7-2020

Spatial panel data models with temporal heterogeneity

Yuhong XU

Singapore Management University, yuhong.xu.2015@phdecons.smu.edu.sg

SINGAPORE MANAGEMENT UNIVERSITY

PHD DISSERTATION

Spatial Panel Data Models with Temporal Heterogeneity

Yuhong Xu

supervised by

Professor ZHENLIN YANG

July 13, 2020

Spatial Panel Data Models with Temporal Heterogeneity

by

Yuhong Xu

Submitted to School of Economics in partial fulfillment of
the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

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

Anthony S.Tay
Associate Professor of Economics
Singapore Management University

Jing Li
Assistant Professor of Economics
Singapore Management University

Wen-Chi Liao,
Associate Professor of Real Estate
National University of Singapore

Singapore Management University
2020

Abstract

This dissertation studies the fixed effects (FE) spatial panel data (SPD) models with temporal heterogeneity (TH), where the regression coefficients and spatial coefficients are allowed to change with time. The FE-SPD model with time-varying coefficients renders the usual transformation method in dealing with the fixed effects inapplicable, and an adjusted quasi score (AQS) method is proposed, which adjusts the concentrated quasi score function with the fixed effects being concentrated out. AQS tests for the lack of temporal heterogeneity (TH) in slope and spatial parameters are first proposed. Then, a set of AQS estimation and inference methods for the FE-SPD model with temporal heterogeneity is developed, when the AQS tests reject the hypothesis of temporal homogeneity. Finally, an attempt is made to extend these methodologies to allow the idiosyncratic errors of the model to be heteroskedastic along the cross-section dimension, where a method called outer-product-of-martingale-differences is proposed to estimate the variance of the AQS functions which in turn gives a robust estimator of the variance-covariance matrix of the AQS estimators.

Asymptotic properties of the AQS tests are examined. Consistency and asymptotic normality of the AQS estimators are examined under both homoscedastic and heteroskedastic errors. Extensive Monte Carlo experiments are conducted and the results show excellent finite sample performance of the proposed AQS tests, the proposed AQS estimators of the full model, and the corresponding estimates of the standard errors. Empirical illustrations are provided.

Contents

Acknowledgment

I want to express my deepest gratitude to Prof. Zhenlin Yang, who instruct me and provide patience and guidance throughout my PhD years. With his instruction and encouragement, the research life is more rewarding than I could have ever imagined. His kindness and understandings impress me. Also I am grateful to Prof. Anthony S.Tay, Prof. Jing Li and Prof. Wen-Chi Liao. They provide useful suggestions and help me a lot in research.

To my family, thanks for supporting me in these years and always being the one I could turn to. Finally, thanks to the two administrative staff, Thor Qiu Ling and Amelia Tan, who help me a lot during my PhD life.

1 Introduction

Temporal heterogeneity is an important feature in spatial panel model (SPD) model but relatively unexplored in the spatial panel literature. In a SPD model, it can occur on spatial parameters, intercept, slope and error variance. Many economic processes, for example, housing decisions, technology adoption, unemployment, welfare participation, price decisions, crime rates, trade flows, etc., exhibit time heterogeneity patterns. Therefore, being able to control *unobserved heterogeneity* may be one of the most important features of a SPD model.

In the SPD model, the strength of the interactions among locations may not stay the same over time. Therefore, techniques based upon constant coefficient models might be inadequate. Models with time-varying coefficients (TVC) should not be ignored, it can enhance the short-run forecasting in terms of accuracy and consistency, and it also allows us to identify influential data observations with estimation of parameters on a period-by-period basis.

In this dissertation, adjusted quasi score (AQS) tests are firstly proposed to test for the lack of temporal heterogeneity in regression slopes and spatial parameters, then a set of AQS-based estimation and inference methods are developed for the FE-SPD with TVC under both homoscedastic and heteroskedastic errors.

Consider the following spatial panel data model (SPD) model with two-way fixed effects where the spatial effects appear in the model in the forms of spatial lag (SL) and spatial error (SE):

$$Y_{nt} = \lambda_t W_n Y_{nt} + X_{nt}\beta_t + c_n + \alpha_t l_n + U_{nt}, \quad U_{nt} = \rho_t M_n U_{nt} + V_{nt}, \tag{1.1}$$

where Y_{nt} is an $n \times 1$ vector of observations on the dependent variable for $t = 1, 2, \ldots, T$; X_{nt} is an $n \times k$ matrix containing the values of k exogenous regressors, W_n is an $n \times n$ spatial weight matrix; M_n is another spatial weight matrix capturing the spatial interactions among the disturbances, which can be the same as W_n. V_{nt} is an $n \times 1$ vector of idiosyncratic errors, possibly are subject to unknown heteroskedasticity; λ_t is the *spatial lag parameters* in period t, ρ_t is the *spatial error parameters* in period t, and β_t is the $k \times 1$ vector of regression coefficients for the tth period; c_n denotes the individual-specific fixed effects or the spatial heterogeneity in intercept. $\{\alpha_t\}$ are the unobserved time-specific effects or the unobserved temporal heterogeneity in the intercept, and l_n is an $n \times 1$ vector of ones.

The above FE-SPD model considerd in this dissertation is fairly general, it allows the existence of temporal heterogeneity in regression and spatial coefficients, settings specific to each chapter will be presented chapter-wisely.

The usual transformation method used to eliminate FE is not applicable here due to the TVC. Therefore, an AQS method is proposed to adjust the concentrated quasi score (CQS) function with the fixed effects being concentrated out. This dissertation contains three topics. Chapter 2 presents the first topic: "Specification Tests for Temporal Heterogeneity in Spatial Panel Data Models with Fixed Effects". It introduces two types of adjusted quasi score tests (naïve and robust) for temporal homogeneity/heterogeneity in regression and spatial coefficients in the SPD models allowing the existence of the time and/or individual specific fixed effects. Asymptotic properties of the AQS tests are examined. The Monte Carlo results show that the robust tests have much superior finite and large sample properties than the naïve tests. The proposed tests are robust against nonnormality, can be used to identify possible existence of temporal heterogeneity and can also be repeatedly applied to identify a 'parsimonious model'. Empirical applications of the proposed tests are given to facilitate the applications of the methods.

Chapter 3 presents the second topic: "Adjusted Quasi-Score Estimation of Spatial Panel Data Models with Time Varying Coefficients". This chapter focuses on the estimation and inference problems for the FE-SPD model with time varying coefficients (TVC), where the individual- and time-specific effects take an additive form, and the temporal heterogeneity occurs on the intercept, slopes, as well as the spatial lag parameters, allowing the spatial errors in the model. The unbiased estimating functions are obtained by adjusting the quasi scores to establish consistency and asymptotic normality of the proposed AQS-estimator, and to give a complete set of inference methods. Monte Carlo evidence for the good finite sample performance of the proposed methods is presented. An empirical illustration is provided.

Chapter 4 presents the third topic: " Heteroskedasticity Robust Estimation of Spatial

Panel Data Models with Temporal Heterogeneity". The presence of social interactions will lead to a more complicated variance structure, therefore we would expect the variances of the error terms to be different in certain applications. With spatial interactions, the homoskedasticity assumptions are quite restrictive in the SPD models. Therefore, an AQS estimation method is proposed to adjust the concentrated score functions with FE being concentrated out, so that the AQS functions obtained are robust against unknown heteroskedasticity. For heteroskedasticity robust inferences, we develop an outer-product-of-martingale-differences (OPMD) method for estimating the variance of the AQS functions, which together with the expected Hessian matrix of the AQS functions give a robust estimator of the VC matrix of the AQS estimators. Consistency and asymptotic normality of the AQS-estimators are examined. Monte Carlo study is conducted and the results show excellent finite sample performance of the AQS-estimators and the corresponding estimates of standard errors.

Chapter 5 concludes this dissertation and discusses possible extensions in the future.

2 Specification Tests for Temporal Heterogeneity in Spatial Panel Data Models with Fixed Effects

In this chapter, we propose *adjusted quasi score* (AQS) tests for testing the existence of temporal heterogeneity in slope and spatial parameters in spatial panel data (SPD) models, allowing for the presence of individual-specific and/or time-specific fixed effects (or in general intercept heterogeneity). The SPD model with *spatial lag* is treated in detail by first considering the model with individual fixed effects only, and then extending it to the model with both individual and time fixed effects. Two types of AQS tests (*naïve and robust*) are proposed, and their asymptotic properties are presented. These tests are then fully extended to SPD models with both *spatial lag* and *spatial error*. Monte Carlo results show that the robust tests have much superior finite and large sample properties than the naive tests. Thus, the proposed robust tests provide reliable tools for identifying possible existence of temporal heterogeneity in regression and spatial coefficients. Empirical applications of the proposed tests are given[1].

2.1 Introduction

Being able to control *unobserved heterogeneity* may be one of the most important features of a panel data (PD) model. Heterogeneity may occur on intercept, slope and error variance. In a spatial PD model (SPD), it may also occur on spatial parameters (Anselin, 1988). Heterogeneity in variance is often referred to as *heteroskedasticity*. Heterogeneity may occur in spatial and/or temporal dimension. When unobserved heterogeneity occurs on the intercept, it gives rise to individual-specific effects and/or time-specific effects, which may appear in the model *additively* or *interactively*. Change point or structural break may be considered as a special case of unobserved heterogeneity.

[1]We thank Editor, Gabriel Ahlfeldt, and two anonymous referees for their constructive comments that have led to significant improvements of this chapter. Thanks are also due to the participants of the XI World Conference of the Spatial Econometrics Association, Singapore, June 2017, and the seminar participants at the Tohoku University, Japan, Dec. 2018, for their helpful comments. Zhenlin Yang gratefully acknowledges the financial support from Singapore Management University under Grant C244/MSS16E003.

Temporal heterogeneity is a common feature in an SPD model. It is an important issue but relatively unexplored in the spatial panel literature. Temporal heterogeneity may occur as a result of a credit crunch or debt, an oil price shock, a tax policy change, a fad or fashion in society, a discovery of a new medicine, and an enaction of new governmental program (Bai, 2010). Many economic processes, for example, housing decisions, technology adoption, unemployment, welfare participation, price decisions, crime rates, trade flows, etc., exhibit time heterogeneity patterns. Values observed at one location depend on the values of neighboring observations at nearby locations. Therefore, one may be interested in the question whether this dependence stays the same over time.

There is a sizable literature on temporal heterogeneity in regular panel data models, mostly on change points or structural breaks, see, Bai (2010), Liao (2008), Feng et al. (2009), to name a few. In spatial models, previous literature has focused more on the spatial heterogeneity (e.g., Aquaro et al., 2015; LeSage et al., 2016, 2017). The literature on temporal heterogeneity in spatial panel data models is rather thin. We are only aware of the following two works, Sengupta (2017) who proposes tests for a structural break in a spatial panel model without fixed effects, and Li (2018) who studies fixed effects SPD models with structural changes. SPD models with temporal heterogeneity also appear in finance literature, see, e.g., Blasques et al. (2016) and Catania and Billé (2017), but under a different setting where the time dimension is much larger than the spatial dimension.

In this chapter, we consider the fixed effects SPD models with temporal heterogeneity in regression and spatial coefficients. We focus on the testing problems. The presence of temporal heterogeneity renders the usual fixed effects estimation method through transformation (Lee and Yu, 2010; Baltagi and Yang, 2013b; Yang et al. 2016) inapplicable in handling the individual-specific fixed effects. A general method, the *adjusted quasi score* (AQS) method, is introduced for constructing tests for temporal homogeneity/heterogeneity on regression coefficients and spatial correlation coefficients in SPD models, allowing for presence of spatial-temporal heterogeneity in the intercepts (or fixed effects). The SPD model with spatial lag dependence is first treated in detail by first considering the model with individual-specific fixed effects only, and then extended to the model with both individual and time specific fixed effects. Two types of AQS tests (*naïve and robust*) are proposed, and their asymptotic properties are presented. These tests are then fully extended to the SPD models with both spatial lag (SL) and spatial error (SE) dependence. Monte Carlo results show that the robust tests have much superior finite and large sample properties than the naive tests. Thus, the proposed robust tests provide reliable tools for practitioners. Two empirical applications of the proposed tests are presented, and a detailed guidance is given to aid applied researchers in their empirical studies.

The rest of the chapter is organized as follows. Section 2.2 presents AQS tests for the panel SL model with one-way and two-way fixed effects, where a general method for constructing non-normality robust AQS tests is outlined. Section 2.3 generalizes these tests to the SPD models with both SL and SE dependence. Section 2.4 presents Monte Carlo results. Section 2.5 presents two empirical applications to give a detailed illustration on how the proposed methods are implemented. Section 2.6 discuss possible extensions and concludes this chapter.

2.2 Tests for Temporal Heterogeneity in Panel SL Model

In this section, we introduce the general AQS method for constructing the specification tests and a method for the practical implementations of these tests, using the simplest panel SL model with one-way FE (i.e., unobserved *spatial heterogeneity* in the intercept). Then, we extend these tests to a panel SL model with two-way FE (i.e., the unobserved *spatiotemporal heterogeneity* in intercepts). Asymptotic properties of the proposed tests are presented. Some key quantities for calculating the test statistics, the Hessian and expected Hessian matrices, and the variance-covariance matrix of the AQS function, are given in Appendix A.2, and proofs of theorems are sketched in Appendix A.3.

2.2.1 Panel SL model with one-way FE

Consider the following panel SL model with individual-specific FE, or one-way FE:

$$Y_{nt} = \lambda_t W_n Y_{nt} + X_{nt}\beta_t + c_n + V_{nt}, \tag{2.1}$$

where Y_{nt} is an $n \times 1$ vector of observations on the dependent variable for $t = 1, 2, \ldots, T$; X_{nt} is an $n \times k$ matrix containing the values of exogenous regressors and possibly their spatial lags, W_n is an $n \times n$ spatial weight matrix; V_{nt} is an $n \times 1$ vector of independent and identically distributed (iid) disturbances with mean zero and variance σ^2; λ_t is the *spatial lag parameter* and β_t is a $k \times 1$ vector of regression coefficients for the tth period; and c_n denotes the individual-specific fixed effects or the spatial heterogeneity in intercept.

Null hypotheses. We are primarily interested in tests for temporal homogeneity (TH) in regression and spatial coefficients, i.e., the tests of the null hypothesis:

$$H_0^{\mathtt{TH}} : \beta_1 = \cdots = \beta_T = \beta \;\text{ and }\; \lambda_1 = \cdots = \lambda_T = \lambda, \tag{2.2}$$

allowing for the presence of unobserved cross-sectional heterogeneity in intercept, i.e., the individual specific fixed effects c_n. If $H_0^{\mathtt{TH}}$ is rejected, one may wish to find the 'cause' of such a rejection instead of fitting the general heterogeneous model (2.1). Natural tests to proceed would be the tests of TH in regression coefficients only (RH), $H_0^{\mathtt{RH}} : \beta_1 = \cdots = \beta_T = \beta$, and the tests of TH in spatial coefficients only (SH): $H_0^{\mathtt{SH}} : \lambda_1 = \cdots = \lambda_T = \lambda$. If $H_0^{\mathtt{RH}}$ is not rejected, then one may infer that the cause of rejection of $H_0^{\mathtt{TH}}$ is the existence of temporal heterogeneity in spatial coefficients; if $H_0^{\mathtt{SH}}$ is not rejected, then the cause of rejection of $H_0^{\mathtt{TH}}$ may be the existence temporal heterogeneity in regression coefficients. In both cases, one would fit a simpler model of heterogeneous spatial coefficients only, or of heterogeneous regression coefficients only. If both $H_0^{\mathtt{RH}}$ and $H_0^{\mathtt{SH}}$ are rejected, one may need to fit the general model (2.1). However, rejection of both $H_0^{\mathtt{RH}}$ and $H_0^{\mathtt{SH}}$ may be due to the existence of *change points* (CPs) in β-coefficients and λ-coefficients, giving rise to a case of particular interest: change point detection in the spirit of Bai (2010) and Li (2018):

$$H_0^{\mathtt{CP}} : \beta_1 = \cdots = \beta_{b_0} \neq \beta_{b_0+1} = \cdots = \beta_T \;\text{ and }\; \lambda_1 = \cdots = \lambda_{\ell_0} \neq \lambda_{\ell_0+1} = \cdots = \lambda_T, \tag{2.3}$$

where $1 < b_0, \ell_0 < T$, and b_0 and ℓ_0 can be the same or different. If $H_0^{\mathtt{CP}}$ is not rejected, one may fit a much simpler model with one CP in β_t at $t = b_0$ and one CP for λ_t at $t = \ell_0$. These discussions can be extended to have more one CP in β_t and λ_t. All of these hypotheses can be put in a general framework and tests can be constructed in a general manner.[1]

Adjusted (quasi) score functions. As λ_t and β_t are allowed to change with t, the usual fixed-effects estimation methods, such as first differencing or orthogonal transformation, cannot be applied. We propose an *adjusted score* (AS) or *adjusted quasi score* (AQS) method for estimating the structural parameters in the model, which proceeds by first eliminating c_n through direct maximization of the loglikelihood function, given the structural parameters, and then adjusting the resulted concentrated (quasi) score function to give a set of estimating functions that are unbiased or asymptotically unbiased so as to achieve asymptotically unbiased estimation. The resulted set of AS or AQS functions then lead to a set of score-type of tests, referred to as the **AQS tests** in this chapter, for identifying temporal heterogeneity in regression coefficients and spatial parameters.

We develop score-type tests as they require only the estimation of the null model. However, the construction of the score-type of tests requires the full quasi score (QS) function, derived from the quasi Gaussian loglikelihood, **as if** $\{V_{nt}\}$ are iid $N(0, \sigma^2 I_n)$:

$$\ell_{\mathtt{SL1}}(\boldsymbol{\theta}, c_n) = -\frac{nT}{2}\ln(2\pi\sigma^2) + \sum_{t=1}^{T}\ln|A_n(\lambda_t)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} V'_{nt}(\lambda_t, \beta_t, c_n)V_{nt}(\lambda_t, \beta_t, c_n), \tag{2.4}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \sigma^2)'$, $\boldsymbol{\beta} = (\beta'_1, \ldots, \beta'_T)'$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$; $A_n(\lambda_t) = I_n - \lambda_t W_n$, I_n is an $n \times n$ identity matrix, and $V_{nt}(\beta_t, \lambda_t, c_n) = A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t - c_n,\ t = 1, \ldots, T$.

First, given $\boldsymbol{\theta}$, $\ell_{\mathtt{SL1}}(\boldsymbol{\theta}, c_n)$ is partially maximized at: $\tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{T}\sum_{t=1}^{T}[A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t]$, which gives the concentrated loglikelihood function of $\boldsymbol{\theta}$ upon substitution:

$$\ell^c_{\mathtt{SL1}}(\boldsymbol{\theta}) = -\frac{nT}{2}\ln(2\pi\sigma^2) + \sum_{t=1}^{T}\ln|A_n(\lambda_t)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\widetilde{V}'_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda})\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}), \tag{2.5}$$

where $\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t - \tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda})$. Then, differentiate $\ell^c_{\mathtt{SL1}}(\boldsymbol{\theta})$ to get the concentrated score (CS) or concentrated quasi score (CQS) function of $\boldsymbol{\theta}$:

$$S^c_{\mathtt{SL1}}(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2}X'_{nt}\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}), \quad t = 1, \ldots, T, \\ \frac{1}{\sigma^2}(W_n Y_{nt})'\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}) - \mathrm{tr}[G_n(\lambda_t)], \quad t = 1, \ldots, T, \\ -\frac{nT}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\widetilde{V}'_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda})\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}), \end{cases} \tag{2.6}$$

[1] Various conditional tests of, e.g., RH given SH, SH given RH, CP on β_t only given SH, and CP on λ_t only given RH, are also of interest, of which, the test of RH given SH is an extension of the well known Chow's (1960) test for a linear regression and Anselin's (1988, Sec. 9.2.2) test for a spatial error model.

where $G_n(\lambda_t) = W_n A_n^{-1}(\lambda_t), t = 1, \ldots, T$.

Let $\theta_0 = (\beta_0', \lambda_0', \sigma_0^2)'$ be the true value of the general parameter vector $\theta = (\beta', \lambda', \sigma^2)'$. We view that Model (2.1) holds only under the true θ_0. The usual expectation and variance operators correspond to θ_0. At the true θ_0, we have $\tilde{c}_n(\beta_0, \lambda_0) = \overline{V}_n + c_n$ and thus $\widetilde{V}_{nt} \equiv \widetilde{V}_{nt}(\beta_0, \lambda_0) = V_{nt} - \overline{V}_n$, where $\overline{V}_n = \frac{1}{T}\sum_{t=1}^T V_{nt}$. Furthermore, $W_n Y_{nt} = G_n(\lambda_{t0})(X_{nt}\beta_{t0} + c_n + V_{nt})$. With these, it is easy to show that,

$$\mathrm{E}[S^c_{\mathrm{SL1}}(\theta_0)] = \big\{0'_{Tk,1},\ -\tfrac{1}{T}\mathrm{tr}[G_n(\lambda_{t0})],\ t = 1, \ldots T,\ -\tfrac{n}{2\sigma_0^2}\big\}',$$

where $0_{m,r}$ denotes an $m \times r$ matrix of zeros. Clearly, $\frac{1}{nT}\mathrm{E}[S^c_{\mathrm{SL1}}(\theta_0)] \nrightarrow 0$, unless $T \to \infty$. A necessary condition for consistent estimation is violated. Therefore, the direct approach does not yield consistent estimators unless T goes to large. Even if T goes large with n, there will be an asymptotic bias of order $O(\frac{1}{T^2})$ for the estimation of $\{\lambda_t\}$, and an asymptotic bias of order $O(\frac{1}{T})$ for the estimation of σ^2.

To have a inference method that is consistent and asymptotically unbiased, CS or CQS function given in (2.6) should be adjusted by subtracting the above bias vector from it, leading to the AS or AQS function as

$$S^\star_{\mathrm{SL1}}(\theta) = \begin{cases} \frac{1}{\sigma^2} X'_{nt}\widetilde{V}_{nt}(\beta, \lambda),\ t = 1, \ldots, T, \\ \frac{1}{\sigma^2}(W_n Y_{nt})'\widetilde{V}_{nt}(\beta, \lambda) - \frac{T-1}{T}\mathrm{tr}[G_n(\lambda_t)],\ t = 1, \ldots, T, \\ -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}'_{nt}(\beta, \lambda)\widetilde{V}_{nt}(\beta, \lambda). \end{cases} \tag{2.7}$$

It is easy to show that $\mathrm{E}[S^\star_{\mathrm{SL1}}(\theta_0)] = 0$, and that $\frac{1}{nT}S^\star_{\mathrm{SL1}}(\theta_0) \xrightarrow{p} 0$ as $n \to \infty$ alone, or both n and T go infinity. Thus, this AQS function gives a set of unbiased estimating functions, and paves the way for developing asymptotically valid score-type tests.[2]

Construction of AQS tests. Denote by $\tilde{\theta}_{\mathrm{SL1}}$ the constrained estimator of θ under H_0.[3] Let $J_{\mathrm{SL1}}(\theta) = -\frac{\partial}{\partial\theta'}S^\star_{\mathrm{SL1}}(\theta)$, $I_{\mathrm{SL1}}(\theta_0) = \mathrm{E}[J_{\mathrm{SL1}}(\theta_0)]$ and $\Sigma_{\mathrm{SL1}}(\theta_0) = \mathrm{Var}[S^\star_{\mathrm{SL1}}(\theta_0)]$,

[2]Solving the estimating equation, $S^\star_{\mathrm{SL1}}(\theta) = 0$, gives the unconstrained AQS estimator of θ. Simplifying this AQS function under the null gives AQS function of the null model, and the constrained estimates of the null model parameters. See the end of section for a general method for estimating the null models.

[3]For testing H_0^{TH} in (2.2), for example, $\tilde{\beta}_{\mathrm{SL1}} = 1_T \otimes \tilde{\beta}_{\mathrm{SL1}}$, $\tilde{\lambda}_{\mathrm{SL1}} = 1_T \otimes \tilde{\lambda}_{\mathrm{SL1}}$, and $\tilde{\theta}_{\mathrm{SL1}} = (\tilde{\beta}'_{\mathrm{SL1}}, \tilde{\lambda}'_{\mathrm{SL1}}, \tilde{\sigma}^2_{\mathrm{SL1}})'$, where $\tilde{\beta}_{\mathrm{SL1}}$ and $\tilde{\lambda}_{\mathrm{SL1}}$ are the estimators of the common β and λ, and 1_T is a $T \times 1$ vector of ones.

with their expressions given in Appendix A.2.1. The *usual* score test, treating $S^{\star}_{\texttt{SL1}}(\theta)$ as a genuine score vector so that the information matrix equality (IME) holds, takes the form:

$$T_{\texttt{SL1}} = S^{\star}_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}})' J^{-1}_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}}) S^{\star}_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}}), \tag{2.8}$$

where $J_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}})$ can be replaced by $I_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}})$ or $\Sigma_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}})$. However, $S^{\star}_{\texttt{SL1}}(\theta)$ is not a genuine score function even if the errors are normal, as it comes from the original score function after some adjustments. In this case, the IME or its generalized version (Cameron and Trivedy, 2005; Wooldridge, 2010) does not hold. Hence, the test statistic $T_{\texttt{SL1}}$ constructed in this *usual* way may not be valid even if the errors are normal, *unless* under 'specific' situations where $I_{\texttt{SL1}}(\theta_0)$ and $\Sigma_{\texttt{SL1}}(\theta_0)$ are asymptotically equivalent, i.e., the IME holds asymptotically. See the discussions below Theorem 2.1 for details.

To address this issue, denoting $k_q = \dim(\theta) = (k+1)T + 1$, we put our testing problem in a general framework with null hypothesis being written as

$$H_0 : C\theta_0 = 0, \tag{2.9}$$

where C is a $k_p \times k_q$ matrix generating k_p linear contrasts in the parameter vector θ.

For example, for testing $H_0^{\texttt{TH}}$ in (2.2), the number of constraints $k_p = (k+1)(T-1)$, and the linear contrast matrix $C = [\texttt{blkdiag}\{C_T^k,\ C_T^1\},\ 0_{k_p,1}]$, where $\texttt{blkdiag}\{\cdots\}$ forms a block diagonal matrix, and C_τ^m is an $m(\tau-1) \times m\tau$ matrix defined as

$$C_\tau^m = [(1_{\tau-1} \otimes I_m),\ -(I_{\tau-1} \otimes I_m)], \tag{2.10}$$

where $\otimes$ is the Kronecker product; for testing $H_0^{\texttt{RH}}$, $C = [C_T^k, 0_{k_p,T}, 0_{k_p,1}]$ and $k_p = (T-1)k$; for testing $H_0^{\texttt{SH}}$, $C = [0_{k_p,kT}, C_T^1, 0_{k_p,1}]$ and $k_p = T-1$; and for testing $H_0^{\texttt{CP}}$ in (2.3), $C = [\texttt{blkdiag}\{C_{b_0}^k, C_{T-b_0}^k, C_{\ell_0}^1,$
$C_{T-\ell_0}^1\}, 0_{k_p}]$ and $k_p = (T-2)(k+1)$. The C matrices for tests of $\texttt{CP}$ on β-coefficients only or tests of $\texttt{CP}$ on λ-coefficients only can be formulated easily. The $\texttt{CP}$-test can be carried out repeatedly until the 'true' change points are detected. In all these and other interesting cases, k_p and C can be easily written out.

The score-type test is constructed based on the AQS function $S^{\star}_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}})$, and its asymptotic variance-covariance (VC) matrix. Denote by $N_0 = n(T-1)$ the *effective* *sample size* to differentiate from the overall sample size $N = nT$. Under mild regularity conditions, such as the $\sqrt{N_0}$-consistency of $\tilde{\theta}_{\texttt{SL1}}$ under the null, we have by Taylor expansion:

$$\tfrac{1}{\sqrt{N_0}} S^{\star}_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}}) = \tfrac{1}{\sqrt{N_0}} S^{\star}_{\texttt{SL1}}(\theta_0) + \tfrac{1}{N_0} I_{\texttt{SL1}}(\theta_0)\sqrt{N_0}(\tilde{\theta}_{\texttt{SL1}} - \theta_0) + o_p(1), \text{ and}$$

$$[\tfrac{1}{N_0} I_{\texttt{SL1}}(\theta_0)]^{-1} \tfrac{1}{\sqrt{N_0}} S^{\star}_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}}) = [\tfrac{1}{N_0} I_{\texttt{SL1}}(\theta_0)]^{-1} \tfrac{1}{\sqrt{N_0}} S^{\star}_{\texttt{SL1}}(\theta_0) + \sqrt{N_0}(\tilde{\theta}_{\texttt{SL1}} - \theta_0) + o_p(1).$$

As $C\theta_0 = 0$ under H_0, we have $C\tilde{\theta}_{\texttt{SL1}} = 0$ (see Wooldridge 2010, p.424). It follows that

$$C[\tfrac{1}{N_0} I_{\texttt{SL1}}(\theta_0)]^{-1} \tfrac{1}{\sqrt{N_0}} S^{\star}_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}}) = C[\tfrac{1}{N_0} I_{\texttt{SL1}}(\theta_0)]^{-1} \tfrac{1}{\sqrt{N_0}} S^{\star}_{\texttt{SL1}}(\theta_0) + o_p(1), \tag{2.11}$$

leading to the asymptotic VC matrix of $C[\frac{1}{N} I_{\texttt{SL1}}(\theta_0)]^{-1} \frac{1}{\sqrt{N}} S^{\star}_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}})$ as

$$\Xi_{\texttt{SL1}}(\theta_0) = C[\tfrac{1}{N_0} I_{\texttt{SL1}}(\theta_0)]^{-1}[\tfrac{1}{N_0}\Sigma_{\texttt{SL1}}(\theta_0)][\tfrac{1}{N_0} I_{\texttt{SL1}}(\theta_0)]^{-1} C'. \tag{2.12}$$

This gives an asymptotically valid and nonnormality robust AQS test:

$$T^{(r)}_{\texttt{SL1}} = \widetilde{S}^{\star\prime}_{\texttt{SL1}} \widetilde{I}^{-1}_{\texttt{SL1}} C' \big(C \widetilde{I}^{-1}_{\texttt{SL1}} \widetilde{\Sigma}_{\texttt{SL1}} \widetilde{I}^{-1}_{\texttt{SL1}} C'\big)^{-1} C \widetilde{I}^{-1}_{\texttt{SL1}} \widetilde{S}^{\star}_{\texttt{SL1}}, \tag{2.13}$$

where $\widetilde{S}^{\star}_{\texttt{SL1}} = S^{\star}_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}})$, $\widetilde{I}_{\texttt{SL1}} = I_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}})$, and $\widetilde{\Sigma}_{\texttt{SL1}} = \Sigma_{\texttt{SL1}}(\tilde{\theta}_{\texttt{SL1}})$.

Remark 2.1 *Although the AQS test given in* (2.13) *is developed based on the panel SL model with 1FE, the general principles behind apply to all models considered in this chapter. It also applies to more complicated spatial models as well as many non-spatial models.*

Asymptotic properties. In studying the asymptotic properties of the proposed tests, we focus on the tests of temporal homogeneity to ease the exposition. Therefore, some of the regularity conditions, i.e., Assumptions 2 and 4, correspond to the null model under $H_0^{\texttt{TH}}$ in (2.2) only. However, these assumptions can be easily relaxed to cater a non-homogeneous null model. Denote $X^{\circ}_{nt} = X_{nt} - \bar{X}_n$, where $\bar{X}_n = \frac{1}{T}\sum_{t=1}^{T} X_{nt}$.

Assumption 1. *The disturbances* $\{v_{it}\}$ *are iid across* i *and* t *with mean zero, variance* σ_0^2*, and* $E|v_{it}|^{4+\epsilon_0} < \infty$ *for some* $\epsilon_0 > 0$.

Assumption 2. *Under* H_0*, the parameter space* Λ *of the common* λ *is compact, and the true value* λ_0 *is in the interior of* Λ*. The matrix* $A_n(\lambda)$ *is invertible for all* $\lambda \in \Lambda$.

Assumption 3. *The elements of X_{nt} are non-stochastic, and are bounded uniformly in n and t, such that $\lim_{N_0 \to \infty} \frac{1}{N_0} \sum_{t=1}^{T} X_{nt}^{\circ\prime} X_{nt}^{\circ}$ exists and nonsingular. The elements of c_n are uniformly bounded.*

Assumption 4. *W_n has zero diagonal elements, and is uniformly bounded in both row and column sums in absolute value. $A_n^{-1}(\lambda)$ is also uniformly bounded in both row and column sums in absolute value for λ in a neighborhood of λ_0.*

Theorem 2.1 *Under Assumptions 1-4, if further, (i) $\tilde{\theta}_{\texttt{SL1}}$ is $\sqrt{N_0}$-consistent for θ_0 under $H_0^{\texttt{TH}}$, and (ii) $I_{\texttt{SL1}}(\theta)$ and $\Xi_{\texttt{SL1}}(\theta)$ are positive definite for θ in a neighborhood of θ_0 when N_0 is large enough, then we have, under $H_0^{\texttt{TH}}$, $T_{\texttt{SL1}}^{(r)} \xrightarrow{D} \chi^2_{k_p}$, as $n \to \infty$.*

Note that in case of testing for temporal homogeneity, $k_p = (T-1)(k+1)$, and that in case of testing for a 'single change' of points, $k_p = (T-2)(k+1)$. It can easily be seen that $T_{\texttt{SL1}}$ is in general not an asymptotic pivotal quantity due to the violation of IME. However, if $I_{\texttt{SL1}}(\theta_0) \asymp \Sigma_{\texttt{SL1}}(\theta_0)$, where $\asymp$ denotes asymptotic equivalence, then $\widetilde{I}_{\texttt{SL1}}^{-1} C' \big(C \widetilde{I}_{\texttt{SL1}}^{-1} \widetilde{\Sigma}_{\texttt{SL1}} \widetilde{I}_{\texttt{SL1}}^{-1} C'\big)^{-1} C \widetilde{I}_{\texttt{SL1}}^{-1} \asymp \widetilde{I}_{\texttt{SL1}}^{-1}$ (see Wooldridge, 2010, p. 424), and hence $T_{\texttt{SL1}}$ becomes valid. This is in fact true when T is also large as seen from the expressions given in Appendix A.2.1, but this case needs an extra care as in Remark 2.2 below.

Remark 2.2 *When $T \to \infty$ as $n \to \infty$, the degrees of freedom (d.f) of the chi-square statistic increase with n. In this case, one may apply the arguments for 'double asymptotics' (see, e.g., Rempala and Wesolowski, 2016) to show that $(T_{\texttt{SL1}}^{(r)} - k_p)/\sqrt{2k_p} \xrightarrow{D} N(0,1)$ as $n/\sqrt{T} \to \infty$. This sample size requirement (n goes large faster than $\sqrt{T}$) is rather weak as it is typical in spatial panels that n is at least as large as T.*

Estimation of null models. The construction of the AQS tests requires estimation of various null models, which could be the homogeneous model as specified by $H_0^{\texttt{TH}}$ in (2.2), the model with homogeneity in β's only, the model with homogeneity in λ's only, or the model with change points as specified by $H_0^{\texttt{CP}}$ in (2.3), etc. Each null model can be estimated by solving the simplified AQS equations by simplifying $S_{\texttt{SL1}}^{\star}(\theta)$ according to the null hypothesis, which is clearly inconvenient to the applied researchers. To facilitate **practical applications** of our methods, a general *Lagrange Multiplier* (LM) method is introduced. Let $l_{\texttt{SL1}}(\boldsymbol{\theta})$ be the objective function to be maximized subject to $C\boldsymbol{\theta}_0 = 0$, with $S^{\star}_{\texttt{SL1}}(\boldsymbol{\theta})$ given in (2.7) being its partial derivatives. Define the Lagrangian

$$\mathcal{L}_{\texttt{SL1}}(\boldsymbol{\theta}) = l_{\texttt{SL1}}(\boldsymbol{\theta}) - \phi'(C\boldsymbol{\theta}),$$

where ϕ is a $k_p \times 1$ vector of Lagrange multipliers. Taking partial derivatives and equating to 0, we have k_q equations $\frac{\partial \mathcal{L}_{\texttt{SL1}}}{\partial \boldsymbol{\theta}} = S^{*}_{\texttt{SL1}}(\boldsymbol{\theta}) - C'\phi = 0_{k_q,1}$. Together with the k_p constraints $C\boldsymbol{\theta} = 0$, we have $k_q + k_p$ equations for the $k_q + k_p$ unknowns $\boldsymbol{\theta}$ and ϕ, leading to

$$\begin{pmatrix} \tilde{\boldsymbol{\theta}}_{\texttt{SL1}} \\ \tilde{\phi}_{\texttt{SL1}} \end{pmatrix} = \arg \left\{ \begin{array}{l} S^{\star}_{\texttt{SL1}}(\boldsymbol{\theta}) - C'\phi = 0_{k_q,1} \\ C\boldsymbol{\theta} = 0_{k_p,1} \end{array} \right\}. \tag{2.14}$$

To further aid the applications, we make the Matlab codes available upon request, or online at `http://www.mysmu.edu/faculty/zlyang/`.

Finally, from the expressions of $I_{\texttt{SL1}}(\boldsymbol{\theta}_0)$ and $\Sigma_{\texttt{SL1}}(\boldsymbol{\theta}_0)$ given in Appendix A.2.1, we see that they both contain c_n, which is estimated by plugging the null estimates $\tilde{\boldsymbol{\beta}}_{\texttt{SL1}}$ and $\tilde{\boldsymbol{\lambda}}_{\texttt{SL1}}$ into $\tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda})$. Furthermore, in case of nonnormality, the VC matrix $\Sigma_{\texttt{SL1}}(\boldsymbol{\theta}_0)$ contains two additional parameters, the skewness γ and excess kurtosis κ of the idiosyncratic errors $V_{n,it}$, and their estimates are obtained by applying Lemma 4.1(a) of Yang et al. (2016). See Sec. 2.5 for a detailed discussion on issues related to practical implementations.

However, as the hypothesis H_0^{TH} given in (2.2) and the corresponding homogeneous model plays an important role in studying the asymptotic properties of the test and in Monte Carlo simulation, an outline is given on how $S^{\star}_{\texttt{SL1}}(\boldsymbol{\theta})$ is simplified and how it leads to constrained AQS estimators with the desired asymptotic properties. Let $\theta = (\beta', \lambda, \sigma^2)'$. The constrained estimate of c_n given (β, λ) becomes $\tilde{c}^{\circ}_n(\beta, \lambda) = A_n(\lambda)\bar{Y}_n - \bar{X}_n\beta$ where $\bar{Y}_n$ and $\bar{X}_n$ are the averages of $\{Y_{nt}\}$ and $\{X_{nt}\}$, respectively. Along the same line leading to (2.7), one can easily show that AQS function for the homogeneous model takes the form:

$$S^{\circ}_{\texttt{SL1}}(\theta) = \begin{cases} \frac{1}{\sigma^2}\sum_{t-1}^{T} X^{\circ\prime}_{nt}\widetilde{V}^{\circ}_{nt}(\beta,\lambda), \\ \frac{1}{\sigma^2}\sum_{t-1}^{T}(W_nY^{\circ}_{nt})'\widetilde{V}^{\circ}_{nt}(\beta,\lambda) - (T-1)\mathrm{tr}[G_n(\lambda)], \\ -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\widetilde{V}^{\circ\prime}_{nt}(\beta,\lambda)\widetilde{V}^{\circ}_{nt}(\beta,\lambda), \end{cases} \tag{2.15}$$

$\widetilde{V}_{nt}^{\circ}(\beta,\lambda) = A_n(\lambda)Y_{nt} - X_{nt}\beta - \tilde{c}_n^{\circ}(\beta,\lambda) = A_n(\lambda)Y_{nt}^{\circ} - X_{nt}^{\circ}\beta$, where $Y_{nt}^{\circ} = Y_{nt} - \bar{Y}_n$ and $X_{nt}^{\circ} = X_{nt} - \bar{X}_n$. Solving the estimating equations, $S_{\texttt{SL1}}^{\circ}(\theta) = 0$, gives the null estimator $\tilde{\theta}_{\texttt{SL1}}$ of θ. The AQS estimation provides an alternative to the QML estimation based on transformation of Lee and Yu (2010). The two can be shown to be asymptotically equivalent, and therefore $\tilde{\theta}_{\texttt{SL1}}$ is $\sqrt{n(T-1)}$-consistent for θ.

2.2.2 Panel SL model with two-way FE

While the unit-specific fixed effects are important to the spatial panel data models, the time-specific effects often cannot be neglected. In this section, we extend our tests to panel SL model with two-way FE (2FE). The model takes the following form:

$$Y_{nt} = \lambda_t W_n Y_{nt} + X_{nt}\beta_t + c_n + \alpha_t 1_n + V_{nt}, \tag{2.16}$$

where $\{\alpha_t\}$ are the unobserved time-specific effects or the unobserved temporal heterogeneity in the intercept, and 1_n is an $n \times 1$ vector of ones. As the spatial parameters and regression coefficients change only with time. One can apply transformation method to eliminate the time-specific effects as is widely applied in the literature, see, e.g., Lee and Yu (2010), Baltagi and Yang (2013b) and Yang et al. (2016). Define $J_n = I_n - \frac{1}{n}1_n 1_n'$. Assume W_n is row-normalized (i.e., row sums are one). Then, $J_n W_n = J_n W_n J_n$. Let $(F_{n,n-1}, \frac{1}{\sqrt{n}}1_n)$ be the orthonormal eigenvector matrix of J_n, where $F_{n,n-1}$ is the $n \times (n-1)$ sub-matrix corresponding to the eigenvalues of one. By *Spectral Theorem*, $J_n = F_{n,n-1}F_{n,n-1}'$. It follows that $F_{n,n-1}'W_n = F_{n,n-1}'W_n F_{n,n-1}F_{n,n-1}'$. Premultiplying $F_{n,n-1}'$ on both sides of (2.16), we have the following transformed model:

$$Y_{nt}^* = \lambda_t W_n^* Y_{nt}^* + X_{nt}^*\beta_t + c_n^* + V_{nt}^*, \; t = 1, \ldots, T, \tag{2.17}$$

where $Y_{nt}^* = F_{n,n-1}'Y_{nt}$, and so are X_{nt}^*, c_n^* and V_{nt}^* defined; and $W_n^* = F_{n,n-1}'W_n F_{n,n-1}$. After the transformation, the overall sample size is $(n-1)T$. Model (2.17) takes an identical form as Model (2.1). Furthermore, $V_{nt}^* \sim (0, \sigma_0^2 I_{n-1})$, which is normal if V_{nt}^* is, and is independent of V_{ns}^*, $s \neq t$.[4] Hence, the steps leading to the score-type tests and the consistent estimation of the null model are similar to those for the SL one-way FE model.

[4] The time-specific effects can also be eliminated by pre-multiplying J_n on both sides of (2.16). However, the resulted disturbances $J_n V_{nt}$ would not be linearly independent over the cross-section dimension.

Define $A_n^*(\lambda_t) = I_{n-1} - \lambda_t W_n^*$, $t = 1, \ldots, T$. The quasi Gaussian loglikelihood function of $\theta = (\beta', \lambda', \sigma^2)'$ and c_n^* of Model (2.17) is

$$\ell_{\mathrm{SL2}}(\theta, c_n^*) = -\tfrac{(n-1)T}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln|A_n^*(\lambda_t)| \\ - \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T V_{nt}^{*\prime}(\lambda_t, \beta_t, c_n^*)V_{nt}^*(\lambda_t, \beta_t, c_n^*), \tag{2.18}$$

where $V_{nt}^*(\beta_t, \lambda_t, c_n^*) = A_n^*(\lambda_t)Y_{nt}^* - X_{nt}^*\beta_t - c_n^*$. Given θ, $\ell_{\mathrm{SL2}}(\theta, c_n^*)$ is maximized at:

$$\tilde{c}_n^*(\beta, \lambda) = \tfrac{1}{T}\textstyle\sum_{t=1}^T [A_n^*(\lambda_t)Y_{nt}^* - X_{nt}^*\beta_t], \tag{2.19}$$

which gives the concentrated loglikelihood function of θ upon substitution:

$$\ell_{\mathrm{SL2}}^c(\theta) = -\tfrac{(n-1)T}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln|A_n^*(\lambda_t)| - \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime}(\beta, \lambda)\widetilde{V}_{nt}^*(\beta, \lambda), \tag{2.20}$$

where $\widetilde{V}_{nt}^*(\beta, \lambda) = A_n^*(\lambda_t)Y_{nt}^* - X_{nt}^*\beta_t - \tilde{c}_n^*(\beta, \lambda)$. Now, define $G_n^*(\lambda_t) = W_n^* A_n^{*-1}(\lambda_t)$. Differentiating $\ell_{\mathrm{SL2}}^c(\theta)$ gives the CS or CQS function of θ of Model (2.17):

$$S_{\mathrm{SL2}}^c(\theta) = \begin{cases} \frac{1}{\sigma^2}X_{nt}^{*\prime}\widetilde{V}_{nt}^*(\beta, \lambda), \quad t = 1, \ldots, T, \\ \frac{1}{\sigma^2}(W_n^* Y_{nt}^*)'\widetilde{V}_{nt}^*(\beta, \lambda) - \mathrm{tr}[G_n^*(\lambda_t)], \quad t = 1, \ldots, T, \\ -\frac{(n-1)T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime}(\beta, \lambda)\widetilde{V}_{nt}^*(\beta, \lambda). \end{cases} \tag{2.21}$$

Takes the expectation of the above score, we have,

$$\mathrm{E}[S_{\mathrm{SL2}}^c(\theta_0)] = \big\{0_{Tk}', \ -\tfrac{1}{T}\mathrm{tr}[G_n^*(\lambda_{t0})], \ t = 1, \ldots T, \ -\tfrac{n-1}{2\sigma_0^2}\big\}',$$

which again shows that model estimation based on maximizing the quasi loglikelihood would not lead to consistent estimates of the model parameters. The CQS function given in (2.21) should be adjusted by recentering, giving the AQS function of Model (2.17):

$$S_{\mathrm{SL2}}^\star(\theta) = \begin{cases} \frac{1}{\sigma^2}X_{nt}^{*\prime}\widetilde{V}_{nt}^*(\beta, \lambda), \ t = 1, \ldots, T, \\ \frac{1}{\sigma^2}(W_n^* Y_{nt}^*)'\widetilde{V}_{nt}^*(\beta, \lambda) - \frac{T-1}{T}\mathrm{tr}[G_n^*(\lambda_t)], \ t = 1, \ldots, T, \\ -\frac{(n-1)(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime}(\beta, \lambda)\widetilde{V}_{nt}^*(\beta, \lambda). \end{cases} \tag{2.22}$$

It is easy to show that $\mathrm{E}[S_{\mathrm{SL2}}^\star(\theta)] = 0$, and that $\frac{1}{nT}S_{\mathrm{SL2}}^\star(\theta_0) \xrightarrow{p} 0$ as $n \to \infty$ alone, or both n and T go infinity. Thus, this AQS function gives a set of unbiased estimating functions, and paves the way for developing asymptotic valid score-type tests. Again, simplifying this AQS function under various null hypotheses gives the AQS functions of the null models and the constrained estimates. See the end of the Section for a general formulation.

Now, the tests concerning $\{\beta_t\}$ and $\{\lambda_t\}$ allow the existence of both unobserved cross-sectional and time-specific heterogeneity in the intercept, i.e., the existence of both individual specific fixed effects and the time specific fixed effects. As the transformed 2FE panel SL model takes an identical form as 1FE panel SL model, the tests developed for 1FE panel SL model extends directly to give tests for the 2FE panel SL model. Let $\tilde{\theta}_{\text{SL2}}$ be the null estimate of θ. Let $I_{\text{SL2}}(\theta_0)$ and $\Sigma_{\text{SL2}}(\theta_0)$ be, respectively, the expected negative Hessian and the VC matrix of $S^{\star}_{\text{SL2}}(\theta_0)$, given in Appendix A.2.2. The AQS test, robust against nonnormality and taking into account of the estimation of fixed effects, is:

$$T^{(r)}_{\text{SL2}} = \widetilde{S}^{\star\prime}_{\text{SL2}}\widetilde{I}^{-1}_{\text{SL2}}C'\big(C\widetilde{I}^{-1}_{\text{SL2}}\widetilde{\Sigma}_{\text{SL2}}\widetilde{I}^{-1}_{\text{SL2}}C'\big)^{-1}C\widetilde{I}^{-1}_{\text{SL2}}\widetilde{S}^{\star}_{\text{SL2}}, \tag{2.23}$$

where $\widetilde{S}^{\star}_{\text{SL2}} = S^{\star}_{\text{SL2}}(\tilde{\theta}_{\text{SL2}})$, $\widetilde{I}_{\text{SL2}} = I_{\text{SL2}}(\tilde{\theta}_{\text{SL2}})$, and $\widetilde{\Sigma}_{\text{SL2}} = \Sigma_{\text{SL2}}(\tilde{\theta}_{\text{SL2}})$. As in the case of 1FE-SL model, when $I_{\text{SL2}}(\theta_0) \asymp \Sigma_{\text{SL2}}(\theta_0)$, $\widetilde{I}^{-1}_{\text{SL2}}C'\big(C\widetilde{I}^{-1}_{\text{SL2}}\widetilde{\Sigma}_{\text{SL2}}\widetilde{I}^{-1}_{\text{SL2}}C'\big)^{-1}C\widetilde{I}^{-1}_{\text{SL2}} \asymp \widetilde{I}^{-1}_{\text{SL2}}$, and hence $T^{(r)}_{\text{SL2}}$ reduces to the *naïve* test: $T_{\text{SL2}} = \widetilde{S}^{\star\prime}_{\text{SL2}}\widetilde{J}^{-1}_{\text{SL2}}\widetilde{S}^{\star}_{\text{SL2}}$, where $\widetilde{J}_{\text{SL2}} = -\frac{\partial}{\partial\theta}S^{\star}_{\text{SL2}}(\tilde{\theta}_{\text{SL2}})$.

Asymptotic properties of these tests can be studied along the same line as the tests for 1FE panel SL model. Again we focus on the test of H_0^{TH} for ease of exposition. The effective sample size becomes $N_0 = (n-1)(T-1)$ due to the 'estimation' of both individual- and time-specific FEs. Let $\Xi_{\text{SL2}}(\theta)$ and $X^{*\circ}_{nt}$ be defined as $\Xi_{\text{SL1}}(\theta)$ and X°_{nt} in SL-one way FE model.

Assumption $3'$: The elements of X_{nt} are nonstochastic, and are bounded uniformly in n and t, such that $\lim_{N_0\to\infty}\frac{1}{N_0}\sum_{t=1}^{T}X^{*\circ\prime}_{nt}X^{*\circ}_{nt}$ exists and is nonsingular.

Theorem 2.2 *Under Assumptions 1-2, $3'$, and 4, if further, (i) $\tilde{\theta}_{\text{SL2}}$ is $\sqrt{N_0}$-consistent for θ_0 under H_0^{TH}, and (ii) $I_{\text{SL2}}(\theta)$ and $\Xi_{\text{SL2}}(\theta)$ are positive definite for θ in a neighborhood of θ_0 when N_0 is large enough, then we have, under H_0^{TH}, $T^{(r)}_{\text{SL2}} \xrightarrow{D} \chi^2_{k_p}$, as $n\to\infty$.*

Note that while the effective sample size for the 2FE-SL model is smaller than that of the 1FE-SL model, the d.f. associated with the test statistics remain the same. As discussed below Theorem 2.1, T_{SL2} is not an asymptotic pivotal quantity unless T is also large. As in Remark 2.2, if T grows with n, $(T^{(r)}_{\text{SL2}} - k_p)/\sqrt{2k_p} \xrightarrow{D} N(0,1)$, as $n/\sqrt{T}\to\infty$.

Estimation of null models. The general constrained root-finding method, the LM procedure, presented at the end of the section for the panel SL model with 1FE directly applies to the panel SL model with 2FE to give constrained estimates of various null models. This greatly facilitates the practical applications. Again, the homogeneous model specified by H_0^{TH} in (2.2) and its AQS estimation play important roles in studying the asymptotic properties and performing Monte Carlo simulations, and therefore an outline is given on the estimation procedures based on the simplified AQS function. The constrained estimate of c_n^*, given (β, λ), becomes $\tilde{c}_n^{*\circ}(\beta, \lambda) = A_n^*(\lambda)\bar{Y}_n^* - \bar{X}_n^*\beta$, where $\bar{Y}_n^*$ and $\bar{X}_n^*$ are the averages of $\{Y_{nt}^*\}$ and $\{X_{nt}^*\}$, respectively. Along the same line leading to (2.15), we have the AS or AQS function for the homogeneous panel SL model with 2FE:

$$S_{\text{SL2}}^{\circ}(\theta) = \begin{cases} \frac{1}{\sigma^2}\sum_{t-1}^{T} X_{nt}^{*\circ\prime}\widetilde{V}_{nt}^{*\circ}(\beta, \lambda), \\ \frac{1}{\sigma^2}\sum_{t-1}^{T}(W_n^* Y_{nt}^{*\circ})'\widetilde{V}_{nt}^{*\circ}(\beta, \lambda) - (T-1)\text{tr}[G_n^*(\lambda)], \\ -\frac{(n-1)(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\widetilde{V}_{nt}^{*\circ\prime}(\beta, \lambda)\widetilde{V}_{nt}^{*\circ}(\beta, \lambda), \end{cases} \tag{2.24}$$

where $\widetilde{V}_{nt}^{*\circ}(\beta, \lambda) = A_n^*(\lambda)Y_{nt}^* - X_{nt}^*\beta - \tilde{c}_n^{*\circ}(\beta, \lambda) = A_n(\lambda)Y_{nt}^{*\circ} - X_{nt}^{*\circ}\beta$, $Y_{nt}^{*\circ} = Y_{nt}^* - \bar{Y}_n^*$ and $X_{nt}^{*\circ} = X_{nt}^* - \bar{X}_n^*$. Solving the estimating equations, $S_{\text{SL2}}^{\circ}(\theta) = 0$, gives the null estimator $\tilde{\theta}_{\text{SL2}}$ of θ. Again, it can be shown to be asymptotically equivalent to the transformation-based QML estimator of Lee and Yu (2010). Thus, $\tilde{\theta}_{\text{SL2}}$ is $\sqrt{(n-1)(T-1)}$-consistent for θ. The estimation of c_n and γ and κ contained in $I_{\text{SL2}}(\boldsymbol{\theta}_0)$ and $\Sigma_{\text{SL2}}(\boldsymbol{\theta}_0)$ proceeds similarly.

2.3 Test for Temporal Heterogeneity in Panel SLE Model

The tests introduced in the earlier section can be easily extended to a more general SPD model where the the disturbances are also subject to spatial interactions, giving an SPD model with both spatial lag and error (SLE) dependence. Again, we first present results for the one-way FE model, and then the results for the two-way FE model.

2.3.1 Panel SLE model with one-way FE

The SLE model with one-way fixed effects has the form:

$$Y_{nt} = \lambda_t W_n Y_{nt} + X_{nt}\beta_t + c_n + U_{nt}, \quad U_{nt} = \rho_t M_n U_{nt} + V_{nt}, \tag{2.25}$$

where M_n is another spatial weight matrix capturing the spatial interactions among the disturbances, which can be the same as W_n, and $\{\rho_t\}$ are the spatial error parameters, possibly changing with time. Again, we are primarily interested in the test for temporal homogeneity, which now corresponds to a test of the following null hypothesis:

$$H_0^{\texttt{TH}}: \beta_1 = \cdots = \beta_T = \beta, \ \ \lambda_1 = \cdots = \lambda_T = \lambda, \ \text{ and } \ \rho_1 = \cdots = \rho_T = \rho. \tag{2.26}$$

If this test is rejected, one would be interested in testing various hypotheses discussed in Sec. 2.1, including $H_0^{\texttt{CP}}$ in (2.3) extended to include the ρ-component, to find out the cause of the rejection. An **interesting test** for the panel SLE model would be the conditional test: $H_0^{\texttt{THC}}: \beta_1 = \cdots = \beta_T = \beta$, and $\lambda_1 = \cdots = \lambda_T = \lambda$, given $\rho_1 = \cdots = \rho_T = \rho$. In this case, the alternative (full) model is a submodel of (2.25) with the disturbance following a homogeneous SAR process: $U_{nt} = \rho M_n U_{nt} + V_{nt}$. We present the most general case here, and give necessary details related to this submodel at the end of Sec.2.3.

