Spatial dynamic models with intertemporal optimization: specification and estimation*

Hanbat Jeong[†] and Lung-fei Lee[‡]

August, 2018

Abstract

In this paper, we introduce 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 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 conduct impulse response and welfare analyses corresponding to changing exogenous characteristics in a region.

Keywords: Dynamic interaction, Intertemporal optimization, Markov perfect equilibrium, SDPD models, QML estimation, Welfare and counterfactual analysis, County's public safety spending

JEL classification: C33, C51, C57

*We would like to thank Professors In Choi, Jason Blevins, Robert de Jong, and Yongcheol Shin for their valuable comments and suggestions. We also thank the audience for comments for our presentations in the Asian Meeting this year in Sogang University, Seoul, Korea, and the China Meeting in Fudan University, Shanghai, China.

[†]Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. E-mail: jeong.181@osu.edu

[‡]Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. E-mail: lee.1777@osu.edu

1 Introduction and motivation

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).[1] 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 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.

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

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.[2]

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 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

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

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.[3] 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.

The paper is organized as follows. Section 2 introduces an economic foundation for our model. In Section 3, we build an econometric model based on the theoretical setting in Section 2. Section 4 presents the QML estimation method and asymptotic properties of that estimator. Section 5 reports our investigation on the

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

finite sample performance of the QML estimator. In Section 6, we apply our model to counties' public safety spending competition. Section 7 concludes. Some detailed derivations of estimating equations and asymptotic analysis of estimation are relegated in Appendices.[4]

Notation and convention : Let $A_n = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$ be an $n \times n$ square matrix. For any $n \times m$ matrix B, $[B]_{ij}$ denotes the (i,j) element of B. We denote the i^{th} unit column vector $(0, \cdots, 1, \cdots, 0)'$ as e_i. For any vector a, $[a]_i$ denotes the i^{th} component. For any $n \times n$ matrix A_n, $Diag(A_n)$ is a diagonal matrix formed by the diagonal of A_n, $\|A_n\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{n} |A_{ij}|$ is its maximum column sum norm, and $\|A_n\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |A_{ij}|$ is its row sum norm. In addition, the spectral norm of A_n is $\|A_n\|_2 = \sqrt{\phi_1(A_n' A_n)}$ where $\phi_1(A_n' A_n)$ denotes the largest eigenvalue of $A_n' A_n$, i.e., $\|A_n\|_2$ is the largest singular value of A_n.[5] Our asymptotic analyses on estimation of Section 4 are based on a large number of time periods T and a large number of cross sectional units n, unless otherwise specified. Convergence in probability and convergence in distribution are denoted respectively as $\xrightarrow{p}$ and $\xrightarrow{d}$.

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.

2.1 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

[4]Due to space limitation, some of the analyses are provided in a supplementary file.

[5]Those matrix norms are induced by corresponding vector norms.

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

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

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 $\boldsymbol{\eta}_n = (\eta_1, \cdots, \eta_n)'$, $X_n = (x_1, \cdots, x_n)'$ 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 $\boldsymbol{\eta}_n$ as a regression function, $\boldsymbol{\eta}_n = X_n\beta_0 + \mathcal{E}_n$, agents' optimized decisions in a perfect information game give rise to the spatial autoregressive (SAR) model

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + \mathcal{E}_n \tag{2}$$

where Y_n is the vector of Nash equilibrium (NE). The system (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) but with additional time lags $Y_{n,t-1}$, diffusion $W_n Y_{n,t-1}$ and individual and time fixed effects:

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + \mathcal{E}_{nt} \tag{3}$$

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

$$u_i\left(Y_{nt}, Y_{n,t-1}, \eta_{it}\right) = \eta_{it} y_{it} + \lambda_0 y_{it} w_{i.} Y_{nt} + \rho_0 y_{it} w_{i.} Y_{n,t-1} - c\left(y_{it}, y_{i,t-1}\right) \tag{4}$$

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 $0 < \gamma_0 < 1$.[6] The η_{it} denotes the $t^{th}-$period index of heterogeneity of agent i containing those exogenous characteristics, which might evolve over time.[7] 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' past decisions and accommodate to the new environment innovated by new technologies.[8] In this setting, individual's learning comes from his/her recent past neighboring decisions.[9] 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.[10] In our

[6] 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

$$\frac{1}{\gamma_{1,0} + \gamma_{2,0}} u_i\left(Y_{nt}, Y_{n,t-1}, \eta_{it}\right) = \tilde{\eta}_{it} y_{it} + \tilde{\lambda}_0 y_{it} w_{i.} Y_{nt} + \tilde{\rho}_0 y_{it} w_{i.} Y_{n,t-1} - \tilde{c}\left(y_{it}, y_{i,t-1}\right)$$

and $\tilde{c}\left(y_{it}, y_{i,t-1}\right) = \frac{\tilde{\gamma}_0}{2}\left(y_{it} - y_{i,t-1}\right)^2 + \frac{1-\tilde{\gamma}_0}{2} y_{it}^2$ where $\tilde{\eta}_{it} = \frac{1}{\gamma_{1,0}+\gamma_{2,0}} \eta_{it}$, $\tilde{\lambda}_0 = \frac{\lambda_0}{\gamma_{1,0}+\gamma_{2,0}}$, $\tilde{\rho}_0 = \frac{\rho_0}{\gamma_{1,0}+\gamma_{2,0}}$, and $\tilde{\gamma}_0 = \frac{\gamma_0}{\gamma_{1,0}+\gamma_{2,0}}$ (i.e., structural parameters are normalized by $\frac{1}{\gamma_{1,0}+\gamma_{2,0}}$).

[7] 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.

[8] 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.

[9] 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.

[10] In this paper, we adopt the specification of the quadratic adjustment cost (the famous study about that is Kennan (1979)).

framework, $c(y_{it}, y_{i,t-1})$ consists of two parts: (i) dynamic adjustment cost, $\frac{\gamma_0}{2}(y_{it} - y_{i,t-1})^2$, 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 $\frac{\gamma_0}{2}(y_{it} - y_{i,t-1})^2$ 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' 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 $u_i(y_{it}, Y_{-i,t}, Y_{n,t-1}, \eta_{it})$ where $Y_{-i,t} = (y_{1t}, \cdots, y_{i-1,t}, y_{i+1,t}, \cdots, y_{nt})'$. It means that agent i knows the optimum choices $Y_{-i,t}$ of others. The first order conditions of such optimization problems give equation (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.[11] Under the linear-quadratic payoff (4), this will result in a new spatial dynamic panel data (SDPD) model.

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

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

Alternatively, Engsted and Haldrup (1994) employ the following quadratic adjustment cost for analyzing the demand for labor,

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

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.

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

where $\sigma(\cdot)$ denotes the $\sigma-$field[12] 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 $\boldsymbol{\eta}_{nt} = (\eta_{1t}, \cdots, \eta_{nt})'$ for each t. Given $(Y_{n0}, \boldsymbol{\eta}_{n1})$, 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

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

where $A_n^{trad} = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$ 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}.[13] Let $\Delta_n = Diag(A_n^{trad})$. Then, the NE vector at $t = 1$ can be characterized by a modified SAR equation:

$$\begin{aligned} Y_{n1}^*(Y_{n0}, \boldsymbol{\eta}_{n1}) &= \lambda_0 W_n Y_{n1}^*(Y_{n0}, \boldsymbol{\eta}_{n1}) + \delta\left[\Delta_n A_n^{trad} - \gamma_0 I_n\right] Y_{n1}^*(Y_{n0}, \boldsymbol{\eta}_{n1}) + (\gamma_0 I_n + \rho_0 W_n) Y_{n0} \\ &\quad + \boldsymbol{\eta}_{n1} + \delta \Delta_n S_n^{-1} E_1(\boldsymbol{\eta}_{2n}) \end{aligned}$$

where $E_t(\cdot)$ denotes the mathematical conditional expectation on $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ 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

$$Y_{n1}^*(Y_{n0}, \boldsymbol{\eta}_{n1}) = R_{n1}^{-1}(\gamma_0 I_n + \rho_0 W_n) Y_{n0} + R_{n1}^{-1}\left(\boldsymbol{\eta}_{n1} + \delta \Delta_n S_n^{-1} E_1(\boldsymbol{\eta}_{2n})\right). \quad (5)$$

From equation (5), we see that taking into account the expected outcomes in the second period, as $\delta > 0$, it brings in the additional spatial influence $\delta \Delta_n A_n^{trad} Y_{n1}^*(Y_{n0}, \boldsymbol{\eta}_{n1})$ and the time influence $\delta\gamma_0 I_n$ due to their effects on possible future outcomes.

[12] 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.

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

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 $\boldsymbol{\eta}_{nt} = (\eta_{1t}, \cdots, \eta_{nt})'$, each agent, say i, is assumed to maximize the expected discounted intertemporal payoff

$$u_i(y_{it}, Y_{-i,t}, Y_{n,t-1}, \eta_{it}) + \sum_{s=1}^{\infty} \delta^s E\left(u_i\left(Y_{n,t+s}, Y_{n,t+s-1}, \eta_{i,t+s}\right)|\mathcal{B}_{it}\right) \tag{6}$$

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.[14]

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 $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ 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).[15] Since the $t^{th}-$period optimal decisions only depend on $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ and, under the Markov assumption other past histories and exogenous characteristics are irrelevant to the current decision-making, $E(\cdot|\mathcal{B}_{it}) = E(\cdot|Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ for all $i = 1, \cdots, n$. Hence, we can simply define the conditional expectation operator $E_t(\cdot)$ by $E_t(\cdot) = E(\cdot|Y_{n,t-1}, \boldsymbol{\eta}_{nt})$. Also, time itself is not payoff-relevant, so we can drop the subscript "t" from agents' optimal policy functions $y_{it}^*(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ (for $i = 1, \cdots, n$) in the definition of MPE.

[14] 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.

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

Definition 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) $y^*_{it}(Y_{n,t-1},\boldsymbol{\eta}_{nt}) = f_i(Y_{n,t-1},\boldsymbol{\eta}_{nt})$,

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

$$V_i(Y_{n,t-1},\boldsymbol{\eta}_{nt}) = \max_{y_{it}} \left\{ u_i\left(y_{it}, Y^*_{-i,t}(Y_{n,t-1},\boldsymbol{\eta}_{nt}), Y_{n,t-1}, \eta_{it}\right) + \delta E_t\left(V_i(y_{it}, Y^*_{-i,t}(Y_{n,t-1},\boldsymbol{\eta}_{nt}), \boldsymbol{\eta}_{n,t+1})\right)\right\} \quad (7)$$

where $Y^*_{-i,t}(Y_{n,t-1},\boldsymbol{\eta}_{nt}) = \Big(y^*_{1t}(Y_{n,t-1},\boldsymbol{\eta}_{nt}),\cdots,y^*_{i-1,t}(Y_{n,t-1},\boldsymbol{\eta}_{nt}), y^*_{i+1,t}(Y_{n,t-1},\boldsymbol{\eta}_{nt}),\cdots,y^*_{nt}(Y_{n,t-1},\boldsymbol{\eta}_{nt})\Big)'$, *and*

(iii) (principle of optimality) the policy function $f_i(\cdot) = y^*_{it}(\cdot)$ *attains the right side of the Bellman equation (7).*

The principle of optimality characterizes the equivalent relationship between the two solutions to the intertemporal choice problem (6) and the functional equation (7). In other words, given $(Y_{n,t-1},\boldsymbol{\eta}_{nt})$,

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

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

Since payoff (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

$$V_i(Y_{n,t-1},\boldsymbol{\eta}_{nt}) = Y'_{n,t-1}Q_iY_{n,t-1} + Y'_{n,t-1}L_i\boldsymbol{\eta}_{nt} + \boldsymbol{\eta}'_{nt}G_i\boldsymbol{\eta}_{nt} + c_i \quad (8)$$

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.[16] To have a well-defined Bellman equation (a recursive relationship), $V_i(\cdot)$ should be a continuous and bounded

[16]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 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).

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 (7) can be characterized by using the maximum operator $\mathcal{T}$:

$$\begin{aligned} V_i(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) &= \mathcal{T}(V_i)(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) \\ &= \max_{y_{it}} \left\{ u_i\left(y_{it}, Y^*_{-i,t}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}), Y_{n,t-1}, \eta_{it}\right) + \delta E_t\left(V_i(y_{it}, Y^*_{-i,t}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}), \boldsymbol{\eta}_{t+1})\right)\right\}, \end{aligned} \tag{9}$$

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.[17] For this, define

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

Assumption 2.1 *We assume*

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

Under Assumption 2.1 (i), we have a linear expectation $E_t\left(\boldsymbol{\eta}^v_{n,t+1}\right) = E\left(\boldsymbol{\eta}^v_{n,t+1} | \boldsymbol{\eta}^v_{nt}\right) = \Pi_n \boldsymbol{\eta}^v_{nt}$ and other parts of histories (e.g., $\boldsymbol{\eta}^v_{n,t-1}$, $\boldsymbol{\eta}^v_{n,t-2}$, $\cdots$) are not relevant.[18] 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 $\boldsymbol{\eta}_{n,t+1}$ 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 2.1 (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.[19]

[17] The detailed arguments can be found in Appendix A and our supplementary file.

[18] 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.

[19] 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.

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.[20] 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

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

For our model, because the payoff function $u_i(\cdot)$ is a linear-quadratic form in Y_{nt} and $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$, 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,

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

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

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

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

$$Y_{nt}^*(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) = (\lambda_0 W_n + \delta Q_n^*) Y_{nt}^*(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) + (\gamma_0 I_n + \rho_0 W_n) Y_{n,t-1} + (I_n + \delta L_n^* \Pi_n) \boldsymbol{\eta}_{nt}, \qquad (10)$$

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

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

$$R_n = S_n - \delta Q_n^*, \text{ where } S_n = I_n - \lambda_0 W_n. \tag{11}$$

So the NE activity vector at time t is

$$Y_{nt}^* (Y_{n,t-1}, \boldsymbol{\eta}_{nt}) = A_n Y_{n,t-1} + B_n \boldsymbol{\eta}_{nt} \tag{12}$$

where $A_n = R_n^{-1} (\gamma_0 I_n + \rho_0 W_n)$ and $B_n = R_n^{-1} (I_n + \delta L_n^* \Pi_n)$. 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.[21] 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 L_n^* can be represented by

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

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.[22] 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^*:

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

which implies

$$\begin{aligned} Y_{nt}^* (Y_{n,t-1}, \boldsymbol{\eta}_{nt}) &= (\lambda_0 W_n + \delta D_{n,1} (\gamma_0 I_n + \rho_0 W_n) - \delta \gamma_0 I_n) Y_{nt}^* (Y_{n,t-1}, \boldsymbol{\eta}_{nt}) \\ &\quad + (\gamma_0 I_n + \rho_0 W_n) Y_{n,t-1} + \left(I_n + \sum_{m=1}^{\infty} \delta^m D_{n,m} \Pi_n^m \right) \boldsymbol{\eta}_{nt} \end{aligned} \tag{13}$$

[21] For illustrative purposes, suppose there is no isolated spatial unit. Then, all elements in Q_n^* are nonzero. In our system equation (10), 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.

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

and

$$R_n = (1+\delta\gamma_0) I_n - \lambda_0 W_n - \delta D_{n,1} (\gamma_0 I_n + \rho_0 W_n) \tag{14}$$

Equation (13) describes a role of future relevant components combined with δ. The additional components $\delta\gamma_0 I_n$ and $-\delta D_{n,1}(\gamma_0 I_n + \rho_0 W_n)$ 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 $e_i' Q_n^* = e_i'(Q_i + Q_i')$ and $e_i' L_n^* = e_i' L_i$ for all $i = 1, \cdots, n$. To explain equation (13), consider the first-order condition of agent i's arbitrary t period problem:

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

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

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.[23] Each Y_{nt} is supposed to be realized as an equilibrium, (i.e., $Y_{nt} = Y_{nt}^*(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$). For estimation, we assume some structures on $\boldsymbol{\eta}_{nt}$. First, $\boldsymbol{\eta}_{nt}$ 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 $\boldsymbol{\eta}_{nt}$. 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 $\mathbf{c}_{n0}^* = (c_{1,0}^*, \cdots, c_{n,0}^*)'$ which is invariant over time, the corresponding Π_n would be an identity matrix, thus

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

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

$$Y_{nt} = (\lambda_0 W_n + \delta Q_n^*)\,Y_{nt} + (\gamma_0 I_n + \rho_0 W_n)\,Y_{n,t-1} + (I_n + \delta L_n^* \Pi_n)\,X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t,0} l_n + \mathcal{E}_{nt} \qquad (15)$$

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 $\boldsymbol{\alpha}_{T0}$, where $\boldsymbol{\alpha}_{T0} = (\alpha_{1,0}, \cdots, \alpha_{T,0})'$ is the vector of time effects.

As special cases of model specification (15), 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[24], we have

$$R_n = \frac{1}{2}\left(1 + \delta\gamma_0 + \sqrt{1 + 2\delta\gamma_0 - \delta\gamma_0^2\,(4-\delta)}\right) I_n. \qquad (16)$$

To obtain validity of (16), $1 + 2\delta\gamma_0 - \delta\gamma_0^2\,(4-\delta) > 0$ 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 $e_i' A_n' \left[-e_i e_i' + \delta\,(Q_i + Q_i')\right] A_n + \gamma_0 e_i' \left[A_n' e_i e_i' + (A_n - I_n)\right]$. 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 (15) is

$$Y_{nt} = A_n Y_{n,t-1} + R_n^{-1}\left[(I_n + \delta L_n^* \Pi_n)\,X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t,0} l_n + \mathcal{E}_{nt}\right] \qquad (17)$$

where $A_n = R_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$ with $R_n = I_n - (\lambda_0 W_n + \delta Q_n^*)$. Stability of system (15) means the spatial-time dependence should be manageable. Note that $Q_n^* = D_{n,1}(\gamma_0 I_n + \rho_0 W_n) - \gamma_0 I_n$, $L_n^* = \sum_{m=1}^{\infty} \delta^{m-1} D_{n,m} \Pi_n^{m-1}$

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

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.[25] 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

$$Y_{nt} = \sum_{j=0}^{\infty} A_n^j R_n^{-1} \left[\left(I_n + \delta L_n^* \Pi_n \right) X_{n,t-j} \beta_0 + \mathbf{c}_{n0} + \alpha_{t-j,0} l_n + \mathcal{E}_{n,t-j} \right]. \quad (18)$$

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.

4 Estimation

4.1 Quasi-maximum likelihood estimation

To estimate equation (15), 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.[26] Let $\theta = \left(\lambda, \gamma, \rho, \beta', \sigma_\epsilon^2 \right)'$ 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 $\boldsymbol{\alpha}_T = (\alpha_1, \cdots, \alpha_T)'$. 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

$$\ln L_{nT}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T) = -\frac{nT}{2} \ln 2\pi - \frac{nT}{2} \ln \sigma_\epsilon^2 + T \ln |R_n(\theta_1)| - \frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^{T} \mathcal{E}_{nt}'(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T) \mathcal{E}_{nt}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T) \quad (19)$$

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

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 $L_n^*(\theta_1)$ for

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

[26] 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.

each θ_1 (i.e., inner loop). As the total number of individual and time fixed effects in $\mathbf{c}_{n0}$ and $\boldsymbol{\alpha}_{T0}$ 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 $\boldsymbol{\alpha}_{T0}$ 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

$$\ln L_{nT,c}(\theta) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln \sigma_\epsilon^2 + T\ln|R_n(\theta_1)| - \frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}\tilde{\mathcal{E}}_{nt}'(\theta) J_n \tilde{\mathcal{E}}_{nt}(\theta) \tag{20}$$

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.[27] From (20), 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.[28]

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

[27] 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.

[28] We introduce those approximation methods for calculating $|R_n(\theta_1)|$ and $R_n^{-1}(\theta_1)$ in our supplementary file.

Assumption 4.3 *The parameter space* Θ *of* θ *is compact. The true parameter* θ_0 *is in* $int(\Theta)$*.*

Assumption 4.4 $\{X_{nt}\}_{t=1}^{T}$, $\{\alpha_{t0}\}_{t=1}^{T}$ *and* $\mathbf{c}_{n0}$ *are conditional upon nonstochastic values with* $\sup_{n,T} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} |x_{it,k}|^{2+\eta} < \infty$ *for all* k, $\sup_{T} \frac{1}{T} \sum_{t=1}^{T} |\alpha_{t0}|^{2+\eta} < \infty$ *and* $\sup_{n} \frac{1}{n} \sum_{i=1}^{n} |c_{i,0}|^{2+\eta} < \infty$ *for some* $\eta > 0$*.*

Assumption 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 $\theta \in int(\Theta)$*, 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 4.6 *We assume that* T *goes to infinity and* n *is an increasing function of* T*.*

Assumption 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 4.2 (i) assumes *i.i.d.* disturbances across i and t for simplicity. Assuming a compact parameter space (Assumption 4.3) is for theoretical analyses (for details, refer to Chapter 4 in Amemiya (1985)). Assumption 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 (15), 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 $\mathbf{c}_{n0}$.[29] Assumption 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 (15) for any $\theta_1 \in \Theta_1$ (Assumption 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 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 $\frac{\partial \ln L_{nT,c}(\theta)}{\partial \theta}$

[29] 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.

and $\frac{\partial^2 \ln L_{nT,c}(\theta)}{\partial\theta\partial\theta'}$ for $\theta \in \Theta$ 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 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 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 4.5 (iv), $\sum_{h=1}^{\infty} \delta^{h-1} D_{n,1}^h \Pi_n^{h-1} = D_{n,1}(1 - \delta D_{n,1}\Pi_n)^{-1}$.[30] 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 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[31]:

$$\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left[B_{y,n}\tilde{Y}_{n,t-1}^{(-)} + D_{nt}\right]' J_n \tilde{\mathcal{E}}_{nt} + \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left[\tilde{\mathcal{E}}_{nt}' B_{q,n}' J_n \tilde{\mathcal{E}}_{nt} - \sigma_{\epsilon,0}^2 tr(B_{q,n})\right] \tag{21}$$

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

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

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

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left[B_{y,n}\tilde{Y}_{n,t-1}^{(-)(u)} + D_{nt}\right]' J_n\mathcal{E}_{nt} + \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left[\mathcal{E}_{nt}' B_{q,n}' J_n\mathcal{E}_{nt} - \sigma_{\epsilon,0}^2 tr(J_n B_{q,n})\right], \tag{22}$$

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

[30] 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.

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

$\sqrt{\frac{T}{n}}\left[\sigma_{\epsilon,0}^{2}\left(tr\left(B_{q,n}\right)-tr\left(J_{n}B_{q,n}\right)\right)\right]$ 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 $\bar{U}_{nT,-1}^{\prime}B_{n}\bar{\mathcal{E}}_{nT}$, and $\bar{\mathcal{E}}_{nT}^{\prime}B_{n}\bar{\mathcal{E}}_{nT}$ 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 B.

Consistency and asymptotic normality

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

$$Q_{nT}\left(\theta\right)=\frac{1}{nT}E\ln L_{nT,c}\left(\theta\right)=-\frac{1}{2}\ln2\pi-\frac{1}{2}\ln\sigma_{\epsilon}^{2}+\frac{1}{n}\ln\left|R_{n}\left(\theta_{1}\right)\right|-\frac{1}{2\sigma_{\epsilon}^{2}}\frac{1}{nT}E\left(\sum_{t=1}^{T}\tilde{\mathcal{E}}_{nt}^{\prime}\left(\theta\right)J_{n}\tilde{\mathcal{E}}_{nt}\left(\theta\right)\right)$$

To show consistency, the first step is verifying uniform convergence of sample average of the log-likelihood function, $\sup_{\theta\in\Theta}\left|\frac{1}{nT}\ln L_{nT,c}\left(\theta\right)-Q_{nT}\left(\theta\right)\right|\xrightarrow{p}0$ as $n,\,T\rightarrow\infty$. After this, we show $Q_{nT}\left(\theta\right)$ is well-behaved at any point θ in Θ by verifying uniform equicontinuity of $Q_{nT}\left(\theta\right)$.[32] Obtaining the identification uniqueness completes the proof of consistency. The assumption below describes the identification uniqueness conditions.

Assumption 4.7 (Identification) *To identify θ_{0}, we assume*

(i) $\lim_{n,T\rightarrow\infty}\left[\frac{1}{n}\ln\left|\sigma_{\epsilon,0}^{2}R_{n}^{-1\prime}R_{n}^{-1}\right|-\frac{1}{n}\ln\left|\sigma_{\epsilon,nT}^{2}(\theta_{1})R_{n}^{-1\prime}\left(\theta_{1}\right)R_{n}^{-1}\left(\theta_{1}\right)\right|\right]\neq0$ *for* $\theta_{1}\neq\theta_{1,0}$ *where*

$$\begin{aligned}\sigma_{\epsilon,nT}^{2}(\theta_{1})=&\frac{1}{nT}\sum_{t=1}^{T}E\left(\tilde{Z}_{nt}\left(\theta_{1}\right)-\tilde{\mathbb{X}}_{nt}\left(\theta_{1}\right)\left[\sum_{s=1}^{T}\tilde{\mathbb{X}}_{ns}^{\prime}\left(\theta_{1}\right)J_{n}\tilde{\mathbb{X}}_{ns}\left(\theta_{1}\right)\right]^{-1}\sum_{s=1}^{T}\tilde{\mathbb{X}}_{ns}^{\prime}\left(\theta_{1}\right)J_{n}\tilde{Z}_{ns}\left(\theta_{1}\right)\right)^{\prime}\\&\times J_{n}\left(\tilde{Z}_{nt}\left(\theta_{1}\right)-\tilde{\mathbb{X}}_{nt}\left(\theta_{1}\right)\left[\sum_{s=1}^{T}\tilde{\mathbb{X}}_{ns}^{\prime}\left(\theta_{1}\right)J_{n}\tilde{\mathbb{X}}_{ns}\left(\theta_{1}\right)\right]^{-1}\sum_{s=1}^{T}\tilde{\mathbb{X}}_{ns}^{\prime}\left(\theta_{1}\right)J_{n}\tilde{Z}_{ns}\left(\theta_{1}\right)\right)\\&+\frac{\sigma_{\epsilon,0}^{2}}{n-1}tr\left(R_{n}^{-1\prime}R_{n}^{\prime}\left(\theta_{1}\right)J_{n}R_{n}\left(\theta_{1}\right)R_{n}^{-1}\right)+o(1),\end{aligned}$$

$\tilde{Z}_{nt}\left(\theta_{1}\right)=\left[R_{n}\left(\theta_{1}\right)R_{n}^{-1}\left(\gamma_{0}I_{n}+\rho_{0}W_{n}\right)-\left(\gamma I_{n}+\rho W_{n}\right)\right]\tilde{Y}_{n,t-1}^{(-)}+R_{n}\left(\theta_{1}\right)R_{n}^{-1}\left[\tilde{\mathbb{X}}_{nt}\beta_{0}+\tilde{\alpha}_{t,0}l_{n}\right]$, *and* $\tilde{\mathbb{X}}_{nt}\left(\theta_{1}\right)=\left(I_{n}+\delta L_{n}^{*}\left(\theta_{1}\right)\Pi_{n}\right)\tilde{X}_{nt}$ *with* $\tilde{\mathbb{X}}_{nt}=\tilde{\mathbb{X}}_{nt}\left(\theta_{1,0}\right)$.

(ii) $\lim_{T\rightarrow\infty}\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{nt}^{\prime}J_{n}\tilde{\mathbb{X}}_{nt}$ *exists and is nonsingular.*

[32] Formally, $\limsup_{T\rightarrow\infty}\sup_{\theta^{\prime}\in\Theta}\sup_{\|\theta-\theta^{\prime}\|\leq\delta}\left|Q_{nT}\left(\theta\right)-Q_{nT}\left(\theta^{\prime}\right)\right|\rightarrow0$ as $\delta\rightarrow0$.

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 $\sigma^2_{\epsilon,nT}(\theta_1, \beta) = \arg\max_{\sigma^2_\epsilon} Q_{nT}\left(\theta_1, \beta, \sigma^2_\epsilon\right)$ and $\beta_{nT}(\theta_1) = \arg\max_{\beta} Q_{nT}\left(\theta_1, \beta, \sigma^2_\epsilon\right)$. Assumption 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

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

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 $J_n \tilde{\mathbb{X}}_{nt}\beta_0$ since $J_n \tilde{Z}_{nt}(\theta_{1,0}) = J_n \tilde{\mathbb{X}}_{nt}\beta_0$. Hence, the first term, $\sigma^2_{\epsilon,nT,1}(\theta_1)$, is a quadratic function of the difference between the two approximation functions for $J_n \tilde{\mathbb{X}}_{nt}\beta_0$ 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 $\theta_{1,0}$.[33] Identifying β_0 is done by Assumption 4.7 (ii), which is analogous to identification of β_0 in a standard linear regression once $\theta_{1,0}$ is identified. When replacing $\tilde{\mathbb{X}}_{nt}$ by $\tilde{X}_{nt}$, we can observe this feature and Assumption 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 4.1 *Suppose Assumptions 4.1 - 4.7 hold. Then,* $\hat{\theta}_{ml,nT} \xrightarrow{p} \theta_0$ *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 4.8 $\liminf_{n,T\to\infty} \phi_{\min}(\Omega_{\theta_0,nT}) > 0$ *and* $\liminf_{n,T\to\infty} \phi_{\min}(\Sigma_{\theta_0,nT}) > 0$ *where* $\phi_{\min}(\cdot)$ *denotes the smallest eigenvalue.*

Due to Assumption 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 $\inf_{n,T} \phi_{\min}(\Sigma_{\theta_0,nT}) > 0$ 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

[33] Detailed comments for identification can be found in the supplementary file.

the mean value theorem, and the central limit theorem for martingale difference arrays to $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{nT,c}^{(u)}(\theta_0)}{\partial \theta}$. The theorem below gives the asymptotic distribution of $\hat{\theta}_{ml,nT}$.

Theorem 4.2 *Suppose Assumptions 4.1 - 4.8 hold. Then,*

$$\sqrt{nT}\left(\hat{\theta}_{ml,nT}-\theta_0\right)+\sqrt{\frac{n}{T}}\Sigma_{\theta_0,nT}^{-1}a_{n,1}(\theta_0)+\sqrt{\frac{T}{n}}\Sigma_{\theta_0,nT}^{-1}a_{n,2}(\theta_0)+O_p\left(\max\left(\sqrt{\frac{n}{T^3}},\sqrt{\frac{T}{n^3}},\sqrt{\frac{1}{T}}\right)\right)$$
$$\xrightarrow{d} N\left(0,\Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1}\right),$$

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 4.2, we have the results: (i) if $\frac{n}{T}\to 0$, $n\left(\hat{\theta}_{ml,nT}-\theta_0\right)+\Sigma_{\theta_0,nT}^{-1}a_{n,2}\left(\theta_0\right)\xrightarrow{p} 0$, (ii) if $\frac{n}{T}\to c\in(0,\infty)$, $\sqrt{nT}\left(\hat{\theta}_{ml,nT}-\theta_0\right)+\sqrt{c}\Sigma_{\theta_0,nT}^{-1}a_{n,1}\left(\theta_0\right)+\sqrt{\frac{1}{c}}\Sigma_{\theta_0,nT}^{-1}a_{n,2}\left(\theta_0\right)\xrightarrow{d} N\left(0,\Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1}\right)$, and (iii) if $\frac{n}{T}\to\infty$, $T\left(\hat{\theta}_{ml,nT}-\theta_0\right)+\Sigma_{\theta_0,nT}^{-1}a_{n,1}\left(\theta_0\right)\xrightarrow{p} 0$. $\hat{\theta}_{ml,nT}$ has an asymptotic bias of order $O\left(\max\left\{\frac{1}{n},\frac{1}{T}\right\}\right)$ due to $-\frac{1}{T}\Sigma_{\theta_0,nT}^{-1}a_{n,1}\left(\theta_0\right)-\frac{1}{n}\Sigma_{\theta_0,nT}^{-1}a_{n,2}\left(\theta_0\right)$. 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

$$\hat{\theta}_{ml,nT}^{c}=\hat{\theta}_{ml,nT}-\frac{1}{T}\left[-\Sigma_{\theta_0,nT}^{-1}a_{n,1}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}}-\frac{1}{n}\left[-\Sigma_{\theta_0,nT}^{-1}a_{n,2}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}},$$ can be valuable.

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

Assumption 4.9 $\sum_{h=0}^{\infty}A_n^h\left(\theta_1\right)$ *and* $\sum_{h=1}^{\infty}hA_n^{h-1}\left(\theta_1\right)$ *are uniformly bounded in either row or column sums uniformly in a neighborhood of* θ_0.

Under Assumption 4.9, we have

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

when $\frac{n}{T^3}\to 0$ and $\frac{T}{n^3}\to 0$. Hence, we can apply the asymptotic equivalence.[34]

[34] That is, if (i)

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

and (ii) $\sqrt{nT}\left(\hat{\theta}_{ml,nT}-\frac{1}{T}\left[-\left(\Sigma_{\theta_0,nT}\right)^{-1}a_{n,1}(\theta_0)\right]-\frac{1}{n}\left[-\Sigma_{\theta_0,nT}^{-1}a_{n,2}(\theta_0)\right]-\theta_0\right)\xrightarrow{d}N\left(0,*\right)$ where $*$ denotes the asymptotic variance derived in Corollary 4.3, we also have $\sqrt{nT}\left(\hat{\theta}_{ml,nT}^{c}-\theta_0\right)\xrightarrow{d}N\left(0,*\right)$.

Corollary 4.3 *Under the additional Assumption 4.9,* $\frac{n}{T^3} \to 0$ *and* $\frac{T}{n^3} \to 0$*, then*

$$\sqrt{nT}\left(\hat{\theta}^c_{ml,nT} - \theta_0\right) \xrightarrow{d} N\left(0, \Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}\right).$$

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 5.

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.[35] 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 4.4 *Suppose Assumptions 4.1 - 4.8 hold. Additionally, assume* $\sum_{t=1}^T \alpha_{t0} = 0$*. Then,*

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

(iii) Assume Assumption 4.9, $\frac{n}{T^3} \to 0$ *and* $\frac{T}{n^3} \to 0$*. For each* i*,* $\sqrt{T}\left(\hat{c}^c_{i,ml} - c_{i,0}\right) \xrightarrow{d} N(0, \sigma^2_{\epsilon,0})$ *where* $\hat{c}^c_{i,ml} = \hat{c}_{i,ml}(\hat{\theta}^c_{ml,nT})$*. For each* t*,* $\sqrt{n}\left(\hat{\alpha}^c_{t,ml} - \alpha_{t0}\right) \xrightarrow{d} N(0, \sigma^2_{\epsilon,0})$ *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 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[36] $\sqrt{T}(\hat{c}_{i,ml} - c_{i,0}) = \frac{1}{\sqrt{T}}\sum_{t=1}^T \epsilon_{it} +$

[35] Identified $\mathbf{c}_{n0}$ are employed to recover agents' time-invariant characteristics η^{iv}_i's and $\alpha_{t,0}$'s represent common economic shocks. For details, see the supplementary file.

[36] 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

$O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}, \frac{\sqrt{T}}{n}\right)\right)$, 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 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 $\boldsymbol{\alpha}_{T,0}$ via (ii) $\hat{r}_{nt}\left(\hat{\theta}^{c}_{ml,nT}\right)$ would be beneficial compared to employing $\hat{r}_{nt}\left(\hat{\theta}_{ml,nT}\right)$.

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.

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 $\left[I_n, W_n, W_n', W_n'W_n, W_n^2, \cdots\right]$ as IVs. Define the sample moment function $g^{\mathbf{L}}_{nT}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T) = \frac{1}{nT}\sum_{t=1}^{T} Z_{nt}'\mathcal{E}_{nt}(\theta, \mathbf{c}_n, \alpha_t)$ and observe $E\left(g^{\mathbf{L}}_{nT}(\theta_0, \mathbf{c}_{n0}, \boldsymbol{\alpha}_{T,0})\right) = \mathbf{0}_{q\times 1}$. Then, the NL2S estimator (NL2SE) can be obtained by minimizing the objective function: $g^{\mathbf{L}\prime}_{nT}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T)\left(\frac{1}{nT}\sum_{t=1}^{T} Z_{nt}'Z_{nt}\right)^{-1} g^{\mathbf{L}}_{nT}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T)$.[37] 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 non-singularity of it. Remaining conditions for consistency and asymptotic normality can be achieved by our suggested assumptions for the QML method.[38] In next section, we compare estimation results by the QML and NL2S methods to investigate whether the NL2S estimation method could work well.

approach of estimating $\mathbf{c}_{n0}$ and $\boldsymbol{\alpha}_{T,0}$, we have the different convergence rate of the QMLE.

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

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

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

$$R_n Y_{nt} = \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + \sum_{k=1}^{K} (I_n + \delta L_n^* \Pi_n) \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n\right) X_{nt,k} + \mathbf{c}_{n0} + \alpha_{t0} l_n + \mathcal{E}_{nt} \quad (23)$$

and the expectation function Π_n is specified based on

$$X_{nt,k} = A_{k,n} X_{n,t-1,k} + \mathbf{c}_{n,k,0} + \alpha_{t,k,0} l_n + V_{nt,k} \quad (24)$$

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 $V_{nt,k}$.[39]

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 $\|\cdot\|_\infty$).[40] 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 mean square error (RMSE) and (iv) 95% coverage

[39] 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.

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

probability (CP).[41] The number of sample repetitions I is 400. The obtained MC results reported in Table 1 with $\delta = 0.95$ are summarized in Subsections 5.1 and 5.2.

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 $\hat{\gamma}_{ml,nT}$ ($\hat{\gamma}^c_{ml,nT}$), $\hat{\sigma}^2_{ml,nT}$ ($\hat{\sigma}^{2,c}_{ml,nT}$), $\hat{\gamma}_{1,ml,nT}$ ($\hat{\gamma}^c_{1,ml,nT}$), $\hat{\rho}_{1,ml,nT}$ ($\hat{\rho}^c_{1,ml,nT}$) 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 increase, contribution of large T for reducing biases is relatively larger compared to that of large n.

(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.

(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 0.95 when we increase n and T. Overall, the results (i), (ii) and (iii) also hold for $\delta = 0.5$.[42]

(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 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.

5.2 The results for specific parameters

(λ_0) In terms of empirical biases and coverage probabilities, $\hat{\lambda}^c_{ml,nT}$ 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 $\hat{\lambda}^c_{ml,nT}$ have relatively low coverage probabilities.

Based on $\hat{\lambda}^{S,c}_{ml,nT}$ when $(n, T) = (49, 10)$, the signs of misspecification biases are positive if $\rho_0 > 0$, but

[41] The 95% coverage probability is defined by

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

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 $\widehat{\Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}}$ denotes a consistent estimate of $\Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}$. We employ $\left[\Sigma^{-1}_{\theta}\Omega_{\theta}\Sigma^{-1}_{\theta}\right]_{\theta=\hat{\theta}_{ml,nT}}$ for $\widehat{\Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}}$.

[42] Those results are reported in the supplementary file.

are negative if $\rho_0 < 0$. From these results, the sign of ρ_0 determines the sign of the misspecification bias of $\hat{\lambda}^{S,c}_{ml,nT}$ 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 $\hat{\gamma}^{c}_{ml,nT}$ 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.

$(\sigma^2_{\epsilon,0})$ 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.

$\left(\rho_{1,0}\right)$ 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.

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.[43] 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 Figure 1, 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 and rich exogenous variables. Additional results and discussions can be found in the supplementary file.

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.

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

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

as well as no Durbin regressor for simplicity. As IVs, we employ $[Y_{n,t-1}, X_{nt}]$ and its transformations by $\left[I_n, W_n, W_n', W_n'W_n, W_n^2\right]$. Under this circumstance, $W_n [Y_{n,t-1}, X_{nt}]$ can play an important role in identifying θ_0.

For each estimation method and parameter value, we report empirical bias, standard deviation, and RMSE as bar graphs (Figure 2).[44] 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 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 $\tilde{\mathcal{E}}_{nt} (\theta)$'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.[45]

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.

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

[45] In the supplement file, we introduce an approximation method based on the Taylor expansion.

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 (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.[46] 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 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.[47] Table 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,

[46] 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.

[47] 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.

amounts of public safety spending, and proportion of expenditures on public safety. The maximal public safety spending is 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%).[48] 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 $\frac{\text{Population}}{\text{Land area (km}^2)}$, 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:

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

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.[49] 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 $d_c = 95$.

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.[50] To achieve a stable process of a decision variable, we consider

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

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

[50] 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 $\delta\left(1 + \bar{\imath}_r\right) = 1$. In a conventional intertemporal consumption-saving model, $\delta\left(1 + \bar{\imath}_r\right) = 1$ means completely smoothed consumption. For the detailed discussion, refer to Chapter 1.3 in Ljungqvist and Sargent (2012).

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 (23) and (24).[51]

The estimation results are summarized in Tables 3.A to 3.D: Tables 3.A, B, C, and D 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 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[52], 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 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.[53]

Here we provide economic interpretations based on the forward-looking agent model with $d_c = 80$. We can recover the cost function: $c\left(y_{it}, y_{i,t-1}\right) = 0.2541\left(y_{it} - y_{i,t-1}\right)^2 + 0.2459 y_{it}^2$. The marginal direct effect of

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

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

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

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.[54] 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

$$\frac{\partial y_{it}}{\partial x_{jt,k}} = \left[R_n^{-1}\left(I_n + \delta \mathbf{D}_{n,k} A_{k,n}\right)\left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n\right)\right]_{ij} \tag{25}$$

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 $\left[R_n^{-1}\left(I_n + \delta \mathbf{D}_{n,k} A_{k,n}\right)\left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n\right)\right]_{ii}$. 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 (25) 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 (17 neighbors) while Dare county has the smallest number of neighbors (3 neighbors). The figure below describes neighbors of the two counties.

[Figure 3 here]

Table 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.[55]

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

[55] Additional comments for this issues can be found in the supplement file.

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:

$$\frac{\partial \left[E_t\left(Y_{n,t+h}\right)\right]_i}{\partial x_{jt,k}} = \left[\sum_{g=0}^{h} A_n^{h-g} R_n^{-1} \left(I_n + \delta \mathbf{D}_{n,k} A_{k,n}\right) \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n\right) A_{k,n}^{g}\right]_{ij}. \tag{26}$$

Using formula (26), 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.

[Figure 4 here]

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 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, 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 $\ddot{X}_{nT,1}$)

$$\ddot{X}_{nT,1} = \begin{bmatrix} x_{1,T,1} & \cdots & x_{j,T,1} + \Delta_x & \cdots & x_{n,T,1} \end{bmatrix}'$$

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

$$\hat{\Delta}_{\mathcal{W}} = \tilde{\mathcal{W}}^F\left(\left\{Y_{nT}, \ddot{X}_{nT,1}\right\}; \hat{\theta}^c_{ml,nT}\right) - \tilde{\mathcal{W}}^F\left(\left\{Y_{nT}, X_{nT,1}\right\}; \hat{\theta}^c_{ml,nT}\right) \tag{27}$$

where $\tilde{\mathcal{W}}^F(\{Y_{nT}, X_{nT,1}\}; \theta)$ 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. $\tilde{\mathcal{W}}^F\left(\left\{Y_{nT}, \tilde{X}_{nT,1}\right\}; \hat{\theta}^c_{ml,nT}\right)$ captures social welfare when a county receives some subsidy while $\tilde{\mathcal{W}}^F\left(\{Y_{nT}, X_{nT,1}\}; \hat{\theta}^c_{ml,nT}\right)$ evaluates social welfare in a given realized economic environment. The difference between $\tilde{\mathcal{W}}^F\left(\left\{Y_{nT}, \tilde{X}_{nT,1}\right\}; \hat{\theta}^c_{ml,nT}\right)$ and $\tilde{\mathcal{W}}^F\left(\{Y_{nT}, X_{nT,1}\}; \hat{\theta}^c_{ml,nT}\right)$ will capture a welfare change corresponding to the change of policy.[56]

For convenience of analysis, we only 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 (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 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.

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

[56] The detailed derivation and specification can be found in the supplement file.

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.

Appendix A: Derivation of the MPE equation

In this appendix, we derive the NE equation by solving equation (7). By the principle of optimality, a solution from the intertemporal choice problem (6) is equivalent to that of the functional equation (7) if the latter exists. For this, we need to verify the existence and uniqueness of $V_i(\cdot)$ satisfying both (6) and (7). 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.[57]

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 $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ and $Y^*_{-i,t}(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$, define the operator $\mathcal{T}$ which maps the j^{th} approximation to the $(j+1)^{th}$ approximation of $V_i(\cdot)$ by

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

for $j = 0, 1, 2, \cdots$. From $V_i^{(j)}(\cdot)$'s, we can also generate $Y^{*(j)}_{nt}(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$'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.

[57] More details can be found in the supplementary document.

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 $\mathcal{C}\left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$ where all possible $Y_{n,t-1} \in \left(\chi_y\right)^n \subseteq \mathbb{R}^n$ and $\boldsymbol{\eta}_{nt} \in \left(\chi_\eta\right)^n \subseteq \mathbb{R}^n$. Note that $\mathcal{C}\left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$ is a well-known Banach space. Under Assumption 2.1, $\left\{V_i^{(j)}(\cdot)\right\}_j \subset \mathcal{C}\left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$ 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 $\mathcal{T}V_i^{(j)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ at $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$. Since $u_i(\cdot)$ is strictly concave with strictly decreasing marginals[58] with respect to large y_{it}, we can guarantee for unique NE decisions.[59]

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 $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ 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 $\mathcal{C}\left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$ for each $i = 1, \cdots, n$ and subsequently a unique NE $Y_{nt}^*(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$.

