comes from

$$\begin{aligned}\mathbf{0}_{n\times 1} &= \frac{\partial\mathcal{W}_{0,t}\left(\mathbf{b}_{nt};Y_{nt}^{*},Y_{n,t-1},\eta_{nt},\tau_{nt}\right)}{\partial\mathbf{b}_{nt}}+\frac{\partial Y_{nt}^{*\prime}}{\partial\mathbf{b}_{nt}}\frac{\partial\mathcal{W}_{0,t}\left(\mathbf{b}_{nt};Y_{nt}^{*},Y_{n,t-1},\eta_{nt},\tau_{nt}\right)}{\partial Y_{nt}^{*}} \\ &\quad +\delta\mathbf{E}_t\frac{\partial Y_{nt}^{*\prime}}{\partial\mathbf{b}_{nt}}\frac{\partial V^L\left(Y_{nt}^{*},\eta_{n,t+1},\tau_{n,t+1}\right)}{\partial Y_{nt}^{*}},\end{aligned}$$

and $\frac{\partial}{\partial Y_{nt}}\mathbf{E}_t Y_{nt}' L_n^{L,\tau}\tau_{n,t+1}=\mathbf{E}_t L_n^{L,\tau}\tau_{n,t+1}=L_n^{L,\tau}\tau_n$.

and

$$F_n = \left(R_n^L\right)^{-1}\left(I_n + \delta\tilde{\pi}_0 B_n' L_n^{L,\tau}\right)$$

where $R_n^L = I_n - B_n'\Phi_{n,0} - \left(\Phi_{n,0} - B_n'\left(S_n^L - \delta Q_n^{L,*}\right)\right)B_n$.

At time t, after observing the leader's optimal decisions on grants ($\mathbf{b}_{nt}^*$), each follower i selects y_{it} by maximizing his/her lifetime payoff: given $(Y_{n,t-1}, \eta_{nt})$

$$u_i\left(y_{it}, Y_{-i,t},; Y_{n,t-1}, b_{it}, \eta_{it}\right) + \mathbf{E}_t\sum_{s=1}^{\infty}\delta^s u_i\left(y_{i,t+s}, Y_{-i,t+s},; Y_{n,t+s-1}, b_{i,t+s}, \eta_{i,t+s}\right). \quad (4.9)$$

Note that the followers choose their current economic actions by rationally expecting (i) the leader's and followers' future optimal actions $\left\{\mathbf{b}_{n,t+s}^*, Y_{n,t+s}^*\right\}_{s=1}^{\infty}$, and (ii) the future exogenous characteristics $\{\eta_{n,t+s}\}_{s=1}^{\infty}$. For example, at time t, a follower expects the $(t+1)^{th}$-period leader's optimal grant decisions by

$$\mathbf{E}_t\mathbf{b}_{n,t+1}^* = D_n Y_{nt} + \pi_0 E_n\eta_{nt} + \tilde{\pi}_0 F_n\tau_{nt}, \quad (4.10)$$

which is a function of Y_{nt}. In the rational expectation equilibrium, there are two notable findings: (i) current followers' actions (Y_{nt}) are affected by the leader's expected future actions ($\mathbf{E}_t\mathbf{b}_{n,t+1}^*$)[140], and (ii) the leader confirms the forecasts $\mathbf{E}_t\mathbf{b}_{n,t+1}^* = D_n Y_{nt} + \pi_0 E_n\eta_{nt} + \tilde{\pi}_0 F_n\tau_{nt}$ and determines $\mathbf{b}_{n,t+1}^* = D_n Y_{nt} + E_n\eta_{n,t+1} + F_n\tau_{n,t+1}$ at time $t+1$. To the leader, that is, using the same D_n, E_n and F_n plays a role as a constraint (commitment). Then, at period t, the expectation error would be $\mathbf{b}_{n,t+1}^* - \mathbf{E}_t\mathbf{b}_{n,t+1}^* = E_n\xi_{n,t+1} + F_n\xi_{n,t+1}^{\tau}$, which is a linear function of the $(t+1)^{th}$-period unexpected exogenous shocks. Since we will specify τ_{nt} (and its affine transformation) as interactive unobserved effects in estimation, the nuisance parameter $\tilde{\pi}_0$ has no meaning. Hence, we assume $\tilde{\pi}_0 = 0$.

[140]That is, each follower knows that he/she can affect $\mathbf{E}_t\mathbf{b}_{n,t+1}^*$ in his/her current decision-making.

Under the stable economic environment, the follower i's lifetime problem can be also shown as the Bellman equation:[141]

$$V_i^F\left(Y_{n,t-1},\mathbf{b}_{nt}^*,\eta_{nt}\right)=\max_{y_{it}}\left\{\begin{array}{c}u_i\left(y_{it},Y_{-i,t}^*,;Y_{n,t-1},b_{it}^*,\eta_{it}\right)\\+\delta\mathbf{E}_tV_i^F\left(y_{it},Y_{-i,t}^*,\mathbf{b}_{n,t+1}^*,\eta_{n,t+1}\right)\end{array}\right\}\tag{4.11}$$

such that $\mathbf{b}_{n,t+1}^*=D_nY_{nt}^*+E_n\eta_{n,t+1}+F_n\tau_{n,t+1}$. Since $u_i(\cdot)$ is LQ in actions of both types of agents, their optimal actions are linear in state variables. For each $i=1,\cdots,n$, hence $V_i^F(\cdot)$ takes a LQ function of its argument:

$$\begin{aligned}V_i^F\left(Y_{n,t-1},\mathbf{b}_{nt}^*,\eta_{nt}\right)&=Y_{n,t-1}'Q_i^FY_{n,t-1}+Y_{n,t-1}'L_i^{F,b}\mathbf{b}_{nt}^*+Y_{n,t-1}'L_i^{F,\eta}\eta_{nt}\\&\quad+\mathbf{b}_{nt}^{*\prime}Q_i^{F,b}\mathbf{b}_{nt}^*+\mathbf{b}_{nt}^{*\prime}L_i^{F,b,\eta}\eta_{nt}+\eta_{nt}'Q_i^{F,\eta}\eta_{nt}+c_i^F\end{aligned}$$

where $n\times n$ matrices Q_i^F, $L_i^{F,b}$, $L_i^{F,\eta}$, $Q_i^{F,b}$, $L_i^{F,b,\eta}$, and $Q_i^{F,\eta}$ and a scalar c_i^F are the solutions to the algebraic matrix Riccati equations. The formulas of them can be found in Appendix C. The first order condition of follower i's lifetime problem is

$$\begin{aligned}0&=\eta_{it}+\phi_{i,0}b_{it}^*+e_i'\left(\gamma_0I_n+\rho_0W_n\right)Y_{n,t-1}-e_i'S_nY_{nt}^*\\&\quad+\delta e_i'\left(Q_i^F+Q_i^{F\prime}\right)Y_{nt}^*+\delta e_i'L_i^{F,b}\mathbf{E}_t\mathbf{b}_{n,t+1}^*+\delta e_i'L_i^{F,\eta}\mathbf{E}_t\eta_{n,t+1}\end{aligned}\tag{4.12}$$

where e_i denotes the i^{th} unit vector, and $S_n=I_n-\lambda_0W_n$. Note that the followers are not able to change the leader's same period actions $\mathbf{b}_{nt}^*$.

Based on (4.12), we characterize A_n, B_n, and C_n. For each $i=1,\cdots,n$, define the $n\times n$ matrices $Q_n^{F,*}$, $L_n^{F,b,*}$, and $L_n^{F,\eta,*}$ such that $e_i'Q_n^{F,*}=e_i'\left(Q_i^F+Q_i^{F\prime}\right)$, $e_i'L_n^{F,b,*}=e_i'L_i^{F,b}$, and $e_i'L_n^{F,\eta,*}=e_i'L_i^{F,\eta}$. Using (4.10), equation (4.12) yields

$$\begin{aligned}\left[S_n-\delta Q_n^{F,*}-\delta L_n^{F,b,*}D_n\right]Y_{nt}^*&=\left(\gamma_0I_n+\rho_0W_n\right)Y_{n,t-1}\\&\quad+\Phi_{n,0}\mathbf{b}_{nt}^*+\left[I_n+\delta\pi_0\left(L_n^{F,b,*}E_n+L_n^{F,\eta,*}\right)\right]\eta_{nt}.\end{aligned}\tag{4.13}$$

[141]When $\tilde{\pi}_0=0$, the arguments of $V_i^F(\cdot)$ are only $Y_{n,t-1}$, $\mathbf{b}_{nt}^*$, and η_{nt}.

Hence, we obtain

$$A_n = \left(R_n^F\right)^{-1}\left(\gamma_0 I_n + \rho_0 W_n\right), B_n = \left(R_n^F\right)^{-1}\Phi_{n,0},$$

and $C_n = \left(R_n^F\right)^{-1}\left[I_n + \delta\pi_0\left(L_n^{F,b,*}E_n + L_n^{F,\eta,*}\right)\right]$ where $R_n^F = S_n - \delta Q_n^{F,*} - \delta L_n^{F,b,*}D_n$. Note that $\mathbf{b}_{nt}^*$ is endogenous since $C_n\eta_{nt}$ and $E_n\eta_{nt}$ might be correlated. Then, the model's representation as exogenous variables is

$$\begin{aligned} R_n^F Y_{nt}^* &= \left(\gamma_0 I_n + \rho_0 W_n + \Phi_{n,0}D_n\right)Y_{n,t-1} \\ &\quad + \left(I_n + \Phi_{n,0}E_n + \delta\pi_0\left(L_n^{F,b,*}E_n + L_n^{F,\eta,*}\right)\right)\eta_{nt} + \Phi_{n,0}F_n\tau_{nt}. \end{aligned} \tag{4.14}$$

4.3 Econometric model

In this section, we establish an econometric model based on rational expectation equilibrium equations (4.8) and (4.13).[142] From this section, we drop the superscript "*" since the observed actions $(\mathbf{b}_{nt}, Y_{nt})$ are assumed to be optimally realized. We want to estimate the structural parameters λ_0, γ_0, ρ_0, and $\{\phi_{i,0}\}_{i=1}^n$ based on a panel data set $\{Y_{nt}, X_{nt}\}_{t=0}^T$, and $\{\mathbf{b}_{nt}\}_{t=1}^T$ with observed (or prespecified) W_n.[143]

First, we consider stability of equations (4.8) and (4.13) to have a well-defined log likelihood function (variance structure). Second, we will characterize the parameters $\phi_{i,0}$ (showing dependencies upon the leader) as a function of geographic distances between the leader and the followers. Next, we will derive the log-likelihood function by giving a structure to η_{nt} and τ_{nt}. After that, we discuss implementing the (quasi) maximum likelihood (QML) estimation method. We assume that there are n spatial

[142]Our estimation is based on equation (4.14). To implement this, note that we also need to recover equation (4.8).

[143]We do not attempt to estimate the time-discounting factor δ for easy identification.

units and T periods in a sample. Our estimation is based on the large n and large T framework.[144]

4.3.1 Stability

A motivation of considering stability of the system is to characterize manageable dependence across space and time. Our economic model is based on the time stable environment with given n. To obtain both space and time stability[145] of equations (4.8) and (4.13), we need to additionally impose the following sufficient conditions[146]: $\|A_n\|_s < 1$, $\|A_n + B_n D_n\|_s < 1$, and $\|B_n E_n + C_n\|_s \leq c_f$ where $\|\cdot\|_s$ denotes the spectral matrix norm, and $c_f > 0$ is an uniformly bounded (in n) constant. Under $\|A_n + B_n D_n\|_s < 1$, and $\max\{\|B_n E_n + C_n\|_s, \|B_n F_n\|_s\} \leq c_f$,

$$Y_{nt} = \sum_{s=0}^{\infty} (A_n + B_n D_n)^s \{(B_n E_n + C_n)\eta_{n,t-s} + B_n F_n \tau_{n,t-s}\}, \tag{4.15}$$

which is useful to capture the variance structure of Y_{nt}. To recover the variance structure of Y_{nt}, observe that verifying the optimal actions of both types of agents is required ($\{A_n, B_n, C_n\}$ as well as $\{D_n, E_n, F_n\}$). Consider the leader's decision variables. If $\|A_n\|_s < 1$, by infinite substitution,

$$\mathbf{b}_{nt} = \sum_{s=1}^{\infty} D_n A_n^{s-1} B_n \mathbf{b}_{n,t-s} + E_n \eta_{nt} + F_n \tau_{nt} + \sum_{s=1}^{\infty} D_n A_n^{s-1} C_n \eta_{n,t-s}. \tag{4.16}$$

By observing (4.16), note that $\mathbf{b}_{nt}$ relies on the entire histories $\{\mathbf{b}_{ns}\}_{s=t-1,t-2,\cdots}$, and all realized exogenous characteristics $\{\eta_{ns}\}_{s=t,t-1,\cdots}$ (and τ_{nt}), which they do not appear the leader's optimization problem.

[144]Note that our theoretical model is based on given n. From this section (estimation part), we consider both large time series observations as well as spatial units.

[145]It means that the equilibrium system is stable regardless of the number of agents n.

[146]When $\delta = 0$, stability conditions can be represented by a function of parameters and eigenvalues of W_n. If $0 < \delta < 1$, however, it is difficult to find simple stability conditions due to highly nonlinearity in parameters.

Now we can evaluate $Var\left(Y_{nt}\right)$ and $Var\left(\mathbf{b}_{nt}\right)$ using equation (4.15) and assuming $\eta_{nt} \sim i.i.d.\left(0, \sigma_0^2 I_n\right)$.[147] The importance of verifying the variance structures is to find additional sources of identification. Note that

$$Var\left(Y_{nt}\right) = \sigma_0^2 \sum_{j=0}^{\infty} \left(A_n + B_n D_n\right)^j \left(B_n E_n + C_n\right) \left(B_n E_n + C_n\right)' \left(A_n' + D_n' B_n'\right)^j ,$$

and

$$\begin{aligned} & Var\left(\mathbf{b}_{nt}\right) \\ = \;& \sigma_0^2 \left[\begin{array}{c} E_n E_n' \\ + \sum_{j=1}^{\infty} D_n \left(A_n + B_n D_n\right)^{j-1} \left(B_n E_n + C_n\right) \\ \cdot \left(B_n E_n + C_n\right)' \left(A_n' + D_n' B_n'\right)^{j-1} D_n' \end{array} \right]. \end{aligned}$$

As a special case, consider $\delta = 0$ (myopic agents). For each period t, the two types of agents play the two-stage game, which can be solved by backward induction. By firstly solving the followers' problems, we have

$$Y_{nt} = S_n^{-1} \left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n,t-1} + S_n^{-1} \Phi_{n,0} \mathbf{b}_{nt} + S_n^{-1} \eta_{nt} \tag{4.17}$$

for time t as the subgame perfect Nash equilibrium (SPNE) equation. Then, $A_n = S_n^{-1} \left(\gamma_0 I_n + \rho_0 W_n\right)$, $B_n = S_n^{-1} \Phi_{n,0}$, and $C_n = S_n^{-1}$. Equation (4.17) follows a reduced form of spatial dynamic panel data (SDPD) models with considering that $\mathbf{b}_{nt}$ is endogenous. The vector of leader's optimal actions is

$$\begin{aligned} & \left[I_n - \Phi_{n,0} \left(S_n S_n'\right)^{-1} \Phi_{n,0}\right] \mathbf{b}_{nt} \\ = \;& \Phi_{n,0} \left(S_n S_n'\right)^{-1} \left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n,t-1} + \Phi_{n,0} \left(S_n S_n'\right)^{-1} \eta_{nt} + \tau_{nt} \end{aligned}$$

since

$$I_n - B_n' \Phi_{n,0} - \Phi_{n,0} B_n + B_n' S_n^L B_n = I_n - \Phi_{n,0} \left(S_n S_n'\right)^{-1} \Phi_{n,0},$$

[147] We will give statistical disturbances as a part of η_{nt}. For τ_{nt}, we will consider a structure of interactive fixed effects.

$$\left(\left(\Phi_{n,0}-B_n'S_n^L\right)S_n^{-1}+B_n'\right)(\gamma_0 I_n+\rho_0 W_n)$$
$$=\ \Phi_{n,0}\left(S_nS_n'\right)^{-1}(\gamma_0 I_n+\rho_0 W_n),$$

and $B_n'+\left(\Phi_{n,0}-B_n'S^L\right)C_n=\Phi_{n,0}\left(S_n^{-1}+S_n^{-1\prime}-S_n^{-1\prime}S_n^LS_n^{-1}\right)=\Phi_{n,0}\left(S_nS_n'\right)^{-1}$ in this case.[148] That is,

$$D_n=\left[I_n-\Phi_{n,0}\left(S_nS_n'\right)^{-1}\Phi_{n,0}\right]^{-1}\Phi_{n,0}\left(S_nS_n'\right)^{-1}(\gamma_0 I_n+\rho_0 W_n),$$

$E_n=\left[I_n-\Phi_{n,0}\left(S_nS_n'\right)^{-1}\Phi_{n,0}\right]^{-1}\Phi_{n,0}\left(S_nS_n'\right)^{-1}$, and $F_n=\left[I_n-\Phi_{n,0}\left(S_nS_n'\right)^{-1}\Phi_{n,0}\right]^{-1}$.

If $\Phi_{n,0}=\phi_0 I_n$, $\|W_n\|_\infty=1$, and $\|W_n\|_1\leq c_w<\infty$, we can find a sufficient condition for stability. Observe that

$$\begin{aligned}A_n+B_nD_n &= S_n^{-1}(\gamma_0 I_n+\rho_0 W_n)+\phi_0^2S_n^{-1}\left(I_n-\phi_0^2\left(S_nS_n'\right)^{-1}\right)^{-1}\\
&\quad\cdot\left(S_nS_n'\right)^{-1}(\gamma_0 I_n+\rho_0 W_n)\\
&= S_n^{-1}\left(I_n+\phi_0^2\left(I_n-\phi_0^2\left(S_nS_n'\right)^{-1}\right)^{-1}\left(S_nS_n'\right)^{-1}\right)(\gamma_0 I_n+\rho_0 W_n)\\
&= S_n^{-1}\left(I_n-\phi_0^2\left(S_nS_n'\right)^{-1}\right)^{-1}(\gamma_0 I_n+\rho_0 W_n)\end{aligned}$$

by using a Neumann series expansion, $\left(I_n-\phi_0^2\left(S_nS_n'\right)^{-1}\right)^{-1}=\sum_{k=0}^{\infty}\phi_0^{2k}\left(S_n'^{-1}S_n^{-1}\right)^k$.[149] As a sufficient condition, note that we need to have $\frac{|\phi_0|^2}{(1-|\lambda_0|)(1-|\lambda_0|c_w)}<1$ to have the

[148]Note that

$$\begin{aligned}S_n^{-1}+S_n^{-1\prime}-S_n^{-1\prime}S_n^LS_n^{-1} &= S_n^{-1}+S_n^{-1\prime}-S_n^{-1\prime}\left(S_n-\lambda_0W_n'\right)S_n^{-1}\\
&= S_n^{-1}+S_n^{-1\prime}-S_n^{-1\prime}\left(I_n-\lambda_0W_n'S_n^{-1}\right)\\
&= \left(I_n-\lambda_0S_n^{-1\prime}W_n'\right)S_n^{-1}\\
&= S_n^{-1\prime}S_n^{-1}=\left(S_nS_n'\right)^{-1}\end{aligned}$$

since $I_n-\lambda_0S_n^{-1\prime}W_n'=S_n^{-1\prime}$.

[149]Then,

$$\begin{aligned}\left\|\left(I_n-\phi_0^2\left(S_nS_n'\right)^{-1}\right)^{-1}\right\|_\infty &\leq \sum_{k=0}^{\infty}|\phi_0|^{2k}\left\|S_n'^{-1}S_n^{-1}\right\|_\infty^k\\
&\leq \sum_{k=0}^{\infty}\left(|\phi_0|^2\left\|S_n^{-1}\right\|_1\left\|S_n^{-1}\right\|_\infty\right)^k\end{aligned}$$

Neumann series expansion. To have a stable system, that is, $|\phi_0|$ is also needed to be manageable. In consequence,

$$\begin{aligned} & \|A_n + B_n D_n\|_\infty \\ = & \left\| S_n^{-1} \left(I_n - \phi_0^2 \left(S_n S_n' \right)^{-1} \right)^{-1} \left(\gamma_0 I_n + \rho_0 W_n \right) \right\|_\infty \\ \leq & \frac{\gamma_0 + |\rho_0|}{1 - |\lambda_0|} \cdot \frac{(1 - |\lambda_0|)(1 - |\lambda_0| c_w)}{(1 - |\lambda_0|)(1 - |\lambda_0| c_w) - |\phi_0|^2} < 1. \end{aligned}$$

Since $\frac{\gamma_0 + |\rho_0|}{1 - |\lambda_0|} < 1$ is a sufficient stability condition of the conventional SDPD model and $\frac{(1-|\lambda_0|)(1-|\lambda_0| c_w)}{(1-|\lambda_0|)(1-|\lambda_0| c_w) - |\phi_0|^2} > 1$, a stability condition of our model with $\delta = 0$ will be stricter than the conventional one.

4.3.2 Topological specification for $\phi_{i,0}$

In this subsection, we specify the unilateral influences from the leader and the followers by the parameters $\phi_{i,0}$'s. We will characterize $\phi_{i,0}$ as a known function of geographic arrangements.[150] Recall that there are $n+1$ spatial units: 0 denotes the leader while $i = 1, \cdots, n$ denote the followers. We assume that they have innate locations and (possibly) unevenly placed in $\mathbf{R}^d$ ($d \geq 1$). Let G be a lattice in $\mathbf{R}^d$, which is a set of potential locations of spatial units.[151] A subset $G_n \subset G$ denotes a set of locations relevant to observed (sample) spatial units. The location function is a one-to-one and onto mapping from $\{0, 1, \cdots, n\}$ to G_n: i.e., $l : \{0, 1, \cdots, n\} \mapsto G_n$ and $l(0)$, $l(1)$, $\cdots$, $l(n) \in G_n$. As a metric, we can consider the Euclidean

$$\leq \frac{1}{1 - |\phi_0|^2 / (1 - |\lambda_0|)(1 - |\lambda_0| c_w)}$$

if $\frac{|\phi_0|^2}{(1-|\lambda_0|)(1-|\lambda_0| c_w)} < 1$.

[150] As a general setting, a nonparametric specification for $\phi_{i,0}$ can be also considered. We leave this issue for future study.

[151] Then, G is countably infinite.

distance: $d(i,j) = \|l(i) - l(j)\|_E$ for $i \neq j$. We set the minimum distance between two different spatial units to be a positive constant for asymptotic inferences (i.e., increasing domain asymptotics).

Based on this setting, we evaluate the distances between the leader and the followers, $\{d(0,i)\}_{i=1}^n$. Note that $\frac{\partial^2 u_i(y_{it}, Y_{-i,t},; Y_{n,t-1}, b_{it}, \eta_{it})}{\partial b_{it} \partial y_{it}} = \phi_{i,0}$, so the parameter $\phi_{i,0}$ describes the direct effect of the leader's action b_{it} on the i's marginal payoff. In the rational expectation equilibrium, the total effect of b_{it} on the i's marginal payoff will be $\phi_{i,0} + \lambda_0 \sum_{j=1}^n w_{ij} B_{n,ji}$ since b_{it} can also affect other followers' current actions in the equilibrium. We can consider the following two specifications[152]: (i) (homogeneous effects) $\phi_{i,0} = \phi_0$ for all $i = 1, \cdots, n$, and (ii) (heterogeneous effects by geographic locations) $\phi_{i,0}(d_i) = \phi_0(d_{\max} - d_i)$ for each $i = 1, \cdots, n$. Under the second specification, $\Phi_{n,0} = \phi_0 \Lambda_d$ where $\Lambda_d = diag(d_{\max} - d_1, \cdots, d_{\max} - d_n)$ (which is known or prespecified by an econometrician).

4.3.3 Derivation of log-likelihood function

For each t, we can consider $\eta_{nt} = X_{nt}\beta_0 + \eta_n^{iv} + \alpha_{t,0} l_n + \mathcal{E}_{nt}$ where X_{nt} is an $n \times K$ matrix of explanatory variables, $\beta_0 = (\beta_{1,0}, \cdots, \beta_{K,0})'$ is a $K-$dimensional vector of parameters, $\eta_n^{iv} = (\eta_1^{iv}, \cdots, \eta_n^{iv})'$ denotes an $n \times 1$ vector of individuals' invariant characteristics, $\alpha_{t,0}$ denotes the t^{th}-period time specific effect, and $\mathcal{E}_{nt} = (\epsilon_{1t}, \cdots, \epsilon_{nt})'$ is an $n \times 1$ vector of disturbances. Assume that (i) $\alpha_{t,0}$ and $\mathcal{E}_{nt}$ are orthogonal to the $(t-1)^{th}$-period information set[153], and (ii) nuisance parameters involving the process X_{nt} (π_0) are prespecified (or already revealed) before estimation. We can define

[152]In our simulation study, it is difficult to identify multiple parameters in $\Phi_{n,0}$ from the log-likelihood function. Hence, we suggest specifying $\Phi_{n,0}$ by one parameter.

[153]It means that π_0's corresponding to $\alpha_{t,0} l_n$ and $\mathcal{E}_{nt}$ are zeros.

individual fixed effects $\mathbf{c}_{n0} = (c_{1,0}, \cdots, c_{n,0})'$.[154] Next, consider the specification of $\Phi_{n,0} F_n \tau_{nt}$. As the simplest case, $\Phi_{n,0} F_n \tau_{nt}$ will be absorbed in $\mathbf{c}_{n0}$ if $\tau_{nt} = \tau_n$ for all t.

Then, the estimation equation generated by equation (4.14) is

$$\begin{aligned} R_n^F Y_{nt} &= \left(\gamma_0 I_n + \rho_0 W_n + \Phi_{n,0} D_n\right) Y_{n,t-1} \\ & \quad + \mathbf{X}_{nt} \beta_0 + R_n^e \left(\mathbf{c}_{n0} + \alpha_{t,0} l_n + \mathcal{E}_{nt}\right) \end{aligned} \tag{4.18}$$

for $t = 1, \cdots, T$, where

$$R_n^e = I_n + \Phi_{n,0} E_n,$$

$$R_n^F = S_n - \delta Q_n^{F,*} - \delta L_n^{F,b,*} D_n,$$

and

$$\mathbf{X}_{nt} = \left(I_n + \Phi_{n,0} E_n + \delta \pi_0 \left(L_n^{F,b,*} E_n + L_n^{F,\eta,*}\right)\right) X_{nt}.$$

For all i and t, assume $\epsilon_{it} \sim i.i.d. \left(0, \sigma_0^2\right)$ where $\sigma_0^2 > 0$. W_n and Λ_d are functions of geographic arrangements, and they are assumed to be prespecified. In implementing equation (4.18) via the (quasi) maximum likelihood method, we do not need $\{\mathbf{b}_{nt}\}$ since (i) we know how to generate $\{\mathbf{b}_{nt}\}$ for each parameter value given state variables and (ii) the components of algebraic matrix Riccati equations do not depend on levels of state variables (due to the LQ payoff assumption).

For parameter values, let $\theta = (\lambda, \gamma, \rho, \phi, \beta', \sigma^2)'$, $\mathbf{c}_n = (c_1, \cdots, c_n)'$, $\alpha_T = (\alpha_1, \cdots, \alpha_T)'$, and θ_0, $\mathbf{c}_{n0}$, $\alpha_{T,0}$ denote the true values. Observe $\dim(\theta) = 5 + K$. Note that the

[154]The $\mathbf{c}_{n0}$ is linearly transformed η_n^{iv}. In detail,

$$\mathbf{c}_{n0} = \left(R_n^e\right)^{-1} \left(I_n + \Phi_{n,0} E_n + \delta \pi_0 \left(L_n^{F,b,*} E_n + L_n^{F,\eta,*}\right)\right) \eta_n^{iv}.$$

asymptotic matrices R_n^F, R_n^e, D_n, E_n, $L_n^{F,b,*}$, and $L_n^{F,\eta,*}$ (which are needed to numerically evaluate) are functions of $\lambda, \gamma, \rho, \phi$, and W_n. For this, let $\theta_1 = (\lambda, \gamma, \rho, \phi)'$ and $\theta_{1,0}$ denote the true parameters. For each θ_1, define $\Phi_n(\phi)$, $R_n^F(\theta_1)$, $R_n^e(\theta_1)$, $D_n(\theta_1)$, $E_n(\theta_1)$, $L_n^{F,b,*}(\theta_1)$, and $L_n^{F,\eta,*}(\theta_1)$ representing those asymptotic matrices evaluated at θ_1. That is, $\Phi_{n,0} = \Phi_n(\phi_0)$, $R_n^F = R_n^F(\theta_{1,0})$, $R_n^e = R_n^e(\theta_1)$, $D_n = D_n(\theta_1)$, $E_n = E_n(\theta_{1,0})$, $L_n^{F,b,*} = L_n^{F,b,*}(\theta_{1,0})$, and $L_n^{F,\eta,*} = L_n^{F,\eta,*}(\theta_{1,0})$. To deal with the incidental parameters $\mathbf{c}_{n0}$ and $\alpha_{T,0}$, we employ the direct estimation approach. To eliminate $\mathbf{c}_{n0}$ in the log-likelihood function, we define $\tilde{Y}_{nt} = Y_{nt} - \bar{Y}_{nT}$, $\tilde{Y}_{n,t-1}^{(-)} = Y_{n,t-1} - \bar{Y}_{nT,-1}$, $\tilde{X}_{nt} = X_{nt} - \bar{X}_{nT}$ and $\tilde{\mathbf{X}}_{nt} = \mathbf{X}_{nt} - \bar{\mathbf{X}}_{nT}$ where $\bar{Y}_{nT} = \frac{1}{T}\sum_{s=1}^T Y_{ns}$, $\bar{Y}_{nT,-1} = \frac{1}{T}\sum_{s=0}^{T-1} Y_{ns}$, $\bar{X}_{nT} = \frac{1}{T}\sum_{s=1}^T X_{ns}$ and $\bar{\mathbf{X}}_{nT} = \frac{1}{T}\sum_{s=1}^T \mathbf{X}_{ns}$. The orthogonal projector $J_n = I_n - \frac{1}{n} l_n l_n'$ is introduced to delete $\alpha_{t,0}$ in the log-likelihood function.

Now we derive the log-likelihood function based on equation (4.18). To derive the proper density function for Y_{nt}, suppose invertibility of R_n^e. Conditional on $Y_{n,t-1}$, X_{nt}, $\mathbf{c}_n$, and α_t, the stochastic component of Y_{nt} is $\left(R_n^F\right)^{-1} R_n^e \mathcal{E}_{nt}$. The concentrated log-likelihood function for θ with $\mathbf{c}_{n0}$ and $\alpha_{T,0}$ concentrated out is

$$\begin{aligned} \ln L_{nT,c}(\theta) &= -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln \sigma^2 + T \ln \left| R_n^F(\theta_1) \right| - T \ln \left| R_n^e(\theta_1) \right| \\ &\quad -\frac{1}{2\sigma^2}\sum_{t=1}^T \tilde{\mathcal{E}}_{nt}(\theta) J_n \tilde{\mathcal{E}}_{nt}(\theta) \end{aligned} \tag{4.19}$$

where

$$\begin{aligned} \tilde{\mathcal{E}}_{nt}(\theta) &= \left(R_n^e(\theta_1)\right)^{-1} R_n^F(\theta_1) \tilde{Y}_{nt} \\ &\quad - \left(R_n^e(\theta_1)\right)^{-1} \left(\gamma I_n + \rho W_n + \Phi_n(\phi_0) D_n(\theta_1)\right) \tilde{Y}_{n,t-1}^{(-)} \\ &\quad - \left(R_n^e(\theta_1)\right)^{-1} \tilde{\mathbf{X}}_{nt}(\theta_1)\beta \end{aligned} \tag{4.20}$$

where

$$\mathbf{X}_{nt}(\theta_1) = \begin{pmatrix} I_n + \Phi_{n,0}(\phi) E_n(\theta_1) \\ +\delta\pi_0 \left(L_n^{F,b,*}(\theta_1) E_n(\theta_1) + L_n^{F,\eta,*}(\theta_1) \right) \end{pmatrix} X_{nt}$$

and $\tilde{\mathbf{X}}_{nt}(\theta_1) = \mathbf{X}_{nt}(\theta_1) - \bar{\mathbf{X}}_{nT}(\theta_1)$. Then, the maximization for parameter searching will be done on the fixed dimensional parameter space Θ: i.e., $\hat{\theta}_{nT} = \arg\max_{\theta\in\Theta} \ln L_{nT,c}(\theta)$. This is the outer loop maximization procedure. For each θ_1, the inner loop (iteration) procedure is to get the numerical solutions to $R_n^F(\theta_1)$, $R_n^e(\theta_1)$, $D_n(\theta_1)$, $E_n(\theta_1)$, $L_n^{F,b,*}(\theta_1)$, and $L_n^{F,\eta,*}(\theta_1)$. For the detailed evaluation method for the inner loop, refer to Appendix C.

4.4 Estimation and statistical properties: the quasi-maximum likelihood (QML) estimation method

Now we study large sample properties of the quasi-maximum likelihood (QML) estimator. By introducing some regularity conditions, we can derive consistency and the asymptotic normality of the QMLE. The framework studying large sample properties is the increasing domain asymptotics. i.e., increasing the sample size is ensured by growing a spatial domain. To support the large sample properties, we will conduct the Monte Carlo simulations in next section.

4.4.1 Regularity assumptions

For asymptotic analysis, define $R_{n,\lambda}^F(\theta_1) = \frac{\partial R_n^F(\theta_1)}{\partial \lambda}$ where Θ_1 be a subparameter space for θ_1. Other (first, second, and third orders) derivatives of $R_n^F(\theta_1)$, $R_n^e(\theta_1)$, $D_n(\theta_1)$, $E_n(\theta_1)$, $L_n^{F,b,*}(\theta_1)$, and $L_n^{F,\eta,*}(\theta_1)$ with respect to an element of θ_1 are similarly defined by adding a relevant subscript. At $\theta_1 = \theta_{1,0}$, we denote $R_{n,\lambda}^F = R_{n,\lambda}^F(\theta_{1,0})$ and so on. For the (unconditional) expectation operator in the estimation part, we use the notation $\mathbf{E}(\cdot)$. Here are regularity conditions for consistency and asymptotic normality.

Assumption 4.4.1 *W_n has zero diagonal elements. W_n is nonstochastic and bounded in row and column sums in absolute value. All components in Λ_d are uniformly bounded*

Assumption 4.4.2 *For all i and t, we assume $\epsilon_{it}\tilde{}\, i.i.d.\left(0, \sigma_0^2\right)$, and $\mathbf{E}\left|\epsilon_{it}\right|^{4+\eta_\epsilon} < \infty$ for some $\eta > 0$.*

Assumption 4.4.3 *Compact parameter space Θ is assumed. $\theta_0 \in int\left(\Theta\right)$.*

Assumption 4.4.4 *We assume that $\{X_{nt}\}_{t=0}^{T}$, $\mathbf{c}_{n0}$, and $\{\alpha_{t,0}\}_{t=1}^{T}$ are conditional on nonstochastic values. For some $\eta > 0$, $\max_k \sup_{n,T} \sum_{i=1}^{n} \sum_{t=1}^{T} \left|x_{it,k}\right|^{2+\eta} < \infty$, $\sup_T \frac{1}{T} \sum_{t=1}^{T} \left|\alpha_{t,0}\right|^{2+\eta} < \infty$, and $\sup_n \frac{1}{n} \sum_{i=1}^{n} \left|c_{i,0}\right|^{2+\eta} < \infty$.*

Assumption 4.4.5 *(i) For $\theta_1 \in \Theta_1$, $R_n^F\left(\theta_1\right)$ and $R_n^e\left(\theta_1\right)$ are nonsingular. The matrices $R_n^F\left(\theta_1\right)$, $R_n^e\left(\theta_1\right)$, $D_n\left(\theta_1\right)$, $E_n\left(\theta_1\right)$, $L_n^{F,b,*}\left(\theta_1\right)$, and $L_n^{F,\eta,*}\left(\theta_1\right)$ are uniformly bounded in both row and column sum norms, uniformly in $\theta_1 \in \Theta_1$.*

(ii) For $\theta_1 \in int\left(\Theta_1\right)$, the existence of the first, second, and third derivatives of $R_n^F\left(\theta_1\right)$, $R_n^e\left(\theta_1\right)$, $D_n\left(\theta_1\right)$, $E_n\left(\theta_1\right)$, $L_n^{F,b,*}\left(\theta_1\right)$, and $L_n^{F,\eta,*}\left(\theta_1\right)$ with respect to θ_1 is assumed. And, they are uniformly bounded in both row and column sum norms, uniformly in $\theta_1 \in \Theta_1$.

(iii) Recall that $A_n + B_n D_n = \left(R_n^F\right)^{-1}\left(\gamma_0 I_n + \rho_0 W_n + \Phi_{n,0} D_n\right)$. $\sum_{h=1}^{\infty} abs\left(\left(A_n + B_n D_n\right)^h\right)$ is uniformly bounded in both row and column sum norms, where $\left[abs\left(A_n + B_n D_n\right)\right]_{ij} = \left|\left[A_n + B_n D_n\right]_{ij}\right|$, for example.

Assumption 4.4.6 *n is an increasing function of T with $T \to \infty$.*

In spatial econometrics, Assumption 4.4.1 is conventional. Uniform boundedness of W_n gives the spatial stability condition. For simplicity, we assume *i.i.d.* disturbances ϵ_{it}'s across i and t by Assumption 4.4.2. In Assumption 4.4.2, assuming the higher than the fourth moment for ϵ_{it} is for a central limit theorem for a LQ form (Kelejian and Prucha (2001)). Assumption 4.4.3 gives compactness of Θ, which is for asymptotic analysis of a nonlinear extremum estimator (Chapter 4 in Amemiya (1985)). Note that X_{nt} and $\alpha_{t,0}$ are stochastic in the theoretical model's environment.[155] Since the expected their future values can be represented as currently realized ones with π_0, by Assumption 4.4.4, we assume that X_{nt}, $\mathbf{c}_{n0}$, and $\alpha_{T,0}$ are conditional upon as constants with the empirical moment restrictions. This assumption is for simplicity of asymptotic analysis, so Assumption 4.4.4 can be relaxed.

To have well-definedness of the model for each $\theta \in \Theta$, we introduce Assumption 4.4.5. By Assumption 4.4.5 (i), invertibility of $R_n^F(\theta_1)$ and $R_n^e(\theta_1)$ implies existence and uniqueness of the system 4.18 (and its correlation structure) for each $\theta_1 \in \Theta_1$. The second part of Assumption 4.4.5 (i) characterizes weak dependencies generated by the model's structure and W_n for each $\theta_1 \in \Theta_1$. Assumption 4.4.5 (ii) is introduced for technical issues to guarantee that $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\hat{\theta}_{nT}\right)}{\partial\theta\partial\theta'} - \mathbf{E}\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}(\theta_0)}{\partial\theta\partial\theta'}\right) = o_p(1)$ uniformly in the set Θ. For simple asymptotic analysis, we introduce Assumption 4.4.5 (iii), which is a sufficient condition of time and space stability. Since we take the direct estimation approach for $\mathbf{c}_{n0}$, and $\alpha_{T,0}$, Assumption 4.4.6 is introduced.

Since we focus on the large n and T framework, it is convenient to have a king vector representation. Let $L = nT$, and the subscript "L" in a vector/matrix denotes a stacked vector/matrix. For example, $Y_L = (Y_{n1}', \cdots, Y_{nT}')'$ and $X_L = (X_{n1}', \cdots, X_{nT}')'$

[155] Since (i) X_{nt} and $\alpha_{t,0}$ are parts of η_{nt}, (ii) η_{nt} follows a linear Markov process, and (iii) economic agents know them, they rationally expect future values of them based on $\mathbf{E}_t(\cdot)$.

become respectively an $L \times 1$ vector and an $L \times K$ matrix. Using this representation, equation (4.18) can be rewritten by

$$\begin{aligned} \left(I_T \otimes (R_n^e)^{-1} R_n^F\right) Y_L &= \left(I_T \otimes (\gamma_0 I_n + \rho_0 W_n + \Phi_{n,0} D_n)\right) Y_{L,-1} \\ &\quad + \left(I_T \otimes (R_n^e)^{-1}\right) \mathbf{X}_L \beta_0 + \mathbf{c}_{L,0} + \alpha_{L,0} + \mathcal{E}_L \end{aligned} \tag{4.21}$$

where $Y_{L,-1} = \left(Y'_{n0}, \cdots, Y'_{n,T-1}\right)'$, $\mathbf{c}_{L,0} = l_T \otimes \mathbf{c}_{n0}$, and $\alpha_{L,0} = (\alpha_{1,0}, \cdots, \alpha_{T,0})' \otimes l_n$. And, the king vector representation of the concentrated log-likelihood function is

$$\begin{aligned} \ln L_{L,c}(\theta) &= -\frac{L}{2} \ln 2\pi - \frac{L}{2} \ln \sigma^2 + T \ln \left| R_n^F(\theta_1) \right| - T \ln \left| R_n^e(\theta_1) \right| \\ &\quad - \frac{1}{2\sigma^2} \mathcal{E}'_L(\theta) (J_T \otimes J_n) \mathcal{E}_L(\theta) \end{aligned} \tag{4.22}$$

where $\mathcal{E}_L(\theta) = \left(\mathcal{E}'_{n1}(\theta), \cdots, \mathcal{E}'_{nT}(\theta)\right)'$, and

$$\begin{aligned} \mathcal{E}_{nt}(\theta) &= \left(R_n^e(\theta_1)\right)^{-1} R_n^F(\theta_1) Y_{nt} \\ &\quad - \left(R_n^e(\theta_1)\right)^{-1} \left(\gamma I_n + \rho W_n + \Phi_n(\phi) D_n(\theta_1)\right) Y_{n,t-1} \\ &\quad - \left(R_n^e(\theta_1)\right)^{-1} \mathbf{X}_{nt}(\theta_1) \beta. \end{aligned}$$

Then,

$$\begin{aligned} \mathcal{E}_L(\theta) &= \left(I_T \otimes \left(R_n^e(\theta_1)\right)^{-1} R_n^F(\theta_1)\right) Y_L \\ &\quad - \left(I_T \otimes \left(R_n^e(\theta_1)\right)^{-1} \left(\gamma I_n + \rho W_n + \Phi_n(\phi) D_n(\theta_1)\right)\right) Y_{L,-1} \\ &\quad - \left(I_T \otimes \left(R_n^e(\theta_1)\right)^{-1}\right) \mathbf{X}_L(\theta_1) \beta. \end{aligned}$$

4.4.2 Consistency

The main purpose of this subsection is to show $\text{plim}_{L \to \infty} \hat{\theta}_L = \theta_0$. Let $Q_L(\theta) = \mathbf{E}\left(\frac{1}{L} \ln L_{L,c}(\theta)\right)$. The first step is verifying $\sup_{\theta \in \Theta} \left| \frac{1}{L} \ln L_{L,c}(\theta) - Q_L(\theta) \right| \to_p 0$ as

$L \to \infty$. Second, uniformly equicontinuity of $Q_L(\theta)$ will be verified. To obtain consistency, lastly, the identification uniqueness conditions are required. By the information inequality in likelihood theory, the assumption below characterizes the identification uniqueness.

Assumption 4.4.7 (Identification) *Assume*

(i) $\lim_{L\to\infty} \left\{ \begin{array}{c} \frac{1}{n} \ln \left| \sigma_0^2 \left(R_n^F\right)^{-1} R_n^e R_n^{e\prime} \left(R_n^F\right)^{-1\prime} \right| \\ -\frac{1}{n} \ln \left| \sigma_L^2(\theta_1) \left(R_n^F(\theta_1)\right)^{-1} R_n^e(\theta_1) R_n^e(\theta_1)' \left(R_n^F\right)^{-1\prime} \left(R_n^F(\theta_1)\right)^{-1\prime} \right| \end{array} \right\} \neq 0$ *for* $\theta_1 \neq \theta_{1,0}$.

(ii) $\lim_{L\to\infty} \frac{1}{L} \mathbf{X}_L' \left(I_T \otimes (R_n^e)^{-1\prime}\right) (J_T \otimes J_n) \left(I_T \otimes (R_n^e)^{-1}\right) \mathbf{X}_L$ *exists and is positive definite.*

Note that Assumption 4.4.7 states the identification conditions focusing on large sample statistical theories. Derivation of the conditions in Assumption 4.4.7 is relegated to Appendix C. Assumption 4.4.7 (i) is a sufficient condition for unique identification of $\theta_{1,0}$. This condition is derived from the expected concentrated log-likelihood function $Q_{L,c}(\theta_1) \equiv Q_L(\theta_1, \beta_L(\theta_1), \sigma_L^2(\theta_1))$ where $\beta_L(\theta_1) = \arg\max_\beta Q_L(\theta)$ and $\sigma_L^2(\theta_1) = \arg\max_{\sigma^2} Q_L(\theta_1, \beta_L(\theta_1), \sigma^2)$. That is, Assumption 4.4.7 (i) is a sufficient condition of the unique identification condition under large samples,
i.e., $\limsup_{L\to\infty} \max_{\theta_1 \in \mathcal{N}^c(\theta_{1,0},\varepsilon)} [Q_{L,c}(\theta_1) - Q_{L,c}(\theta_{1,0})] < 0$ where $\mathcal{N}^c(\theta_{1,0}, \varepsilon)$ denotes the complement of an open neighborhood of Θ_1 of radius $\varepsilon > 0$. Assumption 4.4.7 (ii) is for identifying β_0. Given identified $\theta_{1,0}$, β_0 is identified if there are sufficient variations in the generated regressors $\mathbf{X}_L$. Observe that the two conditions in Assumption 4.4.7 do not depend on normality on $\mathcal{E}_{nt}$. Hence, we can apply the identification conditions in Assumption 4.4.7 to the quasi log-likelihood function.

Here is the theorem stating consistency of $\hat{\theta}_L$. Proof of Theorem 4.4.1 can be found in Appendix C.

Theorem 4.4.1 *Under Assumptions 4.4.1-4.4.7,* $plim_{T\to\infty}\hat{\theta}_L = \theta_0$.

Note that the limiting argument in Theorem 4.4.1 is T due to Assumption 4.4.6, i.e., $T \to \infty$ implies $L \to \infty$.

