6-2022

Spatial panel data models: Unbalanced panel, threshold effect and network structure

Xiaoyu MENG
Singapore Management University, xiaoyu.meng.2017@phdecons.smu.edu.sg

SPATIAL PANEL DATA MODELS: UNBALANCED PANEL, THRESHOLD EFFECT AND NETWORK STRUCTURE

By

XIAOYU MENG

A DISSERTATION

In

ECONOMICS

Presented to the Singapore Management University in Partial Fulfilment

of the Requirements for the Degree of PhD in Economics

2022

____

Supervisor of Dissertation

____

PhD in Economics, Programme Director

SPATIAL PANEL DATA MODELS: UNBALANCED PANEL, THRESHOLD EFFECT AND NETWORK STRUCTURE

by
Xiaoyu Meng

Submitted to the School of Economics in Partial Fulfilment of the
Requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Zhenlin Yang (Supervisor/Chair)
Professor of Economics & Statistics
Singapore Management University

Yichong Zhang
Assistant Professor of Economics
Singapore Management University

Sainan Jin
Associate Professor of Economics
Singapore Management University
1 Jul 2008 to 30 Jun 2015

Professor of Economics
Singapore Management University
1 Jul 2015 to 31 Dec 2021

Professor of Economics
Tsinghua University
(From Jan 2022 onwards)

Xiaoyi Han
Associate Professor of Economics
Xiamen University

Singapore Management University
2022

Spatial Panel Data Models: Unbalanced Panel, Threshold Effect and Network Structure

Xiaoyu Meng

Abstract

This thesis studies the estimation and inference problems for spatial panel data models when the panels are unbalanced, when the panels contain threshold effects, or when the panels contain time-varying network structures. These three scenarios divide the thesis naturally into three chapters.

The first chapter considers estimation and inferences for fixed effects spatial panel data models based on unbalanced panels that result from randomly missing spatial units. The unbalanced nature of the panel data renders the standard method of estimation inapplicable. In this chapter, we proposed an M-estimation method where the estimating functions are obtained by adjusting the concentrated quasi scores to account for the estimation of fixed effects and/or the presence of unknown spatiotemporal heteroscedasticity. The method allows for general time-varying spatial weight matrices without row-normalization, and is able to give full control of the individual and time specific effects for all the spatial units involved in the data. Consistency and asymptotic normality of the proposed estimators are established. Inference methods are introduced and their consistency is proved. Monte Carlo results show excellent finite sample performance of the proposed methods. An empirical application is presented on commodity tax competition among US states.

The second chapter introduces general estimation and inference methods for threshold spatial panel data models with two-way fixed effects (2FE) in a diminishing-threshold-effects framework. A valid objective function is first obtained by a simple adjustment on the concentrated quasi loglikelihood with 2FE being concentrated out, which leads to a consistent estimation of all common parameters including the threshold parameter. We then show that the estimation of threshold parameter has an asymptotically negligible effect on

the asymptotic distribution of the other estimators, and thereby lead to valid inference methods for other common parameters after a bias correction. A likelihood ratio test is proposed for statistical inference on the threshold parameter. We also propose a sup-Wald test for the presence of threshold effects, based on an M-estimation method with the estimating functions being obtained by simply adjusting the concentrated quasi-score functions. Monte Carlo results show that the proposed methods perform well in finite samples. An empirical application is presented on age-of-leader effects on political competitions across Chinese cities.

The third chapter considers the specification and estimation of a three-dimensional (3-D) spatial panel data model with time-varying network structures. The model allows for endogenous and exogenous interaction effects, correlation of unobservables, and most importantly group-specific effects that are allowed to interact with the individual and time specific effects. The time-varying network structures provide information on the identification of various interaction effects but also yield time-varying sociomatrices whose row sums may not be constant, which renders the transformation-based quasi maximum likelihood inapplicable. In this chapter, we propose an adjusted quasi score method where the estimating functions are obtained by adjusting the concentrated quasi scores (with fixed effects being concentrated out) to account for the effects of concentration. The method is able to give full control of general specifications of three-way fixed effects. Consistency and asymptotic normality of the proposed estimators are established. Monte Carlo results show excellent finite sample performance of the proposed methods.

Table of Contents

Acknowledgement

Foremost, I would like to express my deepest appreciation to my advisor, Professor Zhenlin Yang, for keeping give me the encouragement and guidance to undertake this research. For the past four years, we have met almost every week and I have always been able to obtain insightful advice and feedback from these meetings. This experience is invaluable for me to not only complete this dissertation but also know how to become a good researcher. I could not have imagined having a better advisor and mentor for my Ph.D. study.

My sincere thanks also go to the rest of my dissertation committee: Professor Sainan Jin, Professor Yichong Zhang, and Professor Xiaoyi Han, for their invaluable comments and feedback on my research.

I would also like to thank my cohorts and friends. It has always been a pleasant journey with you. I still miss the first year when we discussed the problems from each course all together. The time of playing badminton with you also makes this journey more fulfilling.

I also want express thanks to my family for their unconditional love and support. A special thank you to my grandma, who passed away during my Ph.D. study. I miss you forever.

Last but not least, to my fiancée Yang Lan, thank you for making this journey much more meaningful. I am unable to imagine a more memorable thing in my life than pursuing a Ph.D. degree together with you overseas.

Chapter 1

Unbalanced Spatial Panel Data Models with Fixed Effects

1.1 Introduction

The literature on spatial panel data (SPD) models has been fast-growing since Anselin (1988), due to the fact that the SPD models are able to take into account the spatial interaction effects and control for the unobservable heterogeneity. Most of the works on SPD models are based on "complete" or "balanced" panels, i.e., a set of observations collected on n spatial units over the entire T periods in time (e.g., Baltagi et al., 2003; Lee and Yu, 2010; Baltagi and Yang, 2013a,b; Yang et al., 2016; Liu and Yang, 2020, to mention a few); only a few on "incomplete" or "unbalanced" panels (Wang and Lee, 2013b; Egger et al., 2005; Baltagi et al., 2007; Baltagi et al., 2015). This is in stark contrast to the usual panel model literature, which contains a sizable portion of works on unbalanced panels (e.g., Wansbeek and Kapteyn, 1989; Baltagi and Chang, 1994; Davis, 2001; Baltagi et al., 2001; Antweiler, 2001; Baltagi and Song, 2006; Bai et al., 2015; Wooldridge, 2019, among others), textbook treatments (Baltagi, 2013; Hsiao, 2014; Greene, 2018), and software implementations (`STATA`, `SAS`, and `R`).

Unbalanced panels are likely to be the norm in typical economic empirical settings (Baltagi and Song, 2006), so are the unbalanced spatial panels.

Unbalancedness may be the result of "randomly missing" observations such as early drop-outs, late entrants and lack of economic activities, or "nonrandomly missing" observations such as attrition and sample selection (Baltagi and Song, 2006; Baltagi, 2013, Ch. 9). The key difference between the two missing mechanisms is that in the former analyses can simply be done based on the actual observed data, but in the latter "imputation" may be necessary before formal analyses. Under the random missing mechanism, most of the methods and techniques developed for balanced panels can be adapted to suit the unbalanced panels, but these may not be true or cannot be easily done for spatial panels. Under the nonrandom missing mechanism, the treatments become much more complicated for both regular and spatial panels, in particular the latter.

The limited literature on unbalanced spatial panels contains three interesting empirical studies under the randomly missing mechanism (RMM). Baltagi et al. (2007) studied the third-country effects on foreign direct investment (FDI) based on an unbalanced SPD model with only spatial error effects. Egger et al. (2005) studied US state tax competition based on an unbalanced SPD model with both spatial lag and spatial error. Both papers focus on random effects model and adapt the GMM approach of Kapoor et al. (2007). There are no theoretical studies being given on the properties of these methods and no formal considerations being given on the models with fixed effects. Baltagi et al. (2015) studied hedonic housing prices based on an unbalanced spatial lag pseudo-panel data model with nested random effects by adapting the ML approach of Antweiler (2001). The only two theoretical studies in this literature are Wang and Lee (2013b) and Zhou et al. (2022). In the former study, SPD models with (correlated) random effects are considered where missing data occur only on the response variable. Although the fixed effects model was also treated in their appendix, it requires the spatial weights matrices to be time-invariant that is clearly not satisfied by the unbalanced SPD models. In the latter work, an autoregressive panel data model with spatially correlated error terms is studied under the missing at random (MAR) assumption in the sense that the missing of the response does not depend on the response itself. Related works but under spatial cross-sectional setup include Kelejian and Prucha (2010), LeSage and Pace (2004), Wang and Lee (2013a) and Zhou et al.(2017). Many important and common issues remain for the unbalanced SPD models even under the simpler random missing mechanism, such as fixed effects, heteroscedasticity (spatial and temporal), and serial correlation. It is therefore highly desirable to develop general estimation and inference methods to address these issues.

In this chapter, we consider the unbalanced SPD models with RMM. In a spatial panel framework, by RMM we mean specifically "randomly missing spatial units" in the sense that the spatial units not present in the t-th time period did not make impacts on their 'neighbors' at that time so that analyses can simply be done based on the observed spatial units and their spatial interactions. The popular transformation method (Lee and Yu, 2010) cannot be applied to handle the fixed effects due to the fact that spatial weight matrices are time-varying, and may not be row-normalizable (Liu and Lee, 2010). The heteroscedasticity-robust method of Liu and Yang (2020) cannot be applied either due to a similar reason. The GMM-type methods of Moscone and Tosetti (2011) and Badinger and Egger (2015) cannot be easily adapted either, besides the issues on efficiency, time fixed effects, and incidental parameters problem. Allowing serial correlation in the error term is interesting but has not been considered. We focus on the unbalanced SPD models with both unit- and time-specific fixed effects, where the errors can be homoscedastic or heteroscedastic of unknown form in both cross-sectional and time dimensions, leaving the issue of serial correlation to the extensions section.

We propose a general M-, or *adjusted quasi score* (AQS), estimation method, for estimating the unbalanced SPD models. The method starts from the joint quasi score functions of both the common parameters and fixed effects, then concentrates out the fixed effects to give the concentrated quasi score functions, and then adjusts these concentrated score functions to give a set of unbiased estimating functions for the common parameters – the AQS functions. Solving the AQS equations gives the M-estimators that are shown to be consistent and asymptotically unbiased. We first consider an FE-SPD model with both spatial lag and spatial error effects under homoscedasticity to fix the main ideas behind the proposed methodology. Then, we make a full extension of the methods to allow for unknown heteroscedasticity in the errors across both space and time. For this, a new way of adjusting the concentrated quasi score functions is required to make them robust against the unknown heteroscedasticity. Consistency and asymptotic normality of all these proposed estimators are established. Simple methods of inference are introduced under both homoscedastic and heteroscedastic errors. Monte Carlo results show excellent finite sample performance of the proposed methods. The proposed methods are simple and reliable, and yet quite general, having great extensibility for extra features in the model (e.g., serial correlation and time-varying coefficients), and for different types of models (e.g., models with random effects and interactive fixed effects). Last, an empirical application on commodity tax competition among US states is demonstrated.

The rest of this chapter is organized as follows. Section 1.2 introduces the M-estimation method for estimating an unbalance SPD model with two-way FE under homoscedasticity, studies the consistency and asymptotic normality of the M-estimators, and presents a simple method for standard errors estimation. Section 1.3 makes a full extension of the M-estimation in Section 1.2 by allowing the errors to be heteroscedastic across both space and time. Section 1.4 presents Monte Carlo simulation findings. The empirical application is given in Section 1.5. Finally, Section 1.6 concludes. Proofs of the main results are given in Appendices B and C.

1.2 Unbalanced FE-SPD Model with Homoscedasticity

1.2.1 The model

Consider a study that lasts T periods and involves a total of n spatial units. At time t, only n_t of these n spatial units are available to give observations on their responses and explanatory variables, and the rest are not due to random missing, e.g., early drop-outs, late entries, lack of economic activities, etc., as discussed in the introduction. These spatial units are interconnected with their 'connectivity' changing over time; they typically vary in size, causing the error distributions to be heteroscedastic; and certain unit- and time-specific features may not be observed but must be acknowledged. These give rise to a spatial panel data (SPD) model with unbalanced panels, time-varying spatial weight matrices, unknown heteroscedasticity, and unit- and time-specific fixed effects (FE):

$$Y_t = \lambda_0 W_t Y_t + X_t \beta_0 + D_t \mu_0 + \alpha_{t0} l_{n_t} + U_t, \quad U_t = \rho_0 M_t U_t + V_t, \;\; t = 1, \ldots, T, \tag{1.1}$$

where Y_t is a vector of observations on n_t spatial units at time t, X_t is an $n_t \times k$ matrices containing values of k time-varying exogenous regressors, and $U_t = (u_{1t}, u_{2t}, \ldots, u_{n_t t})'$ and $V_t = (v_{1t}, v_{2t}, \ldots, v_{n_t t})'$ are $n_t \times 1$ vectors of disturbances and idiosyncratic errors, respectively. W_t and M_t are given $n_t \times n_t$ spatial weight matrices. λ_0 and ρ_0 are spatial coefficients, which together with W_t and M_t characterize the spatial lag (SL) effects and the spatial error (SE) effects, respectively. Spatial Durbin terms, $W_t X_t$, can be added. However, there might be overfitting identification problem if the model contains all three spatial effects (Anselin et al., 2008; Lee and Yu, 2016). β_0 is a $k \times 1$ vector of regression coefficients. $\mu_0 = \{\mu_{i0}\}_{i=1}^n$ denotes an $n \times 1$ vector of unit-specific effects and $\alpha_0 = \{\alpha_{t0}\}_{t=1}^T$ a $T \times 1$ vector of time-specific effects. D_t is an $n_t \times n$ 'selection' matrix obtained from the $n \times n$ identity matrix I_n by deleting its rows that correspond to the missing units at time t, and l_{n_t} is an $n_t \times 1$ vector of ones.

Both μ_0 and α_0 are allowed to correlate with the time-varying regressors in an arbitrary manner and hence are considered as fixed effects. When the change in W_t and M_t is due only to the missing spatial units, they can be written as $W_t = D_t W D_t'$ and $M_t = D_t M D_t'$, where W and M are the spatial weight matrices for all the n spatial units involved in the study. The idiosyncratic errors $\{v_{it}\}$ are first treated as independent and identically distributed (iid) across i and over t, and then extended to be independent but not identically distributed (inid).

An important advantage of the modeling strategy of (1.1) is that it allows the full control of the unobserved heterogeneity of all n spatial units, as long as each of the n spatial units is observed at least twice over the entire period of study so that all the n units remain in the model after the fixed effects being concentrated out. Moreover, the spatial weight matrices W_t and M_t are not necessarily row-normalized, and they are allowed to be generally time-varying, catering to both the random-missing mechanism and the genuine time-varying features.

Some generic notations and conventions will be followed. For a square matrix, $|\cdot|$ denotes its determinant and $\mathtt{tr}(\cdot)$ its trace. For a real symmetric matrix, $\gamma_{\min}(\cdot)$ and $\gamma_{\max}(\cdot)$ denote its smallest and largest eigenvalues. For a real $n \times m$ matrix A, A' denotes its transpose, $\|A\|_F$ its Frobenius norm, $\|A\|_1$ its maximum column sum norm, $\|A\|_\infty$ its maximum row sum norm, and $A^\circ = A + A'$. For a real $n \times m$ matrix A with a full column rank, $\mathbb{P}_A = A(A'A)^{-1}A'$ and $\mathbb{Q}_A = I_n - \mathbb{P}_A$ are the two orthogonal projection matrices. The operator $\mathtt{diag}(\cdot)$ forms a diagonal matrix by diagonal elements of a square matrix or elements of a given vector, $\mathtt{diagv}(\cdot)$ forms a column vector using diagonal elements of a square matrix, and $\mathtt{blkdiag}(\cdots)$ forms a block-diagonal matrix by the given submatrices. The usual expectation and variance operators, $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$, correspond to true parameter values with a subscript 0.

1.2.2 Quasi-Maximum likelihood estimation

Define $\mathbf{W} = \mathtt{blkdiag}(W_1, \ldots, W_T)$, $\mathbf{M} = \mathtt{blkdiag}(M_1, \ldots, M_T)$, $\mathbf{D}_\mu = (D_1', \ldots, D_T')'$, and $\mathbf{D}_\alpha = \mathtt{blkdiag}(l_{n_1}, \ldots, l_{n_T})$. Denote $N = \sum_{t=1}^T n_t$, $\mathbf{Y} = (Y_1', \ldots, Y_T')'$, $\mathbf{X} = (X_1', \ldots, X_T')'$, $\mathbf{U} = (U_1', \ldots, U_T')'$, and $\mathbf{V} = (V_1', \ldots, V_T')'$. Model (1.1) is written in the matrix form: $\mathbf{Y} = \lambda_0 \mathbf{W}\mathbf{Y} + \mathbf{X}\beta_0 + \mathbf{D}_\mu \mu_0 + \mathbf{D}_\alpha \alpha_0 + \mathbf{U}$ and $\mathbf{U} = \rho_0 \mathbf{M}\mathbf{U} + \mathbf{V}$. The existing method of estimating an SPD model with fixed effects is to apply orthogonal transformations to wipe out the fixed effects so that the transformed model remains in the same spatial structure and the (quasi) likelihood can be formed (see, e.g., Lee and Yu, 2010; Yang et al., 2016). This method requires that the panel is balanced, spatial weight matrices are time-invariant and row-normalized, and idiosyncratic errors are homoscedastic. However, none of these is met in the current model specification. To overcome this difficulty, we start with the quasi maximum likelihood (QML) method that estimates the common parameters and the fixed effects together. To eliminate the effects of estimating the fixed effects on the estimation of the common parameters, we in next subsection modify the quasi score functions to produce a set of unbiased and consistent estimating equations. For QML estimation, first note that there are $n+T$ fixed effects parameters but only $n+T-1$ of them are identifiable. A zero-sum constraint is put on the $\alpha_t's$ and the QML estimation is based on the following model form:

$$\mathbf{Y} = \lambda_0 \mathbf{W}\mathbf{Y} + \mathbf{X}\beta_0 + \mathbf{D}_\mu \mu_0 + \mathbf{D}_\alpha^\star \alpha_0^\star + \mathbf{U}, \quad \mathbf{U} = \rho_0 \mathbf{M}\mathbf{U} + \mathbf{V}. \tag{1.2}$$

where $\alpha_0^\star = (\alpha_{20}^\star, \ldots, \alpha_{T0}^\star)'$, and $\mathbf{D}_\alpha^\star = [-l_{n_1} l'_{T-1};\ \texttt{blkdiag}(l_{n_2}, \ldots, l_{n_T})]$.

Denote the set of *common* parameters by $\theta = (\beta', \sigma_v^2, \delta')'$ where $\delta = (\lambda, \rho)'$, and the set of *incidental* parameters by $\phi = (\mu', \alpha^{\star\prime})'$. Define $\mathbf{A}_N(\lambda) = I_N - \lambda \mathbf{W}$ and $\mathbf{B}_N(\rho) = I_N - \rho \mathbf{M}$. We have the quasi Gaussian loglikelihood function of θ and ϕ:

$$\ell_N(\theta, \phi) = -\tfrac{N}{2} \ln 2\pi - \tfrac{N}{2} \ln \sigma_v^2 + \ln |\mathbf{A}_N(\lambda)| + \ln |\mathbf{B}_N(\rho)| - \tfrac{1}{2\sigma_v^2} \mathbf{V}'(\beta, \delta, \phi)\mathbf{V}(\beta, \delta, \phi), \tag{1.3}$$

where $\mathbf{V}(\beta, \delta, \phi) = \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta - \mathbf{D}\phi]$, and $\mathbf{D} = [\mathbf{D}_\mu,\ \mathbf{D}_\alpha^\star]$.

Let $\mathbb{D}(\rho) = \mathbf{B}_N(\rho)\mathbf{D}$. Given θ, $\ell_N(\theta, \phi)$ is partially maximized at

$$\hat{\phi}_N(\beta, \delta) = [\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1} \mathbb{D}'(\rho) \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]. \tag{1.4}$$

Substituting $\hat{\phi}_N(\beta, \delta)$ into $\ell_N(\theta, \phi)$ gives the concentrated quasi loglikelihood function for θ:

$$\ell_N^c(\theta) = -\tfrac{N}{2} \ln 2\pi - \tfrac{N}{2} \ln \sigma_v^2 + \ln |\mathbf{A}_N(\lambda)| + \ln |\mathbf{B}_N(\rho)| - \tfrac{1}{2\sigma_v^2} \tilde{\mathbf{V}}'(\beta, \delta)\tilde{\mathbf{V}}(\beta, \delta), \tag{1.5}$$

where $\tilde{\mathbf{V}}(\beta, \delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$ and $\mathbb{Q}_{\mathbb{D}}(\rho)$ is the projection matrix based on $\mathbb{D}(\rho)$. The direct quasi maximum likelihood (QML) estimator $\hat{\theta}_{\texttt{QML}}$ of θ maximizes $\ell_N^c(\theta)$.

However, such a direct estimation of the common parameters θ completely ignores the impact from the estimation of the incidental parameters ϕ, rendering $\hat{\theta}_{\texttt{QML}}$ be inconsistent or asymptotically biased – the well known *incidental parameters problem* of Neyman and Scott (1948). In their study for a balanced FE-SPD model, Lee and Yu (2010) show that the direct QMLEs of β and δ are consistent no matter T is large or small, but their distributions are asymptotically centered only when T is small relative to n. They further show that the QMLE of σ_v^2 is inconsistent and its limiting distribution is degenerate due to the incidental parameters problem when T is finite. Therefore, if the direct QML approach were followed, a bias correction needs to be done to remove the asymptotic bias for valid statistical inferences, which needs one additional condition that $\frac{T}{n^3} \to 0$. To overcome these problems, Lee and Yu (2010) propose a transformation approach to wipe out the fixed effects, taking advantage of the panel being balanced and spatial weight matrices being time-invariant and row-normalized. In our model specification, none of these features holds and the transformation approach fails to work. Therefore, an alternative (and more general) approach is highly desirable.

1.2.3 The M-estimation

The root cause of inconsistency or asymptotic bias for the direct QML estimation is that a necessary condition for consistency of QML estimators, $\text{plim}\frac{1}{N}S_N^c(\theta_0) = 0$, is violated due to the concentration/estimation of the incidental parameters μ and α, where θ_0 denotes the true value of the parameter vector θ, and $S_N^c(\theta) = \frac{\partial}{\partial\theta}\ell_N^c(\theta)$ is a set of the concentrated quasi score (CQS) functions given as (see Appendix B)

$$S_N^c(\theta) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma_v^4}[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - N\sigma_v^2], \\ \frac{1}{\sigma_v^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{F}_N(\lambda)], \\ \frac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{G}_N(\rho)], \end{cases} \tag{1.6}$$

where $\mathbf{F}_N(\lambda) = \mathbf{W}\mathbf{A}_N^{-1}(\lambda)$ and $\mathbf{G}_N(\rho) = \mathbf{M}\mathbf{B}_N^{-1}(\rho)$.

Under mild conditions, maximizing $\ell_N^c(\theta)$ is equivalent to solving $S_N^c(\theta) = 0$. It is easy to show that at the true value θ_0 of θ,

$$\mathrm{E}[S_N^c(\theta_0)] = \begin{cases} 0_k, \\ -\frac{n+T-1}{2\sigma_{v0}^2}, \\ \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho_0)\mathbf{B}_N(\rho_0)\mathbf{F}_N(\lambda_0)\mathbf{B}_N^{-1}(\rho_0)] - \mathtt{tr}[\mathbf{F}_N(\lambda_0)], \\ \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho_0)\mathbf{G}_N(\rho_0)] - \mathtt{tr}[\mathbf{G}_N(\rho_0)], \end{cases} \tag{1.7}$$

and that $\lim_{N\to\infty} \frac{1}{N}\mathrm{E}[S_N^{\mathrm{c}}(\theta_0)] \neq 0$ with a fixed T. This implies that $\operatorname{plim}_{N\to\infty} \frac{1}{N} S_N^{\mathrm{c}}(\theta_0) \neq 0$, and therefore $\hat{\theta}_{\texttt{QML}}$ cannot be consistent when T is fixed. When T goes large with n, consistency can be achieved but one can show that the limiting distribution of $\sqrt{N}(\hat{\theta}_{\texttt{QML}} - \theta_0)$ is a non-centered normal, suggesting that $\hat{\theta}_N$ has a bias of order $\frac{1}{\sqrt{N}}$.

Note that $\mathrm{E}[S_N^{\mathrm{c}}(\theta_0)]$ depends only on the common parameters θ_0 and the observables. It therefore offers a feasible way to analytically correct the CQS functions to give a set of unbiased estimating functions, or the *adjusted quasi score* (AQS) functions, as $S_N^*(\theta_0) = S_N^{\mathrm{c}}(\theta_0) - \mathrm{E}[S_N^{\mathrm{c}}(\theta_0)]$, which takes the form at the general θ:

$$S_N^*(\theta) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma_v^4}\left[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - N_1\sigma_v^2\right], \\ \frac{1}{\sigma_v^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)], \\ \frac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)], \end{cases} \tag{1.8}$$

where $N_1 = N - n - T + 1$, the *effective* sample size after taking into account the estimation of fixed effects. Solving the AQS equations: $S_N^*(\theta) = 0$, gives the M-estimator of θ, i.e., $\hat{\theta}_N^* = \arg\{S_N^*(\theta) = 0\}$. It is easy to verity that $\mathrm{E}[S_N^*(\theta_0)] = 0$ and $\operatorname{plim}\frac{1}{N}S_N^*(\theta_0) = 0$, making it possible for $\hat{\theta}_N^*$ to be $\sqrt{N_1}$-consistent with a proper limiting distribution.

The M-estimation falls in spirit to the "*Modified Equations of Maximum Likelihood*" of Neyman and Scott (1948, Sec. 5), in searching for a potential method to handle the incidental parameters problem. Its generality and versatility in dealing with the incidental parameters problems have been demonstrated by recent works: Baltagi and Yang (2013a,b), Liu and Yang (2015, 2020), Yang (2018), Li and Yang (2020, 2021) and Xu and Yang (2020). Clearly, this approach falls into the M-estimation method, and it is also a method of moments under the '*just identified*' scenario. Therefore, the resulting estimator is also called the MM estimator. Our approach offers a special way of finding the 'right set' of estimating equations or moment conditions. In the special case of a balanced panel with time-invariant and row-normalized spatial weight matrices, our M-estimation is equivalent to the QML method of

Lee and Yu (2010) based on orthogonal transformations, with effective sample size $N_1 = N - n - T + 1 = (n-1)(T-1)$.

The root-finding process for the M-estimation can be simplified by first solving the equations for β and σ_v^2, giving the constrained M-estimators of β and σ_v^2:

$$\hat{\beta}_N^*(\delta) = [\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbf{C}_N(\delta)\mathbf{Y} \quad \text{and} \quad \hat{\sigma}_{v,N}^{*2}(\delta) = \tfrac{1}{N_1}\hat{\mathbf{V}}'(\delta)\hat{\mathbf{V}}(\delta), \tag{1.9}$$

where $\mathbb{X}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{X}$, $\mathbf{C}_N(\delta) = \mathbf{B}_N(\rho)\mathbf{A}_N(\lambda)$ and $\hat{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\hat{\beta}_N^*(\delta), \delta)$. Substituting $\hat{\beta}_N^*(\delta)$ and $\hat{\sigma}_{v,N}^{*2}(\delta)$ back into (1.8) gives the concentrated AQS functions of δ:

$$S_N^{*c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}_{v,N}^{*2}(\delta)}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\hat{\mathbf{V}}(\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)], \\ \frac{1}{\hat{\sigma}_{v,N}^{*2}(\delta)}\hat{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\hat{\mathbf{V}}(\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)]. \end{cases} \tag{1.10}$$

Solving the concentrated estimating (or AQS) equations, $S_N^{*c}(\delta) = 0$, we obtain the unconstrained M-estimator $\hat{\delta}_N^*$ of δ. Thus the unconstrained M-estimators of β and σ_v^2 are $\hat{\beta}_N^* \equiv \hat{\beta}_N^*(\hat{\delta}_N^*)$ and $\hat{\sigma}_{v,N}^{*2} \equiv \hat{\sigma}_{v,N}^{*2}(\hat{\delta}_N^*)$. The M-estimator of θ is thus $\hat{\theta}_N^* = (\hat{\beta}_N^{*\prime}, \hat{\sigma}_{v,N}^{*2}, \hat{\delta}_N^*)'$.

From the above developments, we see that a big advantage of this method is that it provides a consistent estimation of all parameters including σ_v^2 with the joint asymptotic distribution of the M-estimators being centered as long as N is large. Therefore, all the problems associated with the incidental parameters are gone. Furthermore, we do not have any restriction on the proportion of n and T as they go to infinity, and T (or n) can be even fixed. As this method is based on the adjusted quasi score functions, it may inherit the nice properties from the maximum likelihood estimation. It is well known that ML estimators often have better finite-samples properties than GMM/IV estimators. See also Hsiao (2018) for more discussions on the advantages of the quasi-likelihood approach compared with GMM estimation.

1.2.4 Asymptotic properties of the M-estimators

Denote a parametric quantity evaluated at the true parameter values by dropping its argument(s), e.g., $\mathbf{A}_N \equiv \mathbf{A}_N(\lambda_0)$, $\mathbf{B}_N \equiv \mathbf{B}_N(\rho_0)$, and $\mathbf{C}_N \equiv \mathbf{C}_N(\delta_0)$. Let Δ be the parameter space for δ, and Δ_λ and Δ_ρ be the subspaces for λ and ρ, respectively. Consistency and asymptotic normality of the proposed M-estimators for the unbalanced FE-SPD model are established under the following set of regularity conditions.

Assumption A: *The innovations v_{it} are iid for all i and t with mean zero, variance σ_{v0}^2, and $E|v_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.*

Assumption B: *The space Δ is compact, and the true parameters δ_0 lie in its interior.*

Assumption C: (i) *The elements of $\mathbf{X}$ are non-stochastic and bounded, uniformly in i and t, and* (ii) $\lim_{N\to\infty} \frac{1}{N}\mathbb{X}'(\rho)\mathbb{X}(\rho)$ *exists and is non-singular, uniformly in $\rho \in \Delta_\rho$.*

Assumption D: *$\{W_t\}$ and $\{M_t\}$ are known time-varying matrices. $\mathbf{W}$ and $\mathbf{M}$ are such that* (i) *their elements are at most of uniform order h_n^{-1} such that $\frac{h_n}{n} \to 0$, as $n \to \infty$;* (ii) *their diagonal elements are zero; and* (iii) $\|\mathbf{W}\|_\infty$, $\|\mathbf{W}\|_1$, $\|\mathbf{M}\|_\infty$, *and* $\|\mathbf{M}\|_1$ *are all bounded.*

Assumption E: *For $\mathbb{A}(\varpi) = \mathbf{A}_N(\lambda)$ or $\mathbf{B}_N(\rho)$ with $\varpi = \lambda$ or ρ,*

(i) *both $\|\mathbb{A}^{-1}\|_\infty$ and $\|\mathbb{A}^{-1}\|_1$ are bounded;*

(ii) *either $\|\mathbb{A}^{-1}(\varpi)\|_\infty$ or $\|\mathbb{A}^{-1}(\varpi)\|_1$ is bounded, uniformly in $\varpi \in \Delta_\varpi$;*

(iii) $0 < \underline{c}_\varpi \le \inf_{\varpi\in\Delta_\varpi} \gamma_{\min}[\mathbb{A}'(\varpi)\mathbb{A}(\varpi)] \le \sup_{\varpi\in\Delta_\varpi} \gamma_{\max}[\mathbb{A}'(\varpi)\mathbb{A}(\varpi)] \le \bar{c}_\varpi < \infty$;

(iv) $B_s(\rho)D_s[\frac{1}{T}\sum_{t=1}^T D_t'B_t'(\rho)J_t(\rho)B_t(\rho)D_t]^{-1}D_t'B_t'(\rho)$ *is bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$ for all s and t, where $B_t(\rho) = I_{n_t} - \rho M_t$ for $t = 1,\ldots,T$, and $J_t(\rho) = I_{n_1}$ for $t = 1$, and $I_{n_t} - B_t(\rho)l_{n_t}[l_{n_t}'B_t'(\rho)B_t(\rho)l_{n_t}]^{-1}l_{n_t}'B_t'(\rho)$ for $t = 2,\ldots,T$.*

Assumption F: (i) *n is large (T is large or small),* (ii) *$\forall t$, n_t increases with n in the same rate, and* (iii) *all spatial units are observed at least twice over a total of T periods.*

Assumptions A-E are standard in the spatial econometrics literature (see, e.g., Lee and Yu, 2010; Yang, 2018) except Assumption E(iv). With this additional condition, Lemma A.3 shows that $\|\mathbb{Q}_\mathbb{D}(\rho)\|_1$ and $\|\mathbb{Q}_\mathbb{D}(\rho)\|_\infty$ are bounded uniformly in $\rho \in \Delta_\rho$, which is necessary to facilitate the study of the asymptotic properties of the spatial parameter estimators. Assumption E(iv) is not restrictive as it holds for a special balanced panel (see Appendix B). Assumption F(i) allows (a) both n and T are large and (b) n is large and T is finite. Both scenarios encounter the so-called incidental parameters problem of Neyman and Scott (1948) due to the direct estimation of the individual and time fixed effects. The former leads to the asymptotic bias and the latter the inconsistency in the estimation of the structural parameters. As the usual transformation method is inapplicable to handle the incidental parameters problem in the unbalanced panels, a new (M-estimation) method is therefore introduced. Assumption F(ii) requires that each n_t increases with n, indicating that the number of observed individuals should not be too small relative to n in each period. Assumption F(iii) ensures that the spatial structure is complete after μ is concentrated out.

We first prove the consistency of $\hat{\delta}^*_N$. The key step in the proof is to compare $S^{*c}_N(\delta)$ with its population counterpart. Let $\bar{S}^*_N(\theta) = \mathrm{E}[S^*_N(\theta)]$. Given δ, $\bar{S}^*_N(\theta) = 0$ is partially solved at

$$\bar{\beta}^*_N(\delta) = [\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbf{C}_N(\delta)\mathrm{E}(\mathbf{Y}) \quad \text{and} \quad \bar{\sigma}^{*2}_{v,N}(\delta) = \tfrac{1}{N_1}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\bar{\mathbf{V}}(\delta)], \tag{1.11}$$

where $\bar{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\bar{\beta}^*_N(\delta), \delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\bar{\beta}^*_N(\delta)]$. Substituting $\bar{\beta}^*_N(\delta)$ and $\bar{\sigma}^{*2}_{v,N}(\delta)$ into the δ-component of $\bar{S}^*_N(\theta)$, we obtain the population counterpart of $S^{*c}_N(\delta)$ as

$$\bar{S}^{*c}_N(\delta) = \begin{cases} \frac{1}{\bar{\sigma}^{*2}_{v,N}(\delta)}\mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}'_N(\rho)\bar{\mathbf{V}}(\delta)] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}^{-1}_N(\rho)], \\ \frac{1}{\bar{\sigma}^{*2}_{v,N}(\delta)}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\bar{\mathbf{V}}(\delta)] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)]. \end{cases} \tag{1.12}$$

Clearly, $S^{*c}_N(\hat{\delta}^*_N) = 0$ by construction. Also, it is easy to see that $\bar{S}^{*c}_N(\delta_0) = 0$ as $\bar{\beta}^*_N(\delta_0) = \beta_0$ and $\bar{\sigma}^{*2}_{v,N}(\delta_0) = \sigma^2_{v0}$. Thus, by theorem 5.9 of van der Vaart (1998), $\hat{\delta}^*_N$ will be consistent for δ_0 if $\sup_{\delta\in\Delta}\frac{1}{N_1}\left\|S^{*c}_N(\delta) - \bar{S}^{*c}_N(\delta)\right\| \xrightarrow{p} 0$ and the following identification condition holds:

Assumption G: *$inf_{\delta:d(\delta,\delta_0)\geq\epsilon}\left\|\bar{S}^{*c}_N(\delta)\right\| > 0$ for every $\epsilon > 0$, where $d(\delta,\delta_0)$ is a measure of distance between δ and δ_0.*

Assumption G is a high level assumption being put up for simplicity of presentation. It can be shown to be true under some low level conditions. We have (see (B.5), Appendix B),

$$\bar{\sigma}^{*2}_{v,N}(\delta) = \tfrac{1}{N_1}\eta'\mathbf{A}'^{-1}_N \mathcal{Q}'_N(\delta)\mathcal{Q}_N(\delta)\mathbf{A}^{-1}_N\eta + \tfrac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_N(\delta)],$$

where $\mathcal{C}_N(\delta) = \mathbf{C}_N(\delta)(\mathbf{C}'_N\mathbf{C}_N)^{-1}\mathbf{C}'_N(\delta)$, $\mathcal{Q}_N(\delta) = \mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$, and $\eta = \mathbf{X}\beta_0 + \mathbf{D}\phi_0$. A sufficient condition for Assumption G to hold is either (a) or (b) holds, where

(a) $\frac{1}{\bar{\sigma}^{*2}_{v,N}(\delta)}\eta'\mathbf{F}'_N\mathbf{B}'_N(\rho)\mathcal{Q}_N(\delta)\mathbf{A}^{-1}_N\eta + \mathtt{tr}[\frac{\sigma^2_{v0}}{\bar{\sigma}^{*2}_{v,N}(\delta)}\mathcal{P}_1(\delta) - \mathcal{P}_2(\delta)] \neq 0$, for $\delta \neq \delta_0$,

(b) $\frac{1}{\bar{\sigma}^{*2}_{v,N}(\delta)}\eta'\mathbf{A}'^{-1}_N\mathcal{Q}'_N(\delta)\mathbf{G}_N(\rho)\mathcal{Q}_N(\delta)\mathbf{A}^{-1}_N\eta + \mathtt{tr}[\frac{\sigma^2_{v0}}{\bar{\sigma}^{*2}_{v,N}(\delta)}\mathcal{P}_3(\rho)\mathcal{C}_N(\delta) - \mathcal{P}_3(\rho)] \neq 0$, for $\delta \neq \delta_0$,

with $\mathcal{P}_1(\delta) = \mathbf{C}'^{-1}_N\mathbf{C}'_N(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N\mathbf{B}^{-1}_N$, $\mathcal{P}_2(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}^{-1}_N(\rho)$, and $\mathcal{P}_3(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)$. It is easy to see that $\mathcal{Q}_N(\delta_0)\mathbf{A}^{-1}_N\eta = 0$, $\mathcal{C}_N(\delta_0) = I_N$ and $\bar{\sigma}^{*2}_{v,N}(\delta_0) = \sigma^2_{v0}$. Hence the two quantities in (a) and (b) are identical 0 at the true parameter values. Once the consistency of $\hat{\delta}^*_N$ is established, the consistency of $\hat{\beta}^*_N$ and $\hat{\sigma}^{*2}_{v,N}$ follows by Assumptions C-E.

Theorem 1.1. *Suppose Assumptions A-G hold. We have, as* $N \to \infty$, $\hat{\theta}^*_N \xrightarrow{p} \theta_0$.

To derive the asymptotic distribution of $\hat{\theta}^*_N$, we apply the mean value theorem: $0 = S^*_N(\hat{\theta}^*_N) = S^*_N(\theta_0) + \frac{\partial}{\partial\theta'}S^*_N(\bar{\theta})(\hat{\theta}^*_N - \theta_0)$, where $\bar{\theta}$ lies between $\hat{\theta}^*_N$ and θ_0, and its value varies over the rows of $\frac{\partial}{\partial\theta'}S^*_N(\bar{\theta})$. Using $\tilde{\mathbf{V}}(\beta_0,\delta_0) = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$ and $\mathbf{Y} = \mathbf{A}^{-1}_N(\eta + \mathbf{B}^{-1}_N\mathbf{V})$,

$$S^*_N(\theta_0) = \begin{cases} \frac{1}{\sigma^2_{v0}}\mathbb{X}'\mathbf{V}, \\ \frac{1}{2\sigma^4_{v0}}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathbf{V} - N_1\sigma^2_v), \\ \frac{1}{\sigma^2_{v0}}\mathbf{V}'\mathcal{P}_2\mathbf{B}_N\eta + \frac{1}{\sigma^2_{v0}}\mathbf{V}'\mathcal{P}_2\mathbf{V} - \mathtt{tr}(\mathcal{P}_2), \\ \frac{1}{\sigma^2_{v0}}\mathbf{V}'\mathcal{P}_3\mathbf{V} - \mathtt{tr}(\mathcal{P}_3), \end{cases} \tag{1.13}$$

and its asymptotic normality is proved by the central limit theorem (CLT) for linear-quadratic (LQ) forms of Kelejian and Prucha (2001). This together with the proper asymptotic behavior of the 'Hessian' matrix, $\frac{\partial}{\partial\theta'}S^*_N(\theta)$ (given in (B.4), Appendix B), lead to the following theorem.

Theorem 1.2. *Under Assumptions A-G, we have, as* $N \to \infty$,

$$\sqrt{N_1}\big(\hat{\theta}^*_N - \theta_0\big) \xrightarrow{D} N\Big(0, \lim_{N\to\infty}\Sigma^{*-1}_N(\theta_0)\Gamma^*_N(\theta_0)\Sigma^{*\prime-1}_N(\theta_0)\Big),$$

where $\Sigma^*_N(\theta_0) = -\frac{1}{N_1}\mathrm{E}[\frac{\partial}{\partial\theta'}S^*_N(\theta_0)]$ *and* $\Gamma^*_N(\theta_0) = \frac{1}{N_1}\mathrm{Var}[S^*_N(\theta_0)]$, *both assumed to exist and* $\Sigma^*_N(\theta_0)$ *assumed to be positive definite for sufficiently large* N.

1.2.5 Inference based on M-estimation

To conduct inferences for θ based on the proposed M-estimators, consistent estimates of $\Sigma_N^*(\theta_0)$ and $\Gamma_N^*(\theta_0)$ are needed. The analytical expression of $\Sigma_N^*(\theta)$ can easily be obtained from the Hessian matrix $\frac{\partial}{\partial\theta'}S_N^*(\theta)$ that is given in (B.4). As it depends only on the common parameters θ, a simple plug-in estimator $\Sigma_N^*(\hat{\theta}_N^*)$ can be used to consistently estimate $\Sigma_N^*(\theta_0)$. Alternatively, a simpler sample analogue of $\Sigma_N^*(\theta)$ also provides a consistent estimator:

$$\widehat{\Sigma}_N^* = -\frac{1}{N_1}\frac{\partial}{\partial\theta'}S_N^*(\theta)\Big|_{\theta=\hat{\theta}_N^*}. \tag{1.14}$$

The consistency of $\Sigma_N^*(\hat{\theta}_N^*)$ or $\widehat{\Sigma}_N^*$ is proved in the proof of Theorem 1.2.

Now, using Lemma A.5 with iid errors, one derives $\Gamma_N^*(\theta_0)$, which has the distinct elements:

$$\begin{aligned}
N_1\Gamma_{\beta\theta}^* &= \big[\tfrac{1}{\sigma_{v0}^2}\mathbb{X}'\mathbb{X},\ \ \tfrac{\gamma}{2\sigma_{v0}^3}\mathbb{X}'q,\ \ \tfrac{\gamma}{\sigma_{v0}}\mathbb{X}'p_2 + \tfrac{1}{\sigma_{v0}^2}\mathbb{X}'\mathcal{P}_2\mathbf{B}_N\eta,\ \ \tfrac{\gamma}{2\sigma_{v0}}\mathbb{X}'p_3\big],\\
N_1\Gamma_{\sigma_v^2\sigma_v^2}^* &= \tfrac{1}{4\sigma_{v0}^4}(2N_1 + \kappa q'q),\\
N_1\Gamma_{\sigma_v^2\lambda}^* &= \tfrac{\gamma}{2\sigma_{v0}^3}q'\mathcal{P}_2\mathbf{B}_N\eta + \tfrac{1}{2\sigma_{v0}^2}[2\mathtt{tr}(\mathcal{P}_2\mathbb{Q}_{\mathbb{D}}) + \kappa q'p_2],\\
N_1\Gamma_{\sigma_v^2\rho}^* &= \tfrac{1}{2\sigma_{v0}^2}[2\mathtt{tr}(\mathcal{P}_3\mathbb{Q}_{\mathbb{D}}) + \kappa q'p_3],\\
N_1\Gamma_{\lambda\lambda}^* &= \tfrac{1}{\sigma_{v0}^2}\eta'\mathbf{B}_N'\mathcal{P}_2'\mathcal{P}_2\mathbf{B}_N\eta + \tfrac{2\gamma}{\sigma_{v0}}p_2'\mathcal{P}_2\mathbf{B}_N\eta + \mathtt{tr}(\mathcal{P}_2\mathcal{P}_2^\circ) + \kappa p_2'p_2,\\
N_1\Gamma_{\lambda\rho}^* &= \mathtt{tr}(\mathcal{P}_3\mathcal{P}_2^\circ) + \kappa p_2'p_3 + \tfrac{\gamma}{\sigma_{v0}}p_3'\mathcal{P}_2\mathbf{B}_N\eta,\\
N_1\Gamma_{\rho\rho}^* &= \mathtt{tr}(\mathcal{P}_3\mathcal{P}_3^\circ) + \kappa p_3'p_3,
\end{aligned} \tag{1.15}$$

where $p_r = \mathtt{diagv}(\mathcal{P}_r), r = 2, 3$, and $q = \mathtt{diagv}(\mathbb{Q}_{\mathbb{D}})$. This shows clearly that the estimation of $\Gamma_N^*(\theta_0)$ is more complicated as $\Gamma_N^*(\theta_0)$ contains not only the common parameters θ, but also the fixed effects ϕ embedded in η, and the skewness γ and the excess kurtosis κ of the idiosyncratic errors. Thus, the common plug-in approach may not provide a valid estimate.

Let $\hat{\phi}_N^*$ be the M-estimator of ϕ, obtained through (1.4), i.e., $\hat{\phi}_N^* = \hat{\phi}_N(\hat{\beta}_N^*, \hat{\delta}_N^*)$. Let $\Gamma_N^*(\hat{\theta}_N^*) = \Gamma_N^*(\theta)|_{(\theta=\hat{\theta}_N^*, \phi=\hat{\phi}_N^*, \gamma=\hat{\gamma}_N, \kappa=\hat{\kappa}_N)}$ be the plug-in estimator, where $\hat{\gamma}_N$ and $\hat{\kappa}_N$ areconsistent estimators of γ and κ. When both n and T are large, $\Gamma_N^*(\hat{\theta}_N^*)$ would be consistent as $\hat{\phi}_N^*$ is. However, when n is large but T is fixed, $\hat{\phi}_N^*$ (its component $\hat{\mu}_N^*$) is not consistent. Plugging $\hat{\mu}_N^*$ into $\Gamma_N^*(\theta)$ will induce a bias (inconsistency), and a bias correction is necessary.

From the expression of $\Gamma_N^*(\theta_0)$ given above, we see that only the λ-components involve ϕ through η, which may not be consistently estimated by the plug-in method. We can further show that the components of $\Gamma_N^*(\theta_0)$ linear in ϕ can also be consistently estimated by the plug-in method. Therefore, the only term that cannot be consistently estimated by the plug-in method is $\frac{1}{\sigma_{v0}^2}\eta'\mathbf{B}_N'\mathcal{P}_2'\mathcal{P}_2\mathbf{B}_N\eta$ associated with the λ-λ component of $\Gamma_N^*(\theta_0)$. We have the following corollary. See its proof in Appendix B for details on these discussions.

Corollary 1.1. *Under the assumptions of Theorem 1.2, we have,*

$$\Gamma_N^*(\hat{\theta}_N^*) = \Gamma_N^*(\theta_0) + \text{Bias}^*(\delta_0) + o_p(1),$$

where $\text{Bias}^*(\delta_0)$ *is a* $(k+3)\times(k+3)$ *matrix having zero entries everywhere except the* λ-λ *entry, which takes the form* $\frac{1}{N_1}\texttt{tr}(\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}})$.

The result of Corollary 1.1 leads immediately a general consistent estimator of $\Gamma_N^*(\theta_0)$:

$$\widehat{\Gamma}_N^* = \Gamma_N^*(\hat{\theta}_N^*) - \text{Bias}^*(\hat{\delta}_N^*). \tag{1.16}$$

Then, it is only left to find consistent estimators for γ and κ. Since we cannot 'consistently' estimate $\mathbf{V} = \mathbf{B}_N(\mathbf{A}_N\mathbf{Y}-\eta)$ due to the incidental parameters problem, we start from $\tilde{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$, which can be 'consistently' estimated by $\hat{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}(\hat{\rho}_N^*)\mathbf{B}_N(\hat{\rho}_N^*)[\mathbf{A}_N(\hat{\lambda}_N^*)\mathbf{Y}-\mathbf{X}\hat{\beta}_N^*]$. Let q_{jk} be the (j,k)th element of $\mathbb{Q}_{\mathbb{D}}$. Denote the elements of $\mathbf{V}$ by v_j, and the elements of $\tilde{\mathbf{V}}$ by $\tilde{v}_j, j=1,\ldots,N$, where j is the combined index for $i=1,\ldots,n_t$ and $t=1,\ldots,T$. Then, $\tilde{v}_j = q_{j1}v_1 + q_{j2}v_2 + \cdots + q_{jN}v_N$, and thus,

$$\text{E}(\tilde{v}_j^3) = \textstyle\sum_{k=1}^N q_{jk}^3\text{E}(v_k^3) = \sigma_v^3\gamma\sum_{k=1}^N q_{jk}^3,\ j=1,\ldots,N.$$

Summing $\text{E}(\tilde{v}_j^3)$ over j, we obtain $\gamma = \big(\sum_{j=1}^N \text{E}(\tilde{v}_j^3)\big)\big(\sigma_v^3\sum_{j=1}^N\sum_{k=1}^N q_{jk}^3\big)^{-1}$, and its sample analogue gives a consistent estimator of γ:

$$\hat{\gamma}_N = \frac{\sum_{j=1}^N \hat{v}_j^3}{\hat{\sigma}_{v,N}^{*3}\sum_{j=1}^N\sum_{k=1}^N \hat{q}_{jk}^3}. \tag{1.17}$$

where $\hat{v}_j$ is the jth element of $\hat{\mathbf{V}}(\hat{\beta}_N^*,\hat{\lambda}_N^*)$ and $\hat{q}_{jk}$ is the (j,k)th element of $\mathbb{Q}_{\mathbb{D}}(\hat{\rho}_N^*)$. Similarly,

$$\begin{aligned}\text{E}(\tilde{v}_j^4) &= \textstyle\sum_{k=1}^N q_{jk}^4\text{E}(v_k^4) + 3\sigma_v^4\sum_{k=1}^N\sum_{l=1}^N q_{jk}^2q_{jl}^2 - 3\sigma_v^4\sum_{k=1}^N q_{jk}^4 \\ &= \textstyle\sum_{k=1}^N q_{jk}^4\kappa\sigma_v^4 + 3\sigma_v^4\sum_{k=1}^N\sum_{l=1}^N q_{jk}^2q_{jl}^2,\end{aligned}$$

which gives $\kappa = \big(\sum_{j=1}^N \text{E}(\tilde{v}_j^4) - 3\sigma_v^4\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N q_{jk}^2q_{jl}^2\big)\big(\sigma_v^4\sum_{j=1}^N\sum_{k=1}^N q_{jk}^4\big)^{-1}$

by summing $\mathrm{E}(\tilde{v}_j^4)$ over j. Hence, a consistent estimator for κ is

$$\hat{\kappa}_N = \frac{\sum_{j=1}^N \hat{v}_j^4 - 3\hat{\sigma}_{v,N}^{*4} \sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N \hat{q}_{jk}^2 \hat{q}_{jl}^2}{\hat{\sigma}_{v,N}^{*4} \sum_{j=1}^N \sum_{k=1}^N \hat{q}_{jk}^4}. \tag{1.18}$$

Corollary 1.2. *Under Assumptions A-G, we have, as $N \to \infty$,*

(i) $\hat{\gamma}_N \xrightarrow{p} \gamma_0$ *and* $\hat{\kappa}_N \xrightarrow{p} \kappa_0$; *(ii)* $\widehat{\Sigma}_N^* - \Sigma_N^*(\theta_0) \xrightarrow{p} 0$ *and* $\widehat{\Gamma}_N^* - \Gamma_N^*(\theta_0) \xrightarrow{p} 0$; *and therefore* $\widehat{\Sigma}_N^{*-1}\widehat{\Gamma}_N^*\widehat{\Sigma}_N^{*\prime-1} - \Sigma_N^{*-1}(\theta_0)\Gamma_N^*(\theta_0)\Sigma_N^{*\prime-1}(\theta_0) \xrightarrow{p} 0.$

1.3 Unbalanced FE-SPD Model with Heteroscedasticity

Cross-sectional heteroscedasticity is rather common in spatial regression models due to misspecification, peer interaction, aggregation, clustering, etc. (Anselin, 1988; Liu and Yang, 2015). The same is true for SPD or unbalanced SPD models. Robust methods have been introduced for SPD models, but are limited to balanced panels with cross-sectional heteroscedasticity only (Moscone and Tosetti, 2011; Baltagi and Yang, 2013b; Badinger and Egger, 2015; Liu and Yang, 2020). Time-series heteroscedasticity is also important, in particular in short panels (Alvarez and Arellano, 2004; Bai, 2013). Therefore, it is highly desirable to extend the set of estimation and inference methods introduced in Section 1.2 to allow for unknown spatiotemporal heteroscedasticity as specified in the extended assumption below.

Assumption A′: The innovations v_j (or v_{it}) are independently but not identically distributed (inid), i.e., $\{v_j\} \sim inid(0, \sigma_j^2)$, and $E|v_j|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.

Assumption A′ relaxes Assumption A by allowing the variance of the idiosyncratic error to vary freely across cross-section and over time. As $\mathrm{E}[S_N^*(\theta_0)] \neq 0$ under Assumption A′, we need to readjust score functions (1.6) to make it centered under unknown heteroscedasticity.

1.3.1 M-Estimation under unknown heteroscedasticity

Denote $\mathbf{H} = \mathtt{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2)$, and hence $\mathrm{Var}(\mathbf{V}) = \mathbf{H}$. As in Liu and Yang (2015, 2020), we modify the relevant components of the CQS vector $S_N^{\mathrm{c}}(\theta)$ given in (1.6), so that their expectations at the true parameter θ_0 are zero under unknown heteroscedasticity.

First, consider the stochastic element of the λ-component of $S_N^{\rm c}(\theta)$ given in (1.6). Define $\bar{\mathbf{F}}_N(\delta) = \mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)$, and as usual denote $\bar{\mathbf{F}}_N = \bar{\mathbf{F}}_N(\delta_0)$. Using $\tilde{\mathbf{V}}(\beta_0, \delta_0) = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$ and $\mathbf{B}_N\mathbf{W}\mathbf{Y} = \bar{\mathbf{F}}_N\mathbf{C}_N\mathbf{Y}$, and noting $\mathbf{C}_N\mathbf{Y} = \mathbf{B}_N\eta + \mathbf{V}$ and $\eta = \mathbf{X}\beta_0 + \mathbf{D}\phi_0$, we have,

$$
\begin{aligned}
&\mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'\tilde{\mathbf{V}}(\beta_0, \delta_0)] = \mathrm{E}(\mathbf{Y}'\mathbf{C}_N'\bar{\mathbf{F}}_N'\mathbb{Q}_{\mathbb{D}}\mathbf{V}) = \mathtt{tr}(\mathbf{H}\bar{\mathbf{F}}_N'\mathbb{Q}_{\mathbb{D}}) = \mathtt{tr}[\mathbf{H}\ \mathtt{diag}(\bar{\mathbf{F}}_N'\mathbb{Q}_{\mathbb{D}})] \\
&\quad = \mathtt{tr}[\mathbf{H}\ \mathtt{diag}(\bar{\mathbf{F}}_N'\mathbb{Q}_{\mathbb{D}})\ \mathtt{diag}(\mathbb{Q}_{\mathbb{D}})^{-1}\mathbb{Q}_{\mathbb{D}}] = \mathrm{E}(\mathbf{Y}'\mathbf{C}_N'\bar{\mathbb{F}}_N'\mathbb{Q}_{\mathbb{D}}\mathbf{V}),
\end{aligned}
$$

where $\bar{\mathbb{F}}_N' = \bar{\mathbb{F}}_N'(\delta_0)$ and $\bar{\mathbb{F}}_N'(\delta) = \mathtt{diag}[\bar{\mathbf{F}}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$. Taking the difference between the quantities inside the second expectation and the last expectation, we obtain:

$$
\mathbf{Y}'\mathbf{C}_N'(\delta)[\bar{\mathbf{F}}_N'(\delta) - \bar{\mathbb{F}}_N'(\delta)]\tilde{\mathbf{V}}(\beta, \delta), \tag{1.19}
$$

the adjusted λ-component of the CQS functions, having a zero expectation and a zero probability limit upon dividing by N at θ_0 under unknown heteroscedasticity.

Now, consider the stochastic element of the ρ-component of the CQS vector $S_N^{\rm c}(\theta)$ given in (1.6). Similar to the above, we have,

$$
\begin{aligned}
&\mathrm{E}(\tilde{\mathbf{V}}'\mathbf{G}_N\tilde{\mathbf{V}}) = \mathrm{E}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathbf{G}_N\mathbb{Q}_{\mathbb{D}}\mathbf{V}) = \mathtt{tr}(\mathbf{H}\bar{\mathbf{G}}_N\mathbb{Q}_{\mathbb{D}}) = \mathtt{tr}[\mathbf{H}\ \mathtt{diag}(\bar{\mathbf{G}}_N\mathbb{Q}_{\mathbb{D}})] \\
&\quad = \mathtt{tr}[\mathbf{H}\ \mathtt{diag}(\bar{\mathbf{G}}_N\mathbb{Q}_{\mathbb{D}})\ \mathtt{diag}(\mathbb{Q}_{\mathbb{D}})^{-1}\mathbb{Q}_{\mathbb{D}}] = \mathrm{E}(\mathbf{V}'\bar{\mathbb{G}}_N\mathbb{Q}_{\mathbb{D}}\mathbf{V}),
\end{aligned}
$$

where $\bar{\mathbf{G}}_N(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)$ and $\bar{\mathbb{G}}_N(\rho) = \mathtt{diag}[\bar{\mathbf{G}}_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$. Replacing the $\mathbf{V}'$ in the second and last expectations by $[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)$, and taking the difference between the two quantities inside the expectations, we obtain a robust AQS function for ρ:

$$
[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\tilde{\mathbf{V}}(\beta, \delta). \tag{1.20}
$$

The β-component of $S_N^{\rm c}(\theta)$ is automatically robust against the unknown heteroscedasticity. Therefore, the desired AQS functions robust against the unknown heteroscedasticity $\mathbf{H}$ are,

$$
S_N^{\diamond}(\beta, \delta) = \begin{cases}
\mathbb{X}'(\rho)\tilde{\mathbf{V}}(\beta, \delta), \\
\mathbf{Y}'\mathbf{C}_N'(\delta)[\bar{\mathbf{F}}_N'(\delta) - \bar{\mathbb{F}}_N'(\delta)]\tilde{\mathbf{V}}(\beta, \delta), \\
[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\tilde{\mathbf{V}}(\beta, \delta).
\end{cases} \tag{1.21}
$$

Solving $S_N^{\diamond}(\beta, \delta) = 0$ gives the robust M- (RM-) estimators, $\hat{\beta}_N^{\diamond}$ and $\hat{\delta}_N^{\diamond}$, of β

and δ.

Similarly, this root-finding process can be simplified by first solving for β given δ, to give the constrained estimator $\hat{\beta}^{\diamond}_N(\delta)$ and the concentrated robust AQS functions:

$$S^{\diamond c}_N(\delta) = \begin{cases} \mathbf{Y}'\mathbf{C}'_N(\delta)[\bar{\mathbf{F}}'_N(\delta) - \bar{\mathbb{F}}'_N(\delta)]\hat{\mathbf{V}}(\delta), \\ [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\hat{\beta}^{\diamond}_N(\delta)]'\mathbf{B}'_N(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\hat{\mathbf{V}}(\delta), \end{cases} \tag{1.22}$$

where $\hat{\beta}^{\diamond}_N(\delta) = \hat{\beta}^{*}_N(\delta)$ given in (1.9), and $\hat{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\hat{\beta}^{\diamond}_N(\delta), \delta)$. Then, solving $S^{\diamond c}_N(\delta) = 0$, we obtain the RM-estimator $\hat{\delta}^{\diamond}_N$ of δ, and thus the RM-estimator $\hat{\beta}^{\diamond}_N \equiv \hat{\beta}^{\diamond}_N(\hat{\delta}^{\diamond}_N)$ of β.

1.3.2 Asymptotic properties of the robust M-estimators

Similar to the case of the homoscedastic model, we first establish the consistency of $\hat{\delta}^{\diamond}_N$. Then, the consistency of $\hat{\beta}^{\diamond}_N$ follows. Let $\bar{S}^{\diamond}_N(\beta, \delta) = \mathrm{E}[S^{\diamond}_N(\beta, \delta)]$ be the population robust AQS functions. Then, the β-component of $\bar{S}^{\diamond}_N(\beta, \delta) = 0$ is solved at $\bar{\beta}^{\diamond}_N(\delta) = \bar{\beta}^{*}_N(\delta)$ given in (1.11). Upon substitution, we obtain the population counterpart of $S^{\diamond c}_N(\delta)$:

$$\bar{S}^{\diamond c}_N(\delta) = \begin{cases} \mathrm{E}[\mathbf{Y}'\mathbf{C}'_N(\delta)[\bar{\mathbf{F}}'_N(\delta) - \bar{\mathbb{F}}'_N(\delta)]\bar{\mathbf{V}}(\delta)], \\ \mathrm{E}\big\{\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\bar{\beta}^{\diamond}_N(\delta)]'\mathbf{B}'_N(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\bar{\mathbf{V}}(\delta)\big\}, \end{cases} \tag{1.23}$$

where $\bar{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\bar{\beta}^{\diamond}_N(\delta), \delta)$. As $S^{\diamond c}_N(\hat{\delta}^{\diamond}_N)$ and $\bar{S}^{\diamond c}_N(\delta_0)$ are both zero, by theorem 5.9 of van der Vaart (1998) $\hat{\delta}^{\diamond}_N$ will be consistent for δ_0 if $\sup_{\delta \in \Delta} \frac{1}{N_1} \big\| S^{\diamond c}_N(\delta) - \bar{S}^{\diamond c}_N(\delta)\big\| \xrightarrow{p} 0$ and the following identification condition holds:

Assumption G′: *$inf_{\delta: d(\delta,\delta_0) \geq \epsilon} \big\|\bar{S}^{\diamond c}_N(\delta)\big\| > 0$ for every $\epsilon > 0$, where $d(\delta, \delta_0)$ is a measure of distance between δ and δ_0.*

Again, Assumption G′ is put up for simplicity. More primitive conditions under which Assumption G′ holds are that for $\delta \neq \delta_0$ either of the following conditions holds:

(a) $\eta'\mathbf{A}'^{-1}_N\mathbf{C}'_N(\delta)[\bar{\mathbf{F}}'_N(\delta) - \bar{\mathbb{F}}'_N(\delta)]\mathcal{Q}_N(\delta)\mathbf{A}^{-1}_N\eta + \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}^h_N(\delta)\big(\bar{\mathbf{F}}'_N(\delta) - \bar{\mathbb{F}}'_N(\delta)\big)] \neq 0$; or

(b) $\eta'\mathbf{A}'^{-1}_N\mathbf{C}'_N(\delta)\mathbb{M}'_N(\rho)\big[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)\big]\mathcal{Q}_N(\delta)\mathbf{A}^{-1}_N\eta + \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}^h_N(\delta)\big(\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)\big)] \neq 0$,

where $\mathcal{C}^h_N(\delta) = \mathbf{C}_N(\delta)\mathbf{C}^{-1}_N\mathbf{H}\mathbf{C}^{-1\prime}_N\mathbf{C}'_N(\delta)$ and $\mathbb{M}_N(\rho) = \mathbf{I}_N - \mathbf{B}_N(\rho)\mathbf{X}[\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}$

$\mathbb{X}'(\rho)$. Similarly, as $\mathcal{C}_N^h(\delta_0) = \mathbf{H}$ and $\mathcal{Q}_N(\delta_0)\mathbf{A}_N^{-1}\eta = 0$, the two quantities in (a) and (b) are 0 at δ_0.

Denote $\xi = (\beta', \delta')'$ and $\hat{\xi}_N^\diamond = (\hat{\beta}_N^{\diamond\prime}, \hat{\delta}_N^{\diamond\prime})'$. We have the following consistency theorem.

Theorem 1.3. *Under Assumptions A′, B-F and G′, we have, as $N \to \infty$, $\hat{\xi}_N^\diamond \xrightarrow{p} \xi_0$.*

Similarly, the asymptotic normality of $\hat{\xi}_N^\diamond$ can be established, by applying the mean value theorem to each element of $S_N^\diamond(\hat{\xi}_N^\diamond) = 0$ at ξ_0. The robust AQS function at ξ_0 is

$$S_N^\diamond(\xi_0) = \begin{cases} \mathbb{X}'\mathbf{V}, \\ \eta'\mathbf{B}_N'(\bar{\mathbf{F}}_N' - \bar{\mathbb{F}}_N')\mathbb{Q}_{\mathbb{D}}\mathbf{V} + \mathbf{V}'(\bar{\mathbf{F}}_N' - \bar{\mathbb{F}}_N')\mathbb{Q}_{\mathbb{D}}\mathbf{V}, \\ \phi_0'\mathbb{D}'(\bar{\mathbf{G}}_N - \bar{\mathbb{G}}_N)\mathbb{Q}_{\mathbb{D}}\mathbf{V} + \mathbf{V}'(\bar{\mathbf{G}}_N - \bar{\mathbb{G}}_N)\mathbb{Q}_{\mathbb{D}}\mathbf{V}, \end{cases} \tag{1.24}$$

which can also be verified to be asymptotically normal by using the CLT for LQ forms of Kelejian and Prucha (2001). The adjusted Hessian $\frac{\partial}{\partial \xi'} S_N^\diamond(\bar{\xi})$, shown in (C.1) in Appendix C, has a proper asymptotic behavior, for some $\bar{\xi}$ lying between $\hat{\xi}_N^\diamond$ and ξ_0 elementwise. Consequently, the asymptotic distribution for $\hat{\xi}_N^\diamond$ can be established in the following theorem.

Theorem 1.4. *Under the assumptions of Theorem 1.3, we have, as $N \to \infty$,*

$$\sqrt{N_1}\big(\hat{\xi}_N^\diamond - \xi_0\big) \xrightarrow{D} N\Big(0, \lim_{N\to\infty} \Sigma_N^{\diamond-1}(\xi_0)\Gamma_N^\diamond(\xi_0)\Sigma_N^{\diamond\prime-1}(\xi_0)\Big),$$

where $\Sigma_N^\diamond(\xi_0) = -\frac{1}{N_1}\mathrm{E}\big[\frac{\partial}{\partial \xi'} S_N^\diamond(\xi_0)\big]$ and $\Gamma_N^\diamond(\xi_0) = \frac{1}{N_1}\mathrm{Var}\big[S_N^\diamond(\xi_0)\big]$, both assumed to exist and $\Sigma_N^\diamond(\xi_0)$ assumed to be positive definite for sufficiently large N.

1.3.3 Heteroscedasticity robust inferences

Robust inferences for ξ_0 depends on the availability of consistent estimators of $\Sigma_N^\diamond(\xi_0)$ and $\Gamma_N^\diamond(\xi_0)$. Similar to the case of homoscedastic model, $\Sigma_N^\diamond(\xi_0)$ can be estimated by its observed counterpart $\widehat{\Sigma}_N^\diamond = -\frac{1}{N_1}\frac{\partial}{\partial \xi'} S_N^\diamond(\xi)|_{\xi=\hat{\xi}_N^\diamond}$, with detailed expression of $\frac{\partial}{\partial \xi'} S_N^\diamond(\xi)$ being given in (C.1), Appendix C. The consistency of $\widehat{\Sigma}_N^\diamond$ is proved in the proof of Theorem 1.4.

However, the VC matrix $\Gamma_N^\diamond(\xi_0)$ involves the common parameters ξ_0, the fixed effects ϕ_0, and the unknown $\mathbf{H}$, as seen from its distinct elements derived by Lemma A.5:

$$\begin{aligned}
N_1\Gamma^{\diamond}_{\beta\xi} &= [\mathbb{X}'\mathbf{H}\mathbb{X},\ \ \mathbb{X}'\mathbf{H}\mathbb{L}_\lambda\mathbf{B}_N\eta,\ \ \mathbb{X}'\mathbf{H}\mathbb{L}_\rho\mathbb{D}\phi_0],\\
N_1\Gamma^{\diamond}_{\lambda\lambda} &= \eta'\mathbf{B}_N'\mathbb{L}_\lambda'\mathbf{H}\mathbb{L}_\lambda\mathbf{B}_N\eta + \mathtt{tr}(\mathbf{H}\mathbb{L}_\lambda\mathbf{H}\mathbb{L}_\lambda^\circ),\\
N_1\Gamma^{\diamond}_{\lambda\rho} &= \eta'\mathbf{B}_N'\mathbb{L}_\lambda'\mathbf{H}\mathbb{L}_\rho\mathbb{D}\phi_0 + \mathtt{tr}(\mathbf{H}\mathbb{L}_\lambda\mathbf{H}\mathbb{L}_\rho^\circ),\\
N_1\Gamma^{\diamond}_{\rho\rho} &= \phi_0'\mathbb{D}'\mathbb{L}_\rho'\mathbf{H}\mathbb{L}_\rho\mathbb{D}\phi_0 + \mathtt{tr}(\mathbf{H}\mathbb{L}_\rho\mathbf{H}\mathbb{L}_\rho^\circ),
\end{aligned} \tag{1.25}$$

where $\mathbb{L}_\lambda(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)[\bar{\mathbf{F}}_N(\delta) - \bar{\mathbb{F}}_N(\delta)]$ and $\mathbb{L}_\rho(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)[\bar{\mathbf{G}}_N'(\rho) - \bar{\mathbb{G}}_N'(\rho)]$. This makes the estimation of $\Gamma^{\diamond}_N(\xi_0)$ more challenging than the case of homoscedastic model as the number of unknown elements (parameters) in ϕ and $\mathbf{H}$ both grow with the sample size N (a more serious incidental parameters problem). A nice feature of the analytical expression of $\Gamma^{\diamond}_N(\xi_0)$ is that it does not involve 3rd and 4th moments of the errors due to the fact that the key matrices, $\mathbb{L}_\lambda(\delta)$ and $\mathbb{L}_\rho(\delta)$, have zero diagonals. This makes it possible to adopt again the approach of 'plug-in' and 'bias-correction' as in the case of homoscedastic model.

To facilitate the discussion, write $\Gamma^{\diamond}_N(\xi_0)$ as $\Gamma^{\diamond}_N(\xi_0, \phi, \mathbf{H})$. Let $\hat{\phi}^{\diamond}_N$ be the estimator of ϕ obtained from the RM-estimator $\hat{\xi}^{\diamond}_N$ through (1.4). We then define $\Gamma^{\diamond}_N(\hat{\xi}^{\diamond}_N, \hat{\phi}^{\diamond}_N, \mathbf{H})$ as the plug-in estimator of $\Gamma^{\diamond}_N(\xi_0)$ for a given $\mathbf{H}$. We have (similar to Corollary 1.1) the following corollary.

Corollary 1.3. *Under the assumptions of Theorem 1.4, we have,*

$$\Gamma^{\diamond}_N(\hat{\xi}^{\diamond}_N, \hat{\phi}^{\diamond}_N, \mathbf{H}) = \Gamma^{\diamond}_N(\xi_0) + \text{Bias}^{\diamond}_\phi(\delta_0, \mathbf{H}) + o_p(1),$$

where $\text{Bias}^{\diamond}_\phi(\delta_0, \mathbf{H})$ *is a* $(k+2)\times(k+2)$ *matrix with all the* β*-related entries being zero and the* δ *entry of the elements:* $\frac{1}{N_1}\mathtt{tr}(\mathbf{H}\mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_{\mathbb{D}})$*, for* $a, b = \lambda, \rho$*.*

To estimate $\mathbf{H}$ and thus to give a full estimate of $\Gamma^{\diamond}_N(\xi_0, \phi, \mathbf{H})$, note that $\tilde{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$, which can be 'consistently' estimated by $\hat{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}(\hat{\rho}^{\diamond}_N)\mathbf{B}_N(\hat{\rho}^{\diamond}_N)$ $[\mathbf{A}_N(\hat{\lambda}^{\diamond}_N)\mathbf{Y} - \mathbf{X}\hat{\beta}^{\diamond}_N]$. Note also that

$$\text{E}(\tilde{\mathbf{V}} \odot \tilde{\mathbf{V}}) = [\mathbb{Q}_{\mathbb{D}} \odot \mathbb{Q}_{\mathbb{D}}](\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)',$$

where $\odot$ denotes the Hadamard (elementwise) product. A natural set of estimates of the heteroscedasticity parameters $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$ is therefore given as follows:

$$(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_N^2)' = [\mathbb{Q}_{\mathbb{D}}(\hat{\rho}^{\diamond}_N) \odot \mathbb{Q}_{\mathbb{D}}(\hat{\rho}^{\diamond}_N)]^{-}(\hat{\mathbf{V}} \odot \hat{\mathbf{V}}),$$

where $[\ \cdot\ ]^{-}$ denotes a generalized inverse. Therefore, an estimate of $\mathbf{H}$ is $\widehat{\mathbf{H}} = \mathtt{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_N^2)$.

To 'see' the invertibility of $\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)$, we have, $\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho) = I_N - 2I_N \odot \mathbb{P}_{\mathbb{D}}(\rho) + \mathbb{P}_{\mathbb{D}}(\rho) \odot \mathbb{P}_{\mathbb{D}}(\rho)$. By Schur product theorem, the last term is positive semi-definite. In addition, when T is not too small, $I_N - 2I_N \odot \mathbb{P}_{\mathbb{D}}(\rho)$ is positive definite, because the diagonal elements of $\mathbb{P}_{\mathbb{D}}(\rho)$ are of order $O_p(1/T)$ (See the proof of Lemma A.3). Thus, $\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)$ is typically invertible, for ρ in a neighborhood of ρ_0, which is assumed to facilitate the proof of theoretical results. In practice, however, one may just use the generalized inverse of $\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)$.

From (1.25), we see that the elements of $\Gamma(\xi_0, \phi, \mathbf{H})$ take either of the forms: $\mathtt{tr}(\mathbf{H}C_N)$ and $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N)$. It is important to know the effects of replacing $\mathbf{H}$ by $\hat{\mathbf{H}}$ in these two forms.

Lemma 1.1. *Assume $\Pi_N(\rho) = [\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$ exists for ρ in a neighborhood of ρ_0, and is bounded in both row and column sum norms. Let $A_N = [a_{ij}]$ and $B_N = [b_{ij}]$ be square matrices of dimension N with zero diagonals and bounded row and column sum norms. Let $C_N = [c_{ij}]$ be an $N \times N$ matrix with diagonal elements being uniformly bounded. We have,*

(i) $\frac{1}{N}\mathtt{tr}(\widehat{\mathbf{H}}C_N) - \frac{1}{N}\mathtt{tr}(\mathbf{H}C_N) = o_p(1)$,

(ii) $\frac{1}{N}\mathtt{tr}(\widehat{\mathbf{H}}A_N\widehat{\mathbf{H}}B_N) - \frac{2}{N}\mathtt{tr}((A_N \odot B_N)\Pi_N\Lambda(\mathbf{H})\Pi_N) - \frac{1}{N}\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N) = o_p(1)$,

where $\Pi_N = \Pi_N(\rho_0)$, $\Lambda(\mathbf{H}) = \{(q_j'\mathbf{H}q_k)^2\}_{j,k=1}^N$, and q_j' is the jth row of $\mathbb{Q}_{\mathbb{D}}$.

The assumptions on $\Pi_N(\rho)$ in Lemma 1.1 always hold for a balanced panel (See Appendix C). The bias term in Corollary 1.3 needs a further correction when $\mathbf{H}$ is replaced by $\hat{\mathbf{H}}$ as it contains elements of the form $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N)$ with diagonal elements of A_N and B_N not strictly zero. However, the effect of non-zero diagonals is shown to be negligible due to the existence of a lower ranked matrix $\mathbb{P}_{\mathbb{D}}$ and its orthogonality with $\mathbb{Q}_{\mathbb{D}}$. Combining the results of Corollary 1.3 and Lemma 1.1, we have the full estimate of $\Gamma_N^{\diamond}(\xi_0)$:

$$\widehat{\Gamma}_N^{\diamond} = \Gamma_N^{\diamond}(\hat{\xi}_N^{\diamond}, \hat{\phi}_N^{\diamond}, \hat{\mathbf{H}}) - \mathrm{Bias}_{\phi}^{\diamond}(\hat{\delta}_N^{\diamond}, \widehat{\mathbf{H}}) - \mathrm{Bias}_{\mathbf{H}}^{\diamond}(\hat{\delta}_N^{\diamond}, \widehat{\mathbf{H}}), \tag{1.26}$$

where $\mathrm{Bias}_{\mathbf{H}}^{\diamond}(\delta_0, \mathbf{H})$ has entries 0, or $\frac{2}{N_1}\mathtt{tr}((\mathbb{L}_a \odot \mathbb{L}_b^{\circ} - \mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}})\Pi_N\Lambda(\mathbf{H})\Pi_N)$,

$a, b = \lambda, \rho$.

Corollary 1.4. *Under the assumptions of Theorem 1.4, we have as $N \to \infty$,*

$$\widehat{\Sigma}_N^{\diamond} - \Sigma_N^{\diamond}(\xi_0) \xrightarrow{p} 0 \quad \textit{and} \quad \widehat{\Gamma}_N^{\diamond} - \Gamma_N^{\diamond}(\xi_0) \xrightarrow{p} 0,$$

and therefore, $\widehat{\Sigma}_N^{\diamond -1}\widehat{\Gamma}_N^{\diamond}\widehat{\Sigma}_N^{\diamond\prime -1} - \Sigma_N^{\diamond -1}(\xi_0)\Gamma_N^{\diamond}(\xi_0)\Sigma_N^{\diamond\prime -1}(\xi_0) \xrightarrow{p} 0.$

1.4 Monte Carlo Study

Extensive Monte Carlo experiments are carried out to investigate the finite sample performance of the proposed M-estimators, the RM-estimators, and the corresponding standard error estimators of the unbalanced SPD models with two-way fixed effects. To see the effectiveness of the adjustments on the concentrated quasi scores in controlling the effects of estimating the fixed effects, we also include the direct QML estimators in the Monte Carlo study. We choose different values of n and T, and fix the percentage of randomly missing observations at 10%, and make sure that each individual is observed at least twice over the entire period. We consider two data generating processes: unbalanced FE-SPD models with `SL` and `SE` effects or with `SL` and `SD` (spatial Durbin) effects:

$$\texttt{SL-SE Model}: \; Y_t = \lambda W_t Y_t + X_t\beta_1 + D_t\mu + \alpha_t l_{n_t} + U_t, \quad U_t = \rho M_t U_t + V_t, \tag{1.27}$$

$$\texttt{SL-SD Model}: \; Y_t = \lambda W_t Y_t + X_t\beta_1 + W_t X_t \beta_2 + D_t\mu + \alpha_t l_{n_t} + V_t, \tag{1.28}$$

for $t = 1, \ldots, T$. Note that we consider Durbin effects, $W_t X_t$, only in the `SL` model due to the identification issue mentioned earlier. We choose $\beta_1 = 1$, $\beta_2 = 0$ or 0.5, $\lambda = 0.2$ and $\rho = 0.2$. Generate $X_t's$ independently from $N(0, 2^2 I_n)$, and set the individual effects $\mu = \frac{1}{T}\Sigma_{t=1}^T X_t + e$, where $e \sim N(0, I_n)$. Then, omit the "missing" elements of X_t. The time fixed effects α are generated from $N(0, I_T)$. The error (v_{it}) distributions can be (i) normal, (ii) normal mixture (10% $N(0, 4^2)$ and 90% $N(0, 1)$), or (iii) chi-square with 3 degrees of freedom. For the purpose of comparison, we set $\sigma_{v0}^2 = 1$ for homoscedastic case, and set the average of error variances in the heteroscedastic case to 1. Monte Carlo (empirical) means and standard deviations (shown in the brackets) are reported for QMLE, M-estimation and RM-estimation. Further, empirical averages of the standard error estimates (shown in the square brackets) are also reported for M-estimation and RM-estimation, based on the robust VC matrix estimates, $\widehat{\Sigma}_N^{*-1}\widehat{\Gamma}_N^{*}\widehat{\Sigma}_N^{*\prime-1}$ for the M-estimation and $\widehat{\Sigma}_N^{\diamond-1}\widehat{\Gamma}_N^{\diamond}\widehat{\Sigma}_N^{\diamond\prime-1}$ for the RM-estimation. The number of Monte Carlo runs is 1000.

The spatial weights W_t and M_t are first generated as time-varying $n \times n$ matrices according to `rook` contiguity, `queen` contiguity, or `group` interaction scheme, and then their rows and columns corresponding to the missing spatial units are deleted. The groups' sizes in the `group` interaction scheme can be either increasing or fixed as n increases. In the former, we let $K(n) = Round(n^{0.5})$ be the number of groups and then generate $K(n)$ group sizes according to a uniform distribution, and in the latter, we start with six groups of sizes (3,5,7,9,11,15) and then replicate to give a n to be multiples of 50. See Yang (2015) for details in generating these spatial layouts. In the latter case, the variation in group sizes does not shrink to zero as n increases. As a result, the M-estimation would not be consistent under heteroscedasticity (Liu and Yang, 2015, 2020). In this case, the heteroscedasticity is generated as follows: for each group, if the group size is larger than the mean group size, then the variance is set to be the same as the group size, otherwise, the variance is the square of the inverse of the group size (Lin and Lee, 2010).

Tables 1.1a and 1.1b report partial Monte Carlo results for the unbalanced FE-SPD model with `SL` and `SE` effects and homoscedastic errors, for $T = 5$ and 10, respectively. The results show an excellent finite performance of the proposed M-estimation and RM-estimation, as well as their standard error estimators. The proposed M-estimation performs uniformly much better than the QML method in the estimation of σ_v^2, λ and ρ, irrespective of the choices of the spatial weight matrices and the values of n and T. Our M-estimators exhibit a good performance even when the sample size is as small as $n = 50$ and $T = 5$, and improve on average when the sample expands, regardless of the error distributions. The $\sqrt{N_1}$-consistency of the M-estimators is clearly demonstrated by the Monte Carlo sds. Moreover, the robust estimates of standard errors $\hat{sd}$'s are on average very close to the corresponding Monte

Carlo standard errors. By comparing the results of M-estimators and RM-estimators, we cannot see which one beats the other in terms of bias and efficiency for these homoscedastic models.

Tables 1.2a and 1.2b present partial Monte Carlo results for the unbalanced FE-SPD model with `SL` and `SD` effects and homoscedastic errors, for $T = 5$ and 10, respectively. The results again show an excellent performance of the proposed set of estimation and inference methods. As in the case of the `SL-SE` model, the M-estimation and RM-estimation give quite similar results, and both show a clear convergence as sample size increases. Their corresponding standard error estimates also perform very well. In contrast, the QMLE can perform poorly.

Tables 1.3a and 1.3b report partial Monte Carlo results for the unbalanced FE-SPD model with `SL` and `SE` effects and heteroscedastic errors, for $T = 5$ and 10, respectively. The results show an excellent finite sample performance of the proposed RM-estimation and its estimated standard error. In contrast, the QMLE and M-estimation typically provide worse estimates for spatial parameters than RM-estimation. Our RM-estimators perform well even when sample size is quite small, and show convergence to the true value as sample size increases. In addition, $\widehat{sd}$s are very closed to sds for our RM-estimators, consistent with our theoretical expectation.

Tables 1.4 presents partial Monte Carlo results for the unbalanced FE-SPD model with `SL` and `SD` effects and heteroscedastic errors, for $T = 5$ and 10, respectively. The weight matrix is specified as group interaction with a fixed group sizes scheme. We can see a much better finite sample performance for our RM-estimation than QMLE and M-estimation, and the corresponding standard error estimates also have a good performance.

1.5 An Empirical Application

In this section, we present an empirical study to analyze horizontal competition in excise taxes on gasoline, cigarettes, and beers among US states. According to theoretical studies by Kanbur and Keen (1993) and Nielsen (2001), the tax competition is usually induced by cross-border shopping, which causes

Table 1.1a. Empirical mean(sd)[$\hat{sd}$] of the estimators for FE-SPD model with `SL-SE` effects, 10% random missing, **homoscedasticity**, $(\beta_1, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, **T=5**.

	W= Rook, M=Queen			W=Group-I, M=Queen		
	QMLE	M-Est	RM-Est	QMLE	M-Est	RM-Est
			$n = 50$; error = 1, 2, 3, for the three panels below			
β_1	.9998(.039)	1.0007(.039)[.039]	1.0007(.039)[.039]	.9976(.038)	.9986(.038)[.038]	.9986(.038)[.038]
λ	.1848(.063)	.1999(.063)[.062]	.1999(.063)[.062]	.1666(.077)	.1885(.075)[.075]	.1887(.076)[.074]
ρ	.1112(.152)	.1868(.146)[.148]	.1867(.146)[.146]	.1101(.147)	.1889(.141)[.150]	.1889(.141)[.147]
σ_v^2	.7394(.083)	.9829(.110)[.107]	–	.7390(.082)	.9828(.109)[.106]	–
β_1	.9981(.038)	.9989(.038)[.039]	.9989(.038)[.038]	.9980(.038)	.9989(.038)[.038]	.9989(.038)[.038]
λ	.1849(.061)	.1998(.061)[.062]	.1999(.061)[.060]	.1689(.076)	.1909(.074)[.074]	.1909(.074)[.072]
ρ	.1179(.149)	.1933(.143)[.148]	.1932(.144)[.140]	.1121(.148)	.1915(.143)[.150]	.1913(.143)[.143]
σ_v^2	.7358(.172)	.9780(.228)[.215]	–	.7420(.176)	.9867(.234)[.218]	–
β_1	.9981(.038)	.9990(.038)[.039]	.9990(.038)[.038]	.9980(.038)	.9990(.037)[.038]	.9990(.037)[.038]
λ	.1825(.061)	.1976(.061)[.062]	.1976(.061)[.061]	.1688(.078)	.1907(.076)[.074]	.1908(.076)[.073]
ρ	.1165(.150)	.1919(.144)[.148]	.1917(.144)[.143]	.1104(.150)	.1894(.145)[.150]	.1894(.145)[.146]
σ_v^2	.7421(.128)	.9864(.169)[.161]	–	.7380(.129)	.9814(.171)[.161]	–
			$n = 100$; error = 1, 2, 3, for the three panels below			
β_1	1.0010(.027)	1.0011(.026)[.027]	1.0011(.026)[.027]	1.0015(.029)	1.0009(.029)[.029]	1.0009(.029)[.028]
λ	.1922(.043)	.1993(.043)[.042]	.1994(.043)[.042]	.1842(.055)	.1960(.055)[.054]	.1961(.055)[.053]
ρ	.1565(.099)	.1906(.096)[.100]	.1906(.096)[.099]	.1626(.104)	.1954(.101)[.099]	.1954(.101)[.098]
σ_v^2	.7617(.060)	.9942(.078)[.076]	–	.7604(.058)	.9928(.076)[.076]	–
β_1	.9993(.028)	.9994(.028)[.027]	.9994(.028)[.027]	1.0015(.029)	1.0009(.029)[.029]	1.0009(.029)[.028]
λ	.1923(.042)	.1994(.042)[.042]	.1994(.042)[.042]	.1829(.055)	.1948(.054)[.054]	.1948(.054)[.053]
ρ	.1623(.102)	.1962(.099)[.099]	.1962(.099)[.096]	.1588(.100)	.1917(.097)[.099]	.1916(.097)[.097]
σ_v^2	.7624(.128)	.9951(.167)[.160]	–	.7674(.128)	1.0019(.167)[.161]	–
β_1	.9983(.027)	.9984(.027)[.027]	.9984(.027)[.027]	1.0000(.028)	.9994(.028)[.029]	.9994(.028)[.028]
λ	.1937(.043)	.2009(.043)[.042]	.2009(.043)[.042]	.1831(.056)	.1950(.055)[.054]	.1950(.055)[.053]
ρ	.1621(.100)	.1961(.097)[.099]	.1961(.097)[.098]	.1599(.098)	.1928(.095)[.099]	.1929(.096)[.097]
σ_v^2	.7625(.092)	.9951(.120)[.118]	–	.7636(.091)	.9970(.118)[.118]	–
			$n = 200$; error = 1, 2, 3, for the three panels below			
β_1	1.0002(.019)	1.0001(.019)[.019]	1.0001(.019)[.019]	1.0002(.020)	1.0001(.020)[.020]	1.0001(.020)[.019]
λ	.1964(.028)	.1998(.028)[.029]	.1998(.028)[.029]	.1856(.049)	.1955(.049)[.048]	.1955(.049)[.048]
ρ	.1805(.071)	.1947(.069)[.068]	.1948(.069)[.068]	.1829(.069)	.1970(.068)[.068]	.1970(.068)[.068]
σ_v^2	.7703(.042)	.9958(.054)[.053]	–	.7708(.040)	.9966(.052)[.053]	–
β_1	.9997(.019)	.9996(.019)[.019]	.9996(.019)[.019]	1.0001(.020)	.9999(.020)[.020]	.9999(.020)[.019]
λ	.1969(.029)	.2003(.029)[.029]	.2003(.029)[.028]	.1851(.049)	.1950(.049)[.048]	.1950(.049)[.048]
ρ	.1850(.069)	.1991(.067)[.068]	.1991(.067)[.067]	.1864(.068)	.2004(.066)[.068]	.2004(.066)[.067]
σ_v^2	.7679(.089)	.9927(.115)[.114]	–	.7701(.091)	.9956(.118)[.114]	–
β_1	1.0007(.019)	1.0006(.019)[.019]	1.0006(.019)[.019]	1.0002(.019)	1.0000(.019)[.020]	1.0000(.019)[.020]
λ	.1968(.028)	.2002(.028)[.029]	.2002(.028)[.029]	.1861(.049)	.1960(.048)[.048]	.1960(.048)[.048]
ρ	.1840(.069)	.1981(.067)[.068]	.1981(.067)[.067]	.1832(.070)	.1973(.068)[.068]	.1973(.068)[.067]
σ_v^2	.7688(.063)	.9939(.082)[.083]	–	.7736(.066)	1.0002(.085)[.085]	–
			$n = 400$; error = 1, 2, 3, for the three panels below			
β_1	1.0003(.014)	1.0003(.014)[.013]	1.0003(.014)[.013]	1.0003(.013)	1.0003(.013)[.013]	1.0003(.013)[.013]
λ	.1985(.019)	.2001(.019)[.019]	.2001(.019)[.019]	.1875(.041)	.1949(.040)[.042]	.1949(.040)[.042]
ρ	.1936(.049)	.1982(.048)[.047]	.1982(.048)[.047]	.1953(.049)	.1999(.048)[.047]	.1999(.048)[.047]
σ_v^2	.7738(.028)	.9966(.036)[.038]	–	.7734(.029)	.9961(.037)[.038]	–
β_1	1.0001(.013)	1.0000(.013)[.013]	1.0000(.013)[.013]	1.0007(.013)	1.0007(.013)[.013]	1.0007(.013)[.013]
λ	.1985(.019)	.2001(.019)[.020]	.2001(.019)[.019]	.1899(.041)	.1972(.041)[.042]	.1972(.041)[.042]
ρ	.1937(.048)	.1983(.047)[.048]	.1983(.047)[.047]	.1922(.048)	.1969(.047)[.047]	.1969(.047)[.047]
σ_v^2	.7782(.063)	1.0023(.081)[.082]	–	.7767(.062)	1.0004(.080)[.082]	–
β_1	1.0001(.013)	1.0001(.013)[.013]	1.0001(.013)[.013]	.9999(.013)	.9999(.013)[.013]	.9999(.013)[.013]
λ	.1972(.020)	.1988(.020)[.020]	.1987(.020)[.019]	.1921(.042)	.1994(.041)[.042]	.1994(.041)[.042]
ρ	.1944(.050)	.1990(.049)[.047]	.1990(.049)[.047]	.1924(.049)	.1970(.048)[.047]	.1970(.048)[.047]
σ_v^2	.7743(.049)	.9973(.063)[.060]	–	.7729(.046)	.9955(.059)[.060]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 1.1b. Empirical mean(sd)[$\hat{sd}$] of the estimators for FE-SPD model with `SL-SE` effects, 10% random missing, **homoscedasticity**, $(\beta_1, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, **T=10**.

	W=Rook, M=Queen			W=Group-I, M=Queen		
	QMLE	M-Est	RM-Est	QMLE	M-Est	RM-Est
	$n = 50$; error = 1, 2, 3, for the three panels below					
β_1	1.0019(.026)	1.0008(.026)[.026]	1.0008(.026)[.026]	.9993(.025)	.9989(.025)[.026]	.9989(.025)[.026]
λ	.1820(.039)	.1976(.039)[.040]	.1976(.039)[.040]	.1780(.046)	.1971(.045)[.045]	.1971(.046)[.045]
ρ	.1239(.093)	.1974(.091)[.091]	.1974(.092)[.091]	.1210(.093)	.1927(.092)[.091]	.1927(.092)[.091]
σ_v^2	.8641(.062)	.9930(.071)[.071]	–	.8641(.063)	.9936(.072)[.071]	–
β_1	1.0006(.026)	.9995(.026)[.026]	.9995(.026)[.026]	.9986(.025)	.9982(.025)[.026]	.9982(.025)[.025]
λ	.1849(.039)	.2004(.039)[.040]	.2004(.039)[.039]	.1779(.047)	.1968(.046)[.045]	.1969(.047)[.045]
ρ	.1203(.093)	.1941(.091)[.092]	.1940(.091)[.089]	.1235(.093)	.1951(.091)[.091]	.1950(.091)[.089]
σ_v^2	.8625(.144)	.9912(.166)[.156]	–	.8641(.138)	.9935(.158)[.156]	–
β_1	1.0019(.026)	1.0008(.026)[.026]	1.0008(.026)[.026]	1.0003(.026)	.9999(.026)[.026]	.9999(.026)[.026]
λ	.1819(.040)	.1976(.040)[.040]	.1976(.040)[.040]	.1771(.046)	.1960(.045)[.045]	.1961(.045)[.045]
ρ	.1194(.094)	.1931(.093)[.091]	.1931(.093)[.090]	.1211(.093)	.1928(.091)[.091]	.1928(.092)[.090]
σ_v^2	.8667(.105)	.9962(.121)[.113]	–	.8615(.101)	.9906(.116)[.113]	–
	$n = 100$; error = 1, 2, 3, for the three panels below					
β_1	1.0001(.018)	.9997(.018)[.018]	.9997(.018)[.018]	1.0004(.018)	1.0004(.018)[.017]	1.0004(.018)[.017]
λ	.1924(.027)	.1993(.027)[.027]	.1993(.027)[.027]	.1817(.040)	.1963(.039)[.039]	.1964(.039)[.039]
ρ	.1600(.063)	.1952(.062)[.063]	.1952(.062)[.063]	.1638(.064)	.1988(.063)[.063]	.1989(.063)[.063]
σ_v^2	.8792(.044)	.9986(.050)[.050]	–	.8787(.046)	.9981(.052)[.050]	–
β_1	1.0005(.018)	1.0000(.018)[.018]	1.0000(.018)[.018]	1.0001(.018)	1.0000(.018)[.017]	1.0000(.018)[.017]
λ	.1932(.027)	.2001(.027)[.027]	.2000(.027)[.027]	.1838(.040)	.1983(.040)[.039]	.1983(.040)[.039]
ρ	.1634(.062)	.1985(.061)[.063]	.1985(.061)[.062]	.1601(.063)	.1952(.062)[.063]	.1952(.062)[.062]
σ_v^2	.8773(.102)	.9964(.116)[.112]	–	.8780(.101)	.9973(.115)[.113]	–
β_1	1.0005(.018)	1.0001(.018)[.018]	1.0001(.018)[.018]	.9998(.018)	.9998(.018)[.017]	.9998(.018)[.017]
λ	.1923(.027)	.1992(.027)[.027]	.1992(.027)[.027]	.1834(.041)	.1979(.040)[.039]	.1979(.040)[.039]
ρ	.1609(.064)	.1961(.063)[.063]	.1961(.063)[.063]	.1618(.064)	.1969(.063)[.063]	.1969(.063)[.062]
σ_v^2	.8782(.073)	.9975(.083)[.082]	–	.8763(.072)	.9954(.082)[.082]	–
	$n = 200$; error = 1, 2, 3, for the three panels below					
β_1	1.0004(.013)	1.0001(.013)[.013]	1.0001(.013)[.013]	.9996(.013)	.9996(.013)[.012]	.9996(.013)[.012]
λ	.1961(.018)	.1994(.018)[.019]	.1994(.018)[.019]	.1883(.033)	.1986(.033)[.033]	.1986(.033)[.033]
ρ	.1823(.044)	.1985(.044)[.044]	.1986(.044)[.044]	.1834(.044)	.1997(.043)[.044]	.1997(.043)[.044]
σ_v^2	.8826(.030)	.9973(.034)[.035]	–	.8836(.031)	.9986(.035)[.035]	–
β_1	1.0002(.013)	.9999(.013)[.013]	.9999(.013)[.013]	.9998(.013)	.9997(.013)[.012]	.9997(.013)[.012]
λ	.1960(.018)	.1993(.018)[.019]	.1993(.018)[.019]	.1876(.033)	.1979(.033)[.033]	.1980(.033)[.033]
ρ	.1821(.043)	.1984(.043)[.044]	.1984(.043)[.044]	.1808(.045)	.1972(.044)[.044]	.1972(.044)[.044]
σ_v^2	.8820(.071)	.9967(.080)[.080]	–	.8825(.075)	.9973(.084)[.081]	–
β_1	.9996(.012)	.9993(.012)[.013]	.9993(.012)[.013]	1.0005(.012)	1.0005(.012)[.012]	1.0005(.012)[.012]
λ	.1968(.019)	.2000(.019)[.019]	.2000(.019)[.019]	.1878(.033)	.1981(.033)[.033]	.1982(.033)[.033]
ρ	.1818(.044)	.1980(.043)[.044]	.1980(.043)[.044]	.1834(.046)	.1997(.046)[.044]	.1997(.046)[.044]
σ_v^2	.8842(.053)	.9992(.060)[.058]	–	.8829(.051)	.9978(.057)[.058]	–
	$n = 400$; error = 1, 2, 3, for the three panels below					
β_1	1.0004(.009)	1.0003(.009)[.009]	1.0003(.009)[.009]	1.0001(.009)	1.0000(.009)[.009]	1.0000(.009)[.009]
λ	.1982(.014)	.1998(.014)[.014]	.1999(.014)[.014]	.1922(.027)	.1987(.027)[.026]	.1989(.027)[.026]
ρ	.1918(.033)	.1989(.032)[.031]	.1989(.032)[.031]	.1915(.033)	.1986(.033)[.031]	.1986(.033)[.031]
σ_v^2	.8854(.022)	.9982(.024)[.025]	–	.8853(.022)	.9982(.024)[.025]	–
β_1	.9998(.009)	.9997(.009)[.009]	.9997(.009)[.009]	1.0001(.009)	1.0000(.009)[.009]	1.0000(.009)[.009]
λ	.1983(.013)	.1999(.013)[.014]	.2000(.013)[.014]	.1913(.027)	.1978(.027)[.026]	.1981(.027)[.026]
ρ	.1931(.031)	.2001(.030)[.031]	.2001(.030)[.031]	.1905(.032)	.1976(.032)[.031]	.1976(.032)[.031]
σ_v^2	.8847(.050)	.9974(.056)[.057]	–	.8851(.051)	.9979(.057)[.057]	–
β_1	.9995(.009)	.9994(.009)[.009]	.9994(.009)[.009]	.9997(.009)	.9996(.009)[.009]	.9996(.009)[.009]
λ	.1978(.013)	.1994(.013)[.014]	.1996(.013)[.014]	.1926(.026)	.1991(.026)[.026]	.1993(.026)[.026]
ρ	.1931(.031)	.2002(.031)[.031]	.2002(.031)[.031]	.1907(.032)	.1978(.031)[.031]	.1978(.031)[.031]
σ_v^2	.8873(.038)	1.0004(.043)[.042]	–	.8881(.036)	1.0013(.041)[.042]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 1.2a. Empirical mean(sd)[$\hat{sd}$] of estimators for FE-SPD model with SL-SD effects, 10% random missing, **homoscedasticity**, $(\beta_1, \beta_2, \lambda, \sigma_v^2) = (1, 0.5, 0.2, 1)$, **T=5**.