Step 4 (Recovering $V_i(\cdot)$ for each i and $Y_{nt}^*(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$): From the initial iteration with $V_i^{(0)} = 0$, we have $V_i^{(1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) = Y_{n,t-1}' Q_i^{(1)} Y_{n,t-1} + Y_{n,t-1}' L_i^{(1)} \boldsymbol{\eta}_{nt} + \boldsymbol{\eta}_{nt}' G_i^{(1)} \boldsymbol{\eta}_{nt} + c_i^{(1)}$, 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 zero elsewhere. By mathematical induction, we generate the following matrix Riccati equations:

$$Q_i^{(j+1)} = A_n^{(j+1)\prime}\left[\mathcal{I}_i\left(\frac{1}{2}I_n - S_n\right) + \delta Q_i^{(j)}\right] A_n^{(j+1)} + A_n^{(j+1)\prime}\mathcal{I}_i(\gamma_0 I_n + \rho_0 W_n) - \frac{\gamma_0}{2}\mathcal{I}_i, \tag{28}$$

$$Q_n^{*(j+1)} = \left[\left(Q_1^{(j+1)} + Q_1^{(j+1)\prime}\right)e_1, \quad \cdots, \quad \left(Q_n^{(j+1)} + Q_n^{(j+1)\prime}\right)e_n\right]',$$

$$\begin{aligned} L_i^{(j+1)} \quad = \quad & A_n^{(j+1)\prime}\left\{\left[\mathcal{I}_i\left(\frac{1}{2}I_n - S_n\right) + \delta Q_i^{(j)}\right] + \left[\mathcal{I}_i\left(\frac{1}{2}I_n - S_n\right) + \delta Q_i^{(j)}\right]'\right\} B_n^{(j+1)} \\ & + A_n^{(j+1)\prime}\left(\mathcal{I}_i + \delta L_i^{(j)}\Pi_n\right) + (\gamma_0 I_n + \rho_0 W_n)'\mathcal{I}_i B_n^{(j+1)}, \end{aligned} \tag{29}$$

$$L_n^{*(j+1)} = \left[L_1^{(j+1)\prime}e_1, \quad \cdots, \quad L_n^{(j+1)\prime}e_n\right]',$$

$$G_i^{(j+1)} = B_n^{(j+1)\prime}\left[\mathcal{I}_i\left(\frac{1}{2}I_n - S_n\right) + \delta Q_i^{(j)}\right] B_n^{(j+1)} + B_n^{(j+1)\prime}\left(\mathcal{I}_i + \delta L_i^{(j)}\Pi_n\right) + \delta \Pi_n' G_i^{(j)}\Pi_n, \tag{30}$$

and $c_i^{(j+1)} = \delta\left(c_i^{(j)} + tr\left(G_i^{(j)}\Omega_\xi\right)\right)$, where $A_n^{(j+1)} = \left[R_n^{(j+1)}\right]^{-1}(\gamma_0 I_n + \rho_0 W_n)$ and
$B_n^{(j+1)} = \left[R_n^{(j+1)}\right]^{-1}\left(I_n + \delta L_n^{*(j)}\Pi_n\right)$ with $R_n^{(j+1)} = S_n - \delta Q_n^{*(j)}$.

[58] Note that $u_i(\cdot)$ will eventually decrease in y_{it}. This property is important because our maximization problem is not constrained.

[59] Refer to Theorems 3.8 and 4.9 in Stokey et al. (1989).

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,

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

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

$$Y_{nt}^*\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) = \left(\lambda_0 W_n + \delta Q_n^*\right) Y_{nt}^*\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) + \left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n,t-1} + \left(I_i + \delta L_n^* \Pi_n\right) \boldsymbol{\eta}_{nt},$$

which implies that

$$Y_{nt}^*\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) = A_n Y_{n,t-1} + B_n \boldsymbol{\eta}_{nt},$$

where $A_n = R_n^{-1}\left(\gamma_0 I_n + \rho_0 W_n\right)$ and $B_n = R_n^{-1}\left(I_n + \delta L_n^* \Pi_n\right)$ 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, 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

$$\begin{aligned} Y_{nt}^{*(2)}\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) &= A_n^{(2)} Y_{n,t-1} + \left[R_n^{(2)}\right]^{-1}\left(\boldsymbol{\eta}_{nt} + \delta L_n^{*(1)} E_t\left(\boldsymbol{\eta}_{n,t+1}\right)\right) \\ &= A_n^{(2)} Y_{n,t-1} + \left[R_n^{(2)}\right]^{-1}\left(I_n + \delta D_{n,1}^{(2)} \Pi_n\right) \boldsymbol{\eta}_{nt}. \end{aligned}$$

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

$$L_n^{*(j)} = D_{n,1}^{(j+1)} + \delta D_{n,2}^{(j+1)} \Pi_n + \cdots + \delta^{j-1} D_{n,j}^{(j+1)} \Pi_n^{j-1} \tag{31}$$

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

$$\begin{aligned} & Y_{nt}^{*(j+1)}\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) \\ = \;& A_n^{(j+1)} Y_{n,t-1} + \left[R_n^{(j+1)}\right]^{-1}\left(I_n + \delta L_n^{*(j)} \Pi_n\right) \boldsymbol{\eta}_{nt} \\ = \;& A_n^{(j+1)} Y_{n,t-1} + \left[R_n^{(j+1)}\right]^{-1}\left(\boldsymbol{\eta}_{nt} + \delta D_{n,1}^{(j+1)} E_t\left(\boldsymbol{\eta}_{n,t+1}\right) + \delta^2 D_{n,2}^{(j+1)} E_t\left(\boldsymbol{\eta}_{n,t+2}\right) + \cdots + \delta^j D_{n,j}^{(j+1)} E_t\left(\boldsymbol{\eta}_{n,t+j}\right)\right) \\ = \;& A_n^{(j+1)} Y_{n,t-1} + \left[R_n^{(j+1)}\right]^{-1}\left(I_n + \delta D_{n,1}^{(j+1)} \Pi_n + \delta^2 D_{n,2}^{(j+1)} \Pi_n^2 + \cdots + \delta^j D_{n,j}^{(j+1)} \Pi_n^j\right) \boldsymbol{\eta}_{nt}. \end{aligned}$$

The second equality holds due to the law of iterative expectations. For notational convenience, let

$$\begin{aligned} C_i^{(j+1)} &= A_n^{(j+1)\prime}\left\{\left[\mathcal{I}_i\left(\frac{1}{2}I_n - S_n\right) + \delta Q_i^{(j)}\right] + \left[\mathcal{I}_i\left(\frac{1}{2}I_n - S_n\right) + \delta Q_i^{(j)}\right]'\right\} + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i \\ &= A_n^{(j+1)\prime}\left\{\mathcal{I}_i\left[-I_n + \lambda_0\left(W_n + W_n'\right)\right] + \delta\left(Q_i^{(j)} + Q_i^{(j)\prime}\right)\right\} + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i \end{aligned}$$

for $j = 1, 2, \cdots$. And, $C_i^{(1)} = A_n^{(1)\prime}\mathcal{I}_i\left[-I_n + \lambda_0\left(W_n + W_n'\right)\right] + (\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i$, so observe

$$\begin{aligned} & e_i'\left(C_i^{(1)}\left[R_n^{(1)}\right]^{-1} + A_n^{(1)\prime}\mathcal{I}_i\right) \\ = \; & e_i' A_n^{(1)\prime}\mathcal{I}_i\left[-I_n + \lambda_0\left(W_n + W_n'\right)\right] S_n^{-1} + e_i'(\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i S_n^{-1} + e_i' A_n^{(1)\prime}\mathcal{I}_i \\ = \; & e_i' A_n^{(1)\prime}\mathcal{I}_i\left(-I_n + \lambda_0 W_n' S_n^{-1}\right) + e_i' A_n^{(1)\prime}\mathcal{I}_i + e_i'(\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i S_n^{-1} \\ = \; & e_i' A_n^{(1)\prime}\mathcal{I}_i \lambda_0 W_n' S_n^{-1} + e_i'(\gamma_0 I_n + \rho_0 W_n)' \mathcal{I}_i S_n^{-1} \\ = \; & e_i'\left(A_n^{(1)\prime}\mathcal{I}_i \lambda_0 W_n' + (\gamma_0 I_n + \rho_0 W_n)' S_n^{-1\prime} S_n' \mathcal{I}_i\right) S_n^{-1} \\ = \; & e_i' A_n^{(1)\prime}\mathcal{I}_i S_n^{-1} = e_i' D_{n,1}^{(2)}. \end{aligned}$$

By equation (29),

$$\begin{aligned} L_i^{(j)} &= C_i^{(j)} B_n^{(j)} + A_n^{(j)\prime}\mathcal{I}_i + \delta A_n^{(j)\prime} L_i^{(j-1)}\Pi_n \\ &= C_i^{(j)} B_n^{(j)} + \delta A_n^{(j)\prime} C_i^{(j-1)} B_n^{(j-1)}\Pi_n + A_n^{(j)\prime}\mathcal{I}_i + \delta A_n^{(j)\prime} A_n^{(j-1)\prime}\mathcal{I}_i\Pi_n + \delta^2 A_n^{(j)\prime} A_n^{(j-1)\prime} L_i^{(j-2)}\Pi_n^2 \\ &= C_i^{(j)} B_n^{(j)} + \delta A_n^{(j)\prime} C_i^{(j-1)} B_n^{(j-1)}\Pi_n + \cdots + \delta^{j-2} A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(3)\prime} C_i^{(2)} B_n^{(2)} \Pi_n^{j-2} \\ &\quad + A_n^{(j)\prime}\mathcal{I}_i + \delta A_n^{(j)\prime} A_n^{(j-1)\prime}\mathcal{I}_i\Pi_n + \cdots + \delta^{j-2} A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(2)\prime}\mathcal{I}_i \Pi_n^{j-2} \\ &\quad + \delta^{j-1} A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(2)\prime} L_i^{(1)} \Pi_n^{j-1}. \end{aligned}$$

Then, we have

$$\begin{aligned} L_i^{(j)} &= C_i^{(j)} \left[R_n^{(j)}\right]^{-1} \left(I_n + \delta D_{n,1}^{(j)}\Pi_n + \delta^2 D_{n,2}^{(j)}\Pi_n^2 + \cdots + \delta^{j-1} D_{n,j-1}^{(j)}\Pi_n^{j-1}\right) \\ &\quad + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} \left(\delta\Pi_n + \delta^2 D_{n,1}^{(j-1)}\Pi_n^2 + \delta^3 D_{n,2}^{(j-1)}\Pi_n^3 + \cdots + \delta^{j-1} D_{n,j-2}^{(j-1)}\Pi_n^{j-1}\right) + \cdots \\ &\quad + A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(3)\prime} C_i^{(2)} \left[R_n^{(2)}\right]^{-1} \left(\delta^{j-2}\Pi_n^{j-2} + \delta^{j-1} D_{n,1}^{(2)}\Pi_n^{j-1}\right) \\ &\quad + A_n^{(j)\prime}\mathcal{I}_i + \delta A_n^{(j)\prime} A_n^{(j-1)\prime}\mathcal{I}_i\Pi_n + \cdots + \delta^{j-2} A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(2)\prime}\mathcal{I}_i\Pi_n^{j-2} \\ &\quad + \delta^{j-1} A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(2)\prime} L_i^{(1)}\Pi_n^{j-1} \\ &= \left(C_i^{(j)} \left[R_n^{(j)}\right]^{-1} + A_n^{(j)\prime}\mathcal{I}_i\right) \\ &\quad + \delta\left(C_i^{(j)} \left[R_n^{(j)}\right]^{-1} D_{n,1}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} + A_n^{(j)\prime} A_n^{(j-1)\prime}\mathcal{I}_i\right)\Pi_n \\ &\quad + \delta^2 \left(\begin{array}{c} C_i^{(j)} \left[R_n^{(j)}\right]^{-1} D_{n,2}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} D_{n,1}^{(j-1)} + A_n^{(j)\prime} A_n^{(j-1)\prime} C_i^{(j-2)} \left[R_n^{(j-2)}\right]^{-1} \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} A_n^{(j-2)\prime}\mathcal{I}_i \end{array}\right)\Pi_n^2 + \cdots \\ &\quad + \delta^{j-1} \left(\begin{array}{c} C_i^{(j)} \left[R_n^{(j)}\right]^{-1} D_{n,j-1}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} D_{n,j-2}^{(j-1)} + \cdots \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(3)\prime} C_i^{(2)} \left[R_n^{(2)}\right]^{-1} D_{n,1}^{(2)} + A_n^{(j)\prime} A_n^{(j-1)\prime} \cdots A_n^{(1)\prime}\mathcal{I}_i S_n^{-1}. \end{array}\right)\Pi_n^{j-1}. \end{aligned}$$

As $e_i' L_n^{*(j)} = e_i' L_i^{(j)}$, by applying the method of undetermined coefficients based on (31) and by taking e_i', we have

$$e_i' D_{n,1}^{(j+1)} = e_i' \left(C_i^{(j)} \left[R_n^{(j)}\right]^{-1} + A_n^{(j)\prime}\mathcal{I}_i\right),$$

$$e_i' D_{n,2}^{(j+1)} = e_i' \left(C_i^{(j)} \left[R_n^{(j)}\right]^{-1} D_{n,1}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} + A_n^{(j)\prime} A_n^{(j-1)\prime}\mathcal{I}_i\right),$$

$$e_i' D_{n,3}^{(j+1)} = e_i' \left(\begin{array}{c} C_i^{(j)} \left[R_n^{(j)}\right]^{-1} D_{n,2}^{(j)} + A_n^{(j)\prime} C_i^{(j-1)} \left[R_n^{(j-1)}\right]^{-1} D_{n,1}^{(j-1)} + A_n^{(j)\prime} A_n^{(j-1)\prime} C_i^{(j-2)} \left[R_n^{(j-2)}\right]^{-1} \\ + A_n^{(j)\prime} A_n^{(j-1)\prime} A_n^{(j-2)\prime}\mathcal{I}_i \end{array}\right), \ldots$$

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 *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. Note that $e_i' Q_n^{*(j)} = e_i' \left(Q_i^{(j)} + Q_i^{(j)\prime} \right)$ and

$$\begin{aligned}
e_i' \left(Q_i^{(j)} + Q_i^{(j)\prime} \right) &= e_i' A_n^{(j)\prime} \left\{ \mathcal{I}_i \left[-I_n + \lambda_0 \left(W_n + W_n' \right) \right] + \delta \left(Q_i^{(j-1)} + Q_i^{(j-1)\prime} \right) \right\} A_n^{(j)} \\
&\quad + e_i' A_n^{(j)\prime} \mathcal{I}_i \left(\gamma_0 I_n + \rho_0 W_n \right) + e_i' \left(\gamma_0 I_n + \rho_0 W_n \right)' \mathcal{I}_i A_n^{(j)} - \gamma_0 e_i' \\
&= e_i' A_n^{(j)\prime} \left\{ \mathcal{I}_i \left[-I_n + \lambda_0 \left(W_n + W_n' \right) \right] + \delta \left(Q_i^{(j-1)} + Q_i^{(j-1)\prime} \right) \right\} A_n^{(j)} \\
&\quad + e_i' A_n^{(j)\prime} e_i e_i' \left(\gamma_0 I_n + \rho_0 W_n \right) + \gamma_0 e_i' A_n^{(j)} - \gamma_0 e_i' \\
&= e_i' A_n^{(j)\prime} \left\{ \mathcal{I}_i \left[-I_n + \lambda_0 \left(W_n + W_n' \right) \right] + \delta \left(Q_i^{(j-1)} + Q_i^{(j-1)\prime} \right) \right\} \left[R_n^{(j)} \right]^{-1} \left(\gamma_0 I_n + \rho_0 W_n \right) \\
&\quad + e_i' \left(\gamma_0 I_n + \rho_0 W_n \right)' \mathcal{I}_i \left[R_n^{(j)} \right]^{-1} \left(\gamma_0 I_n + \rho_0 W_n \right) + e_i' A_n^{(j)\prime} e_i e_i' \left(\gamma_0 I_n + \rho_0 W_n \right) - \gamma_0 e_i' \\
&= e_i' \left(\begin{array}{c} A_n^{(j)\prime} \left\{ \mathcal{I}_i \left[-I_n + \lambda_0 \left(W_n + W_n' \right) \right] + \delta \left(Q_i^{(j-1)} + Q_i^{(j-1)\prime} \right) \right\} \left[R_n^{(j)} \right]^{-1} \\ + \left(\gamma_0 I_n + \rho_0 W_n \right)' \mathcal{I}_i \left[R_n^{(j)} \right]^{-1} + A_n^{(j)\prime} \mathcal{I}_i \end{array} \right) \left(\gamma_0 I_n + \rho_0 W_n \right) \\
&\quad - e_i' \gamma_0 I_n \\
&= e_i' \left(D_{n,1}^{(j+1)} \left(\gamma_0 I_n + \rho_0 W_n \right) - \gamma_0 I_n \right)
\end{aligned}$$

for $j = 2, 3, \cdots$, since $e_i' \left(\gamma_0 I_n + \rho_0 W_n \right) e_i = \gamma_0$ by $e_i' W_n e_i = w_{ii} = 0$ for all $i = 1, \cdots, n$. ■

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.2 (Stewart (1998)) *Consider a linear operator $I_n - C_n$ satisfies $\lim_{j\to\infty} \left\| C_n^j \right\| = 0$ 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

$$\left\| \frac{\lambda_0}{1 + \delta\gamma_0} W_n + \frac{\delta}{1 + \delta\gamma_0} D_{n,1}^{(j+1)} \left(\gamma_0 I_n + \rho_0 W_n \right) \right\|_\infty < 1.$$

Then, $\left[R_n^{(j+1)} \right]^{-1}$ has the Neumann series expansion,

$$\left[R_n^{(j+1)} \right]^{-1} = \frac{1}{1 + \delta\gamma_0} \left[I_n + \sum_{j=1}^{\infty} \left(\frac{1}{1 + \delta\gamma_0} \right)^j \left(\lambda_0 W_n + \delta\gamma_0 D_{n,1}^{(j+1)} + \delta\rho_0 D_{n,1}^{(j+1)} W_n \right)^j \right].$$

Appendix B: Statistical results

In this section, we list components of asymptotic biases of the QMLE, and provide briefly proofs of Theorems 4.1, 4.2, 4.4 and Corollary 4.3. The detailed proofs can be found in our supplementary file.

7.1 First order derivatives of the log-likelihood function

Note that

$$\tilde{Y}_{nt} = A_n \tilde{Y}^{(-)}_{n,t-1} + \sum_{k=1}^{K} R_n^{-1} \left(I_n + \delta L_n^* \Pi_n \right) \tilde{X}_{nt,k} \beta_0 + R_n^{-1} \left(\tilde{\alpha}_{t,0} l_n + \tilde{\mathcal{E}}_{nt} \right).$$

The components of $\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \theta}$ are

$$\begin{aligned}
\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \lambda} &= \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[-R_{n\lambda} R_n^{-1} \begin{pmatrix} \left(\gamma_0 I_n + \rho_0 W_n\right) \tilde{Y}^{(-)}_{n,t-1} \\ + \left(I_n + \delta L_n^* \Pi_n \right) \tilde{X}_{nt,k} \beta_0 + \tilde{\alpha}_{t0} l_n \end{pmatrix} + \delta L^*_{n\lambda} \Pi_n \tilde{X}_{nt,k} \beta_0 \right]' J_n \tilde{\mathcal{E}}_{nt} \\
&\quad + \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[\tilde{\mathcal{E}}'_{nt} \left(-R_n^{-1\prime} R'_{n\lambda} \right) J_n \tilde{\mathcal{E}}_{nt} - \sigma^2_{\epsilon,0} tr \left(-R_{n\lambda} R_n^{-1} \right) \right],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \gamma} &= \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[-R_{n\gamma} R_n^{-1} \begin{pmatrix} \left(\gamma_0 I_n + \rho_0 W_n\right) \tilde{Y}^{(-)}_{n,t-1} \\ + \left(I_n + \delta L_n^* \Pi_n \right) \tilde{X}_{nt,k} \beta_0 + \tilde{\alpha}_{t0} l_n \end{pmatrix} + \tilde{Y}^{(-)}_{n,t-1} + \delta L^*_{n\gamma} \Pi_n \tilde{X}_{nt,k} \beta_0 \right]' J_n \tilde{\mathcal{E}}_{nt} \\
&\quad + \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[\tilde{\mathcal{E}}'_{nt} \left(-R_n^{-1\prime} R'_{n\gamma} \right) J_n \tilde{\mathcal{E}}_{nt} - \sigma^2_{\epsilon,0} tr \left(-R_{n\gamma} R_n^{-1} \right) \right],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \rho} &= \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[-R_{n\rho} R_n^{-1} \begin{pmatrix} \left(\gamma_0 I_n + \rho_0 W_n\right) \tilde{Y}^{(-)}_{n,t-1} \\ + \left(I_n + \delta L_n^* \Pi_n \right) \tilde{X}_{nt,k} \beta_0 + \tilde{\alpha}_{t0} l_n \end{pmatrix} + W_n \tilde{Y}^{(-)}_{n,t-1} + \delta L^*_{n\rho} \Pi_n \tilde{X}_{nt,k} \beta_0 \right]' J_n \tilde{\mathcal{E}}_{nt} \\
&\quad + \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[\tilde{\mathcal{E}}'_{nt} \left(-R_n^{-1\prime} R'_{n\rho} \right) J_n \tilde{\mathcal{E}}_{nt} - \sigma^2_{\epsilon,0} tr \left(-R_{n\rho} R_n^{-1} \right) \right],
\end{aligned}$$

$$\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \beta_k} = \frac{1}{\sigma^2_{\epsilon,0}} \sum_{t=1}^{T} \left[\left(I_n + \delta L_n^* \Pi_n \right) \tilde{X}_{nt,k} \right]' J_n \tilde{\mathcal{E}}_{nt} \text{ for } k = 1, \cdots, K,$$

$$\frac{\partial \ln L_{nT,c}(\theta_0)}{\partial \sigma^2_\epsilon} = \frac{1}{2\sigma^4_{\epsilon,0}} \sum_{t=1}^{T} \left[\tilde{\mathcal{E}}'_{nt} J_n \tilde{\mathcal{E}}_{nt} - n\sigma^2_{\epsilon,0} \right].$$

7.2 Components of asymptotic biases of QMLEs

Here are the components of $\Delta_{1,nT}$, $\Delta_{2,nT}$, $a_{n,1}(\theta_0)$, and $a_{n,2}(\theta_0)$:

$\Delta^{\lambda}_{1,nT} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left[\left(-R_{n\lambda} R_n^{-1} \left(\gamma_0 I_n + \rho_0 W_n \right) \bar{U}_{nT,-1} \right)' J_n \bar{\mathcal{E}}_{nT} + \bar{\mathcal{E}}'_{nT} (-R_n^{-1\prime} R'_{n\lambda}) J_n \bar{\mathcal{E}}_{nT} \right]$,

$\Delta^{\gamma}_{1,nT} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left[\left(\left(-R_{n\gamma} R_n^{-1} \left(\gamma_0 I_n + \rho_0 W_n \right) + I_n \right) \bar{U}_{nT,-1} \right)' J_n \bar{\mathcal{E}}_{nT} + \bar{\mathcal{E}}'_{nT} (-R_n^{-1\prime} R'_{n\gamma}) J_n \bar{\mathcal{E}}_{nT} \right]$,

$\Delta^{\rho}_{1,nT} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left[\left(\left(-R_{n\rho} R_n^{-1} \left(\gamma_0 I_n + \rho_0 W_n \right) + W_n \right) \bar{U}_{nT,-1} \right)' J_n \bar{\mathcal{E}}_{nT} + \bar{\mathcal{E}}'_{nT} (-R_n^{-1\prime} R'_{n\rho}) J_n \bar{\mathcal{E}}_{nT} \right]$,

$\Delta^{\beta_{1,k}}_{1,nT} = \mathbf{0}_{K\times 1}$, $\Delta^{\sigma^2_\epsilon}_{1,nT} = \frac{1}{2\sigma^4_{\epsilon,0}} \sqrt{\frac{T}{n}} \bar{\mathcal{E}}'_{nT} J_n \bar{\mathcal{E}}_{nT}$,

$\Delta^{\lambda}_{2,nT} = \sqrt{\frac{T}{n}} \left[tr \left(-R_{n\lambda} R_n^{-1} \right) - tr \left(J_n (-R_{n\lambda} R_n^{-1}) \right) \right]$, $\Delta^{\gamma}_{2,nT} = \sqrt{\frac{T}{n}} \left[tr \left(-R_{n\gamma} R_n^{-1} \right) - tr \left(J_n (-R_{n\gamma} R_n^{-1}) \right) \right]$,

$\Delta^{\rho}_{2,nT} = \sqrt{\frac{T}{n}}\left[tr\left(-R_{n\rho}R_n^{-1}\right) - tr\left(J_n(-R_{n\rho}R_n^{-1})\right)\right]$, $\Delta^{\beta}_{2,nT} = \mathbf{0}_{K\times 1}$, and $\Delta^{\sigma^2_\epsilon}_{2,nT} = \sqrt{\frac{T}{n}}\frac{1}{2\sigma^2_{\epsilon,0}}$,

$$a_{n,1}(\theta_0) = \begin{pmatrix} \frac{1}{n}tr\left(J_n(-R_{n\lambda}A_n)\left(\sum_{h=0}^{\infty}A_n^h\right)R_n^{-1}\right) + \frac{1}{n}tr\left(J_n(-R_{n\lambda}R_n^{-1})\right) \\ \frac{1}{n}tr\left(J_n\left(-R_{n\gamma}A_n + I_n\right)\left(\sum_{h=0}^{\infty}A_n^h\right)R_n^{-1}\right) + \frac{1}{n}tr\left(J_n(-R_{n\gamma}R_n^{-1})\right) \\ \frac{1}{n}tr\left(J_n\left(-R_{n\rho}A_n + W_n\right)\left(\sum_{h=0}^{\infty}A_n^h\right)R_n^{-1}\right) + \frac{1}{n}tr\left(J_n(-R_{n\rho}R_n^{-1})\right) \\ \mathbf{0}_{K\times 1} \\ \frac{n-1}{n}\frac{1}{2\sigma^2_{\epsilon,0}} \end{pmatrix},$$

and

$$a_{n,2}(\theta_0) = \begin{pmatrix} \frac{1}{n}l_n'(-R_{n\lambda}R_n^{-1})l_n, & \frac{1}{n}l_n'(-R_{n\gamma}R_n^{-1})l_n, & \frac{1}{n}l_n'(-R_{n\rho}R_n^{-1})l_n, & \mathbf{0}_{1\times K} & \frac{1}{2\sigma^2_{\epsilon,0}} \end{pmatrix}$$

7.3 Sketches of Proofs (Consistency and asymptotic normality)

Sketch of proof of Theorem 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, $\sup_{\theta\in\Theta}\left|\frac{1}{nT}\ln L_{nT,c}(\theta) - Q_{nT}(\theta)\right| \xrightarrow{p} 0$ as $T\to\infty$. 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), $\frac{1}{nT}\sum_{t=1}^{T}\left[\tilde{\mathcal{E}}'_{nt}(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta) - E\left(\tilde{\mathcal{E}}'_{nt}(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)\right)\right] \xrightarrow{p} 0$ uniformly in $\theta\in\Theta$. Since σ^2_ϵ is assumed to be bounded away from zero,

$$\frac{1}{nT}\ln L_{nT,c}(\theta) - Q_{nT}(\theta) = -\frac{1}{2\sigma^2_\epsilon}\frac{1}{nT}\sum_{t=1}^{T}\left[\tilde{\mathcal{E}}'_{nt}(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta) - E\left(\tilde{\mathcal{E}}'_{nt}(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)\right)\right] \xrightarrow{p} 0$$

uniformly in $\theta\in\Theta$.

Secondly, we will show that $Q_{nT}(\theta)$ is uniformly equicontinuous in $\theta\in\Theta$. Note that

$$\frac{1}{nT}\sum_{t=1}^{T}E\left(\tilde{\mathcal{E}}'_{nt}(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)\right) = q_{nT,1}(\theta_1,\beta) + q_{nT,2}(\theta_1) + o(1)$$

where

$$q_{nT,1}(\theta_1,\beta) = \frac{1}{nT}\sum_{t=1}^{T} E\left[\begin{array}{c}\left(R_n(\theta_1)R_n^{-1}(\gamma_0 I_n+\rho_0 W_n)-(\gamma I_n+\rho W_n)\right)\tilde{Y}_{n,t-1}^{(-)}+\tilde{\alpha}_{t,0}R_n(\theta_1)R_n^{-1}l_n \\ +R_n(\theta_1)R_n^{-1}\tilde{\mathbb{X}}_{nt}\beta_0-\tilde{\mathbb{X}}_{nt}(\theta_1)\beta\end{array}\right]' \times J_n\left[\begin{array}{c}\left(R_n(\theta_1)R_n^{-1}(\gamma_0 I_n+\rho_0 W_n)-(\gamma I_n+\rho W_n)\right)\tilde{Y}_{n,t-1}^{(-)}+\tilde{\alpha}_{t,0}R_n(\theta_1)R_n^{-1}l_n \\ +R_n(\theta_1)R_n^{-1}\tilde{\mathbb{X}}_{nt}\beta_0-\tilde{\mathbb{X}}_{nt}(\theta_1)\beta\end{array}\right],$$

and $q_{nT,2}(\theta_1)=\frac{T-1}{nT}\sigma_{\epsilon,0}^2 tr\left(R_n^{-1\prime}R_n'(\theta_1)J_nR_n(\theta_1)R_n^{-1}\right)$. 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 $\theta_1,\theta_2\in\Theta$, $\bar{\theta}$ lies between θ_1 and θ_2, and $h_{nT}(\cdot)$ are uniformly bounded. Uniform boundedness of $h_{nT}(\cdot)$ comes from Assumptions 4.3 - 4.5. By applying Assumption 4.7, we achieve the desired result. ■

Sketch of proof of Theorem 4.2. This proof relies on the Taylor expansion:

$$\sqrt{nT}\left(\hat{\theta}_{ml,nT}-\theta_0\right)=\left(-\frac{1}{nT}\frac{\partial^2\ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta}-\Delta_{1,nT}-\Delta_{2,nT}\right)$$

where $\bar{\theta}_{nT}$ lies between θ_0 and $\hat{\theta}_{ml,nT}$. By Assumptions 4.2 (ii), 4.3 and 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 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 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,

$$\sqrt{nT}\left(\hat{\theta}_{ml,nT}-\theta_0\right)=\underbrace{\left(-\frac{1}{nT}\frac{\partial^2\ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}\right)^{-1}}_{=O_p(1)}\cdot\left(\underbrace{\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{nT,c}^{(u)}(\theta_0)}{\partial\theta}}_{=O_p(1)}-\underbrace{\Delta_{1,nT}}_{=O\left(\sqrt{\frac{n}{T}}\right)+O\left(\sqrt{\frac{n}{T^3}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)}-\underbrace{\Delta_{2,nT}}_{=O\left(\sqrt{\frac{T}{n}}\right)}\right),$$

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 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 $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{(u)}_{nT,c}(\theta_0)}{\partial\theta}$. The stochastic components of $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{(u)}_{nT,c}(\theta_0)}{\partial\theta}$ take a linear-quadratic form, $\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}\xi_{nt,i}$, where $E\left(\xi_{nt,i}|\mathcal{F}_{n,t,i-1}\right) = 0$

$$\mathcal{F}_{n,t,i} = \sigma\left(\epsilon_{11}, \cdots, \epsilon_{n1}, \cdots, \epsilon_{1,t-1}, \cdots, \epsilon_{n,t-1}, \epsilon_{1t}, \cdots, \epsilon_{it}\right), \tag{32}$$

and $\mathcal{F}_{n,0,0} = \{\varnothing, \Omega\}$, 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, $\left\{\left(\xi_{nt,i}, \mathcal{F}_{n,t,i}\right) : i = 1, \cdots, n \text{ and } t = 1, \cdots, T\right\}$. 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).[60] In consequence, we obtain $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{(u)}_{nT,c}(\theta_0)}{\partial\theta} \xrightarrow{d} N\left(0, \Omega_{\theta_0}\right)$ as $T \to \infty$ and have the desired results. ■

Sketch of proof of Corollary 4.3. By Theorem 4.2,

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

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

$$\sqrt{nT}\left(\hat{\theta}^c_{ml,nT} - \theta_0\right) \xrightarrow{d} N\left(0, \Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}\right)$$

if

$$\sqrt{\frac{n}{T}}\left(\left[\Sigma^{-1}_{\theta,nT}a_{n,1}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \Sigma^{-1}_{\theta_0,nT}a_{n,1}(\theta_0)\right) \xrightarrow{p} 0 \tag{33}$$

and

$$\sqrt{\frac{T}{n}}\left(\left[\Sigma^{-1}_{\theta,nT}a_{n,2}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \Sigma^{-1}_{\theta_0,nT}a_{n,2}(\theta_0)\right) \xrightarrow{p} 0. \tag{34}$$

Assumption 4.9, $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$ (with Assumptions 4.3 and 4.5) imply (33) and (34). The detailed arguments can be found in our supplementary file. ■

Sketch of proof of Theorem 4.4. (i) First, note that $\hat{c}_{i,ml} = c_{i,ml}\left(\hat{\theta}_{ml,nT}\right)$. By Theorem 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 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, $\sqrt{T}\left(\hat{c}_{i,ml}(\hat{\theta}_{ml,nT}) - c_{i,0}\right) \xrightarrow{d} N(0, \sigma^2_{\epsilon,0})$ if $\frac{\sqrt{T}}{n} \to 0$; and $\hat{c}_{i,ml}(\hat{\theta}_{ml,nT})$'s are asymptotically independent from each other.

[60] Also, refer to Kelejian and Prucha (2001).

(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 $\sqrt{n}\left(\hat{\alpha}_{t,ml}-\alpha_{t0}\right)\xrightarrow{d}N(0,\sigma_{\epsilon,0}^2)$ 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 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). ■

References

[1] Amemiya, T. 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

[2] Anselin, L. 1988. Spatial Econometrics: Methods and Models. Kluwer Academic, The Netherlands.

[3] Ballester, C., A. Calvō-Armengol, and Y. Zenou 2006. Who's who in networks. wanted: the key player. Econometrica 74, 1403-1417.

[4] 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.

[5] Blackwell, D. 1965. Discounted dynamic programming. Annals of Mathematical Statistics 36, 226-235.

[6] Calvō-Armengol A., E. Patacchini, and Y. Zenou 2009. Peer effects and social networks in education. Review of Economic Studies 76, 1239-1267.

[7] Cliff, A.D. and J.K. Ord 1973. Spatial Autocorrelation. London: Pion Ltd.

[8] Engsted, T., and N. Haldrup 1994. The linear quadratic adjustment cost model and the demand for labor. Journal of Applied Econometrics 9, 145-159.

[9] Fuente, A. 2000. Mathematical methods and models for economists. Cambridge University Press, United Kingdom

[10] 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.

[11] Jackson, O., and Y. Zenou 2014. Games on network. Handbook of Game Theory Vol. 4.

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

[13] 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.

[14] Kennan, J. 1979. The estimation of partial adjustment models with rational expectations. Econometrica 47, 1441-1455.

[15] 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.

[16] Lee, L.F. 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. Econometrica 72, 1899-1925.

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

[18] Lee, S.Y. 2008. Three essays on spatial econometrics and empirical industrial organization, PhD Thesis, The Ohio State University (2008).

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] LeSage, J., and R. K. Pace 2009. Introduction to Spatial Econometrics. Boca Raton, FL, Chapman and Hall/CRC.

[24] 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.

[25] Liu, X., J. H. Kagel and L.F. Lee 2012. Learning from peers in signaling game experiments. Journal of Applied Econometrics 27, 1037-1058.

[26] Ljungqvist, L., and T.J. Sargent 2012. Recursive Macroeconomic Theory. Third edition, The MIT Press.

[27] 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.

[28] 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.

[29] Maskin, E., and J. Tirole 2001. Markov perfect equilibrium I. Observable Actions. Journal of Economic Theory 100 (2), 191-219.

[30] Milgrom P, and Roberts J. 1982. Limit pricing and entry under incomplete information: an equilibrium analysis. Econometrica 50, 443-459.

[31] Murphy, K. M. and Topel, R. H. 1985. Estimation and inference in two-step econometric models. Journal of Business & Economic Statistics 3, 370–379.

[32] Ord, J.K. 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120-126.

[33] Pagan, A. 1984. Econometric issues in the analysis of regressions with generated regressors. International Economic Review 25 (1), 221-247.

[34] Stewart, G. 1998. Matrix Algorithms: Basic Decompositions. Society for Industrial and Applied Mathematics, Philadelphia.

[35] Stokey, N. L., Lucas, R. E. and Prescott E. C. 1989. Recursive Methods in Economic Dynamics. Cambridge, MA: Havard University Press.

[36] Yang, C. and L.F. Lee 2017. Strategical interactions on municipal public safety spending with correlated private information. Regional Science and Urban Economics, forthcoming.

[37] 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.

Table 1 : Performance of $\hat{\theta}_{ml,nT}$ and $\hat{\theta}^{c}_{ml,nT}$ when $\delta = 0.95$

$(n,T)=(49,10)$ $(\lambda,\rho)=(0.2,0.2)$		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	-0.0205	-0.1470	0.0548	0.0397	0.0628	-0.2147	-0.1493	-0.0135	-0.1383
	SD	0.0665	0.0565	0.0786	0.0429	0.0813	0.0740	0.0442	0.0860	0.0612
	RMSE	0.0695	0.1575	0.0957	0.0584	0.1027	0.2271	0.1557	0.0870	0.1512
	CP	0.9300	0.2150	0.8550	0.8425	0.8800	0.1375	0.0600	0.9425	0.3050
$\hat{\theta}^{c}_{ml,nT}$	Bias	0.0031	-0.0311	0.0002	0.0220	0.0263	-0.0452	-0.0258	-0.0044	-0.0361
	SD	0.0712	0.0650	0.0898	0.0428	0.0827	0.0864	0.0491	0.0948	0.0684
	RMSE	0.0712	0.0720	0.0897	0.0481	0.0867	0.0974	0.0554	0.0948	0.0773
	CP	0.9325	0.8400	0.9025	0.9175	0.9300	0.7850	0.8650	0.9075	0.8075
$\hat{\theta}^{S}_{ml,nT}$	Bias	0.0230	-0.1870	0.0205	0.0009	0.0325	-0.4338			
	SD	0.0567	0.0424	0.0695	0.0374	0.0733	0.0395			
	RMSE	0.0612	0.1918	0.0724	0.0373	0.0801	0.4356			
	CP	0.9150	0.0025	0.9450	0.9425	0.9175	0.0000			
$\hat{\theta}^{S,c}_{ml,nT}$	Bias	0.0549	-0.0990	-0.0367	-0.0106	0.0121	-0.3692			
	SD	0.0574	0.0462	0.0746	0.0368	0.0726	0.0441			
	RMSE	0.0794	0.1092	0.0831	0.0383	0.0735	0.3718			
	CP	0.8100	0.3150	0.8925	0.9400	0.9300	0.0000			

$(n,T)=(49,10)$ $(\lambda,\rho)=(0.2,-0.2)$		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	-0.0811	-0.1618	0.0990	0.0279	0.0406	-0.2442	-0.1496	-0.0158	-0.1383
	SD	0.0724	0.0567	0.0870	0.0419	0.0798	0.0736	0.0442	0.0859	0.0612
	RMSE	0.1087	0.1714	0.1317	0.0503	0.0895	0.2550	0.1560	0.0873	0.1512
	CP	0.7800	0.1625	0.7675	0.8900	0.9125	0.0725	0.0600	0.9375	0.3050
$\hat{\theta}^{c}_{ml,nT}$	Bias	-0.0198	-0.0373	0.0243	0.0190	0.0183	-0.0622	-0.0261	-0.0064	-0.0362
	SD	0.0801	0.0665	0.1000	0.0420	0.0822	0.0877	0.0491	0.0949	0.0684
	RMSE	0.0824	0.0762	0.1028	0.0460	0.0841	0.1074	0.0556	0.0950	0.0773
	CP	0.8900	0.8175	0.8800	0.9150	0.9250	0.7100	0.8625	0.9075	0.8075
$\hat{\theta}^{S}_{ml,nT}$	Bias	-0.1125	-0.1994	0.1161	-0.0158	0.0010	-0.4569			
	SD	0.0601	0.0425	0.0731	0.0368	0.0715	0.0379			
	RMSE	0.1275	0.2039	0.1372	0.0400	0.0714	0.4585			
	CP	0.5100	0.0000	0.6075	0.9250	0.9425	0.0000			
$\hat{\theta}^{S,c}_{ml,nT}$	Bias	-0.0809	-0.1063	0.0588	-0.0268	-0.0178	-0.3935			
	SD	0.0612	0.0463	0.0789	0.0363	0.0710	0.0423			
	RMSE	0.1014	0.1159	0.0983	0.0450	0.0731	0.3958			
	CP	0.6900	0.2400	0.8450	0.8800	0.9350	0.0000			

$(n,T)=(49,10)$ $(\lambda,\rho)=(-0.2,0.2)$		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	-0.0154	-0.1503	0.0112	0.0257	0.0300	-0.2346	-0.1497	-0.0164	-0.1383
	SD	0.0722	0.0560	0.0837	0.0417	0.0794	0.0731	0.0443	0.0860	0.0612
	RMSE	0.0737	0.1604	0.0843	0.0489	0.0848	0.2457	0.1561	0.0875	0.1512
	CP	0.9375	0.1875	0.9650	0.9000	0.9200	0.1000	0.0600	0.9375	0.3050
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0042	-0.0334	-0.0013	0.0174	0.0141	-0.0568	-0.0262	-0.0070	-0.0362
	SD	0.0781	0.0645	0.0952	0.0414	0.0806	0.0863	0.0492	0.0950	0.0684
	RMSE	0.0781	0.0725	0.0951	0.0448	0.0817	0.1032	0.0556	0.0951	0.0773
	CP	0.9150	0.8250	0.9075	0.9225	0.9300	0.7375	0.8600	0.9050	0.8050
$\hat{\theta}_{ml,nT}^{S}$	Bias	0.0486	-0.1912	-0.0236	-0.0220	-0.0165	-0.4582			
	SD	0.0584	0.0415	0.0701	0.0363	0.0702	0.0379			
	RMSE	0.0760	0.1956	0.0738	0.0424	0.0721	0.4597			
	CP	0.8700	0.0025	0.9350	0.9000	0.9300	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	0.0749	-0.1042	-0.0438	-0.0315	-0.0321	-0.3927			
	SD	0.0599	0.0449	0.0758	0.0359	0.0697	0.0424			
	RMSE	0.0959	0.1134	0.0874	0.0477	0.0767	0.3950			
	CP	0.7175	0.2300	0.8675	0.8550	0.9250	0.0000			

$(n,T)=(49,10)$ $(\lambda,\rho)=(-0.2,-0.2)$		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	-0.0728	-0.1577	0.0708	0.0287	0.0283	-0.2316	-0.1497	-0.0165	-0.1383
	SD	0.0668	0.0551	0.0799	0.0413	0.0793	0.0734	0.0443	0.0860	0.0612
	RMSE	0.0988	0.1670	0.1066	0.0503	0.0841	0.2429	0.1561	0.0875	0.1512
	CP	0.7825	0.1375	0.8550	0.8800	0.9175	0.0950	0.0600	0.9375	0.3050
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0245	-0.0385	0.0310	0.0189	0.0118	-0.0556	-0.0261	-0.0071	-0.0362
	SD	0.0730	0.0631	0.0914	0.0409	0.0794	0.0862	0.0492	0.0949	0.0684
	RMSE	0.0769	0.0739	0.0964	0.0450	0.0802	0.1025	0.0556	0.0951	0.0773
	CP	0.8950	0.7875	0.8850	0.9100	0.9325	0.7350	0.8600	0.9075	0.8050
$\hat{\theta}_{ml,nT}^{S}$	Bias	-0.0749	-0.1940	0.0868	-0.0131	-0.0103	-0.4424			
	SD	0.0581	0.0417	0.0695	0.0370	0.0726	0.0395			
	RMSE	0.0947	0.1985	0.1111	0.0392	0.0733	0.4441			
	CP	0.6900	0.0000	0.7375	0.9375	0.9375	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	-0.0516	-0.1029	0.0591	-0.0221	-0.0247	-0.3741			
	SD	0.0597	0.0452	0.0759	0.0367	0.0720	0.0443			
	RMSE	0.0788	0.1123	0.0961	0.0428	0.0760	0.3767			
	CP	0.8100	0.2500	0.8400	0.9025	0.9225	0.0000			

$(n,T)=(49,30)$ $(\lambda,\rho)=(0.2,0.2)$		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	-0.0399	-0.0446	0.0373	0.0232	0.0449	-0.0751	-0.0487	0.0020	-0.0571
	SD	0.0391	0.0323	0.0450	0.0241	0.0443	0.0433	0.0271	0.0479	0.0379
	RMSE	0.0558	0.0550	0.0584	0.0334	0.0630	0.0866	0.0557	0.0479	0.0685
	CP	0.8425	0.7025	0.8425	0.8225	0.8375	0.5300	0.4925	0.9275	0.6000
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0018	-0.0048	-0.0028	0.0041	0.0074	-0.0097	-0.0047	0.0033	-0.0070
	SD	0.0399	0.0334	0.0467	0.0239	0.0440	0.0455	0.0281	0.0493	0.0399
	RMSE	0.0399	0.0337	0.0467	0.0242	0.0446	0.0465	0.0284	0.0493	0.0405
	CP	0.9550	0.9250	0.9550	0.9550	0.9575	0.8975	0.9075	0.9300	0.9125
$\hat{\theta}_{ml,nT}^{S}$	Bias	0.0237	-0.1118	-0.0015	-0.0058	0.0319	-0.3841			
	SD	0.0308	0.0233	0.0378	0.0214	0.0391	0.0250			
	RMSE	0.0388	0.1142	0.0378	0.0221	0.0504	0.3849			
	CP	0.8650	0.0000	0.9575	0.9350	0.8675	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	0.0576	-0.0820	-0.0362	-0.0152	0.0117	-0.3542			
	SD	0.0311	0.0239	0.0387	0.0213	0.0389	0.0262			
	RMSE	0.0655	0.0854	0.0529	0.0261	0.0406	0.3552			