Following the same set of notation as in the earlier section, and further denoting $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_T)'$, $\theta = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\rho}', \sigma^2)'$, and $B_n(\rho_t) = I_n - \rho_t M_n, t = 1, \ldots, T$, we have the (quasi) Gaussian loglikelihood for (θ, c_n):

$$\begin{aligned}\ell_{\texttt{SLE1}}(\theta, c_n) = &-\tfrac{nT}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln|A_n(\lambda_t)| + \sum_{t=1}^T \ln|B_n(\rho_t)| \\ &- \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T V_{nt}'(\beta_t, \lambda_t, \rho_t, c_n) V_{nt}(\beta_t, \lambda_t, \rho_t, c_n),\end{aligned} \tag{2.27}$$

where $V_{nt}(\beta_t, \lambda_t, \rho_t, c_n) = B_n(\rho_t)[A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t - c_n], \ t = 1, \ldots, T.$

Similarly to the developments in the previous section, we first eliminate c_n through a direct maximization of the loglikelihood function, given the other model parameters θ, and then adjust the resulted CS or CQS function to eliminate the asymptotic bias or inconsistency. Given θ, $\ell_{\texttt{SLE1}}(\theta, c_n)$ is maximized at

$$\tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\rho}) = \big[\textstyle\sum_{t=1}^T B_n'(\rho_t)B_n(\rho_t)\big]^{-1}\sum_{t=1}^T \big[B_n'(\rho_t)B_n(\rho_t)\big(A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t\big)\big], \tag{2.28}$$

leading to the concentrated (quasi) Gaussian loglikelihood function of θ upon substitution:

$$\ell^c_{\texttt{SLE1}}(\theta) = -\tfrac{nT}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln|A_n(\lambda_t)| + \sum_{t=1}^T \ln|B_n(\rho_t)| \\ - \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T \widetilde{V}'_{nt}(\beta,\lambda,\rho)\widetilde{V}_{nt}(\beta,\lambda,\rho), \tag{2.29}$$

where $\widetilde{V}_{nt}(\beta,\lambda,\rho) = V_{nt}(\beta_t,\lambda_t,\rho_t,\tilde{c}_n(\beta,\lambda,\rho)) = B_n(\rho_t)[A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t - \tilde{c}_n(\beta,\lambda,\rho)]$.

To facilitate the subsequent derivations, denote $U^\circ_{nt}(\beta_t,\lambda_t) = A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t$, and $D_n(\rho_t) = B'_n(\rho_t)B_n(\rho_t)$ and $\mathbb{D}_n(\rho) = \sum_{t=1}^T D_n(\rho_t)$. Then,

$$\widetilde{V}_{nt}(\beta,\lambda,\rho) = B_n(\rho_t)U^\circ_{nt}(\beta_t,\lambda_t) - B_n(\rho_t)\tilde{c}_n(\beta,\lambda,\rho),$$

$\tilde{c}_n(\beta,\lambda,\rho) = \mathbb{D}_n^{-1}(\rho)\sum_{t=1}^T D_n(\rho_t)U^\circ_{nt}(\beta_t,\lambda_t)$, and the key term in (2.29):

$$\textstyle\sum_{t=1}^T \widetilde{V}'_{nt}(\beta,\lambda,\rho)\widetilde{V}_{nt}(\beta,\lambda,\rho) = \sum_{t=1}^T U^{\circ\prime}_{nt}(\beta_t,\lambda_t)D_n(\rho_t)U^\circ_{nt}(\beta_t,\lambda_t) \\ - \big(\textstyle\sum_{t=1}^T D_n(\rho_t)U^\circ_{nt}(\beta_t,\lambda_t)\big)'\mathbb{D}_n^{-1}(\rho)\big(\sum_{t=1}^T D_n(\rho_t)U^\circ_{nt}(\beta_t,\lambda_t)\big).$$

Differentiating $\ell^c_{\texttt{SLE1}}(\theta)$ gives the CS or CQS function of θ:

$$S^c_{\texttt{SLE1}}(\theta) = \begin{cases} \frac{1}{\sigma^2}X'_{nt}B'_n(\rho_t)\widetilde{V}_{nt}(\beta,\lambda,\rho), \quad t = 1,\ldots,T, \\ \frac{1}{\sigma^2}(W_nY_{nt})'B'_n(\rho_t)\widetilde{V}_{nt}(\beta,\lambda,\rho) - \mathrm{tr}[G_n(\lambda_t)], \quad t = 1,\ldots,T, \\ \frac{1}{\sigma^2}\widetilde{V}'_{nt}(\beta,\lambda,\rho)H_n(\rho_t)\widetilde{V}_{nt}(\beta,\lambda,\rho) - \mathrm{tr}[H_n(\rho_t)], \quad t = 1,\ldots,T, \\ -\frac{nT}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}'_{nt}(\beta,\lambda,\rho)\widetilde{V}_{nt}(\beta,\lambda,\rho), \end{cases} \tag{2.30}$$

where $H_n(\rho_t) = M_nB_n^{-1}(\rho_t), \quad t = 1,\ldots,T$.

At the true θ_0, we have, $\tilde{c}_n(\beta_0,\lambda_0,\rho_0) = c_n + \mathbb{D}_n^{-1}\sum_{s=1}^T B'_{ns}V_{ns}$ and hence $\widetilde{V}_{nt} \equiv \widetilde{V}_{nt}(\beta_0,\lambda_0,\rho_0) = V_{nt} - B_{nt}\mathbb{D}_n^{-1}\sum_{s=1}^T B'_{ns}V_{ns}$, and $W_nY_{nt} = G_{nt}(X_{nt}\beta_0 + c_n + B_{nt}^{-1}V_{nt})$, where $B_{nt} = B_n(\rho_{t0})$, $G_{nt} = G_n(\lambda_{t0})$, and $\mathbb{D}_n = \mathbb{D}_n(\rho_0)$. It is easy to show that,

$$\mathrm{E}[S^c_{\texttt{SLE1}}(\theta_0)] = \begin{cases} 0_{Tk,1}, \\ -\mathrm{tr}[\mathbb{D}_n^{-1}(\rho_0)B'_n(\rho_{t0})B_n(\rho_{t0})G_n(\lambda_{t0})], \; t = 1,\ldots T, \\ -\mathrm{tr}[B_n(\rho_{t0})\mathbb{D}_n^{-1}(\rho_0)B'_n(\rho_{t0})H_n(\rho_{t0})], \; t = 1,\ldots T, \\ -\frac{n}{2\sigma_0^2}. \end{cases}$$

Therefore, the AS or AQS function of $\boldsymbol{\theta}$ for Model (2.25) takes the form:

$$
S^{\star}_{\mathtt{SLE1}}(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2} X'_{nt} B'_n(\rho_t) \widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\rho}), \quad t = 1, \dots, T, \\ \frac{1}{\sigma^2} (W_n Y_{nt})' B'_n(\rho_t) \widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\rho}) - \mathrm{tr}[R_{nt}(\boldsymbol{\rho}) G_n(\lambda_t)], \quad t = 1, \dots, T, \\ \frac{1}{\sigma^2} \widetilde{V}'_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\rho}) H_n(\rho_t) \widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\rho}) - \mathrm{tr}[S_{nt}(\boldsymbol{\rho}) H_n(\rho_t)], \quad t = 1, \dots, T, \\ -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^{T} \widetilde{V}'_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\rho}) \widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\rho}), \end{cases} \tag{2.31}
$$

where $R_{nt}(\boldsymbol{\rho}) = I_n - \mathbb{D}_n^{-1}(\boldsymbol{\rho}) B'_n(\rho_t) B_n(\rho_t)$ and $S_{nt}(\boldsymbol{\rho}) = I_n - B_n(\rho_t) \mathbb{D}_n^{-1}(\boldsymbol{\rho}) B'_n(\rho_t)$.

It is easy to show that $\mathrm{E}[S^{\star}_{\mathtt{SLE1}}(\boldsymbol{\theta})] = 0$, and that $\frac{1}{nT} S^{\star}_{\mathtt{SLE1}}(\boldsymbol{\theta}_0) \xrightarrow{p} 0$ as $n \to \infty$ alone, or both n and T go infinity. Thus, this AQS function gives a set of unbiased estimating functions, and paves the way for developing asymptotic valid score-type tests.

Construction of AQS tests. Denote the constrained estimator (under H_0) of $\boldsymbol{\theta}$ by $\tilde{\boldsymbol{\theta}}_{\mathtt{SLE1}}$.[5] To test various hypotheses concerning temporal homogeneity/heterogeneity, one is tempt to use the *naïve* test, $T_{\mathtt{SLE1}} = S^{\star}_{\mathtt{SLE1}}(\tilde{\boldsymbol{\theta}}_{\mathtt{SLE1}})' J^{-1}_{\mathtt{SLE1}}(\tilde{\boldsymbol{\theta}}_{\mathtt{SLE1}}) S^{\star}_{\mathtt{SLE1}}(\tilde{\boldsymbol{\theta}}_{\mathtt{SLE1}})$, treating $S^{\star}_{\mathtt{SLE1}}(\boldsymbol{\theta})$ as a genuine score function, where $J_{\mathtt{SLE1}}(\boldsymbol{\theta}_0) = -\frac{\partial}{\partial \boldsymbol{\theta}'} S^{\star}_{\mathtt{SLE1}}(\boldsymbol{\theta}_0)$, which can be replaced by $I_{\mathtt{SLE1}}(\boldsymbol{\theta}_0) = \mathrm{E}[J_{\mathtt{SLE1}}(\boldsymbol{\theta}_0)]$, or $\Sigma_{\mathtt{SLE1}}(\boldsymbol{\theta}_0) = \mathrm{Var}[S^{\star}_{\mathtt{SLE1}}(\boldsymbol{\theta}_0)]$ (see Appendix A.2.3 for their expressions). Again, $S^{\star}_{\mathtt{SLE1}}(\boldsymbol{\theta})$ is not a genuine score function. Hence, the test constructed in the usual way may not be a valid test statistic, even if the errors are normal.

To give a general robust test, we again, as in the previous section, put our testing problem in a general framework with null hypothesis being written as H_0: $C\boldsymbol{\theta}_0 = 0$, with some modifications on C to include the ρ parameters. The dimensions of C are again denoted as $k_p \times k_q$ with k_p linear contrasts on the parameter vector $\boldsymbol{\theta}$ of dimension $k_q = (k+2)T + 1$. For $H_0^{\mathtt{TH}}$ in (2.26), we have $k_p = (T-1)(k+2)$ and $C = [\mathtt{blkdiag}\{C_T^k, C_T^1, C_T^1\}, 0_{k_p,1}]$, where C_τ^m is defined in (2.10). For tests of $\mathtt{CP}$ in β_t, λ_t and ρ_t at time points b_0, ℓ_0 and r_0, respectively, $k_p = (T-2)(k+2)$ and $C = [\mathtt{blkdiag}\{C_{b_0}^k, C_{T-b_0}^k, C_{\ell_0}^1, C_{T-\ell_0}^1, C_{r_0}^1, C_{T-r_0}^1\}, 0_{k_p,1}]$.

Similarly, the score-type test is based on the AQS function $S^{\star}_{\mathtt{SLE1}}(\tilde{\boldsymbol{\theta}}_{\mathtt{SLE1}})$ evaluated at

[5] In case of testing $H_0^{\mathtt{TH}}$ given in (2.26), the constrained estimators of $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ are, respectively, $\tilde{\boldsymbol{\beta}}_{\mathtt{SLE1}} = 1_T \otimes \tilde{\beta}_{\mathtt{SLE1}}$, $\tilde{\boldsymbol{\lambda}}_{\mathtt{SLE1}} = 1_T \otimes \tilde{\lambda}_{\mathtt{SLE1}}$, and $\tilde{\boldsymbol{\rho}}_{\mathtt{SLE1}} = 1_T \otimes \tilde{\rho}_{\mathtt{SLE1}}$, where $\tilde{\beta}_{\mathtt{SLE1}}$, $\tilde{\lambda}_{\mathtt{SLE1}}$ and $\tilde{\rho}_{\mathtt{SLE1}}$ are the estimators of the common β, λ and ρ, leading to the constrained estimator of $\boldsymbol{\theta}$ as $\tilde{\boldsymbol{\theta}}_{\mathtt{SLE1}} = (\tilde{\boldsymbol{\beta}}'_{\mathtt{SLE1}}, \tilde{\boldsymbol{\lambda}}'_{\mathtt{SLE1}}, \tilde{\boldsymbol{\rho}}'_{\mathtt{SLE1}}, \tilde{\sigma}^2_{\mathtt{SLE1}})'$.

the null estimate $\tilde{\theta}_{\texttt{SLE1}}$ of θ, and the asymptotic VC matrix of $S^{\star}_{\texttt{SLE1}}(\tilde{\theta}_{\texttt{SLE1}})$. Now, the effective sample size is back to $N_0 = n(T-1)$ as for the 1FE panel SL model. Following the fundamental developments in Sec 2.2, we have, under mild regularity conditions such as the $\sqrt{N_0}$-consistency of $\tilde{\theta}_{\texttt{SLE1}}$, an asymptotically valid and nonnormality robust AQS test:

$$T^{(r)}_{\texttt{SLE1}} = \widetilde{S}^{\star\prime}_{\texttt{SLE1}} \widetilde{I}^{-1}_{\texttt{SLE1}} C' \big(C \widetilde{I}^{-1}_{\texttt{SLE1}} \widetilde{\Sigma}_{\texttt{SLE1}} \widetilde{I}^{-1}_{\texttt{SLE1}} C'\big)^{-1} C \widetilde{I}^{-1}_{\texttt{SLE1}} \widetilde{S}^{\star}_{\texttt{SLE1}}, \tag{2.32}$$

where $\widetilde{S}^{\star}_{\texttt{SLE1}} = S^{\star}_{\texttt{SLE1}}(\tilde{\theta}_{\texttt{SLE1}})$, $\widetilde{I}_{\texttt{SLE1}} = I_{\texttt{SLE1}}(\tilde{\theta}_{\texttt{SLE1}})$, and $\widetilde{\Sigma}_{\texttt{SLE1}} = \Sigma_{\texttt{SLE1}}(\tilde{\theta}_{\texttt{SLE1}})$.

Asymptotic properties of the proposed tests are established based on Assumptions 1-4 in Sec. 2.2, and the following additional conditions on M_n and $B_n(\rho)$.

Assumption 5. *Under H_0, the parameter space $\mathbb{P}$ of the common ρ is compact. The true value ρ_0 is in the interior of $\mathbb{P}$. The matrix $B_n(\rho)$ is invertible for all $\rho \in \mathbb{P}$. M_n has zero diagonal elements, and are uniformly bounded in both row and column sums in absolute value. $B_n^{-1}(\rho)$ is uniformly bounded in both row and column sums in absolute value for ρ in a neighborhood of ρ_0.*

Furthermore, the existence and consistency of the constrained estimator $\tilde{\beta}_{\texttt{SLE1}}$ depends on the existence and nonsingularity of $\lim_{n\to\infty} \frac{1}{nT} \sum_{t=1}^{T} X^{\circ\prime}_{nt} B_n' B_n X^{\circ}_{nt}$, which follows from Assumption 2 and the positive definiteness of $B_n' B_n$. Denoting $\Xi_{\texttt{SLE1}}(\theta) = C I^{-1}_{\texttt{SLE1}}(\theta) \Sigma_{\texttt{SLE1}}(\theta) I^{-1}_{\texttt{SLE1}}(\theta) C'$, we have the following theorem.

Theorem 2.3 *Under Assumptions 1-5, if further, (i) $\tilde{\theta}_{\texttt{SEL1}}$ is $\sqrt{N_0}$-consistent for θ_0 under $H_0^{\texttt{TH}}$, and (ii) $I_{\texttt{SLE1}}(\theta)$ and $\Xi_{\texttt{SLE1}}(\theta)$ are positive definite for θ in a neighborhood of θ_0 when N_0 is large enough, then we have, under $H_0^{\texttt{TH}}$, $T^{(r)}_{\texttt{SLE1}} \xrightarrow{D} \chi^2_{k_p}$, as $n \to \infty$.*

Note that the d.f. associated with the test statistics is $k_p = (T-1)(k+2)$ for testing for temporal homogeneity, and $k_p = (T-2)(k+2)$ for testing for a 'single change'. Similarly, if T increases with n it can be shown that $T_{\texttt{SLE1}}$ is not an asymptotic pivotal quantity, and that $(T^{(r)}_{\texttt{SLE1}} - k_p)/\sqrt{2k_p} \xrightarrow{D} N(0,1)$, as $n/\sqrt{T} \to \infty$.

Estimation of null models. The general LM procedure presented previously can be applied to estimate various null (1FE-SLE) models based on $S^{\star}_{\texttt{SLE1}}(\theta)$ and a properly specified linear contrast matrix C. To estimate the homogeneous model for asymptotic analyses and Monte Carlo simulation, let $\theta = (\beta', \lambda, \rho, \sigma^2)'$. Under $H_0^{\texttt{TH}}$, the constrained estimate of c_n given (β, λ) becomes $\tilde{c}_n^{\circ}(\beta, \lambda) = A_n(\lambda)\bar{Y}_n - \bar{X}_n\beta$, and the error vector becomes $\widetilde{V}_{nt}^{\circ}(\beta, \lambda, \rho) = B_n(\rho)[A_n(\lambda)Y_{nt}^{\circ} - X_{nt}^{\circ}\beta]$, where $Y_{nt}^{\circ} = Y_{nt} - \bar{Y}_n$, $X_{nt}^{\circ} = X_{nt} - \bar{X}_n$, and $\bar{Y}_n = \frac{1}{T}\sum_{t=1}^{T} Y_{nt}$ and $\bar{X}_n = \frac{1}{T}\sum_{t=1}^{T} X_{nt}$. The AQS function at $H_0^{\mathtt{TH}}$ takes the form:

$$S_{\mathtt{SLE1}}^{\circ}(\theta) = \begin{cases} \frac{1}{\sigma^2}\sum_{t=1}^{T} X_{nt}^{\circ}B_n'(\rho)\widetilde{V}_{nt}^{\circ}(\beta, \lambda, \rho), \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(W_nY_{nt}^{\circ})'B_n'(\rho)\widetilde{V}_{nt}^{\circ}(\beta, \lambda, \rho) - (T-1)\text{tr}[G_n(\lambda)], \\ \frac{1}{\sigma^2}\sum_{t=1}^{T} \widetilde{V}_{nt}^{\circ\prime}(\beta, \lambda, \rho)H_n(\rho)\widetilde{V}_{nt}^{\circ}(\beta, \lambda, \rho) - (T-1)\text{tr}[H_n(\rho)], \\ -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T} \widetilde{V}_{nt}^{\circ\prime}(\beta, \lambda, \rho)\widetilde{V}_{nt}^{\circ}(\beta, \lambda, \rho). \end{cases} \tag{2.33}$$

Solving the estimating equations, $S_{\mathtt{SLE1}}^{\circ}(\theta) = 0$, gives the null estimator $\tilde{\theta}_{\mathtt{SLE1}}$ of θ, which is shown to be asymptotically equivalent to the transformation-based QML estimator of Lee and Yu (2010), and thus is $\sqrt{n(T-1)}$-consistent. To estimate c_n, γ and κ, refer to the discussions at the end of the discussion for SL-one way FE model.

2.3.2 Panel SLE model with two-way FE

The panel SLE model with two-way fixed effects has the form:

$$Y_{nt} = \lambda_t W_n Y_{nt} + X_{nt}\beta_t + c_n + \alpha_t 1_n + U_{nt}, \quad U_{nt} = \rho_t M_n U_{nt} + V_{nt}, \tag{2.34}$$

which extends Model (2.16) by adding the spatial error dependence term. Applying the same orthonormal transformation as that for Model (2.16), i.e., premultiplying $F_{n,n-1}'$ on both sides of (2.34), and using $J_nW_n = J_nW_nJ_n$, $J_nM_n = J_nM_nJ_n$ and $J_n = F_{n,n-1}F_{n,n-1}'$, we have the following transformed model:

$$Y_{nt}^* = \lambda_t W_n^* Y_{nt}^* + X_{nt}^*\beta_t + c_n^* + U_{nt}^*, \quad U_{nt}^* = \rho_t M_n^* U_{nt}^* + V_{nt}^*, \tag{2.35}$$

where Y_{nt}^*, X_{nt}^*, c_n^*, W_n^* and V_{nt}^* are defined as in Model (2.17), and $M_n^* = F_{n,n-1}'M_nF_{n,n-1}$. After the transformation, the effective sample size becomes $N_0 = (n-1)(T-1)$ as for the 2FE panel SL model. As Model (2.35) takes an identical form as Model (2.25) and the elements of V_{nt}^* are iid normal if the original errors are normal, the steps leading to the score-type test and the steps leading to consistent estimation of the null models are similar. We first present the results for the general model, and then give the necessary details for the submodel with constant ρ at the end of this section and in Appendix A.2.5.

Define $A_n^*(\rho_t) = I_{n-1} - \lambda_t W_n^*$ and $B_n^*(\rho_t) = I_{n-1} - \rho_t M_n^*$, $t = 1, \ldots, T$. Similar to the previous section, we eliminate c_n^* through a direct maximization of the loglikelihood function to give the concentrated loglikelihood function of θ:

$$\ell^c_{\text{SLE2}}(\theta) = -\tfrac{nT}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln|A_n^*(\lambda_t)| + \sum_{t=1}^T \ln|B_n^*(\rho_t)|$$
$$\textit{the} - \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime}(\beta, \lambda, \rho)\widetilde{V}_{nt}^*(\beta, \lambda, \rho) \tag{2.36}$$

where $\widetilde{V}_{nt}^*(\beta, \lambda, \rho) = B_n^*(\rho_t)U_{nt}^{\circ *}(\beta_t, \lambda_t) - B_n^*(\rho_t)\mathbb{D}_n^{*-1}(\rho)\sum_{s=1}^T D_n^*(\rho_s)U_{ns}^{\circ *}(\beta_s, \lambda_s)$, $\mathbb{D}_n^*(\rho) = \sum_{t=1}^T D_n^*(\rho_t)$, $D_n^*(\rho_t) = B_n^{*\prime}(\rho_t)B_n^*(\rho_t)$, and $U_{nt}^{\circ *}(\beta_t, \lambda_t) = A_n^*(\lambda_t)Y_{nt}^* - X_{nt}^*\beta_t$. As in the previous subsection, we can obtain the AS or AQS function of θ for Model (2.34) as

$$S^\star_{\text{SLE2}}(\theta) = \begin{cases} \frac{1}{\sigma^2}X_{nt}^{*\prime}B_n^{*\prime}(\rho_t)\widetilde{V}_{nt}^*(\beta, \lambda, \rho), \quad t = 1, \ldots, T, \\ \frac{1}{\sigma^2}(W_n^*Y_{nt}^*)'B_n^{*\prime}(\rho_t)\widetilde{V}_{nt}^*(\beta, \lambda, \rho) - \text{tr}[R_{nt}^*(\rho)G_n^*(\lambda_t)], \quad t = 1, \ldots, T, \\ \frac{1}{\sigma^2}\widetilde{V}_{nt}^{*\prime}(\beta, \lambda, \rho)H_n^*(\rho_t)\widetilde{V}_{nt}^*(\beta, \lambda, \rho) - \text{tr}[S_{nt}^*(\rho)H_n^*(\rho_t)], \quad t = 1, \ldots, T, \\ -\frac{(n-1)(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime}(\beta, \lambda, \rho)\widetilde{V}_{nt}^*(\beta, \lambda, \rho), \end{cases} \tag{2.37}$$

where $R_{nt}^*(\rho) = I_{n-1} - \mathbb{D}_n^{*-1}(\rho)D_{nt}^*(\rho_t)$, and $S_{nt}^*(\rho) = I_{n-1} - B_{nt}^*(\rho_t)\mathbb{D}_n^{*-1}(\rho)B_{nt}^{*\prime}(\rho_t)$.

Denote the null estimator of θ by $\tilde{\theta}_{\text{SLE2}}$. Let $J_{\text{SLE2}}(\theta) = -\frac{\partial}{\partial\theta'}S^\star_{\text{SLE2}}(\theta)$, $I_{\text{SLE2}}(\theta_0) = \text{E}[J_{\text{SLE2}}(\theta_0)]$ and $\Sigma_{\text{SLE2}}(\theta_0) = \text{Var}[S^\star_{\text{SLE2}}(\theta_0)]$ with their expressions given in Appendix A.2.4. The robust AQS test, taking into account of estimation of fixed effects, has the forms:

$$T^{(r)}_{\text{SLE2}} = \widetilde{S}^{\star\prime}_{\text{SLE2}}\widetilde{I}^{-1}_{\text{SLE2}}C'\left(C\widetilde{I}^{-1}_{\text{SLE2}}\widetilde{\Sigma}_{\text{SLE2}}\widetilde{I}^{-1}_{\text{SLE2}}C'\right)^{-1}C\widetilde{I}^{-1}_{\text{SLE2}}\widetilde{S}^\star_{\text{SLE2}}, \tag{2.38}$$

where $\widetilde{S}^\star_{\text{SLE2}} = S^\star_{\text{SLE2}}(\tilde{\theta}_{\text{SLE2}})$, $\widetilde{I}_{\text{SLE2}} = I_{\text{SLE2}}(\tilde{\theta}_{\text{SLE2}})$, $\widetilde{\Sigma}_{\text{SLE2}} = \Sigma_{\text{SLE2}}(\tilde{\theta}_{\text{SLE2}})$, and the linear contrast matrix C has the same form as that for the 1FE panel SLE model. Similarly, when $I_{\text{SLE2}}(\theta_0) \asymp \Sigma_{\text{SLE2}}(\theta_0)$, $T^{(r)}_{\text{SLE2}}$ reduces to the *naïve* test: $T_{\text{SLE2}} = \widetilde{S}^{\star\prime}_{\text{SLE2}}J^{-1}_{\text{SLE2}}(\tilde{\theta}_{\text{SLE2}})\widetilde{S}^\star_{\text{SLE2}}$.

Let $\Xi_{\text{SLE2}}(\theta)$ be defined similarly as $\Xi_{\text{SLE1}}(\theta)$ for the 1FE panel SLE model.

Theorem 2.4 *Under Assumptions 1-2, 3′, and 4-5, if (i) $\tilde{\theta}_{\text{SLE2}}$ is $\sqrt{N_0}$-consistent for θ_0 under H_0^{TH}, and (ii) $I_{\text{SLE2}}(\theta)$ and $\Xi_{\text{SLE2}}(\theta)$ are positive definite for θ in a neighborhood of θ_0 when N_0 is large enough, then we have, under H_0^{TH}, $T^{(r)}_{\text{SLE2}} \xrightarrow{D} \chi^2_{k_p}$, as $n \to \infty$.*

The d.f. k_p associated with these tests remain the same as that in Theorem 2.3. Similarly, it can be shown that T_{SLE2} is not an asymptotic pivotal quantity, and that $(T^{(r)}_{\text{SLE2}} -$

$k_p)/\sqrt{2k_p} \xrightarrow{D} N(0,1)$, as $n/\sqrt{T} \to \infty$.

Estimation of the null model. Again, the general LM procedure can be adapted to estimated a null (panel SLE-2FE) model based on the AQS function $S^{\star}_{\text{SLE2}}(\theta)$ and a properly specified linear contrast matrix C. To estimate the null model specified by H_0^{TH}, the constrained estimate of c_n given (β, λ) becomes $\tilde{c}_n^{\circ *}(\beta, \lambda) = A_n^*(\lambda)\bar{Y}_n^* - \bar{X}_n^*\beta$ where $\bar{Y}_n^*$ and $\bar{X}_n^*$ are the averages of $\{Y_{nt}^*\}$ and $\{X_{nt}^*\}$, respectively. Along the same line leading to (2.37), one can easily show that AQS function of Model (2.35) at H_0^{TH} takes the form:

$$S^{\circ *}_{\text{SLE2}}(\theta) = \begin{cases} \frac{1}{\sigma^2}\sum_{t=1}^T X_{nt}^{\circ *\prime} B_n^{*\prime}(\rho)\widetilde{V}_{nt}^{\circ *}(\beta, \lambda, \rho), \\ \frac{1}{\sigma^2}\sum_{t=1}^T (W_n^* Y_{nt}^{\circ *})' B_n^{*\prime}(\rho)\widetilde{V}_{nt}^{\circ *}(\beta, \lambda, \rho) - (T-1)\text{tr}[G_n^*(\lambda)], \\ \frac{1}{\sigma^2}\sum_{t=1}^T \widetilde{V}_{nt}^{\circ *\prime}(\beta, \lambda, \rho) H_n^*(\rho)\widetilde{V}_{nt}^{\circ *}(\beta, \lambda, \rho) - (T-1)\text{tr}[H_n^*(\lambda)], \\ -\frac{(n-1)(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}_{nt}^{\circ *\prime}(\beta, \lambda, \rho)\widetilde{V}_{nt}^{\circ *}(\beta, \lambda, \rho), \end{cases} \tag{2.39}$$

$\widetilde{V}_{nt}^{\circ *}(\beta, \lambda, \rho) = B_n^*(\rho)[A_n^*(\lambda)Y_{nt}^* - X_{nt}^*\beta - \tilde{c}_n^{\circ *}(\beta, \lambda)] = B_n^*(\rho)[A_n^*(\lambda)Y_{nt}^{\circ *} - X_{nt}^{\circ *}\beta]$, where $Y_{nt}^{\circ *} = Y_{nt}^* - \bar{Y}_n^*$ and $X_{nt}^{\circ *} = X_{nt}^* - \bar{X}_n^*$. Solving the estimating equations, $S^{\circ *}_{\text{SLE2}}(\theta) = 0$, gives the null estimator $\tilde{\theta}_{\text{SLE2}}$ of $\theta = (\beta', \lambda, \rho, \sigma^2)'$, which is shown to be asymptotically equivalent to the transformation-based estimator of Lee and Yu (2010). Thus, $\tilde{\theta}_{\text{SLE2}}$ is $\sqrt{(n-1)(T-1)}$-consistent for θ. Estimation of c_n, γ and κ proceeds similarly.

A special submodel is the **panel SLE model homogeneous ρ-coefficients**. With two-way FE, the AQS function of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \rho, \sigma^2)'$ is obtained by simplifying $S^{\star}_{\text{SLE2}}(\boldsymbol{\theta})$:

$$S^{\star 0}_{\text{SLE2}}(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2} X_{nt}^{*\prime} B_n^{*\prime}(\rho)\widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho), \quad t = 1, \dots, T, \\ \frac{1}{\sigma^2}(W_n^* Y_{nt}^*)' B_n^{*\prime}(\rho)\widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) - \frac{T-1}{T}\text{tr}[G_n^*(\lambda_t)], \quad t = 1, \dots, T, \\ \frac{1}{\sigma^2}\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) H_n^*(\rho)\widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) - (T-1)\text{tr}[H_n^*(\rho)], \\ -\frac{(n-1)(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)\widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho). \end{cases} \tag{2.40}$$

This provides a channel for carrying out various conditional tests, given the temporal homogeneity in ρ. Necessary details for constructing these tests are provided in Appendix B.5., and these can easily be simplified to give AQS tests for the 1FE model.

Finally, a very special submodel, the SPD model with spatial errors (SE), is also briefly discussed here as it parallels with the panel SL models popular in practical applications.

The AQS function of $\theta = (\boldsymbol{\beta}', \boldsymbol{\rho}', \sigma^2)'$ of the **panel SE model with 2FE** takes the form:

$$S^{\star}_{\texttt{SE2}}(\theta) = \begin{cases} \frac{1}{\sigma^2} X^{*\prime}_{nt} B^{*\prime}_n(\rho_t) \widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\rho}), \quad t = 1, \ldots, T, \\ \frac{1}{\sigma^2} \widetilde{V}^{*\prime}_{nt}(\boldsymbol{\beta}, \boldsymbol{\rho}) H^*_n(\rho_t) \widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\rho}) - \mathrm{tr}[S^*_{nt}(\boldsymbol{\rho}) H^*_n(\rho_t)], \quad t = 1, \ldots, T, \\ -\frac{(n-1)(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^T \widetilde{V}^{*\prime}_{nt}(\boldsymbol{\beta}, \boldsymbol{\rho}) \widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\rho}), \end{cases} \tag{2.41}$$

where $\widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\rho}) = B^*_n(\rho_t) U^{\circ *}_{nt}(\beta_t) - B^*_n(\rho_t) \mathbb{D}^{*-1}_n(\boldsymbol{\rho}) \sum_{s=1}^T D^*_n(\rho_s) U^{\circ *}_{ns}(\beta_s)$, and $U^{\circ *}_{nt}(\beta_t) = Y^*_{nt} - X^*_{nt}\beta_t$. This can be used to perform tests concerning $\{\beta_t\}$ and $\{\rho_t\}$ in the panel SE model with 2FE. The necessary detail for constructing these tests are given in Appendix A.2.6, which can easily be simplified to give the AQS tests for panel SE model with 1FE.

2.4 Monte Carlo Study

Extensive Monte Carlo experiments are conducted to investigate the finite sample performance of the proposed tests, based on the following four data generation processes (DGPs), the SPD models with, respectively, 1FE-SL, 2FE-SL, 1FE-SLE and 2FE-SLE:

$$\begin{aligned} \texttt{DGP1}: \ & Y_{nt} = \lambda_{t0} W_n Y_{nt} + X_{1nt}\beta_{1t0} + X_{2nt}\beta_{2t0} + c_{n0} + V_{nt}, \ t = 1, 2, \ldots, T, \\ \texttt{DGP2}: \ & Y_{nt} = \lambda_{t0} W_n Y_{nt} + X_{1nt}\beta_{1t0} + X_{2nt}\beta_{2t0} + c_{n0} + \alpha_{t0} 1_n + V_{nt}, \ t = 1, 2, \ldots, T. \\ \texttt{DGP3}: \ & Y_{nt} = \lambda_{t0} W_n Y_{nt} + X_{1nt}\beta_{1t0} + X_{2nt}\beta_{2t0} + c_{n0} + U_{nt}, \\ & \qquad U_{nt} = \rho_{t0} M_n U_{nt} + V_{nt}, \ t = 1, 2, \ldots, T. \\ \texttt{DGP4}: \ & Y_{nt} = \lambda_{t0} W_n Y_{nt} + X_{1nt}\beta_{1t0} + X_{2nt}\beta_{2t0} + c_{n0} + \alpha_{t0} 1_n + U_{nt}, \\ & \qquad U_{nt} = \rho_{t0} M_n U_{nt} + V_{nt}, \ t = 1, 2, \ldots, T. \end{aligned}$$

We concentrate on the tests of temporal homogeneity. In all the Monte Carlo experiments for simulating the empirical sizes of the tests, $\beta_t = (\beta_{1t}, \beta_{2t})' = (1, 1)'$, $\lambda_t \in \{0.5, 0, -0.5\}$, and $\rho_t \in \{0.5, 0, -0.5\}$ for all $t = 1, \ldots, T$, $\sigma_0^2 = 1$, $n \in \{50, 100, 200, 500\}$, and $T = \{3, 6\}$. Each set of Monte Carlo results is based on 10,000 Monte Carlo samples for the two SL models, and 5,000 for the two SLE models.

The **weight matrices** are generated based on three different methods: (i) `Rook Contiguity`, (ii) `Queen Contiguity`, and (iii) `Group Interaction`, with details given in Yang (2015a). In spatial layouts (i)-(ii), the degree of spatial interactions (number of neighbors each unit has) is fixed, while in (iii) it may grow with the sample size. This is attained by allowing the number of groups, G, in the sample of spatial units to be directly related to the sample size n, e.g., $G = n^{0.5}$. Hence, the average group size, $m = n/G$, gives a measure of the degree of spatial dependence among the n spatial units. The actual sizes of the groups are generated from a discrete uniform distribution from $.5m$ to $1.5m$.

The two **exogenous regressors** are generated according to `REG1:` $X_{jnt} \overset{iid}{\sim} N(0, I_n)$ for $j = 1, 2$ and $t = 1, \ldots, T$; and `REG2:` the *i*th value of the *j*th regressor in the *g*th group is such that $X_{jt,ig} \overset{iid}{\sim} (2z_g + z_{ig})/\sqrt{10}$, where $(z_g, z_{ig}) \overset{iid}{\sim} N(0, 1)$ when group interaction scheme is followed; $\{X_{jt,ig}\}$ are thus independent across *j* and *t*, but not across i.

The **errors**, $v_{it} = \sigma_0 e_{it}$, are generated according to `err1:` $\{e_{it}\}$ are iid standard normal; `err2:` $\{e_{it}\}$ are iid normal mixture with 10% of values from $N(0, 4)$ and the remaining from $N(0, 1)$, standardized to have mean 0 and variance 1; and `err3:` $\{e_{it}\}$ iid log-normal (i.e., $\log e_{it} \overset{iid}{\sim} N(0, 1)$) standardized to have mean 0 and variance 1.

Partial Monte Carlo results are reported in Tables 2.1 & 2.2 for the panel SL models, and Tables 2.3 & 2.4 for the panel SLE models. The results in Tables 1 & 2 show the following.

(i) The proposed robust test performs very well in general with empirical coverage probabilities all very close to their nominal levels, except that in cases of heavy spatial dependence (`Group Interaction`) and not-so-large n, it can be slightly undersized. As sample size increases, the empirical sizes quickly converge to their nominal levels.

(ii) In contrast, the naïve test can perform quite badly, with empirical sizes being as high as 35% for tests of 10% nominal level, when the erorrs are fairly non-normal (e.g., log-normal). It is interesting to note that the size distortions for the naïve tests also drop as sample size increase.

(iii) A larger T seems lead to a worsened performance for the naïve tests under `Queen Contiguity` but not under `Group Interaction`.

(iv) The finite sample performance of the tests for 1FE panel SL model do not seem to differ much from those for 2FE panel SL model.

From the results for the panel SLE model, reported (in Tables 2.3 & 2.4) and unreported (available from the authors upon request), similar patterns are observed for the finite sample performance of the proposed tests. In summary, the proposed robust tests are reliable and easy to apply, and hence are recommended for the applied researchers. The Monte Carlo experiments for the power of the tests, and the size and power of the other tests, e.g., tests for change points, are also carried out, and the results (available from the authors upon request) show similar patterns.

Table 2.1a. Empirical Sizes of Tests for Temporal Homogeneity in Panel SL Model
One-Way Fixed Effects, `Queen Contiguity`

		$T=3$						$T=6$					
λ	n	T_{SL1}			$T_{\text{SL1}}^{(r)}$			T_{SL1}			$T_{\text{SL1}}^{(r)}$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
		Normal Error											
.5	50	.208	.135	.052	.096	.045	.007	.216	.138	.050	.095	.044	.008
	100	.150	.086	.024	.098	.046	.009	.161	.097	.028	.103	.050	.009
	200	.128	.068	.015	.103	.049	.008	.129	.069	.018	.099	.051	.010
	500	.107	.054	.010	.097	.046	.007	.110	.054	.011	.098	.049	.009
0	50	.204	.135	.053	.102	.048	.008	.214	.137	.050	.095	.046	.009
	100	.147	.086	.025	.099	.048	.008	.160	.096	.027	.105	.051	.009
	200	.127	.069	.015	.104	.049	.009	.127	.068	.018	.100	.049	.010
	500	.111	.056	.011	.100	.048	.008	.109	.056	.012	.099	.050	.010
-.5	50	.204	.133	.055	.102	.048	.008	.212	.136	.051	.097	.046	.009
	100	.147	.086	.025	.099	.049	.008	.160	.097	.027	.103	.050	.009
	200	.129	.068	.015	.103	.048	.009	.127	.070	.017	.100	.050	.010
	500	.108	.055	.012	.101	.048	.009	.110	.056	.012	.100	.049	.010
		Normal Mixture Error											
.5	50	.201	.129	.053	.096	.047	.006	.229	.154	.061	.121	.070	.023
	100	.149	.088	.027	.100	.048	.009	.163	.096	.029	.099	.050	.010
	200	.130	.073	.019	.105	.052	.011	.133	.073	.018	.103	.054	.010
	500	.112	.058	.012	.102	.051	.009	.118	.061	.012	.102	.051	.010
0	50	.197	.126	.052	.099	.047	.007	.229	.150	.061	.103	.053	.011
	100	.149	.087	.028	.102	.049	.010	.161	.094	.029	.099	.048	.010
	200	.129	.073	.019	.105	.052	.010	.132	.073	.018	.104	.054	.011
	500	.111	.059	.012	.103	.051	.010	.120	.061	.012	.102	.053	.009
-.5	50	.193	.129	.052	.097	.048	.008	.231	.151	.062	.103	.053	.012
	100	.150	.088	.028	.101	.050	.010	.162	.094	.030	.101	.050	.010
	200	.130	.073	.019	.104	.052	.011	.132	.073	.018	.103	.053	.011
	500	.113	.059	.013	.102	.051	.010	.118	.062	.013	.101	.052	.010
		Log-normal Error											
.5	50	.180	.119	.045	.089	.043	.008	.211	.145	.060	.100	.054	.017
	100	.149	.087	.027	.097	.047	.009	.164	.102	.032	.101	.057	.012
	200	.133	.071	.018	.097	.045	.009	.147	.087	.030	.101	.055	.014
	500	.127	.071	.018	.100	.051	.011	.142	.078	.030	.101	.050	.011
0	50	.180	.118	.046	.093	.044	.008	.193	.130	.056	.099	.054	.015
	100	.132	.078	.023	.094	.047	.009	.146	.086	.024	.100	.052	.010
	200	.109	.057	.013	.089	.042	.008	.114	.064	.017	.094	.051	.012
	500	.099	.052	.012	.010	.050	.010	.110	.058	.013	.102	.053	.011
-.5	50	.194	.128	.049	.097	.045	.008	.225	.154	.072	.106	.058	.016
	100	.142	.083	.024	.096	.047	.010	.191	.118	.042	.104	.057	.013
	200	.120	.067	.017	.095	.046	.009	.166	.102	.032	.102	.054	.012
	500	.118	.065	.016	.098	.050	.011	.151	.102	.032	.102	.050	.010

Table 2.1b. Empirical Sizes of Tests for Temporal Homogeneity in Panel SL Model
One-Way Fixed Effects, `Group Interaction`

		$T=3$						$T=6$					
λ	n	T_{SL1}			$T_{\text{SL1}}^{(r)}$			T_{SL1}			$T_{\text{SL1}}^{(r)}$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
Normal Error													
.5	50	.222	.144	.057	.086	.034	.004	.219	.136	.048	.085	.039	.007
	100	.150	.089	.025	.088	.039	.006	.165	.094	.028	.089	.042	.007
	200	.124	.067	.018	.092	.042	.008	.128	.070	.016	.094	.045	.008
	500	.110	.059	.014	.097	.049	.011	.113	.057	.012	.095	.048	.009
0	50	.232	.157	.065	.087	.036	.005	.232	.151	.056	.084	.040	.007
	100	.155	.091	.027	.089	.040	.006	.173	.099	.030	.091	.044	.008
	200	.124	.068	.020	.090	.042	.008	.131	.071	.016	.095	.044	.008
	500	.110	.060	.015	.098	.049	.010	.114	.058	.013	.096	.048	.009
-.5	50	.238	.163	.071	.086	.038	.004	.239	.159	.063	.085	.038	.007
	100	.157	.092	.029	.088	.040	.005	.178	.102	.033	.089	.043	.008
	200	.126	.069	.020	.091	.043	.008	.133	.072	.016	.096	.043	.008
	500	.111	.061	.014	.098	.049	.010	.115	.059	.012	.096	.048	.009
Normal Mixture Error													
.5	50	.230	.151	.056	.087	.033	.004	.215	.143	.051	.088	.046	.009
	100	.154	.088	.025	.087	.041	.006	.165	.094	.025	.087	.041	.009
	200	.131	.070	.017	.095	.043	.008	.133	.071	.018	.093	.043	.009
	500	.114	.061	.013	.100	.048	.009	.116	.059	.011	.096	.048	.008
0	50	.241	.163	.068	.088	.036	.005	.231	.155	.061	.088	.046	.008
	100	.157	.092	.029	.089	.041	.006	.170	.098	.029	.089	.041	.008
	200	.133	.070	.018	.095	.044	.008	.133	.072	.019	.094	.042	.009
	500	.114	.059	.014	.099	.048	.010	.133	.072	.019	.094	.042	.009
-.5	50	.259	.181	.081	.093	.043	.007	.270	.186	.083	.096	.050	.010
	100	.168	.103	.033	.096	.046	.007	.193	.118	.040	.093	.046	.010
	200	.136	.075	.020	.097	.045	.009	.142	.079	.023	.094	.045	.010
	500	.116	.060	.015	.098	.048	.009	.117	.059	.012	.097	.048	.008
Log-normal Error													
.5	50	.218	.143	.054	.081	.035	.005	.206	.137	.050	.079	.040	.009
	100	.151	.088	.026	.084	.037	.005	.176	.107	.034	.091	.048	.012
	200	.130	.069	.018	.091	.043	.006	.142	.081	.022	.095	.051	.012
	500	.108	.057	.012	.094	.045	.008	.126	.066	.016	.101	.049	.010
0	50	.227	.151	.064	.084	.036	.006	.243	.166	.075	.087	.045	.010
	100	.152	.091	.029	.088	.040	.006	.185	.122	.046	.097	.049	.013
	200	.137	.077	.019	.096	.047	.008	.136	.078	.025	.097	.052	.011
	500	.107	.059	.014	.098	.048	.009	.115	.057	.014	.098	.048	.010
-.5	50	.263	.188	.086	.093	.043	.008	.350	.259	.139	.106	.057	.015
	100	.179	.114	.042	.101	.049	.010	.260	.186	.090	.105	.054	.014
	200	.161	.096	.029	.107	.056	.010	.185	.114	.043	.103	.052	.013
	500	.123	.067	.018	.100	.051	.010	.131	.072	.021	.101	.051	.010

Table 2.2a. Empirical Sizes of Tests for Temporal Homogeneity in Panel SL Model
Two-Way Fixed Effects, `Queen Contiguity`

		$T=3$						$T=6$					
λ	n	T_{SL2}			$T_{\text{SL2}}^{(r)}$			T_{SL2}			$T_{\text{SL2}}^{(r)}$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
		Normal Error											
.5	50	.192	.123	.047	.093	.045	.007	.228	.148	.059	.100	.050	.010
	100	.140	.080	.023	.096	.048	.009	.157	.094	.029	.102	.050	.011
	200	.120	.064	.015	.098	.049	.009	.128	.068	.017	.101	.052	.009
	500	.103	.051	.013	.098	.048	.011	.105	.056	.012	.095	.049	.010
0	50	.194	.123	.048	.094	.046	.008	.224	.147	.059	.101	.049	.010
	100	.138	.082	.023	.095	.050	.009	.126	.069	.017	.099	.051	.009
	200	.115	.064	.016	.096	.049	.009	.157	.095	.027	.101	.049	.010
	500	.101	.052	.012	.098	.048	.009	.126	.069	.017	.099	.051	.009
-.5	50	.192	.123	.047	.093	.045	.009	.225	.148	.058	.100	.049	.009
	100	.138	.081	.023	.096	.049	.008	.157	.092	.027	.101	.048	.010
	200	.116	.063	.015	.096	.049	.009	.125	.069	.016	.102	.050	.009
	500	.105	.055	.011	.096	.048	.009	.108	.056	.013	.097	.051	.011
		Normal Mixture Error											
.5	50	.198	.131	.052	.100	.048	.008	.232	.155	.063	.106	.054	.013
	100	.140	.080	.025	.096	.047	.010	.165	.100	.030	.107	.055	.012
	200	.124	.067	.016	.101	.051	.009	.132	.071	.019	.104	.051	.013
	500	.110	.055	.013	.100	.050	.010	.106	.056	.012	.097	.051	.010
0	50	.199	.132	.052	.102	.048	.009	.234	.154	.064	.110	.055	.013
	100	.139	.080	.024	.097	.047	.009	.166	.100	.031	.109	.054	.011
	200	.124	.067	.017	.102	.051	.010	.129	.072	.019	.102	.051	.013
	500	.110	.055	.012	.102	.050	.010	.106	.055	.013	.096	.049	.010
-.5	50	.199	.130	.053	.101	.049	.009	.234	.157	.066	.112	.057	.013
	100	.143	.084	.025	.101	.048	.009	.164	.097	.031	.107	.053	.012
	200	.123	.069	.016	.103	.051	.010	.133	.073	.020	.105	.053	.012
	500	.109	.056	.012	.101	.050	.009	.107	.056	.014	.096	.048	.012
		Log-normal Error											
.5	50	.196	.131	.055	.100	.050	.009	.242	.171	.079	.107	.067	.018
	100	.139	.081	.027	.095	.050	.011	.171	.112	.041	.105	.055	.015
	200	.128	.070	.018	.106	.053	.010	.141	.081	.026	.104	.052	.013
	500	.109	.060	.014	.101	.052	.011	.123	.068	.019	.101	.051	.010
0	50	.196	.133	.059	.106	.055	.010	.239	.167	.081	.110	.055	.021
	100	.137	.078	.024	.095	.048	.010	.166	.110	.039	.107	.054	.018
	200	.126	.070	.018	.104	.052	.010	.133	.079	.025	.105	.049	.015
	500	.107	.056	.013	.100	.051	.010	.116	.061	.016	.102	.051	.013
-.5	50	.205	.141	.066	.112	.062	.011	.249	.177	.083	.108	.055	.026
	100	.154	.089	.028	.106	.052	.012	.172	.110	.042	.099	.048	.019
	200	.129	.074	.019	.107	.056	.012	.145	.088	.030	.098	.049	.020
	500	.110	.058	.014	.103	.052	.010	.122	.068	.018	.100	.049	.014

Table 2.2b. Empirical Sizes of Tests for Temporal Homogeneity in Panel SL Model
Two-Way Fixed Effects, `Group Interaction`

		$T=3$						$T=6$					
λ	n	T_{SL2}			$T^{(r)}_{\text{SL2}}$			T_{SL2}			$T^{(r)}_{\text{SL2}}$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
		Normal Error											
.5	50	.226	.148	.059	.086	.038	.005	.223	.142	.052	.087	.040	.007
	100	.155	.090	.025	.090	.036	.006	.166	.095	.029	.089	.043	.007
	200	.124	.070	.018	.091	.044	.006	.131	.073	.016	.093	.045	.008
	500	.112	.060	.015	.097	.050	.010	.114	.057	.013	.096	.047	.010
0	50	.240	.159	.068	.088	.039	.005	.237	.154	.059	.086	.040	.007
	100	.159	.094	.025	.090	.037	.006	.174	.102	.031	.088	.042	.007
	200	.127	.072	.018	.091	.044	.007	.133	.074	.016	.094	.046	.008
	500	.112	.060	.014	.097	.050	.010	.116	.059	.013	.097	.046	.010
-.5	50	.244	.167	.075	.088	.039	.005	.249	.164	.065	.086	.040	.007
	100	.163	.096	.028	.089	.038	.006	.179	.104	.033	.085	.043	.007
	200	.127	.073	.019	.092	.045	.007	.134	.076	.017	.094	.045	.008
	500	.113	.059	.014	.098	.049	.010	.117	.059	.013	.097	.046	.010
		Normal Mixture Error											
.5	50	.232	.150	.058	.080	.034	.005	.222	.144	.055	.082	.041	.008
	100	.159	.090	.024	.088	.039	.006	.164	.095	.027	.083	.041	.008
	200	.130	.072	.018	.095	.045	.007	.133	.071	.017	.089	.043	.010
	500	.114	.059	.014	.097	.048	.009	.118	.060	.012	.098	.047	.009
0	50	.245	.167	.069	.085	.038	.006	.247	.165	.071	.083	.039	.007
	100	.164	.098	.027	.089	.040	.006	.175	.103	.032	.080	.038	.007
	200	.131	.072	.018	.094	.043	.008	.132	.072	.019	.089	.041	.009
	500	.115	.059	.014	.096	.048	.009	.119	.060	.012	.096	.047	.009
-.5	50	.269	.185	.085	.097	.047	.009	.298	.209	.100	.101	.052	.012
	100	.177	.110	.035	.099	.046	.007	.205	.127	.045	.094	.046	.008
	200	.138	.077	.020	.096	.045	.008	.145	.082	.023	.095	.045	.010
	500	.115	.059	.014	.096	.047	.009	.122	.063	.013	.099	.049	.009
		Log-normal Error											
.5	50	.217	.143	.057	.078	.036	.005	.215	.142	.055	.076	.036	.008
	100	.152	.088	.025	.079	.034	.005	.176	.111	.036	.082	.041	.009
	200	.132	.073	.018	.089	.044	.006	.141	.080	.023	.088	.046	.010
	500	.113	.057	.013	.094	.047	.008	.119	.062	.014	.096	.048	.009
0	50	.240	.165	.073	.085	.040	.006	.246	.174	.079	.085	.038	.008
	100	.164	.099	.034	.086	.041	.006	.191	.129	.051	.091	.040	.008
	200	.135	.076	.020	.092	.043	.007	.143	.083	.027	.095	.044	.009
	500	.111	.057	.014	.092	.045	.008	.113	.060	.013	.097	.045	.010
-.5	50	.287	.207	.104	.112	.060	.013	.347	.269	.151	.119	.068	.022
	100	.201	.131	.054	.109	.057	.012	.270	.195	.099	.119	.065	.019
	200	.156	.095	.028	.105	.054	.010	.191	.122	.049	.105	.056	.014
	500	.120	.067	.017	.098	.050	.009	.141	.081	.021	.103	.052	.010

Table 2.3a. Empirical Sizes of Tests for Temporal Homogeneity in Panel SLE Model
One-Way Fixed Effects, `Queen Contiguity`, $\lambda = 0.5$.

		$T=3$						$T=6$					
ρ	n	T_{SLE1}			$T^{(r)}_{\text{SLE1}}$			T_{SLE1}			$T^{(r)}_{\text{SLE1}}$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
Normal Error													
.5	50	.199	.142	.075	.082	.039	.005	.161	.099	.036	.090	.042	.011
	100	.123	.068	.025	.094	.043	.009	.097	.050	.012	.092	.043	.006
	200	.084	.044	.009	.099	.046	.007	.079	.038	.008	.102	.049	.011
	500	.070	.034	.006	.104	.049	.009	.064	.030	.005	.102	.054	.009
0	50	.223	.164	.093	.090	.042	.006	.171	.104	.041	.093	.047	.010
	100	.132	.076	.029	.095	.046	.012	.105	.058	.014	.097	.047	.007
	200	.087	.046	.011	.103	.050	.010	.082	.039	.008	.104	.050	.011
	500	.069	.036	.006	.102	.050	.011	.063	.028	.005	.103	.054	.010
-.5	50	.232	.174	.098	.093	.042	.006	.181	.120	.048	.096	.047	.010
	100	.134	.083	.033	.097	.045	.011	.118	.064	.014	.098	.048	.008
	200	.097	.047	.013	.105	.050	.012	.079	.039	.008	.102	.052	.011
	500	.070	.035	.006	.102	.052	.009	.061	.028	.005	.102	.049	.011
Normal Mixture Error													
.5	50	.196	.139	.072	.081	.037	.004	.168	.106	.044	.092	.047	.008
	100	.121	.070	.025	.087	.040	.008	.107	.057	.017	.096	.053	.012
	200	.084	.043	.011	.092	.046	.006	.082	.044	.010	.101	.052	.013
	500	.071	.035	.008	.099	.052	.012	.070	.036	.009	.097	.046	.014
0	50	.212	.151	.080	.087	.042	.005	.167	.110	.044	.089	.045	.010
	100	.131	.076	.028	.089	.041	.009	.105	.054	.015	.097	.046	.011
	200	.085	.046	.011	.095	.046	.008	.078	.039	.009	.100	.047	.012
	500	.071	.036	.007	.097	.050	.010	.064	.032	.006	.104	.054	.012
-.5	50	.226	.164	.090	.093	.040	.006	.197	.131	.057	.104	.056	.013
	100	.140	.083	.030	.094	.043	.009	.126	.073	.023	.104	.055	.013
	200	.094	.050	.013	.102	.051	.010	.086	.048	.013	.103	.055	.014
	500	.073	.038	.009	.101	.051	.012	.074	.034	.005	.102	.055	.011
Log-normal Error													
.5	50	.150	.102	.046	.083	.038	.006	.169	.108	.044	.092	.048	.010
	100	.115	.075	.035	.091	.044	.010	.106	.058	.015	.098	.051	.010
	200	.109	.067	.027	.095	.046	.009	.073	.036	.008	.090	.046	.010
	500	.089	.050	.016	.100	.049	.011	.064	.032	.006	.104	.052	.012
0	50	.217	.160	.090	.082	.041	.009	.179	.118	.045	.092	.048	.011
	100	.126	.077	.031	.087	.042	.009	.108	.062	.017	.100	.055	.008
	200	.101	.055	.015	.103	.048	.010	.074	.035	.007	.095	.044	.008
	500	.071	.035	.008	.096	.048	.010	.059	.031	.006	.099	.050	.011
-.5	50	.192	.138	.069	.090	.045	.006	.202	.136	.054	.098	.050	.011
	100	.137	.087	.038	.092	.048	.010	.128	.074	.019	.108	.057	.010
	200	.094	.045	.014	.101	.048	.011	.081	.041	.008	.099	.049	.009
	500	.078	.040	.010	.102	.051	.010	.064	.030	.005	.105	.050	.012

Table 2.3b. Empirical Sizes of Tests for Temporal Homogeneity in Panel SLE Model
One-Way Fixed Effects, `Queen Contiguity`, $\lambda = -0.5$.