4.4.3 Asymptotic normality

In the previous subsection, we show that the QML estimator $\hat{\theta}_L$ can be accurate under large L. The next step is to obtain the asymptotic variance of $\hat{\theta}_L$ for statistical inferences. Since deriving the asymptotic distribution relies on the Taylor approximation argument, the main part at this point is studying $\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \theta}$. In Appendix C, we report the formulas of $\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \theta}$. In general, each component of $\frac{1}{\sqrt{L}}\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \theta}$ takes the LQ form of $\mathcal{E}_L$:

$$\begin{aligned} \tilde{\mathbf{s}}_L &= \frac{1}{\sqrt{L}}\left(\left(I_T \otimes B_{y,n}\right) Y_{L,-1} + \left(I_T \otimes C_{x,nt}\right)\right)' \left(J_T \otimes J_n\right) \mathcal{E}_L \\ &\quad + \frac{1}{\sqrt{L}}\left(\mathcal{E}_L' \left(I_T \otimes B_{q,n}'\right)\left(J_T \otimes J_n\right) \mathcal{E}_L - \sigma_0^2 tr\left(I_T \otimes B_{q,n}\right)\right) \end{aligned} \tag{4.23}$$

where $C_{x,nt}$ represents an $n \times 1$ vector of time-evolving nonstochastic components (a linear transformation of X_{nt} or $\alpha_{t,0} l_n$), and $B_{y,n}$ and $B_{q,n}$ are $n \times n$ (uniformly bounded) linear transformation matrices. The LQ form (4.23) says that our asymptotic analysis will be based on the martingale difference arrays for LQ forms. However, $\mathbf{E}\left(\tilde{\mathbf{s}}_L\right)$ and its asymptotic value are not zero. Since it implies the existence of asymptotic biases in $\frac{1}{\sqrt{L}}\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \theta}$ (so is $\hat{\theta}_L$), we need to adjust them for the asymptotically centered confidence intervals. Hence, our bias correction method will rely on calculating the scores' expected values at θ_0.

To specify the asymptotic bias of $\hat{\theta}_L$, we need to consider $\mathbf{E}(\tilde{\mathbf{s}}_L)$: let $\tilde{\mathbf{s}}_L = \tilde{\mathbf{s}}_L^{(u)} - \Delta_{1,L}^s - \Delta_{2,L}^s$ where $\mathbf{E}\left(\tilde{\mathbf{s}}_L^{(u)}\right) = 0$ (mean zero part),

$$\begin{aligned} \Delta_{1,L}^s &= \sqrt{\frac{T}{n}}\left[\left(B_{y,n}\bar{U}_{nT,-1}\right)' J_n \not{\mathcal{E}}_{nt} + \not{\mathcal{E}}'_{nt} B'_{n,q} J_n \not{\mathcal{E}}_{nt}\right], \\ \Delta_{2,L}^s &= \sqrt{\frac{T}{n}}\sigma_0^2\left[tr\left(B_{q,n}\right) - tr\left(J_n B_{q,n}\right)\right], \end{aligned}$$

and $\bar{U}_{L,-1} = \frac{1}{T}\sum_{t=0}^{T-1}\sum_{h=0}^{\infty}\left(A_n + B_n D_n\right)^h \left(B_n E_n + C_n\right)\mathcal{E}_{n,t-h}$. Note that (i) $\sum_{h=0}^{\infty}\left(A_n + B_n D_n\right)^h\left(B_n E_n + C_n\right)\mathcal{E}_{n,t-h}$ is the stochastic component of Y_{nt} for each t; (ii) $\Delta_{1,L}^s$ and $\Delta_{2,L}^s$ denote the sources of asymptotic biases. Using Lemmas 2.1 and 2.2 in the supplementary file of Jeong and Lee (2018), we have

$$\Delta_{1,L}^s = \sqrt{\frac{n}{T}} a_{n,1}^s\left(\theta_0\right) + O\left(\sqrt{\frac{n}{T^3}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

and

$$\Delta_{2,L}^s = \sqrt{\frac{T}{n}} a_{n,2}^s\left(\theta_0\right)$$

where $a_{n,1}^s\left(\theta_0\right) = \frac{1}{n}tr\left(J_n B_{y,n}\left(\sum_{h=0}^{\infty}\left(A_n + B_n D_n\right)^h\right)\left(B_n E_n + C_n\right)\right) + \frac{1}{n}tr\left(J_n B_{q,n}\right) = O(1)$ and $a_{n,2}^s\left(\theta_0\right) = \frac{1}{n}l_n' B_{q,n} l_n = O(1)$. Note that $a_{n,1}^s\left(\theta_0\right)$ and $a_{n,2}^s\left(\theta_0\right)$ are uniformly bounded constants by Assumption 4.4.5 (i). Using those formulations, we obtain $\frac{1}{\sqrt{L}}\frac{\partial \ln L_{L,c}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{L}}\frac{\partial \ln L_{L,c}^{(u)}(\theta_0)}{\partial\theta} - \Delta_{1,L} - \Delta_{2,L}$ where $\Delta_{1,L} = \left(\Delta_{1,L}^{\lambda}, \Delta_{1,L}^{\gamma}, \Delta_{1,L}^{\rho}, \Delta_{1,L}^{\phi}, \Delta_{1,L}^{\beta\prime}, \Delta_{1,L}^{\sigma^2}\right)'$ and $\Delta_{2,L} = \left(\Delta_{2,L}^{\lambda}, \Delta_{2,L}^{\gamma}, \Delta_{2,L}^{\rho}, \Delta_{2,L}^{\phi}, \Delta_{2,L}^{\beta\prime}, \Delta_{2,L}^{\sigma^2}\right)'$. The superscripts in $\Delta_{1,L}$ and $\Delta_{2,L}$ denote the components corresponding to a specific parameter. We can also define the vectors $a_{1,n}\left(\theta_0\right)$ and $a_{2,n}\left(\theta_0\right)$ based on $\Delta_{1,L}$ and $\Delta_{2,L}$.

Next, consider deriving the asymptotic distribution of $\hat{\theta}_L$. For each $\theta \in \Theta$, define $\Sigma_{\theta,L} = -\mathbf{E}\left(\frac{1}{L}\frac{\partial^2 \ln L_{L,c}(\theta)}{\partial\theta\partial\theta'}\right)$, $\Sigma_{\theta_0,L} = -\mathbf{E}\left(\frac{1}{L}\frac{\partial^2 \ln L_{L,c}(\theta_0)}{\partial\theta\partial\theta'}\right)$, $\Omega_{\theta,L} = \mathbf{E}\left(\frac{1}{L}\frac{\partial \ln L_{L,c}(\theta)}{\partial\theta}\frac{\partial \ln L_{L,c}(\theta)}{\partial\theta'}\right)$ and $\Omega_{\theta_0,L} = \mathbf{E}\left(\frac{1}{L}\frac{\partial \ln L_{L,c}(\theta_0)}{\partial\theta}\frac{\partial \ln L_{L,c}(\theta_0)}{\partial\theta'}\right)$. Here is the assumption for $\Sigma_{\theta_0,L}$ and $\Omega_{\theta_0,L}$ to have the well-defined asymptotic variance of $\hat{\theta}_L$. For this assumption, let $\phi_{\min}(M_n)$ denote the smallest eigenvalue of a matrix M_n.

Assumption 4.4.8 $\liminf_{L\to\infty} \phi_{\min}(\Sigma_{\theta_0,L}) > 0$ *and* $\liminf_{L\to\infty} \phi_{\min}(\Omega_{\theta_0,L}) > 0$.

Due to the parts (i) and (ii) of Assumption 4.4.5, $\Sigma_{\theta,L}$ and $\Omega_{\theta,L}$ are continuously differentiable functions in $\theta \in int(\Theta)$. Around θ_0, therefore, $\Sigma_{\theta,L}$ and $\Omega_{\theta,L}$ are non-singular for sufficiently large L. Let $\Sigma_{\theta_0} = \lim_{L\to\infty} \Sigma_{\theta_0,L}$ and $\Omega_{\theta_0} = \lim_{L\to\infty} \Omega_{\theta_0,L}$. In consequence, the limiting distribution of $\hat{\theta}_L$ is obtained.

Theorem 4.4.2 *Under Assumptions 4.4.1-4.4.8,*

$$\begin{aligned} &\sqrt{L}\left(\hat{\theta}_L - \theta_0\right) + \sqrt{\frac{n}{T}}\Sigma_{\theta_0,L}^{-1} a_{n,1}(\theta_0) + \sqrt{\frac{T}{n}}\Sigma_{\theta_0,L}^{-1} a_{n,2}(\theta_0) + o_p(1) \\ \to\ & {}_d N\left(0, \Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1}\right) \end{aligned}$$

as $L \to \infty$.

Here are interpretations of Theorem 4.4.2. First, $\hat{\theta}_L - \theta_0 = O_p\left(\max\left\{\frac{1}{\sqrt{L}}, \frac{1}{n}, \frac{1}{T}\right\}\right)$, which implies the convergence rate of $\hat{\theta}_L$. Even though we can achieve $\hat{\theta}_L - \theta_0 \to_p 0$ as $L \to \infty$, a ratio of n and T plays an important role in characterizing the asymptotic distribution of $\hat{\theta}_L$. Namely, a moderate ratio between n and T is required to have the nondegenerate asymptotic distribution of $\hat{\theta}_L$. Suppose $\frac{n}{T} \to c \in (0,\infty)$. Since $a_{n,1}(\theta_0)$ and $a_{n,2}(\theta_0)$ are of $O(1)$, $\sqrt{L}\left(\hat{\theta}_L - \theta_0\right) + \sqrt{c}\Sigma_{\theta_0,L}^{-1} a_{n,1}(\theta_0) + \sqrt{\frac{1}{c}} a_{n,2}(\theta_0) \to_d N\left(0, \Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1}\right)$. i.e., we can have the nondegenerate asymptotic distribution of $\hat{\theta}_L$ if $\frac{n}{T} \to c$, but the asymptotic biases exist. If $\frac{n}{T} \to 0$ or $\frac{n}{T} \to \infty$, the asymptotic distribution of $\hat{\theta}_L$ will be degenerated. Focusing on the case of $\frac{n}{T} \to c$, the next step is to have a bias corrected estimator.

4.4.4 Bias correction

By Theorem 4.4.2, we define

$$\hat{\theta}_L^c = \hat{\theta}_L - \frac{1}{T}\left[-\Sigma_{\hat{\theta}_L,L}^{-1} a_{n,1}\left(\hat{\theta}_L\right)\right] - \frac{1}{n}\left[-\Sigma_{\hat{\theta}_L,L}^{-1} a_{n,2}\left(\hat{\theta}_L\right)\right], \tag{4.24}$$

which is the bias corrected MLE. Note that the ideal bias correction terms are respectively $-\Sigma_{\theta_0,L}^{-1} a_{n,1}(\theta_0)$ and $-\Sigma_{\theta_0,L}^{-1} a_{n,2}(\theta_0)$. To have a successful bias correction, we need to achieve some conditions for the asymptotic equivalence. To have a successful bias correction, we need to achieve some conditions for the asymptotic equivalence: i.e., $\sqrt{\frac{n}{T}}\left(\Sigma_{\hat{\theta}_L,L}^{-1} a_{n,1}\left(\hat{\theta}_L\right) - \Sigma_{\theta_0,L}^{-1} a_{n,1}(\theta_0)\right) \to_p 0$ and $\sqrt{\frac{T}{n}}\left(\Sigma_{\hat{\theta}_L,L}^{-1} a_{n,2}\left(\hat{\theta}_L\right) - \Sigma_{\theta_0,L}^{-1} a_{n,2}(\theta_0)\right) \to_p 0$ as $L \to \infty$. The assumption below describes the conditions to have the asymptotic equivalence.

Assumption 4.4.9 *(i)* $\frac{n}{T^3} \to 0$ *and* $\frac{T}{n^3} \to 0$.

(ii) In a neighborhood of $\theta_{1,0}$, $\sum_{h=0}^{\infty}\left(A_n(\theta_1) + B_n(\theta_1) D_n(\theta_1)\right)^h$ *and* $\sum_{h=1}^{\infty} h\left(A_n(\theta_1) + B_n(\theta_1) D_n(\theta_1)\right)^{h-1}$ *are uniformly bounded in row and column sums.*

Under the additional assumption (Assumption 4.4.9), we have

$$\sqrt{L}\left(\hat{\theta}_L^c - \theta_0\right) \to_d N\left(0, \Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1}\right) \tag{4.25}$$

as $L \to \infty$. For the details on Assumption 4.4.9, refer to Corollary 4.3 of Jeong and Lee (2018).

4.5 Simulations

This section reports some simulation results to study small sample properties of the QMLE. For $t = 1, \cdots, T$, the DGP for our simulation is specified by

$$R_n^F Y_{nt} = \left(\gamma_0 I_n + \rho_0 W_n + \Phi_{n,0} D_n\right) Y_{n,t-1} + \mathbf{X}_{nt}\beta_0 + R_n^e\left(\mathbf{c}_{n0} + \alpha_{t,0} l_n + \mathcal{E}_{nt}\right) \tag{4.26}$$

where

$$R_n^e = I_n + \Phi_{n,0} E_n,$$

$R_n^F = S_n - \delta Q_n^{F,*} - \delta L_n^{F,b,*} D_n$, and $\mathbf{X}_{nt} = (I_n + \Phi_{n,0} E_n) X_{nt}$. i.e., we set $K = 1$ and $\pi_0 = 0$. For this, X_{nt} is drawn from $i.i.d.N(\mathbf{0}_{n \times 1}, I_n)$.

Our simulation design aligns with the empirical application in Section 4.6 in terms of the sample size and the spatial network W_n. As sample sizes, we consider $(n,T) = (48, 25)$. That is, economic agents in this simulation represent the 48 contiguous U.S. states. For each pair (i,j), let $w_{ij} = \frac{\tilde{w}_{ij}}{\sum_{k=1}^{n} \tilde{w}_{ik}}$ where

$$\tilde{w}_{ij} = \mathbf{1}\{(i,j)\, nbd\} \tag{4.27}$$

if $j \neq i$. We choose $\delta = 0.9$. From the standard normal distributions, we draw $\mathbf{c}_{n0}$, $\alpha_{T,0}$, and $\mathcal{E}_{nt}$. To have a stable functional form of the DGP, we firstly generate the data with $30+T$ periods and take the last T periods as our sample. We consider four sets of the true parameter values: $\theta_0 = (0.1, 0.2, 0.1, 0.4, 1, 1)'$, $(0.1, 0.2, -0.1, 0.4, 1, 1)'$, $(-0.1, 0.2, 0.1, 0.4, 1, 1)'$, and $(-0.1, 0.2, -0.1, 0.4, 1, 1)'$. Four criteria are reported for performance evaluation: (i) bias, (ii) standard deviation (SD), (iii) theoretical standard deviation (T-SD), and (iv) root mean square error (RMSE). We conduct 400 repetitions for each case.

Table 4.1 shows the detailed simulation results. Overall, the estimated theoretical standard deviations are similar to empirically evaluated ones. For all cases of θ_0's, the QMLE and its bias corrected version show similar performance in terms of RMSEs. We detect downward biases in the QMLEs for λ_0, γ_0, ϕ_0 and σ_0^2. In estimating ρ_0, the magnitude of biases in $\hat{\theta}_L$ is small. Our bias correction method can reduce the magnitude of biases except for estimation of ρ_0 and ϕ_0. Since biases in the QMLEs

Table 4.1: Performance of the QML estimator

Case 1: $\theta_0 = (0.1, 0.2, 0.1, 0.4, 1, 1)'$

	$\hat{\theta}_L$					
	λ	γ	ρ	ϕ	β	σ
Bias	-0.0332	-0.0249	0.0052	-0.0125	-0.0028	-0.0634
SD	0.0334	0.0202	0.0407	0.1115	0.0307	0.061
T-SD	0.0313	0.0201	0.0386	0.1046	0.0304	0.0578
RMSE	0.047	0.0321	0.041	0.1122	0.0308	0.088
	$\hat{\theta}_L^c$					
	λ	γ	ρ	ϕ	β	σ
Bias	-0.0123	-0.0016	0	-0.0411	-0.0014	0.0062
SD	0.0317	0.0206	0.0411	0.1071	0.0306	0.0619
T-SD	0.0313	0.0201	0.0386	0.1046	0.0304	0.0578
RMSE	0.034	0.0207	0.0411	0.1147	0.0306	0.0623

Case 2: $\theta_0 = (0.1, 0.2, -0.1, 0.4, 1, 1)'$

	$\hat{\theta}_L$					
	λ	γ	ρ	ϕ	β	σ
Bias	-0.0328	-0.0261	0.0081	-0.0136	-0.0027	-0.0631
SD	0.0334	0.0204	0.0406	0.1137	0.0307	0.0615
T-SD	0.0313	0.0201	0.0388	0.1143	0.0304	0.0584
RMSE	0.0468	0.0332	0.0414	0.1145	0.0308	0.0881
	$\hat{\theta}_L^c$					
	λ	γ	ρ	ϕ	β	σ
Bias	-0.0121	-0.0018	0.0009	-0.0426	-0.0014	0.0066
SD	0.0318	0.0208	0.0412	0.1099	0.0306	0.0625
T-SD	0.0313	0.0201	0.0388	0.1143	0.0304	0.0584
RMSE	0.034	0.0209	0.0412	0.1179	0.0307	0.0628

Case 3: $\theta_0 = (-0.1, 0.2, 0.1, 0.4, 1, 1)'$

	$\hat{\theta}_L$					
	λ	γ	ρ	ϕ	β	σ
Bias	-0.0303	-0.0255	-0.0002	-0.0115	-0.0038	-0.0685
SD	0.034	0.0204	0.0425	0.1209	0.0308	0.0648
T-SD	0.033	0.0202	0.0408	0.1223	0.0304	0.062
RMSE	0.0456	0.0327	0.0425	0.1214	0.031	0.0943
	$\hat{\theta}_L^c$					
	λ	γ	ρ	ϕ	β	σ
Bias	0.0029	-0.0028	-0.0035	-0.0238	-0.0012	-0.0029
SD	0.0318	0.0207	0.0429	0.1114	0.0308	0.0651
T-SD	0.033	0.0202	0.0408	0.1223	0.0304	0.062
RMSE	0.0319	0.0209	0.043	0.1139	0.0308	0.0651

Case 4: $\theta_0 = (-0.1, 0.2, -0.1, 0.4, 1, 1)'$

	$\hat{\theta}_L$					
	λ	γ	ρ	ϕ	β	σ
Bias	-0.03	-0.0266	0.004	-0.0112	-0.0037	-0.0684
SD	0.034	0.0206	0.0421	0.1197	0.0308	0.065
T-SD	0.0329	0.0202	0.0409	0.1222	0.0304	0.0621
RMSE	0.0453	0.0337	0.0423	0.1202	0.0311	0.0944
	$\hat{\theta}_L^c$					
	λ	γ	ρ	ϕ	β	σ
Bias	0.003	-0.0024	-0.0009	-0.0233	-0.0012	-0.003
SD	0.0318	0.0209	0.0426	0.1105	0.0308	0.0653
T-SD	0.0329	0.0202	0.0409	0.1222	0.0304	0.0621
RMSE	0.0319	0.0211	0.0426	0.1129	0.0308	0.0654

Table 4.2: Descriptive statistics: 48 contiguous states in U.S.

Variables	Mean	Standard dev.	Min	Max
Total expenditure	5.7992	1.2951	3.4631	11.6327
Total revenue	5.8646	1.2749	1.5816	14.8746
Grants from the Federal government	1.4947	0.4185	0.5757	4.1586
Population (millions)	6.0305	6.5155	0.4663	39.25
Population density	73.7276	100.0283	469.5999	1.8541
Personal income	35.7279	5.6464	25.7151	59.1866
Unemployment rate	0.0559	0.0187	0.023	0.1361

Note: Sample is 48 contiguous states from 1992 to 2016. Dollar amounts are in thousands and real per capita values adjusted by the GDP deflator with base year 2012.

for ρ_0 and ϕ_0 are not large, we can say that the proposed bias correction method is effective.

4.6 Application

In this section, we apply our econometric model. The federal government is a leader, and the 48 contiguous state governments (excluding Alaska, and Hawaii) are followers (i.e., $n = 48$). Time periods of our data set are 1992 to 2016 (i.e., $T = 25$, so $L = 1,200$). From the United States Census Bureau, we obtain the states' finance and demographic/economic variables. Levels of grants from the federal governments can be also found in this source. For the additional macroeconomic variables (e.g., GDP deflator, interest rates, states' unemployment rates), we utilize the website of the Federal reserve bank of S.t. Louis. All dollar amounts are in thousands and real per capita values adjusted by the GDP deflator (with the base year 2012). Table 4.2 shows the descriptive statistics for the collected variables. Note that the vector of dependent variables (Y_{nt}) represents the states' general expenditures at time t. As explanatory variables (X_{nt}), we employ (i) total revenue, (ii) time differenced population density[156], and (iii) state's unemployment rate. We observe large variations in the variables across states and time periods.

We will estimate payoff functions for general expenditures of state governments by estimating θ_0. After recovering θ_0, we can also recover the leader's payoff by summing followers' payoffs. At each t, the federal government allocates the national resources to the states by deciding levels of grants. Each state government selects an amount of expenditure after observing the level of grants. Among state governments, there might exist spatial interactions. For this, we use the same W_n in Section 4.5. We choose $\delta = 0.956$ by considering the average of long-run interest rates for the sampling periods. For $\Phi_{n,0}$, we consider two specifications: for $i = 1, \cdots, n$ (i) (Specification (1)) $\phi_{i,0} = \phi_0 d_i$, and (ii) (Specification (2)) $\phi_{i,0} = \phi_0$. For each i, we compute d_i by the Haversine formula.[157] The mean and standard deviation of $\{d_i\}$ are respectively 1.4583 and 1.0691.

Table 4.3 summarizes the estimation results. In Table 4.3, we report the bias corrected QML estimates with their corresponding standard deviations. For both specifications, we obtain the similar estimation results. When we consider the sample log likelihood as a goodness-of-fit measure, Specification (1) is better. We find that there is a positive spatial spillover effect in the states' expenditures. The estimates for γ_0 are large, so the dynamic adjustment cost is high. The estimated coefficients

[156] We find that the U.S. population densities are quite persistent. In our model framework, all exogenous (time-varying) characteristics (η^v_{nt} in the theoretical model) should be stationary. To avoid nonstationary variable issues, we use the time differenced population densities.

[157] For example, d_{Ohio} denotes the (thousand) kilometer based distance between Columbus and Washington D.C.

Table 4.3: Model estimation

	Specification (1):	Specification (2)
Total revenue per capita	0.1028*** [0.0098]	0.1026*** [0.0098]
Δ Population density	-0.0041 [0.0139]	-0.0050 [0.0138]
Unemployment rate	-1.8004*** [0.6768]	-1.7490*** [0.6720]
λ	0.0748* [0.0410]	0.0770* [0.0418]
γ	0.8520*** [0.0162]	0.8626*** [0.0608]
ρ	-0.0540 [0.0462]	-0.0512 [0.0460]
ϕ	0.0261 [0.0228]	0.2015 [1.8148]
σ^2	0.0283*** [0.0012]	0.0290*** [0.0040]
Sample log likelihood	445.0063	430.4086
No. of Obs.	1152	1152

Note: Theoretical standard deviations are in parenthesis. Estimates that are significant at the 10 percent, 5 percent, and 1 percent levels are respectively marked by *, **, and ***.

for ρ_0 are negative, but they are not significant even for the 10% significance level. The estimates of ϕ_0 in both specifications are positive. In Specification (1), the central government positively affects the state governments' expenditures and those effects diminish corresponding the geographic distances between the central and state governments. However, estimated coefficients for ϕ_0 are not statistically significant. It seems that there is no effect of the federal grants on the states' expenditures. For the exogenous characteristics, there exists a significant positive effect of the state's total revenue; a significant negative effect of the state's unemployment rate is detected.

4.7 Conclusion and future works

This paper introduces a spatial dynamic panel data (SDPD) model explaining the relationships between two types of forward-looking agents: a leader and multiple followers. In practical applications, they represent the central and local governments.

Motivated by Chapter 19 of Ljungqvist and Sargent (2012), we establish a dynamic Stackelberg game played on a spatial network. Derived optimal actions lead to a spatial econometric model. Next, we introduce how to apply the quasi-maximum likelihood (QML) method for recovering structural parameters. Large and finite sample properties of the QML estimator are investigated.

4.7.1 Future works

Since our model specification also describes the central government behaviors, we can consider aggregate economic shocks directly affecting the central government. Hence, there might exist aggregate economic shocks heterogeneously affecting the state governments' decisions. Instead of having the additive specification for individual and time fixed effects, the interactive fixed effect specification (factor structure) can be considered. In order to specify $\left(\Phi_{n,0} + \delta\tilde{\pi}_0 L_n^{F,b,*}\right) F_n \tau_{nt}$, also, we consider a general specification (factor structure) for individual and time effects: $\Gamma_n f_t$ where $f_t = (f_{t,1}, \cdots, f_{t,r_0})'$ be an r_0-dimensional common factors and $\Gamma_n = \left(\Gamma'_{n,1}, \cdots, \Gamma'_{n,n}\right)'$ (with $\Gamma_{n,i} = (\Gamma_{n,i,1}, \cdots, \Gamma_{n,i,r_0})'$ for each i) denotes an $n \times r_0$ matrix of factor loadings. Note that the dimension of commen factors (r_0) can be multiple. Let $\mathbf{f}_T = (f'_1, \cdots, f'_T)'$ be a $T \times r_0$ matrix of common factors. We can allow flexible correlations between X_{nt} and $\Gamma_n f_t$, so they are considered as parameters.

Also, we have another reason for having the factor structure. Recall that the first order condition of follower i's lifetime problem:

$$\begin{aligned} 0 &= \eta_{it} + \phi_{i,0} b^*_{it} + e'_i \left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n,t-1} - e'_i S_n Y^*_{nt} \\ &\quad + \delta e'_i \left(Q^F_i + Q^{F\prime}_i\right) Y^*_{nt} + \delta e'_i L^{F,b}_i \mathbf{E}_t \mathbf{b}^*_{n,t+1} + \delta e'_i L^{F,\eta}_i \mathbf{E}_t \eta_{n,t+1}. \end{aligned} \tag{4.28}$$

The error specification involves $\eta_{it} + \delta e_i' L_i^{F,\eta} \mathbf{E}_t \eta_{n,t+1}$ and η_{it} might include the time specific shock $\alpha_{t,0}$. Note that the additive specification comes from assuming $\mathbf{E}_t(\alpha_{t+1,0}) = 0$. If $\{\alpha_{t,0}\}$ follows an AR(1) process and the followers rationally expect the future aggregate shock, the unobserved individual and time effects will follow the factor structure. A famous study for this issue in a regression model is Bai (2009) and the application to SDPD models can be found in Shi and Lee (2017). Shi and Lee (2017) treat the interactive individual and time effects as incidental parameters.

Appendix A: Appendix for Chapter 2

A.1: Derivation of the MPE equation

In this appendix, we derive the NE equation by solving equation (2.8). By the principle of optimality, a solution from the intertemporal choice problem (2.7) is equivalent to that of the functional equation (2.8) if the latter exists. For this, we need to verify the existence and uniqueness of $V_i(\cdot)$ satisfying both (2.7) and (2.8). The unknown $V_i(\cdot)$ will be implied by known $u_i(\cdot)$. All mathematical arguments in this part are based on Stokey et al. (1989) and Fuente (2000). Here we present some basic discussions and essential mathematical results.

Step 1 (Formation of $V_i^{(j)}(\cdot)$'s): We choose an arbitrary agent i for our analysis. Consider the period t. For any given $(Y_{n,t-1}, \eta_{nt})$ and $Y^*_{-i,t}(Y_{n,t-1}, \eta_{nt})$, define the operator $\mathcal{T}$ which maps the j^{th} approximation to the $(j+1)^{th}$ approximation of $V_i(\cdot)$ by

$$\begin{aligned} V_i^{(j+1)}(Y_{n,t-1}, \eta_{nt}) &= \mathcal{T}\left(V_i^{(j)}\right)(Y_{n,t-1}, \eta_{nt}) \\ &= \max_{y_{it}} \left\{ \begin{array}{c} u_i\left(y_{it}, Y^{*(j+1)}_{-i,t}(Y_{n,t-1}, \eta_{nt}), Y_{n,t-1}, \eta_{it}\right) \\ +\delta E_t\left(V_i^{(j)}\left(y_{it}, Y^{*(j+1)}_{-i,t}(Y_{n,t-1}, \eta_{nt}), \eta_{n,t+1}\right)\right) \end{array} \right\} \end{aligned}$$

for $j = 0, 1, 2, \cdots$. From $V_i^{(j)}(\cdot)$'s, we can also generate $Y_{nt}^{*(j)}(Y_{n,t-1}, \eta_{nt})$'s ($j = 1, 2, \cdots$). Using $\mathcal{T}$, we generate $V_i^{(j)}(\cdot)$'s (from $V_i^{(0)} = 0$) and corresponding (approximated) MPE equations.

Step 2 (Continuity of $\mathcal{T}$): Note that the domain of $\mathcal{T}$ contains a set of $V_i(\cdot)$'s (i.e., $V_i^{(j)}(\cdot)$'s). Consider a set of continuous and bounded functions $\mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n\right)$ where all possible $Y_{n,t-1} \in (\chi_y)^n \subseteq \mathbf{R}^n$ and $\eta_{nt} \in (\chi_\eta)^n \subseteq \mathbf{R}^n$. Note that $\mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n\right)$ is a well-known Banach space. Under Assumption 2.2.1, $\left\{V_i^{(j)}(\cdot)\right\}_j \subset \mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n\right)$ for any continuous and bounded function $V_i^{(0)}(\cdot)$. Then, we can apply the theorem of maximum, which yields (i) existence of optimal decisions and (ii) continuity of $\mathcal{T}V_i^{(j)}(Y_{n,t-1}, \eta_{nt})$ at $(Y_{n,t-1}, \eta_{nt})$. Since $u_i(\cdot)$ is strictly concave with strictly decreasing marginals[158] with respect to large y_{it}, we can guarantee for unique NE decisions.[159]

Step 3 (Contraction mapping theorem): Since $\mathcal{T}$ is the maximum operator, its arguments $V_i^{(j)}(\cdot)$'s are continuous and bounded functions in $(Y_{n,t-1}, \eta_{nt})$ and $\delta \in (0,1)$, $\mathcal{T}$ satisfies the Blackwell's (1965) sufficient conditions to be a contraction mapping. By the contraction mapping theorem, there exists a unique fixed point $V_i(\cdot)$ in $\mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n\right)$ for each $i = 1, \cdots, n$ and subsequently a unique NE $Y_{nt}^*(Y_{n,t-1}, \eta_{nt})$.

Step 4 (Recovering $V_i(\cdot)$ for each i and $Y_{nt}^*(Y_{n,t-1}, \eta_{nt})$): From the initial iteration with $V_i^{(0)} = 0$, we have $V_i^{(1)}(Y_{n,t-1}, \eta_{nt}) = Y_{n,t-1}' Q_i^{(1)} Y_{n,t-1} + Y_{n,t-1}' L_i^{(1)} \eta_{nt} + \eta_{nt}' G_i^{(1)} \eta_{nt} + c_i^{(1)}$, where $A_n^{(1)} = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, $B_n^{(1)} = S_n^{-1}$, $Q_i^{(1)} = \frac{1}{2}\left(A_n^{(1)\prime} \mathcal{I}_i A_n^{(1)} - \gamma_0 \mathcal{I}_i\right)$, $L_i^{(1)} = A_n^{(1)\prime} \mathcal{I}_i B_n^{(1)}$, $G_i^{(1)} = \frac{1}{2} B_n^{(1)\prime} \mathcal{I}_i B_n^{(1)}$ and $c_i^{(1)} = 0$ with $\mathcal{I}_i$ being a diagonal matrix with only a unit for its i^{th} diagonal element and

[158] Note that $u_i(\cdot)$ will eventually decrease in y_{it}. This property is important because our maximization problem is not constrained.

[159] Refer to Theorems 3.8 and 4.9 in Stokey et al. (1989).

zero elsewhere. By mathematical induction, we generate the following matrix Riccati equations:

$$Q_i^{(j+1)} = A_n^{(j+1)\prime}\left[\mathcal{I}_i\left(\frac{1}{2}I_n - S_n\right) + \delta Q_i^{(j)}\right]A_n^{(j+1)} + A_n^{(j+1)\prime}\mathcal{I}_i\left(\gamma_0 I_n + \rho_0 W_n\right) - \frac{\gamma_0}{2}\mathcal{I}_i, \tag{A.1}$$

$$Q_n^{*(j+1)} = [\left(Q_1^{(j+1)} + Q_1^{(j+1)\prime}\right)e_1, \cdots, \left(Q_n^{(j+1)} + Q_n^{(j+1)\prime}\right)e_n]',$$

$$\begin{aligned} L_i^{(j+1)} &= A_n^{(j+1)\prime}\left\{\begin{array}{c} [\mathcal{I}_i\left(\frac{1}{2}I_n - S_n\right) + \delta Q_i^{(j)}] \\ +[\mathcal{I}_i\left(\frac{1}{2}I_n - S_n\right) + \delta Q_i^{(j)}]' \end{array}\right\}B_n^{(j+1)} \\ &\quad + A_n^{(j+1)\prime}\left(\mathcal{I}_i + \delta L_i^{(j)}\Pi_n\right) + \left(\gamma_0 I_n + \rho_0 W_n\right)'\mathcal{I}_i B_n^{(j+1)}, \end{aligned} \tag{A.2}$$

$$L_n^{*(j+1)} = [L_1^{(j+1)\prime}e_1, \cdots, L_n^{(j+1)\prime}e_n], \tag{A.3}$$

$$G_i^{(j+1)} = B_n^{(j+1)\prime}\left[\mathcal{I}_i\left(\frac{1}{2}I_n - S_n\right) + \delta Q_i^{(j)}\right]B_n^{(j+1)} + B_n^{(j+1)\prime}\left(\mathcal{I}_i + \delta L_i^{(j)}\Pi_n\right) + \delta\Pi_n' G_i^{(j)}\Pi_n, \tag{A.4}$$

and $c_i^{(j+1)} = \delta\left(c_i^{(j)} + tr\left(G_i^{(j)}\Omega_\xi\right)\right)$, where $A_n^{(j+1)} = \left[R_n^{(j+1)}\right]^{-1}\left(\gamma_0 I_n + \rho_0 W_n\right)$ and $B_n^{(j+1)} = \left[R_n^{(j+1)}\right]^{-1}\left(I_n + \delta L_n^{*(j)}\Pi_n\right)$ with $R_n^{(j+1)} = S_n - \delta Q_n^{*(j)}$.

By taking $j \to \infty$, we obtain the asymptotic version of algebraic matrix Riccati equations for Q_n, L_n, G_i's and c_i, i.e., for each i,

$$V_i\left(Y_{n,t-1}, \eta_{nt}\right) = Y_{n,t-1}' Q_i Y_{n,t-1} + Y_{n,t-1}' L_i \eta_{nt} + \eta_{nt}' G_i \eta_{nt} + c_i$$

where $Q_i = \lim_{j\to\infty} Q_i^{(j)}$, $L_i = \lim_{j\to\infty} L_i^{(j)}$, $G_i = \lim_{j\to\infty} G_i^{(j)}$ and $c_i = \lim_{j\to\infty} c_i^{(j)}$. Then, the activity outcomes NE equation will be

$$Y_{nt}^*\left(Y_{n,t-1}, \eta_{nt}\right) = \left(\lambda_0 W_n + \delta Q_n^*\right)Y_{nt}^*\left(Y_{n,t-1}, \eta_{nt}\right) + \left(\gamma_0 I_n + \rho_0 W_n\right)Y_{n,t-1} + \left(I_i + \delta L_n^*\Pi_n\right)\eta_{nt},$$

which implies that

$$Y_{nt}^{*}\left(Y_{n,t-1},\eta_{nt}\right)=A_{n}Y_{n,t-1}+B_{n}\eta_{nt},$$

where $A_n = R_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$ and $B_n = R_n^{-1}(I_n + \delta L_n^{*}\Pi_n)$ with $R_n = S_n - \delta Q_n^{*}$.

From the above expressions, we can also have an alternative representation of Q_n^{*} in the subsequent Proposition A.1.1, which has some similarity on the additional term due to future influence as in the two-period case. First of all, we can have an alternative representation of $B_n^{(j)}$, $j = 1, 2, \cdots$. Note that $B_n^{(1)} = S_n^{-1}$. Consider $B_n^{(2)} = \left[R_n^{(2)}\right]^{-1}\left(I_n + \delta L_n^{*(1)}\Pi_n\right)$. Using $e_i' L_n^{*(1)} = e_i' L_i^{(1)}$ with $L_i^{(1)} = A_n^{(1)\prime}\mathcal{I}_i S_n^{-1}$, we can define $D_{n,1}^{(2)} = Diag\left(A_n^{(1)}\right)B_n^{(1)}$ such that $B_n^{(2)} = \left[R_n^{(2)}\right]^{-1}\left(I_n + \delta D_{n,1}^{(2)}\Pi_n\right)$. This has

$$\begin{aligned} Y_{nt}^{*(2)}\left(Y_{n,t-1},\eta_{nt}\right) &= A_n^{(2)}Y_{n,t-1} + \left[R_n^{(2)}\right]^{-1}\left(\eta_{nt} + \delta L_n^{*(1)}E_t\left(\eta_{n,t+1}\right)\right) \\ &= A_n^{(2)}Y_{n,t-1} + \left[R_n^{(2)}\right]^{-1}\left(I_n + \delta D_{n,1}^{(2)}\Pi_n\right)\eta_{nt}. \end{aligned}$$

Consider iteratively $B_n^{(j+1)} = \left[R_n^{(j+1)}\right]^{-1}\left(I_n + \delta L_n^{*(j)}\Pi_n\right)$ for $j = 2, 3, \cdots$. We can show that

$$L_n^{*(j)} = D_{n,1}^{(j+1)} + \delta D_{n,2}^{(j+1)}\Pi_n + \cdots + \delta^{j-1}D_{n,j}^{(j+1)}\Pi_n^{j-1} \tag{A.5}$$

for some $D_{n,1}^{(j+1)}$, $D_{n,2}^{(j+1)}$, $\cdots$, $D_{n,j}^{(j+1)}$ by the method of undetermined coefficients.

Hence,

$$B_n^{(j+1)} = \left[R_n^{(j+1)}\right]^{-1}\left(I_n + \delta D_{n,1}^{(j+1)}\Pi_n + \delta^2 D_{n,2}^{(j+1)}\Pi_n^2 + \cdots + \delta^j D_{n,j}^{(j+1)}\Pi_n^j\right)$$

so that

$$\begin{aligned} & Y_{nt}^{*(j+1)}\left(Y_{n,t-1},\eta_{nt}\right) \\ = \; & A_n^{(j+1)}Y_{n,t-1} + \left[R_n^{(j+1)}\right]^{-1}\left(I_n + \delta L_n^{*(j)}\Pi_n\right)\eta_{nt} \end{aligned}$$

$$= A_n^{(j+1)} Y_{n,t-1} + \left[R_n^{(j+1)}\right]^{-1} \begin{pmatrix} \eta_{nt} + \delta D_{n,1}^{(j+1)} E_t(\eta_{n,t+1}) + \delta^2 D_{n,2}^{(j+1)} E_t(\eta_{n,t+2}) \\ + \cdots + \delta^j D_{n,j}^{(j+1)} E_t(\eta_{n,t+j}) \end{pmatrix}$$
$$= A_n^{(j+1)} Y_{n,t-1} + \left[R_n^{(j+1)}\right]^{-1} \left(I_n + \delta D_{n,1}^{(j+1)} \Pi_n + \delta^2 D_{n,2}^{(j+1)} \Pi_n^2 + \cdots + \delta^j D_{n,j}^{(j+1)} \Pi_n^j\right) \eta_{nt}.$$

The second equality holds due to the law of iterative expectations. For notational convenience, let

$$\begin{aligned} C_i^{(j+1)} &= A_n^{(j+1)\prime} \left\{ \left[\mathcal{I}_i \left(\frac{1}{2} I_n - S_n\right) + \delta Q_i^{(j)}\right] + \left[\mathcal{I}_i \left(\frac{1}{2} I_n - S_n\right) + \delta Q_i^{(j)}\right]' \right\} \\ &\quad + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i \\ &= A_n^{(j+1)\prime} \left\{ \mathcal{I}_i \left[-I_n + \lambda_0 (W_n + W_n')\right] + \delta \left(Q_i^{(j)} + Q_i^{(j)\prime}\right) \right\} \\ &\quad + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i \end{aligned}$$

for $j = 1, 2, \cdots$. And, $C_i^{(1)} = A_n^{(1)\prime} \mathcal{I}_i \left[-I_n + \lambda_0 (W_n + W_n')\right] + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i$, so observe

$$\begin{aligned} & e_i' \left(C_i^{(1)} \left[R_n^{(1)}\right]^{-1} + A_n^{(1)\prime} \mathcal{I}_i \right) \\ =\ & e_i' A_n^{(1)\prime} \mathcal{I}_i \left[-I_n + \lambda_0 (W_n + W_n')\right] S_n^{-1} + e_i' (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i S_n^{-1} + e_i' A_n^{(1)\prime} \mathcal{I}_i \\ =\ & e_i' A_n^{(1)\prime} \mathcal{I}_i \left(-I_n + \lambda_0 W_n' S_n^{-1}\right) + e_i' A_n^{(1)\prime} \mathcal{I}_i + e_i' (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i S_n^{-1} \\ =\ & e_i' A_n^{(1)\prime} \mathcal{I}_i \lambda_0 W_n' S_n^{-1} + e_i' (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i S_n^{-1} \\ =\ & e_i' \left(A_n^{(1)\prime} \mathcal{I}_i \lambda_0 W_n' + (\gamma_0 I_n + \rho_0 W_n)' S_n^{-1\prime} S_n' \mathcal{I}_i \right) S_n^{-1} \\ =\ & e_i' A_n^{(1)\prime} \mathcal{I}_i S_n^{-1} = e_i' D_{n,1}^{(2)}. \end{aligned}$$

By equation (A.2),

$$\begin{aligned} L_i^{(j)} &= C_i^{(j)} B_n^{(j)} + A_n^{(j)\prime} \mathcal{I}_i + \delta A_n^{(j)\prime} L_i^{(j-1)} \Pi_n \\ &= C_i^{(j)} B_n^{(j)} + \delta A_n^{(j)\prime} C_i^{(j-1)} B_n^{(j-1)} \Pi_n + A_n^{(j)\prime} \mathcal{I}_i + \delta A_n^{(j)\prime} A_n^{(j-1)\prime} \mathcal{I}_i \Pi_n \\ &\quad + \delta^2 A_n^{(j)\prime} A_n^{(j-1)\prime} L_i^{(j-2)} \Pi_n^2 \end{aligned}$$

$$
\begin{aligned}
&= C_i^{(j)} B_n^{(j)} + \delta A_n^{(j)\prime} C_i^{(j-1)} B_n^{(j-1)} \Pi_n + \cdots \\
&\quad + \delta^{j-2} A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(3)\prime} C_i^{(2)} B_n^{(2)} \Pi_n^{j-2} \\
&\quad + A_n^{(j)\prime} \mathcal{I}_i + \delta A_n^{(j)\prime} A_n^{(j-1)\prime} \mathcal{I}_i \Pi_n + \cdots + \delta^{j-2} A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(2)\prime} \mathcal{I}_i \Pi_n^{j-2} \\
&\quad + \delta^{j-1} A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(2)\prime} L_i^{(1)} \Pi_n^{j-1}.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
L_i^{(j)} &= C_i^{(j)} \left[R_n^{(j)}\right]^{-1} \left(I_n + \delta D_{n,1}^{(j)} \Pi_n + \delta^2 D_{n,2}^{(j)} \Pi_n^2 + \cdots + \delta^{j-1} D_{n,j-1}^{(j)} \Pi_n^{j-1}\right) \\
&\quad + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} \begin{pmatrix} \delta \Pi_n + \delta^2 D_{n,1}^{(j-1)} \Pi_n^2 + \delta^3 D_{n,2}^{(j-1)} \Pi_n^3 \\ + \cdots + \delta^{j-1} D_{n,j-2}^{(j-1)} \Pi_n^{j-1} \end{pmatrix} + \cdots \\
&\quad + A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(3)\prime} C_i^{(2)} \left[R_n^{(2)}\right]^{-1} \left(\delta^{j-2} \Pi_n^{j-2} + \delta^{j-1} D_{n,1}^{(2)} \Pi_n^{j-1}\right) \\
&\quad + A_n^{(j)\prime} \mathcal{I}_i + \delta A_n^{(j)\prime} A_n^{(j-1)\prime} \mathcal{I}_i \Pi_n + \cdots + \delta^{j-2} A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(2)\prime} \mathcal{I}_i \Pi_n^{j-2} \\
&\quad + \delta^{j-1} A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(2)\prime} L_i^{(1)} \Pi_n^{j-1} \\
&= \left(C_i^{(j)} \left[R_n^{(j)}\right]^{-1} + A_n^{(j)\prime} \mathcal{I}_i\right) \\
&\quad + \delta \left(C_i^{(j)} \left[R_n^{(j)}\right]^{-1} D_{n,1}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} + A_n^{(j)\prime} A_n^{(j-1)\prime} \mathcal{I}_i\right) \Pi_n \\
&\quad + \delta^2 \begin{pmatrix} C_i^{(j)} \left[R_n^{(j)}\right]^{-1} D_{n,2}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} D_{n,1}^{(j-1)} \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} C_i^{(j-2)} \left[R_n^{(j-2)}\right]^{-1} \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} A_n^{(j-2)\prime} \mathcal{I}_i \end{pmatrix} \Pi_n^2 + \cdots \\
&\quad + \delta^{j-1} \begin{pmatrix} C_i^{(j)} \left[R_n^{(j)}\right]^{-1} D_{n,j-1}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} D_{n,j-2}^{(j-1)} + \cdots \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(3)\prime} C_i^{(2)} \left[R_n^{(2)}\right]^{-1} D_{n,1}^{(2)} \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(1)\prime} \mathcal{I}_i S_n^{-1}. \end{pmatrix} \Pi_n^{j-1}.
\end{aligned}
$$

As $e_i' L_n^{*(j)} = e_i' L_i^{(j)}$, by applying the method of undetermined coefficients based on (A.5) and by taking e_i', we have

$$
e_i' D_{n,1}^{(j+1)} = e_i' \left(C_i^{(j)} \left[R_n^{(j)}\right]^{-1} + A_n^{(j)\prime} \mathcal{I}_i\right),
$$

$$
e_i' D_{n,2}^{(j+1)} = e_i' \left(C_i^{(j)} \left[R_n^{(j)}\right]^{-1} D_{n,1}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} + A_n^{(j)\prime} A_n^{(j-1)\prime} \mathcal{I}_i\right),
$$

$$e_i' D_{n,3}^{(j+1)} = e_i' \left(\begin{array}{c} C_i^{(j)} \left[R_n^{(j)} \right]^{-1} D_{n,2}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)} \right]^{-1} D_{n,1}^{(j-1)} \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} C_i^{(j-2)} \left[R_n^{(j-2)} \right]^{-1} \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} A_n^{(j-2)\prime} \mathcal{I}_i \end{array} \right), \cdots$$

and

$$e_i' D_{n,j}^{(j+1)} = e_i' \left(\begin{array}{c} C_i^{(j)} \left[R_n^{(j)} \right]^{-1} D_{n,j-1}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)} \right]^{-1} D_{n,j-2}^{(j-1)} + \cdots \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(3)\prime} C_i^{(2)} \left[R_n^{(2)} \right]^{-1} D_{n,1}^{(2)} \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(1)\prime} \mathcal{I}_i S_n^{-1}. \end{array} \right)$$

We observe $\left\{ D_{n,1}^{(1)}, D_{n,1}^{(2)}, \cdots \right\}$ (where $D_{n,1}^{(1)} = \mathbf{0}_{n \times n}$) characterize evolution of $\left\{ D_{n,k}^{(j+1)} \right\}_{k,j}$.