	CP	0.5500	0.0625	0.8300	0.8525	0.9450	0.0000			

$(n,T)=(49,30)$ $(\lambda,\rho)=(0.2,-0.2)$		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	-0.0571	-0.0542	0.0363	0.0171	0.0310	-0.0927	-0.0489	-0.0001	-0.0571
	SD	0.0424	0.0323	0.0498	0.0235	0.0435	0.0431	0.0271	0.0480	0.0379
	RMSE	0.0711	0.0631	0.0615	0.0291	0.0534	0.1022	0.0559	0.0480	0.0685
	CP	0.7475	0.5875	0.9000	0.8725	0.8875	0.3575	0.4925	0.9275	0.6000
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0037	-0.0057	-0.0026	0.0030	0.0044	-0.0133	-0.0048	0.0025	-0.0070
	SD	0.0443	0.0337	0.0521	0.0236	0.0438	0.0460	0.0281	0.0495	0.0399
	RMSE	0.0444	0.0341	0.0521	0.0237	0.0440	0.0478	0.0284	0.0495	0.0405
	CP	0.9600	0.9300	0.9500	0.9475	0.9500	0.8825	0.9075	0.9275	0.9125
$\hat{\theta}_{ml,nT}^{S}$	Bias	-0.1131	-0.1210	0.0699	-0.0224	-0.0003	-0.4064			
	SD	0.0324	0.0230	0.0396	0.0210	0.0381	0.0241			
	RMSE	0.1177	0.1231	0.0803	0.0307	0.0380	0.4071			
	CP	0.0800	0.0000	0.6150	0.7800	0.9475	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	-0.0790	-0.0866	0.0406	-0.0308	-0.0169	-0.3759			
	SD	0.0329	0.0236	0.0407	0.0209	0.0379	0.0253			
	RMSE	0.0856	0.0898	0.0575	0.0372	0.0415	0.3768			
	CP	0.3100	0.0350	0.8075	0.6450	0.9500	0.0000			

$(n,T)=(49,30)$ $(\lambda,\rho)=(-0.2,0.2)$		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	-0.0333	-0.0481	0.0125	0.0156	0.0249	-0.0895	-0.0490	-0.0007	-0.0571
	SD	0.0426	0.0322	0.0502	0.0233	0.0427	0.0430	0.0271	0.0482	0.0379
	RMSE	0.0540	0.0578	0.0517	0.0280	0.0494	0.0993	0.0560	0.0481	0.0685
	CP	0.8750	0.6450	0.9475	0.8800	0.9025	0.3925	0.4900	0.9275	0.6000
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0007	-0.0060	-0.0056	0.0025	0.0032	-0.0137	-0.0048	0.0024	-0.0070
	SD	0.0441	0.0331	0.0523	0.0232	0.0426	0.0456	0.0281	0.0497	0.0399
	RMSE	0.0441	0.0336	0.0526	0.0233	0.0427	0.0475	0.0285	0.0497	0.0405
	CP	0.9525	0.9175	0.9350	0.9450	0.9550	0.8800	0.9075	0.9250	0.9125
$\hat{\theta}_{ml,nT}^{S}$	Bias	0.0482	-0.1169	-0.0310	-0.0287	-0.0183	-0.4090			
	SD	0.0314	0.0227	0.0398	0.0206	0.0374	0.0243			
	RMSE	0.0575	0.1191	0.0504	0.0353	0.0416	0.4097			
	CP	0.7200	0.0000	0.8925	0.6700	0.9475	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	0.0767	-0.0868	-0.0465	-0.0356	-0.0319	-0.3760			
	SD	0.0322	0.0232	0.0411	0.0206	0.0374	0.0256			
	RMSE	0.0832	0.0898	0.0620	0.0411	0.0492	0.3769			
	CP	0.3475	0.0275	0.7900	0.5500	0.8725	0.0000			

$(n,T)=(49,30)$ $(\lambda,\rho)=(-0.2,-0.2)$		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	-0.0469	-0.0544	0.0276	0.0151	0.0194	-0.0922	-0.0490	-0.0010	-0.0571
	SD	0.0385	0.0313	0.0455	0.0226	0.0415	0.0435	0.0272	0.0482	0.0379
	RMSE	0.0607	0.0627	0.0532	0.0272	0.0458	0.1019	0.0560	0.0481	0.0685
	CP	0.7775	0.5225	0.9275	0.8800	0.9225	0.3625	0.4875	0.9325	0.6000
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0021	-0.0076	0.0003	0.0025	0.0015	-0.0137	-0.0048	0.0022	-0.0070
	SD	0.0404	0.0322	0.0477	0.0226	0.0412	0.0464	0.0281	0.0497	0.0399
	RMSE	0.0404	0.0330	0.0477	0.0227	0.0412	0.0483	0.0285	0.0497	0.0405
	CP	0.9625	0.9325	0.9450	0.9525	0.9475	0.8850	0.9075	0.9250	0.9125
$\hat{\theta}_{ml,nT}^{S}$	Bias	-0.0800	-0.1160	0.0547	-0.0193	-0.0102	-0.3899			
	SD	0.0308	0.0229	0.0381	0.0208	0.0384	0.0253			
	RMSE	0.0857	0.1183	0.0667	0.0283	0.0397	0.3907			
	CP	0.2525	0.0000	0.7225	0.8175	0.9500	0.0000			
	Bias	-0.0551	-0.0819	0.0363	-0.0256	-0.0212	-0.3549			
	SD	0.0317	0.0234	0.0395	0.0208	0.0384	0.0267			

$\hat{\theta}^{S,c}_{ml,nT}$	RMSE	0.0636	0.0852	0.0536	0.0330	0.0438	0.3559			
	CP	0.5750	0.0450	0.8450	0.7400	0.9275	0.0000			

$(n,T)=(81,10)$ $(\lambda,\rho)=(0.2,0.2)$		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	0.0012	-0.1486	0.0507	0.0323	0.0481	-0.2104	-0.1486	-0.0094	-0.1332
	SD	0.0520	0.0440	0.0621	0.0366	0.0686	0.0557	0.0342	0.0703	0.0447
	RMSE	0.0520	0.1550	0.0801	0.0488	0.0837	0.2177	0.1525	0.0708	0.1405
	CP	0.9325	0.0625	0.8225	0.8250	0.8825	0.0275	0.0100	0.9375	0.1550
$\hat{\theta}^{c}_{ml,nT}$	Bias	0.0104	-0.0309	0.0011	0.0172	0.0167	-0.0421	-0.0245	-0.0001	-0.0367
	SD	0.0552	0.0503	0.0709	0.0359	0.0689	0.0650	0.0381	0.0765	0.0497
	RMSE	0.0561	0.0589	0.0708	0.0398	0.0708	0.0774	0.0452	0.0764	0.0618
	CP	0.9075	0.7950	0.9025	0.9025	0.9250	0.7825	0.8575	0.9025	0.8000
$\hat{\theta}^{S}_{ml,nT}$	Bias	0.0419	-0.1878	0.0164	-0.0052	0.0217	-0.4297			
	SD	0.0458	0.0331	0.0560	0.0319	0.0609	0.0315			
	RMSE	0.0620	0.1907	0.0582	0.0323	0.0646	0.4308			
	CP	0.7975	0.0000	0.9225	0.9075	0.9025	0.0000			
$\hat{\theta}^{S,c}_{ml,nT}$	Bias	0.0617	-0.0984	-0.0367	-0.0148	0.0063	-0.3679			
	SD	0.0460	0.0358	0.0600	0.0314	0.0597	0.0349			
	RMSE	0.0769	0.1047	0.0703	0.0346	0.0600	0.3695			
	CP	0.6875	0.1400	0.8575	0.8900	0.9325	0.0000			

$(n,T)=(81,10)$ $(\lambda,\rho)=(0.2,-0.2)$		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	-0.0620	-0.1609	0.0917	0.0233	0.0326	-0.2351	-0.1488	-0.0110	-0.1332
	SD	0.0577	0.0441	0.0712	0.0359	0.0672	0.0551	0.0342	0.0702	0.0447
	RMSE	0.0847	0.1668	0.1160	0.0428	0.0746	0.2415	0.1527	0.0710	0.1405
	CP	0.7850	0.0350	0.6725	0.8625	0.8900	0.0075	0.0100	0.9325	0.1550
$\hat{\theta}^{c}_{ml,nT}$	Bias	-0.0134	-0.0356	0.0199	0.0159	0.0135	-0.0553	-0.0247	-0.0012	-0.0368
	SD	0.0631	0.0515	0.0813	0.0354	0.0683	0.0655	0.0380	0.0767	0.0497
	RMSE	0.0644	0.0625	0.0836	0.0388	0.0696	0.0857	0.0453	0.0766	0.0618
	CP	0.8975	0.7800	0.8775	0.8975	0.9175	0.7350	0.8575	0.9050	0.7975
$\hat{\theta}^{S}_{ml,nT}$	Bias	-0.0967	-0.1984	0.1097	-0.0204	-0.0060	-0.4514			
	SD	0.0483	0.0330	0.0595	0.0314	0.0593	0.0302			
	RMSE	0.1080	0.2011	0.1248	0.0374	0.0595	0.4524			
	CP	0.4250	0.0000	0.5075	0.8500	0.9300	0.0000			
$\hat{\theta}^{S,c}_{ml,nT}$	Bias	-0.0767	-0.1049	0.0551	-0.0298	-0.0213	-0.3911			
	SD	0.0487	0.0357	0.0638	0.0309	0.0585	0.0335			
	RMSE	0.0909	0.1108	0.0842	0.0429	0.0622	0.3925			
	CP	0.5875	0.0875	0.7875	0.7825	0.9225	0.0000			

$(n,T)=(81,10)$ $(\lambda,\rho)=(-0.2,0.2)$		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	0.0010	-0.1509	0.0062	0.0212	0.0223	-0.2259	-0.1488	-0.0115	-0.1332
	SD	0.0572	0.0431	0.0701	0.0357	0.0670	0.0545	0.0342	0.0703	0.0447
	RMSE	0.0571	0.1569	0.0703	0.0415	0.0705	0.2324	0.1527	0.0711	0.1405
	CP	0.9500	0.0475	0.9425	0.8750	0.9000	0.0100	0.0100	0.9350	0.1550
$\hat{\theta}^{c}_{ml,nT}$	Bias	0.0007	-0.0321	-0.0036	0.0144	0.0093	-0.0504	-0.0247	-0.0017	-0.0368
	SD	0.0614	0.0494	0.0794	0.0349	0.0672	0.0642	0.0380	0.0768	0.0497
	RMSE	0.0614	0.0588	0.0794	0.0377	0.0677	0.0816	0.0453	0.0767	0.0618
	CP	0.9250	0.7975	0.9150	0.9050	0.9200	0.7500	0.8575	0.9050	0.7975
$\hat{\theta}^{S}_{ml,nT}$	Bias	0.0618	-0.1912	-0.0271	-0.0258	-0.0228	-0.4517			
	SD	0.0471	0.0320	0.0590	0.0310	0.0584	0.0302			
	RMSE	0.0777	0.1938	0.0649	0.0403	0.0626	0.4528			
	CP	0.7175	0.0000	0.9050	0.8225	0.9200	0.0000			
	Bias	0.0783	-0.1030	-0.0452	-0.0344	-0.0358	-0.3901			

$\hat{\theta}^{S,c}_{ml,nT}$	SD	0.0478	0.0344	0.0634	0.0305	0.0575	0.0336			
	RMSE	0.0917	0.1085	0.0778	0.0459	0.0677	0.3915			
	CP	0.6050	0.0850	0.8450	0.7325	0.8875	0.0000			

$(n,T)=(81,10)$ $(\lambda,\rho)=(-0.2,-0.2)$		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	-0.0594	-0.1564	0.0596	0.0250	0.0237	-0.2206	-0.1488	-0.0115	-0.1332
	SD	0.0529	0.0422	0.0665	0.0349	0.0658	0.0549	0.0342	0.0703	0.0447
	RMSE	0.0795	0.1620	0.0892	0.0429	0.0698	0.2273	0.1527	0.0711	0.1405
	CP	0.7750	0.0250	0.8200	0.8550	0.9050	0.0100	0.0100	0.9350	0.1550
$\hat{\theta}^{c}_{ml,nT}$	Bias	-0.0202	-0.0362	0.0222	0.0161	0.0085	-0.0481	-0.0247	-0.0017	-0.0368
	SD	0.0575	0.0481	0.0752	0.0340	0.0652	0.0646	0.0380	0.0768	0.0497
	RMSE	0.0609	0.0601	0.0783	0.0376	0.0656	0.0804	0.0453	0.0767	0.0618
	CP	0.8975	0.7825	0.8800	0.8975	0.9175	0.7825	0.8575	0.9050	0.7975
$\hat{\theta}^{S}_{ml,nT}$	Bias	-0.0639	-0.1930	0.0787	-0.0169	-0.0148	-0.4357			
	SD	0.0463	0.0320	0.0577	0.0313	0.0595	0.0313			
	RMSE	0.0788	0.1956	0.0975	0.0355	0.0612	0.4368			
	CP	0.6625	0.0000	0.6675	0.8700	0.9225	0.0000			
$\hat{\theta}^{S,c}_{ml,nT}$	Bias	-0.0488	-0.1014	0.0533	-0.0252	-0.0278	-0.3716			
	SD	0.0470	0.0343	0.0624	0.0309	0.0587	0.0348			
	RMSE	0.0677	0.1070	0.0820	0.0398	0.0649	0.3733			
	CP	0.7550	0.0925	0.7925	0.8300	0.9125	0.0000			

$(n,T)=(81,30)$ $(\lambda,\rho)=(0.2,0.2)$		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	-0.0220	-0.0453	0.0336	0.0184	0.0354	-0.0686	-0.0479	-0.0007	-0.0501
	SD	0.0327	0.0249	0.0354	0.0194	0.0382	0.0333	0.0189	0.0362	0.0272
	RMSE	0.0394	0.0517	0.0488	0.0267	0.0520	0.0763	0.0515	0.0361	0.0570
	CP	0.8625	0.5700	0.8475	0.8075	0.8025	0.4125	0.3075	0.9425	0.5400
$\hat{\theta}^{c}_{ml,nT}$	Bias	-0.0002	-0.0038	0.0008	0.0020	0.0053	-0.0062	-0.0034	0.0008	-0.0071
	SD	0.0333	0.0257	0.0367	0.0192	0.0381	0.0348	0.0196	0.0371	0.0284
	RMSE	0.0332	0.0260	0.0366	0.0193	0.0385	0.0353	0.0199	0.0371	0.0292
	CP	0.9225	0.9250	0.9300	0.9250	0.9400	0.9225	0.9350	0.9425	0.9275
$\hat{\theta}^{S}_{ml,nT}$	Bias	0.0395	-0.1121	-0.0042	-0.0096	0.0258	-0.3789			
	SD	0.0264	0.0179	0.0299	0.0172	0.0340	0.0187			
	RMSE	0.0475	0.1136	0.0301	0.0197	0.0426	0.3794			
	CP	0.6175	0.0000	0.9450	0.8800	0.8375	0.0000			
$\hat{\theta}^{S,c}_{ml,nT}$	Bias	0.0606	-0.0813	-0.0331	-0.0171	0.0110	-0.3526			
	SD	0.0266	0.0184	0.0305	0.0171	0.0339	0.0196			
	RMSE	0.0662	0.0834	0.0450	0.0242	0.0357	0.3532			
	CP	0.3125	0.0075	0.7950	0.8025	0.9225	0.0000			

$(n,T)=(81,30)$ $(\lambda,\rho)=(0.2,-0.2)$		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	-0.0409	-0.0526	0.0362	0.0143	0.0261	-0.0814	-0.0481	-0.0023	-0.0501
	SD	0.0358	0.0247	0.0399	0.0191	0.0379	0.0326	0.0189	0.0363	0.0272
	RMSE	0.0543	0.0581	0.0538	0.0239	0.0460	0.0877	0.0516	0.0363	0.0570
	CP	0.7600	0.4350	0.8500	0.8650	0.8550	0.2600	0.2975	0.9400	0.5400
$\hat{\theta}^{c}_{ml,nT}$	Bias	-0.0030	-0.0045	0.0023	0.0018	0.0045	-0.0085	-0.0035	0.0003	-0.0071
	SD	0.0370	0.0258	0.0416	0.0190	0.0381	0.0344	0.0196	0.0373	0.0284
	RMSE	0.0370	0.0261	0.0416	0.0191	0.0383	0.0354	0.0199	0.0372	0.0292
	CP	0.9375	0.9400	0.9225	0.9250	0.9350	0.9200	0.9350	0.9400	0.9275
$\hat{\theta}^{S}_{ml,nT}$	Bias	-0.0987	-0.1198	0.0695	-0.0251	-0.0043	-0.4001			
	SD	0.0283	0.0175	0.0318	0.0170	0.0332	0.0180			
	RMSE	0.1026	0.1210	0.0764	0.0303	0.0334	0.4005			
	CP	0.0425	0.0000	0.3800	0.6375	0.9300	0.0000			

$\hat{\theta}^{S,c}_{ml,nT}$	Bias	-0.0772	-0.0857	0.0442	-0.0318	-0.0164	-0.3737			
	SD	0.0286	0.0181	0.0325	0.0169	0.0332	0.0188			
	RMSE	0.0823	0.0876	0.0549	0.0360	0.0369	0.3742			
	CP	0.1625	0.0050	0.7275	0.5025	0.9025	0.0000			

$(n,T)=(81,30)$ $(\lambda,\rho)=(-0.2,0.2)$		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	-0.0197	-0.0471	0.0126	0.0132	0.0214	-0.0777	-0.0481	-0.0028	-0.0501
	SD	0.0360	0.0245	0.0394	0.0189	0.0375	0.0324	0.0189	0.0363	0.0272
	RMSE	0.0410	0.0531	0.0414	0.0230	0.0431	0.0841	0.0517	0.0364	0.0570
	CP	0.8825	0.5075	0.9375	0.8725	0.8700	0.3050	0.2975	0.9375	0.5400
$\hat{\theta}^{c}_{ml,nT}$	Bias	-0.0011	-0.0038	-0.0005	0.0016	0.0037	-0.0079	-0.0036	0.0001	-0.0071
	SD	0.0370	0.0254	0.0410	0.0188	0.0374	0.0340	0.0196	0.0373	0.0284
	RMSE	0.0370	0.0256	0.0410	0.0188	0.0376	0.0349	0.0199	0.0373	0.0292
	CP	0.9325	0.9375	0.9250	0.9250	0.9350	0.9125	0.9350	0.9375	0.9275
$\hat{\theta}^{S}_{ml,nT}$	Bias	0.0598	-0.1161	-0.0300	-0.0306	-0.0210	-0.4016			
	SD	0.0278	0.0173	0.0314	0.0168	0.0328	0.0180			
	RMSE	0.0659	0.1173	0.0434	0.0349	0.0389	0.4020			
	CP	0.3775	0.0000	0.8700	0.5200	0.8800	0.0000			
$\hat{\theta}^{S,c}_{ml,nT}$	Bias	0.0776	-0.0852	-0.0423	-0.0365	-0.0315	-0.3737			
	SD	0.0282	0.0178	0.0322	0.0167	0.0328	0.0188			
	RMSE	0.0826	0.0870	0.0531	0.0401	0.0454	0.3741			
	CP	0.1875	0.0050	0.7500	0.3675	0.7900	0.0000			

$(n,T)=(81,30)$ $(\lambda,\rho)=(-0.2,-0.2)$		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	-0.0371	-0.0515	0.0256	0.0138	0.0184	-0.0781	-0.0481	-0.0030	-0.0501
	SD	0.0322	0.0238	0.0366	0.0184	0.0363	0.0322	0.0189	0.0363	0.0272
	RMSE	0.0491	0.0567	0.0446	0.0230	0.0407	0.0845	0.0517	0.0364	0.0570
	CP	0.7800	0.3900	0.8825	0.8625	0.8825	0.3175	0.2975	0.9375	0.5400
$\hat{\theta}^{c}_{ml,nT}$	Bias	-0.0035	-0.0056	0.0042	0.0019	0.0029	-0.0081	-0.0036	-0.0000	-0.0071
	SD	0.0333	0.0245	0.0382	0.0183	0.0361	0.0339	0.0196	0.0373	0.0284
	RMSE	0.0334	0.0251	0.0384	0.0183	0.0361	0.0348	0.0199	0.0373	0.0292
	CP	0.9400	0.9325	0.9475	0.9300	0.9325	0.9200	0.9350	0.9375	0.9275
$\hat{\theta}^{S}_{ml,nT}$	Bias	-0.0696	-0.1144	0.0547	-0.0212	-0.0122	-0.3827			
	SD	0.0269	0.0173	0.0307	0.0170	0.0335	0.0186			
	RMSE	0.0747	0.1157	0.0628	0.0271	0.0356	0.3831			
	CP	0.2175	0.0000	0.5550	0.7300	0.9000	0.0000			
$\hat{\theta}^{S,c}_{ml,nT}$	Bias	-0.0538	-0.0807	0.0403	-0.0267	-0.0207	-0.3531			
	SD	0.0274	0.0177	0.0317	0.0169	0.0335	0.0195			
	RMSE	0.0604	0.0826	0.0512	0.0316	0.0394	0.3537			
	CP	0.4225	0.0050	0.7300	0.6000	0.8800	0.0000			

Table 2: Descriptive statistics: counties in North Carolina

Variables	Mean	Standard deviation	Minimum	Maximum
Public safety spending ($\$\times10^6$)	20.5504	26.3425	0	237.3665
Total revenue ($\$\times10^6$)	126.2299	216.4341	0	1786.4493
Proportion on total expenditure	0.193	0.060	0	0.448
Population ($\times10^3$)	94.4273	140.5532	4.1430	1035.6050
Land area (km^2)	1259.181	497.481	446.701	2457.924
Population density ($/km^2$)	74.7630	99.8999	3.3976	763.9331
Median ages	40.0793	4.5780	23.9	51.3
Median household income ($\$\times10^4$)	4.1410	0.7681	2.5107	7.0620
Distance (km)	248.1450	147.8367	12.2632	751.9034
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.

Table 3.A: Model estimation I. $d_c = 50km$

	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	4935.0	4934.6
BIC	5610.4	5610.0
No. of Obs	1200	1200
No. of "neighbors"	4.3400 (1.4229)	4.3400 (1.4229)
Cutoff distance (km)	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%, 5%, and 1% levels are respectively marked by "*", "**", and "***".

Table 3.B: Model estimation II. $d_c = 65km$

	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 (km)	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%, 5%, and 1% levels are respectively marked by "*", "**", and "***".

Table 3.C: Model estimation III. $d_c = 80km$

	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 (km)	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%, 5%, and 1% levels are respectively marked by "*", "**", and "***".

Table 3.D: Model estimation IV. $d_c = 95km$

	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 (km)	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%, 5%, and 1% levels are respectively marked by “*”, “**”, and “***”.

Table 4. The direct and cumulative effect of increasing the total revenue (per capita) by one thousand dollars

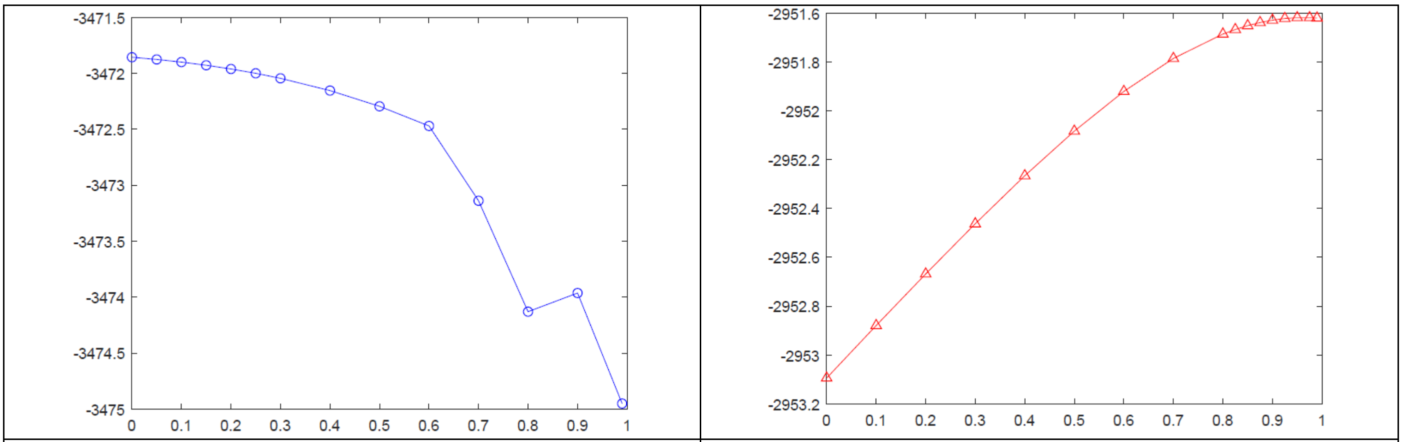
		Iredell county	Dare county
Direct	Own effect	0.1239	0.1239
	Neighboring effect	-0.0039	-0.0222
Cumulative	Own effect	0.1046	0.1045
	Neighboring effect	-0.0030	-0.0167
No. of neighbors		17	3

Table 5. Changes of social welfare if a county’s 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)).

Figure 1.A: Selection of δ via likelihood measures



Myopic	Forward-looking

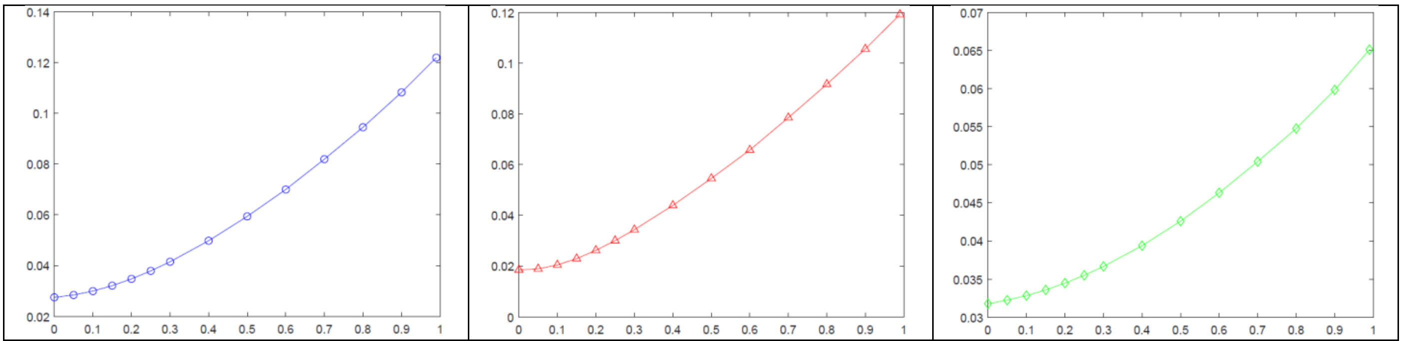
Note: We show two representative cases:

(i) Myopic: $\delta_0 = 0$, $(n, T) = (81, 30)$ and $K = 2$

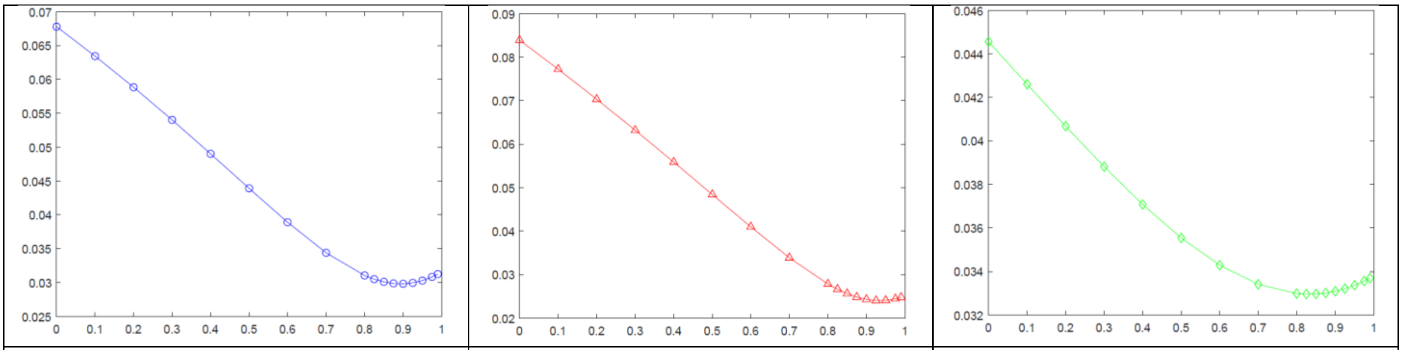
(ii) Forward-looking: $\delta_0 = 0.95$, $(n, T) = (81, 30)$ and $K = 2$.

We set $\lambda_0 = 0.2$, $\gamma_0 = 0.4$, and $\rho_0 = 0.4$, and other circumstances are the same as the main simulation. The x-axis shows δ's while the y-axis reports the sample log-likelihood.

Figure 1.B: RMSEs in estimating λ_0, γ_0, and ρ_0 for misspecified δ



λ_0, myopic model	γ_0, myopic model	ρ_0, myopic model



λ_0, forward-looking model	γ_0, forward-looking model	ρ_0, forward-looking model

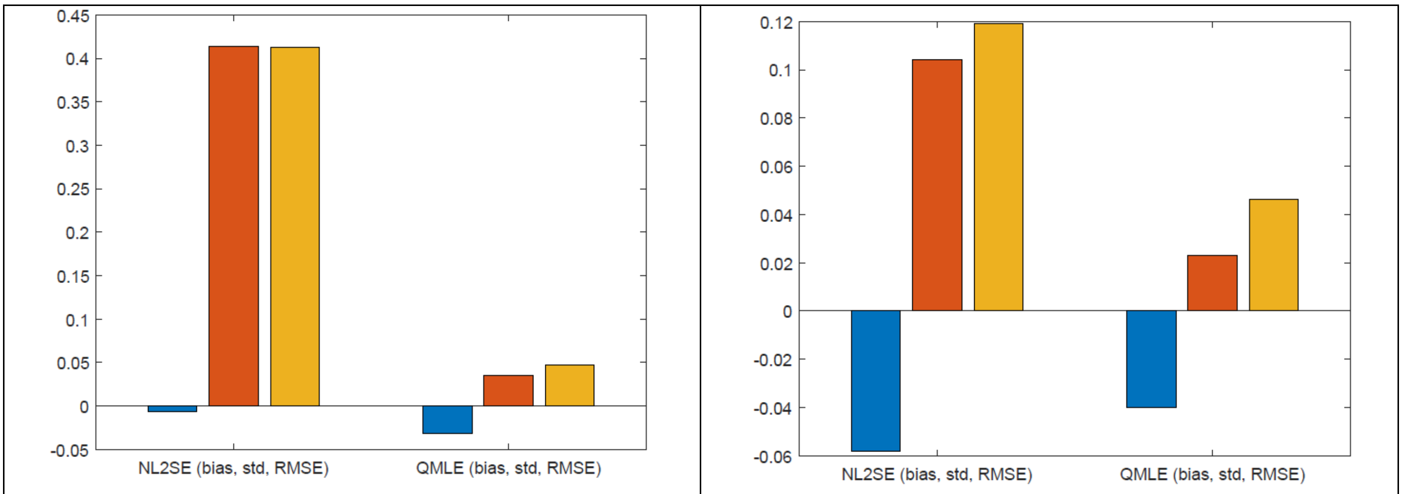
Note: We show two representative cases:

(i) Myopic: $\delta_0 = 0$, $(n,T) = (81,30)$ and $K = 2$

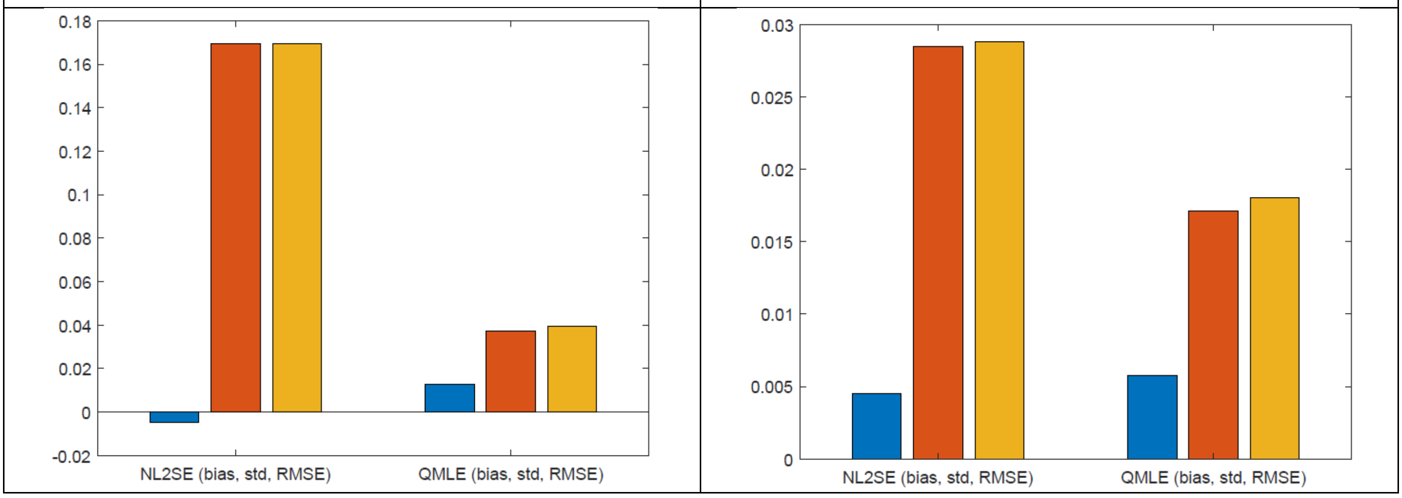
(ii) Forward-looking: $\delta_0 = 0.95$, $(n,T) = (81,30)$ and $K = 2$.

We set $\lambda_0 = 0.2$, $\gamma_0 = 0.4$, and $\rho_0 = 0.4$, and other circumstances are the same as the main simulation. The x-axis shows δ's while the y-axis reports the RMSEs.

Figure 2: Performance comparison: QMLE and NL2SE



λ_0	γ_0



ρ_0	$\beta_{1,1,0}$

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

Figure 3: Neighbors of the two counties (based on $d_c=80$)

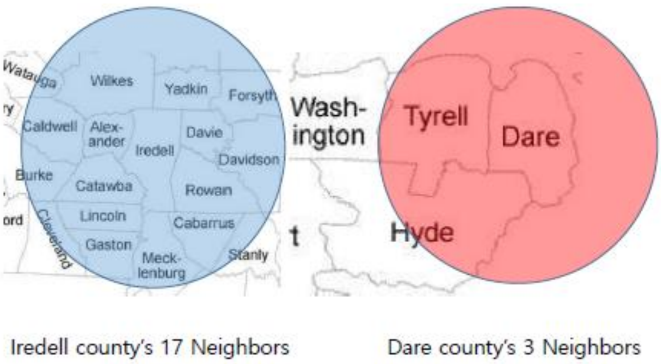
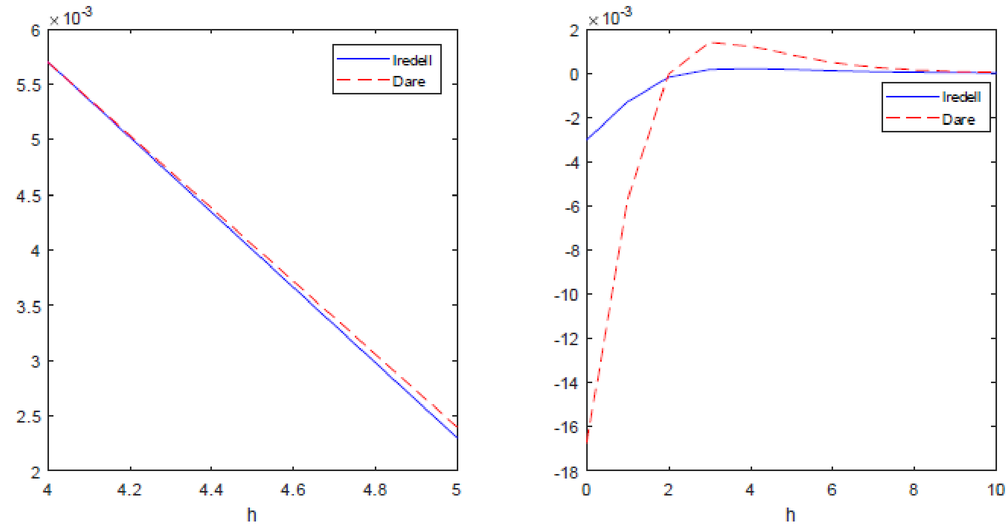


Figure 4: Impulse response functions: own effects (left) and neighboring effects (right)



Supplement to "Spatial dynamic models with intertemporal optimization: specification and estimation"

Hanbat Jeong* and Lung-fei Lee†

August, 2018

Abstract

This document contains some technical and detailed theoretical analyses, and additional Monte Carlo results for Jeong and Lee's (2018).

1 Derivation of the Markov perfect equilibrium (MPE) equation

Recall the Bellman equation of our model is

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

where $Y^*_{-i,t}(Y_{n,t-1},\boldsymbol{\eta}_{nt}) = \left(y^*_{1t}(Y_{n,t-1},\boldsymbol{\eta}_{nt}), \cdots, y^*_{i-1,t}(Y_{n,t-1},\boldsymbol{\eta}_{nt}), y^*_{i+1,t}(Y_{n,t-1},\boldsymbol{\eta}_{nt}), \cdots, y^*_{nt}(Y_{n,t-1},\boldsymbol{\eta}_{nt})\right)'$, for $i = 1, \cdots, n$, and an arbitrary t. Throughout this section, we take the following assumption.

Assumption 1.1 *For each t, $\boldsymbol{\eta}^v_{n,t+1} = \Pi_n \boldsymbol{\eta}^v_{nt} + \xi_{n,t+1}$ where $\|\Pi_n\| < 1$, $\|\cdot\|$ denotes a proper matrix norm, $\boldsymbol{\eta}^v_{nt} = (\eta^v_{1t}, \cdots, \eta^v_{nt})'$, $E_t\left(\xi_{n,t+1}\right) = 0$ and $E_t\left(\xi_{n,t+1}\xi'_{n,t+1}\right) = \Omega_\xi$ which is positive definite.*

1.1 Step 1: Generation of $V_i^{(j)}(\cdot)$'s

Derivations are in our main draft. Since all entries in W_n are finite, by Assumption 1.1, all entries of $Q_i^{(j)}$, $L_i^{(j)}$, $G_i^{(j)}$, and $c_i^{(j)}$ (for each i and for $j = 1, 2, \cdots$) which are functions of W_n and $(\delta, \lambda_0, \gamma_0, \rho_0, \Pi_n, \Omega_\xi)$

*Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. E-mail: jeong.181@osu.edu

†Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. E-mail: lee.1777@osu.edu

generated by Step 1 are finite, and not relevant to $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$. $A_n^{(j)}$ and $B_n^{(j)}$ can be evaluated by using $Q_i^{(j)}$ and $L_i^{(j)}$.

1.2 Step 2: Continuity of $\mathcal{T}$

To investigate $\mathcal{T}$, we review several mathematical results. They are imported from Stokey et al. (1989) and Fuente (2000). We will reproduce those arguments in our framework with simple sketches of proofs. Note that arguments of $\mathcal{T}$ are $V_i^{(j)}(\cdot)$'s $j = 0, 1, 2, \cdots$. The domain of $V_i^{(j)}(\cdot)$'s is a subset of $\mathbb{R}^{2n}$ denoted by $(\chi_y)^n \times (\chi_\eta)^n$ where χ_y, $\chi_\eta \subseteq \mathbb{R}$. By **Step 1**, we can claim that, for any continuous and bounded function $V_i^{(0)}(\cdot)$, $\left\{V_i^{(j)}(\cdot)\right\}_{j=0,1,2,\cdots} \in \mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n\right)$ where $\mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n\right)$ is the set of bounded and continuous functions from $(\chi_y)^n \times (\chi_\eta)^n$ to $\mathbb{R}$ equipped with the sup norm (or the uniform norm),

$$\|V\|_u = \sup_{(Y_{n,t-1}, \boldsymbol{\eta}_t) \in (\chi_y)^n \times (\chi_\eta)^n} |V(Y_{n,t-1}, \boldsymbol{\eta}_{nt})|$$

for $V \in \mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n\right)$. Note that $\mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n\right)$ is a complete normed vector space equipped with $\|\cdot\|_u$. Hence, if $V_i^{(j)} \to V_i^*$ as $j \to \infty$, a candidate limit function $V_i^*(\cdot)$ is also bounded and continuous. The theorem below verifies the continuity of $\mathcal{T}V_i^{(j)}(\cdot)$ and the existence and uniqueness of agent i's optimal action.

Theorem 1.1 (Theorem of the maximum) *For all* $j = 1, 2, \cdots$,

$$u_i\left(y_{it}, Y_{-i,t}^{*(j+1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}), Y_{n,t-1}, \eta_{it}\right) + \delta E_t\left(V_i^{(j)}\left(y_{it}, Y_{-i,t}^{*(j+1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}), \boldsymbol{\eta}_{n,t+1}\right)\right) \tag{2}$$

is a continuous and bounded function from $\underbrace{\chi_y}_{=\text{ decision space}} \times \underbrace{\left((\chi_y)^n \times (\chi_\eta)^n\right)}_{=\text{ state space}}$ *to* $\mathbb{R}$.

The set of optimal decisions

$$\Xi_i^{(j+1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) = \left\{ y_{it} : \left(\begin{array}{c} u_i\left(y_{it}, Y_{-i,t}^{*(j+1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}), Y_{n,t-1}, \eta_{it}\right) \\ +\delta E_t\left(V_i^{(j)}\left(y_{it}, Y_{-i,t}^{*(j+1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}), \boldsymbol{\eta}_{n,t+1}\right)\right) \end{array} \right) = \mathcal{T}V_i^{(j)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) \right\} \tag{3}$$

is a singleton and $y_{it}^{*(j+1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ *is a continuous function of* $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$. *Furthermore,* $\mathcal{T}V_i^{(j)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ *is continuous at* $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$.

Theorem 1.1 is a slightly modified version of Theorem 3.6 in Stokey et al. (1989).

Proof of Theorem 1.1. Choose arbitrary $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ and $j \in \{2, 3, \cdots\}$ and fix them. By proceeding the two stages, we will get the desired results.

Stage 1: Since $u_i(y_{it}, Y_{-i,t}, Y_{n,t-1}, \eta_{it})$ is strictly concave in y_{it}, $\Xi_i^{(1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ is a singleton set. From $j=1$, we inductively generate unique $Y_{nt}^{*(j)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) = A_n^{(j)} Y_{n,t-1} + B_n^{(j)} \boldsymbol{\eta}_{nt}$ for $j = 2, 3, \cdots$. It implies that (i) $\Xi_i^{(j)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) = \left\{ y_{it}^{*(j)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) \right\}$ (i.e., a singleton set), and (ii) $Y_{nt}^{*(j)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ is a linear transformation of $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$.

Stage 2: We want to show that $Y_{nt}^{*(j)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$ is also continuous and bounded on $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$. By showing continuity of $\mathcal{T}V_i^{(j)}(\cdot)$, we finish the proof. Arbitrarily choose a convergent sequence $\left\{ \left(Y_{n,t-1}^{(k)}, \boldsymbol{\eta}_{nt}^{(k)} \right) \right\}_k \subset (\chi_y)^n \times (\chi_\eta)^n$ such that $\left(Y_{n,t-1}^{(k)}, \boldsymbol{\eta}_{nt}^{(k)} \right) \to (Y_{n,t-1}, \boldsymbol{\eta}_{nt})$. For each k, we can choose $\left\{ y_{it}^{(k)} \right\} = \Xi_i^{(j+1)} \left(Y_{n,t-1}^{(k)}, \boldsymbol{\eta}_{nt}^{(k)} \right)$ since $\Xi_i^{(j+1)} \left(Y_{n,t-1}^{(k)}, \boldsymbol{\eta}_{nt}^{(k)} \right)$ is a singleton by **Stage 1**. Since $y_{it}^{(k)}$ is continuous of $\left(Y_{n,t-1}^{(k)}, \boldsymbol{\eta}_{nt}^{(k)} \right)$ by **Stage 1**, $y_{it}^{(k)} \to y_{it}$ as $k \to \infty$ because $\left\{ y_{it}^{(k)} \right\} = \Xi_i^{(j+1)} \left(Y_{n,t-1}^{(k)}, \boldsymbol{\eta}_{nt}^{(k)} \right)$ and $\{y_{it}\} = \Xi_i^{(j+1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$. By continuity of (2),

$$
\begin{aligned}
& \mathcal{T}V_i^{(j)} \left(Y_{n,t-1}^{(k)}, \boldsymbol{\eta}_{nt}^{(k)} \right) \\
= \; & u_i \left(y_{it}^{(k)}, Y_{-i,t}^{*(j+1)}(Y_{n,t-1}^{(k)}, \boldsymbol{\eta}_{nt}^{(k)}), Y_{n,t-1}^{(k)}, \eta_{it}^{(k)} \right) + \delta E_t \left(V_i^{(j)} \left(y_{it}^{(k)}, Y_{-i,t}^{*(j+1)}(Y_{n,t-1}^{(k)}, \boldsymbol{\eta}_{nt}^{(k)}), \boldsymbol{\eta}_{n,t+1} \right) \right) \\
\to \; & u_i \left(y_{it}, Y_{-i,t}^{*(j+1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}), Y_{n,t-1}, \eta_{it} \right) + \delta E_t \left(V_i^{(j)} \left(y_{it}, Y_{-i,t}^{*(j+1)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}), \boldsymbol{\eta}_{n,t+1} \right) \right) \\
= \; & \mathcal{T}V_i^{(j)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt})
\end{aligned}
$$

as $k \to \infty$. Hence, $\left\{ \mathcal{T}V_i^{(j)}(Y_{n,t-1}, \boldsymbol{\eta}_{nt}) \right\}$ is continuous at $(Y_{n,t-1}, \boldsymbol{\eta}_{nt})$.■

1.3 Step 3: Contraction mapping and Banach fixed point theorem

Note that $\mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n \right)$ is a complete normed vector space (=Banach space) equipped with the sup norm $\|\cdot\|_u$. Hence, it is also a metric space with the metric $d(f_1, f_2) = \|f_1 - f_2\|_u$ for any $f_1,\ f_2 \in \mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n \right)$. Based on that, we consider the definition of a contraction mapping.