	W=Queen			W=Group-I		
	QMLE	M-Est	RM-Est	QMLE	M-Est	RM-Est
	$n = 50$; error = 1, 2, 3, for the three panels below					
β_1	1.0056(.041)	.9999(.041)[.041]	.9999(.041)[.041]	1.0130(.043)	1.0060(.043)[.043]	1.0060(.043)[.042]
β_2	.5898(.194)	.5147(.195)[.194]	.5146(.195)[.193]	.6567(.251)	.5636(.241)[.231]	.5637(.242)[.228]
λ	.1276(.125)	.1862(.125)[.123]	.1863(.125)[.122]	.1103(.131)	.1644(.125)[.119]	.1643(.126)[.117]
σ_v^2	.7390(.082)	.9779(.109)[.106]	–	.7398(.081)	.9793(.108)[.106]	–
β_1	1.0081(.040)	1.0024(.040)[.041]	1.0024(.040)[.040]	1.0101(.044)	1.0032(.044)[.042]	1.0031(.044)[.042]
β_2	.5909(.196)	.5158(.197)[.195]	.5156(.197)[.189]	.6426(.238)	.5504(.229)[.229]	.5497(.230)[.220]
λ	.1235(.122)	.1822(.122)[.124]	.1823(.123)[.120]	.1130(.124)	.1668(.119)[.119]	.1673(.120)[.114]
σ_v^2	.7410(.180)	.9806(.238)[.216]	–	.7413(.171)	.9812(.226)[.215]	–
β_1	1.0066(.041)	1.0009(.041)[.041]	1.0009(.041)[.041]	1.0113(.043)	1.0044(.042)[.042]	1.0044(.042)[.042]
β_2	.5904(.192)	.5151(.193)[.195]	.5151(.194)[.192]	.6385(.242)	.5463(.232)[.229]	.5458(.234)[.225]
λ	.1252(.123)	.1840(.123)[.124]	.1841(.123)[.122]	.1183(.126)	.1719(.121)[.118]	.1723(.122)[.116]
σ_v^2	.7436(.128)	.9841(.169)[.160]	–	.7411(.130)	.9809(.172)[.158]	–
	$n = 100$; error = 1, 2, 3, for the three panels below					
β_1	1.0047(.030)	1.0016(.030)[.031]	1.0016(.030)[.031]	1.0038(.027)	1.0012(.027)[.027]	1.0012(.027)[.027]
β_2	.5502(.135)	.5114(.135)[.133]	.5114(.135)[.138]	.5879(.180)	.5307(.175)[.173]	.5306(.176)[.177]
λ	.1621(.085)	.1908(.085)[.084]	.1908(.085)[.088]	.1466(.095)	.1808(.093)[.092]	.1808(.093)[.094]
σ_v^2	.7618(.059)	.9903(.076)[.075]	–	.7630(.058)	.9921(.075)[.075]	–
β_1	1.0053(.031)	1.0021(.031)[.031]	1.0021(.031)[.031]	1.0043(.027)	1.0016(.027)[.027]	1.0016(.027)[.027]
β_2	.5551(.130)	.5163(.130)[.133]	.5162(.130)[.136]	.5956(.189)	.5384(.183)[.173]	.5385(.183)[.175]
λ	.1585(.084)	.1872(.084)[.084]	.1872(.084)[.087]	.1433(.100)	.1775(.097)[.092]	.1774(.097)[.093]
σ_v^2	.7675(.129)	.9977(.168)[.159]	–	.7644(.129)	.9940(.167)[.159]	–
β_1	1.0032(.030)	1.0001(.030)[.030]	1.0001(.030)[.031]	1.0044(.027)	1.0017(.027)[.027]	1.0017(.027)[.027]
β_2	.5535(.136)	.5149(.136)[.133]	.5150(.136)[.136]	.5859(.180)	.5285(.175)[.173]	.5283(.175)[.176]
λ	.1598(.086)	.1884(.085)[.084]	.1883(.085)[.087]	.1465(.096)	.1808(.093)[.092]	.1810(.093)[.094]
σ_v^2	.7616(.091)	.9900(.119)[.116]	–	.7676(.095)	.9981(.123)[.118]	–
	$n = 200$; error = 1, 2, 3, for the three panels below					
β_1	1.0020(.021)	1.0006(.021)[.020]	1.0006(.021)[.021]	1.0027(.020)	1.0011(.020)[.020]	1.0011(.020)[.020]
β_2	.5244(.096)	.5056(.096)[.094]	.5056(.096)[.097]	.5722(.170)	.5257(.165)[.157]	.5257(.165)[.160]
λ	.1824(.058)	.1962(.059)[.057]	.1962(.059)[.060]	.1597(.083)	.1858(.081)[.079]	.1859(.081)[.080]
σ_v^2	.7713(.041)	.9948(.053)[.053]	–	.7726(.041)	.9949(.053)[.053]	–
β_1	1.0013(.020)	.9999(.020)[.020]	.9999(.020)[.021]	1.0026(.020)	1.0011(.020)[.020]	1.0011(.020)[.020]
β_2	.5269(.093)	.5082(.093)[.093]	.5081(.093)[.097]	.5687(.160)	.5224(.156)[.157]	.5222(.156)[.158]
λ	.1808(.057)	.1945(.057)[.057]	.1946(.057)[.060]	.1605(.080)	.1866(.078)[.079]	.1867(.078)[.079]
σ_v^2	.7726(.091)	.9964(.118)[.114]	–	.7740(.089)	.9967(.114)[.113]	–
β_1	1.0017(.020)	1.0002(.020)[.020]	1.0002(.020)[.021]	1.0033(.019)	1.0018(.019)[.020]	1.0018(.019)[.020]
β_2	.5248(.094)	.5060(.094)[.094]	.5060(.094)[.097]	.5759(.164)	.5293(.160)[.158]	.5293(.160)[.159]
λ	.1826(.056)	.1963(.056)[.057]	.1963(.056)[.060]	.1564(.082)	.1827(.080)[.079]	.1827(.080)[.080]
σ_v^2	.7741(.065)	.9984(.083)[.084]	–	.7726(.067)	.9950(.087)[.083]	–
	$n = 400$; error = 1, 2, 3, for the three panels below					
β_1	1.0012(.014)	1.0005(.014)[.014]	1.0005(.014)[.014]	1.0012(.014)	1.0005(.014)[.014]	1.0005(.014)[.014]
β_2	.5120(.065)	.5028(.065)[.064]	.5028(.065)[.067]	.5526(.142)	.5197(.139)[.136]	.5196(.139)[.138]
λ	.1909(.041)	.1980(.041)[.041]	.1980(.041)[.043]	.1702(.069)	.1890(.067)[.067]	.1891(.067)[.068]
σ_v^2	.7761(.030)	.9992(.039)[.038]	–	.7759(.029)	.9989(.038)[.038]	–
β_1	1.0011(.014)	1.0004(.014)[.014]	1.0004(.014)[.014]	1.0010(.014)	1.0003(.014)[.014]	1.0003(.014)[.014]
β_2	.5116(.064)	.5024(.064)[.064]	.5024(.064)[.067]	.5553(.141)	.5224(.139)[.136]	.5225(.139)[.138]
λ	.1904(.042)	.1975(.042)[.041]	.1975(.042)[.043]	.1682(.070)	.1870(.069)[.067]	.1870(.069)[.068]
σ_v^2	.7730(.062)	.9952(.080)[.081]	–	.7756(.064)	.9986(.082)[.082]	–
β_1	1.0010(.014)	1.0003(.014)[.014]	1.0003(.014)[.014]	1.0012(.014)	1.0004(.014)[.014]	1.0004(.014)[.014]
β_2	.5101(.063)	.5009(.063)[.064]	.5009(.063)[.067]	.5551(.140)	.5223(.138)[.136]	.5222(.138)[.138]
λ	.1913(.040)	.1984(.040)[.041]	.1984(.040)[.043]	.1692(.068)	.1879(.066)[.067]	.1880(.067)[.068]
σ_v^2	.7764(.047)	.9996(.061)[.060]	–	.7763(.048)	.9996(.061)[.060]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 1.2b. Empirical mean(sd)[$\hat{sd}$] of estimators for FE-SPD model with `SL-SD` effects, 10% random missing, **homoscedasticity**, $(\beta_1, \beta_2, \lambda, \sigma_v^2) = (1, 0.5, 0.2, 1)$, **T=10**.

	W=Queen			W=Group-I		
	QMLE	M-Est	RM-Est	QMLE	M-Est	RM-Est
			$n = 50$; error = 1, 2, 3, for the three panels below			
β_1	1.0072(.026)	1.0014(.026)[.027]	1.0014(.026)[.027]	1.0083(.029)	1.0016(.029)[.028]	1.0016(.029)[.028]
β_2	.5823(.122)	.5074(.123)[.124]	.5073(.123)[.125]	.6090(.159)	.5238(.152)[.149]	.5237(.152)[.149]
λ	.1368(.078)	.1935(.078)[.080]	.1935(.079)[.081]	.1357(.081)	.1861(.078)[.075]	.1861(.078)[.075]
σ_v^2	.8604(.060)	.9889(.069)[.070]	–	.8607(.062)	.9894(.071)[.071]	–
β_1	1.0069(.027)	1.0011(.027)[.027]	1.0011(.027)[.027]	1.0099(.028)	1.0032(.027)[.028]	1.0032(.027)[.028]
β_2	.5850(.122)	.5104(.123)[.124]	.5104(.123)[.124]	.6074(.156)	.5223(.150)[.149]	.5224(.150)[.148]
λ	.1354(.080)	.1918(.080)[.079]	.1919(.080)[.079]	.1355(.079)	.1858(.075)[.075]	.1857(.075)[.073]
σ_v^2	.8674(.140)	.9970(.161)[.158]	–	.8686(.143)	.9984(.164)[.159]	–
β_1	1.0068(.027)	1.0010(.027)[.027]	1.0010(.027)[.027]	1.0069(.028)	1.0003(.028)[.028]	1.0003(.028)[.028]
β_2	.5827(.124)	.5082(.125)[.124]	.5079(.125)[.125]	.6057(.153)	.5208(.147)[.148]	.5209(.147)[.148]
λ	.1364(.080)	.1929(.080)[.079]	.1931(.080)[.080]	.1370(.078)	.1872(.074)[.074]	.1871(.074)[.074]
σ_v^2	.8607(.105)	.9893(.120)[.113]	–	.8575(.101)	.9857(.116)[.112]	–
			$n = 100$; error = 1, 2, 3, for the three panels below			
β_1	1.0024(.019)	.9997(.019)[.018]	.9997(.019)[.018]	1.0033(.019)	.9999(.019)[.019]	.9999(.019)[.019]
β_2	.5381(.083)	.5025(.083)[.083]	.5025(.083)[.086]	.5815(.135)	.5142(.130)[.127]	.5140(.131)[.128]
λ	.1697(.055)	.1982(.055)[.055]	.1982(.055)[.057]	.1528(.069)	.1927(.067)[.065]	.1928(.067)[.066]
σ_v^2	.8765(.043)	.9956(.049)[.050]	–	.8766(.044)	.9958(.050)[.050]	–
β_1	1.0036(.018)	1.0008(.018)[.018]	1.0008(.018)[.018]	1.0036(.019)	1.0002(.019)[.019]	1.0002(.019)[.019]
β_2	.5403(.083)	.5046(.083)[.083]	.5045(.083)[.086]	.5882(.133)	.5207(.128)[.128]	.5206(.128)[.128]
λ	.1678(.056)	.1963(.056)[.055]	.1963(.056)[.057]	.1480(.069)	.1881(.066)[.066]	.1882(.066)[.066]
σ_v^2	.8771(.101)	.9963(.115)[.113]	–	.8752(.101)	.9942(.114)[.113]	–
β_1	1.0029(.018)	1.0001(.018)[.018]	1.0001(.018)[.018]	1.0045(.019)	1.0011(.019)[.019]	1.0011(.019)[.019]
β_2	.5381(.083)	.5024(.083)[.083]	.5024(.083)[.086]	.5853(.131)	.5179(.127)[.127]	.5179(.126)[.128]
λ	.1686(.055)	.1971(.055)[.055]	.1971(.055)[.057]	.1488(.068)	.1889(.066)[.066]	.1889(.066)[.066]
σ_v^2	.8755(.074)	.9944(.084)[.081]	–	.8755(.075)	.9945(.086)[.081]	–
			$n = 200$; error = 1, 2, 3, for the three panels below			
β_1	1.0017(.013)	1.0003(.013)[.013]	1.0003(.013)[.013]	1.0023(.013)	1.0007(.013)[.013]	1.0007(.013)[.013]
β_2	.5218(.061)	.5034(.061)[.060]	.5034(.061)[.062]	.5592(.109)	.5124(.106)[.104]	.5124(.106)[.106]
λ	.1827(.039)	.1972(.039)[.039]	.1972(.039)[.040]	.1653(.055)	.1926(.054)[.054]	.1925(.054)[.055]
σ_v^2	.8835(.032)	.9979(.036)[.035]	–	.8837(.032)	.9983(.036)[.035]	–
β_1	1.0018(.013)	1.0005(.013)[.013]	1.0005(.013)[.013]	1.0021(.013)	1.0005(.013)[.013]	1.0005(.013)[.013]
β_2	.5202(.060)	.5017(.060)[.060]	.5018(.060)[.062]	.5580(.106)	.5112(.103)[.104]	.5110(.103)[.106]
λ	.1838(.040)	.1982(.040)[.039]	.1982(.040)[.040]	.1661(.055)	.1933(.053)[.054]	.1934(.053)[.055]
σ_v^2	.8816(.073)	.9958(.082)[.080]	–	.8839(.072)	.9985(.082)[.081]	–
β_1	1.0017(.013)	1.0003(.013)[.013]	1.0003(.013)[.013]	1.0019(.013)	1.0004(.013)[.013]	1.0004(.013)[.013]
β_2	.5220(.060)	.5035(.060)[.060]	.5035(.060)[.062]	.5581(.107)	.5113(.104)[.104]	.5112(.104)[.106]
λ	.1834(.039)	.1978(.039)[.039]	.1978(.039)[.040]	.1670(.054)	.1942(.053)[.054]	.1943(.053)[.055]
σ_v^2	.8837(.052)	.9981(.059)[.058]	–	.8844(.051)	.9991(.057)[.058]	–
			$n = 400$; error = 1, 2, 3, for the three panels below			
β_1	1.0007(.009)	1.0000(.009)[.009]	1.0000(.009)[.010]	1.0012(.009)	1.0004(.009)[.009]	1.0004(.009)[.009]
β_2	.5101(.043)	.5004(.043)[.044]	.5004(.043)[.046]	.5476(.100)	.5120(.098)[.091]	.5121(.098)[.093]
λ	.1918(.026)	.1990(.026)[.028]	.1989(.026)[.029]	.1732(.049)	.1935(.048)[.046]	.1934(.048)[.047]
σ_v^2	.8851(.022)	.9980(.025)[.025]	–	.8867(.023)	.9998(.026)[.025]	–
β_1	1.0006(.009)	.9998(.009)[.009]	.9998(.009)[.010]	1.0007(.009)	.9999(.009)[.009]	.9999(.009)[.009]
β_2	.5086(.043)	.4989(.043)[.044]	.5018(.060)[.062]	.5444(.093)	.5088(.091)[.091]	.5088(.091)[.092]
λ	.1934(.026)	.2006(.026)[.028]	.1982(.040)[.040]	.1753(.047)	.1955(.046)[.046]	.1956(.046)[.047]
σ_v^2	.8867(.052)	.9998(.059)[.058]	–	.8853(.051)	.9983(.058)[.057]	–
β_1	1.0009(.009)	1.0002(.009)[.009]	1.0003(.013)[.013]	1.0012(.009)	1.0003(.009)[.009]	1.0003(.009)[.009]
β_2	.5127(.042)	.5030(.042)[.044]	.5035(.060)[.062]	.5489(.094)	.5131(.092)[.092]	.5131(.092)[.093]
λ	.1913(.026)	.1984(.026)[.028]	.1978(.039)[.040]	.1726(.048)	.1930(.047)[.046]	.1930(.047)[.047]
σ_v^2	.8840(.037)	.9967(.042)[.041]	–	.8864(.036)	.9996(.041)[.042]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 1.3a. Empirical mean(sd)[$\hat{sd}$] of estimators for FE-SPD model with `SL-SE` effects, 10% random missing, **heteroscedasticity**, $(\beta_1, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, **T=5**.

	W=M=Group-II			W=Group-II, M=Queen		
	QMLE	M-Est	M-Est	QMLE	M-Est	RM-Est
	$n = 50$; error = 1, 2, 3, for the three panels below					
β_1	1.0001(.042)	1.0001(.042)[.041]	.9993(.042)[.042]	.9974(.039)	.9978(.039)[.040]	.9979(.039)[.039]
λ	.1878(.070)	.1944(.068)[.098]	.1973(.080)[.080]	.1311(.086)	.1735(.083)[.106]	.1894(.089)[.090]
ρ	-.0161(.209)	.0905(.177)[.199]	.1016(.272)[.247]	.1077(.147)	.1845(.141)[.145]	.1889(.141)[.142]
σ_v^2	.7664(.102)	1.0237(.136)[.150]	–	.7717(.102)	1.0264(.136)[.150]	–
β_1	1.0012(.042)	1.0013(.042)[.040]	1.0005(.043)[.041]	.9980(.039)	.9985(.039)[.040]	.9985(.039)[.038]
λ	.1873(.072)	.1938(.069)[.098]	.1956(.082)[.080]	.1343(.083)	.1757(.080)[.105]	.1915(.085)[.088]
ρ	-.0008(.198)	.1036(.168)[.198]	.1235(.248)[.232]	.1086(.146)	.1853(.140)[.146]	.1894(.140)[.135]
σ_v^2	.7606(.217)	1.0154(.290)[.274]	–	.7695(.221)	1.0234(.294)[.277]	–
β_1	.9986(.041)	.9986(.042)[.041]	.9978(.042)[.041]	.9980(.040)	.9984(.040)[.040]	.9985(.040)[.039]
λ	.1843(.070)	.1911(.067)[.099]	.1928(.081)[.080]	.1316(.084)	.1737(.081)[.106]	.1896(.087)[.089]
ρ	-.0072(.205)	.0980(.174)[.198]	.1144(.260)[.238]	.1113(.144)	.1878(.138)[.145]	.1919(.138)[.139]
σ_v^2	.7727(.158)	1.0319(.211)[.212]	–	.7744(.161)	1.0300(.214)[.210]	–
	$n = 100$; error = 1, 2, 3, for the three panels below					
β_1	1.0005(.028)	1.0005(.028)[.028]	1.0002(.028)[.028]	1.0003(.025)	1.0005(.025)[.026]	1.0006(.025)[.026]
λ	.1927(.049)	.1954(.048)[.064]	.1992(.056)[.054]	.1733(.045)	.1849(.045)[.054]	.1960(.046)[.048]
ρ	.0883(.131)	.1304(.120)[.127]	.1573(.167)[.160]	.1665(.095)	.1995(.093)[.095]	.1980(.092)[.094]
σ_v^2	.7572(.071)	.9925(.093)[.105]	–	.7706(.070)	.9985(.090)[.102]	–
β_1	1.0004(.029)	1.0004(.029)[.028]	1.0001(.029)[.028]	.9994(.026)	.9996(.026)[.026]	.9997(.026)[.026]
λ	.1921(.049)	.1948(.048)[.063]	.1985(.055)[.053]	.1736(.045)	.1851(.045)[.054]	.1960(.046)[.047]
ρ	.0884(.129)	.1305(.118)[.128]	.1578(.165)[.157]	.1668(.093)	.1997(.090)[.095]	.1984(.090)[.090]
σ_v^2	.7554(.155)	.9901(.203)[.199]	–	.7648(.153)	.9910(.198)[.194]	–
β_1	.9997(.029)	.9997(.029)[.028]	.9995(.029)[.028]	.9996(.026)	.9998(.026)[.026]	.9999(.026)[.026]
λ	.1914(.049)	.1941(.048)[.063]	.1979(.055)[.054]	.1739(.045)	.1854(.045)[.054]	.1964(.047)[.047]
ρ	.0877(.130)	.1299(.119)[.128]	.1566(.167)[.159]	.1656(.097)	.1985(.095)[.095]	.1971(.094)[.093]
σ_v^2	.7614(.115)	.9979(.150)[.152]	–	.7646(.112)	.9907(.145)[.146]	–
	$n = 200$; error = 1, 2, 3, for the three panels below					
β_1	.9991(.019)	.9991(.019)[.019]	.9990(.019)[.019]	.9987(.020)	.9988(.020)[.019]	.9989(.020)[.019]
λ	.1950(.034)	.1962(.034)[.044]	.2000(.040)[.040]	.1784(.035)	.1853(.034)[.042]	.1980(.036)[.037]
ρ	.1272(.086)	.1441(.082)[.084]	.1763(.112)[.107]	.1862(.071)	.2006(.070)[.068]	.1996(.069)[.069]
σ_v^2	.7657(.050)	.9907(.065)[.073]	–	.7647(.051)	.9883(.065)[.073]	–
β_1	.9996(.019)	.9996(.019)[.019]	.9995(.019)[.019]	.9992(.019)	.9993(.019)[.019]	.9994(.019)[.019]
λ	.1939(.035)	.1951(.035)[.044]	.1984(.041)[.040]	.1773(.035)	.1842(.034)[.042]	.1968(.036)[.037]
ρ	.1345(.083)	.1511(.079)[.083]	.1857(.106)[.105]	.1816(.071)	.1960(.069)[.068]	.1952(.069)[.068]
σ_v^2	.7704(.111)	.9967(.144)[.143]	–	.7660(.110)	.9899(.142)[.141]	–
β_1	.9996(.019)	.9996(.019)[.019]	.9996(.019)[.019]	.9987(.019)	.9988(.019)[.019]	.9990(.019)[.019]
λ	.1947(.035)	.1959(.035)[.044]	.1996(.041)[.040]	.1782(.035)	.1851(.034)[.042]	.1978(.036)[.037]
ρ	.1290(.085)	.1459(.081)[.084]	.1787(.110)[.106]	.1840(.070)	.1984(.069)[.068]	.1975(.068)[.068]
σ_v^2	.7646(.080)	.9893(.104)[.106]	–	.7664(.081)	.9905(.104)[.106]	–
	$n = 400$; error = 1, 2, 3, for the three panels below					
β_1	.9999(.014)	.9999(.014)[.013]	.9999(.014)[.014]	1.0000(.013)	1.0000(.013)[.013]	1.0000(.013)[.013]
λ	.1966(.026)	.1970(.026)[.031]	.1998(.030)[.030]	.1862(.024)	.1894(.024)[.028]	.2002(.025)[.026]
ρ	.1491(.060)	.1550(.058)[.058]	.1892(.075)[.074]	.1945(.047)	.1990(.046)[.047]	.1991(.046)[.047]
σ_v^2	.7849(.034)	1.0110(.044)[.052]	–	.7839(.034)	1.0096(.044)[.052]	–
β_1	.9998(.014)	.9998(.014)[.013]	.9998(.014)[.014]	.9994(.013)	.9994(.013)[.013]	.9994(.013)[.013]
λ	.1968(.027)	.1972(.027)[.031]	.1998(.031)[.030]	.1835(.025)	.1866(.025)[.028]	.1973(.026)[.026]
ρ	.1509(.061)	.1568(.059)[.058]	.1914(.075)[.074]	.1959(.048)	.2003(.047)[.047]	.2004(.047)[.047]
σ_v^2	.7878(.079)	1.0148(.102)[.103]	–	.7842(.080)	1.0100(.103)[.103]	–
β_1	1.0000(.013)	1.0000(.013)[.013]	1.0000(.014)[.014]	1.0005(.014)	1.0005(.014)[.013]	1.0005(.014)[.013]
λ	.1949(.027)	.1953(.027)[.031]	.1980(.031)[.030]	.1861(.024)	.1893(.024)[.028]	.2001(.025)[.026]
ρ	.1500(.059)	.1559(.057)[.058]	.1904(.073)[.074]	.1955(.047)	.1999(.046)[.047]	.2001(.046)[.047]
σ_v^2	.7869(.059)	1.0136(.076)[.078]	–	.7854(.059)	1.0116(.076)[.078]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 1.3b. Empirical mean(sd)[$\hat{sd}$] of estimators for FE-SPD model with `SL-SE` effects, 10% random missing, **heteroscedasticity**, $(\beta_1, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, **T=10**.

	W=M=Group-II			W=Group-II, M=Queen		
	QMLE	M-Est	RM-Est	QMLE	M-Est	RM-Est
	$n = 50$; error = 1, 2, 3, for the three panels below					
β_1	1.0015(.025)	1.0012(.025)[.025]	1.0009(.025)[.024]	1.0024(.026)	1.0007(.026)[.026]	.9999(.026)[.025]
λ	.1899(.039)	.1953(.038)[.054]	.1987(.043)[.042]	.1677(.036)	.1880(.035)[.047]	.1973(.037)[.037]
ρ	.0447(.119)	.1342(.106)[.117]	.1660(.146)[.139]	.1290(.092)	.2001(.091)[.090]	.1988(.091)[.089]
σ_v^2	.8647(.073)	.9940(.083)[.093]	–	.8737(.074)	1.0030(.085)[.094]	–
β_1	.9997(.025)	.9995(.025)[.025]	.9992(.025)[.024]	1.0015(.025)	.9998(.025)[.025]	.9989(.025)[.025]
λ	.1900(.039)	.1954(.037)[.054]	.1990(.042)[.042]	.1686(.035)	.1888(.034)[.047]	.1981(.036)[.037]
ρ	.0425(.118)	.1324(.105)[.119]	.1636(.145)[.137]	.1284(.092)	.1995(.090)[.090]	.1980(.089)[.087]
σ_v^2	.8715(.175)	1.0019(.201)[.192]	–	.8744(.173)	1.0037(.198)[.193]	–
β_1	1.0011(.025)	1.0009(.025)[.025]	1.0006(.025)[.024]	1.0021(.025)	1.0004(.025)[.026]	.9995(.025)[.025]
λ	.1898(.039)	.1952(.037)[.054]	.1987(.042)[.042]	.1677(.035)	.1881(.035)[.047]	.1974(.036)[.037]
ρ	.0461(.115)	.1355(.102)[.117]	.1682(.139)[.138]	.1280(.092)	.1992(.090)[.090]	.1981(.090)[.088]
σ_v^2	.8646(.125)	.9940(.144)[.141]	–	.8751(.123)	1.0047(.141)[.144]	–
	$n = 100$; error = 1, 2, 3, for the three panels below					
β_1	.9994(.018)	.9995(.018)[.018]	.9996(.018)[.018]	.9996(.018)	.9994(.018)[.018]	.9992(.018)[.018]
λ	.1901(.034)	.1936(.034)[.043]	.1980(.040)[.038]	.1760(.034)	.1889(.033)[.038]	.2000(.035)[.035]
ρ	.1077(.082)	.1470(.077)[.080]	.1809(.106)[.101]	.1627(.066)	.1979(.065)[.064]	.1979(.065)[.064]
σ_v^2	.8663(.053)	.9860(.060)[.067]	–	.8717(.054)	.9931(.062)[.067]	–
β_1	.9991(.018)	.9992(.018)[.018]	.9992(.018)[.018]	1.0003(.019)	1.0001(.019)[.018]	.9999(.019)[.018]
λ	.1902(.033)	.1937(.032)[.043]	.1978(.038)[.038]	.1746(.034)	.1875(.033)[.038]	.1985(.035)[.035]
ρ	.1106(.078)	.1497(.073)[.080]	.1848(.101)[.099]	.1605(.066)	.1957(.065)[.064]	.1957(.065)[.063]
σ_v^2	.8659(.124)	.9855(.141)[.139]	–	.8736(.125)	.9953(.143)[.139]	–
β_1	.9999(.018)	1.0000(.018)[.018]	1.0000(.018)[.018]	.9997(.019)	.9995(.019)[.018]	.9993(.019)[.018]
λ	.1922(.033)	.1957(.032)[.043]	.2005(.037)[.038]	.1738(.033)	.1867(.033)[.038]	.1977(.035)[.035]
ρ	.1077(.079)	.1470(.074)[.080]	.1808(.102)[.100]	.1607(.066)	.1958(.065)[.064]	.1958(.064)[.063]
σ_v^2	.8630(.088)	.9822(.100)[.102]	–	.8760(.091)	.9981(.104)[.103]	–
	$n = 200$; error = 1, 2, 3, for the three panels below					
β_1	.9999(.013)	1.0000(.013)[.013]	1.0000(.013)[.013]	1.0004(.013)	1.0001(.013)[.013]	.9997(.013)[.013]
λ	.1938(.026)	.1951(.026)[.030]	.1976(.030)[.029]	.1833(.022)	.1892(.022)[.026]	.1997(.023)[.023]
ρ	.1420(.057)	.1601(.055)[.055]	.1955(.071)[.069]	.1830(.045)	.1995(.044)[.044]	.1999(.044)[.044]
σ_v^2	.8932(.038)	1.0097(.043)[.048]	–	.8922(.036)	1.0082(.041)[.048]	–
β_1	.9999(.013)	1.0000(.013)[.013]	1.0000(.013)[.013]	1.0007(.012)	1.0005(.012)[.013]	1.0000(.012)[.013]
λ	.1958(.026)	.1972(.025)[.030]	.2000(.029)[.029]	.1821(.022)	.1880(.022)[.026]	.1985(.023)[.023]
ρ	.1399(.056)	.1581(.054)[.055]	.1926(.070)[.069]	.1798(.045)	.1964(.045)[.044]	.1968(.045)[.044]
σ_v^2	.8939(.089)	1.0105(.101)[.101]	–	.8912(.090)	1.0070(.102)[.100]	–
β_1	.9993(.013)	.9993(.013)[.013]	.9993(.013)[.013]	1.0006(.013)	1.0003(.013)[.013]	.9999(.013)[.013]
λ	.1955(.026)	.1969(.026)[.030]	.1998(.029)[.029]	.1830(.023)	.1890(.023)[.026]	.1996(.024)[.023]
ρ	.1379(.057)	.1561(.055)[.055]	.1903(.071)[.070]	.1784(.043)	.1949(.043)[.044]	.1953(.043)[.044]
σ_v^2	.8916(.063)	1.0080(.072)[.074]	–	.8956(.065)	1.0121(.074)[.074]	–
	$n = 400$; error = 1, 2, 3, for the three panels below					
β_1	1.0003(.009)	1.0009(.009)[.009]	1.0003(.009)[.009]	.9997(.009)	.9997(.009)[.009]	.9994(.009)[.009]
λ	.1974(.016)	.2490(.018)[.019]	.2010(.018)[.018]	.1860(.016)	.1888(.016)[.018]	.1991(.016)[.017]
ρ	.1533(.037)	.1210(.028)[.037]	.1959(.046)[.047]	.1924(.032)	.1995(.031)[.031]	.2001(.031)[.032]
σ_v^2	.8923(.027)	1.0063(.030)[.034]	–	.8913(.028)	1.0049(.031)[.034]	–
β_1	.9995(.009)	1.0001(.009)[.009]	.9995(.009)[.009]	1.0003(.009)	1.0003(.009)[.009]	1.0000(.009)[.009]
λ	.1965(.017)	.2493(.019)[.019]	.1997(.019)[.018]	.1871(.016)	.1900(.016)[.018]	.2004(.017)[.017]
ρ	.1562(.038)	.1232(.029)[.037]	.1996(.048)[.047]	.1914(.031)	.1985(.030)[.031]	.1991(.030)[.031]
σ_v^2	.8933(.061)	1.0075(.069)[.072]	–	.8923(.063)	1.0060(.071)[.071]	–
β_1	.9998(.009)	1.0004(.009)[.009]	.9998(.009)[.009]	1.0003(.009)	1.0002(.009)[.009]	1.0000(.009)[.009]
λ	.1968(.017)	.2485(.018)[.019]	.2003(.019)[.018]	.1864(.016)	.1892(.016)[.018]	.1996(.016)[.017]
ρ	.1531(.038)	.1208(.029)[.037]	.1958(.047)[.047]	.1922(.031)	.1993(.031)[.031]	.1999(.031)[.032]
σ_v^2	.8960(.045)	1.0105(.051)[.053]	–	.8950(.047)	1.0090(.053)[.053]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 1.4. Empirical mean(sd)[$\hat{sd}$] of estimators for FE-SPD model with `SL-SD` effects, 10% random missing, **heteroscedasticity**, $(\beta_1, \beta_2, \lambda, \sigma_v^2) = (1, 0.2, 0.2, 1)$, `W=Group-II`.

	T=5			T=10		
	QMLE	M-Est	RM-Est	QMLE	M-Est	RM-Est
	$n = 50$; error = 1, 2, 3, for the three panels below					
β_1	1.0113(.045)	1.0054(.045)[.042]	1.0033(.045)[.044]	1.0134(.028)	1.0058(.028)[.027]	1.0024(.028)[.029]
β_2	.7304(.237)	.6103(.223)[.277]	.5672(.256)[.269]	.6962(.153)	.5798(.143)[.172]	.5278(.167)[.190]
λ	.0569(.146)	.1317(.137)[.152]	.1583(.159)[.165]	.0811(.089)	.1517(.084)[.096]	.1831(.099)[.114]
σ_v^2	.7704(.102)	1.0208(.135)[.148]	–	.8676(.074)	.9945(.085)[.093]	–
β_1	1.0124(.045)	1.0065(.044)[.042]	1.0045(.045)[.043]	1.0130(.028)	1.0054(.028)[.027]	1.0021(.028)[.028]
β_2	.7306(.230)	.6121(.216)[.276]	.5705(.247)[.261]	.6972(.151)	.5811(.141)[.172]	.5302(.166)[.187]
λ	.0572(.141)	.1310(.132)[.151]	.1566(.153)[.160]	.0805(.089)	.1508(.083)[.096]	.1816(.099)[.112]
σ_v^2	.7664(.224)	1.0156(.297)[.273]	–	.8662(.171)	.9930(.196)[.189]	–
β_1	1.0084(.044)	1.0025(.044)[.043]	1.0003(.044)[.043]	1.0127(.028)	1.0051(.028)[.027]	1.0017(.029)[.029]
β_2	.7193(.232)	.5998(.218)[.276]	.5557(.249)[.266]	.6969(.156)	.5804(.146)[.172]	.5285(.171)[.189]
λ	.0627(.143)	.1373(.134)[.152]	.1645(.155)[.164]	.0809(.092)	.1514(.086)[.096]	.1828(.102)[.113]
σ_v^2	.7771(.166)	1.0297(.219)[.211]	–	.8710(.122)	.9984(.140)[.142]	–
	$n = 100$; error = 1, 2, 3, for the three panels below					
β_1	1.0090(.032)	1.0056(.032)[.030]	1.0028(.032)[.032]	1.0080(.020)	1.0045(.020)[.019]	1.0011(.020)[.021]
β_2	.6379(.166)	.5817(.161)[.185]	.5351(.185)[.201]	.6174(.106)	.5644(.102)[.118]	.5130(.120)[.125]
λ	.1149(.095)	.1497(.092)[.103]	.1785(.109)[.119]	.1250(.063)	.1588(.061)[.066]	.1914(.073)[.077]
σ_v^2	.7609(.070)	.9942(.091)[.105]	–	.8659(.052)	.9843(.060)[.067]	–
β_1	1.0081(.031)	1.0047(.031)[.029]	1.0019(.031)[.032]	1.0080(.020)	1.0045(.020)[.019]	1.0012(.020)[.021]
β_2	.6336(.161)	.5778(.156)[.185]	.5313(.179)[.196]	.6179(.104)	.5650(.101)[.118]	.5141(.118)[.124]
λ	.1168(.092)	.1513(.089)[.102]	.1801(.105)[.116]	.1247(.063)	.1584(.061)[.067]	.1907(.072)[.076]
σ_v^2	.7653(.160)	.9998(.209)[.200]	–	.8667(.119)	.9853(.136)[.138]	–
β_1	1.0081(.031)	1.0048(.031)[.030]	1.0020(.031)[.032]	1.0077(.020)	1.0042(.020)[.019]	1.0009(.021)[.021]
β_2	.6370(.163)	.5810(.158)[.185]	.5348(.182)[.200]	.6186(.106)	.5656(.103)[.118]	.5150(.121)[.124]
λ	.1150(.095)	.1497(.092)[.103]	.1782(.108)[.118]	.1238(.064)	.1575(.062)[.067]	.1897(.075)[.076]
σ_v^2	.7629(.115)	.9968(.150)[.152]	–	.8691(.090)	.9880(.102)[.102]	–
	$n = 200$; error = 1, 2, 3, for the three panels below					
β_1	1.0067(.021)	1.0050(.021)[.021]	1.0017(.022)[.023]	1.0053(.015)	1.0036(.015)[.014]	1.0005(.015)[.015]
β_2	.5960(.117)	.5690(.116)[.129]	.5193(.133)[.146]	.5831(.077)	.5563(.076)[.082]	.5074(.086)[.093]
λ	.1412(.067)	.1573(.066)[.068]	.1869(.077)[.085]	.1485(.044)	.1651(.044)[.046]	.1953(.051)[.055]
σ_v^2	.7670(.051)	.9909(.066)[.073]	–	.8830(.039)	.9979(.044)[.047]	–
β_1	1.0056(.021)	1.0038(.021)[.021]	1.0005(.022)[.023]	1.0051(.014)	1.0034(.014)[.014]	1.0003(.015)[.015]
β_2	.5952(.113)	.5683(.111)[.129]	.5179(.127)[.144]	.5807(.075)	.5540(.073)[.083]	.5048(.083)[.093]
λ	.1438(.064)	.1598(.063)[.068]	.1898(.074)[.084]	.1495(.043)	.1660(.043)[.046]	.1964(.050)[.055]
σ_v^2	.7684(.110)	.9926(.143)[.141]	–	.8814(.088)	.9962(.099)[.099]	–
β_1	1.0062(.022)	1.0044(.021)[.021]	1.0011(.022)[.023]	1.0049(.015)	1.0032(.015)[.014]	1.0001(.015)[.015]
β_2	.5952(.116)	.5682(.114)[.129]	.5180(.131)[.145]	.5839(.077)	.5572(.076)[.083]	.5081(.086)[.093]
λ	.1434(.066)	.1595(.065)[.068]	.1893(.076)[.084]	.1487(.044)	.1652(.043)[.046]	.1956(.051)[.055]
σ_v^2	.7646(.082)	.9878(.105)[.106]	–	.8848(.063)	1.0000(.071)[.074]	–
	$n = 400$; error = 1, 2, 3, for the three panels below					
β_1	1.0049(.015)	1.0040(.015)[.014]	1.0006(.015)[.016]	1.0039(.010)	1.0030(.010)[.010]	1.0000(.010)[.011]
β_2	.5716(.079)	.5582(.078)[.086]	.5083(.090)[.103]	.5631(.051)	.5498(.051)[.057]	.5031(.057)[.065]
λ	.1562(.046)	.1643(.046)[.048]	.1947(.054)[.062]	.1616(.030)	.1698(.030)[.032]	.1987(.034)[.039]
σ_v^2	.7778(.036)	.9995(.046)[.052]	–	.8933(.027)	1.0073(.031)[.034]	–
β_1	1.0048(.014)	1.0040(.014)[.014]	1.0006(.015)[.015]	1.0041(.010)	1.0032(.010)[.010]	1.0003(.010)[.010]
β_2	.5695(.080)	.5561(.079)[.086]	.5057(.091)[.103]	.5627(.051)	.5495(.051)[.057]	.5030(.057)[.065]
λ	.1579(.047)	.1660(.047)[.048]	.1968(.055)[.061]	.1609(.030)	.1691(.029)[.032]	.1979(.034)[.039]
σ_v^2	.7778(.078)	.9995(.100)[.102]	–	.8948(.064)	1.0089(.072)[.072]	–
β_1	1.0047(.014)	1.0038(.014)[.014]	1.0005(.014)[.015]	1.0038(.010)	1.0030(.010)[.010]	1.0000(.010)[.010]
β_2	.5714(.078)	.5580(.077)[.086]	.5084(.089)[.103]	.5615(.050)	.5483(.049)[.057]	.5017(.056)[.065]
λ	.1560(.046)	.1641(.046)[.048]	.1944(.054)[.062]	.1620(.029)	.1701(.029)[.032]	.1990(.033)[.039]
σ_v^2	.7777(.057)	.9993(.073)[.076]	–	.8905(.047)	1.0041(.053)[.053]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

the tax rates of neighboring states are likely to play a role in the determination of excise tax policy. Based on this theory, Egger et al. (2005) and Devereux et al. (2007) use spatial econometric techniques to identify such phenomenon. Both of them propose to use the 2SLS method to estimate the spatial lag parameter. As the quadratic moment conditions are not used for spatial parameters, the 2SLS approach usually cannot provide efficient estimation results, compared with our M-estimation. In addition, our M-estimation is able to be robust against unknown heteroscedasticity.

We construct a panel of data from 48 US states over 23 years, from 1977 to 1999. Following Devereux et al. (2007), we do not use the two states, Alaska and Hawaii, as they do not share borders with any other states. The missing percentages for three types of tax rates are, respectively, 8.61%, 6.25%, and 6.34%. As the missing percentage is small, we argue the interaction effects of unobserved units on observed ones are asymptotically negligible. That is, with a fixed number of missing, the unbalanced model at least asymptotically fits our random missing mechanism. The spatial neighboring states are defined as those that share a common border. The overall spatial weight matrix W for the total 48 states is row-normalized with zero on the diagonals. In each time period t, the rows and columns of W corresponding to the missing observations will be deleted, yielding W_t. Thus, $\{W_t\}$ are generally time-varying and may not be row-normalized. We follow Egger et al. (2005) and set a number of control variables including state size (population density), spatially weighted size, age dependency ratio, government ideological orientation, lagged sales tax rate, top income tax rate, and public expenditure. The data of tax rates on gasoline, cigarettes, and beers are collected from the World Tax Database (WTD) maintained by the Office of Tax Policy Research at the University of Michigan, and all the other control variables are collected following Egger et al. (2005).

Table 1.5 summarizes the M-estimation results for gasoline, cigarettes, and beer tax rates. For the models with only `SE` effects, the spatial error parameter estimate exhibits a significant positive value for each type of tax rate. This implies that the unobserved stochastic shocks increasing one state's tax rates

Table 1.5. M-estimation results for US state tax competition.

Explanatory Variables	Gasoline			Cigarettes			Beer		
	SE	SL	SL-SE	SE	SL	SL-SE	SE	SL	SL-SE
Spatial lag parameter		0.273	0.235		0.487	0.788		0.072	-0.270
		(0.02)***	(0.02)***		(0.05)***	(0.05)***		(0.03)***	(0.03)***
State size	-0.015	0.002	-0.002	0.342	0.306	0.257	0.084	0.097	0.082
	(0.02)	(0.01)	(0.01)	(0.06)***	(0.03)***	(0.02)***	(0.03)***	(0.01)***	(0.02)***
Weighted size	0.001	-0.010	-0.008	-0.009	-0.072	-0.117	-0.009	-0.018	0.003
of neighbours	(0.01)	(0.06)	(0.06)	(0.03)	(0.18)	(0.13)	(0.01)	(0.08)	(0.10)
Dependency ratio	0.012	-0.047	-0.044	0.313	0.159	-0.050	0.245	0.233	0.338
	(0.07)	(0.01)***	(0.01)***	(0.24)	(0.03)***	(0.03)*	(0.10)***	(0.01)***	(0.01)***
Political orientation	0.029	0.021	0.021	-0.047	-0.055	-0.039	-0.012	-0.018	-0.002
of state governments	(0.01)***	(0.17)	(0.17)	(0.03)	(0.48)	(0.50)	(0.02)	(0.23)	(0.21)
Lagged sales tax rate	-0.180	-0.120	-0.130	1.628	1.256	0.527	0.037	-0.030	0.082
	(0.17)	(0.06)**	(0.06)**	(0.50)***	(0.17)***	(0.16)***	(0.23)	(0.08)	(0.08)
Lagged top income	0.143	0.137	0.133	-0.331	-0.328	-0.271	-0.073	-0.080	-0.067
tax rate	(0.06)***	(0.00)***	(0.00)***	(0.17)*	(0.00)***	(0.00)***	(0.08)	(0.00)***	(0.00)***
Public expenditure	-0.001	-0.001	-0.001	0.002	0.002	0.002	0.001	0.001	0.001
	(0.00)***	(0.04)	(0.05)	(0.00)***	(0.04)	(0.05)	(0.00)***	(0.04)	(0.08)
Spatial error parameter	0.300		0.071	0.492		-0.555	0.190		0.451
	(0.05)***		(0.08)	(0.04)***		(0.12)***	(0.05)***		(0.08)***
Observations	1009	1009	1009	1035	1035	1035	1034	1034	1034

Note: Standard errors of coefficients are in parentheses. *,**,*** represent significance levels 10%, 5% and 1%, respectively.

would also have a positive effect on its neighbors' tax rates. For the models with only SL effects, we identify a significant positive spatial lag parameter for all the tax rates. These findings suggest the existence of the tax competition and consistent with the results in Devereux et al. (2007). When including both SL and SE effects into the models, we still can observe significant positive results for spatial lag parameter estimate for Gasoline and Cigarettes tax rates. In addition, the spatial error parameter estimate for Cigarettes tax rates is significantly negative, which is consistent with that in Egger et al. (2005). However, our approach has advantages in that we are able to make statistical inferences on the SE effect. For the Beer tax rates, we find a significant negative estimate for the SL parameter but a significant positive estimate for the SE parameter.

As different states vary greatly in so many aspects such as history, population structure, and other social-economical characteristics, it is thus natural to believe that the innovations may be heteroscedastic. Therefore, we also report all the empirical findings based on the RM-estimation in Table 1.6. The coefficients are close to those in Table 1.5 with only mild differences in significance level for some estimates.

1.6 Extensions

As discussed in the introduction section, the unbalanced SPD (USPD) models and the associated M-estimations are quite general in that they can be

Table 1.6. RM-estimation results for US state tax competition.

Explanatory Variables	Gasoline			Cigarettes			Beer		
	`SE`	`SL`	`SL-SE`	`SE`	`SL`	`SL-SE`	`SE`	`SL`	`SL-SE`
Spatial lag parameter		0.267	0.231		0.468	0.715		0.071	-0.259
		(0.03)***	(0.03)***		(0.08)***	(0.07)***		(0.02)***	(0.02)***
State size	-0.016	0.001	-0.003	0.341	0.305	0.271	0.084	0.096	0.083
	(0.03)	(0.01)	(0.01)	(0.10)***	(0.02)***	(0.02)***	(0.02)***	(0.01)***	(0.01)***
Weighted size	0.001	-0.009	-0.008	-0.009	-0.069	-0.106	-0.009	-0.018	0.003
of neighbours	(0.01)	(0.06)	(0.06)	(0.02)	(0.17)	(0.17)	(0.01)	(0.13)	(0.16)
Dependency ratio	0.014	-0.046	-0.043	0.323	0.173	0.010	0.245	0.233	0.336
	(0.07)	(0.01)***	(0.01)***	(0.20)	(0.04)***	(0.03)	(0.15)	(0.02)***	(0.02)***
Political orientation	0.029	0.021	0.021	-0.046	-0.055	-0.044	-0.012	-0.018	-0.002
of state governments	(0.01)***	(0.23)	(0.23)	(0.04)	(0.53)	(0.63)	(0.02)	(0.18)	(0.17)
Lagged sales tax rate	-0.177	-0.119	-0.128	1.601	1.224	0.671	0.036	-0.030	0.078
	(0.23)	(0.07)	(0.07)*	(0.57)***	(0.17)***	(0.16)***	(0.19)	(0.06)	(0.06)
Lagged top income	0.144	0.137	0.134	-0.334	-0.332	-0.292	-0.073	-0.080	-0.069
tax rate	(0.07)*	(0.00)***	(0.00)***	(0.16)**	(0.00)***	(0.00)***	(0.06)	(0.00)***	(0.00)***
Public expenditure	-0.001	-0.001	-0.001	0.002	0.002	0.002	0.001	0.001	0.001
	(0.00)***	(0.04)	(0.05)	(0.00)***	(0.06)	(0.12)	(0.00)***	(0.05)	(0.09)
Spatial error parameter	0.289		0.068	0.478		-0.421	0.189		0.436
	(0.05)***		(0.07)	(0.06)***		(0.23)*	(0.06)***		(0.09)***
Observations	1009	1009	1009	1035	1035	1035	1034	1034	1034

Note: Standard errors of coefficients are in parentheses. *,**,*** represent significance levels 10%, 5% and 1%, respectively.

extended to allow for additional features in the model or to different types of unbalanced SPD models. For illustration, we extend the current model to allow errors to be serially correlated, and consider models with random effects (RE). We present some details on the following four extensions: (*i*) USPD model with two-way FE and serial correlation, (*ii*) USPD model with two-way FE, heteroscedasticity and serial correlation, (*iii*) USPD model with two-way RE and serial correlation, and (*iv*) USPD model with two-way RE, heteroscedasticity and serial correlation. For serial correlation, we assume that the model errors follow a stationary AR(1), i.e., $v_{it} = \varrho v_{i,t-1} + e_{it}$ with $|\varrho| < 1$. Cases (*i*), (*ii*) and (*iv*) all encounter incidental parameters problem, the standard methods for balanced panels cannot be applied, and the proposed M-estimation needs to be called for. Case (*iii*) illustrates the simplicity of the proposed modeling strategy in controlling the random effects in the unbalanced SPD models with general time-varying spatial weight matrices and serial correlation.

(i) USPD Model with Two-Way FE and Serial Correlation

Assume $v_{it} = \varrho v_{i,t-1} + e_{it}$ with $|\varrho| < 1$, and $e_{it} \sim \text{iid}(0, \sigma_e^2)$. Denote $\mathbb{K} = \texttt{blkdiag}(D_1, \ldots, D_T)$. It is easy to see that $\text{Var}(\mathbf{V}) = \sigma_e^2 \mathbb{K}(\Omega_V(\varrho) \otimes I_n)\mathbb{K}' \equiv \sigma_e^2 \Upsilon_N(\varrho)$, where

$$\Omega_V(\varrho) = \frac{1}{1-\varrho^2} \begin{bmatrix} 1 & \varrho & \cdots & \varrho^{T-1} \\ \varrho & 1 & \cdots & \varrho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varrho^{T-1} & \varrho^{T-2} & & 1 \end{bmatrix}.$$

Denote $\mathbb{U}_N(\rho, \varrho) = \Upsilon_N^{-1}(\varrho) - \Upsilon_N^{-1}(\varrho)\mathbb{D}(\rho)[\mathbb{D}'(\rho)\Upsilon_N^{-1}(\varrho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)\Upsilon_N^{-1}(\varrho)$. Let $\theta = (\beta', \sigma_e^2, \delta)'$, where $\delta = (\lambda, \rho, \varrho)'$. The concentrated quasi Gaussian loglikelihood function (with ϕ being concentrated) of θ takes the form:

$$\ell_N^c(\theta) = -\tfrac{N}{2}\ln 2\pi - \tfrac{N}{2}\ln\sigma_e^2 - \tfrac{1}{2}\ln|\Upsilon_N(\varrho)| + \ln|\mathbf{A}_N(\lambda)| + \ln|\mathbf{B}_N(\rho)| - \tfrac{1}{2\sigma_e^2}\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta),$$

where $\tilde{\mathbf{V}}(\beta,\delta) = \mathbb{U}_N(\rho,\varrho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$. Hence, the concentrated quasi score (CQS) functions $S_N^c(\theta) = \frac{\partial}{\partial\theta}\ell_N^c(\theta)$ is given as

$$S_N^c(\theta) = \begin{cases} \frac{1}{\sigma_e^2}\mathbf{X}'\mathbf{B}_N'(\rho)\mathbb{U}_N'(\rho,\varrho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma_e^4}[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - N\sigma_e^2], \\ \frac{1}{\sigma_e^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\mathbb{U}_N'(\rho,\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{F}_N(\lambda)], \\ \frac{1}{\sigma_e^2}\tilde{\mathbf{V}}(\beta,\delta)'\mathbf{G}_N(\rho)\Upsilon_N(\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{G}_N(\rho)], \\ \frac{1}{2\sigma_e^2}\tilde{\mathbf{V}}(\beta,\delta)'\dot{\Upsilon}_N(\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \frac{1}{2}\mathtt{tr}[\Upsilon_N^{-1}(\varrho)\dot{\Upsilon}_N(\varrho)], \end{cases}$$

where $\dot{\Upsilon}_N(\varrho) = \frac{\partial}{\partial\varrho}\Upsilon_N(\varrho)$. To remove the effect from estimating FEs, we correct $S_N^c(\theta)$ using $S_N^*(\theta_0) = S_N^c(\theta_0) - \mathrm{E}[S_N^c(\theta_0)]$, which takes the form at the general θ:

$$S_N^*(\theta) = \begin{cases} \frac{1}{\sigma_e^2}\mathbf{X}'\mathbf{B}_N'(\rho)\mathbb{U}_N'(\rho,\varrho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma_e^4}[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - \sigma_e^2\mathtt{tr}(\mathbb{U}_N(\rho,\varrho))], \\ \frac{1}{\sigma_e^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\mathbb{U}_N'(\rho,\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)\Upsilon_N(\varrho)\mathbb{U}_N^2(\rho,\varrho)], \\ \frac{1}{\sigma_e^2}\tilde{\mathbf{V}}(\beta,\delta)'\mathbf{G}_N(\rho)\Upsilon_N(\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{G}_N(\rho)\Upsilon_N(\varrho)\mathbb{U}_N(\rho,\varrho)], \\ \frac{1}{2\sigma_e^2}\tilde{\mathbf{V}}(\beta,\delta)'\dot{\Upsilon}_N(\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \frac{1}{2}\mathtt{tr}[\dot{\Upsilon}_N(\varrho)\mathbb{U}_N(\rho,\varrho)]. \end{cases}$$

Solving the AQS equations: $S_N^*(\theta) = 0$, gives the M-estimator of θ.

(ii) USPD Model with Two-Way FE, heteroscedasticity and Serial Correlation

Now, we consider the case that errors are heteroscedastic across individuals and serially correlated across time, i.e., $v_{it} = \varrho v_{i,t-1} + e_{it}$ with $|\varrho| < 1$ and $e_{it} \sim \text{inid}(0, \sigma_i^2)$. Let $h = \mathtt{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. In this case, we have $\mathrm{Var}(\mathbf{V}) = \mathbf{H}\Upsilon_N(\varrho)$, where $\mathbf{H} = \mathtt{blkdiag}(h_1, \ldots, h_T)$ and h_t is obtained from h by omitting the rows and columns corresponding to the missing units at time t. Following the similar derivations as we do in Subsection 1.3.1, we obtain the desired AQS functions robust against the unknown heteroscedasticity:

$$S_N^{\diamond}(\beta,\delta) = \begin{cases} \mathbf{X}'\mathbf{B}_N'(\rho)\mathbb{U}_N'(\rho,\varrho)\tilde{\mathbf{V}}(\beta,\delta), \\ \mathbf{Y}'\mathbf{A}_N'(\lambda)\mathbf{B}_N'(\rho)[\bar{\mathbf{F}}_N'(\lambda,\rho) - \bar{\mathbb{F}}_N'(\delta)]\tilde{\mathbf{V}}(\beta,\delta), \\ [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)[\bar{\mathbf{G}}_N(\rho,\varrho) - \bar{\mathbb{G}}_N(\rho,\varrho)]\tilde{\mathbf{V}}(\beta,\delta), \\ [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)[\bar{\mathbf{U}}_N(\rho,\varrho) - \bar{\mathbb{U}}_N(\rho,\varrho)]\tilde{\mathbf{V}}(\beta,\delta), \end{cases}$$

where $\bar{\mathbf{F}}_N'(\lambda,\rho) = \mathbf{B}_N^{-1\prime}(\rho)\mathbf{F}_N'(\lambda)\mathbf{B}_N'(\rho)$, $\bar{\mathbf{G}}_N(\rho,\varrho) = \mathbb{U}_N(\rho,\varrho)\mathbf{G}_N(\rho)\Upsilon_N(\varrho)$, $\bar{\mathbf{U}}_N(\rho,\varrho) = \mathbb{U}_N(\rho,\varrho)\dot{\Upsilon}_N(\varrho)$, $\bar{\mathbb{F}}_N'(\delta) = \Upsilon_N^{-1}(\varrho)\texttt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{F}}_N'(\delta)\mathbb{U}_N(\rho,\varrho)]\texttt{diag}[\mathbb{U}_N(\rho,\varrho)]^{-1}$, $\bar{\mathbb{G}}_N(\rho,\varrho) = \Upsilon_N^{-1}(\varrho)\texttt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{G}}_N(\rho,\varrho)\mathbb{U}_N(\rho,\varrho)]\texttt{diag}[\mathbb{U}_N(\rho,\varrho)]^{-1}$, and $\bar{\mathbb{U}}_N(\rho,\varrho) = \Upsilon_N^{-1}(\varrho)\texttt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{U}}_N(\rho,\varrho)\mathbb{U}_N(\rho,\varrho)]\texttt{diag}[\mathbb{U}_N(\rho,\varrho)]^{-1}$.

Solving the robust AQS equations: $S_N^{\diamond}(\beta,\delta) = 0$, gives the M-estimators of β and δ, robust against unknown heteroscedasticity, and allowing serial correlation of AR(1) form.

(iii) USPD Model with Two-Way RE and Serial Correlation

Assume $\mu_i \sim \text{iid}(0,\sigma_\mu^2)$, $\alpha_t \sim \text{iid}(0,\sigma_\alpha^2)$, and they are mutually independent and independent of e_{it}. Then the covariance matrix of the composite error term is

$$\mho_N(\theta_1) = \sigma_\mu^2\mathcal{D}_\mu(\rho) + \sigma_\alpha^2\mathcal{D}_\alpha(\rho) + \sigma_e^2\Upsilon_N(\varrho),$$

with $\theta_1 = (\rho,\varrho,\sigma_e^2,\sigma_\mu^2,\sigma_\alpha^2)'$, $\mathcal{D}_\mu(\rho) = \mathbf{B}_N(\rho)\mathbf{D}_\mu\mathbf{D}_\mu'\mathbf{B}_N'(\rho)$ and $\mathcal{D}_\alpha(\rho) = \mathbf{B}_N(\rho)\mathbf{D}_\alpha\mathbf{D}_\alpha'\mathbf{B}_N'(\rho)$. The quasi Gaussian loglikelihood function of $\theta = (\beta',\lambda,\theta_1')'$ is

$$\begin{aligned} \ell_N(\theta) = & -\tfrac{N}{2}\ln 2\pi - \tfrac{1}{2}\ln|\mho_N(\theta_1)| + \ln|\mathbf{A}_N(\lambda)| + \ln|\mathbf{B}_N(\rho)| \\ & -\tfrac{1}{2}\mathbf{V}'(\beta,\lambda,\rho)\mho_N^{-1}(\theta_1)\mathbf{V}(\beta,\lambda,\rho), \end{aligned}$$

where $\mathbf{V}(\beta,\lambda,\rho) = \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$. The direct QML estimator $\hat{\theta}_{\texttt{QML}}$ of θ maximizes the above equation $\ell_N(\theta)$, and its consistency and asymptotic normality can be easily established.

(iv) USPD Model with Two-Way RE, heteroscedasticity and Serial Correlation

We now extend the model in (*iii*) to allow heteroscedasticity in the errors as in (*ii*) above. Denote $\tilde{\mathbf{V}}(\theta) = \mho_N^{-1}(\theta_1)\mathbf{V}(\beta,\lambda,\rho)$. The quasi score functions assuming homoscedasticity are:

$$S_N(\theta) = \begin{cases} \mathbf{X}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\theta), \\ \mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\theta) - \mathtt{tr}[\mathbf{F}_N(\lambda)], \\ \sigma_e^2\tilde{\mathbf{V}}'(\theta)\mathbf{G}_N(\rho)\Upsilon_N(\varrho)\tilde{\mathbf{V}}(\theta) - \sigma_e^2\mathtt{tr}[\mathbf{G}_N(\rho)\Upsilon_N(\varrho)\mho_N^{-1}(\theta_1)], \\ \frac{\sigma_e^2}{2}\tilde{\mathbf{V}}'(\theta)\dot{\Upsilon}_N(\varrho)\tilde{\mathbf{V}}(\theta) - \frac{\sigma_e^2}{2}\mathtt{tr}[\dot{\Upsilon}_N(\varrho)\mho_N^{-1}(\theta_1)], \\ \frac{1}{2}\tilde{\mathbf{V}}'(\theta)\Upsilon_N(\varrho)\tilde{\mathbf{V}}(\theta) - \frac{1}{2}\mathtt{tr}[\Upsilon_N(\varrho)\mho_N^{-1}(\theta_1)], \\ \frac{1}{2}\tilde{\mathbf{V}}'(\theta)\mathcal{D}_\mu(\rho)\tilde{\mathbf{V}}(\theta) - \frac{1}{2}\mathtt{tr}[\mathcal{D}_\mu(\rho)\mho_N^{-1}(\theta_1)], \\ \frac{1}{2}\tilde{\mathbf{V}}'(\theta)\mathcal{D}_\alpha(\rho)\tilde{\mathbf{V}}(\theta) - \frac{1}{2}\mathtt{tr}[\mathcal{D}_\alpha(\rho)\mho_N^{-1}(\theta_1)]. \end{cases}$$

It is easy to see that $\mathrm{E}[S_N^{\mathrm{c}}(\theta_0)] \neq 0$ when e_{it}'s are heteroscedastic. Therefore, some adjustments on the above quasi score functions are necessary in order to have consistent estimation. Denote $\theta_2 = (\rho, \varrho, \sigma_\mu^2, \sigma_\alpha^2)'$, $\xi = (\beta', \lambda, \theta_2')'$, $\mho_N(\theta_2) = \sigma_\mu^2\mathcal{D}_\mu(\rho) + \sigma_\alpha^2\mathcal{D}_\alpha(\rho) + \Upsilon_N(\varrho)$ and $\tilde{\mathbf{V}}(\xi) = \mho_N^{-1}(\theta_2)\mathbf{V}(\beta, \lambda, \rho)$. Alone the similar ideas of Section, some tedious algebra leads to the AQS functions robust against the unknown heteroscedasticity and allowing serial correlation of AR(1) form:

$$S_N^\diamond(\xi) = \begin{cases} \mathbf{X}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\xi), \\ \mathbf{Y}'\mathbf{A}_N'(\lambda)\mathbf{B}_N'(\rho)[\bar{\mathbf{F}}_N'(\lambda,\rho) - \bar{\mathbb{F}}_N'(\lambda,\theta_2)]\tilde{\mathbf{V}}(\xi) - \mathtt{tr}[\bar{\mathbf{F}}_N(\lambda,\rho) - \bar{\mathbb{F}}_N(\lambda,\theta_2)], \\ \mathbf{V}'(\beta,\lambda,\rho)[\bar{\mathbf{G}}_N(\theta_2) - \bar{\mathbb{G}}_N(\theta_2)]\tilde{\mathbf{V}}(\xi) - \mathtt{tr}[\bar{\mathbf{G}}_N(\theta_2) - \bar{\mathbb{G}}_N(\theta_2)], \\ \mathbf{V}'(\beta,\lambda,\rho)[\bar{\mathbf{U}}_N(\theta_2) - \bar{\mathbb{U}}_N(\theta_2)]\tilde{\mathbf{V}}(\xi) - \mathtt{tr}[\bar{\mathbf{U}}_N(\theta_2) - \bar{\mathbb{U}}_N(\theta_2)], \\ \mathbf{V}'(\beta,\lambda,\rho)[\bar{\mathbf{S}}_\mu(\theta_2) - \bar{\mathbb{S}}_\mu(\theta_2)]\tilde{\mathbf{V}}(\xi) - \mathtt{tr}[\bar{\mathbf{S}}_\mu(\theta_2) - \bar{\mathbb{S}}_\mu(\theta_2)], \\ \mathbf{V}'(\beta,\lambda,\rho)[\bar{\mathbf{S}}_\alpha(\theta_2) - \bar{\mathbb{S}}_\alpha(\theta_2)]\tilde{\mathbf{V}}(\xi) - \mathtt{tr}[\bar{\mathbf{S}}_\alpha(\theta_2) - \bar{\mathbb{S}}_\alpha(\theta_2)], \end{cases}$$

where $\bar{\mathbf{F}}_N'(\lambda,\rho) = \mathbf{B}_N^{-1\prime}(\rho)\mathbf{F}_N'(\lambda)\mathbf{B}_N'(\rho)$, $\bar{\mathbf{G}}_N(\theta_2) = \mho_N^{-1}(\theta_2)\mathbf{G}_N(\rho)\Upsilon_N(\varrho)$,
$\bar{\mathbf{U}}_N(\theta_2) = \mho_N^{-1}(\theta_2)\dot{\Upsilon}_N(\varrho)$, $\bar{\mathbf{S}}_\varpi(\theta_2) = \mho_N^{-1}(\theta_2)\mathcal{D}_\varpi(\rho)$,
$\bar{\mathbb{F}}_N'(\lambda,\theta_2) = \Upsilon_N^{-1}(\varrho)\mathtt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{F}}_N'(\delta)\mho_N^{-1}(\theta_2)]\mathtt{diag}[\mho_N^{-1}(\theta_2)]^{-1}$,
$\bar{\mathbb{G}}_N(\theta_2) = \Upsilon_N^{-1}(\varrho)\mathtt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{G}}_N(\theta_2)\mho_N^{-1}(\theta_2)]\mathtt{diag}[\mho_N^{-1}(\theta_2)]^{-1}$,
$\bar{\mathbb{U}}_N(\theta_2) = \Upsilon_N^{-1}(\varrho)\mathtt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{U}}_N(\theta_2)\mho_N^{-1}(\theta_2)]\mathtt{diag}[\mho_N^{-1}(\theta_2)]^{-1}$,
and $\bar{\mathbb{S}}_\varpi(\theta_2) = \Upsilon_N^{-1}(\varrho)\mathtt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{S}}_\varpi(\theta_2)\mho_N^{-1}(\theta_2)]\mathtt{diag}[\mho_N^{-1}(\theta_2)]^{-1}$, for $\varpi = \mu$ or α.

Solving the robust AQS equations: $S_N^\diamond(\xi) = 0$, gives the M-estimator of ξ, robust against unknown heteroscedasticity, and allowing serial correlation of AR(1) form.

Asymptotic properties of the M-estimators in cases (*i*) and (*ii*) can be studied in a similar way as that in the main text of this chapter, and inferences methods can be developed along the same line. However, formal studies on these cases are still quite involved, and can only be done in a future research work. For the cases (*iii*) and (*iv*), we do not foresee any difficulties in establishing the asymptotic properties of the QML and M-estimators, but developments of the inference methods may encounter some difficulties due to the involvement of three error components which may be all non-normal, and the allowance of unknown heteroscedasticity. Formal studies on these cases are in our future research agenda.

1.7 Conclusion

We consider estimation and inference for an unbalanced spatial panel data model with both individual and time fixed effects, where the unbalancedness is caused by, e.g., late entries, early dropouts, lack of economic activities, such that the missing spatial units at a given time period do not generate any spillover effects on their 'neighbors". Unbalanced spatial panels with fixed effects render the commonly adopted approach, the orthogonal transformation, inapplicable. An *adjusted quasi score* (AQS) is proposed, which adjusts the concentrated quasi scores (with the fixed effects being concentrated out) to remove the effects of estimating these *incidental parameters*. For the statistical inferences, the main difficulty lies with the fact that 'consistent' estimates of the idiosyncratic errors are unavailable due to the incidental parameters problem. A 'plug-in and then bias-correction' method is proposed to give consistent estimates of the standard errors of the M-estimators. The proposed methods are then extended to allow for unknown heteroscedasticity along both the cross-sectional and time dimensions. Monte Carlo results show excellent performance of the proposed estimation and inference methods.

The proposed methods are seen to be very general in handling the unbalanced SPD models in the presence of incidental parameters such as fixed effects and unknown heteroscedasticity, allowing the spatial weight matrices to be time-varying and without row-normalizations. The generality of the proposed methods is further demonstrated by considering the following extensions: the unbalanced SPD model with (*i*) two-way fixed effects (FE) and serial correlation, (*ii*) two-way FE, heteroscedasticity and serial correlation, (*iii*) two-way random effects (RE) and serial correlation, and (*iv*) two-way RE, heteroscedasticity and serial correlation. The current study also sheds light on an interesting but challenging extension: unbalanced SPD models with interactive fixed effects in the spirit of Bai et al. (2015). However, rigorous studies on these extensions can only be done in future works.

Chapter 2

Threshold Spatial Panel Data Models with Fixed Effects

2.1 Introduction

Since Anselin (1988), researchers have been paying increasing attention to spatial panel data (SPD) models for their ability to model the cross-sectional dependence while maintaining full control for the unobservable heterogeneity. See, among others, Baltagi et al. (2003), Lee and Yu (2010), Baltagi and Yang (2013a,b), Yang et al. (2016), Liu and Yang (2020) and Li and Yang (2020,2021). The threshold regression model is another popular specification with wide practical applicability, which divides observations into distinct regimes, depending on the value of an observable variable (threshold variable) – whether or not it exceeds some threshold value. See Hansen (2011) for an overview of the development of threshold models in both econometrics and economics literature. However, the existing studies of threshold models have been limited to the regular panel data models (e.g., Hansen, 1999) until very recently Wei et al. (2021) propose a threshold SPD model aiming to include both important structures in a single model.[1] The combined model can benefit from the advantages of both structures. On one hand, it enables the panel

[1]Related works but under spatial cross-sectional setup include Deng (2018) and Zhu et al. (2020).

threshold model to account for cross-sectional dependence, and on the other hand, the combined model can capture heterogeneous spatial effects under different circumstances. This greatly increases the flexibility of the SPD models.

The spatial models with threshold effects (though rarely studied in the econometrics literature) can offer a wide range of applications. In social science, Schelling (1971) finds a phenomenon of “neighborhood tipping” in the process of residential segregation that only when a recognizable new minority enters a neighborhood in sufficient numbers, will the spillover effects occur to the earlier residents. Strong empirical evidence of this phenomenon is later found by Card et al. (2008). In public economics, Glaeser et al. (1996) study the relationship between crime and social interactions and find that the amount of spillover effects for different types of crimes is distinct, depending on the severity of the crime. In empirical finance, Pesaran and Pick (2017) argue that financial crises spreading from one place to others may have two categories of causes: inter-dependence and financial contagion, distinguished by some threshold effects, and different causes would have varying degrees of impact on other places.

The literature on threshold models has been fast-growing since Tong (1978) due to their broad applicability. Chan (1993) first finds that the asymptotic distribution of least squares (LS) estimator of threshold parameter is a functional of a compound Poission process, depending on many nuisance parameters including the marginal distribution of the covariates and all the regression coefficients. Hence, this theory cannot provide a practical method to make statistical inference on the threshold parameter. Hansen (2000) assumes that the threshold effects decrease with sample size and presents a likelihood ratio (LR) test method for the construction of confidence intervals for the threshold parameter. Seo and Linton (2007) propose a smoothed LS estimation and establish inference theories in both fixed and diminishing threshold effects frameworks. However, both of these studies are at the expense of the convergence rate, compared to Chan (1993). As for the panel threshold models, Hansen (1999) studies a threshold static panel data model under the same diminishing-threshold-effect assumption. He suggests using the classic within transformation to first eliminate the individual-specific fixed effects and then applying the LS method for estimation. Dang et al. (2012) extend the model to include dynamic structure and propose to apply the GMM technique for estimation in a short panel setup. Seo and Shin (2016) further extend the dynamic model by allowing for endogenous threshold variable and regressors in both shrinking and fixed threshold effects frameworks. They develop a general GMM estimator for threshold parameter which follows a normal distribution asymptotically. However, all the above studies maintain the assumption of no cross-sectional dependence, which may be restrictive for some empirical applications. In a recent paper, Miao et al. (2020) try to relax this restriction and propose a panel threshold model with interactive fixed effects. They study the LS estimation in the diminishing-threshold-effect framework.

Spatial models offer an alternative way to model cross-sectional dependence, and in addition they can capture endogenous and contextual spatial interaction effects (Manski, 1993). The sole work in this literature is Wei et al. (2021), which extends Hansen (1999) to allow for spatial autoregressive (SAR) structure in the model. The threshold effects are allowed for both spatial and regression coefficients. However, the estimation method they propose is a two-stage least square (2SLS) estimation (Caner and Hansen, 2004), which is typically inefficient compared to ML-type estimation for spatial models. Besides, the asymptotic properties of the estimators are not studied and the statistical inference of the threshold parameter is not considered. Another related work is Li (2018), which studies an SPD model with structural change, a special case of the threshold SPD model with the threshold variable being simply the time variable. A direct quasi maximum likelihood (QML) approach is proposed for the model estimation, where the *incidental parameters problem* (Neyman and Scott, 1948) due to the estimation of fixed effects is not addressed. Both studies do not consider the time-specific fixed effects, which might be important and have empirical implications in many economic studies (e.g., Ertur and Koch, 2007; Elhorst and Fréret, 2009). Moreover, the additional time fixed effects can make the incidental parameters problem more complicated to deal with in the SPD framework (Lee and Yu, 2010). It is therefore highly desirable to develop a formal study to provide general estimation, testing and inference methods for the threshold SPD models with both individual and time fixed effects.

In this chapter, we study a threshold spatial panel data model with two-way fixed effects, which allows for the presence of threshold effects in both spatial parameter and regression slopes. To facilitate statistical inference on the threshold parameter, we impose the assumption of diminishing-threshold-effect as in Hansen (2000). We propose an ML-type estimation method as it often has better finite-samples properties than GMM/IV approach. The classical within transformation is not suitable for the ML estimation, as it will create linear dependence in the resulting disturbances. Meanwhile, the presence of the threshold effects also renders the orthogonal transformation approach (Lee and Yu, 2010) inapplicable to eliminate fixed effects because spatial weight matrices are often time-varying and may not be row-normalized. Therefore, an *adjusted quasi maximum likelihood* estimation method is proposed. The method starts from the joint quasi Gaussian loglikelihood function of all the parameters, then concentrates out the fixed effects to give the concentrated quasi loglikelihood function, and then adjusts this concentrated loglikelihood function to 'recover' the effect of degrees of freedom loss due to the estimation of the fixed effects parameters. Maximizing the adjusted concentrated loglikelihood function gives the adjusted QML estimators. It is worth mentioning that the incidental parameter problem caused by the estimation of fixed effects cannot be fully removed by this adjustment to the concentrated loglikelihood function. Thus, one main challenging part of our study still lies with the analysis of the threshold estimator in conjunction with the incidental parameters problem. In the contrast, Hansen (1999) does not have such an incidental parameter problem as the within transformation can be applied to eliminate the fixed effects. Another main challenging part is that the nonlinearity resulting from the SAR structure makes the analysis in Hansen (1999) unsuitable for our model.

We find our adjusted QML method can yield consistent estimations for spatial, variance and regression coefficients, no matter T is small or large, which is different from the direct QML approach in Li (2018) where the variance estimate is inconsistent when T is fixed. Under a non-restrictive condition that T does not grow faster than n, the adjusted QML estimator of the threshold parameter is also shown to be consistent. We find the convergence rate of the threshold estimate is related to the threshold diminishing rate and, under this rate, the estimation error of the threshold estimate has an asymptotically negligible effect on the other estimators. Hence, the asymptotic normality of the common estimators except for the threshold estimator is established. Due to the incidental parameters problem, the estimates of spatial parameters are not asymptotically centered when T and n go to infinity at the same rate. A simple bias correction procedure is proposed to remove these asymptotic biases. We also derive the asymptotic distribution of the threshold estimate but it involves the unknown parameters that cannot be estimated correctly. Therefore, we propose an LR test statistic to facilitate inference on the threshold parameter. In contrast to Hansen (1999), the LR statistic is not asymptotically pivotal even under homoskedasticity because its limiting distribution involves the third and fourth moments of the errors. When the errors are normally distributed, these third and fourth moments are zero and thus the LR statistic becomes pivotal. If the normality condition is suspected, a non-parametric technique can be applied to estimate the unknown parameter in the limit distribution of LR statistic. Finally, we follow Hansen (1996) and propose a sup-Wald statistic to test the existence of threshold effects. In view of the fact that the asymptotic distribution of the sup-Wald statistic based on the adjusted QML estimators cannot be approximated via simulation, we instead propose a sup-Wald statistic based on an M-, or *adjusted quasi score* (AQS) estimation, where the estimating function is obtained by adjusting the quasi score functions with the fixed effects being concentrated out. We introduce a bootstrap procedure to obtain the asymptotically correct critical values for the proposed sup-Wald statistic. Monte Carlo results show the excellent performance of the proposed estimators and test statistics.

The practical relevance of allowing for threshold effects in SPD models is illustrated by studying the threshold effect of leaders' age on political competitions across 338 cities in China over the periods 2010 to 2012. The political competitions among city leaders of the same level are identified by the spatial effects across city-level total investments. We find that the competitions are strong for local leaders who are younger than a threshold age but tend to vanish for those who are older than the threshold level, approaching the retirement age.

The outline of this chapter is as follows. Section 2.2 introduces our model and assumptions, discusses the QML estimation and its asymptotic properties, and studies the likelihood ratio test on the threshold value. Section 2.3 studies the hypothesis testing on the presence of threshold effects. Monte Carlo simulation findings are in Section 2.4. Section 2.5 applies our method to study the age-of-leader effects on political competitions across Chinese cities. Section 2.6 discusses some extensions. Section 2.7 concludes. Proofs are collected in the appendices.

Notation. I_m denotes an $m \times m$ identity matrix, $0_{m\times n}$ an $m \times n$ zero matrix, and l_m an $m \times 1$ vector of ones. For a square matrix, $|\cdot|$ denotes its determinant and $\mathtt{tr}(\cdot)$ its trace. For a real symmetric matrix, $\rho_{min}(\cdot)$ denotes its smallest eigenvalue. For a real $n \times m$ matrix A with elements a_{ij}, its Frobenius norm is denoted as $\|A\|$, its maximum column sum norm by $\|A\|_1 = \max_{1\leqslant j\leqslant m} \sum_{i=1}^{n} |a_{ij}|$ and its maximum row sum norm by $\|A\|_\infty = \max_{1\leqslant i\leqslant n} \sum_{j=1}^{m} |a_{ij}|$. The operators $\mathtt{diag}(\cdot)$ forms a diagonal matrix using the diagonal elements of a square matrix or a given vector and $\mathtt{diagv}(\cdot)$ forms a column vector using the diagonal elements of a square matrix. The true value of a parameter is denoted by adding a subscript 0. Finally, $A^s = A + A'$.

2.2 The Model and Adjusted QML Estimation

2.2.1 Threshold SPD model with fixed effects

Consider a total of n spatial units, interconnected at time t through an $n \times n$ spatial weight matrix W_t. There exists a threshold variable q_{it} such that depending on its value the spatial and regression coefficients may differ. Let $d_{it}(\gamma) = \mathbb{1}(q_{it} \leqslant \gamma)$, where $\mathbb{1}(\cdot)$ is the indicator function and γ is the threshold parameter assumed to take values in a bounded set $\Gamma = [\underline{\gamma}, \overline{\gamma}]$. Define

$d_t(\gamma) = \texttt{diag}\{d_{1t}(\gamma), \ldots, d_{nt}(\gamma)\}$. The model we consider takes the vector form:

$$Y_t = \lambda_{10} W_t Y_t + \lambda_{20} d_t(\gamma_0) W_t Y_t + X_t \beta_{10} + d_t(\gamma_0) X_t \beta_{20} + \mu_0 + \alpha_{t0} l_n + V_t, \quad (2.1)$$

$t = 1, \ldots, T$, where $Y_t = (y_{1t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of responses at time t, $X_t = (x_{1t}, \ldots, x_{nt})'$ is an $n \times k$ matrices containing values of k time-varying regressors, and $V_t = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ is an $n \times 1$ vector of idiosyncratic errors. The λ_{10} characterizes the baseline spatial lag effect and β_{10} $(k \times 1)$ the baseline regression coefficients; and λ_{20} and β_{20} $(k \times 1)$ are the corresponding threshold effects.[2] $\mu_0 = \{\mu_{i0}\}_{i=1}^n$ is an $n \times 1$ vector of individual-specific effects and $\alpha_0 = \{\alpha_{t0}\}_{t=1}^T$ is a $T \times 1$ vector of time-specific effects, which are allowed to be correlated with X_t in an arbitrary manner. Therefore, the model is referred to, in this chapter, as the *threshold spatial panel data* (TSPD) model with *two-way fixed effects* (2FE).

The neighborhood structure of the n spatial units at period t is captured by a time-varying spatial weight matrix W_t, and the magnitude of the interaction effects from its neighbors is measured by the spatial lag parameters. Thus, Model (2.1) implies that each spatial unit i in any period t receives a certain level of interaction effects from its neighbors (measured by λ_{10} or $\lambda_{10} + \lambda_{20}$), depending on the level of its threshold variable q_{it}. When γ_0 is known, our model can simply be treated as a second-order SPD model of two spatial lag terms with weight matrices being W_t and $d_t(\gamma_0)W_t$, respectively. However, there is one complication: $d_t(\gamma_0)W_t$ can no longer be time-invariant and row-normalized even if W_t's are. This causes the convenient transformation approach (Lee and Yu, 2010) inapplicable for model estimation and inference. Therefore, an alternative method, allowing W_t to be time-varying without the need of row-normalization, is desired. When γ_0 is unknown and has to be estimated together with other parameters, the involvement of step functions $d_{it}(\gamma_0)$ in the likelihood renders it to be discrete in γ and therefore the standard likelihood inference methods are no longer valid.

As discussed in the introduction, we approach the estimation and inference problems for the TSPD-2FE model by an adjusted QML method, where

[2]Model (2.1) can be extended by including the spatial lag terms of X_t, i.e., the spatial Durbin effects, and their threshold effects without additional technical complications.

the concentrated loglikelihood (with μ and α being concentrated out) is modified to give a consistent estimation of all the common parameters. Then, we show that the estimation of the threshold parameters does not have an impact asymptotically on the joint asymptotic distribution of the other common parameters, thereby leading to bias-corrected estimation and inference methods for these parameters. A likelihood ratio test is proposed for inference for the threshold parameter.

2.2.2 Adjusted QML estimation

Denote $\lambda = (\lambda_1, \lambda_2)'$, $\beta = (\beta_1', \beta_2')'$, $\phi = (\beta', \lambda')'$ and $\theta = (\phi', \sigma^2)'$. Define $A_t(\lambda, \gamma) = I_n - \lambda_1 W_t - \lambda_2 d_t(\gamma) W_t$, $\mathbb{Y}_t(\lambda, \gamma) = A_t(\lambda, \gamma) Y_t$ and $\mathbb{X}_t(\gamma) = [X_t,\ d_t(\gamma) X_t]$. Under exogeneity of $\{X_t\}$ and $\{W_t\}$, we have the quasi Gaussian loglikelihood **as if** $\{v_{it}\}$ are iid $N(0, \sigma_0^2)$,

$$\begin{aligned} \ell_{nT}(\theta, \gamma, \mu, \alpha) = & -\tfrac{nT}{2} \ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln |A_t(\lambda, \gamma)| \\ & - \tfrac{1}{2\sigma^2} \textstyle\sum_{t=1}^T V_t'(\phi, \gamma, \mu, \alpha) V_t(\phi, \gamma, \mu, \alpha), \end{aligned} \tag{2.2}$$

where $V_t(\phi, \gamma, \mu, \alpha) = \mathbb{Y}_t(\lambda, \gamma) - \mathbb{X}_t(\gamma)\beta - \mu - \alpha_t l_n$.

To estimate the individual and time fixed effects, we impose a zero-sum constraint, $\sum_{t=1}^T \alpha_t = 0$, to avoid the unidentification of μ_{i0} and α_{t0} as $\mu_{i0} + \alpha_{t0} = (\mu_{i0} + c) + (\alpha_{t0} - c)$ for an arbitrary c. Then, given ϕ and γ, the first-order conditions for μ and α_t imply

$$\hat{\mu}(\phi, \gamma) = \tfrac{1}{T} \textstyle\sum_{t=1}^T [\mathbb{Y}_t(\lambda, \gamma) - \mathbb{X}_t(\gamma)\beta] \quad \text{and} \quad \hat{\alpha}_t(\phi, \gamma) = \tfrac{1}{n} l_n' [\widetilde{\mathbb{Y}}_t(\lambda, \gamma) - \widetilde{\mathbb{X}}_t(\gamma)\beta],$$

where an important shorthand notation is used: for a sequence of vectors or matrices $\Pi_t, t = 1, \ldots, T$, their time demeaned versions are denoted by $\widetilde{\Pi}_t = \Pi_t - \frac{1}{T} \sum_{t=1}^T \Pi_t$.

Substituting $\hat{\mu}(\phi, \gamma)$ and $\hat{\alpha}_t(\phi, \gamma)$ into $\ell_{nT}(\theta, \gamma, \mu, \alpha)$ for μ and α, respectively, gives the concentrated quasi loglikelihood function for (θ, γ):

$$\ell_{nT}^c(\theta, \gamma) = -\tfrac{nT}{2} \ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln |A_t(\lambda, \gamma)| - \tfrac{1}{2\sigma^2} \textstyle\sum_{t=1}^T \widetilde{V}_t'(\phi, \gamma) J_n \widetilde{V}_t(\phi, \gamma), \tag{2.3}$$

where $J_n = I_n - \frac{1}{n} l_n l_n'$ and $\widetilde{V}_t(\phi, \gamma) = \widetilde{\mathbb{Y}}_t(\lambda, \gamma) - \widetilde{\mathbb{X}}_t(\gamma)\beta$. Maximizing $\ell_{nT}^c(\theta, \gamma)$ gives the *direct* QML estimators $\hat{\theta}_{nT}^d$ and $\hat{\gamma}_{nT}^d$ of θ and γ.

Note that, for a known γ_0, $\ell_{nT}^c(\theta, \gamma_0)$ corresponds to the regular SPD-2FE model studied by Lee and Yu (2010), who show that when T is fixed maximizing $\ell^c_{nT}(\theta, \gamma_0)$ w.r.t. θ only gives consistent estimators for spatial and regression parameters β and λ (in general) but not for the variance parameter σ^2 - the well known *incidental parameters problem* of Neyman and Scott (1948). An intuitive interpretation is that the direct QML estimator of σ^2 failed to 'recover' the effect of degrees of freedom loss due to the estimation of n fixed effects parameters μ. One would expect that the TSPD-2FE model face the same issue in the sense that maximizing $\ell^c_{nT}(\theta, \gamma)$ would not lead to a consistent estimation of σ^2 when T is fixed. However, we find that a simple adjustment on $\ell^c_{nT}(\theta, \gamma)$ will achieve the consistency of joint estimation of θ and γ (see next subsection for a detailed theoretical reasoning), that is

$$\begin{aligned} \ell^*_{nT}(\theta, \gamma) = & -\tfrac{n(T-1)}{2} \ln(2\pi\sigma^2) + \tfrac{T-1}{T} \textstyle\sum_{t=1}^T \ln |A_t(\lambda, \gamma)| \\ & - \tfrac{1}{2\sigma^2} \textstyle\sum_{t=1}^T \widetilde{V}'_t(\phi, \gamma) J_n \widetilde{V}_t(\phi, \gamma). \end{aligned} \tag{2.4}$$

Therefore, the adjusted QML estimators of θ and γ are defined as follows

$$(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) = \underset{(\theta,\gamma)\in\Theta\times\Gamma}{\operatorname{argmax}} \, \ell^*_{nT}(\theta, \gamma),$$

where Θ is the parameter space for θ.[3] To solve the above maximization problem, we first maximize the above objective function to obtain an estimate $\hat{\theta}_{nT}(\gamma)$ of θ for a given γ. Then, we define $\ell^{*c}_{nT}(\gamma) \equiv \ell^*_{nT}(\hat{\theta}_{nT}(\gamma), \gamma)$, and search over Γ for $\hat{\gamma}_{nT}$ that maximizes $\ell^{*c}_{nT}(\gamma)$. Note that the objective function $\ell^{*c}_{nT}(\gamma)$ is a step function with at most nT steps as it depends on γ only through the indicator function $\mathbb{1}\{q_{it} \leqslant \gamma\}$. Thus, the latter maximization problem is reduced to search for $\hat{\gamma}_{nT}$ over $\Gamma_{nT} = \Gamma \cap \{q_{it}, 1 \leqslant i \leqslant n, 1 \leqslant t \leqslant T\}$. When nT is large, Hansen (1999) suggests that the search can be restricted to a grid of N_0 specific quantiles for some $N_0 < nT$, $\Gamma_{N_0} = \{q_{(1)}, \ldots, q_{(N_0)}\}$, where $q_{(j)}$ is the $[\eta + \frac{j-1}{N_0-1}(1-2\eta)]$th quantile of the sample q_{it} and $\eta = 1\%$ or 5%. Then, $\hat{\gamma}_{N_0} = \operatorname{argmax}_{\gamma\in\Gamma_{N_0}} \ell^{*c}_{nT}(\gamma)$ is a good approximation to $\hat{\gamma}_{nT}$. Given $\hat{\gamma}_{nT}$, the QMLE of θ is just $\hat{\theta}_{nT} \equiv \hat{\theta}_{nT}(\hat{\gamma}_{nT})$.

[3]Our adjusted QML approach falls in spirit to the "*Bias-Correction of the Concentrated Likelihood function*" of Arellano and Hahn (2007). Cox and Reid's (1987) adjusted profile likelihood approach also belongs to this category but requires the parameter of interest to be orthogonalized to the nuisance parameters.

2.2.3 Asymptotic properties of the adjusted QML estimators

In this subsection, we study the asymptotic properties of the adjusted QML estimators. We first find the probability limit of $(\hat{\theta}_{nT}, \hat{\gamma}_{nT})$, then exam the convergence rate of $\hat{\gamma}_{nT}$, then derive the asymptotic distribution of $\hat{\theta}_{nT}$ (and that of $\hat{\gamma}_{nT}$), and then we introduce a statistical inference procedure for γ based on a likelihood ratio test.

Denote $G_t(\lambda, \gamma) = W_t A_t^{-1}(\lambda, \gamma)$ and $Z_t(\psi, \gamma) = G_t(\lambda, \gamma)(\mathbb{X}_t(\gamma)\beta + \mu + \alpha_t l_n)$. Let $H_t = [X_t,\ Z_t]$ and $M(\gamma) = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \mathrm{E}(h_{it}h_{it}'|q_{it} = \gamma)$, where h_{it}' is the ith row of H_t. Let $f(\cdot)$ be the probability density function of q_{it}. Throughout this chapter, a parametric quantity at true parameter(s) is denoted by dropping its argument(s), e.g., $A_t = A_t(\lambda_0, \gamma_0)$, $G_t = G_t(\lambda_0, \gamma_0)$, $M = M(\gamma_0)$ and $f = f(\gamma_0)$. To provide rigorous asymptotic analysis of the adjusted QML estimators, we make the following assumptions:

Assumption A: The innovations v_{it} are independent and identically distributed (iid) across i and t, having mean zero, variance σ_0^2, and $E|v_{it}|^{8+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.

Assumption B: (i) The regressors and the threshold variable are exogenous with elements (x_{it}, q_{it}) being iid across i and t, (ii) $\mathrm{E}(\|h_{it}\|^4) < \infty$, (iii) For all $\gamma \in \Gamma$, $\mathrm{E}(\|h_{it}\|^4|q_{it} = \gamma) \leqslant c$ and $f(\gamma) \leqslant c$ for some $c < \infty$, (iv) $M(\gamma)f(\gamma)$ is continuous at $\gamma = \gamma_0$, (v) $0 < Mf < \infty$, (vi) the limit of $\frac{1}{nT}\sum_{t=1}^T \widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{\mathbb{X}}_t(\gamma)$ exists and is nonsingular.

Assumption C: $\{W_t\}$ are exogenous time-varying spatial weight matrices with zero diagonal elements. Both $\|W_t\|_1$ and $\|W_t\|_\infty$ are bounded for all t.

Assumption D: The true λ_0 lies in the interior of a compact space Λ. For each t and $(\gamma, \lambda) \in \Gamma \times \Lambda$, (i) $A_t(\lambda, \gamma)$ is invertible; (ii) both $\|A_t^{-1}(\lambda, \gamma)\|_1$ and $\|A_t^{-1}(\lambda, \gamma)\|_\infty$ are bounded.

Assumption E: n is large, and T can be finite or large but cannot grow faster than n, i.e., $\frac{T}{n} \to a$, where $0 \leqslant a < \infty$.

Assumption F: Threshold effects λ_{20} and β_{20} satisfy that $\lambda_{20} = (nT)^{-\tau}l_0$ and $\beta_{20} = (nT)^{-\tau}b_0$ for some $\tau \in (0, 1/2)$ with $l_0 \in \mathbf{R}$, $l_0 \neq 0$ and $b_0 \in \mathbf{R}^k$, $b_0 \neq 0$.

The iid assumption in A is standard in the spatial econometrics literature (see, e.g., Lee and Yu, 2010; Li, 2018), but the finite eighth moment condition on errors is more stringent than in the standard SPD models, where only a finite fourth moment condition is required. With this stronger assumption, Lemma B.1 in Appendix B shows the weak convergence in function space for some linear-quadratic (LQ) forms that depend on γ through the indication function $\mathbb{1}\{q_{it} \leqslant \gamma\}$, which is crucial for the asymptotic studies on our estimators. Assumption B(i) assumes regressors and threshold variable are both exogenous, which also appears in Hansen (1999, 2000). As $\{h_{it}\}$ can be treated as the model regressors in a reduced form of (2.1) (see (2.12)), Assumption B(ii)–(v) are also common in the threshold literature, corresponding to Assumptions 4, 6 and 7 in Hansen (1999). Assumption B(vi) is the identification condition for β. Assumption C and D are standard in spatial econometrics literature. Assumption E allows (i) both n and T are large and (ii) n is large and T is finite. Both scenarios encounter the incidental parameters problem of Neyman and Scott (1948) due to the estimation of the individual and time fixed effects. The assumption that T cannot grow faster than n is used to establish the consistency of $\hat{\gamma}_{nT}$. Assumption F is in the spirit of Hansen (2000) so that the asymptotic distribution of the threshold estimator is free of nuisance parameters, and thus making statistical inference on γ is possible. On the contrary, if the threshold effects are fixed (i.e., $\tau = 0$), according to Chan (1993), we can expect that the asymptotic distribution of $\hat{\gamma}_{nT}$ will involve nuisance parameters such as the marginal distribution of the x_{it}.

Validity of the objective function $\ell^*_{nT}(\theta,\gamma)$. Based on some of the assumptions above, a critical discussion can then be given on the validity of the objective function $\ell^*_{nT}(\theta,\gamma)$. First, the adjusted score vector with respect to θ has the form:

$$S^*_{\theta,nT}(\theta,\gamma) = \begin{cases} \frac{1}{\sigma^2}\sum_{t=1}^T \widetilde{\mathbb{X}}'_t(\gamma) J_n \widetilde{V}_t(\phi,\gamma), \\ \frac{1}{\sigma^2}\sum_{t=1}^T \widetilde{\mathcal{Y}}'_t J_n \widetilde{V}_t(\phi,\gamma) - \bar{T}\sum_{t=1}^T \texttt{tr}[G_t(\lambda,\gamma)], \\ \frac{1}{\sigma^2}\sum_{t=1}^T \widetilde{\mathcal{Y}}^{\circ\prime}_t(\gamma) J_n \widetilde{V}_t(\phi,\gamma) - \bar{T}\sum_{t=1}^T \texttt{tr}[d_t(\gamma)G_t(\lambda,\gamma)], \\ \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}'_t(\phi,\gamma) J_n \widetilde{V}_t(\phi,\gamma) - \frac{n(T-1)}{2\sigma^2}, \end{cases} \tag{2.5}$$

where $\bar{T} = \frac{T-1}{T}$, $\mathcal{Y}_t = W_t Y_t$ and $\mathcal{Y}^\circ_t(\gamma) = d_t(\gamma)\mathcal{Y}_t$ to denote their time demeaned versions $\widetilde{\mathcal{Y}}_t$ and $\widetilde{\mathcal{Y}}_t^{\circ}(\gamma)$. Then, we find $\text{plim}_{n,T\to\infty}\frac{1}{n(T-1)}S^*_{\theta,nT}(\theta_0,\gamma_0) = 0$ holds, which is a necessary condition for the consistency of $\hat{\theta}_{nT}$, i.e., $\hat{\theta}_{nT} - \theta_0 = o_p(1)$. We have

$$\mathrm{E}[S^*_{\theta,nT}(\theta_0,\gamma_0)] = \big(0_{1\times 2k}, -\tfrac{T-1}{nT}\textstyle\sum_{t=1}^T l_n' G_t l_n, -\tfrac{T-1}{nT}\sum_{t=1}^T l_n' d_t(\gamma_0) G_t l_n, -\tfrac{(T-1)}{2\sigma_0^2}\big).$$

Note that $l_n' G_t l_n$ and $l_n' d_t(\gamma_0) G_t l_n$ are both $O(n)$ for any t, since G_t's are bounded in both row and column sums. Thus, we have $\mathrm{E}[S^*_{\theta,nT}(\theta_0,\gamma_0)] = O(T)$. The adjustment corrects the degrees of freedom loss, making the effective sample size become $n(T-1)$. It follows that $\frac{1}{n(T-1)}S^*_{\theta,nT}(\theta_0,\gamma_0) = O_p(\frac{1}{n})$. Hence, a consistent estimation is possible for all the common parameters based on maximizing $\ell^*_{nT}(\theta,\gamma)$, whether T is small or large.

To proceed with a detailed study on the asymptotic properties of $\hat{\theta}_{nT}$ and $\hat{\gamma}_{nT}$, we first examine their consistency, which follows from the consistency of $\hat{\lambda}_{nT}$ and $\hat{\gamma}_{nT}$ as argued below. Given λ and γ, $\ell^*_{nT}(\theta,\gamma)$ is partially maximized at

$$\hat{\beta}_{nT}(\lambda,\gamma) = [\textstyle\sum_{t=1}^T \widetilde{\mathbb{X}}_t'(\gamma) J_n \widetilde{\mathbb{X}}_t(\gamma)]^{-1}\sum_{t=1}^T \widetilde{\mathbb{X}}_t'(\gamma) J_n \widetilde{\mathbb{Y}}_t(\lambda,\gamma), \text{ and} \quad (2.6)$$

$$\hat{\sigma}^2_{nT}(\lambda,\gamma) = \tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T \big\| J_n[\widetilde{\mathbb{Y}}_t(\lambda,\gamma) - \widetilde{\mathbb{X}}_t(\gamma)\hat{\beta}_{nT}(\lambda,\gamma)]\big\|^2. \quad (2.7)$$

Hence, the concentrated loglikelihood function of λ and γ is[4]

$$\ell^{*c}_{nT}(\lambda,\gamma) = -\tfrac{n(T-1)}{2}(\ln 2\pi + 1) - \tfrac{n(T-1)}{2}\ln\hat{\sigma}^2_{nT}(\lambda,\gamma) + \bar{T}\textstyle\sum_{t=1}^T \ln|A_t(\lambda,\gamma)|. \quad (2.8)$$

The QMLEs $\hat{\lambda}_{nT}$ and $\hat{\gamma}_{nT}$ together maximize the above likelihood function, and the QMLEs of β and σ^2 are respectively, $\hat{\beta}_{nT} \equiv \hat{\beta}_{nT}(\hat{\lambda}_{nT},\hat{\gamma}_{nT})$ and $\hat{\sigma}^2_{nT} \equiv \hat{\sigma}^2_{nT}(\hat{\lambda}_{nT},\hat{\gamma}_{nT})$. It is interesting to note from (B.4) and (B.5) in Appendix B that, under Assumptions B(vi) and F, the consistencies of $\hat{\beta}_{nT}$ and $\hat{\sigma}^2_{nT}$ follows that of $\hat{\lambda}_{nT}$, whether $\hat{\gamma}_{nT}$ is consistent or not.