Proposition A.1.1 *A relationship between $Q_n^{*(j)}$ and $L_n^{*(j)}$ from $D_{n,1}^{(j+1)}$ is*

$$Q_n^{*(j)} = D_{n,1}^{(j+1)} \left(\gamma_0 I_n + \rho_0 W_n \right) - \gamma_0 I_n$$

for $j = 1, 2, \cdots$.

Proof of Proposition A.1.1. Note that $e_i' Q_n^{*(j)} = e_i' \left(Q_i^{(j)} + Q_i^{(j)\prime} \right)$ and

$$\begin{aligned}
& e_i' \left(Q_i^{(j)} + Q_i^{(j)\prime} \right) \\
= \; & e_i' A_n^{(j)\prime} \left\{ \mathcal{I}_i \left[-I_n + \lambda_0 \left(W_n + W_n' \right) \right] + \delta \left(Q_i^{(j-1)} + Q_i^{(j-1)\prime} \right) \right\} A_n^{(j)} \\
& + e_i' A_n^{(j)\prime} \mathcal{I}_i \left(\gamma_0 I_n + \rho_0 W_n \right) + e_i' \left(\gamma_0 I_n + \rho_0 W_n \right)' \mathcal{I}_i A_n^{(j)} - \gamma_0 e_i' \\
= \; & e_i' A_n^{(j)\prime} \left\{ \mathcal{I}_i \left[-I_n + \lambda_0 \left(W_n + W_n' \right) \right] + \delta \left(Q_i^{(j-1)} + Q_i^{(j-1)\prime} \right) \right\} A_n^{(j)} \\
& + e_i' A_n^{(j)\prime} e_i e_i' \left(\gamma_0 I_n + \rho_0 W_n \right) + \gamma_0 e_i' A_n^{(j)} - \gamma_0 e_i' \\
= \; & e_i' A_n^{(j)\prime} \left\{ \mathcal{I}_i \left[-I_n + \lambda_0 \left(W_n + W_n' \right) \right] + \delta \left(Q_i^{(j-1)} + Q_i^{(j-1)\prime} \right) \right\} \left[R_n^{(j)} \right]^{-1} \left(\gamma_0 I_n + \rho_0 W_n \right) \\
& + e_i' \left(\gamma_0 I_n + \rho_0 W_n \right)' \mathcal{I}_i \left[R_n^{(j)} \right]^{-1} \left(\gamma_0 I_n + \rho_0 W_n \right) + e_i' A_n^{(j)\prime} e_i e_i' \left(\gamma_0 I_n + \rho_0 W_n \right) - \gamma_0 e_i' \\
= \; & e_i' \left(\begin{array}{c} A_n^{(j)\prime} \left\{ \mathcal{I}_i \left[-I_n + \lambda_0 \left(W_n + W_n' \right) \right] + \delta \left(Q_i^{(j-1)} + Q_i^{(j-1)\prime} \right) \right\} \left[R_n^{(j)} \right]^{-1} \\ + \left(\gamma_0 I_n + \rho_0 W_n \right)' \mathcal{I}_i \left[R_n^{(j)} \right]^{-1} + A_n^{(j)\prime} \mathcal{I}_i \end{array} \right) \\
& \times \left(\gamma_0 I_n + \rho_0 W_n \right) \\
& - e_i' \gamma_0 I_n \\
= \; & e_i' \left(D_{n,1}^{(j+1)} \left(\gamma_0 I_n + \rho_0 W_n \right) - \gamma_0 I_n \right)
\end{aligned}$$

for $j = 2, 3, \cdots$, since $e_i' (\gamma_0 I_n + \rho_0 W_n) e_i = \gamma_0$ by $e_i' W_n e_i = w_{ii} = 0$ for all $i = 1, \cdots, n$. Q.E.D.

To have a stable system, a sufficient condition is $\left\| A_n^{(j+1)} \right\|_\infty < 1$ for each j. By the following mathematical result, we can check invertibility of $R_n^{(j+1)}$ and the possibility of representing its inverse as a Neumann series.

Proposition A.1.2 (Stewart (1998)) *Consider a linear operator $I_n - C_n$ satisfies $\lim_{j\to\infty} \|C_n^j\| = 0$ where $\|\cdot\|$ denotes a well-defined operator norm. Then, $I_n - C_n$ is invertible and its inverse has a Neumann series expansion:*

$$(I_n - C_n)^{-1} = \sum_{j=0}^{\infty} C_n^j.$$

Hence, for our model, the implied spatial time series process for Y_{nt} to be stable in both space and time dimensions, it suffices to assume that

$$\left\| \frac{\lambda_0}{1+\delta\gamma_0} W_n + \frac{\delta}{1+\delta\gamma_0} D_{n,1}^{(j+1)} (\gamma_0 I_n + \rho_0 W_n) \right\|_\infty < 1.$$

Then, $\left[R_n^{(j+1)}\right]^{-1}$ has the Neumann series expansion,

$$\left[R_n^{(j+1)}\right]^{-1} = \frac{1}{1+\delta\gamma_0} \left[I_n + \sum_{j=1}^{\infty} \left(\frac{1}{1+\delta\gamma_0}\right)^j \left(\lambda_0 W_n + \delta\gamma_0 D_{n,1}^{(j+1)} + \delta\rho_0 D_{n,1}^{(j+1)} W_n\right)^j \right].$$

A.2: Statistical results

In this section, we list components of asymptotic biases of the QMLE, and provide briefly proofs of Theorems 2.4.1, 2.4.2, 2.4.4 and Corollary 2.4.3. The detailed proofs can be found in our supplementary file.

First order derivatives of the log-likelihood function

Note that

$$\tilde{Y}_{nt} = A_n \tilde{Y}_{n,t-1}^{(-)} + \sum_{k=1}^{K} R_n^{-1} (I_n + \delta L_n^* \Pi_n) \tilde{X}_{nt,k} \beta_0 + R_n^{-1} \left(\tilde{\alpha}_{t,0} l_n + \tilde{\mathcal{E}}_{nt} \right).$$

The components of $\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \theta}$ are

$$\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \lambda} = \begin{array}{l} \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[\begin{array}{c} -R_{n\lambda} R_n^{-1} \left(\begin{array}{c} (\gamma_0 I_n + \rho_0 W_n) \tilde{Y}_{n,t-1}^{(-)} \\ + (I_n + \delta L_n^* \Pi_n) \tilde{X}_{nt,k} \beta_0 + \tilde{\alpha}_{t0} l_n \end{array} \right) \\ + \delta L_{n\lambda}^* \Pi_n \tilde{X}_{nt,k} \beta_0 \end{array} \right]' J_n \tilde{\mathcal{E}}_{nt} \\ + \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[\tilde{\mathcal{E}}'_{nt} (-R_n^{-1\prime} R'_{n\lambda}) J_n \tilde{\mathcal{E}}_{nt} - \sigma^2_{\epsilon,0} tr\left(-R_{n\lambda} R_n^{-1}\right) \right] \end{array},$$

$$\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \gamma} = \begin{array}{l} \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[\begin{array}{c} -R_{n\gamma} R_n^{-1} \left(\begin{array}{c} (\gamma_0 I_n + \rho_0 W_n) \tilde{Y}_{n,t-1}^{(-)} \\ + (I_n + \delta L_n^* \Pi_n) \tilde{X}_{nt,k} \beta_0 + \tilde{\alpha}_{t0} l_n \end{array} \right) \\ + \tilde{Y}_{n,t-1}^{(-)} + \delta L_{n\gamma}^* \Pi_n \tilde{X}_{nt,k} \beta_0 \end{array} \right]' J_n \tilde{\mathcal{E}}_{nt} \\ + \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[\tilde{\mathcal{E}}'_{nt} \left(-R_n^{-1\prime} R'_{n\gamma}\right) J_n \tilde{\mathcal{E}}_{nt} - \sigma^2_{\epsilon,0} tr\left(-R_{n\gamma} R_n^{-1}\right) \right] \end{array},$$

$$\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \rho} = \begin{array}{l} \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[\begin{array}{c} -R_{n\rho} R_n^{-1} \left(\begin{array}{c} (\gamma_0 I_n + \rho_0 W_n) \tilde{Y}_{n,t-1}^{(-)} \\ + (I_n + \delta L_n^* \Pi_n) \tilde{X}_{nt,k} \beta_0 + \tilde{\alpha}_{t0} l_n \end{array} \right) \\ + W_n \tilde{Y}_{n,t-1}^{(-)} + \delta L_{n\rho}^* \Pi_n \tilde{X}_{nt,k} \beta_0 \end{array} \right]' J_n \tilde{\mathcal{E}}_{nt} \\ + \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[\tilde{\mathcal{E}}'_{nt} \left(-R_n^{-1\prime} R'_{n\rho}\right) J_n \tilde{\mathcal{E}}_{nt} - \sigma^2_{\epsilon,0} tr\left(-R_{n\rho} R_n^{-1}\right) \right] \end{array},$$

$$\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \beta_k} = \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[(I_n + \delta L_n^* \Pi_n) \tilde{X}_{nt,k} \right]' J_n \tilde{\mathcal{E}}_{nt} \text{ for } k = 1, \cdots, K,$$

$$\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \sigma^2_\epsilon} = \frac{1}{2\sigma^4_{\epsilon,0}} \sum_{t=1}^{T} \left[\tilde{\mathcal{E}}'_{nt} J_n \tilde{\mathcal{E}}_{nt} - n\sigma^2_{\epsilon,0} \right].$$

Components of asymptotic biases of QMLEs

Here are the components of $\Delta_{1,nT}$, $\Delta_{2,nT}$, $a_{n,1}(\theta_0)$, and $a_{n,2}(\theta_0)$:

$$\Delta_{1,nT}^{\lambda} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left[\begin{array}{c} \left(-R_{n\lambda} R_n^{-1} (\gamma_0 I_n + \rho_0 W_n) \bar{U}_{nT,-1}\right)' J_n \bar{\mathcal{E}}_{nT} \\ + \bar{\mathcal{E}}'_{nT} (-R_n^{-1\prime} R'_{n\lambda}) J_n \bar{\mathcal{E}}_{nT} \end{array} \right],$$

$$\Delta_{1,nT}^{\gamma} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left[\begin{array}{c} \left((-R_{n\gamma} R_n^{-1} (\gamma_0 I_n + \rho_0 W_n) + I_n) \bar{U}_{nT,-1}\right)' J_n \bar{\mathcal{E}}_{nT} \\ + \bar{\mathcal{E}}'_{nT} (-R_n^{-1\prime} R'_{n\gamma}) J_n \bar{\mathcal{E}}_{nT} \end{array} \right],$$

$$\Delta_{1,nT}^{\rho} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left[\begin{array}{c} \left((-R_{n\rho} R_n^{-1} (\gamma_0 I_n + \rho_0 W_n) + W_n) \bar{U}_{nT,-1}\right)' J_n \bar{\mathcal{E}}_{nT} \\ + \bar{\mathcal{E}}'_{nT} (-R_n^{-1\prime} R'_{n\rho}) J_n \bar{\mathcal{E}}_{nT} \end{array} \right],$$

$\Delta_{1,nT}^{\beta_{1,k}} = \mathbf{0}_{K\times 1}$, $\Delta_{1,nT}^{\sigma^2_\epsilon} = \frac{1}{2\sigma^4_{\epsilon,0}} \sqrt{\frac{T}{n}} \bar{\mathcal{E}}'_{nT} J_n \bar{\mathcal{E}}_{nT}$,

$\Delta_{2,nT}^{\lambda} = \sqrt{\frac{T}{n}} \left[tr\left(-R_{n\lambda} R_n^{-1}\right) - tr\left(J_n(-R_{n\lambda} R_n^{-1})\right) \right]$,

$\Delta_{2,nT}^{\gamma} = \sqrt{\frac{T}{n}} \left[tr\left(-R_{n\gamma} R_n^{-1}\right) - tr\left(J_n(-R_{n\gamma} R_n^{-1})\right) \right]$,

$\Delta_{2,nT}^{\rho} = \sqrt{\frac{T}{n}} \left[tr\left(-R_{n\rho} R_n^{-1}\right) - tr\left(J_n(-R_{n\rho} R_n^{-1})\right) \right]$,

$\Delta_{2,nT}^{\beta} = \mathbf{0}_{K\times 1}$, and $\Delta_{2,nT}^{\sigma^2_\epsilon} = \sqrt{\frac{T}{n}} \frac{1}{2\sigma^2_{\epsilon,0}}$,

$$a_{n,1}(\theta_0)=\begin{bmatrix}\frac{1}{n}tr\left(J_n(-R_{n\lambda}A_n)\left(\sum_{h=0}^{\infty}A_n^h\right)R_n^{-1}\right)+\frac{1}{n}tr\left(J_n(-R_{n\lambda}R_n^{-1})\right)\\ \frac{1}{n}tr\left(J_n\left(-R_{n\gamma}A_n+I_n\right)\left(\sum_{h=0}^{\infty}A_n^h\right)R_n^{-1}\right)+\frac{1}{n}tr\left(J_n(-R_{n\gamma}R_n^{-1})\right)\\ \frac{1}{n}tr\left(J_n\left(-R_{n\rho}A_n+W_n\right)\left(\sum_{h=0}^{\infty}A_n^h\right)R_n^{-1}\right)+\frac{1}{n}tr\left(J_n(-R_{n\rho}R_n^{-1})\right)\\ \mathbf{0}_{K\times 1}\\ \frac{n-1}{n}\frac{1}{2\sigma_{\epsilon,0}^2}\end{bmatrix},$$

and

$$a_{n,2}(\theta_0)=[\frac{1}{n}l_n'(-R_{n\lambda}R_n^{-1})l_n,\frac{1}{n}l_n'(-R_{n\gamma}R_n^{-1})l_n,\frac{1}{n}l_n'(-R_{n\rho}R_n^{-1})l_n,\mathbf{0}_{1\times K},\frac{1}{2\sigma_{\epsilon,0}^2}]$$

Sketches of Proofs (Consistency and asymptotic normality)

Sketch of proof of Theorem 2.4.1. Consistency can be shown in three steps.

In the first step, we shall show the uniform convergence of sample average of the log-likelihood function, $\sup_{\theta\in\Theta}\left|\frac{1}{nT}\ln L_{nT,c}(\theta)-Q_{nT}(\theta)\right|\rightarrow_p 0$ as $T\rightarrow\infty$. The main component of $\frac{1}{nT}\ln L_{nT,c}(\theta)-Q_{nT}(\theta)$ is $\frac{1}{nT}\sum_{t=1}^{T}\left[\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)-E\left(\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)\right)\right]$. Since (i) θ is bounded in the compact parameter space Θ and $R_n(\theta_1)$, R_n^{-1}, and $L_n^*(\theta_1)$ are uniformly bounded in both row and column sum norms, uniformly in $\theta_1\in\Theta_1$, it follows that $R_n(\theta_1)R_n^{-1}-I_n$ and $L_n^*-L_n^*(\theta_1)$ are also uniformly bounded in row and column sum norms uniformly in $\theta_1\in\Theta_1$. By Lemmas 8 and 15 in Yu et al. (2008), $\frac{1}{nT}\sum_{t=1}^{T}\left[\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)-E\left(\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)\right)\right]\rightarrow_p 0$ uniformly in $\theta\in\Theta$. Since σ_ϵ^2 is assumed to be bounded away from zero,

$$\frac{1}{nT}\ln L_{nT,c}(\theta)-Q_{nT}(\theta)=-\frac{1}{2\sigma_\epsilon^2}\frac{1}{nT}\sum_{t=1}^{T}\left[\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)-E\left(\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)\right)\right]\rightarrow_p 0$$

uniformly in $\theta\in\Theta$.

Secondly, we will show that $Q_{nT}(\theta)$ is uniformly equicontinuous in $\theta\in\Theta$. Note that

$$\frac{1}{nT}\sum_{t=1}^{T} E\left(\mathcal{E}'_{nt}(\theta) J_n \mathcal{E}_{nt}(\theta)\right) = q_{nT,1}(\theta_1, \beta) + q_{nT,2}(\theta_1) + o(1)$$

where

$$q_{nT,1}(\theta_1, \beta) = \frac{1}{nT}\sum_{t=1}^{T} E\left[\begin{array}{c} (R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)) \tilde{Y}_{n,t-1}^{(-)} \\ +\tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ +R_n(\theta_1) R_n^{-1} \tilde{\mathbf{X}}_{nt}\beta_0 - \tilde{\mathbf{X}}_{nt}(\theta_1)\beta \end{array}\right]' \times J_n \left[\begin{array}{c} (R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)) \tilde{Y}_{n,t-1}^{(-)} \\ +\tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ +R_n(\theta_1) R_n^{-1} \tilde{\mathbf{X}}_{nt}\beta_0 - \tilde{\mathbf{X}}_{nt}(\theta_1)\beta \end{array}\right],$$

and $q_{nT,2}(\theta_1) = \frac{T-1}{nT}\sigma_{\epsilon,0}^2 tr\left(R_n^{-1\prime} R_n'(\theta_1) J_n R_n(\theta_1) R_n^{-1}\right)$. For the equicontinuity of $Q_{nT}(\theta)$, we verify (i) $\ln \sigma_\epsilon^2$ is uniformly continuous, (ii) $\frac{1}{n}\ln|R_n(\theta_1)|$ is uniformly equicontinuous, and (iii) $q_{nT,1}(\theta)$ and $q_{nT,2}(\theta_1)$ are uniformly equicontinuous. The basic idea of showing those properties is to verify that each component can be represented by $(\theta_1 - \theta_2)\cdot h_{nT}\left(\bar{\theta}\right)$, where θ_1, $\theta_2 \in \Theta$, $\bar{\theta}$ lies between θ_1 and θ_2, and $h_{nT}(\cdot)$ are uniformly bounded. Uniform boundedness of $h_{nT}(\cdot)$ comes from Assumptions 2.4.3 - 2.4.5. By applying Assumption 2.4.7, we achieve the desired result. Q.E.D.

Sketch of proof of Theorem 2.4.2. This proof relies on the Taylor expansion:

$$\sqrt{nT}\left(\hat{\theta}_{ml,nT} - \theta_0\right) = \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}\right)^{-1}\left(\begin{array}{c}\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta} \\ -\Delta_{1,nT} - \Delta_{2,nT}\end{array}\right)$$

where $\bar{\theta}_{nT}$ lies between θ_0 and $\hat{\theta}_{ml,nT}$. By Assumptions 2.4.2 (ii), 2.4.3 and 2.4.5,

$$\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}\right) - \Sigma_{\theta_0,nT} = \left\|\bar{\theta}_{nT} - \theta_0\right\| \cdot O_p(1) + O_p\left(\frac{1}{\sqrt{nT}}\right).$$

Theorem 2.4.1 implies $\left\|\bar{\theta}_{nT} - \theta_0\right\| = o_p(1)$. Under large T, $\Sigma_{\theta_0,nT}$ is nonsingular in θ around θ_0 by Assumption 2.4.8. These imply $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}$ is of $O_p(1)$ and invertible. Hence,

$$\sqrt{nT}\left(\hat{\theta}_{ml,nT} - \theta_0\right) = \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}\right)^{-1}$$

$$\cdot\left(\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta}-\Delta_{1,nT}-\Delta_{2,nT}\right),$$

which means $\hat{\theta}_{ml,nT}-\theta_0=O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T},\frac{1}{n}\right)\right)$. Note that

$$\begin{aligned}&\sqrt{nT}\left(\hat{\theta}_{ml,nT}-\theta_0\right)+\Sigma_{\theta_0,nT}^{-1}\cdot\left(\Delta_{1,nT}+\Delta_{2,nT}\right)\\&+O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T},\frac{1}{n}\right)\right)\cdot\left(\Delta_{1,nT}+\Delta_{2,nT}\right)\\=&\left(\Sigma_{\theta_0,nT}^{-1}+O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T},\frac{1}{n}\right)\right)\right)\cdot\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta}.\end{aligned}$$

Since (i) $\Sigma_{\theta_0}=\lim_{T\to\infty}\Sigma_{\theta_0,nT}$ exists and is nonsingular by Assumption 2.4.8, (ii) $\Delta_{1,nT}=\sqrt{\frac{n}{T}}a_{n,1}(\theta_0)+O\left(\sqrt{\frac{n}{T^3}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)$ by Lemmas 2.1 and 2.2 in our supplement file, and (iii) $\Delta_{2,nT}=\sqrt{\frac{T}{n}}a_{n,2}(\theta_0)$.

The last task is to investigate $\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta}$. The stochastic components of $\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta}$ take a linear-quadratic form, $\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}\xi_{nt,i}$, where $E\left(\xi_{nt,i}|\mathcal{F}_{n,t,i-1}\right)=0$

$$\mathcal{F}_{n,t,i}=\sigma\left(\epsilon_{11},\cdots,\epsilon_{n1},\cdots,\epsilon_{1,t-1},\cdots,\epsilon_{n,t-1},\epsilon_{1t},\cdots,\epsilon_{it}\right),\tag{A.6}$$

and $\mathcal{F}_{n,0,0}=\{\phi,\Omega\}$, where Ω is the sample space. Let $\mathcal{F}_{n,t,0}=\mathcal{F}_{n,t-1,n}$. Since $\mathcal{F}_{n,t,i-1}\subseteq\mathcal{F}_{n,t,i}$ and $\mathcal{F}_{n,t-1,0}\subseteq\mathcal{F}_{n,t,0}$, we construct the martingale difference arrays, $\{(\xi_{nt,i},\mathcal{F}_{n,t,i}):i=1,\cdots,n,andt=1,\cdots,T\}$. Then, we can apply the martingale central limit theorem to $\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}\xi_{nt,i}$ as Yu et al. (2008).[160] In consequence, we obtain $\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta}\to_d N\left(0,\Omega_{\theta_0}\right)$ as $T\to\infty$ and have the desired results. Q.E.D.

Sketch of proof of Corollary 2.4.3. By Theorem 2.4.2,

$$\begin{aligned}&\sqrt{nT}\left(\hat{\theta}_{ml,nT}-\theta_0\right)+\sqrt{\frac{n}{T}}\Sigma_{\theta_0,nT}^{-1}a_{n,1}(\theta_0)+\sqrt{\frac{T}{n}}\Sigma_{\theta_0,nT}^{-1}a_{n,2}(\theta_0)\\&+O_p\left(\max\left(\sqrt{\frac{n}{T^3}},\sqrt{\frac{T}{n^3}},\frac{1}{\sqrt{T}}\right)\right)\\&\to_d N\left(0,\Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1}\right).\end{aligned}$$

[160] Also, refer to Kelejian and Prucha (2001).

Since $\hat{\theta}^c_{ml,nT} = \hat{\theta}_{ml,nT} - \frac{1}{T}\left[-\Sigma^{-1}_{\theta,nT}a_{n,1}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \frac{1}{n}\left[-\Sigma^{-1}_{\theta,nT}a_{n,2}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}}$,

$$\sqrt{nT}\left(\hat{\theta}^c_{ml,nT} - \theta_0\right) \to_d N\left(0, \Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}\right)$$

if

$$\sqrt{\frac{n}{T}}\left(\left[\Sigma^{-1}_{\theta,nT}a_{n,1}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \Sigma^{-1}_{\theta_0,nT}a_{n,1}(\theta_0)\right) \to_p 0 \tag{A.7}$$

and

$$\sqrt{\frac{T}{n}}\left(\left[\Sigma^{-1}_{\theta,nT}a_{n,2}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \Sigma^{-1}_{\theta_0,nT}a_{n,2}(\theta_0)\right) \to_p 0. \tag{A.8}$$

Assumption 2.4.9, $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$ (with Assumptions 2.4.3 and 2.4.5) imply (A.7) and (A.8). The detailed arguments can be found in our supplementary file. Q.E.D.

Sketch of proof of Theorem 2.4.4. (i) First, note that $\hat{c}_{i,ml} = c_{i,ml}\left(\hat{\theta}_{ml,nT}\right)$. By Theorem 2.4.1 with $\sum_{t=1}^T \alpha_{t0} = 0$, we observe $c_{i,ml}\left(\hat{\theta}_{ml,nT}\right) - c_{i,0} = \frac{1}{T}\sum_{t=1}^T \epsilon_{it} + \left\|\hat{\theta}_{ml,nT} - \theta_0\right\| \cdot O_p(1) = \frac{1}{T}\sum_{t=1}^T \epsilon_{it} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \frac{1}{n}\right)\right)$ by Theorem 2.4.2. Under the rate $\frac{\sqrt{T}}{n} = o(1)$, $\frac{1}{T}\sum_{t=1}^T \epsilon_{it}$ will be the dominant term. Therefore, for each i, $\sqrt{T}\left(\hat{c}_{i,ml}(\hat{\theta}_{ml,nT}) - c_{i,0}\right) \to_d N(0, \sigma^2_{\epsilon,0})$ if $\frac{\sqrt{T}}{n} \to 0$; and $\hat{c}_{i,ml}(\hat{\theta}_{ml,nT})$'s are asymptotically independent from each other.

(ii) Using the same logic, the dominant term of $\sqrt{n}\left(\hat{\alpha}_{t,ml}(\hat{\theta}_{ml,nT}) - \alpha_{t0}\right)$ is $\frac{1}{\sqrt{n}}l_n'\mathcal{E}_{nt}$ if $\frac{\sqrt{n}}{T} = o(1)$. This yields $\sqrt{n}\left(\hat{\alpha}_{t,ml} - \alpha_{t0}\right) \to_d N(0, \sigma^2_{\epsilon,0})$ if $\frac{\sqrt{n}}{T} \to 0$; and the estimates $\hat{\alpha}_{t,ml}$'s for $t = 1, \cdots, T$ are asymptotically independent with each other.

(iii) Under Assumption 2.4.9, $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$, $c_{i,ml}\left(\hat{\theta}^c_{ml,nT}\right) - c_{i,0} = \frac{1}{T}\sum_{t=1}^T \epsilon_{it} + O_p\left(\frac{1}{\sqrt{nT}}\right)$ and $\hat{\alpha}_{t,ml}(\hat{\theta}^c_{ml,nT}) - \alpha_{t0} = \frac{1}{n}l_n'\mathcal{E}_{nt} + O_p\left(\frac{1}{\sqrt{nT}}\right)$ since $\left\|\hat{\theta}^c_{ml,nT} - \theta_0\right\| = O_p\left(\frac{1}{\sqrt{nT}}\right)$. We can apply the same strategies as Parts (i) and (ii). Q.E.D.

Appendix B: Appendix for Chapter 3

B.1. Linear-quadratic (LQ) approximation

Using the LQ perturbation method around the population averages[161], we want to find an approximate solution V_i^e. Observe that

$$u_i\left(Y_{nt}, Y_{n,t-1}, \eta_{it}\right) = \eta_{it} y_{it} + \gamma_0 y_{i,t-1} y_{it} - \frac{\gamma_0}{2} y_{i,t-1}^2 - \frac{1}{2} y_{it}^2 + \lambda_0 \sum_{j=1}^{n} w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt}.$$

Then, the non LQ components are only $w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt}$ for $j \neq i$ such that $d_{ij} \leq d_c$. Hence, we need to approximate the interaction term with the network link, $w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt}$. Let $\tilde{\mathbf{y}}_{t,ij} = [y_{i,t-1} - \bar{y}_i^\circ, y_{j,t-1} - \bar{y}_j^\circ, y_{it} - \bar{y}_i^\circ, y_{jt} - \bar{y}_j^\circ]'$. Around $\left(\bar{y}_i^\circ, \bar{y}_j^\circ\right)$, the second-order Taylor approximation of $w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt}$ (denoted by $w_{t,ij}^{LQ}\left[y_{i,t-1}, y_{j,t-1}, y_{it}, y_{jt}\right]$) is

$$\begin{aligned}
& w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt} \\
\simeq\; & w_{t,ij}^{LQ}\left[y_{i,t-1}, y_{j,t-1}, y_{it}, y_{jt}\right] = c_{ij} + [l_{ij,1}, l_{ij,2}, l_{ij,3}, l_{ij,4}]\tilde{\mathbf{y}}_{t,ij} \\
& + \frac{1}{2}\tilde{\mathbf{y}}_{t,ij}' \begin{bmatrix} q_{ij,11} & q_{ij,12} & q_{ij,13} & q_{ij,14} \\ q_{ij,12} & q_{ij,22} & q_{ij,23} & q_{ij,24} \\ q_{ij,13} & q_{ij,23} & 0 & q_{ij,34} \\ q_{ij,14} & q_{ij,24} & q_{ij,34} & 0 \end{bmatrix} \tilde{\mathbf{y}}_{t,ij}.
\end{aligned}$$

[161] A brief explanation can be found in Judd (1996), Judd's (1998) book Chapter 14.5 and Ljungqvist and Sargent (2002), pp. 143-145.

where $c_{ij} = w_{t,ij}\left[\bar{y}_i^\circ, \bar{y}_j^\circ\right]\bar{y}_i^\circ\bar{y}_j^\circ$, $l_{ij,1} = \frac{\partial w_{t,ij}\left[\bar{y}_i^\circ,\bar{y}_j^\circ\right]}{\partial y_{i,t-1}}\bar{y}_i^\circ\bar{y}_j^\circ$, $l_{ij,2} = \frac{\partial w_{t,ij}\left[\bar{y}_i^\circ,\bar{y}_j^\circ\right]}{\partial y_{j,t-1}}\bar{y}_i^\circ\bar{y}_j^\circ$, $l_{ij,3} = w_{t,ij}\left[\bar{y}_i^\circ, \bar{y}_j^\circ\right]\bar{y}_j^\circ$, $l_{ij,4} = w_{t,ij}\left[\bar{y}_i^\circ, \bar{y}_j^\circ\right]\bar{y}_i^\circ$, $q_{ij,11} = \frac{\partial^2 w_{t,ij}\left[\bar{y}_i^\circ,\bar{y}_j^\circ\right]}{\partial y_{i,t-1}^2}\bar{y}_i^\circ\bar{y}_j^\circ$, $q_{ij,12} = \frac{\partial^2 w_{t,ij}\left[\bar{y}_i^\circ,\bar{y}_j^\circ\right]}{\partial y_{i,t-1}\partial y_{j,t-1}}\bar{y}_i^\circ\bar{y}_j^\circ$, $q_{ij,13} = \frac{\partial w_{t,ij}\left[\bar{y}_i^\circ,\bar{y}_j^\circ\right]}{\partial y_{i,t-1}}\bar{y}_j^\circ$, $q_{ij,14} = \frac{\partial w_{t,ij}\left[\bar{y}_i^\circ,\bar{y}_j^\circ\right]}{\partial y_{j,t-1}}\bar{y}_i^\circ$, $q_{ij,22} = \frac{\partial^2 w_{t,ij}\left[\bar{y}_i^\circ,\bar{y}_j^\circ\right]}{\partial y_{j,t-1}^2}\bar{y}_i^\circ\bar{y}_j^\circ$, $q_{ij,23} = \frac{\partial w_{t,ij}\left[\bar{y}_i^\circ,\bar{y}_j^\circ\right]}{\partial y_{i,t-1}}\bar{y}_j^\circ$, $q_{ij,24} = \frac{\partial w_{t,ij}\left[\bar{y}_i^\circ,\bar{y}_j^\circ\right]}{\partial y_{j,t-1}}\bar{y}_i^\circ$, $q_{ij,34} = w_{t,ij}\left[\bar{y}_i^\circ, \bar{y}_j^\circ\right]$. Note that $q_{ij,33} = q_{ij,44} = 0$ and all components in $w_{t,ij}^{LQ}\left[y_{i,t-1}, y_{j,t-1}, y_{it}, y_{jt}\right]$ are zero if $i = j$.

Then, $u_i^e(\cdot)$ can be written as

$$
\begin{aligned}
& u_i^e\left(Y_{nt}, Y_{n,t-1}, \eta_{it}\right) \\
= \; & \left(\eta_{it} + \gamma_0 y_{i,t-1}\right) y_{it} - \frac{\gamma_0}{2} y_{i,t-1}^2 - \frac{1}{2} y_{it}^2 \\
& + \lambda_0 \sum_{j=1}^{n} \left\{ \begin{array}{c} \tilde{c}_{ij,0} + \tilde{c}_{ij,1} y_{i,t-1} + \tilde{c}_{ij,2} y_{j,t-1} + \tilde{c}_{ij,3} y_{it} + \tilde{c}_{ij,4} y_{jt} \\ + \frac{1}{2} q_{ij,11} y_{i,t-1}^2 + \frac{1}{2} q_{ij,22} y_{j,t-1}^2 + q_{ij,12} y_{i,t-1} y_{j,t-1} \\ + q_{ij,13} y_{i,t-1} y_{it} + q_{ij,23} y_{j,t-1} y_{it} + q_{ij,34} y_{jt} y_{it} \\ + q_{ij,14} y_{i,t-1} y_{jt} + q_{ij,24} y_{j,t-1} y_{jt} \end{array} \right\}
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{c}_{ij,0} &= c_{ij} - \left(l_{ij,1} + l_{ij,3}\right) \bar{y}_i^\circ - \left(l_{ij,2} + l_{ij,4}\right) \bar{y}_j^\circ + \left(\frac{1}{2} q_{ij,11} + q_{ij,13}\right) \left(\bar{y}_i^\circ\right)^2 \\
&\quad + \left(\frac{1}{2} q_{ij,22} + q_{ij,24}\right) \left(\bar{y}_j^\circ\right)^2 \\
&\quad + \left(q_{ij,12} + q_{ij,14} + q_{ij,23} + q_{ij,34}\right) \bar{y}_i^\circ \bar{y}_j^\circ, \\
\tilde{c}_{ij,1} &= l_{ij,1} - \left[q_{ij,11} + q_{ij,13}\right] \bar{y}_i^\circ - \left[q_{ij,12} + q_{ij,14}\right] \bar{y}_j^\circ, \\
\tilde{c}_{ij,2} &= l_{ij,2} - \left[q_{ij,12} + q_{ij,23}\right] \bar{y}_i^\circ - \left[q_{ij,22} + q_{ij,24}\right] \bar{y}_j^\circ, \\
\tilde{c}_{ij,3} &= l_{ij,3} - q_{ij,13} \bar{y}_i^\circ - \left[q_{ij,23} + q_{ij,34}\right] \bar{y}_j^\circ,
\end{aligned}
$$

and $\tilde{c}_{ij,4} = l_{ij,4} - q_{ij,24}\bar{y}_j^\circ - \left[q_{ij,14} + q_{ij,23}\right]\bar{y}_i^\circ$.

We observe that $u_i^e(\cdot)$ is a LQ function of its argument, so $V_i^e\left(Y_{n,t-1}, \eta_{nt}\right)$ will be a LQ function of $\left(Y_{n,t-1}, \eta_{nt}\right)$:

$$
V_i^e\left(Y_{n,t-1}, \eta_{nt}\right) = Y_{n,t-1}' Q_i^e Y_{n,t-1} + Y_{n,t-1}' L_i^e \eta_{nt} + Y_{n,t-1}' D_i^e + \eta_{nt}' E_i^e \eta_{nt} + \eta_{nt}' F_i^e + g_i^e
$$

where Q_i^e, L_i^e, and E_i^e are $n \times n$ matrices, D_i^e and F_i^e are n-dimensional row vectors, and g_i^e is a scalar. A corresponding vector of linear optimal actions is

$$Y_{nt}^e = A_n^e Y_{n,t-1} + B_n^e \eta_{nt} + C_n^e$$

where A_n^e and B_n^e denote $n \times n$ matrices, and C_n^e is an $n \times 1$ vector. The solutions A_n^e, B_n^e, C_n^e, and $\{Q_i^e, L_i^e, D_i^e, E_i^e, F_i^e, g_i^e\}_{i=1}^n$ can be obtained by the following equation:

$$\begin{aligned} V_i^e\left(Y_{n,t-1}, \eta_{nt}\right) &= \max_{y_{it}} \left\{u_i^e\left(y_{it}, Y_{-i,t}, Y_{n,t-1}, \eta_{it}\right) + \delta E_t V_i^e\left(y_{it}, Y_{-i,t}, \eta_{n,t+1}\right)\right\} \\ &= u_i^e\left(Y_{nt}^e, Y_{n,t-1}, \eta_{it}\right) + \delta E_t V_i^e\left(Y_{nt}^e, \eta_{n,t+1}\right. \end{aligned}$$

The components A_n^e, B_n^e, and C_n^e can be characterized by the first order conditions. For notations, let $a_{i.} = (a_{i1}, \cdots, a_{in})$ be a row vector consisting of $a_{i1}, \cdots, a_{in}$. By the first order conditions, we have

$$\begin{aligned} & \left(I_n - \lambda_0 Q_{34,n} - \delta Q_n^{e,*}\right) Y_{nt}^e \\ = \;& \left(\gamma_0 I_n + \lambda_0 diag_{i=1}^n \left(q_{i.,13} l_n\right) + \lambda_0 Q_{23,n}\right) Y_{n,t-1} + \left(I_n + \delta \rho_{\eta,0} L_n^{e,*}\right) \eta_{nt} + \lambda_0 \tilde{C}_{3,n} + \lambda_0 D_n^{e,*} \end{aligned}$$

where for each i $e_{ni}' Q_{34,n} = q_{i.,34}$, $e_{ni}' Q_n^{e,*} = e_{ni}'\left(Q_i^e + Q_i^{e\prime}\right)$, $e_{ni}' Q_{23,n} = q_{i.,23}$, $e_{ni}' L_n^{e,*} = e_{ni}' L_i^e$, $e_{ni}' \tilde{C}_{3,n} = \tilde{c}_{i.,3} l_n$, and $e_{ni}' D_n^{e,*}$. By defining $R_n^e = I_n - \lambda_0 Q_{34,n} - \delta Q_n^{e,*}$, we have

$$\begin{aligned} A_n^e &= \left(R_n^e\right)^{-1}\left(\gamma_0 I_n + \lambda_0 diag_{i=1}^n \left(q_{i.,13} l_n\right) + \lambda_0 Q_{23,n}\right), \\ B_n^e &= \left(R_n^e\right)^{-1}\left(I_n + \delta \rho_{\eta,0} L_n^{e,*}\right), \end{aligned}$$

and $C_n^e = \left(R_n^e\right)^{-1}\left(\lambda_0 \tilde{C}_{3,n} + \delta D_n^{e,*}\right)$. Note that computing Y_{nt}^e just requires evaluating Q_i^e, L_i^e, and D_i^e.

Next, we will provide formulas for the components $\{Q_i^e, L_i^e, D_i^e\}_{i=1}^n$. Define $I_i = e_{ni} e_{ni}'$ for notational convenience. For each i,

$$
\begin{aligned}
Q_i^e &= \gamma_0 A_n^{e\prime} I_i - \frac{\gamma_0}{2} I_i - \frac{1}{2} A_n^{e\prime} I_i A_n^e + \delta A_n^{e\prime} Q_i^e A_n^e + \frac{\lambda_0}{2} l_n^{\prime} q_{i.,11}^{\prime} I_i + \frac{\lambda_0}{2} diag_{j=1}^{n} q_{ij,22} \\
&\quad + \lambda_0 e_{ni} q_{i.,12} + \lambda_0 l_n^{\prime} q_{i.,13}^{\prime} I_i A_n^e + \lambda_0 e_{ni} \left(q_{i.,23} + q_{i.,14} \right) A_n^e \\
&\quad + \lambda_0 A_n^{e\prime} e_{ni} q_{i.,34} A_n^e + \lambda_0 diag_{j=1}^{n} q_{ij,24} A_n^e, \\
L_i^e &= A_n^{e\prime} I_i + \gamma_0 I_i B_n^e - A_n^{e\prime} I_i B_n^e + \delta A_n^{e\prime} \left(Q_i^e + Q_i^{e\prime} \right) B_n^e + \delta \rho_{\eta,0} A_n^{e\prime} L_i \\
&\quad + \lambda_0 q_{i.,13} l_n I_i B_n^e + \lambda_0 e_{ni} q_{i.,23} B_n^e \\
&\quad + \lambda_0 A_n^{e\prime} \left(e_{ni} q_{i.,34} + q_{i.,34}^{\prime} e_{ni}^{\prime} \right) B_n^e + \lambda_0 e_{ni} q_{i.,14} B_n^e + \lambda_0 diag_{j=1}^{n} q_{ij,24} B_n^e,
\end{aligned}
$$

and

$$
\begin{aligned}
D_i^e &= \gamma_0 I_i C_n^e - A_n^{e\prime} I_i C_n^e + \delta A_n^{e\prime} \left(Q_i^e + Q_i^{e\prime} \right) C_n^e + \delta A_n^{e\prime} D_i^e + \lambda_0 \tilde{c}_{i.,1} l_n e_{ni} + \lambda_0 \tilde{c}_{i.,2}^{\prime} \\
&\quad + \lambda_0 A_n^{e\prime} e_{ni} l_n^{\prime} \tilde{c}_{i.,3}^{\prime} + \lambda_0 A_n^{e\prime} \tilde{c}_{i.,4}^{\prime} + \lambda_0 q_{i.,13} l_n I_i C_n^e + \lambda_0 e_{ni} q_{i.,23} C_n^e \\
&\quad + \lambda_0 A_n^{e\prime} \left(e_{ni} q_{i.,34} + q_{i.,34}^{\prime} e_{ni}^{\prime} \right) C_n^e + \lambda_0 e_{ni} q_{i.,14} C_n^e + \lambda_0 diag_{j=1}^{n} q_{ij,24} C_n^e.
\end{aligned}
$$

Hence, $\Delta_n^* = A_n^{e\prime}$, and

$$
\begin{aligned}
& E_t \left(\frac{\partial V_i^e \left(Y_{n,t+1}^e, \eta_{n,t+2} \right)}{\partial Y_{n,t+1}} \right) \\
= \; & E_t \left(\left(Q_i^e + Q_i^{e\prime} \right) Y_{n,t+1}^e + L_i^e \eta_{n,t+2} + D_i^e \right) \\
= \; & \left(Q_i^e + Q_i^{e\prime} \right) A_n^e Y_{nt} + \left(Q_i^e + Q_i^{e\prime} \right) \left(\rho_{\eta,0} B_n^e + \rho_{\eta,0}^2 L_i^e \right) \eta_{nt} + \left(Q_i^e + Q_i^{e\prime} \right) C_n^e + D_i^e
\end{aligned}
$$

for each t and i. Hence, the i^{th}-element of $\nabla V_{n,t+2}^e$ is

$$
\begin{aligned}
& E_t \left(\left(e_{ni}^{\prime} \Delta_n^* \circ \tilde{e}_{ni}^{\prime} \right) \left(\tilde{e}_{ni} \circ \frac{\partial V_i^e \left(Y_{n,t+1}^e, \eta_{n,t+2} \right)}{\partial Y_{n,t+1}} \right) \right) \\
= \; & \left(e_{ni}^{\prime} \Delta_n^* \circ \tilde{e}_{ni}^{\prime} \right) \left\{ \begin{array}{c} \left(Q_i^e + Q_i^{e\prime} \right) A_n^e Y_{nt} + \left(Q_i^e + Q_i^{e\prime} \right) \left(\rho_{\eta,0} B_n^e + \rho_{\eta,0}^2 L_i^e \right) \eta_{nt} \\ + \left(Q_i^e + Q_i^{e\prime} \right) C_n^e + D_i^e \end{array} \right\}.
\end{aligned}
$$

Principle of optimality

A main purpose of this section is to obtain the stationary property of the values $(V_i(\cdot))$ and optimal policy function $(f(\cdot))$. Note that our model does not belong to a linear-quadratic programming since the agent's current decision-making can non-linearly affect his/her own and opponents' future marginal payoffs. To achieve the principle of optimality, a key part is to have a bounded lifetime value given an initial condition $(Y_{n,t-1}, \eta_{nt})$: the infinite sum of current and (expected) future payoffs should be bounded given $(Y_{n,t-1}, \eta_{nt})$. When boundedness of the infinite sum of payoffs is achieved, we can represent the agent's lifetime value and his/her optimal policy function as recursive forms of the state variables, i.e., $V_i(Y_{n,t-1}, \eta_{nt})$ and $f_i(Y_{n,t-1}, \eta_{nt})$ where $V_i(\cdot)$ and $f_i(\cdot)$ do not rely on a specific time t.

As contrary to a myopic (conventional) SDPD model specification, in case of forward-looking models, it is difficult to find a combination of parameters to being a stable space-time process. Hence, we introduce the specification of the agent's choice set using the vector of myopic choices. Note that the agents' values and optimal policy functions are the limit functions of some sequences of functions. The assumption below states that the NE vector from the myopic agent assumption can be an initial guess to calculate those limit functions when $(Y_{n,t-1}, \eta_{nt})$ is given.

Assumption B.11 *Let* $Y_{nt}^B = S_{nt}^{-1}(\gamma_0 Y_{n,t-1} + \eta_{nt})$ *where* $S_{nt} = I_n - \lambda_0 W_{nt}$. *Given* $(Y_{n,t-1}, \eta_{nt})$, *there exists* $M_y > 0$ *such that each* y_{it}^* *is an interior solution in*

$$\Gamma(Y_{n,t-1}, \eta_{nt}) = \left[\min_{i=1,\cdots,n} y_{it}^B - M_y, \max_{i=1,\cdots,n} y_{it}^B + M_y\right] \tag{B.1}$$

where y_{it}^B *denotes the* i^{th}*-element of* Y_{nt}^B.

Assumption B.11 means that we restrict the outcome space for the agents' lifetime problem. However, it does not mean that the forward-looking agents' optimal choices (Y_{nt}^{*}) are similar to the myopic ones (we can take a large but bounded $M_y > 0$). Note that the optimal policy function $f\left(Y_{n,t-1}, \eta_{nt}\right)$ is a unique limit of sequence of functions $\left\{f^{(l)}\left(Y_{n,t-1}, \eta_{nt}\right)\right\}$, i.e., $\lim_{l\to\infty} f^{(l)}\left(Y_{n,t-1}, \eta_{nt}\right) = f\left(Y_{n,t-1}, \eta_{nt}\right)$. Assumption B.11 means that we can achieve the limit $f\left(Y_{n,t-1}, \eta_{nt}\right)$ when we start with $Y_{nt}^{B} = f^{(1)}\left(Y_{n,t-1}, \eta_{nt}\right)$. Invertibility of S_{nt} is for well-definedness of Y_{nt}^{B}, which is a vector of unique maximizers of per period payoffs for all $i = 1, \cdots, n$.[162] Conditional on $\left(Y_{n,t-1}, \eta_{nt}\right)$, hence, all elements of Y_{nt}^{B} become bounded. The existence of $M_y > 0$ is supposed so that $\Gamma\left(Y_{n,t-1}, \eta_{nt}\right) \neq \phi$. Also, restricting each agent's choice to the interior of $\Gamma\left(Y_{n,t-1}, \eta_{nt}\right)$ leads to avoiding explosive his/her lifetime values, so the solution to ICP $V_i^{*}\left(Y_{n,t-1}, \eta_{nt}\right)$ will be bounded.