Definition 1 (Contraction mapping) *Note that $\mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n \right)$ is a metric space with $d(f_1, f_2)$ for any $f_1,\ f_2 \in \mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n \right)$. We say $\mathcal{T}$ is a contraction mapping with modulus $\delta \in (0,1)$ if*

$$
d(\mathcal{T}f_1, \mathcal{T}f_2) \leq \delta d(f_1, f_2) \text{ for any } f_1, f_2 \in \mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n \right).
$$

First, we want to show that $\mathcal{T}$ is a contraction mapping with modulus δ. There is an easy way to check whether $\mathcal{T}$ is a contraction mapping. This is called Blackwell's (1965) sufficient conditions.

Proposition 1.2 (Blackwell's sufficient conditions) *Note that $\mathcal{T} : \mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n \right) \to \mathcal{C}\left((\chi_y)^n \times (\chi_\eta)^n \right)$ be an operator.*[1] *Assume $\mathcal{T}$ satisfies*

[1]Indeed, arguments of $\mathcal{T}$ need not be continuous functions to employ Proposition 1.2.

(i) (Monotonicity) For $f_1,\ f_2 \in \mathcal{C}\left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$ *such that* $f_1\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) \leq f_2\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right)$ *for all* $\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) \in \left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$*, we have*

$$\left(\mathcal{T} f_1\right)\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) \leq \left(\mathcal{T} f_2\right)\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) \text{ for all } \left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) \in \left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right).$$

(ii) (Discounting) There exists $\delta \in (0,1)$ *such that for all* $f \in \mathcal{C}\left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$*,*

$$\left(\mathcal{T}\left(f+c\right)\right)\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) \leq \left(\mathcal{T} f\right)\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) + \delta c \text{ for all } \left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) \in \left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$$

where $\left(f+c\right)\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) = f\left(Y_{n,t-1}, \boldsymbol{\eta}_{nt}\right) + c$*.*

Then, $\mathcal{T}$ *is a contraction mapping with modulus* δ*.*

By properties of max operator and the time-discounting factor δ, our $\mathcal{T}$ satisfies the Blackwell sufficient conditions, and hence is a contraction mapping. Then, we can obtain the following proposition.

Proposition 1.3 (Contraction mapping theorem) *Note that* $\mathcal{C}\left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$ *is a Banach space with* $\|\cdot\|_u$ *and* $\mathcal{T}$ *is a contraction mapping with modulus* $\delta \in (0,1)$*. Then, we obtain*

(i) There exists a unique fixed point $V_i \in \mathcal{C}\left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$ *(That is,* $\mathcal{T} V_i = V_i$*).*

(ii) For any $V_i^{(0)} \in \mathcal{C}\left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$*,* $\left\|\mathcal{T}^j V_i^{(0)} - V_i\right\|_u \leq \delta^j \left\|V_i^{(0)} - V_i\right\|_u$ *for* $j = 0, 1, 2, \cdots$*.*

Proposition 1.3 is also called the Banach fixed point theorem. The main idea of proving this follows three steps. First, by arbitrary choosing $V_i^{(0)}$ from $\mathcal{C}\left(\left(\chi_y\right)^n \times \left(\chi_\eta\right)^n\right)$, a Cauchy sequence $\left\{V_i^{(j)}\right\}_j$ can be inductively generated from $\mathcal{T}$, i.e., $V_i^{(1)} = \mathcal{T} V_i^{(0)}$, $V_i^{(2)} = \mathcal{T} V_i^{(1)}$, $\cdots$. Second, $\mathcal{T} V_i = V_i$ is verified by using the discounting property of $\mathcal{T}$. Regardless of a starting point, $\mathcal{T}$ yields ultimate convergence to a unique fixed point V_i. By Proposition 1.3, for our model, V_i^* is the same as the unique fixed point V_i and the vector of optimal decisions is $Y_{nt}^*(\cdot)$. The way of getting $Y_{nt}^*(\cdot)$ and $V_i^*(\cdot)$'s are in the main text.

2 QML estimation

2.1 Model identification

This subsection will discuss the identification issue. From Rothenberg (1971), θ' and θ'' are observationally equivalent if $L_{nT}\left(\theta' \middle| \left\{Y_{nt}\right\}_{t=1}^T\right) = L_{nT}\left(\theta'' \middle| \left\{Y_{nt}\right\}_{t=1}^T\right)$ a.e. Hence, $\theta_0 \in \Theta$ is identifiable if and only if there is no other $\theta \in \Theta$ is observationally equivalent. Identification of θ_0 in this setting comes from the information inequality: (i) for any $\theta \in \Theta$, $E\left(\ln L_{nT}\left(\theta | \{Y_{nt}\}_{t=1}^{T}\right)\right) \leq E\left(\ln L_{nT}\left(\theta_0 | \{Y_{nt}\}_{t=1}^{T}\right)\right)$, and (ii) $L_{nT}\left(\theta | \{Y_{nt}\}_{t=1}^{T}\right) = L_{nT}\left(\theta_0 | \{Y_{nt}\}_{t=1}^{T}\right)$ a.e. in $\{Y_{nt}\}_{t=1}^{T}$ if and only if

$$E\left(\ln L_{nT}\left(\theta | \{Y_{nt}\}_{t=1}^{T}\right)\right) = E\left(\ln L_{nT}\left(\theta_0 | \{Y_{nt}\}_{t=1}^{T}\right)\right).$$

The information inequality comes from concavity of the logarithmic function. The expected (concentrated) log-likelihood function[2] is

$$Q_{nT}\left(\theta_1, \beta, \sigma_\epsilon^2\right) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma_\epsilon^2 + \frac{1}{n}\ln|R_n(\theta_1)| - \frac{1}{2\sigma_\epsilon^2}\frac{1}{nT}\sum_{t=1}^{T} E\left(\tilde{\mathcal{E}}_{nt}'(\theta) J_n \tilde{\mathcal{E}}_{nt}(\theta)\right)$$

for $\theta \in \Theta$, where $\tilde{\mathcal{E}}_{nt}(\theta) = R_n(\theta_1)\tilde{Y}_{nt} - (\gamma I_n + \rho W_n)\tilde{Y}_{n,t-1}^{(-)} - (I_n + \delta L_n^*(\theta_1)\Pi_n)\tilde{X}_{nt}\beta$. First, note that

$$\sigma_{\epsilon,nT}^2(\theta_1, \beta) = \arg\max_{\sigma_\epsilon^2} Q_{nT}\left(\theta_1, \beta, \sigma_\epsilon^2\right) = \frac{1}{nT}\sum_{t=1}^{T} E\left(\tilde{\mathcal{E}}_{nt}'(\theta) J_n \tilde{\mathcal{E}}_{nt}(\theta)\right),$$

and $\beta_{nT}(\theta_1) = \arg\max_{\beta} Q_{nT}\left(\theta_1, \beta, \sigma_\epsilon^2\right) = \left[\sum_{t=1}^{T}\tilde{\mathbb{X}}_{nt}'(\theta_1) J_n \tilde{\mathbb{X}}_{nt}(\theta_1)\right]^{-1}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{nt}'(\theta_1) J_n E\left(\tilde{Z}_{nt}(\theta_1)\right)$ where $\tilde{\mathbb{X}}_{nt}(\theta_1) = (I_n + \delta L_n^*(\theta_1)\Pi_n)\tilde{X}_{nt}$, $\tilde{\mathbb{X}}_{nt} = \tilde{\mathbb{X}}_{nt}(\theta_{1,0})$ and

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

If $\theta_1 = \theta_{1,0}$, $J_n\tilde{Z}_{nt}(\theta_{1,0}) = J_n\tilde{\mathbb{X}}_{nt}\beta_0$. Hence, $J_n\tilde{Z}_{nt}(\theta_1)$ represents the misspecified $J_n\tilde{\mathbb{X}}_{nt}\beta_0$ if we evaluate it at $\theta_1 \in \Theta_1 \backslash \{\theta_{1,0}\}$. For $\theta_1 \in \Theta_1 \backslash \{\theta_{1,0}\}$, $J_n\left(\tilde{Z}_{nt}(\theta_1) - \tilde{\mathbb{X}}_{nt}\beta_0\right)$ shows the misspecification error. Given the identification of $\theta_{1,0}$, we obtain the identification condition for β_0: $\lim_{n,T\to\infty}\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{nt}' J_n \tilde{\mathbb{X}}_{nt}$ exists and is nonsingular.

Using $\beta_{nT}(\theta_1)$ and $\sigma_{\epsilon,nT}^2(\theta_1) \equiv \sigma_{\epsilon,nT}^2(\theta_1, \beta_{nT}(\theta_1))$, we derive the concentrated expected log-likelihood at θ_1 is

$$Q_{nT,c}(\theta_1) = Q_{nT}\left(\theta_1, \beta_{nT}(\theta_1), \sigma_{\epsilon,nT}^2(\theta_1, \beta_{nT}(\theta_1))\right) = -\frac{1}{2}[\ln 2\pi + 1] - \frac{1}{2}\ln \sigma_{\epsilon,nT}^2(\theta_1) + \frac{1}{n}\ln|R_n(\theta_1)|$$

with $Q_{nT,c}(\theta_{1,0}) = -\frac{1}{2}[\ln 2\pi + 1] - \frac{1}{2}\ln \sigma_{\epsilon,0}^2 + \frac{1}{n}\ln|R_n|$. Then,

$$\begin{aligned} Q_{nT,c}(\theta_1) - Q_{nT,c}(\theta_{1,0}) &= -\left[\ln \sigma_{\epsilon,nT}^2(\theta_1) - \ln \sigma_{\epsilon,0}^2\right] + \frac{1}{n}\left[\ln|R_n(\theta_1)| - \ln|R_n|\right] \\ &= \frac{1}{n}\ln\left|\left(\sigma_{\epsilon,0}^2\right)^{\frac{1}{2}} R_n^{-1}\right| - \frac{1}{n}\ln\left|\left(\sigma_{\epsilon,nT}^2(\theta_1)\right)^{\frac{1}{2}} R_n^{-1}(\theta_1)\right| \\ &= \frac{1}{2}\left\{\frac{1}{n}\ln\left|\sigma_{\epsilon,0}^2 R_n^{-1\prime} R_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_{\epsilon,nT}^2(\theta_1) R_n^{-1}(\theta_1) R_n^{-1}(\theta_1)'\right|\right\}. \end{aligned}$$

[2] Note that we can apply the information inequality to a concentrated log-likelihood function.

Hence, we obtain the unique identification condition for $\theta_{1,0}$ under large n and T:

$$\lim_{n,T\to\infty}\left[\frac{1}{n}\ln\left|\sigma^2_{\epsilon,0}R_n^{-1\prime}R_n^{-1}\right|-\frac{1}{n}\ln\left|\sigma^2_{\epsilon,nT}(\theta_1)R_n^{-1\prime}(\theta_1)R_n^{-1}(\theta_1)\right|\right]\neq 0 \tag{4}$$

for $\theta_1\neq\theta_{1,0}$ where

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

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

and $\sigma^2_{\epsilon,nT,2}(\theta_1)=\frac{\sigma^2_{\epsilon,0}}{n-1}tr\left(R_n^{-1\prime}R'_n(\theta_1)J_nR_n(\theta_1)R_n^{-1}\right)$. We observe that $\sigma^2_{\epsilon,nT}(\theta_1)$ consists of two parts. The first term, $\sigma^2_{\epsilon,nT,1}(\theta_1)$, is a quadratic function of the difference between two approximation functions for $J_n\tilde{\mathbb{X}}_{nt}\beta_0$. The second term, $\sigma^2_{\epsilon,nT,2}(\theta_1)$, comes from $E\left(\tilde{\mathcal{E}}'_{nt}R_n^{-1\prime}R'_n(\theta_1)J_nR_n(\theta_1)R_n^{-1}\tilde{\mathcal{E}}_{nt}\right)$. Note that the identification condition (4) can be written as

$$\lim_{n,T\to\infty}\left[\ln\left(\frac{\sigma^2_{\epsilon,0}}{\sigma^2_{\epsilon,nT,1}(\theta_1)+\sigma^2_{\epsilon,nT,2}(\theta_1)}\right)+\frac{1}{n}\ln\left|R_n^{-1\prime}R_n^{-1}\right|-\frac{1}{n}\ln\left|R_n^{-1\prime}(\theta_1)R_n^{-1}(\theta_1)\right|\right]\neq 0$$

If θ_1 is close to $\theta_{1,0}$, $\sigma^2_{\epsilon,nT,1}(\theta_1)$ is close to zero. Around $\theta_{1,0}$, hence, $\sigma^2_{\epsilon,nT,2}(\theta_1)$ plays a main role in identifying $\theta_{1,0}$.

2.2 Derivation of the concentrated joint log-likelihood function

For estimation, we assume the following structure on $\boldsymbol{\eta}_{nt}$ to derive the joint log-likelihood function.

Assumption 2.1 *(i) For each t, $\boldsymbol{\eta}_{nt}=\boldsymbol{\eta}^{iv}_n+\boldsymbol{\eta}^{v}_{nt}$ where $\boldsymbol{\eta}^{iv}_n=\left(\eta^{iv}_1,\cdots,\eta^{iv}_n\right)'$ and $\boldsymbol{\eta}^{v}_{nt}=(\eta^{v}_{1t},\cdots,\eta^{v}_{nt})'$.*

(ii) $\boldsymbol{\eta}^{v}_{nt}=X_{nt}\beta_{1,0}+W_nX_{nt}\beta_{2,0}+\alpha_{t,0}l_n+\mathcal{E}_{nt}$ where $\beta_{1,0}=\left(\beta_{1,1,0},\cdots,\beta_{1,K,0}\right)'$ and $\beta_{2,0}=\left(\beta_{2,1,0},\cdots,\beta_{2,K,0}\right)'$ are respectively coefficients of X_{nt} and W_nX_{nt}. For each t, $\alpha_{t,0}$ is a period-specific shock and $\mathcal{E}_{nt}=(\epsilon_{1t},\cdots,\epsilon_{nt})'$ is an n-dimensional vector of idiosyncratic shocks.

(iii) The $\alpha_{t,0}$ *and* $\mathcal{E}_{nt}$ *are independently generated across time* t*.*

(iv) $X_{nt,k}$ *is generated by*

$$X_{nt,k} = A_{k,n} X_{n,t-1,k} + \mathbf{c}_{n,k,0} + \alpha_{t,k,0} l_n + V_{nt,k} \tag{5}$$

where $A_{k,n} = \gamma_{k,0} I_n + \rho_{k,0} W_n$ *with* $\max_{k=1,\cdots,K} \|A_{k,n}\|_\infty < 1$*, and* $V_{nt} = (V_{nt,1}, \cdots, V_{nt,K})$ *denotes a disturbance term of* $X_{nt,k}$*, which is independent with the* $(t-1)$*-period agents' information set.*

We assume that (i), (ii), (iii) and (iv) are known to all agents.

Assumption 2.1 (i) means $\boldsymbol{\eta}_{nt}$ is additively separable. By Assumption 2.1 (ii), the time-variant part $\boldsymbol{\eta}_{nt}^v$ is composed of two parts: (i) observable (to econometricians) part $X_{nt}\beta_{1,0} + W_n X_{nt}\beta_{2,0}$ and (ii) unobservable (to econometricians) shocks $\alpha_{t,0} l_n + \mathcal{E}_{nt}$. In general, X_{nt} means own exogenous characteristics while $W_n X_{nt}$ describes rivals' exogenous characteristics (which capture externalities and/or contextual effects). Assumption 2.1 (iii) implies also that for any t, $E_t(\alpha_{t+1}) = 0$ and $E_t(\mathcal{E}_{n,t+1}) = \mathbf{0}_{n\times 1}$. Assumption 2.1 (iv) assumes stationarity of $X_{nt,k}$. By Assumption 2.1 (iv) and supposing $\mathbf{c}_{n,k,0} = h_{1,k}\boldsymbol{\eta}_n^{iv}$ for some coefficient $h_{1,k}$,

$$\begin{aligned}
&\boldsymbol{\eta}_{nt} + \delta E_t(L_n^* \Pi_n(\boldsymbol{\eta}_{n,t+1})) \\
= &\sum_{k=1}^{K}\left(I_n + \sum_{l=1}^{\infty}\delta^l D_{n,l} A_{k,n}^l\right)(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n) X_{nt,k} \\
&+ \underbrace{\left(\left(I_n + \sum_{l=1}^{\infty}\delta^l D_{n,l}\right) + \sum_{k=1}^{K}\sum_{l=1}^{\infty}\delta^l D_{n,l}\left(\sum_{m=0}^{l-1} A_{k,n}^m\right)(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n) h_{1,k}\right)\boldsymbol{\eta}_n^{iv}}_{=\text{time-invariant component part}} + \alpha_{t,0} l_n + \mathcal{E}_{nt}.
\end{aligned}$$

The time-invariant components are absorbed in individual specific effects denoted by $\mathbf{c}_{n,0} = (c_1, \cdots, c_n)'$. For notational convenience, define $\mathbf{D}_{n,k} = \sum_{l=1}^{\infty}\delta^{l-1} D_{n,l} A_{k,n}^{l-1}$ for each k. Then, the part of observables is

$$\begin{aligned}
&\sum_{k=1}^{K}\left(I_n + \sum_{l=1}^{\infty}\delta^l D_{n,l} A_{k,n}^l\right)(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n) X_{nt,k} \qquad (6)\\
= &\underbrace{\sum_{k=1}^{K}(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n) X_{nt,k}}_{\text{first term}} + \underbrace{\sum_{k=1}^{K}\delta \mathbf{D}_{n,k} A_{k,n}(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n) X_{nt,k}}_{\text{second term}}.
\end{aligned}$$

The first term in (6) describes the part of $X_{nt,k}$'s affecting Y_{nt} directly at time t. The second term in (6) captures effects of (discounted) expected future characteristics based on current available information.

For our econometric model, assume that ϵ_{it} ($i = 1, \cdots, n$ and $t = 1, \cdots, T$) has zero mean and finite variance $\sigma^2_{\epsilon,0}$ and also $V_{nt,k}$ has zero mean and finite variance $\sigma^2_{V,k,0}$ for each k. The main parameters are λ_0, γ_0, ρ_0, $\beta_{1,0}$ and $\beta_{2,0}$. Let $\theta_{1,0}$ be the true $\theta_1 = (\lambda, \gamma, \rho)'$. Let $R_n(\theta_1)$ be the spatial-time filter evaluated at θ_1 so that $R_n = R_n(\theta_{1,0})$. The parameters $\gamma_{1,0}, \cdots, \gamma_{K,0}, \rho_{1,0}, \cdots, \rho_{K,0}$ drive the dynamics of $X_{nt,k}$'s. For possible values of those parameters, let $\gamma_X = (\gamma_1, \cdots, \gamma_K)'$ and $\rho_X = (\rho_1, \cdots, \rho_K)'$. Then,

$$\theta = \left(\theta_1', \beta_1', \beta_2', \sigma^2_\epsilon, \gamma_X', \rho_X', \sigma^2_{V,1}, \cdots, \sigma^2_{V,K}\right)' \tag{7}$$

is the set of parameters for estimation, where θ_0 is the true value of θ. The dimension of the parameters is $4 + 5K$. To distinguish the true individual- or time-specific effects, we add the subscript "0" to α_t and $\mathbf{c}_n$.

Hence, the data generating process (DGP) consists of

$$R_n Y_{nt} = (\gamma_0 I_n + \rho_0 W_n) Y_{n,t-1} + \sum_{k=1}^{K} (I_n + \delta \mathbf{D}_{n,k} A_{k,n}) \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n\right) X_{nt,k} + \mathbf{c}_{n0} + \alpha_{t0} l_n + \mathcal{E}_{nt} \tag{8}$$

and

$$X_{nt,k} = A_{k,n} X_{n,t-1,k} + \mathbf{c}_{n,k,0} + \alpha_{t,k,0} l_n + V_{nt,k} \text{ for } k = 1, \cdots, K \tag{9}$$

where $\mathbf{D}_{n,k}$ is a function of $(\theta_{1,0}, \gamma_{X,0}, \rho_{X,0})$. The reduced form of equation (8) is

$$Y_{nt} = A_n Y_{n,t-1} + \sum_{k=1}^{K} R_n^{-1} B_{X,k,n} X_{nt,k} + R_n^{-1} (\mathbf{c}_{n0} + \alpha_{t0} l_n + \mathcal{E}_{nt}) \tag{10}$$

where $A_n = R_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, and $B_{X,k,n} = (I_n + \delta \mathbf{D}_{n,k} A_{k,n}) \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n\right)$ for $k = 1, \cdots, K$. Let $\boldsymbol{\alpha}_T = (\alpha_1, \cdots, \alpha_T)'$ and $\boldsymbol{\alpha}_{T,k} = (\alpha_{1,k}, \cdots, \alpha_{T,k})'$ for $k = 1, \cdots, K$. To derive the log-likelihood function for equation (8), assume that for each i^{th} column (i^{th} individual) of $(\mathcal{E}_{nt}, V_{nt,1}, \cdots, V_{nt,K})' \backsim i.i.d.N\left(\mathbf{0}_{(1+K)\times 1}, diag\left(\sigma^2_{\epsilon,0}, \sigma^2_{V,1,0}, \cdots, \sigma^2_{V,K,0}\right)\right)$. Given (Y_{n0}, X_{n0}), the joint density of $\{Y_{nt}, X_{nt}\}_{t=1}^T$ is

$$\begin{aligned} f(Y_{n1}, \cdots, Y_{nT}, X_{n1}, \cdots, X_{nT}; \theta) &= \prod_{t=1}^{T} f\left(Y_{nt}, X_{nt} | \{Y_{ns}, X_{ns}\}_{s=0}^{t-1}; \theta\right) \\ &= \prod_{t=1}^{T} f\left(Y_{nt} | X_{nt}, \{Y_{ns}, X_{ns}\}_{s=0}^{t-1}; \theta\right) \cdot f\left(X_{nt} | \{Y_{ns}, X_{ns}\}_{s=0}^{t-1}; \theta\right) \\ &= \prod_{t=1}^{T} f(Y_{nt} | X_{nt}, Y_{n,t-1}; \theta) \cdot f(X_{nt} | X_{n,t-1}; \theta) \\ &= \prod_{t=1}^{T} f(Y_{nt} | X_{nt}, Y_{n,t-1}; \theta) \cdot \prod_{k=1}^{K} f(X_{nt,k} | X_{n,t-1,k}; \theta). \end{aligned}$$

The first and second equalities come from the relation between the joint probability and the conditional probabilities. By observing (10), we have the third equality. Since V_{nt,k_1} and V_{nt,k_2} are uncorrelated for $k_1 \neq k_2$, the last equality holds. The corresponding log-likelihood function will be

$$
\begin{aligned}
& \ln L_{nT}\left(\theta, \mathbf{c}_n, \{\mathbf{c}_{n,k}\}_{k=1}^{K}, \boldsymbol{\alpha}_T, \{\boldsymbol{\alpha}_{T,k}\}_{k=1}^{K}\right) \qquad (11) \\
= & -\frac{nT(K+1)}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma_\epsilon^2 - \frac{nT}{2}\sum_{k=1}^{K}\ln\sigma_{V,k}^2 + T\ln|R_n(\theta_1)| \\
& -\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}\mathcal{E}_{nt}'(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T)\mathcal{E}_{nt}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T) - \sum_{t=1}^{T}\sum_{k=1}^{K}\frac{1}{2\sigma_{V,k}^2}V_{nt,k}'(\gamma_k, \rho_k, \mathbf{c}_{n,k}, \boldsymbol{\alpha}_{T,k})V_{nt,k}(\gamma_k, \rho_k, \mathbf{c}_{n,k}, \boldsymbol{\alpha}_{T,k})
\end{aligned}
$$

where $\mathcal{E}_{nt}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T) = R_n(\theta_1)Y_{nt} - (\gamma I_n + \rho W_n)Y_{n,t-1} - \sum_{k=1}^{K}B_{X,k,n}(\theta)X_{nt,k} - \mathbf{c}_n - \alpha_t l_n$, and
$V_{nt,k}(\gamma_k, \rho_k, \mathbf{c}_{n,k}, \boldsymbol{\alpha}_{T,k}) = X_{nt,k} - (\gamma_k I_n + \rho_k W_n)X_{n,t-1,k} - \mathbf{c}_{n,k} - \alpha_{t,k}l_n$ for $k = 1, \cdots, K$. Since $\mathbf{c}_n$ and $\boldsymbol{\alpha}_T$ are linear parameters, we have the following concentrated log-likelihood function

$$
\begin{aligned}
\ln L_{nT,c}(\theta) = & -\frac{nT(K+1)}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma_\epsilon^2 - \frac{nT}{2}\sum_{k=1}^{K}\ln\sigma_{V,k}^2 + T\ln|R_n(\theta_1)| \qquad (12) \\
& -\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta) - \sum_{t=1}^{T}\sum_{k=1}^{K}\frac{1}{2\sigma_{V,k}^2}\tilde{V}_{nt,k}'(\gamma_k, \rho_k)J_n\tilde{V}_{nt,k}(\gamma_k, \rho_k)
\end{aligned}
$$

where $\tilde{\mathcal{E}}_{nt}(\theta) = R_n(\theta_1)\tilde{Y}_{nt} - (\gamma I_n + \rho W_n)\tilde{Y}_{n,t-1}^{(-)} - \sum_{k=1}^{K}B_{X,k,n}(\theta)\tilde{X}_{nt,k}$, and
$\tilde{V}_{nt,k}(\gamma_k, \rho_k) = \tilde{X}_{nt,k} - (\gamma_k I_n + \rho_k W_n)\tilde{X}_{n,t-1,k}^{(-)}$ for $k = 1, \cdots, K$.

2.3 Some notations on derivatives

For further steps, consider the relationships by the mean value theorem, which defines various quantities of $C_{n,k}^{\lambda}(\theta)$, etc., :

$$
R_n(\theta_1)R_n^{-1}B_{X,k,n} - B_{X,k,n}(\theta) = \begin{bmatrix} (\lambda_0 - \lambda)\cdot C_{n,k}^{\lambda}(\bar{\theta}) + (\gamma_0 - \gamma)\cdot C_{n,k}^{\gamma}(\bar{\theta}) + (\rho_0 - \rho)\cdot C_{n,k}^{\rho}(\bar{\theta}) \\ + (\beta_{1,k,0} - \beta_{1,k})\cdot C_{n,k}^{\beta_1,k}(\bar{\theta}) + (\beta_{2,k,0} - \beta_{2,k})\cdot C_{n,k}^{\beta_2,k}(\bar{\theta}) \\ (\gamma_{k,0} - \gamma_k)\cdot C_{n,k}^{\gamma,k}(\bar{\theta}) + (\rho_{k,0} - \rho_k)\cdot C_{n,k}^{\rho,k}(\bar{\theta}) \end{bmatrix}
$$

where $k = 1, \cdots, K$, $\bar{\theta}$ lies between θ and θ_0 and
$C_{n,k}^{\lambda}(\theta) = -R_{n\lambda}(\theta_1)R_n^{-1}B_{X,k,n} + \delta\mathbf{D}_{n\lambda,k}(\theta_1, \gamma_k, \rho_k)A_{k,n}(\gamma_k, \rho_k)\left(\beta_{1,k}I_n + \beta_{2,k}W_n\right)$,
$C_{n,k}^{\gamma}(\theta) = -R_{n\gamma}(\theta_1)R_n^{-1}B_{X,k,n} + \delta\mathbf{D}_{n\gamma,k}(\theta_1, \gamma_k, \rho_k)A_{k,n}(\gamma_k, \rho_k)\left(\beta_{1,k}I_n + \beta_{2,k}W_n\right)$,
$C_{n,k}^{\rho}(\theta) = -R_{n\rho}(\theta_1)R_n^{-1}B_{X,k,n} + \delta\mathbf{D}_{n\rho,k}(\theta_1, \gamma_k, \rho_k)A_{k,n}(\gamma_k, \rho_k)\left(\beta_{1,k}I_n + \beta_{2,k}W_n\right)$,
$C_{n,k}^{\beta_1,k}(\theta) = (I_n + \delta\mathbf{D}_{n,k}(\theta_1, \gamma_k, \rho_k)A_{k,n}(\gamma_k, \rho_k))$ for $k = 1, \cdots, K$,

$C^{\beta_2,k}_{n,k}(\theta) = (I_n + \delta \mathbf{D}_{n,k}(\theta_1, \gamma_k, \rho_k) A_{k,n}(\gamma_k, \rho_k)) W_n$ for $k = 1, \cdots, K$,
$C^{\gamma,k}_{n,k}(\theta) = \delta \left(\mathbf{D}_{n,\gamma_k,k}(\theta_1, \gamma_k, \rho_k) A_{k,n}(\gamma_k, \rho_k) + \mathbf{D}_{n,k}(\theta_1, \gamma_k, \rho_k) \right) \left(\beta_{1,k} I_n + \beta_{2,k} W_n \right)$ for $k = 1, \cdots, K$,
$C^{\rho,k}_{n,k}(\theta) = \delta \left(\mathbf{D}_{n,\rho_k,k}(\theta_1, \gamma_k, \rho_k) A_{k,n}(\gamma_k, \rho_k) + \mathbf{D}_{n,k}(\theta_1, \gamma_k, \rho_k) W_n \right) \left(\beta_{1,k} I_n + \beta_{2,k} W_n \right)$ for $k = 1, \cdots, K$,
and $\bar{\theta}$ lies between θ and θ_0. These defined C's would be used later on.

For $k = 1, \cdots, K$, denote $\mathbf{D}_{n,\lambda\lambda,k}(\theta_1, \gamma_k, \rho_k) = \frac{\partial^2 \mathbf{D}_{n,k}(\theta_1, \gamma_k, \rho_k)}{\partial \lambda^2}$. Other second order derivatives are defined similarly.

2.4 $\Delta_{1,nT}$ and $\Delta_{2,nT}$

Here are the components of $\Delta_{1,nT}$ and $\Delta_{2,nT}$, relevant for asymptotic bias of the QMLE:

$$\Delta^{\lambda}_{1,nT} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left[\begin{array}{c} \left(\begin{array}{c} \left(-R_{n\lambda} R_n^{-1} (\gamma_0 I_n + \rho_0 W_n) \right) \left(\bar{U}_{1,nT,-1} + \bar{U}_{2,nT,-1} \right) \\ + \sum_{k=1}^{K} \left[\begin{array}{c} \left(-R_{n\lambda} R_n^{-1} \right) B_{X,k,n} A_{k,n} \\ + \delta \mathbf{D}_{n\lambda,k} A_{k,n} \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) A_{k,n} \end{array} \right] \bar{U}_{3,nT,k,-1} \end{array} \right)' J_n \bar{\mathcal{E}}_{nT} \\ + \bar{\mathcal{E}}'_{nT} (-R_n^{-1\prime} R'_{n\lambda}) J_n \bar{\mathcal{E}}_{nT} + \sum_{k=1}^{K} \bar{V}'_{nT,k} \left(\begin{array}{c} \left(-R_{n\lambda} R_n^{-1} \right) B_{X,k,n} \\ + \delta \mathbf{D}_{n\lambda,k} A_{k,n} \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) \end{array} \right)' J_n \bar{\mathcal{E}}_{nT} \end{array} \right],$$

$$\Delta^{\gamma}_{1,nT} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left[\begin{array}{c} \left(\begin{array}{c} \left(-R_{n\gamma} R_n^{-1} (\gamma_0 I_n + \rho_0 W_n) + I_n \right) \left(\bar{U}_{1,nT,-1} + \bar{U}_{2,nT,-1} \right) \\ + \sum_{k=1}^{K} \left[\begin{array}{c} \left(-R_{n\gamma} R_n^{-1} \right) B_{X,k,n} A_{k,n} \\ + \delta \mathbf{D}_{n\gamma,k} A_{k,n} \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) A_{k,n} \end{array} \right] \bar{U}_{3,nT,k,-1} \end{array} \right)' J_n \bar{\mathcal{E}}_{nT} \\ + \bar{\mathcal{E}}'_{nT} (-R_n^{-1\prime} R'_{n\gamma}) J_n \bar{\mathcal{E}}_{nT} + \sum_{k=1}^{K} \bar{V}'_{nT,k} \left(\begin{array}{c} \left(-R_{n\gamma} R_n^{-1} \right) B_{X,k,n} \\ + \delta \mathbf{D}_{n\gamma,k} A_{k,n} \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) \end{array} \right)' J_n \bar{\mathcal{E}}_{nT} \end{array} \right],$$

$$\Delta^{\rho}_{1,nT} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left[\begin{array}{c} \left(\begin{array}{c} \left(-R_{n\rho} R_n^{-1} (\gamma_0 I_n + \rho_0 W_n) + W_n \right) \left(\bar{U}_{1,nT,-1} + \bar{U}_{2,nT,-1} \right) \\ + \sum_{k=1}^{K} \left[\begin{array}{c} \left(-R_{n\rho} R_n^{-1} \right) B_{X,k,n} A_{k,n} \\ + \delta \mathbf{D}_{n\rho,k} A_{k,n} \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) A_{k,n} \end{array} \right] \bar{U}_{3,nT,k,-1} \end{array} \right)' J_n \bar{\mathcal{E}}_{nT} \\ + \bar{\mathcal{E}}'_{nT} (-R_n^{-1\prime} R'_{n\rho}) J_n \bar{\mathcal{E}}_{nT} + \sum_{k=1}^{K} \bar{V}'_{nT,k} \left(\begin{array}{c} \left(-R_{n\rho} R_n^{-1} \right) B_{X,k,n} \\ + \delta \mathbf{D}_{n\rho,k} A_{k,n} \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) \end{array} \right)' J_n \bar{\mathcal{E}}_{nT} \end{array} \right],$$

$$\Delta^{\beta_{1,k}}_{1,nT} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left(\left[\left(I_n + \delta \mathbf{D}_{n,k} A_{k,n} \right) A_{k,n} \right] \bar{U}_{3,nT,k,-1} \right)' J_n \bar{\mathcal{E}}_{nT} + \bar{V}'_{nT,k} \left(I_n + \delta \mathbf{D}_{n,k} A_{k,n} \right)' J_n \bar{\mathcal{E}}_{nT},$$

$$\Delta^{\beta_{2,k}}_{1,nT} = \frac{1}{\sigma^2_{\epsilon,0}} \sqrt{\frac{T}{n}} \left(\left[\left(I_n + \delta \mathbf{D}_{n,k} A_{k,n} \right) W_n A_{k,n} W_n \right] \bar{U}_{3,nT,k,-1} \right)' J_n \bar{\mathcal{E}}_{nT} + \bar{V}'_{nT,k} \left[\left(I_n + \delta \mathbf{D}_{n,k} A_{k,n} \right) W_n \right]' J_n \bar{\mathcal{E}}_{nT},$$

$$\Delta^{\sigma^2_{\epsilon}}_{1,nT} = \frac{1}{2\sigma^4_{\epsilon,0}} \sqrt{\frac{T}{n}} \bar{\mathcal{E}}'_{nT} J_n \bar{\mathcal{E}}_{nT},$$

$$\Delta^{\gamma_{X,k}}_{1,nT} = \sqrt{\frac{T}{n}} \left[\begin{array}{c} \frac{1}{\sigma^2_{\epsilon,0}} \left(\delta \left(\mathbf{D}_{n,\gamma_{X,k},k} A_{k,n} + \mathbf{D}_{n,k} \right) \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) A_{k,n} \bar{U}_{3,nT,k,-1} \right)' J_n \bar{\mathcal{E}}_{nT} \\ + \frac{1}{\sigma^2_{\epsilon,0}} \bar{V}'_{nT,k} \left(\delta \left(\mathbf{D}_{n,\gamma_{X,k},k} A_{k,n} + \mathbf{D}_{n,k} \right) \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) \right)' J_n \bar{\mathcal{E}}_{nT} + \frac{1}{\sigma^2_{V,k,0}} \bar{U}'_{3,nT,k,-1} J_n \bar{V}_{nT,k} \end{array} \right],$$

$$\Delta_{1,nT}^{\rho_{X,k}} = \sqrt{\frac{T}{n}} \begin{bmatrix} \frac{1}{\sigma_{\epsilon,0}^2}\left(\delta\left(\mathbf{D}_{n,\rho_{X,k},k}A_{k,n} + \mathbf{D}_{n,k}W_n\right)\left(\beta_{1,k,0}I_n + \beta_{2,k,0}W_n\right)W_nA_{k,n}\bar{U}_{3,nT,k,-1}\right)' J_n\bar{\mathcal{E}}_{nT} \\ +\frac{1}{\sigma_{\epsilon,0}^2}\bar{V}_{nT,k}'\left(\delta\left(\mathbf{D}_{n,\rho_{X,k},k}A_{k,n} + \mathbf{D}_{n,k}W_n\right)\left(\beta_{1,k,0}I_n + \beta_{2,k,0}W_n\right)W_n\right)' J_n\bar{\mathcal{E}}_{nT} \\ +\frac{1}{\sigma_{V,k,0}^2}\bar{U}_{3,nT,k,-1}'W_n'J_n\bar{V}_{nT,k} \end{bmatrix},$$

$\Delta_{1,nT}^{\sigma_{V,k}^2} = \frac{1}{2\sigma_{V,k,0}^4}\sqrt{\frac{T}{n}}\bar{V}_{nT,k}'J_n\bar{V}_{nT,k}$,

$\Delta_{2,nT}^{\lambda} = \sqrt{\frac{T}{n}}\left[tr\left(-R_{n\lambda}R_n^{-1}\right) - tr\left(J_n(-R_{n\lambda}R_n^{-1})\right)\right]$, $\Delta_{2,nT}^{\gamma} = \sqrt{\frac{T}{n}}\left[tr\left(-R_{n\gamma}R_n^{-1}\right) - tr\left(J_n(-R_{n\gamma}R_n^{-1})\right)\right]$,

$\Delta_{2,nT}^{\rho} = \sqrt{\frac{T}{n}}\left[tr\left(-R_{n\rho}R_n^{-1}\right) - tr\left(J_n(-R_{n\rho}R_n^{-1})\right)\right]$, $\Delta_{2,nT}^{\beta_1} = \mathbf{0}_{K\times 1}$, $\Delta_{2,nT}^{\beta_2} = \mathbf{0}_{K\times 1}$, $\Delta_{2,nT}^{\sigma_\epsilon^2} = \sqrt{\frac{T}{n}}\frac{1}{2\sigma_{\epsilon,0}^2}$,

$\Delta_{2,nT}^{\gamma_{X,k}} = 0$, $\Delta_{2,nT}^{\rho_{X,k}} = 0$, and $\Delta_{2,nT}^{\sigma_{V,k}^2} = \sqrt{\frac{T}{n}}\frac{1}{2\sigma_{V,k,0}^2}$ for $k = 1, \cdots, K$.

2.5 Some lemmas for the asymptotic properties of QMLEs

Note that Y_{nt} can be represented by

$$Y_{nt} = \underbrace{\begin{bmatrix} \sum_{h=0}^{\infty}\sum_{g=0}^{\infty}\sum_{k=1}^{K} A_n^hR_n^{-1}B_{X,k,n}A_{k,n}^g\left(\mathbf{c}_{n,k,0} + \alpha_{t-h-g,k,0}l_n\right) \\ +\sum_{h=0}^{\infty}A_n^hR_n^{-1}\left(\mathbf{c}_{n0} + \alpha_{t-h,0}l_n\right) \end{bmatrix}}_{\text{nonstochastic component of } Y_{nt}} + \underbrace{\begin{bmatrix} \sum_{h=0}^{\infty}\sum_{g=0}^{\infty}\sum_{k=1}^{K} A_n^hR_n^{-1}B_{X,k,n}A_{k,n}^gV_{n,t-h-g,k} \\ +\sum_{h=0}^{\infty}A_n^hR_n^{-1}\mathcal{E}_{n,t-h} \end{bmatrix}}_{\text{stochastic component of } Y_{nt}}. \tag{13}$$

Hence, the main stochastic component of Y_{nt} is

$$\sum_{h=0}^{\infty}\sum_{g=0}^{\infty}\sum_{k=1}^{K}A_n^hR_n^{-1}B_{X,k,n}A_{k,n}^gV_{n,t-h-g,k} + \sum_{h=0}^{\infty}A_n^hR_n^{-1}\mathcal{E}_{n,t-h}. \tag{14}$$

To investigate (14), define $\mathbb{U}_{1,nt} = \sum_{h=0}^{\infty}\sum_{g=0}^{\infty}\sum_{k=1}^{K}P_{nh}Q_{ng,k}V_{n,t-h-g,k}$ and $\mathbb{U}_{2,nt} = \sum_{h=0}^{\infty}P_{nh}\mathcal{E}_{n,t-h}$ where $\{P_{nh}\}_{h=0}^{\infty}$ and $\{Q_{ng,k}\}_{g=0}^{\infty}|_{k=1}^{K}$ are $n\times n$ uniformly bounded (in n) matrices. Then, the stochastic component (14) takes the form of $\mathbb{U}_{1,nt} + \mathbb{U}_{2,nt}$. Similarly, the main stochastic component of $X_{nt,k}$ is verified by $\mathbb{U}_{3,nt,k} = \sum_{g=0}^{\infty}P_{X,ng,k}V_{n,t-g,k}$ where $\{P_{X,ng,k}\}_{g=0}^{\infty}|_{k=1}^{K}$ are $n\times n$ uniformly bounded (in n) matrices for each k. The following assumptions and lemmas are fundamental in our asymptotic analysis and similar to Yu et al. (2008) and Lee and Yu (2010).

Assumption 2.2 *For all i, t, and k, $\epsilon_{it} \backsim i.i.d.(0, \sigma_{\epsilon,0}^2)$, $v_{it,k} \backsim i.i.d.(0, \sigma_{V,k,0}^2)$, and ϵ's and v's are independent. Suppose $E|\epsilon_{it}|^{4+\eta} < \infty$ and $\max_{k=1,\cdots,K}E|v_{it,k}|^{4+\eta} < \infty$ for some $\eta > 0$.*

Assumption 2.3 *$\sum_{h=0}^{\infty} abs(P_{nh})$, $\sum_{h=0}^{\infty}\sum_{g=0}^{\infty}\sum_{k=1}^{K} abs(P_{nh})abs(Q_{ng,k})$ and $\sum_{g=0}^{\infty} abs(P_{X,ng,k})$ are uniformly bounded.*

Assumption 2.4 *(i) D_{nt} denotes an $n\times 1$ nonstochastic and all of elements are uniformly bounded in n and t. (ii) Let B_n, $B_{1,n,1},\cdots,B_{1,n,K}$, $B_{2,n,1},\cdots,B_{2,n,K}$ be $n\times n$ nonstochastic and uniformly bounded matrices.*

Assumption 2.5 *T goes to infinity. n is an increasing function of T.*

Note that the first order condition (except for $\frac{\partial \ln L_{nT,c}^{(u)}(\theta_0)}{\partial\sigma_{V,k}^2}$, $k=1,\cdots,K$) at $\theta=\theta_0$ takes the linear quadratic form,

$$\underbrace{\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left[B_{y,n}\tilde{Y}_{n,t-1}+\sum_{k=1}^{K}B_{1,X,k,n}\tilde{X}_{n,t-1,k}+D_{nt}\right]'J_n\tilde{\mathcal{E}}_{nt}}_{\text{Linear term I}}+\underbrace{\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left[\tilde{\mathcal{E}}_{nt}'B_{q,n}'J_n\tilde{\mathcal{E}}_{nt}-\sigma_{\epsilon,0}^2 tr\left(B_{q,n}\right)\right]}_{\text{Quadratic term I}}$$

$$+\underbrace{\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{k=1}^{K}\tilde{V}_{nt,k}'B_{2,X,k,n}'J_n\tilde{\mathcal{E}}_{nt}}_{\text{Cross term I}}+\underbrace{\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{k=1}^{K}\tilde{X}_{n,t-1,k}'B_{3,X,k,n}'J_n\tilde{V}_{nt,k}}_{\text{Linear term II}}$$

where $B_{y,n}$, $B_{q,n}$, $\{B_{1,X,k,n}\}_{k=1}^{K}$, $\{B_{2,X,k,n}\}_{k=1}^{K}$ and $\{B_{3,X,k,n}\}_{k=1}^{K}$ are uniformly bounded (in n). The following lemmas 2.1 and 2.2 describe stochastic orders of linear and/or quadratic terms.