		$T=3$						$T=6$					
ρ	n	$T_{\texttt{SLE1}}$			$T_{\texttt{SLE1}}^{(r)}$			$T_{\texttt{SLE1}}$			$T_{\texttt{SLE1}}^{(r)}$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
						Normal Error							
.5	50	.190	.131	.058	.088	.037	.007	.167	.102	.036	.088	.042	.010
	100	.116	.068	.022	.093	.044	.009	.098	.050	.013	.091	.044	.007
	200	.079	.042	.009	.094	.046	.007	.078	.040	.010	.100	.050	.012
	500	.071	.033	.007	.101	.050	.009	.060	.029	.005	.102	.053	.009
0	50	.209	.149	.073	.091	.040	.006	.169	.104	.040	.094	.043	.010
	100	.125	.073	.027	.099	.050	.011	.102	.056	.013	.093	.047	.006
	200	.084	.043	.010	.098	.048	.010	.079	.040	.008	.104	.051	.010
	500	.072	.033	.007	.103	.050	.011	.059	.029	.005	.096	.054	.010
-.5	50	.225	.162	.085	.095	.040	.006	.172	.111	.044	.094	.043	.010
	100	.131	.081	.031	.101	.050	.011	.109	.059	.013	.099	.047	.008
	200	.089	.044	.013	.105	.049	.011	.082	.039	.009	.104	.052	.010
	500	.069	.032	.007	.100	.049	.010	.057	.030	.005	.096	.049	.012
						Normal Mixture Error							
.5	50	.187	.129	.061	.079	.034	.004	.176	.111	.043	.092	.047	.008
	100	.111	.068	.022	.086	.042	.008	.105	.054	.016	.097	.051	.013
	200	.083	.044	.009	.091	.047	.006	.085	.046	.010	.102	.056	.012
	500	.072	.033	.008	.102	.049	.011	.074	.036	.008	.099	.053	.010
0	50	.200	.140	.071	.086	.039	.006	.166	.105	.041	.090	.047	.010
	100	.126	.074	.027	.092	.042	.009	.103	.056	.016	.095	.049	.011
	200	.079	.045	.009	.095	.047	.008	.076	.041	.010	.098	.050	.012
	500	.071	.035	.008	.100	.049	.010	.064	.031	.007	.101	.050	.012
-.5	50	.218	.156	.080	.088	.041	.007	.191	.124	.052	.100	.054	.013
	100	.136	.079	.031	.096	.045	.008	.119	.068	.021	.105	.055	.013
	200	.087	.048	.013	.098	.048	.009	.088	.048	.014	.106	.057	.014
	500	.073	.037	.009	.103	.053	.011	.075	.034	.007	.104	.053	.011
						Log-normal Error							
.5	50	.175	.125	.063	.084	.036	.009	.174	.110	.043	.092	.046	.010
	100	.138	.087	.038	.089	.042	.010	.099	.055	.016	.098	.050	.011
	200	.096	.048	.014	.096	.045	.008	.075	.037	.008	.098	.046	.011
	500	.075	.038	.009	.101	.052	.011	.066	.028	.006	.100	.053	.013
0	50	.207	.145	.081	.086	.042	.011	.173	.111	.044	.093	.046	.010
	100	.122	.078	.029	.090	.044	.009	.105	.056	.013	.096	.048	.009
	200	.091	.047	.010	.095	.047	.008	.076	.037	.007	.099	.046	.009
	500	.071	.035	.008	.099	.049	.011	.057	.027	.006	.101	.047	.011
-.5	50	.201	.138	.072	.093	.043	.008	.191	.125	.051	.097	.049	.012
	100	.141	.092	.039	.096	.048	.010	.118	.067	.017	.104	.053	.010
	200	.089	.045	.012	.104	.050	.009	.084	.041	.008	.104	.051	.010
	500	.072	.034	.007	.104	.049	.010	.062	.029	.006	.103	.046	.012

Table 2.4a. Empirical Sizes of Tests for Temporal Homogeneity in Panel SLE Model
Two-Way Fixed Effects, `Queen Contiguity`, $\lambda = 0.5$.

		$T=3$						$T=6$					
ρ	n	T_{SLE2}			$T^{(r)}_{\text{SLE2}}$			T_{SLE2}			$T^{(r)}_{\text{SLE2}}$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
		Normal Error											
.5	50	.235	.181	.105	.083	.038	.006	.310	.226	.115	.087	.044	.008
	100	.212	.151	.086	.093	.045	.008	.190	.111	.036	.090	.041	.007
	200	.182	.121	.054	.098	.044	.006	.139	.079	.021	.101	.049	.011
	500	.134	.073	.022	.100	.048	.010	.121	.064	.014	.102	.055	.009
0	50	.272	.208	.117	.088	.043	.007	.314	.224	.111	.094	.045	.010
	100	.217	.143	.070	.094	.046	.011	.197	.116	.036	.093	.043	.008
	200	.161	.097	.032	.100	.051	.008	.142	.083	.022	.103	.050	.011
	500	.125	.065	.017	.105	.049	.011	.119	.064	.014	.102	.053	.010
-.5	50	.302	.233	.136	.094	.042	.005	.321	.239	.114	.092	.045	.009
	100	.209	.142	.062	.095	.046	.011	.205	.128	.042	.096	.047	.009
	200	.153	.090	.029	.102	.050	.009	.151	.081	.023	.098	.052	.010
	500	.119	.064	.015	.102	.054	.009	.115	.061	.014	.103	.051	.010
		Normal Mixture Error											
.5	50	.221	.159	.090	.083	.037	.004	.315	.242	.127	.090	.044	.008
	100	.212	.154	.085	.085	.044	.008	.201	.128	.050	.097	.053	.010
	200	.183	.122	.059	.092	.046	.008	.150	.090	.029	.101	.052	.009
	500	.137	.082	.028	.100	.053	.012	.139	.079	.022	.100	.053	.010
0	50	.269	.201	.114	.089	.043	.005	.315	.235	.124	.092	.052	.012
	100	.212	.149	.075	.089	.045	.009	.189	.118	.043	.096	.047	.010
	200	.158	.098	.033	.096	.048	.008	.143	.078	.025	.099	.050	.013
	500	.121	.070	.016	.099	.050	.010	.120	.063	.016	.102	.053	.012
-.5	50	.285	.225	.137	.093	.046	.008	.380	.286	.164	.103	.056	.011
	100	.229	.161	.083	.100	.047	.010	.229	.152	.061	.108	.060	.012
	200	.166	.102	.036	.101	.053	.009	.176	.106	.034	.104	.058	.012
	500	.132	.070	.018	.106	.054	.012	.136	.075	.020	.097	.050	.010
		Log-normal Error											
.5	50	.239	.181	.105	.085	.039	.006	.314	.232	.123	.091	.043	.008
	100	.222	.154	.086	.090	.043	.007	.196	.117	.041	.095	.047	.009
	200	.185	.126	.056	.096	.047	.008	.138	.079	.020	.097	.047	.009
	500	.138	.074	.024	.102	.049	.011	.123	.064	.016	.105	.052	.010
0	50	.246	.188	.108	.085	.042	.010	.319	.235	.115	.095	.047	.011
	100	.204	.141	.074	.090	.045	.007	.194	.115	.040	.095	.051	.008
	200	.180	.114	.047	.095	.047	.009	.142	.076	.021	.095	.048	.009
	500	.129	.075	.022	.097	.048	.010	.115	.060	.014	.100	.050	.011
-.5	50	.300	.235	.146	.093	.044	.008	.344	.246	.126	.097	.050	.011
	100	.214	.145	.064	.094	.045	.010	.208	.133	.050	.101	.055	.010
	200	.156	.092	.028	.099	.046	.008	.154	.086	.023	.101	.049	.011
	500	.123	.066	.015	.104	.051	.010	.121	.061	.014	.102	.050	.010

Table 2.4b. Empirical Sizes of Tests for Temporal Homogeneity in Panel SLE Model Two-Way Fixed Effects, `Queen Contiguity`, $\lambda = -0.5$.

		$T=3$						$T=6$					
ρ	n	T_{SLE2}			$T^{(r)}_{\text{SLE2}}$			T_{SLE2}			$T^{(r)}_{\text{SLE2}}$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
							Normal Error						
.5	50	.235	.173	.105	.086	.039	.007	.313	.225	.117	.089	.044	.009
	100	.216	.158	.086	.093	.046	.009	.189	.113	.037	.088	.044	.006
	200	.180	.117	.054	.093	.047	.007	.143	.079	.023	.100	.049	.012
	500	.134	.076	.021	.103	.048	.010	.118	.062	.014	.100	.053	.010
0	50	.271	.206	.116	.089	.040	.007	.315	.226	.109	.093	.044	.009
	100	.220	.149	.072	.098	.048	.011	.197	.115	.038	.092	.047	.008
	200	.160	.096	.032	.100	.051	.009	.146	.085	.024	.104	.052	.011
	500	.127	.062	.017	.103	.049	.011	.111	.059	.015	.094	.050	.010
-.5	50	.301	.233	.130	.095	.038	.007	.325	.232	.112	.092	.044	.009
	100	.214	.146	.065	.101	.048	.011	.206	.127	.039	.096	.046	.008
	200	.158	.092	.029	.103	.050	.011	.152	.087	.022	.102	.053	.010
	500	.117	.065	.014	.100	.050	.010	.111	.057	.013	.096	.048	.011
							Normal Mixture Error						
.5	50	.220	.161	.088	.080	.035	.005	.316	.243	.129	.093	.047	.009
	100	.213	.153	.085	.088	.043	.009	.204	.129	.048	.103	.051	.012
	200	.182	.121	.059	.096	.047	.006	.153	.089	.032	.106	.058	.013
	500	.139	.083	.030	.104	.049	.010	.137	.080	.022	.101	.051	.010
0	50	.256	.194	.113	.084	.043	.006	.321	.242	.124	.093	.049	.011
	100	.214	.151	.079	.091	.046	.008	.189	.121	.042	.098	.046	.011
	200	.155	.100	.033	.097	.048	.009	.146	.079	.028	.095	.051	.013
	500	.124	.068	.018	.099	.049	.011	.118	.064	.017	.102	.053	.012
-.5	50	.279	.219	.138	.089	.043	.007	.378	.288	.162	.111	.059	.016
	100	.232	.157	.082	.097	.049	.010	.234	.151	.058	.110	.057	.013
	200	.166	.103	.035	.102	.050	.010	.170	.104	.035	.106	.052	.014
	500	.128	.072	.019	.103	.054	.011	.134	.078	.018	.098	.047	.010
							Log-normal Error						
.5	50	.230	.178	.105	.086	.039	.008	.317	.232	.125	.089	.043	.009
	100	.218	.156	.087	.093	.045	.008	.197	.116	.042	.093	.049	.009
	200	.184	.122	.055	.093	.044	.008	.143	.080	.022	.100	.047	.010
	500	.139	.077	.024	.101	.052	.010	.119	.063	.015	.102	.053	.011
0	50	.242	.184	.107	.087	.043	.011	.315	.230	.113	.095	.046	.010
	100	.202	.142	.074	.091	.043	.010	.196	.115	.039	.093	.047	.008
	200	.176	.114	.046	.098	.046	.010	.141	.082	.023	.099	.049	.010
	500	.128	.074	.024	.098	.050	.011	.110	.055	.013	.102	.050	.010
-.5	50	.298	.230	.138	.095	.042	.008	.332	.245	.127	.097	.050	.010
	100	.220	.146	.067	.100	.045	.011	.212	.129	.048	.100	.052	.010
	200	.156	.092	.029	.100	.046	.009	.154	.089	.022	.105	.051	.011
	500	.123	.065	.015	.104	.051	.009	.115	.060	.015	.100	.048	.011

2.5 Empirical Applications

The specification tests of temporal homogeneity in spatial panel data models proposed in this chapter are demonstrated in empirical settings using two well known data sets: *Public Capital Productivity* (Munnell, 1990) and *Cigarette Demand* (Baltagi and Levin, 1992). We endeavor to provide a detailed guidance to aid applied researchers in their empirical studies. First, a general discussion is given on the issues of spatial interaction and spatiotemporal heterogeneity commonly existed in economic studies.

2.5.1 Spatial interaction and spatiotemporal heterogeneity.

A wide range of empirical studies, such as urban economics, international trade, public finance, industrial organization, real estate analyses and regional economics, deal with spatial interaction. Values observed at one location depend on the values of neighboring observations at nearby locations due to budget spillovers, difference in tax rates, copycatting, network effects, et. However, this dependence may not stay the same over time. There are two major reasons for specifying, estimating, and testing for the time-varying spatial effects in the regression models. One is the growing interest in using theoretical economics that include time-varying spatial effects to analyze economic phenomenon such as externalities, group patterns and some other economic processes, for example, housing decisions, unemployment, price decisions, crime rates, trade flows, etc., which exhibit time heterogeneity patterns. The effects of relevant variables, including interactions among agents, on economic activities are changing over time. This may be due to the change of government policy, an unexpected accident, the change of the benefit from the interactions. The second driver is the need from geographic research and environmental study, where researchers usually face a large set of geocoded data when analyzing the relationships between different variables. Under this situation, due to the spatial interaction and the fact that everything in nature is changing over time, time-varying spatial autoregression model is more outstanding than many other econometric models. Adding the time-varying spatial effects in the regression model may be necessary.

One empirical problem we discuss in this paper is the U.S. *cigarette demand* in state level from 1963-1992 (Baltagi and Levin, 1992). The tax policy on cigarette differs by states, and this leads to substantial cross-state sales. Due to the government interventions (in 1965, 1967, 1971) and the reports about the health hazards of smoking (in 1983), the effects of the spatial lag, spatial error and the variables (price per pack of cigarettes, population, per capita disposable income, and etc) on the US cigarette demand might be subject to the temporal heterogeneity. The other empirical problem we discuss is the U.S. *public capital productivity* in state level from 1970-1986 (Munnell, 1990). The private production of each state may subject to spillover effects of infrastructure improvement from other states. Temporal homogeneity may be in question due to the change in policies and the change of economic environment such as 1973 oil crisis and the 1979 energy crisis. These two data sets have been extensively used in Baltagi (2013) for the illustrations of various standard panel data techniques.

Many other empirical studies have documented the existence of spatial interaction or spatial spillover effects, and these naturally raise the question whether these spillover effects as well as the economic variables effects remain constant over time due to policy change. Case (1991) studied spatial patterns in household demand. Case et al. (1993) showed that the U.S. states' budget expenditure depends on the spending of similar states. Policies have changed over the years, and one might be interested in testing if the *spatial patterns* and *budget spillovers* remain the same over time. Acemoglu et al. (2012) studied the inter-sectoral input-output linkages in the U.S. Baltagi et al. (2016) studied intra-sectoral spillovers in total factor productivity (TFP) across Chinese producers in the chemical industry using a panel data on 12,552 firms over 2004-2006, by modeling spatial spillovers in TFP through contextual effects of observable variables and the spatial dependence of the disturbances. Test of stability/homogeneity of the covariate effects as well as spatial effects may be interesting, perhaps based on extended data.

Therefore, it is highly desirable to have a general procedure to identify the possible existence of temporal heterogeneity in spatial panel data models to aid the applied researchers in their empirical studies. The AQS test we propose may serve the purpose.

We provide a detailed instruction, through two empirical applications, of how to construct AQS-tests for testing certain null hypothesis in an SPD model allowing spatiotemporal heterogeneity in the intercept (fixed effects), i.e., the model specified by (2.1), (2.16), (2.25), or (2.34), based on the AQS function defined by (2.7), (2.22), (2.31), or (2.37). Given a null hypothesis, the linear contrast matrix C is defined, the null model is estimated by solving the LM-equations (defined as in (2.14) for the panel SL model with 1FE), and the corresponding test statistic defined by (2.13), (2.23), (2.32), or (2.38) is computed.

2.5.2 Public capital productivity

Munnell (1990) investigated the productivity of public capital in private production based on data for 48 U.S. states observed over 17 years (1970-1986). Baltagi and Pinnoi (1995) considered a Cobb-Douglas production function of the form:

$$\ln(\texttt{gsp}) = \beta_1 \ln(\texttt{pcap}) + \beta_2 \ln(\texttt{pc}) + \beta_3 \ln(\texttt{emp}) + \beta_4 \texttt{unemp} + \epsilon,$$

with state-specific fixed effects, where '`gsp`' is the gross social product of a given state, '`pcap`', '`pc`' and '`emp`' are the inputs of private capital, public capital, and labor respectively. In order to capture business cycle effects, an additional variable '`unemp`' is also added which indicates the state unemployment rate. The model now is extended by adding the time-specific fixed effects and the spatial effects. The latter is for capturing the possible spill over effects of public capital. The spatial weight matrix (W_n) is specified using a contiguity form where (i, j)th element is indicated as 1 if state i and j share a common border, otherwise 0. The final W_n is row normalized. The data file `Product.csv` and the spatial weights matrix `weight_Product.csv`, and the associated matlab files can be found in the website: `http://www.mysmu.edu/faculty/zlyang/`.

It is well known that 1970-86 is the period that U.S. had experienced several social and economic shocks such as the baby booms in the early 1970s, the oil crises in 1973 and 1979, and economic recession between 1980-82. It is therefore questionable that the above production relationship would remain stable over time. We demonstrate how our AQS test can answer this question, and how it may help detecting change points.

To test $H_0^{\texttt{TH}}$, the temporal homogeneity, assign $k = 4$. Based on full data, $T = 17$, $k_p = (k+1)(T-1) = 80$ and $C = [\texttt{blkdiag}\{C_T^k, C_T^1\}, 0_{k_p,1}]$ for the SL models; and $(k+2)(T-1) = 96$ and $C = [\texttt{blkdiag}\{C_T^k, C_T^1, C_T^1\}, 0_{k_p,1}]$ for the SLE models, where

C_τ^m is defined in (2.10) for $m = 1, k$. To test H_0^{TH} based on first four periods, $T = 4$, $k_p = (k+1)(T-1) = 15$ for the SL models, and $(k+2)(T-1) = 18$ for the SLE models. The C matrices remain in the same forms. Note that k_p is also the degrees of freedom (df) of the chi-squared test statistics, based on which the asymptotic critical values and p-values are found.

Table below summarize the values of the test statistics and their p-values, for the *naïve* tests and the nonnormality robust AQS tests for temporal homogeneity based on both the full dataset and a subset of data, fitted using the four models: 1FE-SL, 2FE-SL, 1FE-SLE and 2FE-SLE. From the table we see that all tests based on full data (t_1–t_{17}) give a clean rejection of the temporal homogeneity hypothesis H_0^{TH}.

Tests for Temporal Homogeneity: Public Capital Productivity

Data	T_{SL1}	$T_{\text{SL1}}^{(r)}$	T_{SL2}	$T_{\text{SL2}}^{(r)}$	$T_{\text{SLE1}}^{(r)}$	$T_{\text{SLE1}}^{(r)}$	T_{SLE2}	$T_{\text{SLE2}}^{(r)}$
t_1–t_{17}	1621	321	3189	328	1971	289	1556	326
	.000	.000	.000	.000	.000	.000	.000	.000
t_1–t_5	215.60	68.14	22.34	18.22	47.18	38.43	33.08	19.57
	.000	.000	.322	.573	.003	.031	.102	.721
t_1–t_4	10.24	9.37	9.59	8.69	11.78	10.61	7.07	11.43
	.804	.857	.845	.893	.858	.910	.990	.875

Note: p-values are in every second row.

As discussed in Section 2.1, a rejection of H_0^{TH} may be due to the existence of change points instead of full heterogeneity. Thus, we break down the panel into sub-periods to test whether H_0^{TH} holds for a smaller panel. Indeed, based on the first four periods (t_1–t_4), all tests do not reject H_0^{TH}, indicating that the panel consisting of the first four periods is fairly homogeneous. Furthermore, based on t_1–t_5, the tests $T_{\text{SL1}}^{(r)}$ and $T_{\text{SLE1}}^{(r)}$ reject H_0^{TH} but $T_{\text{SL2}}^{(r)}$ and $T_{\text{SLE2}}^{(r)}$ do not, suggesting that if temporal heterogeneity in intercepts is not controlled for, the first change point is t_5 or 1974, the year after the first oil crisis. However, $T_{\text{SL2}}^{(r)}$ and $T_{\text{SLE2}}^{(r)}$ do not reject H_0^{TH} **up to** first six periods, meaning that after controlling both spatial and temporal heterogeneity in intercepts, the panel is homogeneous in first six periods but changes in structure from 7th period onwards.[6] Applying the pair of test $T_{\text{SL2}}^{(r)}$ and $T_{\text{SLE2}}^{(r)}$ to

[6]The p-values for these two tests are .513 and .633 based on t_1–t_6, and .000 and .000 based on t_1–t_7, suggesting that the structure has changed since year 7 (or 1977) onwards.

test $H_0^{\mathtt{TH}}$ based on other sub-periods from 1976 onwards, all tests reject $H_0^{\mathtt{TH}}$ at 10% level, except the tests based on the following tow sub-periods: t_7–t_8 and t_{12}–t_{13}. These suggest that there exist multiple change points in this panel, and hence the standard applications of homogeneous penal methods are not valid.[7]

Based on the above results, we recommend the pairs of tests $T_{\mathtt{SL2}}^{(r)}$ and $T_{\mathtt{SLE2}}^{(r)}$ for practical applications as they control both spatial and temporal heterogeneity in intercepts (two-way fixed effects). We can further carry out the tests for detecting change points. However,the tests for temporal homogeneity based on sub-panels have revealed quite a clear picture, we therefore do not pursue `CP` tests in this application.

2.5.3 Cigarette demand.

Second application of the proposed tests uses another well known data set, *the Cigarettes Demand for the United States* (Baltagi and Levin, 1992). It contains a panel of 46 states over 30 time periods (1963-1992). The data file `cigarette.csv`, spatial weight matrix `weight_cigarette.csv`, and the associated matlab codes can be found in the website: `http://www.mysmu.edu/faculty/zlyang/`. Our analysis is based on the response variable Y = Cigarette sales in packs per capita; and the covariates X_1 = Price per pack of cigarettes; X_2 = Population above the age of 16; X_3 = Per capita disposable income; and X_4 = Minimum price in adjoining states per pack of cigarettes. Earlier studies include Hamilton (1972), McGuiness and Cowling (1975), Baltagi and Levin (1986, 1992), Baltagi et al. (2000), and Yang et al. (2006), all under homogeneity assumption and in $\log$-$\log$ form except in Yang et al. (2006) who estimated the Box-Cox functional form. The spatial weight matrix is specified using a contiguity form where (i, j)th element is 1 if state i and j share a common border, otherwise 0, and then row normalized.

Tests for temporal homogeneity/heterogeneity is of particular interest in cigarette demand, due to government's policy interventions (in 1965, 1967, 1971) in attempting reducing the consumptions of cigarettes, and the reports from medial journals as well as Surgeon General warning (in 1983) about the health hazards of smoking (see Baltagi

[7] The relatively much bigger values of the usual or naïve tests show that they are rather unreliable, in line with the Monte Carlo results.

and Levin, 1986). The table below summarize the values of the test statistics and their p-values, for tests of homogeneity based on the full panel or sub-panels and using the log-log form.

Tests for Temporal Homogeneity: Cigarette Demand

	$T^{(r)}_{\text{SL1}}$	$T^{(r)}_{\text{SL2}}$	$T^{(r)}_{\text{SLE1}}$	$T^{(r)}_{\text{SLE2}}$		$T^{(r)}_{\text{SL1}}$	$T^{(r)}_{\text{SL2}}$	$T^{(r)}_{\text{SLE1}}$	$T^{(r)}_{\text{SLE2}}$
t_1–t_{30}	443	517	507	587	t_1–t_{10}	122	118	116	126
	.000	.000	.000	.000		.000	.000	.000	.000
t_{11}–t_{20}	99	90	104	112	t_{21}–t_{30}	135	114	121	106
	.000	.000	.000	.000		.000	.000	.000	.000
t_1–t_3	13.13	9.38	9.68	8.75	t_4–t_5	6.72	6.23	7.86	8.10
	.217	.497	.644	.724		.242	.285	.248	.230
t_1–t_5	43.0	30.7	45.0	40.8	t_5–t_8	21.7	19.2	21.2	17.4
	.002	.060	.006	.018		.116	.204	.271	.495

Note: p-values are in every second row.

From the results we see that all tests based on the full data, and the first, second and last ten years data clearly reject H_0^{TH}, the hypothesis of temporal homogeneity in regression and spatial coefficients. Therefore, the `Cigarette Demand` panel is temporally heterogeneous. Further breaking down the panel and repeatedly applying the set of robust tests, we see that only the sub-panels 1963-65, 1966-67, and 1967-70 are fairly stable, suggesting that panel structures have changed after 1965, 1967, and 1970, in line with the policy interventions in 1965, 1967 and 1971. From the results, we also see that controlling the temporal heterogeneity in intercepts seems increase the stability of the overall model structure as seen from the larger p-values associated with $T^{(r)}_{\text{SL2}}$ and $T^{(r)}_{\text{SLE2}}$.

Furthermore, applying $T^{(r)}_{\text{SL2}}$ to test H_0^{CP} based on data from t_1–t_5 with $b_0 = \ell_0 = 3$ gives a p-value of 0.632 compared with 0.06 from the test of H_0^{TH}given in the table above. This confirms that 1965 is a point after which the structure has changed. Similarly, based on data from t_4–t_9, the p-value is 0.231 for testing H_0^{CP} using $T^{(r)}_{\text{SL2}}$ with $b_0 = \ell_0 = 3$, suggesting that 1968 is another change point. The `CP` tests with multiple change points can be carried out as well based on the general LM procedure we propose.

However, the matlab function `fsolve` that our LM-procedure depends upon may not always perform well. This seems to be an interesting computation problem, and is beyond the scope of this paper. In any situation, one can always repeatedly apply our robust tests for testing temporal homogeneity as they are based up the optimization functions such as `fminbnd` and `fmincon`, which are numerically much more stable stable than `fsolve`.

In summary, our tests show that there exit multiple change points in the the `Cigarette Demand` panel, and hence in real applications, one should base their analyses either on a shorter panel so that a homogeneous SPD model can be used, or a relatively longer panel and the corresponding SPD model with 'specified' change points.

2.6 Conclusion and Discussion

We introduce *adjusted quasi score tests* for temporal homogeneity/heterogeneity in regression and spatial coefficients in spatial panel data models allowing the existence of spatial and temporal heterogeneity in the intercepts of the model. The proposed tests are robust against nonnormality, they are simple and reliable as shown by the Monte Carlo results, and can be repeatedly applied to identity a 'parsimonious model' instead of the model with full temporal heterogeneity. That is, once the null hypothesis of homogeneity is rejected (as in the two empirical applications), one may proceed with further tests of hypotheses with known change points suggested by the data (as in Cigarette Demand application). Thus, the proposed tests provide useful tools for the applied researchers.

The tests can be extended by (i) adding higher-order spatial terms and spatial Durbin terms in the model, (ii) treating individual- and time-specific effects as random effects, or correlated random effects, (iii) allowing spatial-temporal heterogeneity in error variance (i.e., heteroskedasticity), (iv) allowing interactive fixed effects, and (v) by allowing dynamic effects in the model. These extensions are interesting but clearly beyond the scope of the current chapter, which will be in our future research agenda.

3 Adjusted Quasi-Score Estimation of Spatial Panel Data Models with Time Varying Coefficients

In this chapter, an adjusted quasi-score (AQS) method is proposed to estimate the fixed-effects (FE) spatial panel data models with time-varying regression and spatial coefficients. Time FE is first transformed away. The AQS functions are then obtained through adjusting the concentrated quasi scores with individual FE being concentrated out, giving a set of unbiased estimating functions and the AQS estimators that are consistent and asymptotically normal. The AQS estimation strategy naturally allows the spatial weight matrices to change with time as well. Monte Carlo results show that the proposed methods have an excellent finite sample performance. An empirical illustration using cigarette demand data is provided.

3.1 Introduction

Consider the following spatial panel data model (SPD) with two-way fixed effects:

$$Y_{nt} = \lambda_t W_n Y_{nt} + X_{nt}\beta_t + c_n + \alpha_t 1_n + U_{nt}, \quad U_{nt} = \rho M_n U_{nt} + V_{nt}, \tag{3.1}$$

$t = 1, 2, \ldots, T$, where Y_{nt} is an $n \times 1$ vector of observations on the dependent variable; X_{nt} is an $n \times k$ matrix containing the values of k exogenous regressors; W_n is an $n \times n$ spatial weight matrix, and M_n is another spatial weight matrix capturing the spatial interactions among the disturbances, which can be the same as W_n; V_{nt} is an $n \times 1$ vector of independent and identically distributed (iid) errors with mean zero and variance σ^2; λ_t is the *spatial lag* (SL) parameter in period t, ρ is the *spatial error* (SE) parameter, and β_t is the $k \times 1$ vector of regression coefficients in period t; c_n denotes the individual-specific fixed effects (FE) or spatial heterogeneity in intercept, and $\{\alpha_t\}$ are the time-specific fixed effects or unobserved temporal heterogeneity in the intercept; and 1_n is an $n \times 1$ vector of ones.

In the above model setting, both the regression coefficients and the spatial lag parameters are subject to temporal heterogeneity, but not the parameters in the errors. With SL effect, both the mean and variance of a spatial unit are directly affected by some other spatial units, however, with SE effect, only the variance of a spatial unit are directly affected by some other spatial units. Therefore, the model setting given in (3.1) with temporal heterogeneity in SL parameter provides a way of capturing time-varying spatial effects on both mean and variance. Furthermore, the spatial weight matrix W_n or M_n or both may be allowed to change with time as well, making the way of capturing the time-varying spatial effects more flexible. However, to ease the exposition, we first treat them as constant matrices and then indicates the way to relax them latter at the end of the paper, to facilitate the practical applications.[8]

Models with time-varying coefficients (TVC) have the following advantages over models with time-invariant parameters: (i) it enhances the short-run forecasting in terms of accuracy and consistency (Li et al., 2006), (ii) with estimation of time-varying parameters of interest on a period-by-period basis, it allows us to identify influential data observations (Anselin and Florax, 1995). *Temporal heterogeneity* is an important feature in economic behavior: many economic process, for example, housing decisions, welfare participation, trade flows, etc., exhibit time heterogeneity patterns. It may occur as a result of a credit crunch, an oil price shock, a tax policy change, a fad or fashion in society, a discovery of a new medicine, and an enaction of new governmental program (Bai, 2010). However, with TVC in the panel data model with individual FE, the traditional method of estimation based on transformation cannot be applied. A lack of estimation and inference for SPD models with TVC is thus a serious shortcoming.

In this chapter, we consider the estimation and inference for the FE-SPD model with time-varying regression and spatial coefficients, which extends the FE-SPD models with constant coefficients studied by Lee and Yu (2010), Baltagi and Yang (2013b) and Yang et al. (2016). The temporal heterogeneity can occur on the regression slopes. In a spatial panel data model (SPD), it may also occur on the spatial parameters (Anselin, 1988). Literature on the estimation of models with temporal heterogeneity is expanding in recent

[8]Model (3.1) is fairly general, it embeds several submodels popular in the literature. Setting $\rho = 0$, it reduces to an SPD model with SL only. Dropping one of the two FEs, the model is reduced to a one-way FE model. On the other hand, the model can be further extended to include higher-order spatial terms.

years. More and more econometricians realize that economic relationships are changing over time, and therefore, they start to consider models with stochastic parameters, see, e.g., Chow (1984), Nicholls and Pagan(1985), to name a few. The maximum likelihood estimation techniques are popular in the early literature, see, e.g., Cooley and Prescott (1976), where the parameters of the model are subject to permanent and transitory changes over time, but there is no fixed effects in their model setting. Recent literature propose a nonparametric estimation method to estimate models with time-varying parameters. See, e.g., Robinson (1989) and Orbe et al. (2005), where the methods are based on the assumption that the regression coefficients are smoothly varying over time index. There are some other literature dealing with more interesting settings, such as model with seasonal effects (Ferreira et al., 2000) or model with large time dimension (Li and Liao, 2018). Although *temporal heterogeneity* is an important feature in panel data models, it is relatively unexplored in the spatial panel literature.

In this chapter, an *adjusted quasi score* (AQS) method is proposed to estimate the FE-SPD model with time-varying regression coefficients and time-varying spatial lag coefficients, allowing the spatial errors in the model. The AQS functions are obtained through adjusting the concentrated quasi scores with individual-specific FE being concentrated out, after the time-specific effects being transformed away by an orthogonal transformation, leading to a set of unbiased estimating functions and the AQS estimators that are consistent and asymptotically normal. Monte Carlo results show that the proposed methods have an excellent finite sample performance. Empirical evidence on the temporal heterogeneity is presented based the well-known *cigarette demand* data.

The rest of the chapter is organized as follows. Section 3.2 introduces the AQS-estimation method for the general FE-SPD model with time-varying coefficients in (3.1), then specializes the AQS-estimation method to several popular submodels. Section 3.3 presents the consistency and asymptotic properties of the proposed AQS-estimators, with a separate treatment on the scenarios of "large n and large T" and "large n and small T". Section 3.4 presents Monte Carlo results. Section 3.5 presents an empirical illustration. Section 3.6 concludes the chapter with some further discussion.

3.2 AQS-Estimation of FE-SPD Models with TVC

In this section, we present a general framework for estimating the fixed effects (FE) SPD models with time-varying coefficients. The estimation strategy is valid when n is large, but T can be large or small. The basic idea of this approach is to first formulate the Gaussian likelihood function, and then adjust the resulting quasi score function to lead to a set of unbiased estimating functions. We demonstrate the exact cause of inconsistency of the estimators based on likelihood, and to show how one can adjust the quasi scores to give consistent estimators. We first outline the quasi maximum likelihood estimation, and then we introduce the AQS-estimation method for the general model specified in (3.1). Then, we give a discussion on how the general estimation method be specialized to some popular submodels to facilitate the practical applications.

3.2.1 The QML estimation

For the FE-SPD model with TVC specified by model (3.1), when both n and T are large we have to deal with the two sets of incidental parameters, individual FE and time FE, in order to achieved desired asymptotic properties of the parameter estimates. When n is large but T is small and fixed, the model becomes essentially an one-way FE model as time FE can be merged into the time-varying regressors in the form of time dummies.

As the spatial parameters and regression coefficients in (3.1) may change with time, one can apply transformation method to eliminate the time-specific effects only, **provided** that the spatial weight matrices are *row-normalized*. The transformation method is widely applied in the literature, see, e.g., Lee and Yu (2010), Baltagi and Yang (2013b) and Yang et al. (2016). Define $J_n = I_n - \frac{1}{n} l_n l_n'$. Assume W_n and M_n are row-normalized (i.e., row sums are one). Then, $J_n W_n = J_n W_n J_n$ and $J_n M_n = J_n M_n J_n$. Let $(F_{n,n-1}, \frac{1}{\sqrt{n}} l_n)$ be the orthonormal eigenvector matrix of J_n, where $F_{n,n-1}$ is the $n \times (n-1)$ sub-matrix corresponding to the eigenvalues of one. By *Spectral Theorem*, $J_n = F_{n,n-1} F_{n,n-1}'$. It follows that $F_{n,n-1}' W_n = F_{n,n-1}' W_n F_{n,n-1} F_{n,n-1}'$ and $F_{n,n-1}' M_n = F_{n,n-1}' M_n F_{n,n-1} F_{n,n-1}'$. Premultiplying $F_{n,n-1}'$ on both sides of (3.1), we have the following transformed model:

$$Y_{nt}^* = \lambda_t W_n^* Y_{nt}^* + X_{nt}^* \beta_t + c_n^* + U_{nt}^*, \quad U_{nt}^* = \rho_0 M_n^* U_{nt}^* + V_{nt}^*, \tag{3.2}$$

where $Y^*_{nt} = F'_{n,n-1}Y_{nt}$ and similarly are $X^*_{nt}, c^*_n, U^*_{nt}$ and V^*_{nt} defined, $W^*_n = F'_{n,n-1}W_nF_{n,n-1}$ and $M^*_n = F'_{n,n-1}M_nF_{n,n-1}$. After the transformation, the effective sample size becomes $N = (n-1) \times T$. Furthermore, $V^*_{nt} \sim (0, \sigma^2_0 I_{n-1})$, which is normal if V_{nt} is.[9]

Denote $\boldsymbol{\beta} = (\beta'_1, \dots, \beta'_T)'$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_T)'$, and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \rho, \sigma^2)'$. Define $A^*_n(\lambda_t) = I_{n-1} - \lambda_t W^*_n$ and $B^*_n(\rho) = I_{n-1} - \rho M^*_n$, $t = 1, \dots, T$. The quasi Gaussian loglikelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \rho, \sigma^2)'$ and c^*_n of Model (3.2) is

$$\begin{aligned}\ell(\boldsymbol{\theta}, c^*_n) = &- \tfrac{(n-1)T}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln|A^*_n(\lambda_t)| + T\ln|B^*_n(\rho)| \\ &- \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T V^{*\prime}_{nt}(\beta_t, \lambda_t, \rho, c^*_n)V^*_{nt}(\beta_t, \lambda_t, \rho, c^*_n),\end{aligned} \tag{3.3}$$

where $V^*_{nt}(\beta_t, \lambda_t, \rho, c^*_n) = B^*_n(\rho)[A^*_n(\lambda_t)Y^*_{nt} - X^*_{nt}\beta_t - c^*_n],\ t = 1, \dots, T$.

As $\{\lambda_t\}$ and $\{\beta_t\}$ are allowed to change with t, the usual fixed-effects estimation methods, such as first differencing or orthogonal transformation, cannot be applied to eliminate the individual FE. Therefore, we proceed by eliminating c^*_n through direct maximization of the loglikelihood function. First, given $\boldsymbol{\theta}$, $\ell(\boldsymbol{\theta}, c^*_n)$ is partially maximized at:

$$\tilde{c}^*_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) = \tfrac{1}{T}\textstyle\sum_{t=1}^T \big(A^*_n(\lambda_t)Y^*_{nt} - X^*_{nt}\beta_t\big), \tag{3.4}$$

which leads to the concentrated loglikelihood function of $\boldsymbol{\theta}$ upon substitution:

$$\begin{aligned}\ell^c(\boldsymbol{\theta}) = &- \tfrac{(n-1)T}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln|A^*_n(\lambda_t)| + T\ln|B^*_n(\rho)| \\ &- \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T \widetilde{V}^{*\prime}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)\widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho),\end{aligned} \tag{3.5}$$

where $\widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) = V^*_{nt}(\beta_t, \lambda_t, \rho, \tilde{c}^*_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)) = B^*_n(\rho)[A^*_n(\lambda_t)Y^*_{nt} - X^*_{nt}\beta_t - \tilde{c}^*_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)]$. Maximizing $\ell^c(\boldsymbol{\theta})$ gives the QML estimator $\hat{\boldsymbol{\theta}}_{\texttt{QML}}$ of the vector of common parameters $\boldsymbol{\theta}$.

3.2.2 The AQS estimation

Including the fixed effects into the spatial panel data models, we are likely to encounter the incidental parameter problems. Therefore, We propose an *adjusted quasi score* (AQS) method through adjusting the resulting concentrated (quasi) score functions to give a set of

[9] The time-specific effects can also be eliminated by pre-multiplying J_n on both sides of (3.1). However, the resulting disturbances J_nV_{nt} would not be linearly independent over the cross-section dimension.

unbiased estimating functions. By eliminating the asymptotic bias in the AQS functions, the AQS method achieves asymptotically unbiased parameter estimation.

To facilitate the subsequent derivations, denote $U^{\circ}_{nt}(\beta_t, \lambda_t) = A^*_n(\lambda_t)Y^*_{nt} - X^*_{nt}\beta_t$, $D^*_n(\rho) = B^{*\prime}_n(\rho)B^*_n(\rho)$. Then, we have $\widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) = B^*_n(\rho)[U^{\circ}_{nt}(\beta_t, \lambda_t) - \tilde{c}^*_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)]$, $\tilde{c}^*_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) = \frac{1}{T}\sum_{t=1}^T U^{\circ}_{nt}(\beta_t, \lambda_t)$, and the key term in (3.5):

$$\begin{aligned}\sum_{t=1}^T \widetilde{V}^{*\prime}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)\widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) &= \sum_{t=1}^T U^{\circ\prime}_{nt}(\beta_t, \lambda_t)D^*_n(\rho)U^{\circ}_{nt}(\beta_t, \lambda_t)\\ &\quad - \tfrac{1}{T}\big(\sum_{t=1}^T U^{\circ}_{nt}(\beta_t, \lambda_t)\big)'D^*_n(\rho)\big(\sum_{t=1}^T U^{\circ}_{nt}(\beta_t, \lambda_t)\big).\end{aligned}$$

Differentiating $\ell^c(\boldsymbol{\theta})$ gives the CS or CQS function of $\boldsymbol{\theta}$:

$$S^c(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2}X^{*\prime}_{nt}B^{*\prime}_n(\rho)\widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho), \quad t = 1, \dots, T, \\ \frac{1}{\sigma^2}(W^*_nY^*_{nt})'B^{*\prime}_n(\rho)\widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) - \mathrm{tr}[G^*_n(\lambda_t)], \quad t = 1, \dots, T, \\ \frac{1}{\sigma^2}\sum_{t=1}^T \widetilde{V}^{*\prime}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)H^*_n(\rho)\widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) - T\mathrm{tr}[H^*_n(\rho)], \\ -\frac{(n-1)T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}^{*\prime}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)\widetilde{V}^*_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho), \end{cases} \tag{3.6}$$

where $G^*_n(\lambda_t) = W^*_nA^{*-1}_n(\lambda_t)$ and $H^*_n(\rho) = M^*_nB^{*-1}_n(\rho), \ \ t = 1, \dots, T$.

Under mild conditions, maximizing the concentrated loglikelihood $\ell^c(\boldsymbol{\theta})$ is equivalent to solving the estimating equation $S^c(\boldsymbol{\theta}) = 0$. Denote $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}'_0, \boldsymbol{\lambda}'_0, \rho_0, \sigma^2_0)'$ as the true value of the general parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \rho, \sigma^2)'$. It is well known that for a regular quasi-score estimation problem, a necessary condition for the quasi-score estimators to be consistent is that the probability limit of the estimating function (in this case, the averaged conditional quasi score) at the true parameter value is zero, i.e.,

$$\lim_{n\to\infty}\tfrac{1}{nT}S^c(\boldsymbol{\theta}_0) \xrightarrow{p} 0,$$

see, e.g., van der Vaart (1998). However, as shown below this is not the case unless T also goes to infinity. Thus, the concentrated quasi-score estimators are not consistent unless $T \to \infty$. To solve this problem, we first derive $\mathrm{E}[S(\boldsymbol{\theta}_0)]$, and then adjust the quasi score $S^c(\boldsymbol{\theta})$ by centering so that the adjusted quasi score (AQS) vector, $S^\star(\boldsymbol{\theta}_0) = S(\boldsymbol{\theta}_0) - \mathrm{E}[S(\boldsymbol{\theta}_0)]$, is such that $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S^\star(\boldsymbol{\theta}_0) = 0$.

Assume Model (3.2) holds only under the true $\boldsymbol{\theta}_0$ and the usual expectation and variance operators correspond to $\boldsymbol{\theta}_0$. We have, $\tilde{c}^*_n(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \boldsymbol{\rho}_0) = c^*_n + B^{*-1}_n\overline{V}^*_n$ and hence $\widetilde{V}_{nt}^* \equiv \widetilde{V}_{nt}^*(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \rho_0) = V_{nt}^* - \overline{V}_n^*$, where $\overline{V}_n^* = \frac{1}{T}\sum_{t=1}^T V_{nt}^*$. Furthermore, $W_n^* Y_{nt}^* = G_{nt}^*(X_{nt}^*\beta_0 + c_n^* + B_n^{*-1}V_{nt}^*)$, where $B_{n0}^* = B_n^*(\rho_0)$ and $G_{nt0}^* = G_n^*(\lambda_{t0})$. Then, we obtain,

$$\mathrm{E}[S^c(\boldsymbol{\theta}_0)] = \begin{cases} \mathbf{0}_{Tk,1}, \\ -\frac{1}{T}\mathrm{tr}[G_n^*(\lambda_{t0})], \ t = 1, \dots T, \\ -\mathrm{tr}[H_n^*(\rho_0)], \\ -\frac{n-1}{2\sigma_0^2}. \end{cases}$$

where $\mathbf{0}_{m,r}$ denotes an $m \times r$ matrix of zeros.

The above results show that the $(\boldsymbol{\lambda}, \rho, \sigma^2)$ elements of $\mathrm{E}[S^c(\boldsymbol{\theta}_0)]$ are not zero. Hence, $\mathrm{plim}_{n\to\infty}\frac{1}{nT}\frac{\partial}{\partial\boldsymbol{\lambda}}\ell^c(\boldsymbol{\theta}_0)$, $\mathrm{plim}_{n\to\infty}\frac{1}{nT}\frac{\partial}{\partial\rho}\ell^c(\boldsymbol{\theta}_0)$, and $\mathrm{plim}_{n\to\infty}\frac{1}{nT}\frac{\partial}{\partial\sigma^2}\ell^c(\boldsymbol{\theta}_0)$ are all non-zero, suggesting that the estimators based on the direct approach cannot be consistent in general. The direct approach does not yield consistent estimators unless T goes to large. Even if T goes to large with n, there will be an asymptotic bias of order depending on T.[10] Therefore, we adjust the concentrated (quasi) scores given in (3.6) by subtracting the bias vector from it, so as to give a set of unbiased estimating functions, leading to the vector of the *adjusted quasi score* (AQS) functions below:

$$S^\star(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2}X_{nt}^{*\prime}B_n^{*\prime}(\rho)\widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho), \quad t = 1, \dots, T, \\ \frac{1}{\sigma^2}(W_n^*Y_{nt}^*)' B_n^{*\prime}(\rho)\widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) - \frac{T-1}{T}\mathrm{tr}[G_n^*(\lambda_t)], \quad t = 1, \dots, T, \\ \frac{1}{\sigma^2}\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)H_n^*(\rho)\widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) - (T-1)\mathrm{tr}[H_n^*(\rho)], \\ -\frac{(n-1)(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)\widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho), \end{cases} \tag{3.7}$$

It is easy to show that $\mathrm{E}[S^\star(\boldsymbol{\theta})] = 0$, and that $\frac{1}{nT}S^\star(\boldsymbol{\theta}_0) \xrightarrow{p} 0$ as $n \to \infty$ alone, or the finite dimensional components of $\frac{1}{nT}S^\star(\boldsymbol{\theta}_0)$ approach to 0 in probability when both n and T go infinity. The adjusted quasi score(AQS) function above leads to an estimator of $\boldsymbol{\theta}$ that not only is consistent but also has a centered asymptotic distribution, whether T is

[10] To be exact, if $\frac{1}{nT}\mathrm{E}[S(\theta_0)] = O(\frac{1}{T})$, then $\frac{1}{\sqrt{nT}}\mathrm{E}[S(\theta_0)] = O((\frac{n}{T})^{\frac{1}{2}})$, implying $\mathrm{E}[\sqrt{nT}(\tilde{\theta} - \theta_0)] = O((\frac{n}{T})^{\frac{1}{2}})$. The latter says that $\sqrt{nT}(\tilde{\theta} - \theta_0)$ would converge to a non-centered normal if $\frac{n}{T} \to c > 0$. If $\frac{n}{T} \to 0$ (large T case), the asymptotic bias vanishes.

fixed or grows with n. The latter implies that when T grows with n, the estimation based on the AQS functions eliminates the asymptotic bias incurred in the direct approach.

Solving $S^{\star}(\boldsymbol{\theta}) = 0$ leads to the AQS-estimator $\hat{\boldsymbol{\theta}}_{\texttt{AQS}}$ of $\boldsymbol{\theta}$. This root-finding process can be simplified by first solving the equations for $\boldsymbol{\beta}$ and σ^2, given $\delta = (\boldsymbol{\lambda}', \rho)'$, resulting in the constrained AQS estimators of $\boldsymbol{\beta}$ and σ^2 as

$$\hat{\boldsymbol{\beta}}(\delta) = (X_N^{*\prime} B_N^{*\prime}(\rho)\Omega B_N^{*}(\rho) X_N^{*})^{-1} X_N^{*\prime} B_N^{*\prime}(\rho)\Omega B_N^{*}(\rho) A_N^{*}(\boldsymbol{\lambda}) Y_N^{*}, \tag{3.8}$$

$$\hat{\sigma}^2(\delta) = \tfrac{1}{(n-1)(T-1)} \hat{V}_N^{*\prime}(\delta)\hat{V}_N^{*}(\delta), \tag{3.9}$$

where $Y_N^{*} = (Y_{n1}^{*\prime}, \ldots, Y_{nT}^{*\prime})'$, $X_N^{*} = \texttt{blkdiag}(X_{n1}^{*}, \ldots, X_{nT}^{*})$ where $\texttt{blkdiag}(\,)$ forms a block diagonal matrix, $A_N^{*}(\boldsymbol{\lambda}) = \texttt{blkdiag}(A_n^{*}(\lambda_1), \ldots, A_n^{*}(\lambda_T))$, $B_N^{*} = I_T \otimes B_n^{*}(\rho)$ where $\otimes$ denotes Kronecker product, $\Omega = I_N - \frac{1}{T}(1_T 1_T' \otimes I_{n-1})$ and $\hat{V}_N^{*}(\delta) = \widetilde{V}_N^{*}(\hat{\boldsymbol{\beta}}(\delta), \delta) = \Omega B_N^{*}(\rho)[A_N^{*}(\boldsymbol{\lambda}) Y_N^{*} - X_N^{*}\hat{\boldsymbol{\beta}}(\delta)]$. Substituting $\hat{\boldsymbol{\beta}}(\delta)$ and $\hat{\sigma}^2(\delta)$ back into the middle two components of the AQS function in (3.7) gives the concentrated AQS function:

$$S^{\star c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}^2(\delta)} Y_N^{\circ\prime} B_N^{*\prime} \hat{V}_N^{*}(\delta) - \frac{T-1}{T} g_N(\boldsymbol{\lambda}), \\ \frac{1}{\hat{\sigma}^2(\delta)} \hat{V}_N^{*\prime}(\delta)(\delta) H_N^{*}(\rho) \hat{V}_N^{*}(\delta) - (T-1)\mathrm{tr}[H_n^{*}(\rho)], \end{cases} \tag{3.10}$$

where $Y_N^{\circ} = \texttt{blkdiag}(W_n^{*} Y_{n1}^{*}, \ldots, W_n^{*} Y_{nT}^{*})$ and $g_N(\boldsymbol{\lambda}) = \big(\mathrm{tr}[G_n^{*}(\lambda_1)], \ldots, \mathrm{tr}[G_n^{*}(\lambda_T)]\big)'$, and $H_N^{*}(\rho) = I_T \otimes H_n^{*}(\rho)$. Solving the resulting concentrated estimating equations, $S^{\star c}(\delta) = 0$, we obtain the unconstrained AQS estimator $\hat{\delta}_{\texttt{AQS}}$ of δ. The unconstrained AQS estimators of $\boldsymbol{\beta}$ and σ^2 are thus $\hat{\boldsymbol{\beta}}_{\texttt{AQS}} \equiv \hat{\boldsymbol{\beta}}(\hat{\delta}_{\texttt{AQS}})$ and $\hat{\sigma}^2_{\texttt{AQS}} \equiv \hat{\sigma}^2(\hat{\delta}_{\texttt{AQS}})$. Let $\hat{\boldsymbol{\theta}}_{\texttt{AQS}} = (\hat{\boldsymbol{\beta}}'_{\texttt{AQS}}, \hat{\boldsymbol{\lambda}}'_{\texttt{AQS}}, \hat{\rho}, \hat{\sigma}^2_{\texttt{AQS}})'$.

Remark 3.1 *Transformation for eliminating the time FE depends on the assumption that the spatial weight matrices are 'row-normalizable'. However, in real applications, not all spatial weight matrices are row-normalizable. In this case transformation method is totally inapplicable, but our 'concentration and adjustment' strategy remains applicable.*

Remark 3.2 The QML and AQS estimators of $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)$ are equivalent. *By concentrating out σ^2 from* (3.6) *and* (3.7)*, we see that the resultant concentrated versions of these two sets of functions are identical, showing the equivalence.*

Remark 3.3 *Thus, in terms of point estimation, one can simply follow the QML method and then rescale the QMLE of σ^2 by multiplying the factor $\frac{T}{T-1}$. However, in terms of* *inference, one must follow the AQS method in order to have a valid set of 'estimating functions' so as to obtain joint asymptotic distribution and valid estimator of the variance-covariance (VC) matrix of the AQS estimators.*

3.2.3 A discussion on some submodels

To facilitate the practical applications, we give a brief discussion on how the general AQS estimation strategy simplifies to some popular submodels of (3.1), namely, (i) SPD model with both SL and SE (SLE) but one-way individual FE (1FE), (ii) SPD model with SL only and two-way FE (2FE), and (iii) SPD model with SL only and 1FE.