Now we establish a time-invariant functional form of agents' optimal decisions by showing equivalence of the two solutions to the ICP and FE. On one hand, we claim that the solution to ICP $V_i^{*}(\cdot)$ implies that of FE $V_i(\cdot)$. Note that $\Gamma\left(Y_{n,t-1}, \eta_{nt}\right) \neq \phi$. Then, the agent's objective function (lifetime payoff) is well-defined for every point in the feasible choice set. Thus, the agent's lifetime value will not be explosive given $\left(Y_{n,t-1}, \eta_{nt}\right)$. Hence, $V_i^{*}(\cdot)$ satisfies the FE by Theorem 4.2 in Stokey et al. (1989).

On the other hand, we want to know that the solution $V_i(\cdot)$ to the FE satisfies that to the ICP $V_i^{*}(\cdot)$. The maximum operator $\mathcal{T}$, the Bellman equation is characterized by

$$V_i\left(Y_{n,t-1}, \eta_{nt}\right) = \mathcal{T}\left(V_i\right)\left(Y_{n,t-1}, \eta_{nt}\right) = \max_{y_{it}} \left\{ \begin{array}{c} u_i\left(y_{it}, Y_{-i,t}^{*}, Y_{n,t-1}, \eta_{it}\right) \\ +\delta E_t\left(V_i\left(y_{it}, Y_{-i,t}^{*}, \eta_{n,t+1}\right)\right) \end{array} \right\},$$

[162] Due to existence of the cost function $c(\cdot,\cdot)$, we have $\frac{\partial^2 u_{it}}{\partial y_{it}^2} = -1$ (strict concavity of u_{it}), the agent i's t^{th}-period payoff u_{it} eventually decreases in y_{it}. It leads to the existence and uniqueness of Y_{nt}^{B} without explicit constraints.

where the functional solution $V_i(\cdot)$ is the fixed point of $\mathcal{T}$. The solution $V_i(\cdot)$ should be a continuous and bounded function of $(Y_{n,t-1}, \eta_{nt})$ to have a recursive relationship. First, given $(Y_{n,t-1}, \eta_{nt})$ the vector Y_{nt}^B is the unique optimizer of $\{u_{it}\}_{i=1}^n$: for any strategy profile Y_{nt}

$$\begin{aligned} & u_i(Y_{nt}, Y_{n,t-1}, \eta_{it}) \\ \leq\ & u_i\left(Y_{nt}^B, Y_{n,t-1}, \eta_{it}\right) \\ =\ & \gamma_0^2 Y_{n,t-1}' S_{nt}^{-1\prime} e_{ni} e_{ni}' S_{nt}^{-1} Y_{n,t-1} - \frac{\gamma_0}{2} y_{i,t-1}^2 \\ & +2\gamma_0 Y_{n,t-1}' S_{nt}^{-1\prime} e_{ni} e_{ni}' S_{nt}^{-1} \eta_{nt} + \eta_{nt}' S_{nt}^{-1\prime} e_{ni} e_{ni}' S_{nt} \eta_{nt} \end{aligned}$$

for all $i = 1, \cdots, n$, which is bounded in every t. By the monotonicity of integrals, for any Y_{nt} $E_{t-1}(u_i(Y_{nt}, Y_{n,t-1}, \eta_{it})) \leq E_{t-1}\left(u_i\left(Y_{nt}^B, Y_{n,t-1}, \eta_{it}\right)\right) < \infty$ for all t. Hence, for all choices $\{Y_{ns}\}_{s=t}^{\infty}$ satisfying (B.1)

$$u_i(Y_{nt}, Y_{n,t-1}, \eta_{it}) + \sum_{s=1}^{\infty} \delta^s E_t(u_i(Y_{n,t+s}, Y_{n,t+s-1}, \eta_{i,t+s})) \leq \hat{V}_i$$

where $\hat{V}_i = \frac{1}{1-\delta} \sup_t E_{t-1} u_i\left(Y_{nt}^B, Y_{n,t-1}, \eta_{it}\right) < \infty$, which is an upper bound of the agent i's lifetime value. This implies that the solution to i's ICP $V_i^*(Y_{n,t-1}, \eta_{nt})$ is bounded by $\hat{V}_i$. Then, by Theorem 4.14 in Stocky et al. (1989), $V_i(\cdot) = V_i^*(\cdot)$.

Discussions on the second-order condition

In this subsection, we provide discussions on the second-order condition for optimality. By introducing a static (linear-quadratic) network game, we explain the motivations and implications. Assume that (i) there are n agents, (ii) they are interrelated via a spatial network W_n, and (iii) each agent i has his/her exogenous characteristic η_i and chooses y_i by maximizing the linear-quadratic payoff,

$$u_i(y_i, Y_{-i,n}, \eta_i) = (\eta_i + \lambda_0 w_{i.} Y_n) y_i - \frac{1}{2} y_i^2$$

where $Y_n = (y_1, \cdots, y_n)'$, $Y_{-i,n} = (y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_n)'$, and $w_{i.}$ denotes the i^{th}-row of W_n. The first-order condition is

$$\eta_i + \lambda_0 w_{i.} Y_n - y_i = 0. \tag{B.2}$$

Since the second-order condition is $\frac{\partial^2 u_i}{\partial y_i^2} = -1 < 0$ (strictly concavity), the first-order condition is sufficient to characterize the optimum y_i^*. For this, hence, we do not need a restriction on λ_0 and/or W_n.[163]

Now we need to check the second-order condition to employ the first-order condition as a sufficient condition for optimality. Here is the high level assumption for strict concavity of the ICP.

Assumption B.12 *(i) y_{it}^* takes a (first-order) Markov strategy, i.e., $y_{it}^* = y_{it}^* (Y_{n,t-1}, \eta_{nt})$.*

(ii) Assumption B.11 holds

(iii) $\sup_t \phi_{\max} \left(\frac{\partial \mathcal{J}(Y_{n,t-1}, \eta_{nt}, Y_{nt})}{\partial Y_{nt}'} |_{Y_{nt}^*} \right) < 0$ where $\mathcal{J}(\cdot)$ is an $n \times 1$ vector of first-order conditions, and $\phi_{\max}(\cdot)$ denotes the maximum eigenvalue.

Assumption B.12 (i) restricts the state dependency of the optimal choices. Since the ICP is not a linear-quadratic dynamic programming, we introduce a restriction on the agent's choice set by Assumption B.12 (ii). That assumption is a device of having a bounded lifetime value. Assumption B.12 (iii) is strict concavity of the agent i's lifetime problem. Mathematically, it is equivalent that $\frac{\partial \mathcal{J}(Y_{n,t-1}, \eta_{nt}, Y_{nt})}{\partial Y_{nt}'} |_{Y_{nt}^*}$ is

[163] Even though we achieve the optimality of equation (B.2) by $\frac{\partial^2 u_i}{\partial y_i^2} = -1 < 0$, obtaining a unique NE is a different story. Consider the characterization of a unique NE. By the first-order conditions, we have the following system:

$$Y_n^* = \lambda_0 W_n Y_n^* + \eta_n \tag{B.3}$$

where $Y_n^* = (y_1^*, \cdots, y_n^*)'$ and $\eta_n = (\eta_1, \cdots, \eta_n)'$. To achieve uniqueness of Y_n^*, we need to impose $\|\lambda_0 W_n\| < 1$ to being Y_n^* as a unique fixed point of the system (B.3). In summary, we firstly need to check the second-order conditions. And then, try to find conditions specifying a unique NE from the first-order conditions.

negative definite. Under Assumption B.12 (iii), the first-order condition is sufficient to characterize the optimum.

Discussions on the qualities of the LQ approximation

Note that the difference between two solutions, (i) under full rationality and (ii) under bounded rationality, can be captured by studying qualities of the LQ approximation. Here are some conditions that the two solutions are similar. In this analysis, we suppose $\dim(z_{it}) = 1$.

Condition B.11 *(i) For each t, $|\eta_{it}|$ is not large for every i.*

(ii) $d_c > 0$ is not large and uniformly bounded in n.

(iii) $|\psi_0|$, α_e and α_d are small.

(iv) $|\lambda|$ is small.

Consider Condition B.11 (i). Note that η_{nt} drives the dynamics of Y_{nt}. Note that the optimal policy function $f\left(Y_{n,t-1}, \eta_{nt}\right)$ is continuously differentiable to its arguments since the agent's payoff function $u_i\left(Y_{nt}, Y_{n,t-1}, \eta_{it}\right)$ is supposed to be infinitely differentiable (Santos (1991)). It implies that $Y_{nt} = f\left(Y_{n,t-1}, \eta_{nt}\right)$ is a Lipschitz continuous function of η_{nt}. In consequences, the outcome process $\{Y_{nt}\}$ stably evolves around the optimal steady state $\bar{Y}_n^\circ = (\bar{y}_1^\circ, \cdots, \bar{y}_n^\circ)'$ such that $\bar{Y}_n^\circ = f\left(\bar{Y}_n^\circ, \bar{\eta}_n^\circ\right)$. This condition is consistent with the conventional qualification condition of the LQ approximation (e.g., Benigno and Woodford (2012)). In a data set, we can directly precheck the deviations, $y_{it} - \bar{y}_i^\circ$.

Condition B.11 (ii), (iii), and (iv) are specific ones in our model. To investigate them, we firstly consider the LQ approximation of the per period payoff

$u_i\left(Y_{nt}, Y_{n,t-1}, \eta_{it}\right)$. Observe that

$$\begin{aligned} u_i\left(Y_{nt}, Y_{n,t-1}, \eta_{it}\right) &= \eta_{it} y_{it} + \gamma y_{i,t-1} y_{it} - \frac{\gamma}{2} y_{i,t-1}^2 - \frac{1}{2} y_{it}^2 \\ &\quad + \lambda \sum_{j=1}^{n} w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt}. \end{aligned}$$

i.e., the non LQ components are only $\lambda \sum_{j=1}^{n} w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt}$. Hence, our approximation target is to approximate a network link combined with activity interactions, $w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt}$.

Note that $d_c > 0$ characterizes the number of approximation terms (i.e., $w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt} = 0$ if $d_{ij} > d_c$). If d_c is large, so is $\#\left\{j : w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] > 0\right\}$. Hence, considering a sparse W_{nt} can reduce potential numerical approximation errors. If $\#\left\{j : w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] > 0\right\}$ is uniformly bounded in n, the approximation errors do not explode in increasing n since the number of approximated functions is uniformly bounded.[164] Choose such j $(d_{ij} \leq d_c)$ and consider the LQ approximation of $w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt}$. Around $\left(\bar{y}_i^{\circ}, \bar{y}_j^{\circ}\right)$, the second-order Taylor approximation of $w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt}$ (denoted by $w_{t,ij}^{LQ}\left[y_{i,t-1}, y_{j,t-1}, y_{it}, y_{jt}\right]$) is

$$w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt} = w_{t,ij}^{LQ}\left[y_{i,t-1}, y_{j,t-1}, y_{it}, y_{jt}\right] + \Psi_{t,ij}$$

where $\Psi_{t,ij}$ denotes the approximation error.

Recall that $w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt}$ is a real analytic function around $\left(\bar{y}_i^{\circ}, \bar{y}_j^{\circ}\right)$. For notational convenience, $\omega = \left(\omega_1, \omega_2, \omega_3, \omega_4\right)'$, $\omega_1 = y_{i,t-1}$, $\omega_2 = y_{j,t-1}$, $\omega_3 = y_{it}$, and $\omega_4 = y_{jt}$. We set $\tilde{\omega}_1 = y_{i,t-1} - \bar{y}_i$, and $\tilde{\omega}_2$, $\tilde{\omega}_3$, $\tilde{\omega}_4$ and $\tilde{\omega}$ are defined similarly. The approximation error $\Psi_{t,ij}$ can be characterized by the third order Taylor expansion:

$$\Psi_{t,ij} = w_{t,ij}\left[y_{i,t-1}, y_{j,t-1}\right] y_{it} y_{jt} - w_{t,ij}^{LQ}\left[y_{i,t-1}, y_{j,t-1}, y_{it}, y_{jt}\right]$$

[164] In spatial econometric literature, the number of j satisfying $d_{ij} \leq d_c$ for each i is assumed to independent with the number of units n.

$$= \frac{1}{6}\sum_{k_1=1}^{4}\sum_{k_2=1}^{4}\sum_{k_3=1}^{4}\frac{\partial^3 w_{t,ij}\left[\omega_1,\omega_2\right]\omega_3\omega_4}{\partial\omega_{k_1}\partial\omega_{k_2}\partial\omega_{k_3}}|_{\bar{\omega}^*}\tilde{\omega}_{k_1}\tilde{\omega}_{k_2}\tilde{\omega}_{k_3}$$

where $\bar{\omega}^\circ = \left(\bar{y}_i^\circ, \bar{y}_j^\circ, \bar{y}_i^\circ, \bar{y}_j^\circ\right)'$ and $\bar{\omega}^*$ lies between ω and $\bar{\omega}^\circ$. First, observe that $\Psi_{t,ij}$ is a function of $y_{i,t-1} - \bar{y}_i^\circ$, $y_{j,t-1} - \bar{y}_j^\circ$, $y_{it} - \bar{y}_i^\circ$, and $y_{jt} - \bar{y}_j^\circ$. The components of $\frac{\partial^3 w_{t,ij}[\omega_1,\omega_2]\omega_3\omega_4}{\partial\omega_{k_1}\partial\omega_{k_2}\partial\omega_{k_3}}$ are

(i) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{i,t-1}^3}|_{\bar{\omega}^*} = \frac{\partial^3 w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{i,t-1}^3}\bar{y}_i^*\bar{y}_j^*$,

(ii) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{i,t-1}^2\partial y_{j,t-1}}|_{\bar{\omega}^*} = \frac{\partial^3 w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{i,t-1}^2\partial y_{j,t-1}}\bar{y}_i^*\bar{y}_j^*$,

(iii) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{i,t-1}^2\partial y_{it}}|_{\bar{\omega}^*} = \frac{\partial^2 w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{i,t-1}^2}\bar{y}_j^*$,

(iv) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{i,t-1}^2\partial y_{jt}}|_{\bar{\omega}^*} = \frac{\partial^2 w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{i,t-1}^2}\bar{y}_i^*$,

(v) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{i,t-1}\partial y_{j,t-1}^2}|_{\bar{\omega}^*} = \frac{\partial^3 w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{i,t-1}\partial y_{j,t-1}^2}\bar{y}_i^*\bar{y}_j^*$,

(vi) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{i,t-1}\partial y_{it}^2}|_{\bar{\omega}^*} = 0$,

(vii) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{i,t-1}\partial y_{jt}^2}|_{\bar{\omega}^*} = 0$,

(viii) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{i,t-1}\partial y_{j,t-1}\partial y_{it}}|_{\bar{\omega}^*} = \frac{\partial^2 w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{i,t-1}\partial y_{j,t-1}}\bar{y}_j^*$,

(ix) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{i,t-1}\partial y_{j,t-1}\partial y_{jt}}|_{\bar{\omega}^*} = \frac{\partial^2 w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{i,t-1}\partial y_{j,t-1}}\bar{y}_i^*$,

(x) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{i,t-1}\partial y_{it}\partial y_{jt}}|_{\bar{\omega}^*} = \frac{\partial w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{i,t-1}}$,

(xi) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{j,t-1}^3}|_{\bar{\omega}^*} = \frac{\partial^3 w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{j,t-1}^3}\bar{y}_i^*\bar{y}_j^*$,

(xii) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{j,t-1}^2\partial y_{it}}|_{\omega=\bar{\omega}^*} = \frac{\partial^2 w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{j,t-1}^2}\bar{y}_j^*$,

(xiii) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{j,t-1}^2\partial y_{jt}}|_{\omega=\bar{\omega}^*} = \frac{\partial^2 w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{j,t-1}^2}\bar{y}_i^*$,

(xiv) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{j,t-1}\partial y_{it}^2}|_{\bar{\omega}^*} = 0$,

(xv) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{j,t-1}\partial y_{jt}^2}|_{\bar{\omega}^*} = 0$,

(xvi) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{j,t-1}\partial y_{it}\partial y_{jt}}|_{\bar{\omega}^*} = \frac{\partial w_{t,ij}\left[\bar{y}_i^*,\bar{y}_j^*\right]}{\partial y_{j,t-1}}$,

(xvii) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{it}^3}|_{\bar{\omega}^*} = 0$,

(xviii) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{it}^2\partial y_{jt}}|_{\bar{\omega}^*} = 0$,

(xix) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{it}\partial y_{jt}^2}|_{\bar{\omega}^*} = 0$,

(xx) $\frac{\partial^3 w_{t,ij}[y_{i,t-1},y_{j,t-1}]y_{it}y_{jt}}{\partial y_{jt}^3}|_{\bar{\omega}^*} = 0.$

We observe that $\Psi_{t,ij}$ can be affected by $\frac{\partial^k w_{t,ij}[y_{i,t-1},y_{j,t-1}]}{\partial y_{i,t-1}^{k_1} \partial y_{j,t-1}^{k_2}}$ where k_1, $k_2 = 0, 1, 2, 3$, and $k = 1, 2, 3$ such that $k = k_1 + k_2$. It means that the functional forms of $h(\cdot)$ and $h_d(\cdot)$ with parameter $\gamma_{Z,0}$ (Condition B.11 (iii)) affect the quality of the LQ approximation. To see the details, consider $h(z_{it}, z_{jt}) = |z_{it} - z_{jt}|^{-\alpha_e}$ and $h(d_{ij}, d_c) = d_{ij}^{-\alpha_d} \mathbf{1}\{d_{ij} \leq d_c\}$ where α_e and α_d are positive sensitivity parameters. The first order derivative of $w_{t,ij}[y_{i,t-1}, y_{j,t-1}]$ with respect to $y_{i,t-1}$ is

$$\frac{\partial w_{t,ij}[y_{i,t-1}, y_{j,t-1}]}{\partial y_{i,t-1}} = -\psi_0 \alpha_e |z_{it} - z_{jt}|^{-(1+\alpha_e)} sgn(z_{it} - z_{jt}) d_{ij}^{-\alpha_d} \mathbf{1}\{d_{ij} \leq d_c\}.$$

In this casse, there exist three factors influencing the LQ approximation: (i) two sensitivity parameters α_e and α_d, (ii) the levels of distances, $|z_{it} - z_{jt}|$ and d_{ij}, and (iii) the magnitude of effect of $Y_{n,t-1}$ on Z_{nt} (captured by $\gamma_{Z,0}$).

Here is an example.

Example B.12 *Consider a simple linear formation function* $w_{t,ij}[y_{i,t-1}, y_{j,t-1}] = \Phi_i y_{i,t-1} + \Phi_j y_{j,t-1} + Geo_{ij}$ *where* Φ_i *and* Φ_j *denote respectively coefficients of* $y_{i,t-1}$ *and* $y_{j,t-1}$, *and* Geo_{ij} *is a part of* $w_{t,ij}[y_{i,t-1}, y_{j,t-1}]$ *purely constructed by geographical relationships between* i *and* j. *In this case, we observe that the all components above third-order derivatives except for* $\frac{\partial w_{t,ij}[\bar{y}_i^*, \bar{y}_j^*]}{\partial y_{i,t-1}} = \Phi_i$ *and* $\frac{\partial w_{t,ij}[\bar{y}_i^*, \bar{y}_j^*]}{\partial y_{j,t-1}} = \Phi_j$ *(parts (x) and (xvi)) are zero.*

Then, we represent $\Psi_{t,ij} = \Psi_{t,ij}\left(O\left((\max_{k,t}|y_{kt} - \bar{y}_k^\circ|)^3\right), \left\{\frac{\partial^k w_{t,ij}}{\partial y_{i,t-1}^{k_1} \partial y_{j,t-1}^{k_2}}|_{\bar{\omega}^*}\right\}_{k=1,2,3}\right)$ and the LQ approximation error would be shown by

$$\begin{aligned} & u_i(Y_{nt}, Y_{n,t-1}, \eta_{it}) - u_i^e(Y_{nt}, Y_{n,t-1}, \eta_{it}) \\ = {} & \lambda \sum_{j, d_{ij} \leq d_c} \Psi_{t,ij}\left(O\left(\left(\max_{k,t}|y_{kt} - \bar{y}_k^\circ|\right)^3\right), \left\{\frac{\partial^k w_{t,ij}}{\partial y_{i,t-1}^{k_1} \partial y_{j,t-1}^{k_2}}|_{\bar{\omega}^*}\right\}_{k=1,2,3}\right). \end{aligned} \tag{B.4}$$

The magnitude of parameter λ, i.e., $|\lambda|$, linearly controls the approximation error. When $|\lambda|$ is closed to zero, one can expect a small approximation error (Condition B.11 (iv)). In case of Example B.12, the approximation error takes a simpler form,

$$\lambda \sum_{j,d_{ij} \leq d_c} \Psi_{t,ij} \left(O \left(\left(\max_{k,t} |y_{kt} - \bar{y}_k^\circ| \right)^3 \right), \Phi_i, \Phi_j \right).$$

Last, consider evaluating an upper bound of $V_i (Y_{n,t-1}, \eta_{nt}) - V_i^e (Y_{n,t-1}, \eta_{nt})$. For this, we remind the following features.

Remark B.13 *Let $(Y_{n,t-1}, \eta_{nt})$ be the common initial condition of the two dynamic programming. Note that the process η_{nt} exogenously evolves. Let $Y_{nt}^m = f^m (Y_{n,t-1}, \eta_{nt})$ and $Y_{nt}^e = f^e (Y_{n,t-1}, \eta_{nt})$ denote the vector of two Markov strategies. Consider two choice-specific lifetime payoffs by taking Y_{nt}^m and Y_{nt}^e:*[165]

$$\begin{aligned} \mathcal{V}_i^m \left(\{Y_{ns}^m\}_{s=t}^\infty ; Y_{n,t-1}, \eta_{nt} \right) &= u_{it} \left(Y_{nt}^m, Y_{n,t-1}, \eta_{it} \right) \\ &\quad + \sum_{s=1}^\infty \delta^s E_t \left(u_{i,t+s} \left(Y_{n,t+s}^m, Y_{n,t+s-1}^m, \eta_{i,t+s} \right) \right) \end{aligned}$$

and

$$\begin{aligned} \mathcal{V}_i^e \left(\{Y_{ns}^e\}_{s=t}^\infty ; Y_{n,t-1}, \eta_{nt} \right) &= u_{it}^e \left(Y_{nt}^e, Y_{n,t-1}, \eta_{it} \right) \\ &\quad + \sum_{s=1}^\infty \delta^s E_t \left(u_{i,t+s}^e \left(Y_{n,t+s}^e, Y_{n,t+s-1}^e, \eta_{i,t+s} \right) \right). \end{aligned}$$

We consider the difference of the two functions denoted by $\Delta_{\mathcal{V},i}$:

$$\begin{aligned} \Delta_{\mathcal{V},i} &= \mathcal{V}_i^m \left(\{Y_{ns}^m\}_{s=t}^\infty ; Y_{n,t-1}, \eta_{nt} \right) - \mathcal{V}_i^e \left(\{Y_{ns}^e\}_{s=t}^\infty ; Y_{n,t-1}, \eta_{nt} \right) \\ &= \left[u_i \left(Y_{nt}^m, Y_{n,t-1}, \eta_{it} \right) - u_i^e \left(Y_{nt}^e, Y_{n,t-1}, \eta_{it} \right) \right] \\ &\quad + \sum_{s=1}^\infty \delta^s E_t \left(\begin{array}{c} u_i \left(Y_{n,t+s}^m, Y_{n,t+s-1}^m, \eta_{i,t+s} \right) \\ -u_i^e \left(Y_{n,t+s}^e, Y_{n,t+s-1}^e, \eta_{i,t+s} \right) \end{array} \right). \end{aligned}$$

[165] For the uncertain future exogenous characteristics $\eta_{n,t+1}, \eta_{n,t+2}, \cdots$, the conditional expectation $E_t (\cdot)$ is formed based on η_{nt}. Hence, the two lifetime payoffs share the same conditional expectation.

Hence, the issue of characterizing $\Delta_{\mathcal{V},i}$ is the difference of the per period payoff functions, $u_i(\cdot)$ and $u_i^e(\cdot)$ under the difference choices ($Y_{n,t+s}^m$ and $Y_{n,t+s}^e$) as well as initial conditions ($Y_{n,t+s-1}^m$ and $Y_{n,t+s-1}^e$) at time $t+s$. Note that the difference at the initial period, $u_i(Y_{nt}^m, Y_{n,t-1}, \eta_{it}) - u_i^e(Y_{nt}^e, Y_{n,t-1}, \eta_{it})$, is a special case (same initial conditions with different choices). Then, by using (B.4),

$$\begin{aligned}
& u_i\left(Y_{nt}^m, Y_{n,t-1}^m, \eta_{it}\right) - u_i^e\left(Y_{nt}^e, Y_{n,t-1}^e, \eta_{it}\right) \\
= \; & u_i^e\left(Y_{nt}^m, Y_{n,t-1}^m, \eta_{it}\right) - u_i^e\left(Y_{nt}^e, Y_{n,t-1}^e, \eta_{it}\right) \\
& + \lambda \sum_{j, d_{ij} \le d_c} \Psi_{t,ij}\left(O\left(\left(\max_{k,t} |y_{kt} - \bar{y}_k^\circ|\right)^3\right), \left\{\frac{\partial^k w_{t,ij}}{\partial y_{i,t-1}^{k_1} \partial y_{j,t-1}^{k_2}} |_{\bar{\omega}^*}\right\}_{k=1,2,3}\right) \\
= \; & \eta_{it}\left(y_{it}^m - y_{it}^e\right) + \gamma\left(y_{i,t-1}^m y_{it}^m - y_{i,t-1}^e y_{it}^e\right) - \frac{1}{2}\left(\left(y_{it}^m\right)^2 - \left(y_{it}^e\right)^2\right) \\
& - \frac{\gamma}{2}\left(\left(y_{i,t-1}^m\right)^2 - \left(y_{i,t-1}^e\right)^2\right) \\
& + \lambda \sum_{j, d_{ij} \le d_c} \left[w_{t,ij}^{LQ}\left[y_{i,t-1}^m, y_{j,t-1}^m, y_{it}^m, y_{jt}^m\right] - w_{t,ij}^{LQ}\left[y_{i,t-1}^e, y_{j,t-1}^e, y_{it}^e, y_{jt}^e\right]\right] \\
& + \lambda \sum_{j, d_{ij} \le d_c} \Psi_{t,ij}\left(O\left(\left(\max_{i,t} |y_{kt} - \bar{y}_k^\circ|\right)^3\right), \left\{\frac{\partial^k w_{t,ij}}{\partial y_{i,t-1}^{k_1} \partial y_{j,t-1}^{k_2}} |_{\bar{\omega}^*}\right\}_{k=1,2,3}\right).
\end{aligned}$$

Hence, $u_i\left(Y_{nt}^m, Y_{n,t-1}^m, \eta_{it}\right) - u_i^e\left(Y_{nt}^e, Y_{n,t-1}^e, \eta_{it}\right)$ consists of (i) the LQ function of $y_{ls}^m - y_{ls}^e$ for $l = i, j$ and $s = t-1, t$, and (ii) the approximation error of the per period payoff. An upper bound of $y_{ls}^m - y_{ls}^e$ can be characterized by

$$|y_{ls}^m - y_{ls}^e| = |(y_{ls}^m - \bar{y}_l) - (y_{ls}^e - \bar{y}_l)| \le 2 \max_{l=1,\cdots,n} \max_{s=t-1,t} \max\left\{|y_{ls}^m - \bar{y}_l^\circ|, |y_{ls}^e - \bar{y}_l^\circ|\right\}.$$

Observe that the first component of the difference, $\eta_{it}(y_{it}^m - y_{it}^e)$, involves the i's exogenous characteristic. Either $|y_{ls}^m - y_{ls}^e|$ or $|\eta_{it}|$ is large, the LQ approximation would be worse. This is the re-justification of importance of Condition B.11 (i). Hence, $\Delta_{\mathcal{V},i}$ can be characterized by

$$\Delta_{\mathcal{V},i} = \Delta_{\mathcal{V},i}\left(\tilde{\omega}^{\max}, \left(\tilde{\omega}^{\max}\right)^2, \left(\tilde{\omega}^{\max}\right)^3, \left\{\frac{\partial^k w_{t,ij}}{\partial y_{i,t-1}^{k_1} \partial y_{j,t-1}^{k_2}} |_{\bar{\omega}^*}\right\}_{k=1,2,3}, |\eta_{it}|, |\lambda|, d_c\right)$$

where $\tilde{\omega}^{\max} = \max_{k,t} |y_{it} - \bar{y}_i^{\circ}|$.

Approximated Euler equation system $\mathcal{J}^e(Y_{n,t-1}, \eta_{nt}, Y_{nt})$

Now we study errors of using $\mathcal{J}^e(Y_{n,t-1}, \eta_{nt}, Y_{nt})$ at the true parameter values. We reproduce the Euler equation system and relevant assumptions for completeness. For agent i at time t,

$$\begin{aligned} 0 \;=\;& \eta_{it} + \gamma_0 y_{i,t-1} + \lambda_0 w_{t,i.}\left[Y_{n,t-1}\right] Y_{nt}^* - (1+\delta\gamma_0)\, y_{it}^* \\ & + \delta E_t\left[\left(\gamma_0 + \lambda_0 \frac{\partial w_{t+1,i.}\left[Y_{nt}^*\right]}{\partial y_{it}} Y_{n,t+1}^*\right) y_{i,t+1}^*\right] \\ & + \delta E_t \sum_{j\neq i}^{n} \frac{\partial y_{j,t+1}}{\partial y_{it}} \lambda_0 w_{t+1,ij}\left[Y_{nt}^*\right] y_{i,t+1}^* \\ & + \sum_{k=1}^{\infty} \delta^{k+1} E_t \left(\begin{array}{l} \lambda_0 \sum_{j_1,\cdots,j_k\neq i}^{n} \frac{\partial y_{j_1,t+1}}{\partial y_{it}} \cdots \frac{\partial y_{j_k,t+k}}{\partial y_{j_{k-1},t+k-1}} \frac{\partial w_{t+k+1,i.}\left[Y_{n,t+k}^*\right]}{\partial y_{j_k,t+k}} Y_{n,t+k+1}^* y_{i,t+k+1}^* \\ + \lambda_0 \sum_{j_1,\cdots j_{k+1}\neq i}^{n} \frac{\partial y_{j_1,t+1}}{\partial y_{it}} \cdots \frac{\partial y_{j_{k+1},t+k+1}}{\partial y_{j_k,t+k}} w_{t+k+1,i,j_{k+1}}\left[Y_{n,t+k}^*\right] y_{i,t+k+1}^* \end{array}\right). \end{aligned} \tag{B.5}$$

And, the approximated Euler equations are

$$E\left[\begin{array}{c} \left[S_n(W_{nt}) Y_{nt} - \gamma_0 Y_{n,t-1} - X_{nt}\beta_0\right] + \delta\gamma_0\left(Y_{nt} - Y_{n,t+1}^e\right) \\ -\delta\left\{\left[\sum_{p=1}^{P} \lambda_0 \psi_{p,0} M_{n,t+1,p} + \lambda_0 N_{n,t+1}\right] Y_{n,t+1}^e\right\} \\ -\delta^2 \nabla V_{n,t+2}^e - \mathbf{c}_{n0} - \alpha_{t,0} l_n \end{array} |\ell_{t-1}\right] = \mathbf{0}_{n\times 1} \tag{B.6}$$

for $t = 1, \cdots, T$

Assumption B.13 *(i) At θ_0, $\mathbf{c}_{n0}$, and $\alpha_{t,0}$, equation (B.6) holds.*

(ii) $\sup_{n,t} \rho_{\max}\left(\frac{\partial f^e(Y_{n,t-1},\eta_{nt})}{\partial Y_{n,t-1}'}\right) \leq \tau < 1$ *for some* $0 \leq \tau < 1$ *and* $\sup_{n,t} \rho_{\max}\left(\frac{\partial f^e(Y_{n,t-1},\eta_{nt})}{\partial \eta_{nt}'}\right) \leq c_\eta$ *for some* $c_\eta < \infty$ *a.e where* $\rho_{\max}(A_n)$ *stands for the spectral radius of* A_n*.*

First of all, the Euler equation (B.5) involves the infinite sum, so summability helps to obtain a non explosive Euler equation system. For this, it suffices to consider controlling $\sum_{j_1,\cdots,j_k\neq i}^{n} \frac{\partial y_{j_1,t+1}}{\partial y_{it}} \cdots \frac{\partial y_{j_k,t+k}}{\partial y_{j_{k-1},t+k-1}}$ since this summation increases $n-1$ additional components when a time horizon increases by one unit. Under Assumption

B.13 (ii),

$$
\begin{aligned}
& \sum_{j_1,\cdots,j_k \neq i} \frac{\partial y_{j_1,t+1}}{\partial y_{it}} \cdots \frac{\partial y_{j_k,t+k}}{\partial y_{j_{k-1},t+k-1}} \\
\leq \; & \sum_{j_1 \neq i} \cdots \sum_{j_k \neq i} \left| \frac{\partial y_{j_1,t+1}}{\partial y_{it}} \cdots \frac{\partial y_{j_k,t+k}}{\partial y_{j_{k-1},t+k-1}} \right| \\
\leq \; & \sum_{j_1 \neq i} \cdots \sum_{j_{k-1} \neq i} \left| \frac{\partial y_{j_1,t+1}}{\partial y_{it}} \cdots \frac{\partial y_{j_{k-1},t+k-1}}{\partial y_{j_{k-2},t+k-2}} \right| \sum_{j_k \neq i} \left| \frac{\partial y_{j_k,t+k}}{\partial y_{j_{k-1},t+k-1}} \right| \\
\leq \; & \tau^k \to 0
\end{aligned}
$$

as $k \to \infty$. A key is to control the $n-1$ additional components $\sum_{j_k \neq i} \left| \frac{\partial y_{j_k,t+k}}{\partial y_{j_{k-1},t+k-1}} \right|$ by the uniformly bounded constant, $0 < \tau < 1$. Hence, Assumption B.13 (ii) is a device of limiting the marginal change of y_{it} on the remote future marginal payoffs.

Now we consider

$$
\mathcal{J}^e \left(Y_{n,t-1}, \eta_{nt}, Y_{nt} \right) - \mathcal{J} \left(Y_{n,t-1}, \eta_{nt}, Y_{nt} \right)
$$

since $\mathcal{J} \left(Y_{n,t-1}, \eta_{nt}, Y_{nt} \right) = \mathbf{0}_{n \times 1}$. Then, the i^{th}-component of Euler equation error (denoted by ϱ_i) is

$$
\begin{aligned}
\varrho_i \;=\; & \lambda_0 \delta E_t \left\{ y_{i,t+1} \sum_{j \neq i} \left(\left[\Delta_n^* \right]_{ij} - \left[\Delta_{n,t+1}^* \right]_{ij} \right) w_{t+1,ij} \right\} \\
& + \delta^2 E_t \left\{ \sum_{j \neq i} \left(\begin{array}{c} \left(\left[\Delta_n^* \right]_{ij} - \left[\Delta_{n,t+1}^* \right]_{ij} \right) \frac{\partial V_i^e(Y_{n,t+1},\eta_{n,t+2})}{\partial y_{i,t+1}} \\ + \left[\Delta_{n,t+1}^* \right]_{ij} \left(\frac{\partial V_i^e(Y_{n,t+1},\eta_{n,t+2})}{\partial y_{i,t+1}} - \frac{\partial V_i(Y_{n,t+1},\eta_{n,t+2})}{\partial y_{i,t+1}} \right) \end{array} \right) \right\}.
\end{aligned} \tag{B.7}
$$

First, if $|\lambda_0|$ or δ is large, so is the magnitude of ϱ_i. Second, the level $y_{i,t+1}$ has an impact on ϱ_i. Third, if $W_{n,t+1}$ contains many nonzero elements or some signals from i to j at time $t+1$ are strong (large magnitude of $w_{t+1,ij}$), the magnitude of ϱ_i might be amplified. Fourth, if the approximating error of the value function ($\Delta_{\mathcal{V},i}$) is large, ϱ_i would be also large. Last, consider the components $\left[\Delta_{n,t+1}^* \right]_{ij}$ and $\left[\Delta_n^* \right]_{ij} - \left[\Delta_{n,t+1}^* \right]_{ij}$ for $j \neq i$. By Assumption B.13 (ii), $\left\| \Delta_{n,t+1}^* \right\| \leq \tau < 1$. If the column sum norm is take for $\|\cdot\|$, it implies that $\sum_{j=1}^{n}\left|\frac{\partial y_{j,t+1}}{\partial y_{it}}\right| \leq \tau$. To deliver an intuition, suppose that (i) $|\lambda_0|$ is close to zero, (ii) W_{nt}'s are sparse (most of elements are zero), and (iii) elements of W_{nt}'s lie between 0 and 1. Then, due to the adjustment cost parameter $0 < \gamma_0 < 1$, the own dynamic influence $\frac{\partial y_{i,t+1}}{\partial y_{it}}$ might dominate other components $\frac{\partial y_{j,t+1}}{\partial y_{it}}$, $j \neq i$ (i.e., $\Delta_{n,t+1}^{*}$ can be a diagonally dominant matrix[166]). By the envelope theorem, $\frac{\partial y_{i,t+1}}{\partial y_{it}}$ is eliminated from the Euler equation. It implies that $\sum_{j\neq i}^{n}\left|\frac{\partial y_{j,t+1}}{\partial y_{it}}\right|$ is much smaller than τ. Note that the difference $\Delta_n^* - \Delta_{n,t+1}^*$ is $\frac{\partial f(Y_{n,t-1},\eta_{nt})}{\partial Y_{n,t-1}'}|_{\left(\bar{Y}_n^{\circ},\bar{\eta}_n^{\circ}\right)} - \frac{\partial f(Y_{n,t-1},\eta_{nt})}{\partial Y_{n,t-1}'}|_{\left(Y_{n,t-1}^{+},\eta_{nt}^{+}\right)}$. Since $\left(Y_{n,t-1}^{+},\eta_{nt}^{+}\right)$ lies between $(Y_{n,t-1},\eta_{nt})$ and $\left(\bar{Y}_n^{\circ},\bar{\eta}_n^{\circ}\right)$ and $f(\cdot)$ is a C^{∞}-function, the difference $\Delta_n^* - \Delta_{n,t+1}^*$ will be small. Moreover, by the envelope theorem, the (i,i)-components of $\Delta_n^* - \Delta_{n,t+1}^*$ is not considered in the Euler equation.

B.2. Statistical appendix

Recall that $\varsigma_{l(i,t),L} = \varsigma\left(\epsilon_{l(i,t),L}, Z_L^*, X_L, \mathbf{c}_{n0}, \alpha_{T,0}\right)$ for each $l(i,t) \in D$, $\varsigma_{nt} = \left(\varsigma_{l(1,t),L}, \cdots, \varsigma_{l(n,t),L}\right)'$ for each t and $\varsigma_L = (\varsigma_{n1}', \cdots, \varsigma_{nT}')'$. Observe that the baseline random field ς for approximation contains (i) the strictly exogenous variables and (ii) i.i.d. errors. Similarly, we define ς_{nt}^* and ς_L^* from $\varsigma_{l(i,t),L}^* = \varsigma^*\left(z_{l(i,t),L}^*, X_L, \mathbf{c}_{n0}, \alpha_{T,0}\right)$.

Some lemmas and propositions

In this part, we introduce some basic properties of spatial-time NED on ς. Some properties verified by Davidson (1994), Jenish and Prucha (2012), Qu and Lee (2015), and Qu et al. (2017) will be reproduced for completeness of the paper. Our purpose is to get approximability of $y_{l(i,t),L}$ based on neighboring input processes $\varsigma_{l(j,t'),L}$'s such that $\|l(i,t) - l(j,t')\|_{\infty} \leq s$ for some $s > 0$. By Assumption 3.4.6 (i), we firstly guarantee for time series stability of Y_{nt}. Hence, Y_{nt} can be approximated by ς_{nt},

[166]By the definition, it says that $\left|\frac{\partial y_{j,t+1}}{\partial y_{it}}\right| \geq \sum_{j\neq i}^{n}\left|\frac{\partial y_{j,t+1}}{\partial y_{it}}\right|$.

$\varsigma_{n,t-1}, \cdots, \varsigma_{n,t-s}$ for some $s \in \mathbf{Z}_{+}$. For this, define $\mathcal{F}_t(s) = \sigma\left(\{\varsigma_{n,t'}\}_{t'=t-s}^{t}\right)$ for each t. The following Proposition is a useful result describing that the conditional mean (denoted by $E(y|\mathcal{F}_0)$) is the optimal predictor of a random variable y given partial information $\mathcal{F}_0$.

Proposition B.2.1 *(i)* $\|y - E(y|\mathcal{F}_0)\|_{L_2} \leq \|y - \hat{y}\|_{L_2}$ *where* $\hat{y}$ *is any* $\mathcal{F}_0-$*measurable approximation to* y.

(ii) For $p \geq 1$, $\|y - E(y|\mathcal{F}_2)\|_{L_p} \leq 2\|y - E(y|\mathcal{F}_1)\|_{L_p}$ *where* y *is* $\mathcal{F}_0-$*measurable and* $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_0$.

Proof of Proposition B.2.1. See Theorems 10. 12 and 10. 28 in Davidson (1994). Q.E.D.

Note that for each n Y_{nt} is a multivariate time series. Thus, we have the representation, $Y_{nt} = f^e(\varsigma_{nt}, Y_{n,t-1})$ where $f^e(\cdot)$ is a contraction mapping of $Y_{n,t-1}$ by Assumption B.13 (ii). The next Proposition is for the time NED property of Y_{nt}.

Proposition B.2.2 *For each* n, $\{Y_{nt}\}$ *is time* L_2*-NED on* $\{\varsigma_{nt}\}$. *That is,*

$$\|Y_{nt} - E(Y_{nt}|\mathcal{F}_t(s))\|_{L_2} \leq d_{nt}\tilde{\psi}(s)$$

where $\tilde{\psi}(s) = \sum_{l=s+1}^{\infty} \tau^{l-1}$ *and* d_{nt} *is set to be* $\sup_t \left\|\frac{\partial f^e(\varsigma_{nt}, Y_{n,t-1})}{\partial \varsigma_{nt}'}\varsigma_{nt}\right\|_{L_2} \leq c_\eta \sup_t \|\varsigma_{nt}\|_{L_2}$.

Proof of Proposition B.2.2. This property is directly implied by the model's assumption (time stability). Q.E.D.

Note that a key is that $f^e(\cdot)$ is a contraction mapping of $Y_{n,t-1}$ (by Assumption B.13 (ii)), i.e., $\sup_{n,t} \rho_{\max}\left(\frac{\partial f^e(\varsigma_{nt}, Y_{n,t-1})}{\partial Y_{n,t-1}'}\right) \leq \tau < 1$. Hence, Y_{nt} can be approximated by $\varsigma_{nt}, \varsigma_{n,t-1}, \cdots, \varsigma_{n,t-s}$ for some $s \in \mathbf{N}$ and the effects of remote past ones $(\varsigma_{n,t-s-1}, \varsigma_{n,t-s-2}, \cdots)$ diminish when s increases since $\tilde{\psi}(s) = \sum_{l=s+1}^{\infty} \tau^{l-1} \rightarrow 0$ as

$s \to \infty$. This result has the same implication of Wold decomposition theorem: any zero-mean weakly stationary process can have an (infinite) invertible moving average representation. For each $i = 1, \cdots, n$, hence, we can apply the LLN by Theorem 19.13 in Davidson (1994), i.e., $\bar{y}_{i,T} - y_i^{\circ} \to_p 0$ as $T \to \infty$.[167]

Note that $\|l(i,t) - l(j,t')\|_{\infty} \leq s$ for some $s > 0$ is equivalent that $|t - t'| \leq s$ and $\|l(i,0) - l(j,0)\|_{\infty} \leq s$ for some $s > 0$. In the spatial dimension, there might also exist some spatial units $j = j_1, \cdots, j_M$ satisfying $\|l(i,0) - l(j,0)\|_{\infty} \leq s$, which approximate $y_{l(i,t),L}$ well. That is, at most $(M+1)(s+1)$-spatial-time processes $\varsigma_{l(i,t),L}, \varsigma_{l(j_1,t),L}, \cdots, \varsigma_{l(j_M,t),L}, \cdots, \varsigma_{l(i,t-s),L}, \varsigma_{l(j_1,t-s),L}, \cdots, \varsigma_{l(j_M,t-s),L}$ may approximate $y_{l(i,t),L}$ and $(M+1)(s+1)$ is much smaller than L. Thus, we need to verify whether $y_{l(i,t),L}$ is a spatial-NED process given $Y_{n,t-1}$.

Hence, we consider only the spatial dimension (a countable subset of $\mathbf{R}^d$). First, consider the basic topological structure for the space for spatial units, which is introduced by Claim B.1 in Qu and Lee (2015).

Proposition B.2.3 *For any spatial unit i and distance $s \geq 1$, let $B_i(s) = \{j : \|l(i,0) - l(j,0)\|_{\infty} \leq s\}$. Then, there exist at most $c_5 s^d$ spatial units in $B_i(s)$ where $c_5 > 0$.*

Proof of Proposition B.2.3. See Lemma A.1 (ii) in Jenish and Prucha (2009). Q.E.D.

Proposition B.2.3 says that the maximum number of spatial units around i is specified by the distance s.[168] By Assumption 3.2.1, there is a finite threshold d_c

[167] Then, we can justify the LQ perturbation method using the feasible time averages $(\bar{Y}_{n,T}, \bar{\eta}_{n,T})$ instead of $(\bar{Y}_n^{\circ}, \bar{\eta}_n^{\circ})$.

[168] i.e., the number of spatial units of D within radius $s \geq 1$ centered at $i \in \mathbf{R}^d$ is of $O(s^d)$. Note that this order $O(s^d)$ does not depend on specific i.

characterizing a physical neighbor. Then, $1 \leq \inf_{i,n} d_c(i) \leq \sup_{i,n} d_c(i) \leq c_5 d_c^d < \infty$, which means that every spatial unit has at least one neighbor and at most uniformly (in n) bounded number of neighbors. Then, we derive the following proposition.