Lemma 2.1 (Quadratic and cross terms) *Suppose Assumptions 2.2, 2.4 (ii) and 2.5 hold. Then,*

(i) $\frac{1}{nT}\sum_{t=1}^{T}\mathcal{E}_{nt}'B_n\mathcal{E}_{nt}-E\left(\frac{1}{nT}\sum_{t=1}^{T}\mathcal{E}_{nt}'B_n\mathcal{E}_{nt}\right)=O_p\left(\frac{1}{\sqrt{nT}}\right)$ *where* $E\left(\frac{1}{nT}\sum_{t=1}^{T}\mathcal{E}_{nt}'B_n\mathcal{E}_{nt}\right)=\frac{1}{n}\sigma_{\epsilon,0}^2 tr\left(B_n\right)=O\left(1\right)$.

(ii) $\frac{1}{n}\bar{\mathcal{E}}_{nT}'B_n\bar{\mathcal{E}}_{nT}-E\left(\frac{1}{n}\bar{\mathcal{E}}_{nT}'B_n\bar{\mathcal{E}}_{nT}\right)=O_p\left(\frac{1}{\sqrt{nT^2}}\right)$ *where* $E\left(\frac{1}{n}\bar{\mathcal{E}}_{nT}'B_n\bar{\mathcal{E}}_{nT}\right)=\frac{1}{nT}\sigma_{\epsilon,0}^2 tr\left(B_n\right)=O\left(\frac{1}{T}\right)$.

(iii) $\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}V_{nt,k}'B_{1,n,k}\mathcal{E}_{nt}=O_p\left(\frac{1}{\sqrt{nT}}\right)$. *Note that* $E\left(\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}V_{nt,k}'B_{1,n,k}\mathcal{E}_{nt}\right)=0$.

(iv) $\frac{1}{n}\sum_{k=1}^{K}\bar{V}_{nT,k}'B_{1,n,k}\bar{\mathcal{E}}_{nT}=O_p\left(\frac{1}{\sqrt{nT^2}}\right)$. *Note that* $E\left(\frac{1}{n}\sum_{k=1}^{K}\bar{V}_{nT,k}'B_{1,n,k}\bar{\mathcal{E}}_{nT}\right)=0$.

(v) $\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}V_{nt,k}'B_{2,n,k}V_{nt,k}-E\left(\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}V_{nt,k}'B_{2,n,k}V_{nt,k}\right)=O_p\left(\frac{1}{\sqrt{nT}}\right)$ *where*

$$E\left(\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}V_{nt,k}'B_{2,n,k}V_{nt,k}\right)=\frac{1}{n}\sum_{k=1}^{K}\sigma_{V,k,0}^2 tr\left(B_{2,n,k}\right)=O\left(1\right).$$

(vi) $\frac{1}{n}\sum_{k=1}^{K}\bar{V}_{nT,k}'B_{2,n,k}\bar{V}_{nT,k}-E\left(\frac{1}{n}\sum_{k=1}^{K}\bar{V}_{nT,k}'B_{2,n,k}\bar{V}_{nT,k}\right)=O_p\left(\frac{1}{\sqrt{nT^2}}\right)$ *where*

$$E\left(\frac{1}{n}\sum_{k=1}^{K}\bar{V}_{nT,k}'B_{2,n,k}\bar{V}_{nT,k}\right)=\frac{1}{nT}\sum_{k=1}^{K}\sigma_{V,k,0}^2 tr\left(B_{2,n,k}\right)=O\left(\frac{1}{T}\right).$$

Lemma 2.2 (Linear terms) *Suppose Assumptions 2.2, 2.3 and 2.5 hold. Then,*

(i) $\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{U}}'_{1,n,t-1}B_n\tilde{\mathcal{E}}_{nt}=O_p\left(\frac{1}{\sqrt{nT}}\right)$*. Note that* $E\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{U}}'_{1,n,t-1}B_n\tilde{\mathcal{E}}_{nt}\right)=0$.

(ii) $\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{U}}'_{2,n,t-1}B_n\tilde{\mathcal{E}}_{nt}-E\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{U}}'_{2,n,t-1}B_n\tilde{\mathcal{E}}_{nt}\right)=O_p\left(\frac{1}{\sqrt{nT}}\right)$ *where* $E\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{U}}'_{2,n,t-1}B_n\tilde{\mathcal{E}}_{nt}\right)=O\left(\frac{1}{T}\right)$*. Note that*

$$\begin{aligned}\frac{1}{n}E\left(\bar{\mathbb{U}}'_{2,nT,-1}B_n\bar{\mathcal{E}}_{nT}\right) &= \frac{1}{nT^2}\sum_{h=0}^{T-2}(T-1)\sigma^2_{\epsilon,0}tr\left(P'_{nh}B_n\right)-\frac{1}{nT^2}\sum_{h=0}^{T-2}h\sigma^2_{\epsilon,0}tr\left(P'_{nh}B_n\right)\\ &= \frac{1}{T}\underbrace{\frac{1}{n}\sigma^2_{\epsilon,0}tr\left(\sum_{h=0}^{\infty}P'_{nh}B_n\right)}_{=O(1)}+O\left(\frac{1}{T^2}\right)=O\left(\frac{1}{T}\right)+O\left(\frac{1}{T^2}\right).\end{aligned}$$

(iii) $\frac{1}{nT}\sum_{t=1}^{T}D'_{nt}B_n\tilde{\mathcal{E}}_{nt}$ *and* $\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}D'_{nt}B_{1,n,k}\tilde{V}_{nt,k}$ *have* $O_p\left(\frac{1}{\sqrt{nT}}\right)$.

(iv) $\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}\tilde{\mathbb{U}}'_{3,n,t-1,k}B_{1,n,k}\tilde{\mathcal{E}}_{nt}=O_p\left(\frac{1}{\sqrt{nT}}\right)$*. Note that* $E\left(\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}\tilde{\mathbb{U}}'_{3,n,t-1,k}B_{1,n,k}\tilde{\mathcal{E}}_{nt}\right)=0$.

(v) $\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}\tilde{\mathbb{U}}'_{3,n,t-1,k}B_{1,n,k}\tilde{V}_{nt,k}-E\left(\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}\tilde{\mathbb{U}}'_{3,n,t-1,k}B_{1,n,k}\tilde{V}_{nt,k}\right)=O_p\left(\frac{1}{\sqrt{nT}}\right)$ *where* $E\left(\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K}\tilde{\mathbb{U}}'_{3,n,t-1,k}B_{1,n,k}\tilde{V}_{nt,k}\right)=O\left(\frac{1}{T}\right)$*. Note that*

$$\frac{1}{n}\sum_{k=1}^{K}E\left(\bar{\mathbb{U}}'_{3,nT,k,-1}B_{1,n,k}\bar{V}_{nT}\right)=\frac{1}{T}\frac{1}{n}tr\left(\sum_{k=1}^{K}\sum_{h=0}^{\infty}\sigma^2_{V,k,0}P'_{X,nh,k}B_{1,n,k}\right)+O\left(\frac{1}{T^2}\right).$$

To derive the asymptotic distribution of $\hat{\theta}_{ml,nT}$, consider the stochastic component of $\frac{\partial\ln L^{(u)}_{nT,c}(\theta_0)}{\partial\theta}$ (except for $\frac{\partial\ln L^{(u)}_{nT,c}(\theta_0)}{\partial\sigma^2_{V,k}}$, $k=1,\cdots,K$).

$$\mathbf{s}_{nT}=\sum_{t=1}^{T}\begin{pmatrix}\left(\mathbb{U}_{1,n,t-1}+\mathbb{U}_{2,n,t-1}\right)'\mathcal{E}_{nt}+D'_{nt}\mathcal{E}_{nt}+\left(\mathcal{E}'_{nt}B_n\mathcal{E}_{nt}-\sigma^2_{\epsilon,0}tr\left(B_n\right)\right)\\ +\sum_{k=1}^{K}\mathbb{U}'_{3,n,t-1,k}\mathcal{E}_{nt}+\sum_{k=1}^{K}V'_{nt,k}B_{1,n,k}\mathcal{E}_{nt}+\sum_{k=1}^{K}\mathbb{U}'_{3,n,t-1,k}V_{nt,k}\end{pmatrix}=\sum_{t=1}^{T}\sum_{i=1}^{n}\xi_{nt,i}$$

and

$$\begin{aligned}\xi_{nt,i} &= \left(u_{1,i,t-1}+u_{2,i,t-1}+\sum_{k=1}^{K}u_{3,i,t-1,k}+\sum_{k=1}^{K}\sum_{j=1}^{n}b_{1,n,ij,k}v_{jt}+d_{nt,i}\right)\epsilon_{it}\\ &\quad +b_{n,ii}(\epsilon^2_{it}-\sigma^2_{\epsilon,0})+2\epsilon_{it}\sum_{j=1}^{i-1}b_{n,ij}\epsilon_{jt}+\sum_{k=1}^{K}u_{3,i,t-1,k}v_{it,k}\end{aligned}$$

where $u_{1,i,t-1}$, $u_{2,i,t-1}$, $u_{3,i,t-1}$ and $d_{nt,i}$ denotes respectively the i-th element of $\mathbb{U}_{1,n,t-1}$, $\mathbb{U}_{2,n,t-1}$, $\mathbb{U}_{3,n,t-1,k}$ and D_{nt}. $b_{n,ij}$, $b_{1,n,ij,k}$, $b_{2,n,ij,k}$ and $b_{3,n,ij,k}$ denote respectively the (i,j)-component of B_n, $B_{1,n,k}$, $B_{2,n,k}$ and

$B_{3,n,k}$. Also, $\frac{\partial \ln L_{nT,c}^{(u)}(\theta_0)}{\partial \sigma_{V,k}^2}$, $k=1,\cdots,K$ take the quadratic form,

$$\mathbf{s}_{nT,k}^{\sigma_V^2} = \frac{1}{2\sigma_{V,k,0}^4}\sum_{t=1}^{T}\left(V_{nt,k}'J_nV_{nt,k}-(n-1)\sigma_{V,k,0}^2\right) = \sum_{t=1}^{T}\sum_{i=1}^{n}\xi_{nt,i,k}^{\sigma_{V,k}^2}$$

where $\xi_{nt,i,k}^{\sigma_{V,k}^2} = \left(\frac{1-\frac{1}{n}}{2\sigma_{V,k,0}^4}\right)\left(v_{it,k}^2-\sigma_{V,k,0}^2\right)+2v_{it,k}\sum_{j=1}^{i-1}\left(\frac{-\frac{1}{n}}{2\sigma_{V,k,0}^4}\right)v_{jt,k}$ for $k=1,\cdots,K$. Note that the expectations of $\mathbf{s}_{nT}$ and $\mathbf{s}_{nT,k}^{\sigma_V^2}$'s are $E\left(\mathbf{s}_{nT}\right)=\sum_{t=1}^{T}\sum_{i=1}^{n}E\left(\xi_{nt,i}\right)=0$ and $E\left(\mathbf{s}_{nT,k}^{\sigma_V^2}\right)=\sum_{t=1}^{T}\sum_{i=1}^{n}E\left(\xi_{nt,i,k}^{\sigma_{V,k}^2}\right)=0$ for $k=1,\cdots,K$ using the statistical independence between ϵ_{it} and $v_{it,k}$'s. Let $\mu_{\epsilon,0}^{(3)}=E\left(\epsilon_{it}^3\right)$, $\mu_{V,k,0}^{(3)}=E\left(v_{it,k}^3\right)$, $\mu_{\epsilon,0}^{(4)}=E\left(\epsilon_{it}^4\right)$, and $\mu_{V,k,0}^{(4)}=E\left(v_{it,k}^4\right)$ for $k=1,\cdots,K$. Next, consider calculating the variance of $\mathbf{s}_{nT}$: note that

(i-a) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\mathbb{U}_{1,n,t-1}+\mathbb{U}_{2,n,t-1}\right)'\mathcal{E}_{nt}\mathcal{E}_{ns}'\left(\mathbb{U}_{1,n,s-1}+\mathbb{U}_{2,n,s-1}\right)\right)$
$= T\sum_{k=1}^{K}\sigma_{\varepsilon,0}^2\sigma_{V,k,0}^2 tr\left(\sum_{f=0}^{\infty}\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}Q_{n,f-h_2,k}'P_{nh_2}'P_{nh_1}Q_{n,f-h_1,k}\right)+T\sigma_{\varepsilon,0}^4 tr\left(\sum_{h=0}^{\infty}P_{nh}'P_{nh}\right),$

(i-b) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\mathbb{U}_{1,n,t-1}+\mathbb{U}_{2,n,t-1}\right)'\mathcal{E}_{nt}\mathcal{E}_{ns}'D_{ns}\right)=0,$

(i-c) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\mathbb{U}_{1,n,t-1}+\mathbb{U}_{2,n,t-1}\right)'\mathcal{E}_{nt}\cdot\left(\mathcal{E}_{ns}'B_n\mathcal{E}_{ns}-\sigma_{\epsilon,0}^2 tr\left(B_n\right)\right)\right)=0,$

(i-d) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\left(\mathbb{U}_{1,n,t-1}+\mathbb{U}_{2,n,t-1}\right)'\mathcal{E}_{nt}\mathcal{E}_{ns}'\mathbb{U}_{3,n,s-1,k}'\right)$
$= T\sum_{k=1}^{K}\sigma_{\varepsilon,0}^2\sigma_{V,k,0}^2 tr\left(\sum_{f=0}^{\infty}\sum_{h=0}^{\infty}P_{X,nf,k}'P_{nh}Q_{n,f-h,k}\right),$

(i-e) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\left(\mathbb{U}_{1,n,t-1}+\mathbb{U}_{2,n,t-1}\right)'\mathcal{E}_{nt}\cdot\left(V_{ns,k}'B_{1,n,k}\mathcal{E}_{ns}\right)\right)=0,$

(i-f) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\mathbb{U}_{1,n,t-1}+\mathbb{U}_{2,n,t-1}\right)'\mathcal{E}_{nt}\cdot\left(\sum_{k=1}^{K}\mathbb{U}_{3,n,s-1,k}'V_{ns,k}\right)\right)=0,$

(ii-a) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}D_{nt}'\mathcal{E}_{nt}\mathcal{E}_{ns}'D_{nt}\right)=\sigma_{\varepsilon,0}^2\sum_{t=1}^{T}D_{nt}'D_{nt},$

(ii-b) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}D_{nt}'\mathcal{E}_{nt}\cdot\left(\mathcal{E}_{ns}'B_n\mathcal{E}_{ns}-\sigma_{\epsilon,0}^2 tr\left(B_n\right)\right)\right)=\mu_{\epsilon,0}^{(3)}\sum_{t=1}^{T}\sum_{i=1}^{n}d_{nt,i}b_{n,ii},$

(ii-c) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}D_{nt}'\mathcal{E}_{nt}\mathcal{E}_{ns}'\mathbb{U}_{3,n,s-1,k}'\right)=0,$

(ii-d) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}D_{nt}'\mathcal{E}_{nt}\mathcal{E}_{ns}'B_{1,n,k}V_{ns,k}\right)=0,$

(ii-e) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}D_{nt}'\mathcal{E}_{nt}V_{ns,k}\mathbb{U}_{n,s-1,3,k}\right)=0$

(iii-a) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\mathcal{E}_{nt}'B_n\mathcal{E}_{nt}-\sigma_{\epsilon,0}^2 tr\left(B_n\right)\right)\cdot\left(\mathcal{E}_{ns}'B_n\mathcal{E}_{ns}-\sigma_{\epsilon,0}^2 tr\left(B_n\right)\right)\right)$
$= T\left(\left(\mu_{\epsilon,0}^{(4)}-3\sigma_{\epsilon,0}^4\right)\sum_{i=1}^{n}b_{n,ii}^2+\sigma_{\epsilon,0}^4\left(tr\left(B_n^2\right)+tr\left(B_nB_n'\right)\right)\right),$

(iii-b) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\left(\mathcal{E}_{nt}'B_n\mathcal{E}_{nt}-\sigma_{\epsilon,0}^2 tr\left(B_n\right)\right)\cdot\mathcal{E}_{ns}'\mathbb{U}_{n,s-1,3,k}\right)=0,$

(iii-c) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\left(\mathcal{E}_{nt}'B_n\mathcal{E}_{nt}-\sigma_{\epsilon,0}^2 tr\left(B_n\right)\right)\cdot\mathcal{E}_{ns}'B_{1,n,k}V_{ns,k}\right)=0,$

(iii-d) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\left(\mathcal{E}_{nt}'B_n\mathcal{E}_{nt}-\sigma_{\epsilon,0}^2 tr\left(B_n\right)\right)\cdot V_{ns,k}'\mathbb{U}_{n,s-1,3,k}\right)=0,$

(iv-a) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{K}\mathbb{U}_{n,t-1,3,k}'\mathcal{E}_{nt}\mathcal{E}_{ns}'\mathbb{U}_{n,s-1,3,l}\right)=T\sigma_{\epsilon,0}^2\sum_{k=1}^{K}\sigma_{V,k,0}^2 tr\left(\sum_{h=0}^{\infty}P_{X,nh}'P_{X,nh}\right),$

(iv-b) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{K}\mathbb{U}_{n,t-1,3,k}'\mathcal{E}_{nt}\mathcal{E}_{ns}'B_{1,n,l}'V_{ns,l}\right)=0,$

(iv-c) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{K}\mathbb{U}'_{n,t-1,3,k}\mathcal{E}_{nt}V'_{ns,l}\mathbb{U}_{n,s-1,3,l}\right)=0$,

(v-a) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{K}V'_{nt,k}B_{1,n,k}\mathcal{E}_{nt}\mathcal{E}'_{ns}B'_{1,n,l}V_{ns,l}\right)=T\sigma^2_{\epsilon,0}\sum_{k=1}^{K}\sigma^2_{V,k,0}tr\left(B'_{1,n,k}B_{1,n,k}\right)$,

(v-b) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{K}V'_{nt,k}B_{1,n,k}\mathcal{E}_{nt}V'_{ns,l}\mathbb{U}_{n,s-1,3,l}\right)=0$,

(vi-a) $E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{K}\sum_{l=1}^{K}\mathbb{U}'_{n,t-1,3,k}V_{nt,k}V'_{ns,l}\mathbb{U}_{n,s-1,3,l}\right)=T\sum_{k=1}^{K}\sigma^4_{V,k,0}tr\left(\sum_{h=0}^{\infty}P'_{X,nh}P_{X,nh}\right)$.

Then, we obtain the variance of $\mathbf{s}_{nT}$, $\sigma^2_{\mathbf{s}_{nT}}\equiv Var\left(\mathbf{s}_{nT}\right)$:

$$\begin{aligned}\sigma^2_{\mathbf{s}_{nT}} &= T\sum_{k=1}^{K}\sigma^2_{\varepsilon,0}\sigma^2_{V,k,0}tr\left(\sum_{f=0}^{\infty}\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}Q'_{n,f-h_2,k}P'_{nh_2}P_{nh_1}Q_{n,f-h_1,k}\right)+T\sigma^4_{\varepsilon,0}tr\left(\sum_{h=0}^{\infty}P'_{nh}P_{nh}\right) \\ &\quad +2T\sum_{k=1}^{K}\sigma^2_{\varepsilon,0}\sigma^2_{V,k,0}tr\left(\sum_{f=0}^{\infty}\sum_{h=0}^{\infty}P'_{X,nf,k}P_{nh}Q_{n,f-h,k}\right)+\sigma^2_{\varepsilon,0}\sum_{t=1}^{T}D'_{nt}D_{nt} \\ &\quad +2\mu^{(3)}_{\epsilon,0}\sum_{t=1}^{T}\sum_{i=1}^{n}d_{nt,i}b_{n,ii}+T\left(\left(\mu^{(4)}_{\epsilon,0}-3\sigma^4_{\epsilon,0}\right)\sum_{i=1}^{n}b^2_{n,ii}+2\sigma^4_{\epsilon,0}tr\left(B^2_n\right)\right) \\ &\quad +T\sigma^2_{\epsilon,0}\sum_{k=1}^{K}\sigma^2_{V,k,0}tr\left(\left(\sum_{h=0}^{\infty}P'_{X,nh}P_{X,nh}\right)+B'_{1,n,k}B_{1,n,k}\right)+T\sum_{k=1}^{K}\sigma^4_{V,k,0}tr\left(\sum_{h=0}^{\infty}P'_{X,nh}P_{X,nh}\right).\end{aligned} \tag{15}$$

Also, $\sigma^2_{\mathbf{s}^{\sigma^2_V}_{nT,k}}\equiv Var\left(\mathbf{s}^{\sigma^2_V}_{nT,k}\right)=(n-1)T\frac{1}{4\sigma^8_{V,k,0}}\left(\left(\mu^{(4)}_{V,k,0}-3\sigma^4_{V,k,0}\right)\left(\frac{n-1}{n}\right)+2\sigma^4_{V,k,0}\right)$ for $k=1,\cdots,K$. Since $E\left(\mathbf{s}_{nT}\mathbf{s}^{\sigma^2_V}_{nT,k}\right)=0$ for all $k=1,\cdots,K$ and $E\left(\mathbf{s}^{\sigma^2_V}_{nT,k}\mathbf{s}^{\sigma^2_V}_{nT,l}\right)=0$ for all $k\neq l$, we can apply the Cramér-Wold device to verify the asymptotic distribution of the main statistic, $\frac{\partial\ln L^{(u)}_{nT}(\theta_0)}{\partial\theta}$. Here is the detailed proof strategy:

Step 1: The first step is to verify the asymptotic distribution of the univariate random variables, $\mathbf{s}_{nT}$ and $\mathbf{s}^{\sigma^2_V}_{nT,k}$. That is, $\frac{\mathbf{s}_{nT}}{\sigma_{\mathbf{s}_{nT}}}\xrightarrow{d}N(0,1)$ and $\frac{\mathbf{s}^{\sigma^2_V}_{nT,k}}{\sigma_{\mathbf{s}^{\sigma^2_V}_{nT,k}}}\xrightarrow{d}N(0,1)$ for all $k=1,\cdots,K$. Similar to Yu et al. (2008), we apply the central limit theorem of the martingale difference array. The idea of proof is following. At first, we consider the σ-field,

$$\mathcal{F}_{n,t,i}=\sigma\left(\epsilon_{11},\cdots,\epsilon_{n1},\cdots,\epsilon_{1,t-1},\cdots,\epsilon_{n,t-1},\epsilon_{1t},\cdots,\epsilon_{it}\right) \tag{16}$$

and $\mathcal{F}_{n,0,0}=\{\varnothing,\Omega\}$ where Ω denotes the sample space. Let $\mathcal{F}_{n,t,0}=\mathcal{F}_{n,t-1,n}$ as a convention. By using statistical independence between ϵ_{it} and $v_{it,k}$'s, we have $E\left(\xi_{nt,i}|\mathcal{F}_{n,t,i-1}\right)=0$, $E\left(\xi_{nt,i}|\mathcal{F}_{n,t-1,n}\right)=0$, $E\left(\xi^{\sigma^2_{V,k}}_{nt,i,k}|\mathcal{F}_{n,t,i-1}\right)=0$, and $E\left(\xi^{\sigma^2_{V,k}}_{nt,i,k}|\mathcal{F}_{n,t-1,n}\right)=0$. From these with $\mathcal{F}_{n,t,i-1}\subseteq\mathcal{F}_{n,t,i}$ and $\mathcal{F}_{n,t-1,0}\subseteq\mathcal{F}_{n,t,0}$, we construct the martingale difference arrays $\left\{\xi_{nt,i},\mathcal{F}_{n,t,i}:i=1,\cdots,n\text{ and }t=1,\cdots,T\right\}$, and $\left\{\xi^{\sigma^2_{V,k}}_{nt,i,k},\mathcal{F}_{n,t,i}:i=1,\cdots,n\text{ and }t=1,\cdots,T\right\}$ for $k=1,\cdots,K$. To apply the CLT to $\mathbf{s}_{nT}=\sum_{t=1}^{T}\sum_{i=1}^{n}\xi_{nt,i}$

and $\mathbf{s}^{\sigma_V^2}_{nT,k} = \sum_{t=1}^T \sum_{i=1}^n \xi^{\sigma^2_{V,k}}_{nt,i,k}$, we need to check two sufficient conditions: for all k

$$\text{(i) } \frac{1}{\sigma^{2+\eta}_{\mathbf{s}_{nT}}} \sum_{t=1}^T \sum_{i=1}^n E\left|\xi_{nt,i}\right|^{2+\eta} \to 0,\ \frac{1}{\sigma^{2+\eta}_{\mathbf{s}^{\sigma_V^2}_{nT,k}}} \sum_{t=1}^T \sum_{i=1}^n E\left|\xi^{\sigma^2_{V,k}}_{nt,i,k}\right|^{2+\eta} \to 0$$

and (ii) $\frac{1}{\sigma^2_{\mathbf{s}_{nT}}} \sum_{t=1}^T \sum_{i=1}^n E\left(\xi^2_{nt,i}|\mathcal{F}_{n,t,i-1}\right) \xrightarrow{p} 1$, $\frac{1}{\sigma^2_{\mathbf{s}^{\sigma_V^2}_{nT,k}}} \sum_{t=1}^T \sum_{i=1}^n E\left(\xi^{\sigma^2_{V,k}}_{nt,i,k}|\mathcal{F}_{n,t,i-1}\right) \xrightarrow{p} 1$. The first condition is a Liapounov's condition and the second one is for convergence of the conditional variances to the unconditional variances.

Step 2: Note that $\frac{1}{\sqrt{nT}} \frac{\partial \ln L^{(u)}_{nT}(\theta_0)}{\partial \theta}$ consists of $\mathbf{s}_{nT}$ and $\mathbf{s}^{\sigma_V^2}_{nT,k}$. Since we know the variances and covariances of $\mathbf{s}_{nT}$ and $\mathbf{s}^{\sigma_V^2}_{nT,k}$, the Cramér-Wold device can be applied. Then, we have $\frac{1}{\sqrt{nT}} \frac{\partial \ln L^{(u)}_{nT,c}(\theta_0)}{\partial \theta} \xrightarrow{d} N\left(0, \Omega_{\theta_0}\right)$ where $\Omega_{\theta_0} = \lim_{T\to\infty} \Omega_{\theta_0,nT}$.

Lemma 2.3 *Suppose Assumptions 2.2 - 2.5 hold. If the sequence* $\frac{1}{nT}\sigma^2_{\mathbf{s}_{nT}}$ *is bounded away from zero, then* $\frac{\mathbf{s}_{nT}}{\sigma_{\mathbf{s}_{nT}}} \xrightarrow{d} N(0,1)$.

Lemma 2.4 *Suppose Assumptions 2.2 - 2.5 hold. Then,*

(i) $\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{1,n,t-1} B_n \tilde{\mathbb{U}}_{1,n,t-1} - E\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{1,n,t-1} B_n \tilde{\mathbb{U}}_{1,n,t-1} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ *where* $E\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{1,n,t-1} B_n \tilde{\mathbb{U}}_{1,n,t-1} = O(1)$,

(ii) $\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{1,n,t-1} B_n \tilde{\mathbb{U}}_{2,n,t-1} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ *where* $E\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{1,n,t-1} B_n \tilde{\mathbb{U}}_{2,n,t-1} = 0$,

(iii) $\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{2,n,t-1} B_n \tilde{\mathbb{U}}_{2,n,t-1} - E\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{2,n,t-1} B_n \tilde{\mathbb{U}}_{2,n,t-1} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ *where* $E\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{2,n,t-1} B_n \tilde{\mathbb{U}}_{2,n,t-1} = O(1)$,

(iv) $\frac{1}{nT}\sum_{t=1}^T D'_{nt} B_n \tilde{\mathbb{U}}_{1,n,t-1} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ *where* D_{nt} *is an* $n \times 1$ *time-variant deterministic vector with all its elements bounded for all* n *and* t, *and* $E\frac{1}{nT}\sum_{t=1}^T D'_{nt} B_n \tilde{\mathbb{U}}_{1,n,t-1} = 0$,

(v) $\frac{1}{nT}\sum_{t=1}^T D'_{nt} B_n \tilde{\mathbb{U}}_{2,n,t-1} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ *where* $E\frac{1}{nT}\sum_{t=1}^T D'_{nt} B_n \tilde{\mathbb{U}}_{2,n,t-1} = 0$,

(vi) for $k, l = 1, \cdots, K$,

$$\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{3,n,t-1,k} B_n \tilde{\mathbb{U}}_{3,n,t-1,l} - E\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{3,n,t-1,k} B_n \tilde{\mathbb{U}}_{3,n,t-1,l} = O_p\left(\frac{1}{\sqrt{nT}}\right)$$

where $E\frac{1}{nT}\sum_{t=1}^T \tilde{\mathbb{U}}'_{3,n,t-1,k} B_n \tilde{\mathbb{U}}_{3,n,t-1,l} = O(1)$,

(vii) for $k = 1, \cdots, K$, $\frac{1}{nT}\sum_{t=1}^T D'_{nt} B_n \tilde{\mathbb{U}}_{3,n,t-1,k} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ *where* $E\frac{1}{nT}\sum_{t=1}^T D'_{nt} B_n \tilde{\mathbb{U}}_{3,n,t-1,k} = 0$,

(viii) for $k = 1, \cdots, K$,

$$\frac{1}{nT}\sum_{t=1}^{T}\left(\tilde{\mathbb{U}}_{1,n,t-1}+\tilde{\mathbb{U}}_{2,n,t-1}\right)B_n\tilde{\mathbb{U}}_{3,n,t-1,k}-E\frac{1}{nT}\sum_{t=1}^{T}\left(\tilde{\mathbb{U}}_{1,n,t-1}+\tilde{\mathbb{U}}_{2,n,t-1}\right)B_n\tilde{\mathbb{U}}_{3,n,t-1,k}=O_p\left(\frac{1}{\sqrt{nT}}\right)$$

where $E\frac{1}{nT}\sum_{t=1}^{T}\left(\tilde{\mathbb{U}}_{1,n,t-1}+\tilde{\mathbb{U}}_{2,n,t-1}\right)B_n\tilde{\mathbb{U}}_{3,n,t-1,k}=O(1)$.

From Lemma 2.4, we have the following results:

$$\text{(i) } \frac{1}{nT}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}^{(-)\prime}B_n\tilde{Y}_{n,t-1}^{(-)}-E\frac{1}{nT}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}^{(-)\prime}B_n\tilde{Y}_{n,t-1}^{(-)}=O_p\left(\frac{1}{\sqrt{nT}}\right)$$

where $E\frac{1}{nT}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}^{(-)\prime}B_n\tilde{Y}_{n,t-1}^{(-)}=O(1)$ from Lemma 2.4 (vi) and (vii),

$$\text{(ii) } \frac{1}{nT}\sum_{t=1}^{T}\tilde{X}_{n,t-1,k}^{(-)\prime}B_n\tilde{X}_{n,t-1,l}^{(-)\prime}-E\frac{1}{nT}\sum_{t=1}^{T}\tilde{X}_{n,t-1,k}^{(-)\prime}B_n\tilde{X}_{n,t-1,l}^{(-)\prime}=O_p\left(\frac{1}{\sqrt{nT}}\right)$$

where $E\frac{1}{nT}\sum_{t=1}^{T}\tilde{X}_{n,t-1,k}^{(-)\prime}B_n\tilde{X}_{n,t-1,l}^{(-)\prime}=O(1)$ from Lemma 2.4 for $k, l = 1, \cdots, K$, and

$$\text{(iii) } \frac{1}{nT}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}^{(-)\prime}B_n\tilde{X}_{n,t-1,k}^{(-)}-E\frac{1}{nT}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}^{(-)\prime}B_n\tilde{X}_{n,t-1,k}^{(-)}=O_p\left(\frac{1}{\sqrt{nT}}\right)$$

where $E\frac{1}{nT}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}^{(-)\prime}B_n\tilde{X}_{n,t-1,k}^{(-)}=O(1)$ from Lemma 2.4 (iv)-(v) and (vii)-(viii) for $k = 1, \cdots, K$.

2.6 Proofs of theorems: consistency and asymptotic normality

Proof of Theorem 4.1. We firstly show the uniform convergence of $\frac{1}{nT}\ln L_{nT}(\theta)-Q_{nT}(\theta)\xrightarrow{p}0$ uniformly in $\theta\in\Theta$. The main issue is whether the terms

$$\frac{1}{nT}\sum_{t=1}^{T}\left[\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)-E\left(\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)\right)\right]$$

and $\frac{1}{nT}\sum_{t=1}^{T}\left[\tilde{V}_{nt,k}'(\gamma_k,\rho_k)J_n\tilde{V}_{nt,k}(\gamma_k,\rho_k)-E\left(\tilde{V}_{nt,k}'(\gamma_k,\rho_k)J_n\tilde{V}_{nt,k}(\gamma_k,\rho_k)\right)\right]$ for $k = 1, \cdots, K$ converge to zero in probability uniformly in $\theta\in\Theta$. Let

$$\tilde{\mathcal{E}}_{nt}(\theta)=\tilde{\mathcal{E}}_{nt}^{A}(\theta)+R_n(\theta_1)R_n^{-1}\tilde{\mathcal{E}}_{nt}+\sum_{k=1}^{K}\left(R_n(\theta_1)R_n^{-1}B_{X,k,n}-B_{X,k,n}(\theta)\right)\tilde{V}_{nt,k}.$$

where $\tilde{\mathcal{E}}_{nt}^{A}(\theta)=\left[\begin{array}{c}\left(R_n(\theta_1)R_n^{-1}(\gamma_0 I_n+\rho_0 W_n)-(\gamma I_n+\rho W_n)\right)\tilde{Y}_{n,t-1}+\tilde{\alpha}_{t,0}R_n(\theta_1)R_n^{-1}l_n\\ +\sum_{k=1}^{K}\left(R_n(\theta_1)R_n^{-1}B_{X,k,n}-B_{X,k,n}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k}+\tilde{\alpha}_{t,k,0}l_n\right)\end{array}\right]$. Consider

$$
\begin{aligned}
&\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)\\
=\ &\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathcal{E}}_{nt}^{A\prime}(\theta)J_n\tilde{\mathcal{E}}_{nt}^{A}(\theta)+\frac{2}{nT}\sum_{t=1}^{T}\tilde{\mathcal{E}}_{nt}^{A\prime}(\theta)J_nR_n(\theta_1)R_n^{-1}\tilde{\mathcal{E}}_{nt}\\
&+\frac{2}{nT}\sum_{k=1}^{K}\sum_{t=1}^{T}\tilde{\mathcal{E}}_{nt}^{A\prime}(\theta)J_n\left(R_n(\theta_1)R_n^{-1}B_{X,k,n}-B_{X,k,n}(\theta)\right)\tilde{V}_{nt,k}+\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathcal{E}}_{nt}'R_n^{-1\prime}R_n'(\theta_1)J_nR_n(\theta_1)R_n^{-1}\tilde{\mathcal{E}}_{nt}\\
&+\frac{2}{nT}\sum_{k=1}^{K}\sum_{t=1}^{T}\tilde{\mathcal{E}}_{nt}'R_n^{-1\prime}R_n'(\theta_1)J_n\left(R_n(\theta_1)R_n^{-1}B_{X,k,n}-B_{X,k,n}(\theta)\right)\tilde{V}_{nt,k}\\
&+\frac{1}{nT}\sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{t=1}^{T}\tilde{V}_{nt,k}'\left(R_n(\theta_1)R_n^{-1}B_{X,k,n}-B_{X,k,n}(\theta)\right)'J_n\left(R_n(\theta_1)R_n^{-1}B_{X,l,n}-B_{X,l,n}(\theta)\right)\tilde{V}_{nt,l}.
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt,k}'(\gamma_k,\rho_k)J_n\tilde{V}_{nt,k}(\gamma_k,\rho_k)\\
=\ &\frac{1}{nT}\sum_{t=1}^{T}\tilde{X}_{n,t-1,k}'\left((\gamma_{k,0}-\gamma_k)I_n+(\rho_{k,0}-\rho_k)W_n\right)'J_n\left((\gamma_{k,0}-\gamma_k)I_n+(\rho_{k,0}-\rho_k)W_n\right)\tilde{X}_{n,t-1,k}\\
&+\frac{2}{nT}\sum_{t=1}^{T}\tilde{X}_{n,t-1,k}'\left((\gamma_{k,0}-\gamma_k)I_n+(\rho_{k,0}-\rho_k)W_n\right)'J_n\tilde{V}_{nt,k}+\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt,k}'J_n\tilde{V}_{nt,k}.
\end{aligned}
$$

Since (i) θ is bounded in the compact parameter space Θ and (ii) $R_n(\theta_1)$ and R_n^{-1} are uniformly bounded in $\theta\in\Theta$ and (iii) $B_{X,k,n}(\theta)$ and $B_{X.k,n}$ are uniformly bounded in $\theta\in\Theta$, $R_n(\theta_1)R_n^{-1}-I_n$ and $B_{X,k,n}-B_{X,k,n}(\theta)$ (for $k=1,\cdots,K$) are also uniformly bounded in $\theta\in\Theta$. By using Lemmas 8 and 15 in Yu et al. (2008), therefore,

$$
\frac{1}{nT}\sum_{t=1}^{T}\left[\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)-E\left(\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)\right)\right]\xrightarrow{p}0
$$

and $\frac{1}{nT}\sum_{t=1}^{T}\left[\tilde{V}_{nt,k}'(\gamma_k,\rho_k)J_n\tilde{V}_{nt,k}(\gamma_k,\rho_k)-E\left(\tilde{V}_{nt,k}'(\gamma_k,\rho_k)J_n\tilde{V}_{nt,k}(\gamma_k,\rho_k)\right)\right]\xrightarrow{p}0$ uniformly in $\theta\in\Theta$. Since σ_ϵ^2, $\sigma_{V,1}^2$, $\cdots$, $\sigma_{V,K}^2$ are bounded away from zero,

$$
\begin{aligned}
&\frac{1}{nT}\ln L_{nT}(\theta)-Q_{nT}(\theta)\\
=\ &-\frac{1}{2\sigma_\epsilon^2}\frac{1}{nT}\sum_{t=1}^{T}\left[\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)-E\left(\tilde{\mathcal{E}}_{nt}'(\theta)J_n\tilde{\mathcal{E}}_{nt}(\theta)\right)\right]\\
&-\frac{1}{nT}\sum_{k=1}^{K}\frac{1}{2\sigma_{V,k}^2}\sum_{t=1}^{T}\left[\tilde{V}_{nt,k}'(\gamma_k,\rho_k)J_n\tilde{V}_{nt,k}(\gamma_k,\rho_k)-E\left(\tilde{V}_{nt,k}'(\gamma_k,\rho_k)J_n\tilde{V}_{nt,k}(\gamma_k,\rho_k)\right)\right]\xrightarrow{p}0
\end{aligned}
$$

uniformly in $\theta \in \Theta$.

Secondly, we shall show that $Q_{nT}(\theta)$ is uniformly equicontinuous in $\theta \in \Theta$. Note that

$$\frac{1}{nT}\sum_{t=1}^{T} E\left(\tilde{\mathcal{E}}_{nt}'(\theta) J_n \tilde{\mathcal{E}}_{nt}(\theta)\right) = q_{nT,1}(\theta) + q_{nT,2}(\theta_1) + \sum_{k=1}^{K} q_{nT,3,k}(\theta) + o(1)$$

and $\frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{K} E\left(\tilde{V}_{nt,k}'(\gamma_k,\rho_k) J_n \tilde{V}_{nt,k}(\gamma_k,\rho_k)\right) = \sum_{k=1}^{K} q_{nT,4,k}(\gamma_k,\rho_k) + \frac{(n-1)(T-1)}{nT}\sum_{k=1}^{K}\sigma_{V,k,0}^2 + o(1)$ where

$$\begin{aligned} q_{nT,1}(\theta) &= \frac{1}{nT}\sum_{t=1}^{T} E\left(\begin{array}{c} \left(R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ + \sum_{k=1}^{K}\left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right)' \\ &\quad \times J_n \left(\begin{array}{c} \left(R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ + \sum_{k=1}^{K}\left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right), \end{aligned}$$

$$q_{nT,2}(\theta_1) = \frac{T-1}{nT}\sigma_{\epsilon,0}^2 tr\left(R_n^{-1\prime} R_n'(\theta_1) J_n R_n(\theta_1) R_n^{-1}\right),$$

$$q_{nT,3,k}(\theta) = \frac{T-1}{nT}\sigma_{V,k,0}^2 tr\left(\left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)' J_n \left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)\right),$$

and $q_{nT,4,k}(\gamma_k,\rho_k) = \frac{1}{nT}\sum_{t=1}^{T} E\left(\begin{array}{c} \tilde{X}_{n,t-1,k}'\left((\gamma_{k,0}-\gamma_k) I_n + (\rho_{k,0}-\rho_k) W_n\right)' \\ \times J_n\left((\gamma_{k,0}-\gamma_k) I_n + (\rho_{k,0}-\rho_k) W_n\right)\tilde{X}_{n,t-1,k} \end{array}\right)$ for $k = 1, \cdots, K$.

To show the uniform equicontinuity of $Q_{nT}(\theta)$, we should verify (i) $\ln\sigma_\epsilon^2$ is uniformly continuous, (ii) $\ln\sigma_{V,k}^2$, $k = 1, \cdots, K$, are uniformly continuous, (iii) $\frac{1}{n}\ln|R_n(\theta_1)|$ is uniformly equicontinuous, and (iv) $q_{nT,1}(\theta)$, $q_{nT,2}(\theta_1)$, $\{q_{nT,3,k}(\theta)\}_{k=1}^{K}$ and $\{q_{nT,4,k}(\gamma_k,\rho_k)\}_{k=1}^{K}$ are uniformly equicontinuous.

(i) and (ii) hold because σ_ϵ^2, $\sigma_{V,1}^2, \cdots, \sigma_{V,K}^2$ are bounded away from zero in Θ. Consider (iii). For $\theta_{1,1}$, $\theta_{1,2}$ in Θ,

$$\begin{aligned} &\frac{1}{n}\ln|R_n(\theta_{1,1})| - \frac{1}{n}\ln|R_n(\theta_{1,2})| \\ =\ &\frac{1}{n}tr\left(R_{n\lambda}(\bar{\theta}_1) R_n^{-1}(\bar{\theta}_1)\right)\cdot(\lambda_1-\lambda_2) + \frac{1}{n}tr\left(R_{n\gamma}(\bar{\theta}_1) R_n^{-1}(\bar{\theta}_1)\right)\cdot(\gamma_1-\gamma_2) + \frac{1}{n}tr\left(R_{n\rho}(\bar{\theta}_1) R_n^{-1}(\bar{\theta}_1)\right)\cdot(\rho_1-\rho_2) \end{aligned}$$

where $\theta_{1,1} = (\lambda_1,\gamma_1,\rho_1)'$, $\theta_{1,2} = (\lambda_2,\gamma_2,\rho_2)'$ and $\bar{\theta}_1$ lies between $\theta_{1,1}$ and $\theta_{1,2}$. Since $R_{n\lambda}(\theta_1)$, $R_{n\gamma}(\theta_1)$, $R_{n\rho}(\theta_1)$ and $R_n^{-1}(\theta_1)$ are uniformly bounded for all θ_1 in Θ, $\frac{1}{n}tr\left(R_{n\lambda}(\bar{\theta}_1) R_n^{-1}(\bar{\theta}_1)\right)$, $\frac{1}{n}tr\left(R_{n\gamma}(\bar{\theta}_1) R_n^{-1}(\bar{\theta}_1)\right)$ and $\frac{1}{n}tr\left(R_{n\rho}(\bar{\theta}_1) R_n^{-1}(\bar{\theta}_1)\right)$ are bounded. Hence, we have the uniform equicontinuity of $\frac{1}{n}\ln|R_n(\theta_1)|$. Last, we consider (iv). By the Taylor expansion, for θ_a, $\theta_b \in \Theta$

$$q_{nT,1}(\theta_a) - q_{nT,1}(\theta_b) = \frac{\partial q_{1,nT}(\bar{\theta})}{\partial\theta'}(\theta_a - \theta_b)$$

where $\bar{\theta}$ lies between θ_a and θ_b and the components of $\frac{\partial q_{nT,1}(\theta)}{\partial \theta}$ are

$$\frac{\partial q_{nT,1}(\theta)}{\partial \lambda} = \begin{array}{l} E\frac{1}{nT}\sum_{t=1}^{T} 2\left[\begin{array}{c} R_{n\lambda}(\theta_1) R_n^{-1}\left((\gamma_0 I_n + \rho_0 W_n)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} l_n\right) \\ +\sum_{k=1}^{K}\left(R_{n\lambda}(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n,\lambda}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right]' \\ \times J_n\left[\begin{array}{c} \left(R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ +\sum_{k=1}^{K}\left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right], \end{array}$$

where $B_{X,k,n,\lambda}(\theta) = \delta D_{n\lambda,k}(\theta_1, \gamma_k, \rho_k) A_{k,n}(\gamma_k, \rho_k)(\beta_{1,k} I_n + \beta_{2,k} W_n)$,

$$\frac{\partial q_{nT,1}(\theta_1)}{\partial \gamma} = \begin{array}{l} E\frac{1}{nT}\sum_{t=1}^{T} 2\left[\begin{array}{c} \left(R_{n\gamma}(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - I_n\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_{n\gamma}(\theta_1) R_n^{-1} l_n \\ +\sum_{k=1}^{K}\left(R_{n\gamma}(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n,\gamma}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right]' \\ \times J_n\left[\begin{array}{c} \left(R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ +\sum_{k=1}^{K}\left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right], \end{array}$$

where $B_{X,k,n,\gamma}(\theta) = \delta D_{n\gamma,k}(\theta_1, \gamma_k, \rho_k) A_{k,n}(\gamma_k, \rho_k)(\beta_{1,k} I_n + \beta_{2,k} W_n)$,

$$\frac{\partial q_{nT,1}(\theta)}{\partial \rho} = \begin{array}{l} E\frac{1}{nT}\sum_{t=1}^{T} 2\left[\begin{array}{c} \left(R_{n\rho}(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - W_n\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_{n\rho}(\theta_1) R_n^{-1} l_n \\ +\sum_{k=1}^{K}\left(R_{n\rho}(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n,\rho}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right]' \\ \times J_n\left[\begin{array}{c} \left(R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ +\sum_{k=1}^{K}\left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right], \end{array}$$

where $B_{X,k,n,\rho}(\theta) = \delta D_{n\rho,k}(\theta_1, \gamma_k, \rho_k) A_{k,n}(\gamma_k, \rho_k)(\beta_{1,k} I_n + \beta_{2,k} W_n)$,

$$\frac{\partial q_{nT,1}(\theta)}{\partial \beta_{1,k}} = \begin{array}{l} -E\frac{1}{nT}\sum_{t=1}^{T} 2\left[(I_n + \delta \mathbf{D}_{n,k}(\theta_1, \gamma_k, \rho_k) A_{k,n}(\gamma_k, \rho_k))\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right)\right]' \\ \times J_n\left[\begin{array}{c} \left(R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ +\sum_{k=1}^{K}\left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right], \end{array}$$

$$\frac{\partial q_{nT,1}(\theta)}{\partial \beta_{2,k}} = \begin{array}{l} -E\frac{1}{nT}\sum_{t=1}^{T} 2\left[(I_n + \delta \mathbf{D}_{n,k}(\theta_1, \gamma_k, \rho_k) A_{k,n}(\gamma_k, \rho_k)) W_n\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right)\right]' \\ \times J_n\left[\begin{array}{c} \left(R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ +\sum_{k=1}^{K}\left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right], \end{array}$$

$$\frac{\partial q_{nT,1}(\theta)}{\partial \gamma_k} = \begin{array}{l} -E\frac{1}{nT}\sum_{t=1}^{T} 2\left[\begin{array}{c} \delta\left(\mathbf{D}_{n,\gamma_k,k}(\theta_1, \gamma_k, \rho_k) A_{k,n}(\gamma_k, \rho_k) + \mathbf{D}_{n,k}(\theta_1, \gamma_k, \rho_k)\right) \\ \times(\beta_{1,k} I_n + \beta_{2,k} W_n)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right]' \\ \times J_n\left[\begin{array}{c} \left(R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ +\sum_{k=1}^{K}\left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right], \end{array}$$

$$\frac{\partial q_{nT,1}(\theta)}{\partial \rho_k} = \begin{array}{l} -E\frac{1}{nT}\sum_{t=1}^{T} 2\left[\begin{array}{c} \delta\left(\mathbf{D}_{n,\rho_k,k}(\theta_1, \gamma_k, \rho_k) A_{k,n}(\gamma_k, \rho_k) + \mathbf{D}_{n,k}(\theta_1, \gamma_k, \rho_k)\right) W_n \\ \times(\beta_{1,k} I_n + \beta_{2,k} W_n)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right]' \\ \times J_n\left[\begin{array}{c} \left(R_n(\theta_1) R_n^{-1}(\gamma_0 I_n + \rho_0 W_n) - (\gamma I_n + \rho W_n)\right)\tilde{Y}_{n,t-1} + \tilde{\alpha}_{t,0} R_n(\theta_1) R_n^{-1} l_n \\ +\sum_{k=1}^{K}\left(R_n(\theta_1) R_n^{-1} B_{X,k,n} - B_{X,k,n}(\theta)\right)\left(A_{k,n}\tilde{X}_{n,t-1,k} + \tilde{\alpha}_{t,k,0} l_n\right) \end{array}\right], \end{array}$$

$\frac{\partial q_{nT,1}(\theta)}{\partial \sigma_\epsilon^2} = 0$, and $\frac{\partial q_{nT,1}(\theta)}{\partial \sigma_{V,k}^2} = 0$ for $k = 1, \cdots, K$.