SPD model with SLE and 1FE. Until now, we have considered the SPD models with both individual-specific FE and time-specific FE, and the way to handle both sets of FEs is to 'eliminate' them by transformation and concentration with adjustment. In case when n is large and T is small and fixed, the transformation to wipe out the time FE is unnecessary as the model fits into one-way FE model (with individual-specific FE only), which is also popular in empirical applications. In the following, we consider the AQS-estimation for SPD models with individual fixed effects only.

$$Y_{nt} = \lambda_t W_n Y_{nt} + X_{nt}\beta_t + c_n + U_{nt}, \quad U_{nt} = \rho_0 M_n U_{nt} + V_{nt}, \tag{3.11}$$

Model (3.11) takes an identical form as Model (3.2).[11] Hence the steps leading to the AQS estimators are similar. Define $A_n(\lambda_t) = I_n - \lambda_t W_n$, $t = 1, \ldots, T$, $B_n(\rho) = I_n - \rho M_n$. The overall sample size in this model is $N = n \times T$. The AQS vector becomes:

$$S^\star(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2} X'_{nt} B'_n(\rho) \widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho), \quad t = 1, \ldots, T, \\ \frac{1}{\sigma^2} (W_n Y_{nt})' B'_n(\rho) \widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) - \frac{T-1}{T} \mathrm{tr}[(\boldsymbol{\rho}) G_n(\lambda_t)], \quad t = 1, \ldots, T, \\ \frac{1}{\sigma^2} \sum_{t=1}^T \widetilde{V}'_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) H_n(\rho) \widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) - (T-1)\mathrm{tr}[H_n(\rho)], \\ -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^T \widetilde{V}'_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) \widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho), \end{cases} \tag{3.12}$$

[11]Now, X_t may contain the column vector 1_n and β_t may contain α_t to incorporate the time FE when T is small and fixed.

where $\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) = V_{nt}(\beta_t, \lambda_t, \rho, \tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)) = B_n(\rho)[A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t - \tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho)]$, and $\tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho) = \frac{1}{T}\sum_{t=1}^T \big(A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t\big)$. Given $\delta = (\boldsymbol{\lambda}', \rho)'$, the constrained AQS estimators of $\boldsymbol{\beta}$ and σ^2 take the form:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}(\delta) &= (X_N' B_N'(\rho)\Omega B_N(\rho)X_N)^{-1} X_N' B_N'(\rho)\Omega B_N(\rho) A_N(\boldsymbol{\lambda}) Y_N, &(3.13)\\
\hat{\sigma}^2(\delta) &= \tfrac{1}{n(T-1)} \hat{V}_N'(\delta)\hat{V}_N(\delta), &(3.14)
\end{aligned}$$

where $X_N = \texttt{blkdiag}(X_{n1}, \ldots, X_{nT})$, $A_N(\boldsymbol{\lambda}) = \texttt{blkdiag}(A_n(\lambda_1), \ldots, A_n(\lambda_T))$, $B_N(\rho) = I_T \otimes B_n(\rho)$, $Y_N = (Y_{n1}', \ldots, Y_{nT}')'$, $\Omega = I_N - \frac{1}{T}(\ell_T\ell_T' \otimes I_n)$, and $\hat{V}_N(\delta) = \widetilde{V}_N(\hat{\boldsymbol{\beta}}(\delta), \delta) = \Omega B_N(\rho)[A_N(\boldsymbol{\lambda})Y_N - X_N\hat{\boldsymbol{\beta}}(\delta)]$. Substituting $\hat{\boldsymbol{\beta}}(\delta)$ and $\hat{\sigma}^2(\delta)$ back into the middle two components of the AQS function in (3.12) gives the concentrated AQS function:

$$S^{\star c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}^2(\delta)} Y_N^{\circ\prime} B_N' \hat{V}_N(\delta) - \frac{T-1}{T} g_N(\boldsymbol{\lambda}), \\ \frac{1}{\hat{\sigma}^2(\delta)} \hat{V}_N'(\delta) H_N(\rho) \hat{V}_N(\delta) - (T-1)\mathrm{tr}[H_n(\rho)], \end{cases} \tag{3.15}$$

where Y_N° $=\texttt{blkdiag}(W_nY_{n1}, \ldots, W_nY_{nT})$, $g_N(\boldsymbol{\lambda}) = \big(\mathrm{tr}[G_n(\lambda_1)], \ldots, \mathrm{tr}[G_n(\lambda_T)]\big)'$, and $H_N(\rho) = I_T \otimes H_n(\rho)$. Solving $S^{\star c}(\delta) = 0$ gives the unconstrained AQSE $\hat{\delta}$ of δ, and the unconstrained AQSEs of $\boldsymbol{\beta}$ and σ^2 as $\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\hat{\delta})$ and $\hat{\sigma}^2 \equiv \hat{\sigma}^2(\hat{\delta})$.

SPD model with SL and 2FE. The model takes the following form: $Y_{nt} = \lambda_t W_n Y_{nt} + X_{nt}\beta_t + c_n + \alpha_t l_n + V_{nt}$, where W_n is row-normalized. Applying the same orthonormal transformation as that for Model (3.1), we have the following transformed model:

$$Y_{nt}^* = \lambda_t W_n^* Y_{nt}^* + X_{nt}^*\beta_t + c_n^* + V_{nt}^*, \ t = 1, \ldots, T, \tag{3.16}$$

where Y_{nt}^*, X_{nt}^*, c_n^*, W_n^* and V_{nt}^* are defined as in Model (3.2). Now, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \sigma^2)'$. After the transformation, the overall sample size is $N = (n-1)T$ as for the 2FE-SLE model. Following the same steps as in the previous section, we obtain the AQS vector:

$$S^\star(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2} X_{nt}^{*\prime} \widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}), \ t = 1, \ldots, T, \\ \frac{1}{\sigma^2}(W_n^* Y_{nt}^*)' \widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}) - \frac{T-1}{T}\mathrm{tr}[G_n^*(\lambda_t)], \ t = 1, \ldots, T, \\ -\frac{(n-1)(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime}(\boldsymbol{\beta}, \boldsymbol{\lambda}) \widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}). \end{cases} \tag{3.17}$$

where $\widetilde{V}_{nt}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}) = A_n^*(\lambda_t)Y_{nt}^* - X_{nt}^*\beta_t - \tilde{c}_n^*(\boldsymbol{\beta}, \boldsymbol{\lambda})$, $\tilde{c}_n^*(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{T}\sum_{t=1}^T [A_n^*(\lambda_t)Y_{nt}^* - X_{nt}^*\beta_t]$. Again, $A_n^*(\lambda_t) = I_{n-1} - \lambda_t W_n^*$ and $G_n^*(\lambda_t) = W_n^* A_n^{*-1}(\lambda_t)$. The constrained

AQS estimators of $\boldsymbol{\beta}$ and σ^2, given $\boldsymbol{\lambda}$, are: $\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) = (X_N^{*\prime}\Omega X_N^*)^{-1}X_N^{*\prime}\Omega A_N^*(\boldsymbol{\lambda})Y_N^*$ and $\hat{\sigma}^2(\boldsymbol{\lambda}) = \frac{1}{(n-1)(T-1)}\hat{V}_N^{*\prime}(\boldsymbol{\lambda})\hat{V}_N^*(\boldsymbol{\lambda})$, where X_N^*, $A_N^*(\boldsymbol{\lambda})$ and Y_N^* are defined as in 3.8-3.9, $\hat{V}_N^*(\boldsymbol{\lambda}) = \widetilde{V}_N^*(\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = \Omega[A_N^*(\boldsymbol{\lambda})Y_N^* - X_N^*\hat{\boldsymbol{\beta}}(\delta)]$. Substituting $\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})$ and $\hat{\sigma}^2(\boldsymbol{\lambda})$ back into the middle component of the AQS function in (3.17) gives the concentrated AQS function:

$$S^{\star c}(\boldsymbol{\lambda}) = \frac{1}{\hat{\sigma}^2(\boldsymbol{\lambda})}Y_N^{\circ\prime}\hat{V}_N^*(\boldsymbol{\lambda}) - \frac{T-1}{T}g_N(\boldsymbol{\lambda}), \tag{3.18}$$

where Y_N° and $g_N(\boldsymbol{\lambda})$ are defined as in the concentrated AQS function (3.10). We obtain the unconstrained AQS estimators $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ by solving $S^{\star c}(\boldsymbol{\lambda}) = 0$. The unconstrained AQS estimators of $\boldsymbol{\beta}$ and σ^2 are thus $\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\lambda}})$ and $\hat{\sigma}^2 \equiv \hat{\sigma}^2(\hat{\boldsymbol{\lambda}})$.

SPD model with SL and 1FE. Consider the model, $Y_{nt} = \lambda_t W_n Y_{nt} + X_{nt}\beta_t + c_n + V_{nt}$, which takes an identical form as Model (3.16). The AQS vector becomes:

$$S^\star(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2}X_{nt}'\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}), \ t = 1, \ldots, T, \\ \frac{1}{\sigma^2}(W_nY_{nt})'\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}) - \frac{T-1}{T}\mathrm{tr}[G_n(\lambda_t)], \ t = 1, \ldots, T, \\ -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}_{nt}'(\boldsymbol{\beta}, \boldsymbol{\lambda})\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}). \end{cases} \tag{3.19}$$

where $\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t - \tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda})$ and $\tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{T}\sum_{t=1}^T[A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t]$. Given $\boldsymbol{\lambda}$, the constrained AQSEs of $\boldsymbol{\beta}$ and σ^2 are: $\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) = (X_N'\Omega X_N)^{-1}X_N'\Omega A_N(\boldsymbol{\lambda})Y_N$ and $\hat{\sigma}^2(\boldsymbol{\lambda}) = \frac{1}{n(T-1)}\hat{V}_N'(\boldsymbol{\lambda})\hat{V}_N(\boldsymbol{\lambda})$. Here, X_N, $A_N(\boldsymbol{\lambda})$ and Y_N are defined in 3.13-3.14, and $\hat{V}_N(\boldsymbol{\lambda}) = \widetilde{V}_N(\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = \Omega[A_N(\boldsymbol{\lambda})Y_N - X_N\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})]$. Substituting $\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})$ and $\hat{\sigma}^2(\boldsymbol{\lambda})$ back into the middle component of the AQS function in (3.19) gives the concentrated AQS function:

$$S^{\star c}(\boldsymbol{\lambda}) = \frac{1}{\hat{\sigma}^2(\boldsymbol{\lambda})}Y_N^{\circ\prime}\hat{V}_N(\delta) - \frac{T-1}{T}g_N(\boldsymbol{\lambda}), \tag{3.20}$$

where Y_N° and $g_N(\boldsymbol{\lambda})$ are defined as in (3.15). Solving $S^{\star c}(\delta) = 0$ gives the AQSE $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$, and the AQSEs of $\boldsymbol{\beta}$ and σ^2 as $\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\lambda}})$ and $\hat{\sigma}^2 \equiv \hat{\sigma}^2(\hat{\boldsymbol{\lambda}})$.

3.3 Asymptotic properties of the AQS estimators

In this section we study the consistency and asymptotic normality of the proposed AQS estimators for the FE-SPD model with time-varying coefficients. To facilitate the discussions, first recall: $\boldsymbol{\theta}_0$ denotes the true value of the parameter vector $\boldsymbol{\theta}$; a parametric vector/matrix at the true parameter value is differentiated from that at a general parameter value by dropping its argument, e.g., $B_n^* \equiv B_n^*(\rho_0)$; and the usual expectation, variance and covariance operators 'E' 'Var' and 'Cov' correspond to the true parameter vector θ_0. Second, some general notation and convention are as follows: (i) δ denotes the vector of parameters in the concentrated AQS function, and $\mathbf{\Delta}$ the space from which δ takes values; (ii) tr$(\cdot)$, $|\cdot|$ and $\|\cdot\|$ denote, respectively, the trace, determinant, and Frobenius norm of a matrix; (iii) $\gamma_{\max}(A)$ and $\gamma_{\min}(A)$ denote, respectively, the largest and smallest eigenvalues of a real symmetric matrix A; and (iv) $\mathtt{diag}(a_k)$ forms a diagonal matrix using the elements $\{a_k\}$ and $\mathtt{blkdiag}(A_k)$ forms a block-diagonal matrix using the matrices $\{A_k\}$. Proofs of the lemmas and theorems are sketched in Appendices.

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

Assumption B: *The space $\mathbf{\Delta}$ is compact, and the true parameter δ_0 lies in its interior.*

Assumption C: *The time-varying regressors $\{X_{nt},\ t = 1, \ldots, T\}$ are exogenous with respect to v_{it} but are correlated with μ and α in an arbitrary manner, their values are uniformly bounded in n and t, and $\lim_{n\to\infty} \frac{1}{nT} X_N^{*\prime} B_N^{*\prime} B_N^* X_N^*$ exists and is nonsingular.*

Assumption D: *(i) the elements w_{ij} of W_n are at most of order h_n^{-1}, uniformly in all i and j, and $w_{ii} = 0$ for all i; (ii) $h_n/n \to 0$ as $n \to \infty$; (iii) W_n is row-normalized and is uniformly bounded in both row and column sums in absolute value; (iv) The matrix $A_n(\lambda_t)$ is invertible for all $\lambda_t \in \Lambda_t$, $A_n^{-1}(\lambda_{t0})$ is uniformly bounded in both row and column sums, and $A_n^{-1}(\lambda_t)$ is uniformly bounded in either row or column sums, uniformly in $\lambda_t \in \Lambda_t$, where Λ_t is a compact parameter space, $t = 1, \ldots, T$.*

Assumption E: *(i) the elements m_{ij} of M_n are at most of order h_n^{-1}, uniformly in all i and j, and $m_{ii} = 0$ for all i; (ii) $h_n/n \to 0$ as $n \to \infty$; (iii) M_n is row-normalized and is uniformly bounded in both row and column sums in absolute value; (iv) The matrix $B_n(\rho)$ is invertible for all $\rho \in \mathbb{P}$, $B_n^{-1}(\rho_0)$ is uniformly bounded in both row and column sums, and $B_n^{-1}(\rho)$ is uniformly bounded in either row or column sums, uniformly in $\rho \in \mathbb{P}$, where $\mathbb{P}$ denotes a compact parameter space.*

Assumptions A-E are standard in the spatial econometrics literature (see Lee, 2004;

Lee and Yu, 2010; Su and Yang, 2015; and Yang et al., 2016). The existence of 4th moment of idiosyncratic errors in Assumption A is a standard requirement in QML and GMM estimation. The proof of consistency of the 'nonlinear' parameters δ requires the compactness of the parameter space Δ as in Assumption B. Assumptions C, D and E together guarantee the existence and nonsingularity of $\lim_{n\to\infty}\frac{1}{nT}X_N^{*\prime}B_N^{*\prime}(\rho)\Omega B_N^*(\rho)X_N^*$, and with these the consistency of $\hat{\boldsymbol{\beta}}_{\texttt{AQS}}$ and $\hat{\sigma}_{\texttt{AQS}}$ follows immediately that of $\hat{\delta}$. Conditions (i), (iii) and (iv) under Assumptions D and E are standard conditions put on the spatial weight matrices (Lee, 2004; Yang, 2018). Assumption D(ii) and E(ii) further allow the degree of spatial dependence to grow with n (Lee, 2004; Yang, 2018).

The consistency of the AQS estimators $\hat{\boldsymbol{\theta}}_{\texttt{AQS}}$ lies with the consistency of $\hat{\delta}_{\texttt{AQS}}$, as the consistency of $\hat{\boldsymbol{\beta}}_{\texttt{AQS}}$ and $\hat{\sigma}^2_{\texttt{AQS}}$ follows almost immediately that of $\hat{\delta}$ under assumptions C-E. The concentrated estimating function (CEF) $S^{\star c}(\delta)$ and its population counterpart play a major role for the consistency of $\hat{\delta}_{\texttt{AQS}}$ for δ.

Define $\bar{S}^{\star}(\boldsymbol{\theta}) = \mathrm{E}[S^{\star}(\boldsymbol{\theta})]$, the population counterpart of the joint AQS function given in (3.7). Given δ, $\bar{S}^{\star}(\boldsymbol{\theta}) = 0$ is partially solved at:

$$\bar{\boldsymbol{\beta}}(\delta) = (X_N^{*\prime}B_N^{*\prime}(\rho)\Omega B_N^*(\rho)X_N^*)^{-1}X_N^{*\prime}B_N^{*\prime}(\rho)\Omega B_N^*(\rho)A_N^*(\boldsymbol{\lambda})\mathrm{E}(Y_N^*), \tag{3.21}$$

$$\bar{\sigma}^2(\delta) = \tfrac{1}{(n-1)(T-1)}\mathrm{E}[\bar{V}_N^{*\prime}(\delta)\bar{V}_N^*(\delta)], \tag{3.22}$$

where $\bar{V}_N^*(\delta) = \widetilde{V}_N^*|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}(\delta)} = \Omega B_N^*(\rho)[A_N^*(\boldsymbol{\lambda})Y_N^* - X_N^*\bar{\boldsymbol{\beta}}(\delta)]$. Substituting $\bar{\boldsymbol{\beta}}(\delta)$ and $\bar{\sigma}^2(\delta)$ back into the δ-component of $\bar{S}^{\star}(\boldsymbol{\theta})$ leads to the population counterpart of the concentrated AQS function given in (3.10) as

$$\bar{S}^{\star c}(\delta) = \begin{cases} \frac{1}{\bar{\sigma}^2(\delta)}\mathrm{E}[Y_N^{\circ\prime}B_N^{*\prime}(\rho)\bar{V}_N^*(\delta)] - \frac{T-1}{T}g_N(\boldsymbol{\lambda}), \\ \frac{1}{\bar{\sigma}^2(\delta)}\mathrm{E}[\bar{V}_N^{*\prime}(\delta)H_N^*(\rho)\bar{V}_N^*(\delta)] - (T-1)\mathrm{tr}[H_n^*(\rho)]. \end{cases} \tag{3.23}$$

Clearly, the AQS-estimator $\hat{\delta}$ of δ_0 is a zero of $S^{\star c}(\delta)$. It is easy to see that $\bar{S}^{\star c}(\delta_0) = 0$ through $\bar{\boldsymbol{\beta}}(\delta_0) = \boldsymbol{\beta}_0$ and $\bar{\sigma}^2(\delta_0) = \sigma_0^2$, i.e., δ_0 is a zero of $\bar{S}^{\star c}(\delta)$. Thus, by Theorem 5.9 of van der Vaart (1998), $\hat{\delta}_{\texttt{AQS}}$ will be consistent for δ_0 if $\sup_{\delta\in\boldsymbol{\Delta}}\frac{1}{N^*}\left\|S^{\star c}(\delta) - \bar{S}^{\star c}(\delta)\right\| \xrightarrow{p} 0$, and the following identification condition holds, where $N^* = (n-1)(T-1)$.

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

This assumption can be seen to be satisfied by some more primitive (but messier in expressions) conditions. Let $D_N(\delta) = B_N^*(\rho)A_N^*(\boldsymbol{\lambda})$ and $D_N \equiv D_N(\delta_0)$, $f_N = A_N^{*-1}X_N\boldsymbol{\beta}_0 = (f'_{n1},\ldots,f'_{nT})'$, f_N°=$\texttt{blkdiag}(f_{n1},\ldots,f_{nT})$, $D_N^*(\delta) = D_N(\delta) - B_N^*(\rho)A_N^*$, $G_N^* = \mathbf{W}_N A_N^{-1}(\boldsymbol{\lambda})$ and $\mathbf{W}_N = I_T \otimes W_n$. Then, it is easy to see that Assumption F holds if:

$$\lim_{N\to\infty}\tfrac{1}{N}F_N(\delta) \neq 0, \forall \delta \neq \delta_0, \quad \text{where} \tag{3.24}$$

$$F_N(\delta) = \begin{cases} f_N^{\circ\prime}D_N^{*\prime}\bar{G}_N\mathbf{M}(\rho)\Omega D_N(\delta)f_N + \sigma_0^2(g_N^{\circ}(\boldsymbol{\lambda}) - \frac{T-1}{T}g_N(\boldsymbol{\lambda})), \\ f'_N D'_N(\delta)\Omega\mathbf{M}(\rho)H_N^*(\delta)\mathbf{M}(\rho)\Omega D_N(\delta)f_N + \sigma_0^2\mathrm{tr}(H_N^{\circ} - (T-1)H_n^*(\delta)), \end{cases}$$

where $\bar{G}_N = B_N^{*-1\prime}G_N^{*\prime}B_N^{*\prime}(\rho)$, $g_N^{\circ}(\boldsymbol{\lambda}) = \big(\mathrm{tr}[G_n^{\circ}(\lambda_1)],\ldots,\mathrm{tr}[G_n^{\circ}(\lambda_T)]\big)'$ is a vector with elements that capture trace of block diagonal matrices of $G_N^{\circ} = D_N'^{-1}D'_N(\delta)\bar{G}_N\Omega D_N(\delta)D_N^{-1}$ $= \texttt{blkdiag}(G_n^{\circ}(\lambda_1),\ldots,G_n^{\circ}(\lambda_T))$, $H_N^{\circ} = D_N'^{-1}D'_N(\delta)\Omega H_N^*\Omega D_N(\delta)D_N^{-1}$, and $\mathbf{M}(\rho) = I_N - \Omega B_N^*(\rho)X_N^*(X_N^{*\prime}B_N^{*\prime}(\rho)\Omega B_N^*(\rho)X_N^*)^{-1}X_N^{*\prime}B_N^{*\prime}(\rho)\Omega$.

Note that to show $\sup_{\delta\in\boldsymbol{\Delta}}\frac{1}{N^*}\big\|S^{\star c}(\delta) - \bar{S}^{\star c}(\delta)\big\| \xrightarrow{p} 0$, the detailed expressions for $\bar{\sigma}^2(\delta)$ and $\bar{S}^{\star c}(\delta)$ are needed, which can be easily obtained through the following identity:

$$\bar{V}_N^*(\delta) = \mathbf{M}(\rho)\Omega B_N^*(\rho)A_N^*(\boldsymbol{\lambda})Y_N^* + \mathbf{P}(\rho)\Omega B_N^*A_N^*(\boldsymbol{\lambda})\widetilde{Y}_N^*, \tag{3.25}$$

where $\widetilde{Y}_N^* = Y_N^* - \mathrm{E}(Y_N^*)$, $\mathbf{M}(\rho) = I_N - \Omega B_N^*(\rho)X_N^*(X_N^{*\prime}B_N^{*\prime}(\rho)\Omega B_N^*(\rho)X_N^*)^{-1}X_N^{*\prime}B_N^{*\prime}(\rho)\Omega$, and $\mathbf{P}(\rho) = I_N - \mathbf{M}(\rho)$. Also note that the quantities $\mathrm{E}(\widetilde{Y}_N^*)$ and $\mathrm{Var}(\widetilde{Y}_N^*)$, etc., involved in (3.21)-(3.23) are functions of $\boldsymbol{\theta}_0$, but not $\boldsymbol{\theta}$.

The identification for $\boldsymbol{\beta}$ and σ_0^2 follows with the identification of δ by Assumption C, D and E. The compactness of the parameter space of $\boldsymbol{\beta}$ and σ_0^2 is not needed due to the linearity property. We have the following theorem with the detailed proof given in Appendix C.

Theorem 3.1 *Under Assumptions A-F, $\boldsymbol{\theta}_0$ is identified. Furthermore, for the AQS-estimators $\hat{\boldsymbol{\theta}}_{\texttt{AQS}}$ based on the AQS function, $\hat{\boldsymbol{\theta}}_{\texttt{AQS}} \xrightarrow{p} \boldsymbol{\theta}_0$.*

To derive the asymptotic distribution of the AQS estimators $\hat{\boldsymbol{\theta}}_{\texttt{AQS}}$, we start with a Taylor expansion of the joint AQS equations $S^{\star}(\hat{\boldsymbol{\theta}}_{\texttt{AQS}}) = 0$ at $\boldsymbol{\theta}_0$, and then we verify that the AQS function $S^{\star}(\boldsymbol{\theta}_0)$ is asymptotically normal and that the corresponding adjusted Hessian $\frac{\partial}{\partial\boldsymbol{\theta}'}S^{\star}(\bar{\boldsymbol{\theta}})$ has proper asymptotic behavior for some $\bar{\boldsymbol{\theta}}$ lying between $\hat{\boldsymbol{\theta}}_{\texttt{AQS}}$ and $\boldsymbol{\theta}_0$

elementwise. Let $\mathbb{V}_N = (V'_{n1}, \ldots, V'_{nT})'$ be the vector of original errors with elements $\{v_{it}\}$ being iid of mean 0, variance σ^2. We can express $\widetilde{V}^*_{nt}$ and $W^*_n Y^*_{nt}$ in terms of $\mathbb{V}_N$.

Lemma 3.1 *Let z_t be a $T \times 1$ vector of element 1 in the tth position and 0 elsewhere, and define $Z_{Nt} = z_t \otimes I_n$, $\bar{Z}_N = \frac{1}{T}(l_T \otimes I_n)$, and $Z^{\circ}_{Nt} = Z_{Nt} - \bar{Z}_N$. We have,*

$$\widetilde{V}^*_{nt} \equiv \widetilde{V}^*_{nt}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \rho_0) = F'_{n,n-1}(V_{nt} - \bar{V}_N) = F'_{n,n-1} Z^{\circ\prime}_{Nt} \mathbb{V}_N, \tag{3.26}$$

$$W^*_n Y^*_{nt} = G^*_{nt}(X^*_{nt}\beta_0 + c^*_n + B^{*-1}_n V^*_{nt}) = \eta^*_{nt0} + G^*_{nt} B^{*-1}_n F'_{n,n-1} Z'_{Nt} \mathbb{V}_N. \tag{3.27}$$

Using these representations, the AQS function at θ_0 can be written as

$$S^\star(\theta_0) = \begin{cases} \Pi'_{1t}\mathbb{V}_N, \quad t = 1, \ldots, T, \\ \Pi'_{2t}\mathbb{V}_N + \mathbb{V}'_N \Phi_{1t} \mathbb{V}_N - \frac{T-1}{T}\mathrm{tr}(G^*_{nt0}), \quad t = 1, \ldots, T, \\ \mathbb{V}'_N \Phi_{2t} \mathbb{V}_N - (T-1)\mathrm{tr}(H^*_{n0}), \\ \mathbb{V}'_N \Psi \mathbb{V}_N - \frac{(n-1)(T-1)}{2\sigma^2}, \end{cases} \tag{3.28}$$

where $\Pi_{1t} = \frac{1}{\sigma_0^2} Z^{\circ *}_{Nt} B^*_{n0} X^*_{nt}$, $\Pi_{2t} = \frac{1}{\sigma_0^2} Z^{\circ *}_{Nt} B^*_{n0} \eta^*_{nt0}$,$\Phi_{1t} = \frac{1}{\sigma_0^2} Z^*_{Nt} B^{*-1\prime}_{n0} G^{*\prime}_{nt0} B^{*\prime}_{n0} Z^{\circ *\prime}_{Nt}$, $\Phi_2 = \frac{1}{\sigma_0^2}\sum_{t=1}^T Z^{\circ *}_{Nt} H^*_{n0} Z^{\circ *\prime}_{Nt}$, and $\Psi = \frac{1}{2\sigma_0^4}\sum_{t=1}^T Z^{\circ *}_{Nt} Z^{\circ *\prime}_{Nt}$, with $Z^*_{Nt} = Z_{Nt} F_{n,n-1}$ and $Z^{\circ *}_{Nt} = Z^{\circ}_{Nt} F_{n,n-1}$. The above representation for AQS function given in (3.7) at θ_0 in terms of $\mathbb{V}_N = (V'_{n1}, \ldots, V'_{nT})'$ turns out to be very useful. It leads to a simple way for establishing the asymptotic normality and estimating the variance-covariance (VC) matrix of the AQS vector. To further simplify the expressions, denote N^* as the effective sample size, where $N^* = n^* \times T^*$, $n^* = n - 1$ and $T^* = T - 1$.

Theorem 3.2 *Under the assumptions of Theorem 3.1, we have, as $n \to \infty$,*

$$\sqrt{N^*}\big(\hat{\theta}_{\texttt{AQS}} - \theta_0\big) \xrightarrow{D} N\big[0, \lim_{n\to\infty} I^{\circ -1}(\theta_0)\Sigma^{\circ}(\theta_0) I^{\circ -1}(\theta_0)\big],$$

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

The expressions for $I^{\circ}(\theta)$ and $\Sigma^{\circ}(\theta)$ can be found in Appendix B.2, where $I^{\circ}(\theta)$ can be consistently estimated by $-\frac{1}{N^*}\frac{\partial}{\partial \theta'} S^\star(\hat{\theta}_{\texttt{AQS}})$. The quantity $\Sigma^{\circ}(\theta)$ involves the 3rd and 4th cumulants (or skewness and excess kurtosis), $\mu_{(3)}$ and $\mu_{(4)}$, of the original errors

v_{it}. However, only the estimates of the transformed errors are available. Therefore, some details on the methods of estimating $\mu_{(3)}$ and $\mu_{(4)}$ are necessary. The elements of the transformed errors V_{nt}^* may not be totally independent unless the original errors are normal and their 3rd and 4th moments may not be constant. Thus, one needs to work with the original error vector V_{nt} through $V_{nt}^* = F'_{n,n-1}V_{nt}$. and their estimates are obtained by applying Lemma 4.1(a) of Yang et al. (2016).

Case of large n and large T. So far, we focus on the short panels, i.e., panels with large n, and small and fixed T. When T increases with n, the asymptotic arguments leading the consistency and asymptotic normality of the AQS estimator $\hat{\theta}_{\texttt{AQS}}$ are no longer appropriate, as the dimensions of θ_0, $I^{\circ}(\theta_0)$ and $\Sigma^{\circ}(\theta_0)$ grow with the increase of T. A connected phenomenon is that the β_t and λ_t components of $I^{\circ}(\theta_0)$ will approach to zero as $n, T \to \infty$. This raises a issue of convergence rates for the components of the AQS estimator $\hat{\theta}_{\texttt{AQS}}$. While in spatial framework a panel with large n and small T is more popular, it is also important to allow for a panel with large n and large T (but smaller than n). In this sense, to keep out theoretical arguments simple, one can simply apply the so-called 'sequential asymptotics' arguments to extend the results of Theorems (3.1) and (3.2) by letting n goes large first and then T.

To do so, we work on each component, β_t and λ_t, of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. From the information matrix $I(\theta_0) = -\mathrm{E}[\frac{\partial}{\partial\theta'}S^{\star}(\theta_0)]$ given in Appendix B.2, we see that the β_t block of $\frac{1}{nT}I(\theta_0)$ is $\frac{1}{nT}\big(\frac{T-1}{T\sigma_0^2}X_{nt}^{*\prime}D_n^*X_{nt}^*\big)$, which approaches to a zero matrix as $n, T \to \infty$. However, the quantity with a different normalizing factor $\frac{1}{n}$, $\frac{1}{n}\big(\frac{T-1}{T\sigma_0^2}X_{nt}^{*\prime}D_n^*X_{nt}^*\big)$ will converge to a positive definite matrix as $n, T \to \infty$. A similar phenomenon holds for the λ_t component of $\frac{1}{nT}I(\theta_0)$. Furthermore, it is easy to see that the (ρ, σ^2) component of $\frac{1}{nT}I(\theta_0)$ converges to a positive definite matrix as $n, T \to \infty$. These reveal that the convergence rate for $\hat{\beta}_t$ and $\hat{\lambda}_t$ are both $\sqrt{n}$, which the rate of convergence for $\hat{\rho}$ and $\hat{\sigma}^2$ are both $\sqrt{nT}$.

We have the following results.

Theorem 3.3 *Under the assumptions of Theorem 3.2, we have,*

$$(i)\ \sqrt{n}\big(\hat{\beta}_t-\beta_{t0}\big)\xrightarrow{D}N\big(0,\ \Omega_t\big),\ \textit{for each } t,\ \textit{as } n\to\infty,\ \textit{and then } T\to\infty,$$
$$(ii)\ \sqrt{n}\big(\hat{\lambda}_t-\lambda_t0\big)\xrightarrow{D}N\big(0,\ \tau^2_{\lambda_t}\big),\ \textit{for each } t,\ \textit{as } n\to\infty,\ \textit{and then } T\to\infty,$$
$$(iii)\ \sqrt{nT}\big(\hat{\rho}-\rho_0\big)\xrightarrow{D}N\big(0,\ \tau^2_{\rho}\big),\ \textit{as } n,T\to\infty,$$
$$(iv)\ \sqrt{nT}\big(\hat{\sigma}^2-\sigma^2_0\big)\xrightarrow{D}N\big(0,\ \tau^2_{\sigma^2}\big),\ \textit{as } n,T\to\infty,$$

where Ω_t and $\tau^2_{\lambda_t}$ are the limits of the corresponding components of $\frac{1}{T}I^{\circ-1}(\theta_0)\Sigma^{\circ}(\theta_0)I^{\circ-1}(\theta_0)$, and τ^2_{ρ} and $\tau^2_{\sigma^2}$ the limits of the corresponding components of $I^{\circ-1}(\theta_0)\Sigma^{\circ}(\theta_0)I^{\circ-1}(\theta_0)$.

From the results Theorem 3.3, it is clear the joint inference for a finite number of components of $\boldsymbol{\beta}$ can be made by extending the result (i), the joint inference for a finite number of components of $\boldsymbol{\lambda}$ can be made by extending the result (ii), and the joint inference concerning a finite number of components of $\boldsymbol{\theta}$ can be made by extending and combining the results $(i)-(iv)$ of Theorem 3.3. These results provide useful tools for the practical applications in switching from the fixed T scenario to the large T scenario.

3.4 Monte Carlo Simulation

Monte Carlo experiments are carried out to investigate the finite sample performance of (i) the proposed AQS estimators of the FE-SPD model with time-varying coefficients, and (ii) the estimated standard errors of the AQS estimators. We use the following four models in our Monte Carlo experiments, all having two time-varying regressors:

$$\begin{aligned}
&\texttt{SL}_1: && Y_{nt}=\lambda_{t0}W_nY_{nt}+X_{nt}\beta_{t0}+c_{n0}+V_{nt},\\
&\texttt{SL}_2: && Y_{nt}=\lambda_{t0}W_nY_{nt}+X_{nt}\beta_{t0}+c_{n0}+\alpha_{t0}\ell_n+V_{nt},\\
&\texttt{SLE}_1: && Y_{nt}=\lambda_{t0}W_nY_{nt}+X_{nt}\beta_{t0}+c_{n0}+U_{nt},\ U_{nt}=\rho_0M_nU_{nt}+V_{nt},\\
&\texttt{SLE}_2: && Y_{nt}=\lambda_{t0}W_nY_{nt}+X_{nt}\beta_{t0}+c_{n0}+\alpha_{t0}\ell_n+U_{nt},U_{nt}=\rho_0M_nU_{nt}+V_{nt},
\end{aligned}$$

where $t=1,\dots T$.

In the Monte Carlo experiments, we choose $n=(50,150,500)$, and $T=(3,6)$. For $T=3$, we set $\beta'_{10}=(1.0,1.0)$, $\beta'_{20}=(0.75,1.25)$, $\beta'_{30}=(1.25,0.75)$, and $\boldsymbol{\lambda}'_0=(0.5,0.25,0.75)$; and for $T=6$, we set $\boldsymbol{\beta}'_0=(1.0,1.0;0.75,1.25;1.25,0.75;1.0,1.0;0.75,$

$1.25; 1.25, 0.75)$, and $\boldsymbol{\lambda}_0' = (0.5, 0.25, 0.75, 0.5, 0.25, 0.75)$. Finally, $\sigma_0 = 1$ and $\rho_0 = 0.5$. The details of generating idiosyncratic errors, weight matrices and regressors are as follows.

Weight matrices: We use three different methods for generating the spatial weights matrices (i) `Rook Contiguity`, (ii) `Queen Contiguity`, and (iii) `Group Interaction`. The degree of spatial interactions (number of neighbors each unit has) specified by layouts $(i) - (ii)$ are all fixed while in (iii) it may grow with the sample size. This is attained by allowing the number of groups, G, in the sample of spatial units to be directly related to the sample size n, e.g., $G = n^{0.5}$. Hence, the average group size, $m = n/G$, gives a measure of the degree of spatial dependence among the n spatial units. The actual sizes of the groups are generated from a discrete uniform distribution from $.5m$ to $1.5m$.

Regressors: The exogenous regressors are generated according to `REG1`: $X_{knt} \overset{iid}{\sim} N(0, \sigma_\tau^2)$, which are independent across $k = 1, 2$, and $t = 1, \ldots, T$. The σ_τ^2 is the key parameter that controls the variability of the regressors, and thus the signal-to-noise ratio. In case we set σ_τ^2 equals to 1, X_{knt} is generated from standard normal distribution. In case when the spatial dependence is in the form of group interaction, the regressors can also be generated according to `REG2`: the *i*th value of the *k*th regressor in the *g*th group is such that $X_{kt,ig} \overset{iid}{\sim} (2z_g + z_{ig})/\sqrt{10}$, where $(z_g, z_{i,g}) \overset{iid}{\sim} N(0, 1)$ when group interaction scheme is followed; $\{X_{kt,ig}\}$ are thus independent across *k* and *t*, but not across i.

Error Distribution: $v_{it} = \sigma_0 e_{it}$, are generated according to `err1`: $\{e_{it}\}$ are iid standard normal; `err2`: $\{e_{it}\}$ are iid normal mixture with 10% of values from $N(0, 4)$ and the remaining from $N(0, 1)$, standardized to have mean 0 and variance 1; and `err3`: $\{e_{it}\}$ iid chi-square with 3 degrees of freedom, standardized to have mean 0 and variance 1.[12]

Monte Carlo (empirical) means and standard deviations (sds) are reported for the AQS estimators and the QML-estimator of σ^2 (as the AQS and QML estimators of other parameters are numerically identical). Empirical averages of the standard errors (ses): $\widehat{se}$ based on $I^{\circ -1}(\hat{\theta})\widehat{\Sigma}^{\circ}(\hat{\theta})I^{\circ -1}(\hat{\theta})$, are also reported for the proposed AQS estimators. Due

[12]See Yang (2015a) for more details on generating idiosyncratic errors, weight matrices and regressors.

to the space constraint, partial Monte Carlo results are reported in Tables 3.1 & 3.3 for the panel SLE models with 2FE and 1FE, respectively, and Tables 3.2 & 3.4 for the panel SL models with 2FE and 1FE, respectively. The results show the following patterns:

(i) For the case where the QMLE is inconsistent for σ^2, AQSE provides an alternative with consistency and efficiency. The ML-estimator for σ^2 can be quite biased, even if n increases, it does not show a sign of convergence. The AQSE of σ^2 is significantly less biased.

(ii) For the case where the QML and AQS estimates of the parameters are very similar, only the AQS estimates are reported. These parameters includes the time-varying covariate effects $\boldsymbol{\beta}$, time-varying spatial lag effects $\boldsymbol{\lambda}$ and the spatial error effect ρ if the model contains. Denote $S^c(\delta)$ as the concentrated score function under QMLE. The reason of similarity is because the resulting concentrated score functions, one is the concentrated AQS function $S^{\star c}(\delta)$, and the other is $S^c(\delta)$, are the same when equating to zero.

(iii) The estimates of the AQSE-based standard errors perform well with the values are on average very close to the corresponding Monte Carlo sds, except the ses of σ^2 when errors are nonnormal where it is relatively smaller.

(iv) The AQS-estimators of the spatial parameters may converge slower due to: (i)the intrinsic nature of the score-type estimation of the spatial effects, and (ii) a stronger spatial error dependence, such as replacing the weight matrices by `Group interaction`.

(v) The results clearly show that as n and T get larger, the AQS-estimators converge faster. To summarize, the AQS-estimation performs better than the ML-estimation, especially when T is small.

In summary, the proposed AQS-estimation is reliable and easy to apply, and hence is recommended for the applied researchers.

Table 3.1a. Empirical Mean (sd) [se] of AQS-Estimator: SLE Model

Two-Way FE, W_n and M_n: Queen Contiguity, $T = 3$

		$n = 50$	$n = 150$	$n = 500$
(a) Normal Error				
β_1	1.00	1.0000 (.010) [.010]	1.0001 (.005) [.005]	1.0000 (.003) [.003]
	1.00	.9999 (.010) [.009]	1.0000 (.005) [.005]	.9999 (.003) [.003]
β_2	0.75	.7499 (.010) [.009]	.7500 (.005) [.005]	.7500 (.003) [.003]
	1.25	1.2500 (.009) [.008]	1.2500 (.005) [.005]	1.2500 (.003) [.003]
β_3	1.25	1.2500 (.008) [.007]	1.2500 (.006) [.006]	1.2500 (.003) [.003]
	0.75	.7501 (.008) [.007]	.7500 (.005) [.005]	.7501 (.003) [.003]
λ	0.50	.4999 (.010) [.010]	.4999 (.007) [.006]	.4999 (.003) [.003]
	0.25	.2498 (.016) [.015]	.2499 (.007) [.007]	.2500 (.004) [.004]
	0.75	.7500 (.007) [.007]	.7500 (.004) [.004]	.7500 (.002) [.002]
ρ	0.50	.4945 (.155) [.133]	.4992 (.076) [.074]	.5000 (.041) [.040]
σ^2	1.00	.8913 (.133) [.127]	.9643 (.083) [.080]	.9894 (.045) [.045]
σ^2	MLE	.5942 (.089)	.6428 (.055)	.6596 (.030)
(b) Normal Mixture Error				
β_1	1.00	1.0000 (.010) [.010]	1.0000 (.005) [.005]	1.0000 (.003) [.003]
	1.00	.9999 (.010) [.009]	.9999 (.005) [.005]	.9999 (.003) [.003]
β_2	0.75	.7498 (.010) [.009]	.7500 (.005) [.005]	.7500 (.003) [.003]
	1.25	1.2500 (.008) [.008]	1.2500 (.005) [.005]	1.2500 (.003) [.003]
β_3	1.25	1.2499 (.008) [.007]	1.2500 (.006) [.006]	1.2500 (.003) [.003]
	0.75	.7501 (.008) [.007]	.7500 (.005) [.005]	.7501 (.003) [.003]
λ	0.50	.5000 (.010) [.010]	.5000 (.007) [.006]	.5000 (.003) [.003]
	0.25	.2497 (.016) [.015]	.2499 (.007) [.007]	.2500 (.004) [.004]
	0.75	.7499 (.007) [.007]	.7499 (.004) [.004]	.7500 (.002) [.002]
ρ	0.50	.4982 (.153) [.132]	.5013 (.075) [.074]	.5006 (.041) [.040]
σ^2	1.00	.8944 (.266) [.147]	.9630 (.162) [.093]	.9886 (.092) [.051]
σ^2	MLE	.5963 (.177)	.6420 (.108)	.6591 (.061)
(c) Chi-Square-normal Error				
β_1	1.00	.9999 (.010) [.010]	1.0000 (.005) [.005]	1.0000 (.003) [.003]
	1.00	.9998 (.010) [.009]	1.0001 (.005) [.005]	1.0000 (.003) [.003]
β_2	0.75	.7500 (.010) [.009]	.7500 (.005) [.005]	.7501 (.003) [.003]
	1.25	1.2502 (.008) [.008]	1.2499 (.005) [.005]	1.2500 (.003) [.003]
β_3	1.25	1.2500 (.008) [.007]	1.2501 (.006) [.006]	1.2500 (.003) [.003]
	0.75	.7501 (.008) [.007]	.7501 (.005) [.005]	.7500 (.003) [.003]
λ	0.50	.5001 (.010) [.010]	.5001 (.006) [.006]	.5000 (.003) [.003]
	0.25	.2497 (.016) [.015]	.2502 (.007) [.007]	.2500 (.004) [.004]
	0.75	.7499 (.007) [.007]	.7499 (.004) [.004]	.7500 (.002) [.002]
ρ	0.50	.4969 (.151) [.132]	.5013 (.077) [.074]	.4997 (.041) [.040]
σ^2	1.00	.8929 (.201) [.136]	.9637 (.122) [.085]	.9890 (.069) [.048]
σ^2	QMLE	.5953 (.134)	.6425 (.081)	.6593 (.046)

Table 3.1b. Empirical Mean (sd) [se] of AQS-Estimator: `SLE` Model
Two-Way FE, W_n and M_n: Queen Contiguity, $T = 6$

		$n = 50$	$n = 150$	$n = 500$
		(a) Normal Error		
β_1	1.00	.9998 (.009) [.009]	1.0000 (.005) [.005]	1.0000 (.002) [.002]
	1.00	1.0002 (.007) [.007]	1.0000 (.004) [.004]	1.0000 (.002) [.002]
β_2	0.75	.7499 (.008) [.008]	.7499 (.005) [.005]	.7501 (.003) [.003]
	1.25	1.2498 (.008) [.007]	1.2498 (.004) [.004]	1.2500 (.002) [.002]
β_3	1.25	1.2500 (.007) [.007]	1.2500 (.005) [.005]	1.2500 (.003) [.003]
	0.75	.7500 (.008) [.007]	.7501 (.004) [.004]	.7501 (.002) [.002]
β_4	1.00	1.0001 (.009) [.008]	1.0000 (.004) [.004]	.9999 (.003) [.002]
	1.00	.9999 (.007) [.007]	.9999 (.005) [.005]	1.0001 (.002) [.002]
β_5	0.75	.7502 (.008) [.007]	.7500 (.004) [.004]	.7500 (.002) [.002]
	1.25	1.2498 (.008) [.008]	1.2500 (.005) [.005]	1.2500 (.002) [.002]
β_6	1.25	1.2499 (.007) [.007]	1.2500 (.005) [.005]	1.2500 (.003) [.003]
	0.75	.7500 (.007) [.007]	.7500 (.004) [.004]	.7500 (.002) [.002]
λ	0.50	.4997 (.011) [.010]	.4999 (.005) [.005]	.5000 (.003) [.003]
	0.25	.2498 (.012) [.012]	.2500 (.005) [.005]	.2501 (.003) [.003]
	0.75	.7501 (.007) [.006]	.7499 (.004) [.004]	.7500 (.002) [.002]
	0.50	.5000 (.010) [.009]	.5000 (.006) [.006]	.5000 (.003) [.003]
	0.25	.2500 (.009) [.008]	.2500 (.006) [.006]	.2500 (.003) [.004]
	0.75	.7497 (.009) [.008]	.7499 (.004) [.004]	.7500 (.002) [.002]
ρ	0.50	.5137 (.090) [.083]	.5040 (.048) [.047]	.5018 (.026) [.025]
σ^2	1.00	.9182 (.087) [.083]	.9728 (.052) [.051]	.9913 (.029) [.029]
σ^2	QMLE	.7651 (.073)	.8106 (.044)	.8261 (.024)
		(b) Normal Mixture Error		
β_1	1.00	.9998 (.009) [.009]	1.0001 (.005) [.005]	1.0000 (.002) [.002]
	1.00	1.0001 (.007) [.007]	1.0000 (.004) [.004]	1.0000 (.002) [.002]
β_2	0.75	.7498 (.008) [.008]	.7499 (.005) [.005]	.7501 (.003) [.003]
	1.25	1.2498 (.008) [.007]	1.2499 (.004) [.004]	1.2500 (.002) [.002]
β_3	1.25	1.2500 (.007) [.007]	1.2499 (.005) [.005]	1.2501 (.003) [.003]
	0.75	.7501 (.008) [.007]	.7500 (.004) [.004]	.7501 (.002) [.002]
β_4	1.00	1.0002 (.009) [.008]	1.0000 (.004) [.004]	.9999 (.003) [.002]
	1.00	.9999 (.007) [.007]	.9999 (.005) [.005]	1.0001 (.002) [.002]
β_5	0.75	.7502 (.008) [.007]	.7500 (.004) [.004]	.7500 (.002) [.002]
	1.25	1.2500 (.008) [.008]	1.2500 (.005) [.005]	1.2500 (.002) [.002]
β_6	1.25	1.2500 (.007) [.007]	1.2500 (.005) [.005]	1.2500 (.003) [.003]
	0.75	.7499 (.007) [.007]	.7501 (.004) [.004]	.7500 (.002) [.002]
λ	0.50	.4997 (.011) [.010]	.4999 (.005) [.005]	.5000 (.003) [.003]
	0.25	.2497 (.012) [.012]	.2499 (.005) [.005]	.2501 (.003) [.003]
	0.75	.7500 (.007) [.006]	.7500 (.004) [.004]	.7500 (.002) [.002]
	0.50	.5000 (.010) [.009]	.5000 (.006) [.005]	.5001 (.003) [.003]
	0.25	.2501 (.009) [.008]	.2499 (.006) [.006]	.2500 (.004) [.004]
	0.75	.7499 (.009) [.008]	.7499 (.004) [.004]	.7500 (.002) [.002]
ρ	0.50	.5143 (.089) [.083]	.5041 (.048) [.047]	.5018 (.026) [.026]
σ^2	1.00	.9194 (.184) [.109]	.9730 (.113) [.066]	.9929 (.064) [.035]
σ^2	QMLE	.7662 (.153)	.8108 (.094)	.8274 (.053)

Table 3.1b (cont'd). Empirical Mean (sd) [se] of AQS-Estimator:
SLE Model Two-Way FE, W_n and M_n: Queen Contiguity, $T = 6$

		$n = 50$	$n = 150$	$n = 500$
		(c) Chi-Square-normal Error		
β_{11}	1.00	.9999 (.009) [.009]	1.0001 (.005) [.005]	1.0000 (.002) [.002]
	1.00	1.0001 (.007) [.007]	1.0001 (.004) [.004]	1.0000 (.002) [.002]
β_2	0.75	.7500 (.008) [.008]	.7501 (.005) [.005]	.7500 (.003) [.003]
	1.25	1.2500 (.008) [.007]	1.2500 (.004) [.004]	1.2500 (.002) [.002]
β_3	1.25	1.2501 (.007) [.007]	1.2499 (.005) [.005]	1.2500 (.003) [.003]
	0.75	.7499 (.008) [.007]	.7501 (.004) [.004]	.7500 (.002) [.002]
β_4	1.00	.9998 (.009) [.008]	1.0001 (.004) [.004]	1.0000 (.003) [.002]
	1.00	.9999 (.007) [.007]	1.0000 (.005) [.005]	1.0000 (.002) [.002]
β_5	0.75	.7502 (.008) [.007]	.7500 (.004) [.004]	.7500 (.002) [.002]
	1.25	1.2499 (.008) [.008]	1.2500 (.005) [.005]	1.2500 (.002) [.002]
β_6	1.25	1.2501 (.007) [.007]	1.2499 (.005) [.005]	1.2500 (.002) [.003]
	0.75	.7498 (.007) [.007]	.7500 (.004) [.004]	.7500 (.002) [.002]
λ	0.50	.4998 (.011) [.010]	.5000 (.005) [.005]	.5000 (.003) [.003]
	0.25	.2497 (.012) [.012]	.2500 (.005) [.005]	.2499 (.003) [.003]
	0.75	.7499 (.006) [.006]	.7499 (.004) [.004]	.7500 (.002) [.002]
	0.50	.4999 (.010) [.009]	.4999 (.006) [.006]	.5000 (.003) [.003]
	0.25	.2500 (.009) [.008]	.2500 (.006) [.006]	.2500 (.004) [.004]
	0.75	.7497 (.009) [.008]	.7499 (.004) [.004]	.7500 (.002) [.002]
ρ	0.50	.5175 (.087) [.083]	.5034 (.048) [.047]	.5010 (.026) [.026]
σ^2	1.00	.9154 (.138) [.094]	.9728 (.083) [.058]	.9924 (.046) [.032]
σ^2	QMLE	.7628 (.115)	.8106 (.069)	.8270 (.038)

Table 3.2a. Empirical Mean (sd) [se] of AQS-Estimator: SL Model
Two-Way FE, W_n: Rook Contiguity, $T = 3$

		$n = 50$	$n = 150$	$n = 500$
(a) Normal Error				
β_1	1.00	1.0000 (.020) [.019]	1.0001 (.010) [.010]	.9999 (.006) [.006]
	1.00	.9997 (.017) [.016]	.9999 (.012) [.012]	1.0000 (.006) [.006]
β_2	0.75	.7503 (.020) [.019]	.7501 (.010) [.010]	.7499 (.005) [.005]
	1.25	1.2500 (.021) [.020]	1.2498 (.011) [.011]	1.2499 (.006) [.006]
β_3	1.25	1.2502 (.018) [.018]	1.2500 (.010) [.010]	1.2500 (.006) [.006]
	0.75	.7498 (.017) [.016]	.7502 (.011) [.011]	.7500 (.005) [.005]
λ	0.50	.4996 (.015) [.014]	.4999 (.011) [.011]	.5000 (.005) [.005]
	0.25	.2497 (.019) [.019]	.2496 (.013) [.013]	.2500 (.006) [.005]
	0.75	.7491 (.020) [.019]	.7499 (.006) [.006]	.7500 (.003) [.003]
σ^2	1.00	.9075 (.134) [.128]	.9690 (.082) [.079]	.9909 (.045) [.044]
σ^2	QMLE	.6050 (.089)	.6460 (.054)	.6606 (.030)
(b) Normal Mixture Error				
β_1	1.00	1.0002 (.020) [.019]	1.0000 (.010) [.010]	.9999 (.006) [.006]
	1.00	.9999 (.017) [.016]	.9999 (.012) [.012]	.9999 (.006) [.006]
β_2	0.75	.7501 (.020) [.019]	.7502 (.010) [.010]	.7499 (.005) [.005]
	1.25	1.2502 (.021) [.020]	1.2499 (.011) [.011]	1.2500 (.006) [.006]
β_3	1.25	1.2502 (.018) [.018]	1.2499 (.010) [.010]	1.2500 (.006) [.006]
	0.75	.7500 (.017) [.016]	.7500 (.011) [.011]	.7500 (.005) [.005]
λ	0.50	.4996 (.015) [.014]	.5000 (.011) [.011]	.5000 (.005) [.005]
	0.25	.2498 (.019) [.019]	.2498 (.013) [.012]	.2499 (.006) [.005]
	0.75	.7493 (.020) [.019]	.7500 (.006) [.006]	.7500 (.003) [.003]
σ^2	1.00	.9090 (.268) [.149]	.9679 (.162) [.092]	.9901 (.092) [.051]
σ^2	QMLE	.6060 (.179)	.6453 (.108)	.6601 (.061)
(c) Chi-Square-normal Error				
β_1	1.00	1.0001 (.020) [.019]	.9998 (.010) [.010]	1.0000 (.006) [.006]
	1.00	.9999 (.017) [.016]	.9996 (.012) [.012]	1.0000 (.005) [.006]
β_2	0.75	.7505 (.020) [.019]	.7499 (.010) [.010]	.7500 (.005) [.005]
	1.25	1.2503 (.021) [.020]	1.2499 (.011) [.011]	1.2500 (.006) [.006]
β_3	1.25	1.2501 (.019) [.018]	1.2501 (.010) [.010]	1.2501 (.006) [.006]
	0.75	.7495 (.017) [.016]	.7501 (.011) [.011]	.7501 (.005) [.005]
λ	0.50	.4999 (.015) [.014]	.4999 (.011) [.011]	.5000 (.005) [.005]
	0.25	.2494 (.020) [.019]	.2500 (.013) [.012]	.2500 (.006) [.005]
	0.75	.7492 (.020) [.019]	.7498 (.006) [.006]	.7499 (.003) [.003]
σ^2	1.00	.9066 (.202) [.137]	.9688 (.121) [.085]	.9905 (.068) [.047]
σ^2	QMLE	.6044 (.135)	.6459 (.080)	.6603 (.046)

Table 3.2b. Empirical Mean (sd) [se] of AQS-Estimator: SL Model
Two-Way FE, W_n: Rook Contiguity, $T = 6$

		$n = 50$	$n = 150$	$n = 500$
(a) Normal Error				
β_1	1.00	1.0002 (.017) [.017]	1.0001 (.009) [.009]	1.0000 (.005) [.005]
	1.00	.9996 (.015) [.015]	1.0000 (.009) [.009]	.9999 (.005) [.005]
β_2	0.75	.7503 (.019) [.018]	.7502 (.010) [.010]	.7501 (.005) [.005]
	1.25	1.2500 (.015) [.014]	1.2500 (.009) [.009]	1.2501 (.005) [.005]
β_3	1.25	1.2509 (.017) [.017]	1.2499 (.010) [.010]	1.2499 (.005) [.005]
	0.75	.7506 (.016) [.015]	.7501 (.009) [.009]	.7500 (.005) [.005]
β_4	1.00	1.0000 (.015) [.014]	.9998 (.010) [.010]	1.0002 (.005) [.005]
	1.00	.9998 (.015) [.015]	.9997 (.009) [.009]	1.0001 (.005) [.005]
β_5	0.75	.7497 (.018) [.017]	.7500 (.009) [.009]	.7499 (.005) [.005]
	1.25	1.2506 (.016) [.015]	1.2502 (.009) [.009]	1.2500 (.005) [.005]
β_6	1.25	1.2506 (.017) [.016]	1.2503 (.009) [.009]	1.2501 (.005) [.005]
	0.75	.7497 (.017) [.016]	.7503 (.009) [.009]	.7501 (.005) [.005]
λ	0.50	.4998 (.017) [.016]	.4996 (.008) [.008]	.4999 (.005) [.005]
	0.25	.2497 (.014) [.014]	.2497 (.012) [.011]	.2498 (.005) [.005]
	0.75	.7491 (.020) [.019]	.7498 (.006) [.006]	.7500 (.003) [.004]
	0.50	.5003 (.015) [.015]	.5006 (.010) [.010]	.4998 (.005) [.005]
	0.25	.2500 (.017) [.016]	.2497 (.010) [.010]	.2498 (.006) [.006]
	0.75	.7496 (.010) [.009]	.7498 (.007) [.007]	.7500 (.003) [.003]
σ^2	1.00	.9272 (.088) [.083]	.9763 (.052) [.050]	.9923 (.028) [.028]
σ^2	QMLE	.7727 (.073)	.8136 (.044)	.8270 (.023)
(b) Normal Mixture Error				
β_1	1.00	1.0001 (.018) [.016]	1.0001 (.009) [.009]	1.0001 (.005) [.005]
	1.00	.9995 (.015) [.015]	.9999 (.009) [.009]	.9999 (.005) [.005]
β_2	0.75	.7505 (.019) [.018]	.7502 (.010) [.010]	.7501 (.005) [.005]
	1.25	1.2502 (.015) [.014]	1.2502 (.009) [.009]	1.2500 (.005) [.005]
β_3	1.25	1.2507 (.017) [.017]	1.2501 (.010) [.010]	1.2500 (.005) [.005]
	0.75	.7503 (.016) [.015]	.7498 (.009) [.009]	.7501 (.005) [.005]
β_4	1.00	.9998 (.015) [.014]	1.0000 (.011) [.010]	1.0002 (.005) [.005]
	1.00	.9998 (.016) [.015]	.9998 (.009) [.009]	1.0000 (.005) [.005]
β_5	0.75	.7495 (.018) [.017]	.7500 (.009) [.009]	.7499 (.005) [.005]
	1.25	1.2503 (.015) [.015]	1.2501 (.009) [.009]	1.2501 (.005) [.005]
β_6	1.25	1.2506 (.017) [.016]	1.2506 (.009) [.009]	1.2500 (.005) [.005]
	0.75	.7498 (.017) [.016]	.7501 (.009) [.009]	.7500 (.005) [.005]
λ	0.50	.4996 (.017) [.016]	.4998 (.008) [.008]	.4999 (.005) [.005]
	0.25	.2497 (.015) [.014]	.2498 (.012) [.011]	.2498 (.005) [.005]
	0.75	.7493 (.019) [.018]	.7497 (.006) [.006]	.7500 (.004) [.004]
	0.50	.5002 (.016) [.015]	.5007 (.010) [.010]	.4999 (.005) [.005]
	0.25	.2498 (.017) [.016]	.2497 (.010) [.010]	.2498 (.006) [.006]
	0.75	.7496 (.010) [.009]	.7499 (.007) [.007]	.7500 (.003) [.003]
σ^2	1.00	.9249 (.183) [.109]	.9749 (.113) [.065]	.9946 (.066) [.035]
σ^2	QMLE	.7707 (.153)	.8124 (.094)	.8289 (.055)