The above discussions suggest that the consistency of $\hat{\theta}_{nT}$ does not rely on that of $\hat{\gamma}_{nT}$ under the diminishing threshold assumption, and thus it can be established separately. To see it more clearly, we introduce the population

[4] It is worth to mention that the concentrated function of $\ell^c_{nT}(\theta,\gamma)$ in (2.3), $\ell^c_{nT}(\lambda,\gamma) = \frac{T}{T-1}\ell^{*c}_{nT}(\lambda,\gamma) + \frac{nT}{2}\ln\frac{T}{T-1}$. Thus, $\ell^{*c}_{nT}(\lambda,\gamma)$ and $\ell^c_{nT}(\lambda,\gamma)$ yield the same maximizer. However, as discussed above, $\ell^*_{nT}(\theta,\gamma)$ is a valid joint objective function that is possible to provide consistent estimates for all the common parameters.

counterpart of $\ell^{*c}_{nT}(\lambda,\gamma)$, which is $\bar{\ell}^{*c}_{nT}(\lambda,\gamma) = \max_{\beta,\sigma^2} \mathrm{E}[\ell^{*}_{nT}(\theta,\gamma)]$. Given λ and γ, $\mathrm{E}[\ell^{*}_{nT}(\theta,\gamma)]$ is partially maximized at

$$\bar{\beta}_{nT}(\lambda,\gamma) = [\textstyle\sum_{t=1}^{T} \mathrm{E}(\widetilde{\mathbb{X}}'_t(\gamma) J_n \widetilde{\mathbb{X}}_t(\gamma))]^{-1} \sum_{t=1}^{T} \mathrm{E}(\widetilde{\mathbb{X}}'_t(\gamma) J_n \widetilde{\mathbb{Y}}_t(\lambda,\gamma)), \quad \text{and} \tag{2.9}$$

$$\bar{\sigma}^2_{nT}(\lambda,\gamma) = \tfrac{1}{n(T-1)} \textstyle\sum_{t=1}^{T} \mathrm{E}\big\| J_n[\widetilde{\mathbb{Y}}_t(\lambda,\gamma) - \widetilde{\mathbb{X}}_t(\gamma)\bar{\beta}_{nT}(\lambda,\gamma)]\big\|^2. \tag{2.10}$$

Thus, we have, upon substituting $\bar{\beta}_{nT}(\lambda,\gamma)$ and $\bar{\sigma}^2_{nT}(\lambda,\gamma)$ back in $\mathrm{E}[\ell^{*}_{nT}(\theta,\gamma)]$,

$$\bar{\ell}^{*c}_{nT}(\lambda,\gamma) = \tfrac{n(T-1)}{2}(\ln 2\pi + 1) - \tfrac{n(T-1)}{2} \ln \bar{\sigma}^2_{nT}(\lambda,\gamma) + \bar{T} \textstyle\sum_{t=1}^{T} \mathrm{E}(\ln |A_t(\lambda,\gamma)|). \tag{2.11}$$

If γ_0 were known, standard asymptotic arguments (e.g., Theorem 5.9 of van der Vaar, 1998) lead to the consistency of $\hat{\lambda}_{nT}(\gamma_0)$: uniform convergence of $\frac{1}{n(T-1)}[\ell^{*c}_{nT}(\lambda,\gamma_0) - \bar{\ell}^{*c}_{nT}(\lambda,\gamma_0)]$ to 0 in $\lambda \in \Lambda$, and global identification of λ_0 in that it uniquely maximizes the limit of $\frac{1}{n(T-1)}\bar{\ell}^{*c}_{nT}(\lambda,\gamma_0)$ on Λ. When γ_0 is unknown, the convergence of $\frac{1}{n(T-1)}[\ell^{*c}_{nT}(\lambda,\gamma) - \bar{\ell}^{*c}_{nT}(\lambda,\gamma)]$ is still useful to establish the consistency of $\hat{\lambda}_{nT}$, uniformly in $\gamma \in \Gamma$, but it cannot provide useful information to study the asymptotic behavior of $\hat{\gamma}_{nT}$ as the threshold effects become zero at the limit. Therefore, we first show that $\hat{\lambda}_{nT}(\gamma)$ is consistent to λ_0, uniformly in $\gamma \in \Gamma$[5]. It then provides the basis to further establish the consistency of $\hat{\gamma}_{nT}$. Let $\sigma^2_{nT}(\lambda,\gamma) = \frac{\sigma_0^2}{(n-1)T} \sum_{t=1}^{T} \mathtt{tr}(A_t'^{-1} A'_t(\lambda,\gamma) J_n A_t(\lambda,\gamma) A_t^{-1})$ and $\mathbb{H}_t(\gamma) = [X_t,\ Z_t,\ d_t(\gamma)X_t,\ d_t(\gamma)Z_t]$. We provide the identification conditions for λ as follows.

Assumption G: Either (i) the limit of $\frac{1}{n(T-1)} \sum_{t=1}^{T}[\widetilde{\mathbb{H}}'_t(\gamma) J_n \widetilde{\mathbb{H}}_t(\gamma)]$ exists and is nonsingular, uniformly in $\gamma \in \Gamma$; or (ii) the limit of $\frac{1}{n(T-1)} \sum_{t=1}^{T}(\ln |\sigma^2_{nT}(\lambda_0,\gamma) A_t^{-1}(\lambda_0,\gamma) A_t'^{-1}(\lambda_0,\gamma)| - \ln |\sigma^2_{nT}(\lambda,\gamma) A_t^{-1}(\lambda,\gamma) A_t'^{-1}(\lambda,\gamma)|) \neq 0$ for $\lambda \neq \lambda_0$, uniformly in $\gamma \in \Gamma$.

Assumption G generalizes the global identification conditions for SPD models in Lee and Yu (2010) to the models with threshold effects. To gain more intuitions on these assumptions, noting $Y_t = A_t^{-1}(X_t\beta_{10} + d_t(\gamma_0)X_t\beta_{20} + \mu_0 + \alpha_{t0} l_n + V_t)$, we have

$$Y_t = X_t\beta_{10} + Z_t\lambda_{10} + d_t(\gamma_0)X_t\beta_{20} + d_t(\gamma_0)Z_t\lambda_{20} + \mu_0 + \alpha_{t0} l_n + A_t^{-1}V_t, \quad t = 1,\ldots,T, \tag{2.12}$$

[5]One can also see from (B.14) that $\lim_{nT\to\infty} \frac{1}{n(T-1)} \mathrm{E}S_{\theta,nT}(\theta_0,\gamma) = 0$, uniformly in $\gamma \in \Gamma$, a necessary condition to have the consistency of $\hat{\theta}_{nT}(\gamma)$.

because $A_t^{-1} = I_n + \lambda_{10} G_t + \lambda_{20} d_t(\gamma_0) G_t$, which comes from $I_n = A_t + \lambda_{10} W_t + \lambda_{20} d_t(\gamma_0) W_t$ by multiplying A_t^{-1} on both sides of the equation. Clearly, the above equation can be treated as a standard panel data model with regressor $\mathbb{H}_t(\gamma_0)$ in period t. Thus, it is standard to impose the non-singularity or full rank condition on the limit of $\frac{1}{n(T-1)} \sum_{t=1}^{T} \mathrm{E}[\widetilde{\mathbb{H}}_t'(\gamma_0) J_n \widetilde{\mathbb{H}}_t(\gamma_0)]$ to identify ϕ or λ. As discussed above, the consistency of $\hat{\gamma}_{nT}$ has not been established so far. Therefore, we impose Assumption G(i) instead. In addition, note that $\mathrm{Var}(Y_t|X_t) = \sigma_0^2 A_t^{-1} A_t'^{-1}$. Thus, λ can also be identified by the uniqueness of $\mathrm{Var}(\mathbf{Y}|X_1, \ldots, X_T)$, where $\mathbf{Y} = (Y_1', \ldots, Y_T')'$. Again, $\hat{\gamma}_{nT}$ is not consistent at this moment so that we impose Assumption G(ii) instead. With the above identification conditions and the uniform convergence of $\frac{1}{n(T-1)}[\ell_{nT}^{*c}(\lambda, \gamma) - \bar{\ell}_{nT}^{*c}(\lambda, \gamma)]$ to 0 in $\gamma \in \Gamma$, we have the following theorem.

Theorem 2.1. *Suppose Assumptions A-G hold. We have* $\hat{\theta}_{nT} - \theta_0 \xrightarrow{p} 0$.

As the adjustment to the concentrated loglikelihood function in (2.4) can help to 'recover' the degrees of freedom loss caused by the estimation of the incidental parameter μ, we see all the common estimators are consistent even when T is fixed. As discussed above, although $\hat{\lambda}_{nT}$ is shown to be consistent, the convergence of the original objective function $\frac{1}{n(T-1)} \ell_{nT}^{*c}(\lambda, \gamma)$ is still too fast to be useful for studying the limiting behavior of $\hat{\gamma}_{nT}$, when the threshold effects shrink to zero at rate $(nT)^{-\tau}$. However, we find the re-scaled objective function:

$$\ell_{nT}^{**}(\gamma) = \frac{(nT)^{2\tau}}{n(T-1)} [\ell_{nT}^{*c}(\hat{\lambda}_{nT}(\gamma), \gamma) - \ell_{nT}^{*c}(\lambda_0, \gamma_0)] \tag{2.13}$$

can be very useful. Specifically, multiplying $(nT)^{2\tau}$ gives us the non-diminishing threshold effects, while taking the differences removes the terms that are not asymptotically negligible, i.e., those have order bigger than $O_p((nT)^{1-2\tau})$. The consistency of $\hat{\gamma}_{nT}$ follows if the maximizer of $\ell_{nT}^{**}(\gamma)$ has an asymptotically negligible distance from γ_0, i.e., the identification condition for γ. Let $d_t(\gamma_1, \gamma_2) = d_t(\gamma_1) - d_t(\gamma_2)$ and $\mathcal{H}_t(\gamma) = [\mathbb{H}_t(\gamma),\ d_t(\gamma_0, \gamma) H_t]$. Let $\mathbb{C}_t(\gamma)$ be a 3×3 matrix with (a, b)th element being $\mathtt{tr}[\mathbb{C}_{a,t}^s(\gamma) \mathbb{C}_{b,t}^s(\gamma)]$, where $\mathbb{C}_{1,t} = G_t - \frac{1}{nT} \sum_{t=1}^{T} \mathtt{tr}(G_t) I_n$, $\mathbb{C}_{2,t}(\gamma) = d_t(\gamma) G_t - \frac{1}{nT} \sum_{t=1}^{T} \mathtt{tr}(d_t(\gamma) G_t) I_n$, $\mathbb{C}_{3,t}(\gamma) = d_t(\gamma_0, \gamma) G_t - \frac{1}{nT} \sum_{t=1}^{T} \mathtt{tr}(d_t(\gamma_0, \gamma) G_t) I_n$. We introduce the identification condition for γ.

Assumption H: There exists a constant $c > 0$ such that either
(*i*) $\rho_{min}\big(\frac{1}{n(T-1)}\sum_{t=1}^{T}\widetilde{\mathcal{H}}_t'(\gamma)J_n\widetilde{\mathcal{H}}_t(\gamma)\big) \geq c|\gamma-\gamma_0|$, or
(*ii*) $\rho_{min}\big(\frac{1}{n(T-1)}\sum_{t=1}^{T}\mathbb{C}_t(\gamma)\big) \geq c|\gamma-\gamma_0|$.

As shown in Appendix B, to study the asymptotic properties of $\ell_{nT}^{**}(\hat{\gamma}_{nT})$, one has to establish a rough convergence rate for $\hat{\lambda}_{nT}-\lambda_0$. As the objective function is highly nonlinear in the λ and there is no closed form solution for its QMLE, we have to rely on the study of the θ-component of the concentrated quasi score (CQS) function given in (2.5). We start with a Taylor expansion of $S_{\theta,nT}^{*}(\hat{\theta}_{nT},\hat{\gamma}_{nT})=0$ at θ_0, then justify the non-singularity of the limit of the information matrix, which is also guaranteed by Assumption H, and finally study the order of the component CQS function $S_{\theta,nT}^{*}(\theta_0,\hat{\gamma}_{nT})$. By Theorem 2.1, we thus prove $(nT)^{\tau}(\hat{\theta}_{nT}-\theta_0)=O_p(1)$ under Assumption E, regardless of the consistency of $\hat{\gamma}_{nT}$. With this preliminary convergence rate, the limit of $\ell_{nT}^{**}(\hat{\gamma}_{nT})$ can be shown to be less than zero unless $|\hat{\gamma}_{nT}-\gamma_0|=o_p(1)$ under Assumption H. However, the definition of $\hat{\gamma}_{nT}$ implies that we must have $\ell_{nT}^{**}(\hat{\gamma}_{nT})\geq 0$. Consequently, the consistency of $\hat{\gamma}_{nT}$ follows.

Theorem 2.2. *Suppose Assumptions A-H hold. We have* $\hat{\gamma}_{nT}-\gamma_0 \xrightarrow{p} 0$.

To establish the convergence rate for $\hat{\gamma}_{nT}$, one needs a more precise knowledge of the convergence rate for $\hat{\theta}_{nT}$. With the consistency of $\hat{\gamma}_{nT}$ and the Taylor expansion mentioned above, we can further show $(nT)^{\tau}(\hat{\theta}_{nT}-\theta_0)=o_p(1)$. Based on this result, we next establish the convergence rate for $\hat{\gamma}_{nT}$ in the following theorem.

Theorem 2.3. *Under Assumptions A-H,* $a_{nT}(\hat{\gamma}_{nT}-\gamma_0)=O_p(1)$*, where* $a_{nT}=(nT)^{1-2\tau}$.

The theorem shows that the convergence rate of $\hat{\gamma}_{nT}$ is a_{nT}, which is in line with the findings of Hansen (1999). Intuitively speaking, the convergence rate is faster when the threshold effects in the model are larger (threshold diminishing rate is slower as τ is smaller), providing more sample information regarding to the threshold parameter γ and hence more precise estimation of it. Theorem 2.3 is crucial for establishing the asymptotic distribution of $\hat{\theta}_{nT}$. Again, by using the Taylor expansion of the $S^*_{\theta,nT}(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) = 0$ at θ_0, the asymptotic property of $\sqrt{n(T-1)}(\hat{\theta}_{nT} - \theta_0)$ depends on that of $\frac{1}{\sqrt{n(T-1)}} S^*_{\theta,nT}(\theta_0, \hat{\gamma}_{nT})$ provided the non-singularity of the limit of the information matrix. With Theorem 2.3, we can show $\frac{1}{\sqrt{n(T-1)}} \| S^*_{\theta,nT}(\theta_0, \hat{\gamma}_{nT}) - S^*_{\theta,nT}(\theta_0, \gamma_0) \| \xrightarrow{p} 0$. Thus, it is equivalent to studying the the asymptotic property of $\frac{1}{\sqrt{n(T-1)}} S^*_{\theta,nT}(\theta_0, \gamma_0)$, which leads to the limiting distribution of $\hat{\theta}_{nT}(\gamma_0)$ and that of $\hat{\theta}_{nT}$. We have the following theorem.

Theorem 2.4. *Under Assumptions A-H, we have*

(i) $\sqrt{n(T-1)}(\hat{\theta}_{nT} - \theta_0) + \Sigma^{-1}_{nT}\sqrt{a} b_{\theta,nT} \xrightarrow{D} N\big(0, \lim_{nT\to\infty} \Sigma^{-1}_{nT}\Omega_{nT}\Sigma^{-1}_{nT}\big)$,

(ii) $\sqrt{n(T-1)}(\hat{\theta}_{nT} - \hat{\theta}_{nT}(\gamma_0)) \xrightarrow{p} 0$,

where $\Sigma_{nT} = \Sigma_{nT}(\theta_0, \gamma_0)$ *and* $\Omega_{nT} = \Omega_{nT}(\theta_0, \gamma_0)$ *are given in* (B.13) *and* (B.23), *respectively;* $a = \frac{T}{n}$ *and* $b_{\theta,nT}(\theta, \gamma) = \big(0_{1\times 2k}, \frac{\sqrt{T}}{nT}\sum_{t=1}^T l'_n G_t(\lambda, \gamma) l_n, \frac{\sqrt{T}}{nT}\sum_{t=1}^T l'_n d_t(\gamma) G_t(\lambda, \gamma) l_n, \frac{\sqrt{T}}{2\sigma^2}\big)'$.

Theorem 2.4(i) shows that the convergence rate of $\hat{\theta}_{nT}$ is $\sqrt{n(T-1)}$ but has an asymptotic bias when n and T go to infinity proportionally. If T is fixed, then $a = \frac{T}{n} \to 0$ and hence the estimates of all common parameters are asymptotically centered. This result is in contract to the direct approach of Li (2018) where the variance estimate is asymptotically biased no matter T is large or small. As is shown in (B.23) in Appendix B, $\Omega_{nT}(\theta_0, \gamma_0)$ involves not only the excess kurtosis κ_4 of the errors but also the skewness κ_3 and one additional component $\mathbb{B}_{nT}(\gamma_0)$. This is also in contrast with the result for standard SPD model (Lee and Yu, 2010) where the limiting variance only involves κ_4. The reason for this is due to the time-varying feature of the model. Specifically, if both $\{G_t\}$ and $\{d_t(\gamma_0)G_t\}$ are time constant, we will not have κ_3 and $\mathbb{B}_{nT}(\gamma_0)$ in $\Omega_{nT}(\theta_0, \gamma_0)$ anymore. Moreover, when errors $\{v_{it}\}$ are normally distributed ($\kappa_3 = \kappa_4 = 0$) and T is large, we have $\Omega_{nT}(\theta_0, \gamma_0) = \Sigma_{nT}(\theta_0, \gamma_0)$, because $\mathbb{B}_{nT}(\gamma_0) = O(\frac{1}{T})$. From Theorem 2.4($ii$), we can also see that given the convergence rate of $\hat{\gamma}_{nT}$ in Theorem 2.3, we can treat the true value of γ as given. In other words, the estimation error associated with $\hat{\gamma}_{nT}$ has asymptotically negligible effects on the asymptotic property of the QMLEs of the common θ. To sum up, Theorem 2.1 shows that $\hat{\theta}_{nT}(\gamma)$ is consistent for θ_0 for any $\gamma \in \Gamma$, while Theorem 2.4(*ii*) suggests that $\hat{\theta}_{nT} \equiv \hat{\theta}_{nT}(\hat{\gamma}_{nT})$ is asymptotically equivalent to $\hat{\theta}_{nT}(\gamma_0)$. Finally, as $\hat{\theta}_{nT}$ may be asymptotically biased, in practice one can carry out a bias-correction using the results of Theorem 2.4, as in Lee and Yu (2010) for a regular SPD model.

Next, we establish the asymptotic distribution of $\hat{\gamma}_{nT}$. Let

$$\Xi_1 = \lim_{nT\to\infty} \bar{T}[\delta_0' M \delta_0 + l_0^2 \sigma_0^2 (\pi_1 + \pi_2)] \text{ and}$$
$$\Xi_2 = \lim_{nT\to\infty} \bar{T}^2 (2 l_0 \sigma_0 \kappa_3 \delta_0' \pi_3 + l_0^2 \sigma_0^2 \kappa_4 \pi_2),$$

where $\delta_0 = (b_0', l_0)'$, $\pi_1(\gamma) = \frac{1}{nT}\sum_{t=1}^T \sum_{i=1}^n \mathrm{E}(\sum_{j=1}^n g_{ij,t}^2 | q_{it} = \gamma)$, $\pi_2(\gamma) = \frac{1}{nT}\sum_{t=1}^T \sum_{i=1}^n \mathrm{E}(g_{ii,t}^2 | q_{it} = \gamma)$, $\pi_3(\gamma) = \frac{1}{nT}\sum_{t=1}^T \sum_{i=1}^n \mathrm{E}(g_{ii,t} h_{it} | q_{it} = \gamma)$, and $g_{ij,t}$ is the (i,j)th entry of G_t. Under Assumption B(v), it is easy to see that Ξ_1 must be strictly positive. Thus, the following theorem provides the asymptotic distribution of $\hat{\gamma}_{nT}$.

Theorem 2.5. *Under Assumptions A-H, we have*

$$a_{nT}(\hat{\gamma}_{nT} - \gamma_0) \xrightarrow{D} \frac{\sigma_0^2}{f} \frac{\Xi}{\Xi_1^2} \underset{-\infty<r<\infty}{\operatorname{argmax}} \big[-\frac{|r|}{2} + W(r) \big],$$

where $\Xi = \Xi_1 + \Xi_2$ *and* $W(r)$ *as a two-sided standard Brownian motion on the real line, i.e.,* $W(r) = W_a(-r)\mathbb{1}\{r \leqslant 0\} + W_b(r)\mathbb{1}\{r > 0\}$, *and* $W_a(\cdot)$ *and* $W_b(\cdot)$ *are two independent standard Brownian motions on* $[0,\infty)$ *with* $W_a(0) = W_b(0) = 0$.

According to Chan (1993), when the threshold effects are fixed over sample size (i.e., $\tau = 0$), it may be possible to demonstrate that $nT(\hat{\gamma}_{nT} - \gamma_0) = O_p(1)$, but the asymptotic distribution of $nT(\hat{\gamma}_{nT} - \gamma_0)$ might be a functional of a compound Poisson process that depends on the marginal distribution of x_{it}, and hence is not useful for making inference on γ. In contrast, under the shrinking threshold effects assumption, Theorem 2.5 shows that the limiting distribution of $\hat{\gamma}_{nT}$ does not involve this undue component. However, in order to make inference on γ directly through the above theorem, one has to find a consistent estimate for the scale component $\frac{\Xi}{\Xi_1^2 f}$. Note that both Ξ_1 and Ξ_2 involve δ_0 or l_0, neither of which can be estimated accurately without prior knowledge of the nuisance parameter τ. Thus, we propose a likelihood ratio type test as follows.

2.2.4 Inference for γ based on the likelihood ratio test

Following Hansen (1999), we propose a likelihood ratio statistic to test the null hypothesis $H_0 : \gamma = \gamma_0$. Recall $\ell^{*c}_{nT}(\gamma) \equiv \ell^*_{nT}(\hat{\theta}_{nT}(\gamma), \gamma)$. Define

$$LR_{nT}(\gamma) = 2[\ell^{*c}_{nT}(\hat{\gamma}_{nT}) - \ell^{*c}_{nT}(\gamma)]. \tag{2.14}$$

Theorem 2.6. *Under Assumptions A-H, we have*

$$LR_{nT}(\gamma_0) \xrightarrow{D} \varpi^2\, \mho,$$

where $\varpi^2 = 1 + \frac{\Xi_2}{\Xi_1}$*, and* $\mho = \max_{-\infty<r<\infty}[-|r| + 2W(r)]$*, of which the distribution function is characterized by* $P(\mho \leqslant z) = (1 - e^{-z/2})^2$.

Note that ϖ^2 from the above theorem is equal to $\frac{\Xi}{\Xi_1}$, since $\Xi = \Xi_1 + \Xi_2$. Thus, ϖ^2 must be strictly positive, because Ξ_1 is so and $\sigma_0^2\Xi$ is the variance of some LQ form (see Lemma B.3 in Appendix B). In a special case when error terms are iid normally distributed, one has $\Xi_2 = 0$ because $\kappa_3 = \kappa_4 = 0$. It follows that $\varpi^2 = 1$ and the asymptotic distribution of $LR_{nT}(\gamma_0)$ is pivotal. This result is different from Theorem 1 of Hansen (1999), in which only the homoskedasticity assumption is needed to have the scale parameter equal to 1. In the standard panel data model, the corresponding $\sigma_0^2\Xi$ is just the variance of some linear form so that it does not involve the third and fourth moments of errors.

When errors are not normally distributed, ϖ^2 must be estimated consistently. Let $\phi_{20} = (\beta'_{20}, \lambda_{20})'$, the collection of the threshold effects. Then, by Assumption F, we have

$$\frac{\Xi_2}{\Xi_1} = \frac{(nT)^{-2\tau}\Xi_2}{(nT)^{-2\tau}\Xi_1} = \frac{\lim_{nT\to\infty}\bar{T}[2\lambda_{20}\sigma_0\kappa_3\phi'_{20}\pi_3 + \lambda_{20}^2\sigma_0^2\kappa_4\pi_2]}{\lim_{nT\to\infty}[\phi'_{20}M\phi_{20} + \lambda_{20}^2\sigma_0^2(\pi_1 + \pi_2)]}.$$

Note that $\phi'_{20}M\phi_{20} = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \mathrm{E}[(x'_{it}\beta_{20})^2 + 2\lambda_{20}Z_{it}x'_{it}\beta_{20} + Z_{it}^2\lambda_{20}^2|q_{it} = \gamma_0]$ and $\phi'_{20}\pi_3 = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \mathrm{E}[g_{ii,t}(x'_{it}\beta_{20} + Z_{it}\lambda_{20})|q_{it} = \gamma_0]$, where Z_{it} is the ith element of Z_t. As $Y_t = A_t^{-1}(\mathbb{X}_t\beta_0 + \mu_0 + \alpha_{t0}l_n + V_t)$, we have $\mathcal{Y}_t \equiv W_tY_t = Z_t + G_tV_t$, which implies that $\mathrm{E}(Z_{it}x'_{it}|q_{it}) = \mathrm{E}(\mathcal{Y}_{it}x'_{it}|q_{it})$, $\mathrm{E}(g_{ii,t}Z_{it}|q_{it}) = \mathrm{E}(g_{ii,t}\mathcal{Y}_{it}|q_{it})$ and $\mathrm{E}(Z_{it}^2|q_{it}) = \mathrm{E}(\mathcal{Y}_{it}^2|q_{it}) - \sigma_0^2\mathrm{E}(\sum_{j=1}^n g_{ij,t}^2|q_{it})$, where $\mathcal{Y}_{it}$ is the ith element of $\mathcal{Y}_t$. Given these, we have

$$\frac{\Xi_2}{\Xi_1} = \frac{\lim_{nT\to\infty}\sum_{i=1}^n\sum_{t=1}^T \mathrm{E}(\vartheta_{2,it}|q_{it} = \gamma_0)}{\lim_{nT\to\infty}\sum_{i=1}^n\sum_{t=1}^T \mathrm{E}(\vartheta_{1,it}|q_{it} = \gamma_0)},$$

where we define $\vartheta_{1,it} = (x_{it}'\beta_{20})^2 + 2\lambda_{20}\mathcal{Y}_{it}x_{it}'\beta_{20} + \mathcal{Y}_{it}^2\lambda_{20}^2 + \lambda_{20}^2\sigma_0^2 g_{ii,t}^2$ and $\vartheta_{2,it} = \bar{T}[2\lambda_{20}\sigma_0\kappa_3 g_{ii,t}(x_{it}'\beta_{20} + \mathcal{Y}_{it}\lambda_{20}) + \lambda_{20}^2\sigma_0^2\kappa_4 g_{ii,t}^2]$. To find their sample counterparts, we let $\hat{\kappa}_3$ and $\hat{\kappa}_4$ be the consistent estimations of κ_3 and κ_4, which are standard to find in the literature (see Li, 2018), $\hat{g}_{ij,t}$ the ith row and jth column of $G_t(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})$. Thus, their sample counterparts are just $\hat{\vartheta}_{1,it} = (x_{it}'\hat{\beta}_{2,nT})^2 + 2\hat{\lambda}_{2,nT}\mathcal{Y}_{it}x_{it}'\hat{\beta}_{2,nT} + \mathcal{Y}_{it}^2\hat{\lambda}_{2,nT}^2 + \hat{\lambda}_{2,nT}^2\hat{\sigma}_{nT}^2\hat{g}_{ii,t}^2$ and $\hat{\vartheta}_{2,it} = \bar{T}[2\hat{\lambda}_{2,nT}\hat{\sigma}_{nT}\hat{\kappa}_3\hat{g}_{ii,t}(x_{it}'\hat{\beta}_{2,nT} + \mathcal{Y}_{it}\hat{\lambda}_{2,nT}) + \hat{\lambda}_{2,nT}^2\hat{\sigma}_{nT}^2\hat{\kappa}_4\hat{g}_{ii,t}^2]$, respectively. Therefore, we finally propose to estimate ϖ^2 by

$$\hat{\varpi}^2 = 1 + \frac{\sum_{i=1}^n \sum_{t=1}^T K_h(q_{it} - \hat{\gamma}_{nT})\hat{\vartheta}_{2,it}}{\sum_{i=1}^n \sum_{t=1}^T K_h(q_{it} - \hat{\gamma}_{nT})\hat{\vartheta}_{1,it}},$$

where $K_h(u) = h^{-1}k(u/h)$ for some bandwidth $h \to 0$ and kernel function $k(\cdot)$. With this, a test of $H_0 : \gamma = \gamma_0$ rejects at the asymptotic level of α if $LR_{nT}(\gamma_0)/\hat{\varpi}^2$ exceeds $\mho_{1-\alpha}$, where $\mho_{1-\alpha} = -2\ln(1 - \sqrt{1-\alpha})$ is the $1-\alpha$ quantile of $\mho$. From the Table I of Hansen (2000), we have $\mho_{1-\alpha} = 5.94$, 7.35 and 10.59 for $\alpha = 0.1$, 0.05 and 0.01, respectively.

2.3 Testing for the Existence of Threshold Effects

In this section, we introduce a statistic to test whether the threshold effects are statistically significant. The null hypothesis of no threshold effect in the model (2.1) can be represented by $H_0 : \phi_{20} = 0$. However, the threshold parameter γ is not identified under this null hypothesis, so the asymptotic distributions of classical tests are nonstandard and it is impossible to tabulate their critical values. This classical problem was raised by Davies (1977) and has been well investigated by Andrew (1993) and Hansen (1996). We follow Hansen (1996) and study the local power of our test by considering the sequence of Pitman local alternatives: $H_1 : \phi_{20} = \frac{C}{\sqrt{nT}}$, where $C = (C_b', C_l)'$, C_b is a $k \times 1$ vector and C_l is a scalar. In this sequence of alternatives, we see the diminishing rate is faster than the one specified in Assumption F. Meanwhile, according to Hansen (1996), this diminishing rate can also facilitate our study on the distributional theory of the test statistic. Moreover, this sequence of alternatives corresponds to the null hypothesis when $C = 0$.

In this chapter, we consider a Wald-type test statistic. For each $\gamma \in \Gamma$, we construct a classical Wald statistic, $W_{nT}(\gamma)$. The asymptotic distribution of $W_{nT}(\gamma)$ depends on the nuisance parameter γ. However, Hansen (1996) shows in a linear threshold regression framework that a proper transformation of the Wald statistic, such as $\sup_{\gamma\in\Gamma} W_{nT}(\gamma)$, can have an asymptotic distribution free of γ and thus is able to be approximated via simulation. To simulate the asymptotic distribution of the sup-Wald statistic based on our QMLEs for the threshold SPD models, one has to approximate the θ-component of the CQS function, $S^*_{\theta,nT}(\theta_0,\gamma)$ given in (2.5), under the null in the bootstrap world. Note that simulations of the λ_1- and λ_2- components of $S^*_{\theta,nT}(\theta_0,\gamma)$ require a bootstrap sample for $\mathcal{Y}_t = W_t Y_t$, for which we need 'consistent' estimates of the individual fixed effects. But this is not possible when T is fixed due to the incidental parameters problem. In contrast, the standard linear panel data model in Hansen (1999) does not involve such an issue as only the β-component of $S^*_{\theta,nT}(\theta_0,\gamma)$ is needed in the simulation. To deal with this issue, we propose a sup-Wald statistic based on an M-estimator of the model, $\hat{\theta}^{\diamond}_{nT}(\gamma) = \arg\{S^{\diamond}_{\theta,nT}(\theta,\gamma) = 0\}$, where the estimating function $S^{\diamond}_{\theta,nT}(\theta,\gamma)$ is obtained by adjusting $S^*_{\theta,nT}(\theta_0,\gamma)$. To apply the simulation techniques in Hansen (1996, p.419), there are two necessary principals the adjustment should follow: (i) Under the alternatives, $\hat{\theta}^{\diamond}_{nT}(\gamma) - \theta_0 = o_p(1)$, uniformly in $\gamma \in \Gamma$; (ii) Under the null, $S^{\diamond}_{\theta,nT}(\theta_0,\gamma)$ can be approximated using its sample analogue.

For (i), we need that under the alternatives, $\lim_{nT\to\infty} \frac{1}{nT}\mathrm{E}[S^{\diamond}_{\theta,nT}(\theta_0,\gamma)] = 0$ uniformly in $\gamma \in \Gamma$ such that $\mathrm{plim}_{nT\to\infty}\frac{1}{nT}S^{\diamond}_{\theta,nT}(\theta_0,\gamma) = 0$, which is a necessary condition for $\hat{\theta}^{\diamond}_{nT}(\gamma)$ to be consistent. The principal (ii) requires that $S^{\diamond}_{\theta,nT}(\theta_0,\gamma)$ under the null only involves terms that we can calculate without knowledge of μ_0. We next show that the β-component of $S^*_{\theta,nT}(\theta_0,\gamma)$ in (2.5) automatically satisfies these conditions. Note that $V_t(\phi_0,\gamma) = d_t(\gamma_0,\gamma)H_t\phi_{20} + \lambda_{20}d_t(\gamma_0,\gamma)G_tV_t + \mu_0 + \alpha_{t0}l_n + V_t$ by (B.1) from Appendix B. Thus, it is easy to see $\lim_{nT\to\infty}\frac{1}{\sigma_0^2 nT}\mathrm{E}[\sum_{t=1}^T \widetilde{\mathbb{X}}'_t(\gamma)J_n\widetilde{V}_t(\phi_0,\gamma)] = 0$ under the alternatives, so the first principal is satisfied. Meanwhile, we also have $\frac{1}{\sigma_0^2}\sum_{t=1}^T \widetilde{\mathbb{X}}'_t(\gamma)J_n\widetilde{V}_t(\phi_0,\gamma) = \frac{1}{\sigma_0^2}\sum_{t=1}^T \widetilde{\mathbb{X}}'_t(\gamma)J_n\widetilde{V}_t$ under the null. As $\mathbb{X}_t(\gamma)$'s are observed and σ_0^2 can be approximated by its M-estimator, we only need to find bootstrap sample for $\widetilde{V}_t$. Note that $\widetilde{V}_t = \widetilde{\mathbb{Y}}_t - \widetilde{\mathbb{X}}_t\beta_0 - \alpha_{t0}l_n$. Thus, it is nature to use $\widetilde{V}^{\diamond}_t(\gamma) =$

$\widetilde{\mathbb{Y}}_t(\hat{\lambda}^{\diamond}_{nT}(\gamma),\gamma)-\widetilde{\mathbb{X}}_t(\gamma)\hat{\beta}^{\diamond}_{nT}(\gamma)-\hat{\alpha}_t(\hat{\phi}^{\diamond}_{nT}(\gamma),\gamma)l_n$ to approximate $\widetilde{V}_t$, since $\hat{\theta}^{\diamond}_{nT}(\gamma)$ is consistent uniformly in $\gamma\in\Gamma$.

Next, we consider the λ_1-component of (2.5). Recall $Z_t(\psi,\gamma)=G_t(\lambda,\gamma)(\mathbb{X}_t(\gamma)\beta+\mu+\alpha_t l_n)$ and note that $\mathcal{Y}_t=Z_t+G_tV_t$. As discussed above, the difficulty in approximating $\widetilde{\mathcal{Y}}_t$ lies with the fact that a 'consistent' estimate of μ_0 embedded in Z_t is unavailable due to the incidental parameters problem. For this reason, we do adjustment for the λ_1-component as follows. Note that the λ_1-component is essentially an LQ-form. Under the alternatives, we have

$$\begin{aligned}
&\lim_{nT\to\infty}\tfrac{1}{nT}\mathrm{E}[\tfrac{1}{\sigma_0^2}\textstyle\sum_{t=1}^T\widetilde{\mathcal{Y}}_t'J_n\widetilde{V}_t(\phi_0,\gamma)-\bar{T}\sum_{t=1}^T\mathtt{tr}G_t(\lambda_0,\gamma)]\\
=&\lim_{nT\to\infty}\tfrac{1}{nT}\mathrm{E}[\bar{T}\textstyle\sum_{t=1}^T\mathtt{tr}(J_nG_t(\lambda_0,\gamma))-\bar{T}\sum_{t=1}^T\mathtt{tr}G_t(\lambda_0,\gamma)]\\
=&\lim_{nT\to\infty}\tfrac{1}{nT}\mathrm{E}[\tfrac{1}{\sigma_0^2}\textstyle\sum_{t=1}^T(\widetilde{K}_t'(\phi_0,\gamma)+\widetilde{V}_t'(\phi_0,\gamma)J_nG_t(\lambda_0,\gamma))J_n\widetilde{V}_t(\phi_0,\gamma)\\
&-\bar{T}\textstyle\sum_{t=1}^T\mathtt{tr}G_t(\lambda_0,\gamma)],
\end{aligned}$$

for any γ, where $K_t(\phi,\gamma)=G_t(\lambda,\gamma)\mathbb{X}_t(\gamma)\beta$. Then taking the difference between the two quantities inside the second and third expectations and replacing θ_0 by the general parameter θ, we obtain the desired adjusted estimating function for λ_1:

$$\tfrac{1}{\sigma^2}\textstyle\sum_{t=1}^T[\widetilde{K}_t'(\phi,\gamma)+\widetilde{V}_t'(\phi,\gamma)J_nG_t(\lambda,\gamma)]J_n\widetilde{V}_t(\phi,\gamma)-\bar{T}\sum_{t=1}^T\mathtt{tr}(J_nG_t(\lambda,\gamma)),$$

which is still an LQ-form but only involves terms that are approximable. Similarly, the desired adjusted estimating function for λ_2 is:

$$\tfrac{1}{\sigma^2}\textstyle\sum_{t=1}^T[\widetilde{K}_t^{\circ\prime}(\phi,\gamma)+\widetilde{V}_t'(\phi,\gamma)J_nd_t(\gamma)G_t(\lambda,\gamma)]J_n\widetilde{V}_t(\phi,\gamma)-\bar{T}\sum_{t=1}^T\mathtt{tr}(J_nd_t(\gamma)G_t(\lambda,\gamma))],$$

where $K_t^{\circ}(\phi,\gamma)=d_t(\gamma)K_t(\phi,\gamma)$. As for the σ^2-component of (2.5), we have

$$\lim_{nT\to\infty}\tfrac{1}{nT}\mathrm{E}[\tfrac{1}{2\sigma_0^4}\textstyle\sum_{t=1}^T\widetilde{V}_t'(\phi_0,\gamma)J_n\widetilde{V}_t(\phi_0,\gamma)-\frac{n(T-1)}{2\sigma_0^2}]=\lim_{nT\to\infty}\tfrac{1}{nT}[\frac{(n-1)(T-1)}{2\sigma_0^2}-\frac{n(T-1)}{2\sigma_0^2}].$$

Similarly, by taking the difference between the two quantities inside the square brackets and using the general parameter θ instead, we finally obtain the adjusted function for σ^2:

$$\tfrac{1}{2\sigma^4}\textstyle\sum_{t=1}^T\widetilde{V}_t'(\phi,\gamma)J_n\widetilde{V}_t(\phi,\gamma)-\frac{(n-1)(T-1)}{2\sigma^2}.$$

A set of unbiased estimating functions or *adjusted quasi score* (AQS) functions are as follows,

$$S^{\diamond}_{\theta,nT}(\theta,\gamma)=\begin{cases}\frac{1}{\sigma^2}\sum_{t=1}^{T}\widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{V}_t(\phi,\gamma),\\ \frac{1}{\sigma^2}\sum_{t=1}^{T}[\widetilde{K}_t'(\phi,\gamma)+\widetilde{V}_t'(\phi,\gamma)J_nG_t(\lambda,\gamma)]J_n\widetilde{V}_t(\phi,\gamma)\\ \qquad\qquad -\bar{T}\sum_{t=1}^{T}\mathtt{tr}[J_nG_t(\lambda,\gamma)],\\ \frac{1}{\sigma^2}\sum_{t=1}^{T}[\widetilde{K}_t^{\circ\prime}(\phi,\gamma)+\widetilde{V}_t'(\phi,\gamma)J_nd_t(\gamma)G_t(\lambda,\gamma)]J_n\widetilde{V}_t(\phi,\gamma)\\ \qquad\qquad -\bar{T}\sum_{t=1}^{T}\mathtt{tr}[J_nd_t(\gamma)G_t(\lambda,\gamma)],\\ \frac{1}{2\sigma^4}\sum_{t=1}^{T}\widetilde{V}_t'(\phi,\gamma)J_n\widetilde{V}_t(\phi,\gamma)-\frac{(n-1)(T-1)}{2\sigma^2}.\end{cases}\tag{2.15}$$

For each $\gamma\in\Gamma$, the M- or AQS estimator of θ is defined as $\hat{\theta}^{\diamond}_{nT}(\gamma)=\arg\{S^{\diamond}_{\theta,nT}(\theta,\gamma)=0\}$. From the first and last components of $S^{\diamond}_{\theta,nT}(\theta,\gamma)$, we have $\hat{\beta}^{\diamond}_{nT}(\lambda,\gamma)=\hat{\beta}_{nT}(\lambda,\gamma)$ and $\hat{\sigma}^{2\diamond}_{nT}(\lambda,\gamma)=\frac{n}{n-1}\hat{\sigma}^2_{nT}(\lambda,\gamma)$. Thus, under the alternatives, the consistency of $\hat{\theta}^{\diamond}_{nT}(\gamma)$ lies with that of $\hat{\lambda}^{\diamond}_{nT}(\gamma)$ uniformly in γ (see (B.4) and (B.5) in Appendix B). We substitute $\hat{\beta}^{\diamond}_{nT}(\lambda,\gamma)$ and $\hat{\sigma}^{2\diamond}_{nT}(\lambda,\gamma)$ into the λ- components of AQS functions, yielding the concentrated AQS functions, $S^{\diamond c}_{\theta,nT}(\lambda,\gamma)$. To show the consistency of $\hat{\lambda}^{\diamond}_{nT}(\gamma)$, we need to find the population counterpart of $S^{\diamond c}_{\theta,nT}(\lambda,\gamma)$. Let $\bar{S}^{\diamond}_{\theta,nT}(\theta,\gamma)=\mathrm{E}[S^{\diamond}_{\theta,nT}(\theta,\gamma)]$. Given λ and γ, $\bar{S}^{\diamond}_{\theta,nT}(\theta,\gamma)=0$ is partially solved at $\bar{\beta}^{\diamond}_{nT}(\lambda,\gamma)=\bar{\beta}_{nT}(\lambda,\gamma)$ and $\bar{\sigma}^{2\diamond}_{nT}(\lambda,\gamma)=\frac{n}{n-1}\bar{\sigma}^2_{nT}(\lambda,\gamma)$. Substituting them back into the λ- components of $\bar{S}^{\diamond}_{\theta,nT}(\theta,\gamma)$ gives $\bar{S}^{\diamond c}_{\theta,nT}(\lambda,\gamma)$, which is just the population counterpart of $S^{\diamond c}_{\theta,nT}(\lambda,\gamma)$ (The expressions for $S^{\diamond c}_{\theta,nT}(\lambda,\gamma)$ and $\bar{S}^{\diamond c}_{\theta,nT}(\lambda,\gamma)$ can be found in Appendix B). By Theorem 5.9 of van der Vaar (1998), $\hat{\lambda}^{\diamond}_{nT}(\gamma)$ converges to $\lambda_0,\forall\gamma$, if $\sup_{\lambda\in\Lambda}\frac{1}{nT}|S^{\diamond c}_{\theta,nT}(\lambda,\gamma)-\bar{S}^{\diamond c}_{\theta,nT}(\lambda,\gamma)|=o_p(1),\forall\gamma$, and the following identification condition holds:

Assumption G′: $\inf_{\lambda:d(\lambda,\lambda_0)\geq\epsilon}\|\bar{S}^{\diamond c}_{\theta}(\lambda,\gamma)\|>0$ for any γ and $\epsilon>0$, where $\bar{S}^{\diamond c}_{\theta}(\lambda,\gamma)=\lim_{nT\to\infty}\frac{1}{nT}\bar{S}^{\diamond c}_{\theta,nT}(\lambda,\gamma)$ and $d(\lambda,\lambda_0)$ is a measure of distance between λ and λ_0.

Note that Assumption G′ can be shown to be true under some primitive conditions (see Appendix B). Given $\hat{\lambda}^{\diamond}_{nT}(\gamma)$, the M-estimators of β and σ^2 are just $\hat{\beta}^{\diamond}_{nT}(\gamma)\equiv\hat{\beta}^{\diamond}_{nT}(\hat{\lambda}^{\diamond}_{nT}(\gamma),\gamma)$ and $\hat{\sigma}^{2\diamond}_{nT}(\gamma)\equiv\hat{\sigma}^{2\diamond}_{nT}(\hat{\lambda}^{\diamond}_{nT}(\gamma),\gamma)$. Under the alternatives, once the consistency of $\hat{\lambda}^{\diamond}_{nT}(\gamma)$ is established as above, the consistency of $\hat{\beta}^{\diamond}_{nT}(\gamma)$ and $\hat{\sigma}^{2\diamond}_{nT}(\gamma)$ follows by Assumptions B(vi). The following theorem shows that our AQS function $\bar{S}^{\diamond}_{\theta,nT}(\theta,\gamma)$ indeed satisfy the two necessary principals required in Hansen (1996):

Theorem 2.7. *Suppose Assumptions A-E and G' hold. We have,*

(i) under the alternatives, $\hat{\theta}^{\diamond}_{nT}(\gamma) - \theta_0 = o_p(1)$*, uniformly in* $\gamma \in \Gamma$*, and*

(ii) under the null, $S^{\diamond}_{\theta,nT}(\theta_0, \gamma)$ *can be approximated using its sample analogue.*

The proof of result (i) of Theorem 2.7 is in Appendix B. Given (i) and under the null, it is easy to see that $K_t(\phi_0, \gamma)$, $G_t(\lambda_0, \gamma)$ and $\widetilde{V}_t(\phi_0, \gamma)$ can be approximated by $K_t(\hat{\phi}^{\diamond}_{nT}(\gamma), \gamma)$, $G_t(\hat{\lambda}^{\diamond}_{nT}(\gamma), \gamma)$ and $\widetilde{V}^{\diamond}_t(\gamma)$, respectively. Hence, result (ii) also holds. We then define $\Sigma^{\diamond}_{nT}(\gamma, \gamma) = -\frac{1}{nT}\mathrm{E}[\frac{\partial}{\partial \theta'} S^{\diamond}_{\theta,nT}(\theta_0, \gamma)]$ and $\Omega^{\diamond}_{nT}(\gamma, \gamma) = \frac{1}{nT}\mathrm{Var}[S^{\diamond}_{\theta,nT}(\theta_0, \gamma)]$, and the analytical expressions for $\Sigma^{\diamond}_{nT}(\gamma_1, \gamma_2)$ and $\Omega^{\diamond}_{nT}(\gamma_1, \gamma_2)$ are in (B.25) and (B.26), respectively. Letting $\widehat{\mathbb{Q}}_{nT}(\gamma) \equiv \widehat{\Sigma}^{\diamond -1}_{nT}(\gamma, \gamma)\widehat{\Omega}^{\diamond}_{nT}(\gamma, \gamma)\widehat{\Sigma}^{\diamond\prime -1}_{nT}(\gamma, \gamma)$ where $\widehat{\Sigma}^{\diamond}_{nT}(\gamma, \gamma)$ and $\widehat{\Omega}^{\diamond}_{nT}(\gamma, \gamma)$ are their respective plug-in estimators, we have the sup-Wald statistic:

$$\mathrm{sup}W_{nT} \equiv \sup_{\gamma \in \Gamma} W_{nT}(\gamma), \tag{2.16}$$

where $W_{nT}(\gamma) = nT\hat{\theta}^{\diamond\prime}_{nT}(\gamma)\mathbf{L}[\mathbf{L}'\widehat{\mathbb{Q}}_{nT}(\gamma)\mathbf{L}]^{-1}\mathbf{L}'\hat{\theta}^{\diamond}_{nT}(\gamma)$, and $\mathbf{L}$ is a selection matrix defined as

$$\mathbf{L} = \begin{bmatrix} 0_{k\times k} & I_k & 0_{k\times 1} & 0_{k\times 1} & 0_{k\times 1} \\ 0_{1\times k} & 0_{1\times k} & 0 & 1 & 0 \end{bmatrix}'.$$

Theorem 2.8. *Under Assumptions A-E, G', and the alternatives* $H_1 : \phi_{20} = \frac{C}{\sqrt{nT}}$,

$$\mathrm{sup}W_{nT} \xrightarrow{D} \sup_{\gamma \in \Gamma} W^c(\gamma),$$

where $W^c(\gamma) = [\mathbf{L}'\bar{S}(\gamma) + \bar{\Sigma}(\gamma)C]'\bar{Q}(\gamma, \gamma)^{-1}[\mathbf{L}'\bar{S}(\gamma) + \bar{\Sigma}(\gamma)C]$, $\bar{S}(\gamma)$ *is a mean-zero Gaussian process with covariance kernel* $\bar{Q}(\gamma_1, \gamma_2) = \mathbf{L}'\Sigma^{\diamond -1}(\gamma_1, \gamma_2)\Omega^{\diamond}(\gamma_1, \gamma_2)$ $\Sigma^{\diamond -1}(\gamma_1, \gamma_2)\mathbf{L}$*,* $\bar{\Sigma}(\gamma) = \mathbf{L}'\Sigma^{\diamond -1}(\gamma, \gamma)\Sigma^{\diamond}(\gamma, \gamma_0)\mathbf{L}$*,* $\Sigma^{\diamond}(\gamma_1, \gamma_2) = \lim_{nT\to\infty} \Sigma^{\diamond}_{nT}(\gamma_1, \gamma_2)$*, and* $\Omega^{\diamond}(\gamma_1, \gamma_2) = \lim_{nT\to\infty} \Omega^{\diamond}_{nT}(\gamma_1, \gamma_2)$*.*

Clearly, $\sup_{\gamma\in\Gamma} W^0(\gamma) = \sup_{\gamma\in\Gamma} \bar{S}(\gamma)'\mathbf{L}\bar{Q}(\gamma, \gamma)^{-1}\mathbf{L}'\bar{S}(\gamma)$ under the null, $C = 0$. It is a functional of chi-square processes and the asymptotic critical values for which cannot be tabulated in general. In special cases, Andrews (1993) and Li (2018) show the critical values of testing for the existence of structure change depend only on the column dimension of regressors and the parameter space of γ_0 so that they can be tabulated. But it is not the case for general threshold models. We thus follow Hansen (1996) and propose the following bootstrap procedure to approximate the asymptotic null distribution of the test statistic.

1. Calculate $\hat{\alpha}^{\diamond}_{t,nT}(\gamma) = \hat{\alpha}_t(\hat{\phi}^{\diamond}_{nT}(\gamma), \gamma)$ and $\widetilde{V}^{\diamond}_t(\gamma) = \widetilde{\mathbb{Y}}_t(\hat{\lambda}^{\diamond}_{nT}(\gamma), \gamma) - \widetilde{\mathbb{X}}_t(\gamma)\hat{\beta}^{\diamond}_{nT}(\gamma) - \hat{\alpha}^{\diamond}_{t,nT}(\gamma) l_n$. Let $\{\widetilde{v}^{\diamond}_{it}(\gamma)\}$ be the elements of $\widetilde{V}^{\diamond}_t(\gamma)$ and group them by unit: $\widetilde{V}^{\diamond}_i(\gamma) = (\widetilde{v}^{\diamond}_{i1}(\gamma), \ldots, \widetilde{v}^{\diamond}_{iT}(\gamma))$.
2. Let $\mathcal{F}_{nT}$ be the empirical distribution function (EFD) defined by $\{\widetilde{V}^{\diamond}_1(\gamma), \ldots, \widetilde{V}^{\diamond}_n(\gamma)\}$. With replacement, draw a random sample of size n from $\mathcal{F}_{nT}$. Then group them back by time to give a bootstrap sample $\{\widetilde{V}^{b}_t(\gamma)\}$.
3. Calculate $\widehat{S}^{b}_{\theta,nT}(\gamma)$, a value of $S^{\diamond}_{\theta,nT}(\hat{\theta}^{\diamond}_{nT}(\gamma), \gamma)$ with $\widetilde{V}_t(\hat{\phi}^{\diamond}_{nT}(\gamma), \gamma)$ replaced by $\widetilde{V}^{b}_t(\gamma)$.
4. Compute $\sup W^{b}_{nT} \equiv \sup_{\gamma\in\Gamma} \frac{1}{nT} \widehat{S}^{b\prime}_{\theta,nT}(\gamma) \widehat{\Sigma}^{\diamond\prime -1}_{nT}(\gamma,\gamma)\mathbf{L}[\mathbf{L}'\widehat{\mathbb{Q}}_{nT}(\gamma)\mathbf{L}]^{-1}\mathbf{L}'$ $\widehat{\Sigma}^{\diamond -1}_{nT}(\gamma,\gamma)\widehat{S}^{b}_{\theta,nT}(\gamma)$.
5. Repeat steps 2-4 B times.
6. Calculate the bootstrap p-value of the test: $p^{b}_{W} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\{\sup W^{b}_{nT} \geq \sup W_{nT}\}$, and reject the null when p^{b}_{W} is less than the pre-chosen level of significance.

Under the assumptions of Theorem 2.8, $\hat{\theta}^{\diamond}_{nT}(\gamma)$ is consistent to θ_0 on Γ. Therefore, it is easy to see the M-estimator of time fixed effect $\hat{\alpha}^{\diamond}_{t,nT}(\gamma)$ is also consistent for each t, no matter whether T is fixed or large. Hence, $\{\widetilde{v}^{\diamond}_{it}(\gamma)\}$ are consistent estimates of $\{\widetilde{v}_{it}\}$, which are iid across i. When both n and T are large, individual and time fixed effects can be estimated consistently, and thus we can find 'consistent' estimation for idiosyncratic errors $\{v_{it}\}$. Therefore, the block bootstrap procedure above can be simply replaced by boostrapping on $\{\hat{v}_{it}\}$, which are the elements of $\widehat{V}^{\diamond}_t(\gamma) = \mathbb{Y}_t(\hat{\lambda}^{\diamond}_{nT}(\gamma), \gamma) - \mathbb{X}_t(\gamma)\hat{\beta}^{\diamond}_{nT}(\gamma) - \hat{\mu}^{\diamond}_{nT}(\gamma) - \hat{\alpha}^{\diamond}_{t,nT}(\gamma) l_n$ with $\hat{\mu}^{\diamond}_{nT}(\gamma) = \hat{\mu}(\hat{\phi}^{\diamond}_{nT}(\gamma), \gamma)$. Following theorem justifies the asymptotic validity of the above procedure.

Theorem 2.9. *Suppose Assumptions A-E, G′ and the null hypothesis hold,* *we have*

$$\text{sup}W_{nT}^{b} \xrightarrow{D} \sup_{\gamma\in\Gamma} W^{0}(\gamma).$$

The above theorem implies that we can approximate the asymptotic null distribution of the statistic $\text{sup}W_{nT}$ by the EDF of $\{\text{sup}W_{nT}^{b}, b = 1, \ldots, B\}$ for a sufficiently large B. Therefore, we can reject the null at the significance level of α when $p_{W}^{b} < \alpha$.

2.4 Monte Carlo Study

Monte Carlo experiments are carried out to evaluate the finite sample performance of the proposed estimators and test statistics. The following data generating process is used:

$$Y_t = \lambda_1 W_t Y_t + \lambda_2 d_t(\gamma) W_t Y_t + X_t \beta_1 + d_t(\gamma) X_t \beta_2 + \mu + \alpha_t l_n + V_t, \quad t = 1, \ldots, T$$

where the time-varying weight matrices W_t's are generated according to `Queen` contiguity, x_{it} are generated from $N(1, 1)$, the fixed effects μ are generated according to $\frac{1}{T}\Sigma_{t=1}^{T} X_t + e$, where $e \sim N(0, I_N)$, and the time fixed effects α are generated from $N(0, I_T)$. The distributions of the error term can be (*i*) normal, (*ii*) normal mixture (10% $N(0, 4^2)$ and 90% $N(0, 1)$), or (*iii*) chi-square with 3 degrees of freedom. In both (*ii*) and (*iii*), the error distributions are standardized to have mean zero and variance $\sigma^2 = 1$. We set $\beta_1 = 1$, $\lambda_1 = 0.2$, $\beta_2 = \lambda_2 = (nT)^{-0.2}$ and $\gamma = 1$. The number of Monte Carlo runs under each parameter configuration is 1000.

Tables 2.1a and 2.1b report the Monte Carlo results for the 2SLS estimator of Wei et al. (2021), the direct QMLE based on (2.3), and the bias-corrected QMLE (bc-QMLE) based on (2.4), under various combinations of $n = 50, 100, 200$ and $T = 5, 10, 20, 40$. Monte Carlo or empirical estimation biases (bias), and standard deviations (sd) are reported. Further, empirical averages of the robust standard error estimates (rse) are also reported for bc-QMLE. Note that the direct QMLE and bc-QMLE share the same estimated value of γ as discussed in Footnote 4. From the results, we see that the finite sample performance of the 2SLS estimator can be very poor with large biases and large standard deviations. The use of the direct QML method can help improve the estimation of all parameters except for the error variance. This is in line with our theory – the direct QML estimation of σ^2 is inconsistent for when T is fixed due to the incidental parameters problem. In contrast, our bc-QMLE has an excellent finite performance in terms of both consistency and efficiency. All the estimations improve on average when the sample expands, regardless of the error distributions. The $\sqrt{n(T-1)}$ convergence rate of the common QMLEs is clearly demonstrated by the Monte Carlo standard errors. Moreover, the robust estimates of standard errors rses are on average very close to the corresponding Monte Carlo sds and become closer as the sample size increases.

Following Theorem 2.5, we propose to make inference on γ by applying the LR test in Subsection 2.2.4. Table 2.2 reports the empirical size of the LR test. We find the rejection rates for all types of error distributions are close to the nominal levels and improve on average as the sample size becomes large. Table 2.3 reports the results for testing the threshold effects. Under the null, $\beta_{20} = \lambda_{20} = 0$, the rejection rates for all types of error distribution are close to the nominal levels. We also consider the local power of our test in the last two columns of the table. It can be seen that the rejection rates rise quickly as β_{20} and λ_{20} deviate from 0. When $\beta_{20} = \lambda_{20} = 10/\sqrt{nT}$, the power of our test reaches 100% for all the sample sizes.

2.5 An Empirical Application

In this section, we apply our method to study the age-of-leader effects on political competitions across Chinese cities. The tournament competition among Chinese city government leaders has been an important topic in China's economic growth literature (Yao and Zhang, 2015). Local government leaders compete against one another in enhancing local investment and promoting the local economy's growth so as to increase their chances for political promotion. Based on this theory, Yu et al. (2016) document a strong spatial effect for the city-level total investment. Besides, the age of a local leader is another pivotal factor determining the leader's chances of promotion. A leader's chance diminishes quickly as he or she gets older (Yao and Zhang, 2015; Yu et al.,

Table 2.1a. Empirical bias(sd)[$\hat{sd}$] of the estimators for FE-SPD model with threshold effects; W_t=Queen Contiguity.

	2SLS	QMLE	QMLE-bc	2SLS	QMLE	QMLE-bc
		$T=5$			$T=10$	
			n = 100, error = 1, 2, 3, for the three panels below			
β_1	-.0527(0.748)	-.0040(0.090)	-.0063(0.090)[0.090]	.0070(0.901)	.0026(0.060)	.0010(0.060)[0.060]
β_2	.0709(1.217)	.0063(0.137)	.0041(0.136)[0.141]	-.0153(1.087)	-.0098(0.103)	-.0114(0.103)[0.099]
λ_1	.0704(0.728)	-.0238(0.067)	-.0016(0.067)[0.065]	.0138(0.329)	-.0242(0.047)	-.0011(0.047)[0.046]
λ_2	-.0541(1.129)	-.0031(0.044)	-.0034(0.044)[0.044]	.0037(0.625)	-.0006(0.035)	-.0011(0.035)[0.034]
σ^2	.0971(0.079)	-.2135(0.055)	-.0096(0.069)[0.070]	.1087(0.055)	-.1095(0.043)	-.0031(0.048)[0.047]
γ	-.0843(0.416)	-.0114(0.023)	–	-.0143(0.027)	.0006(0.019)	–
β_1	-.0955(0.745)	-.0040(0.092)	-.0064(0.092)[0.090]	.0496(0.868)	-.0007(0.059)	-.0022(0.059)[0.060]
β_2	.1815(1.259)	-.0043(0.148)	-.0065(0.148)[0.141]	-.0635(1.040)	-.0042(0.104)	-.0058(0.104)[0.100]
λ_1	.1151(0.730)	-.0208(0.069)	.0013(0.069)[0.066]	-.0041(0.316)	-.0260(0.046)	-.0028(0.046)[0.046]
λ_2	-.0870(1.134)	-.0031(0.046)	-.0034(0.046)[0.045]	.0350(0.600)	-.0014(0.034)	-.0018(0.034)[0.034]
σ^2	.1003(0.169)	-.2123(0.120)	-.0082(0.151)[0.148]	.1130(0.115)	-.1044(0.096)	.0026(0.108)[0.107]
γ	-.0999(0.433)	-.0114(0.023)	–	-.0159(0.027)	.0002(0.020)	–
β_1	-.0713(0.714)	.0006(0.090)	-.0018(0.090)[0.089]	.0298(0.856)	.0000(0.059)	-.0016(0.059)[0.060]
β_2	.1152(1.221)	.0033(0.148)	.0012(0.148)[0.141]	-.0404(1.029)	-.0056(0.105)	-.0071(0.105)[0.099]
λ_1	.0921(0.668)	-.0248(0.067)	-.0027(0.067)[0.065]	.0066(0.319)	-.0247(0.048)	-.0016(0.048)[0.046]
λ_2	-.0710(1.078)	-.0012(0.045)	-.0016(0.045)[0.045]	.0201(0.593)	-.0021(0.034)	-.0026(0.034)[0.035]
σ^2	.1070(0.126)	-.2086(0.090)	-.0034(0.113)[0.109]	.1044(0.084)	-.1117(0.068)	-.0057(0.077)[0.076]
γ	-.0832(0.406)	-.0103(0.021)	–	-.0130(0.024)	.0019(0.019)	–
			n = 200, error = 1, 2, 3, for the three panels below			
β_1	.0489(0.621)	.0023(0.057)	.0016(0.056)[0.055]	.0248(0.332)	-.0020(0.038)	-.0023(0.038)[0.037]
β_2	-.0537(0.703)	-.0045(0.102)	-.0048(0.102)[0.098]	-.0269(0.424)	.0000(0.066)	-.0003(0.066)[0.065]
λ_1	-.0061(0.183)	-.0162(0.051)	-.0042(0.051)[0.049]	.0041(0.169)	-.0135(0.033)	-.0012(0.033)[0.032]
λ_2	.0337(0.425)	.0027(0.034)	.0025(0.034)[0.033]	.0246(0.271)	-.0019(0.020)	-.0018(0.020)[0.019]
σ^2	.1161(0.060)	-.2109(0.039)	-.0100(0.049)[0.049]	.1133(0.038)	-.1056(0.030)	-.0025(0.033)[0.033]
γ	-.0205(0.069)	.0030(0.019)	–	-.0091(0.019)	-.0017(0.007)	–
β_1	.0203(0.635)	.0011(0.056)	.0004(0.056)[0.055]	.0162(0.325)	-.0014(0.038)	-.0017(0.038)[0.037]
β_2	-.0225(0.717)	-.0031(0.100)	-.0033(0.100)[0.098]	-.0198(0.413)	-.0015(0.066)	-.0018(0.066)[0.065]
λ_1	-.0017(0.186)	-.0111(0.050)	.0009(0.050)[0.049]	.0133(0.171)	-.0140(0.033)	-.0017(0.033)[0.032]
λ_2	.0152(0.434)	.0003(0.034)	.0000(0.034)[0.033]	.0153(0.265)	-.0003(0.019)	-.0003(0.019)[0.019]
σ^2	.1193(0.124)	-.2088(0.091)	-.0074(0.114)[0.107]	.1148(0.080)	-.1051(0.067)	-.0020(0.075)[0.076]
γ	-.0148(0.052)	.0028(0.019)	–	-.0082(0.018)	-.0014(0.009)	–
β_1	.0242(0.643)	.0018(0.053)	.0011(0.053)[0.054]	.0359(0.308)	.0019(0.038)	.0017(0.038)[0.037]
β_2	-.0255(0.728)	-.0033(0.097)	-.0036(0.097)[0.097]	-.0409(0.390)	-.0038(0.066)	-.0041(0.066)[0.064]
λ_1	.0017(0.192)	-.0148(0.050)	-.0027(0.050)[0.049]	.0044(0.161)	-.0131(0.033)	-.0008(0.033)[0.032]
λ_2	.0153(0.440)	.0014(0.034)	.0012(0.034)[0.034]	.0328(0.251)	-.0006(0.019)	-.0006(0.019)[0.020]
σ^2	.1181(0.091)	-.2087(0.065)	-.0073(0.082)[0.078]	.1180(0.063)	-.1025(0.051)	.0009(0.057)[0.055]
γ	-.0180(0.060)	.0051(0.019)	–	-.0081(0.019)	-.0008(0.007)	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square).

Table 2.1b. Empirical bias(sd)[$\hat{sd}$] of the estimators for FE-SPD model with threshold effects; W_t=Queen Contiguity.

	2SLS	QMLE	QMLE-bc	2SLS	QMLE	QMLE-bc
		$T=5$			$T=10$	
		n = 50, error = 1, 2, 3, for the three panels below				
β_1	-.1976(1.170)	-.0029(0.113)	-.0085(0.112)[0.106]	-.1379(0.929)	-.0013(0.082)	-.0030(0.082)[0.080]
β_2	.4325(1.968)	.0097(0.185)	.0158(0.184)[0.167]	.1459(0.996)	.0010(0.139)	.0016(0.139)[0.133]
λ_1	.1591(0.979)	-.0537(0.098)	-.0088(0.098)[0.092]	.0662(0.349)	-.0555(0.066)	-.0093(0.066)[0.063]
λ_2	-.0581(1.096)	.0000(0.059)	-.0024(0.059)[0.057]	-.0855(0.629)	.0002(0.040)	-.0003(0.040)[0.039]
σ^2	.1911(0.124)	-.2296(0.077)	-.0232(0.098)[0.099]	.1088(0.079)	-.1255(0.060)	-.0140(0.067)[0.066]
γ	-.1390(0.567)	-.0142(0.051)	–	-.0160(0.100)	.0045(0.022)	–
β_1	-.2602(1.158)	-.0015(0.115)	-.0070(0.114)[0.105]	-.0910(0.920)	.0019(0.085)	.0003(0.085)[0.080]
β_2	.4917(2.003)	.0019(0.205)	.0080(0.204)[0.166]	.0926(0.959)	-.0004(0.139)	-.0001(0.138)[0.133]
λ_1	.2329(0.994)	-.0546(0.098)	-.0100(0.098)[0.092]	.0544(0.354)	-.0572(0.068)	-.0109(0.068)[0.063]
λ_2	-.1122(1.131)	-.0012(0.060)	-.0035(0.059)[0.057]	-.0554(0.612)	.0014(0.042)	.0009(0.042)[0.039]
σ^2	.1766(0.238)	-.2369(0.172)	-.0324(0.218)[0.194]	.1055(0.155)	-.1243(0.131)	-.0126(0.148)[0.146]
γ	-.1189(0.594)	-.0178(0.089)	–	-.0079(0.087)	.0039(0.023)	–
β_1	-.1699(1.379)	-.0027(0.110)	-.0083(0.110)[0.105]	-.0834(0.941)	-.0022(0.081)	-.0039(0.081)[0.079]
β_2	.5650(2.354)	.0110(0.195)	.0171(0.194)[0.166]	.0937(0.994)	.0052(0.131)	.0057(0.131)[0.132]
λ_1	.1368(1.152)	-.0557(0.100)	-.0109(0.100)[0.092]	.0490(0.369)	-.0545(0.066)	-.0082(0.065)[0.063]
λ_2	.0087(1.285)	.0006(0.059)	-.0016(0.059)[0.057]	-.0497(0.633)	-.0008(0.039)	-.0013(0.039)[0.039]
σ^2	.1910(0.184)	-.2257(0.126)	-.0181(0.160)[0.146]	.1025(0.126)	-.1260(0.097)	-.0146(0.109)[0.106]
γ	-.1571(0.583)	-.0117(0.063)	–	-.0137(0.110)	.0059(0.021)	–
		$T=20$			$T=40$	
		n = 50, error = 1, 2, 3, for the three panels below				
β_1	-.0496(0.461)	.0022(0.054)	-.0004(0.054)[0.052]	.0093(0.391)	-.0008(0.039)	-.0017(0.039)[0.039]
β_2	.0566(0.565)	-.0038(0.093)	-.0026(0.093)[0.089]	-.0123(0.442)	-.0002(0.064)	.0006(0.064)[0.064]
λ_1	.0337(0.244)	-.0522(0.048)	-.0063(0.048)[0.045]	.0072(0.122)	-.0518(0.033)	-.0044(0.033)[0.031]
λ_2	-.0373(0.367)	.0016(0.032)	.0003(0.032)[0.031]	.0076(0.259)	-.0007(0.021)	-.0008(0.021)[0.020]
σ^2	.1068(0.052)	-.0714(0.042)	-.0077(0.044)[0.046]	.0937(0.038)	-.0445(0.032)	-.0047(0.033)[0.032]
γ	-.0091(0.028)	.0002(0.020)	–	-.0054(0.010)	-.0026(0.006)	–
β_1	-.0484(0.447)	.0056(0.052)	.0031(0.052)[0.051]	.0061(0.388)	.0015(0.037)	.0006(0.037)[0.039]
β_2	.0536(0.559)	-.0058(0.092)	-.0046(0.092)[0.089]	-.0100(0.440)	-.0042(0.064)	-.0034(0.064)[0.064]
λ_1	.0367(0.244)	-.0528(0.046)	-.0071(0.046)[0.045]	.0090(0.126)	-.0507(0.033)	-.0033(0.033)[0.032]
λ_2	-.0406(0.365)	.0021(0.032)	.0008(0.032)[0.031]	.0048(0.256)	.0004(0.020)	.0004(0.020)[0.021]
σ^2	.1031(0.118)	-.0747(0.103)	-.0111(0.110)[0.104]	.0987(0.079)	-.0409(0.072)	-.0010(0.075)[0.075]
γ	-.0123(0.059)	.0001(0.019)	–	-.0053(0.011)	-.0026(0.007)	–
β_1	-.0823(0.469)	.0054(0.053)	.0028(0.053)[0.051]	.0086(0.401)	.0016(0.039)	.0007(0.039)[0.038]
β_2	.0993(0.567)	-.0047(0.093)	-.0033(0.092)[0.089]	-.0097(0.459)	-.0020(0.064)	-.0012(0.064)[0.064]
λ_1	.0541(0.255)	-.0516(0.048)	-.0058(0.048)[0.045]	.0069(0.122)	-.0510(0.032)	-.0036(0.032)[0.031]
λ_2	-.0655(0.375)	.0028(0.033)	.0015(0.033)[0.032]	.0059(0.266)	.0001(0.021)	.0000(0.021)[0.021]
σ^2	.1079(0.086)	-.0715(0.073)	-.0078(0.078)[0.076]	.0985(0.059)	-.0409(0.053)	-.0010(0.055)[0.055]
γ	-.0121(0.040)	.0008(0.019)	–	-.0050(0.011)	-.0016(0.006)	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square).

Table 2.2. Empirical sizes of LR test at 0.01, 0.05 and 0.10 levels; W_t = Queen Contiguity.

n	T	Normal errors			Normal mixture			Chi-square		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
50	5	.009	.037	.063	.019	.062	.096	.013	.044	.079
	10	.006	.039	.074	.015	.046	.076	.013	.042	.062
	20	.008	.058	.082	.016	.054	.080	.014	.044	.086
	40	.012	.060	.098	.010	.054	.094	.008	.056	.100
100	5	.008	.037	.062	.012	.049	.078	.011	.039	.063
	10	.016	.059	.095	.017	.064	.106	.008	.042	.082
200	5	.010	.049	.094	.028	.060	.096	.004	.042	.098
	10	.009	.046	.095	.015	.053	.085	.010	.060	.095

Table 2.3. Rejecting frequency of tests for threshold effects at 0.01, 0.05 and 0.10 levels; W_t = Queen Contiguity.

error	n	T	$\lambda_2 = \beta_2 = 0$			$\lambda_2 = \beta_2 = 2/\sqrt{nT}$			$\lambda_2 = \beta_2 = 10/\sqrt{nT}$		
			0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
1	50	5	.014	.034	.064	.110	.256	.366	1.000	1.000	1.000
		10	.016	.044	.068	.092	.248	.344	1.000	1.000	1.000
		20	.016	.046	.082	.140	.272	.400	1.000	1.000	1.000
		40	.010	.058	.102	.098	.264	.358	1.000	1.000	1.000
	100	5	.022	.068	.098	.104	.242	.334	1.000	1.000	1.000
		10	.007	.048	.086	.102	.260	.336	1.000	1.000	1.000
	200	5	.018	.060	.110	.140	.270	.378	1.000	1.000	1.000
		10	.010	.064	.120	.074	.194	.306	1.000	1.000	1.000
2	50	5	.008	.030	.050	.106	.252	.354	1.000	1.000	1.000
		10	.008	.032	.064	.086	.220	.294	1.000	1.000	1.000
		20	.006	.040	.108	.150	.286	.402	1.000	1.000	1.000
		40	.016	.054	.100	.104	.274	.380	1.000	1.000	1.000
	100	5	.007	.034	.065	.118	.262	.336	1.000	1.000	1.000
		10	.011	.036	.074	.100	.234	.348	1.000	1.000	1.000
	200	5	.006	.052	.114	.108	.232	.358	1.000	1.000	1.000
		10	.010	.056	.090	.072	.200	.302	1.000	1.000	1.000
3	50	5	.014	.028	.060	.120	.262	.350	1.000	1.000	1.000
		10	.014	.044	.078	.106	.268	.372	1.000	1.000	1.000
		20	.014	.062	.116	.150	.300	.418	1.000	1.000	1.000
		40	.012	.052	.096	.110	.242	.342	1.000	1.000	1.000
	100	5	.020	.060	.081	.100	.234	.330	1.000	1.000	1.000
		10	.008	.042	.084	.096	.232	.326	1.000	1.000	1.000
	200	5	.014	.068	.100	.092	.256	.358	1.000	1.000	1.000
		10	.008	.046	.078	.100	.212	.328	1.000	1.000	1.000

Note: error = 1(normal), 2(normal mixture), 3(chi-square).

2016). Thus, those leaders who are close to retirement age (60) should have less incentive to join this tournament competition than the young leaders, and thus a weaker spatial effect should be expected among the old leaders. That is, we would expect that the spatial correlation of city-level total investment has a threshold effect based on the leaders' age. In contrast to Yu et al. (2016) who try various cutoff ages to see the change of the spatial correlation over leader's age, our threshold SPD model can directly estimate the threshold age. First, a test of no threshold effects is carried out using the sup-Wald test developed in Section 2.3, and then if this test is rejected, a confidence interval for the threshold parameter is constructed by inverting the LR test given in Subsection 2.2.4.

Model and data. Following the above discussions, we consider the following model:

$$inv_{it} = \lambda_1 \sum_{j=1}^{n} w_{ij,t} inv_{jt} + \lambda_2 \sum_{j=1}^{n} w_{ij,t} inv_{jt} \mathbf{1}\{age_{it} \leqslant \gamma\} + x_{it}\beta + \mu_i + \alpha_t + v_{it},$$

where inv_{it} denotes the total investment of local government of city i in year t, age_{it} denotes the age of the local leader of city i in year t, x_{it} is a vector of time-varying regressors including fiscal revenue, fiscal expenditure, population, manufacturing ratio, GDP per capita and a set of province level variables: fiscal revenue, fiscal expenditure, and public capital investment, μ_i and α_t are the two-way fixed effects, and v_{it} is the idiosyncratic error. We follow Yu et al. (2016) and define those same-province cities whose within-province rankings of GDP per capita are either one place above or below a city as this city's spatial neighbors, because they are the main competitors in the tournament competition. Because there is no theoretical evidence to justify the threshold effects for regression coefficients, they are not included.

We analyze the annual total investments (in RMB) of 338 cities in the 27 provinces in mainland China from 2010 to 2012. Economic data is from *Fiscal Statistics of Cities and Counties in China*, *China City Statistical Yearbook* and *China Statistical Yearbook for Regional Economy* for the period 2010-2012. The ages of leaders are obtained from local government websites. The data is standardized to make all the variables have comparable scales.

Test for the presence of threshold effects. Before the estimation of the model, we conduct hypothesis testing on the presence of the age-of-leader threshold effect. In China's local official system, there are two types of leaders in the local governments, party secretaries and mayors. Party secretaries are mainly responsible for personnel work and overall decision-making while the mayors are for the formulation and implementation of specific economic and social policies so that Yao and Zhang (2013) find the weight of economic performance is lower for the party secretary than for the mayor in the assessment of local leaders. Therefore, we will consider these two groups of leaders, separately. In addition, as Yu et al. (2016) find the age-of-leader effects on political competitions are more clear among old leaders and old leaders have different spatial responsiveness to their young and old neighbors, we also separate the leaders into young and old groups. The old leaders are defined as those whose ages are above the median age (49 for the mayor group and 52 for the party secretary group).

Table 2.4 reports the $\sup W_{nT}$ statistics and the associated bootstrap p-values based on 500 bootstrap replications for both mayor and party chief groups. The row labeled "all vs all" considers the spatial correlation among all the leaders in the group. "old vs all" considers the spatial correlation between old leaders and all their neighbors. Similarly, "old vs old" considers the spatial correlation among all the old leaders in the group, and "old vs young" considers the spatial correlation between old leaders and their young neighbors. Apparently, we can reject the null hypothesis of no threshold effect at the 10% level for "old vs all" and 1% level for "old vs old" patterns in the mayor group.

Estimation results. Table 2.5 reports the regression results for the two scenarios when we can reject the null hypothesis of no threshold effect. "Model 1" and "Model 2" are corresponding to the "old vs all" pattern and "old vs old" pattern in the mayor group, respectively. The estimations of threshold coefficient γ are 54.75 (54 years and 9 months) and 55.33 (55 years and 4 months) for these two models, respectively. We also report the 95% confidence intervals that are based on the likelihood ratio test. The estimations of λ_1 in

Table 2.4. Test for the presence of threshold effect.

	Mayor		Party Chief	
	supW_{nT}	p-value	supW_{nT}	p-value
all vs all	5.114	0.648	6.977	0.388
old vs all	8.325	0.067	7.110	0.157
old vs old	13.27	0.007	7.635	0.221
old vs young	3.418	0.411	6.202	0.120

Table 2.5. Estimates of spatial competitions in government investments among Chinese cites.

Government Investments	Model 1		Model 2	
	Mayor: old vs all		Mayor: old vs old	
Threshold estimate:				
Threshold (γ)	54.75		55.33	
95% confidence interval	[49.33, 57.92]		[49.75, 58.00]	
Spatial effects:				
Base effect (λ_1)	0.012	(0.001)	0.008	(0.002)
Threshold effect (λ_2)	0.071	(0.001)	0.099	(0.002)
Impact of covariates:				
Fiscal revenue	0.308	(0.004)	0.311	(0.004)
Fiscal expenditure	0.127	(0.002)	0.123	(0.002)
Population	0.024	(0.001)	0.025	(0.001)
manufacturing ratio	0.771	(0.004)	0.767	(0.004)
GDP per capita	-0.140	(0.002)	-0.137	(0.002)
Provincial fiscal revenue	0.259	(0.006)	0.251	(0.006)
Provincial fiscal expenditure	-0.129	(0.006)	-0.128	(0.006)
Public capital investment	-0.009	(0.000)	-0.011	(0.000)

Note: The values without parentheses are the QMLE for all the parameters. The values in parentheses are the corresponding standard errors.

these two models suggest that the spatial correlations when the ages of local leaders are beyond the threshold levels are slightly positive (0.012 and 0.008). In the contrast, when the ages of local leaders are below the threshold levels, the spatial correlations among local investments are the estimations of $\lambda_1 + \lambda_2$ and thus become strongly positive as λ_2 are positive with a much larger magnitude. These empirical findings are in line with our theoretical expectation, considering that the city leaders normally take office in their forties or fifties and the mandatory retirement age for them is 60. A more comprehensive study on this topic is of interest as future research.

2.6 Extensions

We have by far focused on a threshold SPD model (2.1) that contains only a spatial lag (SL) structure with additive fixed effects, for ease of exposition. The proposed estimation and inference methods are in fact quite general and can be extended to include additional features in the model such as spatial error dependence, serial correlation, time dynamics, multiple threshold effects, threshold effects on error parameters, interactive fixed effects, etc. An immediate and much-needed extension is the inclusion of spatial error (SE) effect:

$$Y_t = \lambda_{10} W_t Y_t + \lambda_{20} d_t(\gamma_0) W_t Y_t + X_t\beta_{10} + d_t(\gamma_0) X_t \beta_{20} + \mu_0 + \alpha_{t0} l_n + U_t,\ U_t = \rho_0 M_t U_t + V_t,$$

for $t = 1, \ldots, T$, where parameter ρ and weight matrices $\{M_t\}$ together characterize the SE effects, and the other parts are defined in Model (2.1). Let $B_t(\rho) = I_n - \rho M_t$. The quasi Caussian loglikelihood function of all the parameters takes the form

$$\begin{aligned}\ell_{nT}(\theta,\rho,\gamma,\mu,\alpha) = & -\tfrac{nT}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln|A_t(\lambda,\gamma)| + \sum_{t=1}^T \ln|B_t(\rho)| \\ & - \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T V_t'(\phi,\rho,\gamma,\mu,\alpha)V_t(\phi,\rho,\gamma,\mu,\alpha),\end{aligned}$$

where $V_t(\phi,\rho,\gamma,\mu,\alpha) = B_t(\rho)[\mathbb{Y}_t(\lambda,\gamma) - \mathbb{X}_t(\gamma)\beta - \mu - \alpha_t l_n]$. To make μ and α identifiable, we impose $\sum_{t=1}^T B_t'(\rho)B_t(\rho)\alpha_t l_n = 0$. Given (ϕ,ρ,γ), $\ell_{nT}(\theta,\rho,\gamma,\mu,\alpha)$ is partially maximized at

$$\begin{aligned}&\hat{\mu}(\phi,\rho,\gamma) = [\textstyle\sum_{t=1}^T B_t'(\rho)B_t(\rho)]^{-1}\sum_{t=1}^T B_t'(\rho)B_t(\rho)[\mathbb{Y}_t(\lambda,\gamma) - \mathbb{X}_t(\gamma)\beta] \ \text{ and} \\ &\hat{\alpha}_t(\phi,\rho,\gamma) = (l_n'B_t'(\rho)B_t(\rho)l_n)^{-1}l_n'B_t'(\rho)B_t(\rho)[\mathbb{Y}_t(\lambda,\gamma) - \mathbb{X}_t(\gamma)\beta - \hat{\mu}(\phi,\rho,\gamma)].\end{aligned}$$

Thus, the adjusted concentrated quasi loglikelihood function corresponding to (2.4) becomes

$$\begin{aligned}\ell^*_{nT}(\theta,\rho,\gamma) = & -\tfrac{n(T-1)}{2}\ln(2\pi\sigma^2) + \tfrac{T-1}{T}\textstyle\sum_{t=1}^T \ln|A_t(\lambda,\gamma)| + \tfrac{T-1}{T}\sum_{t=1}^T \ln|B_t(\rho)| \\ & - \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T \ddot{V}_t'(\phi,\rho,\gamma)Q_t(\rho)\ddot{V}_t(\phi,\rho,\gamma),\end{aligned}$$

where $Q_t(\rho) = I_n - B_t(\rho)l_n(l_n'B_t'(\rho)B_t(\rho)l_n)^{-1}l_n'B_t'(\rho)$ and $\ddot{V}_t(\phi,\rho,\gamma) = B_t(\rho)$ $[\mathbb{Y}_t(\lambda,\gamma) - \mathbb{X}_t(\gamma)\beta - \hat{\mu}(\phi,\rho,\gamma)]$. In special cases when $\{M_t\}$ are time-invariant and row-normalized, $Q_t(\rho)$ is reduced to J_n as $B_t(\rho)l_n = (1-\rho)l_n$, and $\ddot{V}_t(\phi,\rho,\gamma)$ becomes $B_t(\rho)[\widetilde{\mathbb{Y}}_t(\lambda,\gamma) - \widetilde{\mathbb{X}}_t(\gamma)\beta]$. In general, the adjusted QML estimators of θ, ρ and γ are simply

$$(\hat{\theta}_{nT}, \hat{\rho}_{nT}, \hat{\gamma}_{nT}) = \underset{(\theta,\rho,\gamma)\in\Theta\times\Delta_\rho\times\Gamma}{\text{argmax}} \ \ell^*_{nT}(\theta,\rho,\gamma),$$

where Δ_ρ is the parameter space for ρ.

In practice, we can first maximize the objective function conditional on γ to get $\hat{\theta}_{nT}(\gamma)$ and $\hat{\rho}_{nT}(\gamma)$, and then apply the grid search algorithm in Subsection 2.2.2 to obtain $\hat{\gamma}_{nT}$. With some additional conditions (e.g., both $\|M_t\|_1$ and $\|M_t\|_\infty$ are bounded; both $\|B_t^{-1}(\rho)\|_1$ and $\|B_t^{-1}(\rho)\|_\infty$ are bounded on Δ_ρ; ρ is identifiable), we expect the estimation error of $\hat{\gamma}_{nT}$ still have asymptotically negligible effects on $(\hat{\theta}_{nT}, \hat{\rho}_{nT})$, and thus we can establish similar results to those in Theorems 2.1 - 2.5. Moreover, to construct confidence interval for γ, we construct the LR statistic in the same way as in Subsection 2.2.4,

$$LR_{nT}(\gamma) = 2[\ell^*_{nT}(\hat{\theta}_{nT}, \hat{\rho}_{nT}, \hat{\gamma}_{nT}) - \ell^*_{nT}(\hat{\theta}_{nT}(\gamma), \hat{\rho}_{nT}(\gamma), \gamma)].$$

When errors are normally distributed, the asymptotic distribution of $LR_{nT}(\gamma_0)$ is still pivotal, following the distribution of $\mho$. In this case, the asymptotic $1-\alpha$ confidence interval for γ is the set of values of γ satisfying $LR_{nT}(\gamma) \leqslant \mho_{1-\alpha}$. Finally, to test the presence of threshold effects, we first derive the CQS functions of $\ell^*_{nT}(\theta, \rho, \gamma)$ with respect to θ and ρ, and then adjust them, following the two principals in Section 2.3, to obtain the AQS functions. Thus, the sup-Wald test statistic and bootstrap procedure can be constructed in a similar manner.

Our estimation and inference methods can also be extended to handle models with other additional features. Firstly, extension to allow for serial correlation in the error term (e.g., $v_{it} = \varrho v_{i,t-1} + e_{it}$ with $|\varrho| < 1$) is also straightforward like the above one with the `SE` structure. We expect the arguments and ideas behind estimation and inference methods can still be applied with minor modifications. Secondly, we can generalize our model to the dynamic SPD framework. When T is large, the direct QML approach should provide a consistent estimation for all the parameters, and thus the asymptotic properties of these QMLEs can be derived in a standard manner. When T is fixed, the analysis will become complicated as adjustments to the concentrated QML function are required to deal with the incidental parameters problems coming from both the initial condition and the concentration. Thirdly, extension to include multiple thresholds (Hansen, 1999) is also of theoretical and practical interest. For this extension, our QML approach is still appropriate and the objective function with multiple thresholds corresponding (2.4) is also straightforward to construct. Thus, the adjusted QML estimators of all the parameters including multiple threshold parameters jointly minimize the new objective function. In practice, the grid search over multiple thresholds may require an excessive amount of computation. We recommend using the sequential estimation method with refinement (Bai, 1997; Hansen, 1999) to avoid this computational burden.

Fourthly, our methods can also be extended to include the threshold effects on error parameters, e.g., error variance (Miao et al., 2020). In this case, the threshold effects on error parameters need to be incorporated into the QML function. For example, when error variance has threshold effects, for each observation the variance parameter will appear in the form of $\sigma_1^2 + \sigma_2^2 \mathbb{1}(q_{it} \leqslant \gamma)$, where σ_1^2 is the baseline parameter and σ_2^2 is its threshold effect. Finally, our methods can be extended to allow the individual and time fixed effects to appear in the model interactively. According to Miao et al. (2020), we would expect the concentrated QML estimation (with common factors being concentrated out) can provide a consistent estimation for all the parameters, including threshold parameter and factor loadings, when both n and T are large. Besides, we expect that the estimation error of the threshold estimate still has no asymptotic effect on the asymptotic properties of the other estimators and that the inference methods in this chapter can still be applied. However, formal studies on these extensions are still quite involved and can only be handled in future research.

2.7 Conclusion

In this chapter, we consider estimation and inference for a threshold spatial panel data model with both individual and time fixed effects, where threshold effects are allowed for both spatial and regression parameters. The presence of the threshold effects renders the commonly used orthogonal transformation approach inapplicable to wipe out fixed effects. We propose an *adjusted quasi maximum likelihood* estimation method, where the objective function is obtained by adjusting the concentrated quasi loglikelihood function (with fixed effects being concentrated out) to 'recover' the effect of degrees of freedom loss due to the estimation of these incidental parameters. We study the asymptotic properties of the adjusted QML estimators in the diminishing-threshold-effect framework and propose a likelihood ratio statistic to construct confidence intervals for the threshold parameter. We also consider the hypothesis testing on the presence of threshold effects and a sup-Wald statistic based on an M-estimator is proposed. Monte Carlo results show excellent performance of the proposed estimation and inference methods. We apply our model to study the age-of-leader effects on political competitions across Chinese cities and find competitions only exist among city leaders who are younger than a threshold age.

Chapter 3

Spatial Panel Data Models with Time-Varying Network Structures

3.1 Introduction

One important application of spatial models is to be used as a social interaction model in social economics, where the spatial weight matrix, also called sociomatrix (Liu and Lee, 2010), is used to capture the information on the connections of nodes (individuals) in a social network. It provides methods to model aggregate behavior as the outcome of individual decisions when these decisions are made interactively in the network (Durlauf and Young, 2001). According to Manski (1993), these interaction effects can be separated into endogenous effects, exogenous (contextual) effects, and unobserved correlation effects. Since Lee (2007), researchers have recognized that these effects can be captured by spatial lag term, Durbin term, and group-specific fixed effects, respectively, in spatial autoregressive (SAR) models. Many papers then followed this work and studied these interaction effects using the same model specifications, e.g., Bramoullé et al. (2009), Lin (2010), Liu and Lee (2010), and Lee et al. (2010) under spatial cross-sectional data setup, and Kwok (2019) and Han et al. (2019) under spatial panel data (SPD) setup.

SPD models have received increasing attention from econometricians since

Anselin (1988). See, for example, Baltagi et al. (2003), Lee and Yu (2010), Baltagi and Yang (2013a,b), Yang et al. (2016), and Liu and Yang, (2020). Besides taking into account spatial and social interactions, SPD models are also able to have full control for the unobserved heterogeneity. In this chapter, we reformulate SPD models as social interaction models, where we allow the network structures to be time-varying. That is, each individual could switch social groups or/and their network connections within these social groups may also change over time. This phenomenon is quite common in practice. For example, one major application of the social interaction model is to identify peer effects on student academic achievement (Lin, 2010; Han et al., 2019), where a social group is specified at grade or class level, and one's network connections are made up of friends. Thus, it is common to see that some students change classes over time. In addition, each student may also have different network connections over time in the process of expanding his/her circle of friends. Other major applications include studying the neighborhood influences on consumption behaviors (Case, 1991), welfare participation (Bertrand et al., 2000), and secondary school enrollment decisions (Bobonis and Finan, 2009), all of which are based on neighborhood networks. They are also changeable over time because people might move to different communities. However, this practical feature is not considered in Kwok's (2019) model as he assumes that the sociomatrices and group-specific effects are time-invariant.

Identification issues regarding social interactions models is also a focus of attention in the literature. In a pioneer work, Manski (1993) first pointed out a 'reflection' problem for social interactions in a linear-in-means model, which refers to the difficulty of identification between endogenous and exogenous interaction effects. Lee (2007) considers a SAR model where each node is equally influenced by all the peers in its group but not by itself, as the information on how individuals interact within a group is unobservable. He finds that variations in group sizes can make the identification possible, but it can be weak when all the group sizes are large. Bramoullé et al. (2009) and Lee et al. (2010) then also consider a SAR model but assume the information on network structure is available so that each node is only influenced by its connected nodes (e.g., friends or roommates) in a network (e.g., class, grade or school). They show that the various interaction effects are generally identified by using this information. However, all the above works assume that the sociomatrix is row-normalized such that each row sums to unity, which might have some potential pitfalls (Kelejian and Prucha, 2010) or limitations (e.g., isolated units are not covered) in practice. For this reason, Liu and Lee (2010) abandon this assumption and find the variation in the Bonacich centrality measure (number of connections of each node) can instead yield identification for various interaction effects. As regards to panel social interaction model, Kwok (2019) derives the identification conditions for higher-order SPD models with network structures under the row-normalization assumption. Han et al. (2019) do not provide the identification conditions for their model. In this chapter, we assume that the network information is available and that sociomatrices are not necessary to be row-normalized. Under these assumptions, we derive the identification conditions of various social interaction effects in SPD models with time-varying network structures.

As discussed above, one important advantage of social interaction models in the SPD framework is that they have the ability to have full control of three-dimensional (3-D) unobserved heterogeneity. Besides the common individual- and time-specific effects, the unobserved group-specific effects are additionally used to capture the unobserved correlation effects in a group in that individuals in the same group tend to have similar characteristics or face similar environments (Manski, 1993). However, Kwok (2019) fails to take into account the former two specific effects in his model. Besides, the literature pertaining to multi-dimensional SPD models mainly focuses on random-effects models, e.g., Le Gallo and Pirotte (2017), Baltagi et al. (2003), and Baltagi et al. (2015). Although Han et al. (2019) consider an SPD model with one particular type of three-way fixed effects and develop a Bayesian Markov chain Monte Carlo sampling approach to estimate the model, the well-known *incidental parameters problem* of Neyman and Scott (1948) is ignored in their paper. Hence, there seem to be no formal considerations being given to the SPD models with multi-dimensional fixed effects. In practice, the three types of fixed effects can appear either additively or interactively according to economic theory[1], so that the total number of possible specifications of these fixed effects can be as large as 2^6 (including the one with no fixed effects). In this chapter, we only consider the empirically most meaningful specification for social interaction models as proposed in Han et al. (2019): time-invariant individual-specific effects plus time-varying group-specific effects. For example, it is plausible to assume that the inner abilities of students remain unchanged over time but students in the same classroom may face different environments (e.g., teachers) in different semesters. Although we only focus on this single specification, the main idea behind the proposed methodology can be easily adapted to deal with all the other specifications.

In our model, the number of groups may grow with the sample size (the number of spatial units), and thereby direct estimation of the group-specific effects will give rise to the incidental parameters problem in that some parameter estimates are not consistent or asymptotically biased, similar to the direct estimation of two-way fixed effects (Lee and Yu, 2010). Another standard method is to transform the original model to wipe out the fixed effects. In view of the fact that the spatial matrices are time-varying and their row sums may not be constant (Liu and Lee, 2010), the transformed model would not have a well-defined SAR structure, and thus the (quasi) likelihood function cannot be formed. As an ML-type estimation often has better finite-samples properties than GMM/IV approach, it is desirable to propose a more general ML-type method for estimation.

To tackle the issues mentioned above, we propose a general M-, or *adjusted quasi score* (AQS) method. The method starts from the joint quasi score functions of both the common parameters and fixed effects, then concentrates out the fixed effects to give the concentrated quasi score functions, and then adjusts these concentrated score functions to give a set of unbiased estimating functions (the AQS functions) for the common parameters. Solving these AQS functions gives the AQS estimators of common parameters. Besides all the interaction effects mentioned above, we also allow the disturbances of

[1]See Balazsi et al. (2017) for enormous examples of empirical studies for multi-dimensional fixed effects models.

connected nodes to be spatially correlated to further capture similar preferences of these connected individuals as argued in Moffitt (2001). We show the AQS estimators are consistent and asymptotically normally distributed without asymptotic bias. Simulation results also show our AQS estimators have excellent finite sample performance.

The rest of this chapter is organized as follows. Section 3.2 introduces the SPD model with time-varying network structures, discusses the issues of direct QML estimation, and finally proposes the M-estimation method for estimating the model. The consistency and asymptotic normality of the AQS estimators are also studied in this section. Section 3.3 presents Monte Carlo results and Section 3.4 concludes. Proofs of the main results are given in the Appendices.

Some generic notations and conventions will be followed. I_m denotes an $m \times m$ identity matrix, $0_{m\times n}$ an $m \times n$ zero matrix, and l_m an $m \times 1$ vector of ones. For a square matrix, $|\cdot|$ denotes its determinant and $\texttt{tr}(\cdot)$ its trace. For a real symmetric matrix, $\gamma_{\min}(\cdot)$ and $\gamma_{\max}(\cdot)$ denote, respectively, its smallest and largest eigenvalues. For a real $n \times m$ matrix A, A' denotes its transpose, $A^\circ = A + A'$, $\|A\|_F$ is its Frobenius norm, $\|A\|_1$ its maximum absolute column sum norm, and $\|A\|_\infty$ its maximum absolute row sum norm. For a real $n \times m$ matrix A with a full column rank, $\mathbb{P}_A = A(A'A)^{-1}A'$ denotes the projection matrix into the column space of A, and $\mathbb{Q}_A = I_n - \mathbb{P}_A$ the projection matrix into the space orthogonal to the column space of A. The operator $\texttt{diag}(\cdot)$ forms a diagonal matrix by the diagonal elements of a square matrix or by the elements of a given vector, $\texttt{diagv}(\cdot)$ forms a column vector using the diagonal elements of a square matrix, and $\texttt{blkdiag}(\cdots)$ forms a block-diagonal matrix by placing the given matrices or vectors along the diagonal direction. The usual expectation and variance operators, $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$, correspond to true parameter values with a subscript 0.

3.2 SPD Model with Time-Varying Network Structures

3.2.1 The model with time-invariant grouping

To illustrate the main idea, we first focus on a simple model where individuals are not allowed to switch groups over time so that the group members are fixed in each period. Consider a study that lasts T periods and involves a total of n spatial units. They are divided into G social groups with group sizes $(s_1, \ldots, s_G)$ such that $\sum_{g=1}^{G} s_g = n$. Within the same group, individuals are connected according to the observable network structures which may still be time-varying as individuals may have more or fewer connections over time. The outcome of each individual is subjected to the social interaction effects of these connections. These give rise to a spatial panel data (SPD) model with time-varying network structures:

$$y_{igt} = \lambda \sum_{j=1}^{s_g} w_{ij,gt} y_{jgt} + x_{igt}\beta_1 + \sum_{j=1}^{s_g} w_{ij,gt} x_{jgt}\beta_2 + \mu_{ig} + \gamma_{gt} + u_{igt}, \quad (3.1)$$
$$u_{igt} = \rho \sum_{j=1}^{s_g} m_{ij,gt} u_{jgt} + v_{igt},$$

for $i = 1, \ldots, s_g$, $g = 1, \ldots, G$ and $t = 1, \ldots, T$. Note that y_{igt} is the dependent variable for individual i of group g at time period t. $w_{ij,gt}$ and $m_{ij,gt}$ are the (i, j)th elements of W_{gt} and M_{gt}, respectively, which are the sociomatrices of group g at time period t. In principle, $\{W_{gt}\}$ and $\{M_{gt}\}$ may or may not be the same. x_{igt} is the corresponding $1 \times k$ vector of time-varying exogenous regressors. Idiosyncratic errors $\{v_{jgt}\}$ are assumed to be iid with zero mean and variance σ^2. β_1 is the $k \times 1$ vectors of regression coefficients. In the social interaction literature, λ and β_2 represent the endogenous effect and the contextual effects, respectively, and ρ captures the unobservable correlated effects of connected individuals. μ_{ig} stands for the time-invariant individual-specific effect for unit i in group g. γ_{gt} stands for the time-varying group-specific effect that is shared by all the individuals in group g at time t. In this chapter, they both are allowed to correlate with the regressors in an arbitrary manner and hence are considered as fixed effects (FE).