Since (i) $R_{n\lambda}(\theta_1)$, $R_{n\gamma}(\theta_1)$, $R_{n\rho}(\theta_1)$, R_n^{-1}, $\mathbf{D}_{n,k}(\theta_1,\gamma_k,\rho_k)$, $\mathbf{D}_{n\lambda,k}(\theta_1,\gamma_k,\rho_k)$, $\mathbf{D}_{n\gamma,k}(\theta_1,\gamma_k,\rho_k)$, $\mathbf{D}_{n\rho,k}(\theta_1,\gamma_k,\rho_k)$, $\mathbf{D}_{n,\gamma_k,k}(\theta_1,\gamma_k,\rho_k)$ and $\mathbf{D}_{n,\rho_k,k}(\theta_1,\gamma_k,\rho_k)$ (for all k) are uniformly bounded, (ii) θ is bounded in the compact parameter space Θ and (iii) $E\frac{1}{nT}\sum_{t=1}^T \tilde{Z}_{nt}' B_n \tilde{Z}_{nt} = O(1)$ from Lemma 2.4 where $\tilde{Z}_{nt}$ can take $\tilde{Y}_{n,t-1}$, $\tilde{X}_{n,t-1,k}$'s, $\tilde{\alpha}_{t,0}l_n$ and $\tilde{\alpha}_{t,k,0}l_n$ and B_n is an $n\times n$ uniformly bounded matrix, $\frac{\partial q_{nT,1}(\bar{\theta})}{\partial \theta'}$ is bounded. Thus, $q_{nT,1}(\theta)$ is uniformly equicontinuous. For uniformly equicontinuous of $q_{nT,2}(\theta_1)$, it suffices to show $\frac{1}{n}\sigma_{\epsilon,0}^2 tr\left(R_n^{-1\prime}R_n'(\theta_1)J_nR_n(\theta_1)R_n^{-1}\right)$ is uniformly equicontinuous. By using the expansion of $R_n(\theta_1)R_n^{-1}-I_n$, for $\theta_{1,1}$, $\theta_{1,2}$ in Θ we have

$$\frac{\sigma_{\epsilon,0}^2}{n}\left[tr\left(R_n^{-1\prime}R_n'(\theta_{1,1})J_nR_n(\theta_{1,1})R_n^{-1}\right) - tr\left(R_n^{-1\prime}R_n'(\theta_{1,2})J_nR_n(\theta_{1,2})R_n^{-1}\right)\right]$$

$$= \frac{\sigma_{\epsilon,0}^2}{n}\left[\begin{array}{c} -2(\lambda_1-\lambda_2)tr\left(J_n\left(-R_{n\lambda}(\bar{\theta}_1)R_n^{-1}\right)\right) - 2(\gamma_1-\gamma_2)tr\left(J_n\left(-R_{n\gamma}(\bar{\theta}_1)R_n^{-1}\right)\right) \\ -2(\rho_1-\rho_2)tr\left(J_n\left(-R_{n\rho}(\bar{\theta}_1)R_n^{-1}\right)\right) \\ +(\lambda_1-\lambda_2)(\lambda_1+\lambda_2-2\lambda_0)tr\left(\left(-R_{n\lambda}(\bar{\theta}_1)R_n^{-1}\right)'J_n\left(-R_{n\lambda}(\bar{\theta}_1)R_n^{-1}\right)\right) \\ +2\left(-\lambda_0(\gamma_1-\gamma_2)-\gamma_0(\lambda_1-\lambda_2)+(\lambda_1\gamma_1-\lambda_2\gamma_2)\right)tr\left(\left(-R_{n\lambda}(\bar{\theta}_1)R_n^{-1}\right)'J_n\left(-R_{n\gamma}(\bar{\theta}_1)R_n^{-1}\right)\right) \\ +2\left(-\lambda_0(\rho_1-\rho_2)-\rho_0(\lambda_1-\lambda_2)+(\lambda_1\rho_1-\lambda_2\rho_2)\right)tr\left(\left(-R_{n\lambda}(\bar{\theta}_1)R_n^{-1}\right)'J_n\left(-R_{n\rho}(\bar{\theta}_1)R_n^{-1}\right)\right) \\ +(\gamma_1-\gamma_2)(\gamma_1+\gamma_2-2\gamma_0)tr\left(\left(-R_{n\gamma}(\bar{\theta}_1)R_n^{-1}\right)'J_n\left(-R_{n\gamma}(\bar{\theta}_1)R_n^{-1}\right)\right) \\ +2\left(-\gamma_0(\rho_1-\rho_2)-\rho_0(\gamma_1-\gamma_2)+(\gamma_1\rho_1-\gamma_2\rho_2)\right)tr\left(\left(-R_{n\gamma}(\bar{\theta}_1)R_n^{-1}\right)'J_n\left(-R_{n\rho}(\bar{\theta}_1)R_n^{-1}\right)\right) \\ +(\rho_1-\rho_2)(\rho_1+\rho_2-2\rho_0)tr\left(\left(-R_{n\rho}(\bar{\theta}_1)R_n^{-1}\right)'J_n\left(-R_{n\rho}(\bar{\theta}_1)R_n^{-1}\right)\right) \end{array}\right]$$

where $\bar{\theta}_1$ lies between $\theta_{1,1}$ and $\theta_{1,2}$. Since $R_n(\theta_1)$, $R_{n\lambda}(\theta_1)$, $R_{n\gamma}(\theta_1)$, $R_{n\rho}(\theta_1)$ and R_n^{-1} are uniformly bounded, we obtain the uniform equicontinuity of $q_{nT,2}(\theta_1)$. To show the uniform equicontinuity of $q_{nT,3,k}(\theta)$ for $k = 1, \cdots, K$, it is enough to verify that property of $\frac{\sigma_{V,k,0}^2}{n} tr\left(\left(R_n(\theta)R_n^{-1}B_{X,k,n} - B_{X,k,n}(\theta)\right)'J_n\left(R_n(\theta)R_n^{-1}B_{X,k,n} - B_{X,k,n}(\theta)\right)\right)$. By employing the expansion of $R_n(\theta)R_n^{-1}B_{X,k,n} - B_{X,k,n}(\theta)$, we have the following decomposition: for θ_1, θ_2 in Θ and $k = 1, \cdots, K$, the difference

$$\frac{\sigma_{V,k,0}^2}{n}\left[\begin{array}{c} tr\left(\left(R_n(\theta_1)R_n^{-1}B_{X,k,n} - B_{X,k,n}(\theta_1)\right)'J_n\left(R_n(\theta_1)R_n^{-1}B_{X,k,n} - B_{X,k,n}(\theta_1)\right)\right) \\ -tr\left(\left(R_n(\theta_2)R_n^{-1}B_{X,k,n} - B_{X,k,n}(\theta_2)\right)'J_n\left(R_n(\theta_2)R_n^{-1}B_{X,k,n} - B_{X,k,n}(\theta_2)\right)\right) \end{array}\right]$$

is

$\frac{\sigma_{V,k,0}^2}{n}[(\lambda_1-\lambda_2)(\lambda_1+\lambda_2-2\lambda_0)\cdot tr\left(C_{n,k}^{\lambda}(\bar{\theta})'J_nC_{n,k}^{\lambda}(\bar{\theta})\right)$
$+2\left(-\lambda_0(\gamma_1-\gamma_2)-\gamma_0(\lambda_1-\lambda_2)+(\lambda_1\gamma_1-\lambda_2\gamma_2)\right)\cdot tr\left(C_{n,k}^{\lambda}(\bar{\theta})'J_nC_{n,k}^{\gamma}(\bar{\theta})\right)$
$+2\left(-\lambda_0(\rho_1-\rho_2)-\rho_0(\lambda_1-\lambda_2)+(\lambda_1\rho_1-\lambda_2\rho_2)\right)\cdot tr\left(C_{n,k}^{\lambda}(\bar{\theta})'J_nC_{n,k}^{\rho}(\bar{\theta})\right)$

$$
\begin{aligned}
&+2\left(-\lambda_0\left(\beta_{1,k,1}-\beta_{1,k,2}\right)-\beta_{1,k,0}\left(\lambda_1-\lambda_2\right)+\left(\lambda_1\beta_{1,k,1}-\lambda_2\beta_{1,k,2}\right)\right)\cdot tr\left(C_{n,k}^{\lambda}\left(\bar{\theta}\right)' J_n C_{n,k}^{\beta_1,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\lambda_0\left(\beta_{2,k,1}-\beta_{2,k,2}\right)-\beta_{2,k,0}\left(\lambda_1-\lambda_2\right)+\left(\lambda_1\beta_{2,k,1}-\lambda_2\beta_{2,k,2}\right)\right)\cdot tr\left(C_{n,k}^{\lambda}\left(\bar{\theta}\right)' J_n C_{n,k}^{\beta_2,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\lambda_0\left(\gamma_{k,1}-\gamma_{k,2}\right)-\gamma_{k,0}\left(\lambda_1-\lambda_2\right)+\left(\lambda_1\gamma_{k,1}-\lambda_2\gamma_{k,2}\right)\right)\cdot tr\left(C_{n,k}^{\lambda}\left(\bar{\theta}\right)' J_n C_{n,k}^{\gamma_X,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\lambda_0\left(\rho_{k,1}-\rho_{k,2}\right)-\rho_{k,0}\left(\lambda_1-\lambda_2\right)+\left(\lambda_1\rho_{k,1}-\lambda_2\rho_{k,2}\right)\right)\cdot tr\left(C_{n,k}^{\lambda}\left(\bar{\theta}\right)' J_n C_{n,k}^{\rho_X,k}\left(\bar{\theta}\right)\right)\\
&+\left(\gamma_1-\gamma_2\right)\left(\gamma_1+\gamma_2-2\gamma_0\right)\cdot tr\left(C_{n,k}^{\gamma}\left(\bar{\theta}\right)' J_n C_{n,k}^{\gamma}\left(\bar{\theta}\right)\right)\\
&+2\left(-\gamma_0\left(\rho_1-\rho_2\right)-\rho_0\left(\gamma_1-\gamma_2\right)+\left(\gamma_1\rho_1-\gamma_2\rho_2\right)\right)\cdot tr\left(C_{n,k}^{\gamma}\left(\bar{\theta}\right)' J_n C_{n,k}^{\rho}\left(\bar{\theta}\right)\right)\\
&+2\left(-\gamma_0\left(\beta_{1,k,1}-\beta_{1,k,2}\right)-\beta_{1,k,0}\left(\gamma_1-\gamma_2\right)+\left(\gamma_1\beta_{1,k,1}-\gamma_2\beta_{1,k,2}\right)\right)\cdot tr\left(C_{n,k}^{\gamma}\left(\bar{\theta}\right)' J_n C_{n,k}^{\beta_1,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\gamma_0\left(\beta_{2,k,1}-\beta_{2,k,2}\right)-\beta_{2,k,0}\left(\gamma_1-\gamma_2\right)+\left(\gamma_1\beta_{2,k,1}-\gamma_2\beta_{2,k,2}\right)\right)\cdot tr\left(C_{n,k}^{\gamma}\left(\bar{\theta}\right)' J_n C_{n,k}^{\beta_2,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\gamma_0\left(\gamma_{k,1}-\gamma_{k,2}\right)-\gamma_{k,0}\left(\gamma_1-\gamma_2\right)+\left(\gamma_1\gamma_{k,1}-\gamma_2\gamma_{k,2}\right)\right)\cdot tr\left(C_{n,k}^{\gamma}\left(\bar{\theta}\right)' J_n C_{n,k}^{\gamma_X,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\gamma_0\left(\rho_{k,1}-\rho_{k,2}\right)-\rho_{k,0}\left(\gamma_1-\gamma_2\right)+\left(\gamma_1\rho_{k,1}-\gamma_2\rho_{k,2}\right)\right)\cdot tr\left(C_{n,k}^{\gamma}\left(\bar{\theta}\right)' J_n C_{n,k}^{\rho_X,k}\left(\bar{\theta}\right)\right)\\
&+\left(\rho_1-\rho_2\right)\left(\rho_1+\rho_2-2\rho_0\right)\cdot tr\left(C_{n,k}^{\rho}\left(\bar{\theta}\right)' J_n C_{n,k}^{\rho}\left(\bar{\theta}\right)\right)\\
&+2\left(-\rho_0\left(\beta_{1,k,1}-\beta_{1,k,2}\right)-\beta_{1,k,0}\left(\rho_1-\rho_2\right)+\left(\rho_1\beta_{1,k,1}-\rho_2\beta_{1,k,2}\right)\right)\cdot tr\left(C_{n,k}^{\rho}\left(\bar{\theta}\right)' J_n C_{n,k}^{\beta_1,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\rho_0\left(\beta_{2,k,1}-\beta_{2,k,2}\right)-\beta_{2,k,0}\left(\rho_1-\rho_2\right)+\left(\rho_1\beta_{2,k,1}-\rho_2\beta_{2,k,2}\right)\right)\cdot tr\left(C_{n,k}^{\rho}\left(\bar{\theta}\right)' J_n C_{n,k}^{\beta_2,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\rho_0\left(\gamma_{k,1}-\gamma_{k,2}\right)-\gamma_{k,0}\left(\rho_1-\rho_2\right)+\left(\rho_1\gamma_{k,1}-\rho_2\gamma_{k,2}\right)\right)\cdot tr\left(C_{n,k}^{\rho}\left(\bar{\theta}\right)' J_n C_{n,k}^{\gamma_X,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\rho_0\left(\rho_{k,1}-\rho_{k,2}\right)-\rho_{X,k,0}\left(\rho_1-\rho_2\right)+\left(\rho_1\rho_{k,1}-\rho_2\rho_{k,2}\right)\right)\cdot tr\left(C_{n,k}^{\rho}\left(\bar{\theta}\right)' J_n C_{n,k}^{\rho_X,k}\left(\bar{\theta}\right)\right)\\
&+\left(\beta_{1,k,1}-\beta_{1,k,2}\right)\left(\beta_{1,k,1}+\beta_{1,k,2}-2\beta_{1,k,0}\right)\cdot tr\left(C_{n,k}^{\beta_1,k}\left(\bar{\theta}\right)' J_n C_{n,k}^{\beta_1,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\beta_{1,k,0}\left(\beta_{2,k,1}-\beta_{2,k,2}\right)-\beta_{2,k,0}\left(\beta_{1,k,1}-\beta_{1,k,2}\right)+\left(\beta_{1,k,1}\beta_{2,k,1}-\beta_{1,k,2}\beta_{2,k,2}\right)\right)\cdot tr\left(C_{n,k}^{\beta_1,k}\left(\bar{\theta}\right)' J_n C_{n,k}^{\beta_2,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\beta_{1,k,0}\left(\gamma_{k,1}-\gamma_{k,2}\right)-\gamma_{k,0}\left(\beta_{1,k,1}-\beta_{1,k,2}\right)+\left(\beta_{1,k,1}\gamma_{k,1}-\beta_{1,k,2}\gamma_{k,2}\right)\right)\cdot tr\left(C_{n,k}^{\beta_1,k}\left(\bar{\theta}\right)' J_n C_{n,k}^{\gamma_X,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\beta_{1,k,0}\left(\rho_{k,1}-\rho_{k,2}\right)-\rho_{k,0}\left(\beta_{1,k,1}-\beta_{1,k,2}\right)+\left(\beta_{1,k,1}\rho_{k,1}-\beta_{1,k,2}\rho_{k,2}\right)\right)\cdot tr\left(C_{n,k}^{\beta_1,k}\left(\bar{\theta}\right)' J_n C_{n,k}^{\rho_X,k}\left(\bar{\theta}\right)\right)\\
&+\left(\beta_{2,k,1}-\beta_{2,k,2}\right)\left(\beta_{2,k,1}+\beta_{2,k,2}-2\beta_{2,k,0}\right)\cdot tr\left(C_{n,k}^{\beta_2,k}\left(\bar{\theta}\right)' J_n C_{n,k}^{\beta_2,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\beta_{2,k,0}\left(\gamma_{k,1}-\gamma_{k,2}\right)-\gamma_{k,0}\left(\beta_{2,k,1}-\beta_{2,k,2}\right)+\left(\beta_{2,k,1}\gamma_{X,k,1}-\beta_{2,k,2}\gamma_{X,k,2}\right)\right)\cdot tr\left(C_{n,k}^{\beta_2,k}\left(\bar{\theta}\right)' J_n C_{n,k}^{\gamma_X,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\beta_{2,k,0}\left(\rho_{k,1}-\rho_{k,2}\right)-\rho_{k,0}\left(\beta_{2,k,1}-\beta_{2,k,2}\right)+\left(\beta_{2,k,1}\rho_{X,k,1}-\beta_{2,k,2}\rho_{X,k,2}\right)\right)\cdot tr\left(C_{n,k}^{\beta_2,k}\left(\bar{\theta}\right)' J_n C_{n,k}^{\rho_X,k}\left(\bar{\theta}\right)\right)\\
&+\left(\gamma_{k,1}-\gamma_{k,2}\right)\left(\gamma_{k,1}+\gamma_{k,2}-2\gamma_{k,0}\right)\cdot tr\left(C_{n,k}^{\gamma_X,k}\left(\bar{\theta}\right)' J_n C_{n,k}^{\gamma_X,k}\left(\bar{\theta}\right)\right)\\
&+2\left(-\gamma_{k,0}\left(\rho_{k,1}-\rho_{k,2}\right)-\rho_{k,0}\left(\gamma_{k,1}-\gamma_{k,2}\right)+\left(\gamma_{k,1}\rho_{k,1}-\gamma_{k,2}\rho_{k,2}\right)\right)\cdot tr\left(C_{n,k}^{\gamma_X,k}\left(\bar{\theta}\right)' J_n C_{n,k}^{\rho_X,k}\left(\bar{\theta}\right)\right)\\
&+\left(\rho_{k,1}-\rho_{k,2}\right)\left(\rho_{k,1}+\rho_{k,2}-2\rho_{k,0}\right)\cdot tr\left(C_{n,k}^{\rho_X,k}\left(\bar{\theta}\right)' J_n C_{n,k}^{\rho_X,k}\left(\bar{\theta}\right)\right)].
\end{aligned}
$$

Since $R_n(\theta_1)$, $R_{n\lambda}(\theta_1)$, $R_{n\gamma}(\theta_1)$, $R_{n\rho}(\theta_1)$, R_n^{-1}, $\mathbf{D}_{n,k}(\theta_1,\gamma_k,\rho_k)$, $\mathbf{D}_{n\lambda,k}(\theta_1,\gamma_k,\rho_k)$, $\mathbf{D}_{n\gamma,k}(\theta_1,\gamma_k,\rho_k)$, $\mathbf{D}_{n\rho,k}(\theta_1,\gamma_k,\rho_k)$ $\mathbf{D}_{n,\gamma_{,k},k}(\theta_1,\gamma_k,\rho_k)$ and $\mathbf{D}_{n,\rho_k,k}(\theta_1,\gamma_k,\rho_k)$ (for all $k=1,\cdots,K$) are uniformly bounded for any θ in Θ, we obtain the uniform equicontinuity of $q_{nT,3,k}(\theta)$. Last, the uniform equicontinuity of $\{q_{nT,4,k}(\gamma_k,\rho_k)\}_{k=1}^K$ can be verified because $\frac{1}{nT}\sum_{t=1}^T E\begin{bmatrix}\tilde{X}'_{n,t-1,k}\\ \tilde{X}'_{n,t-1,k}W'_n\end{bmatrix} J_n\begin{bmatrix}\tilde{X}_{n,t-1,k} & W_n\tilde{X}_{n,t-1,k}\end{bmatrix}=O(1)$ for all $k=1,\cdots,K$.

By combining the results by the two steps above and the identification uniqueness assumption, we obtain

$\hat{\theta}_{ml,nT} \xrightarrow{p} \theta_0$. ■

Proof of Theorem 4.2. By the Taylor expansion,

$$\sqrt{nT}\left(\hat{\theta}_{ml,nT}-\theta_0\right)=\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{(u)}_{nT,c}\left(\theta_0\right)}{\partial\theta}-\Delta_{1,nT}-\Delta_{2,nT}\right)$$

where $\bar{\theta}_{nT}$ lies between θ_0 and $\hat{\theta}_{ml,nT}$. Note that $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}(\theta)}{\partial\theta\partial\theta'}$ contains the first and second derivatives of $R_n\left(\theta_1\right)$ and $\{\mathbf{D}_{n,k}\left(\theta_1,\gamma_k,\rho_k\right)\}_{k=1}^{K}$, the difference between $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}(\theta)}{\partial\theta\partial\theta'}$ and $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}(\theta_0)}{\partial\theta\partial\theta'}$ can be characterized by (from the Taylor approximation) $\theta-\theta_0$ multiplied by some function containing up to the third derivatives of $R_n\left(\theta_1\right)$ and $\{\mathbf{D}_{n,k}\left(\theta_1,\gamma_k,\rho_k\right)\}_{k=1}^{K}$. Because we assume existence and uniform boundedness of the first, second and third derivatives of $R_n\left(\theta_1\right)$ and $\{\mathbf{D}_{n,k}\left(\theta_1,\gamma_k,\rho_k\right)\}_{k=1}^{K}$, we can apply the similar strategies of (38) and (39) in Yu et al. (2008), and show that

$$\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}-\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\theta_0\right)}{\partial\theta\partial\theta'}\right)\right)=\left\|\bar{\theta}_{nT}-\theta_0\right\|\cdot O_p\left(1\right) \tag{17}$$

and

$$\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\theta_0\right)}{\partial\theta\partial\theta'}-\Sigma_{\theta_0,nT}\right)=O_p\left(\frac{1}{\sqrt{nT}}\right). \tag{18}$$

Hence,

$$\begin{aligned}-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'} &= \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}-\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\theta_0\right)}{\partial\theta\partial\theta'}\right)\right)\\ &\quad+\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\theta_0\right)}{\partial\theta\partial\theta'}-\Sigma_{\theta_0,nT}\right)+\Sigma_{\theta_0,nT}\\ &= \left\|\bar{\theta}_{nT}-\theta_0\right\|\cdot O_p\left(1\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)+\Sigma_{\theta_0,nT}.\end{aligned}$$

Note that $\left\|\bar{\theta}_{nT}-\theta_0\right\|=o_p\left(1\right)$ by Theorem 4.1 in the main text and $\Sigma_{\theta_0,nT}$ is nonsingular in θ in some neighborhood of θ_0 under large n and T by Assumption 4.8. Hence, $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}$ is invertible and it is of $O_p\left(1\right)$. Then, we have

$$\sqrt{nT}\left(\hat{\theta}_{ml,nT}-\theta_0\right)=\underbrace{\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT,c}\left(\bar{\theta}_{nT}\right)}{\partial\theta\partial\theta'}\right)^{-1}}_{=O_p(1)}\cdot\left(\underbrace{\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{(u)}_{nT,c}\left(\theta_0\right)}{\partial\theta}}_{=O_p(1)}-\underbrace{\Delta_{1,nT}}_{=O\left(\sqrt{\frac{n}{T}}\right)+O\left(\sqrt{\frac{n}{T^3}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)}-\underbrace{\Delta_{2,nT}}_{=O\left(\sqrt{\frac{T}{n}}\right)}\right),$$

which implies $\hat{\theta}_{ml,nT}-\theta_0=O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T},\frac{1}{n}\right)\right)$. Since (i) $\Sigma_{\theta_0}=\lim_{T\to\infty}\Sigma_{\theta_0,nT}$ exists and is nonsingular, (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, (iii) $\Delta_{2,nT}=\sqrt{\frac{T}{n}}a_{n,2}(\theta_0)$ and

(iv) $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{nT}^{(u)}(\theta_0)}{\partial\theta} \xrightarrow{d} N\left(0, \Omega_{\theta_0}\right)$ by Lemma 2.3, we obtain the desired asymptotic distribution result for $\sqrt{nT}\left(\hat{\theta}_{ml,nT} - \theta_0\right)$.

Proof of Corollary 4.3 By Theorem 4.2, we have

$$\sqrt{nT}\left(\hat{\theta}_{ml,nT} - \theta_0\right) + \sqrt{\frac{n}{T}}\Sigma_{\theta_0,nT}^{-1}a_{n,1}\left(\theta_0\right) + \sqrt{\frac{T}{n}}\Sigma_{\theta_0,nT}^{-1}a_{n,2}\left(\theta_0\right) + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{T}{n^3}}, \frac{1}{\sqrt{T}}\right)\right)$$
$$\xrightarrow{d} N\left(0, \Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1}\right).$$

Since $\hat{\theta}_{ml,nT}^{c} = \hat{\theta}_{ml,nT} - \frac{1}{T}\left[-\Sigma_{\theta,nT}^{-1}a_{n,1}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \frac{1}{n}\left[-\Sigma_{\theta,nT}^{-1}a_{n,2}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}}$, $\sqrt{nT}\left(\hat{\theta}_{ml,nT}^{c} - \theta_0\right) \xrightarrow{d} N\left(0, \Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1}\right)$ if

$$\sqrt{\frac{n}{T}}\left(\left[\Sigma_{\theta,nT}^{-1}a_{n,1}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \Sigma_{\theta_0,nT}^{-1}a_{n,1}(\theta_0)\right) \xrightarrow{p} 0 \tag{19}$$

and

$$\sqrt{\frac{T}{n}}\left(\left[\Sigma_{\theta,nT}^{-1}a_{n,2}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \Sigma_{\theta_0,nT}^{-1}a_{n,2}(\theta_0)\right) \xrightarrow{p} 0. \tag{20}$$

Assume $\frac{n}{T^3} \to 0$, $\frac{T}{n^3} \to 0$, (19) and (20) can hold. First, consider (19). From the proof of (i), $\Sigma_{\theta,nT}^{-1}|_{\theta=\hat{\theta}_{ml,nT}} = \Sigma_{\theta_0,nT}^{-1} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \frac{1}{n}\right)\right)$. Hence,

$$\begin{aligned}
&\sqrt{\frac{n}{T}}\left(\left[\Sigma_{\theta,nT}^{-1}a_{n,1}(\theta)\right]|_{\theta=\hat{\theta}_{ml,nT}} - \Sigma_{\theta_0,nT}^{-1}a_{n,1}(\theta_0)\right) \qquad (21)\\
=\;& \sqrt{\frac{n}{T}}\left(\Sigma_{\theta,nT}^{-1}|_{\theta=\hat{\theta}_{ml,nT}} - \Sigma_{\theta_0,nT}^{-1}\right)a_{n,1}(\hat{\theta}_{ml,nT}) + \sqrt{\frac{n}{T}}\Sigma_{\theta_0,nT}^{-1}\left(a_{n,1}(\hat{\theta}_{ml,nT}) - a_{n,1}(\theta_0)\right)\\
=\;& \sqrt{\frac{n}{T}}\cdot O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \frac{1}{n}\right)\right)a_{n,1}(\hat{\theta}_{ml,nT}) + \sqrt{\frac{n}{T}}\Sigma_{\theta_0,nT}^{-1}\left(a_{n,1}(\hat{\theta}_{ml,nT}) - a_{n,1}(\theta_0)\right).
\end{aligned}$$

Since $\hat{\theta}_{ml,nT} - \theta_0 = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \frac{1}{n}\right)\right)$ and $a_{n,1}(\theta_0) = O(1)$, (19) will be valid if $\frac{\partial}{\partial\theta'}a_{n,1}(\bar{\theta}_{nT})$ is stochastically bounded, where $\bar{\theta}_{nT}$ lies between $\hat{\theta}_{ml,nT}$ and θ_0. For this, we will show $\frac{\partial}{\partial\theta'}a_{n,1}(\theta)$ is uniformly bounded in a neighborhood of θ_0. Since $a_{n,1}(\theta)$ includes W_n, $R_n^{-1}(\theta_1)$, $R_{n\lambda}(\theta_1)$, $R_{n\gamma}(\theta_1)$, $R_{n\rho}(\theta_1)$, $A_n(\theta_1)$, $\sum_{h=0}^{\infty}A_n^h(\theta_1)$, $\frac{1}{2\sigma_\epsilon^2}$, $\sum_{h=0}^{\infty}A_{k,n}^h(\gamma_k,\rho_k)$ and $\frac{1}{2\sigma_{V,k}^2}$ for $k = 1,\cdots,K$, therefore, $\frac{\partial}{\partial\theta'}a_{n,1}(\theta)$ consists of W_n,

$R_n^{-1}(\theta_1)$, $R_{n,\lambda\lambda}(\theta_1)$, $R_{n,\lambda\gamma}(\theta_1)$, $R_{n,\lambda\rho}(\theta_1)$, $R_{n,\gamma\gamma}(\theta_1)$, $R_{n,\gamma\rho}(\theta_1)$, $R_{n,\rho\rho}(\theta_1)$, $-\frac{1}{2\sigma_\epsilon^4}$

$$\begin{aligned}
\frac{\partial}{\partial\lambda}R_n^{-1}(\theta_1) &= -R_n^{-1}(\theta_1)\,R_{n\lambda}(\theta_1)\,R_n^{-1}(\theta_1)\,,\ \frac{\partial}{\partial\gamma}R_n^{-1}(\theta_1) = -R_n^{-1}(\theta_1)\,R_{n\gamma}(\theta_1)\,R_n^{-1}(\theta_1)\,,\\
\frac{\partial}{\partial\rho}R_n^{-1}(\theta_1) &= -R_n^{-1}(\theta_1)\,R_{n\rho}(\theta_1)\,R_n^{-1}(\theta_1)\\
\frac{\partial}{\partial\lambda}\sum_{h=0}^{\infty}A_n^h(\theta_1) &= -\sum_{h=1}^{\infty}hA_n^{h-1}(\theta_1)R_{n\lambda}(\theta_1)R_n^{-1}(\theta_1)\,A_n(\theta_1),\\
\frac{\partial}{\partial\gamma}\sum_{h=0}^{\infty}A_n^h(\theta_1) &= \sum_{h=1}^{\infty}hA_n^{h-1}(\theta_1)\left(-R_{n\gamma}(\theta_1)R_n^{-1}(\theta_1)\,A_n(\theta_1)+R_n^{-1}(\theta_1)\right),\\
\frac{\partial}{\partial\rho}\sum_{h=0}^{\infty}A_n^h(\theta_1) &= \sum_{h=1}^{\infty}hA_n^{h-1}(\theta_1)\left(-R_{n\rho}(\theta_1)R_n^{-1}(\theta_1)\,A_n(\theta_1)+R_n^{-1}(\theta_1)\,W_n\right),\\
\frac{\partial}{\partial\gamma_k}\sum_{h=0}^{\infty}A_{k,n}^h(\gamma_k,\rho_k) &= \sum_{h=1}^{\infty}hA_{k,n}^{h-1}(\gamma_k,\rho_k),\\
\frac{\partial}{\partial\rho_k}\sum_{h=0}^{\infty}A_{k,n}^h(\gamma_k,\rho_k) &= \sum_{h=1}^{\infty}hA_{k,n}^{h-1}(\gamma_k,\rho_k)W_n,
\end{aligned}$$

and $-\frac{1}{2\sigma_{V,k}^4}$ for all $k=1,\cdots,K$. By Assumptions 4.3 and 4.5, all these components above are uniformly bounded. It implies that $\frac{\partial}{\partial\theta'}a_{n,1}(\theta)$ is uniformly bounded in a neighborhood of θ_0. Lastly, (20) can be shown. Note that

$$\begin{aligned}
&\sqrt{\frac{T}{n}}\left(\left[\Sigma_{\theta,nT}^{-1}a_{n,2}(\theta)\right]|_{\theta=\hat\theta_{ml,nT}}-\Sigma_{\theta_0,nT}^{-1}a_{n,2}(\theta_0)\right) \qquad (22)\\
=\ &\sqrt{\frac{T}{n}}\left(\Sigma_{\theta,nT}^{-1}|_{\theta=\hat\theta_{ml,nT}}-\Sigma_{\theta_0,nT}^{-1}\right)a_{n,2}(\hat\theta_{ml,nT})+\sqrt{\frac{T}{n}}\Sigma_{\theta_0,nT}^{-1}\left(a_{n,2}(\hat\theta_{ml,nT})-a_{n,2}(\theta_0)\right).
\end{aligned}$$

By the same logic of showing (19), we have to show $\frac{\partial}{\partial\theta'}a_{n,2}(\theta)$ is uniformly bounded in a neighborhood of θ_0. However, this can be directly verified because $a_{n,2}(\theta)$ just contains $R_n^{-1}(\theta_1)$, $R_{n\lambda}(\theta_1)$, $R_{n\gamma}(\theta_1)$, $R_{n\rho}(\theta_1)$, $\frac{1}{2\sigma_\epsilon^2}$, $\frac{1}{2\sigma_{V,k}^2}$ for $k=1,\cdots,K$. As $\bar\theta_{nT}\xrightarrow{p}\theta_0$, all of elements in $\frac{\partial}{\partial\theta'}a_{n,1}(\bar\theta_{nT})$ and $\frac{\partial}{\partial\theta'}a_{n,2}(\bar\theta_{nT})$ are $O_p(1)$. ■

Proof of Theorem 4.4 (i). Since $Y_{nt}=R_n^{-1}((\gamma_0 I_n+\rho_0 W_n)\,Y_{n,t-1}+\sum_{k=1}^K B_{X,k,n}X_{nt,k}+\mathbf{c}_{n0}+\alpha_{t0}l_n+\mathcal{E}_{nt})$,

$$\begin{aligned}
\hat{\mathbf{c}}_{n,ml}(\theta) &= \frac{1}{T}\sum_{t=1}^T(R_n(\theta_1)Y_{nt}-(\gamma I_n+\rho W_n)\,Y_{n,t-1}-\sum_{k=1}^K B_{X,k,n}(\theta)\,X_{nt,k})\\
&= \frac{1}{T}\sum_{t=1}^T\left[\begin{array}{c}\left(R_n(\theta_1)R_n^{-1}(\gamma_0 I_n+\rho_0 W_n)-(\gamma I_n+\rho W_n)\right)Y_{n,t-1}\\ +\sum_{k=1}^K\left(R_n(\theta_1)R_n^{-1}B_{X,k,n}-B_{X,k,n}(\theta)\right)X_{nt,k}+R_n(\theta_1)R_n^{-1}(\mathbf{c}_{n0}+\alpha_{t0}l_n+\mathcal{E}_{nt})\end{array}\right].
\end{aligned}$$

Hence, by the mean value theorem, for each $\theta\in\Theta$,

$\hat{\mathbf{c}}_{n,ml}(\theta)=\frac{1}{T}\sum_{t=1}^T(\lambda_0-\lambda)\cdot\left[\left(-R_{n\lambda}(\bar\theta_1)\,R_n^{-1}\right)((\gamma_0 I_n+\rho_0 W_n)\,Y_{n,t-1}+\mathbf{c}_{n0}+\mathcal{E}_{nt})+\sum_{k=1}^K C_{n,k}^{\lambda}(\bar\theta)\,X_{nt,k}\right]$

$+\frac{1}{T}\sum_{t=1}^{T}\left(\gamma_0-\gamma\right)\cdot\left[\left(-R_{n\gamma}\left(\bar{\theta}_1\right)R_n^{-1}\right)\left(\left(\gamma_0 I_n+\rho_0 W_n\right)Y_{n,t-1}+\mathbf{c}_{n0}+\mathcal{E}_{nt}\right)+Y_{n,t-1}+\sum_{k=1}^{K}C_{n,k}^{\gamma}\left(\bar{\theta}\right)X_{nt,k}\right]$
$+\frac{1}{T}\sum_{t=1}^{T}\left(\rho_0-\rho\right)\cdot\left[\left(-R_{n\rho}\left(\bar{\theta}_1\right)R_n^{-1}\right)\left(\left(\gamma_0 I_n+\rho_0 W_n\right)Y_{n,t-1}+\mathbf{c}_{n0}+\mathcal{E}_{nt}\right)+W_nY_{n,t-1}+\sum_{k=1}^{K}C_{n,k}^{\rho}\left(\bar{\theta}\right)X_{nt,k}\right]$
$+\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\beta_{1,k,0}-\beta_{1,k}\right)\cdot C_{n,k}^{\beta_1,k}\left(\bar{\theta}\right)X_{nt,k}+\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\beta_{2,k,0}-\beta_{2,k}\right)\cdot C_{n,k}^{\beta_2,k}\left(\bar{\theta}\right)X_{nt,k}$
$+\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\gamma_{k,0}-\gamma_k\right)\cdot C_{n,k}^{\gamma_X,k}\left(\bar{\theta}\right)X_{nt,k}+\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\rho_{k,0}-\rho_k\right)\cdot C_{n,k}^{\rho_X,k}\left(\bar{\theta}\right)X_{nt,k}$
$+\mathbf{c}_{n0}+\frac{1}{T}\sum_{t=1}^{T}\mathcal{E}_{nt}$ where $\bar{\theta}$ lies between θ and θ_0 and because $\sum_{t=1}^{T}\alpha_{t0}=0$.

Hence, we obtain

$$\begin{aligned}
&\hat{c}_{i,ml}(\hat{\theta}_{ml,nT})-c_{i,0}\\
=\;&\frac{1}{T}\sum_{t=1}^{T}\left(\lambda_0-\hat{\lambda}_{ml,nT}\right)\cdot\left[\left(-R_{n\lambda}\left(\bar{\theta}_{1,nT}\right)R_n^{-1}\right)\left(\left(\gamma_0 I_n+\rho_0 W_n\right)Y_{n,t-1}+\mathbf{c}_{n0}+\mathcal{E}_{nt}\right)+\sum_{k=1}^{K}C_{n,k}^{\lambda}\left(\bar{\theta}_{nT}\right)X_{nt,k}\right]_i\\
&+\frac{1}{T}\sum_{t=1}^{T}\left(\gamma_0-\hat{\gamma}_{ml,nT}\right)\cdot\left[\begin{array}{c}\left(-R_{n\gamma}\left(\bar{\theta}_{1,nT}\right)R_n^{-1}\right)\left(\left(\gamma_0 I_n+\rho_0 W_n\right)Y_{n,t-1}+\mathbf{c}_{n0}+\mathcal{E}_{nt}\right)\\+Y_{n,t-1}+\sum_{k=1}^{K}C_{n,k}^{\gamma}\left(\bar{\theta}_{nT}\right)X_{nt,k}\end{array}\right]_i\\
&+\frac{1}{T}\sum_{t=1}^{T}\left(\rho_0-\hat{\rho}_{ml,nT}\right)\cdot\left[\begin{array}{c}\left(-R_{n\rho}\left(\bar{\theta}_{1,nT}\right)R_n^{-1}\right)\left(\left(\gamma_0 I_n+\rho_0 W_n\right)Y_{n,t-1}+\mathbf{c}_{n0}+\mathcal{E}_{nt}\right)\\+W_nY_{n,t-1}+\sum_{k=1}^{K}C_{n,k}^{\rho}\left(\bar{\theta}_{nT}\right)X_{nt,k}\end{array}\right]_i\\
&+\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\beta_{1,k,0}-\hat{\beta}_{1,k,ml,nT}\right)\cdot\left[C_{n,k}^{\beta_1,k}\left(\bar{\theta}_{nT}\right)X_{nt,k}\right]_i+\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\beta_{2,k,0}-\hat{\beta}_{2,k,ml,nT}\right)\cdot\left[C_{n,k}^{\beta_2,k}\left(\bar{\theta}_{nT}\right)X_{nt,k}\right]_i\\
&+\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\gamma_{k,0}-\hat{\gamma}_{k,ml,nT}\right)\cdot\left[C_{n,k}^{\gamma_X,k}\left(\bar{\theta}_{nT}\right)X_{nt,k}\right]_i+\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\rho_{k,0}-\hat{\rho}_{k,ml,nT}\right)\cdot\left[C_{n,k}^{\rho_X,k}\left(\bar{\theta}_{nT}\right)X_{nt,k}\right]_i\\
&+\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}
\end{aligned}$$

where $\bar{\theta}_{nT}$ lies between $\hat{\theta}_{ml,nT}$ and θ_0. Since $\hat{\theta}_{ml,nT}-\theta_0=O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T},\frac{1}{n}\right)\right)$ from Theorem 4.1, the dominant term of $c_{i,ml}(\hat{\theta}_{ml,nT})-c_{i,0}$ is $\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}$ for each i and the remainder terms except for $\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}$ is $O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T},\frac{1}{n}\right)\right)$. Therefore, for each i $\sqrt{T}\left(\hat{c}_{i,ml}(\hat{\theta}_{ml,nT})-c_{i,0}\right)\xrightarrow{d}N(0,\sigma_{\epsilon,0}^2)$ and $\hat{c}_{i,ml}(\hat{\theta}_{ml,nT})$'s are asymptotically independent with each other.