Proposition B.2.4 *For any $0 < d_c < \infty$ and $a > 0$,*

$$\lim_{s\to\infty} \sup_{L,l(i,t)\in D} \sum_{l(j,t)\in D, \|l(i,0)-l(j,0)\|_\infty > s} \mathbf{1}\left\{j : \|l(i,0) - l(j,0)\|_\infty \leq ad_c\right\} = 0. \quad \text{(B.8)}$$

Proof of Proposition B.2.4. If $s \leq ad_c$, the number of spatial units satisfying $\|l(i,0) - l(j,0)\|_\infty > s$ and $\|l(i,0) - l(j,0)\|_\infty \leq ad_c$ is at most $c_5 (ad_c)^d$ units by Proposition B.2.3. However, if $s > ad_c$, there is no spatial unit satisfying both $\|l(i,0) - l(j,0)\|_\infty > s$ and $\|l(i,0) - l(j,0)\|_\infty \leq ad_c$. Hence, the quantity

$$\sup_{L,l(i,t)\in D} \sum_{l(j,t)\in D, \|l(i,0)-l(j,0)\|_\infty > s} \mathbf{1}\left\{j : \|l(i,0) - l(j,0)\|_\infty \leq ad_c\right\} = 0$$

if $s > ad_c$ implying (B.8). Q.E.D.

The sum $\sum_{l(j,t)\in D} \mathbf{1}\left\{j : \|l(i,0) - l(j,0)\|_\infty \leq ad_c\right\}$ means the number of neighboring spatial units relevant to the a^{th}-order ($a \in \mathbf{Z}_+$) spatial effects when the unit i is centered.[169] When s becomes large, the a^{th}-order spatial effect will disappear. For $s > ad_c$ where sufficiently large a, and for each $l(i,t) \in D$

$$\begin{aligned} & \left\{l(j,t') \in D : \|l(i,t) - l(j,t')\|_\infty > s\right\} \quad \text{(B.9)} \\ = \ & \left\{l(j,t') : \|l(i,0) - l(j,0)\|_\infty > s, \neq i, |t - t'| \leq s\right\} \\ & \cup \left\{l(j,t') : \|l(i,0) - l(j,0)\|_\infty > s, j \neq i, |t - t'| > s\right\} \\ & \cup \left\{l(j,t') : \|l(i,0) - l(j,0)\|_\infty \leq s, |t - t'| > s\right\}. \end{aligned}$$

[169]Note that $d_c > 0$ controls the direct (first-order) spatial effects. If $2d_c$ is considered, the sum $\sum_{l(j,t)\in D} \mathbf{1}\left\{j : \|l(i,0) - l(j,0)\|_\infty \leq 2d_c\right\}$ means the number of spatial units affected by i's neighbors' neighbors.

By Assumption B.13 (ii), $Y_{nt} = f^e(\varsigma_{nt}, Y_{n,t-1})$ is a contraction mapping of $Y_{n,t-1}$ and a Lipschitz continuous function of ς_{nt}. For the last two cases (a time horizon of two units is far) in (B.9), the dependence between two units is controlled by the weak time dependence (τ). Hence, we focus on the first part of (B.9): a set of spatial units which are far from unit i but near epoch to t. Now the spatial-time NED property of $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ will be satisfied by random fields generated from nonlinear Lipschitz type functionals of spatial processes $\left\{\varsigma_{l(i,t),L}\right\}_{i=1}^{n}$ (at time t).[170] By considering the Lipschitz condition, we want to obtain that small changes in $\left\{\varsigma_{l(i,t),L}\right\}_{i=1}^{n}$ lead to small changes in $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ given $Y_{n,t-1}$. Pick $l(i,t) \in D_L$ and consider $y_{l(i,t),L} = f^e(\varsigma_{nt}, Y_{n,t-1})$ and $y'_{l(i,t),L} = f^e(\varsigma'_{nt}, Y_{n,t-1})$ given the same $Y_{n,t-1}$. Then,

$$\left|y_{l(i,t),L} - y'_{l(i,t),L}\right| \leq \sum_{j=1}^{n} \left|\left[\frac{\partial f^e(\varsigma_{nt}, Y_{n,t-1})}{\partial \varsigma'_{nt}}\right]_{ij}\right| \left|\varsigma_{l(j,t),L} - \varsigma'_{l(j,t),L}\right|.$$

Note that an amount of $\sum_{j=1}^{n} \left|\left[\frac{\partial f(\varsigma_{nt}, Y_{n,t-1})}{\partial \varsigma'_{nt}}\right]_{ij}\right|$ is bounded by c_η by Assumption B.13 (ii).

The next step is to characterize $\left|\left[\frac{\partial f(\varsigma_{nt}, Y_{n,t-1})}{\partial \varsigma'_{nt}}\right]_{ij}\right|$ as a function of $\|l(i,0) - l(j,0)\|_\infty$, d_c and the parameter λ. The purpose of this is to employ Proposition 1 in Jenish and Prucha (2012), which is a tool to justify a weakly dependent random field. For this, define $\bar{W}_n = [\bar{w}_{n,ij}]$ where $\bar{w}_{n,ij} = \sup_t w_{t,ij}$, and note that the column and row sums of $\bar{W}_n$ are uniformly bounded in n by Assumption 3.4.5 (ii). Due to the existence of $d_c > 0$, we also have $\bar{w}_{n,ij} = 0$ if $\|l(i,0) - l(j,0)\|_\infty > d_c$. By Assumption 3.4.6 (iii), $\left\|\bar{\lambda}\bar{W}_n\right\| < 1$ where $\bar{\lambda} = \sup_{\lambda\in\Theta_1} |\lambda| > 0$. Then, $I_n - \bar{\lambda}\bar{W}_n$ is invertible. For $l(i,t)$, $l(j,t') \in D$, we specify the maximum cumulative spatial-time effects between units

[170] That is, we require that $y_{l(i,t),L}$ is a spatial-NED on $\left\{\varsigma_{l(j,t),L}\right\}_{j=1}^{n}$ given $Y_{n,t-1}$.

$l(i,t)$ and $l(j,t')$:

$$w_{l(i,t),l(j,t')} = \begin{cases} \left[\left(I_n - \bar{\lambda}\bar{W}_n\right)^{-1}\right]_{ij} if, t = t' \\ \tau^{|t-t'|} if, i = jand, t \neq t' \\ \tau^{|t-t'|}\left[\left(I_n - \bar{\lambda}\bar{W}_n\right)^{-1}\right]_{ij} if, i \neq j, andt \neq t' \end{cases}. \tag{B.10}$$

If $i \neq j$, the intensity of spatial interactions between $y_{l(i,t),L}$ and $y_{l(j,t),L}$ given $Y_{n,t-1}$ is controlled by a Neumann series expansion based on $\bar{W}_n$, $\bar{W}_n^2$, $\cdots$. If $i \neq j$, the spatial effects are bounded by the potentially maximum spatial influences of the linear SAR model given $Y_{n,t-1}$. If $i \neq j$ with $t \neq t'$, the spatial-time interaction between $l(i,t)$ and $l(j,t')$ would be weaker than the case of $i \neq j$ with $t = t'$ since the spatial-time process $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ is a stable autoregressive time-series process. If $i = j$ with $t \neq t'$, the interaction between two units $l(i,t)$ and $l(i,t')$ (i's own dynamic effect) is controlled by the maximum time influence $\tau^{|t-t'|}$ from Assumption 3.4.6 (i).

Then, $w_{l(i,t),l(j,t')} \geq 0$ is well specified for any $l(i,t)$ and $l(j,t')$ in D. Note that for sufficiently large $a \geq 2$ with $a \in \mathbf{Z}_+$ and $i \neq j$ $\left[I_n + \sum_{k=1}^{a} \bar{\lambda}^k \bar{W}_n^k\right]_{ij} = \bar{\lambda}\bar{w}_{n,ij} + \sum_{k=2}^{a} \bar{\lambda}^k \sum_{m_1=1}^{n} \cdots \sum_{m_{k-1}=1}^{n} \bar{w}_{n,i,m_1} \cdots \bar{w}_{n,m_{k-1},j}$. Note that $\left[\left(I_n - \bar{\lambda}\bar{W}_n\right)^{-1}\right]_{ij} = \lim_{a\to\infty}\left[I_n + \sum_{k=1}^{a} \bar{\lambda}^k \bar{W}_n^k\right]_{ij}$ since $\left\|\bar{\lambda}\bar{W}_n\right\| < 1$. By Qu and Lee (2015), $\left[\bar{W}_n^a\right]_{ij} = 0$ if $\|l(i,0) - l(j,0)\|_\infty > ad_c$. If $\|l(i,0) - l(j,0)\|_\infty > s$ for some large $s > 0$,

$$\begin{aligned} \left[\left(I_n - \bar{\lambda}\bar{W}_n\right)^{-1}\right]_{ij} &= \bar{\lambda}\bar{w}_{n,ij} + \sum_{k=2}^{\left[\frac{s}{d_c}\right]} \bar{\lambda}^k \sum_{m_1=1}^{n} \cdots \sum_{m_{k-1}=1}^{n} \bar{w}_{n,i,m_1} \cdots \bar{w}_{n,m_{k-1},j} \\ &\quad + \sum_{k=\left[\frac{s}{d_c}\right]+1}^{\infty} \bar{\lambda}^k \sum_{m_1=1}^{n} \cdots \sum_{m_{k-1}=1}^{n} \bar{w}_{n,i,m_1} \cdots \bar{w}_{n,m_{k-1},j} \end{aligned} \tag{B.11}$$

$\to 0$ as $s \to \infty$ where $\left[\frac{s}{d_c}\right]$ is the biggest integer that is less or equal than $\frac{s}{d_c}$. Then, for large $s > 0$

$$\sup_{L,l(i,t)\in D} \sum_{l(j,t')\in D, \|l(i,t)-l(j,t')\|_\infty > s} w_{l(i,t),l(j,t')} \tag{B.12}$$

$$
\begin{aligned}
&\leq \sup_{L,l(i,t)\in D} \sum_{l(j,t')\in D, \|l(i,0)-l(j,0)\|_\infty > s, |t-t'|\leq s} w_{l(i,t),l(j,t')} \\
&\quad + \sup_{L,l(i,t)\in D} \sum_{l(j,t')\in D, \|l(i,0)-l(j,0)\|_\infty > s, |t-t'|> s} w_{l(i,t),l(j,t')} \\
&\quad + \sup_{L,l(i,t)\in D} \sum_{l(j,t')\in D, \|l(i,0)-l(j,0)\|_\infty \leq s, |t-t'|> s} w_{l(i,t),l(j,t')} \\
&\leq \sum_{k=\left[\frac{s}{d_c}\right]+1}^{\infty} \bar{\lambda}^k \sum_{m_1=1}^{n} \cdots \sum_{m_{k-1}=1}^{n} \bar{w}_{n,i,m_1} \bar{w}_{n,m_1,m_2} \cdots \bar{w}_{n,m_{k-1},j} \\
&\quad +2 \sup_{L,l(i,t)\in D} \sum_{l(j,t')\in D, |t-t'|>s, j=i} \tau^s
\end{aligned}
$$

$\to 0$ as $s \to \infty$ since $0 < \tau < 1$. The first inequality is due to set relation (B.9). For the second inequality, we use (B.11) and $\tau^{|t-t'|} < \tau^s$ if $|t-t'| > s$.

Using $w_{l(i,t),l(j,t')}$, we have the nonlinear infinite moving average representation[171]

$$
y_{l(i,t),L} = \tilde{f}_{l(i,t),L}\left(\left\{\varsigma_{l(j,t'),L} : l\left(j,t'\right) \in D\right\}\right),
$$

i.e., $y_{l(i,t),L}$ is a nonlinear functional of the random field $\varsigma = \left\{\varsigma_{l(i,t),L} : l\left(i,t\right) \in D\right\}$. That is, $\tilde{f}_{l(i,t),L} : \mathbf{E}^D \to \mathbf{R}$ and $\mathbf{E} \subseteq \mathbf{R}$. Proposition B.2.5 below shows the random field $Y = \left\{y_{l(i,t),L} : l\left(i,t\right) \in D_L, L \geq 1\right\}$ is L_2-NED on ς. The proof of Proposition B.2.5 is a spatial-time extension of Proposition 1 in Jenish and Prucha (2012).

Proposition B.2.5 *Assume Assumptions 3.4.2 and 3.4.4 hold. If*

$$
\left|\tilde{f}_{l(i,t),L}\left(e\right) - \tilde{f}_{l(i,t),L}\left(e'\right)\right| \leq \sum_{l(j,t')\in D} w_{l(i,t),l(j,t')} \left|e_{l(j,t'),L} - e'_{l(j,t'),L}\right| \qquad \text{(B.13)}
$$

where e, $e' \in \mathbf{E}^D$, Y *is* L_2*-NED on* ς *with* $d_{l(i,t),L} = \sup_{L,l(i,t)\in D} \left\|\varsigma_{l(i,t),L}\right\|_{L_2}$ *and*

$$
\psi\left(s\right) = \sup_{L,l(i,t)\in D} \sum_{l(j,t')\in D, \|l(i,t)-l(j,t')\|_\infty > s} w_{l(i,t),l(j,t')}.
$$

Note that $\sup_{L,l(i,t)\in D_L} d_{l(i,t),L} < \infty$ *under the regularity conditions (Assumptions 3.4.2 and 3.4.4). Hence,* Y *is uniformly* L_2*-NED on* ς*.*

[171] As contrary to a spatial process, the domain of input processes should be D (instead of D_L) due to the existence of infinite time lags.

Proof of Proposition B.2.5. As the first step, we need to show $y_{l(i,t),L}$ is $L_{4+\eta_\epsilon}$-bounded. Similar to Proposition B.2.2, we consider the sequence

$$y_{l(i,t),L}^{(s)} = \tilde{f}_{l(i,t),L}\left(\left\{\varsigma_{l(j,t'),L}^{(s)} : l(j,t') \in D\right\}\right)$$

where $\varsigma_{l(j,t'),L}^{(s)} = \varsigma_{l(j,t'),L}$ if $\|l(i,t) - l(j,t')\|_\infty \leq s$; 0 if $\|l(i,t) - l(j,t')\|_\infty > s$. Let $\varsigma^{(s)} = \left\{\varsigma_{l(j,t'),L}^{(s)} : l(j,t') \in D\right\}$ be the random field for approximation. By condition (B.13), for any $s, r \in \mathbf{N}$

$$\begin{aligned}
\left|y_{l(i,t),L}^{(s+r)} - y_{l(i,t),L}^{(s)}\right| &= \left|\tilde{f}_{l(i,t),L}\left(\varsigma^{(s+r)}\right) - \tilde{f}_{l(i,t),L}\left(\varsigma^{(s)}\right)\right| \qquad \text{(B.14)} \\
&\leq \sum_{l(j,t')\in D, s<\|l(i,t)-l(j,t')\|_\infty \leq s+r} w_{l(i,t),l(j,t')}\left|\varsigma_{l(j,t'),L}^{(s)}\right|.
\end{aligned}$$

Hence, by the Minkowski's inequality, for $p = 4 + \eta_\epsilon$

$$\begin{aligned}
\left\|y_{l(i,t),L}^{(s+r)} - y_{l(i,t),L}^{(s)}\right\|_{L_{4+\eta_\epsilon}} &= \left\|\tilde{f}_{l(i,t),L}\left(\varsigma^{(s+r)}\right) - \tilde{f}_{l(i,t),L}\left(\varsigma^{(s)}\right)\right\|_{L_{4+\eta_\epsilon}} \\
&\leq \sup_{L,l(i,t)\in D}\left\|\varsigma_{l(i,t),L}\right\|_{L_{4+\eta_\epsilon}} \cdot \sum_{l(j,t')\in D, s<\|l(i,t)-l(j,t')\|_\infty \leq s+r} w_{l(i,t),l(j,t')} \\
&\leq \sup_{L,l(i,t)\in D}\left\|\varsigma_{l(i,t),L}\right\|_{L_{4+\eta_\epsilon}} \cdot \sum_{l(j,t')\in D, \|l(i,t)-l(j,t')\|_\infty > s} w_{l(i,t),l(j,t')} \to 0
\end{aligned}$$

as $s \to \infty$ since $\sup_{L,l(i,t)\in D}\left\|\varsigma_{l(i,t),L}\right\|_{L_{4+\eta_\epsilon}} < \infty$ by Assumptions 3.4.2 and 3.4.4, and $\sum_{l(j,t')\in D, \|l(i,t)-l(j,t')\|_\infty > s} w_{l(i,t),l(j,t')} \to 0$ as $s \to \infty$ by relation (B.10). Hence, $\left\{y_{l(i,t),L}^{(s)}\right\}_s$ is a Cauchy sequence in the $L_{4+\eta_\epsilon}-$space. Since $L_{4+\eta_\epsilon}-$space is complete, $\lim_{s\to\infty} y_{l(i,t),L}^{(s)} = y_{l(i,t),L}$ exists and that limit point belongs to the $L_{4+\eta_\epsilon}-$space.

Second, we show the NED property. By Proposition B.2.1 (i),

$$\begin{aligned}
&\left\|y_{l(i,t),L} - E\left(y_{l(i,t),L}|\mathcal{F}_{l(i,t),L}(s)\right)\right\|_{L_2} \\
\leq\ &\left\|\tilde{f}_{l(i,t),L}(\varsigma) - \tilde{f}_{l(i,t),L}\left(\varsigma^{(s)}\right)\right\|_{L_2} \\
\leq\ &\sup_{L,l(i,t)\in D}\left\|\varsigma_{l(i,t),L}\right\|_{L_2} \cdot \sum_{l(j,t')\in D, \|l(i,t)-l(j,t')\|_\infty > s} w_{l(i,t),l(j,t')}.
\end{aligned}$$

If we set $\psi(s) = \sum_{l(j,t')\in D, \|l(i,t)-l(j,t')\|_\infty > s} w_{l(i,t),l(j,t')}$ for each s, the NED property is shown by relation (B.10). Q.E.D.

Once we show that Y is NED on ς, we can have the LLN, i.e., $\frac{1}{L}\sum_{l(i,t)\in D_L}\left(y_{l(i,t),L} - E\left(y_{l(i,t),L}\right)\right) = o_p(1)$. To achieve the LLN, it suffices to have L_1-NED of Y on ς. Since Y is L_2-NED on ς, Y is also L_1-NED on ς, so the condition for LLN is achieved. The Proposition below says the LLN.

Proposition B.2.6 *Under the same assumptions in Proposition B.2.5,*

$$\lim_{L\to\infty}\left\|\frac{1}{L}\sum_{l(i,t)\in D_L}\left(y_{l(i,t),L} - E\left(y_{l(i,t),L}\right)\right)\right\|_{L_1} = 0.$$

Since L_1 convergence implies convergence in probability, we have $\frac{1}{L}\sum_{l(i,t)\in D_L}\left(y_{l(i,t),L} - E\left(y_{l(i,t),L}\right)\right) \to_p 0$ *as* $L\to\infty$.

Proof of Proposition B.2.6. First, note that $y_{l(i,t),L}$ is uniformly $L_{4+\eta_\epsilon}$-bounded since $y_{l(i,t),L}$ belongs to the $L_{4+\eta_\epsilon}$-space. This implies $\sup_{L,l(i,t)\in D_L} E\left|y_{l(i,t),L}\right|^4 < \infty$. For each given $s > 0$, define $y^{(s)}_{l(i,t),L} = E\left(y_{l(i,t),L}|\mathcal{F}_{l(i,t),L}(s)\right)$. Second, consider L_4-boundedness of $y^{(s)}_{l(i,t),L}$. By the (conditional) Lyapunov and Jensen's inequalities, we have

$$E\left|y^{(s)}_{l(i,t),L}\right|^4 \le E\left[E\left(\left|y_{l(i,t),L}\right|^4|\mathcal{F}_{l(i,t),L}(s)\right)\right] \le \sup_{L,l(i,t)\in D_L} E\left|y_{l(i,t),L}\right|^4 < \infty$$

for all $s > 0$, and $l(i,t)\in D_L$ with $L \ge 1$. Hence, $y^{(s)}_{l(i,t),L}$ is uniformly L_4-bounded, which implies it is uniformly integrable. Note that for each $s > 0$, $y^{(s)}_{l(i,t),L}$ is a measurable function of $\left\{\varsigma_{l(j,t'),L} : l(j,t')\in D, \|l(i,t) - l(j,t')\|_\infty \le s\right\}$. Then, we can apply the L_1-norm LLN for the spatial-time process $y^{(s)}_{l(i,t),L}$, which is an extension of Theorem 1 in Jenish and

Prucha (2012) (or Theorem 3 in Jenish and Prucha (2009)[172]). Note that showing the LLN relies on the Chebyshev's inequality, so a key of verifying this LLN is to control $Cov\left(y^{(s)}_{l(i,t),L}, y^{(s)}_{l(j,t'),L}\right)$. Then, $Cov\left(y^{(s)}_{l(i,t),L}, y^{(s)}_{l(j,t'),L}\right) = 0$ if $\left\|l\left(i,t\right) - l\left(j,t'\right)\right\|_{\infty} > 2s$ for any $s > 0$. Hence, we obtain for any $s > 0$

$$\left\| \frac{1}{L} \sum_{l(i,t) \in D_L} \left(y^{(s)}_{l(i,t),L} - E\left(y_{l(i,t),L}\right)\right) \right\|_{L_1} \to 0 \tag{B.15}$$

as $L \to \infty$.

By Proposition B.2.5, we have

$$\left\| \frac{1}{L} \sum_{l(i,t) \in D_L} \left(y_{l(i,t),L} - E\left(y_{l(i,t),L} | \mathcal{F}_{l(i,t),L}(s)\right)\right) \right\|_{L_1} \to 0 \tag{B.16}$$

as $s \to \infty$. The results (B.15) and (B.16) with the triangle inequality yield

$$\begin{aligned}
& \lim_{L\to\infty} \left\| \frac{1}{L} \sum_{l(i,t) \in D_L} \left(y_{l(i,t),L} - E\left(y_{l(i,t),L}\right)\right) \right\|_{L_1} \\
= & \lim_{s\to\infty} \lim_{L\to\infty} \left\| \frac{1}{L} \sum_{l(i,t) \in D_L} \left(y_{l(i,t),L} - E\left(y_{l(i,t),L}\right)\right) \right\|_{L_1} \\
\leq & \lim_{s\to\infty} \limsup_{L\to\infty} \left\| \frac{1}{L} \sum_{l(i,t) \in D_L} \left(y_{l(i,t),L} - E\left(y_{l(i,t),L} | \mathcal{F}_{l(i,t),L}(s)\right)\right) \right\|_{L_1} \\
& + \lim_{s\to\infty} \lim_{L\to\infty} \left\| \frac{1}{L} \sum_{l(i,t) \in D_L} \left(E\left(y_{l(i,t),L} | \mathcal{F}_{l(i,t),L}(s)\right) - E\left(y_{l(i,t),L}\right)\right) \right\|_{L_1} \\
= & \ 0.
\end{aligned}$$

This completes the proof. Q.E.D.

A key intuition of the LLN is using the approximated $y_{l(i,t),L}$ based on $\varsigma_{l(j,t'),L}$ such that $\left\|l\left(i,t\right) - l\left(j,t'\right)\right\|_{\infty} \leq s$ for some $s > 0$ (denoted by $y^{(s)}_{l(i,t),L}$). Since $\varsigma_{l(i,t),L}$ is based

[172]Theorem 3 in Jenish and Prucha (2009) is designed for a spatial mixing process. Note that the main input process in our research belongs to α-mixing since it is based on an i.i.d. continuous innovation. By Theorem 14.1 in Davidson (1994), any measurable transformation of finite α-mixing processes is also α-mixing and the size is preserved.

on the i.i.d. disturbance $\epsilon_{l(i,t),L}$, $Cov\left(y^{(s)}_{l(i,t),L}, y^{(s)}_{l(j,t'),L}\right) = 0$ if $\|l(i,t) - l(j,t')\|_\infty > 2s$ for any $s > 0$. And then, we require that $y^{(s)}_{l(i,t),L}$ approaches to $y_{l(i,t),L}$ when $s \to \infty$. For this, the condition $\lim_{s\to\infty} \psi(s) = 0$ is enough. Based on this LLN, we establish pointwise convergence of the sample moment function $g^c_{l(i,t),L}(\theta)$ for each $\theta \in \Theta$. Note that the main component of $g^c_{l(i,t),L}(\theta)$ is $\epsilon_{l(i,t),L}(\theta)$ consisting of $y_{l(i,t),L}$'s and their transformations. Hence, we consider the NED property and kinds of transformations preserving that property. Proposition B.2.7 (i), (ii) and (iii) are respectively spatial-time extensions of Theorems 17.8, 9 and 10 in Davidson (1994).

Proposition B.2.7 *(i) Assume* $\left\|x_{l(i,t),L} - E\left(x_{l(i,t),L}|\mathcal{F}_{l(i,t),L}(s)\right)\right\|_{L_p} \leq d^x_{l(i,t),L}\psi_x(s)$ *and*
$\left\|y_{l(i,t),L} - E\left(y_{l(i,t),L}|\mathcal{F}_{l(i,t),L}(s)\right)\right\|_{L_p} \leq d^y_{l(i,t),L}\psi_y(s)$ *for* $p \geq 1$. *Then,*

$$\left\|x_{l(i,t),L} + y_{l(i,t),L} - E\left(x_{l(i,t),L} + y_{l(i,t),L}|\mathcal{F}_{l(i,t),L}(s)\right)\right\|_{L_p} \leq d_{l(i,t),L}\psi(s)$$

where $d_{l(i,t),L} = \max\left\{d^x_{l(i,t),L}, d^y_{l(i,t),L}\right\}$ *and* $\psi(s) = \psi_x(s) + \psi_y(s)$.

(ii) Assume $\left\|x_{l(i,t),L} - E\left(x_{l(i,t),L}|\mathcal{F}_{l(i,t),L}(s)\right)\right\|_{L_{2p}} \leq d^x_{l(i,t),L}\psi_x(s)$ *and* $\left\|y_{l(i,t),L} - E\left(y_{l(i,t),L}|\mathcal{F}_{l(i,t),L}(s)\right)\right\|_{L_{2p}} \leq d^y_{l(i,t),L}\psi_y(s)$ *for* $p \geq 1$. *Then,*

$$\left\|x_{l(i,t),L}y_{l(i,t),L} - E\left(x_{l(i,t),L}y_{l(i,t),L}|\mathcal{F}_{l(i,t),L}(s)\right)\right\|_{L_p} \leq d_{l(i,t),L}\psi(s)$$

where $d_{l(i,t),L} = \max\left\{\sup_{L,l\in D_L}\left\|x_{l(i,t),L}\right\|_{L_{2p}}, \left\|y_{l(i,t),L}\right\|_{L_{2p}}\right\}\left(d^x_{l(i,t),L} + d^y_{l(i,t),L}\right)$ *and* $\psi(s) = \max\{\psi_x(s), \psi_y(s)\}$.

Proposition B.2.7 says that the summation, and multiplication can preserve the NED property. The remark below describes it and comes from Theorem 17.16 and Example 17.17 in Davidson (1994).

Remark B.2.8 *Assume* $\sup_{L,l(i,t)\in D_L}\left\|x_{l(i,t),L}\right\|_{L_{2r}}<\infty$, $\sup_{L,l(i,t)\in D_L}\left\|y_{l(i,t),L}\right\|_{L_{2r}}<\infty$ *for* $r>2$, *and* $\left\{x_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ *and* $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ *are both* L_2*-NED on* ς*. Then,* $\left\{x_{l(i,t),L}y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ *becomes also* L_2*-NED on* ς*.*

By Proposition B.2.5, we observe that $y_{l(i,t),L}$ belongs to the $L_{4+\eta_\epsilon}$-space if $\sup_{L,l(i,t)\in D_L}\left\|\varsigma_{l(i,t),L}\right\|_{L_{4+\eta_\epsilon}}<\infty$. Note that $\varsigma_{l(i,t),L}$ consists of $\epsilon_{l(i,t),L}$, exogenous part of economic variables, nonstochastic regressors and individual and time dummies. By the conditioning argument, we only consider the moment condition for $\epsilon_{l(i,t),L}$ since the remaining components are assumed to be bounded. By Assumption 3.4.2, we have $\sup_{i,t}E\left|\epsilon_{it}\right|^{4+\eta_\epsilon}<\infty$ for some $\eta_\epsilon>0$ implying that $y_{l(i,t),L}$ is $L_{4+\eta_\epsilon}$ bounded. Assume that $x_{l(i,t),L}$ is a (uniformly bounded) linear transformation of $y_{l(i,t),L}$. Then, $\left\{x_{l(i,t),L}y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ will be $L_{2+\frac{\eta_\epsilon}{2}}$-bounded (by the Minkowski's inequality) and L_2-NED on ς.

Now we establish basic ingredients for consistency of GMM estimator $\hat{\theta}_L$. So, uniform convergence of $\bar{g}_L^c(\theta)$ is required: i.e., $\sup_{\theta\in\Theta}\left|\bar{g}_L^c(\theta)-E\left(\bar{g}_L^c(\theta)\right)\right|\to_p 0$ as $L\to\infty$. As the first step, all components of $\epsilon_{l(i,t),L}(\theta)$ should converge to their expected values for each $\theta\in\Theta$. It relies on the following three propositions. Note that $\left\{y_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ is uniformly $L_{4+\eta_\epsilon}$-bounded and is uniformly L_2-NED on ς. Note that all elements in $\epsilon_{l(i,t),L}(\theta)$ are uniformly bounded transformation of $\left\{y_{l(i,t),L}\right\}_{i=1}^n$, $\left\{y_{l(i,t-1),L}\right\}_{i=1}^n$ and $\left\{\varsigma_{l(i,t),L}\right\}_{i=1}^n$. By Proposition B.2.7 and Remark B.2.8 with Assumptions 3.4.5, 3.4.6, and 3.4.7, the following propositions can be shown.

Proposition B.2.9 $\left\{\sum_{j=1}^n w_{t,ij}y_{l(j,t),L}\right\}_{l(i,t)\in D_L}$ *is also uniformly* $L_{4+\eta_\epsilon}$*-bounded and is uniformly* L_2*-NED on* ς*.*

Proposition B.2.10 *For any* $p = 1, \cdots, P$, $\left\{ [M_{n,t+1,p}(\theta)]_{ii} \cdot y^e_{l(i,t+1),L}(\theta) \right\}_{l(i,t) \in D_L}$ *is uniformly* $L_{4+\eta_\epsilon}$*-Bounded and is uniformly* L_2*-NED on* ς.

Proposition B.2.11 $\left\{ y^e_{l(i,t+1),L}(\theta) \cdot [N_{n,t+1}(\theta)]_{ii} \right\}_{l(i,t) \in D_L}$ *and* $\left\{ [\nabla^e V_{n,t+2}(\theta)]_i \right\}_{l(i,t) \in D_L}$ *are uniformly* $L_{4+\eta_\epsilon}$*–bounded and uniformly* L_2*-NED on* ς.

By Proposition B.2.7 (i), the addition preserves the NED properties. Therefore, for each $\theta \in \Theta$ $\left\{ \epsilon_{l(i,t),L}(\theta) \right\}_{l(i,t) \in D_L}$ is uniformly $L_{4+\eta_\epsilon}$-bounded and uniformly L_2-NED on ς since Θ is closed and bounded (Assumption 2.4.3). Also, $\left\{ \frac{\partial \epsilon_{l(i,t),L}(\theta_0)}{\partial \theta} \right\}_{l(i,t) \in D_L}$ is uniformly $L_{4+\eta_\epsilon}$-bounded and uniformly L_2-NED on ς.

Demeaning operator $J_T \otimes J_n$

Consider the sample moment function $\bar{g}^c_L(\theta)$. Without loss of generality, suppose $m = 1$ since all quadratic moments have the same form. We have

$$\begin{aligned}
\bar{g}^{\mathbf{L},c}_L(\theta) &= \frac{1}{L}\mathbf{q}'_L (J_T \otimes J_n) \mathcal{E}_L(\theta) \\
&= \frac{1}{L}\mathbf{q}'_L \mathcal{E}_L(\theta) - \frac{1}{L}\mathbf{q}'_L \left(\frac{1}{T} l_T l'_T \otimes I_n \right) \mathcal{E}_L(\theta) \\
&\quad -\frac{1}{L}\mathbf{q}'_L \left(I_T \otimes \frac{1}{n} l_n l'_n \right) \mathcal{E}_L(\theta) + \frac{1}{L}\mathbf{q}'_L \frac{1}{L} l_L l'_L \mathcal{E}_L(\theta)
\end{aligned}$$

and

$$\begin{aligned}
&\bar{g}^{\mathbf{Q},c}_L(\theta) = \tfrac{1}{L}\mathcal{E}'_L(\theta) (J_T \otimes J_n) \mathbf{R}_{L,1} (J_T \otimes J_n) \mathcal{E}_L(\theta) \\
&= \tfrac{1}{L}\mathcal{E}'_L(\theta) \mathbf{R}_{L,1} \mathcal{E}_L(\theta) - \tfrac{1}{L}\mathcal{E}'_L(\theta) \mathbf{R}_{L,1} \left(\tfrac{1}{T} l_T l'_T \otimes I_n \right) \mathcal{E}_L(\theta) \\
&-\tfrac{1}{L}\mathcal{E}'_L(\theta) \mathbf{R}_{L,1} \left(I_T \otimes \tfrac{1}{n} l_n l'_n \right) \mathcal{E}_L(\theta) + \tfrac{1}{L}\mathcal{E}'_L(\theta) \mathbf{R}_{L,1} \tfrac{1}{L} l_L l'_L \mathcal{E}_L(\theta) \\
&-\tfrac{1}{L}\mathcal{E}'_L(\theta) \left(\tfrac{1}{T} l_T l'_T \otimes I_n \right) \mathbf{R}_{L,1} \mathcal{E}_L(\theta) + \tfrac{1}{L}\mathcal{E}'_L(\theta) \left(\tfrac{1}{T} l_T l'_T \otimes I_n \right) \mathbf{R}_{L,1} \left(\tfrac{1}{T} l_T l'_T \otimes I_n \right) \mathcal{E}_L(\theta) \\
&+\tfrac{1}{L}\mathcal{E}'_L(\theta) \left(\tfrac{1}{T} l_T l'_T \otimes I_n \right) \mathbf{R}_{L,1} \left(I_T \otimes \tfrac{1}{n} l_n l'_n \right) \mathcal{E}_L(\theta) - \tfrac{1}{L}\mathcal{E}'_L(\theta) \left(\tfrac{1}{T} l_T l'_T \otimes I_n \right) \mathbf{R}_{L,1} \tfrac{1}{L} l_L l'_L \mathcal{E}_L(\theta) \\
&-\tfrac{1}{L}\mathcal{E}'_L(\theta) \left(I_T \otimes \tfrac{1}{n} l_n l'_n \right) \mathbf{R}_{L,1} \mathcal{E}_L(\theta) + \tfrac{1}{L}\mathcal{E}'_L(\theta) \left(I_T \otimes \tfrac{1}{n} l_n l'_n \right) \mathbf{R}_{L,1} \left(\tfrac{1}{T} l_T l'_T \otimes I_n \right) \mathcal{E}'_L(\theta) \\
&+\tfrac{1}{L}\mathcal{E}'_L(\theta) \left(I_T \otimes \tfrac{1}{n} l_n l'_n \right) \mathbf{R}_{L,1} \left(I_T \otimes \tfrac{1}{n} l_n l'_n \right) \mathcal{E}_L(\theta) - \tfrac{1}{L}\mathcal{E}'_L(\theta) \left(I_T \otimes \tfrac{1}{n} l_n l'_n \right) \mathbf{R}_{L,1} \tfrac{1}{L} l_L l'_L \mathcal{E}_L(\theta)
\end{aligned}$$

$+\frac{1}{L}\mathcal{E}_L'(\theta)\frac{1}{L}l_L l_L' \mathbf{R}_{L,1}\mathcal{E}_L(\theta) - \frac{1}{L}\mathcal{E}_L'(\theta)\frac{1}{L}l_L l_L'\mathbf{R}_{L,1}\left(\frac{1}{T}l_T l_T' \otimes I_n\right)\mathcal{E}_L(\theta)$

$-\frac{1}{L}\mathcal{E}_L'(\theta)\frac{1}{L}l_L l_L'\mathbf{R}_{L,1}\left(I_T \otimes \frac{1}{n}l_n l_n'\right)\mathcal{E}_L(\theta) + \frac{1}{L}\mathcal{E}_L'(\theta)\frac{1}{L}l_L l_L'\mathbf{R}_{L,1}\frac{1}{L}l_L l_L'\mathcal{E}_L(\theta).$

For each $\theta \in \Theta$, all components of $\bar{g}_L^{\mathbf{L},c}(\theta)$ and $\bar{g}_L^{\mathbf{Q},c}(\theta)$ should converge to their own expected values. To show them, we can have the alternative representations as summations of $l_T l_T' \otimes I_n$ and $I_T \otimes l_n l_n'$ using properties of the Kronecker product: (a) $l_T l_T' \otimes I_n = \sum_{i=1}^n (l_T \otimes e_{n,i})\left(l_T' \otimes e_{n,i}'\right)$, and (b) $I_T \otimes l_n l_n' = \sum_{t=1}^T (e_{T,t} \otimes l_n)\left(e_{T,t}' \otimes l_n'\right)$.

Lemmas and theorems

Uniform laws of large numbers

For the ULLN (of $S_L^c(\theta)$) and uniform equicontinuity of $(E(S_L^c(\theta)))$, we need to check the three conditions suggested by Assumption 6 in Jenish and Prucha (2012). Proofs can be found in the supplement file. First, the moment functions $g_{l(i,t),L}^c(\theta)$ are required to be p-dominated on Θ for $p = 2$.

Lemma B.2.12 *Under the suggested regularity conditions in the main text,*

$$\sup_{L,l(i,t)\in D_L} E\sup_{\theta\in\Theta}\left|g_{l(i,t),L}^c(\theta)\right|^2 < \infty. \tag{B.17}$$

Proof of B.2.12. Choose arbitrary spatial-time unit $l(i,t) \in D_L$ and fix it. First, consider the linear moment function $g_{l(i,t),L}^{\mathbf{L},c}(\theta) = \mathbf{q}_L'(J_T \otimes J_n)e_{L,l}\epsilon_{l(i,t),L}(\theta)$. Then,

$$\begin{aligned} & E\sup_{\theta\in\Theta}\left|g_{l(i,t),L}^{\mathbf{L},c}(\theta)\right|^2 \\ = & \sup_{\theta\in\Theta}\left|\epsilon_{l(i,t),L}(\theta)\right|^2\left(\begin{array}{c} \frac{(n-1)(T-1)}{nT}\left|q_{l(i,t),L}\right|^2 - \frac{n-1}{nT}\sum_{l(i,t')\in D_L,t\neq t'}\left|q_{l(i,t'),L}\right|^2 \\ -\frac{T-1}{nT}\sum_{l(j,t)\in D_L,j\neq i}\left|q_{l(j,t),L}\right|^2 + \frac{1}{nT}\sum_{l(j,t')\in D_L,j\neq i,t\neq t'}\left|q_{l(j,t'),L}\right|^2 \end{array}\right) \\ \leq & 4\sup_{\theta\in\Theta}\left|\epsilon_{l(i,t),L}(\theta)\right|^2 \cdot \sup_{L,l(i,t)\in D_L}\left|q_{l(i,t),L}\right|^2, \end{aligned}$$

so by the Minkowski's inequality,

$$E\sup_{\theta\in\Theta}\left|g_{l(i,t),L}^{\mathbf{L},c}(\theta)\right|^2 \leq 4\sup_{\theta\in\Theta}\left\|\epsilon_{l(i,t),L}(\theta)\right\|_{L_4}^2 \cdot \sup_{L,l(i,t)\in D_L}\left\|q_{l(i,t),L}\right\|_{L_4}^2$$

The right hand side of inequality above is uniformly bounded by Assumption 2.4.7 (ii) and $\sup_{\theta\in\Theta}\left\|\epsilon_{l(i,t),L}(\theta)\right\|_{L_{4+\eta_\epsilon}}^2 < \infty$ due to the results in Propositions B.2.9, B.2.10, and B.2.11.

Second, consider the quadratic moment function $g_{l(i,t),L}^{\mathbf{Q},c}(\theta) = \sum_{l(j,t')\in D_L}\left[(J_T\otimes J_n)\mathbf{R}_{L,1}(J_T\otimes J_n)\right]_{l,l'}\epsilon_{l(i,t),L}(\theta)\epsilon_{l(j,t'),L}(\theta)$ where $l = l(i,t)$ and $l' = l(j,t')$. By the Minkowski's inequality, we have

$$E\sup_{\theta\in\Theta}\left|g_{l(i,t),L}^{\mathbf{Q},c}(\theta)\right|^2 \leq \left|\sum_{l(j,t')\in D_L}\left[(J_T\otimes J_n)\mathbf{R}_{L,1}(J_T\otimes J_n)\right]_{l,l'}\right|\cdot\sup_{L,l(i,t)\in D_L}\sup_{\theta\in\Theta}\left\|\epsilon_{l(i,t),L}(\theta)\right\|_{L_4}^4.$$

The right hand side of the above inequality is bounded since $\sup_{\theta\in\Theta}\left\|\epsilon_{l(i,t),L}(\theta)\right\|_{L_{4+\eta_\epsilon}}^2 < \infty$ and

$$\begin{aligned}\left|\sum_{l(j,t')\in D_L}\left[(J_T\otimes J_n)\mathbf{R}_{L,1}(J_T\otimes J_n)\right]_{l,l'}\right| &\leq \max_{i=1,\cdots,n}\sum_{j=1}^{n}\left|e_{n,i}'\sum_{s=1}^{T}J_{ts}R_{nt,1}J_{st'}e_{n,j}\right| \\ &\leq \sup_t\max_{i=1,\cdots,n}\sum_{j=1}^{n}\left|\left[R_{nt,1}\right]_{ij}\right| < \infty\end{aligned}$$

by Assumption 3.4.10 (ii). This completes the proof. Q.E.D.

Next, we would like to show the LLN for the components of $g_{l(i,t),L}^{c}(\theta)$ for each $\theta\in\Theta$.

Lemma B.2.13 *For $\theta\in\Theta$, the following results hold.*

(i) $q_{l(i,t),L}'\epsilon_{l(i,t),L}(\theta)$ and $\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(i,t),L}(\theta)\epsilon_{l(j,t'),L}(\theta)$ are uniformly L_1-NED on ς. This implies $\frac{1}{L}\sum_l q_{l(i,t),L}'\epsilon_{l(i,t),L}(\theta) - E\left(\frac{1}{L}\sum_l q_{l(i,t),L}'\epsilon_{l(i,t),L}(\theta)\right) = o_p(1)$ and $\frac{1}{L}\sum_l\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(i,t),L}(\theta)\epsilon_{l(j,t'),L}(\theta) - E\left(\frac{1}{L}\sum_l\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(i,t),L}(\theta)\epsilon_{l(j,t'),L}(\theta)\right) = o_p(1)$.

(ii) $\frac{1}{L}\mathbf{q}_L'\left(\frac{1}{T}l_Tl_T'\otimes I_n\right)\mathcal{E}_L(\theta) - E\left(\frac{1}{L}\mathbf{q}_L'\left(\frac{1}{T}l_Tl_T'\otimes I_n\right)\mathcal{E}_L(\theta)\right) = o_p(1)$, and $\frac{1}{L}\mathcal{E}_L'(\theta)\mathbf{R}_{L,1}\left(\frac{1}{T}l_Tl_T'\otimes I_n\right)\mathcal{E}_L(\theta) - E\left(\frac{1}{L}\mathcal{E}_L'(\theta)\mathbf{R}_{L,1}\left(\frac{1}{T}l_Tl_T'\otimes I_n\right)\mathcal{E}_L(\theta)\right) = o_p(1)$.

(iii) $\frac{1}{L}\mathbf{q}'_L\left(I_T\otimes\frac{1}{n}l_nl'_n\right)\mathcal{E}_L(\theta)-E\left(\frac{1}{L}\mathbf{q}'_L\left(I_T\otimes\frac{1}{n}l_nl'_n\right)\mathcal{E}_L(\theta)\right)=o_p(1)$,

and $\frac{1}{L}\mathcal{E}'_L(\theta)\mathbf{R}_{L,1}\left(I_T\otimes\frac{1}{n}l_nl'_n\right)\mathcal{E}_L(\theta)-E\left(\frac{1}{L}\mathcal{E}'_L(\theta)\mathbf{R}_{L,1}\left(I_T\otimes\frac{1}{n}l_nl'_n\right)\mathcal{E}_L(\theta)\right)=o_p(1)$.

(iv) $\frac{1}{L}\mathbf{q}'_L\frac{1}{L}l_Ll'_L\mathcal{E}_L(\theta)-E\left(\frac{1}{L}\mathbf{q}'_L\frac{1}{L}l_Ll'_L\mathcal{E}_L(\theta)\right)=o_p(1)$,

and $\frac{1}{L}\mathcal{E}'_L(\theta)\mathbf{R}_{L,1}\frac{1}{L}l_Ll'_L\mathcal{E}_L(\theta)-E\left(\frac{1}{L}\mathcal{E}'_L(\theta)\mathbf{R}_{L,1}\frac{1}{L}l_Ll'_L\mathcal{E}_L(\theta)\right)=o_p(1)$.

Proof of Lemma B.2.13. (i) By Parts (ii) and (iii) in Assumption 3.4.9, $\left\{q_{l(i,t),L}\right\}_{l(i,t)\in D_L}$ is uniformly $L_{4+\eta_q}$-bounded and uniformly L_2-NED on ς. Since $\left\{\epsilon_{l(i,t),L}(\theta)\right\}_{l(i,t)\in D_L}$ is uniformly $L_{4+\eta_\epsilon}$-bounded and uniformly L_2-NED on ς, $\left\{q'_{l(i,t),L}\epsilon_{l(i,t),L}(\theta)\right\}_{l(i,t)\in D_L}$ is uniformly L_2-bounded (by the Minkowski's inequality) and uniformly L_1-NED by Proposition B.2.7 (ii) (or uniformly L_2-NED by Remark B.2.8). Then, $\left\|\frac{1}{L}\sum_l q'_{l(i,t),L}\epsilon_{l(i,t),L}(\theta)-E\left(\frac{1}{L}\sum_l q'_{l(i,t),L}\epsilon_{l(i,t),L}(\theta)\right)\right\|_{L_1}\to 0$.

Consider $\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(i,t),L}(\theta)\,\epsilon_{l(j,t'),L}(\theta)=\epsilon_{l(i,t),L}(\theta)\cdot\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(j,t'),L}(\theta)$. Since $\left\{\epsilon_{l(i,t),L}(\theta)\right\}_{l(i,t)\in D_L}$ is uniformly $L_{4+\eta_\epsilon}$-bounded and uniformly L_2-NED on ς, focus on $\left\{\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(j,t'),L}(\theta)\right\}$. Note that $\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(j,t'),L}(\theta)=\sum_{j=1}^n\left[R_{nt,1}\right]_{ij}\epsilon_{l(j,t),L}(\theta)=\sum_{j\in B_i(d_c)}\left[R_{nt,1}\right]_{ij}\epsilon_{l(j,t),L}(\theta)$ by Assumption 3.4.10 (iii). Note that $\sum_{j=1}^n\left|\left[R_{nt,1}\right]_{ij}\right|$ and $\sum_{j=1}^n\mathbf{1}\left\{j:j\in B_i(d_c)\right\}$ are uniformly bounded in i, t and n and $\left\{\epsilon_{l(i,t),L}(\theta)\right\}_{l(i,t)\in D_L}$ is uniformly $L_{4+\eta_\epsilon}$-bounded, so $\left\{\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(j,t'),L}(\theta)\right\}$ is uniformly $L_{4+\eta_\epsilon}$-bounded. Since $\left[R_{nt}\right]_{ij}$ are L_2-NED on ς if $j\in B_i(d_c)$, $\left\{\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(j,t'),L}(\theta)\right\}$ is L_1-NED on ς by Proposition B.2.7 (ii). Then, $\left\|\frac{1}{L}\sum_l\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(i,t),L}(\theta)\,\epsilon_{l(j,t'),L}(\theta)-E\left(\frac{1}{L}\sum_l\sum_{l'}\left[\mathbf{R}_{L,1}\right]_{l,l'}\epsilon_{l(i,t),L}(\theta)\,\epsilon_{l(j,t'),L}(\theta)\right)\right\|_{L_1}\to 0$.