Stacking all the s_g observations in group g yields

$$Y_{gt} = \lambda W_{gt} Y_{gt} + Z_{gt}\beta + \mu_g + \gamma_{gt} l_{s_g} + U_{gt}, \quad U_{gt} = \rho M_{gt} U_{gt} + V_{gt}, \quad (3.2)$$

where $g = 1, \ldots, G$, $t = 1, \ldots, T$, $Y_{gt} = (y_{1gt}, \ldots, y_{s_g,gt})'$, $Z_{gt} = (X_{gt}, W_{gt}X_{gt})$, $X_{gt} = (x'_{1gt}, \ldots, x'_{s_g,gt})'$, $U_{gt} = (u_{1gt}, \ldots, u_{s_g,gt})'$, $V_{gt} = (v_{1gt}, \ldots, v_{s_g,gt})'$, $\beta = (\beta'_1, \beta'_2)'$, and $\mu_g = (\mu_{1g}, \ldots, \mu_{s_g,g})'$. Further stacking all the groups in period

t, we have

$$Y_t = \lambda W_t Y_t + Z_t\beta + \mu + L\gamma_t + U_t, \quad U_t = \rho M_t U_t + V_t, \quad t = 1, \cdots, T, \quad (3.3)$$

where $Y_t = (Y_{1t}', \ldots, Y_{Gt}')'$, $Z_t = (Z_{1t}', \ldots, Z_{Gt}')'$, $U_t = (U_{1t}', \ldots, U_{Gt}')'$, $V_t = (V_{1t}', \ldots, V_{Gt}')'$, $W_t = \texttt{blkdiag}(W_{1t}, \ldots, W_{Gt})$, $L = \texttt{blkdiag}(l_{s_1}, \ldots, l_{s_G})$, $\mu = (\mu_1', \ldots, \mu_G')'$, and lastly, $\gamma_t = (\gamma_{1t}, \ldots, \gamma_{Gt})'$.

As discussed in the introduction, the above model can be adjusted to allow for all the other possible specifications of the three-way fixed effects, depending on the economic theory. The estimation strategy introduced later is flexible to deal with all of these specifications. A major advantage of these models is that they are able to have full control of unobserved heterogeneity along the three dimensions. Another important feature of the above model is that the row sums of time-varying sociomatrices may not be constant. This is commonly seen in network studies because there might be some *isolated* individuals (Bramoullé et al., 2009) whose connection groups are empty, i.e., the row sums in the sociomatrices corresponding to these individuals are zero. But the row sums for the other individuals who have connections will not be zero.

3.2.2 The model with time-varying grouping

The above model can be further generalized to allow units to switch groups over time. In this case, the group members will be changing. We assume there are still n spatial units divided into G social groups in each time period. But the group sizes are now $\{s_{gt}\}$ such that $\sum_{g=1}^{G} s_{gt} = n$ for each t. In this case, the model for y_{igt} is the same as (3.1) except the value ranges of indices i and j now depends on both group g and time t, i.e., $i, j = 1, \ldots, s_{gt}$. Corresponding to (3.2), we stack all the s_{gt} observations in the g-th group at time t,

$$Y_{gt} = \lambda W_{gt} Y_{gt} + Z_{gt}\beta + \mu_g^{(t)} + \gamma_{gt} l_{s_{gt}} + U_{gt}, \quad U_{gt} = \rho M_{gt} U_{gt} + V_{gt}, \quad (3.4)$$

where $g = 1, \ldots, G$, $Y_{gt} = (y_{1gt}, \ldots, y_{s_{gt},gt})'$, $Z_{gt} = (X_{gt}, W_{gt}X_{gt})$, $X_{gt} = (x_{1gt}', \ldots, x_{s_{gt},gt}')'$, $U_{gt} = (u_{1gt}, \ldots, u_{s_{gt},gt})'$, $V_{gt} = (v_{1gt}, \ldots, v_{s_{gt},gt})'$, and $\mu_g^{(t)} = (\mu_{1g}, \ldots, \mu_{s_{gt},g})'$ representing the set of time-invariant individual fixed effects for these s_{gt} units. Then, we stack all the groups in period t and get

$$Y_t = \lambda W_t Y_t + Z_t\beta + \mu^{(t)} + L_t\gamma_t + U_t, \quad U_t = \rho M_t U_t + V_t, \quad t = 1, \cdots, T, \quad (3.5)$$

where $Y_t = (Y'_{1t}, \ldots, Y'_{Gt})'$, $Z_t = (Z'_{1t}, \ldots, Z'_{Gt})'$, $U_t = (U'_{1t}, \ldots, U'_{Gt})'$, $V_t = (V'_{1t}, \ldots, V'_{Gt})'$, $W_t = \texttt{blkdiag}(W_{1t}, \ldots, W_{Gt})$, $L_t = \texttt{blkdiag}(l_{s_{1t}}, \ldots, l_{s_{Gt}})$, $\mu^{(t)} = (\mu_1^{(t)\prime}, \ldots, \mu_G^{(t)\prime})'$ and $\gamma_t = (\gamma_{1t}, \ldots, \gamma_{Gt})'$. It is worth mentioning that although $\mu^{(t)}$ stands for time-invariant individual-specific effects, the elements of it may have various orders over time as the observations are sorted by groups in each time period.

3.2.3 Quasi-maximum likelihood estimation

It is well-known that the maximum likelihood (ML) method is usually more efficient than the 2SLS or GMM approaches, especially when the errors are normally distributed. Therefore, we start with the quasi-maximum likelihood (QML) estimation and then discuss the disadvantages it has. As the main ideas of constructing the loglikelihood functions for the above two models are the same, we discuss them at the same time below. For the model in Subsection 3.2.1, we let $\mathbf{Y} = (Y'_1, \ldots, Y'_T)'$, $\mathbf{W} = \texttt{blkdiag}(W_1, \ldots, W_T)$, $\mathbf{Z} = (Z'_1, \ldots, Z'_T)'$, $\mathbf{U} = (U'_1, \ldots, U'_T)'$, $\mathbf{V} = (V'_1, \ldots, V'_T)'$, $\mathbf{M} = \texttt{blkdiag}(M_1, \ldots, M_T)$, $\gamma = (\gamma'_1, \ldots, \gamma'_T)'$, $\mathbf{D}_\mu = l_T \otimes I_n$, and $\mathbf{D}_\gamma = I_T \otimes L$. Then, we can write model (3.3) into the vector form: $\mathbf{Y} = \lambda \mathbf{W}\mathbf{Y} + \mathbf{Z}\beta + \mathbf{D}_\mu \mu + \mathbf{D}_\gamma \gamma + \mathbf{U}$ and $\mathbf{U} = \rho \mathbf{M}\mathbf{U} + \mathbf{V}$. For the standard two-way FE-SPD model, the transformation approach (e.g., Lee and Yu, 2010; Yang et al., 2016) is usually preferred for eliminating the fixed effects as the transformed model remains in the same spatial structure and thus the (quasi) likelihood can be formed. However, it is hard to find such a transformation to wipe out general three-way fixed effects without affecting the spatial structure. Besides, this method requires that spatial weight matrices (sociomatrices) are time-invariant and row-normalized even for the two-way fixed effects models, and both of these two features are not met in the current model. Therefore, we consider using the direct QML estimation, i.e., we estimate fixed effects together with all the common parameters.

For this, we first obtain the concentrated quasi likelihood function with all fixed effects concentrated out and then maximize the concentrated function to get estimations for the common parameters. In the first step, as dummy matrix

$(\mathbf{D}_\mu, \mathbf{D}_\gamma)$ is rank-deficient, we have to impose some restrictions on the fixed effects parameters to make them concentratable. The two most widely used are either to set their average to zero or to leave out some of the parameters. In this chapter, we follow the latter approach and omit either some individual fixed effects or some group-time interactive fixed effects parameters. For each specific group, it has the same group members over all the periods. Therefore, we will not be able to identify its group effects of all the time periods separated from the individual effects belonging to this group, i.e., $\mu_{ig} + \gamma_{gt} = (\mu_{ig} + c) + (\gamma_{gt} - c)$ for an arbitrary c. To avoid this, we can drop either one μ-dummy or one γ-dummy for this group. For example, we can omit μ_{1g} or γ_{g1} for each group g. Thus, a total of G number of μ-dummies or γ-dummies needs to be dropped. In this chapter, we consider dropping γ-dummies for example. Hence, the group-time effects become $\gamma^\star = (\gamma_2', \ldots, \gamma_T')'$ after omission. The dummy matrix $\mathbf{D}_\gamma$ now changes to $\mathbf{D}_\gamma^\star = [0_{n\times G(T-1)}; I_{T-1} \otimes L]$. Thus, our estimation will be based on the following model:

$$\mathbf{Y} = \lambda \mathbf{W}\mathbf{Y} + \mathbf{Z}\beta + \mathbf{D}\phi + \mathbf{U}, \quad \mathbf{U} = \rho \mathbf{M}\mathbf{U} + \mathbf{V}, \tag{3.6}$$

where $\phi = (\mu', \gamma^{\star\prime})'$ and $\mathbf{D} = (\mathbf{D}_\mu, \mathbf{D}_\gamma^\star)$ with a full column rank, $n + G(T-1)$.

Before we build up the loglikelihood function for the above model, some similar discussion can also be applied to the model in Subsection 3.2.2. Firstly, it is also possible to write the model into vector form: $\mathbf{Y} = \lambda \mathbf{W}\mathbf{Y} + \mathbf{Z}\beta + \mathbf{D}_\mu \mu + \mathbf{D}_\gamma \gamma + \mathbf{U}$ and $\mathbf{U} = \rho \mathbf{M}\mathbf{U} + \mathbf{V}$, where $\mathbf{D}_\mu$ and $\mathbf{D}_\gamma$ become dummy variable matrices that are defined following the observations order in (3.5), and all the other notations are defined same as above. However, the non-identification issues mentioned above now only apply to the groups whose members are fixed over time, i.e., the group size variation over time provides information for identification of the fixed effects. Thus, we can simply omit one γ-dummy for each of these groups. If none of the groups has fixed members, then we can arbitrarily omit one γ-dummy to avoid the simple dummy variable trap. Thus, we can find the estimation model corresponding to (3.6). In this case, the column rank of $\mathbf{D}$ changes to $n + G(T-1) + (G - \bar{r})$, where $\bar{r} = \max\{1, r\}$ and r is the number of groups with fixed members. Thus, model (3.6) can be treated as a special case of this model, where $\bar{r} = G$.

Let $\theta = (\beta', \sigma^2, \delta')'$, where $\delta = (\lambda, \rho)'$. In this chapter, we call θ the set of *common* parameters and ϕ the set of *incidental* parameters. Let $\mathbf{A}_{nT}(\lambda) = I_{nT} - \lambda\mathbf{W}$ and $\mathbf{B}_{nT}(\rho) = I_{nT} - \rho\mathbf{M}$. For both of the above models, we have the quasi Gaussian loglikelihood function of θ and ϕ:

$$\ell_{nT}(\theta, \phi) = -\tfrac{nT}{2}\ln 2\pi - \tfrac{nT}{2}\ln\sigma^2 + \ln|\mathbf{A}_{nT}(\lambda)| + \ln|\mathbf{B}_{nT}(\rho)| - \tfrac{1}{2\sigma^2}\mathbf{V}'(\beta, \delta, \phi)\mathbf{V}(\beta, \delta, \phi), \tag{3.7}$$

where $\mathbf{V}(\beta, \delta, \phi) = \mathbf{B}_{nT}(\rho)[\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\beta - \mathbf{D}\phi]$.

Let $\mathbb{D}(\rho) = \mathbf{B}_{nT}(\rho)\mathbf{D}$. Given θ, $\ell_{nT}(\theta, \phi)$ is partially maximized at

$$\hat{\phi}_{nT}(\beta, \delta) = [\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)\mathbf{B}_{nT}(\rho)[\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\beta]. \tag{3.8}$$

Substituting $\hat{\phi}_{nT}(\beta, \delta)$ into $\ell_{nT}(\theta, \phi)$ gives the concentrated quasi loglikelihood function for θ:

$$\ell^c_{nT}(\theta) = -\tfrac{nT}{2}\ln 2\pi - \tfrac{nT}{2}\ln\sigma^2 + \ln|\mathbf{A}_{nT}(\lambda)| + \ln|\mathbf{B}_{nT}(\rho)| - \tfrac{1}{2\sigma^2}\tilde{\mathbf{V}}'(\beta, \delta)\tilde{\mathbf{V}}(\beta, \delta), \tag{3.9}$$

where $\tilde{\mathbf{V}}(\beta, \delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)[\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\beta]$ and $\mathbb{Q}_{\mathbb{D}}(\rho)$ is the projection matrix based on $\mathbb{D}(\rho)$. The direct quasi maximum likelihood (QML) estimator $\hat{\theta}_{\texttt{QML}}$ of θ maximizes $\ell^c_{nT}(\theta)$. However, Lee and Yu (2010) demonstrate in their work for a two-way FE-SPD model that the direct QMLEs of β and δ are consistent no matter when T is large or small, but their distributions are asymptotically centered only when T is small compared to n. They also show that when T is finite, the QMLE of σ^2 is inconsistent and its limiting distribution is degenerate. The reason for these is that such a direct estimation of the common parameters θ completely ignores the effects from the estimation of the incidental parameters ϕ – the well known *incidental parameters problem* of Neyman and Scott (1948). For the current model specification, as the number of groups is also allowed to be large , the incidental parameters problem could become more complicated. Thus, it is well expected that $\hat{\theta}_{\texttt{QML}}$ will also be inconsistent or asymptotically biased in some scenarios. Therefore, an alternative (and more general) ML-type approach is highly desirable.

3.2.4 Adjusted quasi score estimation

Under mild conditions, maximizing $\ell_{nT}^c(\theta)$ is equivalent to solving $S_{nT}^c(\theta) = 0$, where $S_{nT}^c(\theta) = \frac{\partial}{\partial\theta}\ell_{nT}^c(\theta)$, the set of the concentrated quasi score (CQS) functions. The fundamental reason of inconsistency or asymptotic bias for the direct QML estimation is that a necessary condition for consistency of QML estimators, $\text{plim}\frac{1}{nT}S_{nT}^c(\theta_0) = 0$, is violated due to the concentration of the incidental parameters ϕ. To see it more clearly, we first derive the set of the CQS functions:

$$S_{nT}^c(\theta) = \begin{cases} \frac{1}{\sigma^2}\mathbf{Z}'\mathbf{B}_{nT}'(\rho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma^4}[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - nT\sigma^2], \\ \frac{1}{\sigma^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_{nT}'(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{F}_{nT}(\lambda)], \\ \frac{1}{\sigma^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}_{nT}(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{G}_{nT}(\rho)], \end{cases} \tag{3.10}$$

where $\mathbf{F}_{nT}(\lambda) = \mathbf{W}\mathbf{A}_{nT}^{-1}(\lambda)$ and $\mathbf{G}_{nT}(\rho) = \mathbf{M}\mathbf{B}_{nT}^{-1}(\rho)$. At the true value θ_0 of θ, we have

$$\text{E}[S_{nT}^c(\theta_0)] = \begin{cases} 0_k, \\ -\frac{n+G(T-1)+(G-\bar{r})}{2\sigma_0^2}, \\ \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho_0)\mathbf{B}_{nT}(\rho_0)\mathbf{F}_{nT}(\lambda_0)\mathbf{B}_{nT}^{-1}(\rho_0)] - \mathtt{tr}[\mathbf{F}_{nT}(\lambda_0)], \\ \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho_0)\mathbf{G}_{nT}(\rho_0)] - \mathtt{tr}[\mathbf{G}_{nT}(\rho_0)]. \end{cases} \tag{3.11}$$

Thus, we have $\lim_{nT\to\infty}\frac{1}{nT}\text{E}[S_{nT}^c(\theta_0)] \neq 0$ in general. According to Lemma A.6, this suggests $\text{plim}_{nT\to\infty}\frac{1}{nT}S_{nT}^c(\theta_0) \neq 0$, and therefore $\hat{\theta}_{\mathtt{QML}}$ cannot be consistent in general.

However, we note that $\text{E}[S_{nT}^c(\theta_0)]$ depends only on the common parameters θ_0 and the observables. It therefore offers a feasible way to analytically correct the CQS functions to give a set of unbiased estimating functions, or the *adjusted quasi score* (AQS) functions, as $S_{nT}^*(\theta_0) = S_{nT}^c(\theta_0) - \text{E}[S_{nT}^c(\theta_0)]$, which takes the form at the general θ:

$$S^*_{nT}(\theta) = \begin{cases} \frac{1}{\sigma^2}\mathbf{Z}'\mathbf{B}'_{nT}(\rho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma^4}\big[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - N_1\sigma^2\big], \\ \frac{1}{\sigma^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}'_{nT}(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)\mathbf{F}_{nT}(\lambda)\mathbf{B}^{-1}_{nT}(\rho)], \\ \frac{1}{\sigma^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}_{nT}(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_{nT}(\rho)], \end{cases} \tag{3.12}$$

where $N_1 = (n-G)(T-1)-(G-\bar{r})$, the *effective* sample size after taking into account the estimation of fixed effects. When $\bar{r}$ is larger, there are fewer fixed effects dummies for estimation (concentration) as more dummies are omitted for identification, yielding a larger effective sample size. Solving the AQS equations: $S^*_{nT}(\theta) = 0$, gives the AQS estimator of θ, i.e., $\hat{\theta}^*_{nT} = \arg\{S^*_{nT}(\theta) = 0\}$. It is easy to verity that $\mathrm{E}[S^*_{nT}(\theta_0)] = 0$ and $\mathrm{plim}\frac{1}{nT}S^*_{nT}(\theta_0) = 0$, making it possible for $\hat{\theta}^*_{nT}$ to be $\sqrt{N_1}$-consistent with a proper limiting distribution.

The AQS approach falls in spirit to the "*Modified Equations of Maximum Likelihood*" of Neyman and Scott (1948, Sec. 5) or "*Bias-Correction of the Moment Equation*" of Arellano and Hahn (2007), in searching for a potential method to handle the incidental parameters problem. In the special case of a two-way fixed effects SPD panel with time-invariant and row-normalized spatial weight matrices, our AQS method is equivalent to the QML method of Lee and Yu (2010) based on orthonormal transformations, with an effective sample size $(n-1)(T-1)$.

The root-finding process for the AQS estimation can be simplified by first solving the equations for β and σ^2, giving the constrained AQS estimators of β and σ^2:

$$\hat{\beta}^*_{nT}(\delta) = [\mathbb{Z}'(\rho)\mathbb{Z}(\rho)]^{-1}\mathbb{Z}'(\rho)\mathbf{C}_{nT}(\delta)\mathbf{Y} \quad \text{and} \quad \hat{\sigma}^{*2}_{nT}(\delta) = \tfrac{1}{N_1}\hat{\mathbf{V}}'(\delta)\hat{\mathbf{V}}(\delta), \tag{3.13}$$

where $\mathbb{Z}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)\mathbf{Z}$, $\mathbf{C}_{nT}(\delta) = \mathbf{B}_{nT}(\rho)\mathbf{A}_{nT}(\lambda)$ and $\hat{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\hat{\beta}^*_{nT}(\delta),\delta)$. Substituting $\hat{\beta}^*_{nT}(\delta)$ and $\hat{\sigma}^{*2}_{nT}(\delta)$ back into (3.12) gives the concentrated AQS functions of δ:

$$S^{*c}_{nT}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}^{*2}_{nT}(\delta)}\mathbf{Y}'\mathbf{W}'\mathbf{B}'_{nT}(\rho)\hat{\mathbf{V}}(\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)\mathbf{F}_{nT}(\lambda)\mathbf{B}^{-1}_{nT}(\rho)], \\ \frac{1}{\hat{\sigma}^{*2}_{nT}(\delta)}\hat{\mathbf{V}}'(\delta)\mathbf{G}_{nT}(\rho)\hat{\mathbf{V}}(\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_{nT}(\rho)]. \end{cases} \tag{3.14}$$

Solving the concentrated estimating (or AQS) equations, $S^{*c}_{nT}(\delta) = 0$, we obtain the unconstrained AQS estimator $\hat{\delta}^*_{nT}$ of δ. Thus the unconstrained AQS estimators of β and σ^2 are $\hat{\beta}^*_{nT} \equiv \hat{\beta}^*_{nT}(\hat{\delta}^*_{nT})$ and $\hat{\sigma}^{*2}_{nT} \equiv \hat{\sigma}^{*2}_{nT}(\hat{\delta}^*_{nT})$. The AQS estimator of θ is thus $\hat{\theta}^*_{nT} = (\hat{\beta}^{*\prime}_{nT}, \hat{\sigma}^{*2}_{nT}, \hat{\delta}^*_{nT})'$.

We can see from the preceding developments that a significant feature of this method is that it enables consistent estimation of all parameters, including σ^2, with the joint asymptotic distribution of the AQS estimators being centered as long as N_1 is large. As a result, all the problems associated with incidental parameters problem have been resolved. Furthermore, we have no constraints on the proportions of n and T as they grow to infinity, group sizes can be large or small, and either T can even be fixed. Because the method is based on adjusted quasi-score functions, it may inherit some of the advantages of maximum likelihood estimation, such as efficiency. See also Hsiao (2018) for more discussions on the advantages of the ML-type approach compared with GMM estimation.

3.2.5 Asymptotic properties of the AQS estimators

Denote a parametric quantity evaluated at the true parameter values by dropping its argument(s), e.g., $\mathbf{A}_{nT} \equiv \mathbf{A}_{nT}(\lambda_0)$, $\mathbf{B}_{nT} \equiv \mathbf{B}_{nT}(\rho_0)$, and $\mathbf{C}_{nT} \equiv \mathbf{C}_{nT}(\delta_0)$. Let Δ be the parameter space for δ, and Δ_λ and Δ_ρ be the subspaces for λ and ρ, respectively. Consistency and asymptotic normality of the proposed AQS estimators for the SPD model with time-varying network structures are established under the following set of regularity conditions.

Assumption A: *The innovations v_{igt} are iid for all i and t with mean zero, variance σ_0^2, and $E|v_{igt}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.*

Assumption B: *The space Δ is compact, and the true parameters δ_0 lie in its interior.*

Assumption C: (i) *The elements of* $\mathbf{Z}$ *are non-stochastic and bounded, uniformly in i and t, and* (ii) $\lim_{nT\to\infty} \frac{1}{nT}\mathbb{Z}'(\rho)\mathbb{Z}(\rho)$ *exists and is non-singular, uniformly in $\rho \in \Delta_\rho$.*

Assumption D: $\{W_t\}$ *and* $\{M_t\}$ *are known time-varying matrices.* $\mathbf{W}$ *and* $\mathbf{M}$ *are such that* (i) *their elements are at most of uniform order h_n^{-1} such* *that* $\frac{h_n}{n} \to 0$, *as* $n \to \infty$; (*ii*) *their diagonal elements are zero; and* (*iii*) $\|\mathbf{W}\|_\infty$, $\|\mathbf{W}\|_1$, $\|\mathbf{M}\|_\infty$, *and* $\|\mathbf{M}\|_1$ *are all bounded.*

Assumption E: *For* $\mathbb{A}(\varpi) = \mathbf{A}_{nT}(\lambda)$ *or* $\mathbf{B}_{nT}(\rho)$ *with* $\varpi = \lambda$ *or* ρ,

(*i*) *both* $\|\mathbb{A}^{-1}\|_\infty$ *and* $\|\mathbb{A}^{-1}\|_1$ *are bounded;*

(*ii*) *either* $\|\mathbb{A}^{-1}(\varpi)\|_\infty$ *or* $\|\mathbb{A}^{-1}(\varpi)\|_1$ *is bounded, uniformly in* $\varpi \in \Delta_\varpi$;

(*iii*) $0 < \underline{c}_\varpi \leq \inf_{\varpi\in\Delta_\varpi} \gamma_{\min}[\mathbb{A}'(\varpi)\mathbb{A}(\varpi)] \leq \sup_{\varpi\in\Delta_\varpi} \gamma_{\max}[\mathbb{A}'(\varpi)\mathbb{A}(\varpi)] \leq \bar{c}_\varpi < \infty$;

(*iv*) *both* $\|\mathbb{Q}_{\mathbb{D}}(\rho)\|_1$ *and* $\|\mathbb{Q}_{\mathbb{D}}(\rho)\|_\infty$ *are bounded, uniformly in* $\rho \in \Delta_\rho$.

Assumption F: n *is large, and* T *is large or small. As* nT *goes to infinity,* $\frac{N_1}{nT}$ *tends to an non-zero constant.*

Assumptions A-E are standard in the spatial econometrics literature (see, e.g., Lee and Yu, 2010) except Assumption E(*iv*). This additional condition is necessary to facilitate the study of the asymptotic properties of the spatial estimators. For the time-invariant grouping model, this condition holds as long as

$$B_{gt}(\rho)D_{gt}[\tfrac{1}{GT}\textstyle\sum_{t=1}^{T}\sum_{g=1}^{G} D'_{gt}B'_{gt}(\rho)J_{gt}(\rho)B_{gt}(\rho)D_{gt}]^{-1}D'_{qs}B'_{qs}(\rho)$$

is bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$ for all (g, t, q, s), where $J_{gt}(\rho)$ equals to I_{s_g} for $t = 1$ and $I_{s_g} - B_{gt}(\rho)l_{s_g}[l'_{s_g}B'_{gt}(\rho)B_{gt}(\rho)l_{s_g}]^{-1} l'_{s_g}B'_{gt}(\rho)$ for $t = 2, \dots, T$, and $B_{gt}(\rho) = I_{s_g} - \rho M_{gt}$ for $t = 1, \dots, T$ (see Lemma A.3 for details). Assumption F allows (*a*) both n and T are large and (*b*) n is large and T is finite. Meanwhile, the second part of Assumption F suggests that G can also be either large or small. All the scenarios encounter the so-called incidental parameters problem of Neyman and Scott (1948) due to the direct estimation of the fixed effects. It could lead to the asymptotic bias or inconsistency in the estimation of the common parameters. As the transformation strategy is inapplicable to handle this incidental parameters problem in our model, a new (AQS) method is therefore introduced.

We first prove the consistency of $\hat{\delta}^*_{nT}$. The key step in the proof is to compare $S^{*c}_{nT}(\delta)$ with its population counterpart. Let $\bar{S}^*_{nT}(\theta) = \mathrm{E}[S^*_{nT}(\theta)]$.

Given δ, $\bar{S}^*_{nT}(\theta) = 0$ is partially solved at

$$\bar{\beta}^*_{nT}(\delta) = [\mathbb{Z}'(\rho)\mathbb{Z}(\rho)]^{-1}\mathbb{Z}'(\rho)\mathbf{C}_{nT}(\delta)\mathrm{E}(\mathbf{Y}) \quad \text{and} \quad \bar{\sigma}^{*2}_{nT}(\delta) = \tfrac{1}{N_1}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\bar{\mathbf{V}}(\delta)], \tag{3.15}$$

where $\bar{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\bar{\beta}^*_{nT}(\delta), \delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)[\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\bar{\beta}^*_{nT}(\delta)]$. Substituting $\bar{\beta}^*_{nT}(\delta)$ and $\bar{\sigma}^{*2}_{nT}(\delta)$ into the δ-component of $\bar{S}^*_{nT}(\theta)$, we obtain the population counterpart of $S^{*c}_{nT}(\delta)$ as

$$\bar{S}^{*c}_{nT}(\delta) = \begin{cases} \frac{1}{\bar{\sigma}^{*2}_{nT}(\delta)}\mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}'_{nT}(\rho)\bar{\mathbf{V}}(\delta)] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)\mathbf{F}_{nT}(\lambda)\mathbf{B}^{-1}_{nT}(\rho)], \\ \frac{1}{\bar{\sigma}^{*2}_{nT}(\delta)}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\mathbf{G}_{nT}(\rho)\bar{\mathbf{V}}(\delta)] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_{nT}(\rho)]. \end{cases} \tag{3.16}$$

Clearly, $S^{*c}_{nT}(\hat{\delta}^*_{nT}) = 0$ by construction. Also, it is easy to see that $\bar{S}^{*c}_{nT}(\delta_0) = 0$ as $\bar{\beta}^*_{nT}(\delta_0) = \beta_0$ and $\bar{\sigma}^{*2}_{nT}(\delta_0) = \sigma_0^2$. Thus, by theorem 5.9 of van der Vaart (1998), $\hat{\delta}^*_{nT}$ will be consistent for δ_0 if $\sup_{\delta\in\Delta}\frac{1}{N_1}\left\|S^{*c}_{nT}(\delta) - \bar{S}^{*c}_{nT}(\delta)\right\| \xrightarrow{p} 0$ and the following identification condition holds:

Assumption G: *$inf_{\delta:d(\delta,\delta_0)\geq\epsilon}\left\|\bar{S}^{*c}_{nT}(\delta)\right\| > 0$ for every $\epsilon > 0$, where $d(\delta,\delta_0)$ is a measure of distance between δ and δ_0.*

Assumption G is a high level assumption being put up for simplicity of presentation. It can be shown to be true under some low level conditions. We have (see (B.5), Appendix B),

$$\bar{\sigma}^{*2}_{nT}(\delta) = \tfrac{1}{N_1}\eta'\mathbf{A}'^{-1}_{nT}\mathcal{Q}'_{nT}(\delta)\mathcal{Q}_{nT}(\delta)\mathbf{A}^{-1}_{nT}\eta + \tfrac{\sigma_0^2}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_{nT}(\delta)],$$

where $\mathcal{C}_{nT}(\delta) = \mathbf{C}_{nT}(\delta)(\mathbf{C}'_{nT}\mathbf{C}_{nT})^{-1}\mathbf{C}'_{nT}(\delta)$, $\mathcal{Q}_{nT}(\delta) = \mathbb{Q}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)$, and $\eta = \mathbf{Z}\beta_0 + \mathbf{D}\phi_0$. A sufficient condition for Assumption G to hold is either (a) or (b) holds, where

(a) $\frac{1}{\bar{\sigma}^{*2}_{nT}(\delta)}\eta'\mathbf{F}'_{nT}\mathbf{B}'_{nT}(\rho)\mathcal{Q}_{nT}(\delta)\mathbf{A}^{-1}_{nT}\eta + \mathtt{tr}[\frac{\sigma_0^2}{\bar{\sigma}^{*2}_{nT}(\delta)}\mathcal{P}_1(\delta) - \mathcal{P}_2(\delta)] \neq 0$, for $\delta \neq \delta_0$,

(b) $\frac{1}{\bar{\sigma}^{*2}_{nT}(\delta)}\eta'\mathbf{A}'^{-1}_{nT}\mathcal{Q}'_{nT}(\delta)\mathbf{G}_{nT}(\rho)\mathcal{Q}_{nT}(\delta)\mathbf{A}^{-1}_{nT}\eta + \mathtt{tr}[\frac{\sigma_0^2}{\bar{\sigma}^{*2}_{nT}(\delta)}\mathcal{P}_3(\rho)\mathcal{C}_{nT}(\delta) - \mathcal{P}_3(\rho)] \neq 0$, for $\delta \neq \delta_0$,

with $\mathcal{P}_1(\delta) = \mathbf{C}'^{-1}_{nT}\mathbf{C}'_{nT}(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)\mathbf{F}_{nT}\mathbf{B}^{-1}_{nT}$, $\mathcal{P}_2(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)\mathbf{F}_{nT}(\lambda)\mathbf{B}^{-1}_{nT}(\rho)$, and $\mathcal{P}_3(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)$. It is easy to see that $\mathcal{Q}_{nT}(\delta_0)\mathbf{A}^{-1}_{nT}\eta = 0$, $\mathcal{C}_{nT}(\delta_0) = I_{nT}$ and $\bar{\sigma}^{*2}_{nT}(\delta_0) = \sigma_0^2$. Hence the two quantities in (a) and (b) are identical 0 at the true parameter values. Once the consistency of $\hat{\delta}^*_{nT}$ is established, the consistency of $\hat{\beta}^*_{nT}$ and $\hat{\sigma}^{*2}_{nT}$ follows by Assumptions C-E.

Theorem 3.1. *Suppose Assumptions A-G hold. We have, as $nT \to \infty$,*

$\hat{\theta}^*_{nT} \xrightarrow{p} \theta_0$.

To derive the asymptotic distribution of $\hat{\theta}^*_{nT}$, we apply the mean value theorem: $0 = S^*_{nT}(\hat{\theta}^*_{nT}) = S^*_{nT}(\theta_0) + \frac{\partial}{\partial\theta'}S^*_{nT}(\bar{\theta})(\hat{\theta}^*_{nT} - \theta_0)$, where $\bar{\theta}$ lies between $\hat{\theta}^*_{nT}$ and θ_0, and its value varies over the rows of $\frac{\partial}{\partial\theta'}S^*_{nT}(\bar{\theta})$. Using $\tilde{\mathbf{V}}(\beta_0, \delta_0) = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$ and $\mathbf{Y} = \mathbf{A}^{-1}_{nT}(\eta + \mathbf{B}^{-1}_{nT}\mathbf{V})$,

$$S^*_{nT}(\theta_0) = \begin{cases} \frac{1}{\sigma_0^2}\mathbb{Z}'\mathbf{V}, \\ \frac{1}{2\sigma_0^4}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathbf{V} - N_1\sigma^2), \\ \frac{1}{\sigma_0^2}\mathbf{V}'\mathcal{P}_2\mathbf{B}_{nT}\eta + \frac{1}{\sigma_0^2}\mathbf{V}'\mathcal{P}_2\mathbf{V} - \mathtt{tr}(\mathcal{P}_2), \\ \frac{1}{\sigma_0^2}\mathbf{V}'\mathcal{P}_3\mathbf{V} - \mathtt{tr}(\mathcal{P}_3), \end{cases} \tag{3.17}$$

and its asymptotic normality is proved by the central limit theorem (CLT) for linear-quadratic (LQ) forms of Kelejian and Prucha (2001). This together with the proper asymptotic behavior of the 'Hessian' matrix, $\frac{\partial}{\partial\theta'}S^*_{nT}(\theta)$ (given in (B.4), Appendix B), lead to the following theorem.

Theorem 3.2. *Under Assumptions A-G, we have, as $nT \to \infty$,*

$$\sqrt{N_1}(\hat{\theta}^*_{nT} - \theta_0) \xrightarrow{D} N\Big(0, \lim_{N_1\to\infty} \Sigma^{*-1}_{nT}(\theta_0)\Gamma^*_{nT}(\theta_0)\Sigma^{*\prime-1}_{nT}(\theta_0)\Big),$$

where $\Sigma^*_{nT}(\theta_0) = -\frac{1}{N_1}\mathrm{E}[\frac{\partial}{\partial\theta'}S^*_{nT}(\theta_0)]$ and $\Gamma^*_{nT}(\theta_0) = \frac{1}{N_1}\mathrm{Var}[S^*_{nT}(\theta_0)]$, both assumed to exist and $\Sigma^*_{nT}(\theta_0)$ assumed to be positive definite for sufficiently large N_1.

3.2.6 Inference based on AQS estimation

To conduct inferences for θ based on the proposed AQS estimators, consistent estimates of $\Sigma^*_{nT}(\theta_0)$ and $\Gamma^*_{nT}(\theta_0)$ are needed. The analytical expression of $\Sigma^*_{nT}(\theta)$ can easily be obtained from the Hessian matrix $\frac{\partial}{\partial\theta'}S^*_{nT}(\theta)$ that is given in (B.4). We note that it depends only on the common parameters θ. Therefore, a simple plug-in estimator $\Sigma^*_{nT}(\hat{\theta}^*_{nT})$ can be used to consistently estimate $\Sigma^*_{nT}(\theta_0)$. Alternatively, a simpler sample analogue of $\Sigma^*_{nT}(\theta)$ also provides a consistent estimator:

$$\widehat{\Sigma}^*_{nT} = -\frac{1}{N_1}\frac{\partial}{\partial\theta'}S^*_{nT}(\theta)\Big|_{\theta=\hat{\theta}^*_{nT}}. \tag{3.18}$$

The consistency of $\Sigma^*_{nT}(\hat{\theta}^*_{nT})$ or $\widehat{\Sigma}^*_{nT}$ is proved in the proof of Theorem 3.2.

As for the estimation of $\Gamma^*_{nT}(\theta_0)$, we first derive all the elements of it using

Lemma A.5:

$$
\begin{aligned}
N_1\Gamma^*_{\beta\theta} &= \left[\tfrac{1}{\sigma_0^2}\mathbb{Z}'\mathbb{Z},\ \ \tfrac{\kappa_3}{2\sigma_0^3}\mathbb{Z}'q,\ \ \tfrac{\kappa_3}{\sigma_0}\mathbb{Z}'p_2 + \tfrac{1}{\sigma_0^2}\mathbb{Z}'\mathcal{P}_2\mathbf{B}_{nT}\eta,\ \ \tfrac{\kappa_3}{2\sigma_0}\mathbb{Z}'p_3\right],\\
N_1\Gamma^*_{\sigma^2\sigma^2} &= \tfrac{1}{4\sigma_0^4}(2N_1 + \kappa_4 q'q),\\
N_1\Gamma^*_{\sigma^2\lambda} &= \tfrac{\kappa_3}{2\sigma_0^3}q'\mathcal{P}_2\mathbf{B}_{nT}\eta + \tfrac{1}{2\sigma_0^2}[2\mathtt{tr}(\mathcal{P}_2\mathbb{Q}_{\mathbb{D}}) + \kappa_4 q'p_2],\\
N_1\Gamma^*_{\sigma^2\rho} &= \tfrac{1}{2\sigma_0^2}[2\mathtt{tr}(\mathcal{P}_3\mathbb{Q}_{\mathbb{D}}) + \kappa_4 q'p_3],\\
N_1\Gamma^*_{\lambda\lambda} &= \tfrac{1}{\sigma_0^2}\eta'\mathbf{B}'_{nT}\mathcal{P}'_2\mathcal{P}_2\mathbf{B}_{nT}\eta + \tfrac{2\kappa_3}{\sigma_0}p'_2\mathcal{P}_2\mathbf{B}_{nT}\eta + \mathtt{tr}(\mathcal{P}_2\mathcal{P}_2^\circ) + \kappa_4 p'_2p_2,\\
N_1\Gamma^*_{\lambda\rho} &= \mathtt{tr}(\mathcal{P}_3\mathcal{P}_2^\circ) + \kappa_4 p'_2p_3 + \tfrac{\kappa_3}{\sigma_0}p'_3\mathcal{P}_2\mathbf{B}_{nT}\eta,\\
N_1\Gamma^*_{\rho\rho} &= \mathtt{tr}(\mathcal{P}_3\mathcal{P}_3^\circ) + \kappa_4 p'_3p_3,
\end{aligned}
\tag{3.19}
$$

where $p_r = \mathtt{diagv}(\mathcal{P}_r), r = 2, 3$, and $q = \mathtt{diagv}(\mathbb{Q}_{\mathbb{D}})$. From the above expressions, we see that $\Gamma^*_{nT}(\theta_0)$ contains not only the common parameters θ, but also the incidental parameters ϕ embedded in η, and the skewness κ_3 and the excess kurtosis κ_4 of the idiosyncratic errors. Thus, the common plug-in approach may not provide a valid estimate. To be more specific, let $\Gamma^*_{nT}(\hat{\theta}^*_{nT}) = \Gamma^*_{nT}(\theta)|_{(\theta=\hat{\theta}^*_{nT},\phi=\hat{\phi}^*_{nT},\kappa_3=\hat{\kappa}_{3,nT},\kappa_4=\hat{\kappa}_{4,nT})}$ be the plug-in estimator, where $\hat{\phi}^*_{nT}$ is the AQS estimator of ϕ, obtained through (3.8), i.e., $\hat{\phi}^*_{nT} = \hat{\phi}_{nT}(\hat{\beta}^*_{nT}, \hat{\delta}^*_{nT})$, and $\hat{\kappa}_{3,nT}$ and $\hat{\kappa}_{4,nT}$ are consistent estimators of κ_3 and κ_4. When s_g, $\forall g$, and T are large at the same time, $\Gamma^*_{nT}(\hat{\theta}^*_{nT})$ would be consistent as $\hat{\phi}^*_{nT}$ is. However, when group sizes of some group g is not large, $\hat{\gamma}^*_{gt,nT}$ (component of $\hat{\phi}^*_{nT}$) is not consistent for each t. When T is not large, then $\hat{\phi}^*_{nT}$ (its component $\hat{\mu}^*_{nT}$) is also not consistent. Plugging $\hat{\phi}^*_{nT}$ into $\Gamma^*_{nT}(\theta)$ will induce a bias (inconsistency), and a bias correction is necessary.

From the expression of $\Gamma^*_{nT}(\theta_0)$ given above, we see that only the λ-components involve ϕ through η, which may not be consistently estimated by the plug-in method. We can further show that the components of $\Gamma^*_{nT}(\theta_0)$ linear in ϕ can also be consistently estimated by the plug-in method. Therefore, the only term that cannot be consistently estimated by the plug-in method is $\frac{1}{\sigma_0^2}\eta'\mathbf{B}'_{nT}\mathcal{P}'_2\mathcal{P}_2\mathbf{B}_{nT}\eta$ associated with the λ-λ component of $\Gamma^*_{nT}(\theta_0)$. We have the following corollary. See its proof in Appendix B for details on these discussions.

Corollary 3.1. *Under the assumptions of Theorem 3.2, we have,*

$$\Gamma^*_{nT}(\hat{\theta}^*_{nT}) = \Gamma^*_{nT}(\theta_0) + \text{Bias}^*(\delta_0) + o_p(1),$$

where $\text{Bias}^*(\delta_0)$ *is a* $(k+3)\times(k+3)$ *matrix having zero entries everywhere except the* λ-λ *entry, which takes the form* $\frac{1}{N_1}\mathtt{tr}(\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}})$.

The result of Corollary 3.1 leads immediately a general consistent estimator of $\Gamma^*_{nT}(\theta_0)$:

$$\widehat{\Gamma}^*_{nT} = \Gamma^*_{nT}(\hat{\theta}^*_{nT}) - \text{Bias}^*(\hat{\delta}^*_{nT}). \tag{3.20}$$

Then, it is only left to find consistent estimators for κ_3 and κ_4. Since we cannot 'consistently' estimate $\mathbf{V} = \mathbf{B}_{nT}(\mathbf{A}_{nT}\mathbf{Y} - \eta)$ due to the incidental parameters problem, we start from $\tilde{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$, which can be 'consistently' estimated by $\hat{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}(\hat{\rho}^*_{nT})\mathbf{B}_{nT}(\hat{\rho}^*_{nT})[\mathbf{A}_{nT}(\hat{\lambda}^*_{nT})\mathbf{Y} - \mathbf{Z}\hat{\beta}^*_{nT}]$. Let q_{jk} be the (j,k)th element of $\mathbb{Q}_{\mathbb{D}}$. Denote the elements of $\mathbf{V}$ by v_j, and the elements of $\tilde{\mathbf{V}}$ by $\tilde{v}_j, j = 1,\dots,N$, where j is the combined index for $i = 1,\dots,s_g$, $g = 1,\dots,G$ and $t = 1,\dots,T$. Then, $\tilde{v}_j = q_{j1}v_1 + q_{j2}v_2 + \cdots + q_{jN}v_N$, and thus,

$$\text{E}(\tilde{v}_j^3) = \textstyle\sum_{k=1}^N q_{jk}^3\text{E}(v_k^3) = \sigma^3\kappa_3\sum_{k=1}^N q_{jk}^3,\ j = 1,\dots,N.$$

Summing $\text{E}(\tilde{v}_j^3)$ over j, we obtain $\kappa_3 = \big(\sum_{j=1}^N \text{E}(\tilde{v}_j^3)\big)\big(\sigma^3\sum_{j=1}^N\sum_{k=1}^N q_{jk}^3\big)^{-1}$, and its sample analogue gives a consistent estimator of κ_3:

$$\hat{\kappa}_{3,nT} = \frac{\sum_{j=1}^N \hat{v}_j^3}{\hat{\sigma}^{*3}_{nT}\sum_{j=1}^N\sum_{k=1}^N \hat{q}_{jk}^3}. \tag{3.21}$$

where $\hat{v}_j$ is the jth element of $\hat{\mathbf{V}}(\hat{\beta}^*_{nT}, \hat{\lambda}^*_{nT})$ and $\hat{q}_{jk}$ is the (j,k)th element of $\mathbb{Q}_{\mathbb{D}}(\hat{\rho}^*_{nT})$. Similarly,

$$\begin{aligned}\text{E}(\tilde{v}_j^4) &= \textstyle\sum_{k=1}^N q_{jk}^4\text{E}(v_k^4) + 3\sigma^4\sum_{k=1}^N\sum_{l=1}^N q_{jk}^2q_{jl}^2 - 3\sigma^4\sum_{k=1}^N q_{jk}^4\\ &= \textstyle\sum_{k=1}^N q_{jk}^4\kappa_4\sigma^4 + 3\sigma^4\sum_{k=1}^N\sum_{l=1}^N q_{jk}^2q_{jl}^2,\end{aligned}$$

which gives $\kappa_4 = \big(\sum_{j=1}^N \text{E}(\tilde{v}_j^4) - 3\sigma^4\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N q_{jk}^2q_{jl}^2\big)\big(\sigma^4\sum_{j=1}^N\sum_{k=1}^N q_{jk}^4\big)^{-1}$ by summing $\text{E}(\tilde{v}_j^4)$ over j. Hence, a consistent estimator for κ_4 is

$$\hat{\kappa}_{4,nT} = \frac{\sum_{j=1}^N \hat{v}_j^4 - 3\hat{\sigma}^{*4}_{nT}\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N \hat{q}_{jk}^2\hat{q}_{jl}^2}{\hat{\sigma}^{*4}_{nT}\sum_{j=1}^N\sum_{k=1}^N \hat{q}_{jk}^4}. \tag{3.22}$$

Corollary 3.2. *Under Assumptions A-G, we have, as* $N_1 \to \infty$,

(i) $\hat{\kappa}_{3,nT} \xrightarrow{p} \kappa_{30}$ *and* $\hat{\kappa}_{4,nT} \xrightarrow{p} \kappa_{40}$; *(ii)* $\widehat{\Sigma}^*_{nT} - \Sigma^*_{nT}(\theta_0) \xrightarrow{p} 0$ *and* $\widehat{\Gamma}^*_{nT} - \Gamma^*_{nT}(\theta_0) \xrightarrow{p} 0$;

and therefore $\widehat{\Sigma}_{nT}^{*-1}\widehat{\Gamma}_{nT}^{*}\widehat{\Sigma}_{nT}^{*\prime-1} - \Sigma_{nT}^{*-1}(\theta_0)\Gamma_{nT}^{*}(\theta_0)\Sigma_{nT}^{*\prime-1}(\theta_0) \xrightarrow{p} 0.$

3.3 Monte Carlo Study

Extensive Monte Carlo experiments are carried out to investigate the finite sample performance of the proposed AQS estimators and the corresponding standard error estimators for the SPD models with time-varying network structures. In order to see the effectiveness of the adjustments on the concentrated quasi scores in controlling the effects of estimating the fixed effects, we also include the direct QML estimators in the Monte Carlo study. Both time-invariant and time-varying grouping cases are considered. Hence, we have the following two data generating processes:

$$Y_{gt} = \lambda W_{gt} Y_{gt} + Z_{gt}\beta + \mu_g + \gamma_{gt} l_{s_g} + U_{gt}, \quad U_{gt} = \rho W_{gt} U_{gt} + V_{gt}, \tag{3.23}$$

$$Y_{gt} = \lambda W_{gt} Y_{gt} + Z_{gt}\beta + \mu_g^{(t)} + \gamma_{gt} l_{s_{gt}} + U_{gt}, \quad U_{gt} = \rho W_{gt} U_{gt} + V_{gt}, \tag{3.24}$$

for $g = 1, \ldots, G$ and $t = 1, \ldots, T$, where the first process represents the time-invariant grouping model and the second the time-varying grouping model. We choose $T = 5$ or 10 for both models. For the first model, we consider four combinations with different numbers of groups G and group sizes $\{s_g\}$. The first case contains 5 groups with equal group sizes of $s_g = 10$. For comparison, the second case contains 10 groups with equal group sizes of $s_g = 5$. To study the effect of group sizes, we also consider 5 and 10 groups with equal group sizes of $s_g = 20$ and $s_g = 10$, respectively. Therefore, the first two cases have $n = 50$ and the other two have $n = 100$. For the second model, we let n units randomly be separated into G groups at each time, where $n = 50$ or 100, and $G = 5$ or 10.

For each group at period t, the sociomatrix W_{gt} is generated following Liu and Lee (2010). First, for the ith row of W_{gt}, we generate an integer k_{igt} uniformly at random from the set of integers [0,1,2,3]. Then we set the $(i+1)$th,..., $(i+k_{igt})$th elements of the ith row of W_{gt} to be ones and the rest elements in that row to be zeros, if $i + k_{igt} \leq s_g$; otherwise the entries of ones will be wrapped around such that the first $(i + k_{igt} - m)$ entries of the ith row will be ones. We choose $\beta_1 = 1$, $\beta_2 = 0.5$, $\sigma^2 = 1$, $\lambda = 0.2$ and

$\rho = 0.2$. Generate $X'_{gt}s$ independently from $N(0, I_{s_g})$, and set the individual effects as $\frac{1}{T}\Sigma_{t=1}^{T} X_{gt} + e$, where $e \sim N(0, I_{s_g})$. The group-time fixed effects γ_{gt} are generated from $N(0,1)$. The error (v_{igt}) distributions can be (*i*) normal, (*ii*) normal mixture (10% $N(0, 4^2)$ and 90% $N(0,1)$), or (*iii*) chi-square with 3 degrees of freedom.[2] Monte Carlo (empirical) means and standard deviations (shown in the brackets) are reported for QMLE and AQSE. Further, empirical averages of the standard error estimates (shown in the square brackets) are also reported for AQSE, based on the robust VC matrix estimates, $\widehat{\Sigma}_{nT}^{*-1}\widehat{\Gamma}_{nT}^{*}\widehat{\Sigma}_{nT}^{*\prime-1}$. The number of Monte Carlo runs is 1000.

Tables 3.1a and 3.1b report Monte Carlo results for the time-invariant grouping model, for $T = 5$ and 10, respectively. The results show an excellent finite performance of the proposed AQS estimators, as well as their standard error estimators. The proposed AQS method performs uniformly much better than the QML method in the point estimation of σ^2, λ, and ρ, irrespective of the choices of G, $\{s_g\}$ and T. By comparing the empirical sds of two types of estimators, we see that AQSE is almost as efficient as the QMLE. Our AQS estimators exhibit a good performance even when the sample size is as small as $n = 50$ and $T = 5$, and improve on average when the sample expands, regardless of the error distributions. The $\sqrt{N_1}$-consistency of the AQSEs is clearly demonstrated by the Monte Carlo sds. Moreover, the robust estimates of standard errors $\hat{sd}$'s are on average very close to the corresponding Monte Carlo standard errors.

Tables 3.2a and 3.2b report Monte Carlo results for the time-varying grouping model, for $T = 5$ and 10, respectively. The results again show an excellent finite sample performance of the proposed estimation. The corresponding standard error estimates also perform very well. In contrast, the QMLE typical provide worse estimates than the AQSE. We see the QMLEs of ρ and σ^2 are still far away from their true values even for the largest sample size.

[2] In the cases (*ii*) and (*iii*), the generated errors are standardized to have mean zero and variance σ^2.

Table 3.1a. Empirical mean$(sd)[\hat{sd}]$ of the estimators for SPD model with time-invariant grouping, $(\beta_1, \beta_2, \lambda, \rho, \sigma^2) = (1, 0.5, 0.2, 0.2, 1)$, $T = 5$.

n	G		Normal Errors		Normal Mixture		Chi-Square	
			QMLE	AQSE	QMLE	AQSE	QMLE	AQSE
50	5	β_1	0.9992(0.088)	0.9996(0.085)[0.084]	1.0004(0.087)	1.0004(0.086)[0.084]	0.9997(0.085)	1.0000(0.084)[0.085]
		β_2	0.5024(0.078)	0.5015(0.081)[0.080]	0.5005(0.081)	0.4999(0.083)[0.080]	0.5015(0.079)	0.4997(0.082)[0.080]
		λ	0.2004(0.033)	0.1996(0.042)[0.040]	0.1996(0.033)	0.1996(0.043)[0.040]	0.2006(0.034)	0.2011(0.044)[0.041]
		ρ	-0.0076(0.096)	0.1897(0.097)[0.102]	-0.0078(0.107)	0.1871(0.106)[0.106]	-0.0114(0.104)	0.1814(0.105)[0.103]
		σ^2	0.6803(0.073)	0.9748(0.105)[0.104]	0.6741(0.148)	0.9649(0.210)[0.200]	0.6785(0.111)	0.9719(0.158)[0.150]
	10	β_1	0.9984(0.083)	0.9991(0.081)[0.081]	1.0014(0.083)	1.0023(0.081)[0.081]	1.0028(0.082)	1.0039(0.081)[0.081]
		β_2	0.4989(0.082)	0.4993(0.084)[0.082]	0.5011(0.085)	0.5004(0.085)[0.082]	0.4998(0.084)	0.4996(0.084)[0.083]
		λ	0.1998(0.031)	0.1987(0.038)[0.039]	0.2010(0.034)	0.2011(0.042)[0.038]	0.2009(0.031)	0.2011(0.040)[0.040]
		ρ	-0.0065(0.095)	0.1914(0.093)[0.102]	-0.0089(0.105)	0.1860(0.104)[0.104]	-0.0100(0.104)	0.1858(0.102)[0.105]
		σ^2	0.6807(0.075)	0.9755(0.107)[0.104]	0.6786(0.157)	0.9716(0.224)[0.202]	0.6812(0.114)	0.9760(0.163)[0.150]
100	5	β_1	0.9973(0.052)	0.9968(0.052)[0.050]	0.9980(0.051)	0.9979(0.051)[0.050]	1.0020(0.051)	1.0019(0.051)[0.050]
		β_2	0.4962(0.052)	0.4980(0.053)[0.053]	0.4992(0.055)	0.5010(0.055)[0.054]	0.4985(0.053)	0.5004(0.054)[0.053]
		λ	0.2016(0.023)	0.1997(0.025)[0.025]	0.2014(0.025)	0.1997(0.026)[0.025]	0.2020(0.024)	0.2002(0.025)[0.025]
		ρ	0.1474(0.065)	0.1967(0.053)[0.055]	0.1441(0.070)	0.1941(0.058)[0.057]	0.1487(0.069)	0.1978(0.057)[0.056]
		σ^2	0.7508(0.055)	0.9906(0.073)[0.072]	0.7506(0.117)	0.9904(0.155)[0.149]	0.7495(0.085)	0.9888(0.112)[0.110]
	10	β_1	1.0009(0.050)	1.0008(0.050)[0.050]	0.9972(0.050)	0.9970(0.049)[0.050]	1.0019(0.051)	1.0017(0.051)[0.050]
		β_2	0.4993(0.054)	0.5011(0.054)[0.054]	0.4974(0.055)	0.4994(0.055)[0.055]	0.4977(0.053)	0.5000(0.053)[0.054]
		λ	0.2009(0.024)	0.1991(0.026)[0.025]	0.2008(0.024)	0.1987(0.026)[0.025]	0.2015(0.024)	0.1994(0.025)[0.025]
		ρ	0.1462(0.068)	0.1953(0.056)[0.055]	0.1492(0.070)	0.1985(0.058)[0.057]	0.1494(0.067)	0.1989(0.054)[0.056]
		σ^2	0.7497(0.056)	0.9892(0.074)[0.071]	0.7498(0.114)	0.9892(0.150)[0.150]	0.7527(0.086)	0.9931(0.113)[0.110]
200	5	β_1	0.9999(0.039)	0.9996(0.039)[0.039]	1.0004(0.038)	1.0002(0.038)[0.039]	1.0002(0.041)	0.9999(0.041)[0.039]
		β_2	0.4953(0.037)	0.4979(0.038)[0.039]	0.4981(0.039)	0.5007(0.040)[0.039]	0.4964(0.039)	0.4989(0.039)[0.039]
		λ	0.2025(0.016)	0.2003(0.018)[0.018]	0.2029(0.017)	0.2006(0.018)[0.018]	0.2011(0.018)	0.1988(0.019)[0.018]
		ρ	0.1490(0.048)	0.1993(0.039)[0.040]	0.1484(0.047)	0.1991(0.039)[0.040]	0.1490(0.048)	0.1992(0.040)[0.040]
		σ^2	0.7533(0.037)	0.9939(0.049)[0.051]	0.7527(0.084)	0.9931(0.111)[0.108]	0.7543(0.063)	0.9952(0.083)[0.079]
	10	β_1	1.0025(0.038)	1.0023(0.038)[0.036]	1.0026(0.036)	1.0023(0.036)[0.036]	1.0007(0.036)	1.0004(0.035)[0.036]
		β_2	0.4972(0.037)	0.4996(0.037)[0.037]	0.4970(0.036)	0.4994(0.036)[0.037]	0.4995(0.036)	0.5018(0.037)[0.037]
		λ	0.2031(0.016)	0.2011(0.017)[0.017]	0.2022(0.016)	0.2000(0.017)[0.017]	0.2016(0.017)	0.1996(0.018)[0.017]
		ρ	0.1477(0.046)	0.1982(0.039)[0.039]	0.1476(0.047)	0.1984(0.040)[0.040]	0.1481(0.049)	0.1982(0.041)[0.039]
		σ^2	0.7532(0.037)	0.9937(0.049)[0.051]	0.7535(0.085)	0.9942(0.112)[0.108]	0.7531(0.063)	0.9936(0.083)[0.079]

Table 3.1b. Empirical mean(sd)[$\hat{sd}$] of the estimators for SPD model with time-invariant grouping, $(\beta_1, \beta_2, \lambda, \rho, \sigma^2) = (1, 0.5, 0.2, 0.2, 1)$, $T = 10$.

n	G		Normal Errors		Normal Mixture		Chi-Square	
			QMLE	AQSE	QMLE	AQSE	QMLE	AQSE
50	5	β_1	0.9990(0.051)	0.9993(0.050)[0.049]	0.9994(0.050)	0.9994(0.049)[0.048]	1.0010(0.048)	1.0011(0.048)[0.049]
		β_2	0.5023(0.048)	0.5017(0.048)[0.049]	0.5016(0.049)	0.5008(0.050)[0.049]	0.5016(0.049)	0.5000(0.049)[0.049]
		λ	0.1983(0.021)	0.1984(0.026)[0.024]	0.1986(0.020)	0.1990(0.026)[0.025]	0.1996(0.020)	0.2004(0.024)[0.024]
		ρ	0.0080(0.056)	0.1963(0.066)[0.064]	0.0044(0.059)	0.1934(0.071)[0.071]	0.0045(0.058)	0.1936(0.066)[0.068]
		σ^2	0.7740(0.055)	0.9871(0.070)[0.071]	0.7669(0.116)	0.9781(0.148)[0.147]	0.7734(0.091)	0.9867(0.115)[0.110]
	10	β_1	1.0012(0.051)	1.0014(0.049)[0.050]	0.9999(0.051)	1.0002(0.049)[0.050]	0.9985(0.051)	0.9990(0.050)[0.049]
		β_2	0.5021(0.051)	0.5011(0.052)[0.050]	0.5032(0.049)	0.5023(0.050)[0.051]	0.5016(0.049)	0.5010(0.050)[0.051]
		λ	0.1979(0.021)	0.1980(0.025)[0.024]	0.1985(0.020)	0.1985(0.025)[0.025]	0.1982(0.021)	0.1989(0.026)[0.025]
		ρ	0.0102(0.056)	0.1993(0.064)[0.065]	0.0085(0.059)	0.1962(0.070)[0.069]	0.0040(0.060)	0.1924(0.070)[0.068]
		σ^2	0.7750(0.056)	0.9883(0.071)[0.071]	0.7762(0.119)	0.9895(0.151)[0.149]	0.7724(0.091)	0.9853(0.116)[0.109]
100	5	β_1	1.0003(0.035)	1.0000(0.035)[0.034]	0.9997(0.034)	0.9996(0.034)[0.034]	0.9996(0.035)	0.9994(0.034)[0.034]
		β_2	0.4983(0.035)	0.5002(0.035)[0.035]	0.4989(0.033)	0.5009(0.034)[0.035]	0.4990(0.034)	0.5009(0.035)[0.035]
		λ	0.2013(0.016)	0.1993(0.017)[0.017]	0.2007(0.016)	0.1989(0.017)[0.017]	0.2010(0.016)	0.1992(0.017)[0.017]
		ρ	0.1446(0.041)	0.2003(0.037)[0.036]	0.1439(0.041)	0.1992(0.037)[0.037]	0.1445(0.040)	0.2000(0.037)[0.036]
		σ^2	0.8463(0.042)	0.9934(0.049)[0.048]	0.8491(0.099)	0.9966(0.116)[0.108]	0.8440(0.068)	0.9907(0.080)[0.077]
	10	β_1	0.9999(0.034)	0.9998(0.034)[0.034]	0.9978(0.035)	0.9977(0.035)[0.034]	1.0002(0.033)	1.0002(0.033)[0.034]
		β_2	0.4999(0.033)	0.5016(0.033)[0.034]	0.4967(0.034)	0.4986(0.034)[0.035]	0.4971(0.033)	0.4987(0.034)[0.034]
		λ	0.2011(0.016)	0.1996(0.017)[0.016]	0.2016(0.015)	0.2001(0.017)[0.016]	0.2021(0.015)	0.2006(0.016)[0.016]
		ρ	0.1426(0.038)	0.1979(0.034)[0.035]	0.1416(0.040)	0.1971(0.036)[0.036]	0.1406(0.040)	0.1967(0.037)[0.036]
		σ^2	0.8457(0.042)	0.9928(0.050)[0.048]	0.8500(0.096)	0.9978(0.112)[0.107]	0.8467(0.069)	0.9939(0.081)[0.077]
200	5	β_1	1.0014(0.023)	1.0012(0.023)[0.024]	0.9997(0.025)	0.9995(0.025)[0.024]	1.0011(0.025)	1.0008(0.025)[0.024]
		β_2	0.4983(0.025)	0.5002(0.025)[0.024]	0.4973(0.024)	0.4990(0.024)[0.024]	0.4993(0.024)	0.5010(0.024)[0.024]
		λ	0.2016(0.011)	0.1999(0.012)[0.011]	0.2014(0.011)	0.1997(0.012)[0.012]	0.2010(0.011)	0.1993(0.012)[0.012]
		ρ	0.1437(0.028)	0.1992(0.025)[0.025]	0.1431(0.028)	0.1988(0.025)[0.026]	0.1447(0.028)	0.2000(0.025)[0.025]
		σ^2	0.8492(0.028)	0.9969(0.033)[0.034]	0.8476(0.065)	0.9950(0.076)[0.076]	0.8480(0.048)	0.9954(0.056)[0.055]
	10	β_1	1.0002(0.023)	0.9999(0.023)[0.024]	0.9998(0.024)	0.9995(0.024)[0.024]	0.9995(0.025)	0.9993(0.025)[0.024]
		β_2	0.4980(0.025)	0.4998(0.025)[0.024]	0.4986(0.024)	0.5004(0.024)[0.024]	0.4980(0.023)	0.4998(0.023)[0.024]
		λ	0.2017(0.011)	0.1999(0.012)[0.011]	0.2017(0.010)	0.1999(0.011)[0.012]	0.2018(0.011)	0.2001(0.011)[0.012]
		ρ	0.1428(0.027)	0.1987(0.025)[0.025]	0.1458(0.027)	0.2012(0.025)[0.025]	0.1425(0.028)	0.1982(0.025)[0.025]
		σ^2	0.8483(0.029)	0.9958(0.034)[0.034]	0.8479(0.066)	0.9953(0.077)[0.077]	0.8499(0.048)	0.9977(0.056)[0.055]

Table 3.2a. Empirical mean(sd)[$\hat{sd}$] of the estimators for SPD model with time-varying grouping, $(\beta_1, \beta_2, \lambda, \rho, \sigma^2) = (1, 0.5, 0.2, 0.2, 1)$, $T = 5$.

n	G		Normal Errors		Normal Mixture		Chi-Square	
			QMLE	AQSE	QMLE	AQSE	QMLE	AQSE
50	5	β_1	0.9984(0.077)	0.9973(0.075)[0.072]	1.0027(0.071)	1.0018(0.069)[0.072]	1.0006(0.073)	1.0004(0.071)[0.072]
		β_2	0.4963(0.079)	0.4989(0.080)[0.076]	0.4960(0.076)	0.4983(0.077)[0.075]	0.4990(0.072)	0.5021(0.073)[0.077]
		λ	0.2031(0.033)	0.1999(0.040)[0.037]	0.2035(0.034)	0.2003(0.040)[0.037]	0.2025(0.031)	0.1992(0.037)[0.038]
		ρ	-0.0198(0.113)	0.1836(0.098)[0.095]	-0.0131(0.118)	0.1832(0.101)[0.100]	-0.0155(0.109)	0.1848(0.091)[0.102]
		σ^2	0.6631(0.074)	0.9697(0.106)[0.105]	0.6575(0.162)	0.9598(0.234)[0.198]	0.6676(0.116)	0.9762(0.169)[0.154]
	10	β_1	0.9979(0.098)	1.0042(0.085)[0.086]	0.9991(0.097)	1.0019(0.086)[0.086]	0.9889(0.095)	1.0005(0.085)[0.086]
		β_2	0.5192(0.099)	0.4959(0.091)[0.089]	0.5157(0.101)	0.4868(0.091)[0.089]	0.5140(0.103)	0.4914(0.091)[0.089]
		λ	0.1810(0.039)	0.2019(0.049)[0.044]	0.1831(0.040)	0.2088(0.051)[0.046]	0.1805(0.039)	0.2058(0.050)[0.045]
		ρ	-0.3467(0.113)	0.1598(0.122)[0.124]	-0.3625(0.125)	0.1363(0.144)[0.137]	-0.3671(0.112)	0.1458(0.133)[0.129]
		σ^2	0.4627(0.070)	0.9427(0.125)[0.123]	0.4593(0.121)	0.9334(0.232)[0.209]	0.4615(0.090)	0.9446(0.176)[0.166]
100	5	β_1	1.0021(0.048)	1.0018(0.049)[0.049]	1.0001(0.048)	0.9998(0.048)[0.049]	0.9994(0.052)	0.9992(0.052)[0.049]
		β_2	0.5011(0.052)	0.5031(0.053)[0.053]	0.4977(0.052)	0.5001(0.053)[0.052]	0.4978(0.053)	0.5001(0.053)[0.052]
		λ	0.2013(0.025)	0.1994(0.026)[0.025]	0.2007(0.023)	0.1985(0.025)[0.025]	0.2025(0.025)	0.2005(0.027)[0.025]
		ρ	0.1468(0.068)	0.1969(0.055)[0.056]	0.1494(0.066)	0.1995(0.054)[0.058]	0.1444(0.072)	0.1948(0.058)[0.057]
		σ^2	0.7503(0.056)	0.9899(0.074)[0.071]	0.7436(0.115)	0.9810(0.152)[0.148]	0.7493(0.087)	0.9887(0.115)[0.109]
	10	β_1	0.9997(0.053)	0.9990(0.052)[0.054]	1.0013(0.054)	0.9997(0.054)[0.054]	0.9995(0.054)	0.9985(0.053)[0.054]
		β_2	0.4945(0.054)	0.4984(0.055)[0.055]	0.4948(0.054)	0.4993(0.054)[0.055]	0.4931(0.055)	0.4981(0.055)[0.056]
		λ	0.2030(0.023)	0.1985(0.028)[0.026]	0.2042(0.022)	0.1995(0.027)[0.027]	0.2037(0.022)	0.1992(0.027)[0.027]
		ρ	-0.0126(0.077)	0.1935(0.070)[0.069]	-0.0105(0.083)	0.1932(0.074)[0.075]	-0.0111(0.077)	0.1917(0.071)[0.071]
		σ^2	0.6703(0.052)	0.9837(0.074)[0.076]	0.6662(0.108)	0.9770(0.157)[0.149]	0.6713(0.081)	0.9847(0.118)[0.111]
200	5	β_1	1.0022(0.036)	1.0018(0.036)[0.036]	0.9995(0.035)	0.9992(0.035)[0.036]	0.9994(0.036)	0.9992(0.036)[0.036]
		β_2	0.4962(0.036)	0.4984(0.037)[0.037]	0.4978(0.036)	0.5002(0.037)[0.037]	0.4988(0.036)	0.5013(0.036)[0.037]
		λ	0.2020(0.017)	0.1997(0.018)[0.018]	0.2016(0.017)	0.1993(0.019)[0.018]	0.2018(0.017)	0.1993(0.019)[0.018]
		ρ	0.1458(0.046)	0.1970(0.039)[0.039]	0.1481(0.048)	0.1989(0.041)[0.040]	0.1505(0.047)	0.2007(0.039)[0.040]
		σ^2	0.7543(0.038)	0.9952(0.050)[0.051]	0.7549(0.082)	0.9960(0.108)[0.108]	0.7531(0.061)	0.9935(0.080)[0.080]
	10	β_1	1.0026(0.035)	1.0023(0.035)[0.036]	0.9985(0.035)	0.9983(0.035)[0.036]	0.9999(0.035)	0.9996(0.035)[0.036]
		β_2	0.4984(0.036)	0.5007(0.037)[0.036]	0.4973(0.037)	0.4996(0.038)[0.036]	0.4984(0.037)	0.5009(0.038)[0.036]
		λ	0.2017(0.017)	0.1995(0.018)[0.017]	0.2016(0.016)	0.1995(0.017)[0.017]	0.2012(0.016)	0.1990(0.017)[0.017]
		ρ	0.1456(0.046)	0.1969(0.038)[0.039]	0.1476(0.048)	0.1981(0.040)[0.040]	0.1509(0.048)	0.2011(0.040)[0.039]
		σ^2	0.7545(0.039)	0.9954(0.051)[0.051]	0.7518(0.081)	0.9918(0.107)[0.108]	0.7508(0.063)	0.9905(0.083)[0.079]

Table 3.2b. Empirical mean(sd)[$\hat{sd}$] of the estimators for SPD model with time-varying grouping, $(\beta_1, \beta_2, \lambda, \rho, \sigma^2) = (1, 0.5, 0.2, 0.2, 1)$, $T = 10$.

n	G		Normal Errors		Normal Mixture		Chi-Square	
			QMLE	AQSE	QMLE	AQSE	QMLE	AQSE
50	5	β_1	1.0021(0.049)	1.0010(0.047)[0.049]	0.9996(0.048)	0.9985(0.047)[0.049]	0.9986(0.051)	0.9980(0.050)[0.049]
		β_2	0.4996(0.053)	0.5010(0.054)[0.052]	0.5024(0.052)	0.5031(0.052)[0.052]	0.4979(0.053)	0.5000(0.053)[0.052]
		λ	0.2020(0.021)	0.1994(0.027)[0.026]	0.1997(0.022)	0.1976(0.027)[0.026]	0.2017(0.022)	0.1990(0.027)[0.026]
		ρ	0.0190(0.061)	0.1969(0.063)[0.064]	0.0204(0.065)	0.1970(0.065)[0.068]	0.0216(0.061)	0.1984(0.063)[0.064]
		σ^2	0.7709(0.054)	0.9872(0.069)[0.071]	0.7690(0.121)	0.9849(0.155)[0.148]	0.7658(0.090)	0.9802(0.114)[0.108]
	10	β_1	0.9978(0.057)	1.0001(0.052)[0.054]	0.9958(0.060)	0.9986(0.054)[0.054]	0.9950(0.060)	1.0003(0.055)[0.055]
		β_2	0.5196(0.063)	0.5002(0.058)[0.057]	0.5216(0.063)	0.5004(0.059)[0.058]	0.5190(0.063)	0.5002(0.061)[0.057]
		λ	0.1857(0.023)	0.2018(0.030)[0.029]	0.1829(0.024)	0.2012(0.033)[0.030]	0.1825(0.023)	0.1998(0.032)[0.029]
		ρ	-0.2419(0.065)	0.1811(0.084)[0.084]	-0.2457(0.073)	0.1751(0.095)[0.095]	-0.2461(0.068)	0.1831(0.088)[0.086]
		σ^2	0.5818(0.054)	0.9772(0.085)[0.083]	0.5762(0.097)	0.9669(0.160)[0.155]	0.5817(0.077)	0.9800(0.123)[0.119]
100	5	β_1	0.9998(0.034)	0.9995(0.034)[0.033]	1.0015(0.033)	1.0014(0.033)[0.033]	1.0021(0.032)	1.0020(0.032)[0.033]
		β_2	0.4972(0.034)	0.4987(0.035)[0.034]	0.5003(0.033)	0.5023(0.034)[0.035]	0.4990(0.034)	0.5009(0.034)[0.034]
		λ	0.2015(0.016)	0.1998(0.017)[0.017]	0.2008(0.016)	0.1989(0.017)[0.017]	0.2017(0.015)	0.2001(0.017)[0.017]
		ρ	0.1425(0.040)	0.1980(0.037)[0.036]	0.1444(0.040)	0.2001(0.036)[0.037]	0.1428(0.042)	0.1981(0.039)[0.036]
		σ^2	0.8456(0.041)	0.9927(0.048)[0.048]	0.8472(0.094)	0.9945(0.111)[0.107]	0.8443(0.066)	0.9911(0.077)[0.077]
	10	β_1	0.9997(0.036)	0.9995(0.036)[0.036]	1.0016(0.037)	1.0007(0.036)[0.036]	0.9994(0.037)	0.9988(0.037)[0.036]
		β_2	0.4973(0.036)	0.5006(0.037)[0.036]	0.5002(0.036)	0.5024(0.037)[0.036]	0.4978(0.036)	0.5002(0.037)[0.036]
		λ	0.2018(0.015)	0.1989(0.019)[0.018]	0.2020(0.015)	0.1996(0.019)[0.019]	0.2015(0.016)	0.1991(0.019)[0.018]
		ρ	0.0236(0.041)	0.1997(0.044)[0.045]	0.0249(0.044)	0.1977(0.048)[0.048]	0.0242(0.043)	0.2000(0.046)[0.046]
		σ^2	0.7758(0.039)	0.9949(0.049)[0.051]	0.7739(0.091)	0.9918(0.117)[0.108]	0.7748(0.063)	0.9935(0.081)[0.079]
200	5	β_1	1.0004(0.024)	1.0002(0.024)[0.024]	1.0009(0.024)	1.0005(0.024)[0.024]	1.0017(0.025)	1.0014(0.025)[0.024]
		β_2	0.4973(0.023)	0.4992(0.023)[0.024]	0.4979(0.023)	0.4997(0.023)[0.024]	0.4980(0.024)	0.4998(0.025)[0.024]
		λ	0.2021(0.011)	0.2004(0.012)[0.012]	0.2024(0.011)	0.2006(0.012)[0.012]	0.2014(0.011)	0.1997(0.012)[0.012]
		ρ	0.1441(0.028)	0.1996(0.026)[0.025]	0.1423(0.027)	0.1983(0.025)[0.026]	0.1428(0.028)	0.1986(0.026)[0.026]
		σ^2	0.8489(0.030)	0.9964(0.035)[0.034]	0.8490(0.067)	0.9966(0.079)[0.077]	0.8505(0.048)	0.9984(0.056)[0.055]
	10	β_1	1.0005(0.024)	1.0002(0.024)[0.023]	1.0007(0.023)	1.0005(0.023)[0.023]	1.0004(0.023)	1.0001(0.023)[0.023]
		β_2	0.4989(0.022)	0.5007(0.022)[0.024]	0.4982(0.023)	0.5001(0.023)[0.024]	0.4975(0.024)	0.4992(0.024)[0.024]
		λ	0.2015(0.011)	0.1998(0.012)[0.012]	0.2017(0.011)	0.2000(0.012)[0.012]	0.2020(0.011)	0.2002(0.012)[0.012]
		ρ	0.1430(0.027)	0.1988(0.025)[0.025]	0.1432(0.028)	0.1991(0.026)[0.026]	0.1432(0.028)	0.1991(0.026)[0.025]
		σ^2	0.8479(0.029)	0.9953(0.034)[0.034]	0.8490(0.067)	0.9967(0.079)[0.077]	0.8482(0.048)	0.9957(0.057)[0.055]

3.4 Conclusion and Discussion

We consider estimation and inference for a spatial panel data (SPD) model with time-varying network structures, which allows for endogenous interaction, exogenous interactions, correlation of unobservables, and most importantly three-dimensional fixed effects. The time-varying network structures render the orthogonal transformation inapplicable, and thus an *adjusted quasi score* (AQS) is proposed, which adjusts the concentrated quasi scores (with the fixed effects being concentrated out) to remove the effects of estimating these *incidental parameters*. Although we focus on an empirically most meaningful specification of the three-way fixed effects, where group-specific effects are time-varying additive with time-invariant individual-specific effects, the proposed estimation strategy can be easily extended to handle all the other possible specifications. For the statistical inferences, the main difficulty lies with the fact that 'consistent' estimates of the idiosyncratic errors are unavailable due to the incidental parameters problem. A 'plug-in and then bias-correction' method is proposed to give consistent estimates of the standard errors of the AQS estimators.

The proposed methods are seen to be very general in handling the SPD models with multi-dimensional unobserved heterogeneity and the presence of generally time-varying spatial weight matrices without row-normalization. The current study also sheds light on an interesting but challenging extension: dynamic SPD models with time-varying network structures. Especially when T is fixed, the analysis will become much more complicated as adjustments to the concentrated quasi score functions are required to deal with the incidental parameters problems coming from both the initial condition and the concentration. Rigorous studies on this extension can only be done in future research.

Appendices

Appendices for Chapter 1

We collect all the technical proofs for the main results in Chapter 1 here. Three appendices are provided. Appendix A provides some basic lemmas that are used in the other appendices. Appendix B and Appendix C present proofs for results in Section 1.2 and Section 1.3, respectively.

Appendix A: Some basic lemmas

The following lemmas are essential to the proofs of the main results in this chapter.

Lemma A.1. *(Kelejian and Prucha, 1999; Lee, 2002): Let* $\{A_N\}$ *and* $\{B_N\}$ *be two sequences of* $N \times N$ *matrices that are uniformly bounded in both row and column sums. Let* C_N *be a sequence of conformable matrices whose elements are uniformly* $O(h_n^{-1})$*. Then,*

(i) the sequence $\{A_N B_N\}$ *are uniformly bounded in both row and column sums,*

(ii) the elements of A_N *are uniformly bounded and* $tr(A_N) = O(N)$*, and*

(iii) the elements of $A_N C_N$ *and* $C_N A_N$ *are uniformly* $O(h_n^{-1})$*.*

Lemma A.2. *(Lemma A.3, Lee, 2004): For* $\mathbf{W}$ *and* $\mathbf{A}_N(\lambda)$ *defined in Model* (1.2)*, if* $\|\mathbf{W}\|$ *and* $\|\mathbf{A}_N^{-1}\|$ *are uniformly bounded, where* $\|\cdot\|$ *is a matrix norm, then* $\|\mathbf{A}_N^{-1}(\lambda)\|$ *is uniformly bounded in a neighborhood of* λ_0*.*

Lemma A.3. *Under Assumptions C-E, we have*

(i) $\mathbb{Q}_{\mathbb{D}}(\rho)$ *is uniformly bounded in both row and column sums, uniformly in* $\rho \in \Delta_\rho$*;*

(ii) $\mathbb{Q}_{\mathbb{X}}(\rho)$ *is uniformly bounded in both row and column sums, uniformly in* $\rho \in \Delta_\rho$*.*

Proof of Lemma A.3: Proof is simpler using a $\mathbf{D}_\alpha^\star$ under the constraint $\alpha_1 = 0$.

Proof of (i)**.**

Let $\mathbb{D}_\mu(\rho) = \mathbf{B}_N(\rho)\mathbf{D}_\mu$, $\mathbb{D}_\alpha(\rho) = \mathbf{B}_N(\rho)\mathbf{D}_\alpha^\star$, $\mathcal{D}_{11}(\rho) = \mathbb{D}_\mu'(\rho)\mathbb{D}_\mu(\rho)$, $\mathcal{D}_{12}(\rho) = \mathbb{D}_\mu'(\rho)\mathbb{D}_\alpha(\rho)$, $\mathcal{D}_{22}(\rho) = \mathbb{D}_\alpha'(\rho)\mathbb{D}_\alpha(\rho)$ and $\mathcal{F}(\rho) = \mathbb{D}_\mu'(\rho)\mathbb{Q}_{\mathbb{D}_\alpha}(\rho)\mathbb{D}_\mu(\rho)$. Using the inverse formula of a partioned matrix, one has

$$[\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1} = \begin{bmatrix} \mathcal{D}_{11}(\rho) & \mathcal{D}_{12}(\rho) \\ \mathcal{D}_{12}'(\rho) & \mathcal{D}_{22}(\rho) \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \mathcal{F}^{-1}(\rho) & -\mathcal{F}^{-1}(\rho)\mathcal{D}_{12}(\rho)\mathcal{D}_{22}^{-1}(\rho) \\ -\mathcal{D}_{22}^{-1}(\rho)\mathcal{D}_{12}'(\rho)\mathcal{F}^{-1}(\rho) & \mathcal{D}_{22}^{-1}(\rho) + \mathcal{D}_{22}^{-1}(\rho)\mathcal{D}_{12}'(\rho)\mathcal{F}^{-1}(\rho)\mathcal{D}_{12}(\rho)\mathcal{D}_{22}^{-1}(\rho) \end{bmatrix}.$$

Plugging this into $\mathbb{Q}_{\mathbb{D}}(\rho)$, we obtain after some algebra,

$$\mathbb{Q}_{\mathbb{D}}(\rho) = \mathbb{Q}_{\mathbb{D}_\alpha}(\rho) - \mathbb{Q}_{\mathbb{D}_\alpha}(\rho)\mathbb{D}_\mu(\rho)[\mathbb{D}_\mu'(\rho)\mathbb{Q}_{\mathbb{D}_\alpha}(\rho)\mathbb{D}_\mu(\rho)]^{-1}\mathbb{D}_\mu'(\rho)\mathbb{Q}_{\mathbb{D}_\alpha}(\rho). \quad \text{(A.1)}$$

Given the special structure of $\mathbb{D}_\alpha(\rho)$, one has $\mathbb{Q}_{\mathbb{D}_\alpha}(\rho) = \texttt{blkdiag}(J_1(\rho), \ldots, J_T(\rho))$, where $J_1(\rho) = I_{n_1}$ and $J_t(\rho) = I_{n_t} - \frac{1}{n_t}B_t(\rho)l_{n_t}[\frac{1}{n_t}l_{n_t}'B_t'(\rho)B_t(\rho)l_{n_t}]^{-1}l_{n_t}'B_t'(\rho)$ for $t = 2, \cdots, T$. By Assumption D, the limit of $\frac{1}{n_t}l_{n_t}'B_t'(\rho)B_t(\rho)l_{n_t}$ is bounded away from zero and the elements of $B_t(\rho)l_{n_t}l_{n_t}'B_t'(\rho)$ are uniformly bounded, uniformly in $\rho \in \Delta_\rho$ for each t. Therefore, $J_t(\rho)$ must be uniformly bounded in both row and column sums, uniformly in $\rho \in \Delta_\rho$ for all t. Hence, $\mathbb{Q}_{\mathbb{D}_\alpha}(\rho)$ is also uniformly bounded in both row and column sums, uniformly in $\rho \in \Delta_\rho$.

We next consider the second term on the RHS of equation (A.1). We denote it as $\bar{\mathcal{Q}}(\rho)$, which can be partitioned into $T \times T$ blocks with (s,t)th block being

$$\bar{\mathcal{Q}}_{s,t}(\rho) = -\tfrac{1}{T}J_s(\rho)B_s(\rho)D_s[\tfrac{1}{T}\textstyle\sum_{t=1}^T D_t'B_t'(\rho)J_t(\rho)B_t(\rho)D_t]^{-1}D_t'B_t'(\rho)J_t(\rho).$$

By assuming that $B_s(\rho)D_s[\frac{1}{T}\sum_{t=1}^T D_t'B_t'(\rho)J_t(\rho)B_t(\rho)D_t]^{-1}D_t'B_t'(\rho)$ is uniformly bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$, for all s and t, we have that the row and column sums of each $\bar{\mathcal{Q}}_{s,t}(\rho)$ must have uniform order $O(1/T)$, uniformly in $\rho \in \Delta_\rho$. As there are T blocks in each row or in each column of $\bar{\mathcal{Q}}(\rho)$, we must have $\bar{\mathcal{Q}}(\rho)$ is bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$. Consequently, $\mathbb{Q}_{\mathbb{D}}(\rho)$ is bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$.

Proof of (ii)**.** Let $Z_N(\rho) = [\frac{1}{N}\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}$ with its (j,k)th element being denoted by $z_{jk}(\rho)$. From Assumption C(ii), $Z_N(\rho)$ converges to a finite limit uniformly in $\rho \in \Delta_\rho$. Therefore, there exists a constant c_z such that $|z_{jk}(\rho)| \leq c_z$ uniformly in $\rho \in \Delta_\rho$ for large enough N. Note that $\mathbb{X}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{X}$. As the elements of $\mathbf{X}$ are uniformly bounded (Assumption C(i)), and $\mathbf{B}_N(\rho)$ and $\mathbb{Q}_{\mathbb{D}}(\rho)$ are bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$, the elements of $\mathbb{X}(\rho)$ are also uniformly bounded, uniformly in $\rho \in \Delta_\rho$. Hence, there exists a constant c_x such that $|x_{jk}(\rho)| \leq c_x$ uniformly in $\rho \in \Delta_\rho$, where $x_{jk}(\rho)$ is the (j,k)th element of $\mathbb{X}(\rho)$. Let $p_{jl}(\rho)$ be the (j,l)th element of $\mathbb{P}_{\mathbb{X}}(\rho) = \frac{1}{N}\mathbb{X}(\rho)[\frac{1}{N}\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)$. It follows that uniformly in $\rho \in \Delta_\rho$, $\sum_{j=1}^N |p_{jl}(\rho)| \leq \frac{1}{N}\sum_{j=1}^N \sum_{r=1}^k \sum_{s=1}^k |z_{rs}(\rho)x_{jr}(\rho)x_{ls}(\rho)| \leq k^2 c_z c_x^2$ for all $l = 1, 2, \ldots, N$. Similarly, uniformly in $\rho \in \Delta_\rho$, we have $\sum_{l=1}^N |p_{jl}(\rho)| \leq \frac{1}{N}\sum_{l=1}^N \sum_{r=1}^k \sum_{s=1}^k |z_{rs}(\rho)x_{jr}(\rho)x_{ls}(\rho)| \leq k^2 c_z c_x^2$ for all $j = 1, 2, \ldots, N$. That is, $\mathbb{P}_{\mathbb{X}}(\rho)$ is bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$. Consequently, $\mathbb{Q}_{\mathbb{X}}(\rho) = I_N - \mathbb{P}_{\mathbb{X}}(\rho)$ is also bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$. ■

Lemma A.4. *Suppose that* $\{A_N\}$ *and* $\{B_N\}$ *are two sequences of* $N \times N$ *matrices that are uniformly bounded in either row or column sums. Under Assumptions C-E,* $\mathtt{tr}[A_N \mathbb{P}_{\mathbb{X}}(\rho) B_N] = O(1)$*, uniformly in* $\rho \in \Delta_\rho$*.*

Proof of Lemma A.4: *From the proof of Lemma A.3, the elements of* $\mathbb{X}(\rho)$ *and the elements of* $[\frac{1}{N}\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}$ *are uniformly bounded, uniformly in* $\rho \in \Delta_\rho$*. If* A_N *and* B_N *are bounded in row (column) sum norm, then* $A_N B_N$ *is also bounded in row (column) sum norm. Thus, Lemma A.6 of Lee (2004) implies that the elements of* $\frac{1}{N}\mathbb{X}'(\rho)A_N B_N \mathbb{X}(\rho)$ *are uniformly bounded. It follows that* $\mathtt{tr}[A_N \mathbb{P}_{\mathbb{X}}(\rho) B_N] = \mathtt{tr}[(\frac{1}{N}\mathbb{X}'(\rho)\mathbb{X}(\rho))^{-1}\frac{1}{N}\mathbb{X}'(\rho)A_N B_N \mathbb{X}(\rho)] = O(1)$*, uniformly in* $\rho \in \Delta_\rho$ *because the number of regressors* k *is fixed.* ■

Lemma A.5. *(Lemma A.2, Lin and Lee, 2010; Lemma A.3, Liu and Yang, 2015): Let* $A_N = [a_{ij}]$ *and* $B_N = [b_{ij}]$ *be two square matrices of dimension* N *and* c_N *be an* $N \times 1$ *vector of elements* c_i*. Assume that innovations* $\{v_j\}$

have zero mean and are mutually independent, i.e. $v_j \sim inid(0, \sigma_j^2)$*. Letting* $\mathbf{H} = \mathtt{diag}\{\sigma_1^2, \cdots, \sigma_N^2\}$ *and* $\mathbf{V} = (v_1, \cdots, v_N)'$*, we have,*

(*i*) $\mathrm{E}(\mathbf{V}'A_N\mathbf{V}) = \mathtt{tr}(\mathbf{H}A_N) = \sum_{i=1}^N a_{ii}\sigma_i^2$,

(*ii*) $\mathrm{E}(\mathbf{V}'A_N\mathbf{V} \cdot c_N'\mathbf{V}) = \sum_{i=1}^N a_{ii}c_i\mathrm{E}(v_i^3)$,

(*iii*) $\mathrm{E}(\mathbf{V}'A_N\mathbf{V} \cdot \mathbf{V}'B_N\mathbf{V}) = \sum_{i=1}^N a_{ii}b_{ii}[\mathrm{E}(v_i^4) - 3\sigma_i^4] + \mathtt{tr}(\mathbf{H}A_N)\mathtt{tr}(\mathbf{H}B_N) + \mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N^\circ)$,

(*iv*) $\mathrm{Var}(\mathbf{V}'A_N\mathbf{V}) = \sum_{i=1}^N a_{ii}^2[\mathrm{E}(v_i^4) - 3\sigma_i^4] + \mathtt{tr}(\mathbf{H}A_N\mathbf{H}A_N^\circ)$.

Lemma A.6. *(Lemma A.3, Lin and Lee, 2010, extended): Let* $\{A_N\}$ *be a sequence of* $N \times N$ *matrices such that either* $\|A_N\|_\infty$ *or* $\|A_N\|_1$ *is bounded. Suppose that the elements of* A_N *are* $O(h_n^{-1})$ *uniformly in all* i *and* j*. Let innovation vector* $\mathbf{V}$ *be defined as in Lemma A.5. Let* c_N *be an* $N \times 1$ *vector with elements of uniform order* $O(h_n^{-1/2})$*. Then*

(*i*) $\mathrm{E}(\mathbf{V}'A_N\mathbf{V}) = O(\frac{N}{h_n})$, (*ii*) $\mathrm{Var}(\mathbf{V}'A_N\mathbf{V}) = O(\frac{N}{h_n})$,

(*iii*) $\mathbf{V}'A_N\mathbf{V} = O_p(\frac{N}{h_n})$, (*iv*) $\mathbf{V}'A_N\mathbf{V} - \mathrm{E}(\mathbf{V}'A_N\mathbf{V}) = O_p((\frac{N}{h_n})^{\frac{1}{2}})$,

(*v*) $c_N'A_N\mathbf{V} = O_p((\frac{N}{h_n})^{\frac{1}{2}})$, **if** $\|A_N\|_1$ *is bounded.*

Proof of Lemma A.6: Firstly, Lemma A.8 of Lee (2004) implies that $\mathtt{tr}(\mathbf{H}A_N)$, $\mathtt{tr}(A_NA_N')$, $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}A_N)$ and $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}A_N')$ are all $O(\frac{N}{h_n})$. As $\sum_{i=1}^N a_{ii}^2 \leq \mathtt{tr}(A_NA_N')$, we also have $\sum_{i=1}^N a_{ii}^2 = O(\frac{N}{h_n})$. These and Lemma A.5 show that $\mathrm{E}(\mathbf{V}'A_N\mathbf{V}) = \mathtt{tr}(\mathbf{H}A_N) = O(\frac{N}{h_n})$ and $\mathrm{Var}(\mathbf{V}'A_N\mathbf{V}) = \sum_{i=1}^N a_{ii}^2[\mathrm{E}(v_i^4) - 3\sigma_i^4] + \mathtt{tr}[\mathbf{H}A_N(\mathbf{H}A_N' + \mathbf{H}A_N)] = O(\frac{N}{h_n})$. As $\mathrm{E}[(\mathbf{V}'A_N\mathbf{V})^2] = \mathrm{Var}(\mathbf{V}'A_N\mathbf{V}) + \mathrm{E}^2(\mathbf{V}'A_N\mathbf{V}) = O((\frac{N}{h_n})^2)$, we have $P(\frac{h_n}{N}|\mathbf{V}'A_N\mathbf{V}| \geq M) \leq \frac{1}{M^2}(\frac{h_n}{N})^2\mathrm{E}[(\mathbf{V}'A_N\mathbf{V})^2] = O(1)$, by the generalized Chebyshev's inequality. It follows that $\mathbf{V}'A_N\mathbf{V} = O_p(\frac{N}{h_n})$. Moreover, by Chebyshev's inequality, $P((\frac{h_n}{N})^{\frac{1}{2}}|\mathbf{V}'A_N\mathbf{V} - \mathrm{E}(\mathbf{V}'A_N\mathbf{V})| \geq M) \leq \frac{1}{M^2}\frac{h_n}{N}\mathrm{Var}(\mathbf{V}'A_N\mathbf{V}) = O(1)$. This implies that $\mathbf{V}'A_N\mathbf{V} - \mathrm{E}(\mathbf{V}'A_N\mathbf{V}) = O_p((\frac{N}{h_n})^{\frac{1}{2}})$. Finally, as the elements of c_N have uniform order $O(h_n^{-1/2})$, there exists a constant $\bar{c}$ such that $|c_j| \leq \frac{\bar{c}}{h_n^{1/2}}$ for all j. Hence, we have by the boundedness of $\|A_N\|_1$,

$$\begin{aligned}\mathrm{Var}[(\tfrac{h_n}{N})^{\frac{1}{2}}c_N'A_N\mathbf{V}] &= \tfrac{h_n}{N}\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N c_jc_ka_{ji}a_{ki}\sigma_i^2 \\ &\leq \bar{c}^2(\tfrac{1}{N}\sum_{i=1}^N\sigma_i^2)(\sum_{j=1}^N|a_{ji}|)(\sum_{k=1}^N|a_{ki}|) = O(1).\end{aligned}$$

It follows that $c_N'A_N\mathbf{V} = O_p((\frac{N}{h_n})^{\frac{1}{2}})$, by Chebyshev's inequality. ■

Appendix B: Proofs for Section 1.2

In proving the theorems, the following facts are used: (i) the eigenvalues of a projection matrix are either 0 or 1; (ii) the eigenvalues of a positive definite (p.d.) matrix are strictly positive; (iii) $\gamma_{min}(A)\mathtt{tr}(B) \leq \mathtt{tr}(AB) \leq \gamma_{max}(A)\mathtt{tr}(B)$ for symmetric matrix A and positive semi-definite (p.s.d.) matrix B; (iv) $\gamma_{max}(A+B) \leq \gamma_{max}(A) + \gamma_{max}(B)$ for symmetric matrices A and B; and (v) $\gamma_{max}(AB) \leq \gamma_{max}(A)\gamma_{max}(B)$ for p.s.d. matrices A and B.

The validity of Assumption E(iv) under a balanced panel:

For a balanced panel with a time-invariant and row-normalized spatial weight matrix, we have for all t, $n_t = n$, $D_t = I_n$, $M_t = M$, and $B_t(\rho) = I_n - \rho M \equiv B(\rho)$. As $M \times l_n = l_n$, $J_t(\rho) = I_n - \frac{1}{n} l_n l_n'$, $t = 2, \ldots, T$. Thus, we are able to get $B_s(\rho)D_s[\frac{1}{T}\sum_{t=1}^{T} D_t' B_t'(\rho) J_t(\rho) B_t(\rho) D_t]^{-1} D_t' B_t'(\rho) = (I_n - \frac{T-1}{nT} l_n l_n')^{-1}$. As $I_n - \frac{T-1}{nT} l_n l_n'$ is strictly diagonally dominant in rows and columns, its inverse is bounded in row and column sum norms (Varah, 1975).

Derivation of the AQS functions and the Hessian matrix:

Writing the key quantity in the concentrated quasi loglikelihood function (1.5) as $\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) = [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$, and using the facts that for an invertible matrix $A(\lambda)$, we have $\frac{\partial}{\partial\lambda}\ln|A(\lambda)| = \mathtt{tr}[A^{-1}(\lambda)\frac{\partial}{\partial\lambda}A(\lambda)]$ and $\frac{\partial}{\partial\lambda}A^{-1}(\lambda) = -A^{-1}(\lambda)[\frac{\partial}{\partial\lambda}A(\lambda)]A^{-1}(\lambda)$, it is straightforward to derive $S_N^c(\theta)$. However, the derivation of the ρ-component is complicated and some intermediate results are useful. First,

$$\begin{aligned} &\frac{\partial}{\partial\rho}[\mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)] \\ =\; &-\mathbf{M}'\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho) - \mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{M} + \mathbf{B}_N'(\rho)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho) \\ &+ \mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho), \end{aligned}$$

where $\dot{\mathbb{Q}}_{\mathbb{D}}(\rho) = \frac{\partial}{\partial\rho}\mathbb{Q}_{\mathbb{D}}(\rho)$. With $\frac{\partial}{\partial\rho}\mathbb{D}(\rho) = -\mathbf{MD} = -\mathbf{G}_N(\rho)\mathbb{D}(\rho)$, we have

$$\dot{\mathbb{Q}}_{\mathbb{D}}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)\mathbb{P}_{\mathbb{D}}(\rho) + \mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho). \tag{B.1}$$

This leads to $-\frac{\partial}{\partial\rho}[\mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)] = \mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N^{\circ}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho) \equiv \Psi(\rho)$, the ρ-component of the CQS function (1.6), and the ρ-component of the

AQS function (1.8):

$$S_\rho^*(\theta) = \frac{1}{2\sigma_v^2}[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\Psi(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)]. \quad \text{(B.2)}$$

This is expressed in terms of $\Psi(\rho)$ and $\mathbf{G}_N^\circ(\rho)$ to facilitate the derivations of the ρ-related terms of the Hessian matrix $\frac{\partial}{\partial\rho}\Psi(\rho)$. Again, the (ρ,ρ) term of $\frac{\partial}{\partial\rho}\Psi(\rho)$ is most complicate. For a comformable vector a, we have by taking use of (B.1) and after some tedious algebra,

$$a'[\tfrac{\partial}{\partial\rho}\Psi(\rho)]a = 2a'\mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)[\mathbf{G}_N^\circ(\rho)\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}_N^\circ(\rho) - \mathbf{G}_N'(\rho)\mathbf{G}_N(\rho)]\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)a. \quad \text{(B.3)}$$

With the set of AQS functions $S_N^*(\theta)$ given in (1.8) and (B.1)-(B.3), we obtain the components of the Hessian matrix $H_N^*(\theta) = \frac{\partial}{\partial\theta'}S_N^*(\theta)$:

$$\begin{aligned}
H_{\beta\beta}^*(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{X}'(\rho)\mathbb{X}(\rho),\\
H_{\beta\sigma_v^2}^*(\theta) &= -\tfrac{1}{\sigma_v^4}\mathbb{X}'(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H_{\sigma_v^2\beta}^{*\prime},\\
H_{\beta\lambda}^*(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{X}'(\rho)\mathbb{Y}(\rho) \;=\; H_{\lambda\beta}^{*\prime},\\
H_{\beta\rho}^*(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{X}'(\rho)\mathbf{G}_N^\circ(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H_{\rho\beta}^{*\prime},\\
H_{\sigma_v^2\sigma_v^2}^*(\theta) &= -\tfrac{1}{\sigma_v^6}\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) + \tfrac{1}{2\sigma_v^4}N_1,\\
H_{\sigma_v^2\lambda}^*(\theta) &= -\tfrac{1}{\sigma_v^4}\mathbb{Y}'(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H_{\lambda\sigma_v^2}^{*\prime},\\
H_{\sigma_v^2\rho}^*(\theta) &= -\tfrac{1}{2\sigma_v^4}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}_N^\circ(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H_{\rho\sigma_v^2}^{*\prime},\\
H_{\lambda\lambda}^*(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{Y}'(\rho)\mathbb{Y}(\rho) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N^2(\lambda)\mathbf{B}_N^{-1}(\rho)],\\
H_{\lambda\rho}^*(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{Y}'(\rho)\mathbf{G}_N^\circ(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{F}_N(\lambda)\mathbb{R}_N(\rho)],\\
H_{\rho\lambda}^*(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{Y}'(\rho)\mathbf{G}_N^\circ(\rho)\tilde{\mathbf{V}}(\beta,\delta),\\
H_{\rho\rho}^*(\theta) &= \tfrac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathcal{R}_{1N}(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathcal{R}_{2N}(\rho)],
\end{aligned} \quad \text{(B.4)}$$

where $\mathbb{Y}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{W}\mathbf{Y}$, $\mathbb{R}_N(\rho) = \mathbf{B}_N^{-1}(\rho)\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}_N^\circ(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)$, $\mathcal{R}_{1N}(\rho) = \mathbf{G}_N^\circ(\rho)\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}_N^\circ(\rho) - \mathbf{G}_N'(\rho)\mathbf{G}_N(\rho)$ and $\mathcal{R}_{2N}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)[\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}_N^\circ(\rho) + \mathbf{G}_N(\rho)]$.