Consider (ii). Using $\tilde{\alpha}_{t0} = \alpha_{t0}$ from $\sum_{t=1}^{T} \alpha_{t0} = 0$, note that

$$
\begin{aligned}
& \hat{\alpha}_{t,ml}(\hat{\theta}_{ml,nT}) - \alpha_{t0} \\
= & \left(\lambda_0 - \hat{\lambda}_{ml,nT}\right) \cdot \frac{1}{n} l_n' \left[\left(-R_{n\lambda}\left(\bar{\theta}_{1,nT}\right) R_n^{-1}\right) \left(\left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n,t-1} + \alpha_{t0} + \mathcal{E}_{nt}\right) + \sum_{k=1}^{K} C_{n,k}^{\lambda}\left(\bar{\theta}_{nT}\right) X_{nt,k} \right] \\
& + \left(\gamma_0 - \hat{\gamma}_{ml,nT}\right) \cdot \frac{1}{n} l_n' \left[\left(-R_{n\gamma}\left(\bar{\theta}_{1,nT}\right) R_n^{-1}\right) \left(\left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n,t-1} + \alpha_{t0} + \mathcal{E}_{nt}\right) + Y_{n,t-1} + \sum_{k=1}^{K} C_{n,k}^{\gamma}\left(\bar{\theta}_{nT}\right) X_{nt,k} \right] \\
& + \left(\rho_0 - \hat{\rho}_{ml,nT}\right) \cdot \frac{1}{n} l_n' \left[\begin{array}{c} \left(-R_{n\rho}\left(\bar{\theta}_{1,nT}\right) R_n^{-1}\right) \left(\left(\gamma_0 I_n + \rho_0 W_n\right) Y_{n,t-1} + \alpha_{t0} + \mathcal{E}_{nt}\right) \\ + W_n Y_{n,t-1} + \sum_{k=1}^{K} C_{n,k}^{\rho}\left(\bar{\theta}_{nT}\right) X_{nt,k} \end{array} \right] \\
& + \sum_{k=1}^{K} \left(\beta_{1,k,0} - \hat{\beta}_{1,k,ml,nT}\right) \cdot \frac{1}{n} l_n' \left[C_{n,k}^{\beta_1,k}\left(\bar{\theta}_{nT}\right) X_{nt,k} \right] + \sum_{k=1}^{K} \left(\beta_{2,k,0} - \hat{\beta}_{2,k,ml,nT}\right) \cdot \frac{1}{n} l_n' \left[C_{n,k}^{\beta_2,k}\left(\bar{\theta}_{nT}\right) X_{nt,k} \right] \\
& + \sum_{k=1}^{K} \left(\gamma_{k,0} - \hat{\gamma}_{k,ml,nT}\right) \cdot \frac{1}{n} l_n' \left[C_{n,k}^{\gamma_X,k}\left(\bar{\theta}_{nT}\right) X_{nt,k} \right] + \sum_{k=1}^{K} \left(\rho_{k,0} - \hat{\rho}_{k,ml,nT}\right) \cdot \frac{1}{n} l_n' \left[C_{n,k}^{\rho_X,k}\left(\bar{\theta}_{nT}\right) X_{nt,k} \right] + \frac{1}{n} l_n' \mathcal{E}_{nt}
\end{aligned}
$$

where $\bar{\theta}_{nT}$ lies between $\hat{\theta}_{ml,nT}$ and θ_0. Since $\hat{\theta}_{ml,nT} - \theta_0 = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \frac{1}{n}\right)\right)$, the dominant term of $\sqrt{n}\left(\hat{\alpha}_{t,ml}(\hat{\theta}_{ml,nT}) - \alpha_{t0}\right)$ is $\frac{1}{\sqrt{n}} l_n' \mathcal{E}_{nt}$. This yields (i) $\sqrt{n}\left(\hat{\alpha}_{t,ml} - \alpha_{t0}\right) \xrightarrow{d} N(0, \sigma_{\epsilon,0}^2)$ and (ii) the estimates $\hat{\alpha}_{t,ml}$'s for $t = 1, \cdots, T$ are asymptotically independent with each other.

Verifying (iii) can be done by applying $\hat{\theta}_{ml,nT}^{c} - \theta_0 = O_p\left(\frac{1}{\sqrt{nT}}\right)$ if $\frac{n}{T^3}$ and $\frac{T}{n^3} \to 0$. ■

2.7 Calculation of $R_n\left(\theta_1\right)$ and $\ln\left|R_n\left(\theta_1\right)\right|$

For the inner loop (evaluation of spatial-time filter $R_n\left(\theta_1\right)$)

Note that $R_n\left(\theta_1\right)$ is the spatial-time filter of our model. Since a component of the inner loop is to evaluate $R_n\left(\theta_1\right)$, we need a numerical approximation method. If we take a high order approximation, a computation cost of getting $R_n\left(\theta_1\right)$ increases explosively when n is large. In Table A.1, we provide the performance of iterations and several approximations for $R_n\left(\theta_1\right)$. For this experiment, we use (row-normalized) rook and queen matrices for W_n and fix $n = 49$, $\delta = 0.95$ and $\gamma_0 = 0.4$. We consider four combinations of (λ_0, ρ_0): $(0.2, 0.2)$, $(0.2, -0.2)$, $(-0.2, 0.2)$ and $(-0.2, -0.2)$. If $|\lambda_0| + \gamma_0 + |\rho_0|$ is small, convergence speed will be rapid. To measure performance of iterations, we use the relative norm, $\left\| R_n^{(j+1)}(\theta_{1,0}) - R_n^{(j)}(\theta_{1,0}) \right\|_s / \left\| R_n^{(j)}(\theta_{1,0}) \right\|_s$ where $\|\cdot\|_s$ denotes the spectral norm and $R_n^{(1)}(\theta_{1,0})$ is set to be $I_n - \lambda_0 W_n$. By Table A.1, we observe that numerical errors decrease when taking more iterations for all cases. By taking the second iteration for approximation, we observe the dramatic reduction of numerical errors. If signs of λ_0 and ρ_0 are different, it seems convergence speed becomes rapid.

For the outer loop (parameter searching)

In evaluating $\ln L_{nT,c}(\theta)$, a demanding part is to calculate $\ln|R_n(\theta_1)|$ at each $\theta_1 \in \Theta_1$. Even though we use conventional SDPD models, evaluating the log-determinant is computationally burdensome when (i) n is very large or (ii) there are multiple spatial weighting matrices, or (iii) we have a nonlinear specification (i.e., $\ln|I_n - W_n(\lambda)|$ where $W_n(\lambda)$ is a nonlinear function of λ). Calculating $\ln|R_n(\theta_1)|$ is more demanding because (i) it is a highly nonlinear function of λ, γ, and ρ, and (ii) it also contains infinite-order polynomials of W_n. Hence, developing technology for calculating $\ln|R_n(\theta_1)|$ might be meaningful because (i) it can reduce computation costs and (ii) might suggest an alternative way under large n.

One approach is to change $\ln|R_n(\theta_1)|$ by a function of trace. Consider the form of $\ln|A_n|$ where $A_n = B_n + C_n = B_n(I_n + D_n)$ where $D_n = B_n^{-1}C_n$. Then, we have

$$|A_n| = |B_n| \cdot \exp\left(tr\left(\ln\left(I_n + D_n\right)\right)\right),$$

where $\ln(I_n + D_n)$ is the matrix logarithm of $I_n + D_n$.[3] If $\|D_n\| < 1$ where $\|\cdot\|$ is a proper matrix norm, $\ln(I_n + D_n)$ can be represented by $\ln(I_n + D_n) = -\sum_{j=1}^{\infty}\frac{(-D_n)^j}{j}$ and we have $\ln|A_n| = \ln|B_n| - \sum_{j=1}^{\infty}\frac{tr\left((-D_n)^j\right)}{j}$. Then, a feasible approximation $\ln|A_n| \simeq \ln|B_n| - \sum_{j=1}^{J}\frac{tr\left((-D_n)^j\right)}{j}$ can be employed in practice where J is a chosen positive integer.

For example, consider the approximation of $\ln|I_n - \lambda W_n|$ where $|\lambda| < 1$ with a row normalized W_n. By using $|I_n - \lambda W_n| = \exp\left(tr\left(\ln\left(I_n - \lambda W_n\right)\right)\right)$ and $\ln(I_n - \lambda W_n) = -\sum_{j=1}^{\infty}\frac{\lambda^j W_n^j}{j}$, we have $\ln|I_n - \lambda W_n| = -\sum_{j=1}^{\infty}\frac{\lambda^j tr\left(W_n^j\right)}{j}$. The details can be found in LeSage and Pace (2009). We can apply the same strategy to our model. Because we know $R_n(\theta_1) = (1+\delta\gamma)I_n - \lambda W_n - \delta D_{n,1}(\theta_1)(\gamma I_n + \rho W_n)$, the decomposition

$$|R_n(\theta_1)| = |(1+\delta\gamma)I_n| \cdot \exp\left(tr\left(\ln\left(I_n - F_n(\theta_1)\right)\right)\right)$$

where $F_n(\theta_1) = \frac{\lambda}{1+\delta\gamma}W_n + \frac{\delta}{1+\delta\gamma}D_{n,1}(\theta_1)(\gamma I_n + \rho W_n)$. If $\|F_n(\theta_1)\| < 1$, we obtain $\ln(I_n - F_n(\theta_1)) = -\sum_{j=1}^{\infty}\frac{F_n^j(\theta_1)}{j}$. It implies $\ln|R_n(\theta_1)| = n \cdot \ln(1+\delta\gamma) - \sum_{j=1}^{\infty}\frac{tr\left(F_n^j(\theta_1)\right)}{j}$. In Table A.2, we present the performance of feasible approximations,

$$\left(\ln|R_n(\theta_1)|\right)_{(J)} = n \cdot \ln(1+\delta\gamma) - \sum_{j=1}^{J}\frac{tr\left(F_n^j(\theta_1)\right)}{j} \tag{23}$$

where $J = 1, 2, 3, 4$ and 5. We evaluate the performance of those approximations by considering $\sup_{\theta_1\in\Theta_1}\left|\left(\frac{1}{n}\ln|R_n(\theta_1)|\right)_{(J)} - \frac{1}{n}\ln|R_n(\theta_1)|\right|$ for $J = 1, \cdots, 5$ where Θ_1 is set to be $[-0.2, 0.2] \times [0, 0.4] \times$

[3] See LeSage and Pace (2009), pp. 96-97.

$[-0.2, 0.2]$, $n = 49$, W_n is considered as rook and queen matrix and $\ln |R_n(\theta_1)|$ is directly calculated through Matlab. We observe the approximations will be finer corresponding to increasing the order J. This result generally holds for large n.

3 Some additional simulations

In this section, we introduce more simulation results to support our empirical analyses. Via Table A.3, first, we report simulation results with the same setting in the main draft but $\delta = 0.5$. The next issue is capturing the true time-discounting factor (denoted by δ_0). In recent structural estimation analyses, time-discounting factor δ_0 is usually a primitive parameter. For dynamic discrete choice models, Komarova et al. (2017) argue that identifying δ_0 is possible under the limited model specification (linear-in-parameter assumption). Under a general model specification, we have difficulty in identifying δ_0 since a log-likelihood function might be flat around δ_0. Instead of estimating δ_0, hence, its value might often be selected by economic reasonings (e.g., long-run interest rates or capital-output ratio) in the empirical macroeconomics literature.. Since there is no general guidance in selecting δ_0 in the statistical aspect, we want to give some practical evidence to determine an appropriate (well fitted to data) time-discounting factor in a forward-looking SDPD model.

By Rothenberg (1971), identification under likelihood theory is based on the information inequality:

$$E\left(\ln L_{nT,c}(\theta, \delta)\right) \leq E\left(\ln L_{nT,c}(\theta_0, \delta_0)\right) \text{ for any } \theta \in \Theta \text{ and } \delta \in [0, 1).$$

Identification uniqueness is achieved if (θ_0, δ_0) is the unique maximizer of $E\left(\ln L_{nT,c}(\theta, \delta)\right)$: by the strict information inequality,

$$E\left(\ln L_{nT,c}(\theta, \delta)\right) < E\left(\ln L_{nT,c}(\theta_0, \delta_0)\right) \text{ for all } (\theta, \delta) \neq (\theta_0, \delta_0). \tag{24}$$

By doing simulations, we evaluate four likelihood measures for different values of δ: (i) average empirical joint log-likelihood ($\overline{E \ln L}$), (ii) average empirical partial log-likelihood ($\overline{E \ln L_1}$), (iii) Akaike information criterion (AIC), and (iv) Bayesian information criterion (BIC).[4] The four measures are based on the (concentrated)

[4]The formulas of the four measures are

$$\overline{E \ln L}(\delta) = \frac{1}{I}\sum_{l=1}^{I} \ln L_{nT,c}\left(\left[\hat{\theta}_{nT,ml}\right]_l, \delta\right), \ \overline{E \ln L_1}(\delta) = \frac{1}{I}\sum_{l=1}^{I} \ln L^P_{nT,c}\left(\left[\hat{\theta}_{nT,ml}\right]_l, \delta\right),$$

$$AIC(\delta) = \frac{1}{I}\sum_{l=1}^{I}\left(-2\ln L_{nT,c}\left(\left[\hat{\theta}_{nT,ml}\right]_l, \delta\right) + 2\cdot(4+5K)\right), \text{ and } BIC(\delta) = \frac{1}{I}\sum_{l=1}^{I}\left(-2\ln L_{nT,c}\left(\left[\hat{\theta}_{nT,ml}\right]_l, \delta\right) + \ln(nT)\cdot(4+5K)\right)$$

where I denotes the number of repetitions (set to be 400), $\ln L^P_{nT,c}(\cdot)$ represents the log-likelihood function only relevant to the main SAR equation, and $4 + 5K$ denotes the model's dimension.

log-likelihood, so our purpose is to verify the strict information inequality (24) via simulations.

We consider the same DGP process in Monte Carlo simulations of the main text. Throughout the experiment, we fix $\lambda_0 = 0.2$, $\gamma_0 = 0.4$, $\rho_0 = 0.2$, and all remaining parameters are set to be the same in the main draft. We consider the six scenarios and evaluate the suggested measures at various δ's (i.e., misspecified δ's).[5] The first case represents a forward-looking model ($\delta_0 = 0.95$), small number of observations ($n = 49$ and $T = 10$), and a low signal-to-noise ratio (SNR) ($K = 1$); The second model is generated by $\delta_0 = 0.95$, $(n, T) = (49, 10)$, and a high SNR ($K = 2$); the third scenario is generated by $\delta_0 = 0.95$, relatively large number of observations ($n = 81$ and $T = 30$), and $K = 1$; the fourth model sets $\delta_0 = 0.95$, $(n, T) = (81, 30)$ and $K = 2$; the fifth model is a myopic model, $(n, T) = (81, 30)$ and $K = 1$; The last one is generated by $\delta_0 = 0$ with $(n, T) = (81, 30)$ and $K = 2$. First, we evaluate and compare the likelihood measures for different δ's. This investigation gives validity of employing the suggested likelihood measures when we select a proper δ among possible candidate values. For all cases, we observe that the four measures are equivalent in the sense of selecting δ_0. Second, we try to evaluate effects of misspecified δ on estimating the main structural parameters λ_0, γ_0, and ρ_0. For that purpose, the bias-corrected QMLE is considered and its RMSEs are evaluated across various δ's. Simulation results are reported via Table A.4 and Figure A.1.

- Model 1: $\delta_0 = 0.95$, $K = 1$ and $(n, T) = (49, 10)$

The $\overline{E \ln L}$ indicates that $\delta = 0.9$ is the best model. However, some irregular zig-zag patterns are observed in $\overline{E \ln L}$. If we compare the cases of $\delta = 0$ and 0.95, the $\overline{E \ln L}$ chooses the myopic model. It implies that choosing δ_0 by the likelihood measures may not work in this case. For λ_0, the RMSE takes a U-shape in δ and is minimized at $\delta = 0.5$. For γ_0, the RMSE is minimized at $\delta = 0.9$ while the case of $\delta = 0.925$ shows the best performance for ρ_0. However, it is hard to observe a regular pattern of effects.

- Model 2: $\delta_0 = 0.95$, $K = 2$ and $(n, T) = (49, 10)$

In the sense of $\overline{E \ln L}$, the model with $\delta = 0.99$ is the best. We observe $\overline{E \ln L}$ tends to increase from $\delta = 0$ to $\delta = 0.99$. Around the true value $\delta_0 = 0.95$, however, some irregular behaviors of $\overline{E \ln L}$ are observed. It means we can distinguish between the two models, (i) myopic and (ii) forward-looking models, while the true δ_0 is difficult to be identified via $\overline{E \ln L}$. The behaviors of RMSEs are similar to those of Model 1. However, irregular patterns disappear relative to the Model 1's case.

[5] Since we consider Durbin regressors, the exact number of exogenous variables is 4 if $K = 2$. For this model, $\beta_{1,2,0} = \beta_{2,2,0}$ are selected as the true values.

- Model 3: $\delta_0 = 0.95$, $K = 1$ and $(n, T) = (81, 30)$

The third model describes that many observations are available relative to the first case in both time and space dimensions. Under large finite samples, the likelihood measures can distinguish the true forward-looking model from the myopic one: $\overline{E \ln L}(0) = -1,799.54 < -1,798.69 = \overline{E \ln L}(0.95)$. However, we do not observe a specific relationship between $\overline{E \ln L}(\delta)$ and δ. Also, the likelihood measures show almost similar values around the true $\delta_0 = 0.95$ (if $\delta = 0.975$ and 0.99). In this case, it is hard to choose a correct time-discounting factor among economically reasonable δ's (e.g., $0.95 \leq \delta_0 < 1$). The RMSEs of parameters λ_0, and γ_0 are respectively minimized at the true value $\delta_0 = 0.95$. The RMSE for ρ_0 is minimized at $\delta = 0.7$. However, the RMSEs take similar values for $0.7 \leq \delta < 1$. Compared to Models 1 and 2, having more observations gives evidence of identifying δ_0 by the likelihood measures and good performance of the QMLEs around δ_0.

- Model 4: $\delta_0 = 0.95$, $K = 2$ and $(n, T) = (81, 30)$

Via Model 4, we perform an experiment on a high signal-to-noise ratio case by including (significant) exogenous regressors. Compared to Models 1, 2 and 3, the likelihood measures show smooth behaviors (no zig-zag pattern) and are optimized around the true δ_0. It means the more transparent relationship between $\overline{E \ln L}(\delta)$ and δ. Hence, we can conclude that the likelihood measures perform well in identifying δ_0 if we have sufficient observations with rich exogenous variables. On estimating λ_0, γ_0, and ρ_0, the RMSEs are minimized around the true value δ_0 except for the case of ρ_0.

- Model 5: $\delta_0 = 0$, $K = 1$ and $(n, T) = (81, 30)$

By Models 5 and 6, we consider identification of δ_0 and misspecification errors if the myopic model ($\delta_0 = 0$) is the true one. The $\overline{E \ln L}(\delta)$ is optimized at $\delta_0 = 0$ and becomes far from the true one if a large δ is selected (i.e., $\overline{E \ln L}(\delta)$ tends to be a decreasing function of δ). Even for the case $K = 1$ (relatively low signal), it seems that considering likelihood measures is good to identify δ_0 if the true model indicates myopic economic agents. For all parameters, the RMSEs are minimized at the true values and they increase corresponding to increasing δ. In case of the myopic model, therefore, identifying δ_0 can be done via the likelihood measures and the misspecification errors are consistent with econometric theory.

- Model 6: $\delta_0 = 0$, $K = 2$ and $(n, T) = (81, 30)$

Model 6 describes the similar DGP process to Model 5 but the relatively high signal. We observe similar behaviors of $\overline{E \ln L}(\delta)$ and RMSEs to those of Model 5.

4 Interpretations of our model and some empirical tools

This section introduces some practical and useful tools that are employed in our empirical analyses. Detailed forms and derivations of those tools and measures are included.

4.1 Cumulative effects

The cumulative effects of $x_{jt,k}$ on y_{it} can be calculated by $\frac{\partial y_{it}}{\partial x_{jt,k}} = \left[R_n^{-1}\left(I_n + \delta \mathbf{D}_{n,k} A_{k,n}\right)\left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n\right)\right]_{ij}$.

4.2 Empirical tool I: Rational forecasting

One useful property of employing dynamic models is providing a prediction of future economic variables. For forecasting horizons $h = 1, 2, \cdots$,

$$\begin{aligned} E_t\left(Y_{n,t+h}\right) &= A_n^{h+1} Y_{n,t-1} + \sum_{k=1}^{K}\sum_{g=0}^{h} A_n^{h-g} R_n^{-1} B_{X,k,n} A_{k,n}^{g} X_{nt,k} + \sum_{u=0}^{h} A_n^{u} R_n^{-1} \mathbf{c}_{n0} \\ &\quad + \sum_{k=1}^{K}\sum_{g=1}^{h}\sum_{f=0}^{g-1}\left(A_n^{h-g} R_n^{-1} B_{X,k,n} A_{k,n}^{f}\right)\mathbf{c}_{n,k,0} + A_n^{h} R_n^{-1}\left(\alpha_{t,0} l_n + \mathcal{E}_{nt}\right) \end{aligned}$$

By employing $\{E_t(Y_{n,t+h})\}_{h=1}^{\infty}$, we can forecast the expected agents' future actions on the MPE. In contrast to the forecasts from conventional dynamic panel data model (including traditional SDPD models), our forecasts reflect economic agents' forward-looking behaviors.

4.3 Empirical tool II: Impulse response functions

For $h = 1, 2, \cdots$, $\frac{\partial\left[E_t\left(Y_{n,t+h}\right)\right]_i}{\partial x_{jt,k}} = \left[\sum_{g=0}^{h} A_n^{h-g} R_n^{-1} B_{X,k,n} A_{k,n}^{g}\right]_{ij}$. Since Y_{nt} is linearly transformed by $X_{nt,k}$'s, the impulse response functions only depend on W_n and the parameters.

4.4 Empirical tool III: Welfare analyses

Last, we suggest some concepts to give policy implications. Note that $\boldsymbol{\eta}_n^{iv}$ is recovered by $\mathbf{c}_{n0}$:

$$\boldsymbol{\eta}_n^{iv} = \left(\left(I_n + \sum_{l=1}^{\infty} \delta^l \mathbf{D}_{n,l} \right) + \sum_{k=1}^{K} \sum_{l=1}^{\infty} \delta^l \mathbf{D}_{n,l} \left(\sum_{m=0}^{l-1} A_{k,n}^m \right) \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) h_{1,k} \right)^{-1} \mathbf{c}_{n0}.$$

Given a panel data set $\{Y_{nt}, X_{nt}\}_{t=0}^{T}$, and W_n with the identified parameters, λ_0, γ_0, ρ_0, $\beta_{1,0}$, $\beta_{2,0}$, $\gamma_{X,0}$, $\rho_{X,0}$, $\mathbf{c}_{n0}$ and $\{\alpha_{t0}\}_{t=1}^{T}$, we have the agent i's recovered per period payoff at time t,

$$\begin{aligned} \hat{u}_i \left(Y_{nt}, Y_{n,t-1}, X_{nt}, \boldsymbol{\eta}_n^{iv}, \alpha_{t0}; \theta_0 \right) &= e_i' \left(\sum_{k=1}^{K} \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) X_{nt,k} + \boldsymbol{\eta}_n^{iv} + \alpha_{t0} l_n \right) y_{it} \\ &\quad + \lambda_0 y_{it} w_{i.} Y_{nt} + \rho_0 y_{it} w_{i.} Y_{n,t-1} - \frac{\gamma_0}{2} (y_{it} - y_{i,t-1})^2 - \frac{1-\gamma_0}{2} y_{it}^2 \end{aligned} \tag{25}$$

for $i = 1, \cdots, n$ and $t = 1, \cdots, T$. By (25), we can also specify the (approximated) i's lifetime value :

$$\hat{V}_i \left(\left\{ Y_{nt}, Y_{n,t-1}, X_{nt}, \boldsymbol{\eta}_n^{iv}, \alpha_{t0} \right\}_{t=1}^{T}; \theta_0 \right) = \sum_{t=1}^{T} \delta^{t-1} \hat{u}_i \left(Y_{nt}, Y_{n,t-1}, X_{nt}, \boldsymbol{\eta}_n^{iv}, \alpha_{t0}; \theta_0 \right). \tag{26}$$

Using the same logic, a measure of social welfare is

$$\hat{\mathcal{W}} \left(\left\{ Y_{nt}, Y_{n,t-1}, X_{nt}, \boldsymbol{\eta}_n^{iv}, \alpha_{t0} \right\}_{t=1}^{T}; \theta_0 \right) = \sum_{i=1}^{n} \hat{V}_i \left(\left\{ Y_{nt}, Y_{n,t-1}, X_{nt}, \boldsymbol{\eta}_n^{iv}, \alpha_{t0} \right\}_{t=1}^{T}; \theta_0 \right). \tag{27}$$

Hence, we can evaluate (i) the i's immediate payoff $\hat{u}_i(\cdot)$, (ii) his/her lifetime value $\hat{V}_i(\cdot)$ and (iii) social welfare $\hat{\mathcal{W}}(\cdot)$.

In many applications, we would like to know effects of some exogenous characteristics on the i's lifetime payoff as well as social welfare. Given $\{Y_{nT}, X_{nT}\}$, we define

$$X_{n,T+h,k}^F \left(X_{nT,k}; \gamma_{k,0}, \rho_{k,0}, \mathbf{c}_{n,k,0}, \alpha_{t,k,0} \right) = A_{k,n}^h X_{nT,k} + \sum_{s=0}^{h-1} A_{k,n}^s \mathbf{c}_{n,k,0} + A_{k,n}^h \alpha_{T,0} l_n \tag{28}$$

for $h = 1, 2, \cdots$, and $k = 1, \cdots, K$ and

$$Y_{n,T+h}^F \left(Y_{nT}, X_{nT,1}, \cdots, X_{nT,K}; \theta_0 \right) = A_n^h Y_{nT} + \sum_{k=1}^{K} \sum_{s=1}^{h} A_n^{h-s} R_n^{-1} B_{X,k,n} X_{n,T+s,k}^F + \sum_{s=0}^{h-1} A_n^s R_n^{-1} \mathbf{c}_{n0} \tag{29}$$

for $h = 1, 2, \cdots$. Using the generated $\left\{ Y_{n,T+h}^F, X_{n,T+h}^F \right\}_{h=1}^{\infty}$, we re-define the measures (25), (26) and (27) by

$$\begin{aligned} & \tilde{u}_i^F \left(Y_{n,T+h}^F, Y_{n,T+h-1}^F, X_{n,T+h}^F, \boldsymbol{\eta}_n^{iv}, \alpha_{T,0}, \{Y_{nT}, X_{nT}\}; \theta_0 \right) \\ = \; & e_i' \left(\sum_{k=1}^{K} \left(\beta_{1,k,0} I_n + \beta_{2,k,0} W_n \right) X_{n,T+h}^F + \boldsymbol{\eta}_n^{iv} + \alpha_{T,0} l_n \right) e_i' Y_{n,T+h}^F \\ & + \lambda_0 e_i' Y_{n,T+h}^F w_{i.} Y_{n,T+h}^F + \rho_0 \left(e_i' Y_{n,T+h}^F \right) w_{i.} Y_{n,T+h-1}^F - \frac{\gamma_0}{2} (e_i' Y_{n,T+h}^F - e_i' Y_{n,T+h-1}^F)^2 - \frac{1-\gamma_0}{2} \left(e_i' Y_{n,T+h}^F \right)^2 \end{aligned} \tag{30}$$

where $Y^F_{nT} = Y_{nT}$, $i = 1, \cdots, n$ and $h = 1, 2, \cdots$,

$$\tilde{V}_i^F\left(\{Y_{nT}, X_{nT}\}; \theta_0\right) = \sum_{h=0}^{H} \delta^{t-1} \hat{u}_i \left(Y^F_{n,T+h}, Y^F_{n,T+h-1}, X^F_{n,T+h}, \boldsymbol{\eta}_n^{iv}, \alpha_{T0}, \{Y_{nT}, X_{nT}\}; \theta_0\right) \tag{31}$$

for $i = 1, \cdots, n$ and

$$\tilde{\mathcal{W}}^F\left(\{Y_{nT}, X_{nT}\}; \theta_0\right) = \sum_{i=1}^{n} \tilde{V}_i^F\left(\{Y_{nT}, X_{nT}\}; \theta_0\right) \tag{32}$$

for some sufficiently large $H \geq 1$. From (30), (31), and (32), we conduct a welfare analysis. For example, consider that a policy change $x_{j,T,k}$ for the individual j by Δ_x. Let $\ddot{X}_{nT,k} = \begin{bmatrix} x_{1,T,k} & \cdots & x_{j,T,k} + \Delta_x & \cdots & x_{n,T,k} \end{bmatrix}'$ where $\ddot{X}_{nT} = \left(X_{nT,1}, \cdots, \ddot{X}_{nT,k}, \cdots, X_{nT,K}\right)$. Using $\ddot{X}_{nT}$, we can evaluate $\tilde{V}_i^F\left(\left\{Y_{nT}, \ddot{X}_{nT}\right\}; \theta_0\right)$ and $\tilde{\mathcal{W}}^F\left(\left\{Y_{nT}, \ddot{X}_{nT}\right\}; \theta_0\right)$. Hence, the (expected) effects of that policy on i's lifetime value and social welfare can be specified by the following differences:

$$\Delta_{V_i} = \tilde{V}_i^F\left(\left\{Y_{nT}, \ddot{X}_{nT}\right\}; \theta_0\right) - \tilde{V}_i^F\left(\{Y_{nT}, X_{nT}\}; \theta_0\right) \text{ and } \Delta_{\mathcal{W}} = \tilde{\mathcal{W}}^F\left(\left\{Y_{nT}, \ddot{X}_{nT}\right\}; \theta_0\right) - \tilde{\mathcal{W}}^F\left(\{Y_{nT}, X_{nT}\}; \theta_0\right).$$

References

[1] Blackwell, D. (1965): "Discounted dynamic programming", Annals of Mathematical Statistics 36, 226-235.

[2] Fuente, A. (2000): "Mathematical methods and models for economists", Cambridge University Press, United Kingdom.

[3] 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.

[4] 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.

[5] LeSage, J., and R. K. Pace (2009): "Introduction to spatial econometrics", Boca Raton, FL, Chapman and Hall/CRC.

[6] Stokey, N. L., Lucas, R. E. and Prescott E. C. (1989): "Recursive methods in economic dynamics, Cambridge, MA: Havard University Press.

[7] 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.

Table A.1 : Performance of several approximations for $R_n(\theta_1)$

Case 1: $(\lambda_0, \gamma_0, \rho_0) = (0.2, 0.4, 0.2)$	A rook matrix	A queen matrix	Case 2: $(\lambda_0, \gamma_0, \rho_0) = (0.2, 0.4, -0.2)$	A rook matrix	A queen matrix
$\|R_n^{(2)}(\theta_1) - R_n^{(1)}(\theta_1)\| / \|R_n^{(1)}(\theta_1)\|$	0.2618	0.2498	$\|R_n^{(2)}(\theta_1) - R_n^{(1)}(\theta_1)\| / \|R_n^{(1)}(\theta_1)\|$	0.2390	0.2602
$\|R_n^{(3)}(\theta_1) - R_n^{(2)}(\theta_1)\| / \|R_n^{(2)}(\theta_1)\|$	0.0191	0.0222	$\|R_n^{(3)}(\theta_1) - R_n^{(2)}(\theta_1)\| / \|R_n^{(2)}(\theta_1)\|$	0.0198	0.0213
$\|R_n^{(4)}(\theta_1) - R_n^{(3)}(\theta_1)\| / \|R_n^{(3)}(\theta_1)\|$	0.0040	0.0046	$\|R_n^{(4)}(\theta_1) - R_n^{(3)}(\theta_1)\| / \|R_n^{(3)}(\theta_1)\|$	0.0020	0.0021
$\|R_n^{(5)}(\theta_1) - R_n^{(4)}(\theta_1)\| / \|R_n^{(4)}(\theta_1)\|$	0.0036	0.0040	$\|R_n^{(5)}(\theta_1) - R_n^{(4)}(\theta_1)\| / \|R_n^{(4)}(\theta_1)\|$	0.0005	0.0004
$\|R_n^{(6)}(\theta_1) - R_n^{(5)}(\theta_1)\| / \|R_n^{(5)}(\theta_1)\|$	0.0030	0.0025	$\|R_n^{(6)}(\theta_1) - R_n^{(5)}(\theta_1)\| / \|R_n^{(5)}(\theta_1)\|$	0.0003	0.0002
$\|R_n^{(7)}(\theta_1) - R_n^{(6)}(\theta_1)\| / \|R_n^{(6)}(\theta_1)\|$	0.0002	0.0001	$\|R_n^{(7)}(\theta_1) - R_n^{(6)}(\theta_1)\| / \|R_n^{(6)}(\theta_1)\|$	0.0000	0.0000
Case 3: $(\lambda_0, \gamma_0, \rho_0) = (-0.2, 0.4, 0.2)$	**A rook matrix**	**A queen matrix**	**Case 4: $(\lambda_0, \gamma_0, \rho_0) = (-0.2, 0.4, -0.2)$**	**A rook matrix**	**A queen matrix**
$\|R_n^{(2)}(\theta_1) - R_n^{(1)}(\theta_1)\| / \|R_n^{(1)}(\theta_1)\|$	0.2390	0.2120	$\|R_n^{(2)}(\theta_1) - R_n^{(1)}(\theta_1)\| / \|R_n^{(1)}(\theta_1)\|$	0.2618	0.2634
$\|R_n^{(3)}(\theta_1) - R_n^{(2)}(\theta_1)\| / \|R_n^{(2)}(\theta_1)\|$	0.0198	0.0201	$\|R_n^{(3)}(\theta_1) - R_n^{(2)}(\theta_1)\| / \|R_n^{(2)}(\theta_1)\|$	0.0191	0.0209
$\|R_n^{(4)}(\theta_1) - R_n^{(3)}(\theta_1)\| / \|R_n^{(3)}(\theta_1)\|$	0.0020	0.0021	$\|R_n^{(4)}(\theta_1) - R_n^{(3)}(\theta_1)\| / \|R_n^{(3)}(\theta_1)\|$	0.0040	0.0037
$\|R_n^{(5)}(\theta_1) - R_n^{(4)}(\theta_1)\| / \|R_n^{(4)}(\theta_1)\|$	0.0005	0.0005	$\|R_n^{(5)}(\theta_1) - R_n^{(4)}(\theta_1)\| / \|R_n^{(4)}(\theta_1)\|$	0.0036	0.0012
$\|R_n^{(6)}(\theta_1) - R_n^{(5)}(\theta_1)\| / \|R_n^{(5)}(\theta_1)\|$	0.0003	0.0003	$\|R_n^{(6)}(\theta_1) - R_n^{(5)}(\theta_1)\| / \|R_n^{(5)}(\theta_1)\|$	0.0030	0.0004
$\|R_n^{(7)}(\theta_1) - R_n^{(6)}(\theta_1)\| / \|R_n^{(6)}(\theta_1)\|$	0.0000	0.0000	$\|R_n^{(7)}(\theta_1) - R_n^{(6)}(\theta_1)\| / \|R_n^{(6)}(\theta_1)\|$	0.0002	0.0000

Table A.2 : Performance of several approximations for $\ln|R_n(\theta_1)|$

Approximation order (J)	A rook matrix	A queen matrix
$J=1$	0.0191	0.0118
$J=2$	0.0031	0.0023
$J=3$	0.0009	0.0006
$J=4$	0.0002	0.0002
$J=5$	0.0001	0.0000

Note: For *J*, refer to equation (23) in Section 2.7.

Table A.3 : Performance of $\hat{\theta}_{ml,nT}$ and $\hat{\theta}^{c}_{ml,nT}$ when $\delta = 0.5$

$(n,T)=(49,10)$ $(\lambda,\rho)=(0.2,0.2)$		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	-0.0282	-0.1352	0.0499	0.0314	0.0452	-0.1709	-0.1496	-0.0159	-0.1383
	SD	0.0620	0.0499	0.0760	0.0440	0.0846	0.0645	0.0443	0.0862	0.0612
	RMSE	0.0680	0.1441	0.0908	0.0540	0.0958	0.1826	0.1560	0.0875	0.1512
	CP	0.9200	0.1925	0.8850	0.8875	0.9125	0.2075	0.0600	0.9400	0.3050
$\hat{\theta}^{c}_{ml,nT}$	Bias	-0.0007	-0.0289	-0.0021	0.0150	0.0153	-0.0369	-0.0261	-0.0066	-0.0362
	SD	0.0639	0.0555	0.0846	0.0434	0.0840	0.0736	0.0491	0.0952	0.0684
	RMSE	0.0639	0.0625	0.0845	0.0458	0.0852	0.0822	0.0556	0.0953	0.0773
	CP	0.9425	0.8525	0.9125	0.9325	0.9400	0.8100	0.8625	0.9125	0.8075
$\hat{\theta}^{S}_{ml,nT}$	Bias	-0.0022	-0.1590	0.0302	0.0097	0.0284	-0.3086			
	SD	0.0576	0.0427	0.0707	0.0412	0.0801	0.0483			
	RMSE	0.0575	0.1646	0.0768	0.0422	0.0849	0.3124			
	CP	0.9375	0.0225	0.9300	0.9450	0.9250	0.0000			
$\hat{\theta}^{S,c}_{ml,nT}$	Bias	0.0301	-0.0668	-0.0228	-0.0020	0.0082	-0.2294			
	SD	0.0583	0.0466	0.0764	0.0405	0.0791	0.0539			
	RMSE	0.0656	0.0814	0.0796	0.0405	0.0794	0.2357			
	CP	0.8950	0.6000	0.9150	0.9475	0.9400	0.0000			

$(n,T)=(49,10)$ $(\lambda,\rho)=(0.2,-0.2)$		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	-0.0608	-0.1474	0.0922	0.0262	0.0354	-0.1892	-0.1497	-0.0164	-0.1383
	SD	0.0661	0.0504	0.0816	0.0437	0.0845	0.0641	0.0443	0.0861	0.0612
	RMSE	0.0897	0.1558	0.1231	0.0509	0.0915	0.1997	0.1561	0.0875	0.1512
	CP	0.8300	0.1300	0.7775	0.8900	0.9175	0.1425	0.0600	0.9400	0.3050
$\hat{\theta}^{c}_{ml,nT}$	Bias	-0.0126	-0.0334	0.0223	0.0133	0.0106	-0.0476	-0.0261	-0.0070	-0.0362
	SD	0.0693	0.0567	0.0910	0.0433	0.0848	0.0739	0.0491	0.0951	0.0684
	RMSE	0.0704	0.0658	0.0936	0.0452	0.0853	0.0878	0.0556	0.0953	0.0773
	CP	0.9075	0.8325	0.8925	0.9375	0.9325	0.7550	0.8650	0.9050	0.8075
$\hat{\theta}^{S}_{ml,nT}$	Bias	-0.0813	-0.1705	0.1030	0.0019	0.0141	-0.3262			
	SD	0.0601	0.0431	0.0739	0.0409	0.0796	0.0471			
	RMSE	0.1011	0.1758	0.1267	0.0409	0.0807	0.3295			
	CP	0.6950	0.0125	0.6925	0.9475	0.9375	0.0000			
$\hat{\theta}^{S,c}_{ml,nT}$	Bias	-0.0492	-0.0730	0.0423	-0.0099	-0.0064	-0.2479			
	SD	0.0611	0.0471	0.0800	0.0403	0.0788	0.0526			
	RMSE	0.0784	0.0868	0.0904	0.0415	0.0789	0.2534			
	CP	0.8525	0.5350	0.8825	0.9425	0.9325	0.0000			

$(n,T)=(49,10)$ $(\lambda,\rho)=(-0.2,0.2)$		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	-0.0218	-0.1364	0.0089	0.0236	0.0263	-0.1853	-0.1497	-0.0166	-0.1383
	SD	0.0646	0.0493	0.0778	0.0433	0.0834	0.0638	0.0443	0.0861	0.0612
	RMSE	0.0681	0.1450	0.0782	0.0493	0.0873	0.1960	0.1561	0.0876	0.1512
	CP	0.9100	0.1625	0.9625	0.8950	0.9225	0.1425	0.0600	0.9400	0.3050
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0032	-0.0302	-0.0028	0.0123	0.0081	-0.0446	-0.0262	-0.0072	-0.0362
	SD	0.0675	0.0547	0.0864	0.0428	0.0830	0.0735	0.0492	0.0952	0.0684
	RMSE	0.0675	0.0624	0.0864	0.0444	0.0833	0.0859	0.0556	0.0954	0.0773
	CP	0.9225	0.8325	0.9150	0.9350	0.9400	0.7650	0.8625	0.9050	0.8050
$\hat{\theta}_{ml,nT}^{S}$	Bias	0.0166	-0.1613	-0.0114	-0.0033	-0.0001	-0.3284			
	SD	0.0581	0.0419	0.0703	0.0404	0.0780	0.0472			
	RMSE	0.0603	0.1667	0.0711	0.0405	0.0779	0.3317			
	CP	0.9425	0.0175	0.9550	0.9450	0.9400	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	0.0421	-0.0706	-0.0267	-0.0132	-0.0163	-0.2469			
	SD	0.0596	0.0454	0.0762	0.0399	0.0773	0.0528			
	RMSE	0.0729	0.0839	0.0807	0.0420	0.0789	0.2525			
	CP	0.8825	0.5500	0.9000	0.9350	0.9325	0.0000			

$(n,T)=(49,10)$ $(\lambda,\rho)=(-0.2,-0.2)$		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	-0.0486	-0.1457	0.0657	0.0241	0.0231	-0.1863	-0.1497	-0.0167	-0.1383
	SD	0.0623	0.0491	0.0762	0.0433	0.0838	0.0646	0.0443	0.0861	0.0612
	RMSE	0.0790	0.1537	0.1005	0.0495	0.0868	0.1971	0.1561	0.0876	0.1512
	CP	0.8600	0.1250	0.8550	0.8975	0.9300	0.1500	0.0600	0.9375	0.3050
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0104	-0.0355	0.0290	0.0125	0.0047	-0.0459	-0.0262	-0.0073	-0.0362
	SD	0.0650	0.0545	0.0856	0.0427	0.0829	0.0743	0.0492	0.0952	0.0684
	RMSE	0.0657	0.0650	0.0903	0.0444	0.0830	0.0872	0.0556	0.0953	0.0773
	CP	0.9075	0.8050	0.9050	0.9375	0.9375	0.7525	0.8625	0.9050	0.8050
$\hat{\theta}_{ml,nT}^{S}$	Bias	-0.0535	-0.1675	0.0753	0.0011	0.0021	-0.3189			
	SD	0.0585	0.0422	0.0705	0.0409	0.0798	0.0483			
	RMSE	0.0792	0.1728	0.1031	0.0409	0.0798	0.3225			
	CP	0.8075	0.0225	0.8075	0.9475	0.9400	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	-0.0297	-0.0720	0.0452	-0.0085	-0.0135	-0.2357			
	SD	0.0601	0.0459	0.0774	0.0405	0.0790	0.0542			
	RMSE	0.0670	0.0854	0.0896	0.0414	0.0800	0.2418			
	CP	0.8775	0.5325	0.8600	0.9400	0.9400	0.0050			

$(n,T)=(49,30)$ $(\lambda,\rho)=(0.2,0.2)$		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	-0.0394	-0.0402	0.0341	0.0183	0.0355	-0.0632	-0.0489	-0.0001	-0.0571
	SD	0.0350	0.0276	0.0426	0.0249	0.0452	0.0393	0.0271	0.0482	0.0379
	RMSE	0.0527	0.0487	0.0546	0.0308	0.0574	0.0744	0.0559	0.0482	0.0685
	CP	0.8125	0.6675	0.8550	0.8625	0.8675	0.5575	0.4900	0.9275	0.6000
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0028	-0.0047	-0.0028	0.0030	0.0055	-0.0088	-0.0048	0.0027	-0.0070
	SD	0.0355	0.0283	0.0439	0.0246	0.0448	0.0414	0.0281	0.0498	0.0399
	RMSE	0.0356	0.0287	0.0440	0.0247	0.0451	0.0423	0.0285	0.0498	0.0405
	CP	0.9550	0.9300	0.9575	0.9575	0.9550	0.8925	0.9075	0.9300	0.9125
$\hat{\theta}_{ml,nT}^{S}$	Bias	-0.0022	-0.0782	0.0123	0.0029	0.0286	-0.2472			
	SD	0.0312	0.0233	0.0387	0.0236	0.0428	0.0306			
	RMSE	0.0312	0.0816	0.0405	0.0237	0.0514	0.2490			
	CP	0.9550	0.0775	0.9350	0.9475	0.8850	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	0.0321	-0.0472	-0.0219	-0.0068	0.0081	-0.2103			
	SD	0.0315	0.0240	0.0397	0.0234	0.0425	0.0321			
	RMSE	0.0449	0.0529	0.0453	0.0243	0.0432	0.2127			
	CP	0.8225	0.4550	0.9225	0.9350	0.9575	0.0000			

$(n,T)=(49,30)$ $(\lambda,\rho)=(0.2,-0.2)$		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	-0.0485	-0.0483	0.0342	0.0152	0.0277	-0.0728	-0.0490	-0.0007	-0.0571
	SD	0.0370	0.0276	0.0451	0.0247	0.0452	0.0391	0.0271	0.0482	0.0379
	RMSE	0.0609	0.0556	0.0566	0.0289	0.0530	0.0826	0.0560	0.0482	0.0685
	CP	0.7625	0.5575	0.8925	0.8825	0.9000	0.4625	0.4925	0.9300	0.6000
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0038	-0.0054	-0.0019	0.0023	0.0036	-0.0109	-0.0048	0.0025	-0.0070
	SD	0.0379	0.0284	0.0468	0.0246	0.0451	0.0415	0.0281	0.0498	0.0399
	RMSE	0.0381	0.0289	0.0467	0.0246	0.0451	0.0429	0.0285	0.0498	0.0405
	CP	0.9550	0.9250	0.9500	0.9525	0.9450	0.8825	0.9075	0.9275	0.9125
$\hat{\theta}_{ml,nT}^{S}$	Bias	-0.0822	-0.0868	0.0537	-0.0053	0.0126	-0.2627			
	SD	0.0323	0.0231	0.0398	0.0233	0.0423	0.0300			
	RMSE	0.0883	0.0898	0.0669	0.0239	0.0441	0.2644			
	CP	0.2700	0.0325	0.7225	0.9425	0.9400	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	-0.0474	-0.0508	0.0223	-0.0145	-0.0055	-0.2254			
	SD	0.0329	0.0237	0.0410	0.0232	0.0421	0.0315			
	RMSE	0.0576	0.0560	0.0466	0.0273	0.0424	0.2275			
	CP	0.7075	0.3775	0.9375	0.8750	0.9475	0.0000			