Table 3.2b (cont'd). Empirical Mean (sd) [se] of AQS-Estimator: SL Model Two-Way FE, W_n: Rook Contiguity, $T = 6$

		$n = 50$	$n = 150$	$n = 500$
(c) Chi-Square-normal Error				
β_1	1.00	.9997 (.017) [.016]	.9997 (.009) [.009]	1.0002 (.005) [.005]
	1.00	.9999 (.014) [.015]	.9999 (.009) [.009]	1.0000 (.005) [.005]
β_2	0.75	.7499 (.019) [.018]	.7507 (.010) [.010]	.7502 (.005) [.005]
	1.25	1.2504 (.015) [.014]	1.2501 (.009) [.009]	1.2498 (.005) [.005]
β_3	1.25	1.2498 (.017) [.017]	1.2499 (.010) [.010]	1.2501 (.005) [.005]
	0.75	.7498 (.016) [.015]	.7498 (.009) [.009]	.7502 (.005) [.005]
β_4	1.00	1.0001 (.015) [.014]	.9997 (.010) [.010]	1.0000 (.005) [.005]
	1.00	.9999 (.016) [.015]	1.0003 (.009) [.009]	1.0000 (.005) [.005]
β_5	0.75	.7500 (.018) [.017]	.7497 (.009) [.009]	.7500 (.005) [.005]
	1.25	1.2499 (.016) [.015]	1.2501 (.009) [.009]	1.2500 (.005) [.005]
β_6	1.25	1.2500 (.016) [.016]	1.2496 (.009) [.009]	1.2500 (.005) [.005]
	0.75	.7503 (.017) [.016]	.7501 (.009) [.009]	.7501 (.005) [.005]
λ	0.50	.5004 (.016) [.016]	.4998 (.008) [.008]	.5000 (.005) [.005]
	0.25	.2498 (.014) [.014]	.2496 (.011) [.011]	.2499 (.005) [.005]
	0.75	.7491 (.020) [.019]	.7501 (.006) [.006]	.7500 (.004) [.004]
	0.50	.4993 (.015) [.015]	.4999 (.010) [.010]	.5000 (.005) [.005]
	0.25	.2498 (.017) [.016]	.2503 (.010) [.010]	.2500 (.006) [.006]
	0.75	.7498 (.010) [.009]	.7501 (.007) [.007]	.7501 (.003) [.003]
σ^2	1.00	.9250 (.139) [.094]	.9743 (.082) [.057]	.9942 (.046) [.031]
σ^2	QMLE	.7708 (.116)	.8119 (.068)	.8285 (.038)

Table 3.3a. Empirical Mean (sd) [se] of AQS-Estimator: SLE Model
One-Way FE, W_n and M_n: Group Interaction, $T = 3$

		$n = 50$	$n = 150$	$n = 500$
		(a) Normal Error		
β_1	1.00	1.0627 (.419) [.412]	1.0341 (.224) [.226]	1.0203 (.123) [.125]
	1.00	0.9711 (.604) [.597]	1.0501 (.229) [.231]	1.0184 (.141) [.141]
β_2	0.75	0.7662 (.388) [.378]	0.7734 (.195) [.195]	0.7632 (.116) [.116]
	1.25	1.3415 (.435) [.441]	1.2659 (.259) [.260]	1.2632 (.142) [.144]
β_3	1.25	1.3356 (.493) [.485]	1.2846 (.260) [.263]	1.2683 (.142) [.143]
	0.75	0.8190 (.389) [.374]	0.7697 (.256) [.254]	0.7628 (.117) [.116]
λ	0.50	0.4264 (.166) [.170]	0.4657 (.090) [.092]	0.4872 (.052) [.053]
	0.25	0.1470 (.220) [.231]	0.2024 (.130) [.131]	0.2340 (.073) [.073]
	0.75	0.7087 (.090) [.088]	0.7303 (.051) [.051]	0.7444 (.024) [.025]
ρ	0.50	0.4436 (.199) [.193]	0.4777 (.104) [.105]	0.4883 (.065) [.064]
σ^2	1.00	0.9018 (.137) [.133]	0.9688 (.082) [.080]	0.9906 (.045) [.045]
σ^2	QMLE	0.6012 (.091)	0.6459 (.055)	0.6604 (.030)
		(b) Normal Mixture Error		
β_1	1.00	1.0629 (.419) [.411]	1.0317 (.225) [.225]	1.0208 (.124) [.125]
	1.00	0.9716 (.613) [.594]	1.0485 (.229) [.231]	1.0178 (.140) [.141]
β_2	0.75	0.7688 (.389) [.377]	0.7737 (.198) [.195]	0.7645 (.117) [.115]
	1.25	1.3357 (.439) [.440]	1.2642 (.257) [.259]	1.2650 (.143) [.144]
β_3	1.25	1.3350 (.496) [.485]	1.2900 (.261) [.264]	1.2676 (.140) [.143]
	0.75	0.8158 (.392) [.373]	0.7695 (.255) [.254]	0.7624 (.116) [.116]
λ	0.50	0.4263 (.168) [.173]	0.4665 (.091) [.093]	0.4877 (.052) [.053]
	0.25	0.1541 (.215) [.230]	0.2037 (.129) [.130]	0.2336 (.073) [.073]
	0.75	0.7056 (.092) [.090]	0.7300 (.051) [.051]	0.7447 (.024) [.025]
ρ	0.50	0.4409 (.198) [.198]	0.4772 (.105) [.106]	0.4876 (.064) [.065]
σ^2	1.00	0.9037 (.269) [.188]	0.9680 (.162) [.123]	0.9899 (.092) [.071]
σ^2	QMLE	0.6012 (.091)	0.6459 (.055)	0.6604 (.030)
		(c) Chi-Square-normal Error		
β_1	1.00	1.0590 (.413) [.414]	1.0282 (.224) [.225]	1.0191 (.123) [.125]
	1.00	0.9870 (.614) [.595]	1.0479 (.231) [.230]	1.0189 (.138) [.141]
β_2	0.75	0.7661 (.387) [.379]	0.7761 (.200) [.195]	0.7609 (.116) [.116]
	1.25	1.3276 (.430) [.440]	1.2681 (.259) [.259]	1.2644 (.142) [.144]
β_3	1.25	1.3285 (.489) [.489]	1.2831 (.265) [.264]	1.2627 (.139) [.143]
	0.75	0.8187 (.396) [.378]	0.7672 (.260) [.253]	0.7605 (.116) [.116]
λ	0.50	0.4214 (.168) [.175]	0.4662 (.089) [.093]	0.4874 (.051) [.053]
	0.25	0.1537 (.223) [.229]	0.1985 (.132) [.131]	0.2338 (.072) [.073]
	0.75	0.7056 (.091) [.089]	0.7297 (.051) [.051]	0.7451 (.024) [.025]
ρ	0.50	0.4474 (.196) [.196]	0.4767 (.104) [.106]	0.4889 (.066) [.064]
σ^2	1.00	0.9044 (.205) [.159]	0.9682 (.122) [.100]	0.9904 (.069) [.057]
σ^2	QMLE	0.6012 (.091)	0.6459 (.055)	0.6604 (.030)

Table 3.3b. Empirical Mean (sd) [se] of AQS-Estimator: `SLE` Model
One-Way FE, W_n and M_n: Group Interaction, $T = 6$

		$n = 50$	$n = 150$	$n = 500$
		(a) Normal Error		
β_1	1.00	1.0510 (.364) [.352]	1.0244 (.191) [.192]	1.0067 (.117) [.118]
	1.00	0.9941 (.526) [.513]	1.0222 (.219) [.218]	1.0133 (.116) [.116]
β_2	0.75	0.7642 (.332) [.323]	0.7768 (.178) [.178]	0.7618 (.105) [.103]
	1.25	1.3237 (.377) [.374]	1.2653 (.274) [.269]	1.2654 (.136) [.138]
β_3	1.25	1.3165 (.438) [.426]	1.2949 (.206) [.209]	1.2663 (.119) [.121]
	0.75	0.7971 (.334) [.318]	0.7753 (.166) [.166]	0.7608 (.107) [.109]
β_4	1.00	1.0369 (.383) [.372]	1.0290 (.184) [.179]	1.0128 (.110) [.111]
	1.00	1.0804 (.289) [.287]	1.0129 (.206) [.203]	1.0143 (.111) [.112]
β_5	0.75	0.8167 (.292) [.290]	0.7854 (.170) [.173]	0.7598 (.117) [.118]
	1.25	1.3655 (.308) [.316]	1.2976 (.191) [.192]	1.2597 (.118) [.120]
β_6	1.25	1.3013 (.435) [.423]	1.2975 (.223) [.226]	1.2610 (.122) [.126]
	0.75	0.7633 (.404) [.393]	0.7584 (.178) [.173]	0.7582 (.103) [.104]
λ	0.50	0.4331 (.144) [.138]	0.4661 (.084) [.082]	0.4895 (.047) [.047]
	0.25	0.1527 (.189) [.192]	0.1992 (.129) [.125]	0.2348 (.067) [.067]
	0.75	0.7130 (.074) [.069]	0.7348 (.039) [.039]	0.7441 (.023) [.024]
	0.50	0.4364 (.131) [.131]	0.4701 (.082) [.079]	0.4884 (.047) [.048]
	0.25	0.1663 (.165) [.174]	0.2117 (.104) [.108]	0.2344 (.068) [.069]
	0.75	0.7169 (.072) [.069]	0.7347 (.041) [.041]	0.7454 (.021) [.021]
ρ	0.50	0.4939 (.093) [.096]	0.4944 (.055) [.057]	0.4942 (.039) [.039]
σ^2	1.00	0.9273 (.088) [.086]	0.9768 (.052) [.051]	0.9927 (.028) [.028]
σ^2	QMLE	0.7728 (.074)	0.8140 (.044)	0.8273 (.024)
		(b) Normal Mixture Error		
β_1	1.00	1.0566 (.360) [.353]	1.0282 (.190) [.192]	1.0059 (.117) [.118]
	1.00	0.9893 (.520) [.511]	1.0270 (.219) [.218]	1.0130 (.116) [.117]
β_2	0.75	0.7652 (.332) [.323]	0.7810 (.181) [.179]	0.7621 (.106) [.104]
	1.25	1.3196 (.368) [.374]	1.2670 (.274) [.269]	1.2630 (.135) [.138]
β_3	1.25	1.3181 (.441) [.430]	1.2988 (.208) [.211]	1.2652 (.121) [.122]
	0.75	0.7959 (.334) [.318]	0.7758 (.167) [.167]	0.7605 (.108) [.109]
β_4	1.00	1.0376 (.384) [.374]	1.0277 (.181) [.180]	1.0133 (.111) [.111]
	1.00	1.0792 (.285) [.289]	1.0134 (.207) [.203]	1.0132 (.111) [.112]
β_5	0.75	0.8127 (.291) [.291]	0.7834 (.171) [.173]	0.7614 (.118) [.118]
	1.25	1.3616 (.301) [.318]	1.2963 (.190) [.193]	1.2619 (.119) [.120]
β_6	1.25	1.3047 (.440) [.427]	1.2998 (.230) [.229]	1.2606 (.122) [.126]
	0.75	0.7577 (.407) [.394]	0.7622 (.179) [.173]	0.7582 (.103) [.104]
λ	0.50	0.4311 (.141) [.142]	0.4653 (.084) [.083]	0.4892 (.046) [.047]
	0.25	0.1601 (.185) [.195]	0.1975 (.128) [.126]	0.2362 (.066) [.067]
	0.75	0.7114 (.077) [.073]	0.7345 (.039) [.039]	0.7443 (.024) [.024]
	0.50	0.4354 (.134) [.135]	0.4712 (.080) [.080]	0.4882 (.047) [.049]
	0.25	0.1671 (.165) [.176]	0.2123 (.105) [.108]	0.2336 (.069) [.069]
	0.75	0.7159 (.073) [.073]	0.7346 (.042) [.042]	0.7455 (.021) [.022]
ρ	0.50	0.4901 (.097) [.101]	0.4949 (.057) [.058]	0.4941 (.039) [.039]
σ^2	1.00	0.9292 (.187) [.151]	0.9765 (.113) [.096]	0.9943 (.064) [.055]
σ^2	QMLE	0.7743 (.156)	0.8138 (.094)	0.8286 (.053)

Table 3.3b (cont'd). Empirical Mean (sd) [se] of AQS-Estimator:
SLE Model One-Way FE, W_n and M_n: Group Interaction, $T = 6$

		$n = 50$	$n = 150$	$n = 500$
		(c) Chi-Square-normal Error		
β_1	1.00	1.0479 (.358) [.351]	1.0215 (.192) [.192]	1.0084 (.118) [.118]
	1.00	1.0059 (.525) [.512]	1.0234 (.219) [.218]	1.0113 (.116) [.117]
β_2	0.75	0.7685 (.336) [.323]	0.7743 (.183) [.179]	0.7578 (.104) [.103]
	1.25	1.3128 (.366) [.374]	1.2645 (.272) [.269]	1.2653 (.138) [.138]
β_3	1.25	1.3174 (.432) [.427]	1.2926 (.206) [.210]	1.2650 (.119) [.121]
	0.75	0.7971 (.343) [.323]	0.7747 (.170) [.167]	0.7596 (.110) [.108]
β_4	1.00	1.0332 (.378) [.372]	1.0385 (.179) [.179]	1.0133 (.112) [.111]
	1.00	1.0735 (.293) [.290]	1.0219 (.205) [.203]	1.0149 (.112) [.112]
β_5	0.75	0.8133 (.292) [.290]	0.7853 (.174) [.173]	0.7582 (.117) [.118]
	1.25	1.3464 (.299) [.317]	1.2960 (.187) [.192]	1.2615 (.118) [.120]
β_6	1.25	1.3150 (.427) [.418]	1.2978 (.223) [.229]	1.2587 (.126) [.125]
	0.75	0.7557 (.408) [.395]	0.7592 (.178) [.174]	0.7555 (.102) [.104]
λ	0.50	0.4315 (.141) [.142]	0.4672 (.082) [.082]	0.4888 (.047) [.047]
	0.25	0.1596 (.190) [.193]	0.1981 (.127) [.125]	0.2350 (.067) [.067]
	0.75	0.7112 (.074) [.070]	0.7357 (.038) [.038]	0.7444 (.024) [.024]
	0.50	0.4336 (.137) [.133]	0.4684 (.080) [.079]	0.4882 (.048) [.049]
	0.25	0.1713 (.163) [.175]	0.2119 (.105) [.107]	0.2362 (.068) [.069]
	0.75	0.7150 (.074) [.072]	0.7362 (.040) [.041]	0.7458 (.021) [.022]
ρ	0.50	0.4929 (.093) [.098]	0.4933 (.056) [.058]	0.4938 (.039) [.039]
σ^2	1.00	0.9260 (.139) [.116]	0.9769 (.082) [.073]	0.9935 (.046) [.041]
σ^2	QMLE	0.7717 (.116)	0.8141 (.069)	0.8279 (.038)

Table 3.4a. Empirical Mean (sd) [se] of AQS-Estimator: `SL` Model
One-Way FE, W_n: Rook Contiguity, $T = 3$

		$n = 50$	$n = 150$	$n = 500$
(a) Normal Error				
β_1	1.00	.9999 (.019) [.018]	1.0001 (.010) [.010]	1.0000 (.006) [.005]
	1.00	.9998 (.017) [.016]	.9999 (.012) [.012]	1.0000 (.006) [.006]
β_2	0.75	.7503 (.020) [.019]	.7501 (.010) [.010]	.7501 (.006) [.006]
	1.25	1.2500 (.021) [.020]	1.2498 (.011) [.011]	1.2500 (.005) [.005]
β_3	1.25	1.2502 (.018) [.017]	1.2500 (.010) [.010]	1.2501 (.006) [.006]
	0.75	.7499 (.017) [.016]	.7502 (.011) [.011]	.7498 (.006) [.005]
λ	0.50	.4996 (.015) [.014]	.4999 (.011) [.011]	.4999 (.005) [.005]
	0.25	.2497 (.019) [.019]	.2496 (.013) [.012]	.2500 (.005) [.005]
	0.75	.7491 (.020) [.019]	.7499 (.006) [.006]	.7500 (.003) [.003]
σ^2	1.00	.9096 (.132) [.127]	.9691 (.081) [.079]	.9909 (.045) [.044]
σ^2	QMLE	.6064 (.088)	.6461 (.054)	.6606 (.030)
(b) Normal Mixture Error				
β_1	1.00	1.0000 (.019) [.018]	1.0000 (.010) [.010]	.9999 (.005) [.005]
	1.00	1.0000 (.017) [.016]	.9999 (.012) [.012]	1.0000 (.006) [.006]
β_2	0.75	.7500 (.020) [.019]	.7502 (.010) [.010]	.7501 (.006) [.006]
	1.25	1.2502 (.021) [.020]	1.2499 (.011) [.011]	1.2500 (.005) [.005]
β_3	1.25	1.2502 (.018) [.017]	1.2499 (.010) [.010]	1.2500 (.006) [.006]
	0.75	.7501 (.017) [.016]	.7500 (.011) [.011]	.7499 (.006) [.005]
λ	0.50	.4996 (.015) [.014]	.5000 (.011) [.011]	.4999 (.005) [.005]
	0.25	.2498 (.019) [.018]	.2498 (.013) [.012]	.2500 (.005) [.005]
	0.75	.7493 (.019) [.019]	.7500 (.006) [.006]	.7500 (.003) [.003]
σ^2	1.00	.9109 (.268) [.184]	.9680 (.162) [.122]	.9902 (.092) [.071]
σ^2	QMLE	.6072 (.178)	.6454 (.108)	.6601 (.061)
(c) Chi-Square-normal Error				
β_1	1.00	1.0002 (.019) [.018]	.9998 (.010) [.010]	.9999 (.005) [.005]
	1.00	.9998 (.017) [.016]	.9996 (.012) [.012]	1.0000 (.006) [.006]
β_2	0.75	.7506 (.020) [.019]	.7500 (.010) [.010]	.7500 (.006) [.006]
	1.25	1.2502 (.020) [.020]	1.2499 (.011) [.011]	1.2501 (.005) [.005]
β_3	1.25	1.2501 (.018) [.017]	1.2501 (.010) [.010]	1.2500 (.006) [.006]
	0.75	.7496 (.017) [.016]	.7501 (.011) [.011]	.7499 (.005) [.005]
λ	0.50	.4998 (.015) [.014]	.4999 (.011) [.011]	.5000 (.005) [.005]
	0.25	.2495 (.020) [.019]	.2500 (.013) [.012]	.2499 (.005) [.005]
	0.75	.7492 (.020) [.019]	.7498 (.006) [.006]	.7500 (.003) [.003]
σ^2	1.00	.9085 (.201) [.154]	.9690 (.121) [.099]	.9906 (.068) [.057]
σ^2	QMLE	.6057 (.134)	.6460 (.080)	.6604 (.045)

Table 3.4b. Empirical Mean (sd) [se] of AQS-Estimator: SL Model
One-Way FE, W_n: Rook Contiguity, $T = 6$

		$n = 50$	$n = 150$	$n = 500$
		(a) Normal Error		
β_1	1.00	1.0003 (.017) [.016]	1.0001 (.009) [.009]	1.0000 (.005) [.005]
	1.00	.9996 (.015) [.015]	1.0000 (.009) [.009]	.9999 (.005) [.005]
β_2	0.75	.7503 (.019) [.018]	.7502 (.010) [.010]	.7501 (.005) [.005]
	1.25	1.2500 (.015) [.014]	1.2500 (.009) [.009]	1.2501 (.005) [.005]
β_3	1.25	1.2509 (.017) [.016]	1.2499 (.010) [.010]	1.2499 (.005) [.005]
	0.75	.7506 (.016) [.015]	.7501 (.009) [.009]	.7501 (.005) [.005]
β_4	1.00	.9999 (.014) [.014]	.9998 (.010) [.010]	1.0002 (.005) [.005]
	1.00	.9999 (.015) [.015]	.9997 (.009) [.009]	1.0001 (.005) [.005]
β_5	0.75	.7497 (.018) [.017]	.7500 (.009) [.009]	.7499 (.005) [.005]
	1.25	1.2506 (.016) [.015]	1.2502 (.009) [.009]	1.2500 (.005) [.005]
β_6	1.25	1.2507 (.017) [.016]	1.2503 (.009) [.009]	1.2501 (.005) [.005]
	0.75	.7497 (.016) [.016]	.7503 (.009) [.009]	.7501 (.005) [.005]
λ	0.50	.4996 (.012) [.011]	.4996 (.008) [.008]	.4999 (.005) [.005]
	0.25	.2497 (.014) [.014]	.2496 (.011) [.011]	.2498 (.005) [.005]
	0.75	.7491 (.017) [.017]	.7498 (.006) [.006]	.7500 (.003) [.004]
	0.50	.5002 (.015) [.015]	.5007 (.009) [.009]	.4998 (.005) [.005]
	0.25	.2500 (.017) [.016]	.2497 (.010) [.010]	.2498 (.006) [.006]
	0.75	.7496 (.009) [.008]	.7498 (.006) [.006]	.7500 (.003) [.003]
σ^2	1.00	.9291 (.087) [.083]	.9764 (.052) [.050]	.9923 (.028) [.028]
σ^2	QMLE	.7743 (.073)	.8136 (.043)	.8269 (.023)
		(b) Normal Mixture Error		
β_1	1.00	1.0002 (.017) [.016]	1.0001 (.009) [.009]	1.0001 (.005) [.005]
	1.00	.9996 (.015) [.014]	.9999 (.009) [.009]	.9999 (.005) [.005]
β_2	0.75	.7505 (.019) [.018]	.7502 (.010) [.010]	.7501 (.005) [.005]
	1.25	1.2502 (.015) [.014]	1.2502 (.009) [.009]	1.2500 (.005) [.005]
β_3	1.25	1.2508 (.017) [.016]	1.2501 (.010) [.010]	1.2500 (.005) [.005]
	0.75	.7503 (.016) [.015]	.7498 (.009) [.009]	.7501 (.005) [.005]
β_4	1.00	.9998 (.015) [.014]	1.0000 (.011) [.010]	1.0002 (.005) [.005]
	1.00	.9998 (.016) [.015]	.9998 (.009) [.009]	1.0000 (.005) [.005]
β_5	0.75	.7495 (.018) [.017]	.7500 (.009) [.009]	.7499 (.005) [.005]
	1.25	1.2503 (.015) [.015]	1.2501 (.009) [.009]	1.2501 (.005) [.005]
β_6	1.25	1.2507 (.017) [.015]	1.2506 (.009) [.009]	1.2500 (.005) [.005]
	0.75	.7498 (.016) [.015]	.7501 (.009) [.009]	.7500 (.005) [.005]
λ	0.50	.4995 (.012) [.011]	.4998 (.008) [.008]	.4999 (.005) [.005]
	0.25	.2497 (.015) [.014]	.2497 (.011) [.011]	.2498 (.005) [.005]
	0.75	.7493 (.017) [.017]	.7497 (.006) [.006]	.7500 (.004) [.004]
	0.50	.5002 (.015) [.015]	.5008 (.009) [.009]	.4999 (.005) [.005]
	0.25	.2497 (.017) [.016]	.2497 (.010) [.010]	.2498 (.006) [.006]
	0.75	.7496 (.009) [.008]	.7499 (.006) [.006]	.7500 (.003) [.003]
σ^2	1.00	.9267 (.183) [.148]	.9749 (.113) [.096]	.9946 (.066) [.055]
σ^2	QMLE	.7723 (.153)	.8124 (.094)	.8288 (.055)

Table 3.4b (cont'd). Empirical Mean (sd) [se] of AQS-Estimator:
SL Model One-Way FE, W_n: Rook Contiguity, $T = 6$

		$n = 50$	$n = 150$	$n = 500$
(c) Chi-Square-normal Error				
β_{11}	1.00	.9998 (.017) [.016]	.9997 (.009) [.009]	1.0002 (.005) [.005]
	1.00	1.0000 (.014) [.014]	.9999 (.009) [.009]	1.0000 (.005) [.005]
β_{12}	0.75	.7499 (.018) [.018]	.7507 (.010) [.010]	.7502 (.005) [.005]
	1.25	1.2504 (.015) [.014]	1.2501 (.009) [.009]	1.2498 (.005) [.005]
β_{13}	1.25	1.2497 (.016) [.016]	1.2499 (.010) [.010]	1.2501 (.005) [.005]
	0.75	.7498 (.015) [.015]	.7498 (.009) [.009]	.7502 (.005) [.005]
β_{14}	1.00	1.0002 (.015) [.014]	.9997 (.010) [.010]	1.0000 (.005) [.005]
	1.00	.9999 (.016) [.015]	1.0003 (.009) [.009]	1.0000 (.005) [.005]
β_{15}	0.75	.7500 (.018) [.017]	.7497 (.009) [.009]	.7500 (.005) [.005]
	1.25	1.2499 (.016) [.015]	1.2501 (.009) [.009]	1.2501 (.005) [.005]
β_{16}	1.25	1.2500 (.016) [.016]	1.2497 (.009) [.009]	1.2500 (.005) [.005]
	0.75	.7503 (.016) [.015]	.7501 (.009) [.009]	.7501 (.005) [.005]
λ_1	0.50	.5003 (.012) [.011]	.4999 (.008) [.008]	.5000 (.005) [.005]
	0.25	.2498 (.014) [.014]	.2497 (.011) [.011]	.2499 (.005) [.005]
	0.75	.7494 (.017) [.016]	.7501 (.006) [.006]	.7500 (.004) [.004]
	0.50	.4993 (.015) [.015]	.4999 (.009) [.009]	.5000 (.005) [.005]
	0.25	.2499 (.017) [.016]	.2503 (.010) [.010]	.2500 (.006) [.006]
	0.75	.7498 (.009) [.008]	.7501 (.006) [.006]	.7501 (.003) [.003]
σ^2	1.00	.9266 (.139) [.113]	.9744 (.081) [.072]	.9942 (.046) [.041]
σ^2	QMLE	.7722 (.115)	.8120 (.068)	.8285 (.038)

3.5 An Empirical Application

In this section, we apply the proposed AQS-estimation and inference methods for the FE-SPD model with TVC to investigate US cigarettes demand from 1963 to 1992. During this period, The federal government attempted to reduce the consumption of cigarettes through (i) the imposition of warning labels in 1965, (ii) the Fairness Doctrine Act to cigarette advertising in June 1967, (iii) the Congressional ban of broadcast advertising of cigarettes in 1971. A lot of researches are carried out, such as Hamilton (1972), McGuiness and Cowling (1975), Baltagi and Levin (1986, 1992), and Baltagi et al. (2000) to help the policy makers to evaluate the effectiveness of the above policies.

In this study, we estimate the time-varying coefficients in cigarette demand models based on panel data from 46 American states over the period 1963 to 1992. We find (i) significant effect of price on cigarette consumption (ii) Insignificant income effect in short run due to the habit persistence (iii) Significant minimum neighbouring price in short run. (iv) support for the effectiveness of warning labels, Fairness Doctrine Act and advertising ban. The specification test proposed by Xu and Yang (2020a) is useful in supporting our findings, temporal heterogeneity pattern is observed in parameter estimation.

We fit the data to the general model (3.1) and several sub-models. We report results of the general model, SLE two-way FE model, and the sub-model, SLE one-way FE model only. Final conclusions are made based on the general model (3.1). For data sources, see Baltagi and Levin (1986). The response variable Y is Cigarette sales in packs per capita; The spatial lag term W_nY captures the demand of cigarette among states, reflecting how demand of the neighbouring states affect the own demand of a state. The time-varying regressors X contain a set of state level variables: X_1 = Price per pack of cigarettes; X_2 = Population above the age of 16; X_3 = Per capita disposable income; and X_4 = Minimum price in adjoining states per pack of cigarettes. Similar as in Baltagi and Levin (1992), the state-specific effects can represent any state-specific characteristic and the year-specific effects can be justified given numerous policy interventions. The spatial weight matrix is specified using a contiguity form where (i, j)th element is 1 if state i and j share a common border, otherwise 0, and then row normalized. Table 3.5a, 3.5b and 3.5c summarize the main empirical findings, fitted using the two models: SLE one-way and SLE two-way FE-SPD models, and are estimated using (a) data from the first three years; (b) data from the first six years; and (c) the full data. The case (a) is different from the other two cases in without the consideration of government interventions.

Table 3.5a presents the estimation results that use the data from the first three years. First, the ML-estimators are very similar as the AQS-estimators, which shows that the empirical results are in consistent with the theory and Monte Carlo results. Thus, it is recommended to use AQS-estimation method. Several patterns are observed: (i) the price effect and the parameter estimate for X_4, the minimum price in adjoining states, are significant in both one-way and two-way FE models; (ii) the parameter estimates for spatial lag effects, spatial error effect and X_2, which captures the population above the age of 16, are insignificant in both one-way and two-way FE models; and (iii) the income effect is significant in one-way FE model but insignificant in two-way FE model. One possible reason for the difference is the omission bias of ignoring time-specific effects, like the Surgeon General report, health warnings, health report and increasing taxation which can deter cigarette consumption. The effect of these anti-smoking policies on smoking and income are negative. There would be income loss for not only cigarette industry, but also related industries in the cigarette supply chain, from manufacturing, transportation, to selling. The above policies across different states are mostly state-invariant and would be controlled by the year dummies. Therefore, the estimators of income effects in one-way FE model does not account for year-specific effects and captures the omission bias of anti-smoking policies.

Table 3.5b presents the estimation results that use the data from the first six years. Again, as the estimates of the time-varying coefficients are very similar, only the AQS estimates are reported. Some similar patterns are observed, such as: (i) the price effect are significant in both one-way and two-way FE models, and (ii) the income effect is significant in one-way FE model but insignificant in two-way FE model. Some different patterns are also observed: in one-way FE model, the neighboring price is significant in 5 out of 6 years while in two-way FE model, it is significant in only one out of the 6 years. It indicates that with the change of observation period T, different estimation results can be obtained. Further extension of estimation period is necessary.

Table 3.5c presents the estimation results using the full data. Besides the significant price effect in both models, we observe that (i) in one-way FE model, the neighboring price is significant in 23 out of 30 years while in two-way FE model, it is significant in only 9 out of 30 years, and (ii) the income effect is significant in one-way FE model but is significant in only 16 out of 30 years in two-way FE model.

The following conclusions are made based on estimation results from the general model, SLE two-way FE model. Due to the significant price effect, an effective option to deter the cigarette consumption is to increase the cigarette taxation. Baltagi and Levin (1992) also argue that cigarette taxation can be used as a policy to combat smoking. The insignificant income effect in short run may be due to the habit persistence effect. It is hard to quit smoking immediately as cigarette is an addictive product, therefore, even there are interventions to deter cigarette consumption, people who smoke would keep smoking in the short run without consideration of current income. However, the income effect can be significant in the long run if people have enough time and preparation to learn how to quit smoking. The insignificant income effect is also found in Baltagi and Levin (1986). The effect of minimum neighbouring price is significantly positive in short run. Simply put, higher cigarette price in state A will encourage consumers in that state to search for cheaper cigarettes in neighbouring states. However, the minimum neighbouring price can be insignificant in long run. Baltagi and Levin (1992) point out that individual states can raise revenues by increasing their state tax on cigarettes. Therefore, in the long run, if the adjoining states of state A raise their cigarette tax, state A may follow suit in raising its state tax to increase the revenue with some reduction in consumption.

The specification test proposed by Xu and Yang(2020a) is used to test the effectiveness of the three major policies in 1965, 1967 and 1971. This AQS-based test is denoted by $T^{(r)}$ in the below table. The null hypothesis thus is:

$$H_0 : \beta_1 = \cdots = \beta_T, \quad \text{and} \quad \lambda_1 = \cdots = \lambda_T \tag{3.29}$$

We test H_0 based on different periods, the results are presented in the following table.

AQS-Tests For Detection of Change Points

	$T^{(r)}_{\texttt{SLE1}}$	$T^{(r)}_{\texttt{SLE2}}$		$T^{(r)}_{\texttt{SLE1}}$	$T^{(r)}_{\texttt{SLE2}}$		$T^{(r)}_{\texttt{SLE1}}$	$T^{(r)}_{\texttt{SLE2}}$
t_1–t_3	9.48	8.45	t_4–t_5	7.87	8.11	t_1–t_5	43.4	39.0
	0.487	0.585		0.164	0.150		0.002	0.007
t_1–t_6	47.9	43.4	t_6–t_8	15.9	12.5	t_1–t_{10}	106	119
	0.004	0.013		0.103	0.254		0.000	0.000

Note: p-values are in every second row.

We break down the panel and repeatedly applying the set of robust AQS-based specification test for detecting change points. In the first ten periods, we see that only the sub-panels 1963-65, 1966-67, and 1968-70 are fairly stable, suggesting that panel structures have changed after 1965, 1967, and 1970, in line with the policy interventions in 1965, 1967 and 1971. The test indicates that parameter estimates are subject to temporal heterogeneity, and the change points are in 1965, 1967, and 1971.

In conclusion, our empirical results finds that taxation can be used as a policy instrument to deter state cigarette consumption. The effectiveness of cigarette taxation depends upon several factors: (i) price elasticity of demand for cigarette; (ii) whether the adjoining states follow suit in raising their state taxes for higher revenues. A temporary reduction of one's money to push him to quit smoking might not be effective. The insignificant income effect in the short run shows that income is not the first consideration of deciding whether smoke or not due to the habit persistence effect. Finally, our results find support for the effectiveness of subsidized anti-smoking messages and advertising ban.

Table 3.5a. AQS Estimator (se) based on Cigarette Demand Data, $T = 3$

t	X_1	X_2	X_3	X_4	Spatial lag
	SLE model with one-way FE.				
1	$-.5628$ (.0083)	.0855 (.0538)	.2148 (.0247)	.1687 (.0374)	.1137 (.1440)
2	$-.6169$ (.0086)	.0824 (.0536)	.1772 (.0245)	.2916 (.0319)	.1267 (.1398)
3	$-.5998$ (.0090)	.0877 (.0533)	.2042 (.0271)	.2604 (.0296)	.0848 (.1449)
	$\hat{\rho}_{\mathtt{QML}} = .0629$		$\hat{\rho}_{\mathtt{AQS}}$ (se) = .0559 (.1668)		
	$\hat{\sigma}^2_{\mathtt{QML}} = .0003$		$\hat{\sigma}^2_{\mathtt{AQS}}$ (se) = .0004 (.0000)		
	SLE model with two-way FE.				
1	$-.5544$ (.0909)	.0738 (.2158)	.1790 (.1508)	.3163 (.1774)	.2903 (.2923)
2	$-.6359$ (.0919)	.0716 (.2152)	.1323 (.1512)	.3625 (.1624)	.2928 (.2913)
3	$-.6372$ (.0903)	.0760 (.2147)	.1494 (.1572)	.3152 (.1587)	.2490 (.3006)
	$\hat{\rho}_{\mathtt{QML}} = -.1522$		$\hat{\rho}_{\mathtt{AQS}}$ (se) = $-.1522$ (.3587)		
	$\hat{\sigma}^2_{\mathtt{QML}} = .0003$		$\hat{\sigma}^2_{\mathtt{AQS}}$ (se) = .0004 (.0001)		

Note: as the estimates of the time-varying coefficients are very similar,
only the AQS estimates are reported.

Table 3.5b. AQS Estimator (se) based on Cigarette Demand Data, $T = 6$

t	X_1	X_2	X_3	X_4	Spatial lag
	SLE model with one-way FE.				
1	$-.3142$ (.0075)	.3054 (.0203)	.2696 (.0168)	$-.2868$ (.0164)	.0765 (.0387)
2	$-.4113$ (.0070)	.3004 (.0203)	.2269 (.0173)	$-.1030$ (.0138)	.0886 (.0369)
3	$-.4263$ (.0066)	.3067 (.0202)	.2404 (.0192)	$-.0706$ (.0107)	.0451 (.0380)
4	$-.7165$ (.0055)	.2925 (.0200)	.2731 (.0202)	.0882 (.0088)	.1091 (.0337)
5	$-.6602$ (.0067)	.3008 (.0199)	.2290 (.0207)	.1292 (.0090)	.1008 (.0348)
6	$-.4363$ (.0082)	.2937 (.0198)	.1917 (.0207)	.0037 (.0095)	.1012 (.0380)
	$\hat{\rho}_{\mathtt{QML}} = -.0452$		$\hat{\rho}_{\mathtt{AQS}}$ (se) = $-.0457$ (.0466)		
	$\hat{\sigma}^2_{\mathtt{QML}} = .0009$		$\hat{\sigma}^2_{\mathtt{AQS}}$ (se) = .0011 (.0000)		
	SLE model with two-way FE.				
1	$-.2889$ (.1132)	.4415 (.1531)	.2310 (.1372)	$-.3516$ (.1651)	.3348 (.1395)
2	$-.4306$ (.1124)	.4329 (.1529)	.2013 (.1404)	$-.1482$ (.1461)	.3480 (.1345)
3	$-.4805$ (.1121)	.4419 (.1524)	.2152 (.1455)	$-.0732$ (.1266)	.3071 (.1368)
4	$-.8329$ (.0911)	.4229 (.1514)	.2694 (.1514)	.1377 (.1029)	.3851 (.1265)
5	$-.8005$ (.0998)	.4363 (.1513)	.2179 (.1551)	.2174 (.1105)	.3869 (.1292)
6	$-.5462$ (.1078)	.4301 (.1508)	.1597 (.1583)	.0828 (.1082)	.3799 (.1360)
	$\hat{\rho}_{\mathtt{QML}} = -.4407$		$\hat{\rho}_{\mathtt{AQS}}$ (se) = $-.4407$ (.1685)		
	$\hat{\sigma}^2_{\mathtt{QML}} = .0011$		$\hat{\sigma}^2_{\mathtt{AQS}}$ (se) = .0014 (.0002)		

Note: as the estimates of the time-varying coefficients are very similar,
only the AQS estimates are reported.

Table 3.5c. AQS Estimator (se) based on Cigarette Demand Data, $T = 30$

t	X_1	X_2	X_3	X_4	Spatial lag
			SLE model with one-way FE.		
1	−.8240 (.0243)	−.0086 (.0011)	.8930 (.0090)	−.8531 (.0423)	−.4446 (.0069)
2	−.8985 (.0250)	−.0124 (.0012)	.8132 (.0088)	−.6361 (.0345)	−.4169 (.0068)
3	−.8605 (.0231)	−.0060 (.0011)	.8406 (.0096)	−.6458 (.0270)	−.4970 (.0073)
4	−1.2331 (.0197)	−.0173 (.0011)	.8646 (.0093)	−.3286 (.0230)	−.4717 (.0061)
5	−.9994 (.0232)	−.0228 (.0011)	.8673 (.0096)	−.5405 (.0258)	−.4863 (.0063)
6	−.8239 (.0298)	−.0464 (.0012)	.8619 (.0095)	−.6838 (.0341)	−.4597 (.0063)
7	−.7315 (.0208)	−.0399 (.0011)	.8008 (.0100)	−.6511 (.0303)	−.4648 (.0068)
8	−.8631 (.0187)	−.0223 (.0011)	.7326 (.0108)	−.3718 (.0205)	−.4779 (.0069)
9	−.8628 (.0174)	−.0080 (.0011)	.6922 (.0107)	−.2720 (.0219)	−.4954 (.0073)
10	−.9487 (.0140)	−.0149 (.0011)	.6445 (.0102)	−.1562 (.0163)	−.4159 (.0073)
11	−1.0271 (.0167)	.0101 (.0011)	.5638 (.0094)	−.0172 (.0174)	−.3681 (.0077)
12	−1.0364 (.0182)	.0095 (.0011)	.5645 (.0107)	−.0202 (.0202)	−.3536 (.0082)
13	−.9816 (.0191)	−.0003 (.0012)	.5703 (.0105)	−.0741 (.0194)	−.3417 (.0083)
14	−1.0087 (.0221)	−.0093 (.0012)	.5531 (.0107)	−.0455 (.0189)	−.2854 (.0082)
15	−1.0604 (.0240)	.0004 (.0012)	.5066 (.0111)	.0259 (.0187)	−.2360 (.0086)
16	−1.0764 (.0288)	.0129 (.0012)	.4512 (.0110)	.1281 (.0219)	−.2132 (.0090)
17	−.9065 (.0286)	.0114 (.0012)	.4490 (.0110)	.0374 (.0227)	−.2786 (.0090)
18	−.7274 (.0284)	.0151 (.0012)	.3321 (.0111)	.0459 (.0219)	−.2239 (.0100)
19	−.7621 (.0286)	.0101 (.0012)	.3712 (.0113)	−.0326 (.0234)	−.1932 (.0100)
20	−.8257 (.0289)	.0122 (.0012)	.3747 (.0115)	.0483 (.0331)	−.2078 (.0102)
21	−.5882 (.0285)	.0213 (.0012)	.2851 (.0124)	−.0855 (.0351)	−.1381 (.0110)
22	−.5188 (.0306)	.0147 (.0012)	.2775 (.0125)	−.1523 (.0373)	−.1139 (.0115)
23	−.3407 (.0293)	.0263 (.0012)	.1820 (.0119)	−.1524 (.0329)	−.1147 (.0117)
24	−.4309 (.0315)	.0394 (.0012)	.2228 (.0119)	−.1320 (.0276)	−.1471 (.0121)
25	−.2564 (.0258)	.0362 (.0012)	.1385 (.0106)	−.1928 (.0273)	−.0902 (.0127)
26	−.3477 (.0277)	.0376 (.0012)	.2078 (.0103)	−.2378 (.0269)	−.0976 (.0117)
27	−.3324 (.0284)	.0405 (.0012)	.2050 (.0108)	−.2539 (.0308)	−.0975 (.0120)
28	−.3752 (.0254)	.0406 (.0012)	.1243 (.0133)	−.2014 (.0324)	.0595 (.0130)
29	−.4202 (.0242)	.0226 (.0012)	.1356 (.0126)	−.2525 (.0286)	.1705 (.0119)
30	−.5355 (.0313)	.0222 (.0012)	.0432 (.0151)	−.0017 (.0328)	.2346 (.0115)
	$\hat{\rho}_{\texttt{QML}} = .6453$		$\hat{\rho}_{\texttt{AQS}}$ (se) = .6453 (.0014)		
	$\hat{\sigma}^2_{\texttt{QML}} = .0057$		$\hat{\sigma}^2_{\texttt{AQS}}$ (se) = .0059 (.0000)		

Note: as the estimates of the time-varying coefficients are very similar, only the AQS estimates are reported.

Table 3.5c (cont'd). AQS Estimator (se) based on Cigarette Demand Data, $T = 30$

t	X_1	X_2	X_3	X_4	Spatial lag
			SLE model with two-way FE.		
1	$-.4586$ (.1513)	$-.0589$ (.0284)	.8208 (.0862)	$-.4353$ (.2794)	$-.3097$ (.0911)
2	$-.5646$ (.1627)	$-.0627$ (.0286)	.7518 (.0873)	$-.3469$ (.2507)	$-.3347$ (.0874)
3	$-.5870$ (.1548)	$-.0568$ (.0286)	.7628 (.0933)	$-.5094$ (.2113)	$-.4187$ (.0881)
4	$-.9520$ (.1334)	$-.0669$ (.0285)	.8453 (.0935)	$-.1920$ (.1524)	$-.4218$ (.0751)
5	$-.7534$ (.1449)	$-.0704$ (.0285)	.8464 (.0941)	$-.3168$ (.1656)	$-.4491$ (.0763)
6	$-.4882$ (.1551)	$-.0906$ (.0289)	.8404 (.0958)	$-.4802$ (.2033)	$-.4072$ (.0779)
7	$-.6421$ (.1301)	$-.0771$ (.0286)	.7979 (.0993)	$-.2016$ (.1817)	$-.3694$ (.0811)
8	$-.6950$ (.1173)	$-.0714$ (.0285)	.7655 (.1028)	$-.2227$ (.1413)	$-.4012$ (.0810)
9	$-.7496$ (.1125)	$-.0610$ (.0286)	.6364 (.1053)	$-.2461$ (.1542)	$-.4642$ (.0842)
10	$-.9005$ (.0991)	$-.0676$ (.0286)	.5782 (.1094)	$-.1769$ (.1248)	$-.4185$ (.0814)
11	-1.0169 (.1078)	$-.0448$ (.0285)	.4655 (.1010)	$-.0787$ (.1237)	$-.4257$ (.0837)
12	-1.0307 (.1125)	$-.0433$ (.0286)	.4750 (.1113)	$-.0565$ (.1375)	$-.4173$ (.0856)
13	-1.0022 (.1160)	$-.0524$ (.0288)	.4398 (.1110)	$-.1305$ (.1320)	$-.4327$ (.0846)
14	-1.0717 (.1246)	$-.0592$ (.0289)	.3885 (.1156)	$-.1022$ (.1369)	$-.3734$ (.0865)
15	-1.1314 (.1298)	$-.0489$ (.0288)	.2977 (.1158)	$-.0696$ (.1332)	$-.3782$ (.0878)
16	-1.1712 (.1429)	$-.0375$ (.0289)	.1956 (.1173)	.0280 (.1510)	$-.3544$ (.0935)
17	-1.0854 (.1434)	$-.0361$ (.0288)	.1598 (.1102)	$-.0767$ (.1483)	$-.4816$ (.0905)
18	$-.9386$ (.1434)	$-.0312$ (.0289)	.0266 (.1058)	$-.0896$ (.1567)	$-.4570$ (.0978)
19	$-.8923$ (.1429)	$-.0393$ (.0291)	.1234 (.1071)	$-.1141$ (.1575)	$-.3586$ (.0989)
20	$-.9970$ (.1433)	$-.0373$ (.0291)	.0770 (.1056)	$-.0673$ (.1946)	$-.3594$ (.0982)
21	$-.8419$ (.1435)	$-.0213$ (.0291)	.0140 (.1061)	$-.2699$ (.1998)	$-.3135$ (.1044)
22	$-.6996$ (.1480)	$-.0322$ (.0292)	.0629 (.1050)	$-.2255$ (.2094)	$-.2629$ (.1080)
23	$-.5515$ (.1474)	$-.0185$ (.0291)	$-.0265$ (.1006)	$-.2809$ (.1909)	$-.3072$ (.1127)
24	$-.7725$ (.1549)	$-.0083$ (.0293)	$-.0116$ (.1036)	$-.4011$ (.1916)	$-.3548$ (.1084)
25	$-.5907$ (.1436)	$-.0091$ (.0292)	$-.0748$ (.0968)	$-.3754$ (.1842)	$-.2818$ (.1155)
26	$-.6769$ (.1436)	$-.0120$ (.0291)	.0356 (.0950)	$-.4234$ (.1859)	$-.2989$ (.1096)
27	$-.7196$ (.1441)	$-.0103$ (.0290)	.0026 (.0964)	$-.4630$ (.2002)	$-.2737$ (.1088)
28	$-.6425$ (.1357)	$-.0075$ (.0289)	$-.0836$ (.1089)	$-.6648$ (.1991)	$-.1809$ (.1159)
29	$-.7632$ (.1326)	$-.0121$ (.0292)	$-.1821$ (.1107)	$-.4626$ (.1625)	$-.1444$ (.1130)
30	$-.8968$ (.1486)	$-.0182$ (.0286)	$-.2586$ (.1184)	$-.5401$ (.1862)	$-.0855$ (.1164)
	$\hat{\rho}_{\mathtt{QML}} = .6495$		$\hat{\rho}_{\mathtt{AQS}}$ (se) = .6495 (.0360)		
	$\hat{\sigma}^2_{\mathtt{QML}} = .0040$		$\hat{\sigma}^2_{\mathtt{AQS}}$ (se) = .0041 (.0002)		

Note: as the estimates of the time-varying coefficients are very similar, only the AQS estimates are reported.

3.6 Conclusion and Discussion

We introduce a general strategy (AQS-estimation) for estimating fixed effects spatial panel data models with time-varying coefficients and two major spatial effects: the spatial lag and spatial error. Based on the adjusted quasi score function, the proposed AQS-estimation method is robust in the sense that it is allowing errors to be nonnormal. The common parameter estimates from the AQS-estimation approach are consistent, and the asymptotic distributions are properly centered. Typically, a consistent estimate requires either n or both n and T to be large. But the convergence rates of the estimators need to be adjusted when T is large. Monte Carlo results are provided to illustrate finite sample properties of the various estimators. Empirical case is also provided to support further applications.

An empirical illustration is presented to help empirical researchers to apply our methods in models with time-varying coefficients. Our methods are quite useful in prediction, meanwhile in finding and analyzing the most influential observations from a large dataset. When applying the AQS-estimation method to models with time-varying coefficients, our analysis should focus on the trend of the temporal change instead of the estimator in one certain period. Estimators that are under temporal homogeneity assumptions may not provide this kind of information to us. The proposed methods are simple and general, which makes them very attractive to practitioners. It would be also interesting to extend our methods to allow for interactive fixed effects, heteroskedasticity in time and cross-section, serial correlation, etc. However, these are clearly beyond the scope of this chapter and will be dealt with in future works.

4 Heteroskedasticity Robust Estimation of Spatial Panel Data Models with Temporal Heterogeneity

In the presence of temporal heterogeneity or TH (time-varying regression and spatial coefficients), the usual transformation-based methods for estimating the fixed effects (FE) spatial panel data (SPD) models are inapplicable. The presence of cross-sectional heteroskedasticity (CH) causes the usual methods of estimating spatial econometric models to be inconsistent. In this chapter, a new set of estimation and inference methods is developed based on the *adjusted quasi scores* (AQS) that simultaneously take care of the three major issues, FE, TH and CH, in the estimation of FE-SPD model. Consistency and asymptotic normality of the robust AQS-estimators are established. The AQS functions are decomposed into a vector martingal diferences so that the *outer-product-of-martingale-difference* (OPMD) gives a consistent estimate of variance of the AQS functions which in turn gives consistent estimates of variance-covariance matrix of the AQS-estimators. Monte Carlo results show that the proposed set of estimation and inference methods has good finite sample performance.

4.1 Introduction

Recently, spatial models are receiving substantial attentions in economic studies. The values observed at one location depend on the values observed at the neighboring locations, giving rise to the so-called spatial dependence. However, this dependence may not be temporally homogeneous similar to the situations where the covariate effects are temporally heterogeneous. Therefore, the spatial model considered in this paper is subject to temporal heterogeneity (TH). Models with time varying coefficients (TVC) have superiority over models with fixed parameters in terms of forecasting and identifying influential data observations. Such an allowance of TVC is important in economic study but also complicates the estimation procedure. Therefore, specialized techniques are desired.

The quasi maximum likelihood (QML) estimation method is the most conventional estimation method in spatial panel data (SPD) models. However, for the SPD model with fixed effects (FE), the direct QML estimates of some parameter is inconsistent due to the incidental parameters problem of Neyman and Scott (1948), see also Lancaster (2000). The usual transformation-based method is inapplicable due to the time-varying nature of the regression and spatial coefficients, and perhaps the time-varying nature of the spatial weight matrices. Xu an d Yang (2020a,b) developed *adjusted quasi score* (AQS) method for testing and estimation of the FE-SPD model with TVC.

Cross-sectional heteroskedasticity (CH) may be another important feature of spatial data due to the fact that the spatial units vary greatly in size. Anselin (1988) raised the issue of heteroskedasticity in spatial models, which may occur more naturally in the presence of peer interactions. The mix of aggregate and non aggregate data in the model may cause errors to be heteroskedastic. See, e.g., Glaeser et al. (1996), LeSage and Pace (2009). Spatial units are often heterogeneous in important characteristics, e.g., size, and hence the homoskedasticity assumptions may not hold in many situations, and therefore, a lack of an estimation theory that allows for heteroskedasticity is a serious shortcoming (Kelejian and Prucha, 2010). The presence of social interactions will inflate the variance of aggregated level data with the extent depending on the strength and structure of the interactions, leading to a more complicated variance structure, therefore we would expect the variances of the error terms to be different in certain applications (Lin and Lee, 2010). With spatial interactions, the homoskedasticity assumptions are quite restrictive in the SPD models. The QML estimators and the corresponding asymptotic distributions derived under the homoskedasticity assumptions are generally inappropriate. Therefore, it is highly desirable to develop a set of estimation and inference methods for the FE-SPD model with TVC that are robust against unknown CH.

Recent spatial econometrics literature has seen many attempts in providing estimation and inference methods robust against unknown heteroskedasticity. See LeSage (1997) for a Bayesian approach; Lin and Lee (2010), Kelejian and Prucha (2010), and Badinger and Egger (2011, 2015) for GMM or 2SLS methods; Jin and Lee (2012), Baltagi and Yang (2013b), Liu and Yang (2015, 2020), and Li and Yang (2020) for likelihood-based approaches. Lin and Lee (2010) provide heteroskedasticity robust GMM estimators by modifying certain moment conditions. Liu and Yang (2015) introduce a modified QML estimation method for a spatial autoregressive model robust against unknown heteroskedasticity and propose an outer-product-of-gradients method to estimate the variance of the score function which in turn leads to a consistent estimate of variance of the modified QML estimators. Liu and Yang (2020) extend these methods to a homogeneous FE-SPD model. Yang (2018), referring these methods as *adjusted quasi score* (AQS) or M-estimation method for model estimation and *outer-product-of-martingale-difference* (OPMD) method for variance-covariance (VC) matrix estimation, present AQS estimators and OPMD standard errors for a fixed effects spatial dynamic panel data model with homoskedastic errors, and Li and Yang (2020) extend these methods to allow errors to be cross-sectionally heteroskedastic of unknown form.

Inspired by Liu and Yang (2015, 2020), Yang (2018) and Li and Yang (2020), in this chapter we develop an AQS estimation method for the FE-SPD model with time-varying regression and spatial coefficients by adjusting the concentrated score functions with FE being concentrated out, so that the AQS functions obtained are robust against unknown heteroskedasticity. For heteroskedasticity robust inferences, we develop an OPMD method for estimating the variance of the AQS functions, which together with the expected Hessian matrix of the AQS functions give a robust estimator of the VC matrix of the AQS estimators. Monte Carlo results show that the AQS estimators and OPMD-based standard error estimates perform very well.

The rest of the chapter is organized as follows. Section 4.2 introduces the heteroskedasticity robust AQS method for estimating the FE-SPD model with TVC, the OPMD-based estimator of VC matrix, and their asymptotic properties. The time-varying nature of the coefficients renders separate considerations of asymptotics under large n and small T, and large n and large T. Section 4.3 presents the Monte Carlo results. Section 4.4 concludes the chapter. Technical details are given in Appendix.