Proof of Theorem 1.1: By theorem 5.9 of van der Vaart (1998), we only need to show $\sup_{\delta\in\delta}\frac{1}{N_1}\left\|S_N^{*c}(\delta) - \bar{S}_N^{*c}(\delta)\right\| \xrightarrow{p} 0$ under the assumptions in Theorem 1.1. From (1.10) and (1.12), the consistency of $\hat{\delta}_N^*$ follows from:

(a) $\inf_{\delta\in\Delta}\bar{\sigma}_{v,N}^{*2}(\delta)$ is bounded away from zero,

(b) $\sup_{\delta\in\Delta}\left|\hat{\sigma}_{v,N}^{*2}(\delta) - \bar{\sigma}_{v,N}^{*2}(\delta)\right| = o_p(1)$,

(c) $\sup_{\delta\in\Delta}\frac{1}{N_1}\left|\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\hat{\mathbf{V}}(\delta) - \mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\bar{\mathbf{V}}(\delta)]\right| = o_p(1)$,

(d) $\sup_{\delta\in\Delta}\frac{1}{N_1}\big|\hat{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\hat{\mathbf{V}}(\delta)-\mathrm{E}[\bar{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\bar{\mathbf{V}}(\delta)]\big|=o_p(1)$.

Proof of (a). From (1.11), $\bar\beta^*_N(\delta)=[\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\mathrm{E}(\mathbf{Y})$ as $\mathbb{X}(\rho)=\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{X}$ and $\mathbb{Q}_{\mathbb{D}}(\rho)$ is idempotent. Thus, $\bar{\mathbf{V}}(\delta)=\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\mathbf{Y}-\mathbb{X}(\rho)\bar\beta^*_N(\delta)=\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\mathbf{Y}+\mathbb{P}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]$. By the orthogonality between $\mathbb{Q}_{\mathbb{D}}(\rho)$ and $\mathbb{P}_{\mathbb{D}}(\rho)$ and using $\mathbf{Y}=\mathbf{A}_N^{-1}(\eta+\mathbf{B}_N^{-1}\mathbf{V})$, we have,

$$\begin{aligned}\bar\sigma^{*2}_{v,N}(\delta)&=\tfrac{1}{N_1}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\bar{\mathbf{V}}(\delta)]\\&=\tfrac{1}{N_1}\mathrm{E}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}]+\tfrac{1}{N_1}\mathrm{E}\big\{[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]'\mathbf{P}(\delta)[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]\big\}\qquad\text{(B.5)}\\&=\tfrac{1}{N_1}\mathrm{E}(\mathbf{Y})'\mathbf{Q}(\delta)\mathrm{E}(\mathbf{Y})+\tfrac{1}{N_1}\mathrm{E}\big\{[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]'[\mathbf{Q}(\delta)+\mathbf{P}(\delta)][\mathbf{Y}-\mathrm{E}(\mathbf{Y})]\big\}\\&=\tfrac{1}{N_1}\mathrm{E}(\mathbf{Y})'\mathbf{Q}(\delta)\mathrm{E}(\mathbf{Y})+\tfrac{1}{N_1}\mathrm{E}\big\{[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]'\mathbf{C}'_N(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]\big\}\\&=\tfrac{1}{N_1}\eta'\mathbf{A}'^{-1}_N\mathbf{Q}(\delta)\mathbf{A}_N^{-1}\eta+\tfrac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_N(\delta)],\end{aligned}$$

where $\mathbf{Q}(\delta)=\mathbf{C}'_N(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$ and $\mathbf{P}(\delta)=\mathbf{C}'_N(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{P}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$. The first term can be written in the form of $a'(\delta)a(\delta)$ for an $N\times 1$ vector function of δ, and thus is non-negative, uniformly in $\delta\in\Delta$. For the second term,

$$\begin{aligned}&\tfrac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_N(\delta)]\ge\tfrac{\sigma^2_{v0}}{N_1}\gamma_{\min}[\mathcal{C}_N(\delta)]\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)]=\sigma^2_{v0}\gamma_{\min}[\mathcal{C}_N(\delta)]\\&\ge\sigma^2_{v0}\gamma_{\max}(\mathbf{A}'_N\mathbf{A}_N)^{-1}\gamma_{\max}(\mathbf{B}'_N\mathbf{B}_N)^{-1}\gamma_{\min}[\mathbf{A}'_N(\lambda)\mathbf{A}_N(\lambda)]\gamma_{\min}[\mathbf{B}'_N(\rho)\mathbf{B}_N(\rho)]>0,\end{aligned}$$

uniformly in $\delta\in\Delta$, by Assumption E(iii). It follows that $\inf_{\delta\in\Delta}\bar\sigma^{*2}_{v,N}(\delta)>0$.

Proof of (b). From (1.9), $\hat\beta^*_N(\delta)=[\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\mathbf{Y}$. Then, $\hat{\mathbf{V}}(\delta)=\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y}-\mathbf{X}\hat\beta^*_N(\delta)]=\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\mathbf{Y}$ and $\hat\sigma^{*2}_{v,N}(\delta)=\frac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}$. From (B.5), $\bar\sigma^{*2}_{v,N}(\delta)=\frac{1}{N_1}\mathrm{E}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}]+\frac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbf{C}'^{-1}_N\mathbf{P}(\delta)\mathbf{C}^{-1}_N]$. Thus,

$$\hat\sigma^{*2}_{v,N}(\delta)-\bar\sigma^{*2}_{v,N}(\delta)=\tfrac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}-\mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]-\tfrac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbf{C}'^{-1}_N\mathbf{P}(\delta)\mathbf{C}^{-1}_N].$$

For the second term, $0\le\frac{1}{N_1}\mathtt{tr}[\mathbf{C}'^{-1}_N\mathbf{P}(\delta)\mathbf{C}^{-1}_N]\le\frac{1}{N_1}\gamma_{\max}[\mathcal{C}_N(\delta)]\gamma^2_{\max}[\mathbb{Q}_{\mathbb{D}}(\rho)]$ $\mathtt{tr}[\mathbb{P}_{\mathbb{X}}(\rho)]=o(1)$, because $\mathtt{tr}[\mathbb{P}_{\mathbb{X}}(\rho)]=k$, $\gamma_{\max}[\mathbb{Q}_{\mathbb{D}}(\rho)]=1$ and, by Assumption E(iii), $\gamma_{\max}[\mathcal{C}_N(\delta)]\le\gamma_{\min}(\mathbf{A}'_N\mathbf{A}_N)^{-1}\gamma_{\min}(\mathbf{B}'_N\mathbf{B}_N)^{-1}\gamma_{\max}[\mathbf{A}'_N(\lambda)\mathbf{A}_N(\lambda)]$ $\gamma_{\max}[\mathbf{B}'_N(\rho)\mathbf{B}_N(\rho)]<\infty$. Therefore, one has $\sup_{\delta\in\Delta}|\frac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbf{C}'^{-1}_N\mathbf{P}(\delta)\mathbf{C}^{-1}_N]|=o(1)$. For the first term, we prove the uniform convergence: $\sup_{\delta\in\Delta}|\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}-\mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]|=o_p(1)$, which follows from pointwise convergence of $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}-$

$\text{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]$ to zero for each $\delta \in \Delta$ and the stochastic equicontinuity of $\frac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}$, according to Andrews (1992). We have,

$$\begin{aligned}&\tfrac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \text{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]\\ =\ &\tfrac{2}{N_1}\mathbf{V}'\mathbf{C}_N^{-1\prime}\mathbf{Q}(\delta)\mathbf{A}_N^{-1}\eta + \tfrac{1}{N_1}[\mathbf{V}'\mathbf{C}_N^{-1\prime}\mathbf{Q}(\delta)\mathbf{C}_N^{-1}\mathbf{V} - \sigma_{v0}^2\mathtt{tr}(\mathbf{C}_N^{-1\prime}\mathbf{Q}(\delta)\mathbf{C}_N^{-1})].\end{aligned}$$

By Assumption E, and Lemmas A.1 and A.3, one shows that $\mathbf{C}_N^{-1\prime}\mathbf{Q}(\delta)\mathbf{A}_N^{-1}$and $\mathbf{C}_N^{-1\prime}\mathbf{Q}(\delta)\mathbf{C}_N^{-1}$ are bounded in both row and column sum norms, for each $\delta \in \Delta$. Further, the elements of η are uniformly bounded. Thus, the pointwise convergence of the first term follows from Lemma A.6 (v), and the pointwise convergence of the second term follows from Lemma A.6 (iv). Therefore, $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \text{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})] \xrightarrow{p} 0$, for each $\delta \in \Delta$.

Next, let δ_1 and δ_2 be in Δ. We have by the mean value theorem (MVT):

$$\tfrac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta_2)\mathbf{Y} - \tfrac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta_1)\mathbf{Y} = \tfrac{1}{N_1}\mathbf{Y}'[\tfrac{\partial}{\partial\delta'}\mathbf{Q}(\bar{\delta})]\mathbf{Y}(\delta_2 - \delta_1),$$

where $\bar{\delta}$ lies between δ_1 and δ_2. It follows that $\frac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}$ is stochastically equicontinuous if $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\varpi}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$, $\varpi = \lambda, \rho$. We only show $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\rho}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$ as the proof of $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\lambda}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$ is similar and simpler. Note that

$$\begin{aligned}\tfrac{\partial}{\partial\rho}\mathbf{Q}(\delta) = &-\mathbf{C}_N'(\delta)\mathbf{G}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta) + \mathbf{C}_N'(\delta)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\\ &+\mathbf{C}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\dot{\mathbb{Q}}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta) + \mathbf{C}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\\ &-\mathbf{C}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)\mathbf{C}_N(\delta),\end{aligned}$$

where $\dot{\mathbb{Q}}_{\mathbb{X}}(\rho) = \frac{\partial}{\partial\rho}\mathbb{Q}_{\mathbb{X}}(\rho)$. Using (B.1), we have after some algebra, $\dot{\mathbb{X}}(\rho) = \frac{\partial}{\partial\rho}\mathbb{X}(\rho) = \mathbb{G}_N(\rho)\mathbb{X}(\rho)$ where $\mathbb{G}_N(\rho) = \mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}_N'(\rho) - \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)$, which gives

$$\dot{\mathbb{Q}}_{\mathbb{X}}(\rho) = -\mathbb{P}_{\mathbb{X}}(\rho)\mathbb{G}_N'(\rho)\mathbb{Q}_{\mathbb{X}}(\rho) - \mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{G}_N(\rho)\mathbb{P}_{\mathbb{X}}(\rho). \tag{B.6}$$

For a comformable vector a and taking use (B.1) and (B.6), we have after some algebra,

$$a'[\tfrac{\partial}{\partial\rho}\mathbf{Q}(\delta)]a = -2a'\bar{\mathbf{Q}}(\delta)a, \tag{B.7}$$

where $\bar{\mathbf{Q}}(\delta) = \mathcal{Q}_N'(\delta)\mathbf{G}_N(\rho)\mathcal{Q}_N(\delta)$ and $\mathcal{Q}_N(\delta) = \mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$. Some rearrangements lead to $\bar{\mathbf{Q}}(\delta) = \mathcal{Q}_N'(\delta)\mathbf{M}\bar{\mathbb{Q}}_{\mathbb{D}}(\rho)\bar{\mathbb{Q}}_{\mathbb{X}}(\rho)\mathbf{A}_N(\lambda)$, where we define $\bar{\mathbb{Q}}_{\mathbb{D}}(\rho) = I_N - \mathbf{D}[\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)\mathbf{B}_N(\rho)$ and $\bar{\mathbb{Q}}_{\mathbb{X}}(\rho) = I_N - \mathbf{X}[\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)$ $\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)$. Following exactly the same way as we prove Lemma A.3, we show that $\bar{\mathbb{Q}}_{\mathbb{D}}(\rho)$ and $\bar{\mathbb{Q}}_{\mathbb{X}}(\rho)$ are also uniformly bounded in both row and column sums, uniformly in $\rho \in \Delta_\rho$. This implies that both $\|\bar{\mathbf{Q}}(\delta)\|_1$ and $\|\bar{\mathbf{Q}}(\delta)\|_\infty$ are bounded uniformly in $\delta \in \Delta$. As $\mathbf{Y} = \mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$, Lemma A.1 and Lemma A.6 imply

$$\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\rho}\mathbf{Q}(\delta)]\mathbf{Y} = -\frac{2}{N_1}\mathbf{Y}'\bar{\mathbf{Q}}(\delta)\mathbf{Y} = -\frac{2}{N_1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})'\mathbf{A}_N'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$$
$$= -\frac{2}{N_1}\eta'\mathbf{A}_N'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{A}_N^{-1}\eta - \frac{4}{N1}\eta'\mathbf{A}_N'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{C}_N^{-1}\mathbf{V} - \frac{2}{N_1}\mathbf{V}'\mathbf{C}_N'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{C}_N^{-1}\mathbf{V} = O_p(1),$$

uniformly in $\delta \in \Delta$. Thus, $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\rho}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$. Following the similar analysis, one also has $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\lambda}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$. Therefore, $\sup_{\delta\in\Delta}|\hat{\sigma}^{*2}_{v,N}(\delta) - \bar{\sigma}^{*2}_{v,N}(\delta)| = o_p(1)$.

Proof of (c). By the expressions of $\hat{\mathbf{V}}(\lambda)$ and $\bar{\mathbf{V}}(\delta)$ given above, we have

$$\frac{1}{N_1}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\hat{\mathbf{V}}(\delta) - \frac{1}{N_1}\mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\bar{\mathbf{V}}(\delta)]$$
$$= \frac{1}{N_1}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\mathcal{Q}_N(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\mathcal{Q}_N(\delta)\mathbf{Y})]$$
$$- \frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{W}'\mathbf{B}_N'(\rho)\mathcal{P}_N(\delta)\mathbf{C}_N^{-1}],$$

where $\mathcal{P}_N(\delta) = \mathbb{P}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$. The first term is similar in form to $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]$ from **(b)**, and its uniform convergence is shown in a similar way. Furthermore, by Lemma A.4, it is easy to see that the second term is $o(1)$ uniformly in $\delta \in \Delta$.

Proof of (d). Again, using the expressions of $\bar{\mathbf{V}}(\delta)$ and $\hat{\mathbf{V}}(\delta)$, we have

$$\frac{1}{N_1}\hat{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\hat{\mathbf{V}}(\delta) - \frac{1}{N_1}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\bar{\mathbf{V}}(\delta)]$$
$$= \frac{1}{N_1}[\mathbf{Y}'\bar{\mathbf{Q}}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\bar{\mathbf{Q}}(\delta)\mathbf{Y})] - \frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathcal{P}_N'(\delta)\mathbf{G}_N^\circ(\rho)\mathcal{Q}_N(\delta)\mathbf{C}_N^{-1}]$$
$$- \frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathcal{P}_N'(\delta)\mathbf{G}_N(\rho)\mathcal{P}_N(\delta)\mathbf{C}_N^{-1}].$$

Therefore, the uniform convergence of the first term can also be shown similarly as we do for $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]$ since they have similar forms. By Lemma A.4, the remaining two terms are easily seen to be $o(1)$, uniformly in $\delta \in \Delta$. ■

Proof of Theorem 1.2: Applying the MVT to each element of $S_N^*(\hat{\theta}_N^*)$, we have

$$0 = \frac{1}{\sqrt{N_1}}S_N^*(\hat{\theta}_N^*) = \frac{1}{\sqrt{N_1}}S_N^*(\theta_0) + \left[\frac{1}{N_1}\frac{\partial}{\partial\theta'}S_N^*(\theta)\Big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}\right]\sqrt{N_1}(\hat{\theta}_N^* - \theta_0), \tag{B.8}$$

where $\{\bar{\theta}_r\}$ are on the line segment between $\hat{\theta}_N^*$ and θ_0. The result of the theorem follows if

(a) $\frac{1}{\sqrt{N_1}} S_N^*(\theta_0) \xrightarrow{D} N[0, \lim_{N\to\infty} \Gamma_N^*(\theta_0)]$,

(b) $\frac{1}{N_1}[\frac{\partial}{\partial\theta'} S_N^*(\theta)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}} - \frac{\partial}{\partial\theta'} S_N^*(\theta_0)] = o_p(1)$, and

(c) $\frac{1}{N_1}[\frac{\partial}{\partial\theta'} S_N^*(\theta_0) - \mathrm{E}(\frac{\partial}{\partial\theta'} S_N^*(\theta_0))] = o_p(1)$.

Proof of (a). From (1.13), we see that the elements of $S_N^*(\theta_0)$ are linear-quadratic forms in $\mathbf{V}$. Thus, for every non-zero $(k+3)\times 1$ vector of constants a, $a'S_N^*(\theta_0)$ is of the form:

$$a'S_N^*(\theta_0) = b_N'\mathbf{V} + \mathbf{V}'\Phi_N\mathbf{V} - \sigma_v^2 \mathtt{tr}(\Phi_N),$$

for suitably defined non-stochastic vector b_N and matrix Φ_N. Based on Assumptions A-F, it is easy to verify (by Lemma A.1 and Lemma A.3(i)) that b_N and matrix Φ_N satisfy the conditions of the CLT for LQ form of Kelejian and Prucha (2001), and hence the asymptotic normality of $\frac{1}{\sqrt{N_1}} a'S_N^*(\theta_0)$ follows. By Cramér-Wold device, $\frac{1}{\sqrt{N_1}} S_N^*(\theta_0) \xrightarrow{D} N[0, \lim_{N\to\infty} \Gamma_N^*(\theta_0)]$, where elements of $\Gamma_N^*(\theta_0)$ are given in (1.15).

Proof of (b). The Hessian matrix $H_N^*(\theta) = \frac{\partial}{\partial\theta'} S_N^*(\theta)$ is given in (B.4). By Assumptions D and E, and Lemma A.1 and Lemma A.3(i), $\mathbb{R}_N(\rho_0)$, $\mathcal{R}_{1N}(\rho_0)$ and $\mathcal{R}_{2N}(\rho_0)$ are all bounded in row and column sum norms. With these and $\mathbf{Y} = \mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$, Lemma A.6 leads to $\frac{1}{N_1} H_N^*(\theta_0) = O_p(1)$. Thus, $\frac{1}{N_1} H_N^*(\bar{\theta}) = O_p(1)$ since $\bar{\theta} \xrightarrow{p} \theta_0$ due to $\hat{\theta}_N^* \xrightarrow{p} \theta_0$, where for ease of exposition, $H_N^*(\bar{\theta})$ is used to denote $\frac{\partial}{\partial\theta'} S_N^*(\theta)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}$. As $\bar{\sigma}_v^2 \xrightarrow{p} \sigma_{v0}^2$, we have $\bar{\sigma}_v^{-r} = \sigma_{v0}^{-r} + o_p(1)$, for $r = 2, 4, 6$. As σ_v^{-r} appears in $H_N^*(\theta)$ multiplicatively, $\frac{1}{N_1} H_N^*(\bar{\theta}) = \frac{1}{N_1} H_N^*(\bar{\beta}, \bar{\lambda}, \bar{\rho}, \sigma_{v0}^2) + o_p(1)$. Thus, the proof of **(b)** is equivalent to the proof of

$$\frac{1}{N_1}[H_N^*(\bar{\beta}, \bar{\lambda}, \bar{\rho}, \sigma_{v0}^2) - H_N^*(\theta_0)] \xrightarrow{p} 0,$$

or the proofs of $\frac{1}{N_1}[H_N^{*\mathtt{S}}(\bar{\beta}, \bar{\lambda}, \bar{\rho}, \sigma_{v0}^2) - H_N^{*\mathtt{S}}(\theta_0)] \xrightarrow{p} 0$ and $\frac{1}{N_1}[H_N^{*\mathtt{NS}}(\bar{\delta}) - H_N^{*\mathtt{NS}}(\delta_0)] \xrightarrow{p} 0$, where $H_N^{*\mathtt{S}}$ and $H_N^{*\mathtt{NS}}$ denote, respectively, the stochastic and non-stochastic parts of H_N^*.

For the stochastic part, we see from (B.4) that all the components of $H_N^{*\mathtt{S}}(\beta, \lambda, \rho, \sigma_{v0}^2)$ are linear, bilinear or quadratic in β and λ, but nonlinear in ρ. Hence, with an application of the MVT on $H_N^{*\mathtt{S}}(\bar{\beta}, \bar{\lambda}, \bar{\rho}, \sigma_{v0}^2)$ w.r.t $\bar{\rho}$ 'variable', we can write $\frac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\bar\rho,\sigma_{v0}^2) - H_N^{*\mathtt{S}}(\theta_0)]$ as

$$\tfrac{1}{N_1}[\tfrac{\partial}{\partial\rho}H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\dot\rho,\sigma_{v0}^2)](\bar\rho-\rho_0) + \tfrac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\rho_0,\sigma_{v0}^2) - H_N^{*\mathtt{S}}(\theta_0)],$$

where $\dot\rho$ lies between $\bar\rho$ and ρ_0. Thus, it suffices to show (i) $\frac{1}{N_1}\frac{\partial}{\partial\rho}H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\dot\rho,\sigma_{v0}^2) = O_p(1)$, and (ii) $\frac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\rho_0,\sigma_{v0}^2) - H_N^{*\mathtt{S}}(\theta_0)] = o_p(1)$.

We select one of the most complicated components, $H_{\rho\lambda}^{*\mathtt{S}}(\theta) = -\frac{1}{\sigma_v^2}\mathbb{Y}'(\rho)\mathbf{G}_N^{\circ}(\rho)\tilde{\mathbf{V}}(\beta,\delta)$, to illustrate the general idea in the proof. We have, after some algebra,

$$\tfrac{1}{N_1}\tfrac{\partial}{\partial\rho}H_{\rho\lambda}^{*\mathtt{S}}(\bar\beta,\bar\lambda,\dot\rho,\sigma_{v0}^2) = \tfrac{2}{N_1\sigma_{v0}^2}\mathbb{Y}'(\dot\rho)\mathcal{R}_{1N}(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_N(\dot\rho)(\mathbf{A}_N(\bar\lambda)\mathbf{Y}-\mathbf{X}\bar\beta),$$

$$\tfrac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\rho_0,\sigma_{v0}^2) - H_N^{*\mathtt{S}}(\theta_0)] = \tfrac{1}{N_1\sigma_{v0}^2}\mathbb{Y}'\mathbf{G}_N^{\circ}\mathbb{Y}(\bar\lambda-\lambda_0) + \tfrac{1}{N_1\sigma_{v0}^2}\mathbb{Y}'\mathbf{G}_N^{\circ}\mathbb{X}(\bar\beta-\beta_0).$$

By Lemmas A.1 and A.6, it is easy to show that $\frac{1}{N_1}\mathbb{Y}'\mathbf{G}_N^{\circ}\mathbb{Y} = O_p(1)$ and $\frac{1}{N_1}\mathbb{Y}'\mathbf{G}_N^{\circ}\mathbb{X} = O_p(1)$. Therefore, (ii) holds. To prove (i), we have

$$\begin{aligned}&\mathbb{Y}'(\dot\rho)\mathcal{R}_{1N}(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_N(\dot\rho)(\mathbf{A}_N(\bar\lambda)\mathbf{Y}-\mathbf{X}\bar\beta)\\ =&(\mathbf{A}_N^{-1}\eta + \mathbf{C}_N^{-1}\mathbf{V})'\mathcal{H}_N(\dot\rho)[\mathbf{A}_N(\bar\lambda)\mathbf{A}_N^{-1}\eta + \mathbf{A}_N(\bar\lambda)\mathbf{C}_N^{-1}\mathbf{V} - \mathbf{X}\bar\beta]\end{aligned}$$

where $\mathcal{H}_N(\dot\rho) = \mathbf{W}'\mathbf{B}_N'(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathcal{R}_{1N}(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_N(\dot\rho)$. Lemma A.2 implies $\mathbf{B}_N^{-1}(\dot\rho)$ embedded in $\mathcal{H}_N(\dot\rho)$ is uniformly bounded in both row and column sums since $\dot\rho-\rho_0 = o_p(1)$. Therefore, it is easy to see the above equation is $O_p(N)$ by Lemma A.6 and then result (i) follows.

For the non-stochastic part, we illustrate the proof using the most complicate $\lambda\lambda$-term. Noting that the non-stochastic part is nonlinear in both $\bar\lambda$ and $\bar\rho$, we have by the MVT,

$$\begin{aligned}&\tfrac{1}{N_1}[H_{\lambda\lambda}^{*\mathtt{NS}}(\bar\delta) - H_{\lambda\lambda}^{*\mathtt{NS}}(\delta_0)] = -\tfrac{1}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\bar\rho)\mathbf{B}_N(\bar\rho)\mathbf{F}_N^2(\bar\lambda)\mathbf{B}_N^{-1}(\bar\rho) - \mathbb{Q}_{\mathbb{D}}\mathbf{B}_N\mathbf{F}_N^2\mathbf{B}_N^{-1}]\\ =&\ -(\bar\lambda-\lambda_0)\tfrac{1}{N_1}\mathtt{tr}[2\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_N(\dot\rho)\mathbf{F}_N^3(\dot\lambda)\mathbf{B}_N^{-1}(\dot\rho)] - (\bar\rho-\rho_0)\tfrac{1}{N_1}\mathtt{tr}[\mathbf{F}_N^2(\dot\lambda)\mathbb{R}_N(\dot\rho)],\end{aligned}$$

where $\dot\lambda$ lies between $\bar\lambda$ and λ_0 and $\dot\rho$ lies between $\bar\rho$ and ρ_0. Again, by Lemma A.2, we conclude that both $\mathbf{A}_N^{-1}(\dot\lambda)$ and $\mathbf{B}_N^{-1}(\dot\rho)$ are uniformly bounded in both row and column sums. Therefore, the terms inside the trace both have elements that are uniformly bounded. As $\bar\delta-\delta_0 = o_p(1)$, we have $\frac{1}{N_1}[H_{\lambda\lambda}^{*\mathtt{NS}}(\bar\delta) - H_{\lambda\lambda}^{*\mathtt{NS}}(\delta_0)] = o_p(1)$.

Proof of (c). Since $\mathbf{Y} = \mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$, the Hessian matrix at true θ_0 are seen to be linear combinations of terms linear or quadratic in $\mathbf{V}$, and constants. The constant terms are canceled out. Other terms are shown to be

$o_p(1)$ based on Lemma A.6. For example,
$\frac{1}{N_1}[H^*_{\rho\rho}(\rho_0) - \mathrm{E}(H^*_{\rho\rho}(\rho_0))] = \frac{1}{N_1\sigma^2_{v0}}[\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathrm{E}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V})] = o_p(1)$.
■

Proof of Corollary 1.1: Note that $\Gamma^*_N(\hat{\theta}^*_N) = \Gamma^*_N(\theta)|_{(\theta=\hat{\theta}^*_N, \phi=\hat{\phi}^*_N, \gamma=\hat{\gamma}_N, \kappa=\hat{\kappa}_N)}$. As $\hat{\theta}^*_N$, $\hat{\gamma}_N$ and $\hat{\kappa}_N$ are consistent estimators for θ_0, γ and κ, plugging these estimators into $\Gamma^*_N(\theta)$ will not bring additional bias to the estimation of $\Gamma^*_N(\theta_0)$. However, due to incidental parameters problem, the $\hat{\mu}^*_N$ component of $\hat{\phi}^*_N$ is not consistent for the estimation of μ_0 when T is fixed. The estimation bias caused by replacing ϕ_N by $\hat{\phi}^*_N$ can be derived as follow. Recall (1.4),

$$\hat{\phi}_N(\beta, \delta) = [\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta].$$

Thus, the unconstrained estimate of ϕ_0 is just $\hat{\phi}^*_N = \hat{\phi}_N(\hat{\beta}^*_N, \hat{\delta}^*_N)$. Note $\mathbf{A}_N(\hat{\lambda}^*_N)\mathbf{Y} - \mathbf{X}\hat{\beta}^*_N = \mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0 - \mathbf{W}\mathbf{Y}(\hat{\lambda}^*_N - \lambda_0) - \mathbf{X}(\hat{\beta}^*_N - \beta_0)$. Applying the MVT on each row of $\mathbf{D}\hat{\phi}^*_N$ with respect to the $\hat{\rho}^*_N$-element, we have,

$$\begin{aligned} \mathbf{D}\hat{\phi}^*_N &= \mathbf{D}[\mathbb{D}'(\hat{\rho}^*_N)\mathbb{D}(\hat{\rho}^*_N)]^{-1}\mathbb{D}'(\hat{\rho}^*_N)\mathbf{B}_N(\hat{\rho}^*_N)[\mathbf{A}_N(\hat{\lambda}^*_N)\mathbf{Y} - \mathbf{X}\hat{\beta}^*_N] \\ &= \mathbf{B}^{-1}_N(\hat{\rho}^*_N)\mathbb{P}_{\mathbb{D}}(\hat{\rho}^*_N)\mathbf{B}_N(\hat{\rho}^*_N)[\mathbf{A}_N(\hat{\lambda}^*_N)\mathbf{Y} - \mathbf{X}\hat{\beta}^*_N] \\ &= [\mathbf{B}^{-1}_N\mathbb{P}_{\mathbb{D}}\mathbf{B}_N - \mathbb{R}_N(\bar{\rho})(\hat{\rho}^*_N - \rho_0)][\mathbf{A}_N(\hat{\lambda}^*_N)\mathbf{Y} - \mathbf{X}\hat{\beta}^*_N] \\ &= \mathbf{D}\phi_0 + \mathbf{B}^{-1}_N\mathbb{P}_{\mathbb{D}}\mathbf{V} - \mathbf{B}^{-1}_N\mathbb{P}_{\mathbb{D}}\mathbf{B}_N[\mathbf{W}\mathbf{Y}(\hat{\lambda}^*_N - \lambda_0) + \mathbf{X}(\hat{\beta}^*_N - \beta_0)] \\ &\quad - \mathbb{R}_N(\bar{\rho})[\mathbf{A}_N(\hat{\lambda}^*_N)\mathbf{Y} - \mathbf{X}\hat{\beta}^*_N](\hat{\rho}^*_N - \rho_0), \end{aligned} \tag{B.9}$$

where $\bar{\rho}$ lies between $\hat{\rho}^*_N$ and ρ_0 and changes over the rows of $\mathbb{R}_N(\bar{\rho})$, and $\mathbb{R}_N(\rho)$ is given below (B.4). From its expression, $\Gamma^*_N(\theta)$ is seen to have components that are either linear or quadratic in $\mathbf{D}\phi$. Let d_N be a non-stochastic N-vector with elements being of uniform order $O(1)$ or $O(h_n^{-1})$. Using (B.9), the terms of $\Gamma^*_N(\hat{\theta}^*_N)$ linear in $\mathbf{D}\hat{\phi}^*_N$ can be represented as

$$\begin{aligned} \tfrac{1}{N_1}d'_N\mathbf{D}\hat{\phi}^*_N &= \tfrac{1}{N_1}d'_N\mathbf{D}\phi_0 + \tfrac{1}{N_1}d'_N\mathbf{B}^{-1}_N\mathbb{P}_{\mathbb{D}}\mathbf{V} \\ &\quad - \tfrac{1}{N_1}d'_N\mathbf{B}^{-1}_N\mathbb{P}_{\mathbb{D}}\mathbf{B}_N[\mathbf{W}\mathbf{Y}(\hat{\lambda}^*_N - \lambda_0) + \mathbf{X}(\hat{\beta}^*_N - \beta_0)] \\ &\quad + \tfrac{1}{N_1}d'_N\mathbb{R}_N(\bar{\rho})[\mathbf{A}_N(\hat{\lambda}^*_N)\mathbf{Y} - \mathbf{X}\hat{\beta}^*_N](\hat{\rho}^*_N - \rho_0) = \tfrac{1}{N_1}d'_N\mathbf{D}\phi_0 + o_p(1), \end{aligned}$$

where the last equation holds because of the consistency of $\hat{\theta}^*_N$ and Lemma A.6, using $\mathbf{Y} = \mathbf{A}^{-1}_N(\eta + \mathbf{B}^{-1}_N\mathbf{V})$. Hence, we can conclude that the terms of $\Gamma^*_N(\theta_0)$ linear in ϕ_0 can be consistently estimated by simply replacing ϕ_0 with

$\hat{\phi}^*_N$.

The only term quadratic in ϕ_0 is contained in $\Gamma^*_{\lambda\lambda}(\theta_0)$, $\frac{1}{N_1\sigma^2_{v0}}\phi_0'\mathbb{D}'\mathcal{P}_2'\mathcal{P}_2\mathbb{D}\phi_0$. Its plug-in estimator is $\frac{1}{N_1\hat{\sigma}^{*2}_{v,N}}\hat{\phi}^{*\prime}_N\mathbb{D}'(\hat{\rho}^*_N)\mathcal{P}_2'(\hat{\delta}^*_N)\mathcal{P}_2(\hat{\delta}^*_N)\mathbb{D}(\hat{\rho}^*_N)\hat{\phi}^*_N$. Using (B.9), $\hat{\theta}^*_N - \theta_0 = o_p(1)$ and Lemma A.6, we show that this estimator is biased:

$$
\begin{aligned}
&\tfrac{1}{N_1\hat{\sigma}^{*2}_{v,N}}\hat{\phi}^{*\prime}_N\mathbb{D}'(\hat{\rho}^*_N)\mathcal{P}_2'(\hat{\delta}^*_N)\mathcal{P}_2(\hat{\delta}^*_N)\mathbb{D}(\hat{\rho}^*_N)\hat{\phi}^*_N \\
&= \tfrac{1}{N_1\hat{\sigma}^{*2}_{v,N}}\phi_0'\mathbb{D}'(\hat{\rho}^*_N)\mathcal{P}_2'(\hat{\delta}^*_N)\mathcal{P}_2(\hat{\delta}^*_N)\mathbb{D}(\hat{\rho}^*_N)\phi_0 \\
&\quad + \tfrac{1}{N_1\hat{\sigma}^{*2}_{v,N}}\mathbf{V}'\mathbb{P}_{\mathbb{D}}\mathbf{B}_N^{-1\prime}\mathbf{B}_N'(\hat{\rho}^*_N)\mathcal{P}_2'(\hat{\delta}^*_N)\mathcal{P}_2(\hat{\delta}^*_N)\mathbf{B}_N(\hat{\rho}^*_N)\mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{V} + o_p(1) \\
&= \tfrac{1}{N_1\sigma^2_{v0}}\phi_0'\mathbb{D}'\mathcal{P}_2'\mathcal{P}_2\mathbb{D}\phi_0 + \tfrac{1}{N_1\sigma^2_{v0}}\mathbf{V}'\mathbb{P}_{\mathbb{D}}\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}}\mathbf{V} + o_p(1) \\
&= \tfrac{1}{N_1\sigma^2_{v0}}\phi_0'\mathbb{D}'\mathcal{P}_2'\mathcal{P}_2\mathbb{D}\phi_0 + \tfrac{1}{N_1}\mathtt{tr}[\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}}] + o_p(1).
\end{aligned}
$$

We see that the bias term, $\frac{1}{N_1}\mathtt{tr}[\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}}]$, involves only the common parameters that can be consistently estimated. Thus, a bias correction can easily be made. Define

$$
\text{Bias}^*_{\lambda\lambda}(\delta) = \tfrac{1}{N_1}\mathtt{tr}[\mathcal{P}_2'(\delta)\mathcal{P}_2(\delta)\mathbb{P}_{\mathbb{D}}(\rho)]. \tag{B.10}
$$

This gives the bias matrix of $\Gamma^*_N(\hat{\theta}^*_N)$, which is a matrix of the same dimension as $\Gamma^*_N(\hat{\theta}^*_N)$, and has the sole non-zero element $\text{Bias}^*_{\lambda\lambda}(\delta_0)$ corresponding to the $\Gamma^*_{\lambda\lambda}(\hat{\theta}^*_N)$ component. ■

Proof of Corollary 1.2.

Proof of (i). Note: $\mathbf{V} = \mathbf{B}_N(\mathbf{A}_N\mathbf{Y} - \eta)$, $\tilde{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$ and $\hat{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}(\hat{\rho}^*_N)\mathbf{B}_N(\hat{\rho}^*_N)$ $[\mathbf{A}_N(\hat{\lambda}^*_N)\mathbf{Y} - \mathbf{X}\hat{\beta}^*_N]$ with respective elements $\{v_j\}$, $\{\tilde{v}_j\}$ and $\{\hat{v}_j\}$, and $\mathbb{Q}_{\mathbb{D}}$ has elements $\{q_{jh}\}$, $j, h = 1, \ldots, N$, where j and h are the combined indices for $i = 1, \ldots, n_t$ and $t = 1, \ldots, T$.

Consistency of $\hat{\gamma}_N$. As $\hat{\sigma}^*_{v,N} - \sigma_{v0} = o_p(1)$ and $\hat{\rho}^*_N - \rho_0 = o_p(1)$, the denominators of $\hat{\gamma}_N$ and γ agree asymptotically. Thus, $\hat{\gamma}_N$ is consistent if $\frac{1}{N}\sum_{j=1}^N[\hat{v}_j^3 - \mathrm{E}(\tilde{v}_j^3)] \xrightarrow{p} 0$, or

$$
(a)\ \tfrac{1}{N}\textstyle\sum_{j=1}^N[\tilde{v}_j^3 - \mathrm{E}(\tilde{v}_j^3)] \xrightarrow{p} 0, \text{ and } \quad (b)\ \tfrac{1}{N}\textstyle\sum_{j=1}^N(\hat{v}_j^3 - \tilde{v}_j^3) \xrightarrow{p} 0.
$$

To prove (a), note that $\tilde{v}_j = \sum_{h=1}^N q_{jh} v_h$. Thus, we have,

$$\begin{aligned}
&\tfrac{1}{N}\textstyle\sum_{j=1}^N[\tilde{v}_j^3 - \mathrm{E}(\tilde{v}_j^3)] \\
&= \tfrac{1}{N}\textstyle\sum_{j=1}^N\sum_{h=1}^N q_{jh}^3[v_h^3 - \mathrm{E}(v_h^3)] + \tfrac{3}{N}\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2 q_{jm} v_l^2 v_m \\
&\quad + \tfrac{6}{N}\textstyle\sum_{j=1}^N\sum_{m=1}^N\sum_{\substack{l\neq m\\ l=1}}^N\sum_{\substack{h\neq m,l\\ h=1}}^N q_{jm}q_{jl}q_{jh}v_m v_l v_h \equiv K_1 + K_2 + K_3.
\end{aligned}$$

First, consider K_1 term. By Lemma A.3, $\mathbb{Q}_{\mathbb{D}}$ is uniformly bounded in both row and column sums. This implies that the elements of $\mathbb{Q}_{\mathbb{D}}$ are uniformly bounded. Therefore, there exists a constant $\bar{q}$ such that $|q_{jh}| \leq \bar{q}$ for all j and h. Given these, we have $\sum_{j=1}^N q_{jh}^3 \leq \sum_{j=1}^N |q_{jh}|^3 \leq \bar{q}^2 \sum_{j=1}^N |q_{jh}| < \infty$. Also note $\{v_i\}$ are iid by Assumption A. Thus, Khinchine's weak law of large number (WLLN) (Feller, 1968, pp. 243-244) implies that K_1 converges to zero in probability as sample size increases.

For the other two terms, we have by switching the order of summations when needed,

$$\begin{aligned}
K_2 &= \tfrac{3}{N}\textstyle\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2 q_{jm}(v_l^2 - \sigma_v^2)v_m + \tfrac{3}{N}\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2 q_{jm}\sigma_v^2 v_m, \\
&= \tfrac{3}{N}\textstyle\sum_{m=1}^N(v_m^2 - \sigma_v^2)(\sum_{j=1}^N\sum_{l=1}^{m-1} q_{jm}^2 q_{jl} v_l) + \tfrac{3}{N}\sum_{m=1}^N v_m[\sum_{j=1}^N\sum_{l=1}^{m-1} q_{jl}^2 q_{jm} \\
&\quad (v_l^2 - \sigma_v^2)] + \tfrac{3}{N}\textstyle\sum_{m=1}^N\sum_{j=1}^N\sum_{\substack{l\neq m\\ l=1}}^N q_{jl}^2 q_{jm}\sigma_v^2 v_m, \\
K_3 &= \tfrac{18}{N}\textstyle\sum_{m=1}^N v_m(\sum_{j=1}^N\sum_{l=1}^{m-1}\sum_{\substack{h\neq l\\ h=1}}^{m-1} q_{jm}q_{jl}q_{jh}v_l v_h) \equiv \tfrac{1}{N}\sum_{m=1}^N g_{4,m}.
\end{aligned}$$

Therefore, we have $K_2 = \frac{1}{N}\sum_{m=1}^N(g_{1,m} + g_{2,m} + g_{3,m})$ and $K_3 = \frac{1}{N}\sum_{m=1}^N g_{4,m}$, where

$$\begin{aligned}
g_{1,m} &= 3(v_m^2 - \sigma_v^2)\textstyle\sum_{j=1}^N\sum_{l=1}^{m-1} q_{jm}^2 q_{jl} v_l, \\
g_{2,m} &= 3v_m\textstyle\sum_{j=1}^N\sum_{l=1}^{m-1} q_{jl}^2 q_{jm}(v_l^2 - \sigma_v^2), \\
g_{3,m} &= 3\textstyle\sum_{j=1}^N\sum_{\substack{l\neq m\\ l=1}}^N q_{jl}^2 q_{jm}\sigma_v^2 v_m, \\
g_{4,m} &= v_m\textstyle\sum_{j=1}^N\sum_{l=1}^{m-1}\sum_{\substack{h\neq l\\ h=1}}^{m-1} q_{jm}q_{jl}q_{jh}v_l v_h.
\end{aligned}$$

Let $\{\mathcal{G}_m\}$ be the increasing sequence of σ-fields generated by $(v_1, \cdots, v_j, j = 1, \cdots, m)$, $m = 1, \cdots, N$. Then, $\mathrm{E}[(g_{1,m}, g_{2,m}, g_{3,m}, g_{4,m})|\mathcal{G}_{m-1}] = 0$; hence, $\{(g_{1,m}, g_{2,m}, g_{3,m}, g_{4,m})', \mathcal{G}_m\}$ form a vector martingale difference (M.D.) sequence. As $\mathbb{Q}_{\mathbb{D}}$ is bounded in row and column sum norms, by Assumption A, it is easy to see that $\mathrm{E}|g_{s,m}|^{1+\epsilon} < \infty$, for $s = 1, 2, 3, 4$ and $\epsilon > 0$. Hence, $\{g_{1,m}\}$, $\{g_{2,m}\}$, $\{g_{3,m}\}$ and $\{g_{4,m}\}$ are uniformly integrable, and the WLLN of

Davidson (1994, Theorem 19.7) applies to give $K_2 \xrightarrow{p} 0$ and $K_3 \xrightarrow{p} 0$.

To prove (b), using the notation $\tilde{\mathbf{V}}(\xi) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$ in (1.5) where $\xi = (\beta', \delta')'$, we have $\tilde{\mathbf{V}} = \tilde{\mathbf{V}}(\xi_0)$ and $\hat{\mathbf{V}} = \tilde{\mathbf{V}}(\hat{\xi}_N^*)$. Let $\mathbf{S}(\xi) = \frac{\partial}{\partial \xi'}\tilde{\mathbf{V}}(\xi)$, we have

$$\mathbf{S}(\xi) = \{-\mathbb{X}(\rho),\ \ -\mathbb{Y}(\rho),\ \ [\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho) - \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{M}][\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]\},$$

where expressions of $\mathbb{Y}(\rho)$ and $\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)$ are in (B.1) and (B.4), respectively. Let $s_j'(\xi)$ be the jth row of $\mathbf{S}(\xi)$. We have by the MVT, for each $j = 1, 2, \ldots, N$,

$$\hat{v}_j \equiv \tilde{v}_j(\hat{\xi}_N^*) = \tilde{v}_j(\xi_0) + s_j'(\bar{\xi})(\hat{\xi}_N^* - \xi_0) = \tilde{v}_j + \psi_j'(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|), \quad \text{(B.11)}$$

where $\bar{\xi}$ lies between $\hat{\xi}_N^*$ and ξ_0, and $\psi_j' = \text{plim}_{N\to\infty} s_j'(\bar{\xi})$, which is easily shown to be $O_p(1)$ as follow. Consider the first k (the number of regressors) elements of ψ_j' first. They are the limits of the jth row of $-\mathbb{X}(\bar{\rho})$, which are just the jth row of $-\mathbb{X}$ because $\bar{\rho} \xrightarrow{p} \rho_0$, implied by $\hat{\rho}_N^* - \rho_0 = o_p(1)$. Hence, we conclude that the first k elements of ψ_j' are $O(1)$, for each $j = 1, 2, \ldots, N$. For the remaining two elements in each ψ_j', they are the limits of elements from the last two columns of $\mathbf{S}(\bar{\xi})$. It is easy to see the limits of the last two columns of $\mathbf{S}(\bar{\xi})$ are just $-\mathbb{Y}$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}][\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0]$. Using $\mathbf{Y} = \mathbf{A}_N^{-1}\eta + \mathbf{C}_N^{-1}\mathbf{V}$, we have $-\mathbb{Y} = \mathcal{P}_2\mathbf{B}_N\eta + \mathcal{P}_2\mathbf{V}$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}][\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0] = [\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{D}\phi_0 + [\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{B}_N^{-1}\mathbf{V}$. By Lemma A.1, we have the elements of $\mathcal{P}_2\mathbf{B}_N\eta$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{D}\phi_0$ are uniformly bounded, and $\mathcal{P}_2$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{B}_N^{-1}$ are uniformly bounded in both row and column sum norms. Hence, it is easy to see each element of $-\mathbb{Y}$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}][\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0]$ are $O_p(1)$, i.e., the last two elements in ψ_j' are also $O_p(1)$, for each $j = 1, 2, \ldots, N$.

As $\tilde{v}_j = O_p(1)$, $\psi_j' = O_p(1)$ and $\hat{\xi}_N^* - \xi_0 = O_p(\frac{1}{\sqrt{N_1}})$, we have by (B.11), $\hat{v}_j^3 = \tilde{v}_j^3 + 3\tilde{v}_j^2\psi_j'(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|)$. It follows that

$$\begin{aligned}\tfrac{1}{N}\textstyle\sum_{j=1}^N(\hat{v}_j^3 - \tilde{v}_j^3) &= \tfrac{3}{N}\textstyle\sum_{j=1}^N \tilde{v}_j^2\psi_j'(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|)\\ &= \tfrac{3\sigma_v^2}{N}\textstyle\sum_{j=1}^N(\sum_{k=1}^N q_{jk}^2\psi_j')(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|) = o_p(1),\end{aligned}$$

as $\frac{1}{N}\sum_{j=1}^N(\sum_{k=1}^N q_{jk}^2\psi_j') = (\sum_{k=1}^N q_{jk}^2)\frac{1}{N}(\sum_{j=1}^N \psi_j') = O(1)$.

Consistency of $\hat{\kappa}_N$. As $\hat{\sigma}_{v,N}^* - \sigma_{v0} = o_p(1)$ and $\hat{\rho}_N^* - \rho_0 = o_p(1)$, the result follows if $\frac{1}{N}\sum_{j=1}^N[\hat{v}_j^4 - \mathrm{E}(\tilde{v}_j^4)] \xrightarrow{p} 0$. This amounts to show that

$$(c)\ \tfrac{1}{N}\sum_{j=1}^{N}[\tilde{v}_j^4 - \mathrm{E}(\tilde{v}_j^4)] \xrightarrow{p} 0 \ \text{ and } \ (d)\ \tfrac{1}{N}\sum_{j=1}^{N}(\hat{v}_j^4 - \tilde{v}_j^4) \xrightarrow{p} 0.$$

To prove (c), we have

$$\begin{aligned}
&\tfrac{1}{N}\textstyle\sum_{j=1}^{N}\tilde{v}_j^4 - \tfrac{1}{N}\sum_{j=1}^{N}\mathrm{E}(\tilde{v}_j^4)\\
=&\tfrac{1}{N}\textstyle\sum_{j=1}^{N}\sum_{h=1}^{N} q_{jh}^4[v_h^4 - \mathrm{E}(v_h^4)] + \tfrac{3}{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{\substack{m\neq l\\ m=1}}^{N} q_{jl}^2 q_{jm}^2 (v_l^2 v_m^2 - \sigma_v^4)\\
&+\tfrac{4}{N}\textstyle\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{\substack{m\neq l\\ m=1}}^{N} q_{jl}^3 q_{jm} v_l^3 v_m + \tfrac{6}{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{\substack{m\neq l\\ m=1}}^{N}\sum_{\substack{h\neq m,l\\ m=1}}^{N} q_{jl}^2 q_{jm} q_{jh} v_l^2 v_m v_h\\
&+\tfrac{1}{N}\textstyle\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{\substack{m\neq l\\ m=1}}^{N}\sum_{\substack{h\neq m,l\\ m=1}}^{N}\sum_{\substack{p\neq m,l,h\\ m=1}}^{N} q_{jl} q_{jm} q_{jh} q_{jp} v_l v_m v_h v_p \equiv \sum_{r=1}^{5} R_r.
\end{aligned}$$

By using WLLN of Davidson (1994, Theorem 19.7) for M.D. arrays as in the proof of (a), we have $R_r = o_p(1)$ for $r = 1, 3, 4, 5$. For R_2, noting that $v_l^2 v_m^2 - \sigma_v^4 = (v_l^2 - \sigma_v^2)(v_m^2 - \sigma_v^2) + \sigma_v^2(v_m^2 - \sigma_v^2) + \sigma_v^2(v_l^2 - \sigma_v^2)$, we have

$$\begin{aligned}
R_2 =& \tfrac{6}{N}\textstyle\sum_{l=1}^{N}(v_l^2 - \sigma_v^2)[\sum_{j=1}^{N}\sum_{m=1}^{l-1} q_{jl}^2 q_{jm}^2 (v_m^2 - \sigma_v^2)]\\
&+\tfrac{6}{N}\textstyle\sum_{l=1}^{N}[\sum_{j=1}^{N}\sum_{\substack{m\neq l\\ m=1}}^{N} q_{jl}^2 q_{jm}^2 \sigma_v^2 (v_l^2 - \sigma_v^2)] \equiv \tfrac{6}{N}\sum_{l=1}^{N}(f_l + f_{2,l}).
\end{aligned}$$

Since $\mathrm{E}[f_l|\mathcal{G}_{l-1}] = 0$ and $\{f_{2,l}\}$ are independent, it is easy to see they both form an M.D. sequence. In addition, it is easily seen that $\mathrm{E}|f_{s,l}|^{1+\epsilon} < \infty$, for $s = 1, 2$ and $\epsilon > 0$, so that $\{f_l\}$ and $\{f_{2,l}\}$ are uniformly integrable. Therefore, the WLLN of Davidson (1994, Theorem 19.7) also implies that $\frac{6}{N}\sum_{l=1}^{N} f_l = o_p(1)$ and $\frac{6}{N}\sum_{l=1}^{N} f_{2,l} = o_p(1)$.

To prove (d), we have by (B.11) $\hat{v}_j^4 = \tilde{v}_j^4 + 4\tilde{v}_j^3\psi_j'(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|)$. It follows that

$$\begin{aligned}
\tfrac{1}{N}\textstyle\sum_{j=1}^{N}(\hat{v}_j^4 - \tilde{v}_j^4) &= \tfrac{4}{N}\textstyle\sum_{j=1}^{N}\tilde{v}_j^3\psi_j'(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|)\\
&= \tfrac{4\sigma_v^3\gamma}{N}\textstyle\sum_{j=1}^{N}(\sum_{k=1}^{N} q_{jk}^3\psi_j')(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|) = o_p(1).
\end{aligned}$$

Proof of (ii). The consistency of $\widehat{\Sigma}_N^*$ to $\Sigma_N^*(\theta_0)$ can be shown similarly as what we do in the proof of Theorem 1.2 for results (b) and (c). For $\widehat{\Gamma}_N^* - \Gamma_N^*(\theta_0) \xrightarrow{p} 0$, we only need to show that $\mathrm{Bias}^*(\hat{\delta}_N^*) - \mathrm{Bias}^*(\delta_0) = o_p(1)$, based on Corollary 1.1. That is to show

$$\tfrac{1}{N_1}\{\mathtt{tr}[\mathcal{P}_2'(\hat{\delta}_N^*)\mathcal{P}_2(\hat{\delta}_N^*)\mathbb{P}_{\mathbb{D}}(\hat{\rho}_N^*)] - \mathtt{tr}(\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}})\} = o_p(1),$$

which can be easily proved by using the MVT as we do for $\frac{1}{N_1}[H_{\lambda\lambda}^{*\mathtt{NS}}(\bar{\delta}) - H_{\lambda\lambda}^{*\mathtt{NS}}(\delta_0)]$ in the proof of Theorem 1.2 (b). ∎

Appendix C: Proofs for Section 1.3

The validity of assumptions on $\Pi_N(\rho)$ in Lemma 1.1 under a balanced panel:

Following the first part in Appendix B, we have $\mathbb{Q}_{\mathbb{D}}(\rho) = (I_T - \frac{l_T l_T'}{T}) \otimes (I_n - \frac{l_n l_n'}{n})$, where $\otimes$ denotes the Kronecker product. Thus, $[\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$ exists if $T > 2$ by Schur product theorem. Further, $|(\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho))_{ii}| - \sum_{j \neq i} |(\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho))_{ij}| = \frac{(n-1)(T-1)[(n-2)(T-2)-2]}{n^2 T^2} > c > 0, \forall i,\ T > 2$. As $\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)$ is symmetric, we conclude it is strictly diagonally dominant in both rows and columns. Hence, Theorem 1 and Corollary 1 of Varah (1975) imply that $\|\Pi_N(\rho)\|_1$ and $\|\Pi_N(\rho)\|_\infty$ are both bounded.

Derivation of the Hessian matrix for robust AQS functions:

With the set of robust AQS functions $S_N^\diamond(\xi)$ given in (1.21), we obtain the components of the Hessian matrix $H_N^\diamond(\xi) = \frac{\partial}{\partial \xi'} S_N^\diamond(\xi)$:

$$
\begin{aligned}
H^\diamond_{\beta\beta}(\xi) &= -\mathbb{X}'(\rho)\mathbb{X}(\rho), \qquad H^\diamond_{\beta\lambda}(\xi) = -\mathbb{X}'(\rho)\mathbb{Y}(\rho),\\
H^\diamond_{\beta\rho}(\xi) &= -\mathbb{X}'(\rho)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta), \quad H^\diamond_{\lambda\beta}(\xi) = -\mathbf{Y}'\mathbf{C}'_N(\delta)\mathbb{L}'_\lambda(\delta)\mathbb{X}(\rho),\\
H^\diamond_{\lambda\lambda}(\xi) &= -\mathbb{Y}'(\rho)\mathbb{Y}(\rho) + \mathbf{Y}'\mathbf{W}'\mathbf{B}'_N(\rho)\bar{\mathbb{F}}'_N(\delta)\tilde{\mathbf{V}}(\beta,\delta)\\
&\quad -\mathbf{Y}'\mathbf{C}'_N(\delta)[\bar{\mathbb{F}}'_{N\lambda}(\delta)\tilde{\mathbf{V}}(\beta,\delta) - \bar{\mathbb{F}}'_N(\delta)\mathbb{Y}(\rho)],\\
H^\diamond_{\lambda\rho}(\xi) &= -\mathbb{Y}'(\rho)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathbf{Y}'\mathbf{C}'_N(\delta)[-\mathbf{G}'_N(\delta)\bar{\mathbb{F}}'_N(\delta)\\
&\quad +\bar{\mathbb{F}}'_{N\rho}(\delta) + \bar{\mathbb{F}}'_N(\delta)\mathbb{G}_N(\rho)]\tilde{\mathbf{V}}(\beta,\delta),\\
H^\diamond_{\rho\beta}(\xi) &= -[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}'_N(\rho)\mathbb{L}'_\rho(\rho)\mathbb{X}(\rho) - \tilde{\mathbf{V}}'(\beta,\delta)\mathbb{L}_\rho(\rho)\mathbf{B}_N(\rho)\mathbf{X},\\
H^\diamond_{\rho\lambda}(\xi) &= -\mathbb{Y}'(\rho)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) + \mathbf{Y}'\mathbf{W}'\mathbf{B}'_N(\rho)\bar{\mathbb{G}}_N(\rho)\tilde{\mathbf{V}}(\beta,\delta)\\
&\quad +[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}'_N(\rho)\bar{\mathbb{G}}_N(\rho)\mathbb{Y}(\rho),\\
H^\diamond_{\rho\rho}(\xi) &= \tilde{\mathbf{V}}'(\beta,\delta)\mathcal{R}_{1N}(\rho)\tilde{\mathbf{V}}(\beta,\delta) - [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}'_N(\rho)[-\mathbf{G}'_N(\delta)\bar{\mathbb{G}}_N(\rho)\\
&\quad +\bar{\mathbb{G}}_{N\rho}(\rho) + \bar{\mathbb{G}}_N(\rho)\mathbb{G}_N(\rho)]\tilde{\mathbf{V}}(\beta,\delta),
\end{aligned} \tag{C.1}
$$

where $\bar{\mathbb{F}}'_{N\lambda}(\delta) = \mathtt{diag}[\mathbf{B}_N^{-1\prime}(\rho)\mathbf{F}_N'^2(\lambda)\mathbf{B}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$,
$\bar{\mathbb{F}}'_{N\rho}(\delta) = \mathtt{diag}[\mathcal{K}_{1N}(\delta)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1} - \bar{\mathbb{F}}'_N(\delta)\mathtt{diag}[\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$,
$\mathcal{K}_{1N}(\delta) = \bar{\mathbf{F}}'_N(\delta)\mathbf{G}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho) - \bar{\mathbf{F}}'_N(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}'_N(\rho) + \bar{\mathbf{F}}'_N(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_N(\rho)\mathbb{P}_{\mathbb{D}}(\rho)$,
$\bar{\mathbb{G}}_{N\rho}(\rho) = \mathtt{diag}[\mathcal{K}_{2N}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1} - \bar{\mathbb{G}}_N(\rho)\mathtt{diag}[\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$,
$\mathcal{K}_{2N}(\rho) = [\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)\mathbb{P}_{\mathbb{D}}(\rho) + \mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)]\mathbf{G}^\circ_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho) + \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}^2_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)$.

Proof of Theorem 1.3. Since the consistency of $\hat{\beta}_N^{\diamond}$ follows almost immediately that of $\hat{\delta}_N^{\diamond}$ under Assumptions C and E, we only need to prove that $\hat{\delta}_N^{\diamond}$ is consistent to δ_0. By theorem 5.9 of van der Vaart (1998), $\hat{\delta}_N^{\diamond}$ will be consistent for δ_0 if $\sup_{\delta\in\Delta}\frac{1}{N_1}\left\|S_N^{\diamond c}(\delta)-\bar{S}_N^{\diamond c}(\delta)\right\| \xrightarrow{p} 0$.

Let $\mathbb{L}_\lambda(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)[\bar{\mathbf{F}}_N(\delta) - \bar{\mathbb{F}}_N(\delta)]$, $\mathbb{L}_\rho(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)[\bar{\mathbf{G}}'_N(\rho) - \bar{\mathbb{G}}'_N(\rho)]$ and $\mathbb{N}_N(\rho) = I_N - \mathbb{M}_N(\rho)$. Note that $\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^{\diamond}(\delta)] = \mathbb{M}_N(\rho)\mathbf{C}_N(\delta)\mathbf{Y}$ and $\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\bar{\beta}_N^{\diamond}(\delta)] = \mathbb{M}_N(\rho)\mathbf{C}_N(\delta)\mathbf{Y} + \mathbb{N}_N(\rho)\mathbf{C}_N(\delta)[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]$. Recall $\hat{\mathbf{V}}(\delta) = \mathcal{Q}_N(\delta)\mathbf{Y}$ and $\bar{\mathbf{V}}(\delta) = \mathcal{Q}_N(\delta)\mathbf{Y} + \mathcal{P}_N(\delta)[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]$. With Assumption G$'$, the consistency of $\hat{\delta}_N^{\diamond}$ follows if:

(*i*) $\sup_{\delta\in\Delta}\frac{1}{N_1}\left|\mathbf{Y}'\mathbf{Q}_r^h(\delta)\mathbf{Y} - \mathrm{E}[\mathbf{Y}'\mathbf{Q}_r^h(\delta)\mathbf{Y}]\right| = o_p(1)$, for $r = 1, 2$;

(*ii*) $\sup_{\delta\in\Delta}\frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}_s^h(\delta)\mathbf{C}_N^{-1}] = o(1)$, for $s = 1, 2, 3$;

where $\mathbf{Q}_1^h(\delta) = \mathbf{C}'_N(\delta)\mathbb{L}'_\lambda(\delta)\mathcal{Q}_N(\delta)$, $\mathbf{Q}_2^h(\delta) = \mathbf{C}'_N(\delta)\mathbb{M}'_N(\rho)\mathbb{L}'_\rho(\rho)\mathcal{Q}_N(\delta)$, $\mathbf{P}_1^h(\delta) = \mathbf{C}'_N(\delta)\mathbb{L}'_\lambda(\delta)\mathcal{P}_N(\delta)$, $\mathbf{P}_2^h(\delta) = \mathbf{C}'_N(\delta)\mathbb{L}'_\rho(\rho)\mathcal{P}_N(\delta)$ and $\mathbf{P}_3^h(\delta) = \mathbf{C}'_N(\delta)\mathbb{N}'_N(\rho)\mathbb{L}'_\rho(\rho)\mathcal{Q}_N(\delta)$.

Note that $\mathbf{Q}_1^h(\delta) = \mathbf{C}'_N(\delta)[\bar{\mathbf{F}}'_N(\delta) - \bar{\mathbb{F}}'_N(\delta)]\mathcal{Q}_N(\delta) = \mathbf{W}'\mathbf{B}'_N(\rho)\mathcal{Q}_N(\delta) - \mathbf{C}'_N(\delta)\bar{\mathbb{F}}'_N(\delta)\mathcal{Q}_N(\delta)$. As $\bar{\mathbb{F}}'_N(\delta)$ is a diagonal matrix which is naturally bounded in both row and column sums, uniformly in $\delta\in\Delta$, we conclude $\mathbf{Q}_1^h(\delta)$ is bounded in both row and column sum norms, uniformly in $\delta\in\Delta$, by Lemma A.1. Similarly, $\mathbf{Q}_2^h(\delta) = \mathbf{C}'_N(\delta)\mathbb{M}'_N(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\mathcal{Q}_N(\delta) = \bar{\mathbf{Q}}(\delta) - \mathbf{C}'_N(\delta)$ $\mathbb{M}'_N(\rho)\bar{\mathbb{G}}_N(\rho)\mathcal{Q}_N(\delta)$ is also bounded in both row and column sum norms, uniformly in $\delta\in\Delta$. Hence, $\mathbf{Q}_1^h(\delta)$ and $\mathbf{Q}_2^h(\delta)$ have forms similar to $\mathbf{Q}(\delta)$. The proof of (*i*) thus follows that of Theorem 1.1 **(b)**. For (*ii*), noting that $\mathcal{P}_N(\delta) = \mathbb{P}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$, we have $\sup_{\delta\in\Delta}\frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}_s^h(\delta)\mathbf{C}_N^{-1}] = o(1), s = 1, 2$, by Lemma A.4. For the final result, we have,

$$\begin{aligned}\tfrac{1}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}_3^h(\delta)\mathbf{C}_N^{-1}] &= -\tfrac{1}{N_1}\mathtt{tr}[\mathbf{C}'_N(\delta)\mathbb{N}'_N(\rho)\mathbb{L}'_\rho(\rho)\mathcal{Q}_N(\delta)\mathrm{Var}(\mathbf{Y})] \\ &= -\tfrac{1}{N_1}\mathtt{tr}\big[(\tfrac{1}{N_1}\mathbb{X}'(\rho)\mathbb{X}(\rho))^{-1}(\tfrac{1}{N_1}\mathbf{X}'\mathbf{B}'_N(\rho)\mathbb{L}'_\rho(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_N^h(\delta)\mathbb{X}(\rho))\big].\end{aligned}$$

Assumption C implies that the elements of $[\frac{1}{N_1}\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}$ are uniformly bounded for large enough N, uniformly in $\rho\in\Delta_\rho$. Lemma A.1 and Lemma A.3 together imply the term between $\mathbf{X}'$ and $\mathbb{X}(\rho)$ are uniformly bounded in both row and column sums, uniformly in $\delta\in\Delta$. Hence, the elements of the second part in the trace are also uniformly bounded. As the number of regressors k is finite, the quantity $\frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}_3^h(\delta)\mathbf{C}_N^{-1}]$ will shrink to zero as N goes large, uniformly in $\delta \in \Delta$. These complete the proof of the theorem. ∎

Proof of Theorem 1.4. Applying the MVT on each row of $S_N^\diamond(\hat{\xi}_N^\diamond)$, we have,

$$0 = \frac{1}{\sqrt{N_1}} S_N^\diamond(\hat{\xi}_N^\diamond) = \frac{1}{\sqrt{N_1}} S_N^\diamond(\xi_0) + \left[\frac{1}{N_1}\frac{\partial}{\partial \xi'} S_N^\diamond(\xi)\Big|_{\xi=\bar{\xi}_r \text{ in } r\text{th row}}\right]\sqrt{N_1}(\hat{\xi}_N^\diamond - \xi_0),$$

where $\{\bar{\xi}_r\}$ are on the line segment between $\hat{\xi}_N^\diamond$ and ξ_0. The result of the theorem follows if

(*a*) $\frac{1}{\sqrt{N_1}} S_N^\diamond(\xi_0) \xrightarrow{D} N[0,\ \lim_{N\to\infty}\Gamma_N^\diamond(\xi_0)]$,

(*b*) $\frac{1}{N_1}\left[\frac{\partial}{\partial \xi'} S_N^\diamond(\xi)\big|_{\xi=\bar{\xi}_r \text{ in } r\text{th row}} - \frac{\partial}{\partial \xi'} S_N^\diamond(\xi_0)\right] = o_p(1)$, and

(*c*) $\frac{1}{N_1}\left[\frac{\partial}{\partial \xi'} S_N^\diamond(\xi_0) - \mathrm{E}(\frac{\partial}{\partial \xi'} S_N^\diamond(\xi_0))\right] = o_p(1)$.

Proof of (a). From (1.24), we see that the elements of $S_N^\diamond(\xi_0)$ are linear-quadratic forms in $\mathbf{V}$. Thus, for every non-zero $(k+2)\times 1$ vector of constants a, $a'S_N^\diamond(\xi_0)$ has form:

$$a'S_N^\diamond(\xi_0) = b_N'\mathbf{V} + \mathbf{V}'\Phi_N\mathbf{V} - \sigma_v^2\mathtt{tr}(\Phi_N),$$

for suitably defined non-stochastic vector b_N and matrix Φ_N. Again, by Assumptions A-F it is easy to verify that b_N and matrix Φ_N satisfy the conditions of the CLT for LQ form of Kelejian and Prucha (2001), and hence the asymptotic normality of $\frac{1}{\sqrt{N_1}}a'S_N^\diamond(\xi_0)$ follows. By Cramér-Wold device, $\frac{1}{\sqrt{N_1}} S_N^\diamond(\xi_0) \xrightarrow{D} N[0, \lim_{N\to\infty}\Gamma_N^\diamond(\theta_0)]$, where $\Gamma_N^\diamond(\theta_0)$ is given in (1.25).

Proof of (b). The Hessian matrix $H_N^\diamond(\xi) = \frac{\partial}{\partial \xi'} S_N^\diamond(\xi)$ is given in (C.1). As $\bar{\mathbb{F}}_{N\lambda}'(\delta_0)$, $\bar{\mathbb{F}}_{N\rho}'(\delta_0)$ and $\bar{\mathbb{G}}_{N\rho}(\rho_0)$ are diagonal matrices with uniformly bounded elements, it is easy to see that $\frac{1}{N_1}H_N^\diamond(\xi_0) = O_p(1)$ by Lemma A.6, and hence, $\frac{1}{N_1}H_N^\diamond(\bar{\xi}) = O_p(1)$. Here again for ease of exposition we simply use $H_N^\diamond(\bar{\xi})$ to denote $\frac{\partial}{\partial \xi'} S_N^\diamond(\xi)\big|_{\xi=\bar{\xi}_r \text{ in } r\text{th row}}$. As $H_N^\diamond(\bar{\xi})$ is linear or quadratic in $\bar{\beta}$ and nonlinear in $\bar{\delta}$, we have by applying the MVT on the $\bar{\delta}$-components:

$$\frac{1}{N_1}H_N^\diamond(\bar{\xi}) - \frac{1}{N_1}H_N^\diamond(\xi_0) = \frac{1}{N_1}\frac{\partial}{\partial \delta'}H_N^\diamond(\bar{\beta},\dot{\delta})(\bar{\delta}-\delta_0) + \frac{1}{N_1}[H_N^\diamond(\bar{\beta},\delta_0) - H_N^\diamond(\theta_0)].$$

Similar to the proof of Theorem 2.2 (*b*), we show that $\frac{1}{N_1}\frac{\partial}{\partial \delta'}H_N^\diamond(\bar{\beta},\dot{\delta}) = O_p(1)$. The second term is seen to contain elements either linear or quadratic in $\bar{\beta}-\beta_0$

with the matrices in the linear or quadratic terms being $O_p(1)$. Hence, the desired result follows as $\bar{\xi} - \xi_0 = o_p(1)$.

Proof of (c). Since $\mathbf{Y} = \mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$, all components of $H_N^{\diamond}(\xi_0)$ are linear or quadratic in $\mathbf{V}$. Thus, under the assumptions of the theorem the result **(c)** is proved using Lemma A.6. We provide details of the proof using the most complicate term, $H_{\rho\rho}^{\diamond}(\xi_0)$. Let $\Xi_N = -\mathbf{G}_N'\bar{\mathbb{G}}_N + \bar{\mathbb{G}}_{N\rho} + \bar{\mathbb{G}}_N\mathbb{G}_N$. By Lemma A.1, it is easy to see that Ξ_N is uniformly bounded in both row and column sums in absolute value. Hence, we have

$$
\begin{aligned}
&\tfrac{1}{N_1}[H_{\rho_0\rho_0}^{\diamond}(\xi_0) - \mathrm{E}(H_{\rho_0\rho_0}^{\diamond}(\xi_0))]\\
&= \tfrac{1}{N_1}[\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathrm{E}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V})] - \tfrac{1}{N_1}(\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0)'\mathbf{B}_N'\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V}\\
&\quad + \tfrac{1}{N_1}\mathrm{E}[(\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0)'\mathbf{B}_N'\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V}]\\
&= \tfrac{1}{N_1}[\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathrm{E}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V})] - \tfrac{1}{N_1}[\phi_0'\mathbb{D}'\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathrm{E}(\phi_0'\mathbb{D}'\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V})]\\
&\quad - \tfrac{1}{N_1}[\mathbf{V}'\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathrm{E}(\mathbf{V}'\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V})] \;=\; o_p(1).
\end{aligned}
$$

The proofs for the other terms are done in a similar manner, and the details are omitted. ■

Proof of Corollary 1.3: Just like the homoskedasiticiy case, plugging $\hat{\phi}_N^{\diamond}$ in $\Gamma_N^{\diamond}(\xi)$ induces a bias for terms quadratic in ϕ, and a bias correction is necessary. From (1.25), we see that the terms of $\Gamma_N^{\diamond}(\xi)$ that are quadratic in ϕ are the (λ, ρ) terms and are of the form: $\phi'\mathbb{D}'(\rho)\mathbb{L}_a'(\delta)\mathbf{H}\mathbb{L}_b(\delta)\mathbb{D}(\rho)\phi,\ a, b = \lambda, \rho$, recalling $\eta = \mathbf{X}\beta_0 + \mathbf{D}\phi_0$ and $\mathbb{D}(\rho)\mathbf{B}_N(\rho)\mathbf{D}$.

By applying the MVT on $\hat{\rho}_N^{\diamond}$-variable in the key quantity $\mathbf{D}\hat{\phi}_N^{\diamond}$, we have after some algebra,

$$
\begin{aligned}
\mathbf{D}\hat{\phi}_N^{\diamond} = \mathbf{D}\phi_0 + \mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{V} - \mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{B}_N[\mathbf{W}\mathbf{Y}(\hat{\lambda}_N^{\diamond} - \lambda_0) + \mathbf{X}(\hat{\beta}_N^{\diamond} - \beta_0)]\\
- \mathbb{R}_N(\dot{\rho})[\mathbf{A}_N(\hat{\lambda}_N^{\diamond})\mathbf{Y} - \mathbf{X}\hat{\beta}_N^{\diamond}](\hat{\rho}_N^{\diamond} - \rho_0),
\end{aligned}
$$

where $\dot{\rho}$ lies between $\hat{\rho}_N^{\diamond}$ and ρ_0. Plugging $\mathbf{D}\hat{\phi}_N^{\diamond}$ and other parameter estimates in these quadratic terms, we have,

$$\begin{aligned}
&\tfrac{1}{N_1}\hat{\phi}_N^{\diamond\prime}\mathbb{D}'(\hat{\rho}_N^{\diamond})\mathbb{L}_a'(\hat{\delta}_N^{\diamond})\mathbf{H}\mathbb{L}_b(\hat{\delta}_N^{\diamond})\mathbb{D}(\hat{\rho}_N^{\diamond})\hat{\phi}_N^{\diamond}\\
&= \tfrac{1}{N_1}\phi_0'\mathbb{D}'(\hat{\rho}_N^{\diamond})\mathbb{L}_a'(\hat{\delta}_N^{\diamond})\mathbf{H}\mathbb{L}_b(\hat{\delta}_N^{\diamond})\mathbb{D}(\hat{\rho}_N^{\diamond})\phi_0\\
&\quad + \tfrac{1}{N_1}\mathbf{V}'\mathbb{P}_{\mathbb{D}}\mathbf{B}_N^{-1\prime}\mathbf{B}_N'(\hat{\rho}_N^{\diamond})\mathbb{L}_a'(\hat{\delta}_N^{\diamond})\mathbf{H}\mathbb{L}_b(\hat{\delta}_N^{\diamond})\mathbf{B}_N(\hat{\rho}_N^{\diamond})\mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{V} + o_p(1)\\
&= \tfrac{1}{N_1}\phi_0'\mathbb{D}'\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{D}\phi_0 + \tfrac{1}{N_1}\mathtt{tr}\big[\mathbf{H}\mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_{\mathbb{D}}\big] + o_p(1),
\end{aligned}$$

Define

$$\text{Bias}_{ab}^{\diamond}(\delta,\mathbf{H}) = \tfrac{1}{N_1}\mathtt{tr}\big[\mathbf{H}\mathbb{P}_{\mathbb{D}}(\rho)\mathbb{L}_a'(\delta)\mathbf{H}\mathbb{L}_b(\delta)\mathbb{P}_{\mathbb{D}}(\rho)\big],$$

for $a, b = \lambda, \rho$. Hence, the bias matrix for $\Gamma_N^{\diamond}(\hat{\xi}_N^{\diamond})$ can be written as

$$\text{Bias}_{\phi}^{\diamond}(\delta_0,\mathbf{H}) = \begin{bmatrix} 0 & 0 & 0\\ 0 & \text{Bias}_{\lambda\lambda}^{\diamond}(\delta_0,\mathbf{H}) & \text{Bias}_{\lambda\rho}^{\diamond}(\delta_0,\mathbf{H})\\ 0 & \text{Bias}_{\rho\lambda}^{\diamond}(\delta_0,\mathbf{H}) & \text{Bias}_{\rho\rho}^{\diamond}(\delta_0,\mathbf{H})\end{bmatrix},$$

leading to the result of Corollary 1.3. ■

Proof of Lemma 1.1: Using $\tilde{\mathbf{V}}(\xi) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$ defined in (1.5), let $\tilde{\mathbf{V}} = \tilde{\mathbf{V}}(\xi_0)$ and $\hat{\mathbf{V}} = \tilde{\mathbf{V}}(\hat{\xi}_N^{\diamond})$ and denote their elements by $\{\tilde{v}_j\}$ and $\{\hat{v}_j\}$, respectively. Following (B.11), we have $\hat{v}_j \equiv \tilde{v}_j(\hat{\xi}_N^{\diamond}) = \tilde{v}_j + \psi_j'(\hat{\xi}_N^{\diamond} - \xi_0) + o_p(\|\hat{\xi}_N^{\diamond} - \xi_0\|)$, and in vector form,

$$\hat{\mathbf{V}} = \tilde{\mathbf{V}} + \Psi_N(\hat{\xi}_N^{\diamond} - \xi_0) + o_p(\|\hat{\xi}_N^{\diamond} - \xi_0\|),$$

where $\Psi_N = (\psi_1, \psi_2, \ldots, \psi_N)'$, with ψ_j being defined below (B.11).

Define $\dot{\Pi}_N(\rho) = \frac{\partial}{\partial\rho}\Pi_N(\rho) = -2\Pi_N(\rho)[\dot{\mathbb{Q}}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)]\Pi_N(\rho)$. It is easy to see that $\|\dot{\Pi}_N(\rho)\|_1$ and $\|\dot{\Pi}_N(\rho)\|_\infty$ are bounded in a neighborhood of ρ_0. Let Π_{jh} and $\dot{\Pi}_{jh}$ be the respective elements of Π_N and $\dot{\Pi}_N$. Hence, we have by the MVT, for each $j, h = 1, 2, \ldots, N$, $\Pi_{jh}(\hat{\rho}_N^{\diamond}) = \Pi_{jh} + \dot{\Pi}_{jh}(\bar{\rho})(\hat{\rho}_N^{\diamond} - \rho_0) = \Pi_{jh} + \dot{\Pi}_{jh}(\hat{\rho}_N^{\diamond} - \rho_0) + o_p(\|\hat{\rho}_N^{\diamond} - \rho_0\|)$, where $\bar{\rho}$ lies between $\hat{\rho}_N^{\diamond}$ and ρ_0. In matrix form, we have

$$\Pi_N(\hat{\rho}_N^{\diamond}) = \Pi_N + \dot{\Pi}_N(\hat{\rho}_N^{\diamond} - \rho_0) + o_p(\|\hat{\rho}_N^{\diamond} - \rho_0\|).$$

Define $\hat{h} = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_N^2)' = \Pi_N(\hat{\rho}_N^{\diamond})(\hat{\mathbf{V}} \odot \hat{\mathbf{V}})$ and $\tilde{h} = \Pi_N(\tilde{\mathbf{V}} \odot \tilde{\mathbf{V}})$. As the elements of $\tilde{\mathbf{V}}$ are $O_p(1)$, rows of Ψ_N are $O_p(1)$, elements of Π_N and $\dot{\Pi}_N$ are

$O(1)$, and $\hat{\xi}_N^* - \xi_0 = O_p(\frac{1}{\sqrt{N_1}})$, we have,

$$\hat{h} = \tilde{h} + 2\Pi_N(\tilde{\mathbf{V}} \odot \Psi_N(\hat{\xi}_N^\diamond - \xi_0)) + \dot{\Pi}_N(\tilde{\mathbf{V}} \odot \tilde{\mathbf{V}})(\hat{\rho}_N^\diamond - \rho_0) + o_p(\|\hat{\xi}_N^\diamond - \xi_0\|). \tag{C.2}$$

Proof of (i). Let $c_N = (c_{11}, \cdots, c_{NN})'$ and $h = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)'$. We have,

$$\tfrac{1}{N}[\mathtt{tr}(\hat{\mathbf{H}}C_N) - \mathtt{tr}(\mathbf{H}C_N)] = \tfrac{1}{N}c_N'(\hat{h} - h) = \tfrac{1}{N}c_N'(\hat{h} - \tilde{h}) + \tfrac{1}{N}c_N'(\tilde{h} - h).$$

The result follows if both terms above are $o_p(1)$. For the first term, we have, using (C.2),

$$\begin{aligned}
&\tfrac{1}{N}\bar{c}_N'(\hat{h} - \tilde{h}) \\
=&\tfrac{2}{N}\bar{c}_N'\Pi_N(\tilde{\mathbf{V}} \odot \Psi_N(\hat{\xi}_N^\diamond - \xi_0)) + \tfrac{1}{N}\bar{c}_N'\dot{\Pi}_N(\tilde{\mathbf{V}} \odot \tilde{\mathbf{V}})(\hat{\rho}_N^\diamond - \rho_0) + o_p(\|\hat{\xi}_N^\diamond - \xi_0\|) \\
=&\tfrac{2}{N}\textstyle\sum_{j=1}^N c_{jj}(\sum_{h=1}^N \Pi_{jh}\tilde{v}_h\psi_h')(\hat{\xi}_N^\diamond - \xi_0) + \tfrac{1}{N}\sum_{j=1}^N c_{jj}(\sum_{h=1}^N \dot{\Pi}_{jh}\sum_{k=1}^N q_{hk}^2\sigma_k^2)(\hat{\rho}_N^\diamond - \rho_0) \\
&+ o_p(\|\hat{\xi}_N^\diamond - \xi_0\|) = o_p(1).
\end{aligned}$$

For the second term, we have after some algebra,

$$\tilde{h} = \Pi_N[(\mathbb{Q}_{\mathbb{D}} \odot \mathbb{Q}_{\mathbb{D}})(\mathbf{V} \odot \mathbf{V}) + \zeta] = \mathbf{V} \odot \mathbf{V} + \Pi_N\varepsilon, \tag{C.3}$$

where ε is an $N \times 1$ vector with j-th element $\varepsilon_j = \sum_{k=1}^N v_k\zeta_{jk}$, where $\zeta_{jk} = 2q_{jk}\sum_{l=1}^{k-1} q_{jl}v_l, k \geq 2$, and $\zeta_{j1} = 0$. As ζ_{jk} is $(v_1, \ldots v_{k-1})$-measurable, $\{v_k\zeta_{jk}\}$ form an M.D. sequence. Thus, each ε_j is a sum of M.D.s. Hence, we have

$$\tfrac{1}{N}\bar{c}_N'(\tilde{h} - h) = \tfrac{1}{N}\bar{c}_N'(\mathbf{V} \odot \mathbf{V} - h) + \tfrac{1}{N}\bar{c}_N'\Pi_N\zeta = o_p(1),$$

where $\frac{1}{N}\bar{c}_N'(\mathbf{V} \odot \mathbf{V} - h) = o_p(1)$ by Lemma A.6(v) and $\frac{1}{N}\bar{c}_N'\Pi_N\zeta = o_p(1)$ by WLLN of Davidson (1994, Theorem 19.7) for M.D. arrays.

Proof of (ii). Note that $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N) = h'(A_N \odot B_N)h$. We have,

$$\begin{aligned}
&\tfrac{1}{N}\mathtt{tr}(\widehat{\mathbf{H}}A_N\widehat{\mathbf{H}}B_N) - \tfrac{1}{N}\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N) = \tfrac{1}{N}\hat{h}'(A_N \odot B_N)\hat{h} - \tfrac{1}{N}h'(A_N \odot B_N)h \\
=&\tfrac{1}{N}(\hat{h}'(A_N \odot B_N)\hat{h} - \tilde{h}'(A_N \odot B_N)\tilde{h}) + \tfrac{1}{N}(\tilde{h}'(A_N \odot B_N)\tilde{h} - h'(A_N \odot B_N)h).
\end{aligned} \tag{C.4}$$

The first term of (C.4) can be written as

$$\tfrac{1}{N}(\hat{h}'(A_N \odot B_N)\hat{h} - \tilde{h}'(A_N \odot B_N)\tilde{h}) = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3,$$

where $\mathcal{T}_1 = \frac{1}{N}(\hat{h} - \tilde{h})'(A_N \odot B_N)(\hat{h} - \tilde{h})$, $\mathcal{T}_2 = \frac{1}{N}(\hat{h} - \tilde{h})'(A_N \odot B_N)\tilde{h}$, and

$\mathcal{T}_3 = \frac{1}{N}(\hat{h} - \tilde{h})'(A_N \odot B_N)'\tilde{h}$. Note that A_N and B_N are uniformly bounded in both row and column sum norms, $A_N \odot B_N$ is also uniformly bounded in both row and column sum norms. Hence, using (C.2), $\tilde{\mathbf{V}} = O_p(1)$, $\Psi_N = O_p(1)$ and $\hat{\xi}_N^* - \xi_0 = O_p(\frac{1}{\sqrt{N_1}})$, we can easily show that $\mathcal{T}_r = o_p(1)$, for $r = 1, 2, 3$, as we show $\frac{1}{N}\vec{c}_N'(\hat{h} - \tilde{h}) = o_p(1)$ in the proof of (i). Thus, the first term in (C.4) is $o_p(1)$.

For the second term in (C.4), we have similarly to the first term,

$$\frac{1}{N}(\tilde{h}'(A_N \odot B_N)\tilde{h} - h'(A_N \odot B_N)h) = \mathcal{T}_4 + \mathcal{T}_5 + \mathcal{T}_6,$$

where $\mathcal{T}_4 = \frac{1}{N}(\tilde{h} - h)'(A_N \odot B_N)(\tilde{h} - h)$, $\mathcal{T}_5 = \frac{1}{N}(\tilde{h} - h)'(A_N \odot B_N)h$ and $\mathcal{T}_6 = \frac{1}{N}(\tilde{h} - h)'(A_N \odot B_N)'h$. For the $\mathcal{T}_5$ term, we have by (C.3),

$$\mathcal{T}_5 = \frac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot B_N)h + \frac{1}{N}\varepsilon'\Pi_N(A_N \odot B_N)h = o_p(1),$$

by Lemma A.6(v) and WLLN for M.D. arrays of Davidson (1994, Theorem 19.7). The $\mathcal{T}_6$ term is similar to $\mathcal{T}_5$ and the result follows, i.e., $\mathcal{T}_6 = o_p(1)$.

Thus, it is left to study the limit of $\mathcal{T}_4$. Again, by (C.3) we have,

$$\begin{aligned}\mathcal{T}_4 = &\frac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot B_N)\Pi_N\varepsilon + \frac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot B_N)'\Pi_N\varepsilon \\ &+ \frac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot B_N)(\mathbf{V} \odot \mathbf{V} - h) + \frac{1}{N}\varepsilon'\Pi_N(A_N \odot B_N)\Pi_N\varepsilon \\ \equiv &\mathcal{T}_{4a} + \mathcal{T}_{4b} + \mathcal{T}_{4c} + \mathcal{T}_{4d}. \end{aligned} \tag{C.5}$$

Consider first the term $\mathcal{T}_{4a}$. Denote $\Omega = (A_N \odot B_N)\Pi_N$ with elements $\{\omega_{jk}\}$. We have,

$$\begin{aligned}\mathcal{T}_{4a} = &\frac{1}{N}\sum_{j=1}^N\sum_{k=1}^N \omega_{jk}\varepsilon_j(v_k^2 - \sigma_k^2) \\ = &\frac{1}{N}\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N \omega_{jk}q_{jl}q_{jm}(v_k^2 - \sigma_k^2)v_l v_m \\ = &\frac{1}{N}\sum_{k=1}^N((v_k^2 - \sigma_k^2)\sum_{j=1}^N\sum_{l=1}^{k-1}\sum_{\substack{m\neq l\\ m=1}}^{k-1} \omega_{jk}q_{jl}q_{jm}v_l v_m) \\ &+ \frac{2}{N}\sum_{l=1}^N(v_l\sum_{j=1}^N\sum_{k=1}^{l-1}\sum_{\substack{m\neq k\\ m=1}}^{l-1} \omega_{jk}q_{jl}q_{jm}v_m(v_k^2 - \sigma_k^2)) \\ &+ \frac{2}{N}\sum_{k=1}^N((v_k^3 - \mathrm{E}v_k^3)\sum_{j=1}^N\sum_{m=1}^{k-1} \omega_{jk}q_{jk}q_{jm}v_m) \\ &+ \frac{2}{N}\sum_{m=1}^N(v_m\sum_{j=1}^N\sum_{k=1}^{m-1} \omega_{jk}q_{jk}q_{jm}(v_k^3 - \mathrm{E}v_k^3)) \\ &+ \frac{2}{N}\sum_{m=1}^N(v_m\sum_{j=1}^N\sum_{\substack{k\neq m\\ k=1}}^N \omega_{jk}q_{jk}q_{jm}(\mathrm{E}v_k^3 - \sigma_k^2)),\end{aligned}$$

which is seen to be the average of M.D. sequence and thus is $o_p(1)$ by Theorem 19.7 of Davidson (1994). Similarly, we show that $\mathcal{T}_{4b} = \frac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot$

$B_N)'\Pi_N\varepsilon = o_p(1)$.

For the term $\mathcal{T}_{4c}$, as $\mathrm{E}(\mathbf{V}\odot\mathbf{V}) = h$, we have $\mathrm{E}(\mathcal{T}_{4c}) = \frac{1}{N}\mathtt{tr}((A_N \odot B_N)\mathrm{Var}(\mathbf{V}\odot\mathbf{V})) = 0$. Thus, Lemma A.6($iv$) implies that $\mathcal{T}_{4c} = \frac{1}{N}(\mathbf{V}\odot\mathbf{V}-h)'(A_N\odot B_N)(\mathbf{V}\odot\mathbf{V}-h) \xrightarrow{p} 0$.

Now, for the last term of (C.5), $\mathcal{T}_{4d} = \frac{1}{N}\varepsilon'\Pi_N(A_N\odot B_N)\Pi_N\varepsilon$, we have by taking the advantage that each element of ε is a sum of an M.D. sequence,

$$\mathrm{E}(\varepsilon\varepsilon') = 2(\mathbb{Q}_{\mathbb{D}}\mathbf{H}\mathbb{Q}_{\mathbb{D}})\odot(\mathbb{Q}_{\mathbb{D}}\mathbf{H}\mathbb{Q}_{\mathbb{D}}) - 2(\mathbb{Q}_{\mathbb{D}}\odot\mathbb{Q}_{\mathbb{D}})\mathbf{H}\mathbf{H}(\mathbb{Q}_{\mathbb{D}}\odot\mathbb{Q}_{\mathbb{D}}). \tag{C.6}$$

This gives,

$$\begin{aligned} &\mathrm{E}(\varepsilon'\Pi_N(A_N\odot B_N)\Pi_N\varepsilon) \qquad\qquad (\mathrm{C.7})\\ &= 2\mathtt{tr}((A_N\odot B_N)\Pi_N\Lambda(\mathbf{H})\Pi_N) - 2\mathtt{tr}((A_N\odot B_N)\mathbf{H}^2),\\ &= 2\mathtt{tr}((A_N\odot B_N)\Pi_N\Lambda(\mathbf{H})\Pi_N),\end{aligned}$$

where $\Lambda(\mathbf{H}) = (\mathbb{Q}_{\mathbb{D}}\mathbf{H}\mathbb{Q}_{\mathbb{D}})\odot(\mathbb{Q}_{\mathbb{D}}\mathbf{H}\mathbb{Q}_{\mathbb{D}})$, and the last equation takes use of the fact that the diagonal elements of A_N and B_N are zero.

Finally, to show that $\mathcal{T}_{4d} - \mathrm{E}(\mathcal{T}_{4d}) = o_p(1)$, denote $\chi_N = \Pi_N(A_N\odot B_N)\Pi_N$ with elements $\{\chi_{jk}\}$. It is easy to show that $\{\chi_{jk}\}$ are uniformly bounded, and let $|\chi_{lm}| \le \bar{\chi} < \infty$. We have,

$$\begin{aligned} &\mathrm{Var}(\varepsilon'\Pi_N(A_N\odot B_N)\Pi_N\varepsilon)\\ &= 8\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N\sum_{m=1}^N\sum_{h=1}^N\sum_{\substack{p=1\\p\neq h}}^N\sum_{s=1}^N\sum_{\substack{r=1\\r\neq s}}^N\\ &\qquad \chi_{jk}\chi_{lm}q_{jh}q_{jp}q_{lh}q_{lp}q_{ks}q_{kr}q_{ms}q_{mr}\mathrm{E}(v_h^2v_p^2v_s^2v_r^2)\\ &\le 8\bar{q}^2\bar{\chi}c\sum_{m=1}^N(\sum_{j=1}^N|\chi_{jk}|)(\sum_{k=1}^N|q_{kr}|)(\sum_{l=1}^N|q_{lp}|)(\sum_{h=1}^N|q_{lh}|)\\ &\qquad (\sum_{p=1}^N|q_{jp}|)(\sum_{s=1}^N|q_{ms}|)(\sum_{r=1}^N|q_{mr}|) = O(N),\end{aligned}$$

where the inequality holds because $\mathrm{E}(v_h^2v_p^2v_s^2v_r^2)$ equals either $\mathrm{E}(v_h^2v_s^2)\mathrm{E}(v_p^2v_r^2)$ or $\mathrm{E}(v_h^2v_r^2)\mathrm{E}(v_p^2v_s^2)$ since $h \neq p$ and $s \neq r$, and either of them is less than a constant $c < \infty$, e.g., $\mathrm{E}(v_h^2v_r^2) \le \mathrm{E}^{\frac{1}{2}}(v_h^4)\mathrm{E}^{\frac{1}{2}}(v_r^4) \le c$. Therefore, by Chebyshev's inequality,

$$\begin{aligned} &P(\tfrac{1}{N}|\varepsilon'\Pi_N(A_N\odot B_N)\Pi_N\varepsilon - \mathrm{E}(\varepsilon'\Pi_N(A_N\odot B_N)\Pi_N\varepsilon)| \ge M)\\ &\le \tfrac{1}{M^2}\tfrac{1}{N^2}\mathrm{Var}(\varepsilon'\Pi_N(A_N\odot B_N)\Pi_N\varepsilon) = o(1).\end{aligned}$$

It follows that $\frac{1}{N}\varepsilon'\Pi_N(A_N\odot B_N)\Pi_N\varepsilon - \frac{1}{N}\mathrm{E}(\varepsilon'\Pi_N(A_N\odot B_N)\Pi_N\varepsilon) \xrightarrow{p} 0$. Therefore, we have shown that $\mathcal{T}_4 = \frac{2}{N}\texttt{tr}((A_N \odot B_N)\Pi_N \Lambda(\mathbf{H})\Pi_N) + o_p(1)$. It follows that

$$\frac{1}{N}\texttt{tr}(\widehat{\mathbf{H}} A_N \widehat{\mathbf{H}} B_N^{\circ}) - \frac{1}{N}\texttt{tr}(\mathbf{H} A_N \mathbf{H} B_N^{\circ}) = \textstyle\sum_{r=1}^{6} \mathcal{T}_r$$
$$= \frac{2}{N}\texttt{tr}((A_N \odot B_N)\Pi_N \Lambda(\mathbf{H})\Pi_N) + o_p(1),$$

completing the proof of Lemma 1.1. ■

Proof of Corollary 1.4: The consistency of $\widehat{\Sigma}_N^{\diamond}$ to $\Sigma_N^{\diamond}(\xi_0)$ is implied by results (b) and (c) in the proof of Theorem 1.4. To show $\widehat{\Gamma}_N^{\diamond} - \Gamma_N^{\diamond}(\xi_0) \xrightarrow{p} 0$, we argue as follows:

(a) The transition from $\Gamma_N^{\diamond}(\xi_0, \phi_0, \mathbf{H})$ to $\Gamma_N^{\diamond}(\hat{\xi}_N^{\diamond}, \phi_0, \mathbf{H})$ does not incur cost asymptotically;

(b) The cost of transition from $\Gamma_N^{\diamond}(\hat{\xi}_N^{\diamond}, \phi_0, \mathbf{H})$ to $\Gamma_N^{\diamond}(\hat{\xi}_N^{\diamond}, \hat{\phi}_N^{\diamond}, \mathbf{H})$ is captured by $\text{Bias}_{\phi}^{\diamond}(\hat{\delta}_N^{\diamond}, \mathbf{H})$;

(c) The effect of replacing $\mathbf{H}$ in $\frac{1}{N_1}\texttt{tr}(\mathbf{H}\mathbb{L}_a \mathbf{H}\mathbb{L}_b^{\circ}), a, b = \lambda, \rho$, is captured by $\frac{2}{N_1}\texttt{tr}((\mathbb{L}_a \odot \mathbb{L}_b^{\circ})\Pi_N \Lambda(\mathbf{H})\Pi_N), a, b = \lambda, \rho$;

(d) It is left to show that the cost of transition from $\text{Bias}_{\phi}^{\diamond}(\hat{\delta}_N^{\diamond}, \mathbf{H})$ to $\text{Bias}_{\phi}^{\diamond}(\hat{\delta}_N^{\diamond}, \hat{\mathbf{H}})$ is captured by $-\frac{2}{N_1}\texttt{tr}((\mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b \mathbb{P}_{\mathbb{D}})\Pi_N \Lambda(\mathbf{H})\Pi_N), a, b = \lambda, \rho$.

The non-zero entries in $\text{Bias}_{\phi}^{\diamond}(\delta_0, \mathbf{H})$ are of the form $\frac{1}{N_1}\texttt{tr}(\mathbf{H}\mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_{\mathbb{D}})$, for $a, b = \lambda, \rho$, as given in Corollary 1.3. Applying result (C.7) with $A_N = \mathbb{P}_{\mathbb{D}}\mathbb{L}_a'$ and $B_N = \mathbb{L}_b\mathbb{P}_{\mathbb{D}}$, we have,

$$\frac{1}{N_1}\texttt{tr}\big[\mathbb{P}_{\mathbb{D}}(\hat{\rho}_N^{\diamond})\mathbb{L}_a'(\hat{\delta}_N^{\diamond})\hat{\mathbf{H}}\mathbb{L}_b(\hat{\delta}_N^{\diamond})\mathbb{P}_{\mathbb{D}}(\hat{\rho}_N^{\diamond})\hat{\mathbf{H}} - \mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_{\mathbb{D}}\mathbf{H}\big]$$
$$= \frac{1}{N_1}\texttt{tr}\big[\mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\hat{\mathbf{H}}\mathbb{L}_b\mathbb{P}_{\mathbb{D}}\hat{\mathbf{H}} - \mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_{\mathbb{D}}\mathbf{H}\big] + o_p(1) \ \text{ (by the MVT)}$$
$$= \frac{2}{N_1}\texttt{tr}((\mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}})\Pi_N \Lambda(\mathbf{H})\Pi_N) + \frac{1}{N}\texttt{tr}((\mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}})\mathbf{H}^2) + o_p(1),$$
$$= \frac{2}{N_1}\texttt{tr}((\mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}})\Pi_N \Lambda(\mathbf{H})\Pi_N) + o_p(1),$$

for $a, b = \lambda, \rho$. Although the diagonal elements of $\mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}}$ may not be zero uniformly, their magnitudes are typically small so that the second term of the second last equation is negligible. The detail is tedious and thus is omitted. Under balanced panel data model considered in the first part of Appendix B, we can easily show $\texttt{diag}(\mathbb{L}_a\mathbb{P}_{\mathbb{D}}) = O(\frac{1}{n})$ for $a = \lambda, \rho$. Then, it follows that $\widehat{\Sigma}_N^{\diamond -1}\widehat{\Gamma}_N^{\diamond}\widehat{\Sigma}_N^{\diamond -1} - \Sigma_N^{\diamond -1}(\xi_0)\Gamma_N^{\diamond}(\xi_0)\Sigma_N^{\diamond -1}(\xi_0) \xrightarrow{p} 0$. ■

Appendices for Chapter 2

All the technical proofs for the main results in Chapter 2 are collected here. There are in total three appendices. Appendix A provides some basic lemmas that are used throughout the other appendices. Appendix B presents proofs for the main theorems in this chapter. These proofs rely on some technical lemmas whose proofs are put in Appendix C.

Appendix A: Some basic lemmas

The following lemmas are essential to the proofs of the main results in this chapter.

Lemma A.1. *(Lee, 2002): Let $\{A_n\}$ and $\{B_n\}$ be two sequences of $n \times n$ matrices that are uniformly bounded in both row and column sums. Let C_n be a sequence of conformable matrices whose elements are uniformly $O(h_n^{-1})$. Then*

(i) the sequence $\{A_n B_n\}$ are uniformly bounded in both row and column sums,

(ii) the elements of A_n are uniformly bounded and $\mathtt{tr}(A_n) = O(n)$, and

(iii) the elements of $A_n C_n$ and $C_n A_n$ are uniformly $O(h_n^{-1})$.