$(n,T)=(49,30)$ $(\lambda,\rho)=(-0.2,0.2)$		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	-0.0310	-0.0429	0.0104	0.0130	0.0215	-0.0742	-0.0490	-0.0009	-0.0571
	SD	0.0362	0.0272	0.0454	0.0243	0.0443	0.0393	0.0271	0.0483	0.0379
	RMSE	0.0476	0.0508	0.0465	0.0276	0.0491	0.0840	0.0560	0.0482	0.0685
	CP	0.8600	0.6200	0.9475	0.8975	0.9275	0.4375	0.4900	0.9325	0.6000
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0009	-0.0057	-0.0054	0.0019	0.0025	-0.0114	-0.0048	0.0024	-0.0070
	SD	0.0373	0.0278	0.0471	0.0241	0.0441	0.0417	0.0281	0.0499	0.0399
	RMSE	0.0373	0.0284	0.0473	0.0242	0.0441	0.0432	0.0285	0.0498	0.0405
	CP	0.9650	0.9175	0.9325	0.9525	0.9475	0.8800	0.9075	0.9275	0.9125
$\hat{\theta}_{ml,nT}^{S}$	Bias	0.0159	-0.0824	-0.0143	-0.0104	-0.0018	-0.2668			
	SD	0.0311	0.0227	0.0400	0.0229	0.0415	0.0302			
	RMSE	0.0349	0.0855	0.0424	0.0251	0.0415	0.2685			
	CP	0.9325	0.0350	0.9425	0.9175	0.9475	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	0.0436	-0.0509	-0.0285	-0.0178	-0.0161	-0.2256			
	SD	0.0319	0.0232	0.0413	0.0228	0.0415	0.0319			
	RMSE	0.0540	0.0560	0.0502	0.0289	0.0444	0.2278			
	CP	0.7450	0.3525	0.8925	0.8550	0.9500	0.0000			

$(n,T)=(49,30)$ $(\lambda,\rho)=(-0.2,-0.2)$		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	-0.0369	-0.0489	0.0248	0.0124	0.0169	-0.0760	-0.0490	-0.0011	-0.0571
	SD	0.0342	0.0271	0.0422	0.0240	0.0438	0.0399	0.0271	0.0483	0.0379
	RMSE	0.0503	0.0559	0.0489	0.0269	0.0469	0.0858	0.0560	0.0482	0.0685
	CP	0.8275	0.4950	0.9275	0.9025	0.9275	0.4400	0.4900	0.9325	0.6000
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0012	-0.0070	-0.0001	0.0018	0.0010	-0.0114	-0.0048	0.0023	-0.0070
	SD	0.0353	0.0276	0.0440	0.0238	0.0435	0.0425	0.0281	0.0498	0.0399
	RMSE	0.0353	0.0285	0.0440	0.0239	0.0435	0.0439	0.0285	0.0498	0.0405
	CP	0.9650	0.9300	0.9475	0.9475	0.9500	0.8775	0.9075	0.9275	0.9125
$\hat{\theta}_{ml,nT}^{S}$	Bias	-0.0593	-0.0842	0.0405	-0.0055	0.0020	-0.2539			
	SD	0.0310	0.0230	0.0383	0.0230	0.0422	0.0309			
	RMSE	0.0669	0.0873	0.0557	0.0236	0.0422	0.2557			
	CP	0.5375	0.0350	0.8375	0.9425	0.9425	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	-0.0338	-0.0483	0.0202	-0.0124	-0.0099	-0.2112			
	SD	0.0319	0.0235	0.0398	0.0230	0.0421	0.0326			
	RMSE	0.0464	0.0537	0.0446	0.0261	0.0432	0.2137			
	CP	0.7975	0.4050	0.9275	0.9075	0.9475	0.0000			

$(n,T)=(81,10)$ $(\lambda,\rho)=(0.2,0.2)$		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	-0.0078	-0.1361	0.0451	0.0245	0.0318	-0.1647	-0.1488	-0.0111	-0.1332
	SD	0.0493	0.0387	0.0610	0.0376	0.0711	0.0496	0.0342	0.0703	0.0447
	RMSE	0.0498	0.1415	0.0758	0.0449	0.0778	0.1720	0.1527	0.0711	0.1405
	CP	0.9250	0.0400	0.8450	0.8825	0.9000	0.0650	0.0100	0.9325	0.1550
$\hat{\theta}_{ml,nT}^{c}$	Bias	0.0062	-0.0280	-0.0021	0.0103	0.0070	-0.0334	-0.0247	-0.0014	-0.0368
	SD	0.0504	0.0427	0.0678	0.0367	0.0698	0.0566	0.0380	0.0768	0.0497
	RMSE	0.0507	0.0510	0.0678	0.0381	0.0701	0.0657	0.0453	0.0767	0.0618
	CP	0.9125	0.8100	0.9125	0.9175	0.9425	0.8225	0.8575	0.9075	0.7975
$\hat{\theta}_{ml,nT}^{S}$	Bias	0.0166	-0.1596	0.0255	0.0035	0.0170	-0.3030			
	SD	0.0465	0.0332	0.0571	0.0352	0.0669	0.0384			
	RMSE	0.0493	0.1630	0.0625	0.0353	0.0689	0.3054			
	CP	0.9200	0.0000	0.9000	0.9250	0.9150	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	0.0366	-0.0660	-0.0232	-0.0062	0.0019	-0.2273			
	SD	0.0468	0.0359	0.0617	0.0345	0.0654	0.0427			
	RMSE	0.0594	0.0751	0.0659	0.0350	0.0654	0.2313			
	CP	0.8300	0.4525	0.9025	0.9150	0.9350	0.0000			

$(n,T)=(81,10)$ $(\lambda,\rho)=(0.2,-0.2)$		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	-0.0426	-0.1462	0.0846	0.0210	0.0268	-0.1802	-0.1488	-0.0114	-0.1332
	SD	0.0528	0.0389	0.0662	0.0374	0.0705	0.0489	0.0342	0.0702	0.0447
	RMSE	0.0678	0.1513	0.1074	0.0429	0.0753	0.1867	0.1527	0.0711	0.1405
	CP	0.8575	0.0225	0.6850	0.8875	0.8950	0.0325	0.0100	0.9350	0.1550
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0070	-0.0316	0.0179	0.0097	0.0055	-0.0422	-0.0247	-0.0016	-0.0368
	SD	0.0547	0.0435	0.0733	0.0367	0.0701	0.0563	0.0380	0.0768	0.0497
	RMSE	0.0551	0.0537	0.0753	0.0379	0.0702	0.0703	0.0453	0.0767	0.0618
	CP	0.9225	0.7925	0.8950	0.9225	0.9350	0.7725	0.8575	0.9100	0.7975
$\hat{\theta}_{ml,nT}^{S}$	Bias	-0.0651	-0.1692	0.0958	-0.0032	0.0062	-0.3196			
	SD	0.0483	0.0332	0.0599	0.0349	0.0659	0.0375			
	RMSE	0.0810	0.1724	0.1129	0.0350	0.0661	0.3218			
	CP	0.6775	0.0000	0.5825	0.9025	0.9325	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	-0.0448	-0.0713	0.0381	-0.0134	-0.0106	-0.2450			
	SD	0.0487	0.0360	0.0643	0.0344	0.0649	0.0416			
	RMSE	0.0661	0.0799	0.0747	0.0368	0.0656	0.2485			
	CP	0.8075	0.4050	0.8550	0.8975	0.9350	0.0000			

$(n,T)=(81,10)$ $(\lambda,\rho)=(-0.2,0.2)$		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	-0.0074	-0.1365	0.0042	0.0191	0.0189	-0.1754	-0.1488	-0.0116	-0.1332
	SD	0.0516	0.0377	0.0654	0.0370	0.0697	0.0488	0.0342	0.0702	0.0447
	RMSE	0.0521	0.1416	0.0654	0.0416	0.0721	0.1820	0.1527	0.0711	0.1405
	CP	0.9550	0.0350	0.9400	0.8925	0.9050	0.0425	0.0100	0.9325	0.1550
$\hat{\theta}_{ml,nT}^{c}$	Bias	0.0006	-0.0286	-0.0050	0.0089	0.0034	-0.0390	-0.0247	-0.0017	-0.0368
	SD	0.0534	0.0415	0.0721	0.0361	0.0687	0.0561	0.0380	0.0768	0.0497
	RMSE	0.0534	0.0504	0.0722	0.0372	0.0687	0.0683	0.0453	0.0767	0.0618
	CP	0.9475	0.8125	0.9175	0.9225	0.9325	0.7950	0.8575	0.9100	0.7975
$\hat{\theta}_{ml,nT}^{S}$	Bias	0.0293	-0.1611	-0.0152	-0.0075	-0.0067	-0.3203			
	SD	0.0468	0.0321	0.0592	0.0344	0.0648	0.0375			
	RMSE	0.0552	0.1643	0.0610	0.0352	0.0651	0.3225			
	CP	0.8825	0.0000	0.9300	0.8950	0.9275	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	0.0452	-0.0691	-0.0283	-0.0164	-0.0204	-0.2436			
	SD	0.0475	0.0345	0.0637	0.0339	0.0637	0.0417			
	RMSE	0.0656	0.0772	0.0696	0.0376	0.0668	0.2472			
	CP	0.8000	0.3900	0.8900	0.8850	0.9325	0.0000			

$(n,T)=(81,10)$ $(\lambda,\rho)=(-0.2,-0.2)$		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	-0.0367	-0.1442	0.0550	0.0201	0.0184	-0.1751	-0.1488	-0.0117	-0.1332
	SD	0.0494	0.0373	0.0629	0.0366	0.0689	0.0492	0.0342	0.0702	0.0447
	RMSE	0.0615	0.1490	0.0835	0.0417	0.0713	0.1819	0.1527	0.0711	0.1405
	CP	0.8550	0.0200	0.8250	0.8850	0.9100	0.0450	0.0100	0.9350	0.1550
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0074	-0.0333	0.0211	0.0094	0.0014	-0.0397	-0.0247	-0.0018	-0.0368
	SD	0.0508	0.0411	0.0698	0.0357	0.0678	0.0567	0.0380	0.0768	0.0497
	RMSE	0.0513	0.0529	0.0729	0.0369	0.0677	0.0691	0.0453	0.0767	0.0618
	CP	0.9350	0.7900	0.8775	0.9200	0.9300	0.7800	0.8575	0.9100	0.7975
$\hat{\theta}_{ml,nT}^{S}$	Bias	-0.0429	-0.1662	0.0664	-0.0030	-0.0024	-0.3105			
	SD	0.0466	0.0321	0.0581	0.0346	0.0653	0.0382			
	RMSE	0.0633	0.1693	0.0881	0.0347	0.0653	0.3128			
	CP	0.8025	0.0000	0.7475	0.9000	0.9300	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	-0.0276	-0.0702	0.0389	-0.0117	-0.0166	-0.2323			
	SD	0.0474	0.0346	0.0631	0.0341	0.0644	0.0425			
	RMSE	0.0548	0.0783	0.0740	0.0360	0.0664	0.2362			
	CP	0.8700	0.3900	0.8525	0.8975	0.9325	0.0000			

$(n,T)=(81,30)$ $(\lambda,\rho)=(0.2,0.2)$		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	-0.0224	-0.0408	0.0306	0.0142	0.0278	-0.0555	-0.0480	-0.0025	-0.0501
	SD	0.0297	0.0213	0.0333	0.0201	0.0393	0.0299	0.0189	0.0364	0.0272
	RMSE	0.0372	0.0460	0.0452	0.0245	0.0480	0.0630	0.0516	0.0364	0.0570
	CP	0.8500	0.5025	0.8575	0.8725	0.8575	0.4900	0.3025	0.9400	0.5400
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0007	-0.0037	0.0005	0.0012	0.0042	-0.0054	-0.0035	0.0001	-0.0071
	SD	0.0300	0.0218	0.0343	0.0198	0.0390	0.0313	0.0196	0.0374	0.0284
	RMSE	0.0300	0.0221	0.0343	0.0199	0.0392	0.0317	0.0199	0.0374	0.0292
	CP	0.9300	0.9300	0.9350	0.9225	0.9375	0.9100	0.9350	0.9375	0.9275
$\hat{\theta}_{ml,nT}^{S}$	Bias	0.0136	-0.0786	0.0093	-0.0006	0.0226	-0.2403			
	SD	0.0268	0.0179	0.0303	0.0190	0.0372	0.0229			
	RMSE	0.0300	0.0806	0.0317	0.0190	0.0435	0.2414			
	CP	0.8925	0.0075	0.9375	0.9150	0.8675	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	0.0349	-0.0464	-0.0189	-0.0085	0.0075	-0.2080			
	SD	0.0270	0.0185	0.0310	0.0188	0.0371	0.0239			
	RMSE	0.0441	0.0499	0.0363	0.0206	0.0378	0.2093			
	CP	0.7050	0.2425	0.8825	0.8900	0.9300	0.0000			

$(n,T)=(81,30)$ $(\lambda,\rho)=(0.2,-0.2)$		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	-0.0327	-0.0469	0.0337	0.0121	0.0225	-0.0625	-0.0481	-0.0029	-0.0501
	SD	0.0317	0.0211	0.0362	0.0200	0.0393	0.0294	0.0189	0.0364	0.0272
	RMSE	0.0455	0.0514	0.0494	0.0234	0.0453	0.0691	0.0517	0.0364	0.0570
	CP	0.7800	0.3800	0.8500	0.8925	0.8725	0.3900	0.2975	0.9400	0.5400
$\hat{\theta}_{ml,nT}^{c}$	Bias	-0.0025	-0.0044	0.0025	0.0011	0.0036	-0.0069	-0.0035	-0.0000	-0.0071
	SD	0.0322	0.0218	0.0374	0.0199	0.0393	0.0310	0.0196	0.0374	0.0284
	RMSE	0.0323	0.0222	0.0374	0.0199	0.0394	0.0317	0.0199	0.0374	0.0292
	CP	0.9275	0.9400	0.9225	0.9200	0.9300	0.9175	0.9350	0.9350	0.9275
$\hat{\theta}_{ml,nT}^{S}$	Bias	-0.0673	-0.0855	0.0529	-0.0082	0.0081	-0.2552			
	SD	0.0283	0.0176	0.0320	0.0189	0.0368	0.0224			
	RMSE	0.0730	0.0873	0.0618	0.0206	0.0376	0.2561			
	CP	0.2675	0.0025	0.6225	0.9025	0.9200	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	-0.0454	-0.0498	0.0259	-0.0156	-0.0052	-0.2227			
	SD	0.0285	0.0182	0.0328	0.0188	0.0367	0.0234			
	RMSE	0.0536	0.0530	0.0418	0.0244	0.0371	0.2239			
	CP	0.5850	0.1775	0.8650	0.8275	0.9300	0.0000			

$(n,T)=(81,30)$ $(\lambda,\rho)=(-0.2,0.2)$		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	-0.0189	-0.0419	0.0109	0.0108	0.0185	-0.0623	-0.0481	-0.0032	-0.0501
	SD	0.0314	0.0208	0.0357	0.0198	0.0389	0.0293	0.0189	0.0364	0.0272
	RMSE	0.0367	0.0467	0.0373	0.0225	0.0430	0.0689	0.0517	0.0365	0.0570
	CP	0.8725	0.4625	0.9425	0.9000	0.8925	0.3875	0.2975	0.9425	0.5400
$\hat{\theta}_{ml,nT}^c$	Bias	-0.0010	-0.0037	-0.0006	0.0009	0.0031	-0.0067	-0.0036	-0.0001	-0.0071
	SD	0.0321	0.0213	0.0369	0.0196	0.0387	0.0308	0.0196	0.0374	0.0284
	RMSE	0.0320	0.0216	0.0369	0.0196	0.0388	0.0315	0.0199	0.0374	0.0292
	CP	0.9325	0.9325	0.9275	0.9200	0.9300	0.9200	0.9350	0.9425	0.9275
$\hat{\theta}_{ml,nT}^S$	Bias	0.0270	-0.0815	-0.0133	-0.0125	-0.0044	-0.2576			
	SD	0.0276	0.0174	0.0315	0.0186	0.0364	0.0223			
	RMSE	0.0386	0.0833	0.0341	0.0224	0.0366	0.2586			
	CP	0.8200	0.0050	0.9125	0.8550	0.9225	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	0.0443	-0.0493	-0.0241	-0.0188	-0.0155	-0.2227			
	SD	0.0281	0.0178	0.0324	0.0185	0.0364	0.0234			
	RMSE	0.0524	0.0524	0.0403	0.0264	0.0395	0.2239			
	CP	0.5875	0.1575	0.8925	0.7925	0.9025	0.0000			

$(n,T)=(81,30)$ $(\lambda,\rho)=(-0.2,-0.2)$		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	-0.0272	-0.0465	0.0234	0.0108	0.0155	-0.0631	-0.0481	-0.0032	-0.0501
	SD	0.0292	0.0205	0.0340	0.0195	0.0383	0.0294	0.0189	0.0364	0.0272
	RMSE	0.0399	0.0508	0.0412	0.0223	0.0412	0.0696	0.0517	0.0365	0.0570
	CP	0.8125	0.3600	0.8925	0.9000	0.8950	0.4100	0.2975	0.9425	0.5400
$\hat{\theta}_{ml,nT}^c$	Bias	-0.0018	-0.0052	0.0039	0.0010	0.0023	-0.0069	-0.0036	-0.0002	-0.0071
	SD	0.0298	0.0210	0.0354	0.0193	0.0380	0.0309	0.0196	0.0374	0.0284
	RMSE	0.0298	0.0216	0.0356	0.0193	0.0381	0.0317	0.0199	0.0374	0.0292
	CP	0.9375	0.9375	0.9475	0.9275	0.9250	0.9150	0.9350	0.9425	0.9275
$\hat{\theta}_{ml,nT}^S$	Bias	-0.0492	-0.0824	0.0399	-0.0074	0.0001	-0.2448			
	SD	0.0271	0.0174	0.0310	0.0188	0.0367	0.0227			
	RMSE	0.0561	0.0842	0.0505	0.0201	0.0367	0.2458			
	CP	0.5100	0.0050	0.7450	0.8975	0.9275	0.0000			
$\hat{\theta}_{ml,nT}^{S,c}$	Bias	-0.0330	-0.0470	0.0241	-0.0134	-0.0091	-0.2088			
	SD	0.0276	0.0178	0.0320	0.0187	0.0367	0.0238			
	RMSE	0.0430	0.0503	0.0400	0.0229	0.0378	0.2101			
	CP	0.7150	0.1900	0.8700	0.8525	0.9150	0.0000			

Table A.4 : Likelihood measures for identifying δ_0 and misspecification errors

Model 1: δ_0=0.95, K=1 and (n,T)=(49,10)

δ	$\overline{E\ln L}$	$\overline{E\ln L_1}$	*AIC*	*BIC*	RMSE λ	RMSE γ	RMSE ρ
0	-313.816	-105.899	763.632	1048.851	0.0730	0.1021	0.0822
0.1	-313.808	-105.890	763.616	1048.835	0.0710	0.0973	0.0825
0.2	-313.800	-105.882	763.600	1048.819	0.0693	0.0926	0.0830
0.3	-313.792	-105.874	763.585	1048.804	0.0678	0.0879	0.0837
0.4	-313.831	-105.891	763.662	1048.881	0.0669	0.0835	0.0846
0.5	-313.890	-105.933	763.780	1049.000	0.0668	0.0795	0.0862
0.6	-313.844	-105.900	763.688	1048.907	0.0669	0.0758	0.0875
0.7	-314.032	-106.030	764.063	1049.283	0.0688	0.0728	0.0895
0.8	-313.801	-105.863	763.602	1048.821	0.0701	0.0707	0.0911
0.825	-313.883	-105.918	763.766	1048.985	0.0710	0.0704	0.0918
0.85	-313.835	-105.895	763.671	1048.890	0.0716	0.0701	0.0921
0.875	-313.847	-105.906	763.694	1048.914	0.0727	0.0699	0.0926
0.9	-313.769	-105.845	763.539	1048.758	0.0733	0.0697	0.0933
0.925	-313.822	-105.875	763.644	1048.864	0.0696	0.0940	0.0475
0.95	-313.896	-105.908	763.792	1049.011	0.0758	0.0698	0.0946
0.975	-313.913	-105.929	763.826	1049.046	0.0771	0.0701	0.0956
0.99	-313.772	-105.846	763.544	1048.764	0.0773	0.0699	0.0956

Model 2: δ_0=0.95, K=2 and (n,T)=(49,10)

δ	$\overline{E\ln L}$	$\overline{E\ln L_1}$	*AIC*	*BIC*	RMSE λ	RMSE γ	RMSE ρ
0	-519.571	-104.429	1185.141	1491.333	0.0783	0.0997	0.0766
0.1	-519.547	-104.405	1185.094	1491.286	0.0758	0.0945	0.0767
0.2	-519.524	-104.381	1185.047	1491.239	0.0735	0.0893	0.0770
0.3	-519.500	-104.357	1185.001	1491.192	0.0714	0.0842	0.0775
0.4	-519.478	-104.333	1184.955	1491.147	0.0697	0.0792	0.0782
0.5	-519.456	-104.309	1184.912	1491.104	0.0686	0.0744	0.0792
0.6	-519.436	-104.287	1184.873	1491.064	0.0682	0.0702	0.0804
0.7	-519.419	-104.267	1184.838	1491.030	0.0689	0.0666	0.0819
0.8	-519.405	-104.250	1184.810	1491.002	0.0707	0.0639	0.0837
0.825	-519.402	-104.246	1184.804	1490.996	0.0713	0.0634	0.0842
0.85	-519.400	-104.243	1184.799	1490.991	0.0721	0.0630	0.0847
0.875	-519.397	-104.240	1184.795	1490.986	0.0729	0.0627	0.0852
0.9	-519.524	-104.258	1185.047	1491.239	0.0738	0.0629	0.0857
0.925	-519.394	-104.234	1184.787	1490.979	0.0748	0.0624	0.0864
0.95	-519.480	-104.285	1184.961	1491.152	0.0759	0.0631	0.0871
0.975	-519.391	-104.229	1184.783	1490.974	0.0771	0.0625	0.0876
0.99	-519.391	-104.228	1184.782	1490.974	0.0778	0.0626	0.0880

Model 3: δ_0=0.95, K=1 and (n,T)=(81,30)

δ	$\overline{E\ln L}$	$\overline{E\ln L_1}$	*AIC*	*BIC*	RMSE λ	RMSE γ	RMSE ρ
0	-1799.543	-647.517	3839.086	4534.564	0.0623	0.0835	0.0447
0.1	-1799.422	-647.395	3838.843	4534.321	0.0585	0.0770	0.0429
0.2	-1799.302	-647.275	3838.604	4534.082	0.0546	0.0703	0.0412
0.3	-1799.483	-647.446	3838.965	4534.443	0.0506	0.0633	0.0397
0.4	-1799.849	-647.770	3839.698	4535.175	0.0471	0.0563	0.0384
0.5	-1799.748	-647.663	3839.497	4534.974	0.0432	0.0492	0.0373
0.6	-1801.327	-649.095	3842.654	4538.132	0.0413	0.0426	0.0364
0.7	-1799.594	-647.461	3839.189	4534.666	0.0368	0.0359	0.0357
0.8	-1800.353	-648.154	3840.706	4536.184	0.0361	0.0361	0.0361
0.825	-1799.709	-647.566	3839.419	4534.896	0.0350	0.0350	0.0361
0.85	-1799.451	-647.313	3838.902	4534.379	0.0347	0.0347	0.0361
0.875	-1799.629	-647.478	3839.259	4534.737	0.0347	0.0347	0.0363
0.9	-1799.887	-647.701	3839.775	4535.252	0.0353	0.0353	0.0364
0.925	-1799.272	-647.154	3838.544	4534.021	0.0346	0.0346	0.0367
0.95	-1798.690	-646.650	3837.381	4532.859	0.0339	0.0339	0.0368
0.975	-1798.690	-646.649	3837.379	4532.857	0.0344	0.0344	0.0371
0.99	-1798.690	-646.649	3837.380	4532.858	0.0348	0.0348	0.0373

Model 4: δ_0=0.95, K=2 and (n,T)=(81,30)

δ	$\overline{E\ln L}$	$\overline{E\ln L_1}$	*AIC*	*BIC*	RMSE λ	RMSE γ	RMSE ρ
0	-2953.095	-646.914	6156.189	6880.645	0.0678	0.0839	0.0446
0.1	-2952.879	-646.699	6155.759	6880.215	0.0634	0.0772	0.0426
0.2	-2952.668	-646.487	6155.336	6879.792	0.0588	0.0703	0.0407
0.3	-2952.463	-646.279	6154.925	6879.381	0.0540	0.0632	0.0388
0.4	-2952.266	-646.081	6154.532	6878.988	0.0490	0.0559	0.0371
0.5	-2952.083	-645.895	6154.166	6878.622	0.0439	0.0484	0.0355
0.6	-2951.921	-645.730	6153.841	6878.297	0.0389	0.0410	0.0343
0.7	-2951.786	-645.591	6153.571	6878.027	0.0344	0.0339	0.0334
0.8	-2951.686	-645.488	6153.371	6877.827	0.0310	0.0278	0.0330
0.825	-2951.667	-645.468	6153.334	6877.790	0.0305	0.0266	0.0330
0.85	-2951.651	-645.452	6153.303	6877.759	0.0301	0.0256	0.0330
0.875	-2951.639	-645.438	6153.277	6877.733	0.0299	0.0248	0.0330
0.9	-2951.629	-645.427	6153.258	6877.713	0.0298	0.0243	0.0331
0.925	-2951.622	-645.419	6153.244	6877.700	0.0299	0.0240	0.0332
0.95	-2951.619	-645.415	6153.237	6877.693	0.0303	0.0241	0.0334
0.975	-2951.619	-645.414	6153.237	6877.693	0.0308	0.0244	0.0336
0.99	-2951.620	-645.415	6153.240	6877.696	0.0312	0.0248	0.0337

Model 5: δ_0=0, K=1 and (n,T)=(81,30)

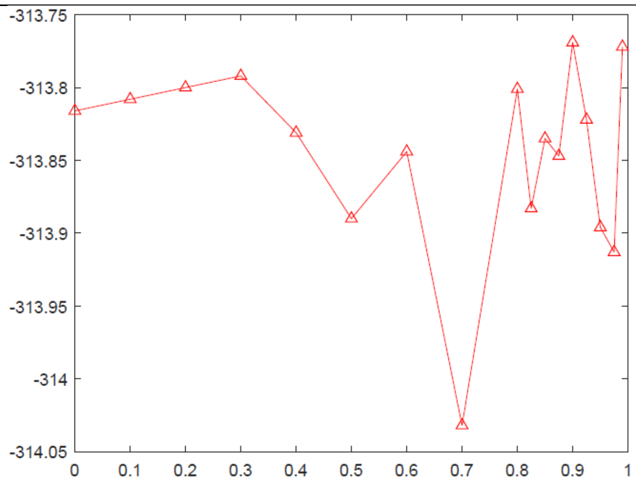
δ	$\overline{E\ln L}$	$\overline{E\ln L_1}$	*AIC*	*BIC*	RMSE λ	RMSE γ	RMSE ρ
0	-2318.808	-1166.782	4877.617	5573.094	0.0275	0.0185	0.0318
0.05	-2318.823	-1166.797	4877.647	5573.124	0.0284	0.0189	0.0322
0.1	-2318.841	-1166.815	4877.682	5573.159	0.0300	0.0205	0.0328
0.15	-2318.861	-1166.835	4877.722	5573.200	0.0321	0.0230	0.0336
0.2	-2318.885	-1166.858	4877.769	5573.247	0.0348	0.0262	0.0345
0.25	-2318.911	-1166.885	4877.823	5573.300	0.0379	0.0301	0.0355
0.3	-2318.942	-1166.915	4877.884	5573.361	0.0415	0.0343	0.0367
0.4	-2319.015	-1166.988	4878.030	5573.508	0.0498	0.0439	0.0394
0.5	-2319.108	-1167.079	4878.215	5573.693	0.0594	0.0546	0.0426
0.6	-2319.450	-1167.388	4878.901	5574.378	0.0700	0.0658	0.0463
0.7	-2319.360	-1167.329	4878.720	5574.197	0.0819	0.0785	0.0504
0.8	-2319.813	-1167.745	4879.627	5575.104	0.0945	0.0917	0.0548
0.9	-2320.515	-1168.452	4881.029	5576.507	0.1082	0.1056	0.0598
0.99	-2320.972	-1168.880	4881.945	5577.422	0.1219	0.1192	0.0651

Model 6: δ_0=0, K=2 and (n,T)=(81,30)

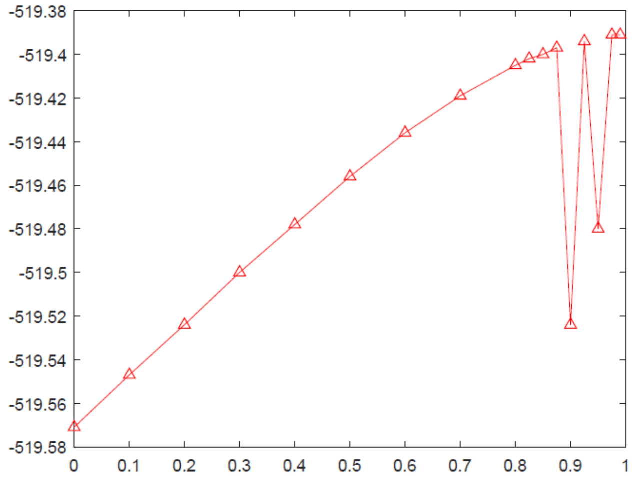
δ	$\overline{E\ln L}$	$\overline{E\ln L_1}$	*AIC*	*BIC*	RMSE λ	RMSE γ	RMSE ρ
0	-3471.856	-1165.676	7193.712	7918.168	0.0253	0.0170	0.0294
0.05	-3471.876	-1165.695	7193.751	7918.207	0.0262	0.0178	0.0298
0.1	-3471.899	-1165.719	7193.799	7918.255	0.0279	0.0197	0.0303
0.15	-3471.928	-1165.747	7193.855	7918.311	0.0303	0.0226	0.0310
0.2	-3471.961	-1165.780	7193.922	7918.378	0.0333	0.0262	0.0319
0.25	-3472.000	-1165.818	7193.999	7918.455	0.0368	0.0304	0.0329
0.3	-3472.044	-1165.863	7194.089	7918.545	0.0408	0.0349	0.0341
0.4	-3472.154	-1165.971	7194.308	7918.764	0.0500	0.0450	0.0369
0.5	-3472.295	-1166.110	7194.590	7919.045	0.0605	0.0560	0.0402
0.6	-3472.470	-1166.283	7194.940	7919.396	0.0723	0.0680	0.0441
0.7	-3473.139	-1166.665	7196.278	7920.734	0.0851	0.0808	0.0484
0.8	-3474.130	-1167.342	7198.261	7922.717	0.0990	0.0944	0.0532
0.9	-3473.964	-1167.428	7197.927	7922.383	0.1145	0.1091	0.0590
0.99	-3474.952	-1168.197	7199.903	7924.359	0.1288	0.1231	0.0643

Figure A.1: Likelihood measures for identifying δ_0 and misspecification errors with figures

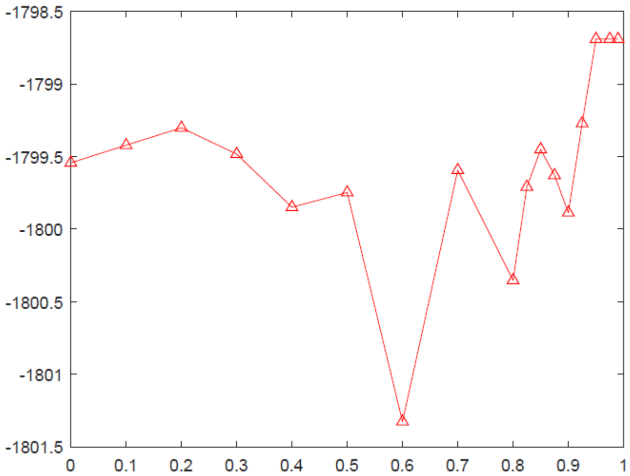
- Average sample log likelihood across different δ's



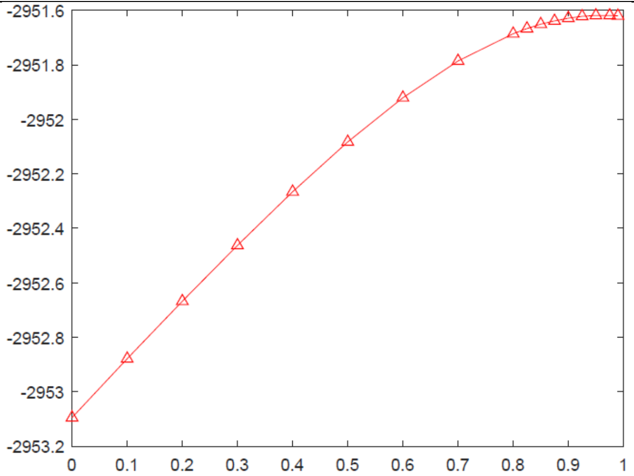
Model 1: δ_0=0.95, K=1 and (n,T)=(49,10)



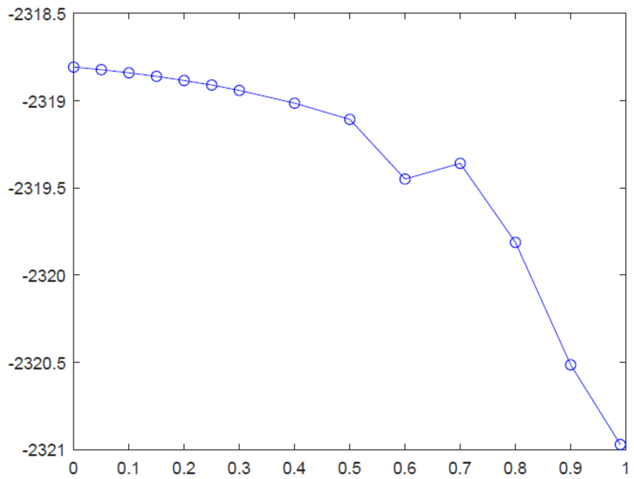
Model 2: δ_0=0.95, K=2 and (n,T)=(49,10)



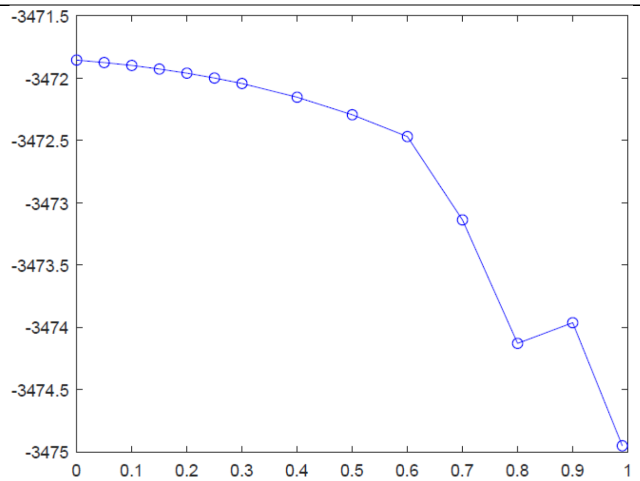
Model 3: δ_0=0.95, K=1 and (n,T)=(81,30)



Model 4: δ_0=0.95, K=2 and (n,T)=(81,30)



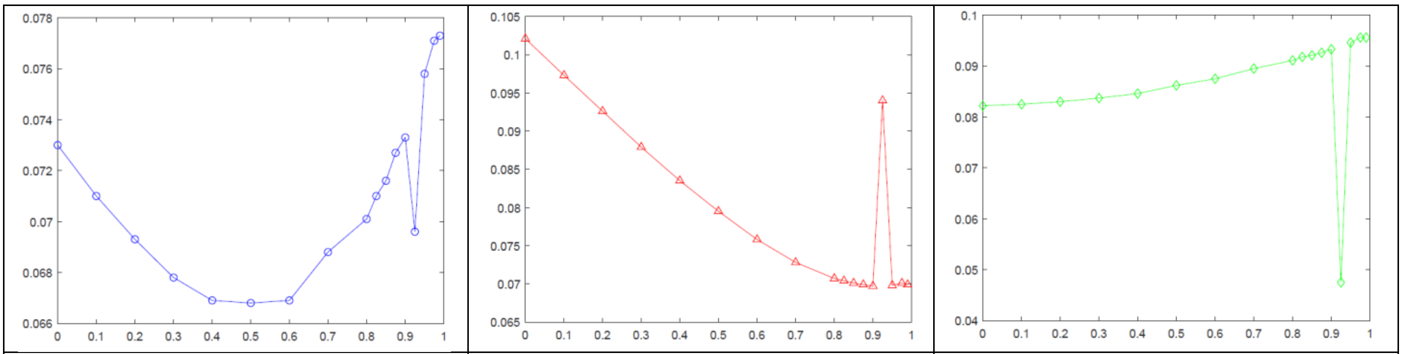
Model 5: δ_0=0, K=1 and (n,T)=(81,30)



Model 6: δ_0=0, K=2 and (n,T)=(81,30)

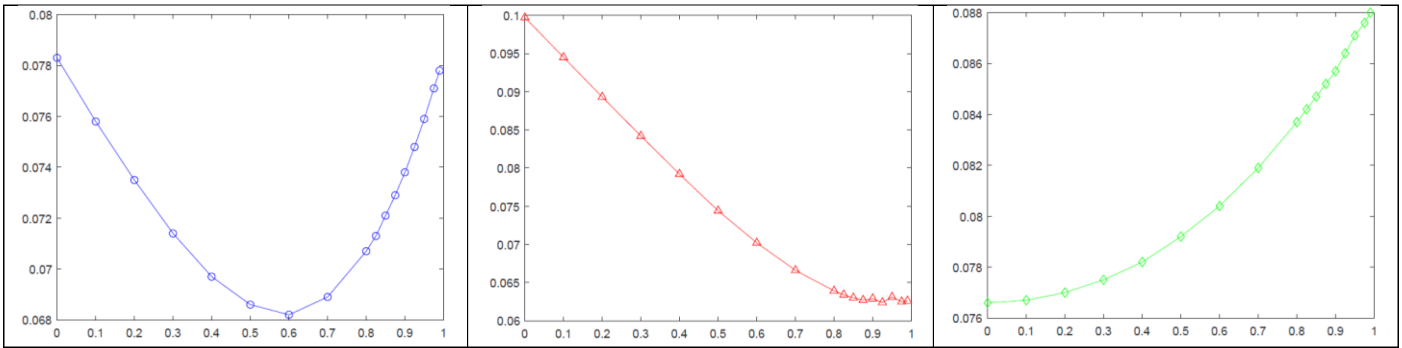
- Misspecification errors (in terms of RMSE) for misspecified δ_0

Model 1: δ_0=0.95, K=1 and (n,T)=(49,10)



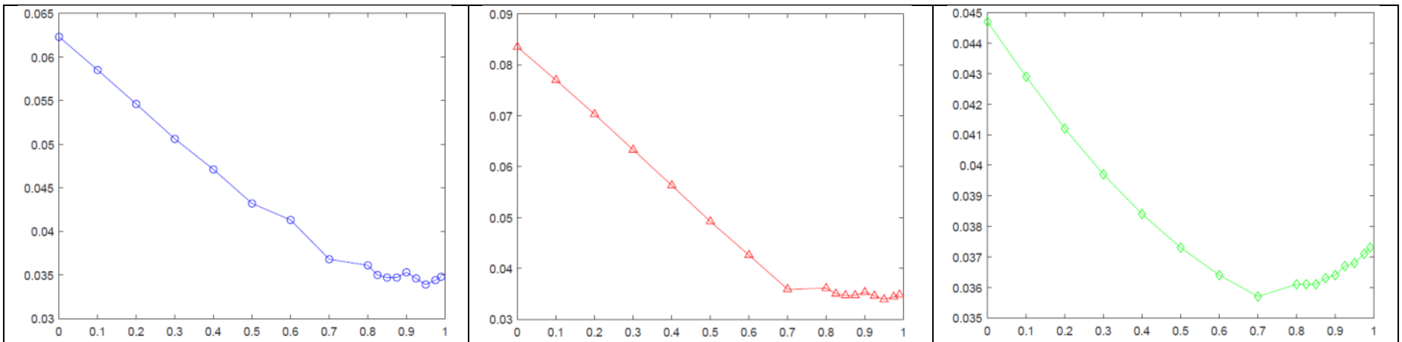
λ_0	γ_0	ρ_0

Model 2: δ_0=0.95, K=2 and (n,T)=(49,10)



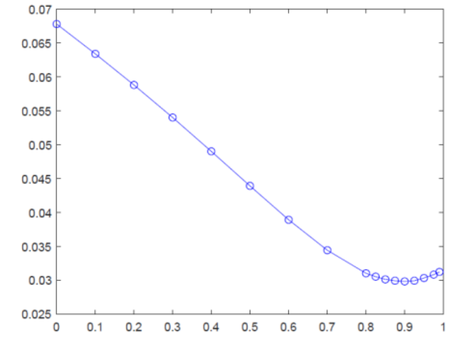
λ_0	γ_0	ρ_0

Model 3: δ_0=0.95, K=1 and (n,T)=(81,30)

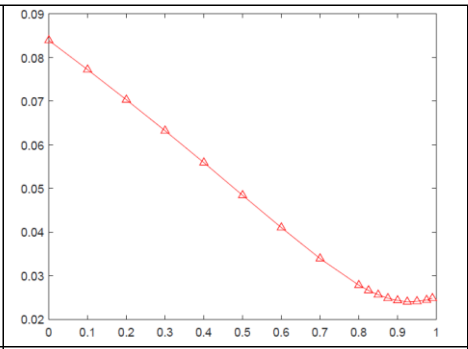


λ_0	γ_0	ρ_0

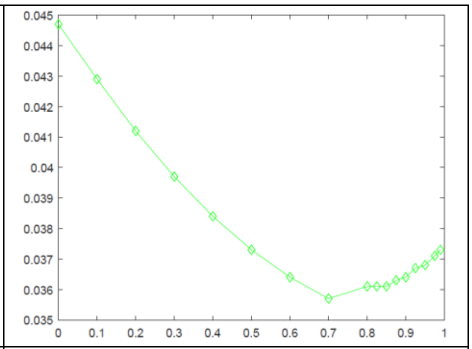
Model 4: δ_0=0.95, K=2 and (n,T)=(81,30)



λ_0

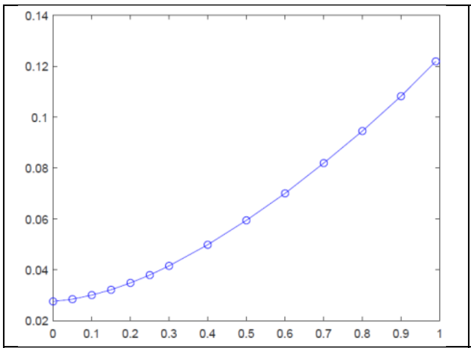


γ_0

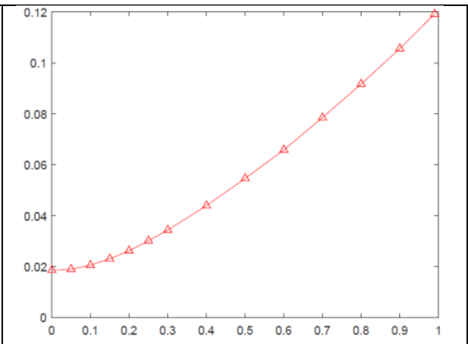


ρ_0

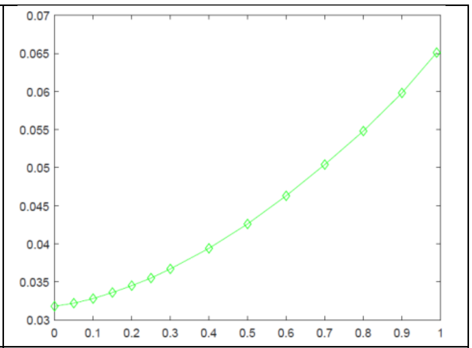
Model 5: δ_0=0, K=1 and (n,T)=(81,30)



λ_0

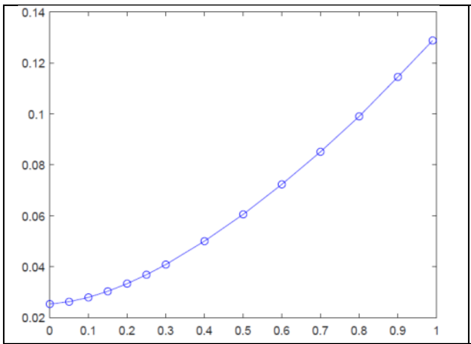


γ_0

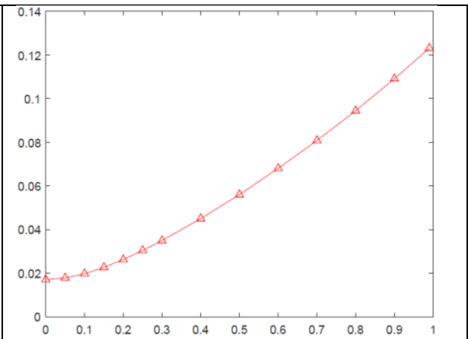


ρ_0

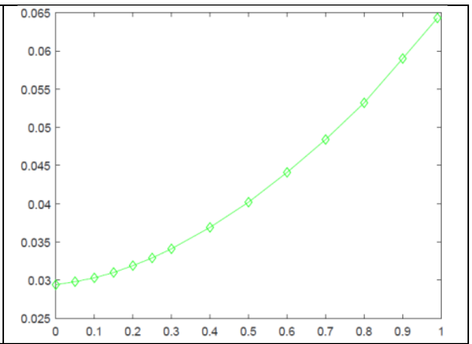
Model 6: δ_0=0, K=2 and (n,T)=(81,30)



λ_0



γ_0



ρ_0