4.2 Robust AQS-Estimation of FE-SPD Model with TVC

The basic idea of the AQS-estimation method is to first formulate the quasi Gaussian likelihood function, and then adjust the quasi score function to give a set of robust and unbiased estimating functions. The idea behind the OPMD-based standard error estimates is to decompose the AQS function into the sum of a vector martingale difference (M.D.) sequence so that the 'average' of the outer products of the elements of the M.D. sequence gives a consistent estimate of variance of the AQS functions, which in turn gives consistent estimates of variance-covariance (VC) matrix of the AQS-estimators.

We first outline a general framework for estimating the FE-SPD model with TVC allowing cross-sectional heteroskedasticity of unknown form. Then, we present the asymptotic properties of the robust AQS estimators. Finally, we introduce the OPMD-based standard error estimators and their consistency is studied. Lemmas and the proofs of the theorems are sketched in Appendices.

4.2.1 Robust Estimation

The model. Consider the following spatial panel data (SPD) model with time-varying coefficients (TVC) and individual-specific FE:

$$Y_{nt} = \lambda_t W_n Y_{nt} + X_{nt}\beta_t + \mathbf{c}_n + V_{nt}, \tag{4.1}$$

for $t = 1, 2, \ldots, T$, where, for a given t, $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of observations on the response variable, X_{nt} is an $n \times k$ matrix containing the values of k nonstochastic, individually and time varying regressors, $V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ is an $n \times 1$ vector of errors where $\{v_{it}\}$ are independent and identically distributed (iid) across t for each i, and independent but not (necessarily) identically distributed (inid) across i for each t, with mean 0 and variance $\sigma_0^2 r_{n,i}$, $i = 1, \ldots, n$ where $r_{n,i} > 0$ and $\frac{1}{n}\sum_{i=1}^{n} r_{n,i} = 1$.[13] $\mathbf{c}_n$ is an $n \times 1$ vector of unobserved spatial heterogeneity in the intercept or simply unobserved individual-specific effects that may be correlated with time-varying regressors, W_n is an $n \times n$ spatial weight matrix, λ_t is the *spatial lag parameter* for period t and β_t is the $k \times 1$ vector of regression coefficients for period t.

As λ_t and β_t are allowed to change with t, the usual fixed-effects estimation methods, such as first differencing or orthogonal transformation, cannot be applied. For an FE-SPD

[13] Note that σ_0 is the average of $\mathrm{Var}(v_{it})$. Under homoskedasticity, $r_{n,i} = 1$, $\forall i$. For generality, we allow $r_{n,i}$ to depend on n, for each i. This parameterization gives a nonparametric version of Breusch and Pagan (1979) and is useful as it allows the estimation of the average scale parameter.

model with TVC and homoskedastic errors, Xu and Yang (2020a) propose specification tests for testing the temporal homogeneity in regression and spatial coefficients. When the 'temporal homogeneity' is rejected, one may need to proceed to estimate the full model with TVC, which is considered in Xu and Yang (2020b). However, when the model errors are heteroskedastic, these testing and estimating strategies are invalid. In this paper, we extend the *adjusted quasi score* (AQS) estimation method of Xu and Yang (2020b) to give AQS estimators that are robust against unknown and cross-sectional heteroskedasticity (CH). For inference, the method given in Xu and Yang (2020b) based on Hessian and variance of the AQS functions is again invalid. We develop an OPMD estimator of the VC matrix of the robust AQS estimators.

Denote $\boldsymbol{\beta} = (\beta_1', \ldots, \beta_T')'$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$, and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \sigma^2)'$. The joint quasi Gaussian loglikelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \sigma^2)'$ and c_n is

$$\ell(\boldsymbol{\theta}, c_n) = -\tfrac{nT}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln|A_n(\lambda_t)| - \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T V_{nt}'(\beta_t, \lambda_t, c_n)V_{nt}(\beta_t, \lambda_t, c_n),$$

where $V_{nt}(\beta_t, \lambda_t, c_n) = A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t - c_n$ and $A_n(\lambda_t) = I_n - \lambda_t W_n$, for $t = 1, \ldots, T$.

We eliminate c_n through a direct maximization of the above loglikelihood function. Given $\boldsymbol{\theta}$, $\ell(\boldsymbol{\theta}, c_n)$ is partially maximized at: $\tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{T}\sum_{t=1}^T \big(A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t\big)$, leading to the concentrated loglikelihood function of $\boldsymbol{\theta}$ upon substitution:

$$\ell^c(\boldsymbol{\theta}) = -\tfrac{nT}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{t=1}^T \ln|A_n(\lambda_t)| - \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^T \widetilde{V}_{nt}'(\boldsymbol{\beta}, \boldsymbol{\lambda})\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}), \qquad (4.2)$$

where $\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = V_{nt}(\beta_t, \lambda_t, \tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda})) = A_n(\lambda_t)Y_{nt} - X_{nt}\beta_t - \tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda})$. Maximizing $\ell^c(\boldsymbol{\theta})$ gives the quasi maximum likelihood estimator of the common parameter vector $\boldsymbol{\theta}$.

However, due to the estimation/elimination of the fixed effects c_n, and due to the existence of unknown CH, the QML estimator cannot be consistent. This can be seen as follows. Differentiating $\ell^c(\boldsymbol{\theta})$ gives *concentrated quasi score* (CQS) function of $\boldsymbol{\theta}$:

$$S^c(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2}X_{nt}'\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}), \quad t = 1, \ldots, T, \\ \frac{1}{\sigma^2}\tilde{\eta}_{nt}'(\boldsymbol{\beta}, \boldsymbol{\lambda})\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}) + \frac{1}{\sigma^2}\widetilde{V}_{nt}'(\boldsymbol{\beta}, \boldsymbol{\lambda})G_n(\lambda_t)\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}) - \mathrm{tr}[G_n(\lambda_t)], \quad t = 1, \ldots, T, \\ -\frac{nT}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^T \widetilde{V}_{nt}'(\boldsymbol{\beta}, \boldsymbol{\lambda})\widetilde{V}_{nt}(\boldsymbol{\beta}, \boldsymbol{\lambda}), \end{cases} \qquad (4.3)$$

where $G_n(\lambda_t) = W_n A_n^{-1}(\lambda_t)$ and $\tilde{\eta}_{nt} = G_n(\lambda_t)(X_{nt}\beta_t + \tilde{c}_n(\boldsymbol{\beta}, \boldsymbol{\lambda}))$, for $t = 1, \ldots, T$.

Denote θ_0 as the true values of the parameters, where $\theta_0 = (\beta_0', \lambda_0', \sigma_0^2)'$. A necessary condition to ensure the consistency of the quasi-score estimators $\hat{\theta}$ is that the probability limit of the estimating function at the true parameter value is zero, i.e.,

$$\lim_{n\to\infty} \tfrac{1}{nT} S^c(\theta_0) \xrightarrow{p} 0,$$

see van der Vaart (1998). The condition is generally not true even when the errors are homoskedastic as shown in Xu and Yang (2020a, b), and with unknown heteroskedasticity this necessary condition is violated even more seriously.

Assume Model (4.1) holds only under θ_0, and the usual expectation and variance operators correspond to θ_0. At the true values of the parameters, we have, $\tilde{c}_n(\beta_0, \lambda_0) = c_n + B_n^{-1}\overline{V}_n$ and hence $\widetilde{V}_{nt} \equiv \widetilde{V}_{nt}(\beta_0, \lambda_0) = V_{nt} - \overline{V}_n$, where $\overline{V}_n = \frac{1}{T}\sum_{t=1}^T V_{nt}$. Denote $G_{nt} = G_n(\lambda_{t0})$. Let $R_n = \mathtt{diag}(r_n)$ and $r_n = (r_{n,1}, \ldots, r_{n,n})'$. It is easy to show that,

$$E(\tilde{\eta}_{nt0}'\widetilde{V}_{nt} + \widetilde{V}_{nt}'G_{nt0}\widetilde{V}_{nt}) = \tfrac{T-1}{T}\sigma_0^2 \mathrm{tr}(R_n G_{nt}), \tag{4.4}$$

$$E\big(\textstyle\sum_{t=1}^T \widetilde{V}_{nt}'\widetilde{V}_{nt}\big) = \sigma_0^2 n(T-1). \tag{4.5}$$

These results show that the (λ, σ^2) elements of the $\frac{1}{nT}\mathrm{E}[S^c(\theta_0)]$, and hence the same elements of $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S^c(\theta_0)$, are not zero so that the QML estimator or quasi-score estimators, cannot be consistent in general.

Our idea is to modify the quasi-score functions of (4.3) so that its expectation at the true parameters θ_0 is zero even under unknown heteroskedasticity.[14] No adjustment for β elements in $S^c(\theta)$ is needed, since it has zero expectation and zero probability limit under unknown heteroskedasticity. The modification is trivial for the σ^2 component, we directly subtract the result in (4.5) from the σ^2 component of (4.3) to obtain the adjusted quasi score function for σ^2 component. Instead of using the result in (4.4) directly to adjust the λ component in $S^c(\theta)$, we find modification term in $\widetilde{V}_{nt}$ with expectation being the same

[14] Making the expectation of an estimating function to be zero leads potentially to a finite sample bias corrected estimation. This is in line with Baltagi and Yang (2013a,b) in constructing standardized or heteroskedasticity robust LM tests with finite sample improvements. See also Kelejian and Prucha (2001, 2010) and Lin and Lee (2010) for some useful methods in handling the linear-quadratic forms of heteroskedastic random vectors.

as the term inside the expectation of (4.4). We have the following:

$$E(\widetilde{V}_{nt}'\mathtt{diag}(G_{nt})\widetilde{V}_{nt}) = \tfrac{T-1}{T}\sigma_0^2\mathrm{tr}(R_nG_{nt}). \tag{4.6}$$

Subtracting the term inside the expectation of (4.6) from the term inside the expectation of (4.4) gives an adjusted quasi score function for $\boldsymbol{\lambda}$ component.

Through the above adjustments, we obtain the desired AQS function $S^{\star}(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$, which upon dividing by nT has zero expectation and zero probability limit under unknown CH. It provides a set of unbiased and heteroskedasticity robust estimating functions:

$$S^{\star}(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2}X_{nt}'\widetilde{V}_{nt}(\boldsymbol{\beta},\boldsymbol{\lambda},\rho), \quad t=1,\dots,T, \\ \frac{1}{\sigma^2}\tilde{\eta}_{nt}'\widetilde{V}_{nt}(\boldsymbol{\beta},\boldsymbol{\lambda}) + \frac{1}{\sigma^2}\widetilde{V}_{nt}'(\boldsymbol{\beta},\boldsymbol{\lambda})G_n^{\circ}(\lambda_t)\widetilde{V}_{nt}(\boldsymbol{\beta},\boldsymbol{\lambda}), \quad t=1,\dots,T, \\ -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\widetilde{V}_{nt}'(\boldsymbol{\beta},\boldsymbol{\lambda})\widetilde{V}_{nt}(\boldsymbol{\beta},\boldsymbol{\lambda}), \end{cases} \tag{4.7}$$

where $G_n^{\circ}(\lambda_t) = G_n(\lambda_t) - \mathtt{diag}(G_n(\lambda_t))$. It is easy to show that $\mathrm{E}[S^{\star}(\boldsymbol{\theta}_0)] = 0$, and that $\frac{1}{nT}S^{\star}(\boldsymbol{\theta}_0) \xrightarrow{p} 0$ as $n \to \infty$ alone, or the finite dimensional components of $\frac{1}{nT}S^{\star}(\boldsymbol{\theta}_0)$ approach to 0 in probability when both n and T go infinity. Solving $S^{\star}(\boldsymbol{\theta}) = 0$ leads to the robust AQS-estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ that not only is consistent but also has a centered asymptotic distribution, whether T is fixed or grows with n.

The root-finding process can be further simplified by first solving the equations for $\boldsymbol{\beta}$ and σ^2, given $\boldsymbol{\lambda}$, resulting in the constrained AQS-estimators of $\boldsymbol{\beta}$ and σ^2 as

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) = (X_N'\Omega X_N)^{-1}X_N'\Omega A_N(\boldsymbol{\lambda})Y_N, \tag{4.8}$$

$$\hat{\sigma}^2(\boldsymbol{\lambda}) = \tfrac{1}{n(T-1)}\hat{V}_N'(\boldsymbol{\lambda})\hat{V}_N(\boldsymbol{\lambda}), \tag{4.9}$$

where $X_N = \mathtt{blkdiag}(X_{n1},\dots,X_{nT})$, $Y_N = (Y_{n1}',\dots,Y_{nT}')'$, $\Omega = I_N - \frac{1}{T}(1_T1_T' \otimes I_n)$, $A_N(\boldsymbol{\lambda}) = \mathtt{blkdiag}(A_n(\lambda_1),\dots,A_n(\lambda_T))$, $\hat{V}_N(\boldsymbol{\lambda}) = \widetilde{V}_N(\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}),\boldsymbol{\lambda}) = \Omega[A_N(\boldsymbol{\lambda})Y_N - X_N\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})] = (\hat{V}_{n1}'(\boldsymbol{\lambda}),\dots,\hat{V}_{nT}'(\boldsymbol{\lambda}))'$. Substituting $\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})$ and $\hat{\sigma}^2(\boldsymbol{\lambda})$ back into the middle component of the AQS function in (4.7) gives the concentrated AQS function:

$$S^{\star c}(\boldsymbol{\lambda}) = \tfrac{1}{\hat{\sigma}^2(\boldsymbol{\lambda})}\hat{\eta}_N'(\boldsymbol{\lambda})\hat{V}_N(\boldsymbol{\lambda}) + \tfrac{1}{\hat{\sigma}^2(\boldsymbol{\lambda})}\hat{V}_N^{\circ\prime}(\boldsymbol{\lambda})G_N^{\circ}(\boldsymbol{\lambda})\hat{V}_N(\boldsymbol{\lambda}), \tag{4.10}$$

where $\hat{\eta}_N(\boldsymbol{\lambda}) = \mathtt{blkdiag}(\hat{\eta}_{n1}(\boldsymbol{\lambda}),\dots,\hat{\eta}_{nT}(\boldsymbol{\lambda}))$, $\hat{V}_N^{\circ}(\boldsymbol{\lambda}) = \mathtt{blkdiag}(\hat{V}_{n1}(\boldsymbol{\lambda}),\dots,\hat{V}_{nT}(\boldsymbol{\lambda}))$, $G_N^{\circ}(\boldsymbol{\lambda}) = \mathtt{blkdiag}$

$(G^{\circ}_{n1}(\boldsymbol{\lambda}), \ldots, G^{\circ}_{nT}(\boldsymbol{\lambda}))$, $\hat{\eta}_{nt}(\boldsymbol{\lambda}) = \tilde{\eta}_{nt}(\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}), \boldsymbol{\lambda})$, and $\hat{V}_{nt}(\boldsymbol{\lambda}) = \widetilde{V}_{nt}(\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}), \boldsymbol{\lambda}))$. Solving the resulting concentrated estimating equation, $S^{\star c}(\boldsymbol{\lambda}) = 0$, we obtain the unconstrained AQS-estimator $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$. The unconstrained AQS-estimators of $\boldsymbol{\beta}$ and σ^2 are thus $\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\lambda}})$ and $\hat{\sigma}^2 \equiv \hat{\sigma}^2(\hat{\boldsymbol{\lambda}})$. Denote $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\sigma}^2, \hat{\boldsymbol{\lambda}}')'$.

4.2.2 Asymptotic Properties of Robust AQS-estimators

We now study the consistency and asymptotic normality of the heteroskedasticity robust AQS-estimators for the FE-SPD model with time varying parameters. We focus on the short panels first, i.e., panels with large n and small T. Then we present the consistency and asymptotic normality of the robust AQS-estimators when T is large. Lemmas and proofs of the theorems are sketched in Appendices.

To facilitate the discussions, some notation and convention are reviewed and new ones are introduced: a parametric function at the true parameter value is differentiated from that at a general parameter value by dropping its argument, e.g., $A_{nt} \equiv A_n(\lambda_{t0})$ and $G_{nt} \equiv G_n(\lambda_{t0})$; the common expectation, variance and covariance operators 'E' 'Var' and 'Cov' correspond to the true parameter vector $\boldsymbol{\theta}_0$; Λ_t denotes the parameter space from which λ_t takes values, and $\boldsymbol{\Lambda}$ is the product space formed by $\{\Lambda_t\}$ from which $\boldsymbol{\lambda}$ takes values; tr$(\cdot)$, $|\cdot|$ and $\|\cdot\|$ denote, respectively, the trace, determinant, and Frobenius norm of a matrix; $\gamma_{\max}(A)$ and $\gamma_{\min}(A)$ denote, respectively, the largest and smallest eigenvalues of a real symmetric matrix A; and $\mathtt{diag}(a_k)$ forms a diagonal matrix using the elements $\{a_k\}$ and $\mathtt{blkdiag}(A_k)$ forms a block-diagonal matrix using the matrices $\{A_k\}$. Furthermore, $\{A_{ts}\}$ forms a new matrix using the sub-matrices A_{ts} for $t, s = 1, \ldots, T$.

There are two factors that cause the inconsistency of QML estimators based on the concentrated loglikelihood function given in (4.2) or equivalently the quasi score function given in (4.3). One is the incidental parameters problem (Neyman and Scott, 1948) induced by direct estimation of the fixed effects c_n. The other is that the existence of cross-sectional heteroskedasticity of completely unknown form induces another set of incidental parameters that further bias the quasi score function.

To see this, let $\mathbf{g}_{nt} = (\mathbf{g}_{nt,1}, \ldots, \mathbf{g}_{nt,n})' = \mathtt{diagv}(G_{nt})$, $\bar{\mathbf{g}}_{nt} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_{nt,i}$, for $t = 1, \ldots, T$. Let $\mathrm{Cov}(\mathbf{g}_{nt}, r_n)$ denote the sample covariance between the two vectors

g_{nt} and r_n. We have, similarly to Lin and Lee (2010), for $t = 1, \ldots, T$,

$$\frac{1}{n}\text{tr}(R_n G_{nt} - G_{nt}) + o_p(1) = \frac{1}{n}\sum_{i=1}^{n}(r_{n,i} - 1)(\mathbf{g}_{nt,i} - \bar{\mathbf{g}}_{nt}) + o_p(1)$$
$$= \text{Cov}(\mathbf{g}_{nt}, r_n) + \text{o}_\text{p}(1) \qquad (4.11)$$

Therefore, for $\hat{\boldsymbol{\theta}}$ to be consistent, it is necessary that as $n \to \infty$, $\text{Cov}(\mathbf{g}_{nt}, r_n) \to 0$. In other words, when $\lim_{n\to\infty}\text{Cov}(\mathbf{g}_{nt}, r_n) \neq 0$, $\hat{\boldsymbol{\theta}}$ cannot be consistent. By Cauchy-Schwartz inequality, this condition is satisfied if $\text{Var}(\mathbf{g}_{nt}) \to 0$, which is equivalent to $\text{Var}(k_n) \to 0$, where k_n is the vector of number of neighbours for each unit.[15]

Assumption A: *The disturbances $\{v_{it}\}$ are such that (i) iid across t but inid across i with $E(v_{it}) = 0$, (ii) $Var(v_{it}) = \sigma_0^2 r_{n,i}$, where $0 < r_{n,i} \leqslant c < \infty$ and $\frac{1}{n}\sum_{i=1}^{n} r_{n,i} = 1$. (iii) $E|v_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.*

Assumption B: *The space $\boldsymbol{\Lambda}$ is compact, and the true parameter $\boldsymbol{\lambda}_0$ lies in its interior.*

Assumption C: *The time-varying regressors $\{X_{nt},\ t = 1, \ldots, T\}$ are exogenous with respect to $\{v_{it}\}$ and are correlated with c_n in an arbitrary manner; their values are uniformly bounded in n and t, and $\lim_{N\to\infty}\frac{1}{N}X_N' X_N$ exists and is nonsingular.*

Assumption D: *(i) the elements w_{ij} of W_n are at most of order h_n^{-1}, uniformly in all i and j, and $w_{ii} = 0$ for all i; (ii) $h_n/n \to 0$ as $n \to \infty$; (iii) $\{W_n\}$ is row-normalized and is uniformly bounded in both row and column sums in absolute value; (iv) The matrix $A_n(\lambda_t)$ is invertible for all $\lambda_t \in \Lambda_t$, $A_n^{-1}(\lambda_{t0})$ is uniformly bounded in both row and column sums, and $A_n^{-1}(\lambda_t)$ is uniformly bounded in either row or column sums, uniformly in $\lambda_t \in \Lambda$, $t = 1, \ldots, T$.*

Assumption A extends Xu and Yang (2020b) to allow for unknown CH. Consistent estimation of $\boldsymbol{\lambda}$ requires the compactness of $\boldsymbol{\Lambda}$ in Assumption B. Under Assumptions C and D, the consistency of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ follows almost immediately that of $\hat{\boldsymbol{\lambda}}$. Conditions

[15] According to Lin and Lee (2010), this condition is satisfied if almost all the diagonal elements of the matrix G_{nt} are equal. For each period, $G_{nt} = W_n + \lambda_t W_n^2 + \lambda_t^2 W_n^3 + \ldots$, if $|\lambda_t| < 1$ and $w_{n,ij} < 1$. Anselin (2003) noted that the diagonal elements of $W_n^\iota, \iota \geqslant 2$ are inversely related to k_n. When W_n is row-normalized and symmetric, $\texttt{diag}(W_n^2) = k_{n,i}^{-1}$. In many spatial layouts such as Rook, Queen, group interactions where the variation in groups sizes becomes small when n gets large, etc, we can find the vanishing $\text{Var}(k_n)$, that is $\text{Var}(k_n) = o(1)$. See Yang (2010), and Liu and Yang (2015).

(i), (iii) and (iv) under Assumptions D are standard conditions put on the spatial weight matrices (Lee, 2004; Yang, 2018). Assumption D(ii) further allow the degree of spatial dependence to grow with n (Lee, 2004; Yang, 2018). Therefore, the concentrated estimating function (CEF) $S^{\star c}(\boldsymbol{\lambda})$ and its population counterpart play the key role in establishing the consistency of the AQS estimator $\hat{\theta}$ of θ.

Define the population counterpart of the AQS functions given in (4.7) as $\bar{S}^{\star}(\boldsymbol{\theta})$, where $\bar{S}^{\star}(\boldsymbol{\theta}) = \mathrm{E}[S^{\star}(\theta)]$. Given $\boldsymbol{\lambda}$, $\bar{S}^{\star}(\boldsymbol{\theta}) = 0$ is partially solved at:

$$\bar{\boldsymbol{\beta}}(\boldsymbol{\lambda}) = (X_N'\Omega X_N)^{-1}X_N'\Omega A_N(\boldsymbol{\lambda})\mathrm{E}(Y_N), \tag{4.12}$$

$$\bar{\sigma}^2(\boldsymbol{\lambda}) = \tfrac{1}{n(T-1)}\mathrm{E}[\bar{V}_N'(\boldsymbol{\lambda})\bar{V}_N(\boldsymbol{\lambda})], \tag{4.13}$$

where $\bar{V}_N(\boldsymbol{\lambda}) = \widetilde{V}_N|_{\beta=\bar{\beta}(\lambda)} = \Omega[A_N(\boldsymbol{\lambda})Y_N - X_N\bar{\boldsymbol{\beta}}(\boldsymbol{\lambda})]$, which can be expressed as another useful form to obtain detailed expressions for $\bar{\sigma}^2(\boldsymbol{\lambda})$ and thus $\bar{S}^{\star c}(\boldsymbol{\lambda})$:

$$\bar{V}_N(\boldsymbol{\lambda}) = \mathbf{M}\Omega A_N(\boldsymbol{\lambda})Y_N + \mathbf{P}\Omega A_N(\boldsymbol{\lambda})\widetilde{Y}_N, \tag{4.14}$$

where $\widetilde{Y}_N = Y_N - \mathrm{E}(Y_N)$, $\mathbf{M} = I_N - \Omega X_N(X_N'\Omega X_N)^{-1}X_N'\Omega$, and $\mathbf{P} = I_N - \mathbf{M}$.

Substituting $\bar{\boldsymbol{\beta}}(\boldsymbol{\lambda})$ and $\bar{\sigma}^2(\boldsymbol{\lambda})$ back into the $\boldsymbol{\lambda}$-component of $\bar{S}^{\star}(\theta)$ leads to the population counterpart of the CEF given in (4.10), as

$$\bar{S}^{\star c}(\boldsymbol{\lambda}) = \tfrac{1}{\bar{\sigma}^2(\boldsymbol{\lambda})}\mathrm{E}[\bar{\eta}_N'(\boldsymbol{\lambda})\bar{V}_N(\boldsymbol{\lambda})] + \tfrac{1}{\bar{\sigma}^2(\boldsymbol{\lambda})}\mathrm{E}[\bar{V}_N^{\circ\prime}(\boldsymbol{\lambda})G_N^{\circ}(\boldsymbol{\lambda})\bar{V}_N(\boldsymbol{\lambda})] \tag{4.15}$$

where $\bar{\eta}_N(\boldsymbol{\lambda}) = \texttt{blkdiag}(\bar{\eta}_{n1}(\boldsymbol{\lambda}), \ldots, \bar{\eta}_{nT}(\boldsymbol{\lambda}))$, $\bar{V}_N^{\circ}(\boldsymbol{\lambda}) = \texttt{blkdiag}(\bar{V}_{n1}(\boldsymbol{\lambda}), \ldots, \bar{V}_{nT}(\boldsymbol{\lambda}))$, $\bar{\eta}_{nt}(\boldsymbol{\lambda}) = \tilde{\eta}_{nt}(\bar{\boldsymbol{\beta}}(\boldsymbol{\lambda}), \boldsymbol{\lambda})$, and $\bar{V}_{nt}(\boldsymbol{\lambda}) = \tilde{V}_{nt}(\bar{\boldsymbol{\beta}}(\boldsymbol{\lambda}), \boldsymbol{\lambda}), t = 1\ldots T$.

Clearly, the AQS-estimator $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}_0$ is a zero of $S^{\star c}(\boldsymbol{\lambda})$, and $\boldsymbol{\lambda}_0$ is a zero of $\bar{S}^{\star c}(\boldsymbol{\lambda})$ as $\bar{\boldsymbol{\beta}}(\boldsymbol{\lambda}_0) = \boldsymbol{\beta}_0$ and $\bar{\sigma}^2(\boldsymbol{\lambda}_0) = \sigma_0^2$, i.e., $\boldsymbol{\lambda}_0$ is a zero of $\bar{S}^{\star c}(\boldsymbol{\lambda})$. Denote the overall sample size as $N = nT$ and the effective sample size as $N^* = n(T-1)$. Thus, by Theorem 5.9 of van der Vaart (1998), consistency of $\hat{\boldsymbol{\lambda}}$ follows from (a) the uniform convergence: $\sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \frac{1}{N^*}\left\|S^{\star c}(\boldsymbol{\lambda}) - \bar{S}^{\star c}(\boldsymbol{\lambda})\right\| \xrightarrow{p} 0$, and (b) the following identification uniqueness condition:

Assumption E: $\inf_{\boldsymbol{\lambda}:\, d(\boldsymbol{\lambda},\boldsymbol{\lambda}_0)\geq\varepsilon} \left\|\bar{S}^{\star c}(\boldsymbol{\lambda})\right\| > 0$ *for every* $\varepsilon > 0$*, where* $d(\boldsymbol{\lambda}, \boldsymbol{\lambda}_0)$ *is a measure of distance between* $\boldsymbol{\lambda}_0$ *and* $\boldsymbol{\lambda}$*.*

Theorem 4.1 *Under Assumptions A-E,* θ_0 *is identified. Furthermore, for the AQS-estimators* $\hat{\theta}$ *based on the AQS function,* $\hat{\theta} \xrightarrow{p} \theta_0$*.*

The derivation of the asymptotic distribution of the AQS-estimators $\hat{\theta}$ starts with a Taylor expansion of the joint AQS function $S^{\star}(\hat{\theta}) = 0$ at θ_0, and then we verify that the AQS functions $S^{\star}(\theta_0)$ is asymptotically normal and that the corresponding adjusted Hessian $\frac{\partial}{\partial\theta'}S^{\star}(\bar{\theta})$ has proper asymptotic behavior for some $\bar{\theta}$ lying between $\hat{\theta}$ and θ_0 elementwise. The central limit theorem (CLT) for linear-quadratic forms by Kelejian and Prucha (2001) would be sufficient to establish the asymptotic properties. Detained proof can be found in the Appendix. Let $\widetilde{V}_N = (\widetilde{V}'_{n1}, \ldots, \widetilde{V}'_{nT})'$ be the vector of elements $\{\widetilde{V}_{it}\}$, where the representation for the AQS functions given in (4.7) in terms of $\widetilde{V}_N$ is crucial in developing an OPMD method for estimating the robust VC matrix. More details will be discussed in Sec. 4.2.3.

Lemma 4.1 *Let z_t be a $T \times 1$ vector of element 1 in the tth position and 0 elsewhere, and define $Z_{Nt} = z_t \otimes I_n$.*

$$\widetilde{V}_{nt} \equiv \widetilde{V}_{nt}(\beta_0, \lambda_0) = Z'_{Nt}\widetilde{V}_N \tag{4.16}$$

Using the representation given in Lemma 4.1, the AQS function at θ_0 can be written as

$$S^{\star}(\theta_0) = \begin{cases} \Pi'_{1t}\widetilde{V}_N, \quad t = 1, \ldots, T, \\ \Pi'_{2t}\widetilde{V}_N + \widetilde{V}'_N\Phi_{1t}\widetilde{V}_N, \quad t = 1, \ldots, T, \\ \widetilde{V}'_N\Phi_2\widetilde{V}_N - \frac{n(T-1)}{2\sigma_0^2}, \end{cases} \tag{4.17}$$

where $\Pi_{1t} = \frac{1}{\sigma_0^2}Z_{Nt}X_{nt}$, $\Pi_{2t} = \frac{1}{\sigma_0^2}Z_{Nt}\tilde{\eta}_{nt0}$, $\Phi_{1t} = \frac{1}{\sigma_0^2}Z_{Nt}G^{\circ}_{nt0}Z'_{Nt}$, and $\Phi_2 = \frac{1}{2\sigma_0^4}\sum_{t=1}^{T} Z_{Nt}Z'_{Nt}$. The above representation for AQS functions given in (4.7) at θ_0 in terms of $\widetilde{V}_N = (\widetilde{V}'_{n1}, \ldots, \widetilde{V}'_{nT})'$ turns out to be very useful in establishing the asymptotic normality and estimating the variance-covariance (VC) matrix of the AQS vector.

Case of large n and small T. When T is small and fixed, the number of parameters, i.e., the munber of elements in the vector θ is fixed. Therefore, standard asymptotic results hold. We have the following theorem.

Theorem 4.2 *Under the assumptions of Theorem 4.1, we have, as $n \to \infty$, T is fixed,*

$$\sqrt{N^*}(\hat{\theta} - \theta_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty} I^{\circ-1}(\theta_0)\Sigma^{\circ}(\theta_0)I^{\circ-1}(\theta_0)\big],$$

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

Case of large n and large T. Although the short panels are more popular in the spatial empirical applications, the large panels, i.e., panels with large n and large T, are also important. Now, we focus on the large panels. As the dimensions of θ_0, $I^{\circ}(\theta_0)$ and $\Sigma^{\circ}(\theta_0)$ grow with the increase of T, the asymptotic arguments of the AQS estimator under small T case are no longer appropriate. Reflecting on the β_t and λ_t components of $I^{\circ}(\theta_0)$, where they will approach to zero as $n, T \to \infty$. This raises a issue of convergence rates for the components of the AQS estimator $\hat{\theta}$. To keep out theoretical arguments simple, we simply extend the results of Theorems (4.1) and (4.2) by letting n goes large first and then T, but T is smaller than n.

Adjustments are made on each component, β_t and λ_t, of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. From the information matrix $I(\theta_0) = -\mathrm{E}[\frac{\partial}{\partial\theta'}S^{\star}(\theta_0)]$ given in Appendix B, we see that the β_t block of $\frac{1}{nT}I(\theta_0)$ is $\frac{1}{nT}\big(\frac{T-1}{T\sigma_0^2}X'_{nt}X_{nt}\big)$, which approaches to a zero matrix as $n, T \to \infty$. However, the quantity with a different normalizing factor $\frac{1}{n}$, $\frac{1}{n}\big(\frac{T-1}{T\sigma_0^2}X'_{nt}X_{nt}\big)$ will converge to a positive definite matrix as $n, T \to \infty$. A similar phenomenon holds for the λ_t component of $\frac{1}{nT}I(\theta_0)$. As for the σ^2 component of $\frac{1}{nT}I(\theta_0)$, it is easy to see that it converges to a positive definite matrix as $n, T \to \infty$. These reveal that the convergence rate for $\hat{\beta}_t$ and $\hat{\lambda}_t$ are both $\sqrt{n}$, but the rate of convergence for $\hat{\sigma}^2$ is $\sqrt{nT}$.

We have the following results.

Theorem 4.3 *Under the assumptions of Theorem 4.2, we have,*

$$(i)\ \ \sqrt{n}\big(\hat{\beta}_t - \beta_{t0}\big) \xrightarrow{D} N\big(0,\ \Omega_t\big),\ \ \textit{for each } t,\ \textit{as } n \to \infty,\ \textit{and then } T \to \infty,$$
$$(ii)\ \sqrt{n}\big(\hat{\lambda}_t - \lambda_t 0\big) \xrightarrow{D} N\big(0,\ \tau^2_{\lambda_t}\big),\ \ \textit{for each } t,\ \textit{as } n \to \infty,\ \textit{and then } T \to \infty,$$
$$(iii)\ \sqrt{nT}\big(\hat{\sigma}^2 - \sigma_0^2\big) \xrightarrow{D} N\big(0,\ \tau^2_{\sigma^2}\big),\ \ \textit{as } n, T \to \infty,$$

where Ω_t and $\tau^2_{\lambda_t}$ are the limits of the corresponding components of $\frac{1}{T}I^{\circ -1}(\theta_0)\Sigma^{\circ}(\theta_0)I^{\circ -1}(\theta_0)$, and $\tau^2_{\sigma^2}$ the limits of the corresponding components of $I^{\circ -1}(\theta_0)\Sigma^{\circ}(\theta_0)I^{\circ -1}(\theta_0)$.

From the results Theorem 4.3, it is clear the joint inference for a finite number of components of $\boldsymbol{\beta}$ can be made by extending the result (i), the joint inference for a finite number of components of $\boldsymbol{\lambda}$ can be made by extending the result (ii), and the joint inference concerning a finite number of components of $\boldsymbol{\theta}$ can be made by extending and combining the results $(i)-(iii)$ of Theorem 4.3. These results provide useful tools for the practical applications in switching from the fixed T scenario to the large T scenario.

4.2.3 OPMD estimation of robust VC matrix

Valid inference requires consistent estimators of $I^{\circ}(\boldsymbol{\theta}_0)$ and $\Sigma^{\circ}(\boldsymbol{\theta}_0)$. Clearly, $\hat{I}^{\circ} = -\frac{1}{N^*}E[\frac{\partial}{\partial\boldsymbol{\theta}'}S^{\star}(\boldsymbol{\theta}_0)|_{\boldsymbol{\theta}_0=\hat{\boldsymbol{\theta}}}]$ or $I^{\circ}(\hat{\boldsymbol{\theta}})$ gives a consistent estimate of $I^{\circ}(\boldsymbol{\theta}_0)$ where the analytical expression of $I^{\circ}(\boldsymbol{\theta}_0)$ is given in Appendix C.2; However, the estimation of $\Sigma^{\circ}(\boldsymbol{\theta}_0)$ run into difficulties. The analytical expression of this quantity cannot be used as it contains 2nd, 3rd and 4th moments of idiosyncratic errors v_{it} that all change with i and hence the usual plug-in method does not apply. For the case of large n and small T, we may use the idea of Yang (2018) to give an OPMD estimate of $\Sigma^{\circ}(\boldsymbol{\theta}_0)$, taking the advantage that $\widetilde{V}_N$ can be estimated and are independent across i for each t. However, for the case of large n and large T, this method is invalid, as when T is large, the dependence over t in the transformed errors $\widetilde{V}_N$ cannot be ignored, and a new method is desired for the estimation of $\Sigma^{\diamond}(\boldsymbol{\theta}_0)$.

Case if large n and small T. From (4.17) we see that the AQS function $S^{\star}(\boldsymbol{\theta}_0)$ contains two types of elements:

$$\Pi'\widetilde{V}_N, \quad \text{and } \widetilde{V}_N'\Phi\widetilde{V}_N,$$

where Π and Φ are nonstochastic matrices (depending on $\boldsymbol{\theta}_0$) with Π being $nT\times p$ or $nT\times 1$, and Φ being $nT\times nT$. Partition Π according to $t=1,\ldots,T$, and denote the partitioned matrices by Π_t. Partition Φ according to $t,s=1,\ldots,T$, and denote the partitioned matrices by Φ_{ts}. Define $\Phi_{t+}=\sum_{s=1}^{T}\Phi_{ts}$, $t=1,\ldots,T$. For a square matrix A, let A^u, A^l and A^d be, respectively, its upper-triangular, lower-triangular, and diagonal matrix such that $A=A^u+A^l+A^d$. Let $\{\mathcal{F}_{n,i}\}$ be the increasing sequence of σ-fields generated by $(v_{j1},\ldots,v_{jT}, j=1,\ldots,i), i=1,\ldots,n$, $n\geqslant 1$. Clearly, $\mathcal{F}_{n,i-1}\subseteq\mathcal{F}_{n,i}$. Following lemma shows that $(\Pi'\widetilde{V}_N,\ \widetilde{V}_N'\Phi\widetilde{V}_N-\mathrm{E}(\widetilde{V}_N'\Phi\widetilde{V}_N)$ can be written as a sum of n uncorrelated terms, which turn out to be a vector martingale difference (M.D.) arrays.

Hence, the average of outer-product-of-martingale-differences (OPMD) give a consistent estimator of the variance of $\frac{1}{\sqrt{N^*}}(\Pi'\widetilde{V}_N, \widetilde{V}_N'\Phi\widetilde{V}_N)$.

Lemma 4.2 *Consider Model (4.1), the general* Π *is* $nT \times p$*, and denote* Π_{it} *as the ith row of* Π_t*. Define*

$$g_{\pi i} = \sum_{t=1}^T \Pi_{it}'\widetilde{V}_{it}, \tag{4.18}$$

$$g_{\Phi i} = \sum_{t=1}^T (\widetilde{V}_{it}\xi_{it} + \widetilde{V}_{it}\widetilde{V}_{it}^* - d_{it}), \tag{4.19}$$

where $\{\xi_{it}\} = \xi_t = \sum_{s=1}^T (\Phi_{st}^{u\prime} + \Phi_{ts}^l)\widetilde{V}_s$*,* $\widetilde{V}_t^* = \sum_{s=1}^T \Phi_{ts}^d \widetilde{V}_s$*,* $d_{it} = \frac{T-1}{T}\sigma_{v0}^2 r_{n,i}\Phi_{ii,tt}$*. Then,*

$$\begin{aligned} \Pi'\widetilde{V}_N &= \sum_{i=1}^n g_{\pi i}, \\ \widetilde{V}_N'\Phi\widetilde{V}_N - \mathrm{E}(\widetilde{V}_N'\Phi\widetilde{V}_N) &= \sum_{i=1}^n g_{\Phi i}, \end{aligned}$$

and $\{(g_{\pi i}', \ g_{\Phi i})', \mathcal{F}_{n,i}\}_{i=1}^n$ *form a vector M.D. sequence.*

Now, following Lemma 4.2, for each $\Pi_r, r = 1, 2,$ defined in (4.17), define $g_{\pi rti}$ according to (4.18); and for each $\Phi_r, r = 1, 2,$ define $g_{\Phi 1ti}$ and $g_{\Phi 2i}$ according to (4.19), respectively. For $t = 1, ..., T$, define

$$g_i = (g_{\pi 1ti}', \ g_{\pi 2ti} + g_{\Phi 1ti}, \ g_{\Phi 2i})'.$$

Then, $S^\star(\theta_0) = \sum_{i=1}^n g_i$, and $\{g_i, \mathcal{F}_{n,i}\}$ form a vector M.D. sequence. Let $\Sigma(\theta_0) = \mathrm{Var}[S^\star(\theta_0)]$, it follows that $\Sigma(\theta_0) = \sum_{i=1}^n \mathrm{E}(g_i g_i')$. The 'average' of the outer products of the estimated $g_i's$, i.e.,

$$\widehat{\Sigma}^\circ = \frac{1}{N^*}\sum_{i=1}^n \hat{g}_i\hat{g}_i', \tag{4.20}$$

which gives a consistent estimator of $\Sigma^\circ(\theta_0)$, where $\hat{g}_i$ is obtained by replacing θ_0 in g_i by $\hat{\theta}$ and $\widetilde{V}_N$ in g_i by its observed counterpart $\widehat{V}_N$.

Theorem 4.4 *Under the assumptions of Theorem 4.2, we have, as* $n \to \infty$ *(T fixed),*

$$\widehat{\Sigma}^\circ - \Sigma^\circ(\theta_0) = \frac{1}{N^*}\sum_{i=1}^n \left[\hat{g}_i\hat{g}_i' - \mathrm{E}(g_i g_i')\right] \xrightarrow{p} 0,$$

and hence, $I^{\circ-1}(\hat{\theta})\widehat{\Sigma}^\circ I^{\circ-1}(\hat{\theta}) - I^{\circ-1}(\theta_0)\Sigma^\circ(\theta_0)I^{\circ-1}(\theta_0) \xrightarrow{p} 0.$

Case of large n and large T. Summing over time dimension in $\Pi'\widetilde{V}_N$ and $\widetilde{V}_N'\Phi\widetilde{V}_N$ ignores the dependence of the elements of $\widetilde{V}_N$ over time. This is fine if T is fixed and the asymptotics depend only on n. However, when T is also large, the asymptotics depend on both n and T, and hence the dependence of the elements of $\widetilde{V}_N$ over t cannot be ignored. Let $j = 1, \ldots, N$ be the combined index for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. In the following, we decompose $\Pi'\widetilde{V}_N$ and $\widetilde{V}_N'\Phi\widetilde{V}_N$ in a different way:

Lemma 4.3 *Let Π, ξ_t and $\widetilde{V}_t^*$ be defined in Lemma 4.2. Define*

$$\begin{aligned} g_{\pi j}^{\diamond} &= \Pi_{N,j}'\widetilde{V}_{N,j}, &(4.21)\\ g_{\Phi j}^{\diamond} &= \widetilde{V}_{N,j}\xi_{N,j} + \widetilde{V}_{N,j}\widetilde{V}_{N,j}^* - d_{N,j}, &(4.22)\end{aligned}$$

where $\Pi_{N,j}$ is the jth row of Π; $\xi_{N,j}$ is the jth element of $\xi_N = (\xi_1', \ldots, \xi_T')$; $\widetilde{V}_{N,j}^*$ is the jth element of $\widetilde{V}_N^* = (\widetilde{V}_1^{*\prime}, \ldots, \widetilde{V}_T^{*\prime})$; and $d_{N,j}$ is the jth element of $\{d_{it}\}$. We freely switch between the single index j and the double indices (i, t). Thus, notations in (4.21)-(4.22) are interchangeable with the notations in (4.18)-(4.19). Then,

$$\begin{aligned} \Pi'\widetilde{V}_N &= \textstyle\sum_{j=1}^N g_{\pi j}^{\diamond},\\ \widetilde{V}_N'\Phi\widetilde{V}_N - E(\widetilde{V}_N'\Phi\widetilde{V}_N) &= \textstyle\sum_{j=1}^N g_{\Phi j}^{\diamond}.\end{aligned}$$

Now, following Lemma 4.3, for each $\Pi_r, r = 1, 2$, defined in (4.17), define $g_{\pi rtj}^{\diamond}$ according to (4.21); and for each $\Phi_r, r = 1, 2$, define $g_{\Phi 1tj}^{\diamond}$ and $g_{\Phi 2j}^{\diamond}$ according to (4.22), respectively. The AQS function can be written as $S^\star(\theta_0) = \sum_{j=1}^N s_{N,j}$, where

$$s_{N,j} = \begin{cases} g_{\pi 1tj}^{\diamond}, \\ g_{\pi 2tj}^{\diamond} + g_{\Phi 1tj}^{\diamond}, \quad t = 1, \ldots, T, \\ g_{\Phi 2j}^{\diamond}, \end{cases} \tag{4.23}$$

Dependence among the elements of $\widetilde{V}_N$ across t may exist, implying that $S_{N,it}$ and $S_{N,is}$ may be correlated and that the OPMD estimate of $\Sigma(\theta_0) = Var[S^\star(\theta_0)]$ under the small T case may not be strictly valid. As $S_{N,j}$ are uncorrelated across i for each t, we have:

$$\begin{aligned}\Sigma(\theta_0) &= \mathrm{Var}(\textstyle\sum_{i=1}^n \sum_{t=1}^T s_{N,it}) = \sum_{i=1}^n \mathrm{Var}(\sum_{t=1}^T s_{N,it}),\\ &= \textstyle\sum_{j=1}^N \mathrm{E}(s_{N,j}s_{N,j}') + 2\sum_{i=1}^n \sum_{t=2}^T \sum_{s=1}^{t-1} \mathrm{E}(s_{N,it}s_{N,is}').\end{aligned}$$

This provides the following estimation of the robust VC matrix when T is large.

$$\hat{\Sigma}(\theta) = \sum_{j=1}^{N} \hat{s}_{N,j}\hat{s}'_{N,j} + \sum_{i=1}^{n} \hat{\mathbf{r}}_{N,i},$$

where $\hat{\mathbf{r}}_{N,i} = 2\sum_{t=2}^{T}\sum_{s=1}^{t-1}(\hat{s}_{N,it}\hat{s}'_{N,is})$, which provides a correction on the cross-t correlations of the elements $\{s_{N,j}\}$ of the AQS function.

Theorem 4.5 *Under Assumption A-E, we have, as $n, T \to \infty$,*

$$\frac{1}{N}[\widehat{\Sigma} - \Sigma(\theta_0)] \xrightarrow{p} 0,$$

and hence, $I^{\diamond -1}(\hat{\theta})\widehat{\Sigma}^{\diamond}I^{\diamond -1}(\hat{\theta}) - I^{\diamond -1}(\theta_0)\Sigma^{\diamond}(\theta_0)I^{\diamond -1}(\theta_0) \xrightarrow{p} 0.$

4.3 Monte Carlo Study

Monte Carlo experiments are carried out to investigate the finite sample performance of (i) the proposed AQS-estimators of the FE-SPD model with TVC and unknown heteroskedasticity and (ii) the OPMD-based standard errors estimates of the AQS-estimators. The model we use in our Monte Carlo experiment is the SL-1FE SPD model, having two time-varying regressors:

$$Y_{nt} = \lambda_{t0}W_nY_{nt} + X_{nt}\beta_{t0} + c_{n0} + V_{nt},$$

where $t = 1, \ldots, T$.

In the Monte Carlo experiments, we choose $n = (50, 100, 200, 500)$, and T is initially set to be 3. We set $\beta'_{10} = (1.0, 1.0)$, $\beta'_{20} = (0.75, 1.25)$, $\beta'_{30} = (1.25, 0.75)$. As for the setting of λ_0, we consider several cases and set (i) $\lambda'_0 = (-0.5, -0.25, -0.75)$; (ii) $\lambda'_0 = (0.5, -0.25, -0.75)$; (iii) $\lambda'_0 = (0.5, 0.25, -0.75)$; (iv) $\lambda'_0 = (0.5, 0.25, 0.75)$. Finally, $\sigma_0 = 1$. The details of generating idiosyncratic errors, weight matrices, cross-sectional heteroskedasticity and regressors are as follows. Each set of Monte Carlo results is based on 2,000 Monte Carlo samples.

Spatial Weight matrices: The spatial weight matrices are generated according to group interaction schemes, neighbors occur in groups where each group member is spatially related to one another resulting in a symmetric W_n matrix. To ensure the heteroskedasticity effect does not fade as n increases (so that the regular QML-estimators are inconsistent), the degree of spatial dependence is fixed with respect to n. This is attained by fixing the possible group sizes in the Group Interaction scheme.

Heteroskedasticity: Similar to Lin and Lee (2010), the heteroskedasticity R_n is generated in two different ways, both emphasizes a nonlinear variance structure. {R1}: if the group size is smaller than the average group size, then $r_{n,i}$ is constructed to be the same as group size, otherwise, it is the square of the inverse of the group size. In this case, the variance function is increasing and then decreasing with the group size. {R2}: if the group size is larger than the average group size, then $r_{n,i}$ is constructed to be the same as group size, otherwise, it is the square of the inverse of the group size. In this second case, the variance function is decreasing and then increasing with the group size.

Regressors: The exogenous regressors are generated according to `REG1`: $X_{knt} \overset{iid}{\sim} N(0,1)$, which are independent across $k = 1, 2$, and $t = 1, \ldots, T$. In case when the spatial dependence is in the form of group interaction, the regressors can also be generated according to `REG2`: the *i*th value of the *k*th regressor in the *g*th group is such that $X_{kt,ig} \overset{iid}{\sim} (2z_g + z_{ig})/\sqrt{10}$, where $(z_g, z_{i,g}) \overset{iid}{\sim} N(0,1)$ when group interaction scheme is followed; $\{X_{kt,ig}\}$ are thus independent across *k* and *t*, but not across i.

Error Distribution: $v_{it} = \sigma_0 r_{ni} e_{it}$, are generated according to `err1`: $\{e_{it}\}$ are iid standard normal; `err2`: $\{e_{it}\}$ are iid normal mixture with 10% of values from $N(0,4)$ and the remaining from $N(0,1)$, standardized to have mean 0 and variance 1; and `err3`: $\{e_{it}\}$ iid chi-square with 3 degrees of freedom, standardized to have mean 0 and variance 1.[16]

Monte Carlo (empirical) means and standard deviations (sds) are reported for the QML-estimators and the AQS-estimators. Empirical averages of the standard errors (ses) are also reported. Due to the space constraint, partial Monte Carlo results are reported. The main results observed from the Monte Carlo experiments are summarized as follows:

(i) The QML-estimators (QMLEs) are inconsistent from Table 4.1-4.4, the AQS estimators (AQSEs) provide a useful consistent alternative with significantly less bias, and the OPMD-based standard error estimates for AQSEs are also consistent.

[16] See Yang (2015a) for more details on generating idiosyncratic errors, weight matrices and regressors.

(ii) The QMLEs for the spatial parameters are inconsistent in Table 4.1-4.3 and are likely to be consistent in Table 4.4. In both cases, AQSEs perform better than the QMLEs. The consistency (robustness) of the AQSE is clearly demonstrated by the Monte Carlo results and the corresponding values of the OPMD-based standard error estimates are very close to their Monte Carlo counterparts in general.

(iii) The QMLEs for the covariate effects are less affected by the unknown heteroskedasticity. The AQSEs for the covariate effects perform well as well.

(iv) The cases with larger T (unreported for brevity) were also investigated. Monte Carlo results show that the the pattern of inconsistency still remains for the QMLEs, but the proposed AQSEs and the OPMD-based estimate for the standard errors are still consistent and continue to perform well with significantly less bias, irrespective of whether the errors are normal or non-normal.

Table 4.1. Empirical Mean(sd)$[se]^*$ of CQS-Estimator, AQS-Estimator

SL One-Way Model, $T = 3$

θ		Normal Error					Normal Mixture					Chi-Square				
		QMLE		AQS-Est			QMLE		AQS-Est			QMLE		AQS-Est		
							$n = 50$									
β_{11}	1.00	.9927	(.170)	.9967	(.170)	[.151]	.9948	(.170)	.9987	(.170)	[.146]	.9953	(.170)	.9996	(.170)	[.150]
β_{21}	1.00	1.0268	(.178)	1.0363	(.175)	[.156]	1.0219	(.180)	1.0321	(.175)	[.149]	1.0328	(.175)	1.0413	(.168)	[.150]
β_{12}	0.75	.7385	(.137)	.7433	(.137)	[.124]	.7408	(.143)	.7453	(.142)	[.119]	.7357	(.148)	.7406	(.146)	[.123]
β_{22}	1.25	1.2940	(.102)	1.2839	(.099)	[.098]	1.2922	(.106)	1.2828	(.102)	[.094]	1.2915	(.102)	1.2819	(.098)	[.095]
β_{13}	1.25	1.2972	(.170)	1.2761	(.168)	[.138]	1.2943	(.170)	1.2748	(.165)	[.131]	1.2994	(.176)	1.2787	(.171)	[.139]
β_{23}	0.75	.7772	(.142)	.7672	(.140)	[.120]	.7763	(.144)	.7672	(.141)	[.115]	.7778	(.133)	.7691	(.133)	[.115]
λ_1	0.50	.4836	(.076)	.4818	(.075)	[.066]	.4861	(.077)	.4839	(.074)	[.063]	.4810	(.076)	.4795	(.073)	[.064]
λ_2	-0.25	-.2980	(.133)	-.2872	(.129)	[.119]	-.2965	(.139)	-.2865	(.132)	[.114]	-.2952	(.125)	-.2850	(.121)	[.113]
λ_3	-0.75	-.8292	(.211)	-.7982	(.206)	[.178]	-.8246	(.218)	-.7959	(.206)	[.170]	-.8319	(.203)	-.8023	(.198)	[.170]
σ_2^2	1.00	.5821	(.160)	.8762	(.241)	[.270]	.5842	(.314)	.8801	(.475)	[.380]	.5908	(.251)	.8894	(.378)	[.333]
							$n = 100$									
β_{11}	1.00	1.0159	(.095)	1.0133	(.093)	[.090]	1.0147	(.094)	1.0125	(.091)	[.086]	1.0128	(.095)	1.0101	(.093)	[.088]
β_{21}	1.00	.9996	(.102)	1.0098	(.100)	[.096]	1.0003	(.101)	1.0105	(.098)	[.092]	.9993	(.101)	1.0092	(.100)	[.094]
β_{12}	0.75	.7612	(.088)	.7555	(.087)	[.082]	.7589	(.088)	.7535	(.088)	[.080]	.7614	(.088)	.7558	(.088)	[.081]
β_{22}	1.25	1.2577	(.100)	1.2558	(.099)	[.094]	1.2604	(.099)	1.2585	(.098)	[.090]	1.2597	(.097)	1.2578	(.097)	[.092]
β_{13}	1.25	1.2808	(.105)	1.2599	(.103)	[.097]	1.2788	(.104)	1.2594	(.100)	[.093]	1.2790	(.109)	1.2580	(.105)	[.098]
β_{23}	0.75	.7706	(.086)	.7564	(.085)	[.077]	.7667	(.084)	.7535	(.082)	[.074]	.7683	(.081)	.7542	(.081)	[.075]
λ_1	0.50	.5013	(.051)	.4933	(.050)	[.050]	.5004	(.052)	.4924	(.050)	[.047]	.5012	(.052)	.4933	(.051)	[.049]
λ_2	-0.25	-.2816	(.100)	-.2672	(.097)	[.094]	-.2802	(.098)	-.2664	(.094)	[.090]	-.2804	(.101)	-.2659	(.096)	[.093]
λ_3	-0.75	-.8141	(.127)	-.7742	(.122)	[.123]	-.8108	(.133)	-.7734	(.121)	[.117]	-.8105	(.141)	-.7701	(.130)	[.125]
σ_2^2	1.00	.6266	(.121)	.9441	(.183)	[.219]	.6219	(.241)	.9376	(.364)	[.326]	.6274	(.187)	.9457	(.283)	[.274]

Note: $[se]^*$: Empirical averages of the standard errors, only for robust AQS-estimators

W_n are generated from Group Interaction scheme, replication number = 2000.