Lemma A.2. *(Lemma B.4, Yang, 2015): Let $\{A_n\}$ be a sequence of $n \times n$ matrices that are uniformly bounded in both row and column sums. Suppose that the elements $a_{n,ij}$ of A_n are bounded uniformly in all i and j, and $a_{n,ii} \neq 0$ for some i. Let v_n be a random n-vector of iid elements with mean zero, variance σ^2 and finite 4th moment, and b_n a random n-vector independent of v_n such that $\{\mathrm{E}(b_{ni}^2)\}$ are bounded. Then*

(i) $\mathrm{E}(v_n' A_n v_n) = O(n)$, *(ii)* $\mathrm{Var}(v_n' A_n v_n) = O(n)$,

(iii) $\mathrm{Var}(v_n' A_n v_n + b_n' v_n) = O(n)$, *(iv)* $v_n' A_n v_n = O_p(n)$,

(v) $v_n' A_n v_n - \mathrm{E}(v_n' A_n v_n) = O_p(n^{\frac{1}{2}})$, *(vi)* $v_n' A_n b_n = O_p(n^{\frac{1}{2}})$.

Lemma A.3. *(Lemma A.5, Yang, 2018): Let $\{\Phi_n\}$ be a sequence of $n \times n$ matrices with row and column sums uniformly bounded, and elements of uniform order $O(h_n^{-1})$. Let $v_n = (v_1, \ldots, v_n)'$ be a random vector of inid elements with mean zero, variance σ^2, and finite $(4 + 2\epsilon_0)$th moment for some*

$\epsilon_0 > 0$. Let $b_n = \{b_{ni}\}$ be an $n \times 1$ random vector, independent of v_n, such that (i) $\{\mathrm{E}(b_{ni}^2)\}$ are of uniform order $O(h_n^{-1})$, (ii) $sup_i E|b_{ni}|^{2+\epsilon_0} < \infty$, (iii) $\frac{h_n}{n}\sum_{i=1}^n[\phi_{n,ii}(b_{ni} - \mathrm{E}b_{ni})] = o_p(1)$ where $\{\phi_{n,ii}\}$ are the diagonal elements of Φ_n, and (iv) $\frac{h_n}{n}\sum_{i=1}^n[b_{ni}^2 - \mathrm{E}(b_{ni}^2)] = o_p(1)$. Define the bilinear-quadratic form:

$$Q_n = b_n' v_n + v_n'\Phi_n v_n - \sigma^2 \mathtt{tr}(\Phi_n),$$

and let $\sigma_{Q_n}^2$ be the variance of Q_n. If $\lim_{n\to\infty} h_n^{1+2/\epsilon_0}/n = 0$ and $\{\frac{h_n}{n}\sigma_{Q_n}^2\}$ are bounded away from zero, then $Q_n/\sigma_{Q_n} \xrightarrow{D} N(0,1)$.

Lemma A.4. *(Lemma 1, Hansen, 1996): If $\{w_i\}$ are iid,* $\mathrm{E}[\Psi(w_i)] < \infty$, *and w_i has a continuous distribution, then*

$$\sup_{\gamma\in\Gamma}\Big\|\frac{1}{n}\sum_{i=1}^n \Psi(w_i)\mathbb{1}\{w_i \leqslant \gamma\} - \mathrm{E}[\Psi(w_i)\mathbb{1}\{w_i \leqslant \gamma\}]\Big\| \longrightarrow 0 \ \ a.s.$$

Appendix B: Proofs of the theorems

This appendix presents proofs of the main theorems in this chapter. For ease of exposition, notations (existing and new) frequently used in the proofs are listed here:

- $A^{[k]}$ denotes the submatrix of A, consisting of the first k rows and columns,
  $h^{[k]}$ denotes the subvector of h, consisting of the first k elements;
- $d_{it}(\gamma) = \mathbb{1}(q_{it} \leqslant \gamma)$, $d_t(\gamma) = \mathtt{diag}\{d_{1t}(\gamma), \ldots, d_{nt}(\gamma)\}$, $d_t(\gamma_1,\gamma_2) = d_t(\gamma_1) - d_t(\gamma_2)$;
- $\beta = (\beta_1', \beta_2')'$, $\lambda = (\lambda_1, \lambda_2)'$, $\phi = (\beta', \lambda')'$, $\delta_0 = (b_0', l_0)'$, $\phi_2 = (\beta_2', \lambda_2)'$, $\omega = (\lambda', \gamma)'$;
- $A_t(\omega) = I_n - \lambda_1 W_t - \lambda_2 d_t(\gamma) W_t$, $G_t(\omega) = W_t A_t^{-1}(\omega)$;
- $Z_t = G_t(\mathbb{X}_t\beta_0 + \mu_0 + \alpha_{t0} l_n)$, $H_t = [X_t, Z_t]$, $\mathcal{Y}_t = W_t Y_t$, $\mathcal{V}_t = G_t V_t$, $R_t = \mathtt{diagv}(G_t)$;
- $\Pi_t^\circ(\gamma) = d_t(\gamma)\Pi_t$, and $\Pi_t^\star(\gamma_1,\gamma_2) = d_t(\gamma_1,\gamma_2)\Pi_t$, for $\Pi_t = X_t$, Z_t, H_t, $\mathcal{V}_t$, $\mathcal{Y}_t$ or R_t;
- $\mathbb{A}_t(\gamma) = [A_t, A_t^\circ(\gamma)]$, for $A_t = X_t, Z_t, H_t, \mathcal{V}_t$ or R_t.

The proof of main theorems can be greatly facilitated by Lemmas B.1−B.4

given below. Proofs of these lemmas are lengthy, not in the central focus, and thus are put in Appendix C.

Lemma B.1. *Under the assumptions of Theorem 2.1, we have,*

$$\mathcal{J}_{1,nT}(\gamma) = \tfrac{1}{\sqrt{nT}} \textstyle\sum_{t=1}^{T} \widetilde{H}_t^{\circ\prime}(\gamma) J_n \widetilde{V}_t \;\Rightarrow\; \mathcal{J}_1(\gamma),$$
$$\mathcal{J}_{2,nT}(\gamma) = \tfrac{1}{\sqrt{nT}} \textstyle\sum_{t=1}^{T} [\widetilde{V}_t' J_n \widetilde{\mathcal{V}}_t^{\circ}(\gamma) - \bar{T}\sigma_0^2 \mathtt{tr}(J_n d_t(\gamma) G_t)] \;\Rightarrow\; \mathcal{J}_2(\gamma),$$

where "$\Rightarrow$" denotes weak convergence with respect to the uniform metric, and both $\mathcal{J}_1(\gamma)$ and $\mathcal{J}_2(\gamma)$ are mean-zero Gaussian processes with almost surely continuous sample paths.

Lemma B.2. *Under the assumptions of Theorem 2.2, we have*

$$\mathcal{F}_{nT}(v) = \tfrac{a_{nT}}{nT} \textstyle\sum_{t=1}^{T} \delta_0' \widetilde{H}_t^{\star\prime}(\gamma_0, \gamma_0 + v/a_{nT}) J_n \widetilde{H}_t^{\star}(\gamma_0, \gamma_0 + v/a_{nT}) \delta_0 \;\Rightarrow\; \bar{T}\delta_0' M \delta_0 f|v|,$$
$$\mathcal{K}_{nT}(v) = \tfrac{a_{nT}}{nT} l_0^2 \textstyle\sum_{t=1}^{T} \widetilde{\mathcal{V}}_t^{\star\prime}(\gamma_0, \gamma_0 + v/a_{nT}) J_n \widetilde{\mathcal{V}}_t^{\star}(\gamma_0, \gamma_0 + v/a_{nT}) \;\Rightarrow\; l_0^2 \sigma_0^2 \bar{T} \pi_1 f|v|,$$
$$\mathcal{L}_{nT}(v) = \tfrac{a_{nT}}{nT} l_0^2 \textstyle\sum_{t=1}^{T} \mathtt{tr}[(d_t(\gamma_0, \gamma_0 + v/a_{nT}) G_t)^2] \;\Rightarrow\; l_0^2 \pi_2 f|v|,$$

where v is on a compact set $\Upsilon = [-\bar{v}, \bar{v}]$.

Lemma B.3. *Under the assumptions of Theorem 2.2, we have*

$$\mathcal{R}_{nT}(v) = \sqrt{a_{nT}} \big[\delta_0' \mathcal{J}_{1,nT}^{\star}(\gamma_0 + v/a_{nT}, \gamma_0) + l_0 \mathcal{J}_{2,nT}^{\star}(\gamma_0 + v/a_{nT}, \gamma_0)\big] \Rightarrow B(v)$$

where v is on a compact set $\Upsilon = [-\bar{v}, \bar{v}]$, $\mathcal{J}_{r,nT}^{\star}(\gamma_1, \gamma_2) = \mathcal{J}_{r,nT}(\gamma_1) - \mathcal{J}_{r,nT}(\gamma_2)$ for $r = 1, 2$, $B(v) = \sqrt{\sigma_0^2 \Xi f} W(v)$ and $W(v)$ is a standard Brownian motion.

Lemma B.4. *Under the assumptions of Theorem 2.3, there exist constants $B > 0$, $0 < k < \infty$, and $0 < l < \infty$, such that for all $\eta > 0$, and $\epsilon > 0$, there exists a $\bar{v} < \infty$ such that for large enough (n, T), $\mathcal{N}_{nT} = \big\{\gamma : \frac{\bar{v}}{a_{nT}} \leqslant |\gamma - \gamma_0| \leqslant B\big\}$, and $r = 1, 2, 3$ and $s = 1, 2$,*

$$(a)\; P\Bigg(\inf_{\gamma \in \mathcal{N}_{nT}} \tfrac{D_{r,nT}(\gamma)}{|\gamma - \gamma_0|} < (1-\eta)k\Bigg) \leqslant \epsilon, \quad (b)\; P\Bigg(\sup_{\gamma \in \mathcal{N}_{nT}} \tfrac{\|F_{s,nT}(\gamma)\|}{|\gamma - \gamma_0|} > (1+\eta)l\Bigg) \leqslant \epsilon,$$
$$(c)\; P\Bigg(\sup_{\gamma \in \mathcal{N}_{nT}} \tfrac{|K_{s,nT}(\gamma)|}{|\gamma - \gamma_0|} > (1+\eta)l\Bigg) \leqslant \epsilon, \quad (d)\; P\Bigg(\sup_{\gamma \in \mathcal{N}_{nT}} \tfrac{|L_{r,nT}(\gamma)|}{|\gamma - \gamma_0|} > (1+\eta)l\Bigg) \leqslant \epsilon,$$
$$(e)\; P\Bigg(\sup_{\gamma \in \mathcal{N}_{nT}} \tfrac{\|P_{r,nT}(\gamma)\|}{|\gamma - \gamma_0|} > \eta\Bigg) \leqslant \epsilon, \quad (f)\; P\Bigg(\sup_{\gamma \in \mathcal{N}_{nT}} \tfrac{\|\mathcal{J}_{s,nT}(\gamma) - \mathcal{J}_{s,nT}(\gamma_0)\|}{\sqrt{a_{nT}}|\gamma - \gamma_0|} > \eta\Bigg) \leqslant \epsilon,$$

where $D_{1,nT}(\gamma) = \delta_0' F_{1,nT}(\gamma) \delta_0$, $D_{2,nT}(\gamma) = l_0^2 K_{1,nT}(\gamma)$, $D_{3,nT}(\gamma) = l_0^2 L_{1,nT}(\gamma)$,

$F_{1,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T} \widetilde{H}_t^{\star\prime}(\gamma_0,\gamma)J_n\widetilde{H}_t^{\star}(\gamma_0,\gamma)$, $F_{2,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T} \widetilde{\mathbb{H}}_t'(\gamma_0)J_n\widetilde{H}_t^{\star}(\gamma_0,\gamma)$,
$K_{1,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T} \widetilde{\mathcal{V}}_t^{\star\prime}(\gamma_0,\gamma)J_n\widetilde{\mathcal{V}}_t^{\star}(\gamma_0,\gamma)$, $K_{2,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T} \widetilde{\mathbb{V}}_t'(\gamma_0)J_n\widetilde{\mathcal{V}}_t^{\star}(\gamma_0,\gamma)$,
$L_{1,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T} \mathtt{tr}[(d_t(\gamma_0,\gamma)G_t)^2]$, $L_{2,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T} \mathtt{tr}[d_t(\gamma_0,\gamma)G_t]$,
$L_{3,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T} \mathtt{tr}[l_n l_n' d_t(\gamma_0,\gamma)G_t]$, $P_{1,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T} \widetilde{H}_t^{\star\prime}(\gamma_0,\gamma)J_n\widetilde{\mathcal{V}}_t^{\star}(\gamma_0,\gamma)$,
$P_{2,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T} \widetilde{\mathbb{H}}_t'(\gamma_0)J_n\widetilde{\mathcal{V}}_t^{\star}(\gamma_0,\gamma)$, $P_{3,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T} \widetilde{\mathbb{V}}_t'(\gamma_0)J_n\widetilde{H}_t^{\star}(\gamma_0,\gamma)$.

Proof of Theorem 2.1: We first prove convergence of $\hat{\beta}_{nT}(\hat{\lambda}_{nT},\gamma)$ and $\hat{\sigma}_{nT}^2(\hat{\lambda}_{nT},\gamma)$, uniformly in $\gamma \in \Gamma$. We have, $A_t(\omega)A_t^{-1} = I_n + (\lambda_{10}-\lambda_1)G_t + (\lambda_{20}-\lambda_2)d_t(\gamma)G_t + \lambda_{20}d_t(\gamma_0,\gamma)G_t$, noting that $A_t^{-1} = I_n + \lambda_{10}G_t + \lambda_{20}d_t(\gamma_0)G_t$. By $Y_t = A_t^{-1}(\mathbb{X}_t\beta_0 + \mu_0 + \alpha_{t0}l_n + V_t)$ and $\mathbb{X}_t\beta_0 = \mathbb{X}_t(\gamma)\beta_0 + X_t^{\star}(\gamma_0,\gamma)\beta_{20}$, we have

$$\mathbb{Y}_t(\omega) = \mathbb{X}_t(\gamma)\beta_0 + \mathcal{D}_t(\gamma)\phi^{\dagger} + \zeta_t(\gamma)\lambda^{\dagger} + \mu_0 + \alpha_{t0}l_n + V_t, \tag{B.1}$$

where $\mathcal{D}_t(\gamma) = (\mathbb{Z}_t(\gamma), H_t^{\star}(\gamma_0,\gamma))$, $\phi^{\dagger} = ((\lambda_0-\lambda)', \phi_{20}')'$, $\zeta_t(\gamma) = (\mathbb{V}_t(\gamma), \mathcal{V}_t^{\star}(\gamma_0,\gamma))$ and $\lambda^{\dagger} = ((\lambda_0-\lambda)', \lambda_{20})'$. Combining it with (2.6) and (2.7), we have

$$\hat{\beta}_{nT}(\omega) = \beta_0 + \mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{D}_{2,nT}(\gamma)\phi^{\dagger} + \mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{V}_{5,nT}(\gamma)\lambda^{\dagger} + \mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{V}_{4,nT}(\gamma), \text{ and} \tag{B.2}$$

$$\begin{aligned}\hat{\sigma}_{nT}^2(\omega) = {} & \mathcal{V}_{1,nT} + 2\mathcal{V}_{2,nT}(\gamma)\lambda^{\dagger} + \lambda^{\dagger\prime}\mathcal{V}_{3,nT}(\gamma)\lambda^{\dagger} + \mathcal{V}_{4,nT}'(\gamma)\mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{V}_{4,nT}(\gamma) \\ & + 2\mathcal{V}_{4,nT}'(\gamma)\mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{V}_{5,nT}(\gamma)\lambda^{\dagger} + \lambda^{\dagger\prime}\mathcal{V}_{5,nT}'(\gamma)\mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{V}_{5,nT}(\gamma)\lambda^{\dagger} \\ & + \phi^{\dagger\prime}[\mathcal{D}_{1,nT}(\gamma) - \mathcal{D}_{2,nT}'(\gamma)\mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{D}_{2,nT}(\gamma)]\phi^{\dagger} \\ & + 2\phi^{\dagger\prime}[\mathcal{V}_{6,nT}(\gamma) - \mathcal{D}_{2,nT}'(\gamma)\mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{V}_{4,nT}(\gamma)] \\ & + 2\phi^{\dagger\prime}[\mathcal{V}_{7,nT}(\gamma) - \mathcal{D}_{2,nT}'(\gamma)\mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{V}_{5,nT}(\gamma)]\lambda^{\dagger},\end{aligned} \tag{B.3}$$

where $\mathcal{P}_{1,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{\mathbb{X}}_t(\gamma)$, $\mathcal{D}_{1,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{\mathcal{D}}_t'(\gamma)J_n\widetilde{\mathcal{D}}_t(\gamma)$,
$\mathcal{D}_{2,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{\mathcal{D}}_t(\gamma)$, $\mathcal{V}_{1,nT} = \frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{V}_t'J_n\widetilde{V}_t$,
$\mathcal{V}_{2,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{V}_t'J_n\widetilde{\zeta}_t(\gamma)$, $\mathcal{V}_{3,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{\zeta}_t'(\gamma)J_n\widetilde{\zeta}_t(\gamma)$,
$\mathcal{V}_{4,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{V}_t$, $\mathcal{V}_{5,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{\zeta}_t(\gamma)$,
$\mathcal{V}_{6,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{\mathcal{D}}_t'(\gamma)J_n\widetilde{V}_t$, $\mathcal{V}_{7,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{\mathcal{D}}_t'(\gamma)J_n\widetilde{\zeta}_t(\gamma)$.

Under Assumption B(vi), the limit of $\mathcal{P}_{1,nT}(\gamma)$ exists and is nonsingular. In addition, we have, uniformly in $\gamma \in \Gamma$, $\mathcal{V}_{4,nT}(\gamma)$ and $\mathcal{V}_{6,nT}(\gamma)$ are $O_p((nT)^{-1/2})$ by Lemma B.1; $\mathcal{V}_{5,nT}(\gamma)$ and $\mathcal{V}_{7,nT}(\gamma)$ are $o_p(1)$ by Lemma A.4; $\mathcal{D}_{1,nT}(\gamma)$, $\mathcal{D}_{2,nT}(\gamma)$, $\mathcal{V}_{2,nT}(\gamma)$ and $\mathcal{V}_{3,nT}(\gamma)$ are all $O_p(1)$ by Lemma A.4, as their expectations are all $O(1)$. Besides, $\mathcal{V}_{1,nT} - \sigma_0^2 = o_p(1)$ by Lemma A.2 and

$\phi_{20} = O((nT)^{-\tau})$ by Assumption F. These together lead to

$$\hat{\beta}_{nT}(\omega) = \beta_0 + \mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{P}_{2,nT}(\gamma)(\lambda_0 - \lambda) + o_p(1), \quad \text{and} \tag{B.4}$$

$$\hat{\sigma}_{nT}^2(\omega) = \sigma_0^2 + 2\mathcal{V}_{21,nT}(\gamma)(\lambda_0 - \lambda) + (\lambda_0 - \lambda)'\big[\mathcal{V}_{31,nT}(\gamma) + \mathcal{P}_{3,nT}(\gamma) - \mathcal{P}_{2,nT}'(\gamma)\mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{P}_{2,nT}(\gamma)\big](\lambda_0 - \lambda) + o_p(1), \tag{B.5}$$

where $\mathcal{V}_{21,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^T \widetilde{V}_t' J_n \widetilde{\mathbb{V}}_t(\gamma)$, $\mathcal{V}_{31,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^T \widetilde{\mathbb{V}}_t'(\gamma) J_n \widetilde{\mathbb{V}}_t(\gamma)$, $\mathcal{P}_{2,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^T \widetilde{\mathbb{X}}_t'(\gamma) J_n \widetilde{\mathbb{Z}}_t(\gamma)$, and $\mathcal{P}_{3,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^T \widetilde{\mathbb{Z}}_t'(\gamma) J_n \widetilde{\mathbb{Z}}_t(\gamma)$. These imply that

$$\hat{\beta}_{nT}(\hat{\lambda}_{nT}, \gamma) = \beta_0 + o_p(1) \text{ and } \hat{\sigma}_{nT}^2(\hat{\lambda}_{nT}, \gamma) = \sigma_0^2 + o_p(1), \tag{B.6}$$

uniformly in $\gamma \in \Gamma$, as long as $\hat{\lambda}_{nT} = \lambda_0 + o_p(1)$.

Therefore, to show the consistency of $\hat{\theta}_{nT}$, we only need to show the consistency of $\hat{\lambda}_{nT}$. By Theorem 2.5 of Newey and McFadden (1994), the consistency of $\hat{\lambda}_{nT}$ follows if

(a) $\sup_{\omega \in \Lambda \times \Gamma} \frac{1}{n(T-1)} |\ell_{nT}^{*c}(\omega) - \bar{\ell}_{nT}^{*c}(\omega)| = o_p(1)$,

(b) $\lim_{nT \to \infty} \frac{1}{n(T-1)} \bar{\ell}_{nT}^{*c}(\omega)$ is uniformly equicontinuous in λ for any γ,

(c) λ_0 uniquely maximizes $\lim_{nT \to \infty} \frac{1}{n(T-1)} \bar{\ell}_{nT}^{*c}(\omega)$ over $\omega \in \Lambda \times \Gamma$.

Proof of (a): Note that, from (2.8) and (2.11),

$$\frac{1}{n(T-1)}[\ell_{nT}^{*c}(\omega) - \bar{\ell}_{nT}^{*c}(\omega)] = -\tfrac{1}{2}[\ln \hat{\sigma}_{nT}^2(\omega) - \ln \bar{\sigma}_{nT}^2(\omega)] + \frac{1}{n(T-1)}\sum_{t=1}^T [\ln |A_t(\omega)| - \mathrm{E}(\ln |A_t(\omega)|)].$$

For the second term, Lemma A.4 implies that $\sup_{\gamma \in \Gamma} \frac{1}{n(T-1)}\sum_{t=1}^T [\ln |A_t(\omega)| - \mathrm{E}(\ln |A_t(\omega)|)] = o_p(1)$ for any given λ. Hence, we have $\sup_{\omega \in \Lambda \times \Gamma} \frac{1}{n(T-1)}\sum_{t=1}^T [\ln |A_t(\omega)| - \mathrm{E}(\ln |A_t(\omega)|)] = o_p(1)$. For the first term, we firstly show $\bar{\sigma}_{nT}^2(\omega)$ is bounded away from zero uniformly in $\omega \in \Lambda \times \Gamma$ so that we have $\ln \hat{\sigma}_{nT}^2(\omega) - \ln \bar{\sigma}_{nT}^2(\omega) = \ln[1 + \bar{\sigma}_{nT}^{-2}(\omega)(\hat{\sigma}_{nT}^2(\omega) - \bar{\sigma}_{nT}^2(\omega))]$, and then prove $\hat{\sigma}_{nT}^2(\omega) - \bar{\sigma}_{nT}^2(\omega) = o_p(1)$ uniformly in $\omega \in \Lambda \times \Gamma$. From (B.1), one can also see $\mathbb{Y}_t(\omega) = \mathbb{X}_t(\gamma)\beta_0 + \mathcal{D}_t(\gamma)\phi^\dagger + \mu_0 + \alpha_{t0} l_n + A_t(\omega) A_t^{-1} V_t$. Thus, by (2.10), we have

$$\bar{\sigma}_{nT}^2(\omega) = \tfrac{n-1}{n}\sigma_{nT}^2(\omega) + \phi^{\dagger\prime}[\mathrm{E}\mathcal{D}_{1,nT}(\gamma) - \mathrm{E}\mathcal{D}_{2,nT}'(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\mathrm{E}\mathcal{D}_{2,nT}(\gamma)]\phi^\dagger, \tag{B.7}$$

where $\sigma_{nT}^2(\omega)$ is above Assumption G. For the second term, the quantity in the square bracket is a Schur complement of $\frac{1}{n(T-1)}\sum_{t=1}^T \mathrm{E}[\widetilde{\mathcal{H}}_t'(\gamma)J_n\widetilde{\mathcal{H}}_t(\gamma)]$, where $\mathcal{H}_t(\gamma)=[\mathbb{H}_t(\gamma),\ H_t^\star(\gamma_0,\gamma)]$. Thus, the quantity in the square bracket must be positive semi-definite, using the properties of the Schur complement. Therefore, the second term is non-negative. For the first term, we have

$$\tfrac{n-1}{n}\sigma_{nT}^2(\omega)=\ddot{\sigma}_{nT}^2(\omega)-\tfrac{\sigma_0^2}{n^2T}\textstyle\sum_{t=1}^T l_n'A_t(\omega)A_t^{-1}A_t'^{-1}A_t'(\omega)l_n=\ddot{\sigma}_{nT}^2(\omega)+O_p(\tfrac{1}{n}),$$

where $\ddot{\sigma}_{nT}^2(\omega)=\frac{\sigma_0^2}{nT}\sum_{t=1}^T \mathtt{tr}(A_t'^{-1}A_t'(\omega)A_t(\omega)A_t^{-1})$. Note that

$$\begin{aligned}&-\tfrac{1}{2}[\ln\ddot{\sigma}_{nT}^2(\omega)-\ln(\bar{T}\sigma_0^2)]+\tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T[\ln|A_t(\omega)|-\ln|A_t|]\\ =&-\tfrac{1}{2}\big[\ln(\tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T\mathtt{tr}(A_t'^{-1}A_t'(\omega)A_t(\omega)A_t^{-1}))\\ &-\ln(\textstyle\prod_{t=1}^T|A_t'^{-1}A_t'(\omega)A_t(\omega)A_t^{-1}|)^{\frac{1}{n(T-1)}}\big]\leqslant 0,\end{aligned}$$

due to the fact that arithmetic mean is no less than geometric means. As $\sigma_{nT}^2(\omega)=\ddot{\sigma}_{nT}^2(\omega)+O_p(\frac{1}{n})$, the above inequality implies

$$-\tfrac{1}{2}\ln\sigma_{nT}^2(\omega)\leqslant-\tfrac{1}{2}\ln\sigma_0^2-\tfrac{1}{2}\ln\bar{T}+\tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T[(\ln|A_t|)-(\ln|A_t(\omega)|)]+O_p(\tfrac{1}{n})=O_p(1). \tag{B.8}$$

Hence, we conclude that $\sigma_{nT}^2(\omega)$ is bounded away from zero on $\Lambda\times\Gamma$ and so is $\bar{\sigma}_{nT}^2(\omega)$.

Thus, it is left to show $\hat{\sigma}_{nT}^2(\omega)-\bar{\sigma}_{nT}^2(\omega)=o_p(1)$, uniformly in $\omega\in\Lambda\times\Gamma$. Firstly, using $A_t(\omega)A_t^{-1}=I_n+(\lambda_{10}-\lambda_1)G_t+(\lambda_{20}-\lambda_2)d_t(\gamma)G_t+\lambda_{20}d_t(\gamma_0,\gamma)G_t$, we have

$$\sigma_{nT}^2(\omega)=\sigma_0^2+\tfrac{2n\sigma_0^2}{n-1}\mathcal{G}_{1,nT}'(\gamma)\lambda^\dagger+\tfrac{n\sigma_0^2}{n-1}\lambda^{\dagger\prime}\mathbb{G}_{1,nT}(\gamma)\lambda^\dagger, \tag{B.9}$$

where $\mathcal{G}_{1,nT}(\gamma)=\frac{1}{nT}\sum_{t=1}^T[\mathtt{tr}(J_nG_t),\mathtt{tr}(J_nd_t(\gamma)G_t),\mathtt{tr}(J_nd_t(\gamma_0,\gamma)G_t)]'$ and

$$\mathbb{G}_{1,nT}(\gamma)=\frac{1}{nT}\sum_{t=1}^T\begin{bmatrix}\mathtt{tr}(J_nG_tG_t'), & \mathtt{tr}(J_nd_t(\gamma)G_tG_t'), & \mathtt{tr}(J_nd_t(\gamma_0,\gamma)G_tG_t'),\\ \sim, & \mathtt{tr}(J_nd_t(\gamma)G_tG_t'd_t(\gamma)), & \mathtt{tr}(J_nd_t(\gamma_0,\gamma)G_tG_t'd_t(\gamma)),\\ \sim, & \sim, & \mathtt{tr}(J_nd_t(\gamma_0,\gamma)G_tG_t'd_t(\gamma_0,\gamma)),\end{bmatrix}.$$

By plugging (B.9) into (B.7), we have

$$\begin{aligned}\bar{\sigma}_{nT}^2(\omega)=&\tfrac{n-1}{n}\sigma_0^2+2\sigma_0^2\mathrm{E}\mathcal{G}_{1,nT}'(\gamma)\lambda^\dagger+\sigma_0^2\lambda^{\dagger\prime}\mathrm{E}\mathbb{G}_{1,nT}(\gamma)\lambda^\dagger\\ &+\phi^{\dagger\prime}[\mathrm{E}\mathcal{D}_{1,nT}(\gamma)-\mathrm{E}\mathcal{D}_{2,nT}'(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\mathrm{E}\mathcal{D}_{2,nT}(\gamma)]\phi^\dagger,\end{aligned} \tag{B.10}$$

Note that elements of $\mathrm{E}\mathcal{G}_{1,nT}(\gamma)$ and $\mathrm{E}\mathbb{G}_{1,nT}(\gamma)$ are uniformly bounded on Γ by Assumption C and D, and $\lambda_{20}=O((nT)^{-\tau})$ by Assumption F. Thus, corresponding to (B.5), we have

$$\bar{\sigma}^2_{nT}(\omega) = \sigma^2_0 + 2\sigma^2_0 \mathrm{E}\mathcal{G}^{[2]\prime}_{1,nT}(\gamma)(\lambda_0 - \lambda) + (\lambda_0 - \lambda)' \big[\sigma^2_0 \mathrm{E}\mathbb{G}^{[2]}_{1,nT}(\gamma) + \mathrm{E}\mathcal{P}_{3,nT}(\gamma) - \mathrm{E}\mathcal{P}'_{2,nT}(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\mathrm{E}\mathcal{P}_{2,nT}(\gamma)\big](\lambda_0 - \lambda) + o(1). \tag{B.11}$$

It is easy to see that $\mathrm{E}\mathcal{V}_{21,nT}(\gamma) = \sigma^2_0 \mathrm{E}\mathcal{G}^{[2]\prime}_{1,nT}(\gamma)$ and $\mathrm{E}\mathcal{V}_{31,nT}(\gamma) = \sigma^2_0 \mathrm{E}\mathbb{G}^{[2]}_{1,nT}(\gamma)$. Thus, Lemma A.4 implies that $\mathcal{V}_{21,nT}(\gamma) - \sigma^2_0 \mathrm{E}\mathcal{G}^{[2]\prime}_{1,nT}(\gamma) \xrightarrow{a.s.} 0$, $\mathcal{V}_{31,nT}(\gamma) - \sigma^2_0 \mathrm{E}\mathbb{G}^{[2]}_{1,nT}(\gamma) \xrightarrow{a.s.} 0$ and $\mathcal{P}_{r,nT}(\gamma) - \mathrm{E}\mathcal{P}_{r,nT}(\gamma) \xrightarrow{a.s.} 0$ for $r = 1, 2, 3$, uniformly in $\gamma \in \Gamma$. These convergences are also uniform on Λ because λ appears simply as linear or quadratic factors in these terms. Therefore, we have $\hat{\sigma}^2_{nT}(\omega) - \bar{\sigma}^2_{nT}(\omega) = o_p(1)$, uniformly in $\omega \in \Lambda \times \Gamma$.

Proof of (b): Recall from (2.11) that

$$\tfrac{1}{n(T-1)}\bar{\ell}^{*c}_{nT}(\omega) = -\tfrac{1}{2}(\ln 2\pi + 1) - \tfrac{1}{2}\ln \bar{\sigma}^2_{nT}(\omega) + \tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T \mathrm{E}(\ln|A_t(\omega)|),$$

From (B.11), we see that the limit of $\bar{\sigma}^2_{nT}(\omega)$ are uniformly equicontinuous on Λ given γ, as they are linear or quadratic in λ with the corresponding vector and matrices, $2\sigma^2_0 \mathrm{E}\mathcal{G}^{[2]\prime}_{1,nT}(\gamma)$, $\sigma^2_0 \mathrm{E}\mathbb{G}^{[2]}_{1,nT}(\gamma)$ and $\mathrm{E}\mathcal{P}_{3,nT}(\gamma) - \mathrm{E}\mathcal{P}'_{2,nT}(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}$ $\mathrm{E}\mathcal{P}_{2,nT}(\gamma)$, being bounded. To see the uniform equicontinuity of $\frac{1}{n(T-1)}\sum_{t=1}^T$ $\mathrm{E}(\ln|A_t(\omega)|)$ on Λ, a Taylor expansion around λ_0 gives,

$$\begin{aligned}&\tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T \mathrm{E}(\ln|A_t(\omega)|)\\ =\ &\tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T \mathrm{E}[(\ln|A_t(\lambda_0,\gamma)|) + \mathtt{tr}(G_t(\dot{\lambda},\gamma))(\lambda_1 - \lambda_{10})\\ &+ \mathtt{tr}(d_t(\gamma)G_t(\dot{\lambda},\gamma))(\lambda_2 - \lambda_{20})],\end{aligned}$$

where $\dot{\lambda}$ lies between λ and λ_0. As $\frac{1}{n(T-1)}\sum_{t=1}^T \mathtt{tr}(G_t(\dot{\lambda},\gamma))$ and $\frac{1}{n(T-1)}\sum_{t=1}^T$ $\mathtt{tr}(d_t(\gamma)G_t(\dot{\lambda},\gamma))$ are uniformly bounded by Assumptions C and D for any $\dot{\lambda}$ and γ, $\frac{1}{n(T-1)}\sum_{t=1}^T \mathrm{E}(\ln|A_t(\omega)|)$ is also uniformly equicontinuous on Λ for any γ.

Proof of (c): Letting $\check{\sigma}_{nT}^2(\omega) = \frac{n-1}{n}\mathrm{E}\sigma_{nT}^2(\omega)$, we have

$$
\begin{aligned}
& \tfrac{1}{n(T-1)}[\bar{\ell}_{nT}^{*c}(\omega) - \bar{\ell}_{nT}^{*c}(\omega_0)] \\
= \; & -\tfrac{1}{2}[\ln \bar{\sigma}_{nT}^2(\omega) - \ln \check{\sigma}_{nT}^2(\omega)] - \tfrac{1}{2}[\ln \check{\sigma}_{nT}^2(\omega) - \ln \bar{\sigma}_{nT}^2(\lambda_0,\gamma)] \\
& - \tfrac{1}{2}[\ln \bar{\sigma}_{nT}^2(\lambda_0,\gamma) - \ln \bar{\sigma}_{nT}^2(\omega_0)] \\
& + \tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T \mathrm{E}[\ln|A_t(\omega)| - \ln|A_t(\lambda_0,\gamma)|] \\
& + \tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T \mathrm{E}[\ln|A_t(\lambda_0,\gamma)| - \ln|A_t|] \\
= \; & -\tfrac{1}{2}[\ln \bar{\sigma}_{nT}^2(\omega) - \ln \check{\sigma}_{nT}^2(\omega)] - \tfrac{1}{2}[\ln \check{\sigma}_{nT}^2(\omega) - \ln \bar{\sigma}_{nT}^2(\lambda_0,\gamma)] \\
& + \tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T \mathrm{E}[\ln|A_t(\omega)| - \ln|A_t(\lambda_0,\gamma)|] + o(1),
\end{aligned}
$$

where the last equation holds because $\bar{\sigma}_{nT}^2(\lambda_0,\gamma) - \bar{\sigma}_{nT}^2(\omega_0) = o_p(1)$ by (B.11) and $\ln|A_t(\lambda_0,\gamma)| - \ln|A_t| = \ln|A_t(\lambda_0,\gamma)A_t^{-1}| = \ln|I_n + \lambda_{20}d_t(\gamma_0,\gamma)G_t| = o_p(1)$. Thus, it amounts to showing that the remain three terms are always negative for $\lambda \neq \lambda_0$ for any γ.

For the first term, using (B.7) and $\phi_{20} = O((nT)^{-\tau})$, we have

$$
\begin{aligned}
& -\tfrac{1}{2}[\ln \bar{\sigma}_{nT}^2(\omega) - \ln \check{\sigma}_{nT}^2(\omega)] \\
= & -\tfrac{1}{2}\ln\left[1 + \check{\sigma}_{nT}^{-2}(\omega)\phi^{\dagger\prime}(\mathrm{E}\mathcal{D}_{1,nT}(\gamma) - \mathrm{E}\mathcal{D}_{2,nT}'(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\mathrm{E}\mathcal{D}_{2,nT}(\gamma))\phi^{\dagger}\right] \\
= & -\tfrac{1}{2}\ln\big\{1 + \check{\sigma}_{nT}^{-2}(\omega)(\lambda_0-\lambda)'[\mathrm{E}\mathcal{P}_{3,nT}(\gamma) \\
& \quad - \mathrm{E}\mathcal{P}_{2,nT}'(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\mathrm{E}\mathcal{P}_{2,nT}(\gamma)](\lambda_0-\lambda)\big\} + o(1).
\end{aligned}
$$

The quantity in the square bracket is the Schur complement of $\frac{1}{n(T-1)}\sum_{t=1}^T \mathrm{E}[\widetilde{\mathbb{H}}_t'(\gamma)J_n\widetilde{\mathbb{H}}_t(\gamma)]$ so that it is positive semi-definite. Thus, the limit of the above equation is non-positive.

For the second and third terms, noting that $\bar{\sigma}_{nT}^2(\lambda_0,\gamma) = \sigma_0^2 + o(1)$ by (B.11), we have

$$
\begin{aligned}
& -\tfrac{1}{2}[\ln \check{\sigma}_{nT}^2(\omega) - \ln \bar{\sigma}_{nT}^2(\lambda_0,\gamma)] + \tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T[\mathrm{E}(\ln|A_t(\omega)|) - \mathrm{E}(\ln|A_t(\lambda_0,\gamma)|)] \\
= \; & -\tfrac{1}{2}[\ln \sigma_{nT}^2(\omega) - \ln \sigma_0^2] + \tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T[(\ln|A_t(\omega)|) - (\ln|A_t(\lambda_0,\gamma)|)] + o_p(1),
\end{aligned}
$$

the limit of which is also non-positive by (B.8). Together, we have

$$
\begin{aligned}
& \tfrac{1}{n(T-1)}[\bar{\ell}_{nT}^{*c}(\omega) - \bar{\ell}_{nT}^{*c}(\omega_0)] \\
= \; & -\tfrac{1}{2}\ln\left[1 + \sigma_{nT}^{-2}(\omega)(\lambda_0-\lambda)'(\mathcal{P}_{3,nT}(\gamma) - \mathcal{P}_{2,nT}'(\gamma)\mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{P}_{2,nT}(\gamma))(\lambda_0-\lambda)\right] \\
& -\tfrac{1}{2}[\ln \sigma_{nT}^2(\omega) - \ln \sigma_0^2] + \tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^T[(\ln|A_t(\omega)|) - (\ln|A_t(\lambda_0,\gamma)|)] + o_p(1).
\end{aligned}
$$

As discussed above, we have $\lim_{nT\to\infty}\frac{1}{n(T-1)}[\bar{\ell}^{*c}_{nT}(\omega)-\bar{\ell}^{*c}_{nT}(\omega_0)]\leqslant 0$. From the partition matrix formula, $\lim_{nT\to\infty}\frac{1}{n(T-1)}\sum_{t=1}^{T}[\widetilde{\mathbb{H}}'_t(\gamma)J_n\widetilde{\mathbb{H}}_t(\gamma)]$ is non-singular if and only if $\lim_{nT\to\infty}\mathcal{P}_{1,nT}(\gamma)$ and $\lim_{nT\to\infty}[\mathcal{P}_{3,nT}(\gamma)-\mathcal{P}'_{2,nT}(\gamma)\mathcal{P}^{-1}_{1,nT}(\gamma)\mathcal{P}_{2,nT}(\gamma)]$ are non-singular. Hence, Assumption G(i) implies $\lim_{nT\to\infty}[\mathcal{P}_{3,nT}(\gamma)-\mathcal{P}'_{2,nT}(\gamma)\mathcal{P}^{-1}_{1,nT}(\gamma)\mathcal{P}_{2,nT}(\gamma)]$ is positive definite so that the limit of $\frac{1}{n(T-1)}[\bar{\ell}^{*c}_{nT}(\omega)-\bar{\ell}^{*c}_{nT}(\omega_0)]$ is strictly less than zero unless $\lambda=\lambda_0$, i.e., λ_0 is the unique maximizer of $\frac{1}{n(T-1)}\bar{\ell}^{*c}_{nT}(\omega)$. If Assumption G($i$) fails, identification requires that the limit of $-\frac{1}{2}[\ln\sigma^2_{nT}(\omega)-\ln\sigma_0^2]+\frac{1}{n(T-1)}\sum_{t=1}^{T}[(\ln|A_t(\omega)|)-(\ln|A_t(\lambda_0,\gamma)|)]$ is strictly less than zero for any γ and $\lambda\neq\lambda_0$, which is equivalent to Assumption G(ii) because $\sigma^2_{nT}(\lambda_0,\gamma)=\sigma_0^2+o_p(1)$ by (B.9) and $\lambda_{20}=O((nT)^{-\tau})$. ■

Proof of Theorem 2.2: In this proof, we show the consistency of $\hat{\gamma}_{nT}$ in two steps:

(a) We derive a preliminary convergence rate for $\hat{\theta}_{nT}$, $(nT)^{\tau}(\hat{\theta}_{nT}-\theta_0)=O_p(1)$;

(b) Based on the convergence rate, we then establish the consistency of $\hat{\gamma}_{nT}$.

Proof of (a): For $S^*_{\theta,nT}(\theta,\gamma)$ given in (2.5), applying the mean value theorem (MVT) to each element of $S^*_{\theta,nT}(\hat{\theta}_{nT},\hat{\gamma}_{nT})$, we have

$$0=S^*_{\theta,nT}(\hat{\theta}_{nT},\hat{\gamma}_{nT})=S^*_{\theta,nT}(\theta_0,\hat{\gamma}_{nT})+\left[\frac{\partial}{\partial\theta'}S^*_{\theta,nT}(\theta,\hat{\gamma}_{nT})\Big|_{\theta=\bar{\theta}_r\text{ in }r\text{th row}}\right](\hat{\theta}_{nT}-\theta_0),$$

where $\{\bar{\theta}_r\}$ are on the line segment between $\hat{\theta}_{nT}$ and θ_0. In the following arguments, we use $H^*_{nT}(\bar{\theta},\gamma)$ to denote $-\frac{\partial}{\partial\theta'}S^*_{\theta,nT}(\theta,\gamma)\big|_{\theta=\bar{\theta}_r\text{ in }r\text{th row}}$ for simplicity. Thus, we have

$$(nT)^{\tau}(\hat{\theta}_{nT}-\theta_0)=\left[\frac{1}{n(T-1)}H^*_{nT}(\bar{\theta},\hat{\gamma}_{nT})\right]^{-1}\frac{(nT)^{\tau}}{n(T-1)}S^*_{\theta,nT}(\theta_0,\hat{\gamma}_{nT}).\qquad\text{(B.12)}$$

Therefore, the proof of the result in (a) is equivalent to showing for any given γ,

(i) $\frac{1}{n(T-1)}[H^*_{nT}(\bar{\theta},\gamma)-H^*_{nT}(\theta_0,\gamma)]=o_p(1)$,

(ii) $\frac{1}{n(T-1)}[H^*_{nT}(\theta_0,\gamma)-\mathrm{E}(H^*_{nT}(\theta_0,\gamma))]=o_p(1)$,

(iii) The limit of $\frac{1}{n(T-1)}\mathrm{E}[H^*_{nT}(\theta_0,\gamma)]$ is non-singular,

(iv) $\frac{(nT)^{\tau}}{n(T-1)}S^*_{\theta,nT}(\theta_0,\gamma)=O_p(1)$.

The Hessian matrix $H^*_{nT}(\theta,\gamma)$ has the following components:

$$
\begin{aligned}
H^*_{\beta\theta} &= \tfrac{1}{\sigma^2}\textstyle\sum_{t=1}^T[\widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{\mathbb{X}}_t(\gamma),\ \ \widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{\mathcal{Y}}_t,\ \ \widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{\mathcal{Y}}_t^\circ(\gamma),\ \ \widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{V}_t(\phi,\gamma)],\\
H^*_{\lambda_1\lambda_1} &= \tfrac{1}{\sigma^2}\textstyle\sum_{t=1}^T\widetilde{\mathcal{Y}}_t'J_n\widetilde{\mathcal{Y}}_t + \bar{T}\sum_{t=1}^T\mathtt{tr}(G_t^2(\omega)),\\
H^*_{\lambda_1\lambda_2} &= \tfrac{1}{\sigma^2}\textstyle\sum_{t=1}^T\widetilde{\mathcal{Y}}_t'J_n\widetilde{\mathcal{Y}}_t^\circ(\gamma) + \bar{T}\sum_{t=1}^T\mathtt{tr}(d_t(\gamma)G_t^2(\omega)),\\
H^*_{\lambda_1\sigma^2} &= \tfrac{1}{\sigma^4}\textstyle\sum_{t=1}^T\widetilde{\mathcal{Y}}_t'J_n\widetilde{V}_t(\phi,\gamma),\\
H^*_{\lambda_2\lambda_2} &= \tfrac{1}{\sigma^2}\textstyle\sum_{t=1}^T\widetilde{\mathcal{Y}}_t^{\circ\prime}(\gamma)J_n\widetilde{\mathcal{Y}}_t^\circ(\gamma) + \bar{T}\sum_{t=1}^T\mathtt{tr}[(d_t(\gamma)G_t(\omega))^2],\\
H^*_{\lambda_2\sigma^2} &= \tfrac{1}{\sigma^4}\textstyle\sum_{t=1}^T\widetilde{\mathcal{Y}}_t^{\circ\prime}(\gamma)J_n\widetilde{V}_t(\phi,\gamma),\\
H^*_{\sigma^2\sigma^2} &= \tfrac{1}{2\sigma^6}[2\textstyle\sum_{t=1}^T\widetilde{V}_t'(\phi,\gamma)J_n\widetilde{V}_t(\phi,\gamma) - n(T-1)\sigma^2].
\end{aligned}
$$

To prove (i), we note that $\widetilde{\mathcal{Y}}_t = \widetilde{Z}_t + \widetilde{\mathcal{V}}_t$, $\widetilde{\mathcal{Y}}_t^\circ(\gamma) = \widetilde{Z}_t^\circ(\gamma) + \widetilde{\mathcal{V}}_t^\circ(\gamma)$ and $J_n\widetilde{V}_t(\phi,\gamma) = J_n\widetilde{\mathbb{Y}}_t(\omega) - J_n\widetilde{\mathbb{X}}_t(\gamma)\beta = J_n\widetilde{\mathbb{X}}_t(\gamma)(\beta_0-\beta) + J_n\widetilde{\mathcal{D}}_t(\gamma)\phi^\dagger + J_n\widetilde{\zeta}_t(\gamma)\lambda^\dagger + J_n\widetilde{V}_t$ by (B.1). Hence, for any given γ, $\frac{1}{n(T-1)}H^*_{nT}(\bar{\theta},\gamma) = O_p(1)$ by Lemma A.1 and A.2. As $\hat{\theta}_{nT} - \theta_0 \xrightarrow{p} 0$, we have $\bar{\theta} - \theta_0 = o_p(1)$. Noting that σ^{-p} appears in $H^*_{nT}(\theta)$ multiplicatively for $p = 2,4,6$ and $\bar{\sigma}^{-p} = \sigma_0^{-p} + o_p(1)$, we have $\frac{1}{n(T-1)}H^*_{nT}(\bar{\theta},\gamma) = \frac{1}{n(T-1)}H^*_{nT}(\bar{\phi},\sigma_0^2,\gamma) + o_p(1)$. Thus, it is equivalent to showing $\frac{1}{n(T-1)}[H^*_{nT}(\bar{\phi},\sigma_0^2,\gamma) - H^*_{nT}(\theta_0,\gamma)] \xrightarrow{p} 0$. As the proofs for all the components in $H^*_{nT}(\bar{\phi},\sigma_0^2,\gamma)$ are similar, we only show one of them for example,

$$
\begin{aligned}
&\tfrac{1}{n(T-1)}[H^*_{\lambda_1\sigma^2}(\bar{\phi},\sigma_0^2,\gamma) - H^*_{\lambda_1\sigma^2}(\theta_0,\gamma)] = \tfrac{1}{n(T-1)\sigma_0^4}\textstyle\sum_{t=1}^T\widetilde{\mathcal{Y}}_t'J_n[\widetilde{V}_t(\bar{\phi},\gamma) - \widetilde{V}_t(\phi_0,\gamma)]\\
&= -\tfrac{1}{n(T-1)\sigma_0^4}\textstyle\sum_{t=1}^T(\widetilde{Z}_t + \widetilde{\mathcal{V}}_t)'J_n[\widetilde{\mathbb{X}}_t(\gamma)(\bar{\beta}-\beta_0) + (\widetilde{\mathbb{Z}}_t(\gamma) + \widetilde{\mathbb{V}}_t(\gamma))(\bar{\lambda}-\lambda_0)] = o_p(1),
\end{aligned}
$$

by Lemma A.1 and A.2, and $\bar{\theta} - \theta_0 = o_p(1)$.

To prove (ii), we note that $\widetilde{\mathcal{Y}}_t = \widetilde{Z}_t + \widetilde{\mathcal{V}}_t$, $\widetilde{\mathcal{Y}}_t^\circ(\gamma) = \widetilde{Z}_t^\circ(\gamma) + \widetilde{\mathcal{V}}_t^\circ(\gamma)$ and $J_n\widetilde{V}_t(\phi_0,\gamma) = J_n\widetilde{V}_t + J_n\widetilde{H}_t^\star(\gamma_0,\gamma)\phi_{20} + \lambda_{20}J_n\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)$ by (B.1). Hence, $\frac{1}{n(T-1)}[H^*_{nT}(\theta_0,\gamma) - \mathrm{E}(H^*_{nT}(\theta_0,\gamma))] = o_p(1)$ is directly followed by Lemma A.4 and $\phi_{20} = O((nT)^{-\alpha})$, similar to the proof of Theorem 2.1.

To prove (iii), using the facts that $\lambda_{20} = O((nT)^{-\tau})$ and the elements of $G_t(\lambda_0,\gamma)d_t(\gamma,\gamma_0)G_t$ are uniformly bounded, we have $\frac{1}{n(T-1)}\sum_{t=1}^T\mathtt{tr}[G_t(\lambda_0,\gamma) - G_t] = \frac{1}{n(T-1)}\sum_{t=1}^T\mathtt{tr}[G_t(\lambda_0,\gamma)(I_n - A_t(\lambda_0,\gamma)A_t^{-1})] = \frac{\lambda_{20}}{n(T-1)}\sum_{t=1}^T\mathtt{tr}[G_t(\lambda_0,\gamma)\, d_t(\gamma,\gamma_0)G_t] = O_p((nT)^{-\tau})$. Meanwhile, for any $n\times n$ matrix Π_t with bounded row and column sum norms, $\frac{1}{n(T-1)}\sum_{t=1}^T[\mathtt{tr}(J_n\Pi_t) - \mathtt{tr}\Pi_t] = O_p(\frac{1}{n})$. Thus, one shows that $\frac{1}{n(T-1)}\mathrm{E}[H^*_{nT}(\theta_0,\gamma)] - \Sigma_{nT}(\theta_0,\gamma) = o(1)$ for any γ, where

$$\Sigma_{nT}(\theta_0,\gamma) = \begin{bmatrix} \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}_{1,nT}(\gamma), & \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}_{2,nT}(\gamma), & 0_{2k\times 1}, \\ \sim, & \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}_{3,nT}(\gamma) + \bar{T}\mathrm{E}\mathbb{S}_{nT}(\gamma), & \frac{\bar{T}}{\sigma_0^2}\mathrm{E}\mathcal{S}_{nT}(\gamma), \\ \sim, & \sim, & \frac{1}{2\sigma_0^4}, \end{bmatrix}, \tag{B.13}$$

with $\mathbb{S}_{nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T}[\mathtt{tr}(G_tG_t^s), \mathtt{tr}[G_t^\circ(\gamma)G_t^s]; \mathtt{tr}[G_t^{\circ s}(\gamma)G_t], \mathtt{tr}[G_t^\circ(\gamma)G_t^{\circ s}(\gamma)]]$, $\mathcal{S}_{nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T}[\mathtt{tr}(G_t),\ \mathtt{tr}(G_t^\circ(\gamma))]'$ and $G_t^\circ(\gamma) = d_t(\gamma)G_t$.

Therefore, it amounts to prove the limit of $\Sigma_{nT}(\theta_0,\gamma)$ is nonsingular on Γ, which follows if $\Sigma_{nT}(\theta_0,\gamma)p = 0$ implies $p = 0$, where $p = (p_1', p_2', p_3)$, p_1 is a $2k\times 1$ vector, p_2 a 2×1 vector, and p_3 a scalar. The first row block of the linear equation system $\Sigma_{nT}(\theta_0,\gamma)p = 0$ is $p_1 = -\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\mathrm{E}\mathcal{P}_{2,nT}(\gamma)p_2$, while the last row shows $p_3 = -2\sigma_0^2\mathrm{E}\mathcal{S}'_{nT}(\gamma)p_2$. Substituting them into the remain equation of the linear system gives us

$$[\tfrac{1}{\sigma_0^2}(\mathrm{E}\mathcal{P}_{3,nT}(\gamma) - \mathrm{E}\mathcal{P}'_{2,nT}(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\mathrm{E}\mathcal{P}_{2,nT}(\gamma)) + \tfrac{1}{2n(T-1)}\textstyle\sum_{t=1}^{T}\mathrm{E}\mathbb{C}_t^{[2]}(\gamma)]p_2 = 0,$$

where $\mathbb{C}_t^{[2]}(\gamma)$ is the submatrix of $\mathbb{C}_t(\gamma)$ by deleting its third row and column. As shown before, the first term in the square bracket is positive semi-definite. Meanwhile, $\frac{1}{n(T-1)}\sum_{t=1}^{T}\mathrm{E}\mathbb{C}_t^{[2]}(\gamma)$ is also positive semi-definite because

$$\tfrac{1}{n(T-1)}\textstyle\sum_{t=1}^{T} z'\mathbb{C}_t^{[2]}(\gamma)z = \tfrac{1}{2n(T-1)}\textstyle\sum_{t=1}^{T}\mathtt{tr}[(z_1\mathbb{C}_{1,t}^s + z_2\mathbb{C}_{2,t}^s(\gamma))'(z_1\mathbb{C}_{1,t}^s + z_2\mathbb{C}_{2,t}^s(\gamma))] \geq 0$$

for all $z = (z_1, z_2)'$ in $\mathbf{R}^2$. Under Assumption H, either $\mathcal{P}_{3,nT}(\gamma) - \mathcal{P}'_{2,nT}(\gamma)\mathcal{P}_{1,nT}^{-1}(\gamma)$ $\mathcal{P}_{2,nT}(\gamma)$ or $\frac{1}{n(T-1)}\sum_{t=1}^{T}\mathrm{E}\mathbb{C}_t^{[2]}(\gamma)$ is strictly positive definite. Therefore, we must have $p_2 = 0$ from the above equation, implying both $p_1 = 0$ and $p_3 = 0$ by the first and last equations of the linear system. Hence, the non-singularity of $\lim_{nT\to\infty}\Sigma_{nT}(\theta_0,\gamma)$ follows.

To prove (iv), using $\widetilde{\mathcal{Y}}_t = \widetilde{Z}_t + \widetilde{\mathcal{V}}_t$, $\widetilde{\mathcal{Y}}_t^\circ(\gamma) = \widetilde{Z}_t^\circ(\gamma) + \widetilde{\mathcal{V}}_t^\circ(\gamma)$ and $J_n\widetilde{V}_t(\phi_0,\gamma) = J_n\widetilde{V}_t + J_n\widetilde{H}_t^\star(\gamma_0,\gamma)\phi_{20} + \lambda_{20}J_n\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)$, we can split $S^*_{\theta,nT}(\theta_0,\gamma)$ into four components,

$$S^*_{\theta,nT}(\theta_0,\gamma) = S^{*u}_{\theta,nT}(\theta_0,\gamma) + \textstyle\sum_{r=1}^{3} B_{r,nT}(\theta_0,\gamma), \tag{B.14}$$

where

$$S^{*u}_{\theta,nT}(\theta_0,\gamma)=\begin{cases}\frac{1}{\sigma_0^2}\sum_{t=1}^T \widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{V}_t,\\ \frac{1}{\sigma_0^2}\sum_{t=1}^T \widetilde{Z}_t'J_n\widetilde{V}_t+\frac{1}{\sigma_0^2}\sum_{t=1}^T[\widetilde{\mathcal{V}}_t'J_n\widetilde{V}_t-\sigma_0^2\bar{T}\mathtt{tr}(J_nG_t)],\\ \frac{1}{\sigma_0^2}\sum_{t=1}^T \widetilde{Z}_t^{\circ\prime}(\gamma)J_n\widetilde{V}_t+\frac{1}{\sigma_0^2}\sum_{t=1}^T[\widetilde{\mathcal{V}}_t^{\circ\prime}(\gamma)J_n\widetilde{V}_t-\sigma_0^2\bar{T}\mathtt{tr}(J_nd_t(\gamma)G_t)],\\ \frac{1}{2\sigma_0^4}\sum_{t=1}^T[\widetilde{V}_t'J_n\widetilde{V}_t-(n-1)\bar{T}\sigma_0^2],\end{cases}$$

$$B_{1,nT}(\theta_0,\gamma)=\begin{cases}\frac{1}{\sigma_0^2}\sum_{t=1}^T \widetilde{\mathbb{X}}_t'(\gamma)J_n[\widetilde{H}_t^\star(\gamma_0,\gamma)\phi_{20}+\lambda_{20}\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)],\\ \frac{1}{\sigma_0^2}\sum_{t=1}^T(\widetilde{Z}_t+\widetilde{\mathcal{V}}_t)'J_n[\widetilde{H}_t^\star(\gamma_0,\gamma)\phi_{20}+\lambda_{20}\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)],\\ \frac{1}{\sigma_0^2}\sum_{t=1}^T[\widetilde{Z}_t^\circ(\gamma)+\widetilde{\mathcal{V}}_t^\circ(\gamma)]'J_n[\widetilde{H}_t^\star(\gamma_0,\gamma)\phi_{20}+\lambda_{20}\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)]\\ \frac{1}{2\sigma_0^4}\sum_{t=1}^T[\widetilde{H}_t^\star(\gamma_0,\gamma)\phi_{20}+\lambda_{20}\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)]'J_n[2\widetilde{V}_t+\widetilde{H}_t^\star(\gamma_0,\gamma)\phi_{20}\\ \qquad\qquad+\lambda_{20}\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)],\end{cases}$$

$B_{2,nT}(\theta_0,\gamma)=-\big(0_{1\times 2k},\bar{T}\sum_{t=1}^T\mathtt{tr}(J_n(G_t(\lambda_0,\gamma)-G_t)),\bar{T}\sum_{t=1}^T\mathtt{tr}(J_nd_t(\gamma)(G_t(\lambda_0,\gamma)-G_t)),0\big)'$, and $B_{3,nT}(\theta_0,\gamma)=-\big(0_{1\times 2k},\bar{T}\sum_{t=1}^T\frac{1}{n}l_n'G_t(\lambda_0,\gamma)l_n,\bar{T}\sum_{t=1}^T\frac{1}{n}l_n'd_t(\gamma) G_t(\lambda_0,\gamma)l_n,\frac{T-1}{2\sigma_0^2}\big)'$.

Note that $\frac{(nT)^\tau}{n(T-1)}=\frac{(nT)^{\tau-1}T}{T-1}$. Thus, it suffices to show that $(nT)^{\tau-1}S^{*u}_{\theta,nT}(\theta_0,\gamma)$ and $(nT)^{\tau-1}B_{r,nT}(\theta_0,\gamma)$ for $r=1,2,3$ are all bounded for any γ. By Lemma A.3 and Lemma B.1, $S^{*u}_{\theta,nT}(\theta_0,\gamma)=O_p(\sqrt{nT})$, uniformly in $\gamma\in\Gamma$. Since $\tau\in(0,\frac{1}{2})$, $(nT)^{\tau-1}S^{*u}_{\theta,nT}(\theta_0,\gamma)=\frac{(nT)^{\tau-\frac{1}{2}}}{\sqrt{nT}}S^{*u}_{\theta,nT}(\theta_0,\gamma)=o_p(1)$. As for $B_{1,nT}(\theta_0,\gamma)$, note that $\phi_{20}=(nT)^{-\tau}\delta_0$, where $\delta_0=(b_0',l_0)'$, by Assumption F. Thus, it is easy to see that $(nT)^{\tau-1}B_{1,nT}(\theta_0,\gamma)=O_p(1)$ uniformly in $\gamma\in\Gamma$. We show the third component of $B_{1,nT}(\theta_0,\gamma)$ for example as the others can be shown similarly. By Lemma A.4, we have

$$\begin{aligned}&\frac{(nT)^{\tau-1}}{\sigma_0^2}\sum_{t=1}^T[\widetilde{Z}_t^\circ(\gamma)+\widetilde{\mathcal{V}}_t^\circ(\gamma)]'J_n[\widetilde{H}_t^\star(\gamma_0,\gamma)\phi_{20}+\lambda_{20}\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)]\\ &=\frac{1}{\sigma_0^2nT}\sum_{t=1}^T[\widetilde{Z}_t^\circ(\gamma)+\widetilde{\mathcal{V}}_t^\circ(\gamma)]'J_n[\widetilde{H}_t^\star(\gamma_0,\gamma)\delta_0+l_0\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)]\\ &=\frac{1}{\sigma_0^2nT}\sum_{t=1}^T\mathrm{E}[\widetilde{Z}_t^{\circ\prime}(\gamma)J_n\widetilde{H}_t^\star(\gamma_0,\gamma)\delta_0+\widetilde{\mathcal{V}}_t^{\circ\prime}(\gamma)J_n\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)l_0]+o_p(1)=O_p(1).\end{aligned}$$

Similarly, we also have $(nT)^{\tau-1}B_{2,nT}(\theta_0,\gamma)=O_p(1)$ uniformly in $\gamma\in\Gamma$. We show one of the two non-zero elements in $B_{2,nT}(\theta_0,\gamma)$ for example, as the the other can be shown similarly. Noting that $G_t(\lambda_0,\gamma)-G_t=G_t(\lambda_0,\gamma)(I_n-$

$A_t(\lambda_0,\gamma)A_t^{-1}) = \lambda_{20}G_t(\lambda_0,\gamma)d_t(\gamma,\gamma_0)G_t$, one has

$$
\begin{aligned}
&(nT)^{\tau-1}\bar{T}\sum_{t=1}^T \mathtt{tr}(J_n d_t(\gamma)(G_t(\lambda_0,\gamma)-G_t)) \\
=\ & \tfrac{l_0\bar{T}}{nT}\sum_{t=1}^T \mathtt{tr}(J_n d_t(\gamma)G_t(\lambda_0,\gamma)d_t(\gamma,\gamma_0)G_t) = O_p(1). \qquad \text{(B.15)}
\end{aligned}
$$

Finally, we show $(nT)^{\tau-1}B_{3,nT}(\theta_0,\gamma) = o_p(1)$. Because the nonzero elements in $B_{3,nT}(\theta_0,\gamma)$ is either $O(T)$ or $O_p(T)$, the elements of $(nT)^{\tau-1}B_{3,nT}(\theta_0,\gamma)$ are either $O(\frac{T^\tau}{n^{1-\tau}})$ or $O_p(\frac{T^\tau}{n^{1-\tau}})$. As $\frac{T}{n}\to a$ and $\tau\in(0,\frac{1}{2})$, $\frac{T^\tau}{n^{1-\tau}} = \frac{a^\tau}{n^{1-2\tau}} = o(1)$. Thus, the desired result holds.

Proof of (b): Note that

$$
\begin{aligned}
\tfrac{(nT)^{2\tau}}{n(T-1)}[\ell_{nT}^{*c}(\hat{\omega}_{nT}) - \ell_{nT}^{*c}(\omega_0)] =\ & -\tfrac{(nT)^{2\tau}}{2}[\ln\hat{\sigma}_{nT}^2 - \ln\hat{\sigma}_{nT}^2(\omega_0)] \\
& + \tfrac{(nT)^{2\tau}}{n(T-1)}\sum_{t=1}^T \ln|A_t(\hat{\omega}_{nT})A_t^{-1}|.
\end{aligned}
$$

Let $\hat{\lambda}_{nT}^\dagger = ((\lambda_0-\hat{\lambda}_{nT})',\lambda_{20})'$ and $\hat{\phi}_{nT}^\dagger = ((\lambda_0-\hat{\lambda}_{nT})',\phi_{20}')'$. By (a) and Assumption F, $\hat{\phi}_{nT}^\dagger,\hat{\lambda}_{nT}^\dagger = O((nT)^{-\tau})$. Thus, using (B.3) and $\hat{\sigma}_{nT}^2(\omega_0) = \mathcal{V}_{1,nT}$, we have

$$
\begin{aligned}
\hat{\sigma}_{nT}^2 - \hat{\sigma}_{nT}^2(\omega_0) =\ & 2\mathcal{V}_{2,nT}(\hat{\gamma}_{nT})\hat{\lambda}_{nT}^\dagger + \hat{\lambda}_{nT}^{\dagger\prime}\mathcal{V}_{3,nT}(\hat{\gamma}_{nT})\hat{\lambda}_{nT}^\dagger \\
& + \hat{\phi}_{nT}^{\dagger\prime}[\mathcal{D}_{1,nT}(\hat{\gamma}_{nT}) - \mathcal{D}_{2,nT}'(\hat{\gamma}_{nT})\mathcal{P}_{1,nT}^{-1}(\hat{\gamma}_{nT})\mathcal{D}_{2,nT}(\hat{\gamma}_{nT})]\hat{\phi}_{nT}^\dagger \\
& + o_p((nT)^{-2\tau}),
\end{aligned}
$$

where the first term is $O_p((nT)^{-\tau})$, and the second and third are $O_p((nT)^{-2\tau})$, by Lemma A.4 and Lemma B.1. Then, using the Taylor expansion for the logarithm, we have

$$
\begin{aligned}
& -\tfrac{(nT)^{2\tau}}{2}[\ln\hat{\sigma}_{nT}^2 - \ln\hat{\sigma}_{nT}^2(\omega_0)] \\
=\ & -\tfrac{(nT)^{2\tau}}{2\hat{\sigma}_{nT}^2(\omega_0)}\{2\mathcal{V}_{2,nT}(\hat{\gamma}_{nT})\hat{\lambda}_{nT}^\dagger + \hat{\lambda}_{nT}^{\dagger\prime}\mathcal{V}_{3,nT}(\hat{\gamma}_{nT})\hat{\lambda}_{nT}^\dagger \\
& + \hat{\phi}_{nT}^{\dagger\prime}[\mathcal{D}_{1,nT}(\hat{\gamma}_{nT}) - \mathcal{D}_{2,nT}'(\hat{\gamma}_{nT})\mathcal{P}_{1,nT}^{-1}(\hat{\gamma}_{nT})\mathcal{D}_{2,nT}(\hat{\gamma}_{nT})]\hat{\phi}_{nT}^\dagger\} \\
& + \tfrac{(nT)^{2\tau}}{\hat{\sigma}_{nT}^4(\omega_0)}\hat{\lambda}_{nT}^{\dagger\prime}\mathcal{V}_{2,nT}'(\hat{\gamma}_{nT})\mathcal{V}_{2,nT}(\hat{\gamma}_{nT})\hat{\lambda}_{nT}^\dagger + o_p(1) \\
=\ & -(nT)^{2\tau}[\mathcal{G}_{1,nT}'(\hat{\gamma}_{nT})\hat{\lambda}_{nT}^\dagger + \tfrac{1}{2}\hat{\lambda}_{nT}^{\dagger\prime}\mathbb{G}_{1,nT}(\hat{\gamma}_{nT})\hat{\lambda}_{nT}^\dagger] \\
& - (nT)^{2\tau}\hat{\lambda}_{nT}^{\dagger\prime}\mathcal{G}_{1,nT}(\hat{\gamma}_{nT})\mathcal{G}_{1,nT}'(\hat{\gamma}_{nT})\hat{\lambda}_{nT}^\dagger \\
& - \tfrac{(nT)^{2\tau}}{2\sigma_0^2}\hat{\phi}_{nT}^{\dagger\prime}[\mathcal{D}_{1,nT}(\hat{\gamma}_{nT}) - \mathcal{D}_{2,nT}'(\hat{\gamma}_{nT})(\mathcal{P}_{1,nT}(\hat{\gamma}_{nT}))^{-1}\mathcal{D}_{2,nT}(\hat{\gamma}_{nT})]\hat{\phi}_{nT}^\dagger + o_p(1), \qquad \text{(B.16)}
\end{aligned}
$$

because $\hat{\sigma}_{nT}^2(\omega_0) - \sigma_0^2 = o_p(1)$, $\mathcal{V}_{2,nT}(\gamma) - \sigma_0^2\mathcal{G}_{1,nT}'(\gamma) = o_p(1)$ and $\mathcal{V}_{3,nT}(\gamma) -$

$\sigma_0^2\mathbb{G}_{1,nT}(\gamma) = o_p(1)$, uniformly in $\gamma \in \Gamma$.

As for $\frac{(nT)^{2\tau}}{n(T-1)}\sum_{t=1}^T \ln|A_t(\hat{\omega}_{nT})A_t^{-1}|$, one has $A_t(\hat{\omega}_{nT})A_t^{-1} = I_n + (\lambda_{10} - \hat{\lambda}_{1,nT})G_t + (\lambda_{20} - \hat{\lambda}_{2,nT})d_t(\hat{\gamma}_{nT})G_t + \lambda_{20}d_t(\gamma_0, \hat{\gamma}_{nT})G_t$ and $\hat{\lambda}_{nT}^\dagger = O((nT)^{-\tau})$. Thus, we have $\ln|A_t(\hat{\omega}_{nT})A_t^{-1}| = \mathtt{tr}[\ln(A_t(\hat{\omega}_{nT})A_t^{-1})] = \mathtt{tr}\big[\sum_{m=1}^\infty(-1)^{m+1} \frac{(A_t(\hat{\omega}_{nT})A_t^{-1}-I_n)^m}{m}\big]$ by Theorem 2.8 of Hall (2015). Therefore, we have

$$\frac{(nT)^{2\tau}}{n(T-1)}\sum_{t=1}^T \ln|A_t(\hat{\omega}_{nT})A_t^{-1}| = (nT)^{2\tau}\mathcal{G}_{2,nT}'(\hat{\gamma}_{nT})\hat{\lambda}_{nT}^\dagger - \frac{(nT)^{2\tau}}{2}\hat{\lambda}_{nT}^{\dagger\prime}\mathbb{G}_{2,nT}(\hat{\gamma}_{nT})\hat{\lambda}_{nT}^\dagger + o_p(1), \quad \text{(B.17)}$$

where $\mathcal{G}_{2,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^T[\mathtt{tr}(G_t), \mathtt{tr}(d_t(\gamma)G_t), \mathtt{tr}(d_t(\gamma_0,\gamma)G_t)]'$ and

$$\mathbb{G}_{2,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^T \begin{bmatrix} \mathtt{tr}(G_t^2), & \mathtt{tr}(d_t(\gamma)G_t^2), & \mathtt{tr}(d_t(\gamma_0,\gamma)G_t^2), \\ \sim, & \mathtt{tr}(G_t^2(\gamma)), & \mathtt{tr}(d_t(\gamma_0,\gamma)G_t d_t(\gamma)G_t), \\ \sim, & \sim, & \mathtt{tr}((d_t(\gamma_0,\gamma)G_t)^2), \end{bmatrix}.$$

Note that $\mathcal{G}_{1,nT}(\gamma) = \mathcal{G}_{2,nT}(\gamma) + O_p(\frac{1}{n})$ and thus $(nT)^\tau[\mathcal{G}_{1,nT}(\hat{\gamma}_{nT}) - \mathcal{G}_{2,nT}(\hat{\gamma}_{nT})] = O_p(\frac{T^\tau}{n^{1-\tau}}) = o_p(1)$. Besides, $\mathbb{G}_{1,nT}(\gamma) = \mathbb{G}_{3,nT}(\gamma) + O_p(\frac{1}{n})$, where

$$\mathbb{G}_{3,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^T \begin{bmatrix} \mathtt{tr}(G_tG_t'), & \mathtt{tr}(d_t(\gamma)G_tG_t'), & \mathtt{tr}(d_t(\gamma_0,\gamma)G_tG_t'), \\ \sim, & \mathtt{tr}(d_t(\gamma)G_tG_t'), & \mathtt{tr}(d_t(\gamma_0,\gamma)G_tG_t'd_t(\gamma)), \\ \sim, & \sim, & \mathtt{tr}(d_t(\gamma_0,\gamma)G_tG_t'), \end{bmatrix}.$$

Combining these with (B.16) and (B.17), we have

$$\begin{aligned} &\frac{(nT)^{2\tau}}{n(T-1)}[\ell_{nT}^{*c}(\hat{\omega}_{nT}) - \ell_{nT}^{*c}(\omega_0)] \\ = &-\frac{(nT)^{2\tau}}{2}\hat{\lambda}_{nT}^{\dagger\prime}[\mathbb{G}_{2,nT}(\hat{\gamma}_{nT}) + \mathbb{G}_{3,nT}(\hat{\gamma}_{nT}) - 2\mathcal{G}_{2,nT}(\hat{\gamma}_{nT})\mathcal{G}_{2,nT}'(\hat{\gamma}_{nT})]\hat{\lambda}_{nT}^\dagger \\ &-\frac{(nT)^{2\tau}}{2\sigma_0^2}\hat{\phi}_{nT}^{\dagger\prime}[\mathcal{D}_{1,nT}(\hat{\gamma}_{nT}) - \mathcal{D}_{2,nT}'(\hat{\gamma}_{nT})\mathcal{P}_{1,nT}^{-1}(\hat{\gamma}_{nT})\mathcal{D}_{2,nT}(\hat{\gamma}_{nT})]\hat{\phi}_{nT}^\dagger + o_p(1). \end{aligned} \quad \text{(B.18)}$$

Firstly, it is easy to see that $\mathbb{G}_{2,nT}(\gamma) + \mathbb{G}_{3,nT}(\gamma) - 2\mathcal{G}_{2,nT}(\gamma)\mathcal{G}_{2,nT}'(\gamma) = \frac{1}{2n(T-1)}\sum_{t=1}^T \mathbb{C}_t(\gamma)$, which is positive semi-definite as $\frac{1}{2n(T-1)}\sum_{t=1}^T z'\mathbb{C}_t(\gamma)z = \frac{1}{2n(T-1)}\sum_{t=1}^T \mathtt{tr}[(z_1\mathbb{C}_{1,t}^s + z_2\mathbb{C}_{2,t}^s(\gamma) + z_3\mathbb{C}_{3,t}^s(\gamma))'(z_1\mathbb{C}_{1,t}^s + z_2\mathbb{C}_{2,t}^s(\gamma) + z_3\mathbb{C}_{3,t}^s(\gamma))] \geq 0$ for all $z = (z_1, z_2, z_3)'$ in $\mathbf{R}^3$. Thus, the first term of (B.18) is non-positive. Secondly, for a comformable vector d, $d'[\mathcal{D}_{1,nT}(\gamma) - \mathcal{D}_{2,nT}'(\gamma)\mathcal{P}_{1,nT}^{-1}(\gamma)\mathcal{D}_{2,nT}(\gamma)]d$ can be written into the form of $a'Qa$ with some $nT \times 1$ vector a and $nT \times nT$ idempotent matrix Q, so that the second term of (B.18) is also non-positive. Therefore, we have $\lim_{nT\to\infty}\frac{(nT)^{2\tau}}{n(T-1)}[\ell_{nT}^{*c}(\hat{\omega}_{nT}) - \ell_{nT}^{*c}(\omega_0)] \leqslant 0$. Under Assumption H($i$), we have $\rho_{min}\big(\mathcal{D}_{1,nT}(\hat{\gamma}_{nT})-\mathcal{D}'_{2,nT}(\hat{\gamma}_{nT})\mathcal{P}^{-1}_{1,nT}(\hat{\gamma}_{nT})\mathcal{D}_{2,nT}(\hat{\gamma}_{nT})\big) \geq \rho_{min}\big(\frac{1}{n(T-1)}\sum_{t=1}^{T} \widetilde{\mathcal{H}}'_t(\gamma)J_n\widetilde{\mathcal{H}}_t(\gamma)\big) \geq c|\hat{\gamma}_{nT}-\gamma_0|$ by Theorem 5 of Smith (1992). It follows that

$$\lim_{nT\to\infty} \tfrac{(nT)^{2\tau}}{n(T-1)}[\ell^{*c}_{nT}(\hat{\omega}_{nT}) - \ell^{*c}_{nT}(\omega_0)] \leqslant -\tfrac{1}{2\sigma_0^2}c|\hat{\gamma}_{nT}-\gamma_0|\|(nT)^{\tau}\hat{\phi}^{\dagger}_{nT}\|^2.$$

By the definition of $\hat{\omega}_{nT}$, we have $\frac{(nT)^{2\tau}}{n(T-1)}(\ell^{*c}_{nT}(\hat{\omega}_{nT}) - \ell^{*c}_{nT}(\omega_0)) \geq 0$. Hence, we must have that $|\hat{\gamma}_{nT}-\gamma_0| = o_p(1)$. Similarly, Assumption H(ii) can also guarantee that $\hat{\gamma}_{nT}-\gamma_0 \xrightarrow{p} 0$. ■

Proof of Theorem 2.3: We first show that $(nT)^{\tau}(\hat{\theta}_{nT}-\theta_0) = o_p(1)$. Given the results (i)-(iv) from the proof of Theorem 2.2, we only need to show that $(nT)^{\tau-1}B_{r,nT}(\theta_0,\hat{\gamma}_{nT}) = o_p(1)$ for $r=1,2$, which is directly implied by the consistency of $\hat{\gamma}_{nT}$. Then, let B, k, l and $\mathcal{N}_{nT}$ be defined in Lemma B.4, and $M \equiv \max\big(k,l,\|\delta_0\|,|l_0|,1,\sigma_0^2\big)$. Pick $\eta,\kappa > 0$ small enough such that $\max(\eta,\kappa) < M$ and

$$\mathcal{M}_0 \equiv -\tfrac{1}{2}k - \tfrac{k}{\sigma_0^2+\kappa} + \tfrac{1}{2}(M\eta+6M^3\kappa) + \tfrac{1}{\sigma_0^2-\kappa}(4M\eta+8M^2\eta+18M^3\kappa+4M^4\kappa) < 0.$$

Let E_{nT} be the joint event that (1) $|\hat{\gamma}_{nT}-\gamma_0| \leqslant B$, (2) $\frac{T^{\tau}}{n^{1-\tau}} < \kappa$, (3) $(nT)^{\tau}|\hat{\theta}_{nT}-\theta_0| \leqslant \kappa$, (4) $\inf_{\gamma\in\mathcal{N}_{nT}} \frac{D_{r,nT}(\gamma)}{|\gamma-\gamma_0|} > (1-\eta)k$, (5) $\sup_{\gamma\in\mathcal{N}_{nT}} \frac{\|F_{s,nT}(\gamma)\|}{|\gamma-\gamma_0|} < (1+\eta)l$, (6) $\sup_{\gamma\in\mathcal{N}_{nT}} \frac{|K_{s,nT}(\gamma)|}{|\gamma-\gamma_0|} < (1+\eta)l$, (7) $\sup_{\gamma\in\mathcal{N}_{nT}} \frac{|L_{r,nT}(\gamma)|}{|\gamma-\gamma_0|} < (1+\eta)l$, (8) $\sup_{\gamma\in\mathcal{N}_{nT}} \frac{\|P_{r,nT}(\gamma)\|}{|\gamma-\gamma_0|} < \eta$, (9) $\sup_{\gamma\in\mathcal{N}_{nT}} \frac{\|\mathcal{J}_{s,nT}(\gamma)-\mathcal{J}_{s,nT}(\gamma_0)\|}{\sqrt{a_{nT}}|\gamma-\gamma_0|} < \eta$, for $s=1,2$ and $r=1,2,3$, and (10) will be established later.

Let $\ell^{\ddagger}_{nT}(\gamma) = \ell^{*}_{nT}(\hat{\theta}_{nT},\gamma)$. We have,

$$\begin{aligned}\ell^{\ddagger}_{nT}(\gamma) - \ell^{\ddagger}_{nT}(\gamma_0) &= \bar{T}\textstyle\sum_{t=1}^{T}\ln|A_t(\hat{\lambda}_{nT},\gamma)| - \bar{T}\sum_{t=1}^{T}\ln|A_t(\hat{\lambda}_{nT},\gamma_0)| \\ &\quad - \tfrac{1}{2\hat{\sigma}^2_{nT}}[\textstyle\sum_{t=1}^{T}\widetilde{V}'_t(\hat{\phi}_{nT},\gamma)J_n\widetilde{V}_t(\hat{\phi}_{nT},\gamma) - \sum_{t=1}^{T}\widetilde{V}'_t(\hat{\phi}_{nT},\gamma_0)J_n\widetilde{V}_t(\hat{\phi}_{nT},\gamma_0)].\end{aligned} \quad \text{(B.19)}$$

We consider the difference of the first two terms at first. By Theorem 2.8 of Hall (2015),

$$\begin{aligned}&\bar{T}\textstyle\sum_{t=1}^{T}\ln|A_t(\hat{\lambda}_{nT},\gamma)| - \bar{T}\sum_{t=1}^{T}\ln|A_t(\hat{\lambda}_{nT},\gamma_0)| \\ &= \bar{T}\hat{\lambda}_{2,nT}\textstyle\sum_{t=1}^{T}\mathtt{tr}\big[d_t(\gamma_0,\gamma)G_t(\hat{\lambda}_{nT},\gamma_0)\big] \\ &\quad - \tfrac{\bar{T}}{2}\hat{\lambda}^2_{2,nT}\textstyle\sum_{t=1}^{T}\mathtt{tr}\big\{[d_t(\gamma_0,\gamma)G_t(\hat{\lambda}_{nT},\gamma_0)]^2\big\} + \ldots \\ &= \mathcal{A}_1(\gamma)+\mathcal{A}_2(\gamma)+\mathcal{A}_3(\gamma),\end{aligned} \quad \text{(B.20)}$$

where $\mathcal{A}_1(\gamma) = \bar{T}\hat{\lambda}_{2,nT}\sum_{t=1}^{T}\mathtt{tr}\big[d_t(\gamma_0,\gamma)G_t\big]$, $\mathcal{A}_2(\gamma) = -\frac{\bar{T}}{2}\hat{\lambda}^2_{2,nT}\sum_{t=1}^{T}\mathtt{tr}[(d_t(\gamma_0,\gamma)$

$G_t)^2]$ and $\mathcal{A}_3(\gamma)$ stands for all the remaining terms. Noting that $G_t(\hat{\lambda}_{nT},\gamma_0) - G_t = G_t(\hat{\lambda}_{nT},\gamma_0)[(\hat{\lambda}_{1,nT} - \lambda_{10})G_t + (\hat{\lambda}_{2,nT} - \lambda_{20})d_t(\gamma_0)G_t]$, the elements of $[d_t(\gamma_0,\gamma)G_t(\hat{\lambda}_{nT},\gamma_0)]^r$, $r \geq 1$, are uniformly bounded by Lemma A.1, and $\hat{\lambda}_{2,nT} = O_p((nT)^{-\tau})$, implied by $\hat{\lambda}_{2,nT} - \lambda_{20} = o_p(1)$ and $\lambda_{20} = O((nT)^{-\tau})$, we can easily see the series of terms in $\mathcal{A}_3(\gamma)$ have smaller order than $\mathcal{A}_2(\gamma)$, uniformly in $\gamma \in \Gamma$.

To simplify the last term in (B.19), we first derive

$$
\begin{aligned}
V_t(\hat{\phi}_{nT},\gamma) &= \mathbb{Y}_t(\hat{\lambda}_{nT},\gamma) - \mathbb{X}_t(\gamma)\hat{\beta}_{nT} \\
&= V_t + \mu_0 + \alpha_{t0} l_n + \mathbb{H}_t(\gamma_0)(\phi_0 - \hat{\phi}_{nT}) + H_t^\star(\gamma_0,\gamma)\hat{\phi}_{2,nT} \\
&\quad + \mathbb{V}_t(\gamma_0)(\lambda_0 - \hat{\lambda}_{nT}) + \hat{\lambda}_{2,nT}\mathcal{V}_t^\star(\gamma_0,\gamma),
\end{aligned}
$$

and hence

$$
\begin{aligned}
&\textstyle\sum_{t=1}^T \widetilde{V}_t'(\hat{\phi}_{nT},\gamma) J_n \widetilde{V}_t(\hat{\phi}_{nT},\gamma) - \sum_{t=1}^T \widetilde{V}_t'(\hat{\phi}_{nT},\gamma_0) J_n \widetilde{V}_t(\hat{\phi}_{nT},\gamma_0) \\
=&\textstyle\ \sum_{t=1}^T [\widetilde{V}_t(\hat{\phi}_{nT},\gamma) + \widetilde{V}_t(\hat{\phi}_{nT},\gamma_0)]' J_n [\widetilde{V}_t(\hat{\phi}_{nT},\gamma) - \widetilde{V}_t(\hat{\phi}_{nT},\gamma_0)] \\
=&\textstyle\ 2\sum_{t=1}^T \widetilde{V}_t' J_n \widetilde{H}_t^\star(\gamma_0,\gamma)\hat{\phi}_{2,nT} + 2\hat{\lambda}_{2,nT}\sum_{t=1}^T \widetilde{V}_t' J_n \widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma) \\
&\textstyle + \sum_{t=1}^T \hat{\phi}_{2,nT}' \widetilde{H}_t^{\star\prime}(\gamma_0,\gamma) J_n \widetilde{H}_t^\star(\gamma_0,\gamma)\hat{\phi}_{2,nT} \\
&\textstyle + 2\hat{\lambda}_{2,nT}\sum_{t=1}^T \hat{\phi}_{2,nT}' \widetilde{H}_t^{\star\prime}(\gamma_0,\gamma) J_n \widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma) \\
&\textstyle + 2\sum_{t=1}^T (\phi_0 - \hat{\phi}_{nT})' \widetilde{\mathbb{H}}_t'(\gamma_0) J_n \widetilde{H}_t^\star(\gamma_0,\gamma)\hat{\phi}_{2,nT} \\
&\textstyle + 2\hat{\lambda}_{2,nT}\sum_{t=1}^T (\phi_0 - \hat{\phi}_{nT})' \widetilde{\mathbb{H}}_t'(\gamma_0) J_n \widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma) \\
&\textstyle + 2\sum_{t=1}^T (\lambda_0 - \hat{\lambda}_{nT})' \widetilde{\mathbb{V}}_t'(\gamma_0) J_n \widetilde{H}_t^\star(\gamma_0,\gamma)\hat{\phi}_{2,nT} \\
&\textstyle + 2\hat{\lambda}_{2,nT}\sum_{t=1}^T (\lambda_0 - \hat{\lambda}_{nT})' \widetilde{\mathbb{V}}_t'(\gamma_0) J_n \widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma) \\
&\textstyle + \hat{\lambda}_{2,nT}^2 \sum_{t=1}^T \widetilde{\mathcal{V}}_t^{\star\prime}(\gamma_0,\gamma) J_n \widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma) \equiv \sum_{s=1}^9 \mathcal{B}_s(\gamma).
\end{aligned}
\tag{B.21}
$$

From (B.20) and (B.21), one has

$$
\frac{\ell_{nT}^\ddagger(\gamma) - \ell_{nT}^\ddagger(\gamma_0)}{a_{nT}(\gamma - \gamma_0)} \leqslant -\frac{\sum_{s=1}^9 \mathcal{B}_s(\gamma) - 2\hat{\sigma}_{nT}^2\mathcal{A}_1(\gamma)}{2\hat{\sigma}_{nT}^2 a_{nT}(\gamma - \gamma_0)} + \frac{\mathcal{A}_2(\gamma)}{a_{nT}(\gamma - \gamma_0)} + \frac{\mathcal{A}_3(\gamma)}{a_{nT}(\gamma - \gamma_0)}. \tag{B.22}
$$

As is shown latter, $\frac{\mathcal{A}_2(\gamma)}{a_{nT}(\gamma-\gamma_0)}$ are uniformly bounded on the set E_{nT}. This implies that $\frac{\mathcal{A}_3(\gamma)}{a_{nT}(\gamma-\gamma_0)}$ will shrink to zero as sample increase. Therefore, we let $\frac{\mathcal{A}_3(\gamma)}{a_{nT}(\gamma-\gamma_0)} \leqslant \kappa$ be the event (10) of E_{nT}. Fix $\epsilon > 0$, one can choose $\bar{v}$ for large enough (n,T) such that $P(E_{nT}) \geqslant 1 - \epsilon$, by Theorem 2.2, Assumption E,

$(nT)^\tau(\hat{\theta}_{nT}-\theta_0)=o_p(1)$ shown at the beginning and Lemma B.4. Suppose $\gamma\in[\gamma_0+\bar{v}/a_{nT},\gamma_0+B]$ and E_{nT} holds. Let $\hat{l}_{nT}=(nT)^\tau\hat{\lambda}_{2,nT}$ and $\hat{b}_{nT}=(nT)^\tau\hat{\beta}_{2,nT}$ so that $\|\hat{\delta}_{nT}-\delta_0\|\leqslant\kappa$, where $\hat{\delta}_{nT}=(\hat{l}'_{nT},\hat{b}_{nT})'$, by event (3). Besides, we have $\sigma_0^2-\kappa\leqslant\sigma_0^2-\kappa(nT)^{-\tau}\leqslant\hat{\sigma}^2_{nT}\leqslant\sigma_0^2+\kappa(nT)^{-\tau}\leqslant\sigma_0^2+\kappa$. Given these, we are going to study each term in the right-hand side of inequality (B.22).

By events (1), (3), (4) and (7), we have

$$\begin{aligned}\frac{\mathcal{A}_2(\gamma)}{a_{nT}(\gamma-\gamma_0)}=&-\frac{\lambda_{20}^2\bar{T}\sum_{t=1}^T\texttt{tr}[(d_t(\gamma_0,\gamma)G_t)^2]}{2a_{nT}(\gamma-\gamma_0)}\\&-\frac{(\hat{\lambda}_{2,nT}-\lambda_{20})(\hat{\lambda}_{2,nT}+\lambda_{20})\bar{T}\sum_{t=1}^T\texttt{tr}[(d_t(\gamma_0,\gamma)G_t)^2]}{2a_{nT}(\gamma-\gamma_0)}\\\leqslant&-\frac{D_{3,nT}(\gamma)}{2(\gamma-\gamma_0)}+\frac{|\hat{l}_{nT}-l_0||\hat{l}_{nT}+l_0||L_{1,nT}(\gamma)|}{2(\gamma-\gamma_0)}\\\leqslant&-\frac{1}{2}(1-\eta)k+\frac{1}{2}\kappa(2|l_0|+\kappa)(1+\eta)l\\\leqslant&-\frac{1}{2}k+\frac{1}{2}(M\eta+6M^3\kappa).\end{aligned}$$

By events (1), (2), (3), (7) and (9), we have

$$\begin{aligned}&-\frac{\mathcal{B}_1(\gamma)+\mathcal{B}_2(\gamma)-2\hat{\sigma}^2_{nT}\mathcal{A}_1(\gamma)}{2\hat{\sigma}^2_{nT}a_{nT}(\gamma-\gamma_0)}\\=&-\frac{\sum_{t=1}^T\widetilde{V}_t'J_n\widetilde{H}_t^\star(\gamma_0,\gamma)\hat{\phi}_{2,nT}}{\hat{\sigma}^2_{nT}a_{nT}(\gamma-\gamma_0)}\\&-\frac{\hat{\lambda}_{2,nT}\left\{\sum_{t=1}^T\widetilde{V}_t'J_n\widetilde{\mathcal{V}}_t^\star(\gamma_0,\gamma)-\sigma_0^2\bar{T}\sum_{t=1}^T\texttt{tr}[J_nd_t(\gamma_0,\gamma)G_t]\right\}}{\hat{\sigma}^2_{nT}a_{nT}(\gamma-\gamma_0)}\\&+\frac{\sigma_0^2\bar{T}\hat{\lambda}_{2,nT}\sum_{t=1}^T\texttt{tr}[l_nl_n'd_t(\gamma_0,\gamma)G_t]}{\hat{\sigma}^2_{nT}na_{nT}(\gamma-\gamma_0)}-\frac{(\sigma_0^2-\hat{\sigma}^2_{nT})\bar{T}\hat{\lambda}_{2,nT}\sum_{t=1}^T\texttt{tr}[d_t(\gamma_0,\gamma)G_t]}{\hat{\sigma}^2_{nT}a_{nT}(\gamma-\gamma_0)}\\\leqslant&\frac{\|\hat{\delta}_{nT}\|\|\mathcal{J}_{1,nT}(\gamma)-\mathcal{J}_{1,nT}(\gamma_0)\|}{\hat{\sigma}^2_{nT}\sqrt{a_{nT}}(\gamma-\gamma_0)}+\frac{|\hat{l}_{nT}|\|\mathcal{J}_{2,nT}(\gamma)-\mathcal{J}_{2,nT}(\gamma_0)\|}{\hat{\sigma}^2_{nT}\sqrt{a_{nT}}(\gamma-\gamma_0)}\\&+\frac{T^\tau|\hat{l}_{nT}|\sigma_0^2|L_{3,nT}(\gamma)|}{n^{1-\tau}\hat{\sigma}^2_{nT}(\gamma-\gamma_0)}+\frac{\kappa|\hat{l}_{nT}||L_{2,nT}(\gamma)|}{\hat{\sigma}^2_{nT}(\gamma-\gamma_0)}\\\leqslant&\frac{(\|\delta_0\|+\kappa)\eta+(|l_0|+\kappa)[\eta+\kappa\sigma_0^2(1+\eta)l+\kappa(1+\eta)l]}{\hat{\sigma}^2_{nT}}\leqslant\frac{4M\eta+4M^3\kappa+4M^4\kappa}{\sigma_0^2-\kappa}.\end{aligned}$$

Next, by events (1), (3), (4), and (5), we have

$$
\begin{aligned}
-\frac{\mathcal{B}_3(\gamma)}{2\hat{\sigma}_{nT}^2 a_{nT}(\gamma-\gamma_0)} =& -\frac{\sum_{t=1}^T \phi_{20}' \widetilde{H}_t^{\star\prime}(\gamma_0,\gamma) J_n \widetilde{H}_t^{\star}(\gamma_0,\gamma)\phi_{20}}{2\hat{\sigma}_{nT}^2 a_{nT}(\gamma-\gamma_0)} \\
&- \frac{\sum_{t=1}^T (\hat{\phi}_{2,nT}-\phi_{20})' \widetilde{H}_t^{\star\prime}(\gamma_0,\gamma) J_n \widetilde{H}_t^{\star}(\gamma_0,\gamma)(\hat{\phi}_{2,nT}+\phi_{20})}{2\hat{\sigma}_{nT}^2 a_{nT}(\gamma-\gamma_0)} \\
\leqslant& -\frac{D_{1,nT}(\gamma)}{2\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} + \|\hat{\delta}_{nT}-\delta_0\|\|\hat{\delta}_{nT}+\delta_0\|\frac{F_{1,nT}}{2\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} \\
\leqslant& \frac{-(1-\eta)k + \kappa(2\|\delta_0\|+\kappa)(1+\eta)l}{2\hat{\sigma}_{nT}^2} \\
\leqslant& -\frac{k}{2\sigma_0^2+2\kappa} + \frac{3M^3\kappa}{\sigma_0^2-\kappa}.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&-\frac{\mathcal{B}_4(\gamma)+\mathcal{B}_6(\gamma)+\mathcal{B}_7(\gamma)}{2\hat{\sigma}_{nT}^2 a_{nT}(\gamma-\gamma_0)} \\
\leqslant& \|\hat{\delta}_{nT}\||\hat{l}_{nT}|\frac{\|P_{1,nT}(\gamma)\|}{\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} + \kappa\frac{|\hat{l}_{nT}|\|P_{2,nT}(\gamma)\| + \|\hat{\delta}_{nT}\|\|P_{3,nT}(\gamma)\|}{\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} \\
\leqslant& \frac{(\|\delta_0\|+\kappa)(|l_0|+\kappa)\eta + \kappa(|l_0|+\kappa)\eta + \kappa(\|\delta_0\|+\kappa)\eta}{\hat{\sigma}_{nT}^2} \leqslant \frac{8M^2\eta}{\sigma_0^2-\kappa},
\end{aligned}
$$

by events (1), (3) and (8).

Then, we have

$$
\begin{aligned}
-\frac{\mathcal{B}_5(\gamma)+\mathcal{B}_8(\gamma)}{2\hat{\sigma}_{nT}^2 a_{nT}(\gamma-\gamma_0)} \leqslant& \frac{\kappa\|\hat{\delta}_{nT}\|\|F_{2,nT}(\gamma)\|}{\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} + \frac{\kappa|\hat{l}_{nT}|\|K_{2,nT}(\gamma)\|}{\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} \\
\leqslant& \frac{\kappa[(\|\delta_0\|+\kappa)+(|l_0|+\kappa)](1+\eta)l}{\hat{\sigma}_{nT}^2} \leqslant \frac{8M^3\kappa}{\sigma_0^2-\kappa},
\end{aligned}
$$

by events (1), (3), (5) and (6).

Finally, by events (1), (3), (4) and (7), we have

$$
\begin{aligned}
&-\frac{\mathcal{B}_9(\gamma)}{2\hat{\sigma}_{nT}^2 a_{nT}(\gamma-\gamma_0)} = \\
&-\frac{[\lambda_{20}^2 + (\hat{\lambda}_{2,nT}-\lambda_{20})(\hat{\lambda}_{2,nT}+\lambda_{20})]\sum_{t=1}^T \widetilde{\mathcal{V}}_t^{\star\prime}(\gamma_0,\gamma) J_n \widetilde{\mathcal{V}}_t^{\star}(\gamma_0,\gamma)}{2\hat{\sigma}_{nT}^2 a_{nT}(\gamma-\gamma_0)} \\
\leqslant& -\frac{D_{2,nT}(\gamma)}{2\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} + \frac{|\hat{l}_{nT}-l_0||\hat{l}_{nT}+l_0||K_{1,nT}(\gamma)|}{2\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} \\
\leqslant& -\frac{k}{2\sigma_0^2+2\kappa} + \frac{\kappa(2|l_0|+\kappa)(1+\eta)l}{2\sigma_0^2-2\kappa} \\
\leqslant& -\frac{k}{2\sigma_0^2+2\kappa} + \frac{3M^3\kappa}{\sigma_0^2-\kappa}.
\end{aligned}
$$

Together, we can get

$$
\frac{\ell_{nT}^{\ddagger}(\gamma)-\ell_{nT}^{\ddagger}(\gamma_0)}{a_{nT}(\gamma-\gamma_0)} \leqslant \mathcal{M}_0 < 0.
$$

Thus, we have shown that on the set E_{nT} with probability large than $1-\epsilon$, if $\gamma \in [\gamma_0+\bar{v}/a_{nT}, \gamma_0+B]$, then $\ell^{\ddagger}_{nT}(\gamma)-\ell^{\ddagger}_{nT}(\gamma_0) < 0$. We can similarly show that if $\gamma \in [\gamma_0-B, \gamma_0-\bar{v}/a_{nT}]$ then $\ell^{\ddagger}_{nT}(\gamma)-\ell^{\ddagger}_{nT}(\gamma_0) < 0$. Since $\ell^{\ddagger}_{nT}(\hat{\gamma}_{nT})-\ell^{\ddagger}_{nT}(\gamma_0) \geqslant 0$, this implies that $|\hat{\gamma}_{nT}-\gamma_0| \leqslant \bar{v}/a_{nT}$ is with probability larger than $1-\epsilon$. That is, $a_{nT}(\hat{\gamma}_{nT}-\gamma_0) = O_p(1)$. ■

Proof of Theorem 2.4: We first show (i) $\sqrt{n(T-1)}(\hat{\theta}_{nT}-\theta_0)+\Sigma^{-1}_{nT}\sqrt{a}b_{\theta,nT} \xrightarrow{D} N\big(0, \lim\limits_{nT\to\infty} \Sigma^{-1}_{nT}\Omega_{nT}\Sigma^{-1}_{nT}\big)$, and then (ii) $\sqrt{n(T-1)}(\hat{\theta}_{nT}-\hat{\theta}_{nT}(\gamma_0)) \xrightarrow{p} 0$.