(ii) First, note that

$\frac{1}{L}\mathbf{q}'_L\left(\frac{1}{T}l_Tl'_T\otimes I_n\right)\mathcal{E}_L(\theta)=\frac{1}{n}\sum_{i=1}^n\left(\frac{1}{T}\mathbf{q}'_L\left(l_T\otimes e_{ni}\right)\right)\cdot\left(\frac{1}{T}\left(l'_T\otimes e'_{ni}\right)\mathcal{E}_L(\theta)\right)=\frac{1}{n}\sum_{i=1}^n\bar{q}'_{i.}\bar{\epsilon}_{i.}(\theta)$

where $\bar{q}_{i.} = \frac{1}{T}\sum_{t=1}^{T} q_{l(i,t),L}$ and $\bar{\epsilon}_{i.}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\epsilon_{l(i,t),L}(\theta)$. Observe

$$\begin{aligned} Var\left(\frac{1}{n}\sum_{i=1}^{n}\bar{q}_{i.}'\bar{\epsilon}_{i.}(\theta)\right) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov\left(\bar{q}_{i.}'\bar{\epsilon}_{i.}(\theta), \bar{q}_{j.}'\bar{\epsilon}_{j.}(\theta)\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\in B_i(ad_c)} Cov\left(\bar{q}_{i.}'\bar{\epsilon}_{i.}(\theta), \bar{q}_{j.}'\bar{\epsilon}_{j.}(\theta)\right) for some a \in \mathbf{N} \\ &\leq \frac{1}{n}c_5 (ad_c)^d \max_{j\in B_i(ad_c)} Cov\left(\bar{q}_{i.}'\bar{\epsilon}_{i.}(\theta), \bar{q}_{j.}'\bar{\epsilon}_{j.}(\theta)\right) \\ &\leq \frac{1}{n}c_5 (ad_c)^d \sqrt{\sup_{i,t,L} E\left(q_{l(i,t),L}^4\right)}\sqrt{\sup_{i,t,L} E\left(\epsilon_{l(i,t),L}(\theta)^4\right)} = O\left(\frac{1}{n}\right). \end{aligned}$$

This result implies $\frac{1}{n}\sum_{i=1}^{n}\bar{q}_{i.}'\bar{\epsilon}_{i.}(\theta) - E\left(\frac{1}{n}\sum_{i=1}^{n}\bar{q}_{i.}'\bar{\epsilon}_{i.}(\theta)\right) = O_p\left(\frac{1}{\sqrt{n}}\right)$ by the Chebyshev's inequality.

Observe $\frac{1}{L}\mathcal{E}_L'(\theta)\mathbf{R}_{L,1}\left(\frac{1}{T}l_T l_T' \otimes I_n\right)\mathcal{E}_L(\theta) = \frac{1}{n}\sum_{i=1}^{n}\tilde{r}_{i.}(\theta)\bar{\epsilon}_{i.}(\theta)$ where $\tilde{r}_{i.}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{n}[R_{nt,1}]_{ji}\epsilon_{l(j,t),L}(\theta)$. Assumption 3.4.10 (iii) says $\tilde{r}_{i.}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j\in B_i(ad_c)}[R_{nt,1}]_{ji}\epsilon_{l(j,t),L}(\theta)$. Then, $Var\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{r}_{i.}(\theta)\bar{\epsilon}_{i.}(\theta)\right) = O\left(\frac{1}{n}\right)$. Hence, we also have $\frac{1}{n}\sum_{i=1}^{n}\tilde{r}_{i.}(\theta)\bar{\epsilon}_{i.}(\theta) - E\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{r}_{i.}(\theta)\bar{\epsilon}_{i.}(\theta)\right) = O_p\left(\frac{1}{\sqrt{n}}\right)$ by the Chebyshev's inequality, which it is the desired result.

(iii) Consider the time-series average forms of $\frac{1}{L}\mathbf{q}_L'\left(I_T \otimes \frac{1}{n}l_n l_n'\right)\mathcal{E}_L(\theta)$ and $\frac{1}{L}\mathcal{E}_L'(\theta)\mathbf{R}_{L,1}\left(I_T \otimes \frac{1}{n}l_n l_n'\right)\mathcal{E}_L(\theta)$:

$$\begin{aligned} \frac{1}{L}\mathbf{q}_L'\left(I_T \otimes \frac{1}{n}l_n l_n'\right)\mathcal{E}_L(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{n}\left(e_{T,t}' \otimes l_n'\right)\mathbf{q}_L\right)\left(\frac{1}{n}\left(e_{T,t}' \otimes l_n'\right)\mathcal{E}_L(\theta)\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta) \end{aligned}$$

and

$$\begin{aligned} \frac{1}{L}\mathcal{E}_L'(\theta)\mathbf{R}_{L,1}\left(I_T \otimes \frac{1}{n}l_n l_n'\right)\mathcal{E}_L(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{n}\left(e_{T,t}' \otimes l_n'\right)\mathbf{R}_{L,1}\mathcal{E}_L(\theta)\right)\left(\frac{1}{n}\left(e_{T,t}' \otimes l_n'\right)\mathcal{E}_L(\theta)\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\tilde{r}_{.t}(\theta)\bar{\epsilon}_{.t}(\theta) \end{aligned}$$

where $\bar{q}_{.t} = \frac{1}{n}\sum_{i=1}^{n} q_{l(i,t),L}$, $\bar{\epsilon}_{.t}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\epsilon_{l(i,t),L}(\theta)$, and $\tilde{r}_{.t}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}[R_{nt,1}]_{ij}\epsilon_{l(j,t),L}(\theta)$.

First, consider showing $\frac{1}{T}\sum_{t=1}^{T}\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta) - E\left(\frac{1}{T}\sum_{t=1}^{T}\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\right) = o_p(1)$. Recall that $\left\{q_{l(i,t),L}'\epsilon_{l(i,t),L}(\theta)\right\}_{l(i,t)\in D_L}$ is uniformly spatial-time L_2-NED on ς, so this implies $\{\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\}$ is uniformly time L_2-NED on $\{\varsigma_{nt}\}$.

That is, $\sup_t \left\|\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta) - E\left(\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\,|\mathcal{F}_t(s)\right)\right\|_{L_2} \leq C_{qe}\tau^{*s}$ implying

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta) - \frac{1}{T}\sum_{t=1}^{T}E\left(\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\,|\mathcal{F}_t(s)\right)\right\|_{L_2} \leq C_{qe}\tau^{*s} \tag{B.18}$$

for some $0 < C_{qe} < \infty$ uniformly in t and T, and for some $0 < \tau^* < 1$. For notational convenience, let $\tilde{q}_{e,t} = E\left(\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\,|\mathcal{F}_t(s)\right)$. Note that $\tilde{q}_{e,t}$ is a function of $\varsigma_{n,t-s},\cdots,\varsigma_{n,t+s}$, so $\tilde{q}_{e,t_1}$ and $\tilde{q}_{e,t_2}$ will not be correlated if $|t_1-t_2| > 2s$. Then,

$$\begin{aligned}
E\left|\frac{1}{T}\sum_{t=1}^{T}\left[E\left(\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\,|\mathcal{F}_t(s)\right) - E\left(\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\right)\right]\right|^2 &= \frac{1}{T}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}Cov\left(\tilde{q}_{e,t_1},\tilde{q}_{e,t_2}\right) \\
&= \frac{1}{T}\sum_{t_1=1}^{T}\sum_{|t_1-t_2|\leq 2s}Cov\left(\tilde{q}_{e,t_1},\tilde{q}_{e,t_2}\right) \\
&\leq \frac{4s}{T}\sup_{L,i,t}E\left(\left|q_{l(i,t),L}\epsilon_{l(i,t),L}(\theta)\right|^2\right) \leq \frac{s}{T}C_{qe,1}
\end{aligned}$$

for some $C_{qe,1} \geq 4\sup_{L,i,t}E\left(\left|q_{l(i,t),L}\epsilon_{l(i,t),L}(\theta)\right|^2\right)$. By the Chebyshev's inequality, we have

$$\frac{1}{T}\sum_{t=1}^{T}\left[E\left(\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\,|\mathcal{F}_t(s)\right) - E\left(\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\right)\right] = O_p\left(\sqrt{\frac{s}{T}}\right). \tag{B.19}$$

By combining results (B.18) and (B.19),

$$\begin{aligned}
\left|\frac{1}{T}\sum_{t=1}^{T}\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta) - E\left(\frac{1}{T}\sum_{t=1}^{T}\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\right)\right| &\leq \left|\frac{1}{T}\sum_{t=1}^{T}\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta) - \frac{1}{T}\sum_{t=1}^{T}E\left(\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\,|\mathcal{F}_t(s)\right)\right| \\
&\quad + \left|\frac{1}{T}\sum_{t=1}^{T}\left[E\left(\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\,|\mathcal{F}_t(s)\right) - E\left(\bar{q}_{.t}'\bar{\epsilon}_{.t}(\theta)\right)\right]\right| \\
&= O_p\left(\tau^{*s} + \sqrt{\frac{s}{T}}\right) \to 0
\end{aligned}$$

as $T\to\infty$ if we set $s = \sqrt{T}$.

Second, observe that $\tilde{r}_{.t}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\in B_i(ad_c)}[R_{nt,1}]_{ij}\,\epsilon_{l(j,t),L}(\theta)$, $[R_{nt,1}]_{ij}$ are measurable functions of $\left\{\varsigma^*_{l(j,t'),L}\right\}_{j=1}^{n}|_{t'=1}^{t}$ and $\sigma\left(\left\{\varsigma^*_{l(j,t'),L}\right\}_{j=1}^{n}|_{t'=1}^{t}\right) \subseteq \mathcal{F}_{l(i,t),L}(s)$ for any $s > 0$. It implies $\tilde{r}_{.t}(\theta)$ is a cross-section average of a linear combination of $\epsilon_{l(j,t),L}(\theta)$'s (with finite components). By using similar arguments verifing the linear moment part, we have the desired result.

(iv) First, consider the linear moment part, $\frac{1}{L}\mathbf{q}'_L l_L \frac{1}{L} l'_L \mathcal{E}_L(\theta)$. Note that

$$\begin{aligned} Cov\left(\frac{1}{L}\mathbf{q}'_L l_L, \frac{1}{L} l'_L \mathcal{E}_L(\theta)\right) &= \frac{1}{L^2}\sum_{l}\sum_{l'} Cov\left(q_{l(i,t),L}\epsilon_{l(j,t'),L}(\theta)\right) \\ &\leq \frac{1}{L}\sup_{L,l(i,t)\in D_L}\sum_{l'} Cov\left(q_{l(i,t),L}\epsilon_{l(j,t'),L}(\theta)\right) = O\left(\frac{1}{L}\right), \end{aligned}$$

so $\frac{1}{L}\mathbf{q}'_L l_L \frac{1}{L} l'_L \mathcal{E}_L(\theta) - E\left(\frac{1}{L}\mathbf{q}'_L l_L \frac{1}{L} l'_L \mathcal{E}_L(\theta)\right) = O_p\left(\frac{1}{\sqrt{L}}\right)$ by the Chebyshev's inequality. By using the similar arguments, we have

$\frac{1}{L}\mathcal{E}'_L(\theta)\mathbf{R}_{L,1} l_L \frac{1}{L} l'_L \mathcal{E}_L(\theta) - E\left(\frac{1}{L}\mathcal{E}'_L(\theta)\mathbf{R}_{L,1} l_L \frac{1}{L} l'_L \mathcal{E}_L(\theta)\right) = O_p\left(\frac{1}{\sqrt{L}}\right)$. Q.E.D.

For each $\theta \in \Theta$, by Lemma B.2.13, hence, we have

$$\left|\frac{1}{L}\sum_{l(i,t)\in D_L} g^c_{l(i,t),L}(\theta) - E\left(\frac{1}{L}\sum_{l(i,t)\in D_L} g^c_{l(i,t),L}(\theta)\right)\right| \to_p 0$$

as $L \to \infty$.This implies $|S^c_L(\theta) - E(S^c_L(\theta))| \to_p 0$ as $L \to \infty$ for each $\theta \in \Theta$. To obtain consistency of $\hat{\theta}_L$, the pointwise LLN should be extended to the ULLN. To achieve smoothness of $g^c_{l(i,t),L}(\theta)$ on the parameter space Θ, hence, a sufficient condition is that $g^c_{l(i,t),L}(\theta)$ should satisfy the Lipschitz condition in $\theta \in \Theta$: for $l(i,t) \in D_L$, $L \geq 1$, and θ, $\theta' \in \Theta$

$$\left|g^c_{l(i,t),L}(\theta) - g^c_{l(i,t),L}(\theta')\right| \leq L_{l(i,t),L}\cdot|\theta - \theta'|$$

a.s., and $\limsup_{L\to\infty}\frac{1}{L}\sum_{l(i,t)\in D_L} EL^{\eta}_{l(i,t),L} < \infty$ for some $\eta > 0$. The lemma below verifies this.

Lemma B.2.14 *For any $l(i,t) \in D_L$, $g^c_{l(i,t),L}(\theta)$ is Lipschitz in $\theta \in \Theta$.*

Proof of B.2.14. Choose θ, $\theta' \in \Theta$ arbitrary with $\theta \neq \theta'$ and fix them. For each $l(i,t) \in D_L$, the Taylor approximation of $g^{c}_{l(i,t),L}(\theta)$ around θ' consists of

$$\begin{aligned} g^{\mathbf{L},c}_{l(i,t),L}(\theta) &= g^{\mathbf{L},c}_{l(i,t),L}(\theta') + \left[(J_T \otimes J_n)\,\mathbf{q}_L\right]'_{l.} \left[L_{\lambda,it}\left(\bar{\theta}\right), L_{\gamma,it}\left(\bar{\theta}\right), L'_{\psi,it}\left(\bar{\theta}\right), L'_{\beta,it}\left(\bar{\theta}\right)\right] \\ &\quad \cdot (\theta - \theta'), \end{aligned}$$

and

$$\begin{aligned} g^{\mathbf{Q},c}_{l(i,t),L}(\theta) &= g^{\mathbf{Q},c}_{l(i,t),L}(\theta') \\ &\quad + \sum_{l'} \left[(J_T \otimes J_n)\,\mathbf{R}_{L,l}\,(J_T \otimes J_n)\right]_{l.l'} \epsilon_{l(j,t),L}(\theta) \\ &\quad \cdot \left[L_{\lambda,it}\left(\bar{\theta}\right), L_{\gamma,it}\left(\bar{\theta}\right), L'_{\psi,it}\left(\bar{\theta}\right), L'_{\beta,it}\left(\bar{\theta}\right)\right] \cdot (\theta - \theta') \end{aligned}$$

where $\bar{\theta}$ lies between θ and θ'. Observe that $L_{\lambda,it}(\theta)$, $L_{\gamma,it}(\theta)$, $L_{\psi,it}(\theta)$, and $L_{\beta,it}(\theta)$ are uniformly L_4-bounded for any $\theta \in \Theta$ (due to the compact parameter space assumption). Hence, the Lipschitz condition in θ is satisfied for both $g^{\mathbf{L},c}_{l(i,t),L}(\theta)$ and $g^{\mathbf{Q},c}_{l(i,t),L}(\theta)$. Q.E.D.

This condition yields L_0 stochastic equicontinuity of $g^{c}_{l(i,t),L}(\theta)$ with respect to $\theta \in \Theta$ by Proposition 1 in Jenish and Prucha (2009). Formally, for any $\varepsilon > 0$

$$\limsup_{L\to\infty} \frac{1}{L} \sum_{l(i,t)\in D_L} P\left(\sup_{\|\theta-\theta'\|\leq\eta} \left|g^{c}_{l(i,t),L}(\theta) - g^{c}_{l(i,t),L}(\theta')\right| > \varepsilon\right) \to 0$$

as $\eta \to 0$. By Theorem 2 in Jenish and Prucha (2009), therefore, we obtain (i) uniform convergence

$$\sup_{\theta\in\Theta} \left|\bar{g}^{c}_{L}(\theta) - E\left(\bar{g}^{c}_{L}(\theta)\right)\right| \to_p 0,$$

and (ii) uniform equicontinuous of $\{E\left(\bar{g}^{c}_{L}(\theta)\right)\}$ on Θ,

i.e., $\limsup_{L\to\infty} \sup_{\theta'\in\Theta} \sup_{\|\theta-\theta'\|\leq\eta} \left|E\left(\bar{g}^{c}_{L}(\theta)\right) - E\left(\bar{g}^{c}_{L}(\theta')\right)\right| \to 0$ as $\eta \to 0$. For uniform convergence of $S^{c}_{L}(\theta)$, we produce the following lemma, which is a modified version of Lemma 3.3 in Pötscher and Prucha (1997).

Lemma B.2.15 *Let $\{\vartheta_L(\cdot,\cdot)\}$ be uniformly equicontinuous on $\mathbf{R}^q \times \Theta$.*

(a) If $\sup_{\theta\in\Theta} |\bar{g}_L^c(\theta) - E(\bar{g}_L^c(\theta))| \to_p 0$ as $L \to \infty$, then

$$\sup_{\theta\in\Theta} |\vartheta_L(\bar{g}_L^c(\theta), \theta) - \vartheta_L(E(\bar{g}_L^c(\theta)), \theta)| \to_p 0$$

as $L \to \infty$.

(b) If $\{E(\bar{g}_L^c(\theta))\}$ is uniformly equicontinuous on Θ, then so $\{\vartheta_L(E(\bar{g}_L^c(\theta)), \theta)\}$ is.

Proof of Lemma B.2.15. See Lemma 3.3 in Pötscher and Prucha (1997). Q.E.D.

The lemma below is not only for the identification assumption (Assumption 3.4.9), but it also differentiate the linear moment part (at θ_0), $\frac{1}{\sqrt{L}}\mathbf{q}_L'(J_T \otimes J_n)\mathcal{E}_L$, to (i) the mean zero part and (ii) the asymptotic bias part.

Lemma B.2.16 *For IV matrix $\mathbf{q}_{nt}$, $plim_{L\to\infty}\frac{1}{L}\mathbf{q}_L'(J_T \otimes J_n)\mathcal{E}_L = \mathbf{0}_{q\times 1}$.*

In $\frac{1}{\sqrt{L}}\mathbf{q}_L'(J_T \otimes J_n)\mathcal{E}_L$, the mean zero part is $\frac{1}{\sqrt{L}}\sum_{i=1}^n\sum_{t=1}^T q_{it}\epsilon_{it} - \frac{1}{n\sqrt{L}}\sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^T q_{jt}\epsilon_{it}$ while the asymptotic bias part is $-\frac{1}{T\sqrt{L}}\sum_{i=1}^n\sum_{t=1}^T\sum_{t'=1}^T q_{it'}\epsilon_{it} + \frac{1}{L\sqrt{L}}\sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^T\sum_{t'=1}^T q_{jt'}\epsilon_{it}$.

Proof of Lemma B.2.16. Note that

$$\begin{aligned}\frac{1}{L}\mathbf{q}_L'(J_T \otimes J_n)\mathcal{E}_L &= \frac{1}{L}\sum_{i=1}^n\sum_{t=1}^T q_{it}\epsilon_{it} - \frac{1}{TL}\sum_{i=1}^n\sum_{t=1}^T\sum_{t'=1}^T q_{it'}\epsilon_{it} \\ &\quad -\frac{1}{nL}\sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^T q_{jt}\epsilon_{it} + \frac{1}{L^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^T\sum_{t'=1}^T q_{jt'}\epsilon_{it}.\end{aligned}$$

Assumption 3.4.9 (i) implies $E(\mathbf{q}_{nt}'\mathcal{E}_{nt}) = E(\mathbf{q}_{nt}'E(\mathcal{E}_{nt}|\ell_{t-1})) = \mathbf{0}_{q\times 1}$ by the law of iterated expectation. Consider the first term. Note that

$$E\left(\frac{1}{L}\sum_{i=1}^n\sum_{t=1}^T q_{it}\epsilon_{it}\right) = \frac{1}{L}\sum_{i=1}^n\sum_{t=1}^T E(q_{it}E(\epsilon_{it}|\ell_{t-1})) = \mathbf{0}_{q\times 1}.$$

Since $E\left(\frac{1}{L}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}q_{it}\epsilon_{it}\epsilon_{jt'}q'_{jt'}\right)=\frac{\sigma_0^2}{L}\sum_{i=1}^{n}\sum_{t=1}^{T}E\left(q_{it}q'_{it}\right)=O\left(1\right)$ by Assumption 3.4.9 (ii), the first term is $O_p\left(\frac{1}{\sqrt{L}}\right)$ by the Chebyshev's inequality.

The expected value of the second term is

$$\begin{aligned}&E\left(\frac{1}{TL}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}q_{it'}\epsilon_{it}\right)\\ =\ &\frac{1}{TL}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'\leq t}E\left(q_{it'}E\left(\epsilon_{it}|\ell_{t-1}\right)\right)+\frac{1}{TL}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'>t}E\left(q_{it'}\epsilon_{it}\right)=\mathbf{0}_{q\times 1}+O\left(\frac{1}{T}\right)\end{aligned}$$

since $\sum_{t'>t}E\left(q_{it'}\epsilon_{it}\right)=O\left(1\right)$. We observe $\frac{1}{TL}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'>t}E\left(q_{it'}\epsilon_{it}\right)$ yields the asymptotic bias stemming from using J_T. Note that $\frac{1}{TL}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}q_{it'}\epsilon_{it}=\frac{1}{\sqrt{TL}}\frac{1}{\sqrt{TL}}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}q_{it'}\epsilon_{it}$. Consider a stochastic order of the variance part of $\frac{1}{\sqrt{TL}}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}q_{it'}\epsilon_{it}$:

$$E\left(\frac{1}{TL}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t'_1=1}^{T}\sum_{t'_2=1}^{T}q_{it'_1}\epsilon_{it_1}\epsilon_{jt_2}q_{jt'_2}\right)=\frac{1}{TL}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}E\left(\epsilon_{it}^2q_{it_1}q'_{it_2}\right)=O\left(1\right)$$

since $\sum_{t_1=1}^{T}E\left(\epsilon_{it}^2q_{it_1}q'_{it_2}\right)$ is uniformly bounded in i, t, t_2 and T.
Thus, $\frac{1}{TL}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}q_{it'}\epsilon_{it}=O_p\left(\frac{1}{\sqrt{TL}}\right)$ by the Chebyshev's inequality.

Consider the third term. Note that

$$E\left(\frac{1}{nL}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}q_{jt}\epsilon_{it}\right)=\frac{1}{nL}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}E\left(q_{jt}E\left(\epsilon_{it}|\ell_{t-1}\right)\right)=\mathbf{0}_{q\times 1},$$

which means that using J_n incorporated with the IVs $\mathbf{q}_{nt}$ does not generate the asymptotic bias. Note that

$$E\left(\frac{1}{nL}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\epsilon_{it}^2q_{j_1t}q'_{j_2t}\right)=\frac{\sigma_0^2}{nL}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}E\left(q_{j_1t}q'_{j_2t}\right)=O\left(1\right)$$

since $\sum_{j_1=1}^{n}\sigma_0^2E\left(q_{j_1t}q'_{j_2t}\right)$ is uniformly bounded in j_2, t and n.
Then, $\frac{1}{nL}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}q_{jt}\epsilon_{it}=O_p\left(\frac{1}{\sqrt{nL}}\right)$.

Last, consider the fourth term. The expected value of that term is

$$E\left(\frac{1}{L^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T} q_{jt'}\epsilon_{it}\right) = \frac{1}{L^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{t'\le t} E\left(q_{jt'}E\left(\epsilon_{it}|\ell_{t-1}\right)\right) + \frac{1}{L^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{t'>t} E\left(q_{jt'}\epsilon_{it}\right) = O\left(\frac{1}{L}\right)$$

since $\sum_{j=1}^{n}\sum_{t'>t} E\left(q_{jt'}\epsilon_{it}\right)$ is uniformly bounded in i, t and n.

And, $\frac{1}{L}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left(q_{jt'}\epsilon_{it}\right) = O\left(1\right)$, which implies $\frac{1}{L^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T} q_{jt'}\epsilon_{it} = O_p\left(\frac{1}{L}\right)$. This completes the proof. Q.E.D.

The mean zero part characterizes the asymptotic distribution of $\hat{\theta}_L$, which it can be written as $\frac{1}{\sqrt{L}}\sum_{t=1}^{T}\mathbf{q}_{nt}'J_n\mathcal{E}_{nt}$. The asymptotic bias part is $\frac{1}{\sqrt{L}}\sum_{t=1}^{T}\bar{\mathbf{q}}_{nT}'J_n\mathcal{E}_{nt}$ where $\bar{\mathbf{q}}_{nT} = \frac{1}{T}\sum_{t'=1}^{T}\mathbf{q}_{nt'}$ and its stochastic order is of $O_p\left(\frac{1}{T}\right)$. Hence, we obtain

$$\begin{aligned}\sqrt{L}\bar{g}_L^{\mathbf{L},c}\left(\theta_0\right) &= \frac{1}{\sqrt{L}}\sum_{t=1}^{T}\mathbf{q}_{nt}'J_n\mathcal{E}_{nt} - \frac{1}{\sqrt{L}}\sum_{t=1}^{T}\bar{\mathbf{q}}_{nT}'J_n\mathcal{E}_{nt} \\ &= \frac{1}{\sqrt{L}}\sum_{t=1}^{T}\mathbf{q}_{nt}'J_n\mathcal{E}_{nt} - \sqrt{\frac{n}{T}}\frac{1}{n}\sum_{t=1}^{T}E\left(\bar{\mathbf{q}}_{nT}'J_n\mathcal{E}_{nt}\right) \\ &\quad -\frac{1}{\sqrt{L}}\left[\sum_{t=1}^{T}\left[\bar{\mathbf{q}}_{nT}'J_n\mathcal{E}_{nt} - E\left(\bar{\mathbf{q}}_{nT}'J_n\mathcal{E}_{nt}\right)\right]\right],\end{aligned}$$

where $\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{t=1}^{T}E\left(\bar{\mathbf{q}}_{nT}'J_n\mathcal{E}_{nt}\right) = \sqrt{\frac{n}{T}}O\left(1\right)$ by Lemma B.2.16 and

$$\begin{aligned}\frac{1}{\sqrt{L}}\left[\sum_{t=1}^{T}\left[\bar{\mathbf{q}}_{nT}'J_n\mathcal{E}_{nt} - E\left(\bar{\mathbf{q}}_{nT}'J_n\mathcal{E}_{nt}\right)\right]\right] &= \frac{1}{\sqrt{T}}\frac{1}{\sqrt{TL}}\left[\sum_{t'=1}^{T}\sum_{t=1}^{T}\left[\mathbf{q}_{nt'}'J_n\mathcal{E}_{nt} - E\left(\mathbf{q}_{nt'}'J_n\mathcal{E}_{nt}\right)\right]\right] \\ &= O_p\left(\frac{1}{\sqrt{T}}\right)\end{aligned}$$

by the Chebyshev's inequality. Hence, we have

$$\sqrt{L}\bar{g}_L^{\mathbf{L},c}\left(\theta_0\right) = \frac{1}{\sqrt{L}}\sum_{t=1}^{T}\mathbf{q}_{nt}'J_n\mathcal{E}_{nt} - \sqrt{\frac{n}{T}}b_{nT}^{\mathbf{L}}\left(\theta_0,\sigma_0^2\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

where $b_{nT}^{\mathbf{L}}\left(\theta_0,\sigma_0^2\right) = \frac{1}{n}\sum_{t=1}^{T}E\left(\bar{\mathbf{q}}_{nT}'J_n\mathcal{E}_{nt}\right) = O\left(1\right)$.

Consider the order of asymptotic bias from using the quadratic moments.

Lemma B.2.17 *For* $l = 1, \cdots, m$, $\frac{\sigma_0^2}{\sqrt{L}} tr\left(\mathbf{R}_{L,l}\left(J_T \otimes J_n\right)\right) = O_p\left(\sqrt{\frac{T}{n}}\right)$.

Proof of Lemma B.2.17. Observe

$$\begin{aligned} \frac{\sigma_0^2}{\sqrt{L}} tr\left(\mathbf{R}_{L,l}\left(J_T \otimes J_n\right)\right) &= \frac{\sigma_0^2}{\sqrt{L}} tr\left(\mathbf{R}_{L,l}\right) - \frac{\sigma_0^2}{T\sqrt{L}} tr\left(\mathbf{R}_{L,l}\left(l_T l_T' \otimes J_n\right)\right) \\ &\quad - \frac{\sigma_0^2}{n\sqrt{L}} tr\left(\mathbf{R}_{L,l}\left(I_T \otimes l_n l_n'\right)\right) + \frac{\sigma_0^2}{L\sqrt{L}} tr\left(\mathbf{R}_{L,l} l_L l_L'\right), \end{aligned}$$

$tr\left(\mathbf{R}_{L,l}\right) = 0$, and $tr\left(\mathbf{R}_{L,l}\left(l_T l_T' \otimes J_n\right)\right) = 0$. The third term $-\frac{\sigma_0^2}{n\sqrt{L}} tr\left(\mathbf{R}_{L,l}\left(I_T \otimes l_n l_n'\right)\right) = -\frac{\sigma_0^2}{n\sqrt{L}} \sum_{t=1}^T \sum_{i=1}^n \sum_{j=1}^n \left[R_{nt,l}\right]_{ij} = O_p\left(\sqrt{\frac{T}{n}}\right)$ since $R_{nt,l}$'s column and row sums are uniformly bounded in n and t. By the same argument, $\frac{\sigma_0^2}{L\sqrt{L}} tr\left(\mathbf{R}_{L,l} l_L l_L'\right) = \frac{\sigma_0^2}{L\sqrt{L}} \sum_{t=1}^T \sum_{i=1}^n \sum_{j=1}^n \left[R_{nt,l}\right]_{ij} = O_p\left(\frac{1}{\sqrt{L}}\right)$. Q.E.D.

The implication of this result is that the asymptotic bias part originated from using quadratic moments is

$$\begin{aligned} & -\frac{\sigma_0^2}{n\sqrt{L}} tr\left(\mathbf{R}_{L,l}\left(I_T \otimes l_n l_n'\right)\right) + O_p\left(\frac{1}{\sqrt{L}}\right) \\ = & -\sqrt{\frac{T}{n}} \frac{\sigma_0^2}{L} \sum_{t=1}^T \sum_{i=1}^n \sum_{j=1}^n \left[R_{nt,l}\right]_{ij} + O_p\left(\frac{1}{\sqrt{L}}\right). \end{aligned}$$

So, define $b_{nT,l}^{\mathbf{Q}}\left(\sigma_0^2\right) = \frac{\sigma_0^2}{L} \sum_{t=1}^T \sum_{i=1}^n \sum_{j=1}^n \left[R_{nt,l}\right]_{ij}$ for $l = 1, \cdots, m$ and is of $O_p(1)$.

Central limit theorem for a linear quadratic form of martingale difference arrays

Consider the asymptotic distribution of $\mathbf{U}_L = \frac{1}{\sqrt{L}} A_L' \mathcal{E}_L + \frac{1}{\sqrt{L}}\left[\mathcal{E}_L' B_L \mathcal{E}_L - \sigma_0^2 tr\left(B_L\right)\right]$ where A_L is an $L \times 1$ vector of predetermined variables and B_L is an $L \times L$ matrix whose components are measurable functions of ς_L^*. The A_L and B_L represent respectively $\tilde{\mathbf{q}}_{L,k_a}$ and $\tilde{\mathbf{R}}_{L,k_a}$ for $k_a = 1, \cdots, K_a$. We can assume B_L is symmetric since it can be replaced by $\frac{1}{2}\left(B_L + B_L'\right)$ and $\mathcal{E}_L' B_L \mathcal{E}_L = \mathcal{E}_L' \frac{1}{2}\left(B_L + B_L'\right) \mathcal{E}_L$. Let $l = l(i,t)$ and $l' = l(j,t')$ for notational convenience. Note that

$$\mathbf{U}_L = \frac{1}{\sqrt{L}} \sum_{l=1}^L \left(\left[A_L\right]_l \epsilon_l + \epsilon_l \sum_{l'=1}^L \left[B_L\right]_{l,l'} \epsilon_{l'} - E\left(\epsilon_l \sum_{l'=1}^L \left[B_L\right]_{l,l'} \epsilon_{l'} | \varsigma_L^*\right)\right) \quad \text{(B.20)}$$

$$= \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \left([A_L]_l \epsilon_l + 2\epsilon_l \sum_{l'=1}^{l-1} [B_L]_{l,l'} \epsilon_{l'} + [B_L]_{l,l} \left(\epsilon_l^2 - \sigma_0^2 \right) \right) = \sum_{l=1}^{L} \mathbf{u}_{L,l}$$

where $\mathbf{u}_{L,l} = \frac{1}{\sqrt{L}} \left[[A_L]_l \epsilon_l + 2\epsilon_l \sum_{l'=1}^{l-1} [B_L]_{l,l'} \epsilon_{l'} + [B_L]_{l,l} (\epsilon_l^2 - \sigma_0^2) \right]$ for each $l = 1, \cdots, L$. Note that

$E \left(\frac{1}{\sqrt{L}} A'_L \mathcal{E}_L | \varsigma_L^* \right) = 0$, $E \left(\frac{1}{\sqrt{L}} [\mathcal{E}'_L B_L \mathcal{E}_L - \sigma_0^2 tr(B_L)] | \varsigma_L^* \right) = 0$,

$Var \left(\frac{1}{\sqrt{L}} A'_L \mathcal{E}_L | \varsigma_L^* \right) = \frac{1}{L} \sigma_0^2 E(A'_L A_L | \varsigma_L^*)$,

$Cov \left(\frac{1}{\sqrt{L}} A'_L \mathcal{E}_L, \frac{1}{\sqrt{L}} [\mathcal{E}'_L B_L \mathcal{E}_L - \sigma_0^2 tr(B_L)] | \varsigma_L^* \right) = \frac{1}{L} \mu_3 E(A'_L | \varsigma_L^*) vec_D(B_L)$ where $\mu_3 = E(\epsilon_l^3)$, and

$$\begin{aligned} \frac{1}{L} Cov (\mathcal{E}'_L B_{L,1} \mathcal{E}_L, \mathcal{E}'_L B_{L,2} \mathcal{E}_L | \varsigma_L^*) &= \frac{1}{L} \left(\mu_4 - 3\sigma_0^4 \right) vec'_D(B_{L,1}) vec_D(B_{L,2}) \\ &\quad + \frac{1}{L} \sigma_0^4 tr \left(B_{L,1} \left(B_{L,2} + B'_{L,2} \right) \right) \end{aligned}$$

where $\mu_4 = E(\epsilon_l^4)$. Let $\sigma^2_{\mathbf{U}_L} = Var(\mathbf{U}_L | \varsigma_L^*)$. Hence,

$$\begin{aligned} \sigma^2_{\mathbf{U}_L} &= \frac{1}{L} \sigma_0^2 E(A'_L A_L | \varsigma_L^*) + \frac{1}{L} \left(\mu_4 - 3\sigma_0^4 \right) vec'_D(B_L) vec_D(B_L) \\ &\quad + \frac{2}{L} \sigma_0^4 tr \left(B_L^2 \right) + \frac{2}{L} \mu_3 E(A'_L | \varsigma_L^*) vec_D(B_L). \end{aligned} \tag{B.21}$$

The exogenous component (ς_l^*) from the network formation process can be captured by a sigma field $\mathcal{C}_l \subset \mathcal{F}$. i.e., ς_l^* is measurable with respect to $\mathcal{C}_l$. Then, $\{\varsigma_l^*\}_{l=1}^{L}$ is specified by $\mathcal{C} = \mathcal{C}_1 \vee \cdots \vee \mathcal{C}_L$ where $\vee$ is the notation for the sigma field generated by the union of two sigma fields.[173] Now we want to establish $\mathcal{C}$-stable convergence of $\mathbf{U}_L$, which says the joint limiting distribution for $(\mathbf{U}_L, \varsigma^*)$ where ς^* is any $\mathcal{C}$-measurable random variable (or vector or matrix). Since $\mathcal{C}$-stable convergence belongs to a joint convergence concept, $\mathcal{C}$-stable convergence implies convergence in distribution. Here is the formal definition of convergence in distribution $\mathcal{C}$-stably stated by Definition 2 in Kuersteiner and Prucha (2013).

[173]Kuersteiner and Prucha (2013, 2018) consider $\mathcal{C}$ as a representation of common economic shocks, i.e, $\mathcal{C} = \vee_{t=1}^{T} \mathcal{C}_t$ and each $\mathcal{C}_t$ represents t^{th}-period common economic shocks. So, we extend this notion.

Definition B.2.18 *Let* $\mathbf{U}$ *be a random variable on* $(\Omega, \mathcal{F}, P)$ *and* $\mathcal{C} \subset \mathcal{F}$. *We say* $\mathbf{U}_L \rightarrow_d \mathbf{U}$ $\mathcal{C}$*-stably if for all* $\varsigma^* \in \mathcal{C}$ *and all* $U \in \mathcal{B}(\mathbf{R})$ *with* $P(\mathbf{U} \in \partial U) = 0$, $P(\{\mathbf{U}_L \in U\} \cap \varsigma^*) \rightarrow P(\{\mathbf{U} \in U\} \cap \varsigma^*)$ *as* $L \rightarrow \infty$ *where* $\mathcal{B}(\mathbf{R})$ *denotes the Borel* σ*-field on* $\mathbf{R}$ *and* ∂U *is the boundary of* U.

Proposition A.1 in Kuersteiner and Prucha (2013) introduces the alternative definitions of $\mathcal{C}$-stable convergence. One of the alternative definitions verifying the CLT is related to the characteristic function: for any $t \in \mathbf{R}$ and $\mathcal{C}$-measurable p-essentially bounded random variable ς^*, $E[\varsigma^* \exp(\mathbf{i}t\mathbf{U}_L)] \rightarrow E[\varsigma^* \exp(\mathbf{i}t\mathbf{U})]$ as $L \rightarrow \infty$. The key of showing this is using the law of iterated expectations, i.e., $E[\varsigma^* \exp(\mathbf{i}t\mathbf{U}_L)] = E[\varsigma^* E[\exp(\mathbf{i}t\mathbf{U}_L|\mathcal{C})]]$ and $E[\varsigma^* \exp(\mathbf{i}t\mathbf{U})] = E[\varsigma^* E[\exp(\mathbf{i}t\mathbf{U}|\mathcal{C})]]$. In consequence, we will have $\mathbf{U}_L|\mathcal{C} \rightarrow_d N(0, \sigma^2_{\mathbf{U}})$ a.s. where $\sigma^2_{\mathbf{U}} = \text{plim}_{L\rightarrow\infty}\sigma^2_{\mathbf{U}_L}$ and it is equivalent that $\lim_{L\rightarrow\infty} E(\exp(\mathbf{i}t\mathbf{U}_L)|\mathcal{C}) = \exp(-t^2\sigma^2_{\mathbf{U}}/2)$ a.s. However, the unconditional distribution will be $\lim_{L\rightarrow\infty} E(\exp(\mathbf{i}t\mathbf{U}_L)) = E(\exp(-t^2\sigma^2_{\mathbf{U}}/2))$ a.s. (i.e., a mixed Gaussian distribution).

Now we establish the CLT for (B.20). Note that

$$\sigma^2_{\mathbf{U}} = plim_{L\rightarrow\infty} \left[\begin{array}{c} \frac{\sigma_0^2}{L} \sum_{l=1}^L [A_L]_l^2 + \frac{(\mu_4 - \sigma_0^4)}{L} \sum_{l=1}^L [B_L]_{ll}^2 \\ + \frac{4\sigma_0^4}{L} \sum_{l=1}^L \sum_{l'=1}^{l-1} [B_L]_{ll'}^2 + \frac{2\mu_3}{L} \sum_{l=1}^L [A_L]_l [B_L]_{ll} \end{array} \right] \tag{B.22}$$

by using (B.21) and $tr\left(B^2_{L,1}\right) = \sum_{l=1}^L \sum_{l'=1}^L [B_L]^2_{ll'} = \sum_{l=1}^L [B_L]^2_{ll} + 2\sum_{l=1}^L \sum_{l'=1}^{l-1} [B_L]^2_{ll'}$. Observe that $\sigma^2_{\mathbf{U}_L}$ and $\sigma^2_{\mathbf{U}}$ are $\mathcal{C}$-measurable.

Lemma B.2.19 *Assume* $\sigma^2_{\mathbf{U}_L} > 0$ *for sufficiently large* L, *and* σ_0^2, μ_3 *and* μ_4 *are constant.*[174] *Then,* $\mathbf{U}_L \rightarrow_d \sigma_{\mathbf{U}} \cdot \xi$ $\mathcal{C}$*-stably as* $L \rightarrow \infty$ *where* ξ $N(0,1)$ *is independent of* $\sigma_{\mathbf{U}}$ *(which is* $\mathcal{C}$*-measurable).*

[174]Or, we can assume σ_0^2, μ_3 and μ_4 are $\mathcal{C}$-measurable.

Proof of Lemma B.2.19. The first step is to construct increasing sub-σ-fields of $\mathcal{F}$ starting with $\mathcal{C}$. For $l = 1, \cdots, L$ define

$$\begin{aligned}
\mathcal{G}_{L,1} &= \sigma(\epsilon_{1,1}) \vee \mathcal{C}, ... \\
\mathcal{G}_{L,n+1} &= \sigma(\epsilon_{1,1}, \cdots, \epsilon_{n,1}, \epsilon_{1,2}) \vee \mathcal{C}, ... \\
\mathcal{G}_{L,l} &= \sigma(\epsilon_{1,1}, \cdots, \epsilon_{n,1}, \epsilon_{1,2}, \cdots, \epsilon_{n,2}, \epsilon_{1,t}, \cdots \epsilon_{it}) \vee \mathcal{C}
\end{aligned}$$

for $l = l(i,t)$ with $\mathcal{G}_{L,0} = \mathcal{C}$. Then, we observe (i) $\mathcal{G}_{L,0} \subset \mathcal{G}_{L,1} \subset \cdots \subset \mathcal{G}_{L,l-1} \subset \mathcal{G}_{L,l} \subset \cdots$, (ii) $\mathbf{u}_{L,l}$ is $\mathcal{G}_{L,l}$-measurable and (iii) $E(\mathbf{u}_{L,l}|\mathcal{G}_{L,l-1}) = 0$ (i.e., $\mathbf{u}_{L,l}$'s are martingale differences). These establish the martingale difference array, $\{(\mathbf{u}_{L,l}, \mathcal{G}_{L,l}) : 1 \leq l \leq L, L \geq 1\}$.

As the second step, we shall show the Liapounov type condition: i.e., $\sum_{l=1}^{L} E|\mathbf{u}_{L,l}|^{2+\eta_U} \to 0$ as $L \to \infty$ for some $\eta_U > 0$. Pick any $p, q > 0$ such that $\frac{1}{p} + \frac{1}{q} = 1$. By the triangle inequality, we have

$$|\mathbf{u}_{L,l}| \leq \frac{1}{\sqrt{L}} \left(\begin{array}{c} |[A_L]_l|^{p\frac{1}{p}} \cdot |\epsilon_l|^{q\frac{1}{q}} + |[B_L]_{ll}|^{\frac{1}{p}} \cdot |[B_L]_{ll}|^{\frac{1}{q}} \cdot |\epsilon_l^2 - \sigma_0^2|^{q\frac{1}{q}} \\ +2^{q\frac{1}{q}} |\epsilon_l|^{q\frac{1}{q}} \cdot \sum_{l'=1}^{l-1} |[B_L]_{ll'}|^{\frac{1}{p}} \cdot |[B_L]_{ll'}|^{\frac{1}{q}} |\epsilon_{l'}|^{q\frac{1}{q}} \end{array} \right),$$

which takes a form of $E|X \cdot Y|$. By the Hölder inequality,

$$\begin{aligned}
|\mathbf{u}_{L,l}| \leq\ & \frac{1}{\sqrt{L}} \left[|[A_L]_l|^p + \sum_{l'=1}^{l} |[B_L]_{ll'}| \right]^{\frac{1}{p}} \\
& \cdot \left[|\epsilon_l|^q + |[B_L]_{ll}| \cdot \left|\epsilon_l^2 - \sigma_0^2\right|^q + 2^q |\epsilon_l|^q \cdot \left(\sum_{l'=1}^{l-1} |[B_L]_{ll'}| \, |\epsilon_{l'}|^q \right) \right]^{\frac{1}{q}}
\end{aligned}$$

implying

$$\begin{aligned}
E|\mathbf{u}_{L,l}|^q \leq\ & \frac{1}{L^{1+\frac{\eta_U}{2}}} E \left[|[A_L]_l|^p + \sum_{l'=1}^{l} |[B_L]_{ll'}| \right]^{\frac{q}{p}} (= partI) \\
& \cdot \left[\begin{array}{c} E|\epsilon_l|^q + |[B_L]_{ll}| \cdot E|\epsilon_l^2 - \sigma_0^2|^q \\ +2^q E|\epsilon_l|^q \cdot \left(\sum_{l'=1}^{l-1} |[B_L]_{ll'}| \, E|\epsilon_{l'}|^q \right) \end{array} \right] (= partII)
\end{aligned}$$

Take $q = 2 + \eta_U$ for some small $\eta_U > 0$. Consider the part II in the above inequality. Since $\sup_l |\epsilon_l|^{4+\eta_\epsilon} < \infty$, $|[B_L]_{ll}|$ and $\sum_{l'=1}^{L} |[B_L]_{ll'}|$ are uniformly bounded in l and

L, the part II is uniformly bounded in l and L. Consider the part I. Since we set $q = 2 + \eta_U$, $\frac{q}{p} = 1 + \eta_U$. By applying the c_r-inequality, we have

$$partI = E\left[\left|[A_L]_l\right|^p + \sum_{l'=1}^{l} \left|[B_L]_{ll'}\right|\right]^{1+\eta_U} \leq 2^{\eta_U}\left(E\left|[A_L]_l\right|^{2+\eta_U} + c_B\right)$$

for some $c_B > 0$. Since $[A_L]_l$ is a part of $\mathbf{q}_L$, $E\left|[A_L]_l\right|^4 = O(1)$ implying $E\left|[A_L]_l\right|^{2+\eta_U} = O(1)$, so part I is of $O(1)$. In consequence, $\sum_{l=1}^{L} E\left|\mathbf{u}_{L,l}\right|^{2+\eta_U} = \frac{1}{L^{1+\frac{\eta_U}{2}}}\sum_{l=1}^{L} O(1) = O\left(L^{-\frac{\eta_U}{2}}\right) \to 0$ as $L \to \infty$.