Table 4.1 (cont'd). Empirical Mean(sd)$[se]^*$ of CQS-Estimator, AQS-Estimator

SL One-Way Model, $T = 3$

θ		Normal Error					Normal Mixture					Chi-Square				
		QMLE		AQS-Est			QMLE		AQS-Est			QMLE		AQS-Est		
							$n = 200$									
β_{11}	1.00	.9982	(.061)	1.0050	(.061)	[.060]	.9980	(.062)	1.0049	(.061)	[.058]	.9986	(.062)	1.0054	(.061)	[.059]
β_{21}	1.00	.9932	(.067)	1.0039	(.066)	[.065]	.9933	(.068)	1.0040	(.066)	[.064]	.9959	(.066)	1.0067	(.065)	[.064]
β_{12}	0.75	.7553	(.072)	.7506	(.071)	[.070]	.7550	(.072)	.7504	(.071)	[.068]	.7548	(.075)	.7501	(.074)	[.070]
β_{22}	1.25	1.2583	(.070)	1.2518	(.070)	[.069]	1.2602	(.070)	1.2538	(.069)	[.067]	1.2614	(.071)	1.2548	(.070)	[.068]
β_{13}	1.25	1.2695	(.068)	1.2525	(.067)	[.065]	1.2673	(.069)	1.2503	(.067)	[.064]	1.2701	(.068)	1.2530	(.066)	[.064]
β_{23}	0.75	.7605	(.049)	.7492	(.048)	[.046]	.7611	(.049)	.7499	(.047)	[.045]	.7634	(.047)	.7522	(.046)	[.045]
λ_1	0.50	.5005	(.029)	.4969	(.029)	[.028]	.5003	(.030)	.4967	(.029)	[.027]	.5003	(.029)	.4967	(.028)	[.028]
λ_2	-0.25	-.2616	(.064)	-.2538	(.063)	[.062]	-.2620	(.064)	-.2543	(.063)	[.060]	-.2625	(.064)	-.2547	(.062)	[.061]
λ_3	-0.75	-.7800	(.089)	-.7524	(.086)	[.084]	-.7798	(.091)	-.7523	(.086)	[.082]	-.7850	(.088)	-.7573	(.084)	[.082]
σ_2^2	1.00	.6475	(.087)	.9738	(.131)	[.165]	.6482	(.177)	.9751	(.266)	[.262]	.6470	(.132)	.9731	(.198)	[.212]
							$n = 500$									
β_{11}	1.00	.9977	(.038)	1.0011	(.037)	[.038]	.9985	(.038)	1.0019	(.037)	[.038]	.9972	(.040)	1.0004	(.039)	[.038]
β_{21}	1.00	.9943	(.044)	1.0008	(.043)	[.044]	.9953	(.044)	1.0018	(.043)	[.043]	.9933	(.045)	.9998	(.044)	[.044]
β_{12}	0.75	.7580	(.037)	.7528	(.036)	[.036]	.7569	(.037)	.7519	(.036)	[.035]	.7559	(.038)	.7508	(.037)	[.036]
β_{22}	1.25	1.2616	(.046)	1.2538	(.045)	[.046]	1.2608	(.046)	1.2531	(.045)	[.045]	1.2593	(.046)	1.2515	(.045)	[.046]
β_{13}	1.25	1.2698	(.043)	1.2513	(.043)	[.044]	1.2693	(.045)	1.2511	(.044)	[.043]	1.2689	(.044)	1.2501	(.043)	[.044]
β_{23}	0.75	.7592	(.030)	.7495	(.029)	[.030]	.7590	(.031)	.7495	(.030)	[.030]	.7607	(.030)	.7509	(.030)	[.030]
λ_1	0.50	.5010	(.018)	.4989	(.018)	[.018]	.5006	(.019)	.4985	(.018)	[.018]	.5017	(.019)	.4996	(.019)	[.018]
λ_2	-0.25	-.2642	(.043)	-.2544	(.042)	[.043]	-.2630	(.043)	-.2534	(.042)	[.042]	-.2616	(.043)	-.2520	(.042)	[.043]
λ_3	-0.75	-.7808	(.053)	-.7530	(.051)	[.053]	-.7805	(.056)	-.7531	(.052)	[.053]	-.7805	(.054)	-.7523	(.051)	[.053]
σ_2^2	1.00	.6576	(.055)	.9888	(.083)	[.108]	.6547	(.108)	.9845	(.163)	[.177]	.6567	(.084)	.9876	(.127)	[.141]

Note: $[se]^*$: Empirical averages of the standard errors, only for robust AQS-estimators

W_n are generated from Group Interaction scheme, replication number = 2000.

Table 4.2. Empirical Mean(sd)$[se]^*$ of CQS-Estimator, AQS-Estimator

SL One-Way Model, $T = 3$

θ		Normal Error					Normal Mixture					Chi-Square				
		QMLE		AQS-Est			QMLE		AQS-Est			QMLE		AQS-Est		
							$n = 50$									
β_{11}	1.00	.9906	(.170)	.9920	(.170)	[.150]	.9934	(.171)	.9947	(.171)	[.145]	.9931	(.171)	.9943	(.171)	[.149]
β_{21}	1.00	1.0425	(.153)	1.0237	(.148)	[.133]	1.0405	(.153)	1.0221	(.148)	[.125]	1.0453	(.143)	1.0283	(.141)	[.125]
β_{12}	0.75	.7370	(.137)	.7441	(.136)	[.124]	.7390	(.143)	.7458	(.142)	[.118]	.7346	(.147)	.7416	(.146)	[.123]
β_{22}	1.25	1.3001	(.101)	1.2833	(.099)	[.098]	1.2987	(.105)	1.2826	(.101)	[.094]	1.2970	(.100)	1.2813	(.097)	[.095]
β_{13}	1.25	1.3002	(.170)	1.2757	(.168)	[.138]	1.2978	(.170)	1.2747	(.165)	[.131]	1.3018	(.176)	1.2782	(.171)	[.139]
β_{23}	0.75	.7818	(.143)	.7666	(.140)	[.120]	.7814	(.145)	.7670	(.141)	[.115]	.7829	(.135)	.7688	(.134)	[.115]
λ_1	-0.50	-.5693	(.194)	-.5394	(.187)	[.167]	-.5653	(.193)	-.5363	(.185)	[.157]	-.5728	(.181)	-.5455	(.178)	[.157]
λ_2	-0.25	-.3050	(.133)	-.2865	(.129)	[.118]	-.3041	(.137)	-.2863	(.132)	[.113]	-.3018	(.125)	-.2843	(.121)	[.113]
λ_3	-0.75	-.8381	(.211)	-.7967	(.206)	[.179]	-.8344	(.215)	-.7952	(.206)	[.170]	-.8400	(.201)	-.8008	(.198)	[.170]
σ_2^2	1.00	.5775	(.158)	.8714	(.239)	[.269]	.5800	(.311)	.8759	(.472)	[.378]	.5876	(.249)	.8865	(.376)	[.333]
							$n = 100$									
β_{11}	1.00	1.0294	(.090)	1.0116	(.089)	[.085]	1.0275	(.089)	1.0104	(.087)	[.081]	1.0262	(.092)	1.0083	(.090)	[.083]
β_{21}	1.00	1.0231	(.093)	1.0073	(.091)	[.086]	1.0229	(.092)	1.0073	(.090)	[.082]	1.0218	(.092)	1.0060	(.090)	[.084]
β_{12}	0.75	.7609	(.088)	.7557	(.087)	[.082]	.7586	(.089)	.7537	(.088)	[.080]	.7612	(.088)	.7561	(.088)	[.081]
β_{22}	1.25	1.2631	(.100)	1.2560	(.099)	[.094]	1.2656	(.099)	1.2586	(.098)	[.090]	1.2651	(.097)	1.2579	(.097)	[.092]
β_{13}	1.25	1.2848	(.105)	1.2602	(.103)	[.097]	1.2828	(.104)	1.2597	(.100)	[.093]	1.2828	(.109)	1.2582	(.105)	[.098]
β_{23}	0.75	.7703	(.086)	.7565	(.085)	[.077]	.7665	(.085)	.7536	(.082)	[.074]	.7684	(.081)	.7545	(.081)	[.075]
λ_1	-0.50	-.5483	(.141)	-.5166	(.134)	[.133]	-.5491	(.144)	-.5178	(.133)	[.125]	-.5464	(.144)	-.5150	(.135)	[.128]
λ_2	-0.25	-.2841	(.100)	-.2672	(.097)	[.094]	-.2826	(.097)	-.2663	(.094)	[.089]	-.2827	(.100)	-.2658	(.096)	[.093]
λ_3	-0.75	-.8167	(.127)	-.7742	(.122)	[.124]	-.8134	(.133)	-.7733	(.121)	[.117]	-.8127	(.141)	-.7700	(.130)	[.125]
σ_2^2	1.00	.6240	(.121)	.9411	(.182)	[.218]	.6190	(.239)	.9342	(.362)	[.324]	.6248	(.187)	.9428	(.283)	[.273]

Note: $[se]^*$: Empirical averages of the standard errors, only for robust AQS-estimators

W_n are generated from Group Interaction scheme, replication number = 2000.

Table 4.2(cont'd). Empirical Mean(sd)$[se]^*$ of CQS-Estimator, AQS-Estimator

SL One-Way Model, $T = 3$

θ		Normal Error					Normal Mixture					Chi-Square				
		QMLE		AQS-Est			QMLE		AQS-Est			QMLE		AQS-Est		
							$n = 200$									
β_{11}	1.00	1.0164	(.057)	1.0040	(.056)	[.055]	1.0160	(.058)	1.0037	(.056)	[.053]	1.0167	(.058)	1.0043	(.057)	[.055]
β_{21}	1.00	1.0161	(.063)	1.0029	(.062)	[.061]	1.0157	(.063)	1.0026	(.062)	[.060]	1.0187	(.064)	1.0055	(.063)	[.061]
β_{12}	0.75	.7542	(.072)	.7508	(.071)	[.070]	.7539	(.072)	.7506	(.071)	[.068]	.7536	(.075)	.7503	(.074)	[.069]
β_{22}	1.25	1.2609	(.071)	1.2519	(.070)	[.069]	1.2628	(.070)	1.2539	(.069)	[.067]	1.2640	(.071)	1.2549	(.070)	[.068]
β_{13}	1.25	1.2704	(.068)	1.2524	(.067)	[.065]	1.2682	(.069)	1.2502	(.067)	[.064]	1.2709	(.068)	1.2529	(.066)	[.064]
β_{23}	0.75	.7623	(.049)	.7492	(.048)	[.047]	.7628	(.049)	.7498	(.047)	[.045]	.7652	(.047)	.7522	(.047)	[.045]
λ_1	-0.50	-.5342	(.081)	-.5082	(.079)	[.077]	-.5341	(.081)	-.5083	(.077)	[.074]	-.5344	(.080)	-.5084	(.077)	[.076]
λ_2	-0.25	-.2658	(.064)	-.2539	(.063)	[.062]	-.2661	(.064)	-.2544	(.063)	[.060]	-.2667	(.063)	-.2547	(.062)	[.061]
λ_3	-0.75	-.7860	(.089)	-.7523	(.086)	[.084]	-.7858	(.090)	-.7521	(.086)	[.082]	-.7911	(.087)	-.7572	(.083)	[.082]
σ_2^2	1.00	.6457	(.087)	.9724	(.131)	[.165]	.6465	(.176)	.9740	(.266)	[.262]	.6452	(.131)	.9718	(.198)	[.211]
							$n = 500$									
β_{11}	1.00	1.0151	(.035)	1.0005	(.035)	[.035]	1.0158	(.035)	1.0013	(.034)	[.035]	1.0148	(.036)	1.0002	(.036)	[.035]
β_{21}	1.00	1.0155	(.040)	1.0002	(.039)	[.039]	1.0164	(.040)	1.0011	(.039)	[.039]	1.0149	(.040)	.9995	(.040)	[.039]
β_{12}	0.75	.7600	(.037)	.7528	(.036)	[.036]	.7590	(.037)	.7519	(.036)	[.035]	.7580	(.038)	.7508	(.037)	[.036]
β_{22}	1.25	1.2661	(.046)	1.2538	(.045)	[.046]	1.2654	(.046)	1.2532	(.045)	[.045]	1.2638	(.046)	1.2515	(.046)	[.046]
β_{13}	1.25	1.2723	(.044)	1.2513	(.043)	[.044]	1.2719	(.045)	1.2512	(.044)	[.043]	1.2714	(.044)	1.2502	(.043)	[.044]
β_{23}	0.75	.7616	(.030)	.7495	(.030)	[.030]	.7615	(.031)	.7496	(.030)	[.030]	.7631	(.030)	.7509	(.030)	[.030]
λ_1	-0.50	-.5302	(.051)	-.5025	(.049)	[.050]	-.5312	(.051)	-.5037	(.048)	[.050]	-.5286	(.052)	-.5009	(.051)	[.050]
λ_2	-0.25	-.2712	(.043)	-.2544	(.042)	[.043]	-.2700	(.043)	-.2534	(.041)	[.042]	-.2687	(.043)	-.2520	(.042)	[.043]
λ_3	-0.75	-.7899	(.053)	-.7530	(.052)	[.053]	-.7895	(.056)	-.7531	(.052)	[.053]	-.7896	(.054)	-.7523	(.051)	[.053]
σ_2^2	1.00	.6560	(.055)	.9883	(.083)	[.108]	.6530	(.108)	.9839	(.162)	[.177]	.6552	(.084)	.9872	(.127)	[.141]

Note: $[se]^*$: Empirical averages of the standard errors, only for robust AQS-estimators

W_n are generated from Group Interaction scheme, replication number = 2000.

Table 4.3. Empirical Mean(sd)$[se]^*$ of CQS-Estimator, AQS-Estimator

SL One-Way Model, $T = 3$

θ		Normal Error					Normal Mixture					Chi-Square				
		QMLE		AQS-Est			QMLE		AQS-Est			QMLE		AQS-Est		
							$n = 50$									
β_{11}	1.00	.9946	(.170)	.9969	(.170)	[.151]	.9965	(.170)	.9988	(.170)	[.146]	.9973	(.169)	.9998	(.170)	[.149]
β_{21}	1.00	1.0240	(.178)	1.0362	(.175)	[.157]	1.0193	(.179)	1.0320	(.175)	[.149]	1.0297	(.174)	1.0412	(.168)	[.150]
β_{12}	0.75	.7390	(.134)	.7465	(.134)	[.122]	.7415	(.140)	.7485	(.140)	[.117]	.7359	(.144)	.7435	(.143)	[.122]
β_{22}	1.25	1.2857	(.107)	1.2876	(.104)	[.103]	1.2849	(.110)	1.2869	(.106)	[.100]	1.2838	(.106)	1.2857	(.102)	[.100]
β_{13}	1.25	1.2951	(.171)	1.2761	(.168)	[.138]	1.2923	(.171)	1.2747	(.165)	[.131]	1.2977	(.177)	1.2789	(.171)	[.139]
β_{23}	0.75	.7744	(.142)	.7676	(.140)	[.120]	.7738	(.144)	.7677	(.141)	[.115]	.7751	(.132)	.7695	(.133)	[.115]
λ_1	0.50	.4854	(.076)	.4817	(.075)	[.066]	.4877	(.077)	.4838	(.074)	[.063]	.4828	(.076)	.4795	(.073)	[.064]
λ_2	0.25	.2266	(.084)	.2255	(.083)	[.076]	.2268	(.087)	.2257	(.084)	[.073]	.2276	(.082)	.2266	(.079)	[.073]
λ_3	-0.75	-.8237	(.212)	-.7984	(.205)	[.178]	-.8194	(.219)	-.7962	(.207)	[.170]	-.8267	(.203)	-.8027	(.197)	[.170]
σ_2^2	1.00	.5843	(.161)	.8791	(.242)	[.271]	.5869	(.317)	.8837	(.478)	[.382]	.5924	(.252)	.8913	(.379)	[.334]
							$n = 100$									
β_{11}	1.00	1.0112	(.096)	1.0135	(.094)	[.091]	1.0101	(.095)	1.0127	(.092)	[.087]	1.0084	(.095)	1.0104	(.094)	[.089]
β_{21}	1.00	.9977	(.102)	1.0098	(.100)	[.096]	.9984	(.101)	1.0104	(.098)	[.092]	.9975	(.101)	1.0092	(.100)	[.094]
β_{12}	0.75	.7557	(.089)	.7567	(.089)	[.084]	.7536	(.089)	.7547	(.089)	[.082]	.7562	(.089)	.7570	(.089)	[.082]
β_{22}	1.25	1.2524	(.102)	1.2575	(.101)	[.096]	1.2552	(.100)	1.2602	(.100)	[.092]	1.2548	(.100)	1.2595	(.099)	[.094]
β_{13}	1.25	1.2800	(.105)	1.2597	(.103)	[.097]	1.2780	(.104)	1.2592	(.100)	[.093]	1.2783	(.110)	1.2578	(.105)	[.098]
β_{23}	0.75	.7689	(.086)	.7565	(.085)	[.078]	.7650	(.085)	.7536	(.083)	[.074]	.7664	(.080)	.7542	(.081)	[.075]
λ_1	0.50	.5018	(.051)	.4932	(.050)	[.050]	.5009	(.052)	.4924	(.050)	[.047]	.5017	(.052)	.4933	(.051)	[.049]
λ_2	0.25	.2386	(.063)	.2388	(.061)	[.060]	.2393	(.061)	.2392	(.059)	[.057]	.2391	(.061)	.2396	(.060)	[.059]
λ_3	-0.75	-.8087	(.127)	-.7741	(.122)	[.123]	-.8054	(.134)	-.7733	(.121)	[.118]	-.8053	(.142)	-.7701	(.130)	[.125]
σ_2^2	1.00	.6281	(.121)	.9456	(.183)	[.219]	.6234	(.241)	.9391	(.365)	[.326]	.6292	(.188)	.9476	(.284)	[.275]

Note: $[se]^*$: Empirical averages of the standard errors, only for robust AQS-estimators

W_n are generated from Group Interaction scheme, replication number = 2000.

Table 4.3(cont'd). Empirical Mean(sd)$[se]^*$ of CQS-Estimator, AQS-Estimator

SL One-Way Model, $T = 3$

θ		Normal Error					Normal Mixture					Chi-Square				
		QMLE		AQS-Est			QMLE		AQS-Est			QMLE		AQS-Est		
							$n = 200$									
β_{11}	1.00	.9971	(.061)	1.0050	(.061)	[.060]	.9970	(.062)	1.0050	(.061)	[.058]	.9976	(.062)	1.0055	(.061)	[.059]
β_{21}	1.00	.9929	(.067)	1.0039	(.066)	[.065]	.9929	(.068)	1.0040	(.066)	[.064]	.9956	(.066)	1.0067	(.065)	[.064]
β_{12}	0.75	.7474	(.074)	.7513	(.074)	[.073]	.7472	(.075)	.7512	(.074)	[.071]	.7469	(.076)	.7509	(.076)	[.072]
β_{22}	1.25	1.2482	(.074)	1.2526	(.073)	[.072]	1.2503	(.073)	1.2547	(.072)	[.070]	1.2515	(.074)	1.2558	(.073)	[.071]
β_{13}	1.25	1.2689	(.068)	1.2525	(.067)	[.065]	1.2666	(.069)	1.2503	(.067)	[.064]	1.2694	(.068)	1.2531	(.066)	[.064]
β_{23}	0.75	.7596	(.049)	.7493	(.048)	[.047]	.7602	(.049)	.7499	(.047)	[.045]	.7625	(.047)	.7522	(.046)	[.045]
λ_1	0.50	.5013	(.029)	.4968	(.029)	[.028]	.5011	(.030)	.4966	(.029)	[.027]	.5011	(.029)	.4967	(.028)	[.028]
λ_2	0.25	.2498	(.041)	.2472	(.040)	[.040]	.2494	(.041)	.2468	(.040)	[.039]	.2491	(.041)	.2466	(.040)	[.039]
λ_3	-0.75	-.7777	(.089)	-.7525	(.086)	[.084]	-.7775	(.092)	-.7524	(.086)	[.082]	-.7827	(.088)	-.7575	(.084)	[.083]
σ_2^2	1.00	.6479	(.087)	.9742	(.131)	[.165]	.6487	(.177)	.9756	(.266)	[.262]	.6474	(.132)	.9737	(.198)	[.212]
							$n = 500$									
β_{11}	1.00	.9950	(.038)	1.0010	(.037)	[.038]	.9958	(.038)	1.0019	(.037)	[.038]	.9944	(.040)	1.0005	(.039)	[.038]
β_{21}	1.00	.9924	(.044)	1.0008	(.043)	[.044]	.9934	(.044)	1.0018	(.043)	[.043]	.9914	(.045)	.9998	(.044)	[.043]
β_{12}	0.75	.7492	(.038)	.7530	(.037)	[.037]	.7484	(.038)	.7522	(.037)	[.037]	.7473	(.039)	.7510	(.038)	[.037]
β_{22}	1.25	1.2485	(.048)	1.2541	(.047)	[.048]	1.2481	(.047)	1.2536	(.046)	[.047]	1.2463	(.048)	1.2518	(.047)	[.047]
β_{13}	1.25	1.2693	(.043)	1.2513	(.043)	[.044]	1.2689	(.045)	1.2511	(.044)	[.043]	1.2685	(.044)	1.2501	(.043)	[.044]
β_{23}	0.75	.7581	(.030)	.7495	(.029)	[.030]	.7580	(.031)	.7496	(.030)	[.030]	.7596	(.030)	.7509	(.030)	[.030]
λ_1	0.50	.5026	(.018)	.4989	(.018)	[.019]	.5022	(.019)	.4985	(.018)	[.018]	.5033	(.019)	.4996	(.019)	[.018]
λ_2	0.25	.2502	(.027)	.2472	(.027)	[.027]	.2507	(.027)	.2477	(.026)	[.027]	.2516	(.027)	.2487	(.026)	[.027]
λ_3	-0.75	-.7770	(.053)	-.7530	(.051)	[.053]	-.7767	(.057)	-.7531	(.052)	[.053]	-.7768	(.054)	-.7524	(.051)	[.053]
σ_2^2	1.00	.6580	(.055)	.9892	(.083)	[.108]	.6550	(.108)	.9848	(.163)	[.177]	.6570	(.084)	.9878	(.127)	[.141]

Note: $[se]^*$: Empirical averages of the standard errors, only for robust AQS-estimators

W_n are generated from Group Interaction scheme, replication number = 2000.

Table 4.4. Empirical Mean(sd)$[se]^*$ of CQS-Estimator, AQS-Estimator

SL One-Way Model, $T = 3$

θ		Normal Error					Normal Mixture					Chi-Square				
		QMLE		AQS-Est			QMLE		AQS-Est			QMLE		AQS-Est		
$n = 50$																
β_{11}	1.00	.9992	(.166)	.9981	(.166)	[.153]	.9995	(.167)	.9984	(.167)	[.149]	.9988	(.163)	.9977	(.163)	[.151]
β_{21}	1.00	1.0418	(.137)	1.0293	(.138)	[.131]	1.0438	(.140)	1.0314	(.141)	[.128]	1.0470	(.138)	1.0348	(.139)	[.128]
β_{12}	0.75	.7335	(.145)	.7382	(.144)	[.138]	.7317	(.146)	.7364	(.146)	[.136]	.7345	(.142)	.7392	(.142)	[.136]
β_{22}	1.25	1.2872	(.140)	1.2768	(.141)	[.129]	1.2880	(.138)	1.2778	(.139)	[.125]	1.2893	(.138)	1.2790	(.139)	[.128]
β_{13}	1.25	1.3092	(.152)	1.2906	(.154)	[.150]	1.3149	(.153)	1.2964	(.154)	[.147]	1.3019	(.153)	1.2835	(.154)	[.146]
β_{23}	0.75	.7861	(.109)	.7752	(.109)	[.103]	.7893	(.107)	.7785	(.107)	[.101]	.7827	(.109)	.7719	(.110)	[.101]
λ_1	0.50	.4777	(.048)	.4844	(.049)	[.048]	.4768	(.049)	.4834	(.050)	[.047]	.4768	(.050)	.4833	(.050)	[.047]
λ_2	0.25	.2239	(.072)	.2310	(.073)	[.068]	.2229	(.071)	.2299	(.071)	[.066]	.2234	(.071)	.2304	(.072)	[.067]
λ_3	0.75	.7364	(.028)	.7406	(.028)	[.027]	.7353	(.028)	.7395	(.028)	[.027]	.7368	(.028)	.7410	(.028)	[.027]
σ_2^2	1.00	.6130	(.108)	.9181	(.162)	[.170]	.6123	(.219)	.9170	(.328)	[.277]	.6075	(.160)	.9098	(.240)	[.217]
$n = 100$																
β_{11}	1.00	1.0254	(.093)	1.0139	(.093)	[.088]	1.0259	(.093)	1.0144	(.093)	[.087]	1.0261	(.091)	1.0147	(.091)	[.088]
β_{21}	1.00	1.0265	(.089)	1.0157	(.090)	[.088]	1.0275	(.091)	1.0167	(.092)	[.086]	1.0255	(.088)	1.0148	(.089)	[.087]
β_{12}	0.75	.7594	(.081)	.7564	(.081)	[.079]	.7601	(.082)	.7570	(.082)	[.078]	.7584	(.083)	.7554	(.083)	[.078]
β_{22}	1.25	1.2596	(.114)	1.2549	(.114)	[.109]	1.2622	(.117)	1.2576	(.117)	[.107]	1.2627	(.112)	1.2581	(.113)	[.107]
β_{13}	1.25	1.2781	(.114)	1.2622	(.115)	[.110]	1.2777	(.112)	1.2617	(.113)	[.107]	1.2798	(.110)	1.2640	(.111)	[.109]
β_{23}	0.75	.7660	(.078)	.7572	(.078)	[.075]	.7642	(.078)	.7553	(.078)	[.073]	.7675	(.076)	.7588	(.076)	[.075]
λ_1	0.50	.4829	(.046)	.4898	(.047)	[.046]	.4827	(.047)	.4896	(.048)	[.045]	.4835	(.046)	.4903	(.046)	[.045]
λ_2	0.25	.2355	(.050)	.2417	(.050)	[.049]	.2351	(.050)	.2413	(.050)	[.048]	.2345	(.051)	.2406	(.052)	[.049]
λ_3	0.75	.7430	(.021)	.7467	(.021)	[.020]	.7433	(.020)	.7471	(.020)	[.020]	.7426	(.020)	.7463	(.021)	[.020]
σ_2^2	1.00	.6408	(.080)	.9598	(.120)	[.130]	.6410	(.161)	.9602	(.241)	[.225]	.6389	(.120)	.9569	(.179)	[.174]

Note: $[se]^*$: Empirical averages of the standard errors, only for robust AQS-estimators

W_n are generated from Group Interaction scheme, replication number = 2000.

Table 4.4 (cont'd). Empirical Mean(sd)$[se]^*$ of CQS-Estimator, AQS-Estimator

$\texttt{SL}$ One-Way Model, $T = 3$

θ		Normal Error					Normal Mixture					Chi-Square				
		QMLE		AQS-Est			QMLE		AQS-Est			QMLE		AQS-Est		
							$n = 200$									
β_{11}	1.00	1.0149	(.056)	1.0060	(.056)	[.056]	1.0164	(.057)	1.0075	(.057)	[.056]	1.0136	(.057)	1.0047	(.057)	[.056]
β_{21}	1.00	1.0133	(.071)	1.0042	(.071)	[.071]	1.0145	(.070)	1.0055	(.070)	[.071]	1.0159	(.071)	1.0068	(.071)	[.071]
β_{12}	0.75	.7546	(.051)	.7522	(.051)	[.049]	.7551	(.051)	.7527	(.051)	[.049]	.7566	(.050)	.7543	(.050)	[.049]
β_{22}	1.25	1.2611	(.062)	1.2551	(.062)	[.061]	1.2611	(.061)	1.2552	(.062)	[.060]	1.2612	(.062)	1.2553	(.062)	[.061]
β_{13}	1.25	1.2687	(.072)	1.2553	(.073)	[.072]	1.2686	(.073)	1.2553	(.073)	[.072]	1.2693	(.073)	1.2559	(.073)	[.072]
β_{23}	0.75	.7628	(.047)	.7537	(.047)	[.048]	.7635	(.048)	.7544	(.047)	[.047]	.7625	(.047)	.7534	(.047)	[.048]
λ_1	0.50	.4907	(.025)	.4965	(.026)	[.027]	.4898	(.026)	.4956	(.026)	[.026]	.4903	(.026)	.4961	(.026)	[.026]
λ_2	0.25	.2413	(.032)	.2461	(.033)	[.032]	.2409	(.032)	.2457	(.032)	[.032]	.2408	(.033)	.2455	(.034)	[.032]
λ_3	0.75	.7450	(.012)	.7483	(.012)	[.013]	.7450	(.013)	.7483	(.013)	[.013]	.7451	(.013)	.7483	(.013)	[.013]
σ_2^2	1.00	.6537	(.058)	.9794	(.086)	[.095]	.6513	(.118)	.9758	(.176)	[.169]	.6534	(.085)	.9789	(.127)	[.132]
							$n = 500$									
β_{11}	1.00	1.0127	(.038)	1.0027	(.038)	[.037]	1.0135	(.037)	1.0035	(.037)	[.037]	1.0135	(.039)	1.0036	(.039)	[.037]
β_{21}	1.00	1.0131	(.039)	1.0026	(.039)	[.039]	1.0133	(.038)	1.0029	(.039)	[.039]	1.0147	(.039)	1.0043	(.040)	[.039]
β_{12}	0.75	.7581	(.035)	.7535	(.035)	[.035]	.7573	(.036)	.7527	(.036)	[.035]	.7567	(.035)	.7521	(.035)	[.035]
β_{22}	1.25	1.2639	(.045)	1.2560	(.046)	[.045]	1.2634	(.044)	1.2555	(.045)	[.045]	1.2607	(.043)	1.2528	(.043)	[.045]
β_{13}	1.25	1.2688	(.048)	1.2542	(.048)	[.047]	1.2687	(.048)	1.2540	(.048)	[.047]	1.2678	(.047)	1.2530	(.047)	[.047]
β_{23}	0.75	.7596	(.032)	.7512	(.032)	[.032]	.7598	(.032)	.7513	(.032)	[.032]	.7598	(.033)	.7513	(.032)	[.032]
λ_1	0.50	.4924	(.016)	.4984	(.016)	[.017]	.4922	(.016)	.4982	(.016)	[.016]	.4917	(.017)	.4977	(.017)	[.017]
λ_2	0.25	.2403	(.023)	.2467	(.024)	[.024]	.2404	(.024)	.2469	(.024)	[.024]	.2415	(.023)	.2479	(.024)	[.024]
λ_3	0.75	.7457	(.008)	.7491	(.008)	[.008]	.7456	(.008)	.7490	(.008)	[.008]	.7459	(.008)	.7494	(.008)	[.008]
σ_2^2	1.00	.6620	(.037)	.9918	(.055)	[.062]	.6624	(.074)	.9922	(.110)	[.112]	.6615	(.054)	.9909	(.081)	[.086]

Note: $[se]^*$: Empirical averages of the standard errors, only for robust AQS-estimators

W_n are generated from Group Interaction scheme, replication number = 2000.

4.4 Conclusion

In this paper, we propose a new estimation and inference method for the fixed effects spatial panel data (FE-SPD) model with time varying coefficients and unknown heteroskedasticity and non-normality of the disturbances. Traditional QML estimators are inconsistent in general when allowing for the unkown heteroskedasiticy, therefore we propose robust adjusted quasi score (AQS) methods, which leads to a set of unbiased and robust estimating equations. For the robust statistic inferences, we propose an outer-product-of-martingale-differences (OPMD) method to estimate the variance of the AQS functions, which together with the expected negative Hessian matrices, leading to robust estimator of the variance-covariance (VC) matrix of the AQS estimators. The Monte Carlo results show that both the AQS-estimators and the OPMD-based standard error estimators perform very well, both are robust against unknown heteroskedasticity and non-normality.

The studies in this paper provide a useful tool for applied researchers who are investigating economic process, for example, housing decisions, unemployment, price decisions, crime rates, trade flows, etc., exhibit time heterogeneity patterns and unknown heteroskedasticity. In case of FE-SPD model with temporal heterogeneity, this paper proposes an AQS strategy for robust estimation and inferences. For future studies on more general models, where the two-way fixed effects, can be interactive or additive, are added, the AQS method may be able to provide an alternative to estimate. It would also be interesting that the studies can be extended by including (i) higher-order spatial terms, (ii) serial correlation (iii) dynamic effects in the model. These extensions are interesting but clearly beyond the scope of the current paper, which will be in our future research agenda.

5 Conclusion and Further Research

This dissertation studies the fixed effects spatial panel data (FE-SPD) models with temporal heterogeneity. Generally, we firstly propose an AQS-test to detect the exsitence of temporal heterogeneity, and then we propose a set of AQS-based estimation and inference methods for FE-SPD models with time-varying coefficients (TVC), with extension to allow for unknown heteroskedasticity and non-normality.

The robust AQS-tests have excellent performance and allow researchers to control unobserved temporal heterogeneity in regression slope and spatial parameters. In a spatial panel data model, the temporal heterogeneity may occur only in certain spatial units, not all the spatial units. However, the AQS-tests proposed in this dissertation cannot identify which spatial units are subject to temporal heterogeneity and which are not. Therefore, a more efficient specification test can be developed in the future.

The AQS-estimators are consistent under both homoscedastic and heteroskedastic errors, therefore it provides useful tools for the applied researchers. In an empirical application, when the observation period T is very large, estimating parameters on a period-by-period basis would lead to a large set of results. It is better to apply the AQS-test to detect the break points firstly, therefore the estimations between the two neighbouring break points are based on the assumption of temporal homogeneity. Under this way, we can avoid a big table containing too much results and it also allows us to learn the temporal pattern easily since we can see the structure breaks directly.

Researchers who want to learn the temporal pattern of an empirical application can start from the AQS-based specification test, once it rejects the hypothesis of temporal homogeneity, they can apply the AQS-based estimation and inference methods. As most of the previous literature are based on the assumption of temporal homogeneity, therefore it would be more interesting and meaningful to compare the results under different temporal assumptions.

Time heterogeneity pattern is an important feature in cunrrent economic process, for example, housing decisions, unemployment, crime rates and trade flows. The study provides a useful tool for applied researchers who are investigating these economic activities.

The dissertation can be extended in several directions. For future research, we can allow for (i) higher-order spatial terms (ii) interactive fixed effects (iii) dynamic effects (iv) serial correlation in the model to apply our methods in more practical applications.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., Tahbaz-Salehi, A., 2012. The network origins of aggregate fluctuations. *Econometrica* 80, 1977-2016.

Andrews, D.W.K., 1992. Generic uniform Convergence. *Econometric Theory* 8, 241-257.

Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. The Netherlands: Kluwer Academic Publishers.

Anselin, L., Florax, R. J. (1995). New directions in spatial econometrics: Introduction. In New directions in spatial econometrics (pp. 3-18). Springer, Berlin, Heidelberg.

Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26, 153-166.

Aquaro, M., Bailey, N., Pesaran, M. H., 2015. Quasi Maximum Likelihood Estimation of Spatial Models with Heterogeneous Coefficients. USC-INET Research Paper No. 15-17 .

Badinger, H., Egger, P., 2011. Estimation of higher-order spatial autoregressive cross-section models with heteroskedastic disturbances. *Papers in Regional Science* 90, 213-235.

Badinger, H., Egger, P., 2015. Fixed effects and random effects estimation of higher-order spatial autoregressive models with spatial autoregressive and heteroskedastic disturbances. *Spatial Economic Analysis* 10, 11-35.

Bai, J., 2010. Common breaks in means and variances for panel data. *Journal of Econometrics* 157, 78-92.

Baltagi, B. H., 2013. *Econometric analysis of panel data, 5th ed.* John Wiley and Sons.

Baltagi, B. H., Egger, P. H., Kesina, M., 2016. Firm-level productivity spillovers in China's chemical industry: a spatial Hausman-Taylor approach. *Journal of Applied Econometrics* 31, 214-248.S

Baltagi, B. H., Griffin, J. M., Xiong, W., 2000. To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* 82, 117-126.

Baltagi, B. H., Levin, D., 1986. Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation, and advertising reconsidered. *Review of Economics and Statistics* 68, 148-155.

Baltagi, B. H., Levin, D., 1992. Cigarette taxation: Raising revenue and reducing consumption. *Structure Change and Economic Dynamics* 3, 321-335.

Baltagi, B. H., Pinnoi, N., 1995. Public Capital Stock and State Productivity Growth: Further Evidence from an Error Components Model. *Empirical Economics* 20, 331-359.

Baltagi, B. H., Yang, Z. L., 2013a. Standardized LM tests for spatial error dependence in linear or panel regressions. *The Econometrics Journal* 16, 103-134.

Baltagi, B. H., Yang, Z. L., 2013b. Heteroskedasticity and non-normality robust LM tests for spatial dependence. *Regional Science and Urban Economics* 43, 725-739.

Bernstein, D. S., 2009. *Matrix Mathematics: Theory, Facts, and Formulas. 2nd edition*. Princeton University Press, Princeton.

Blasques, F., Koopman S. J., Lucas, A., Schaumburg, J., 2016. Spillover dynamics for system risk measurement using spatial financial time series models. *Journal of Econometrics* 195, 211-223.

Breusch, T., Pagan, A., 1979. A simple test for heteroskedasticity and random coefficient variation. *Econometrica 47* 1287-1294.

Cameron, A. C., Trivedi, P. K., 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Case, A. C., 1991. Spatial Patterns in Household Demand. *Econometrica* 59, 953-965.

Case, A. C., Rosen, H. S., Hines, J. R., 1993. Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States. *Journal of Public Economics* 52, 285-307.

Catania L., Billé, A. G., 2017. Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics* 32, 1178-1196.

Chow, G. C., 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28, 591-605.

Chow, G. C. (1984). Random and changing coefficient models. Handbook of econometrics, 2, 1213-1245.

Cooley, T. F., Prescott, E. C. (1976). Estimation in the presence of stochastic parameter variation. Econometrica: journal of the Econometric Society, 167-184.

Davidson, J., 1994. *Stochastic Limit Theory*. Oxford University Press, Oxford.

Feng, Q., Kao, C., Lazarová, S., 2009. Estimation of change points in panel models. *Working paper*, Syracuse University.

Ferreira, E., Nunez-Anton, V., Rodriguez-Poo, J. (2000). Semiparametric approaches to signal extraction problems in economic time series. Computational Statistics Data Analysis, 33(3), 315-333.

Glaeser, E. L., Sacerdote, B., Scheinkman, J. A. (1996). Crime and social interactions. *The Quarterly Journal of Economics* 111, 507-548.

Hamilton, J. L., 1972. The demand for cigarettes: advertising, the health care, and the cigarette advertising ban. *Review of Economics and Statistics* 54, 401-411.

Jin, F., Lee, L. F., 2012. Approximated likelihood and root estimators for spatial interaction in spatial autoregressive models. *Regional Science and Urban Economics* 42 446-458.

Kelejian H. H., Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.

Kelejian, H. H., Prucha, I. R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics, 104, 219-257.

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.

Lancaster, T., 2000. The incidental parameter problem since 1948. *Journal of Econometrics* 95, 391-413.

Lee, L. F., 2002. Consistency and efficiency of least squares estimation for mixed regressive spatial autoregressive models. *Econometric Theory* 18, 252-277.

Lee, L. F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.

Lee, L. F., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154, 165-185.

LeSage, J. P., 1997. Bayesian estimation of spatial autoregressive models. *International Regional Science Review* 20, 113-129.

LeSage, J., Pace, R. K., 2009. *Introduction to spatial econometrics*. Chapman and Hall/CRC.

LeSage, J. P., Vance, C., Chih, Y. Y., 2016. A Bayesian heterogeneous coefficients spatial autoregressive panel data model of retail fuel price rivalry. *Ruhr Economic Papers*, No. 617

LeSage, J. P., Vance, C., Chih, Y. Y., 2017. Bayesian heterogeneous coefficients spatial autoregressive panel data model of retail fuel duopoly pricing. *Regional Science and Urban Economics* 62, 46-55.

Li, G., Wong, K. K., Song, H., Witt, S. F. (2006). Tourism demand forecasting: A time-varying parameter error correction model. Journal of Travel Research, 45(2), 175-185.

Li, J., Liao, Z. (2018). Uniform Nonparametric Inference for Time Series.

Li, K. P., 2018. Spatial panel data models with structural changes. *MPRA Working Paper* No. 85388.

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

Liao, W., 2008. Bayesian Inference of Structural Breaks in Time Varying Volatility Models. *NYU PhD Thesis*.

Lin, X., Lee, L.-F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157, 34-52.

Liu, S. F. and Yang, Z. L. (2015). Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and normality. *Regional Science and Urban Economics* 52, 50-70.

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.

McGuiness, T., Cowling, K., 1975. Advertising and the aggregate demand for cigarettes. *European Economic Review* 6, 311-328.

Munnell, A. H., 1990. How does public infrastructure affect regional economic performance? *New England Economic Review*, Sep./Oct., 11-32.

Neyman, J., Scott, E. L., 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16, 1-32.

Nicholls, D. F., Pagan, A. R. (1985). 16 Varying coefficient regression. Handbook of statistics, 5, 413-449.

Orbe, S., Ferreira, E., Rodriguez-Poo, J. (2005). Nonparametric estimation of time-varying parameters under shape restrictions. Journal of Econometrics, 126(1), 53-77.

Rempala, G. A., Wesolowski, J., 2016. Double symptotics for the chi-square statistic. *Statistics and Probability Letters* 119, 317-325.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In Statistical Analysis and Forecasting of Economic Structural Change (pp. 253-264). Springer, Berlin, Heidelberg.

Sengupta, A., 2017. Testing for a structural break in a spatial panel model. *Econometrics* 5, 1-17.

Su, L. J., Yang Z. L., 2015. QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185,230-258.

van der Vaart, A. W., 1998. *Asymptotic Statistics.* Cambridge University Press.

Wooldridge, J. M., 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.

Xu, Y. H., Yang, Z. L., 2020a. 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

Xu, Y. H., Yang, Z. L., 2020b. Adjusted Quasi-Score Estimation of Spatial Panel Data Models with Time Varying Coefficients. Working paper, Singapore Management University.

Yang, Z. L., 2010. A robust LM test for spatial error components. *Regional Science and Urban Economics* 40, 299-310.

Yang, Z. L., 2015a. A general method for third-order bias and variance corrections on a nonlinear estimator. *Journal of Econometrics* 186, 178-200.

Yang, Z. L., 2015b. LM tests of spatial dependence based on bootstrap critical values. *Journal of Econometrics* 185, 33-59.

Yang, Z. (2018). Unified M-estimation of fixed-effects spatial dynamic models with short panels. *Journal of econometrics* 205, 423-447.

Yang, Z. L., Li, C., Tse, Y. K., 2006. Functional form and spatial dependence in dynamic panels. *Economics Letters* 91, 138-145.

Yang, Z., Yu, J., Liu, S. F., 2016. Bias correction and refined inferences for fixed effects spatial panel data models. *Regional Science and Urban Economics* 61, 52-72.

A Appendix to Chapter 2

A.1 Some Basic Lemmas

Lemma A.1.1 (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 bounded. 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 bounded.

Lemma A.1.2 (Yang, 2015b, Lemma A.1, extended). *For $t = 1, 2$, let A_{nt} be $n \times n$ matrices and c_{nt} be $n \times 1$ vectors. Let ε_n be an $n \times 1$ random vector of iid elements with mean zero, variance σ^2, and finite 3rd and 4th cumulants μ_3 and μ_4. Let a_{nt} be the vector of diagonal elements of A_{nt}. Define $Q_{nt} = c_{nt}'\varepsilon_n + \varepsilon_n' A_{nt}\varepsilon_n$, $t = 1, 2$. Then, for $t, s = 1, 2$,*

$$\begin{aligned}&\mathrm{Cov}(Q_{nt}, Q_{ns}) \equiv f(A_{nt}, c_{nt}; A_{ns}, c_{ns})\\ =\ &\sigma^4 \mathrm{tr}[(A_{nt}' + A_{nt})A_{ns})] + \mu_3 a_{nt}' c_{ns} + \mu_3 c_{nt}' a_{ns} + \mu_4 a_{nt}' a_{ns} + \sigma^2 c_{nt}' c_{ns}.\end{aligned}$$

Lemma A.1.3 (CLT for Linear-Quadratic Forms, Kelejian and Prucha, 2001). *Let A_n, a_n, c_n and ε_n be as in Lemma A.2. Assume (i) A_n is bounded uniformly in row and column sums, (ii) $n^{-1}\sum_{i=1}^{n} |c_{n,i}^{2+\eta_1}| < \infty, \eta_1 > 0$, and (iii) $E|\varepsilon_{n,i}^{4+\eta_2}| < \infty$, $\eta_2 > 0$. Then,*

$$\frac{\varepsilon_n' A_n \varepsilon_n + c_n'\varepsilon_n - \sigma^2 \mathrm{tr}(A_n)}{\{\sigma^4 \mathrm{tr}(A_n' A_n + A_n^2) + \mu_4 a_n' a_n + \sigma^2 c_n' c_n + 2\mu_3 a_n' c_n\}^{\frac{1}{2}}} \xrightarrow{D} N(0, 1).$$

A.2 Hessian, Expected Hessian and VC Matrices

Notation. For $t, s = 1, \ldots, T$, `blkdiag`$\{A_t\}$ forms a block-diagonal matrix by placing A_t diagonally, $\{A_t\}$ forms a matrix by stacking A_t horizontally, and $\{B_{ts}\}$ forms a matrix by the component matrices B_{ts}. The expected negative Hessian $I_{\varpi}(\theta_0)$ and the VC matrix $\Sigma_{\varpi}(\theta_0)$ of the AQS function, ϖ=SL1, SL2, SLE1, SLE2, are both partitioned according to the slope parameters $\boldsymbol{\beta}$, the spatial lag parameters $\boldsymbol{\lambda}$, spatial error parameters $\boldsymbol{\rho}$ (if existing in the model), and the error variance σ^2, with the sub-matrices denoted by, e.g., $I_{\beta\beta}$, $I_{\beta\lambda}$, $\Sigma_{\beta\beta}$, $\Sigma_{\beta\lambda}$. Furthermore, $\mathtt{diag}(\cdot)$ forms a diagonal matrix and $\mathtt{diagv}(\cdot)$ a column vector, based on the diagonal elements of a square matrix.

Parametric quantities, e.g., $A_n(\lambda_{t0})$ and $B_n(\rho_{t0})$, evaluated at the true parameters are denoted as A_{nt} and B_{nt}. For a matrix A_n, denote $A_n^s = A_n + A_n'$. The bold **0** represents generically a vector or a matrix of zeros, to distinguish from the scalar 0.

Let $\mathbb{V}_N = (V_{n1}', \ldots, V_{nT}')'$ be the vector of original errors with elements $\{v_{it}\}$ being iid of mean 0, variance σ^2, skewness γ and excess kurtosis κ. We present here results sufficient for the implementation of the tests introduced in the paper. More details can be found in a `Supplementary Appendix` available at: `http://www.mysmu.edu/faculty/zlyang/`.

A.2.1. Panel SL model with one-way FE. The negative Hessian matrix $J_{\mathtt{SL1}}(\theta_0)$ is given in the `Supplementary Appendix`. Its expectation $I_{\mathtt{SL1}}(\theta_0)$ has the components:

$$I_{\beta\beta} = \mathtt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}X_{nt}'X_{nt}\big\} - \big\{\tfrac{1}{T\sigma_0^2}X_{nt}'X_{ns}\big\},\ I_{\lambda\beta} = \mathtt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta_{nt}'X_{nt}\big\} - \big\{\tfrac{1}{T\sigma_0^2}\eta_{nt}'X_{ns}\big\},$$
$$I_{\lambda\lambda} = \mathtt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta_{nt}'\eta_{nt} + \tfrac{T-1}{T}\mathrm{tr}(G_{nt}^sG_{nt})\big\} - \big\{\tfrac{1}{T\sigma_0^2}\eta_{nt}'\eta_{ns}\big\},\ \ I_{\sigma^2\lambda} = \big\{\tfrac{T-1}{T\sigma_0^2}\mathrm{tr}(G_{nt})\big\},$$
$$I_{\sigma^2\beta} = \mathbf{0},\ I_{\sigma^2\sigma^2} = \tfrac{n(T-1)}{2\sigma_0^4},\ \text{ where } \eta_{nt} = G_{nt}(X_{nt}\beta_{t0} + c_n) \text{ and } G_{nt}^s = G_{nt} + G_{nt}'.$$

The VC matrix $\Sigma_{\mathtt{SL1}}(\theta_0)$ is obtained by applying Lemma A.1.2 with ε replaced by $\mathbb{V}_N$, c_{nt} by Π_{1t} and Π_{2t}, and A_{nt} by Φ_t and Ψ:

$$\Sigma_{\mathtt{SL1}}(\theta_0) = \begin{pmatrix} \{f(\mathbf{0},\Pi_{1t};\mathbf{0},\Pi_{1s})\}, & \{f(\mathbf{0},\Pi_{1t};\Phi_s,\Pi_{2s})\}, & \{f(\mathbf{0},\Pi_{1t};\Psi,\mathbf{0})\} \\ \sim, & \{f(\Phi_t,\Pi_{2t};\Phi_s,\Pi_{2s})\}, & \{f(\Phi_t,\Pi_{2t};\Psi,\mathbf{0})\} \\ \sim, & \sim, & f(\Psi,\mathbf{0};\Psi,\mathbf{0}) \end{pmatrix},$$

where $\Pi_{1t} = \frac{1}{\sigma_0^2}Z_{Nt}^{\circ}X_{nt}$, $\Pi_{2t} = \frac{1}{\sigma_0^2}Z_{Nt}^{\circ}\eta_{nt}$, $\Phi_t = \frac{1}{\sigma_0^2}Z_{Nt}G_{nt}'Z_{Nt}^{\circ\prime}$, and $\Psi = \frac{1}{2\sigma^4}\sum_{t=1}^T Z_{Nt}^{\circ}Z_{Nt}^{\circ\prime}$; $Z_{Nt}^{\circ} = Z_{Nt} - \bar{Z}_N$, $Z_{Nt} = z_t \otimes I_n$, $\bar{Z}_N = \frac{1}{T}(l_T \otimes I_n)$, and z_t be a $T \times 1$ vector of element 1 in the tth position and 0 elsewhere.

A.2.2. Panel SL model with two-way FE. The negative Hessian matrix $J_{\mathtt{SL2}}(\theta_0)$

is given in the `Supplementary Appendix`. Its expectation $I_{\text{SL2}}(\theta_0)$ has the components:

$$\begin{aligned} I_{\beta\beta} &= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}X_{nt}^{*\prime}X_{nt}^*\big\} - \big\{\tfrac{1}{T\sigma_0^2}X_{nt}^{*\prime}X_{ns}^*\big\},\ I_{\lambda\beta} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta_{nt}^{*\prime}X_{nt}^*\big\} - \big\{\tfrac{1}{T\sigma_0^2}\eta_{nt}^{*\prime}X_{ns}^*\big\},\\ I_{\lambda\lambda} &= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta_{nt}^{*\prime}\eta_{nt}^* + \tfrac{T-1}{T}\text{tr}(G_{nt}^{*s}G_{nt}^*)\big\} - \big\{\tfrac{1}{T\sigma_0^2}\eta_{nt}^{*\prime}\eta_{ns}^*\big\},\ I_{\sigma^2\lambda} = \big\{\tfrac{T-1}{T\sigma_0^2}\text{tr}(G_{nt}^*)\big\}, \end{aligned}$$

$I_{\sigma^2\beta} = \mathbf{0}$, $I_{\sigma^2\sigma^2} = \frac{(n-1)(T-1)}{2\sigma_0^4}$, where $\eta_{nt}^* = G_{nt}^*(X_{nt}^*\beta_{t0} + c_n^*)$ and $G_{nt}^{*s} = G_{nt}^* + G_{nt}^{*\prime}$.

$\Sigma_{\text{SL2}}(\theta_0)$ has an identical form as $\Sigma_{\text{SL1}}(\theta_0)$ with the relevant quantities replaced by $\Pi_{1t} = \frac{1}{\sigma_0^2}Z_{Nt}^{\circ *}X_{nt}^*$, $\Pi_{2t} = \frac{1}{\sigma_0^2}Z_{Nt}^{\circ}F_{n,n-1}\eta_{nt}^*$, $\Phi_t = \frac{1}{\sigma_0^2}Z_{Nt}^{*}G_{nt}^{*\prime}Z_{Nt}^{\circ *\prime}$, and $\Psi = \frac{1}{2\sigma^4}\sum_{t=1}^T Z_{Nt}^{\circ *}Z_{Nt}^{\circ *\prime}$, where $Z_{Nt}^* = Z_{Nt}F_{n,n-1}$ and $Z_{Nt}^{\circ *} = Z_{Nt}^{\circ}F_{n,n-1}$.

A.2.3. Panel SLE model with one-way FE. The negative Hessian matrix $J_{\text{SEL1}}(\theta_0)$ is in the `Supplementary Appendix`. Its expectation $I_{\text{SEL1}}(\theta_0)$ has the components:

$$\begin{aligned} I_{\beta\beta} &= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}X_{nt}'D_{nt}X_{nt}\big\} - \big\{\tfrac{1}{\sigma_0^2}X_{nt}'D_{nt}\mathbb{D}_n^{-1}D_{ns}X_{ns}\big\};\\ I_{\lambda\beta} &= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta_t'D_{nt}X_{nt}\big\} - \big\{\tfrac{1}{\sigma_0^2}\eta_t'D_{nt}\mathbb{D}_n^{-1}D_{ns}X_{ns}\big\},\quad I_{\rho\beta} = 0_{Tk}\\ I_{\lambda\lambda} &= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta_{nt}'D_{nt}\eta_{nt} + \text{tr}(S_{nt}\bar{G}_{nt}^s\bar{G}_{nt})\big\} - \big\{\tfrac{1}{\sigma_0^2}\eta_{nt}'D_{nt}\mathbb{D}_n^{-1}D_{ns}\eta_{ns}\big\},\\ I_{\lambda\rho} &= \texttt{blkdiag}\big\{\text{tr}(\bar{G}_{nt}'S_{nt}H_{nt}^s)\big\};\quad I_{\sigma^2\sigma^2} = -\tfrac{n(T-1)}{2\sigma_0^4} + \tfrac{1}{\sigma_0^4}\textstyle\sum_{t=1}^T\text{tr}(S_{nt})\\ I_{\rho\lambda} &= \texttt{blkdiag}\big\{\text{tr}(\bar{G}_{nt}'S_{nt}H_{nt}^sS_{nt})\big\} - \big\{\text{tr}(G_{ns}'D_{ns}\mathbb{D}_n^{-1}\dot{D}_{nt}\mathbb{D}_n^{-1})\big\}\\ I_{\rho\rho} &= \texttt{blkdiag}\big\{\text{tr}(H_{nt}^sS_{nt}H_{nt} - B_{nt}\mathbb{D}_n^{-1}\dot{D}_{nt}B_{nt}^{-1}H_{nt})\big\} + \big\{\text{tr}(B_{nt}\mathbb{D}_n^{-1}\dot{D}_{ns}\mathbb{D}_n^{-1}B_{nt}'H_{nt})\big\}\\ I_{\sigma^2\beta} &= \mathbf{0},\quad I_{\sigma^2\lambda} = \big\{\tfrac{1}{\sigma_0^2}\text{tr}(R_{nt}G_{nt})\big\},\quad I_{\sigma^2\rho} = \tfrac{1}{\sigma_0^2}\text{tr}(S_{nt}H_{nt}). \end{aligned}$$

where $\dot{D}_{nt} = -\frac{d}{d\rho_{t0}}D_{nt} = M_n'B_{nt} + B_{nt}'M_n$, and $\bar{G}_{nt} = B_{nt}G_{nt}B_{nt}^{-1}$.