Proof of (i)**:** Similar to (B.12) and (B.14), we have

$$\sqrt{n(T-1)}(\hat{\theta}_{nT}-\theta_0) = \big[\tfrac{1}{n(T-1)}H^{*}_{nT}(\bar{\theta},\hat{\gamma}_{nT})\big]^{-1}\tfrac{1}{\sqrt{n(T-1)}}S^{*}_{\theta,nT}(\theta_0,\hat{\gamma}_{nT}),$$

where $S^{*}_{\theta,nT}(\theta_0,\hat{\gamma}_{nT}) = S^{*u}_{\theta,nT}(\theta_0,\hat{\gamma}_{nT}) + \sum_{r=1}^{3} B_{r,nT}(\theta_0,\hat{\gamma}_{nT})$. From the proof of Theorem 2.2, we see that $\frac{1}{n(T-1)}H^{*}_{nT}(\bar{\theta},\hat{\gamma}_{nT}) - \Sigma_{nT} = o_p(1)$ as $\hat{\gamma}_{nT}-\gamma_0 \xrightarrow{p} 0$ and $\bar{\theta}-\theta_0 \xrightarrow{p} 0$ implied by $\hat{\theta}_{nT}-\theta_0 = o_p(1)$. Meanwhile, Lemma A.3 and Lemma B.1 imply that $\frac{1}{\sqrt{n(T-1)}}S^{*u}_{\theta,nT}(\theta_0,\hat{\gamma}_{nT})$ converges to a mean zero Gaussian process with variance $\frac{1}{n(T-1)}\text{Var}[S^{*u}_{\theta,nT}]$ also because $\hat{\gamma}_{nT}-\gamma_0 \xrightarrow{p} 0$. The derivation of the covariance (VC) matrix of $S^{*u}_{\theta,nT}$ is straightforward following Lemma B.5 of Yang (2015), but is complicated when written into summation over time. Some intermediate results are useful to derive the final expression. We focus on the variance of one specific quadratic term, as the derivations for the other variances or covariances are similar or less difficult. Let $\mathbf{Q}_{nT} = (I_T - \frac{l_T l_T'}{T}) \otimes (I_n - \frac{l_n l_n'}{n})$, $\mathbf{G}_{nT} = \texttt{blkdiag}(G_1, G_2, \ldots, G_T)$ and $\mathbf{V} = (V_1', V_2', \ldots, V_T')'$, where $\otimes$ stands for Kronecker product and $\texttt{blkdiag}(\cdots)$ forms a block-diagonal matrix by the given submatrices. Let κ_3 be the skewness and κ_4 the excess kurtosis of the idiosyncratic errors. Hence, the variance of $\frac{1}{\sqrt{n(T-1)}}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{V}$ is just $\frac{\kappa_4\sigma_0^4}{n(T-1)}\text{E}[\texttt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT})'\texttt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT})] + \frac{\sigma_0^4}{n(T-1)}\text{E}[\texttt{tr}(\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{Q}_{nT}\mathbf{G}_{nT} + \mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{G}'_{nT})]$ by Lemma B.5 of Yang (2015). After some algebra, we have

$$\begin{aligned}
&\tfrac{1}{n(T-1)}\texttt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT})'\texttt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT}) \\
&= \tfrac{\bar{T}^2}{n(T-1)}\textstyle\sum_{t=1}^{T}\texttt{diagv}(J_nG_t)'\texttt{diagv}(J_nG_t) = \tfrac{\bar{T}^2}{n(T-1)}\sum_{t=1}^{T}R_t'R_t + O_p(\tfrac{1}{n}),
\end{aligned}$$

$$\begin{aligned}
\tfrac{1}{n(T-1)}\texttt{tr}(\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{G}'_{nT}) &= \tfrac{\bar{T}}{n(T-1)}\textstyle\sum_{t=1}^{T}\texttt{tr}(J_nG_tG_t') \\
&= \tfrac{\bar{T}}{n(T-1)}\textstyle\sum_{t=1}^{T}\texttt{tr}(G_tG_t') + O_p(\tfrac{1}{n}), \text{ and}
\end{aligned}$$

$$
\begin{aligned}
&\tfrac{1}{n(T-1)}\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{Q}_{nT}\mathbf{G}_{nT}) \\
&= \tfrac{T-2}{nT(T-1)}\sum_{t=1}^{T}\mathtt{tr}(J_nG_tJ_nG_t) + \tfrac{1}{n(T-1)}\mathtt{tr}[(\tfrac{1}{T}\sum_{t=1}^{T}J_nG_t)^2] \\
&= \tfrac{\bar{T}}{n(T-1)}\sum_{t=1}^{T}\mathtt{tr}(J_nG_tJ_nG_t)+\tfrac{1}{n(T-1)}\mathtt{tr}[(\tfrac{1}{T}\sum_{t=1}^{T}J_nG_t)^2-\tfrac{1}{T}\sum_{t=1}^{T}J_nG_tJ_nG_t] \\
&= \tfrac{\bar{T}}{n(T-1)}\sum_{t=1}^{T}\mathtt{tr}(J_nG_tJ_nG_t) + \tfrac{1}{nT^2(T-1)}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathtt{tr}[J_n(G_t-G_k)J_nG_t] \\
&= \tfrac{\bar{T}}{n(T-1)}\sum_{t=1}^{T}\mathtt{tr}(G_tG_t) + \tfrac{1}{nT^2(T-1)}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathtt{tr}[(G_t-G_k)G_t] + O_p(\tfrac{1}{n}).
\end{aligned}
$$

With these results, we have $\frac{1}{n(T-1)}\mathrm{Var}[S^{*u}_{\theta,nT}] = \Omega_{nT}(\theta_0,\gamma_0) + o(1)$, where $\Omega_{nT}(\theta_0,\gamma_0) = \Sigma_{nT}(\theta_0,\gamma_0) + \Gamma_{nT}(\theta_0,\gamma_0)$. $\Sigma_{nT}(\theta_0,\gamma_0)$ is given in (B.13) and $\Gamma_{nT}(\theta_0,\gamma_0)$ has the expression:

$$
\begin{bmatrix}
0_{2k\times 2k}, & \frac{\bar{T}\kappa_3}{\sigma_0}\mathrm{E}\mathcal{R}_{1,nT}(\gamma_0), & 0_{2k\times 1} \\
\sim, & \frac{2\bar{T}\kappa_3}{\sigma_0}\mathrm{E}\mathcal{R}_{2,nT}(\gamma_0) + \kappa_4\bar{T}^2\mathrm{E}\mathcal{R}_{3,nT}(\gamma_0) + \mathrm{E}\mathbb{B}_{nT}(\gamma_0), & \frac{\kappa_4\bar{T}^2}{2\sigma_0^2}\mathrm{E}R_{nT}(\gamma_0) \\
\sim, & \sim, & \frac{\kappa_4\bar{T}}{4\sigma_0^4}
\end{bmatrix}, \tag{B.23}
$$

where $\mathcal{R}_{1,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T}\widetilde{\mathbb{X}}_t'(\gamma)J_n\widetilde{\mathbb{R}}_t(\gamma)$, $\mathcal{R}_{2,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T}\widetilde{\mathbb{Z}}_t'(\gamma)J_n\widetilde{\mathbb{R}}_t(\gamma)$,

$\mathcal{R}_{3,nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T}\mathbb{R}_t'(\gamma)\mathbb{R}_t(\gamma)$, $\quad R_{nT}(\gamma) = \frac{1}{n(T-1)}\sum_{t=1}^{T}\mathbb{R}_t'(\gamma)l_n$,

$\mathbb{B}_{nT}(\gamma) = [\mathbb{B}_{11,nT},\ \mathbb{B}_{12,nT}(\gamma);\ \mathbb{B}_{12,nT}(\gamma),\ \mathbb{B}_{22,nT}(\gamma)]$,

$\mathbb{B}_{11,nT} = \frac{1}{nT^2(T-1)}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathtt{tr}[(G_t-G_k)G_t]$,

$\mathbb{B}_{12,nT}(\gamma) = \frac{1}{nT^2(T-1)}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathtt{tr}[(d_t(\gamma)G_t - d_k(\gamma)G_k)G_t]$,

$\mathbb{B}_{22,nT}(\gamma) = \frac{1}{nT^2(T-1)}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathtt{tr}[(d_t(\gamma)G_t - d_k(\gamma)G_k)d_t(\gamma)G_t]$.

Next, we see that $\frac{1}{\sqrt{n(T-1)}}B_{3,nT}(\theta_0,\hat{\gamma}_{nT}) - \sqrt{a}b_{\theta,nT} \xrightarrow{p} 0$ as $\hat{\gamma}_{nT} - \gamma_0 \xrightarrow{p} 0$. Hence, it is left to show $\frac{1}{\sqrt{n(T-1)}}B_{r,nT}(\theta_0,\hat{\gamma}_{nT}) = o_p(1)$ for $r = 1,2$. For $B_{1,nT}(\theta_0,\hat{\gamma}_{nT})$, we only study its third and fourth components for example, as the other two components can be studied in a similar manner. By $\phi_{20} = (nT)^{-\tau}\delta_0$ and $\lambda_{20} = (nT)^{-\tau}l_0$, the third component of $\frac{1}{\sqrt{n(T-1)}}B_{1,nT}(\theta_0,\hat{\gamma}_{nT})$ equals to

$$(nT)^{\tau-1/2}\bar{T}^{-1/2}\tfrac{a_{nT}}{\sigma_0^2 nT}\sum_{t=1}^{T}[\widetilde{Z}_t^{\circ}(\hat{\gamma}_{nT})+\widetilde{\mathcal{V}}_t^{\circ}(\hat{\gamma}_{nT})]'J_n[\widetilde{H}_t^{\star}(\gamma_0,\hat{\gamma}_{nT})\delta_0+\widetilde{\mathcal{V}}_t^{\star}(\gamma_0,\hat{\gamma}_{nT})l_0].$$

As $\hat{\gamma}_{nT} = \gamma_0 + \hat{v}_{nT}/a_{nT}$ by Theorem 2.3, these terms in $\frac{a_{nT}}{nT}\sum_{t=1}^{T}[\widetilde{Z}_t^{\circ}(\hat{\gamma}_{nT}) + \widetilde{\mathcal{V}}_t^{\circ}(\hat{\gamma}_{nT})]'J_n[\widetilde{H}_t^{\star}(\gamma_0,\hat{\gamma}_{nT})\delta_0 + \widetilde{\mathcal{V}}_t^{\star}(\gamma_0,\hat{\gamma}_{nT})l_0]$ have similar form to $\mathcal{F}_{nT}(\hat{v}_{nT})$ or $\mathcal{K}_{nT}(\hat{v}_{nT})$ from Lemma B.2 so that we can show they are all $O_p(1)$, following the proof of Lemma B.2. As $(nT)^{\tau-1/2} = o(1)$ by Assumption F, the third component of $\frac{1}{\sqrt{n(T-1)}}B_{1,nT}(\theta_0,\hat{\gamma}_{nT})$ is $o_p(1)$. Similarly, the fourth component of $\frac{1}{\sqrt{n(T-1)}}B_{1,nT}(\theta_0,\hat{\gamma}_{nT})$ equals to

$$\frac{\bar{T}^{-1/2}}{2\sigma_0^4(nT)^{1/2}}\big[\frac{a_{nT}}{nT}\sum_{t=1}^T \delta_0'\widetilde{H}_t^{\star\prime}(\gamma_0,\hat{\gamma}_{nT})J_n\widetilde{H}_t^{\star}(\gamma_0,\hat{\gamma}_{nT})\delta_0+\frac{a_{nT}}{nT}l_0^2\sum_{t=1}^T \widetilde{\mathcal{V}}_t^{\star\prime}(\gamma_0,\hat{\gamma}_{nT})J_n \widetilde{\mathcal{V}}_t^{\star}(\gamma_0,\hat{\gamma}_{nT})+\frac{2\sqrt{a_{nT}}}{\sqrt{nT}}\sum_{t=1}^T \widetilde{V}_t'J_n\widetilde{H}_t^{\star}(\gamma_0,\hat{\gamma}_{nT})\delta_0+\frac{2a_{nT}}{nT}\sum_{t=1}^T l_0\widetilde{\mathcal{V}}_t^{\star\prime}(\gamma_0,\hat{\gamma}_{nT})J_n \widetilde{H}_t^{\star}(\gamma_0,\hat{\gamma}_{nT})\delta_0+(nT)^{\tau}\frac{2a_{nT}}{nT}\sum_{t=1}^T \widetilde{V}_t'J_n\widetilde{\mathcal{V}}_t^{\star}(\gamma_0,\hat{\gamma}_{nT})l_0\big].$$

The first two terms in the square bracket are $O_p(1)$ by Lemma B.2, the third is $O_p(1)$ by Lemma B.3, and the fourth and fifth without $(nT)^{\tau}$ can be easily shown to be $O_p(1)$, following the proof of Lemma B.2. Therefore, the fourth component of $\frac{1}{\sqrt{n(T-1)}}B_{1,nT}(\theta_0,\hat{\gamma}_{nT})$ is also $o_p(1)$. The other components of $\frac{1}{\sqrt{n(T-1)}}B_{1,nT}(\theta_0,\hat{\gamma}_{nT})$ can be shown to be $o_p(1)$ similarly.

Finally, we show all the components of $\frac{1}{\sqrt{n(T-1)}}B_{2,nT}(\theta_0,\hat{\gamma}_{nT})$ are also $o_p(1)$. Consider its second non-zero element for example, as the other can be shown similarly. Similar to (B.15),

$$\begin{aligned} &-\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T \mathtt{tr}[J_n d_t(\hat{\gamma}_{nT})(G_t(\lambda_0,\hat{\gamma}_{nT})-G_t)] \\ =&\ \frac{(nT)^{-\tau}l_0}{\sqrt{n(T-1)}}\sum_{t=1}^T \mathtt{tr}[J_n d_t(\hat{\gamma}_{nT})G_t(\lambda_0,\hat{\gamma}_{nT})d_t(\gamma_0,\hat{\gamma}_{nT})G_t)] \\ =&\ (nT)^{\tau-1/2}\bar{T}^{-1/2}\frac{l_0 a_{nT}}{nT}\sum_{t=1}^T \mathtt{tr}[J_n d_t(\hat{\gamma}_{nT})G_t(\lambda_0,\hat{\gamma}_{nT})d_t(\gamma_0,\hat{\gamma}_{nT})G_t]=o_p(1), \end{aligned}$$

because $(nT)^{\tau-1/2}=o(1)$, and $\frac{l_0 a_{nT}}{nT}\sum_{t=1}^T \mathtt{tr}[J_n d_t(\hat{\gamma}_{nT})G_t(\lambda_0,\hat{\gamma}_{nT})d_t(\gamma_0,\hat{\gamma}_{nT})G_t]$ has similar form to $\mathcal{L}_{nT}(\hat{v}_{nT})$ from Lemma B.2 and thus can be shown to be $O_p(1)$ in a similar manner. By the continuous mapping theorem (CMT), the result in (i) follows.

Proof of (ii)**:** When γ_0 were known, it is easy to see that the QMLE $\hat{\theta}_{nT}(\gamma_0)$ is consistent to θ_0. Thus, by the mean value theorem, we also have

$$\sqrt{n(T-1)}(\hat{\theta}_{nT}(\gamma_0)-\theta_0)=\big[\frac{1}{n(T-1)}H_{nT}^*(\dot{\theta},\gamma_0)\big]^{-1}\frac{1}{\sqrt{n(T-1)}}S_{\theta,nT}^*(\theta_0,\gamma_0),$$

where $H_{nT}^*(\dot{\theta},\gamma)$ denotes $-\frac{\partial}{\partial\theta'}S_{\theta,nT}^*(\theta,\gamma)\big|_{\theta=\dot{\theta}_r \text{ in } r\text{th row}}$ and $\{\dot{\theta}_r\}$ are on the line segment between $\hat{\theta}_{nT}(\gamma_0)$ and θ_0. As $\dot{\theta}-\theta_0\xrightarrow{p}0$ implied by $\hat{\theta}_{nT}(\gamma_0)-\theta_0=o_p(1)$, $\frac{1}{n(T-1)}H_{nT}^*(\dot{\theta},\gamma_0)-\Sigma_{nT}=o_p(1)$. Thus, it is equivalent to showing that $\frac{1}{\sqrt{n(T-1)}}[S_{\theta,nT}^*(\theta_0,\hat{\gamma}_{nT})-S_{\theta,nT}^*]=o_p(1)$. Noting that $S_{\theta,nT}^*=S_{\theta,nT}^{*u}+B_{3,nT}$, we have

$$\frac{1}{\sqrt{n(T-1)}}[S_{\theta,nT}^*(\theta_0,\hat{\gamma}_{nT})-S_{\theta,nT}^*]=\frac{1}{\sqrt{n(T-1)}}[S_{\theta,nT}^{*u}(\theta_0,\hat{\gamma}_{nT})-S_{\theta,nT}^{*u}]+o_p(1),$$

because $\frac{1}{\sqrt{n(T-1)}} B_{r,nT}(\theta_0, \hat{\gamma}_{nT}) = o_p(1)$, for $r = 1, 2$, shown in (i), and meanwhile $\frac{1}{\sqrt{n(T-1)}}[B_{3,nT}(\theta_0, \hat{\gamma}_{nT}) - B_{3,nT}] = o_p(1)$ is directly implied by $\hat{\gamma}_{nT} - \gamma_0 = o_p(1)$. For the non-zero components of $\frac{1}{\sqrt{n(T-1)}}[S^{*u}_{\theta,nT}(\theta_0, \hat{\gamma}_{nT}) - S^{*u}_{\theta,nT}]$, they are $O_p(\frac{1}{\sqrt{a_{nT}}})$ by Lemma B.3, completing the proof. ■

Proof of Theorem 2.5: Let $\mathcal{Q}_{nT}(v) = \ell^{\ddagger}_{nT}(\gamma_0 + v/a_{nT}) - \ell^{\ddagger}_{nT}(\gamma_0)$ and $\mathcal{Q}(v) = \frac{1}{2\sigma_0^2}[-\Xi_1 f|v| + 2\sqrt{\sigma_0^2 \Xi f} W(v)]$. We first show $\mathcal{Q}_{nT}(v) \Rightarrow \mathcal{Q}(v)$ on any compact set $\Upsilon = [-\bar{v}, \bar{v}]$.

For ease of presentation, we follow the notations used in the proof of Theorem 2.4 and define $\mathcal{B}^*_s(v) = \mathcal{B}_s(\gamma_0 + v/a_{nT})$, for $s = 1$ to 9, and $\mathcal{A}^*_m(v) = \mathcal{A}_m(\gamma_0 + v/a_{nT})$, for $m = 1, 2, 3$. As discussed in the proof of Theorem 2.4, the proof of Lemma B.2 implies that $F_{2,nT}(\gamma_0 + v/a_{nT})$, $K_{2,nT}(\gamma_0 + v/a_{nT})$, $L_{2,nT}(\gamma_0 + v/a_{nT})$ and $L_{3,nT}(\gamma_0 + v/a_{nT})$ are both $O_p(\frac{1}{a_{nT}})$, and $P_{r,nT}(\gamma_0 + v/a_{nT})$ for $r = 1, 2, 3$ are all $o_p(\frac{1}{a_{nT}})$. Given these, we can easily see that $\sum_{s=4}^{8} \mathcal{B}^*_s(v) = o_p(1)$, since $(nT)^{\tau}(\hat{\phi}_{nT} - \phi_0) = o_p(1)$ and $(nT)^{\tau}\hat{\phi}_{2,nT} = O_p(1)$. Similarly, we have $\mathcal{B}^*_3(v) = \mathcal{F}_{nT}(v) + o_p(1)$, $\mathcal{B}^*_9(v) = \mathcal{K}_{nT}(v) + o_p(1)$, $\mathcal{A}^*_2(v) = -\frac{1}{2}\bar{T}\mathcal{L}_{nT}(v) + o_p(1)$, $\mathcal{A}^*_3(v) = o_p(1)$, and finally

$$
\begin{aligned}
&\mathcal{B}^*_1(v) + \mathcal{B}^*_2(v) - 2\hat{\sigma}^2_{nT}\mathcal{A}^*_1(v) \\
= &-2\textstyle\sum_{t=1}^{T} \widetilde{V}_t' J_n \widetilde{H}^{\circ}_t(\gamma_0 + v/a_{nT}, \gamma_0)\hat{\phi}_{2,nT} - 2\hat{\lambda}_{2,nT}\sum_{t=1}^{T} \widetilde{V}_t' J_n \widetilde{\mathcal{V}}^{\circ}_t(\gamma_0 + v/a_{nT}, \gamma_0) \\
&- 2\hat{\sigma}^2_{nT}\hat{\lambda}_{2,nT}\bar{T}\textstyle\sum_{t=1}^{T} \mathtt{tr}(d_t(\gamma_0, \gamma_0 + v/a_{nT})G_t) \\
= &-2\mathcal{R}_{nT}(v) + 2\hat{l}_{nT}(nT)^{\tau}(\sigma_0^2 - \hat{\sigma}^2_{nT})a_{nT}\bar{T}L_{2,nT}(\gamma_0 + v/a_{nT}) \\
&- 2\hat{l}_{nT}\tfrac{T^{\tau}}{n^{1-\tau}}\hat{\sigma}^2_{nT}a_{nT}\bar{T}L_{3,nT}(\gamma_0 + v/a_{nT}) + o_p(1) = -2\mathcal{R}_{nT}(v) + o_p(1),
\end{aligned}
$$

where we use $\frac{T^{\tau}}{n^{1-\tau}} = o(1)$ by Assumption E.

Then, from (B.19), (B.20) and (B.21), we have

$$
\begin{aligned}
\mathcal{Q}_{nT}(v) = &-\tfrac{1}{2\hat{\sigma}^2_{nT}}[\textstyle\sum_{s=1}^{9} \mathcal{B}^*_s(v) - 2\hat{\sigma}^2_{nT}\mathcal{A}^*_1(v)] + \mathcal{A}^*_2(v) + \mathcal{A}^*_3(v) \\
= &-\tfrac{1}{2\hat{\sigma}^2_{nT}}[\mathcal{F}_{nT}(v) + \mathcal{K}_{nT}(v) - 2\mathcal{R}_{nT}(v)] - \tfrac{\bar{T}}{2}\mathcal{L}_{nT}(v) + o_p(1).
\end{aligned}
$$

Using Lemma B.2, Lemma B.3 and $\hat{\sigma}^2_{nT} - \sigma_0^2 = o_p(1)$, we finally get $\mathcal{Q}_{nT}(v) \Rightarrow \mathcal{Q}(v)$.

By Theorem 2.3, $a_{nT}(\hat{\gamma}_{nT} - \gamma_0) = \underset{v}{\operatorname{argmax}}\, \mathcal{Q}_{nT}(v) = O_p(1)$. The functional $\mathcal{Q}(v)$ is continuous and has a unique maximum; $\lim_{|v|\to\infty} \mathcal{Q}(v) = -\infty$ almost surely since $\lim_{v\to\infty} B(v)/v = 0$ almost surely. Therefore, the conditions of Theorem 2.7 of Kim and Pollard (1990) are satisfied, which implies

$$a_{nT}(\hat{\gamma}_{nT} - \gamma_0) \xrightarrow{D} \underset{-\infty<v<\infty}{\operatorname{argmax}} \mathcal{Q}(v).$$

We make the change-of-variable $v = \frac{\sigma_0^2}{f}\frac{\Xi}{\Xi_1^2}r$ and then rewrite the asymptotic distribution as

$$\begin{aligned}
\underset{-\infty<v<\infty}{\operatorname{argmax}} \mathcal{Q}(v) &= \underset{-\infty<v<\infty}{\operatorname{argmax}} [-\Xi_1 f|v| + 2\sqrt{\sigma_0^2 \Xi f} W(v)] \\
&= \tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2} \underset{-\infty<r<\infty}{\operatorname{argmax}} [-\tfrac{\sigma_0^2\Xi}{\Xi_1}|r| + 2\sqrt{\sigma_0^2\Xi f} W(\tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2}r)] \\
&= \tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2} \underset{-\infty<r<\infty}{\operatorname{argmax}} [-\tfrac{\sigma_0^2\Xi}{\Xi_1}|r| + 2\tfrac{\sigma_0^2\Xi}{\Xi_1} W(r)] \\
&= \tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2} \underset{-\infty<r<\infty}{\operatorname{argmax}} [-\tfrac{|r|}{2} + W(r)].
\end{aligned}$$

■

Proof of Theorem 2.6: By Theorem 2.3, we can write $\hat{\gamma}_{nT} = \gamma_0 + \frac{\hat{v}_{nT}}{a_{nT}}$. Note that

$$\begin{aligned}
LR_{nT}(\gamma_0) &= 2[\ell_{nT}^{*c}(\hat{\gamma}_{nT}) - \ell_{nT}^{*c}(\gamma_0)] \\
&= 2[\ell_{nT}^{*c}(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) - \ell_{nT}^{*c}(\hat{\theta}_{nT}(\gamma_0), \gamma_0)] \\
&= 2[\ell_{nT}^{*c}(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) - \ell_{nT}^{*c}(\hat{\theta}_{nT}, \gamma_0)] + o_p(1) \quad \text{(Theorem 2.4)} \\
&= 2\mathcal{Q}_{nT}(\hat{v}_{nT}) + o_p(1) \xrightarrow{D} 2\sup_v \mathcal{Q}(v).
\end{aligned}$$

This limiting distribution equals, by the change-of-variable $v = \frac{\sigma_0^2}{f}\frac{\Xi}{\Xi_1^2}r$,

$$\begin{aligned}
\tfrac{1}{\sigma_0^2}\sup_v[-\Xi_1 f|v| + 2\sqrt{\sigma_0^2\Xi f}W(v)] &= \tfrac{1}{\sigma_0^2}\sup_r[-\Xi_1 f|\tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2}r| + 2\sqrt{\sigma_0^2\Xi f}W(\tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2}r)] \\
&= \tfrac{\Xi}{\Xi_1}\sup_r[-|r| + 2W(r)] = \varpi^2\,\mho.
\end{aligned}$$

To find the distribution of $\mho$, note that $\mho = 2\max(\mho_1, \mho_2)$, where $\mho_1 = \sup_{r\leqslant 0}[-|r|/2 + W(r)]$ and $\mho_1 = \sup_{r\geqslant 0}[-|r|/2 + W(r)]$. $\mho_1$ and $\mho_2$ are iid exponential random variables with distribution function $P(\mho_1 \leqslant x) = 1 - e^{-x}$. It follows that $P(\mho \leqslant x) = P(2\max(\mho_1, \mho_2) \leqslant x) = P(\mho_1 \leqslant x/2)P(\mho_2 \leqslant x/2) = (1 - e^{-x})^2$. ■

Proof of Theorem 2.7: As discussed in Section 2.3, we only show result (i) of Theorem 2.7 as (ii) follows (i) directly. Under Assumptions B(vi) and the alternatives, the consistency of the proposed AQS estimator $\hat{\theta}_{nT}^{\diamond}(\gamma)$ lies with that of $\hat{\lambda}^{\diamond}_{nT}(\gamma)$, uniformly in γ. In order to show the consistency of $\hat{\lambda}^{\diamond}_{nT}(\gamma)$, the key step in the proof is to compare $S^{\diamond c}_{\theta,nT}(\omega)$ with its population counterpart. Substituting both $\hat{\beta}^{\diamond}_{nT}(\omega) = \hat{\beta}_{nT}(\omega)$ and $\hat{\sigma}^{2\diamond}_{nT}(\omega) = \frac{n}{n-1}\hat{\sigma}^{2}_{nT}(\omega)$ into the λ-component of AQS functions gives the concentrated AQS functions:

$$S^{\diamond c}_{\theta,nT}(\omega) = \begin{cases} \frac{1}{\hat{\sigma}^{2\diamond}_{nT}(\omega)} \sum_{t=1}^{T} [\widehat{K}^{\diamond\prime}_t(\omega) + \widehat{V}^{\diamond\prime}_t(\omega) J_n G_t(\omega)] J_n \widehat{V}^{\diamond}_t(\omega) - \bar{T}\sum_{t=1}^{T} \mathtt{tr}[J_n G_t(\omega)], \\ \frac{1}{\hat{\sigma}^{2\diamond}_{nT}(\omega)} \sum_{t=1}^{T} [\widehat{K}^{\circ\diamond\prime}_t(\omega) + \widehat{V}^{\diamond\prime}_t(\omega) J_n d_t(\gamma) G_t(\omega)] J_n \widehat{V}^{\diamond}_t(\omega) \\ \qquad\qquad -\bar{T}\sum_{t=1}^{T} \mathtt{tr}[J_n d_t(\gamma) G_t(\omega)]. \end{cases}$$

where $\widehat{K}^{\diamond}_t(\omega) = \widetilde{K}_t(\hat{\beta}^{\diamond}_{nT}(\omega),\omega)$, $\widehat{K}^{\circ\diamond}_t(\omega) = \widetilde{K}^{\circ}_t(\hat{\beta}^{\diamond}_{nT}(\omega),\omega)$ and $\widehat{V}^{\diamond}_t(\omega) = \widetilde{V}_t(\hat{\beta}^{\diamond}_{nT}(\omega),\omega)$. For each $\gamma \in \Gamma$, solving the resulted concentrated estimating equations, $S^{\diamond c}_{\theta,nT}(\omega) = 0$, we obtain the M-estimator $\hat{\lambda}^{\diamond}_{nT}(\gamma)$ of λ_0. Thus the M-estimators of β and σ^2 are $\hat{\beta}^{\diamond}_{nT}(\gamma) \equiv \hat{\beta}^{\diamond}_{nT}(\hat{\lambda}^{\diamond}_{nT}(\gamma),\gamma)$ and $\hat{\sigma}^{2\diamond}_{nT}(\gamma) \equiv \hat{\sigma}^{2\diamond}_{nT}(\hat{\lambda}^{\diamond}_{nT}(\gamma),\gamma)$.

Substituting both $\bar{\beta}^{\diamond}_{nT}(\omega) = \bar{\beta}_{nT}(\omega)$ and $\bar{\sigma}^{2\diamond}_{nT}(\omega) = \frac{n}{n-1}\bar{\sigma}^{2}_{nT}(\omega)$ back into the λ-component of $\bar{S}^{\diamond}_{\theta,nT}(\theta,\gamma)$, we get the population counterpart of $S^{\diamond c}_{\theta,nT}(\omega)$ as

$$\bar{S}^{\diamond c}_{\theta,nT}(\omega) = \begin{cases} \frac{1}{\bar{\sigma}^{2\diamond}_{nT}(\omega)} \sum_{t=1}^{T} \mathrm{E}\{[\bar{K}^{\diamond\prime}_t(\omega) + \bar{V}^{\diamond\prime}_t(\omega) J_n G_t(\omega)] J_n \bar{V}^{\diamond}_t(\omega)\} \\ \qquad\qquad -\bar{T}\sum_{t=1}^{T} \mathrm{E}\{\mathtt{tr}[J_n G_t(\omega)]\}, \\ \frac{1}{\bar{\sigma}^{2\diamond}_{nT}(\omega)} \sum_{t=1}^{T} \mathrm{E}\{[\bar{K}^{\circ\diamond\prime}_t(\omega) + \bar{V}^{\diamond\prime}_t(\omega) J_n d_t(\gamma) G_t(\omega)] J_n \bar{V}^{\diamond}_t(\omega)\} \\ \qquad\qquad -\bar{T}\sum_{t=1}^{T} \mathrm{E}\{\mathtt{tr}[J_n d_t(\gamma) G_t(\omega)]\}. \end{cases}$$

where $\bar{K}^{\diamond}_t(\omega) = \widetilde{K}_t(\bar{\beta}^{\diamond}_{nT}(\omega),\omega)$, $\bar{K}^{\circ\diamond}_t(\omega) = \widetilde{K}^{\circ}_t(\bar{\beta}^{\diamond}_{nT}(\omega),\omega)$ and $\bar{V}^{\diamond}_t(\omega) = \widetilde{V}_t(\bar{\beta}^{\diamond}_{nT}(\omega),\omega)$.

By (B.1) and (B.11), $\bar{\beta}^{\diamond}_{nT}(\omega) = \bar{\beta}_{nT}(\omega) = \beta_0 + \mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\mathrm{E}\mathcal{P}_{2,nT}(\gamma)(\lambda_0 - \lambda) + o(1)$,

$$\begin{aligned} \bar{\sigma}^{2\diamond}_{nT}(\omega) &= \sigma_0^2 + 2\sigma_0^2 \mathrm{E}\mathcal{G}^{[2]\prime}_{1,nT}(\gamma)(\lambda_0 - \lambda) + (\lambda_0 - \lambda)'\big[\sigma_0^2 \mathrm{E}\mathbb{G}^{[2]}_{1,nT}(\gamma) \\ &\qquad + \mathrm{E}\mathcal{P}_{3,nT}(\gamma) - \mathrm{E}\mathcal{P}'_{2,nT}(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\mathrm{E}\mathcal{P}_{2,nT}(\gamma)\big](\lambda_0 - \lambda) + o(1), \quad \text{and} \\ J_n\bar{V}^{\diamond}_t(\omega) &= J_n[\widetilde{\mathbb{Y}}_t(\omega) - \widetilde{\mathbb{X}}_t(\gamma)\bar{\beta}^{\diamond}_{nT}(\omega)] \\ &= J_n[\widetilde{\mathbb{Z}}_t(\gamma)(\lambda_0 - \lambda) + \widetilde{\mathbb{V}}_t(\gamma)(\lambda_0 - \lambda) + \widetilde{V}_t \\ &\qquad - \widetilde{\mathbb{X}}_t(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\mathrm{E}\mathcal{P}_{2,nT}(\gamma)(\lambda_0 - \lambda)] + o_p(1). \end{aligned}$$

Hence, the first component of $\bar{S}^{\diamond c}_{\theta}(\omega)$ equals to

$$\bar{S}^{\diamond c}_{\theta,1}(\omega) = \lim_{nT\to\infty} \tfrac{1}{\bar{\sigma}^{2\diamond}_{nT}(\omega)}\Big\{\big[\mathrm{E}\mathcal{P}_{4,nT}(\omega) + \mathrm{E}\mathcal{P}_{9,nT}(\omega) - \mathrm{E}\mathcal{P}_{5,nT}(\omega)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}$$
$$\mathrm{E}\mathcal{P}_{2,nT}(\gamma)\big](\lambda_0-\lambda) + (\lambda_0-\lambda)'\big[\mathrm{E}\mathcal{P}_{6,nT}(\omega) + \mathrm{E}\mathcal{P}_{10,nT}(\omega)$$
$$+ \mathrm{E}\mathcal{P}'_{2,nT}(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\big(\mathrm{E}\mathcal{P}_{4,nT}(\omega) - \mathrm{E}\mathcal{P}_{8,nT}(\omega)\big)$$
$$+ \mathrm{E}\mathcal{P}'_{2,nT}(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}\big(\mathrm{E}\mathcal{P}_{7,nT}(\omega) - \mathrm{E}\mathcal{P}_{5,nT}(\omega)\big)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}$$
$$\mathrm{E}\mathcal{P}_{2,nT}(\gamma)\big](\lambda_0-\lambda) + \bar{T}(\sigma_0^2 - \bar{\sigma}^{2\diamond}_{nT}(\omega))\textstyle\sum_{t=1}^T \mathrm{E}[\mathtt{tr}(J_nG_t(\omega))]\Big\},$$

where $\mathcal{P}_{4,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{K}'_t(\beta_0,\omega)J_n\widetilde{\mathbb{Z}}_t(\gamma), \mathcal{P}_{5,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{K}'_t(\beta_0,\omega)J_n\widetilde{\mathbb{X}}_t(\gamma)$,
$\mathcal{P}_{6,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{\mathbb{Z}}'_t(\gamma)J_nG_t(\omega)J_n\widetilde{\mathbb{Z}}_t(\gamma),\ \mathcal{P}_{7,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{\mathbb{X}}'_t(\gamma)J_nG_t(\omega)J_n\widetilde{\mathbb{X}}_t(\gamma)$,
$\mathcal{P}_{8,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{\mathbb{X}}'_t(\gamma)J_nG^s_t(\omega)J_n\widetilde{\mathbb{Z}}_t(\gamma), \mathcal{P}_{9,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{V}'_tJ_nG^s_t(\omega)J_n\widetilde{\mathbb{V}}_t(\gamma)$,
$\mathcal{P}_{10,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{\mathbb{V}}'_t(\gamma)J_nG_t(\omega)J_n\widetilde{\mathbb{V}}_t(\gamma)$,
and the second component equals to

$$\bar{S}^{\diamond c}_{\theta,2}(\omega) = \lim_{nT\to\infty} \tfrac{1}{\bar{\sigma}^{2\diamond}_{nT}(\omega)}\Big\{[\mathrm{E}\mathcal{P}^{\circ}_{4,nT}(\omega) + \mathrm{E}\mathcal{P}^{\circ}_{9,nT}(\omega) - \mathrm{E}\mathcal{P}^{\circ}_{5,nT}(\omega)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}$$
$$\mathrm{E}\mathcal{P}_{2,nT}(\gamma)](\lambda_0-\lambda) + (\lambda_0-\lambda)'[\mathrm{E}\mathcal{P}^{\circ}_{6,nT}(\omega) + \mathrm{E}\mathcal{P}^{\circ}_{10,nT}(\omega)$$
$$+ \mathrm{E}\mathcal{P}'_{2,nT}(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}(\mathrm{E}\mathcal{P}^{\circ}_{4,nT}(\omega) - \mathrm{E}\mathcal{P}^{\circ}_{8,nT}(\omega))$$
$$+ \mathrm{E}\mathcal{P}'_{2,nT}(\gamma)\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}(\mathrm{E}\mathcal{P}^{\circ}_{7,nT}(\omega) - \mathrm{E}\mathcal{P}^{\circ}_{5,nT}(\omega))\mathrm{E}(\mathcal{P}_{1,nT}(\gamma))^{-1}$$
$$\mathrm{E}\mathcal{P}_{2,nT}(\gamma)](\lambda_0-\lambda) + \bar{T}(\sigma_0^2 - \bar{\sigma}^{2\diamond}_{nT}(\omega))\textstyle\sum_{t=1}^T \mathrm{E}[\mathtt{tr}(J_nG^{\circ}_t(\omega))]\Big\},$$

where $\mathcal{P}^{\circ}_{4,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{K}^{\circ\prime}_t(\beta_0,\omega)J_n\widetilde{\mathbb{Z}}_t(\gamma), \mathcal{P}^{\circ}_{5,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{K}^{\circ\prime}_t(\beta_0,\omega)J_n\widetilde{\mathbb{X}}_t(\gamma)$,
$\mathcal{P}^{\circ}_{6,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{\mathbb{Z}}'_t(\gamma)J_nG^{\circ}_t(\omega)J_n\widetilde{\mathbb{Z}}_t(\gamma),\quad \mathcal{P}^{\circ}_{7,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{\mathbb{X}}'_t(\gamma)J_nG^{\circ}_t(\omega)J_n\widetilde{\mathbb{X}}_t(\gamma)$,
$\mathcal{P}^{\circ}_{8,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{\mathbb{X}}'_t(\gamma)J_nG^{\circ s}_t(\omega)J_n\widetilde{\mathbb{Z}}_t(\gamma),\ \mathcal{P}^{\circ}_{9,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{V}'_tJ_nG^{\circ s}_t(\omega)J_n\widetilde{\mathbb{V}}_t(\gamma)$,
$\mathcal{P}^{\circ}_{10,nT}(\omega) = \frac{1}{nT}\sum_{t=1}^T \widetilde{\mathbb{V}}'_t(\gamma)J_nG^{\circ}_t(\omega)J_n\widetilde{\mathbb{V}}_t(\gamma),\ G^{\circ}_t(\omega) = d_t(\gamma)G_t(\omega)$.

Clearly, $S^{\diamond c}_{\theta,nT}(\hat{\lambda}^{\diamond}_{nT}(\gamma),\gamma) = 0$ for any γ by construction. It is also easy to see that $\bar{S}^{\diamond c}_{\theta}(\lambda_0,\gamma) = 0$ for any γ, implied by $\bar{S}^{\diamond c}_{\theta,1}(\lambda_0,\gamma) = 0$ and $\bar{S}^{\diamond c}_{\theta,2}(\lambda_0,\gamma) = 0$. By Theorem 5.9 of van der Vaar (1998), $\hat{\lambda}^{\diamond}_{nT}(\gamma)$ will be consistent for λ_0 if $\sup_{\lambda\in\Lambda}\frac{1}{nT}|S^{\diamond c}_{\theta,nT}(\omega) - \bar{S}^{\diamond c}_{\theta,nT}(\omega)| = o_p(1)$ and Assumption G$'$. The low level condition for Assumption G$'$ is either $\bar{S}^{\diamond c}_{\theta,1}(\omega) \neq 0$ for $\lambda \neq \lambda_0$ or $\bar{S}^{\diamond c}_{\theta,2}(\omega) \neq 0$ for $\lambda \neq \lambda_0$, uniformly in $\gamma \in \Gamma$. Thus, it is left to show that $\sup_{\lambda\in\Lambda}\frac{1}{nT}|S^{\diamond c}_{\theta,nT}(\omega) - \bar{S}^{\diamond c}_{\theta,nT}(\omega)| =$

$o_p(1)$. By (B.1), (B.4) and (B.5), $\hat{\beta}_{nT}(\omega) = \beta_0 + \mathcal{P}^{-1}_{1,nT}(\gamma)\mathcal{P}_{2,nT}(\gamma)(\lambda_0 - \lambda) + o_p(1)$,

$$\begin{aligned}
\hat{\sigma}^{2\diamond}_{nT}(\omega) &= \sigma_0^2 + 2\mathcal{V}_{21,nT}(\gamma)(\lambda_0-\lambda) + (\lambda_0-\lambda)'\big[\mathcal{V}_{31,nT}(\gamma) \\
&\qquad + \mathcal{P}_{3,nT}(\gamma) - \mathcal{P}'_{2,nT}(\gamma)\mathcal{P}^{-1}_{1,nT}(\gamma)\mathcal{P}_{2,nT}(\gamma)\big](\lambda_0-\lambda) + o_p(1), \quad \text{and} \\
J_n\widehat{V}^{\diamond}_t(\omega) &= J_n[\widetilde{\mathbb{Y}}_t(\omega) - \widetilde{\mathbb{X}}_t(\gamma)\hat{\beta}^{\diamond}_{nT}(\omega)] \\
&= J_n[\widetilde{\mathbb{Z}}_t(\gamma)(\lambda_0-\lambda) + \widetilde{\mathbb{V}}_t(\gamma)(\lambda_0-\lambda) + \widetilde{V}_t \\
&\qquad - \widetilde{\mathbb{X}}_t(\gamma)\mathcal{P}^{-1}_{1,nT}(\gamma)\mathcal{P}_{2,nT}(\gamma)(\lambda_0-\lambda)] + o_p(1).
\end{aligned}$$

With these, we can show $\sup_{\lambda\in\Lambda}\frac{1}{nT}|S^{\diamond c}_{\theta,nT}(\omega) - \bar{S}^{\diamond c}_{\theta,nT}(\omega)| = o_p(1)$ as we do for Theorem 2.1. ■

Proof of Theorem 2.8: Note that $K_t(\phi,\gamma) = G_t(\lambda,\gamma)\mathbb{X}_t(\gamma)\beta$ and $K^{\circ}_t(\phi,\gamma) = d_t(\gamma)K_t(\phi,\gamma)$. Let $K_t(\gamma) = K_t(\phi_0,\gamma)$, $K^{\circ}_t(\gamma) = d_t(\gamma)K_t(\phi_0,\gamma)$, $\mathbb{K}_t(\gamma) = [K_t(\gamma), K^{\circ}_t(\gamma)]$ and $G^{\circ}_t(\gamma) = d_t(\gamma)G_t$.

Applying the MVT to each element of $S^{\diamond}_{\theta,nT}(\hat{\theta}_{nT}(\gamma),\gamma)$, one has

$$0 = S^{\diamond}_{\theta,nT}(\hat{\theta}_{nT}(\gamma),\gamma) = S^{\diamond}_{\theta,nT}(\theta_0,\gamma) + \big[\tfrac{\partial}{\partial\theta'}S^{\diamond}_{\theta,nT}(\theta,\hat{\gamma}_{nT})\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}\big](\hat{\theta}_{nT}(\gamma)-\theta_0),$$

where $\{\bar{\theta}_r\}$ are on the line segment between $\hat{\theta}_{nT}(\gamma)$ and θ_0. In the following arguments, we use $H^{\diamond}_{nT}(\bar{\theta},\gamma)$ to denote $-\frac{\partial}{\partial\theta'}S^{\diamond}_{\theta,nT}(\theta,\gamma)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}$ for simplicity. Thus, we have

$$\sqrt{nT}(\hat{\theta}^{\diamond}_{nT}(\gamma)-\theta_0) = \big[\tfrac{1}{nT}H^{\diamond}_{nT}(\bar{\theta},\gamma)\big]^{-1}\tfrac{1}{\sqrt{nT}}S^{\diamond}_{\theta,nT}(\theta_0,\gamma). \tag{B.24}$$

Note that $\bar{\theta}-\theta_0 \xrightarrow{p} 0$, implied by Theorem 2.7. Hence, based on the proof of Theorem 2.2, we also have $\frac{1}{nT}H^{\diamond}_{nT}(\bar{\theta},\gamma) - \Sigma^{\diamond}_{nT}(\gamma,\gamma) \xrightarrow{p} 0$, where

$$\Sigma^{\diamond}_{nT}(\gamma_1,\gamma_2) = \begin{bmatrix} \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}^{\diamond}_{1,nT}(\gamma_1,\gamma_2), & \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}^{\diamond}_{2,nT}(\gamma_1,\gamma_2), & 0_{2k\times 1}, \\ \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}^{\diamond\prime}_{3,nT}(\gamma_2,\gamma_1), & \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}^{\diamond}_{4,nT}(\gamma_1,\gamma_2) + \mathrm{E}\mathbb{S}^{\diamond}_{nT}(\gamma_1,\gamma_2), & \frac{\bar{T}}{\sigma_0^2}\mathrm{E}\mathcal{S}^{\diamond}_{nT}(\gamma_1), \\ \sim, & \frac{\bar{T}}{\sigma_0^2}\mathrm{E}\mathcal{S}^{\diamond\prime}_{nT}(\gamma_2), & \frac{\bar{T}}{2\sigma_0^4}, \end{bmatrix}, \tag{B.25}$$

with

$$
\begin{aligned}
\mathcal{P}^{\diamond}_{1,nT}(\gamma_1,\gamma_2) &= \tfrac{1}{nT}\sum_{t=1}^{T}\widetilde{\mathbb{X}}'_t(\gamma_1)J_n\widetilde{\mathbb{X}}_t(\gamma_2),\ \mathcal{P}^{\diamond}_{2,nT}(\gamma_1,\gamma_2) = \tfrac{1}{nT}\sum_{t=1}^{T}\widetilde{\mathbb{X}}'_t(\gamma_1)J_n\widetilde{\mathbb{Z}}_t(\gamma_2),\\
\mathcal{P}^{\diamond}_{3,nT}(\gamma_1,\gamma_2) &= \tfrac{1}{nT}\sum_{t=1}^{T}\widetilde{\mathbb{X}}'_t(\gamma_1)J_n\widetilde{\mathbb{K}}_t(\gamma_2),\ \mathcal{P}^{\diamond}_{4,nT}(\gamma_1,\gamma_2) = \tfrac{1}{nT}\sum_{t=1}^{T}\widetilde{\mathbb{K}}'_t(\gamma_1)J_n\widetilde{\mathbb{Z}}_t(\gamma_2),\\
\mathcal{S}^{\diamond}_{nT}(\gamma) &= \tfrac{1}{nT}\sum_{t=1}^{T}[\mathtt{tr}(G_t),\mathtt{tr}(G^{\circ}_t(\gamma))]',\\
\mathbb{S}^{\diamond}_{nT}(\gamma_1,\gamma_2) &= [\mathbb{S}^{\diamond}_{11,nT},\ \mathbb{S}^{\diamond}_{12,nT}(\gamma_2);\ \mathbb{S}^{\diamond}_{12,nT}(\gamma_1),\ \mathbb{S}^{\diamond}_{22,nT}(\gamma_1,\gamma_2)],\\
\mathbb{S}^{\diamond}_{11,nT} &= \tfrac{1}{nT}\sum_{t=1}^{T}[\bar{T}\mathtt{tr}(G^s_tG_t) + \tfrac{1}{T^2}\sum_{k=1}^{T}\mathtt{tr}((G^s_t - G^s_k)G_t)],\\
\mathbb{S}^{\diamond}_{12,nT}(\gamma) &= \tfrac{1}{nT}\sum_{t=1}^{T}[\bar{T}\mathtt{tr}(G^s_tG^{\circ}_t(\gamma)) + \tfrac{1}{T^2}\sum_{k=1}^{T}\mathtt{tr}((G^s_t - G^s_k)G^{\circ}_t(\gamma))],\ \text{and}\\
\mathbb{S}^{\diamond}_{22,nT}(\gamma_1,\gamma_2) &= \tfrac{1}{nT}\sum_{t=1}^{T}[\bar{T}\mathtt{tr}(G^{\circ s}_t(\gamma_1)G^{\circ}_t(\gamma_2)) + \tfrac{1}{T^2}\sum_{k=1}^{T}\mathtt{tr}((G^{\circ s}_t(\gamma_1) - G^{\circ s}_k(\gamma_1))G^{\circ}_t(\gamma_2))].
\end{aligned}
$$

Note that $J_n\widetilde{V}_t(\phi_0,\gamma) = J_n\widetilde{V}_t + J_n\widetilde{\psi}^{\star}_t(\gamma_0,\gamma)$, where $\widetilde{\psi}^{\star}_t(\gamma_0,\gamma) = \widetilde{H}^{\star}_t(\gamma_0,\gamma)\phi_{20} + \lambda_{20}\widetilde{\mathcal{V}}^{\star}_t(\gamma_0,\gamma)$. Similar to (B.14), $S^{\diamond}_{\theta,nT}(\theta_0,\gamma) = S^{\diamond u}_{\theta,nT}(\theta_0,\gamma) + B^{\diamond}_{nT}(\theta_0,\gamma)$, where

$$
S^{\diamond u}_{\theta,nT}(\theta_0,\gamma) = \begin{cases}
\frac{1}{\sigma_0^2}\sum_{t=1}^{T}\widetilde{\mathbb{X}}'_t(\gamma)J_n\widetilde{V}_t,\\
\frac{1}{\sigma_0^2}\sum_{t=1}^{T}\widetilde{K}'_t(\gamma)J_n\widetilde{V}_t + \frac{1}{\sigma_0^2}\sum_{t=1}^{T}[\widetilde{V}'_tJ_nG_t(\lambda_0,\gamma)J_n\widetilde{V}_t - \sigma_0^2\bar{T}\mathtt{tr}(J_nG_t(\lambda_0,\gamma))],\\
\frac{1}{\sigma_0^2}\sum_{t=1}^{T}\widetilde{K}^{\circ\prime}_t(\gamma)J_n\widetilde{V}_t + \frac{1}{\sigma_0^2}\sum_{t=1}^{T}[\widetilde{V}'_tJ_nd_t(\gamma)G_t(\lambda_0,\gamma)J_n\widetilde{V}_t\\
\qquad\qquad -\sigma_0^2\bar{T}\mathtt{tr}(J_nd_t(\gamma)G_t(\lambda_0,\gamma))],\\
\frac{1}{2\sigma_0^4}\sum_{t=1}^{T}\widetilde{V}'_tJ_n\widetilde{V}_t - \frac{(n-1)(T-1)}{2\sigma_0^2},
\end{cases}
$$

and $B^{\diamond}_{nT}(\theta_0,\gamma)$ has four components:

$$
\begin{aligned}
B^{\diamond}_{\beta,nT}(\theta_0,\gamma) &= \tfrac{1}{\sigma_0^2}\sum_{t=1}^{T}\widetilde{\mathbb{X}}'_t(\gamma)J_n\widetilde{\psi}^{\star}_t(\gamma_0,\gamma),\\
B^{\diamond}_{\lambda_1,nT}(\theta_0,\gamma) &= \tfrac{1}{\sigma_0^2}\sum_{t=1}^{T}\widetilde{K}'_t(\gamma)J_n\widetilde{\psi}^{\star}_t(\gamma_0,\gamma) + \tfrac{1}{\sigma_0^2}\sum_{t=1}^{T}\widetilde{\psi}^{\star\prime}_t(\gamma_0,\gamma)J_nG_t(\lambda_0,\gamma)J_n\widetilde{\psi}^{\star}_t(\gamma_0,\gamma)\\
&\quad + \tfrac{1}{\sigma_0^2}\sum_{t=1}^{T}\widetilde{V}'_tJ_nG^s_t(\lambda_0,\gamma)J_n\widetilde{\psi}^{\star}_t(\gamma_0,\gamma),\\
B^{\diamond}_{\lambda_2,nT}(\theta_0,\gamma) &= \tfrac{1}{\sigma_0^2}\sum_{t=1}^{T}\widetilde{K}^{\circ\prime}_t(\gamma)J_n\widetilde{\psi}^{\star}_t(\gamma_0,\gamma)\\
&\quad + \tfrac{1}{\sigma_0^2}\sum_{t=1}^{T}\widetilde{\psi}^{\star\prime}_t(\gamma_0,\gamma)J_nd_t(\gamma)G_t(\lambda_0,\gamma)J_n\widetilde{\psi}^{\star}_t(\gamma_0,\gamma)\\
&\quad + \tfrac{1}{\sigma_0^2}\sum_{t=1}^{T}\widetilde{V}'_tJ_n[d_t(\gamma)G_t(\lambda_0,\gamma) + G'_t(\lambda_0,\gamma)d_t(\gamma)]J_n\widetilde{\psi}^{\star}_t(\gamma_0,\gamma),\\
B^{\diamond}_{\sigma^2,nT}(\theta_0,\gamma) &= \tfrac{1}{2\sigma_0^4}\sum_{t=1}^{T}\widetilde{\psi}^{\star\prime}_t(\gamma_0,\gamma)J_n[2\widetilde{V}_t + \widetilde{\psi}^{\star}_t(\gamma_0,\gamma)].
\end{aligned}
$$

Following the proof of Lemma B.1, we can also show $\frac{1}{\sqrt{nT}}S^{\diamond u}_{\theta,nT}(\theta_0,\gamma)$ will converge, uniformly in $\gamma\in\Gamma$, to a Gaussian process with mean zero and covariance $\Omega^{\diamond}_{nT}(\gamma,\gamma)$. Follow the notation defined in the proof of Theorem 2.4, one of the quadratic forms in $\frac{1}{\sqrt{nT}}S^{\diamond u}_{\theta,nT}(\theta_0,\gamma)$ is $\frac{1}{\sqrt{nT}}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{G}_{nT}(\lambda_0,\gamma)\mathbf{Q}_{nT}\mathbf{V}$, the variance of which is

$$
\begin{aligned}
&\tfrac{\kappa_4\sigma_0^4}{nT}\mathrm{E}[\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT}(\lambda_0,\gamma)\mathbf{Q}_{nT})'\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT}(\lambda_0,\gamma)\mathbf{Q}_{nT})] \\
&\quad + \tfrac{\sigma_0^4}{nT}\mathrm{E}[\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{G}_{nT}(\lambda_0,\gamma)\mathbf{Q}_{nT}\mathbf{G}_{nT}^s(\lambda_0,\gamma))] \\
&= \tfrac{\kappa_4\sigma_0^4}{nT}\mathrm{E}[\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{Q}_{nT})'\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{Q}_{nT})] \\
&\quad + \tfrac{\sigma_0^4}{nT}\mathrm{E}[\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{Q}_{nT}\mathbf{G}_{nT}^s)] + o(1),
\end{aligned}
$$

where the equation holds because $G_t(\lambda_0,\gamma) - G_t = \lambda_{20}G_t(\lambda_0,\gamma)d_t(\gamma,\gamma_0)G_t$ for each t and $\lambda_{20} = O_p((nT)^{-1/2})$ under the alternatives. In the proof of Theorem 2.4, we have already derived the summation form of the second term. For the first term, note that $\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{Q}_{nT}$ can be partitioned into T^2 blocks and only the T diagonal blocks are useful for $\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{Q}_{nT})$. These T diagonal blocks are $\bar{T}J_nG_1 - \frac{1}{T}(J_nG_1 - \frac{1}{T}\sum_{k=1}^T J_nG_k), \ldots, \bar{T}J_nG_T - \frac{1}{T}(J_nG_T - \frac{1}{T}\sum_{k=1}^T J_nG_k)$. Thus, we have

$$
\begin{aligned}
&\tfrac{1}{nT}\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{Q}_{nT})'\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G}_{nT}\mathbf{Q}_{nT}) \\
&= \tfrac{\bar{T}^2}{nT}\textstyle\sum_{t=1}^T \mathtt{diagv}(J_nG_t)'\mathtt{diagv}(J_nG_t) \\
&\quad - \tfrac{2\bar{T}}{nT^2}\textstyle\sum_{t=1}^T \mathtt{diagv}(J_nG_t)'\mathtt{diagv}(J_nG_t - \frac{1}{T}\sum_{k=1}^T J_nG_k) \\
&\quad + \tfrac{1}{nT^3}\textstyle\sum_{t=1}^T \mathtt{diagv}(J_nG_t - \frac{1}{T}\sum_{k=1}^T J_nG_k)'\mathtt{diagv}(J_nG_t - \frac{1}{T}\sum_{k=1}^T J_nG_k) \\
&= \tfrac{\bar{T}^2}{nT}\textstyle\sum_{t=1}^T R_t'R_t - \frac{2\bar{T}}{nT^2}\sum_{t=1}^T R_t'\widetilde{R}_t + \frac{1}{nT^3}\sum_{t=1}^T \widetilde{R}_t'\widetilde{R}_t + O_p(\frac{1}{n}).
\end{aligned}
$$

With these, the other variances or covariances in $\Omega_{nT}^\diamond(\gamma,\gamma)$ can derived in a similar manner. Thus, we obtain the following expression:

$$
\Omega_{nT}^\diamond(\gamma_1,\gamma_2) = \Omega_{1,nT}^\diamond(\gamma_1,\gamma_2) + \Omega_{2,nT}^\diamond(\gamma_1,\gamma_2), \tag{B.26}
$$

where

$$
\Omega_{1,nT}^\diamond(\gamma_1,\gamma_2) = \begin{bmatrix} \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}_{1,nT}^\diamond(\gamma_1,\gamma_2), & \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}_{3,nT}^\diamond(\gamma_1,\gamma_2), & 0_{2k\times 1}, \\ \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}_{3,nT}^{\diamond\prime}(\gamma_2,\gamma_1), & \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{P}_{5,nT}^\diamond(\gamma_1,\gamma_2) + \mathrm{E}\mathbb{S}_{nT}^\diamond(\gamma_1,\gamma_2), & \frac{\bar{T}}{\sigma_0^2}\mathrm{E}\mathcal{S}_{nT}^\diamond(\gamma_1), \\ 0_{1\times 2k}, & \frac{\bar{T}}{\sigma_0^2}\mathrm{E}\mathcal{S}_{nT}^{\diamond\prime}(\gamma_2), & \frac{\bar{T}}{2\sigma_0^4}, \end{bmatrix},
$$

$$
\Omega_{2,nT}^\diamond(\gamma_1,\gamma_2) =
$$

$$
\begin{bmatrix} 0_{2k\times 2k}, & \frac{\kappa_3(T-2)}{\sigma_0 T}\mathrm{E}\mathcal{R}_{1,nT}^\diamond(\gamma_1,\gamma_2), & 0_{2k\times 1}, \\ \frac{\kappa_3(T-2)}{\sigma_0 T}\mathrm{E}\mathcal{R}_{1,nT}^{\diamond\prime}(\gamma_2,\gamma_1), & \frac{2\kappa_3(T-2)}{\sigma_0 T}\sigma_0\mathrm{E}\mathcal{R}_{2,nT}^\diamond(\gamma_1,\gamma_2) + \kappa_4\mathrm{E}\mathcal{R}_{3,nT}^\diamond(\gamma_1,\gamma_2), & \frac{\kappa_4\bar{T}}{2\sigma_0^2}\mathrm{E}R_{nT}^\diamond(\gamma_1), \\ \sim, & \frac{\kappa_4\bar{T}}{2\sigma_0^2}\mathrm{E}R_{nT}^{\diamond\prime}(\gamma_2), & \frac{\kappa_4\bar{T}^2}{4\sigma_0^4}, \end{bmatrix},
$$

$$
\begin{aligned}
\mathcal{P}^{\diamond}_{5,nT}(\gamma_1,\gamma_2) &= \tfrac{1}{nT}\textstyle\sum_{t=1}^T \widetilde{\mathbb{K}}'_t(\gamma_1)J_n\widetilde{\mathbb{K}}_t(\gamma_2),\\
\mathcal{R}^{\diamond}_{1,nT}(\gamma_1,\gamma_2) &= \tfrac{1}{nT}\textstyle\sum_{t=1}^T \widetilde{\mathbb{X}}'_t(\gamma_1)J_n\widetilde{\mathbb{R}}_t(\gamma_2),\\
\mathcal{R}^{\diamond}_{2,nT}(\gamma_1,\gamma_2) &= \tfrac{1}{nT}\textstyle\sum_{t=1}^T \widetilde{\mathbb{K}}'_t(\gamma_1)J_n\widetilde{\mathbb{R}}_t(\gamma_2),\\
R^{\diamond}_{nT}(\gamma) &= \tfrac{1}{nT}\textstyle\sum_{t=1}^T [\bar{T}R'_tl_n - \tfrac{1}{T}\widetilde{R}'_tl_n,\ \bar{T}R^{\circ\prime}_t(\gamma)l_n - \tfrac{1}{T}\widetilde{R}^{\circ\prime}_t(\gamma)l_n]',\\
\mathcal{R}^{\diamond}_{3,nT}(\gamma_1,\gamma_2) &= [\mathcal{R}^{\diamond}_{31,nT},\ \mathcal{R}^{\diamond}_{32,nT}(\gamma_2);\ \mathcal{R}^{\diamond}_{32,nT}(\gamma_1),\ \mathcal{R}^{\diamond}_{33,nT}(\gamma_1,\gamma_2)],\\
\mathcal{R}^{\diamond}_{31,nT} &= \tfrac{1}{nT}\textstyle\sum_{t=1}^T [\bar{T}^2R'_tR_t - 2\tfrac{\bar{T}}{T}R'_t\widetilde{R}_t + \tfrac{1}{T^2}\widetilde{R}'_t\widetilde{R}_t],\\
\mathcal{R}^{\diamond}_{32,nT}(\gamma) &= \tfrac{1}{nT}\textstyle\sum_{t=1}^T [\bar{T}^2R'_tR^{\circ}_t(\gamma) - 2\tfrac{\bar{T}}{T}R'_t\widetilde{R}^{\circ}_t(\gamma_2) + \tfrac{1}{T^2}\widetilde{R}'_t\widetilde{R}^{\circ}_t(\gamma)],\ \text{ and}\\
\mathcal{R}^{\diamond}_{33,nT}(\gamma_1,\gamma_2) &= \tfrac{1}{nT}\textstyle\sum_{t=1}^T [\bar{T}^2R^{\circ\prime}_t(\gamma_1)R^{\circ}_t(\gamma_2) - 2\tfrac{\bar{T}}{T}R^{\circ\prime}_t(\gamma_1)\widetilde{R}^{\circ}_t(\gamma_2) + \tfrac{1}{T^2}\widetilde{R}^{\circ\prime}_t(\gamma_1)\widetilde{R}^{\circ}_t(\gamma_2)].
\end{aligned}
$$

By Lemma A.4 and under the alternatives, one shows that $\frac{1}{\sqrt{nT}}B^{\diamond}_{nT}(\theta_0,\gamma) = [\Sigma^{\diamond}_{nT}(\gamma,\gamma_0) - \Sigma^{\diamond}_{nT}(\gamma,\gamma)]\mathbf{L}C + o_p(1)$. With (B.24), $\mathbf{L}'\theta_0 = \phi_{20}$, $\sqrt{nT}\phi_{20} = C$ and $C = \mathbf{L}'\mathbf{L}C$, we have

$$
\begin{aligned}
\sqrt{nT}\mathbf{L}'\hat{\theta}^{\diamond}_{nT}(\gamma) &= \sqrt{nT}\mathbf{L}'\theta_0 + \mathbf{L}'\big[\tfrac{1}{nT}H^{\diamond}_{nT}(\bar{\theta},\gamma)\big]^{-1}\tfrac{1}{\sqrt{nT}}S^{\diamond}_{\theta,nT}(\theta_0,\gamma)\\
&= \mathbf{L}'\mathbf{L}C + \mathbf{L}'\Sigma^{\diamond-1}_{nT}(\gamma,\gamma)\tfrac{1}{\sqrt{nT}}S^{\diamond u}_{\theta,nT}(\theta_0,\gamma)\\
&\quad + \mathbf{L}'\Sigma^{\diamond-1}_{nT}(\gamma,\gamma)[\Sigma^{\diamond}_{nT}(\gamma,\gamma_0) - \Sigma^{\diamond}_{nT}(\gamma,\gamma)]\mathbf{L}C + o_p(1)\\
&\Rightarrow \mathbf{L}'\bar{S}(\gamma) + \bar{\Sigma}(\gamma)C.
\end{aligned}
$$

Given the uniform convergence of $\hat{\theta}^{\diamond}_{nT}(\gamma)$ to θ_0, it is also standard to show that $\widehat{\mathbb{Q}}_{nT}(\gamma) - \Sigma^{\diamond-1}(\gamma,\gamma)\Omega^{\diamond}(\gamma,\gamma)\Sigma^{\diamond-1}(\gamma,\gamma) \xrightarrow{p} 0$, based on the proof of Theorem 2.2. Therefore, we have $W_{nT}(\gamma) \Rightarrow W^c(\gamma)$ by the CMT. ∎

Proof of Theorem 2.9: Let

$$
S^{\diamond b}_{\theta,nT}(\gamma) = \begin{cases}
\frac{1}{\sigma_0^2}\sum_{t=1}^T \widetilde{\mathbb{X}}'_t(\gamma)J_n\widetilde{V}^b_t(\gamma),\\
\frac{1}{\sigma_0^2}\sum_{t=1}^T [\beta'_0\widetilde{K}'_t(\lambda_0,\gamma) + \widetilde{V}^{b\prime}_t(\gamma)J_nG_t(\lambda_0,\gamma)]J_n\widetilde{V}^b_t(\gamma)\\
\qquad\qquad -\bar{T}\sum_{t=1}^T \mathtt{tr}[J_nG_t(\lambda_0,\gamma)],\\
\frac{1}{\sigma_0^2}\sum_{t=1}^T [\beta'_0\widetilde{K}^{\circ\prime}_t(\lambda_0,\gamma) + \widetilde{V}^{b\prime}_t(\gamma)J_nd_t(\gamma)G_t(\lambda_0,\gamma)]J_n\widetilde{V}^b_t(\gamma)\\
\qquad\qquad -\bar{T}\sum_{t=1}^T \mathtt{tr}[J_nd_t(\gamma)G_t(\lambda_0,\gamma)],\\
\frac{1}{2\sigma_0^4}\sum_{t=1}^T \widetilde{V}^{b\prime}_t(\gamma)J_n\widetilde{V}^b_t(\gamma) - \frac{(n-1)(T-1)}{2\sigma_0^2}.
\end{cases}
$$

Thus, it suffices to show that (a) $\frac{1}{\sqrt{nT}}\widehat{S}^{\diamond}_{\theta,nT}(\gamma) - \frac{1}{\sqrt{nT}}S^{\diamond b}_{\theta,nT}(\gamma) = o_p(1)$, uniformly in $\gamma \in \Gamma$, and (b) $\frac{1}{\sqrt{nT}}S^{\diamond b}_{\theta,nT}(\gamma) \Rightarrow \frac{1}{\sqrt{nT}}S^{\diamond u}_{\theta,nT}(\theta_0,\gamma)$. The proof of Theorem 2.8 implies that $\hat{\theta}^{\diamond}_{nT}(\gamma) - \theta_0 = O_p((nT)^{-1/2})$, uniformly in $\gamma \in \Gamma$. Thus, the proof of (a) is straightforward by applying the mean value theorem. As for (b), we first note that $\mathrm{E}S^{\diamond b}_{\theta,nT}(\gamma) = 0$ and $\frac{1}{nT}\mathrm{Var}S^{\diamond b}_{\theta,nT}(\gamma) = \Omega^{\diamond}_{nT}(\gamma,\gamma) + o(1)$, for any γ. So $\frac{1}{\sqrt{nT}}S^{\diamond b}_{\theta,nT}(\gamma)$ is a mean-zero Gaussian process with covariance function kernel $\Omega^{\diamond}_{nT}(\gamma,\gamma)$ asymptotically. This can be extended to any finite collection of γ to yield the convergence of finite dimensional distributions. In addition, the stochastic equicontinuity can also be established by similar arguments to the proof of Lemma B.1. Therefore, we have $\frac{1}{\sqrt{nT}}S^{\diamond b}_{\theta,nT}(\gamma) \Rightarrow \frac{1}{\sqrt{nT}}S^{\diamond u}_{\theta,nT}(\theta_0,\gamma)$.

■

Appendix C: Supplementary proofs

Lemma C.1 and Lemma C.2 are first established to help prove the main results.

Lemma C.1. *There is a $c_1 < \infty$ such that for $\underline{\gamma} \leqslant \gamma_1 \leqslant \gamma_2 \leqslant \bar{\gamma}$ and $1 \leqslant r \leqslant 4$,*

$$(i)\quad \mathrm{E}h^r_{it}(\gamma_1,\gamma_2) \leqslant c_1|\gamma_2-\gamma_1|, \quad (ii)\ \mathrm{E}f^r_{it}(\gamma_1,\gamma_2) \leqslant c_1|\gamma_2-\gamma_1|,$$
$$(iii)\ \mathrm{E}k^r_{it}(\gamma_1,\gamma_2) \leqslant c_1|\gamma_2-\gamma_1|, \quad (iv)\ \mathrm{E}l^r_{it}(\gamma_1,\gamma_2) \leqslant c_1|\gamma_2-\gamma_1|.$$

where $h_{it}(\gamma_1,\gamma_2) = \|h_{it}\||d_{it}(\gamma_2,\gamma_1)|$, $f_{it}(\gamma_1,\gamma_2) = \|h_{it}v_{it}\||d_{it}(\gamma_2,\gamma_1)|$,
$k_{it}(\gamma_1,\gamma_2) = |v^2_{it}-\sigma^2_0||g_{ii,t}||d_{it}(\gamma_2,\gamma_1)|$, $l_{it}(\gamma_1,\gamma_2) = |v_{it}|\sum_{j\neq i}^n |g_{ij,t}||v_{jt}||d_{it}(\gamma_2,\gamma_1)|$.

Proof: We only show (i), as the others can be shown similarly. We have

$$\mathrm{E}[Zd_{it}(\gamma)] = \mathrm{E}[\mathrm{E}(Z|q_{it})d_{it}(\gamma)] = \int_{-\infty}^{\gamma} \mathrm{E}(Z|q_{it})dF(q_{it})$$

for any random variable Z, where $F(\cdot)$ denotes the CDF of q_{it}. Hence, we have $\frac{d}{d\gamma}\mathrm{E}[Zd_{it}(\gamma)] = \mathrm{E}(Z|q_{it}=\gamma)f(\gamma)$. Thus by the Jensen inequality and Assumption B(iii), one has

$$\frac{d}{d\gamma}\mathrm{E}(\|h_{it}\|^r d_{it}(\gamma)) = \mathrm{E}(\|h_{it}\|^r|q_{jt}=\gamma)f(\gamma) \leqslant [\mathrm{E}(\|h_{it}\|^4|q_{it}=\gamma)]^{r/4}f(\gamma) \leqslant c^{1+r/4}.$$

Since $d_{jt}(\gamma_2) - d_{jt}(\gamma_1)$ equals either zero or one,

$$\mathrm{E}[\|h_{it}\|^r|d_{it}(\gamma_2)-d_{jt}(\gamma_1)|] = \mathrm{E}[\|h_{it}\|^r d_{it}(\gamma_2)] - \mathrm{E}[\|h_{it}\|^r d_{it}(\gamma_1)] \leqslant c_1|\gamma_2-\gamma_1|,$$

for some $c_1 < \infty$, by a first-order Taylor series expansion, establishing (i). Assume this c_1 is large enough so that results (ii)-(iv) also hold.

Lemma C.2. *There is a $c_2 < \infty$ such that for all $\underline{\gamma} \leqslant \gamma_1 \leqslant \gamma_2 \leqslant \bar{\gamma}$,*

$$\mathrm{E}\Big|\tfrac{1}{\sqrt{nT}}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}[h_{it}^2(\gamma_1,\gamma_2)-\mathrm{E}h_{it}^2(\gamma_1,\gamma_2)]\Big|^2 \leqslant c_2|\gamma_2-\gamma_1|, \qquad (C.1)$$

$$\mathrm{E}\Big|\tfrac{1}{\sqrt{nT}}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}[f_{it}^2(\gamma_1,\gamma_2)-\mathrm{E}f_{it}^2(\gamma_1,\gamma_2)]\Big|^2 \leqslant c_2|\gamma_2-\gamma_1|, \qquad (C.2)$$

$$\mathrm{E}\Big|\tfrac{1}{\sqrt{nT}}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}[k_{it}^2(\gamma_1,\gamma_2)-\mathrm{E}k_{it}^2(\gamma_1,\gamma_2)]\Big|^2 \leqslant c_2|\gamma_2-\gamma_1|, \qquad (C.3)$$

$$\mathrm{E}\Big|\tfrac{1}{\sqrt{nT}}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}[l_{it}^2(\gamma_1,\gamma_2)-\mathrm{E}l_{it}^2(\gamma_1,\gamma_2)]\Big|^2 \leqslant c_2|\gamma_2-\gamma_1|. \qquad (C.4)$$

Proof: We only show (C.4) when $r = 2$, as the proofs of the others are similar and less difficult, using Lemma C.1. As $l_{it}(\gamma_1,\gamma_2)$ are independent across t, we have

$$\begin{aligned}
&\mathrm{E}\Big|\tfrac{1}{\sqrt{nT}}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}[l_{it}^2(\gamma_1,\gamma_2)-\mathrm{E}l_{it}^2(\gamma_1,\gamma_2)]\Big|^2\\
&= \tfrac{1}{nT}\textstyle\sum_{t=1}^{T}\mathrm{E}\Big|\sum_{i=1}^{n}[l_{it}^2(\gamma_1,\gamma_2)-\mathrm{E}l_{it}^2(\gamma_1,\gamma_2)]\Big|^2\\
&= \tfrac{1}{nT}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}\{\mathrm{E}[l_{it}^2(\gamma_1,\gamma_2)l_{jt}^2(\gamma_1,\gamma_2)]-\mathrm{E}l_{it}^2(\gamma_1,\gamma_2)\mathrm{E}l_{jt}^2(\gamma_1,\gamma_2)\}\\
&= \tfrac{1}{nT}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{n}\{\mathrm{E}l_{it}^4(\gamma_1,\gamma_2)-[\mathrm{E}l_{it}^2(\gamma_1,\gamma_2)]^2\}\\
&\quad+\tfrac{1}{nT}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\{\mathrm{E}[l_{it}^2(\gamma_1,\gamma_2)l_{jt}^2(\gamma_1,\gamma_2)]-\mathrm{E}l_{it}^2(\gamma_1,\gamma_2)\mathrm{E}l_{jt}^2(\gamma_1,\gamma_2)\}\\
&\equiv I_1(\gamma_1,\gamma_2)+I_2(\gamma_1,\gamma_2).
\end{aligned}$$

It is easy to verify that $I_1(\gamma_1,\gamma_2) \leqslant \frac{2}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\mathrm{E}[l_{it}^4(\gamma_1,\gamma_2)] \leqslant 2c_1|\gamma_2-\gamma_1|$. Further,

$$\begin{aligned}
I_2(\gamma_1,\gamma_2) = {}& \tfrac{l_0^4}{nT}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{l\neq i}^{n}\sum_{k\neq i}^{n}\sum_{m\neq j}^{n}\sum_{p\neq j}^{n}\Big\{\mathrm{E}(|g_{il,t}||g_{ik,t}||g_{jm,t}||g_{jp,t}|)\\
&\mathrm{E}|d_{it}(\gamma_2,\gamma_1)|\mathrm{E}|d_{jt}(\gamma_2,\gamma_1)|\big[\mathrm{E}(|v_{it}^2||v_{lt}^2||v_{kt}^2||v_{jt}^2||v_{mt}^2||v_{pt}^2|)\\
&-\mathrm{E}(|v_{it}^2||v_{lt}^2||v_{kt}^2|)\mathrm{E}(|v_{jt}^2||v_{mt}^2||v_{pt}^2|)\big]\Big\}.
\end{aligned}$$

Consider the term with highest order in error term, i.e., $l = k = m = p$, as the analyses of other terms are similar and less difficult. This term equals to

$$\begin{aligned}
&\tfrac{l_0^4}{nT}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{l\neq i,j}^{n}\mathrm{E}(|g_{il,t}|^2|g_{jl,t}|^2)\mathrm{E}(|d_{it}(\gamma_2,\gamma_1)||d_{jt}(\gamma_2,\gamma_1)|)\\
&\qquad\qquad\mathrm{E}|v_{it}^2|\mathrm{E}|v_{jt}^2|[\mathrm{E}|v_{lt}^8|-(\mathrm{E}|v_{lt}^4|)^2]\\
&\leqslant \tfrac{l_0^4}{nT}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{n}\mathrm{E}[(\sum_{l=1}^{n}|g_{il,t}|^2)(\sum_{j=1}^{n}|g_{jl,t}|^2)]\mathrm{E}|d_{it}(\gamma_2,\gamma_1)|\mathrm{E}|v_{it}^2|\mathrm{E}|v_{jt}^2|\mathrm{E}|v_{lt}^8|\\
&\leqslant c|\gamma_2-\gamma_1|,
\end{aligned}$$

for some constant $c < \infty$. This is because we have $\mathrm{E}(|d_{it}(\gamma_2,\gamma_1)||d_{jt}(\gamma_2,\gamma_1)|) \leqslant \mathrm{E}^{\frac{1}{2}}|d_{it}(\gamma_2,\gamma_1)|\mathrm{E}^{\frac{1}{2}}|d_{jt}(\gamma_2,\gamma_1)| = \mathrm{E}|d_{it}(\gamma_2,\gamma_1)| \leqslant c_1|\gamma_2-\gamma_1|$ based on (i) of Lemma C.1. Let c be large enough, and hence we can similarly show all the other non-zero terms in $I_2(\gamma_1,\gamma_2)$ are also bounded by $c|\gamma_2-\gamma_1|$. Thus, the desired result follows.

Proof of Lemma B.1: Firstly, we define $J_{1,nT}(\gamma) = \frac{1}{\sqrt{nT}}\sum_{t=1}^{T} H_t^{\circ\prime}(\gamma)V_t$ and $J_{2,nT}(\gamma) = \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}[V_t'\mathcal{V}_t^{\circ}(\gamma)-\sigma_0^2\mathtt{tr}(d_t(\gamma)G_t)]$. As the analysis of $\mathcal{J}_{s,nT}(\gamma)$ is tedious but follows the similar arguments to that of $J_{s,nT}(\gamma)$ for $s=1,2$, we show the uniform convergences of $J_{s,nT}(\gamma)$ instead. Lemma C.1 implies that $\mathrm{E}[\|h_{it}\|^4 d_{it}(\gamma)] < \infty$ for each γ. Meanwhile, it is easy to see that $\{d_t(\gamma)G_t\}$ are matrices with bounded row and column sum norms by Lemma A.1. Hence, $J_{1,nT}(\gamma)$ and $J_{2,nT}(\gamma)$ both converge pointwise to a Gaussian distribution by the central limit theorem (CLT) in Lemma A.3. This can be extended to any finite collection of γ to yield the convergence of the finite-dimensional distributions.

Thus, it is left to establish the tightness of $J_{s,nT}(\gamma)$ for $s=1,2$. We do this by verifying the conditions for Theorem 15.5 of Billingsley (1968). In the following, we claim that there are finite constants c_3 and c_4 such that for all $\gamma_1 \in \Gamma$, $\eta > 0$ and $\varphi \geqslant (nT)^{-1}$, if $\sqrt{nT} \geqslant c_4/\eta$,

$$P\bigg(\sup_{\gamma_1\leqslant\gamma\leqslant\gamma_1+\varphi}|J_{s,nT}(\gamma)-J_{s,nT}(\gamma_1)| > \eta\bigg) \leqslant c_3\varphi^2\eta^{-4},$$

Now suppose the above results are ture for $s=1,2$. Then, fix $\epsilon > 0$ and $\eta > 0$, and let $\varphi = \epsilon\eta^4/c_3$. The above results imply there is a large enough nT such that for any $\gamma_1 \in \Gamma$,

$$P\bigg(\sup_{\gamma_1\leqslant\gamma\leqslant\gamma_1+\varphi}|J_{s,nT}(\gamma)-J_{s,nT}(\gamma_1)| > \eta\bigg) \leqslant c_3\varphi^2\eta^{-4} = \varphi\epsilon,$$

establishing the conditions for Theorem 15.5 of Billingsley (1968).

Now, we turn to show the above claim. It is easy to see that that $J_{1,nT}(\gamma) - J_{1,nT}(\gamma_1) = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}[h_{it}v_{it}d_{it}(\gamma,\gamma_1)]$ and $J_{2,nT}(\gamma) - J_{2,nT}(\gamma_1) = J_{21,nT}(\gamma) - J_{21,nT}(\gamma_1)$, where $J_{21,nT}(\gamma) = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}[(v_{it}^2-\sigma_0^2)g_{ii,t}d_{it}(\gamma,\gamma_1)]$ and $J_{22,nT}(\gamma) - J_{22,nT}(\gamma_1) = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}[v_{it}\sum_{j\neq i}^{n} g_{ij,t}v_{jt}d_{it}(\gamma,\gamma_1)]$. Thus, their proofs are similar using the results of Lemma C.1 and C.2. We show $J_{1,nT}(\gamma)$ for example.

Since $\varphi \geqslant (nT)^{-1}$, we can let m be an integer satisfying $nT\varphi/2 \leqslant m \leqslant nT\varphi$. Set $\varphi_m = \varphi/m$. For $k = 1, \ldots, m+1$, set $\gamma_k = \gamma_1 + (k-1)\varphi_m$, $f_{it,k} = f_{it}(\gamma_k, \gamma_{k+1})$, and $f_{it,jk} = f_{it}(\gamma_k, \gamma_j)$. We let $F_{nT,k} = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T f_{it,k}$, and thus for $\gamma_k \leqslant \gamma \leqslant \gamma_{k+1}$,

$$|J_{1,nT}(\gamma) - J_{1,nT}(\gamma_1)| \leqslant \sqrt{nT}F_{nT,k} \leqslant \sqrt{nT}|F_{nT,k} - \mathrm{E}F_{nT,k}| + \sqrt{nT}\mathrm{E}F_{nT,k}.$$

It follows that

$$\begin{aligned}
&\sup_{\gamma_1 \leqslant \gamma \leqslant \gamma_1 + \varphi} |J_{1,nT}(\gamma) - J_{1,nT}(\gamma_1)| \\
&\leqslant \max_{1 \leqslant k \leqslant m} \sup_{\gamma_k \leqslant \gamma \leqslant \gamma_{k+1}} |J_{1,nT}(\gamma_k) - J_{1,nT}(\gamma_1) + J_{1,nT}(\gamma) - J_{1,nT}(\gamma_k)| \\
&\leqslant \max_{2 \leqslant k \leqslant m+1} |J_{1,nT}(\gamma_k) - J_{1,nT}(\gamma_1)| + \max_{1 \leqslant k \leqslant m} \sqrt{nT}|F_{nT,k} - \mathrm{E}F_{nT,k}| \\
&\quad + \max_{1 \leqslant k \leqslant m} \sqrt{nT}\mathrm{E}F_{nT,k}. \qquad \text{(C.5)}
\end{aligned}$$

In the following analysis, we consider to bound each term of the above equation to show the final result. For any $1 \leqslant j < k \leqslant m+1$, by the Burkholder's inequality (see Hall and Heyde (1980, p.23)) for some constant $\bar{c}_1 < \infty$,

$$\begin{aligned}
&\mathrm{E}|J_{1,nT}(\gamma_k) - J_{1,nT}(\gamma_j)|^4 \\
&\leqslant \bar{c}_1 \mathrm{E}|\tfrac{1}{nT}\textstyle\sum_{i=1}^n \sum_{t=1}^T f_{it,jk}^2|^2 \\
&= \bar{c}_1 \mathrm{E}|\tfrac{1}{nT}\textstyle\sum_{i=1}^n \sum_{t=1}^T (f_{it,jk}^2 - \mathrm{E}f_{it,jk}^2) + \tfrac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \mathrm{E}f_{it,jk}^2|^2.
\end{aligned}$$

By Minkowski's inequality, (iv) of Lemma C.1 and (C.4), the above expression is bounded by

$$\begin{aligned}
&\bar{c}_1 \big[(\mathrm{E}|\tfrac{1}{nT}\textstyle\sum_{i=1}^n \sum_{t=1}^T (f_{it,jk}^2 - \mathrm{E}f_{it,jk}^2)|^2)^{1/2} + c_1(k-j)\varphi_m\big]^2 \\
&\leqslant \bar{c}_1[(\tfrac{c_2(k-j)\varphi_m}{nT})^{1/2} + c_1(k-j)\varphi_m]^2 \leqslant \bar{c}_1(c_1 + \sqrt{c_2})^2(k-j)^2\varphi_m^2,
\end{aligned}$$

where we use the fact that $(nT)^{-1} \leqslant \varphi_m$ and $(k-j)^{1/2} \leqslant (k-j)$. Given the above result, Theorem 12.2 of Billingsley (1968, p. 94) implies that there is a finite constant $\bar{c}_2$ such that

$$P\Big(\max_{2 \leqslant k \leqslant m+1} |J_{1,nT}(\gamma_k) - J_{1,nT}(\gamma_1)| > \eta/3\Big) \leqslant 81\bar{c}_2(m\varphi_m)^2\eta^{-4} = 81\bar{c}_2\varphi^2\eta^{-4}, \qquad \text{(C.6)}$$

which bounds the first term of (C.5).

Next, we consider the second term of (C.5). By Lemma C.1, Lemma C.2

and $(nT)^{-1} \leqslant \varphi_m$,

$$\begin{aligned} \mathrm{E}|\sqrt{nT}(F_{nT,k} - \mathrm{E}F_{nT,k})|^4 &= \mathrm{E}\big|\tfrac{1}{\sqrt{nT}}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}(f_{it,k} - \mathrm{E}f_{it,k})\big|^4 \\ &\leqslant \tfrac{1}{(nT)^2}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}\mathrm{E}f_{it,k}^4 + 3[\tfrac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathrm{E}f_{it,k}^2]^2 \\ &\leqslant \tfrac{1}{nT}c_1\varphi_m + 3c_1^2\varphi_m^2 \leqslant\ (c_1 + 3c_1^2)\varphi_m^2. \end{aligned}$$

By Markov's inequality, the above inequality implies

$$\begin{aligned} P\Big(\max_{1\leqslant k\leqslant m}\sqrt{nT}|F_{nT,k} - \mathrm{E}F_{nT,k}| > \eta/3\Big) &\leqslant \textstyle\sum_{k=1}^{m} P\Big(\sqrt{nT}|F_{nT,k} - \mathrm{E}F_{nT,k}| > \eta/3\Big) \\ &\leqslant 81m(c_1 + 3c_1^2)\varphi_m^2\eta^{-4} \leqslant\ 81(c_1 + 3c_1^2)\varphi^2\eta^{-4}, \end{aligned}$$

where the final equality uses $m\varphi_m = \varphi$ and $\varphi_m \leqslant \varphi$.

Finally, we consider the last term of (C.5). By (iv) of Lemma C.1 and $\varphi_m \leqslant \frac{2}{nT}$,

$$\sqrt{nT}\mathrm{E}F_{nT,k} = \sqrt{nT}\mathrm{E}f_{it,k} \leqslant \sqrt{nT}c_1\varphi_m \leqslant 2c_1(nT)^{-1/2}.$$

Aggregating the above results for the three terms of (C.5), we have if $2c_1(nT)^{-1/2} \leqslant \eta/3$,

$$P\left(\sup_{\gamma_1\leqslant\gamma\leqslant\gamma_1+\varphi}|J_{1,nT}(\gamma) - J_{1,nT}(\gamma_1)| > \eta\right) \leqslant\ 81(\bar{c}_2 + c_1 + 3c_1^2)\varphi^2\eta^{-4}. \quad \text{(C.7)}$$

By setting $c_3 = 81(\bar{c}_2 + c_1 + 3c_1^2)$ and $c_4 = 6c_1$, we achieve the desired result. ∎

Proof of Lemma B.2: We show the result for $\mathcal{F}_{nT}(v)$, as the other two can be shown similarly. For notation simplicity, let $m_{it} = \delta_0' h_{it}$ and $m_{it}(v) = \delta_0' h_{it} d_{it}(\gamma_0, \gamma_0 + v/a_{nT})$. Hence,

$$\begin{aligned} \mathcal{F}_{nT}(v) = &\tfrac{a_{nT}}{nT}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T} m_{it}^2(v) - \tfrac{a_{nT}}{nT^2}\sum_{i=1}^{n}\sum_{k=1}^{T}\sum_{t=1}^{T} m_{it}(v)m_{ik}(v) \\ &- \tfrac{a_{nT}}{n^2T}\textstyle\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T} m_{it}(v)m_{jt}(v) \\ &+ \tfrac{a_{nT}}{n^2T^2}\textstyle\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{T}\sum_{t=1}^{T} m_{it}(v)m_{jk}(v) \equiv\ \sum_{s=1}^{4}\mathcal{F}_{s,nT}(v). \end{aligned} \quad \text{(C.8)}$$

Consider the case where v is positive first. Observe that since $\gamma_1 = \gamma_0 + v/a_{nT} \to \gamma_0$,

$$a_{nT}P(\gamma_0 < q_{it} \leqslant \gamma_1) = v\frac{P(q_{it} \leqslant \gamma_1) - P(q_{it} \leqslant \gamma_0)}{\gamma_1 - \gamma_0} \to f|v| \quad \text{(C.9)}$$

as simple size increases. Symmetrically, we can show that $a_{nT}P(\gamma_1 < q_{it} \leqslant \gamma_0) \to f|v|$, when v is negative. In the following argument, we only consider the case where v is positive, as the negative case can be study symmetrically. Thus,

$$\begin{aligned}\mathrm{E}\mathcal{F}_{1,nT}(v) &= \tfrac{a_{nT}}{nT}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}\mathrm{E}[m_{it}^2\mathbb{1}\{\gamma_0 < q_{it} \leqslant \gamma_1\}] \\ &= \tfrac{1}{nT}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}\mathrm{E}(m_{it}^2|\gamma_0 < q_{it} \leqslant \gamma_1)a_{nT}P(\gamma_0 < q_{it} \leqslant \gamma_1) \\ &\to \tfrac{1}{nT}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}\mathrm{E}(m_{it}^2|q_{it} = \gamma_0)f|v| = \delta_0' M\delta_0 f|v|.\end{aligned}$$

Besides, by (C.1),

$$\begin{aligned}\mathrm{E}|\mathcal{F}_{1,nT}(v) - \mathrm{E}\mathcal{F}_{1,nT}(v)|^2 &\leqslant \tfrac{a_{nT}^2}{nT}\|\delta_0\|^4\mathrm{E}\Big|\tfrac{1}{\sqrt{nT}}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}[h_{it}^2(\gamma_0,\gamma_1) - \mathrm{E}h_{it}^2(\gamma_0,\gamma_1)]\Big|^2 \\ &\leqslant \tfrac{a_{nT}}{nT}\|\delta_0\|^4 c_2|v| = o(1).\end{aligned}$$

Hence, the Markov's inequality implies that $\mathcal{F}_{1,nT}(v) - \delta_0' M\delta_0 f|v| \xrightarrow{p} 0$.

We next consider the second term of (C.8). By (C.9), for $i \neq j$ or $t \neq k$, $a_{nT}P(\gamma_0 < q_{it} \leqslant \gamma_1, \gamma_0 < q_{jk} \leqslant \gamma_1) = a_{nT}P(\gamma_0 < q_{it} \leqslant \gamma_1)P(\gamma_0 < q_{jk} \leqslant \gamma_1) \to 0$. Hence,

$$\begin{aligned}&\mathrm{E}\mathcal{F}_{2,nT}(v) = \tfrac{a_{nT}}{nT^2}\textstyle\sum_{i=1}^{n}\sum_{k=1}^{T}\sum_{t=1}^{T}\mathrm{E}[m_{it}m_{ik}\mathbb{1}\{\gamma_0 < q_{it} \leqslant \gamma_1\}\mathbb{1}\{\gamma_0 < q_{ik} \leqslant \gamma_1\}] \\ &= \tfrac{1}{nT^2}\textstyle\sum_{i=1}^{n}\sum_{k=1}^{T}\sum_{t=1}^{T}\mathrm{E}(m_{it}m_{ik}|\gamma_0 < q_{it} \leqslant \gamma_1, \gamma_0 < q_{ik} \leqslant \gamma_1)a_{nT} \\ &\qquad\qquad P(\gamma_0 < q_{it} \leqslant \gamma_1, \gamma_0 < q_{ik} \leqslant \gamma_1) \\ &= \tfrac{1}{nT^2}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}\mathrm{E}(m_{it}^2|\gamma_0 < q_{it} \leqslant \gamma_1)a_{nT}P(\gamma_0 < q_{it} \leqslant \gamma_1) + o_p(1) \to \tfrac{1}{T}\delta_0' M\delta_0 f|v|.\end{aligned}$$

Similarly, we have

$$\mathrm{E}|\mathcal{F}_{2,nT}(v) - \mathrm{E}\mathcal{F}_{2,nT}(v)|^2 \leqslant \mathrm{E}\mathcal{F}_{2,nT}^2(v) = \tfrac{a_{nT}^2}{n^2T^4}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}\mathrm{E}m_{it}^4(v) + o_p(1) \to 0.$$

Hence, the Markov's inequality implies $\mathcal{F}_{2,nT}(v) - \frac{1}{T}\delta_0' M\delta_0 f|v| \xrightarrow{p} 0$.

Similarly, we have

$$\tfrac{a_{nT}}{n^2T}\textstyle\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}m_{it}(v)m_{jt}(v) = \tfrac{1}{n}\delta_0' M\delta_0 f|v| + o_p(1) \xrightarrow{p} 0$$

and

$$\tfrac{a_{nT}}{n^2T^2}\textstyle\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{T}\sum_{t=1}^{T}m_{it}(v)m_{jk}(v) = \tfrac{1}{nT}\delta_0' M\delta_0 f|v| + o_p(1) \xrightarrow{p} 0.$$

Since $\mathcal{F}_{nT}(v)$ is monotonically increasing on $[0, \bar{v}]$ and decreasing on $[-\bar{v}, 0]$, and the limit function is continuous, the convergence is uniform over Υ. ■

Proof of Lemma B.3: The uniform convergence follows if

(a) The finite dimensional distributions of $\mathcal{R}_{nT}(v)$ converge to those of $B(v)$;

(b) $\mathcal{R}_{nT}(v)$ is tight.

We show (a) first. According to Assumptions A to D, the conditions for the CLT in Lemma A.3 are well established. Hence, for any given $v \in \Upsilon$, we have $\mathcal{R}_{nT}(v) \xrightarrow{D} N(0, \sigma^2_{\mathcal{R}}(v))$, where $\sigma^2_{\mathcal{R}}(v)$ is the variance of $\mathcal{R}_{nT}(v)$. Then, it is left to show $\sigma^2_{\mathcal{R}}(v) = |v|\Xi f$. Let $\Pi^{\diamond}_t(v) = d_t(\gamma_0 + v/a_{nT}, \gamma_0)C_t$ for $\Pi_t = H_t$, R_t or G_t. By Lemma B.5 of Yang (2015), we have

$$\begin{aligned}\sigma^2_{\mathcal{R}}(v) &= \sigma_0^2 \mathrm{E}\mathcal{F}_{nT}(v) + 2l_0\sigma_0^3\kappa_3\bar{T}\tfrac{a_{nT}}{nT}\textstyle\sum_{t=1}^T \mathrm{E}[\delta_0' \widetilde{H}^{\diamond\prime}_t(v) J_n \widetilde{R}^{\diamond}_t(v)] \\ &\quad + \bar{T}^2 l_0^2\sigma_0^4\kappa_4\tfrac{a_{nT}}{nT}\textstyle\sum_{t=1}^T R^{\diamond\prime}_t(v)R^{\diamond}_t(v) \\ &\quad + \bar{T} l_0^2\sigma_0^4\tfrac{a_{nT}}{nT}\textstyle\sum_{t=1}^T \mathtt{tr}[J_n G^{\diamond}_t(v)(G^{\diamond}_t(v)J_n + J_n G^{\diamond}_t(v))] \\ &\quad + l_0^2\sigma_0^4\tfrac{a_{nT}}{nT^3}\textstyle\sum_{t=1}^T\sum_{k=1}^T \mathtt{tr}[J_n(G^{\diamond}_t(v) - G^{\diamond}_k(v))J_n G^{\diamond}_t(v)] \equiv \textstyle\sum_{s=1}^5 \mathcal{C}_s.\end{aligned}$$

By Lemma B.2, we have $(\mathcal{C}_1 + \mathcal{C}_4) - \sigma_0^2\Xi_1 f|v| \xrightarrow{p} 0$. Similar to the proof of Lemma B.2, we can also show $(\mathcal{C}_2 + \mathcal{C}_3) - \sigma_0^2\Xi_2 f|v| \xrightarrow{p} 0$ and $\mathcal{C}_5 \xrightarrow{p} 0$. Hence, we conclude that $\mathcal{R}_{nT}(v) \xrightarrow{D} N(0, \Xi f|v|)$. This argument can be extended to include any finite collection $[v_1, \ldots, v_k]$ to yield the convergence of the finite dimensional distributions of $\mathcal{R}_{nT}(v)$ to those of $B(v)$.