In the third step, we need to verify that the conditional variance converges to $\sigma^2_{\mathbf{U}}$: $\sum_{l=1}^{L} E\left[\mathbf{u}^2_{L,l}|\mathcal{G}_{L,l-1}\right] \to_d \sigma^2_{\mathbf{U}}$. Note that

$$\mathbf{u}^2_{L,l} = \frac{1}{L}\left[\begin{array}{c} \left([A_L]_l\,\epsilon_l + 2\epsilon_l\sum_{l'=1}^{l-1}[B_L]_{ll'}\,\epsilon_{l'}\right)^2 + [B_L]^2_{ll}\left(\epsilon_l^2 - \sigma_0^2\right)^2 \\ +2\,[B_L]_{ll}\left(\epsilon_l^2 - \sigma_0^2\right)\left([A_L]_l\,\epsilon_l + 2\epsilon_l\sum_{l'=1}^{l-1}[B_L]_{ll'}\,\epsilon_{l'}\right) \end{array}\right],$$

$$\begin{aligned} & E\left[\left([A_L]_l\,\epsilon_l + 2\epsilon_l\sum_{l'=1}^{l-1}[B_L]_{ll'}\,\epsilon_{l'}\right)^2|\mathcal{G}_{L,l-1}\right] \\ = \;& [A_L]^2_l\,\sigma_0^2 + 4\,[A_L]_l\,\sigma_0^2\sum_{l'=1}^{l-1}[B_L]_{ll'}\,\epsilon_{l'} + 4\sigma_0^2\sum_{l'=1}^{l-1}\sum_{l''=1}^{l-1}[B_L]_{ll'}\,[B_L]_{ll''}\,\epsilon_{l''} \\ = \;& \sigma_0^2\left([A_L]_l + 2\sum_{l'=1}^{l-1}[B_L]_{ll'}\,\epsilon_{l'}\right)^2, \end{aligned}$$

$$E\left[[B_L]^2_{ll}\left(\epsilon_l^2 - \sigma_0^2\right)^2|\mathcal{G}_{L,l-1}\right] = [B_L]^2_{ll}\left(\mu_4 - \sigma_0^4\right),$$

and

$$\begin{aligned} & E\left[2\,[B_L]_{ll}\left(\epsilon_l^2 - \sigma_0^2\right)\left([A_L]_l\,\epsilon_l + 2\epsilon_l\sum_{l'=1}^{l-1}[B_L]_{ll'}\,\epsilon_{l'}\right)|\mathcal{G}_{L,l-1}\right] \\ = \;& 2\mu_3\,[B_L]_{ll}\left([A_L]_l + 2\sum_{l'=1}^{l-1}[B_L]_{ll'}\,\epsilon_{l'}\right). \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{l=1}^{L} E\left[\mathbf{u}^2_{L,l}|\mathcal{G}_{L,l-1}\right] \;=\; & \frac{\sigma_0^2}{L}\sum_{l=1}^{L}\left([A_L]_l + 2\sum_{l'=1}^{l-1}[B_L]_{ll'}\,\epsilon_{l'}\right)^2 + \frac{1}{L}\sum_{l=1}^{L}[B_L]^2_{ll}\left(\mu_4 - \sigma_0^4\right) \\ & + \frac{1}{L}\sum_{l=1}^{L} 2\mu_3\,[B_L]_{ll}\left([A_L]_l + 2\sum_{l'=1}^{l-1}[B_L]_{ll'}\,\epsilon_{l'}\right), \end{aligned}$$

which converges to $\sigma^2_{\mathbf{U}}$ stated in (B.22).

The multivariate case can be considered by applying the Cramér-Wold device, i.e., if $\mathbf{U}_L$ is a multivariate random variable, we can consider an arbitrary linear combination $c'\mathbf{U}_L$ with $c'c = 1$ which becomes a univariate random variable. Q.E.D.

The first step is to build increasing sub-σ-fields of $\mathcal{F}$ starting with $\mathcal{C}$. For $l = 1, \cdots, L$ define

$$\begin{aligned} \mathcal{G}_{L,1} &= \sigma(\epsilon_{1,1}) \vee \mathcal{C}, \cdots, \mathcal{G}_{L,n+1} = \sigma(\epsilon_{1,1}, \cdots, \epsilon_{n,1}, \epsilon_{1,2}) \vee \mathcal{C}, ... \\ \mathcal{G}_{L,l} &= \sigma(\epsilon_{1,1}, \cdots, \epsilon_{n,1}, \epsilon_{1,2}, \cdots, \epsilon_{n,2}, \epsilon_{1,t}, \cdots \epsilon_{it}) \vee \mathcal{C} \end{aligned}$$

for $l = l(i,t)$ with $\mathcal{G}_{L,0} = \mathcal{C}$. Then, we observe (i) $\mathcal{G}_{L,0} \subset \mathcal{G}_{L,1} \subset \cdots \subset \mathcal{G}_{L,l-1} \subset \mathcal{G}_{L,l} \subset \cdots$, (ii) $\mathbf{u}_{L,l}$ is $\mathcal{G}_{L,l}$-measurable, and (iii) $E(\mathbf{u}_{L,l}|\mathcal{G}_{L,l-1}) = 0$ (i.e., $\mathbf{u}_{L,l}$'s are martingale differences). These establish the martingale difference array, $\{(\mathbf{u}_{L,l}, \mathcal{G}_{L,l}) : 1 \leq l \leq L, L \geq 1\}$. By showing (i) $\sum_{l=1}^{L} E|\mathbf{u}_{L,l}|^{2+\eta_U} \to 0$ as $L \to \infty$ for some $\eta_U > 0$, and
(ii) $\sum_{l=1}^{L} E\left[\mathbf{u}^2_{L,l}|\mathcal{G}_{L,l-1}\right] \to_d \sigma^2_{\mathbf{U}}$, we finish the proof.

We observe that the limiting distribution of $\mathbf{U}_L$ is the mixed normal $\sigma_{\mathbf{U}} \cdot N(0,1)$ and two components $\sigma_{\mathbf{U}}$ and $N(0,1)$ are independent. As a special case, the limiting distribution of $\mathbf{U}_L$ will be $N(0, \sigma^2_{\mathbf{U}})$ if $\sigma^2_{\mathbf{U}}$ is nonstochastic. Even though $\sigma^2_{\mathbf{U}_L}$ and $\sigma^2_{\mathbf{U}}$ can be stochastic ($\mathcal{C}$-measurable), we have $\left(\mathbf{U}_L/\sigma^2_{\mathbf{U}_L}\right)^2 \to_d \chi^2_1$ as $L \to \infty$ by using $\sigma^2_{\mathbf{U}_L} \to_p \sigma^2_{\mathbf{U}}$ and asymptotic independence between $\mathbf{U}_L$ and $\sigma^2_{\mathbf{U}_L}$.

Lemma B.2.20 *Suppose the same assumptions of Lemma B.2.19. Then,* $(\mathbf{U}_L/\sigma_{\mathbf{U}_L})^2 \to_d \chi^2_1$ *as* $L \to \infty$.

Proof of Lemma B.2.20. Note that $1/\sigma^2_{\mathbf{U}_L} \to_p 1/\sigma^2_{\mathbf{U}}$ $\mathcal{C}$-measurable by continuous mapping theorem and $0 < 1/\sigma^2_{\mathbf{U}} < \infty$ since $\sigma^2_{\mathbf{U}_L} > 0$ is assumed for sufficiently large

L. Then, $\mathbf{U}_L/\sigma_{\mathbf{U}_L} = \mathbf{U}_L/\sigma_{\mathbf{U}} + o_p(1)$ implying $\mathbf{U}_L/\sigma_{\mathbf{U}_L} \to_d N(0,1)$ as $L \to \infty$. Hence, by the continuous mapping theorem, we have the desired result. Q.E.D.

Note that conventional test statistics (e.g., Wald statistic) may involve $(\mathbf{U}_L/\sigma_{\mathbf{U}_L})^2$.

Proof of Theorem 3.4.2. As a first step, consider uniform convergence of $S_L^c(\theta)$. i.e.,,

$\sup_{\theta\in\Theta}\left|S_L^c(\theta) - \bar{S}_L^c(\theta)\right| \to_p 0$ as $L \to \infty$ where $\bar{S}_L^c(\theta) = [E\bar{g}_L^c(\theta)]' a_0' a_0 [E\bar{g}_L^c(\theta)]$ for each $\theta \in \Theta$, which is the nonstochastic analogue of $S_L^c(\theta)$. By the triangle inequality,

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|S_L^c(\theta) - \bar{S}_L^c(\theta)\right| \\
\leq\ &\sup_{\theta\in\Theta}\left|\bar{g}_L^{c\prime}(\theta)\, a_0' a_0 \bar{g}_L^c(\theta) - [E\bar{g}_L^c(\theta)]' a_0' a_0 [E\bar{g}_L^c(\theta)]\right| + \sup_{\theta\in\Theta}\left|\bar{g}_L^{c\prime}(\theta)\,(a_L' a_L - a_0' a_0)\, \bar{g}_L^c(\theta)\right| \\
\leq\ &\sup_{\theta\in\Theta}\left|\bar{g}_L^{c\prime}(\theta)\, a_0' a_0 \bar{g}_L^c(\theta) - [E\bar{g}_L^c(\theta)]' a_0' a_0 [E\bar{g}_L^c(\theta)]\right| + \|a_L' a_L - a_0' a_0\|_\infty \cdot \sup_{\theta\in\Theta}\left|\bar{g}_L^c(\theta)\right|^2.
\end{aligned}$$

Consider the first term of the right hand side of the second inequality. Observe that

$$\sup_{\theta\in\Theta} E\left|\bar{g}_L^c(\theta)\right| \leq E\sup_{\theta\in\Theta}\left|\bar{g}_L^c(\theta)\right| \leq \frac{1}{L}\sum_{l(i,t)\in D_L} E\sup_{\theta\in\Theta}\left|g_{l(i,t),L}^c(\theta)\right|,$$

which is uniformly bounded by Lemma B.2.12. Hence, by applying Lemma B.2.15 (i), we have the first term converges uniformly in $\theta \in \Theta$:

$\sup_{\theta\in\Theta}\left|\bar{g}_L^{c\prime}(\theta)\, a_L' a_L \bar{g}_L^c(\theta) - [E\bar{g}_L^c(\theta)]' a_L' a_L [E\bar{g}_L^c(\theta)]\right| \to_p 0.$

Consider the second term of the left hand side of the second inequality. Note that

$$\begin{aligned}
E\sup_{\theta\in\Theta}\left|\bar{g}_L^c(\theta)\right|^2 &\leq \frac{1}{L^2}\sum_{l(i,t),l(j,t')\in D_L} E\left[\sup_{\theta\in\Theta}\left|g_{l(i,t),L}^c(\theta)\right| \cdot \sup_{\theta\in\Theta}\left|g_{l(j,t'),L}^{c\prime}(\theta)\right|\right] \\
&\leq \frac{1}{L^2}\sum_{l(i,t),l(j,t')\in D_L} E\left[\left(\sup_{\theta\in\Theta}\left|g_{l(i,t),L}^c(\theta)\right|\right)^2\right]^{\frac{1}{2}} \cdot E\left[\left(\sup_{\theta\in\Theta}\left|g_{l(j,t'),L}^{c\prime}(\theta)\right|\right)^2\right]^{\frac{1}{2}},
\end{aligned}$$

which is uniformly bounded by Lemma B.2.12. The second inequality above holds due to the Hölder's inequality. Hence, the second term converges uniformly in $\theta \in \Theta$ since $\sup_{\theta\in\Theta}\left|\bar{g}_L^c(\theta)\right|^2 = O_p(1)$ and we have $\|a_L' a_L - a_0' a_0\|_\infty \to_p 0$ by the assumption on a_L.

Now consider uniform equicontinuity of $\left\{\bar{S}_L^c(\theta)\right\}$ on Θ. By Lemma B.2.14, we can achieve uniform equicontinuity of $\{E(\bar{g}_L^c(\theta))\}$ on Θ. By applying Lemma B.2.15 (ii), we have the desired result. Last, with identification uniqueness obtained by Assumption 3.4.9 (iv), we have $\hat{\theta}_L \to_p \theta_0$ as $L \to \infty$. Q.E.D.

Proof of Theorem 3.4.3. In order to show asymptotic normality, we have the Taylor expansion:

$$\begin{aligned}
\sqrt{L}\left(\hat{\theta}_L - \theta_0\right) &= -\left[\frac{\partial \bar{g}_L^{c\prime}\left(\hat{\theta}_L\right)}{\partial \theta} a_L' a_L \frac{\partial \bar{g}_L^c\left(\bar{\theta}_L\right)}{\partial \theta'}\right]^{-1} \frac{\partial \bar{g}_L^{c\prime}\left(\hat{\theta}_L\right)}{\partial \theta} a_L' a_L \sqrt{L} \bar{g}_L^c(\theta_0) \\
&= -\left[\frac{\partial \bar{g}_L^{c\prime}\left(\hat{\theta}_L\right)}{\partial \theta} a_L' a_L' \frac{\partial \bar{g}_L^c\left(\bar{\theta}_L\right)}{\partial \theta'}\right]^{-1} \frac{\partial \bar{g}_L^{c\prime}\left(\hat{\theta}_L\right)}{\partial \theta} a_L' \\
&\quad \times \left[\begin{array}{c} \sqrt{L} a_L \bar{g}_L^{c,(u)}(\theta_0) - \sqrt{\frac{n}{T}} a_L^{(q)} b_L^{\mathbf{L}}(\theta_0, \sigma_0^2) \\ -\sqrt{\frac{T}{n}} \sum_{l=1}^m a_L^{(l)} b_{L,l}^{\mathbf{Q}}(\sigma_0^2) + o_p(1) \end{array}\right]
\end{aligned}$$

where $\bar{\theta}_L$ lies between $\hat{\theta}_L$ and θ_0. First, note that

$$\frac{\partial \bar{g}_L^c\left(\hat{\theta}_L\right)}{\partial \theta'} = \left[\frac{\partial \bar{g}_L^c\left(\hat{\theta}_L\right)}{\partial \theta'} - \frac{\partial \bar{g}_L^c(\theta_0)}{\partial \theta'}\right] + \left[\frac{\partial \bar{g}_L^c(\theta_0)}{\partial \theta'} - G_L\right] + G_L.$$

The first term is $o_p(1)$ by Theorem 4.1 and the continuous mapping theorem while the second term is $o_p(1)$ by applying the LLN to $\frac{\partial \bar{g}_L^c(\theta_0)}{\partial \theta'}$. Hence, $\frac{\partial \bar{g}_L^c\left(\hat{\theta}_L\right)}{\partial \theta'} = G_L + o_p(1)$. By Assumption 3.4.11, $G_L' a_L' a_L G_L$ is nonsingluar for sufficiently large L, so $\frac{\partial \bar{g}_L^{c\prime}\left(\hat{\theta}_L\right)}{\partial \theta} a_L' a_L \frac{\partial \bar{g}_L^c\left(\bar{\theta}_L\right)}{\partial \theta'}$ is invertible for large L and it is of $O_p(1)$.

Hence, $\sqrt{L}\left(\hat{\theta}_L - \theta_0\right) = O_p(1) \cdot \left(O_p(1) + O\left(\sqrt{\frac{n}{T}}\right) + O_p\left(\sqrt{\frac{T}{n}}\right) + o_p(1)\right)$, which implies $\hat{\theta}_L - \theta_0 = O_p\left(\max\left\{\frac{1}{\sqrt{L}}, \frac{1}{n}, \frac{1}{T}\right\}\right)$ and

$$\begin{aligned}
\sqrt{L}\left(\hat{\theta}_L - \theta_0\right) &= \left[G_L' a_L' a_L G_L + o_p(1)\right]^{-1} \\
&\quad \times G_L' a_L' \left[\begin{array}{c} \sqrt{L} a_L \bar{g}_L^{c,(u)}(\theta_0) - \sqrt{\frac{n}{T}} a_L^{(q)} b_L^{\mathbf{L}}(\theta_0, \sigma_0^2) \\ -\sqrt{\frac{T}{n}} \sum_{l=1}^m a_L^{(l)} b_{L,l}^{\mathbf{Q}}(\sigma_0^2) + o_p(1) \end{array}\right].
\end{aligned}$$

It leads that

$$\sqrt{L}\left(\hat{\theta}_L - \theta_0\right) + \sqrt{\frac{n}{T}} \left[G_L' a_L' a_L G_L\right]^{-1} G_L' a_L' a_L^{(q)} b_L^{\mathbf{L}}\left(\theta_0, \sigma_0^2\right)$$

$$+\sqrt{\frac{T}{n}}\left[G_L' a_L' a_L G_L\right]^{-1} G_L' a_L' \sum_{l=1}^{m} a_L^{(l)} b_{L,l}^{\mathbf{Q}}\left(\sigma_0^2\right)+o_p(1)$$
$$= \left[G_L' a_L' a_L G_L\right]^{-1} G_L' a_L' \sqrt{L} a_L \bar{g}_L^{c,(u)}\left(\theta_0\right) \rightarrow_d plim_{L\rightarrow\infty} \Omega_L\left(a_L' a_L\right)^{\frac{1}{2}} \cdot \xi_*$$

where $\Omega_L\left(a_L' a_L\right)=\left[G_L' a_L' a_L G_L\right]^{-1} G_L' a_L' a_L \Sigma_L a_L' a_L G_L\left[G_L' a_L' a_L G_L\right]^{-1}$, and $\xi_* \tilde{} N\left(\mathbf{0}_{(2+P+K)\times 1}, I_{2+P+K}\right)$ which is independent with $\Omega_L\left(a_L' a_L\right)$ since $a_L \rightarrow_p a_0$ and $\sqrt{L} a_0 \bar{g}_L^{c,(u)}\left(\theta_0\right) \rightarrow_d \left(a_0 \Sigma_0 a_0'\right)^{-\frac{1}{2}} \cdot \xi_*$ by Lemma B.2.19. Observe $\Omega_0\left(a_0' a_0\right)= \left[G_0' a_0' a_0 G_0\right]^{-1} G_0' a_0' a_0 \Sigma_0 a_0' a_0 G_0\left[G_0' a_0' a_0 G_0\right]^{-1}$ is $\mathcal{C}$-measurable. This completes the proof. Q.E.D.

B.3. Spatial network formation

In contrast to conventional network formation models, an econometrician might not observe the realized spatial network links. Hence, entries of W_{nt} are usually prespecified by a researcher. Motivated by the gravity model in international trade literature, we employ the following Cobb-Douglas specification for the spatial network link at $t+1$, $w_{t+1,ij}$: for $i \neq j$

$$w_{t+1,ij}=d_{ij}^{-\alpha_d} E_{t+1,ij}^{-\alpha_e}\left(\frac{y_{jt}}{y_{it}}\right)^{\alpha_w} \cdot \mathbf{1}\left\{(i,j)\, nbd\right\}$$

where α_d, α_e and α_w are coefficients. Note that $w_{t+1,ij}$ can be interpreted as the intensity (or amount) of the signal from i to j at time $t+1$. The expected signs of α_d, α_e, and α_w are positive. $\alpha_d>0$ (or $\alpha_e>0$) means that the geographic (or economic distance) between i and j has a negative impact on $w_{t+1,ij}$. If agent j chooses a large amount of health expenditure relative to i at time t (i.e., large $\frac{y_{jt}}{y_{it}}$), the intensity of the signal from i to j ($w_{t+1,ij}$) becomes large at time $t+1$.

Since we attempt to estimate α_d, α_e, and α_w with unobserved $w_{t+1,ij}$, we consider a proxy of the true spatial network link. By regional policy examples, demographic/economic flows can describe intensities of spatial network interactions. Figlio et al. (1999) consider the state-to-state migration flows to identify which states are neighbors. By the United States Census Bureau, we collect the data on the state-to-state annual migration flows (from 2005 to 2016). To estimate α_d, α_e, and α_w, we consider the following specification by taking logarithm:

$$\ln w_{t+1,ij}^{migration} = constant - \alpha_d \ln d_{ij} - \alpha_e \ln E_{t+1,ij} + \alpha_w \ln \left(\frac{y_{jt}}{y_{it}} \right) + \zeta_{t+1,ij} \qquad \text{(B.23)}$$

where $w_{t+1,ij}^{migration}$ is the number of residents who live in state i at time $t+1$ but lived in state j 1 year ago, and $\zeta_{t+1,ij}$ denotes a statistical error. For the statistical error $\zeta_{t+1,ij}$, we consider two specifications: (i) $\{\zeta_{t,ij}\}$ follows i.i.d. disturbances, and (ii) $\zeta_{t,ij} = c_i + c_j + \alpha_t + \zeta_{t,ij}^*$ where c_i is the i's individual fixed effect, α_t is the t^{th}-period time effect, and $\zeta_{t,ij}^*$ denotes the i.i.d. disturbance.

Table 3.3 shows the estimation results. For both specifications, all the coefficients are significant under the 10% significance level. For interpretations, focus on specification (2). First, the intensity of interaction decreses by 2.0421% when d_{ij} increases by 1%. Second, increasing $E_{t,ij}$ by 1% leads to decreasing $w_{t,ij}$ by 0.1313%. Those two observations can show the yardstick competition because $w_{t,ij}^{migration}$ will be large if i and j are (geographically or economically) similar. Note that the third observation can show the welfare motivated move. If state j spends more money on health relative to that of i at time t (i.e., high $\left(\frac{y_{jt}}{y_{it}}\right)$), the migration flow from i to j at time $t+1$ ($w_{t+1,ij}^{migration}$) will be large. Then, we can investigate the marginal effect of changing y_{it} (or y_{jt}) on $w_{t+1,ij}$. Observe that

$$\frac{\partial w_{t+1,ij}}{\partial y_{it}} = (-\alpha_e) \frac{w_{t+1,ij}}{E_{t+1,ij}} \psi_0 sgn \left(z_{i,t+1} - z_{j,t+1} \right) + (-\alpha_w) \frac{w_{t+1,ij}}{y_{it}}$$

where $\psi_0 sgn \left(z_{i,t+1} - z_{j,t+1} \right) = \frac{\partial E_{t+1,ij}}{\partial y_{it}}$. The first part comes from changing $E_{t+1,ij}$ via changing $z_{i,t+1}$. The second part is originated from changing the relative health expenditure $\frac{y_{jt}}{y_{it}}$.

Appendix C: Appendix for Chapter 4

C.1. Solutions to the algebraic matrix Riccati equations

Theoretical foundations of the infinite horizon problem with spatial interactions can be found in Jeong and Lee (2018). At first, we apply the backward induction method. Assume that the initial condition $(Y_{n,t-1}, \eta_{nt})$ is given and fixed. For the followers, an arbitrary follower i is chosen and fixed. Throughout this section, the superscript (j) denotes the iteration numbers.

First iterated components

We set $V_i^{F,(0)} = 0$ and $V^{L,(0)} = 0$ as the initial iteration. Let $\mathcal{I}_i = e_i e_i'$. For the derivations, the following representation for $u_i\left(Y_{nt}, Y_{n,t-1}, \eta_{it}\right)$ is useful:

$$\begin{aligned} & u_i\left(Y_{nt}, Y_{n,t-1}, \mathbf{b}_{nt}^*, \eta_{it}\right) \\ = & \left(\eta_{nt} + \Phi_{n,0}\mathbf{b}_{nt}^* + \left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n,t-1} + \left(\lambda_0 W_n - \frac{1}{2} I_n\right) Y_{nt}\right)' \mathcal{I}_i Y_{nt} \\ & -\frac{\gamma_0}{2} Y_{n,t-1}' \mathcal{I}_i Y_{n,t-1}. \end{aligned}$$

Consider the first iterated value for the follower i:

$$\begin{aligned} & V_i^{F,(1)}\left(Y_{n,t-1}, \mathbf{b}_{nt}^{*,(1)}, \eta_{nt}\right) \qquad \text{(C.1)} \\ = & \ Y_{n,t-1}' Q_i^{F,(1)} Y_{n,t-1} + Y_{n,t-1}' L_i^{F,b,(1)} \mathbf{b}_{nt}^{*,(1)} + Y_{n,t-1}' L_i^{F,\eta,(1)} \eta_{nt} \end{aligned}$$

$$+\mathbf{b}_{nt}^{*,(1)\prime}Q_i^{F,b,(1)}\mathbf{b}_{nt}^{*,(1)}+\mathbf{b}_{nt}^{*,(1)\prime}L_i^{F,b,\eta,(1)}\eta_{nt}+\eta_{nt}'Q_i^{F,\eta,(1)}\eta_{nt}+c_i^{F,(1)}$$

where $Q_i^{F,(1)}=\frac{1}{2}\left(A_n^{(1)\prime}\mathcal{I}_iA_n^{(1)}-\gamma_0\mathcal{I}_i\right)$, $L_i^{F,b,(1)}=A_n^{(1)\prime}\mathcal{I}_iB_n^{(1)}$, $L_i^{F,\eta,(1)}=A_n^{(1)\prime}\mathcal{I}_iC_n^{(1)}$, $Q_i^{F,b,(1)}=\frac{1}{2}B_n^{(1)\prime}\mathcal{I}_iB_n^{(1)}$, $L_i^{F,b,\eta,(1)}=B_n^{(1)\prime}\mathcal{I}_iC_n^{(1)}$, $Q_i^{F,\eta,(1)}=C_n^{(1)\prime}\mathcal{I}_iC_n^{(1)}$, $c_i^{F,(1)}=0$, $R_n^{F,(1)}=S_n$, $A_n^{(1)}=\left(R_n^{F,(1)}\right)^{-1}(\gamma_0I_n+\rho_0W_n)$, $B_n^{(1)}=\left(R_n^{F,(1)}\right)^{-1}\Phi_{n,0}$, and $C_n^{(1)}=\left(R_n^{F,(1)}\right)^{-1}$. For the above, $S_n^{-1}=I_n+\lambda_0W_nS_n^{-1}$ is employed. $\mathbf{b}_{nt}^{*,(1)}$ denotes the maximizer of the leader's optimization problem at the first iteration. Note that $Y_{nt}^{*,(1)}=A_n^{(1)}Y_{n,t-1}+B_n^{(1)}\mathbf{b}_{nt}^{*,(1)}+C_n^{(1)}\eta_{nt}$ is the vector of maximizers satisfying (C.1).

By the backward induction method, the next step is to verify the first iterated value function of the leader, $V^{L,(1)}$:

$$\begin{aligned}
&V^{L,(1)}(Y_{n,t-1},\eta_{nt},\tau_{nt})\\
=\ &Y_{n,t-1}'Q_n^{L,(1)}Y_{n,t-1}+Y_{n,t-1}'L_n^{L,(1)}\eta_{nt}+Y_{n,t-1}'L_n^{L,\tau,(1)}\tau_{nt}\\
&+\eta_{nt}'Q_n^{L,\eta,(1)}\eta_{nt}+\tau_{nt}'Q_n^{L,\tau,(1)}\tau_{nt}+\tau_{nt}'L_n^{L,\tau,\eta,(1)}\eta_{nt}+c_n^{L,(1)}
\end{aligned}$$

where $c_n^{L,(1)}=0$, $Q_n^{L,(1)}=\sum_{k=1}^nQ_{n,k}^{L,(1)}$, $L_n^{L,(1)}=\sum_{k=1}^nL_{n,k}^{L,(1)}$, $L_n^{L,\tau,(1)}=\sum_{k=1}^nL_{n,k}^{L,\tau,(1)}$, $Q_n^{L,\eta,(1)}=\sum_{k=1}^nQ_{n,k}^{L,\eta,(1)}$, $Q_n^{L,\tau,(1)}=\sum_{k=1}^nQ_{n,k}^{L,\tau,(1)}$, $L_n^{L,\tau,\eta,(1)}=\sum_{k=1}^nL_{n,k}^{L,\tau,\eta,(1)}$,

$$\begin{aligned}
Q_{n,k}^{L,(1)}=\ &\left(\gamma_0I_n+\rho_0W_n+\Phi_{n,0}D_n^{(1)}\right)'\mathcal{I}_k\left(A_n^{(1)}+B_n^{(1)}D_n^{(1)}\right)\\
&+\left(A_n^{(1)}+B_n^{(1)}D_n^{(1)}\right)'\left(\lambda_0W_n-\frac{1}{2}I_n\right)'\mathcal{I}_k\left(A_n^{(1)}+B_n^{(1)}D_n^{(1)}\right)-\frac{\gamma_0}{2}\mathcal{I}_k\\
&-\frac{1}{2}D_n^{(1)\prime}\mathcal{I}_kD_n^{(1)},
\end{aligned}$$

$$\begin{aligned}
L_{n,k}^{L,(1)}=\ &\left(A_n^{(1)}+B_n^{(1)}D_n^{(1)}\right)'\mathcal{I}_k\left(I_n+\Phi_{n,0}E_n^{(1)}\right)\\
&+\left(\gamma_0I_n+\rho_0W_n+\Phi_{n,0}D_n^{(1)}\right)'\mathcal{I}_k\left(C_n^{(1)}+B_n^{(1)}E_n^{(1)}\right)\\
&-\left(A_n^{(1)}+B_n^{(1)}D_n^{(1)}\right)'\mathcal{I}_kS_n^L\left(C_n^{(1)}+B_n^{(1)}E_n^{(1)}\right)-D_n^{(1)\prime}\mathcal{I}_kE_n^{(1)},
\end{aligned}$$

$$\begin{aligned}
L_{n,k}^{L,\tau,(1)}=\ &D_n^{(1)\prime}\Phi_{n,0}\mathcal{I}_kB_n^{(1)}F_n^{(1)}+\left(A_n^{(1)}+B_n^{(1)}D_n^{(1)}\right)'\mathcal{I}_k\Phi_{n,0}F_n^{(1)}\\
&-\left(A_n^{(1)}+B_n^{(1)}D_n^{(1)}\right)'\mathcal{I}_kS_n^LB_n^{(1)}F_n^{(1)}-D_n^{(1)\prime}\mathcal{I}_kF_n^{(1)}+D_n^{(1)\prime}\mathcal{I}_k,
\end{aligned}$$

$$Q_{n,k}^{L,\eta,(1)} = \left(I_n + \Phi_{n,0} E_n^{(1)}\right)' \mathcal{I}_k \left(C_n^{(1)} + B_n^{(1)} E_n^{(1)}\right) + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_k B_n^{(1)} F_n^{(1)}$$
$$+ \left(C_n^{(1)} + B_n^{(1)} E_n^{(1)}\right)' \left(\lambda_0 W_n - \frac{1}{2} I_n\right)' \mathcal{I}_k \left(C_n^{(1)} + B_n^{(1)} E_n^{(1)}\right)$$
$$-\frac{1}{2} E_n^{(1)\prime} \mathcal{I}_k E_n^{(1)},$$

$$Q_{n,k}^{L,\tau,(1)} = F_n^{(1)\prime} \Phi_{n,0} \mathcal{I}_k B_n^{(1)} F_n^{(1)} + \left(B_n^{(1)} F_n^{(1)}\right)' \left(\lambda_0 W_n - \frac{1}{2} I_n\right)' \mathcal{I}_k B_n^{(1)} F_n^{(1)}$$
$$-\frac{1}{2} F_n^{(1)\prime} \mathcal{I}_k F_n^{(1)} + \mathcal{I}_k F_n^{(1)},$$

$$L_{n,k}^{L,\tau,\eta,(1)} = \left(B_n^{(1)} F_n^{(1)}\right)' \mathcal{I}_k \left(I_n + \Phi_{n,0} E_n^{(1)}\right) + F_n^{(1)\prime} \Phi_{n,0} \mathcal{I}_k \left(C_n^{(1)} + B_n^{(1)} E_n^{(1)}\right)$$
$$- \left(B_n^{(1)} F_n^{(1)}\right)' \mathcal{I}_k S_n^L \left(C_n^{(1)} + B_n^{(1)} E_n^{(1)}\right) - F_n^{(1)\prime} \mathcal{I}_k E_n^{(1)} + \mathcal{I}_k E_n^{(1)},$$

$$R_n^{L,(1)} = I_n - B_n^{(1)\prime} \Phi_{n,0} - \Phi_{n,0} B_n^{(1)} + B_n^{(1)\prime} S_n^L B_n^{(1)},$$
$$D_n^{(1)} = \left(R_n^{L,(1)}\right)^{-1} \left[\left(\Phi_{n,0} - B_n^{(1)\prime} S_n^L\right) A_n^{(1)} + B_n^{(1)\prime} (\gamma_0 I_n + \rho_0 W_n)\right],$$

$$E_n^{(1)} = \left(R_n^{L,(1)}\right)^{-1} \left[\left(\Phi_{n,0} - B_n^{(1)\prime} S_n^L\right) C_n^{(1)} + B_n^{(1)\prime}\right],$$

and $F_n^{(1)} = \left(R_n^{L,(1)}\right)^{-1}$. Note that $\mathbf{b}_{nt}^{*,(1)} = D_n^{(1)} Y_{n,t-1} + E_n^{(1)} \eta_{nt} + F_n^{(1)} \tau_{nt}$ for the relations above.

Limiting algebraic matrix Riccati equations

Starting with the first iterated results, we generate the following matrix Riccati equations by mathematical induction. First of all, we consider the following Bellman equation (recursive relation):

$$V_i^F \left(Y_{n,t-1}, \mathbf{b}_{nt}^*, \eta_{nt}\right)$$
$$= \left(\eta_{nt} + \Phi_{n,0} \mathbf{b}_{nt}^* + (\gamma_0 I_n + \rho_0 W_n) Y_{n,t-1} + \left(\lambda_0 W_n - \frac{1}{2} I_n\right) Y_{nt}^*\right)' \mathcal{I}_i Y_{nt}^*$$

$$-\frac{\gamma_0}{2} Y'_{n,t-1} \mathcal{I}_i Y_{n,t-1}$$

$$+\delta \mathbf{E}_t \left[\begin{array}{c} Y^{*\prime}_{nt} Q^F_i Y^*_{nt} + Y^{*\prime}_{nt} L^{F,b}_i \mathbf{b}^*_{n,t+1} + Y^{*\prime}_{nt} L^{F,\eta}_i \eta_{n,t+1} \\ +\mathbf{b}^{*\prime}_{n,t+1} Q^{F,b}_i \mathbf{b}^*_{n,t+1} + \mathbf{b}^{*\prime}_{n,t+1} L^{F,b,\eta}_i \eta_{n,t+1} + \eta'_{n,t+1} Q^{F,\eta}_i \eta_{n,t+1} + c^F_i \end{array} \right]$$

where $\mathbf{b}^*_{n,t+1} = D_n Y^*_{nt} + E_n \eta_{n,t+1} + F_n \tau_{n,t+1}$, $Y^*_{nt} = A_n Y_{n,t-1} + B_n Y_{n,t-1} + C_n \eta_{nt}$, $A_n = \lim_{j\to\infty} A^{(j)}_n$, $B_n = \lim_{j\to\infty} B^{(j)}_n$, $C_n = \lim_{j\to\infty} C^{(j)}_n$, $D_n = \lim_{j\to\infty} D^{(j)}_n$, $E_n = \lim_{j\to\infty} E^{(j)}_n$, $F_n = \lim_{j\to\infty} F^{(j)}_n$, $Q^F_i = \lim_{j\to\infty} Q^{F,(j)}_i$, $L^{F,b}_i = \lim_{j\to\infty} L^{F,b,(j)}_i$, $L^{F,\eta}_i = \lim_{j\to\infty} L^{F,\eta,(j)}_i$, $Q^{F,b}_i = \lim_{j\to\infty} Q^{F,b,(j)}_i$, $L^{F,b,\eta}_i = \lim_{j\to\infty} L^{F,b,\eta,(j)}_i$, $Q^{F,\eta}_i = \lim_{j\to\infty} Q^{F,\eta,(j)}_i$, and $c^F_i = \lim_{j\to\infty} c^{F,(j)}_i$. Since both types of agents have the LQ payoff functions, all components of the algebraic Riccati equations (practical as well as limit versions) do not rely on the state variables, $(Y_{n,t-1}, \eta_{nt})$.[175] Then, we have the recursive formulations

$$\begin{aligned}
Q^F_i &= (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i A_n - \frac{\gamma_0}{2}\mathcal{I}_i + A'_n \left(\lambda_0 W_n - \frac{1}{2} I_n\right)' \mathcal{I}_i A_n \\
&\quad +\delta \left[A'_n Q^F_i A_n + A'_n D'_n Q^{F,b}_i D_n A_n \right], \\
L^{F,b}_i &= A'_n \mathcal{I}_i \Phi_{n,0} + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i B_n - A'_n S^L_n \mathcal{I}_i B_n \\
&\quad +\delta \left[\begin{array}{c} A'_n \left(Q^F_i + Q^{F\prime}_i\right) B_n + A'_n \left(L^{F,b}_i D_n + D'_n L^{F,b\prime}_i\right) B_n \\ +A'_n D'_n \left(Q^{F,b}_i + Q^{F,b\prime}_i\right) D_n B_n \end{array} \right],
\end{aligned}$$

$$\begin{aligned}
L^{F,\eta}_i &= A'_n \mathcal{I}_i + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i C_n - A'_n S^L_n \mathcal{I}_i C_n \\
&\quad +\delta \left[\begin{array}{c} A'_n \left(Q^F_i + Q^{F\prime}_i\right) C_n \\ +A'_n \left(L^{F,b}_i (D_n C_n + \pi_0 E_n) + D'_n L^{F,b\prime}_i C_n\right) \\ +\pi_0 A'_n L^{F,\eta}_i + A'_n D'_n \left(Q^{F,b}_i + Q^{F,b\prime}_i\right)(D_n C_n + \pi_0 E_n) \\ +\pi_0 A'_n D'_n L^{F,b,\eta}_i \end{array} \right],
\end{aligned}$$

$$\begin{aligned}
Q^{F,b}_i &= B'_n \mathcal{I}_i \Phi_{n,0} + B'_n \left(\lambda_0 W_n - \frac{1}{2} I_n\right)' \mathcal{I}_i B_n \\
&\quad +\delta \left[B'_n Q^F_i B_n + B'_n L^{F,b}_i D_n B_n + B'_n D'_n Q^{F,b}_i D_n B_n \right],
\end{aligned}$$

[175] That is, each component of the algebraic matrix Riccati equations is a function of the parameters and W_n.

$$L_i^{F,b,\eta} = B_n'\mathcal{I}_i + \Phi_{n,0}\mathcal{I}_i C_n - B_n' S_n^L \mathcal{I}_i C_n,$$
$$+\delta \left[\begin{array}{c} B_n'\left(Q_i^F + Q_i^{F\prime}\right)C_n \\ +B_n'\left(L_i^{F,b}\left(D_n C_n + \pi_0 E_n\right) + D_n' L_i^{F,b\prime} C_n\right) \\ +\pi_0 B_n' L_i^{F,\eta} + B_n' D_n'\left(Q_i^{F,b} + Q_i^{F,b\prime}\right)\left(D_n C_n + \pi_0 E_n\right) \\ +\pi_0 B_n' D_n' L_i^{F,b,\eta} \end{array} \right],$$

and

$$Q_i^{F,\eta} = C_n'\mathcal{I}_i + C_n'\left(\lambda_0 W_n - \frac{1}{2}I_n\right)'\mathcal{I}_i C_n$$
$$+\delta \left[\begin{array}{c} C_n' Q_i^F C_n + C_n' L_i^{F,b}\left(D_n C_n + \pi_0 E_n\right) + \pi_0 C_n' L_i^{F,\eta} \\ +C_n' D_n' Q_i^{F,b} D_n C_n + \pi_0 C_n' D_n'\left(Q_i^{F,b} + Q_i^{F,b\prime}\right)E_n \\ +\pi_0\left(D_n C_n + \pi_0 E_n\right)' L_i^{F,b,\eta} + \pi_0^2 Q_i^{F,\eta} \end{array} \right].$$

Define

$$Q_n^{F,*} = [\left(Q_1^F + Q_1^{F\prime}\right)e_1, \cdots, \left(Q_n^F + Q_n^{F\prime}\right)e_n],$$
$$L_n^{F,b,*} = [L_1^{F,b\prime}e_1, \cdots, L_n^{F,b\prime}e_n],$$

and

$$L_n^{F,\eta,*} = [L_1^{F,\eta\prime}e_1, \cdots, L_n^{F,\eta\prime}e_n].$$

Then,

$$R_n^F = S_n - \delta Q_n^{F,*} - \delta L_n^{F,b,*} D_n,$$
$$A_n = \left(R_n^F\right)^{-1}\left(\gamma_0 I_n + \rho_0 W_n\right), B_n = \left(R_n^F\right)^{-1}\Phi_{n,0},$$

and

$$C_n = \left(R_n^F\right)^{-1}\left[I_n + \delta\pi_0\left(L_n^{F,b,*}E_n + L_n^{F,\eta,*}\right)\right].$$

To obtain A_n, B_n, and C_n, we need to evaluate only $\left\{Q_i^F, L_i^{F,b}, L_i^{F,\eta}\right\}_{i=1}^n$. i.e., other components in $V_i^F\left(Y_{n,t-1}, \mathbf{b}_{nt}^*, \eta_{nt}\right)$ are not needed.

$$
\begin{aligned}
& V^{L,(1)}\left(Y_{n,t-1},\eta_{nt},\tau_{nt}\right) \\
= \;& Y_{n,t-1}^{\prime}Q_n^{L,(1)}Y_{n,t-1}+Y_{n,t-1}^{\prime}L_n^{L,(1)}\eta_{nt}+Y_{n,t-1}^{\prime}L_n^{L,\tau,(1)}\tau_{nt} \\
& +\eta_{nt}^{\prime}Q_n^{L,\eta,(1)}\eta_{nt}+\tau_{nt}^{\prime}Q_n^{L,\tau,(1)}\tau_{nt}+\tau_{nt}^{\prime}L_n^{L,\tau,\eta,(1)}\eta_{nt}+c_n^{L,(1)}
\end{aligned}
$$

To obtain $Q_n^{F,*}$, $L_n^{F,b,*}$, and $L_n^{F,\eta,*}$, however, we should verify D_n and E_n, which characterize the leader's optimal actions. In calculating D_n and E_n, the following recursive relation is employed: given $(Y_{n,t-1},\eta_{nt},\tau_{nt})$

$$
V^L\left(Y_{n,t-1},\eta_{nt},\tau_{nt}\right)=\max_{\mathbf{b}_{nt}}\left\{\begin{array}{c}\mathcal{W}_{0,t}\left(\mathbf{b}_{nt};Y_{nt}^*,Y_{n,t-1},\eta_{nt},\tau_{nt}\right)\\ +\delta\mathbf{E}_t\left[\begin{array}{c}Y_{nt}^{*\prime}Q_n^LY_{nt}^*+Y_{nt}^{*\prime}L_n^L\eta_{n,t+1}+Y_{nt}^{*\prime}L_n^{L,\tau}\tau_{n,t+1}\\ +\eta_{n,t+1}^{\prime}Q_n^{L,\eta}\eta_{n,t+1}+\tau_{n,t+1}^{\prime}Q_n^{L,\tau}\tau_{n,t+1}\\ \tau_{n,t+1}^{\prime}L_n^{L,\tau,\eta}\eta_{n,t+1}+c_n^L\end{array}\right]\end{array}\right\}
$$

subject to $Y_{nt}^*=A_nY_{n,t-1}+B_n\mathbf{b}_{nt}+C_n\eta_{nt}$, where $Q_{n,k}^L=\lim_{j\to\infty}Q_{n,k}^{L,(j)}$, $L_{n,k}^L=\lim_{j\to\infty}L_{n,k}^{L,(j)}$, $L_{n,k}^{L,\tau}=\lim_{j\to\infty}L_{n,k}^{L,\tau,(j)}$, $Q_{n,k}^{L,\eta}=\lim_{j\to\infty}Q_{n,k}^{L,\eta,(j)}$, $Q_{n,k}^{L,\tau}=\lim_{j\to\infty}Q_{n,k}^{L,\tau,(j)}$, $L_{n,k}^{L,\tau,\eta}=\lim_{j\to\infty}L_{n,k}^{L,\tau,\eta,(j)}$, $Q_n^L=\sum_{k=1}^nQ_{n,k}^L$, $L_n^L=\sum_{k=1}^nL_{n,k}^L$, $L_n^{L,\tau}=\sum_{k=1}^nL_{n,k}^{L,\tau}$, $Q_n^{L,\eta}=\sum_{k=1}^nQ_{n,k}^{L,\eta}$, $Q_n^{L,\tau}=\sum_{k=1}^nQ_{n,k}^{L,\tau}$, $L_n^{L,\tau,\eta}=\sum_{k=1}^nL_{n,k}^{L,\tau,\eta}$ and $c_n^L=\lim_{j\to\infty}c_n^{L,(j)}$. Then, we have the recursive relations:

$$
\begin{aligned}
Q_{n,k}^L \;=\; & \left(\gamma_0I_n+\rho_0W_n+\Phi_{n,0}D_n\right)^{\prime}\mathcal{I}_k\left(A_n+B_nD_n\right) \\
& +\left(A_n+B_nD_n\right)^{\prime}\left(\lambda_0W_n-\frac{1}{2}I_n\right)^{\prime}\mathcal{I}_k\left(A_n+B_nD_n\right)-\frac{\gamma_0}{2}\mathcal{I}_k \\
& -\frac{1}{2}D_n^{\prime}\mathcal{I}_kD_n+\delta\left[\left(A_n+B_nD_n\right)^{\prime}Q_{n,k}^L\left(A_n+B_nD_n\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
L_{n,k}^L \;=\; & \left(A_n+B_nD_n\right)^{\prime}\mathcal{I}_kR_n^e \\
& +\left(\gamma_0I_n+\rho_0W_n+\Phi_{n,0}D_n\right)^{\prime}\mathcal{I}_k\left(C_n+B_nE_n\right) \\
& -\left(A_n+B_nD_n\right)^{\prime}\mathcal{I}_kS_n^L\left(C_n+B_nE_n\right)-D_n^{\prime}\mathcal{I}_kE_n \\
& +\delta\left[\begin{array}{c}\left(A_n+B_nD_n\right)^{\prime}\left(Q_{n,k}^L+Q_{n,k}^{L\prime}\right)\left(B_nE_n+C_n\right)\\ +\pi_0\left(A_n+B_nD_n\right)^{\prime}L_{n,k}^L\end{array}\right],
\end{aligned}
$$

$$
\begin{aligned}
L_{n,k}^{L,\tau} &= D_n'\Phi_{n,0}\mathcal{I}_k B_n F_n + \left(A_n^{(1)} + B_n^{(1)} D_n^{(1)}\right)' \mathcal{I}_k \Phi_{n,0} F_n \\
&\quad - \left(A_n + B_n D_n\right)' \mathcal{I}_k S_n^L B_n F_n - D_n' \mathcal{I}_k F_n + D_n' \mathcal{I}_k \\
&\quad + \delta \left(A_n + B_n D_n\right)' \left(Q_{n,k}^L + Q_{n,k}^{L\prime}\right) B_n F_n,
\end{aligned}
$$

$$
\begin{aligned}
Q_{n,k}^{L,\eta} &= \left(R_n^e\right)' \mathcal{I}_k \left(C_n + B_n E_n\right) + \left(\gamma_0 I_n + \rho_0 W_n\right)' \mathcal{I}_k B_n F_n \\
&\quad + \left(C_n + B_n E_n\right)' \left(\lambda_0 W_n - \frac{1}{2} I_n\right)' \mathcal{I}_k \left(C_n + B_n E_n\right) - \frac{1}{2} E_n' \mathcal{I}_k E_n \\
&\quad + \delta \left[\begin{array}{l} \left(B_n E_n + C_n\right)' Q_{n,k}^{L,\eta} \left(B_n E_n + C_n\right) \\ + \pi_0 \left(B_n E_n + C_n\right)' L_{n,k}^L + \pi_0^2 Q_{n,k}^{L,\eta} \end{array} \right],
\end{aligned}
$$

$$
\begin{aligned}
Q_{n,k}^{L,\tau} &= F_n' \Phi_{n,0} \mathcal{I}_k B_n F_n + \left(B_n F_n\right)' \left(\lambda_0 W_n - \frac{1}{2} I_n\right)' \mathcal{I}_k B_n F_n \\
&\quad - \frac{1}{2} F_n' \mathcal{I}_k F_n + \mathcal{I}_k F_n \\
&\quad + \delta \left(B_n F_n\right)' Q_{n,k}^L B_n F_n,
\end{aligned}
$$

and

$$
\begin{aligned}
L_{n,k}^{L,\tau,\eta} &= \left(B_n F_n\right)' \mathcal{I}_k R_n^e + F_n' \Phi_{n,0} \mathcal{I}_k \left(C_n + B_n E_n\right) \\
&\quad - \left(B_n F_n\right)' \mathcal{I}_k S_n^L \left(C_n + B_n E_n\right) - F_n' \mathcal{I}_k E_n + \mathcal{I}_k E_n \\
&\quad + \delta \left[\left(B_n F_n\right)' \left(Q_{n,k}^L + Q_{n,k}^{L\prime}\right) \left(C_n + B_n E_n\right) + \pi_0 \left(B_n F_n\right)' L_{n,k}^L \right].
\end{aligned}
$$

Note that $R_n^e = I_n + \Phi_{n,0} E_n$. The resulting D_n and E_n are

$$
D_n = \left(R_n^L\right)^{-1} \left[\left(\Phi_{n,0} - B_n' \left(S_n^L - \delta Q_n^{L,*}\right)\right) A_n + B_n' \left(\gamma_0 I_n + \rho_0 W_n\right) \right],
$$

$$
E_n = \left(R_n^L\right)^{-1} \left[B_n' \left(I_n + \delta \pi_0 L_n^L\right) + \left(\Phi_{n,0} - B_n' \left(S_n^L - \delta Q_n^{L,*}\right)\right) C_n \right]
$$

and $F_n = \left(R_n^L\right)^{-1}$ where $R_n^L = I_n - B_n' \Phi_{n,0} - \Phi_{n,0} B_n + B_n' \left(S_n^L - \delta Q_n^{L,*}\right) B_n$.