We now show (b). By Lemma B.1, for all $\gamma_j \in \Gamma$, $\eta > 0$ and $\varphi \geqslant (nT)^{-1}$, there exist finite constant c_3 and c_4 such that if $\eta \geqslant c_4/\sqrt{nT}$,

$$P\Bigg(\sup_{\gamma_j \leqslant \gamma \leqslant \gamma_j + \varphi} \|\delta_0'(\mathcal{J}^{\star}_{1,nT}(\gamma, \gamma_j)) + l_0(\mathcal{J}^{\star}_{2,nT}(\gamma, \gamma_j))\| > \eta\Bigg) \leqslant \tfrac{1}{\eta^4}c_3\varphi^2. \tag{C.10}$$

Fix $\epsilon > 0, \eta_1 > 0$. Set $\varphi_1 = \epsilon\eta_1^4/c_3$, $\varphi = \varphi_1/a_{nT}, \eta = \eta_1/\sqrt{a_{nT}}$ and $N_1 = (\max(\varphi^{-1/2}, c_4/\eta_1))^{1/\tau}$. Hence, for $nT \geqslant N_1$, we have $\varphi = \frac{\epsilon\eta^4}{nTc_3}(nT)^{2\tau} \geqslant \frac{\epsilon\eta^4}{nT\varphi c_3} = (nT)^{-1}$ and $\eta \geqslant c_4/\sqrt{nT}$. Set $\gamma_1 = \gamma_0 + v_1/a_{nT}$. By (C.10), for $nT \geqslant N_1$,

$$\begin{aligned}&P\Bigg(\sup_{v_1 \leqslant v \leqslant v_1 + \varphi_1} |\mathcal{R}_{nT}(v) - \mathcal{R}_{nT}(v_1)| > \eta_1\Bigg) \\ &= P\Bigg(\sup_{\gamma_1 \leqslant \gamma \leqslant \gamma_1 + \varphi} \|\delta_0'\mathcal{J}^{\star}_{1,nT}(\gamma, \gamma_j) + l_0\mathcal{J}^{\star}_{2,nT}(\gamma, \gamma_1)\| > \eta\Bigg) \\ &\leqslant \tfrac{1}{\eta_1^4}c_3 a_{nT}^2(\varphi_1/a_{nT})^2 = \varphi_1\epsilon.\end{aligned}$$

As discussed in the proof of Lemma B.1, this shows that (b) holds. ■

Proof of Lemma B.4: Firstly, we show (a) when $r = 1$, and the proofs of the other results in (a)-(d) are similar and thus omitted. Note that $D_{1,nT}(\gamma)$ is just a linear transformation of $D_{11,nT}(\gamma) = \frac{1}{nT}\sum_{t=1}^{T}\delta_0' H_t^{\star\prime}(\gamma_0,\gamma)H_t^{\star}(\gamma_0,\gamma)\delta_0$. It suffices to show

$$P\left(\sup_{\gamma\in\mathcal{N}_{nT}}\frac{D_{11,nT}(\gamma)}{|\gamma-\gamma_0|} < (1-\eta)k\right) \leqslant \epsilon.$$

Without loss of generality (WLOG), we assume $\gamma > \gamma_0$, as a symmetric argument can be established for the case of $\gamma < \gamma_0$. Hence,

$$d\mathrm{E}D_{11,nT}(\gamma)/d\gamma = \delta_0' M(\gamma)f(\gamma)\delta_0.$$

Since $\delta_0' M(\gamma)f(\gamma)\delta_0 > 0$ (Assumption B(v)) and $\delta_0' M(\gamma)f(\gamma)\delta_0$ is continuous at γ_0 (Assumption B(iv)), then there is a B sufficiently small such that

$$k = \min_{|\gamma-\gamma_0|\leqslant B}\delta_0' M(\gamma)f(\gamma)\delta_0 > 0,$$

Because $\mathrm{E}D_{11,nT}(\gamma_0) = 0$, a first-order Taylor series expansion about γ_0 yields

$$\inf_{|\gamma-\gamma_0|\leqslant B}\mathrm{E}D_{11,nT}(\gamma) \geqslant k|\gamma-\gamma_0|. \tag{C.11}$$

Then, (C.1) implies

$$\begin{aligned}\mathrm{E}|D_{11,nT}(\gamma)-\mathrm{E}D_{11,nT}(\gamma)|^2 &\leqslant \|\delta_0\|^4\mathrm{E}|\tfrac{1}{nT}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}[h_{it}^2(\gamma_1,\gamma_2)-\mathrm{E}h_{it}^2(\gamma_1,\gamma_2)]|^2\\ &\leqslant \|\delta_0\|^4(nT)^{-1}c_2|\gamma-\gamma_0|. \end{aligned} \tag{C.12}$$

For any η and ϵ, set

$$b = \frac{1-\eta/2}{1-\eta} > 1, \text{ and} \tag{C.13}$$

$$\bar{v} = \frac{8\|\delta_0\|^4 c_2}{\epsilon\eta^2k^2(1-1/b)}. \tag{C.14}$$

We may assume that (n, T) is large enough so that $\frac{\bar{v}}{a_{nT}} \leqslant B$, else the inequality (a) is trival. For $l = 1, 2, \ldots, \bar{N}+1$, set $\gamma_j = \gamma_0 + \bar{v}b^{j-1}/a_{nT}$, where $\bar{N}$ is the integer such that $\gamma_{\bar{N}} - \gamma_0 = \bar{v}b^{\bar{N}-1}/a_{nT} \leqslant B$ and $\gamma_{\bar{N}+1} - \gamma_0 = \bar{v}b^{\bar{N}}/a_{nT} > B$. (Note that $\bar{N} \geqslant 1$ since $\frac{\bar{v}}{a_{nT}} \leqslant B$.)

Markov's inequality, (C.11) and (C.12) yield

$$
\begin{aligned}
P\Big(\sup_{1\leqslant j\leqslant \bar N}\Big|\frac{D_{11,nT}(\gamma_j)-\mathrm{E}D_{11,nT}(\gamma_j)}{\mathrm{E}D_{11,nT}(\gamma_j)}\Big|>\frac{\eta}{2}\Big) &\leqslant \frac{4}{\eta^2}\sum_{j=1}^{\bar N}\frac{\mathrm{E}|D_{11,nT}(\gamma_j)-\mathrm{E}D_{11,nT}(\gamma_j)|^2}{|\mathrm{E}D_{11,nT}(\gamma_j)|^2}\\
&\leqslant \frac{4}{\eta^2}\sum_{j=1}^{\bar N}\frac{\|\delta_0\|^4(nT)^{-1}c_2}{k^2|\gamma_j-\gamma_0|}\\
&\leqslant (nT)^{-2\tau}\frac{4\|\delta_0\|^4c_2}{\eta^2k^2\bar v}\sum_{j=1}^{\infty}\frac{1}{b^{j-1}}\\
&\leqslant \frac{4\|\delta_0\|^4c_2}{\eta^2k^2\bar v(1-1/b)}=\frac{\epsilon}{2}, \qquad \text{(C.15)}
\end{aligned}
$$

where the final equation is based on (C.14). Thus, with probability exceeding $1-2/\epsilon$, $\left|\frac{D_{11,nT}(\gamma_j)}{\mathrm{E}D_{11,nT}(\gamma_j)}-1\right|\leqslant\frac{\eta}{2}$ for all $1\leqslant j\leqslant\bar N$. So for any $\gamma\in[\gamma_0+\bar v/a_{nT},\gamma_0+B]$, there is some $1\leqslant j\leqslant\bar N$ such that $\gamma_j<\gamma<\gamma_{j+1}$ and

$$
\frac{D_{11,nT}(\gamma)}{|\gamma-\gamma_0|}\geqslant\frac{D_{11,nT}(\gamma_j)}{\mathrm{E}D_{11,nT}(\gamma_j)}\frac{\mathrm{E}D_{11,nT}(\gamma_j)}{|\gamma_{j+1}-\gamma_0|}\geqslant(1-\frac{\eta}{2})\frac{k|\gamma_j-\gamma_0|}{|\gamma_{j+1}-\gamma_0|}=(1-\frac{\eta}{2})\frac{k}{b}
$$

with probability exceeding $1-\epsilon/2$, according to (C.15). Based on the definition of b, (C.13), the above inequality can be simplified as $\frac{D_{11,nT}(\gamma)}{|\gamma-\gamma_0|}\geqslant(1-\eta)k$. Since this event has probability exceeding $1-\epsilon/2$, we have established

$$
P\Big(\inf_{\gamma\in\mathcal{N}_{nT}}\frac{D_{11,nT}(\gamma)}{|\gamma-\gamma_0|}<(1-\eta)k\Big)\leqslant\frac{\epsilon}{2}.
$$

A symmetric argument applies to the case $-B\leqslant\gamma-\gamma_0\leqslant-\frac{\bar v}{a_{nT}}$.

Secondly, we show the results in (e). WLOG, we assume $\gamma>\gamma_0$. Let $\gamma_j=\gamma_0+\bar v b^{j-1}/a_{nT}$ for $l=1,2,\dots,\bar N+1$, where b and $\bar N$ are defined as before. By definition, it is seen that there are at most $\log_b(a_{nT}B/\bar v)+2$ points in the interval $\gamma-\gamma_0\in[\frac{\bar v}{a_{nT}},B]$, i.e., $\bar N\leqslant\log_b(a_{nT}B/\bar v)+2$. Then, for $r=1,2,3$,

$$
P\Big(\sup_{\gamma\in\mathcal{N}_{nT}}\frac{\|P_{r,nT}(\gamma)\|}{|\gamma-\gamma_0|}>\eta\Big)=P\Big(\max_{1\leqslant j\leqslant\bar N}\frac{\|P_{r,nT}(\gamma_j)\|}{|\gamma_j-\gamma_0|}>\eta\Big)\leqslant\sum_{j=1}^{\bar N}P\Big(\frac{\|P_{r,nT}(\gamma_j)\|}{|\gamma_j-\gamma_0|}>\eta\Big).
$$

Following the proof of Lemma C.2, for any j, we have $\mathrm{E}\|P_{r,nT}(\gamma_j)\|^2\leqslant\frac{c_2}{nT}|\gamma_j-\gamma_0|$. Thus, Chebyshev inequality implies that

$$
\sum_{j=1}^{\bar N}P\Big(\frac{\|P_{r,nT}(\gamma_j)\|}{|\gamma_j-\gamma_0|}>\eta\Big)\leqslant\sum_{j=1}^{\bar N}\frac{\mathrm{E}\|P_{r,nT}(\gamma_j)\|^2}{|\gamma_j-\gamma_0|^2\eta^2}\leqslant\sum_{j=1}^{\infty}\frac{c_2a_{nT}}{nT\bar v b^{j-1}\eta^2}\leqslant\frac{c_2(nT)^{-2\tau}}{\bar v(1-1/b)\eta^2}\to 0.
$$

A symmetric argument establishes a similar result for $\gamma<\gamma_0$.

Finally, we consider the two results in (f). As their proofs follow the same manner, we use general notation $\mathcal{J}_{s,nT}(\gamma)$ to denote either of them. Fix $\eta > 0$. For $j = 1, 2, \ldots$, set $\gamma_j - \gamma_0 = \bar{v}2^{j-1}/a_{nT}$, where $\bar{v} < \infty$ will be determined later. By the similar analysis as used in the proof of Lemma B.1, for all $\gamma_j \in \Gamma$, $\eta > 0$ and $\varphi \geqslant (nT)^{-1}$, there exist $c_3, c_4 < \infty$ such that if $\eta \geqslant c_4/\sqrt{nT}$,

$$\mathrm{E}\|\mathcal{J}_{s,nT}(\gamma_j) - \mathcal{J}_{s,nT}(\gamma_0)\|^2 \leqslant c_1|\gamma_j - \gamma_0|, \quad \text{and} \tag{C.16}$$

$$P\biggl(\sup_{\gamma_j \leqslant \gamma \leqslant \gamma_j + \varphi} \|\mathcal{J}_{s,nT}(\gamma) - \mathcal{J}_{s,nT}(\gamma_j)\| > \eta\biggr) \leqslant c_3\varphi^2\eta^{-4}. \tag{C.17}$$

Next, we do the following decomposition

$$\begin{aligned}
&\sup_{\gamma \in \mathcal{N}_{nT}} \frac{\|\mathcal{J}_{s,nT}(\gamma) - \mathcal{J}_{s,nT}(\gamma_0)\|}{\sqrt{a_{nT}}|\gamma - \gamma_0|} \\
&= \sup_j \sup_{\gamma_j \leqslant \gamma \leqslant \gamma_{j+1}} \frac{\|\mathcal{J}_{s,nT}(\gamma) - \mathcal{J}_{s,nT}(\gamma_0)\|}{\sqrt{a_{nT}}|\gamma_j - \gamma_0|} \frac{|\gamma_j - \gamma_0|}{|\gamma - \gamma_0|} \\
&\leqslant \sup_j \sup_{\gamma_j \leqslant \gamma \leqslant \gamma_{j+1}} \frac{\|\mathcal{J}_{s,nT}(\gamma) - \mathcal{J}_{s,nT}(\gamma_j)\|}{\sqrt{a_{nT}}|\gamma_j - \gamma_0|} + \sup_j \frac{\|\mathcal{J}_{s,nT}(\gamma_j) - \mathcal{J}_{s,nT}(\gamma_0)\|}{\sqrt{a_{nT}}|\gamma_j - \gamma_0|}.
\end{aligned} \tag{C.18}$$

For the first term of (C.18), we set $\varphi_j = \gamma_{j+1} - \gamma_j$ and $\eta_j = \sqrt{a_{nT}}|\gamma_j - \gamma_0|\eta/2$, and then

$$\begin{aligned}
&P\biggl(\sup_j \sup_{\gamma_j \leqslant \gamma \leqslant \gamma_{j+1}} \frac{\|\mathcal{J}_{s,nT}(\gamma) - \mathcal{J}_{s,nT}(\gamma_j)\|}{\sqrt{a_{nT}}|\gamma_j - \gamma_0|} > \eta/2\biggr) \\
&\leqslant \sum_{j=1}^{\infty} P\biggl(\sup_{\gamma_j \leqslant \gamma \leqslant \gamma_j + \varphi_j} \|\mathcal{J}_{s,nT}(\gamma) - \mathcal{J}_{s,nT}(\gamma_j)\| > \eta_j\biggr).
\end{aligned}$$

Note that if $\bar{v} \geqslant 1$, then $\varphi_j \geqslant 1/a_{nT} \geqslant 1/n$. In addition , if $\bar{v} \geqslant 12c_1/\eta$, then $\eta_j = \bar{v}2^{j-2}\eta/\sqrt{a_{nT}} \geqslant c_4/\sqrt{a_{nT}} \geqslant c_4/\sqrt{nT}$. Thus, if $\bar{v} \geqslant \max(1, 12c_1/\eta)$, using (C.17), the right hand side of above inequality is bounded by

$$\sum_{j=1}^{\infty} \frac{c_3\varphi_j^2}{\eta_j^4} = \sum_{j=1}^{\infty} \frac{16c_3|\gamma_{j+1} - \gamma_j|^2}{a_{nT}^2|\gamma_j - \gamma_0|^4\eta^4} = \frac{64c_3}{3\bar{v}^2\eta^4}.$$

For the second term of (C.18), Markov's inequality and (C.16) imply

$$\begin{aligned}
P\biggl(\sup_j \frac{\|\mathcal{J}_{s,nT}(\gamma_j) - \mathcal{J}_{s,nT}(\gamma_0)\|}{\sqrt{a_{nT}}|\gamma_j - \gamma_0|} > \eta/2\biggr) &\leqslant \sum_{j=1}^{\infty} P\biggl(\frac{\|\mathcal{J}_{s,nT}(\gamma_j) - \mathcal{J}_{s,nT}(\gamma_0)\|}{\sqrt{a_{nT}}|\gamma_j - \gamma_0|} > \eta/2\biggr) \\
&\leqslant \sum_{j=1}^{\infty} \frac{4\mathrm{E}\|\mathcal{J}_{s,nT}(\gamma_j) - \mathcal{J}_{s,nT}(\gamma_0)\|^2}{a_{nT}|\gamma_j - \gamma_0|^2\eta^2} \\
&\leqslant \sum_{j=1}^{\infty} \frac{4c_1|\gamma_j - \gamma_0|}{a_{nT}|\gamma_j - \gamma_0|^2\eta^2} = \frac{8c_1}{\bar{v}\eta^2}.
\end{aligned}$$

Together, if $\bar{v} \geqslant \max(1, 12c_1/\eta)$ we have

$$P\left(\sup_{\gamma \in \mathcal{N}_{nT}} \frac{\|\mathcal{J}_{s,nT}(\gamma) - \mathcal{J}_{s,nT}(\gamma_0)\|}{\sqrt{a_{nT}}|\gamma - \gamma_0|} > \eta\right) \leqslant \frac{64c_3}{3\bar{v}^2\eta^4} + \frac{8c_1}{\bar{v}\eta^2},$$

which can be made arbitrarily small by picking suitably large $\bar{v}$. Thus, results in (f) hold. ∎

Appendices for Chapter 3

We collect all the technical proofs for the main results in Chapter 3 here. Two appendices are provided in this part. Appendix A provides some basic lemmas that are used in the other appendix. Appendix B presents proofs for results in Section 3.2.

Appendix A: Some basic lemmas

The following lemmas are essential to the proofs of the main results in this chapter.

Lemma A.1. *(Kelejian and Prucha, 1999; Lee, 2002): Let $\{A_N\}$ and $\{B_N\}$ be two sequences of $N \times N$ matrices that are uniformly bounded in both row and column sums. Let C_N be a sequence of conformable matrices whose elements are uniformly $O(h_n^{-1})$. Then,*

(i) the sequence $\{A_N B_N\}$ are uniformly bounded in both row and column sums,

(ii) the elements of A_N are uniformly bounded and $tr(A_N) = O(N)$, and

(iii) the elements of $A_N C_N$ and $C_N A_N$ are uniformly $O(h_n^{-1})$.

Lemma A.2. *(Lemma A.3, Lee, 2004): For $\mathbf{W}$ and $\mathbf{A}_{nT}(\lambda)$ defined in Model* (3.6)*, if $\|\mathbf{W}\|$ and $\|\mathbf{A}_{nT}^{-1}\|$ are uniformly bounded, where $\|\cdot\|$ is a matrix norm, then $\|\mathbf{A}_{nT}^{-1}(\lambda)\|$ is uniformly bounded in a neighborhood of λ_0.*

Lemma A.3. *Under Assumptions C-E, we have*

(i) $\mathbb{Q}_{\mathbb{D}}(\rho)$ is uniformly bounded in both row and column sums, uniformly in $\rho \in \Delta_\rho$;

(ii) $\mathbb{Q}_{\mathbb{Z}}(\rho)$ is uniformly bounded in both row and column sums, uniformly in $\rho \in \Delta_\rho$.

Proof : Under Assumptions C-E, it can be shown similarly to Lemma A.3 of Chapter 1. ■

Lemma A.4. *(Lemma A.4, Chapter 1): Suppose that* $\{A_N\}$ *and* $\{B_N\}$ *are two sequences of* $N \times N$ *matrices that are uniformly bounded in either row or column sums. Under Assumptions C-E,* $\mathtt{tr}[A_N \mathbb{P}_{\mathbb{Z}}(\rho) B_N] = O(1)$*, uniformly in* $\rho \in \Delta_\rho$.

Lemma A.5. *(Lemma B.5, Yang 2015): Let* A_N *and* D_N *be* $N \times N$ *matrices,* b_N *an* $N \times 1$ *vector, and* ϵ_N *an* $N \times 1$ *random vector of iid elements with mean zero, variance* σ^2*, skewness* γ*, and excess kurtosis* κ*. Let* $Q_N = \epsilon_N' A_N \epsilon_N + b_N' \epsilon_N$ *and* $S_N = \epsilon_N' D_N \epsilon_N$*. Then*

(i) $\mathrm{E}(Q_N) = \sigma^2 \mathtt{tr}(A_N)$ *and* $\mathrm{E}(S_N) = \sigma^2 \mathtt{tr}(D_N)$,

(ii) $\mathrm{Var}(Q_N) = \sigma^4 \mathtt{tr}[A_N(A_N + A_N')] + \sigma^4 \kappa a_N' a_N + \sigma^2 b_N' b_N + 2\sigma^3 \gamma a_N' b_N$,

(iii) $\mathrm{Var}(S_N) = \sigma^4 \mathtt{tr}[D_N(D_N + D_N')] + \sigma^4 \kappa d_N' d_N$,

(iv) $\mathrm{Cov}(Q_N, S_N) = \sigma^4 \mathtt{tr}[A_N(D_N + D_N')] + \sigma^4 \kappa a_N' d_N + \sigma^3 \gamma b_N' d_N$.

where a_N *and* d_N *are column vectors of diagonal elements of* A_N *and* D_N*, respectively.*

Lemma A.6. *(Lemma A.3, Lin and Lee, 2010, extended): Let* $\{A_N\}$ *be a sequence of* $N \times N$ *matrices such that either* $\|A_N\|_\infty$ *or* $\|A_N\|_1$ *is bounded. Suppose that the elements of* A_N *are* $O(h_n^{-1})$ *uniformly in all* i *and* j*. Let innovation vector* $\mathbf{V}$ *be defined as in Lemma A.5. Let* c_N *be an* $N \times 1$ *vector with elements of uniform order* $O(h_n^{-1/2})$*. Then*

(i) $\mathrm{E}(\mathbf{V}' A_N \mathbf{V}) = O(\frac{N}{h_n})$, (ii) $\mathrm{Var}(\mathbf{V}' A_N \mathbf{V}) = O(\frac{N}{h_n})$,

(iii) $\mathbf{V}' A_N \mathbf{V} = O_p(\frac{N}{h_n})$, (iv) $\mathbf{V}' A_N \mathbf{V} - \mathrm{E}(\mathbf{V}' A_N \mathbf{V}) = O_p((\frac{N}{h_n})^{\frac{1}{2}})$,

(v) $c_N' A_N \mathbf{V} = O_p((\frac{N}{h_n})^{\frac{1}{2}})$, **if** $\|A_N\|_1$ *is bounded.*

Appendix B: Proofs of the theorems

In proving the theorems, the following facts are used: (i) the eigenvalues of a projection matrix are either 0 or 1; (ii) the eigenvalues of a positive definite (p.d.) matrix are strictly positive; (iii) $\gamma_{min}(A)\mathtt{tr}(B) \leq \mathtt{tr}(AB) \leq \gamma_{max}(A)\mathtt{tr}(B)$ for symmetric matrix A and positive semi-definite (p.s.d.) matrix B; (iv) $\gamma_{max}(A+B) \leq \gamma_{max}(A) + \gamma_{max}(B)$ for symmetric matrices A and B; and (v) $\gamma_{max}(AB) \leq \gamma_{max}(A)\gamma_{max}(B)$ for p.s.d. matrices A and B.

Derivation of the AQS functions and the Hessian matrix:

Writing the key quantity in the concentrated quasi loglikelihood (3.9) as $\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) = [\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\beta]'\mathbf{B}'_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)[\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\beta]$, and using the facts that for an invertible matrix $A(\lambda)$, $\frac{\partial}{\partial\lambda}\ln|A(\lambda)| = \mathtt{tr}[A^{-1}(\lambda)\frac{\partial}{\partial\lambda}A(\lambda)]$ and $\frac{\partial}{\partial\lambda}A^{-1}(\lambda) = -A^{-1}(\lambda)[\frac{\partial}{\partial\lambda}A(\lambda)]A^{-1}(\lambda)$, it is straightforward to derive $S^c_{nT}(\theta)$. However, the derivation of the ρ-component is complicated and some intermediate results are useful. First, $\frac{\partial}{\partial\rho}[\mathbf{B}'_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)] = -\mathbf{M}'\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho) - \mathbf{B}'_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{M} + \mathbf{B}'_{nT}(\rho)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho) + \mathbf{B}'_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)$, where $\dot{\mathbb{Q}}_{\mathbb{D}}(\rho) = \frac{\partial}{\partial\rho}\mathbb{Q}_{\mathbb{D}}(\rho)$. With $\frac{\partial}{\partial\rho}\mathbb{D}(\rho) = -\mathbf{M}[\mathbf{D}_\mu, \mathbf{D}^\star_\gamma] = -\mathbf{G}_{nT}(\rho)\mathbb{D}(\rho)$, we have

$$\dot{\mathbb{Q}}_{\mathbb{D}}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_{nT}(\rho)\mathbb{P}_{\mathbb{D}}(\rho) + \mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}'_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho). \tag{B.1}$$

This leads to $-\frac{\partial}{\partial\rho}[\mathbf{B}'_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)] = \mathbf{B}'_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho) \equiv \Psi(\rho)$, the ρ-component of the CQS function (3.10), and the ρ-component of the AQS function (3.12):

$$S^*_\rho(\theta) = \tfrac{1}{2\sigma^2}[\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\beta]'\Psi(\rho)[\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\beta] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_{nT}(\rho)]. \tag{B.2}$$

This is expressed in terms of $\Psi(\rho)$ and $\mathbf{G}^\circ_{nT}(\rho)$ to facilitate the derivations of the ρ-related terms of the Hessian matrix $\frac{\partial}{\partial\rho}\Psi(\rho)$. Again, the (ρ,ρ) term of $\frac{\partial}{\partial\rho}\Psi(\rho)$ is most complicate. For a comformable vector a, we have by taking use of (B.1) and after some tedious algebra,

$$a'[\tfrac{\partial}{\partial\rho}\Psi(\rho)]a = 2a'\mathbf{B}'_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)[\mathbf{G}^\circ_{nT}(\rho)\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_{nT}(\rho) - \mathbf{G}'_{nT}(\rho)\mathbf{G}_{nT}(\rho)]\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)a. \tag{B.3}$$

With the set of AQS functions $S^*_{nT}(\theta)$ given in (3.12) and (B.1)-(B.3), we obtain the components of the Hessian matrix $H^*_{nT}(\theta) = \frac{\partial}{\partial\theta'}S^*_{nT}(\theta)$:

$$
\begin{aligned}
H^*_{\beta\beta}(\theta) &= -\tfrac{1}{\sigma^2}\mathbb{Z}'(\rho)\mathbb{Z}(\rho),\\
H^*_{\beta\sigma^2}(\theta) &= -\tfrac{1}{\sigma^4}\mathbb{Z}'(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H^{*\prime}_{\sigma^2\beta},\\
H^*_{\beta\lambda}(\theta) &= -\tfrac{1}{\sigma^2}\mathbb{Z}'(\rho)\mathbb{Y}(\rho) \;=\; H^{*\prime}_{\lambda\beta},\\
H^*_{\beta\rho}(\theta) &= -\tfrac{1}{\sigma^2}\mathbb{Z}'(\rho)\mathbf{G}^\circ_{nT}(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H^{*\prime}_{\rho\beta},\\
H^*_{\sigma^2\sigma^2}(\theta) &= -\tfrac{1}{\sigma^6}\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) + \tfrac{1}{2\sigma^4}N_1,\\
H^*_{\sigma^2\lambda}(\theta) &= -\tfrac{1}{\sigma^4}\mathbb{Y}'(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H^{*\prime}_{\lambda\sigma^2},\\
H^*_{\sigma^2\rho}(\theta) &= -\tfrac{1}{2\sigma^4}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}^\circ_{nT}(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H^{*\prime}_{\rho\sigma^2},\\
H^*_{\lambda\lambda}(\theta) &= -\tfrac{1}{\sigma^2}\mathbb{Y}'(\rho)\mathbb{Y}(\rho) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)\mathbf{F}^2_{nT}(\lambda)\mathbf{B}^{-1}_{nT}(\rho)],\\
H^*_{\lambda\rho}(\theta) &= -\tfrac{1}{\sigma^2}\mathbb{Y}'(\rho)\mathbf{G}^\circ_{nT}(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{F}_{nT}(\lambda)\mathbb{R}_{nT}(\rho)],\\
H^*_{\rho\lambda}(\theta) &= -\tfrac{1}{\sigma^2}\mathbb{Y}'(\rho)\mathbf{G}^\circ_{nT}(\rho)\tilde{\mathbf{V}}(\beta,\delta),\\
H^*_{\rho\rho}(\theta) &= \tfrac{1}{\sigma^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathcal{R}_{1N}(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathcal{R}_{2N}(\rho)],
\end{aligned}
\tag{B.4}
$$

where $\mathbb{Y}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)\mathbf{WY}$, $\mathbb{R}_{nT}(\rho) = \mathbf{B}^{-1}_{nT}(\rho)\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_{nT}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)$, $\mathcal{R}_{1N}(\rho) = \mathbf{G}^\circ_{nT}(\rho)\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_{nT}(\rho) - \mathbf{G}'_{nT}(\rho)\mathbf{G}_{nT}(\rho)$ and $\mathcal{R}_{2N}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_{nT}(\rho)$ $[\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_{nT}(\rho) + \mathbf{G}_{nT}(\rho)]$.

Proof of Theorem 3.1: By theorem 5.9 of van der Vaart (1998), we only need to show $\sup_{\delta\in\delta}\frac{1}{N_1}\left\|S^{*c}_{nT}(\delta) - \bar{S}^{*c}_{nT}(\delta)\right\| \xrightarrow{p} 0$ under the assumptions in Theorem 3.1. From (3.14) and (3.16), the consistency of $\hat{\delta}^*_{nT}$ follows from:

(a) $\inf_{\delta\in\Delta}\bar{\sigma}^{*2}_{nT}(\delta)$ is bounded away from zero,

(b) $\sup_{\delta\in\Delta}\left|\hat{\sigma}^{*2}_{nT}(\delta) - \bar{\sigma}^{*2}_{nT}(\delta)\right| = o_p(1)$,

(c) $\sup_{\delta\in\Delta}\frac{1}{N_1}\left|\mathbf{Y}'\mathbf{W}'\mathbf{B}'_{nT}(\rho)\hat{\mathbf{V}}(\delta) - \mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}'_{nT}(\rho)\bar{\mathbf{V}}(\delta)]\right| = o_p(1)$,

(d) $\sup_{\delta\in\Delta}\frac{1}{N_1}\left|\hat{\mathbf{V}}'(\delta)\mathbf{G}_{nT}(\rho)\hat{\mathbf{V}}(\delta) - \mathrm{E}[\bar{\mathbf{V}}'(\delta)\mathbf{G}_{nT}(\rho)\bar{\mathbf{V}}(\delta)]\right| = o_p(1)$.

Proof of (a). From (3.15), $\bar{\beta}^*_{nT}(\delta) = [\mathbb{Z}'(\rho)\mathbb{Z}(\rho)]^{-1}\mathbb{Z}'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)\mathrm{E}(\mathbf{Y})$ as $\mathbb{Z}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)\mathbf{Z}$ and $\mathbb{Q}_{\mathbb{D}}(\rho)$ is idempotent. Thus, $\bar{\mathbf{V}}(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)\mathbf{Y} - \mathbb{Z}(\rho)\bar{\beta}^*_{nT}(\delta) = \mathbb{Q}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)\mathbf{Y} + \mathbb{P}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]$. By the orthogonality between $\mathbb{Q}_{\mathbb{D}}(\rho)$ and $\mathbb{P}_{\mathbb{D}}(\rho)$ and using $\mathbf{Y} = \mathbf{A}^{-1}_{nT}(\eta + \mathbf{B}^{-1}_{nT}\mathbf{V})$, we have,

$$
\begin{aligned}
\bar{\sigma}^{*2}_{nT}(\delta) &= \tfrac{1}{N_1}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\bar{\mathbf{V}}(\delta)] \\
&= \tfrac{1}{N_1}\mathrm{E}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}] + \tfrac{1}{N_1}\mathrm{E}\big\{[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]'\mathbf{P}(\delta)[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]\big\} \qquad \text{(B.5)} \\
&= \tfrac{1}{N_1}\mathrm{E}(\mathbf{Y})'\mathbf{Q}(\delta)\mathrm{E}(\mathbf{Y}) + \tfrac{1}{N_1}\mathrm{E}\big\{[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]'[\mathbf{Q}(\delta)+\mathbf{P}(\delta)][\mathbf{Y}-\mathrm{E}(\mathbf{Y})]\big\} \\
&= \tfrac{1}{N_1}\mathrm{E}(\mathbf{Y})'\mathbf{Q}(\delta)\mathrm{E}(\mathbf{Y}) + \tfrac{1}{N_1}\mathrm{E}\big\{[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]'\mathbf{C}'_{nT}(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)[\mathbf{Y}-\mathrm{E}(\mathbf{Y})]\big\} \\
&= \tfrac{1}{N_1}\eta'\mathbf{A}'^{-1}_{nT}\mathbf{Q}(\delta)\mathbf{A}^{-1}_{nT}\eta + \tfrac{\sigma_0^2}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_{nT}(\delta)],
\end{aligned}
$$

where $\mathbf{Q}(\delta) = \mathbf{C}'_{nT}(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)$ and $\mathbf{P}(\delta) = \mathbf{C}'_{nT}(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{P}_{\mathbb{Z}}(\rho)$ $\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)$. The first term can be written in the form of $a'(\delta)a(\delta)$ for an $N\times 1$ vector function of δ, and thus is non-negative, uniformly in $\delta\in\Delta$. For the second term,

$$
\tfrac{\sigma_0^2}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_{nT}(\delta)] \geq \tfrac{\sigma_0^2}{N_1}\gamma_{\min}[\mathcal{C}_{nT}(\delta)]\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)] = \sigma_0^2\gamma_{\min}[\mathcal{C}_{nT}(\delta)]
$$
$$
\geq \sigma_0^2\gamma_{\max}(\mathbf{A}'_{nT}\mathbf{A}_{nT})^{-1}\gamma_{\max}(\mathbf{B}'_{nT}\mathbf{B}_{nT})^{-1}\gamma_{\min}[\mathbf{A}'_{nT}(\lambda)\mathbf{A}_{nT}(\lambda)]\gamma_{\min}[\mathbf{B}'_{nT}(\rho)\mathbf{B}_{nT}(\rho)] > 0,
$$

uniformly in $\delta\in\Delta$, by Assumption E(iii). It follows that $\inf_{\delta\in\Delta}\bar{\sigma}^{*2}_{nT}(\delta) > 0$.

Proof of (b). From (3.13), $\hat{\beta}^*_{nT}(\delta) = [\mathbb{Z}'(\rho)\mathbb{Z}(\rho)]^{-1}\mathbb{Z}'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)\mathbf{Y}$. Then, $\hat{\mathbf{V}}(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)[\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\hat{\beta}^*_{nT}(\delta)] = \mathbb{Q}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)\mathbf{Y}$ and $\hat{\sigma}^{*2}_{nT}(\delta) = \frac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}$. From (B.5), $\bar{\sigma}^{*2}_{nT}(\delta) = \frac{1}{N_1}\mathrm{E}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}] + \frac{\sigma_0^2}{N_1}\mathtt{tr}[\mathbf{C}'^{-1}_{nT}\mathbf{P}(\delta)\mathbf{C}^{-1}_{nT}]$. Thus,

$$
\hat{\sigma}^{*2}_{nT}(\delta) - \bar{\sigma}^{*2}_{nT}(\delta) = \tfrac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})] - \tfrac{\sigma_0^2}{N_1}\mathtt{tr}[\mathbf{C}'^{-1}_{nT}\mathbf{P}(\delta)\mathbf{C}^{-1}_{nT}].
$$

For the second term, we have, $0 \leq \frac{1}{N_1}\mathtt{tr}[\mathbf{C}'^{-1}_{nT}\mathbf{P}(\delta)\mathbf{C}^{-1}_{nT}] \leq \frac{1}{N_1}\gamma_{\max}[\mathcal{C}_{nT}(\delta)]\gamma^2_{\max}[\mathbb{Q}_{\mathbb{D}}(\rho)]$ $\mathtt{tr}[\mathbb{P}_{\mathbb{Z}}(\rho)] = o(1)$, because $\mathtt{tr}[\mathbb{P}_{\mathbb{Z}}(\rho)] = k$, $\gamma_{\max}[\mathbb{Q}_{\mathbb{D}}(\rho)] = 1$ and, by Assumption E(iii), $\gamma_{\max}[\mathcal{C}_{nT}(\delta)] \leq \gamma_{\min}(\mathbf{A}'_{nT}\mathbf{A}_{nT})^{-1}\gamma_{\min}(\mathbf{B}'_{nT}\mathbf{B}_{nT})^{-1}\gamma_{\max}[\mathbf{A}'_{nT}(\lambda)\mathbf{A}_{nT}(\lambda)]$ $\gamma_{\max}[\mathbf{B}'_{nT}(\rho)\mathbf{B}_{nT}(\rho)] < \infty$. Therefore, one has $\sup_{\delta\in\Delta}|\frac{\sigma_0^2}{N_1}\mathtt{tr}[\mathbf{C}'^{-1}_{nT}\mathbf{P}(\delta)\mathbf{C}^{-1}_{nT}]| = o(1)$. For the first term, we prove the uniform convergence: $\sup_{\delta\in\Delta}|\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]| = o_p(1)$, which follows from pointwise convergence of $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]$ to zero for each $\delta\in\Delta$ and the stochastic equicontinuity of

$\frac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}$, according to Andrews (1992). We have,

$$
\begin{aligned}
&\tfrac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})] \\
&= \tfrac{1}{N_1}(\eta + \mathbf{B}_{nT}^{-1}\mathbf{V})'\mathbf{A}_{nT}'^{-1}\mathbf{Q}(\delta)\mathbf{A}_{nT}^{-1}(\eta + \mathbf{B}_{nT}^{-1}\mathbf{V}) \\
&\quad - \tfrac{1}{N_1}\mathrm{E}[(\eta + \mathbf{B}_{nT}^{-1}\mathbf{V})'\mathbf{A}_{nT}'^{-1}\mathbf{Q}(\delta)\mathbf{A}_{nT}^{-1}(\eta + \mathbf{B}_{nT}^{-1}\mathbf{V})] \\
&= \tfrac{2}{N_1}\mathbf{V}'\mathbf{C}_{nT}^{-1\prime}\mathbf{Q}(\delta)\mathbf{A}_{nT}^{-1}\eta + \tfrac{1}{N_1}[\mathbf{V}'\mathbf{C}_{nT}^{-1\prime}\mathbf{Q}(\delta)\mathbf{C}_{nT}^{-1}\mathbf{V} - \sigma_0^2\mathtt{tr}(\mathbf{C}_{nT}^{-1\prime}\mathbf{Q}(\delta)\mathbf{C}_{nT}^{-1})].
\end{aligned}
$$

By Assumption E, and Lemmas A.1 and A.3, one shows that $\mathbf{C}_{nT}^{-1\prime}\mathbf{Q}(\delta)\mathbf{A}_{nT}^{-1}$and $\mathbf{C}_{nT}^{-1\prime}\mathbf{Q}(\delta)\mathbf{C}_{nT}^{-1}$ are bounded in both row and column sum norms, for each $\delta \in \Delta$. Further, the elements of η are uniformly bounded. Thus, the pointwise convergence of the first term follows from Lemma A.6 (v), and the pointwise convergence of the second term follows from Lemma A.6 (iv). Therefore, $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})] \xrightarrow{p} 0$, for each $\delta \in \Delta$.

Next, let δ_1 and δ_2 be in Δ. We have by the mean value theorem (MVT):

$$\tfrac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta_2)\mathbf{Y} - \tfrac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta_1)\mathbf{Y} = \tfrac{1}{N_1}\mathbf{Y}'[\tfrac{\partial}{\partial\delta'}\mathbf{Q}(\bar{\delta})]\mathbf{Y}(\delta_2 - \delta_1),$$

where $\bar{\delta}$ lies between δ_1 and δ_2. It follows that $\frac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}$ is stochastically equicontinuous if $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\varpi}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$, $\varpi = \lambda, \rho$. We only show $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\rho}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$ as the proof of $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\lambda}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$ is similar and simpler. Note that

$$
\begin{aligned}
\tfrac{\partial}{\partial\rho}\mathbf{Q}(\delta) = &- \mathbf{C}_{nT}'(\delta)\mathbf{G}_{nT}'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta) \\
&+ \mathbf{C}_{nT}'(\delta)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta) \\
&+ \mathbf{C}_{nT}'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\dot{\mathbb{Q}}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta) + \mathbf{C}_{nT}'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{Z}}(\rho)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta) \\
&- \mathbf{C}_{nT}'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_{nT}(\rho)\mathbf{C}_{nT}(\delta),
\end{aligned}
$$

where $\dot{\mathbb{Q}}_{\mathbb{Z}}(\rho) = \frac{\partial}{\partial\rho}\mathbb{Q}_{\mathbb{Z}}(\rho)$. Using (B.1), we have after some algebra, $\dot{\mathbb{Z}}(\rho) = \frac{\partial}{\partial\rho}\mathbb{Z}(\rho) = \mathbb{G}_{nT}(\rho)\mathbb{Z}(\rho)$ where $\mathbb{G}_{nT}(\rho) = \mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}_{nT}'(\rho) - \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_{nT}(\rho)$, which gives

$$\dot{\mathbb{Q}}_{\mathbb{Z}}(\rho) = -\mathbb{P}_{\mathbb{Z}}(\rho)\mathbb{G}_{nT}'(\rho)\mathbb{Q}_{\mathbb{Z}}(\rho) - \mathbb{Q}_{\mathbb{Z}}(\rho)\mathbb{G}_{nT}(\rho)\mathbb{P}_{\mathbb{Z}}(\rho). \tag{B.6}$$

For a comformable vector a and taking use (B.1) and (B.6), we have after some algebra,

$$a'[\tfrac{\partial}{\partial\rho}\mathbf{Q}(\delta)]a = -2a'\bar{\mathbf{Q}}(\delta)a, \tag{B.7}$$

where $\bar{\mathbf{Q}}(\delta) = \mathcal{Q}_{nT}'(\delta)\mathbf{G}_{nT}(\rho)\mathcal{Q}_{nT}(\delta)$ and $\mathcal{Q}_{nT}(\delta) = \mathbb{Q}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)$. Rearrange $\bar{\mathbf{Q}}(\delta) = \mathcal{Q}'_{nT}(\delta)\mathbf{M}\bar{\mathbb{Q}}_{\mathbb{D}}(\rho)\bar{\mathbb{Q}}_{\mathbb{Z}}(\rho)\mathbf{A}_{nT}(\lambda)$, where $\bar{\mathbb{Q}}_{\mathbb{D}}(\rho) = I_{nT} - \mathbf{D}[\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)\mathbf{B}_{nT}(\rho)$ and $\bar{\mathbb{Q}}_{\mathbb{Z}}(\rho) = I_{nT} - \mathbf{Z}[\mathbb{Z}'(\rho)\mathbb{Z}(\rho)]^{-1}\mathbb{Z}'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)$. Following exactly the same way as we prove Lemma A.3, we show that $\bar{\mathbb{Q}}_{\mathbb{D}}(\rho)$ and $\bar{\mathbb{Q}}_{\mathbb{Z}}(\rho)$ are also uniformly bounded in both row and column sums, uniformly in $\rho \in \Delta_\rho$. This implies that both $\|\bar{\mathbf{Q}}(\delta)\|_1$ and $\|\bar{\mathbf{Q}}(\delta)\|_\infty$ are bounded uniformly in $\delta \in \Delta$. As $\mathbf{Y} = \mathbf{A}_{nT}^{-1}(\eta + \mathbf{B}_{nT}^{-1}\mathbf{V})$, Lemma A.1 and Lemma A.6 imply

$$\tfrac{1}{N_1}\mathbf{Y}'[\tfrac{\partial}{\partial\rho}\mathbf{Q}(\delta)]\mathbf{Y} = -\tfrac{2}{N_1}\mathbf{Y}'\bar{\mathbf{Q}}(\delta)\mathbf{Y} = -\tfrac{2}{N_1}(\eta + \mathbf{B}_{nT}^{-1}\mathbf{V})'\mathbf{A}_{nT}'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{A}_{nT}^{-1}(\eta + \mathbf{B}_{nT}^{-1}\mathbf{V})$$
$$= -\tfrac{2}{N_1}\eta'\mathbf{A}_{nT}'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{A}_{nT}^{-1}\eta - \tfrac{4}{N1}\eta'\mathbf{A}_{nT}'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{C}_{nT}^{-1}\mathbf{V} - \tfrac{2}{N_1}\mathbf{V}'\mathbf{C}_{nT}'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{C}_{nT}^{-1}\mathbf{V} = O_p(1),$$

uniformly in $\delta \in \Delta$. Thus, $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\rho}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$. Following the similar analysis, one also has $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\lambda}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$. Therefore, $\sup_{\delta\in\Delta}|\hat{\sigma}_{nT}^{*2}(\delta) - \bar{\sigma}_{nT}^{*2}(\delta)| = o_p(1)$.

Proof of (c). By the expressions of $\hat{\mathbf{V}}(\lambda)$ and $\bar{\mathbf{V}}(\delta)$ given above, we have

$$\tfrac{1}{N_1}\mathbf{Y}'\mathbf{W}'\mathbf{B}_{nT}'(\rho)\hat{\mathbf{V}}(\delta) - \tfrac{1}{N_1}\mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_{nT}'(\rho)\bar{\mathbf{V}}(\delta)]$$
$$= \tfrac{1}{N_1}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_{nT}'(\rho)\mathcal{Q}_{nT}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{W}'\mathbf{B}_{nT}'(\rho)\mathcal{Q}_{nT}(\delta)\mathbf{Y})]$$
$$- \tfrac{\sigma_0^2}{N_1}\mathtt{tr}[\mathbf{C}_{nT}'^{-1}\mathbf{W}'\mathbf{B}_{nT}'(\rho)\mathcal{P}_{nT}(\delta)\mathbf{C}_{nT}^{-1}],$$

where $\mathcal{P}_{nT}(\delta) = \mathbb{P}_{\mathbb{Z}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_{nT}(\delta)$. The first term is similar in form to $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]$ from **(b)**, and its uniform convergence is shown in a similar way. Furthermore, by Lemma A.4, it is easy to see that the second term is $o(1)$ uniformly in $\delta \in \Delta$.

Proof of (d). Again, using the expressions of $\bar{\mathbf{V}}(\delta)$ and $\hat{\mathbf{V}}(\delta)$, we have

$$\tfrac{1}{N_1}\hat{\mathbf{V}}'(\delta)\mathbf{G}_{nT}(\rho)\hat{\mathbf{V}}(\delta) - \tfrac{1}{N_1}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\mathbf{G}_{nT}(\rho)\bar{\mathbf{V}}(\delta)]$$
$$= \tfrac{1}{N_1}[\mathbf{Y}'\bar{\mathbf{Q}}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\bar{\mathbf{Q}}(\delta)\mathbf{Y})] - \tfrac{\sigma_0^2}{N_1}\mathtt{tr}[\mathbf{C}_{nT}'^{-1}\mathcal{P}_{nT}'(\delta)\mathbf{G}_{nT}^{\circ}(\rho)\mathcal{Q}_{nT}(\delta)\mathbf{C}_{nT}^{-1}]$$
$$- \tfrac{\sigma_0^2}{N_1}\mathtt{tr}[\mathbf{C}_{nT}'^{-1}\mathcal{P}_{nT}'(\delta)\mathbf{G}_{nT}(\rho)\mathcal{P}_{nT}(\delta)\mathbf{C}_{nT}^{-1}].$$

Therefore, the uniform convergence of the first term can also be shown similarly as we do for $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]$ since they have similar forms. By Lemma A.4, the remaining two terms are easily seen to be $o(1)$, uniformly in $\delta \in \Delta$. ■

Proof of Theorem 3.2: Applying the MVT to each element of $S_{nT}^{*}(\hat{\theta}_{nT}^{*})$, we have

$$0 = \frac{1}{\sqrt{N_1}} S^*_{nT}(\hat{\theta}^*_{nT}) = \frac{1}{\sqrt{N_1}} S^*_{nT}(\theta_0) + \left[\frac{1}{N_1} \frac{\partial}{\partial \theta'} S^*_{nT}(\theta) \Big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}} \right] \sqrt{N_1}(\hat{\theta}^*_{nT} - \theta_0), \tag{B.8}$$

where $\{\bar{\theta}_r\}$ are on the line segment between $\hat{\theta}^*_{nT}$ and θ_0. The result of the theorem follows if

(a) $\frac{1}{\sqrt{N_1}} S^*_{nT}(\theta_0) \xrightarrow{D} N[0, \lim_{N_1 \to \infty} \Gamma^*_{nT}(\theta_0)]$,

(b) $\frac{1}{N_1}[\frac{\partial}{\partial \theta'} S^*_{nT}(\theta)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}} - \frac{\partial}{\partial \theta'} S^*_{nT}(\theta_0)] = o_p(1)$, and

(c) $\frac{1}{N_1}[\frac{\partial}{\partial \theta'} S^*_{nT}(\theta_0) - \mathrm{E}(\frac{\partial}{\partial \theta'} S^*_{nT}(\theta_0))] = o_p(1)$.

Proof of (a). From (3.17), we see that the elements of $S^*_{nT}(\theta_0)$ are linear-quadratic forms in $\mathbf{V}$. Thus, for every non-zero $(k+3) \times 1$ vector of constants a, $a' S^*_{nT}(\theta_0)$ is of the form:

$$a' S^*_{nT}(\theta_0) = b'_{nT} \mathbf{V} + \mathbf{V}' \Phi_{nT} \mathbf{V} - \sigma^2 \mathtt{tr}(\Phi_{nT}),$$

for suitably defined non-stochastic vector b_{nT} and matrix Φ_{nT}. Based on Assumptions A-F, it is easy to verify (by Lemma A.1 and Lemma A.3(i)) that b_{nT} and matrix Φ_{nT} satisfy the conditions of the CLT for LQ form of Kelejian and Prucha (2001), and hence the asymptotic normality of $\frac{1}{\sqrt{N_1}} a' S^*_{nT}(\theta_0)$ follows. By Cramér-Wold device, $\frac{1}{\sqrt{N_1}} S^*_{nT}(\theta_0) \xrightarrow{D} N[0, \lim_{N_1 \to \infty} \Gamma^*_{nT}(\theta_0)]$, where elements of $\Gamma^*_{nT}(\theta_0)$ are given in (3.19).

Proof of (b). The Hessian matrix $H^*_{nT}(\theta) = \frac{\partial}{\partial \theta'} S^*_{nT}(\theta)$ is given in (B.4). By Assumptions D and E, and Lemma A.1 and Lemma A.3(i), $\mathbb{R}_{nT}(\rho_0)$, $\mathcal{R}_{1N}(\rho_0)$ and $\mathcal{R}_{2N}(\rho_0)$ are all bounded in row and column sum norms. With these and $\mathbf{Y} = \mathbf{A}^{-1}_{nT}(\eta + \mathbf{B}^{-1}_{nT} \mathbf{V})$, Lemma A.6 leads to $\frac{1}{N_1} H^*_{nT}(\theta_0) = O_p(1)$. Thus, $\frac{1}{N_1} H^*_{nT}(\bar{\theta}) = O_p(1)$ since $\bar{\theta} \xrightarrow{p} \theta_0$ due to $\hat{\theta}^*_{nT} \xrightarrow{p} \theta_0$, where for ease of exposition, $H^*_{nT}(\bar{\theta})$ is used to denote $\frac{\partial}{\partial \theta'} S^*_{nT}(\theta)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}$. As $\bar{\sigma}^2 \xrightarrow{p} \sigma^2_0$, we have $\bar{\sigma}^{-r} = \sigma_0^{-r} + o_p(1)$, for $r = 2, 4, 6$. As σ^{-r} appears in $H^*_{nT}(\theta)$ multiplicatively, $\frac{1}{N_1} H^*_{nT}(\bar{\theta}) = \frac{1}{N_1} H^*_{nT}(\bar{\beta}, \bar{\lambda}, \bar{\rho}, \sigma^2_0) + o_p(1)$. Thus, the proof of **(b)** is equivalent to the proof of

$$\frac{1}{N_1}[H^*_{nT}(\bar{\beta}, \bar{\lambda}, \bar{\rho}, \sigma^2_0) - H^*_{nT}(\theta_0)] \xrightarrow{p} 0,$$

or the proofs of $\frac{1}{N_1}[H^{*\mathtt{S}}_{nT}(\bar{\beta}, \bar{\lambda}, \bar{\rho}, \sigma^2_0) - H^{*\mathtt{S}}_{nT}(\theta_0)] \xrightarrow{p} 0$ and $\frac{1}{N_1}[H^{*\mathtt{NS}}_{nT}(\bar{\delta}) - H^{*\mathtt{NS}}_{nT}(\delta_0)] \xrightarrow{p} 0$, where $H^{*\mathtt{S}}_{nT}$ and $H^{*\mathtt{NS}}_{nT}$ denote, respectively, the stochastic and non-stochastic parts of H^*_{nT}.

For the stochastic part, we see from (B.4) that all the components of $H^{*\mathtt{S}}_{nT}(\beta,\lambda,\rho,\sigma_0^2)$ are linear, bilinear or quadratic in β and λ, but nonlinear in ρ. Hence, with an application of the MVT on $H^{*\mathtt{S}}_{nT}(\bar\beta,\bar\lambda,\bar\rho,\sigma_0^2)$ w.r.t $\bar\rho$ 'variable', we can write $\frac{1}{N_1}[H^{*\mathtt{S}}_{nT}(\bar\beta,\bar\lambda,\bar\rho,\sigma_0^2)-H^{*\mathtt{S}}_{nT}(\theta_0)]$ as

$$\tfrac{1}{N_1}[\tfrac{\partial}{\partial\rho}H^{*\mathtt{S}}_{nT}(\bar\beta,\bar\lambda,\dot\rho,\sigma_0^2)](\bar\rho-\rho_0)+\tfrac{1}{N_1}[H^{*\mathtt{S}}_{nT}(\bar\beta,\bar\lambda,\rho_0,\sigma_0^2)-H^{*\mathtt{S}}_{nT}(\theta_0)],$$

where $\dot\rho$ lies between $\bar\rho$ and ρ_0. Therefore, we only need to show that (i) $\frac{1}{N_1}\frac{\partial}{\partial\rho}H^{*\mathtt{S}}_{nT}(\bar\beta,\bar\lambda,\dot\rho,\sigma_0^2)=O_p(1)$, and (ii) $\frac{1}{N_1}[H^{*\mathtt{S}}_{nT}(\bar\beta,\bar\lambda,\rho_0,\sigma_0^2)-H^{*\mathtt{S}}_{nT}(\theta_0)]=o_p(1)$.

We select one of the most complicated components, $H^{*\mathtt{S}}_{\rho\lambda}(\theta)=-\frac{1}{\sigma^2}\mathbb{Y}'(\rho)\mathbf{G}^\circ_{nT}(\rho)\tilde{\mathbf{V}}(\beta,\delta)$, to illustrate the general idea in the proof. We have, after some algebra,

$$\tfrac{1}{N_1}\tfrac{\partial}{\partial\rho}H^{*\mathtt{S}}_{\rho\lambda}(\bar\beta,\bar\lambda,\dot\rho,\sigma_0^2)=\tfrac{2}{N_1\sigma_0^2}\mathbb{Y}'(\dot\rho)\mathcal{R}_{1N}(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_{nT}(\dot\rho)(\mathbf{A}_{nT}(\bar\lambda)\mathbf{Y}-\mathbf{Z}\bar\beta),$$

$$\tfrac{1}{N_1}[H^{*\mathtt{S}}_{nT}(\bar\beta,\bar\lambda,\rho_0,\sigma_0^2)-H^{*\mathtt{S}}_{nT}(\theta_0)]=\tfrac{1}{N_1\sigma_0^2}\mathbb{Y}'\mathbf{G}^\circ_{nT}\mathbb{Y}(\bar\lambda-\lambda_0)+\tfrac{1}{N_1\sigma_0^2}\mathbb{Y}'\mathbf{G}^\circ_{nT}\mathbb{Z}(\bar\beta-\beta_0).$$

By Lemmas A.1 and A.6, it is easy to show that $\frac{1}{N_1}\mathbb{Y}'\mathbf{G}^\circ_{nT}\mathbb{Y}=O_p(1)$ and $\frac{1}{N_1}\mathbb{Y}'\mathbf{G}^\circ_{nT}\mathbb{Z}=O_p(1)$. Therefore, (ii) holds. To prove (i), we have

$$\begin{aligned}&\mathbb{Y}'(\dot\rho)\mathcal{R}_{1N}(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_{nT}(\dot\rho)(\mathbf{A}_{nT}(\bar\lambda)\mathbf{Y}-\mathbf{Z}\bar\beta)\\=&(\mathbf{A}^{-1}_{nT}\eta+\mathbf{C}^{-1}_{nT}\mathbf{V})'\mathcal{H}_{nT}(\dot\rho)[\mathbf{A}_{nT}(\bar\lambda)\mathbf{A}^{-1}_{nT}\eta+\mathbf{A}_{nT}(\bar\lambda)\mathbf{C}^{-1}_{nT}\mathbf{V}-\mathbf{Z}\bar\beta]\end{aligned}$$

where $\mathcal{H}_{nT}(\dot\rho)=\mathbf{W}'\mathbf{B}'_{nT}(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathcal{R}_{1N}(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_{nT}(\dot\rho)$. Lemma A.2 implies $\mathbf{B}^{-1}_{nT}(\dot\rho)$ embedded in $\mathcal{H}_{nT}(\dot\rho)$ is uniformly bounded in both row and column sums since $\dot\rho-\rho_0=o_p(1)$. Therefore, it is easy to see the above equation is $O_p(N)$ by Lemma A.6 and then result (i) follows.

For the non-stochastic part, we illustrate the proof using the most complicate $\lambda\lambda$-term. Noting that the non-stochastic part is nonlinear in both $\bar\lambda$ and $\bar\rho$, we have by the MVT,

$$\begin{aligned}&\tfrac{1}{N_1}[H^{*\mathtt{NS}}_{\lambda\lambda}(\bar\delta)-H^{*\mathtt{NS}}_{\lambda\lambda}(\delta_0)]=-\tfrac{1}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\bar\rho)\mathbf{B}_{nT}(\bar\rho)\mathbf{F}^2_{nT}(\bar\lambda)\mathbf{B}^{-1}_{nT}(\bar\rho)-\mathbb{Q}_{\mathbb{D}}\mathbf{B}_{nT}\mathbf{F}^2_{nT}\mathbf{B}^{-1}_{nT}]\\=&-(\bar\lambda-\lambda_0)\tfrac{1}{N_1}\mathtt{tr}[2\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_{nT}(\dot\rho)\mathbf{F}^3_{nT}(\dot\lambda)\mathbf{B}^{-1}_{nT}(\dot\rho)]-(\bar\rho-\rho_0)\tfrac{1}{N_1}\mathtt{tr}[\mathbf{F}^2_{nT}(\dot\lambda)\mathbb{R}_{nT}(\dot\rho)],\end{aligned}$$

where $\dot\lambda$ lies between $\bar\lambda$ and λ_0 and $\dot\rho$ lies between $\bar\rho$ and ρ_0. Again, by Lemma A.2, we conclude that both $\mathbf{A}^{-1}_{nT}(\dot\lambda)$ and $\mathbf{B}^{-1}_{nT}(\dot\rho)$ are uniformly bounded in both row and column sums. Therefore, the terms inside the trace both have elements that are uniformly bounded. As $\bar\delta-\delta_0=o_p(1)$, we have $\frac{1}{N_1}[H^{*\mathtt{NS}}_{\lambda\lambda}(\bar\delta)-$

$H^{*\texttt{NS}}_{\lambda\lambda}(\delta_0)] = o_p(1)$.

Proof of (c). Since $\mathbf{Y} = \mathbf{A}_{nT}^{-1}(\eta + \mathbf{B}_{nT}^{-1}\mathbf{V})$, the Hessian matrix at true θ_0 are seen to be linear combinations of terms linear or quadratic in $\mathbf{V}$, and constants. The constant terms are canceled out. Other terms are shown to be $o_p(1)$ based on Lemma A.6. For example,

$\frac{1}{N_1}[H^*_{\rho\rho}(\rho_0) - \mathrm{E}(H^*_{\rho\rho}(\rho_0))] = \frac{1}{N_1\sigma_0^2}[\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathrm{E}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V})] = o_p(1)$. ■

Proof of Corollary 3.1:

Note that $\Gamma^*_{nT}(\hat{\theta}^*_{nT}) = \Gamma^*_{nT}(\theta)|_{(\theta=\hat{\theta}^*_{nT},\phi=\hat{\phi}^*_{nT},\kappa_3=\hat{\kappa}_{3,nT},\kappa_4=\hat{\kappa}_{4,nT})}$. As $\hat{\theta}^*_{nT}$, $\hat{\kappa}_{3,nT}$ and $\hat{\kappa}_{4,nT}$ are consistent estimators for θ_0, κ_3 and κ_4, plugging these estimators into $\Gamma^*_{nT}(\theta)$ will not bring additional bias to the estimation of $\Gamma^*_{nT}(\theta_0)$. However, due to incidental parameters problem, the $\hat{\mu}^*_{nT}$ component of $\hat{\phi}^*_{nT}$ is not consistent for the estimation of μ_0 when T is fixed. The estimation bias caused by replacing ϕ_{nT} by $\hat{\phi}^*_{nT}$ can be derived as follow. Recall (3.8),

$$\hat{\phi}_{nT}(\beta,\delta) = [\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)\mathbf{B}_{nT}(\rho)[\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\beta].$$

Thus, the unconstrained estimate of ϕ_0 is just $\hat{\phi}^*_{nT} = \hat{\phi}_{nT}(\hat{\beta}^*_{nT}, \hat{\delta}^*_{nT})$. Note $\mathbf{A}_{nT}(\hat{\lambda}^*_{nT})\mathbf{Y} - \mathbf{Z}\hat{\beta}^*_{nT} = \mathbf{A}_{nT}\mathbf{Y} - \mathbf{Z}\beta_0 - \mathbf{W}\mathbf{Y}(\hat{\lambda}^*_{nT} - \lambda_0) - \mathbf{Z}(\hat{\beta}^*_{nT} - \beta_0)$. Applying the MVT on each row of $\mathbf{D}\hat{\phi}^*_{nT}$ with respect to the $\hat{\rho}^*_{nT}$-element, we have,

$$\begin{aligned}
\mathbf{D}\hat{\phi}^*_{nT} &= \mathbf{D}[\mathbb{D}'(\hat{\rho}^*_{nT})\mathbb{D}(\hat{\rho}^*_{nT})]^{-1}\mathbb{D}'(\hat{\rho}^*_{nT})\mathbf{B}_{nT}(\hat{\rho}^*_{nT})[\mathbf{A}_{nT}(\hat{\lambda}^*_{nT})\mathbf{Y} - \mathbf{Z}\hat{\beta}^*_{nT}] \\
&= \mathbf{B}_{nT}^{-1}(\hat{\rho}^*_{nT})\mathbb{P}_{\mathbb{D}}(\hat{\rho}^*_{nT})\mathbf{B}_{nT}(\hat{\rho}^*_{nT})[\mathbf{A}_{nT}(\hat{\lambda}^*_{nT})\mathbf{Y} - \mathbf{Z}\hat{\beta}^*_{nT}] \\
&= [\mathbf{B}_{nT}^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{B}_{nT} - \mathbb{R}_{nT}(\bar{\rho})(\hat{\rho}^*_{nT} - \rho_0)][\mathbf{A}_{nT}(\hat{\lambda}^*_{nT})\mathbf{Y} - \mathbf{Z}\hat{\beta}^*_{nT}] \\
&= \mathbf{D}\phi_0 + \mathbf{B}_{nT}^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{V} - \mathbf{B}_{nT}^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{B}_{nT}[\mathbf{W}\mathbf{Y}(\hat{\lambda}^*_{nT} - \lambda_0) + \mathbf{Z}(\hat{\beta}^*_{nT} - \beta_0)] \\
&\quad - \mathbb{R}_{nT}(\bar{\rho})[\mathbf{A}_{nT}(\hat{\lambda}^*_{nT})\mathbf{Y} - \mathbf{Z}\hat{\beta}^*_{nT}](\hat{\rho}^*_{nT} - \rho_0),
\end{aligned} \tag{B.9}$$

where $\bar{\rho}$ lies between $\hat{\rho}^*_{nT}$ and ρ_0 and changes over the rows of $\mathbb{R}_{nT}(\bar{\rho})$, and $\mathbb{R}_{nT}(\rho)$ is given below (B.4). From its expression, $\Gamma^*_{nT}(\theta)$ is seen to have components that are either linear or quadratic in $\mathbf{D}\phi$. Let d_{nT} be a non-stochastic nT-vector with elements being of uniform order $O(1)$ or $O(h_n^{-1})$. Using (B.9), the terms of $\Gamma^*_{nT}(\hat\theta^*_{nT})$ linear in $\mathbf{D}\hat\phi^*_{nT}$ can be represented as

$$
\begin{aligned}
&\tfrac{1}{N_1}d'_{nT}\mathbf{D}\hat\phi^*_{nT}\\
&= \tfrac{1}{N_1}d'_{nT}\mathbf{D}\phi_0 + \tfrac{1}{N_1}d'_{nT}\mathbf{B}^{-1}_{nT}\mathbb{P}_{\mathbb{D}}\mathbf{V} - \tfrac{1}{N_1}d'_{nT}\mathbf{B}^{-1}_{nT}\mathbb{P}_{\mathbb{D}}\mathbf{B}_{nT}[\mathbf{W}\mathbf{Y}(\hat\lambda^*_{nT}-\lambda_0)+\mathbf{Z}(\hat\beta^*_{nT}-\beta_0)]\\
&\quad + \tfrac{1}{N_1}d'_{nT}\mathbb{R}_{nT}(\bar\rho)[\mathbf{A}_{nT}(\hat\lambda^*_{nT})\mathbf{Y}-\mathbf{Z}\hat\beta^*_{nT}](\hat\rho^*_{nT}-\rho_0) = \tfrac{1}{N_1}d'_{nT}\mathbf{D}\phi_0 + o_p(1),
\end{aligned}
$$

where the last equation holds because of the consistency of $\hat\theta^*_{nT}$ and Lemma A.6, using $\mathbf{Y} = \mathbf{A}^{-1}_{nT}(\eta + \mathbf{B}^{-1}_{nT}\mathbf{V})$. Hence, we can conclude that the terms of $\Gamma^*_{nT}(\theta_0)$ linear in ϕ_0 can be consistently estimated by simply replacing ϕ_0 with $\hat\phi^*_{nT}$.

The only term that is quadratic in ϕ_0 is contained in $\Gamma^*_{\lambda\lambda}(\theta_0)$, which is $\frac{1}{N_1\sigma_0^2}\phi_0'\mathbb{D}'\mathcal{P}_2'\mathcal{P}_2\mathbb{D}\phi_0$. Its plug-in estimator is $\frac{1}{N_1\hat\sigma^{*2}_{nT}}\hat\phi^{*\prime}_{nT}\mathbb{D}'(\hat\rho^*_{nT})\mathcal{P}_2'(\hat\delta^*_{nT})\mathcal{P}_2(\hat\delta^*_{nT})\mathbb{D}(\hat\rho^*_{nT})\hat\phi^*_{nT}$. Using (B.9), $\hat\theta^*_{nT}-\theta_0 = o_p(1)$ and Lemma A.6, we show that this estimator is biased/inconsistent:

$$
\begin{aligned}
&\tfrac{1}{N_1\hat\sigma^{*2}_{nT}}\hat\phi^{*\prime}_{nT}\mathbb{D}'(\hat\rho^*_{nT})\mathcal{P}_2'(\hat\delta^*_{nT})\mathcal{P}_2(\hat\delta^*_{nT})\mathbb{D}(\hat\rho^*_{nT})\hat\phi^*_{nT}\\
&= \tfrac{1}{N_1\hat\sigma^{*2}_{nT}}\phi_0'\mathbb{D}'(\hat\rho^*_{nT})\mathcal{P}_2'(\hat\delta^*_{nT})\mathcal{P}_2(\hat\delta^*_{nT})\mathbb{D}(\hat\rho^*_{nT})\phi_0\\
&\quad + \tfrac{1}{N_1\hat\sigma^{*2}_{nT}}\mathbf{V}'\mathbb{P}_{\mathbb{D}}\mathbf{B}^{-1\prime}_{nT}\mathbf{B}'_{nT}(\hat\rho^*_{nT})\mathcal{P}_2'(\hat\delta^*_{nT})\mathcal{P}_2(\hat\delta^*_{nT})\mathbf{B}_{nT}(\hat\rho^*_{nT})\mathbf{B}^{-1}_{nT}\mathbb{P}_{\mathbb{D}}\mathbf{V} + o_p(1)\\
&= \tfrac{1}{N_1\sigma_0^2}\phi_0'\mathbb{D}'\mathcal{P}_2'\mathcal{P}_2\mathbb{D}\phi_0 + \tfrac{1}{N_1\sigma_0^2}\mathbf{V}'\mathbb{P}_{\mathbb{D}}\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}}\mathbf{V} + o_p(1)\\
&= \tfrac{1}{N_1\sigma_0^2}\phi_0'\mathbb{D}'\mathcal{P}_2'\mathcal{P}_2\mathbb{D}\phi_0 + \tfrac{1}{N_1}\mathtt{tr}[\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}}] + o_p(1).
\end{aligned}
$$

We see that the bias term, $\frac{1}{N_1}\mathtt{tr}[\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}}]$, involves only the common parameters that can be consistently estimated. Thus, a bias correction can easily be made. Define

$$
\text{Bias}^*_{\lambda\lambda}(\delta) = \tfrac{1}{N_1}\mathtt{tr}[\mathcal{P}_2'(\delta)\mathcal{P}_2(\delta)\mathbb{P}_{\mathbb{D}}(\rho)]. \tag{B.10}
$$

This gives the bias matrix of $\Gamma^*_{nT}(\hat\theta^*_{nT})$, which is a matrix of the same dimension as $\Gamma^*_{nT}(\hat\theta^*_{nT})$, and has the sole non-zero element $\text{Bias}^*_{\lambda\lambda}(\delta_0)$ corresponding to the $\Gamma^*_{\lambda\lambda}(\hat\theta^*_{nT})$ component. ■

Proof of Corollary 3.2.

Proof of (i). Note: $\hat{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}(\hat\rho^*_{nT})\mathbf{B}_{nT}(\hat\rho^*_{nT})[\mathbf{A}_{nT}(\hat\lambda^*_{nT})\mathbf{Y}-\mathbf{Z}\hat\beta^*_{nT}]$, $\mathbf{V} = \mathbf{B}_{nT}(\mathbf{A}_{nT}\mathbf{Y}-\eta)$, $\tilde{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$ and with respective elements $\{\hat v_j\}$, $\{v_j\}$ and $\{\tilde v_j\}$, and $\mathbb{Q}_{\mathbb{D}}$ has elements $\{q_{jh}\}$, $j,h = 1,\ldots,N$, where j and h are the combined indices for $i = 1, \dots, s_g$, $g = 1, \dots, G$ and $t = 1, \dots, T$.

Consistency of $\hat{\kappa}_{3,nT}$. As $\hat{\sigma}^*_{nT} - \sigma_0 = o_p(1)$ and $\hat{\rho}^*_{nT} - \rho_0 = o_p(1)$, the denominators of $\hat{\kappa}_{3,nT}$ and κ_3 agree asymptotically. Thus, $\hat{\kappa}_{3,nT}$ is consistent if $\frac{1}{nT}\sum_{j=1}^N [\hat{v}_j^3 - \mathrm{E}(\tilde{v}_j^3)] \xrightarrow{p} 0$, or

$$(a)\ \frac{1}{nT}\sum_{j=1}^N [\tilde{v}_j^3 - \mathrm{E}(\tilde{v}_j^3)] \xrightarrow{p} 0, \text{ and } \quad (b)\ \frac{1}{nT}\sum_{j=1}^N (\hat{v}_j^3 - \tilde{v}_j^3) \xrightarrow{p} 0.$$

To prove (a), note that $\tilde{v}_j = \sum_{h=1}^N q_{jh} v_h$. Thus, we have,

$$\begin{aligned}
&\frac{1}{nT}\sum_{j=1}^N [\tilde{v}_j^3 - \mathrm{E}(\tilde{v}_j^3)] \\
&= \frac{1}{nT}\sum_{j=1}^N \sum_{h=1}^N q_{jh}^3 [v_h^3 - \mathrm{E}(v_h^3)] + \frac{3}{N}\sum_{j=1}^N \sum_{l=1}^N \sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2 q_{jm} v_l^2 v_m \\
&\quad + \frac{6}{N}\sum_{j=1}^N \sum_{m=1}^N \sum_{\substack{l\neq m\\ l=1}}^N \sum_{\substack{h\neq m,l\\ h=1}}^N q_{jm} q_{jl} q_{jh} v_m v_l v_h \equiv K_1 + K_2 + K_3.
\end{aligned}$$

First, consider K_1 term. By Lemma A.3, $\mathbb{Q}_{\mathbb{D}}$ is uniformly bounded in both row and column sums. This implies that the elements of $\mathbb{Q}_{\mathbb{D}}$ are uniformly bounded. Therefore, there exists a constant $\bar{q}$ such that $|q_{jh}| \leq \bar{q}$ for all j and h. Given these, we have $\sum_{j=1}^N q_{jh}^3 \leq \sum_{j=1}^N |q_{jh}|^3 \leq \bar{q}^2 \sum_{j=1}^N |q_{jh}| < \infty$. Also note $\{v_i\}$ are iid by Assumption A. Thus, Khinchine's weak law of large number (WLLN) (Feller, 1968, pp. 243-244) implies that K_1 converges to zero in probability as sample size increases.

For the other two terms, we have by switching the order of summations when needed,

$$\begin{aligned}
K_2 &= \frac{3}{N}\sum_{j=1}^N \sum_{l=1}^N \sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2 q_{jm}(v_l^2 - \sigma^2) v_m + \frac{3}{N}\sum_{j=1}^N \sum_{l=1}^N \sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2 q_{jm}\sigma^2 v_m, \\
&= \frac{3}{N}\sum_{m=1}^N (v_m^2 - \sigma^2)(\sum_{j=1}^N \sum_{l=1}^{m-1} q_{jm}^2 q_{jl} v_l) \\
&\quad + \frac{3}{N}\sum_{m=1}^N v_m[\sum_{j=1}^N \sum_{l=1}^{m-1} q_{jl}^2 q_{jm}(v_l^2 - \sigma^2)], \\
&\quad + \frac{3}{N}\sum_{m=1}^N \sum_{j=1}^N \sum_{\substack{l\neq m\\ l=1}}^N q_{jl}^2 q_{jm}\sigma^2 v_m, \\
K_3 &= \frac{18}{N}\sum_{m=1}^N v_m(\sum_{j=1}^N \sum_{l=1}^{m-1} \sum_{\substack{h\neq l\\ h=1}}^{m-1} q_{jm} q_{jl} q_{jh} v_l v_h) \equiv \frac{1}{nT}\sum_{m=1}^N g_{4,m}.
\end{aligned}$$

Therefore, we have $K_2 = \frac{1}{nT}\sum_{m=1}^N (g_{1,m} + g_{2,m} + g_{3,m})$ and $K_3 = \frac{1}{nT}\sum_{m=1}^N g_{4,m}$, where

$$
\begin{aligned}
g_{1,m} &= 3(v_m^2 - \sigma^2) \textstyle\sum_{j=1}^{N} \sum_{l=1}^{m-1} q_{jm}^2 q_{jl} v_l, \\
g_{2,m} &= 3 v_m \textstyle\sum_{j=1}^{N} \sum_{l=1}^{m-1} q_{jl}^2 q_{jm} (v_l^2 - \sigma^2), \\
g_{3,m} &= 3 \textstyle\sum_{j=1}^{N} \sum_{\substack{l \neq m \\ l=1}}^{N} q_{jl}^2 q_{jm} \sigma^2 v_m, \\
g_{4,m} &= v_m \textstyle\sum_{j=1}^{N} \sum_{l=1}^{m-1} \sum_{\substack{h \neq l \\ h=1}}^{m-1} q_{jm} q_{jl} q_{jh} v_l v_h.
\end{aligned}
$$

Let $\{\mathcal{G}_m\}$ be the increasing sequence of σ-fields generated by $(v_1, \cdots, v_j, j = 1, \cdots, m)$, $m = 1, \cdots, N$. Then, $\mathrm{E}[(g_{1,m}, g_{2,m}, g_{3,m}, g_{4,m}) | \mathcal{G}_{m-1}] = 0$; hence, $\{(g_{1,m}, g_{2,m}, g_{3,m}, g_{4,m})', \mathcal{G}_m\}$ form a vector martingale difference (M.D.) sequence. As $\mathbb{Q}_{\mathbb{D}}$ is bounded in row and column sum norms, by Assumption A, it is easy to see that $\mathrm{E}|g_{s,m}|^{1+\epsilon} < \infty$, for $s = 1, 2, 3, 4$ and $\epsilon > 0$. Hence, $\{g_{1,m}\}$, $\{g_{2,m}\}$, $\{g_{3,m}\}$ and $\{g_{4,m}\}$ are uniformly integrable, and the WLLN of Davidson (1994, Theorem 19.7) applies to give $K_2 \xrightarrow{p} 0$ and $K_3 \xrightarrow{p} 0$.

To prove (b), using the notation $\tilde{\mathbf{V}}(\xi) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho)[\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\beta]$ in (3.9) where $\xi = (\beta', \delta')'$, we have $\tilde{\mathbf{V}} = \tilde{\mathbf{V}}(\xi_0)$ and $\hat{\mathbf{V}} = \tilde{\mathbf{V}}(\hat{\xi}_{nT}^*)$. Let $\mathbf{S}(\xi) = \frac{\partial}{\partial \xi'} \tilde{\mathbf{V}}(\xi)$, we have

$$
\mathbf{S}(\xi) = \{-\mathbb{Z}(\rho), \;\; -\mathbb{Y}(\rho), \;\; [\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbf{B}_{nT}(\rho) - \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{M}][\mathbf{A}_{nT}(\lambda)\mathbf{Y} - \mathbf{Z}\beta]\},
$$

where expressions of $\mathbb{Y}(\rho)$ and $\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)$ are in (B.1) and (B.4), respectively. Let $s_j'(\xi)$ be the jth row of $\mathbf{S}(\xi)$. We have by the MVT, for each $j = 1, 2, \ldots, N$,

$$
\hat{v}_j \equiv \tilde{v}_j(\hat{\xi}_{nT}^*) = \tilde{v}_j(\xi_0) + s_j'(\bar{\xi})(\hat{\xi}_{nT}^* - \xi_0) = \tilde{v}_j + \psi_j'(\hat{\xi}_{nT}^* - \xi_0) + o_p(\|\hat{\xi}_{nT}^* - \xi_0\|), \tag{B.11}
$$

where $\bar{\xi}$ lies between $\hat{\xi}_{nT}^*$ and ξ_0, and $\psi_j' = \mathrm{plim}_{N_1 \to \infty} s_j'(\bar{\xi})$, which is easily shown to be $O_p(1)$ as follow. Consider the first k (the number of regressors) elements of ψ_j' first. They are the limits of the jth row of $-\mathbb{Z}(\bar{\rho})$, which are just the jth row of $-\mathbb{Z}$ because $\bar{\rho} \xrightarrow{p} \rho_0$, implied by $\hat{\rho}_{nT}^* - \rho_0 = o_p(1)$. Hence, we conclude that the first k elements of ψ_j' are $O(1)$, for each $j = 1, 2, \ldots, N$. For the remaining two elements in each ψ_j', they are the limits of elements from the last two columns of $\mathbf{S}(\bar{\xi})$. It is easy to see the limits of the last two columns of $\mathbf{S}(\bar{\xi})$ are just $-\mathbb{Y}$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_{nT} - \mathbb{Q}_{\mathbb{D}}\mathbf{M}][\mathbf{A}_{nT}\mathbf{Y} - \mathbf{Z}\beta_0]$. Using $\mathbf{Y} = \mathbf{A}_{nT}^{-1}\eta + \mathbf{C}_{nT}^{-1}\mathbf{V}$, we have $-\mathbb{Y} = \mathcal{P}_2\mathbf{B}_{nT}\eta + \mathcal{P}_2\mathbf{V}$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_{nT} - \mathbb{Q}_{\mathbb{D}}\mathbf{M}][\mathbf{A}_{nT}\mathbf{Y} - \mathbf{Z}\beta_0] = [\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_{nT} - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{D}\phi_0 + [\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_{nT} - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{B}_{nT}^{-1}\mathbf{V}$. By

Lemma A.1, we have the elements of $\mathcal{P}_2\mathbf{B}_{nT}\eta$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_{nT} - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{D}\phi_0$ are uniformly bounded, and $\mathcal{P}_2$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_{nT} - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{B}_{nT}^{-1}$ are uniformly bounded in both row and column sum norms. Hence, it is easy to see each element of $-\mathbb{Y}$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_{nT} - \mathbb{Q}_{\mathbb{D}}\mathbf{M}][\mathbf{A}_{nT}\mathbf{Y} - \mathbf{Z}\beta_0]$ are $O_p(1)$, i.e., the last two elements in ψ_j' are also $O_p(1)$, for each $j = 1, 2, \ldots, N$.

As $\tilde{v}_j = O_p(1)$, $\psi_j' = O_p(1)$ and $\hat{\xi}_{nT}^* - \xi_0 = O_p(\frac{1}{\sqrt{N_1}})$, we have by (B.11), $\hat{v}_j^3 = \tilde{v}_j^3 + 3\tilde{v}_j^2\psi_j'(\hat{\xi}_{nT}^* - \xi_0) + o_p(\|\hat{\xi}_{nT}^* - \xi_0\|)$. It follows that

$$\begin{aligned}\tfrac{1}{nT}\textstyle\sum_{j=1}^N(\hat{v}_j^3 - \tilde{v}_j^3) &= \tfrac{3}{N}\textstyle\sum_{j=1}^N \tilde{v}_j^2\psi_j'(\hat{\xi}_{nT}^* - \xi_0) + o_p(\|\hat{\xi}_{nT}^* - \xi_0\|)\\ &= \tfrac{3\sigma^2}{N}\textstyle\sum_{j=1}^N(\sum_{k=1}^N q_{jk}^2\psi_j')(\hat{\xi}_{nT}^* - \xi_0) + o_p(\|\hat{\xi}_{nT}^* - \xi_0\|) = o_p(1),\end{aligned}$$

as $\frac{1}{nT}\sum_{j=1}^N(\sum_{k=1}^N q_{jk}^2\psi_j') = (\sum_{k=1}^N q_{jk}^2)\frac{1}{nT}(\sum_{j=1}^N \psi_j') = O(1)$.

Consistency of $\hat{\kappa}_{4,nT}$. As $\hat{\sigma}_{nT}^* - \sigma_0 = o_p(1)$ and $\hat{\rho}_{nT}^* - \rho_0 = o_p(1)$, the result follows if $\frac{1}{nT}\sum_{j=1}^N[\hat{v}_j^4 - \mathrm{E}(\tilde{v}_j^4)] \xrightarrow{p} 0$. This amounts to show that

$$(c)\ \tfrac{1}{nT}\textstyle\sum_{j=1}^N[\tilde{v}_j^4 - \mathrm{E}(\tilde{v}_j^4)] \xrightarrow{p} 0 \ \text{ and }\ (d)\ \tfrac{1}{nT}\textstyle\sum_{j=1}^N(\hat{v}_j^4 - \tilde{v}_j^4) \xrightarrow{p} 0.$$

To prove (c), we have

$$\begin{aligned}&\tfrac{1}{nT}\textstyle\sum_{j=1}^N \tilde{v}_j^4 - \tfrac{1}{nT}\sum_{j=1}^N \mathrm{E}(\tilde{v}_j^4)\\ =&\tfrac{1}{nT}\textstyle\sum_{j=1}^N\sum_{h=1}^N q_{jh}^4[v_h^4 - \mathrm{E}(v_h^4)] + \tfrac{3}{N}\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2q_{jm}^2(v_l^2v_m^2 - \sigma^4)\\ &+ \tfrac{4}{N}\textstyle\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^3q_{jm}v_l^3v_m + \tfrac{6}{N}\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N\sum_{\substack{h\neq m,l\\ m=1}}^N q_{jl}^2q_{jm}q_{jh}v_l^2v_mv_h\\ &+ \tfrac{1}{nT}\textstyle\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N\sum_{\substack{h\neq m,l\\ m=1}}^N\sum_{\substack{p\neq m,l,h\\ m=1}}^N q_{jl}q_{jm}q_{jh}q_{jp}v_lv_mv_hv_p \equiv \sum_{r=1}^5 R_r.\end{aligned}$$

By using WLLN of Davidson (1994, Theorem 19.7) for M.D. arrays as in the proof of (a), we have $R_r = o_p(1)$ for $r = 1, 3, 4, 5$. For R_2, noting that $v_l^2v_m^2 - \sigma^4 = (v_l^2 - \sigma^2)(v_m^2 - \sigma^2) + \sigma^2(v_m^2 - \sigma^2) + \sigma^2(v_l^2 - \sigma^2)$, we have

$$\begin{aligned}R_2 = &\tfrac{6}{N}\textstyle\sum_{l=1}^N(v_l^2 - \sigma^2)[\sum_{j=1}^N\sum_{m=1}^{l-1} q_{jl}^2q_{jm}^2(v_m^2 - \sigma^2)]\\ &+ \tfrac{6}{N}\textstyle\sum_{l=1}^N[\sum_{j=1}^N\sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2q_{jm}^2\sigma^2(v_l^2 - \sigma^2)] \equiv \tfrac{6}{N}\sum_{l=1}^N(f_l + f_{2,l}).\end{aligned}$$

Since $\mathrm{E}[f_l|\mathcal{G}_{l-1}] = 0$ and $\{f_{2,l}\}$ are independent, it is easy to see they both form an M.D. sequence. In addition, it is easily seen that $\mathrm{E}|f_{s,l}|^{1+\epsilon} < \infty$, for $s = 1, 2$ and $\epsilon > 0$, so that $\{f_l\}$ and $\{f_{2,l}\}$ are uniformly integrable. Therefore, the WLLN of Davidson (1994, Theorem 19.7) also implies that $\frac{6}{N}\sum_{l=1}^N f_l = o_p(1)$ and $\frac{6}{N}\sum_{l=1}^N f_{2,l} = o_p(1)$.

To prove (d), we have by (B.11) $\hat{v}_j^4 = \tilde{v}_j^4 + 4\tilde{v}_j^3 \psi_j'(\hat{\xi}_{nT}^* - \xi_0) + o_p(\|\hat{\xi}_{nT}^* - \xi_0\|)$. It follows that

$$\frac{1}{nT}\sum_{j=1}^N(\hat{v}_j^4 - \tilde{v}_j^4) = \frac{4}{N}\sum_{j=1}^N \tilde{v}_j^3 \psi_j'(\hat{\xi}_{nT}^* - \xi_0) + o_p(\|\hat{\xi}_{nT}^* - \xi_0\|)$$
$$= \frac{4\sigma^3\kappa_3}{N}\sum_{j=1}^N(\sum_{k=1}^N q_{jk}^3 \psi_j')(\hat{\xi}_{nT}^* - \xi_0) + o_p(\|\hat{\xi}_{nT}^* - \xi_0\|) = o_p(1).$$

Proof of (ii). The consistency of $\widehat{\Sigma}_{nT}^*$ to $\Sigma_{nT}^*(\theta_0)$ can be shown similarly as what we do in the proof of Theorem 3.2 for results (b) and (c). For $\widehat{\Gamma}_{nT}^* - \Gamma_{nT}^*(\theta_0) \xrightarrow{p} 0$, we only need to show that $\text{Bias}^*(\hat{\delta}_{nT}^*) - \text{Bias}^*(\delta_0) = o_p(1)$, based on Corollary 3.1. That is to show

$$\frac{1}{N_1}\{\mathtt{tr}[\mathcal{P}_2'(\hat{\delta}_{nT}^*)\mathcal{P}_2(\hat{\delta}_{nT}^*)\mathbb{P}_{\mathbb{D}}(\hat{\rho}_{nT}^*)] - \mathtt{tr}(\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}})\} = o_p(1),$$

which can be easily proved by using the MVT as we do for $\frac{1}{N_1}[H_{\lambda\lambda}^{*\mathtt{NS}}(\bar{\delta}) - H_{\lambda\lambda}^{*\mathtt{NS}}(\delta_0)]$ in the proof of Theorem 3.2 (b). ■

References

[1] Alvarez, J., Arellano, M., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121-1159.

[2] Andrews, D. W., 1992. Generic uniform convergence. *Econometric Theory* 8, 241-257.

[3] Andrews, D. W., 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 821-856.

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

[5] Anselin, L., Le Gallo, J., & Jayet, H., 2008. Spatial panel econometrics. In *The econometrics of panel data: Fundamentals and Recent Development in Theory and Practice* Springer, Berlin, Heidelberg. 625-660.

[6] Antweiler, W., 2001. Nested random effects estimation in unbalanced panel data. *Journal of econometrics* 101, 295-313.

[7] Arellano, M., Hahn, J., 2007. Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs* 43, 381.

[8] Badinger, H., Egger, P., 2015. Fixed effects and random effects estimation of higher-order spatial autoregressive models with spatial autoregressive and heteroscedastic disturbances. *Spatial Economic Analysis* 10, 11-35.

[9] Bai, J., 1997. Estimation of a change point in multiple regression models. *Review of Economics and Statistics* 79, 551-563.

[10] Bai, J., 2009. Panel data models with interactive fixed effects. *Econometrica* 77, 1229-1279.

[11] Bai, J., 2013. Likelihood approach to dynamic panel models with interactive effects. *Working paper*, Columbia University.

[12] Bai, J., Liao, Y., & Yang, J., 2015. Unbalanced panel data models with interactive effects. In: Badi H. Baltagi (Eds). *The Oxford Handbook of Panel Data*.

[13] Bai, J., Ng, S., 2013. Principal components estimation and identification of static factors. *Journal of Econometrics 176*, 18-29.

[14] Balazsi, L., Matyas, L., & Wansbeek, T., 2017. *Fixed effects models.* In The Econometrics of Multi-Dimensional Panels (pp. 1-34). Springer, Cham.

[15] Baltagi, B. H., 2013. *Econometric Analysis of Panel Data*, 5th edition. John Wiley & Sons Ltd.

[16] Baltagi, B. H., Bresson, G., & Etienne, J. M., 2015. Hedonic housing prices in Paris: an unbalanced spatial lag pseudo-panel model with nested random effects. *Journal of Applied Econometrics* 30, 509-528.

[17] Baltagi, B. H., Chang, Y. J., 1994. Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model. *Journal of Econometrics* 62, 67-89.

[18] Baltagi, B. H., Egger, P., & Pfaffermayr, M., 2003. A generalized design for bilateral trade flow models. *Economics letters* 80, 391-397.

[19] Baltagi, B. H., Egger, P., & Pfaffermayr, M., 2007. Estimating models of complex FDI: Are there third-country effects? *Journal of econometrics* 140, 260-281.

[20] Baltagi, B. H., Song, 2006. Unbalanced panel data: A survey. *Statistical Papers* 47, 493-523.

[21] Baltagi, B. H., Song, S. H., & Jung, B. C., 2001. The unbalanced nested error component regression model. *Journal of econometrics* 101, 357-381.

[22] Baltagi, B. H., Song, S. H., Koh, W., 2003. Testing panel data regression models with spatial error correlation. *Journal of econometrics* 117, 123-150.

[23] Baltagi, B. H., Yang, Z. L., 2013a. Standardized LM tests for spatial error dependence in linear or panel regressions. *Econometrics Journal* 16, 103-134.

[24] Baltagi, B. H., Yang, Z. L., 2013b. Heteroskedasticity and non-normality robust LM tests of spatial dependence. *Regional Science and Urban Economics* 43, 725-739.

[25] Bertrand, M., Luttmer, E. F., & Mullainathan, S., 2000. Network effects and welfare cultures. *The Quarterly Journal of Economics* 115(3), 1019-1055.

[26] Billingsley, P., 1968. *Convergence of probability measures.* John Wiley & Sons.

[27] Bobonis, G. J., Finan, F., 2009. Neighborhood peer effects in secondary school enrollment decisions. *The Review of Economics and Statistics* 91(4), 695-716.

[28] Bramoullé, Y., Djebbari, H., & Fortin, B., 2009. Identification of peer effects through social networks. *Journal of econometrics* 150, 41-55.

[29] Caner, M., Hansen, B. E., 2004. Instrumental variable estimation of a threshold model. *Econometric Theory*, 813-843.

[30] Card, D., Mas, A., & Rothstein, J., 2008. Tipping and the Dynamics of Segregation. *The Quarterly Journal of Economics* 123, 177-218.

[31] Case, A. C., 1991. Spatial patterns in household demand. *Econometrica*, 953-965.

[32] Chan, K. S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of statistics* 21, 520-533.

[33] Cox, D. R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society: Series B (Methodological)* 49, 1-18.

[34] Dang, V. A., Kim, M., & Shin, Y., 2012. Asymmetric capital structure adjustments: New evidence from dynamic panel threshold models. *Journal of Empirical Finance* 19, 465-482.

[35] Davidson, J., 1994. *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

[36] Davies, R. B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247-254.

[37] Davis, P., 2001. Estimating multi-way error components models with unbalanced data structures using instrumental variables. *Journal of Econometrics* 106, 67-95.

[38] Deng, Y., 2018. Estimation for the spatial autoregressive threshold model. *Economics Letters* 171, 172-175.

[39] Devereux, M. P., Lockwood, B., & Redoano, M., 2007. Horizontal and vertical indirect tax competition: Theory and some evidence from the USA. *Journal of Public Economics* 91, 451-479.

[40] Durlauf, S. N., Young, H. P., 2001. *The new social economics*. Social dynamics, 1-14.

[41] Egger, P., Pfaffermayr, M., & Winner H., 2005. An unbalanced spatial panel data approach to US state tax competition. *Economics Letter* 88, 329-335.

[42] Elhorst, J. P., Fréret, S., 2009. Evidence of political yardstick competition in France using a two-regime spatial Durbin model with fixed effects. *Journal of Regional Science* 49, 931-951.

[43] Ertur, C., Koch, W., 2007. Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of applied econometrics* 22, 1033-1062.

[44] Feller, W., 1968. *An introduction to probability theory and its applications*. New York: Wiley.

[45] Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A., 1996. Crime and social interactions. *The Quarterly journal of economics* 111, 507-548.

[46] Greene, W. H., 2018. *Econometric Analysis*, 8th Edition. Pearson.

[47] Hall, B. C., 2015. *Lie groups, Lie algebras, and representations: an elementary introduction*. New York: Springer.

[48] Hall, P., Heyde, C. C., 1980. *Martingale Limit Theory and its Applications*. Acad. Press, New York.

[49] Han, X., Hsieh, C. S., & Ko, S. I. (2021). Spatial modeling approach for dynamic network formation and interactions. *Journal of Business & Economic Statistics* 39, 120-135.

[50] Hansen, B. E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 413-430.

[51] Hansen, B. E., 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of econometrics* 93, 345-368.

[52] Hansen, B. E., 2000. Sample splitting and threshold estimation. *Econometrica* 68, 575-603.

[53] Hansen, B. E., 2011. Threshold autoregression in economics. *Statistics and its Interface* 4, 123-127.

[54] Hsiao, C., 2014. *Analysis of panel data*. Cambridge university press.

[55] Hsiao, C., 2018. Panel models with interactive effects. *Journal of Econometrics* 206, 645-673.

[56] Kanbur, R., Keen, M., 1993. Jeux sans frontieres: Tax competition and tax coordination when countries differ in size. *The American Economic Review*, 877-892.

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

[58] Kelejian, H. H., Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review* 40, 509-533.

[59] Kelejian, H. H., Prucha, I. R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of econometrics* 104, 219-257.

[60] Kelejian, H. H., Prucha, I. R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of econometrics* 157, 53-67.

[61] Kim, J., Pollard, D., 1990. Cube root asymptotics. *The Annals of Statistics* 18, 191-219.

[62] Le Gallo, J., Pirotte, A., 2017. *Models for spatial panels.* In The Econometrics of Multi-dimensional Panels (pp. 263-289). Springer, Cham.

[63] Lee, L. F., 2002. Consistency and efficiency of least squares estimation for mixed regressive spatial autoregressive models. *Econometric Theory* 18, 252-277.

[64] Lee, L. F., 2004. A supplement to 'Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models'. *Working paper*, Department of Economics, Ohio State University.

[65] Lee, L. F., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of econometrics* 140, 333-374.

[66] Lee, L. F., Liu, X., & Lin, X., 2010. Specification and estimation of social interaction models with network structures. *The Econometrics Journal* 13, 145-176.

[67] Lee, L. F., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154, 165-185.

[68] Lee, L. F., Yu, J., 2016. Identification of spatial Durbin panel models. *Journal of Applied Econometrics* 31, 133-162.

[69] LeSage, J. P., Pace, R. K., 2004. Models for spatially dependent missing data. *The Journal of Real Estate Finance and Economics* 29, 233-254.

[70] Li, K., 2018. Spatial panel data models with structural change. *Working Paper*. Available at `https://mpra.ub.uni-muenchen.de/85388/1/MPRA_paper_85388.pdf`.

[71] Li, L. Y., Yang, Z. L., 2020. M-Estimation of Fixed Effects Spatial Dynamic Panel Data Models with Small T and Unknown Heteroskedasticity. *Regional Science and Urban Economics*. In Press, `https://doi.org/10.1016/j.regsciurbeco.2020.103520`.

[72] Li, L., Yang, Z., 2021. Spatial dynamic panel data models with correlated random effects. *Journal of Econometrics* 221, 424-454.

[73] Lin, X., 2010. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics* 28, 825-860.

[74] Lin, X., Lee, L. F., 2010. GMM estimation of spatial autoregressive models with unknown heteroscedasticity. *Journal of Econometrics* 157, 34-52.

[75] Liu, X., Lee, L. F., 2010. GMM estimation of social interaction models with centrality. *Journal of Econometrics* 159, 99-115.

[76] Liu, S. F., Yang, Z. L., 2015. Modified QML estimation of spatial autoregressive models with unknown heteroscedasticity and nonnormality. *Regional Science and Urban Economics* 52, 50-70.

[77] Liu, S. F. Yang, Z. L., 2020. Robust estimation and inference of spatial panel data models with fixed effects. *Japanese Journal of Statistics and Data Science*, forthcoming.

[78] Manski, C. F., 1993. Identification of endogenous social effects: The reflection problem. *The review of economic studies* 60(3), 531-542.

[79] Miao, K., Li, K., & Su, L., 2020. Panel threshold models with interactive fixed effects. *Journal of Econometrics* 219, 137-170.

[80] Moffitt, R. A., 2001. Policy interventions, low-level equilibria, and social interactions. *Social dynamics* 4, 6-17.

[81] Moscone, F., Tosetti, E., 2011. GMM estimation of spatial panels with fixed effects and unknown heteroscedasticity. *Regional Science and Urban Economics* 41, 487-497.

[82] Newey, W. K., McFadden, D., 1994., Large sample estimation and hypothesis testing. *Handbook of econometrics* 4, 2111-2245.

[83] Neyman, J., Scott, E. L., 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16, 1-32.

[84] Nielsen, S. B., 2001. A simple model of commodity taxation and cross-border shopping. *Scandinavian Journal of Economics*, 103, 599-623.

[85] Pesaran, M. H., Pick, A., 2007. Econometric issues in the analysis of contagion. *Journal of Economic Dynamics and Control* 31, 1245-1277.

[86] Schelling, T. C., 1971. Dynamic models of segregation. *Journal of mathematical sociology* 1, 143-186.

[87] Seo, M. H., Linton, O., 2007. A smoothed least squares estimator for threshold regression models. *Journal of Econometrics* 141, 704-735.

[88] Seo, M. H., Shin, Y., 2016. Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics* 195, 169-186.

[89] Smith, R. L., 1992. Some interlacing properties of the Schur complement of a Hermitian matrix. *Linear algebra and its applications* 177, 137-144.

[90] Tong, H., 1978., In: Chen, C.H. (Ed.), *On a Threshold Model in Pattern Recognition and Signal Processing*. Sijhoff & Noordhoff, Amsterdam.

[91] van der Vaart, A. W., 1998. *Asymptotic Statistics*. Cambridge University Press.

[92] Varah, J. M., 1975. A lower bound for the smallest singular value of a matrix. *Linear Algebra and its Applications* 11, 3-5.

[93] Wang, W., Lee, L. F., 2013a. Estimation of spatial autoregressive models with randomly missing data in the dependent variable. *Econometrics Journal* 16, 73-102.

[94] Wang, W., Lee, L. F., 2013b. Estimation of spatial panel data models with randomly missing data in the dependent variable. *Regional Science and Urban Economics* 43, 521-538.

[95] Wansbeek, T., Kapteyn, A., 1989. Estimation of the error-components model with incomplete panels. *Journal of Econometrics* 41, 341-361.

[96] Wei, L., Zhang, C., Su, J. J., & Yang, L., 2021. Panel threshold spatial Durbin models with individual fixed effects. *Economics Letters* 201, 109778.

[97] Wooldridge, J. M., 2019. Correlated random effects models with unbalanced panels. *Journal of Econometrics* 211 137-150.

[98] Xu, Y. H., Yang, Z. L., 2020. Specification tests for temporal heterogeneity in spatial panel data models with fixed effects. *Regional Science and Urban Economics.* In Press, `https://doi.org/10.1016/j.regsciurbeco.2019.103488`.

[99] Yang, Z. L., 2015. A general method for third-order bias and variance correction on a nonlinear estimator. *Journal of Econometrics* 186, 178-200.

[100] Yang, Z. L., 2018. Unified M-estimation of Fixed-Effects Spatial Dynamic Models with Short Panels. *Journal of Econometrics* 205, 423-447.

[101] Yang, Z. L., Yu, J. H, & Liu, S. F., 2016. Bias correction and refined inferences for fixed effects spatial panel data models. *Regional Science and Urban Economics* 61, 52-72.

[102] Yao, Y., Zhang, M., 2013. Performance of Officials and the Promotion Tournament: Evidence from Chinese Cities. *Economic Research Journal* 1, 137-150.

[103] Yao, Y., Zhang, M., 2015. Subnational leaders and economic growth: evidence from Chinese cities. *Journal of Economic Growth* 20, 405-436.

[104] Yu, J., Zhou, L. A., & Zhu, G., 2016. Strategic interaction in political competition: Evidence from spatial effects across Chinese cities. *Regional Science and Urban Economics* 57, 23-37.

[105] Zhou, J., Liu, J., Wang, F., & Wang, H., 2022. Autoregressive Model With Spatial Dependence and Missing Data.*Journal of Business and Economic Statistics* 40, 28-34.

[106] Zhou, J., Tu, Y., Chen, Y., & Wang, H., 2017. Estimating spatial autocorrelation with sampled network data. *Journal of Business and Economic Statistics* 35, 130-138.

[107] Zhu, Y., Han, X., & Chen, Y., 2020. Bayesian estimation and model selection of threshold spatial Durbin model. *Economics Letters* 188, 108956.