Practical computation

Based on the above limit versions, we introduce a computing method for the algebraic Riccati equations. In this part, hence, we only report the components relevant to estimation. Note that the first iterated components were already revealed. Suppose that the j^{th} iterated components ($j = 1, 2, \cdots$) are revealed. Then, our first interest is to verify the $(j+1)^{th}$-iterated components for the followers' lifetime problems:

$$\begin{aligned} Q_i^{F,(j+1)} &= (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i A_n^{(j+1)} - \frac{\gamma_0}{2}\mathcal{I}_i \\ &+ A_n^{(j+1)\prime}\left(\lambda_0 W_n - \frac{1}{2}I_n\right)' \mathcal{I}_i A_n^{(j+1)} \\ &+ \delta\left[A_n^{(j+1)\prime} Q_i^{F,(j)} A_n^{(j+1)} + A_n^{(j+1)\prime} D_n^{(j)\prime} Q_i^{F,b,(j)} D_n^{(j)} A_n^{(j+1)}\right], \end{aligned}$$

$$\begin{aligned} L_i^{F,b,(j+1)} &= A_n^{(j+1)\prime}\mathcal{I}_i \Phi_{n,0} + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i B_n^{(j+1)} - A_n^{(j+1)\prime} S_n^L \mathcal{I}_i B_n^{(j+1)} \\ &+ \delta \begin{bmatrix} A_n^{(j+1)\prime}\left(Q_i^{F,(j)} + Q_i^{F,(j)\prime}\right) B_n^{(j+1)} \\ +A_n^{(j+1)\prime}\left(L_i^{F,b,(j)} D_n^{(j)} + D_n^{(j)\prime} L_i^{F,b,(j)\prime}\right) B_n^{(j+1)} \\ +A_n^{(j+1)\prime} D_n^{(j)\prime}\left(Q_i^{F,b,(j)} + Q_i^{F,b,(j)\prime}\right) D_n^{(j)} B_n^{(j+1)} \end{bmatrix}, \end{aligned}$$

$$\begin{aligned} & L_i^{F,\eta,(j+1)} \\ = & A_n^{(j+1)\prime}\mathcal{I}_i + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i C_n^{(j+1)} - A_n^{(j+1)\prime} S_n^L \mathcal{I}_i C_n^{(j+1)} \\ & + \delta \begin{bmatrix} A_n^{(j+1)\prime}\left(Q_i^{F,(j)} + Q_i^{F,(j)\prime}\right) C_n^{(j+1)} \\ +A_n^{(j+1)\prime}\left(L_i^{F,b,(j)}\left(D_n^{(j)} C_n^{(j+1)} + \pi_0 E_n^{(j)}\right) + D_n^{(j)\prime} L_i^{F,b,(j)\prime} C_n^{(j+1)}\right) \\ +\pi_0 A_n^{(j+1)\prime} L_i^{F,\eta,(j)} \\ +A_n^{(j+1)\prime} D_n^{(j)\prime}\left(Q_i^{F,b,(j)} + Q_i^{F,b,(j)\prime}\right)\left(D_n^{(j)} C_n^{(j+1)} + \pi_0 E_n^{(j)}\right) \\ +\pi_0 A_n^{(j+1)\prime} D_n^{(j)\prime} L_i^{F,b,\eta,(j)} \end{bmatrix}, \end{aligned}$$

$$\begin{aligned} Q_i^{F,b,(j+1)} &= B_n^{(j+1)\prime}\mathcal{I}_i \Phi_{n,0} + B_n^{(j+1)\prime}\left(\lambda_0 W_n - \frac{1}{2}I_n\right)' \mathcal{I}_i B_n^{(j+1)} \\ &+ \delta \begin{bmatrix} B_n^{(j+1)\prime} Q_i^{F,(j)} B_n^{(j+1)} + B_n^{(j+1)\prime} L_i^{F,b,(j)} D_n^{(j)} B_n^{(j+1)} \\ +B_n^{(j+1)\prime} D_n^{(j)\prime} Q_i^{F,b,(j)} D_n^{(j)} B_n^{(j+1)} \end{bmatrix}, \end{aligned}$$

and

$$
\begin{aligned}
& L_i^{F,b,\eta,(j+1)} \\
= \; & B_n^{(j+1)\prime}\mathcal{I}_i + \Phi_{n,0}\mathcal{I}_i C_n^{(j+1)} - B_n^{(j+1)\prime} S_n^L \mathcal{I}_i C_n^{(j+1)}, \\
& +\delta \begin{bmatrix} B_n^{(j+1)\prime}\left(Q_i^{F,(j)} + Q_i^{F,(j)\prime}\right) C_n^{(j+1)} \\ +B_n^{(j+1)\prime}\left(L_i^{F,b,(j)}\left(D_n^{(j)} C_n^{(j+1)} + \pi_0 E_n^{(j)}\right) + D_n^{(j)\prime} L_i^{F,b,(j)\prime} C_n^{(j+1)}\right) \\ +\pi_0 B_n^{(j+1)\prime} L_i^{F,\eta,(j)} \\ +B_n^{(j+1)\prime} D_n^{(j)\prime}\left(Q_i^{F,b,(j)} + Q_i^{F,b,(j)\prime}\right)\left(D_n^{(j)} C_n^{(j+1)} + \pi_0 E_n^{(j)}\right) \\ +\pi_0 B_n^{(j+1)\prime} D_n^{(j)\prime} L_i^{F,b,\eta,(j)} \end{bmatrix},
\end{aligned}
$$

where $R_n^{F,(j+1)} = S_n - \delta Q_n^{F,*,(j)} - \delta L_n^{F,b,*,(j)} D_n^{(j)}$, $A_n^{(j+1)} = \left(R_n^{F,(j+1)}\right)^{-1}(\gamma_0 I_n + \rho_0 W_n)$, $B_n^{(j+1)} = \left(R_n^{F,(j+1)}\right)^{-1}\Phi_{n,0}$, and $C_n^{(j+1)} = \left(R_n^{F,(j+1)}\right)^{-1}\left[I_n + \delta\pi_0\left(L_n^{F,b,*,(j)} E_n^{(j)} + L_n^{F,\eta,*,(j)}\right)\right]$. Note that we omit reporting some components which are not relevant to the followers' optimal actions.

Second, we calculate the $(j+1)^{th}$-iterated components relevant to the leader's optimal actions:

$$
\begin{aligned}
& Q_{n,k}^{L,(j+1)} \\
= \; & \left(\gamma_0 I_n + \rho_0 W_n + \Phi_{n,0} D_n^{(j+1)}\right)' \mathcal{I}_k \left(A_n^{(j+1)} + B_n^{(j+1)} D_n^{(j+1)}\right) \\
& + \left(A_n^{(j+1)} + B_n^{(j+1)} D_n^{(j+1)}\right)' \left(\lambda_0 W_n - \frac{1}{2} I_n\right)' \mathcal{I}_k \left(A_n^{(j+1)} + B_n^{(j+1)} D_n^{(j+1)}\right) \\
& - \frac{\gamma_0}{2}\mathcal{I}_k - \frac{1}{2} D_n^{(j+1)\prime}\mathcal{I}_k D_n^{(j+1)} \\
& + \delta\left[\left(A_n^{(j+1)} + B_n^{(j+1)} D_n^{(j+1)}\right)' Q_{n,k}^{L,(j)} \left(A_n^{(j+1)} + B_n^{(j+1)} D_n^{(j+1)}\right)\right],
\end{aligned}
$$

and

$$
\begin{aligned}
& L_{n,k}^{L,(j+1)} \\
= \; & \left(A_n^{(j+1)} + B_n^{(j+1)} D_n^{(j+1)}\right)' \mathcal{I}_k \left(I_n + \Phi_{n,0} E_n^{(j+1)}\right) \\
& + \left(\gamma_0 I_n + \rho_0 W_n + \Phi_{n,0} D_n^{(j+1)}\right)' \mathcal{I}_k \left(C_n^{(j+1)} + B_n^{(j+1)} E_n^{(j+1)}\right)
\end{aligned}
$$

$$
-\left(A_n^{(j+1)}+B_n^{(j+1)}D_n^{(j+1)}\right)'\mathcal{I}_k S_n^L\left(C_n^{(j+1)}+B_n^{(j+1)}E_n^{(j+1)}\right)-D_n^{(j+1)\prime}\mathcal{I}_k E_n^{(j+1)}
$$
$$
+\delta\left[\begin{array}{c}\left(A_n^{(j+1)}+B_n^{(j+1)}D_n^{(j+1)}\right)'\left(Q_{n,k}^{L,(j)}+Q_{n,k}^{L,(j)\prime}\right)\left(B_n^{(j+1)}E_n^{(j+1)}+C_n^{(j+1)}\right)\\ +\pi_0\left(A_n^{(j+1)}+B_n^{(j+1)}D_n^{(j+1)}\right)'L_{n,k}^{L,(j)}\end{array}\right]
$$

where $Q_n^{L,(j+1)}=\sum_{k=1}^n Q_{n,k}^{L,(j+1)}$, $L_n^{L,(j+1)}=\sum_{k=1}^n L_{n,k}^{L,(j+1)}$, $R_n^{L,(j+1)}=I_n-B_n^{(j+1)\prime}\Phi_{n,0}-\Phi_{n,0}B_n^{(j+1)}+B_n^{(j+1)\prime}\left(S_n^L-\delta Q_n^{L,*,(j)}\right)B_n^{(j+1)}$,

$$
D_n^{(j+1)}=\left(R_n^{L,(j+1)}\right)^{-1}\left[\begin{array}{c}\left(\Phi_{n,0}-B_n^{(j+1)\prime}\left(S_n^L-\delta Q_n^{L,*,(j)}\right)\right)A_n^{(j+1)}\\ +B_n^{(j+1)\prime}\left(\gamma_0 I_n+\rho_0 W_n\right)\end{array}\right],
$$

and

$$
E_n^{(j+1)}=\left(R_n^{L,(j+1)}\right)^{-1}\left[\begin{array}{c}B_n^{(j+1)\prime}\left(I_n+\delta\pi_0 L_n^{L,(j)}\right)\\ +\left(\Phi_{n,0}-B_n^{(j+1)\prime}\left(S_n^L-\delta Q_n^{L,*,(j)}\right)\right)C_n^{(j+1)}\end{array}\right].
$$

C.2. Large sample properties

Consistency

In this subsection, we discuss proving Theorem 4.4.1. The first step is to verify the uniform convergence of sample average of the (concentrated) log-likelihood function. i.e., $\sup_{\theta\in\Theta}\left|\frac{1}{L}\ln L_{L,c}(\theta)-Q_L(\theta)\right|\to_p 0$ as $L\to\infty$. Note that σ^2 is bounded away from zero. Then, the main part of $\frac{1}{L}\ln L_{L,c}(\theta)-Q_L(\theta)$ is

$$
\frac{1}{L}\left\{\mathcal{E}_L'(\theta)\left(J_T\otimes J_n\right)\mathcal{E}_L(\theta)-\mathbf{E}\left[\mathcal{E}_L'(\theta)\left(J_T\otimes J_n\right)\mathcal{E}_L(\theta)\right]\right\}.
$$

Observe

$$
\begin{aligned}
&\left(J_T\otimes J_n\right)\mathcal{E}_L(\theta)\\
=\;&\left(J_T\otimes J_n\right)\left[\begin{array}{c}\left(I_T\otimes\left(R_n^e(\theta_1)\right)^{-1}R_n^F(\theta_1)\left(R_n^F\right)^{-1}\left(\gamma_0 I_n+\rho_0 W_n+\Phi_{n,0}D_n\right)\right)\\ -\left(I_T\otimes\left(R_n^e(\theta_1)\right)^{-1}\left(\gamma I_n+\rho W_n+\Phi_{n,0}D_n(\theta_1)\right)\right)\end{array}\right]Y_{L,-1}\\
&+\left(J_T\otimes J_n\right)\left(I_T\otimes\left(R_n^e(\theta_1)\right)^{-1}R_n^F(\theta_1)\left(R_n^F\right)^{-1}\right)\mathbf{X}_L\beta_0\\
&+\left(J_T\otimes J_n\right)\left(I_T\otimes\left(R_n^e(\theta_1)\right)^{-1}R_n^F(\theta_1)\left(R_n^F\right)^{-1}R_n^e\right)\left(\alpha_{L,0}+\mathcal{E}_L\right).
\end{aligned}
$$

Note that $\mathbf{c}_{L,0}$ is eliminated by the projector J_T. By Assumptions 4.4.1-4.4.5, all components in $(J_T \otimes J_n)\,\mathcal{E}_L(\theta)$ are bounded nonstochastic components or stochastic components with their bounded variances and their uniformly bounded linear transformations in n uniformly in $\theta \in \Theta$. As $L \to \infty$, it implies that

$$\frac{1}{L}\left\{\mathcal{E}_L'(\theta)\,(J_T \otimes J_n)\,\mathcal{E}_L(\theta) - \mathbf{E}\left[\mathcal{E}_L'(\theta)\,(J_T \otimes J_n)\,\mathcal{E}_L(\theta)\right]\right\} \to_p 0$$

uniformly in $\theta \in \Theta$.

The second step is showing uniform equicontinuity of $\{Q_L(\theta)\}$ in $\theta \in \Theta$. A starting point of showing uniform equicontinuity of $\{Q_L(\theta)\}$ is to utilize the Taylor approximation argument by continuous differentiability of $Q_L(\theta)$ obtained from Assumption 4.4.5 (i) and (ii). For θ_a, $\theta_b \in \Theta$, we can represent each component of $Q_L(\theta)$ by $(\theta_a - \theta_b) \cdot h_L\left(\bar{\theta}\right)$ where $\bar{\theta}$ lies between θ_a and θ_b and $h_L(\cdot)$ is uniformly bounded in L and in Θ. Then, we can guarantee for the smooth function class $\{Q_L(\theta)\}$ in $\theta \in \Theta$.

Model identification

To finish the proof of consistency, we need to achieve identification uniqueness. The key identification conditions will be derived based on the information inequality (Rothenberg (1971)). For this, two definitions in Rothenberg (1971) are reproduced.

Definition C.2.1 *For $\theta \in \Theta$, $L_L\left(\theta \middle| \{Y_{nt}\}_{t=0}^{T}\right)$ denotes the density function when we have data $\{Y_{nt}\}_{t=0}^{T}$.*

*(i) θ' and θ'' in Θ are **observationally equivalent** if*
$L_L\left(\theta' \middle| \{Y_{nt}\}_{t=0}^{T}\right) = L_L\left(\theta'' \middle| \{Y_{nt}\}_{t=0}^{T}\right)$ *a.e.*

*(ii) $\theta_0 \in \Theta$ is **identifiable** iff there is no observationally equivalent $\theta \in \Theta$.*

We apply Definition C.2.1 to our case, the concentrated log likelihood function $\ln L_{L,c}(\theta)$ since the information inequality argument is valid for the concentrated log-likelihood function. Even though our log-likelihood function is derived from normality on $\mathcal{E}_{nt}$, we will show that identification is also valid for the quasi log-likelihood estimation method (i.e., free to the normal distribution assumption). Then, unique identification of θ_0 is based on two relations: (i) $\mathbf{E}\left(\ln L_{L,c}\left(\theta\right)\right) \leq \mathbf{E}\left(\ln L_{L,c}\left(\theta_0\right)\right)$, and (ii) $L_{L,c}\left(\theta\right) = L_{L,c}\left(\theta_0\right)$ a.e. in $\{Y_{nt}\}_{t=0}^{T}$ iff $\mathbf{E}\left(\ln L_{L,c}\left(\theta\right)\right) = \mathbf{E}\left(\ln L_{L,c}\left(\theta_0\right)\right)$.

Recall that

$$\begin{aligned} Q_L\left(\theta_1, \beta, \sigma^2\right) &= -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 + \frac{1}{n}\ln\left|R_n^F\left(\theta_1\right)\right| \\ &\quad -\frac{1}{n}\ln\left|R_n^e\left(\theta_1\right)\right| - \frac{1}{2\sigma^2}\frac{1}{L}\mathbf{E}\left[\mathcal{E}_L'\left(\theta\right)\left(J_T\otimes J_n\right)\mathcal{E}_L\left(\theta\right)\right] \end{aligned}$$

where

$$\begin{aligned} \mathcal{E}_L\left(\theta\right) &= \left(I_T\otimes\left(R_n^e\left(\theta_1\right)\right)^{-1}R_n^F\left(\theta_1\right)\right)Y_L \\ &\quad -\left(I_T\otimes\left(R_n^e\left(\theta_1\right)\right)^{-1}\left(\gamma I_n + \rho W_n + \Phi_n\left(\phi\right)D_n\left(\theta_1\right)\right)\right)Y_{L,-1} \\ &\quad -\left(I_T\otimes\left(R_n^e\left(\theta_1\right)\right)^{-1}\right)\mathbf{X}_L\left(\theta_1\right)\beta. \end{aligned}$$

Note that $\mathbf{X}_L = \mathbf{X}_L\left(\theta_{1,0}\right)$. Let $\beta_L\left(\theta_1\right) = \arg\max_\beta Q_L\left(\theta_1, \beta, \sigma^2\right)$. Then,

$$\begin{aligned} \beta_L\left(\theta_1\right) &= \left[\mathbf{X}_L\left(\theta_1\right)'\left(I_T\otimes\left(R_n^e\left(\theta_1\right)\right)^{-1\prime}\right)\left(J_T\otimes J_n\right)\left(I_T\otimes\left(R_n^e\left(\theta_1\right)\right)^{-1}\right)\mathbf{X}_L\left(\theta_1\right)\right]^{-1} \\ &\quad \times\mathbf{X}_L\left(\theta_1\right)'\left(I_T\otimes\left(R_n^e\left(\theta_1\right)\right)^{-1\prime}\right)\left(J_T\otimes J_n\right)\mathbf{E}\left[\mathbf{Z}_L\left(\theta_1\right)\right] \end{aligned}$$

where

$$\begin{aligned} \mathbf{Z}_L\left(\theta_1\right) &= \left[\begin{array}{c} \left(I_T\otimes\left(R_n^e\left(\theta_1\right)\right)^{-1}R_n^F\left(\theta_1\right)\left(R_n^F\right)^{-1}\left(\gamma_0 I_n + \rho_0 W_n + \Phi_{n,0}D_n\right)\right) \\ -\left(I_T\otimes\left(R_n^e\left(\theta_1\right)\right)^{-1}\left(\gamma I_n + \rho W_n + \Phi_n\left(\phi\right)D_n\left(\theta_1\right)\right)\right) \end{array}\right]Y_{L,-1} \\ &\quad +\left(I_T\otimes\left(R_n^e\left(\theta_1\right)\right)^{-1}R_n^F\left(\theta_1\right)\left(R_n^F\right)^{-1}\right)\mathbf{X}_L\beta_0. \end{aligned}$$

If $\theta_1 = \theta_{1,0}$, $\mathbf{Z}_L\left(\theta_{1,0}\right) = \left(I_T\otimes\left(R_n^e\right)^{-1}\right)\mathbf{X}_L\beta_0$ and $\beta_L\left(\theta_{1,0}\right) = \beta_0$. Then, $\left(J_T\otimes J_n\right)\mathbf{E}\left[\mathbf{Z}_L\left(\theta_1\right) - \left(I_T\otimes\left(R_n^e\right)^{-1}\right)\mathbf{X}_L\beta_0\right]$ for $\theta_1 \in \Theta_1\backslash\left\{\theta_{1,0}\right\}$ is the main part of the misspecification error for $(J_T \otimes J_n)\left(I_T \otimes (R_n^e)^{-1}\right)\mathbf{X}_L\beta_0$. If identification for $\theta_{1,0}$ is done, we need to have the existence and positive definiteness of $\lim_{L\to\infty}\frac{1}{L}\mathbf{X}_L'\left(I_T \otimes (R_n^e)^{-1\prime}\right)(J_T \otimes J_n)\left(I_T \otimes (R_n^e)^{-1}\right)\mathbf{X}_L$. Observe that identification of β_0 does not rely on normality of $\mathcal{E}_{nt}$. It comes from (i) $\text{plim}_{L\to\infty}\hat{\theta}_{1,nT} = \theta_{1,0}$ (large sample) and (ii) sufficient variations in $(J_T \otimes J_n)\left(I_T \otimes (R_n^e)^{-1}\right)\mathbf{X}_L$. Using $\beta_L(\theta_1)$, we define $\mathcal{E}_L(\theta_1) = \mathcal{E}_L(\theta_1, \beta_L(\theta_1))$ for each $\theta_1 \in \Theta_1$.

Consider unique identification of $\theta_{1,0}$. For this, note that

$$\sigma_L^2(\theta_1) = \arg\max_{\sigma^2} Q_L\left(\theta_1, \beta_L(\theta_1), \sigma^2\right) = \frac{1}{L}\mathbf{E}\left[\mathcal{E}_L'(\theta_1)(J_T \otimes J_n)\mathcal{E}_L(\theta_1)\right].$$

Then, the concentrated expected log-likelihood function for θ_1 is

$$\begin{aligned} Q_{L,c}(\theta_1) &\equiv Q_L\left(\theta_1, \beta_L(\theta_1), \sigma_L^2(\theta_1)\right) \\ &= -\frac{1}{2}(\ln 2\pi + 1) - \frac{1}{2}\ln\sigma_L^2(\theta_1) + \frac{1}{n}\ln\left|R_n^F(\theta_1)\right| - \frac{1}{n}\ln\left|R_n^e(\theta_1)\right|. \end{aligned}$$

At $\theta_1 = \theta_{1,0}$, $Q_{L,c}(\theta_{1,0}) = -\frac{1}{2}(\ln 2\pi + 1) - \frac{1}{2}\ln\sigma_0^2 + \frac{1}{n}\ln\left|R_n^F\right| - \frac{1}{n}\ln\left|R_n^e\right|$. The identification uniqueness condition is derived by the sequence of nonstochastic functions $\{Q_{L,c}(\theta_1)\}$: for arbitrary $\varepsilon > 0$,

$$\limsup_{L\to\infty}\max_{\theta_1 \in \mathcal{N}^c(\theta_{1,0},\varepsilon)}\left[Q_{L,c}(\theta_1) - Q_{L,c}(\theta_{1,0})\right] < 0,$$

where $\mathcal{N}^c(\theta_{1,0}, \varepsilon)$ denotes the complement of an open neighborhood of Θ_1 of radius $\varepsilon > 0$. Consider

$$\begin{aligned} & Q_{L,c}(\theta_1) - Q_{L,c}(\theta_{1,0}) \\ = & -\left(\ln\sigma_L^2(\theta_1) - \ln\sigma_0^2\right) \\ & + \frac{1}{n}\left(\ln\left|R_n^F(\theta_1)\right| - \ln\left|R_n^F\right|\right) - \frac{1}{n}\left(\ln\left|R_n^e(\theta_1)\right| - \ln\left|R_n^e\right|\right) \\ = & \frac{1}{n}\ln\left|\sigma_0\left(R_n^F\right)^{-1}R_n^e\right| - \frac{1}{n}\ln\left|\sigma_L(\theta_1)\left(R_n^F(\theta_1)\right)^{-1}R_n^e(\theta_1)\right| \end{aligned}$$

$$= \frac{1}{2}\left\{\begin{array}{c} \frac{1}{n}\ln\left|\sigma_0^2\left(R_n^F\right)^{-1}R_n^eR_n^{e\prime}\left(R_n^F\right)^{-1\prime}\right| \\ -\frac{1}{n}\ln\left|\sigma_L^2\left(\theta_1\right)\left(R_n^F\left(\theta_1\right)\right)^{-1}R_n^e\left(\theta_1\right)R_n^e\left(\theta_1\right)'\left(R_n^F\right)^{-1\prime}\left(R_n^F\left(\theta_1\right)\right)^{-1\prime}\right| \end{array}\right\}.$$

Under large L, the unique identification condition for $\theta_{1,0}$ is achieved if

$$\lim_{L\to\infty}\left\{\begin{array}{c} \frac{1}{n}\ln\left|\sigma_0^2\left(R_n^F\right)^{-1}R_n^eR_n^{e\prime}\left(R_n^F\right)^{-1\prime}\right| \\ -\frac{1}{n}\ln\left|\sigma_L^2\left(\theta_1\right)\left(R_n^F\left(\theta_1\right)\right)^{-1}R_n^e\left(\theta_1\right)R_n^e\left(\theta_1\right)'\left(R_n^F\right)^{-1\prime}\left(R_n^F\left(\theta_1\right)\right)^{-1\prime}\right| \end{array}\right\}\neq 0 \tag{C.2}$$

for $\theta_1 \neq \theta_{1,0}$ is satisfied. Observe that unique identification condition for $\theta_{1,0}$ by (C.2) does not rely on the normal distribution assumption on $\mathcal{E}_{nt}$.

Asymptotic normality

As the intermediate procedures of deriving the asymptotic distribution of $\hat{\theta}_L$, we report the components of $\frac{1}{\sqrt{L}}\frac{\partial \ln L_{L,c}(\theta_0)}{\partial\theta}$ and asymptotic bias parts.

First-order derivatives of the concentrated log-likelihood function

For notational convenience, we define $F_n(\theta_1)$ satisfying $\mathbf{X}_{nt}(\theta_1) = F_n(\theta_1)X_{nt}$ for each $\theta_1 \in \Theta$. At $\theta_{1,0}$, we denote $F_n = F_n(\theta_1)$. Observe that

$$\begin{aligned}(J_T\otimes J_n)Y_L &= (J_T\otimes J_n)\left(I_T\otimes\left(R_n^F\right)^{-1}(\gamma_0 I_n+\rho_0 W_n+\Phi_{n,0}D_n)\right)Y_{L,-1} \\ &\quad +(J_T\otimes J_n)\left(I_T\otimes\left(R_n^F\right)^{-1}F_n\right)X_L\beta_0 \\ &\quad +(J_T\otimes J_n)\left(I_T\otimes\left(R_n^F\right)^{-1}R_n^e\right)(\alpha_{L,0}+\mathcal{E}_L).\end{aligned}$$

A useful formula here is $\frac{\partial\ln|A_n(x)|}{\partial x} = tr\left(A_n^{-1}(x)\frac{\partial A_n(x)}{\partial x}\right)$ and $\frac{\partial A^{-1}(x)}{\partial x} = -A^{-1}(x)\frac{\partial A(x)}{\partial x}A^{-1}(x)$. A subscript to each each parameter value represents a partial derivative, e.g., $R_{n,\lambda}^e(\theta_1) = \frac{\partial R_n^e(\theta_1)}{\partial\lambda}$. Then,

$$\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \lambda} = \begin{array}{l} \frac{1}{\sigma_0^2}\left[\begin{array}{c} I_T \otimes \left(\begin{array}{c} -\left[(R_n^e)^{-1} R_n^F\right]_\lambda \left(R_n^F\right)^{-1}(\gamma_0 I_n + \rho_0 W_n + \Phi_{n,0} D_n) \\ +\left[(R_n^e)^{-1}(\gamma_0 I_n + \rho_0 W_n + \Phi_{n,0} D_n)\right]_\lambda \end{array}\right) Y_{L,-1} \\ +\left(I_T \otimes -\left[(R_n^e)^{-1} R_n^F\right]_\lambda \left(R_n^F\right)^{-1} F_n + \left[(R_n^e)^{-1} F_n\right]_\lambda\right) X_L \beta_0 \\ +\left(I_T \otimes -\left[(R_n^e)^{-1} R_n^F\right]_\lambda \left(R_n^F\right)^{-1} R_n^e\right)\alpha_{L,0} \end{array}\right]' , \\ \cdot (J_T \otimes J_n)\mathcal{E}_L \\ +\frac{1}{\sigma_0^2}\left[\begin{array}{c} \mathcal{E}_L'\left(I_T \otimes -R_n^{e\prime}\left(R_n^F\right)^{-1\prime}\left[(R_n^e)^{-1} R_n^F\right]_\lambda'\right)(J_T \otimes J_n)\mathcal{E}_L \\ -T\sigma_0^2 tr\left(-\left(R_n^F\right)^{-1} R_{n,\lambda}^F\right) + T\sigma_0^2 tr\left(-(R_n^e)^{-1} R_{n,\lambda}^e\right) \end{array}\right] \end{array}$$

$$\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \beta} = \frac{1}{\sigma_0^2} X_L'\left(I_T \otimes \left(R_n^F\right)^{-1} F_n\right)'(J_T \otimes J_n)\mathcal{E}_L,$$

$$\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \sigma^2} = \frac{1}{2\sigma_0^4}\left[\mathcal{E}_L'(J_T \otimes J_n)\mathcal{E}_L - L\sigma_0^2\right].$$

Other parts, $\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \gamma}$, $\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \rho}$, and $\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \phi}$, take the similar form to $\frac{\partial \ln L_{L,c}(\theta_0)}{\partial \lambda}$ (i.e., a LQ form of $\mathcal{E}_L$).

Asymptotic distribution of $\hat{\theta}_L$

Deriving the asymptotic distribution of $\hat{\theta}_L$ follows the conventional argument for an extremum estimator. The argument is almost similar to that of Jeong and Lee (2018). Here is a sketch of the argument. By the Taylor expansion, we have

$$\sqrt{L}\left(\hat{\theta}_L - \theta_0\right) = \left(-\frac{1}{L}\frac{\partial^2 \ln L_{L,c}\left(\bar{\theta}_L\right)}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{1}{\sqrt{L}}\frac{\partial \ln L_{L,c}^{(u)}(\theta_0)}{\partial\theta} - \Delta_{1,L} - \Delta_{2,L}\right) \quad \text{(C.3)}$$

where $\bar{\theta}_L$ lies between θ_0 and $\hat{\theta}_L$. Since $\hat{\theta}_L \to_p \theta_0$ and $-\frac{1}{L}\frac{\partial^2 \ln L_{L,c}(\theta)}{\partial\theta\partial\theta'}$ is a continuous function of θ, we have $-\frac{1}{L}\frac{\partial^2 \ln L_{L,c}\left(\bar{\theta}_L\right)}{\partial\theta\partial\theta'} - \Sigma_{\theta_0,L} = o_p(1)$. By Lemmas 2.1 and 2.2 in the supplementary file of Jeong and Lee (2018), $\Delta_{1,L} = \sqrt{\frac{n}{T}}a_{n,1}(\theta_0) + o_p(1)$ and $\Delta_{2,L} = \sqrt{\frac{T}{n}}a_{n,2}(\theta_0)$. Hence, (C.3) can be rewritten as

$$\sqrt{L}\left(\hat{\theta}_L - \theta_0\right) = \left(\Sigma_{\theta_0,L} + o_p(1)\right)^{-1}\left(\begin{array}{c} \frac{1}{\sqrt{L}}\frac{\partial \ln L_{L,c}^{(u)}(\theta_0)}{\partial\theta} - \sqrt{\frac{n}{T}}a_{n,1}(\theta_0) \\ -\sqrt{\frac{T}{n}}a_{n,2}(\theta_0) + o_p(1) \end{array}\right) \quad \text{(C.4)}$$

$$\Leftrightarrow \quad \sqrt{L}\left(\hat{\theta}_L - \theta_0\right) + \sqrt{\frac{n}{T}}\Sigma_{\theta_0,L}^{-1}a_{n,1}(\theta_0) + \sqrt{\frac{T}{n}}\Sigma_{\theta_0,L}^{-1}a_{n,2}(\theta_0) + o_p(1)$$

$$= \quad \Sigma_{\theta_0,L}^{-1} \frac{1}{\sqrt{L}} \frac{\partial \ln L_{L,c}^{(u)}(\theta_0)}{\partial \theta}.$$

Note that (C.4) implies that $\hat{\theta}_L - \theta_0 = O_p\left(\max\left\{\frac{1}{\sqrt{L}}, \frac{1}{n}, \frac{1}{T}\right\}\right)$ (i.e., convergence rate of $\hat{\theta}_L$).

The last part is to characterize the asymptotic distribution of $\frac{1}{\sqrt{L}} \frac{\partial \ln L_{L,c}^{(u)}(\theta_0)}{\partial \theta}$. We observe that the components in $\frac{1}{\sqrt{L}} \frac{\partial \ln L_{L,c}^{(u)}(\theta_0)}{\partial \theta}$ take a LQ form of $\mathcal{E}_L$. By applying the martingale CLT for a LQ form (Yu et al. (2008)), we have $\frac{1}{\sqrt{L}} \frac{\partial \ln L_{L,c}^{(u)}(\theta_0)}{\partial \theta} \rightarrow_d N(0, \Omega_{\theta_0})$. By the Slutsky's lemma, we can finish the argument.

Bibliography

[1] Amemiya, T. 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

[2] Anderson, J, and E. van Wincoop 2003. Gravity with Gravitas: A Solution to the Border Puzzle. American Economic Review 93, 170-192.

[3] Anselin, L. 1988. Spatial Econometrics: Methods and Models. Kluwer Academic, The Netherlands.

[4] Ballester, C., A. Calvō-Armengol, and Y. Zenou 2006. Who's who in networks. wanted: the key player. Econometrica 74, 1403-1417.

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

[6] Bai, J. 2009. Panel data models with interactive fixed effects. Econometrica 77 (4), 1229-1279.

[7] Barro, R. J. 1990. Government spending in a simple model of endogenous growth. Journal of Political Economy 98 (5), 103-125.

[8] Benigno, P., and M. Woodford 2012. Linear-quadratic approximation of optimal policy problems, Journal of Economic Theory 147, 1-42.

[9] Blackwell, D. 1965. Discounted dynamic programming. Annals of Mathematical Statistics 36, 226-235.

[10] Bloom, D. E., D. Canning, and J. Sevilla 2004. The effect of health on economic growth: a production function approach. World Development 32 (1), 1-13.

[11] Bramoullé, Y, R. Kranton, and M. D'Amours 2014. Strategic Interaction and Networks. American Economic Review, 104 (3), 898-930.

[12] Brett, C, and J. Pinkse 2000. The determinants of municipal tax rates in British Columbia. Canadian Journal of Economics, 33 (3), 695-714.

[13] Brueckner, J. K. 2003. Strategic interaction among governments: an overview of empirical studies. International Regional Science Review 26 (2), 175-188.

[14] Calvō-Armengol A., E. Patacchini, and Y. Zenou 2009. Peer effects and social networks in education. Review of Economic Studies 76, 1239-1267.

[15] Case, A.C., H.S. Rosen, and J.R. Hines, Jr 1993. Budget spillovers and fiscal policy interdependence: evidence from the states. Journal of Public economics 52, 285-307.

[16] Cliff, A.D. and J.K. Ord 1973. Spatial Autocorrelation. London: Pion Ltd.

[17] Davidson, J. 1994. Stochastic Limit Theory. Oxford: Oxford University Press.

[18] de Paula, A. 2013. Econometric analysis of games with multiple equilibria. Annual Review of Economics 5, 107–131.

[19] Engsted, T., and N. Haldrup 1994. The linear quadratic adjustment cost model and the demand for labor. Journal of Applied Econometrics 9, 145-159.

[20] Figlio, D. N., V. W. Kolpin, and W. E. Reid 1999. Do states play welfare games? Journal of Urban Economics 46, 437-454.

[21] Fuente, A. 2000. Mathematical methods and models for economists. Cambridge University Press, United Kingdom

[22] Fuhrer, J.C. 2000. Habit formation in consumption and its implications for monetary-policy models. American Economic Review 90, 367-390.

[23] Goldsmith-Pinkham, P., and G.W. Imbens 2013. Social networks and the identification of peer effects. Journal of Business and Economic Statistics 31 (3), 253-264.

[24] Hahn J. and G. Kuersteiner 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. Econometrica 70, 1639-1657.

[25] Han, X., and L.F. Lee 2016. Bayesian analysis of spatial panel autoregressive models with time-varying endogenous spatial weight matrices, common factors, and random coefficients. Journal of Business and Economic Statistics 34 (4), 642-660.

[26] Han, X., Chih-Sheng Hsieh, and S.I.M Ko 2019. Spatial modeling approach for dynamic network formation and interactions. Journal of Business and Economic Statistics, forthcoming.

[27] Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029-1054.

[28] Hansen, L.P., and K.J. Singleton 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50, 1269-1286.

[29] Hsieh, Chih-Sheng and L.F. Lee 2017. Specification and estimation of network formation and network interaction models with the exponential probability distribution. working paper.

[30] Jackson, O., and Y. Zenou 2014. Games on network. Handbook of Game Theory Vol. 4.

[31] Jenish, N., and I. R. Prucha 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. Journal of Econometrics 150, 86-98.

[32] Jenish, N., and I. R. Prucha 2012. On spatial processes and asymptotic inference under near-epoch dependence. Journal of Econometrics 170, 178-190.

[33] Jeong, H., and L.F. Lee 2018. Spatial dynamic models with intertemporal strategic interactions: specification and estimation. The Ohio State University, working paper.

[34] Johnsson, I., and H. R. Moon 2017. Estimation of peer effects in endogenous social networks: control function approach. working paper.

[35] Judd, K. L. 1996. Approximation, Perturbation, and Projection Methods in Economic Analysis. In H. Amman, D. Kendrick, & J. Rust, (Eds.), Handbook of computational economics. Amsterdam: Elsevier.

[36] Judd, K. L. 1998. Numerical Methods in Economics. Cambridge. MIT.

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

[38] Kelejian, H., and I. R. Prucha 2001. On the asymptotic distribution of the Moran I test statistic with applicatons. Journal of Econometrics 104, 219-257.

[39] Kelejian, H., and Piras, G. 2014. Estimation of spatial models with endogenous spatial weighting matrices, and an application to a demand model for cigarettes. Regional Science and Urban Economics 46, 140-149.

[40] Kennan, J. 1979. The estimation of partial adjustment models with rational expectations. Econometrica 47, 1441-1455.

[41] Komarova, T., F. Sanches, and D.S. Junior 2017. Joint analysis of the discount factor and payoff parameters in dynamic discrete choice models. working paper.

[42] Kuersteiner, G. M., and I.R. Prucha 2013. Limit theory for panel data models with cross sectional dependence and sequential exogeneity. Journal of Econometrics 174, 107-126.

[43] Kuersteiner, G. M., and I.R. Prucha 2018. Dynamic spatial panel models: networks, common shocks, and sequential exogeneity. University of Maryland, working paper.

[44] Kydland, F. E., and E. C. Prescott 1982. Time to build and aggregate fluctuations. Econometrica 50, 1345-1370.

[45] Lee, L.F. 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. Econometrica 72, 1899-1925.

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

[47] Lee, S.Y. 2008. Three essays on spatial econometrics and empirical industrial organization, PhD Thesis, The Ohio State University (2008).

[48] Lee, L.F., and J. Yu 2010. A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory 26, 564-597.

[49] Lee, L.F., and J. Yu 2012. QML estimation of spatial dynamic panel data models with time varying spatial weights matrices. Spatial Economic Analysis 7 (1), 31-74.

[50] Lee, L.F., and J. Yu 2014. Efficient GMM estimation of spatial dynamic panel data models with fixed effects. Journal of Econometrics 180, 174-197.

[51] Lee, L.F., and J. Yu 2015. Spatial panels. Chapter 5 in The Oxford Handbook of Panel Data, edited by Badi H. Baltagi, Oxford University Press, Oxford.

[52] LeSage, J.P., Y-Y Chih, and C. Vance 2019. Markov Chain Monte Carlo estimation of spatial dynamic panel models for large samples. Computational Statistics and Data Analysis, forthcoming.

[53] LeSage, J., and R. K. Pace 2009. Introduction to Spatial Econometrics. Boca Raton, FL, Chapman and Hall/CRC.

[54] Li, H., C.A. Calder, and N. Cressie 2007. Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. Geographical Analysis 39, 357-375.

[55] Li, T., and S.P. Sethi 2017. A review of dynamic Stakelberg game models. Discrete and Continuous Dynamical Systems Series B 22 (1), 125-159.

[56] Lucas, R. 1976. Econometric policy evaluation: a critique. In Brunner, K.; Meltzer, A. The Phillips curve and labor markets. Carnegie-Rochester conference series on public policy. 1. New York: American Elsevier. pp. 19-46. ISBN 0-444-11007-0.

[57] Liu, X., J. H. Kagel and L.F. Lee 2012. Learning from peers in signaling game experiments. Journal of Applied Econometrics 27, 1037-1058.

[58] Ljungqvist, L., and T.J. Sargent 2012. Recursive Macroeconomic Theory. Third edition, The MIT Press.

[59] Magill, M 1977. A local analysis of N-sector capital accumulation under uncertainty. Journal of Economic Theory 15, 211-218.

[60] Maskin, E., and J. Tirole 1988a. A theory of dynamic oligopoly, I: Overview and quantity competition with large fixed costs. Econometrica 56 (3), 549-569.

[61] Maskin, E., and J. Tirole 1988b. A theory of dynamic oligopoly, II: Price competition, kinked demand curves, and edgeworth cycles. Econometrica 56 (3), 571-599.

[62] Maskin, E., and J. Tirole 2001. Markov perfect equilibrium I. Observable Actions. Journal of Economic Theory 100 (2), 191-219.

[63] Milgrom P, and Roberts J. 1982. Limit pricing and entry under incomplete information: an equilibrium analysis. Econometrica 50, 443-459.

[64] Moran, P.A.P. 1948. The interpretation of statistical maps. Biometrika 35, 255-260.

[65] Murphy, K. M. and Topel, R. H. 1985. Estimation and inference in two-step econometric models. Journal of Business & Economic Statistics 3, 370–379.

[66] Ord, J.K. 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120-126.

[67] Pagan, A. 1984. Econometric issues in the analysis of regressions with generated regressors. International Economic Review 25 (1), 221-247.

[68] Pötscher, B.M., and Prucha, I.R. 1997. Dynamic Nonlinear Econometric Models. Springer-Verlag, New York.

[69] Qu, X. and L.F. Lee 2015. Estimating a spatial autoregressive model with an endogenous spatial weight matrix. Journal of Econometrics 184, 209–232.

[70] Qu, X. and L.F. Lee 2018. Estimation of a SAR model with endogenous spatial weights constructed by bilateral variables. working paper.

[71] Qu, X., L.F. Lee, and J. Yu 2017. QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. Journal of Econometrics 197, 173-201.

[72] Revelli, Federico 2005. On spatial public finance empirics", International Tax and Public Finance 12, 475-492.

[73] Rothenberg, T. J. 1971. Identification in parametric models, Econometrica 39 (3), 577-591.

[74] Rubinstein, A. 1998. Modeling Bounded Rationality. MIT Press, Cambridge, Massachusetts.

[75] Santos, M. S. 1991. Smoothness of the policy function in discrete time economic models. Econometrica 59 (5), 1365-1382.

[76] Shi, W., and L.F. Lee 2017. Spatial dynamic panel data models with interactive fixed effects. Journal of Econometrics 197, 323-347.

[77] Stewart, G. 1998. Matrix Algorithms: Basic Decompositions. Society for Industrial and Applied Mathematics, Philadelphia.

[78] Stokey, N. L., Lucas, R. E. and Prescott E. C. 1989. Recursive Methods in Economic Dynamics. Cambridge, MA: Havard University Press.

[79] Ushchev, P., and Y. Zenou 2018. Social Norms in Networks. working paper.

[80] Xu, X., and L.F. Lee 2015. Maximum likelihood estimation of a spatial autoregressive Tobit model. Journal of Econometrics 188, 264-280.

[81] Xu, X., and L.F. Lee 2016. Estimation of a binary choice game with network links. working paper.

[82] Xu, X., and L.F. Lee 2018. Theoretical foundations for spatial econometric research. Regional Science and Urban Economics, forthcoming.

[83] Yang, C. and L.F. Lee 2017. Strategical interactions on municipal public safety spending with correlated private information. Regional Science and Urban Economics, forthcoming.

[84] Yu, J., R. de Jong, and L.F. Lee 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. Journal of Econometrics 146, 118-134.