The VC matrix $\Sigma_{\text{SEL1}}(\theta_0)$ is obtained by applying Lemma A.1.2 with ε replaced by $\mathbb{V}_N$, c_{nt} by Π_{1t}, or Π_{2t}, and A_{nt} by Φ_{1t}, Φ_{2t}, or Ψ:

$$\Sigma_{\text{SLE1}}(\theta_0) =$$

$$\begin{pmatrix} \{f(\mathbf{0},\Pi_{1t};\mathbf{0},\Pi_{1s})\}, & \{f(\mathbf{0},\Pi_{1t};\Phi_{1s},\Pi_{2s})\}, & \{f(\mathbf{0},\Pi_{1t};\Phi_{2s},\mathbf{0})\}, & \{f(\mathbf{0},\Pi_{1t};\Psi,\mathbf{0})\}\\ \sim, & \{f(\Phi_{1t},\Pi_{2t};\Phi_{1s},\Pi_{2s})\}, & \{f(\Phi_{1t},\Pi_{2t};\Phi_{2s},\mathbf{0})\}, & \{f(\Phi_{1t},\Pi_{2t};\Psi,\mathbf{0})\}\\ \sim, & \sim, & \{f(\Phi_{2t},\mathbf{0};\Phi_{2s},\mathbf{0})\}, & \{f(\Phi_{2t},\mathbf{0};\Psi,\mathbf{0})\}\\ \sim, & \sim, & \sim, & f(\Psi,\mathbf{0};\Psi,\mathbf{0}) \end{pmatrix},$$

where $\Pi_{1t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\diamond} B_{nt} X_{nt}$, $\Pi_{2t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\diamond} B_{nt} \eta_{nt}$, $\Phi_{1t} = \frac{1}{\sigma_0^2} Z_{Nt} B_{nt}^{-1\prime} G_{nt}^{\prime} B_{nt}^{\prime} Z_{Nt}^{\diamond\prime}$, $\Phi_{2t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\diamond} H_{nt} Z_{Nt}^{\diamond\prime}$, $\Psi = \frac{1}{2\sigma_0^4} \sum_{t=1}^{T} Z_{Nt}^{\diamond} Z_{Nt}^{\diamond\prime}$, with $Z_{Nt}^{\diamond\prime} = [Z_{Nt}^{\prime} - B_{nt} \mathbb{D}_n^{-1} (l_T^{\prime} \otimes I_n) \mathbb{B}_N]$ and $\mathbb{B}_N = \texttt{blkdiag}(B_{n1}, \ldots, B_{nT})$.

A.2.4. Panel SLE model with two-way FE. The negative Hessian matrix $J_{\texttt{SEL2}}(\theta_0)$ is in the `Supplementary Appendix`. Its expectation $I_{\texttt{SEL2}}(\theta_0)$ has the components:

$$I_{\beta\beta} = \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} X_{nt}^{*\prime} D_{nt}^{*} X_{nt}^{*}\} - \{\tfrac{1}{\sigma_0^2} X{*\prime}_{nt} D_{nt}^{*} \mathbb{D}_n^{*-1} D_{ns}^{*} X_{ns}^{*}\};$$

$$I_{\lambda\beta} = \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} \eta_t^{*\prime} D_{nt}^{*} X_{nt}^{*}\} - \{\tfrac{1}{\sigma_0^2} \eta_t^{*\prime} D_{nt}^{*} \mathbb{D}_n^{*-1} D_{ns}^{*} X_{ns}^{*}\}; \quad I_{\rho\beta} = 0;$$

$$I_{\lambda\lambda} = \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} \eta_{nt}^{*\prime} D_{nt}^{*} \eta_{nt}^{*} + \text{tr}(S_{nt}^{*} \bar{G}_{nt}^{*s} \bar{G}_{nt}^{*})\} - \{\tfrac{1}{\sigma_0^2} \eta_{nt}^{*\prime} D_{nt}^{*} \mathbb{D}_n^{*-1} D_{ns}^{*} \eta_{ns}^{*}\};$$

$$I_{\lambda\rho} = \texttt{blkdiag}\{\text{tr}(\bar{G}_{nt}^{*\prime} S_{nt}^{*} H_{nt}^{*s})\}; \quad I_{\sigma^2\sigma^2} = -\tfrac{(n-1)(T-1)}{2\sigma_0^4} + \tfrac{1}{\sigma_0^4} \textstyle\sum_{t=1}^{T} \text{tr}(S_{nt}^{*});$$

$$I_{\rho\lambda} = \texttt{blkdiag}\{\text{tr}(\bar{G}_{nt}^{*\prime} S_{nt}^{*} H_{nt}^{*s} S_{nt}^{*})\} - \{\text{tr}(G_{ns}^{*\prime} D_{ns}^{*} \mathbb{D}_n^{*-1} \dot{D}_{nt}^{*} \mathbb{D}_n^{*-1})\};$$

$$I_{\rho\rho} = \texttt{blkdiag}\{\text{tr}(H_{nt}^{*s} S_{nt}^{*} H_{nt}^{*} - B_{nt}^{*} \mathbb{D}_n^{*-1} \dot{D}_{nt}^{*} B_{nt}^{*-1} H_{nt}^{*})\} + \{\text{tr}(B_{nt}^{*} \mathbb{D}_n^{*-1} \dot{D}_{ns}^{*} \mathbb{D}_n^{*-1} B_{nt}^{*\prime} H_{nt}^{*})\};$$

$$I_{\sigma^2\beta} = \mathbf{0}; \quad I_{\sigma^2\lambda} = \{\tfrac{1}{\sigma_0^2} \text{tr}(R_{nt}^{*} G_{nt}^{*})\}; \quad I_{\sigma^2\rho} = \{\tfrac{1}{\sigma_0^2} \text{tr}(S_{nt}^{*} H_{nt}^{*})\},$$

where $\dot{D}_{nt}^{*} = -\frac{d}{d\rho_{t0}} D_{nt}^{*} = M_n^{*\prime} B_{nt}^{*} + B_{nt}^{*\prime} M_n^{*}$, and $\bar{G}_{nt}^{*} = B_{nt}^{*} G_{nt}^{*} B_{nt}^{*-1}$.

The VC matrix $\Sigma_{\texttt{SLE2}}(\theta_0)$ takes an identical form as $\Sigma_{\texttt{SLE1}}(\theta_0)$, but with $\Pi_{1t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\diamond *} B_{nt}^{*} X_{nt}^{*}$, $\Pi_{2t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\diamond *} B_{nt}^{*} \eta_{nt}^{*}$, $\Phi_{1t} = \frac{1}{\sigma_0^2} Z_{Nt}^{*} B_{nt}^{*-1\prime} G_{nt}^{*\prime} B_{nt}^{*\prime} Z_{Nt}^{\diamond *\prime}$, $\Phi_{2t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\diamond *} H_{nt}^{*} Z_{Nt}^{\diamond *\prime}$, and $\Psi = \frac{1}{2\sigma_0^4} \sum_{t=1}^{T} Z_{Nt}^{\diamond *} Z_{Nt}^{\diamond *\prime}$, where $Z_{Nt}^{*} = Z_{Nt} F_{n,n-1}$ and $Z_{Nt}^{\diamond *} = Z_{Nt}^{\diamond} F_{n,n-1}$.

A.2.5. Panel SLE model with two-way FE and homogeneous ρ. The expected negative Hessian corresponding to the AQS function given in (2.40) has components:

$$\begin{aligned}
I_{\beta\beta} &= \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} X_{nt}^{*\prime} D_n^{*} X_{nt}^{*}\} - \{\tfrac{1}{T\sigma_0^2} X_{nt}^{*\prime} D_n^{*} X_{ns}^{*}\},, \\
I_{\lambda\beta} &= \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} \eta_{nt}^{*\prime} D_n^{*} X_{nt}^{*}\} - \{\tfrac{1}{T\sigma_0^2} \eta_{nt}^{*\prime} D_n^{*} X_{ns}^{*}\}, \quad I_{\rho\beta} = \mathbf{0} \\
I_{\lambda\lambda} &= \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} \eta_{nt}^{*\prime} D_n^{*} \eta_{nt}^{*} + \tfrac{T-1}{T} \text{tr}(\bar{G}_{nt}^{*s} \bar{G}_{nt}^{*})\} - \{\tfrac{1}{T\sigma_0^2} \eta_{nt}^{*\prime} D_n^{*} \eta_{ns}^{*}\}, \\
I_{\lambda\rho} &= \{\tfrac{T-1}{T} \text{tr}(\bar{G}_{nt}^{*\prime} H_n^{*s})\}, \quad I_{\rho\rho} = (T-1) \text{tr}(H_n^{*s} H_n^{*}), \\
I_{\sigma^2\beta} &= 0_{tk}^{\prime}, \quad I_{\sigma^2\lambda} = \{\tfrac{T-1}{T\sigma_0^2} \text{tr}(G_{nt}^{*})\}, \quad I_{\sigma^2\rho} = \tfrac{T-1}{\sigma_0^2} \text{tr}(H_n^{*}), \quad I_{\sigma^2\sigma^2} = \tfrac{n(T-1)}{2\sigma_0^4}.
\end{aligned}$$

The VC matrix of the AQS function given in (2.40) is obtained by applying Lemma A.1.2 with ε_n replaced by $\mathbb{V}_N$, c_{nt} by $\Pi_{1t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\circ *} B_n^{*} X_{nt}^{*}$, or $\Pi_{2t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\circ *} B_n^{*} \eta_{nt}^{*}$, and A_{nt} by $\Phi_{1t} = \frac{1}{\sigma_0^2} Z_{Nt}^{*} B_n^{*-1\prime} G_{nt}^{*\prime} B_n^{*\prime} Z_{Nt}^{\circ *\prime}$, or $\Phi_2 = \frac{1}{\sigma_0^2} \sum_{t=1}^{T} Z_{Nt}^{\circ *} H_n^{*} Z_{Nt}^{\circ *\prime}$, or $\Psi = \frac{1}{2\sigma_0^4} \sum_{t=1}^{T} Z_{Nt}^{\circ *} Z_{Nt}^{\circ *\prime}$, where $Z_{Nt}^{*} = Z_{Nt} F_{n,n-1}$, and $Z_{Nt}^{\circ *} = Z_{Nt}^{\circ} F_{n,n-1}$:

$$
\begin{pmatrix}
\{f(\mathbf{0},\Pi_{1t};\mathbf{0},\Pi_{1s})\}, & \{f(\mathbf{0},\Pi_{1t};\Phi_{1s},\Pi_{2s})\}, & \{f(\mathbf{0},\Pi_{1t};\Phi_2,\mathbf{0})\}, & \{f(\mathbf{0},\Pi_{1t};\Psi,\mathbf{0})\} \\
\sim, & \{f(\Phi_{1t},\Pi_{2t};\Phi_{1s},\Pi_{2s})\}, & \{f(\Phi_{1t},\Pi_{2t};\Phi_2,\mathbf{0})\}, & \{f(\Phi_{1t},\Pi_{2t};\Psi,\mathbf{0})\} \\
\sim, & \sim, & f(\Phi_2,\mathbf{0};\Phi_2,\mathbf{0}), & f(\Phi_2,\mathbf{0};\Psi,\mathbf{0}) \\
\sim, & \sim, & \sim, & f(\Psi,\mathbf{0};\Psi,\mathbf{0})
\end{pmatrix}.
$$

A.2.6. Panel SE model with two-way FE The expected negative Hessian matrix corresponding to the AQS function given in (2.41) has the components:

$$
\begin{aligned}
&I_{\beta\beta} = \mathtt{blkdiag}\{\tfrac{1}{\sigma_0^2}X_{nt}^{*\prime}D_{nt}^{*}X_{nt}^{*}\} - \{\tfrac{1}{\sigma_0^2}X_{nt}^{*\prime}D_{nt}^{*}\mathbb{D}_n^{*-1}D_{ns}^{*}X_{ns}^{*}\}, \quad I_{\rho\beta} = \mathbf{0}; \\
&I_{\rho\rho} = \mathtt{blkdiag}\{\mathrm{tr}(H_{nt}^{*s}S_{nt}^{*}H_{nt}^{*} - B_{nt}^{*}\mathbb{D}_n^{*-1}\dot{D}_{nt}B_{nt}^{*-1}H_{nt}^{*})\} + \{\mathrm{tr}(B_{nt}^{*}\mathbb{D}_n^{*-1}\dot{D}_{ns}\mathbb{D}_n^{*-1}B_{nt}^{*\prime}H_{nt}^{*})\}; \\
&I_{\sigma^2\beta} = \mathbf{0}; \quad I_{\sigma^2\rho} = \{\tfrac{1}{\sigma_0^2}\mathrm{tr}(S_{nt}^{*}H_{nt}^{*})\}; \quad I_{\sigma^2\sigma^2} = -\tfrac{(n-1)(T-1)}{2\sigma_0^4} + \tfrac{1}{\sigma_0^4}\textstyle\sum_{t=1}^{T}\mathrm{tr}(S_{nt}^{*}).
\end{aligned}
$$

Applying Lemma A.1.2 with ε_n being replaced by $\mathbb{V}_N$, c_{nt} by $\Pi_t = \frac{1}{\sigma_0^2}Z_{Nt}^{\diamond *}B_{nt}^{*}X_{nt}^{*}$, and A_{nt} by Φ_t or Ψ, we obtain the corresponding VC matrix of the AQS function (2.41):

$$
\begin{pmatrix}
\{f(\mathbf{0},\Pi_t;\mathbf{0},\Pi_s)\}, & \{f(\mathbf{0},\Pi_t;\Phi_s,\mathbf{0})\}, & \{f(\mathbf{0},\Pi_t;\Psi,\mathbf{0})\} \\
\sim, & \{f(\Phi_t,\mathbf{0};\Phi_s,\mathbf{0})\}, & \{f(\Phi_t,\mathbf{0};\Psi,\mathbf{0})\} \\
\sim, & \sim, & f(\Psi,\mathbf{0};\Psi,\mathbf{0})
\end{pmatrix},
$$

where $\Phi_t = \frac{1}{\sigma_0^2}Z_{Nt}^{\diamond *}H_{nt}^{*}Z_{Nt}^{\diamond *\prime}$, $\Psi = \frac{1}{2\sigma^4}\sum_{t=1}^{T}Z_{Nt}^{\diamond *}Z_{Nt}^{\diamond *\prime}$, and $Z_{Nt}^{\diamond *} = Z_{Nt}^{\diamond}F_{n,n-1}$.

A.3 Proof of the Theorems

The four theorems share some similar features. We provide here only the proof of the most general Theorem 2.4. The detailed proofs of all theorems can be found in the `Supplementary Appendix`, available at: `http://www.mysmu.edu/faculty/zlyang/`.

Proof of Theorem 2.4: Consider the AQS function $S_{\mathtt{SLE2}}^{\star}(\theta)$ given in (2.37). We need to show that $\frac{1}{\sqrt{N_0}}S_{\mathtt{SLE2}}^{\star}(\theta_0) \xrightarrow{D} N\big(0,\ \lim_{N_0\to\infty}\frac{1}{N_0}\Sigma_{\mathtt{SLE2}}(\theta_0)\big)$, as $N_0 \to \infty$. We have

$$
\begin{aligned}
&\widetilde{V}_{nt}^{*} \equiv \widetilde{V}_{nt}^{*}(\beta_0,\lambda_0,\rho_0) = V_{nt}^{*} - B_{nt}^{*}\mathbb{D}_n^{*-1}\textstyle\sum_{s=1}^{T}B_{ns}^{*\prime}V_{ns}^{*} = F_{n,n-1}'Z_{Nt}^{\diamond\prime}\mathbb{V}_N, \text{ and} \\
&W_n^{*}Y_{nt}^{*} = G_{nt}^{*}(X_{nt}^{*}\beta_{t0} + c_n^{*} + B_{nt}^{-*1}V_{nt}^{*}) = \eta_{nt}^{*} + G_{nt}^{*}B_{nt}^{*-1}F_{n,n-1}'Z_{Nt}'\mathbb{V}_N.
\end{aligned}
$$

Hence, the AQS function at true θ_0 can be written as

$$S^*_{\mathtt{SLE2}}(\theta_0) = \begin{cases} \Pi'_{1t}\mathbb{V}_N, \quad t = 1, \ldots, T, \\ \Pi'_{2t}\mathbb{V}_N + \mathbb{V}'_N\Phi_{1t}\mathbb{V}_N - \mathrm{tr}(R^*_{nt}G^*_{nt}), \quad t = 1, \ldots, T, \\ \mathbb{V}'_N\Phi_{2t}\mathbb{V}_N - \mathrm{tr}(S^*_{nt}H^*_{nt}), \quad t = 1, \ldots, T, \\ \mathbb{V}'_N\Psi\mathbb{V}_N - \frac{(n-1)(T-1)}{2\sigma^2}, \end{cases} \tag{A.3.1}$$

where $\Pi_{1t} = \frac{1}{\sigma_0^2}Z^{\diamond *}_{Nt}B^*_{nt}X^*_{nt}$, $\Pi_{2t} = \frac{1}{\sigma_0^2}Z^{\diamond *}_{Nt}B^*_{nt}\eta^*_{nt}$, $\Phi_{1t} = \frac{1}{\sigma_0^2}Z^*_{Nt}B^{*-1\prime}_{nt}G^{*\prime}_{nt}B^{*\prime}_{nt}Z^{\diamond *\prime}_{Nt}$, $\Phi_{2t} = \frac{1}{\sigma_0^2}Z^{\diamond *}_{Nt}H^*_{nt}Z^{\diamond *\prime}_{Nt}$, and $\Psi = \frac{1}{2\sigma_0^4}\sum_{t=1}^T Z^{\diamond *}_{Nt}Z^{\diamond *\prime}_{Nt}$, with $Z^*_{Nt} = Z_{Nt}F_{n,n-1}$ and $Z^{\diamond *}_{Nt} = Z^{\diamond}_{Nt}F_{n,n-1}$; $Z_{Nt} = z_t \otimes I_n$ and z_t is a $T \times 1$ vector with tth element being 1 and other elements being zero; and $Z^{\diamond\prime}_{Nt} = [Z'_{Nt} - B_{nt}\mathbb{D}_n^{-1}(l'_T \otimes I_n)\mathbb{B}_N]$ and $\mathbb{B}_N = \mathtt{blkdiag}(B_{n1}, \ldots, B_{nT})$.

First, as the elements of X_{nt} are non-stochastic and uniformly bounded (by Assumption 3), the row and column sums of B^*_{nt} are uniformly bounded in absolute values by Assumption 5 and Lemma A.1.1. It follows that the elements of Π_{1t} are uniformly bounded. By Assumption 4 and Lemma A.1.1(i), G_{nt} is uniformly bounded in both row and column sums. By Lemma A.2 of Lee and Yu (2010),

$$(I_n - \lambda F'_{n,n-1}W_nF_{n,n-1})^{-1} = F'_{n,n-1}(I_n - \lambda W_n)^{-1}F_{n,n-1}. \tag{A.3.2}$$

We have $A^{*-1}_{nt} = F'_{n,n-1}A^{-1}_{nt}F_{n,n-1}$. Thus, G^*_{nt} is uniformly bounded in both row and column sums by Lemma A.1.1(iii), and the elements of $\eta^*_{nt} = G^*_{nt}(X^*_{nt}\beta_{t0} + c^*_n)$ are uniformly bounded by Assumption 3. It follows that the elements of Π_{2t} are uniformly bounded. Similarly, $B^{*-1}_{nt} = F'_{n,n-1}B^{-1}_{nt}F_{n,n-1}$, and therefore the elements of H^*_{nt} is uniformly bounded in both row and column sums. With these and the definitions of Z_{Nt} and $Z^{\diamond}_{Nt}$, it is easy to show that Φ_{1t}, Φ_{2t} and Ψ are uniformly bounded in both row and column sums. Thus, under Assumptions 1-5, the central limit theorem (CLT) of linear-quadratic (LQ) form of Kelejian and Prucha (2001) or its simplified version (under iid errors) given in Lemma A.1.3 can be applied to each element of $S^{\star}_{\mathtt{SLE2}}(\theta_0)$ to establish its asymptotic normality. Then, an application of Cramér-Wold device under a finite T gives, $\frac{1}{\sqrt{N_0}}S^{\star}_{\mathtt{SLE2}}(\theta_0) \xrightarrow{D} N\big(0,\ \lim_{N_0\to\infty}\frac{1}{N_0}\Sigma_{\mathtt{SLE2}}(\theta_0)\big)$, as $N_0 \to \infty$. Then, by (2.11) and (2.12),

$$C[\tfrac{1}{N_0}I_{\mathtt{SLE2}}(\theta_0)]^{-1}\tfrac{1}{\sqrt{N_0}}S^{\star}_{\mathtt{SLE2}}(\tilde{\theta}_{\mathtt{SLE2}}) \xrightarrow{D} N\big(0,\ \lim_{N_0\to\infty}\Xi_{\mathtt{SLE2}}(\theta_0)\big).$$

It left to show that, as $N_0 \to \infty$,

(a) $\frac{1}{N_0}[I_{\texttt{SLE2}}(\tilde{\theta}_{\texttt{SLE2}}) - I_{\texttt{SLE2}}(\theta_0)] \xrightarrow{p} \mathbf{0}$,

(b) $\frac{1}{N_0}[\Sigma_{\texttt{SLE2}}(\tilde{\theta}_{\texttt{SLE2}}) - \Sigma_{\texttt{SLE2}}(\theta_0)] \xrightarrow{p} \mathbf{0}$.

Under the $\sqrt{N_0}$-consistency of $\tilde{\theta}_{\texttt{SLE2}}$ and with the analytical expressions of $I_{\texttt{SLE2}}(\theta_0)$ and $\Sigma_{\texttt{SLE2}}(\theta_0)$ given in Appendix A.2.4, the proofs of these results are repeated applications of the mean value theorem (MVT) to each component of $\frac{1}{N_0}[I_{\texttt{SLE2}}(\tilde{\theta}_{\texttt{SLE2}}) - I_{\texttt{SLE2}}(\theta_0)]$ and each component of $\frac{1}{N_0}[\Sigma_{\texttt{SLE2}}(\tilde{\theta}_{\texttt{SLE2}}) - \Sigma_{\texttt{SLE2}}(\theta_0)]$.

To show (a), we pick a typical element of $I_{\texttt{SLE2}}(\theta_0)$ given in Appendix A.2.4,

$$I_{\lambda\lambda} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta_{nt}^{*\prime}D_{nt}^{*}\eta_{nt}^{*} + \mathrm{tr}(S_{nt}^{*}\bar{G}_{nt}^{*s}\bar{G}_{nt}^{*})\big\} - \big\{\tfrac{1}{\sigma_0^2}\eta_{nt}^{*\prime}D_{nt}^{*}\mathbb{D}_n^{*-1}D_{ns}^{*}\eta_{ns}^{*}\big\}$$

to show that $\frac{1}{N_0}(\widetilde{I}_{\lambda\lambda} - I_{\lambda\lambda}) \xrightarrow{p} 0$. The proofs for the other components follow similarly. Recall: $\eta_{nt}^{*} = G_{nt}^{*}(X_{nt}^{*}\beta_{t0} + c_n^{*})$, $\mathbb{D}_n^{*}(\rho) = \sum_{t=1}^{T} D_n^{*}(\rho_t)$, $D_n^{*}(\rho_t) = B_n^{*\prime}(\rho_t)B_n^{*}(\rho_t)$, $B_n^{*}(\rho_t) = I_{n-1} - \rho_t M_n^{*}$, $S_{nt}^{*}(\rho) = I_{n-1} - B_{nt}^{*}(\rho_t)\mathbb{D}_n^{*-1}(\rho)B_{nt}^{*\prime}(\rho_t)$, and $\bar{G}_{nt}^{*} = B_{nt}^{*}G_{nt}^{*}B_{nt}^{*-1}$.

By Assumptions 4 and 5 and Lemma A.1.1(i), it is straightforward to show the two matrices, $D_n^{*}(\rho_t)$ and $\bar{G}_{nt}^{*}(\lambda_t, \rho_t)$, are uniformly bounded in both row and column sums in a neighborhood of $(\lambda_{t0}, \rho_{t0})$ for each t, and so are their derivatives. Clearly with the properties of $D_n^{*}(\rho_t)$ and a finite T, $\mathbb{D}_n^{*}(\rho)$ is uniformly bounded in both row and column sums in a neighborhood of ρ_0, and so are its derivatives.

By Assumption 5 and Lemma A.1.1(i), $D_n^{*-1}(\rho_t)$ is uniformly bounded in both row and column sums in a neighborhood of ρ_{t0} for each t, and so are its derivatives. By a matrix result that for two invertible matrices A_n and B_n, $(A_n + B_n)^{-1} = A_n^{-1} + \frac{1}{1+c}A_n^{-1}B_nA_n^{-1}$, where $c = \mathrm{tr}(B_nA_n^{-1})$, we infer that for a finite T, $\mathbb{D}_n^{*}(\rho)$ is uniformly bounded in both row and column sums in a neighborhood of ρ_0, and so are its derivatives. It follows that $S_{nt}^{*}(\rho)$ is uniformly bounded in both row and column sums in a neighborhood of ρ_0, and so are its derivatives. Noting that $\tilde{I}_{\lambda\lambda} = I_{\lambda\lambda}(\tilde{\theta}_{\texttt{SLE2}})$ and $I_{\lambda\lambda} = I_{\lambda\lambda}(\theta_0)$, we have by MVT, for each component of $I_{\lambda\lambda}(\theta)$ denoted as $I_{\lambda\lambda,ts}(\theta)$, $t, s = 1, \ldots, T$,

$$\tfrac{1}{N_0}I_{\lambda\lambda,ts}(\tilde{\theta}_{\texttt{SLE2}}) = \tfrac{1}{N_0}I_{\lambda\lambda,ts}(\theta_0) + [\tfrac{1}{N_0}\tfrac{\partial}{\partial\theta'}I_{\lambda\lambda,ts}(\bar{\theta})](\tilde{\theta}_{\texttt{SLE2}} - \theta_0),$$

where $\bar{\theta}$ lies elementwise between $\tilde{\theta}_{\texttt{SLE2}}$ and θ_0, with $\bar{\theta}$ being $\sqrt{N_0}$-consistent as $\tilde{\theta}_{\texttt{SLE2}}$ is. With the above argument and Lemma A.1(ii), we have $\frac{1}{N_0}\frac{\partial}{\partial\theta'}I_{\lambda\lambda,ts}(\bar{\theta}) = O_p(1)$.

Therefore, $\frac{1}{N_0}[I_{\boldsymbol{\lambda}\boldsymbol{\lambda},ts}(\tilde{\boldsymbol{\theta}}_{\texttt{SLE2}}) - I_{\boldsymbol{\lambda}\boldsymbol{\lambda},ts}(\boldsymbol{\theta}_0)] = o_p(1)$ for each (t,s), and $\frac{1}{N_0}[I_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\tilde{\boldsymbol{\theta}}_{\texttt{SLE2}}) - I_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\boldsymbol{\theta}_0)] = o_p(1)$. Note that the easily proved results such as $\frac{1}{N_0}(\tilde{c}_n\tilde{G}_{nt}\tilde{c}_n - c_nG_{nt}c_n) \xrightarrow{p} 0$, has been used. The proofs of the other components of $\frac{1}{N_0}[I_{\texttt{SLE2}}(\tilde{\boldsymbol{\theta}}_{\texttt{SLE2}}) - I_{\texttt{SLE2}}(\boldsymbol{\theta}_0)] \xrightarrow{p} \mathbf{0}$ proceeds similarly.

To show (b), we again choose the most complicated term, $f(\Phi_{1t}, \Pi_{2t}; \Phi_{1s}, \Pi_{2s})$ that corresponds to $\boldsymbol{\lambda}$, to show in details where the quantities involved are given at the end of Appendix A.2.4: $\Pi_{1t} = \frac{1}{\sigma_0^2}Z_{Nt}^{\diamond*}B_{nt}^*X_{nt}^*$, $\Pi_{2t} = \frac{1}{\sigma_0^2}Z_{Nt}^{\diamond*}B_{nt}^*\eta_{nt}^*$, $\Phi_{1t} = \frac{1}{\sigma_0^2}Z_{Nt}^*B_{nt}^{*-1\prime}G_{nt}^{*\prime}B_{nt}^{*\prime}Z_{Nt}^{\diamond*\prime}$, and $\Phi_{2t} = \frac{1}{\sigma_0^2}Z_{Nt}^{\diamond*}H_{nt}^*Z_{Nt}^{\diamond*\prime}$, where $Z_{Nt}^* = Z_{Nt}F_{n,n-1}$ and $Z_{Nt}^{\diamond*} = Z_{Nt}^{\diamond}F_{n,n-1}$.

Applying Lemma A.1.2 with A_{nt} replaced by Φ_{1t}, a_{nt} by $\phi_{1t} = \texttt{diagv}(\Phi_{1t})$, and c_{nt} by Π_{2t} (similarly for the quantities with subscript s), and noting that $\mu_3 = \gamma$ and $\mu_4 = \kappa$, we obtain the covariance between the λ_t- and λ_s-components of the AQS function:

$$f(\Phi_{1t}, \Pi_{2t}; \Phi_{1s}, \Pi_{2s}) = \sigma_0^4\mathrm{tr}[(\Phi_{1t}' + \Phi_{1t})\Phi_{1s}] + \gamma\phi_{1t}'\Pi_{2s} + \gamma\Pi_{2t}'\phi_{1s} + \kappa\phi_{1t}'\phi_{1s} + \sigma_0^2\Pi_{2t}'\Pi_{2s}.$$

Applying MVT and following the similar arguments as in (a), the convergence of the relevant terms can easily be proved, e.g., $\frac{1}{N_0}\{\mathrm{tr}[(\widetilde{\Phi}_{1t}' + \widetilde{\Phi}_{1t})\widetilde{\Phi}_{1s}] - \mathrm{tr}[(\Phi_{1t}' + \Phi_{1t})\Phi_{1s}]\} = o_p(1)$, $\frac{1}{N_0}[\phi_{1t}'\Pi_{2s} - \phi_{1t}'\Pi_{2s}] = o_p(1)$, etc. Furthermore, $\tilde{\sigma}_{\texttt{SLE2}}^2 - \sigma_0^2 = o_p(1)$, and hence $\tilde{\sigma}_{\texttt{SLE2}}^4 - \sigma_0^4 = o_p(1)$; for the estimates obtained from Lemma 4.1(a) of Yang et al. (2016), it is easy to show that $\tilde{\gamma} - \gamma \xrightarrow{p} 0$ and $\tilde{\kappa} - \kappa \xrightarrow{p} 0$. It follows that

$$[\tilde{f}(\widetilde{\Phi}_{1t}, \widetilde{\Pi}_{2t}; \widetilde{\Phi}_{1s}, \widetilde{\Pi}_{2s}) - f(\Phi_{1t}, \Pi_{2t}; \Phi_{1s}, \Pi_{2s})] = o_p(1).$$

Similarly, the convergence of the other elements of $\frac{1}{N_0}[\Sigma_{\texttt{SLE2}}(\tilde{\boldsymbol{\theta}}_{\texttt{SLE2}}) - \Sigma_{\texttt{SLE2}}(\boldsymbol{\theta}_0)]$ is proved.

B Appendix to Chapter 3

B.1 Some Basic Lemmas

The following lemmas are essential for the derivations and proofs of theoretical results, given in the subsequent appendices.

Lemma B.1.1 *(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 bounded. 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 bounded.*

Lemma B.1.2 *(Lee, 2004, p.1918): For W_n and A_{nt} defined in Model (3.1), if $\|W_n\|$ and $\|A_{nt0}^{-1}\|$ are uniformly bounded, where $\|\cdot\|$ is a matrix norm, then $\|A_{nt}^{-1}\|$ is uniformly bounded in a neighborhood of λ_{t0}.*

Lemma B.1.3 *(Lee, 2004, p.1918): Let X_n be an $n \times p$ matrix. If the elements X_n are uniformly bounded and $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular, then $P_n = X_n(X_n'X_n)^{-1}X_n'$ and $M_n = I_n - P_n$ are uniformly bounded in both row and column sums.*

Lemma B.1.4 *(Lee and Yu, 2010): For $W_n^* = F_{n,n-1}' W_n F_{n,n-1}$, when W_n is row normalized, $|I_{n-1} - \lambda_t W_n^*| = \frac{1}{1-\lambda_t}|I_n - \lambda_t W_n|$ and $(I_{n-1} - \lambda_t W_n^*)^{-1} = F_{n,n-1}'(I_n - \lambda_t W_n)^{-1} F_{n,n-1}$.*

Lemma B.1.5 *(Lemma B.4, Yang, 2015a, extended): Let $\{A_n\}$ be a sequence of $n \times n$ matrices that are uniformly bounded in either row or column sums. Suppose that the elements $a_{n,ij}$ of A_n are $O(h_n^{-1})$ uniformly in all i and j. 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 constant n-vector of elements of uniform order $O(h_n^{-1/2})$. Then*

$(i)\ \ \mathrm{E}(v_n' A_n v_n) = O(\frac{n}{h_n})$, $\quad$ $(ii)\ \mathrm{Var}(v_n' A_n v_n) = O(\frac{n}{h_n})$,

$(iii)\ \mathrm{Var}(v_n' A_n v_n + b_n' v_n) = O(\frac{n}{h_n})$, $\quad$ *(iv)* $v_n' A_n v_n = O_p(\frac{n}{h_n})$,

(v) $\ v_n' A_n v_n - \mathrm{E}(v_n' A_n v_n) = O_p((\frac{n}{h_n})^{\frac{1}{2}})$, $\quad$ *(vi)* $v_n' A_n b_n = O_p((\frac{n}{h_n})^{\frac{1}{2}})$,

the results (iii) and (vi) remain valid if b_n is a random n-vector independent of v_n such that $\{\mathrm{E}(b_{ni}^2)\}$ are of uniform order $O(h_n^{-1})$.

Lemma B.1.6 *(Yang, 2015b, Lemma A.1, extended). For $t = 1, 2$, let A_{nt} be $n \times n$ matrices and c_{nt} be $n \times 1$ vectors. Let ε_n be an $n \times 1$ random vector of iid elements with mean zero, variance σ^2, and finite 3rd and 4th cumulants $\mu_{(3)}$ and $\mu_{(4)}$. Let a_{nt} be the vector of diagonal elements of A_{nt}. Define $Q_{nt} = c_{nt}'\varepsilon_n + \varepsilon_n' A_{nt}\varepsilon_n$, $t = 1, 2$. Then, for $t, s = 1, 2$,*

$$\mathrm{Cov}(Q_{nt}, Q_{ns}) \equiv f(A_{nt}, c_{nt}; A_{ns}, c_{ns})$$

$$= \sigma^4 \mathrm{tr}[(A_{nt}' + A_{nt})A_{ns})] + \mu_3 a_{nt}' c_{ns} + \mu_3 c_{nt}' a_{ns} + \mu_4 a_{nt}' a_{ns} + \sigma^2 c_{nt}' c_{ns}. \tag{B.1.1}$$

Various useful special cases of (B.1.1) *are as follows:*

$(i)\ \ \mathrm{Cov}(c_{n1}'\varepsilon_n, Q_{n2}) = f(\mathbf{0}, c_{n1}; A_{n2}, c_{n2}) = \mu_3 c_{n1}' a_{n2} + \sigma^2 c_{n1}' c_{n2}$,

where c_{n1} can be an $n \times k$ matrix with $k \geqslant 1$;

$(ii)\ \ \mathrm{Var}(Q_{n1}) = f(A_{n1}, c_{n1}; A_{n1}, c_{n1}) = \sigma^4 \mathrm{tr}[(A_{n1}' + A_{n1})A_{n1})] + 2\mu_3 a_{n1}' c_{n1}$
$+ \mu_4 a_{n1}' a_{n1} + \sigma^2 c_{n1}' c_{n1}$;

$(iii)\ \mathrm{Var}(\varepsilon_n' A_{n1} \varepsilon_n) = f(A_{n1}, \mathbf{0}; A_{n1}, \mathbf{0}) = \sigma^4 \mathrm{tr}[(A_{n1}' + A_{n1})A_{n1})] + \mu_4 a_{n1}' a_{n1}$.

Lemma B.1.7 *(CLT for Linear-Quadratic Forms, Kelejian and Prucha, 2001). Let A_n, a_n, c_n and ε_n be as in Lemma A.6. Assume (i) A_n is bounded uniformly in row and column sums, (ii) $n^{-1}\sum_{i=1}^n |c_{n,i}^{2+\eta_1}| < \infty, \eta_1 > 0$, and (iii) $E|\varepsilon_{n,i}^{4+\eta_2}| < \infty$, $\eta_2 > 0$. Then,*

$$\frac{\varepsilon_n' A_n \varepsilon_n + c_n'\varepsilon_n - \sigma^2 \mathrm{tr}(A_n)}{\{\sigma^4 \mathrm{tr}(A_n' A_n + A_n^2) + \mu_4 a_n' a_n + \sigma^2 c_n' c_n + 2\mu_3 a_n' c_n\}^{\frac{1}{2}}} \xrightarrow{D} N(0, 1).$$

B.2 Hessian, Expected Hessian and VC Matrices

Notation. For $t, s = 1, \ldots, T$, $\texttt{blkdiag}\{A_t\}$ forms a block-diagonal matrix by placing A_t diagonally, $\{A_t\}$ forms a matrix by stacking A_t horizontally, and $\{B_{ts}\}$ forms a matrix by the component matrices B_{ts}. The expected negative Hessian $I_{\varpi}(\theta_0)$ and the VC matrix $\Sigma_{\varpi}(\theta_0)$ of the AQS function, ϖ=SL1, SL2, SLE1, SLE2, are both partitioned according to the slope parameters $\boldsymbol{\beta}$, the spatial lag parameters $\boldsymbol{\lambda}$, spatial error parameters ρ (if existing in the model), and the error variance σ^2, with the sub-matrices denoted by, e.g., $I_{\beta\beta}$, $I_{\beta\lambda}$, $\Sigma_{\beta\beta}$, $\Sigma_{\beta\lambda}$. Furthermore, $\texttt{diag}(\cdot)$ forms a diagonal matrix and $\texttt{diagv}(\cdot)$ a column vector, based on the diagonal elements of a square matrix.

Parametric quantities, e.g., $A_n(\lambda_{t0})$ and $B_n(\rho_0)$, evaluated at the true parameters are denoted as A_{nt} and B_n. For a matrix A_n, denote $A_n^s = A_n + A_n'$. The bold **0** represents generically a vector or a matrix of zeros, to distinguish from the scalar 0.

Let $\mathbb{V}_N = (V_{n1}', \ldots, V_{nT}')'$ be the vector of original errors with elements $\{v_{it}\}$ being iid of mean 0, variance σ^2, skewness γ and excess kurtosis κ. We present here results sufficient for the implementation of the estimation introduced in the paper. To estimate the VC matrices, we follow the method proposed by Xu and Yang (2020a).

B.2.1 Panel SLE model with two-way FE. The negative Hessian matrix $J_{\texttt{SLE2}}(\theta_0)$ has the components:

$$
\begin{aligned}
&J_{\boldsymbol{\beta\beta}} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2} X_{nt}^{*\prime} D_n^* X_{nt}^*\big\} - \big\{\tfrac{1}{T\sigma_0^2} X_{nt}^{*\prime} D_n^* X_{ns}^*\big\},\\
&J_{\boldsymbol{\beta\lambda}} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2} X_{nt}^{*\prime} D_n^* W_n^* Y_{nt}^*\big\} - \big\{\tfrac{1}{T\sigma_0^2} X_{nt}^{*\prime} D_n^* W_n^* Y_{ns}^*\big\},\\
&J_{\boldsymbol{\beta\rho}} = \big\{\tfrac{1}{\sigma_0^2} X_{nt}^{*\prime} B_n^{*\prime} H_n^{*\prime} \widetilde{V}_{nt}^* + \tfrac{1}{\sigma_0^2} X_{nt}^{*\prime} B_n^{*\prime} H_n^* \widetilde{V}_{nt}^*\big\}\\
&J_{\boldsymbol{\lambda\lambda}} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}(W_n^* Y_{nt}^*)' D_n^* (W_n^* Y_{nt}^*) + \tfrac{T-1}{T}\mathrm{tr}(G_{nt}^{*2})\big\} - \big\{\tfrac{1}{T\sigma_0^2}(W_n^* Y_{nt}^*)' D_n^* (W_n^* Y_{ns}^*)\big\},\\
&J_{\boldsymbol{\lambda\rho}} = \big\{\tfrac{1}{\sigma_0^2}(W_n^* Y_{nt}^*)' B_n^{*\prime} H_n^{*\prime} \widetilde{V}_{nt}^* + \tfrac{1}{\sigma_0^2}(W_n^* Y_{nt}^*)' B_n^{*\prime} H_n^* \widetilde{V}_{nt}^*\big\}\\
&J_{\boldsymbol{\rho\rho}} = \big\{\tfrac{1}{\sigma^2}\textstyle\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime} H_n^{*\prime} H_n^* \widetilde{V}_{nt}^* + (T-1)\mathrm{tr}(H_n^{*2})\big\}\\
&J_{\sigma^2\boldsymbol{\beta}} = \big\{\tfrac{1}{\sigma_0^4} X_{nt}^{*\prime} B_n^{*\prime} \widetilde{V}_{nt}^*\big\},\ J_{\sigma^2\boldsymbol{\lambda}} = \big\{\tfrac{1}{\sigma_0^4}(W_n^* Y_{nt}^*)' B_n^{*\prime} \widetilde{V}_{nt}^*\big\},\\
&J_{\sigma^2\boldsymbol{\rho}} = \big\{\tfrac{1}{\sigma_0^4}\textstyle\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime} H_n^* \widetilde{V}_{nt}^*\big\},\ J_{\sigma^2\sigma^2} = -\tfrac{n(T-1)}{2\sigma^4} + \tfrac{1}{\sigma^6}\textstyle\sum_{t=1}^T \widetilde{V}_{nt}^{*\prime} \widetilde{V}_{nt}^*.
\end{aligned}
$$

Its expectation, $I_{\mathtt{SLE2}}(\theta_0)$, has the components:

$$
\begin{aligned}
&I_{\beta\beta} \;= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}X_{nt}^{*\prime}D_n^*X_{nt}^*\big\} - \big\{\tfrac{1}{T\sigma_0^2}X_{nt}^{*\prime}D_n^*X_{ns}^*\big\};\\
&I_{\beta\lambda} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}X_{nt}^{*\prime}D_n^*\eta_{nt}^*\big\} - \big\{\tfrac{1}{T\sigma_0^2}X_{nt}^{*\prime}D_n^*\eta_{ns}^*\big\}, I_{\rho\rho} \;= \big\{(T-1)\mathrm{tr}[H_n^{*s}H_n^*]\big\}\\
&I_{\lambda\lambda} \;= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta_{nt}^{*\prime}D_n^*\eta_{nt}^* + \tfrac{T-1}{T}\mathrm{tr}[\bar{G}_{nt}^{*s}\bar{G}_{nt}^*]\big\} - \big\{\tfrac{1}{T\sigma_0^2}\eta_{nt}^{*\prime}D_n^*\eta_{ns}^*\big\}, \quad I_{\rho\beta} = 0_{Tk}, \quad I_{\sigma^2\beta} = \mathbf{0},\\
&I_{\lambda\rho} \;= \big\{\tfrac{T-1}{T}\mathrm{tr}[\bar{G}_{nt}^{*\prime}H_n^{*s}]\big\}, \quad I_{\sigma^2\lambda} = \big\{\tfrac{T-1}{T\sigma_0^2}\mathrm{tr}[G_{nt}^*]\big\}, \; I_{\sigma^2\rho} = \tfrac{T-1}{\sigma_0^2}\mathrm{tr}[H_n^*], \quad I_{\sigma^2\sigma^2} = \tfrac{n(T-1)}{2\sigma_0^4}
\end{aligned}
$$

where $\eta_{nt}^* = G_{nt}^*(X_{nt}^*\beta_{t0} + c_n^*)$ and $\bar{G}_{nt}^* = B_n^*G_{nt}^*B_n^{*-1}$.

The VC matrix $\Sigma_{\mathtt{SLE2}}(\theta_0)$ is obtained by applying Lemma B.1.6 with ε replaced by $\mathbb{V}_N$, c_{nt} by Π_{1t} or Π_{2t}, and A_{nt} by Φ_{1t}, Φ_2 or Ψ:

$$
\Sigma_{\mathtt{SLE2}}(\theta_0) = \begin{pmatrix}
\{f(\mathbf{0},\Pi_{1t};\mathbf{0},\Pi_{1s})\}, & \{f(\mathbf{0},\Pi_{1t};\Phi_{1s},\Pi_{2s})\}, & \{f(\mathbf{0},\Pi_{1t};\Phi_2,\mathbf{0})\}, & \{f(\mathbf{0},\Pi_{1t};\Psi,\mathbf{0})\}\\
\sim, & \{f(\Phi_{1t},\Pi_{2t};\Phi_{1s},\Pi_{2s})\}, & \{f(\Phi_{1t},\Pi_{2t};\Phi_2,\mathbf{0})\}, & \{f(\Phi_{1t},\Pi_{2t};\Psi,\mathbf{0})\}\\
\sim, & \sim, & \{f(\Phi_2,\mathbf{0};\Phi_2,\mathbf{0})\}, & \{f(\Phi_2,\mathbf{0};\Psi,\mathbf{0})\}\\
\sim, & \sim, & \sim, & \{f(\Psi,\mathbf{0};\Psi,\mathbf{0})\}
\end{pmatrix}.
$$

where Π_{1t}, Π_{2t}, Φ_{1t}, Φ_2 and Ψ are already defined in (3.28).

B.2.2. Panel SL model with two-way FE. The negative Hessian matrix $J_{\mathtt{SL2}}(\theta_0)$ takes the following form:

$$
\begin{aligned}
&J_{\beta\beta} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}X_{nt}^{*\prime}X_{nt}^*\big\} - \big\{\tfrac{1}{T\sigma_0^2}X_{nt}^{*\prime}X_{ns}^*\big\},\\
&J_{\lambda\beta} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}(W_n^*Y_{nt}^*)'X_{nt}^*\big\} - \big\{\tfrac{1}{T\sigma_0^2}(W_n^*Y_{nt}^*)'X_{ns}^*\big\},\\
&J_{\lambda\lambda} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}(W_n^*Y_{nt}^*)'(W_n^*Y_{nt}^*) + \tfrac{T-1}{T}\mathrm{tr}(G_{nt}^{*2})\big\} - \big\{\tfrac{1}{T\sigma_0^2}(W_n^*Y_{nt}^*)'(W_n^*Y_{ns}^*)\big\},\\
&J_{\sigma^2\beta} = \big\{\tfrac{1}{\sigma_0^4}\tilde{V}_{nt}^{*\prime}X_{nt}^*\big\}, \; J_{\sigma^2\lambda} \;= \big\{\tfrac{1}{\sigma_0^4}(W_n^*Y_{nt}^*)'\tilde{V}_{nt}^*\big\}, \; J_{\sigma^2\sigma^2} = -\tfrac{(n-1)(T-1)}{2\sigma_0^4} + \tfrac{1}{\sigma_0^6}\textstyle\sum_{t=1}^T \tilde{V}_{nt}^{*\prime}\tilde{V}_{nt}^*.
\end{aligned}
$$

Its expectation $I_{\mathtt{SL2}}(\theta_0)$ contains the components:

$$
\begin{aligned}
&I_{\beta\beta} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}X_{nt}^{*\prime}X_{nt}^*\big\} - \big\{\tfrac{1}{T\sigma_0^2}X_{nt}^{*\prime}X_{ns}^*\big\}, \; I_{\lambda\beta} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta_{nt}^{*\prime}X_{nt}^*\big\} - \big\{\tfrac{1}{T\sigma_0^2}\eta_{nt}^{*\prime}X_{ns}^*\big\},\\
&I_{\lambda\lambda} = \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta_{nt}^{*\prime}\eta_{nt}^* + \tfrac{T-1}{T}\mathrm{tr}(G_{nt}^{*s}G_{nt}^*)\big\} - \big\{\tfrac{1}{T\sigma_0^2}\eta_{nt}^{*\prime}\eta_{ns}^*\big\}, \; I_{\sigma^2\lambda} = \big\{\tfrac{T-1}{T\sigma_0^2}\mathrm{tr}(G_{nt}^*)\big\},\\
&I_{\sigma^2\beta} = \mathbf{0}, \quad I_{\sigma^2\sigma^2} = \tfrac{(n-1)(T-1)}{2\sigma_0^4},
\end{aligned}
$$

where $\eta_{nt}^* = G_{nt}^*(X_{nt}^*\beta_{t0} + c_n^*)$ and $G_{nt}^{*s} = G_{nt}^* + G_{nt}^{*\prime}$.

The VC matrix $\Sigma_{\mathtt{SL2}}(\boldsymbol{\theta}_0)$ is obtained by applying Lemma B.1.6 with ε replaced by

$\mathbb{V}_N$, c_{nt} by Π_{1t} and Π_{2t}, and A_{nt} by Φ_t and Ψ:

$$\Sigma_{\texttt{SL2}}(\theta_0) = \begin{pmatrix} \{f(\mathbf{0}, \Pi_{1t}; \mathbf{0}, \Pi_{1s})\}, & \{f(\mathbf{0}, \Pi_{1t}; \Phi_s, \Pi_{2s})\}, & \{f(\mathbf{0}, \Pi_{1t}; \Psi, \mathbf{0})\} \\ \sim, & \{f(\Phi_t, \Pi_{2t}; \Phi_s, \Pi_{2s})\}, & \{f(\Phi_t, \Pi_{2t}; \Psi, \mathbf{0})\} \\ \sim, & \sim, & f(\Psi, \mathbf{0}; \Psi, \mathbf{0}) \end{pmatrix}.$$

where $\Pi_{1t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\circ *} X_{nt}^*$, $\Pi_{2t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\circ *} \eta_{nt0}^*$, $\Phi_t = \frac{1}{\sigma_0^2} Z_{Nt}^* G_{nt0}^{*\prime} Z_{Nt}^{\circ *\prime}$, and $\Psi = \frac{1}{2\sigma^4} \sum_{t=1}^T Z_{Nt}^{\circ *} Z_{Nt}^{\circ *\prime}$, with $Z_{Nt}^* = Z_{Nt} F_{n,n-1}$ and $Z_{Nt}^{\circ *} = Z_{Nt}^{\circ} F_{n,n-1}$. $Z_{Nt}^{\circ} = Z_{Nt} - \bar{Z}_N$, $Z_{Nt} = z_t \otimes I_n$, $\bar{Z}_N = \frac{1}{T}(l_T \otimes I_n)$, and z_t is a $T \times 1$ vector of element 1 in the tth position and 0 elsewhere.

B.2.3. Panel SLE model with one-way FE. The negative Hessian matrix $J_{\texttt{SLE1}}(\theta_0)$ has the components:

$$J_{\boldsymbol{\beta\beta}} = \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} X_{nt}' D_n X_{nt}\} - \{\tfrac{1}{T\sigma_0^2} X_{nt}' D_n X_{ns}\},$$
$$J_{\boldsymbol{\beta\lambda}} = \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} X_{nt}' D_n W_n Y_{nt}\} - \{\tfrac{1}{T\sigma_0^2} X_{nt}' D_n W_n Y_{ns}\},$$
$$J_{\boldsymbol{\beta\rho}} = \{\tfrac{1}{\sigma_0^2} X_{nt}' B_n' H_n' \widetilde{V}_{nt} + \tfrac{1}{\sigma_0^2} X_{nt}' B_n' H_n \widetilde{V}_{nt}\}$$
$$J_{\boldsymbol{\lambda\lambda}} = \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2}(W_n Y_{nt})' D_n (W_n Y_{nt}) + \tfrac{T-1}{T}\mathrm{tr}(G_{nt}^2)\} - \{\tfrac{1}{T\sigma_0^2}(W_n Y_{nt})' D_n (W_n Y_{ns})\},$$
$$J_{\boldsymbol{\lambda\rho}} = \{\tfrac{1}{\sigma_0^2}(W_n Y_{nt})' B_n' H_n' \widetilde{V}_{nt} + \tfrac{1}{\sigma_0^2}(W_n Y_{nt})' B_n' H_n \widetilde{V}_{nt}\}$$
$$J_{\boldsymbol{\rho\rho}} = \{\tfrac{1}{\sigma^2} \textstyle\sum_{t=1}^T \widetilde{V}_{nt}' H_n' H_n \widetilde{V}_{nt} + (T-1)\mathrm{tr}(H_n^2)\}$$
$$J_{\sigma^2\boldsymbol{\beta}} = \{\tfrac{1}{\sigma_0^4} X_{nt}' B_n' \widetilde{V}_{nt}\},\ J_{\sigma^2\boldsymbol{\lambda}} = \{\tfrac{1}{\sigma_0^4}(W_n Y_{nt})' B_n' \widetilde{V}_{nt}\},$$
$$J_{\sigma^2\boldsymbol{\rho}} = \{\tfrac{1}{\sigma_0^4} \textstyle\sum_{t=1}^T \widetilde{V}_{nt}' H_n \widetilde{V}_{nt}\},\ J_{\sigma^2\sigma^2} = -\tfrac{n(T-1)}{2\sigma^4} + \tfrac{1}{\sigma^6} \textstyle\sum_{t=1}^T \widetilde{V}_{nt}' \widetilde{V}_{nt}.$$

Its expectation $I_{\texttt{SLE1}}(\theta_0)$ has the components:

$$I_{\beta\beta} = \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} X_{nt}' D_n X_{nt}\} - \{\tfrac{1}{T\sigma_0^2} X_{nt}' D_n X_{ns}\},$$
$$I_{\beta\lambda} = \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} X_{nt}' D_n \eta_{nt}\} - \{\tfrac{1}{T\sigma_0^2} X_{nt}' D_n \eta_s\},\ I_{\boldsymbol{\rho\rho}} = \{(T-1)\mathrm{tr}[H_n^s H_n]\}$$
$$I_{\lambda\lambda} = \texttt{blkdiag}\{\tfrac{1}{\sigma_0^2} \eta_{nt}' D_n \eta_{nt} + \tfrac{T-1}{T}\mathrm{tr}[\bar{G}_{nt}^s \bar{G}_{nt}]\} - \{\tfrac{1}{T\sigma_0^2} \eta_{nt}' D_n \eta_{ns}\},\quad I_{\boldsymbol{\rho}\beta} = 0_{Tk},\quad I_{\sigma^2\beta} = \mathbf{0},$$
$$I_{\lambda\boldsymbol{\rho}} = \tfrac{T-1}{T}\mathrm{tr}[\bar{G}_{nt}' H_n^s],\quad I_{\sigma^2\lambda} = \{\tfrac{T-1}{T\sigma_0^2}\mathrm{tr}[G_{nt}]\},\quad I_{\sigma^2\boldsymbol{\rho}} = \tfrac{T-1}{\sigma_0^2}\mathrm{tr}[H_n],\quad I_{\sigma^2\sigma^2} = \tfrac{n(T-1)}{2\sigma_0^4}$$

where $\eta_{nt} = G_{nt}(X_{nt}\beta_{t0} + c_n)$ and $\bar{G}_{nt} = B_n G_{nt} B_n^{-1}$.

The VC matrix $\Sigma_{\texttt{SLE1}}(\theta_0)$ takes an identical form as $\Sigma_{\texttt{SLE2}}(\theta_0)$ but with $\Pi_{1t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\circ} B_{n0} X_{nt}$, $\Pi_{2t} = \frac{1}{\sigma_0^2} Z_{Nt}^{\circ} B_{n0} \eta_{nt0}$, $\Phi_{1t} = \frac{1}{\sigma_0^2} Z_{Nt} B_{n0}^{-1\prime} G_{nt0}' B_{n0}' Z_{Nt}^{\circ\prime}$, $\Phi_2 = \frac{1}{\sigma_0^2} \sum_{t=1}^T Z_{Nt}^{\circ} H_{n0} Z_{Nt}^{\circ\prime}$, and $\Psi = \frac{1}{2\sigma_0^4} \sum_{t=1}^T Z_{Nt}^{\circ} Z_{Nt}^{\circ\prime}$.

B.2.4. Panel SL model with one-way FE. The negative Hessian matrix $J_{\mathtt{SL1}}(\theta_0)$ has the components:

$$
\begin{aligned}
J_{\boldsymbol{\beta\beta}} &= \mathtt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}X'_{nt}X_{nt}\big\} - \big\{\tfrac{1}{T\sigma_0^2}X'_{nt}X_{ns}\big\},\\
J_{\boldsymbol{\lambda\beta}} &= \mathtt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}(W_nY_{nt})'X_{nt}\big\} - \big\{\tfrac{1}{T\sigma_0^2}(W_nY_{nt})'X_{ns}\big\},\\
J_{\boldsymbol{\lambda\lambda}} &= \mathtt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}(W_nY_{nt})'(W_nY_{nt}) + \tfrac{T-1}{T}\mathrm{tr}(G_{nt}^2)\big\} - \big\{\tfrac{1}{T\sigma_0^2}(W_nY_{nt})'(W_nY_{ns})\big\},\\
J_{\sigma^2\boldsymbol{\beta}} &= \big\{\tfrac{1}{\sigma_0^4}\tilde{V}'_{nt}X_{nt}\big\},\ J_{\sigma^2\boldsymbol{\lambda}} = \big\{\tfrac{1}{\sigma_0^4}(W_nY_{nt})'\tilde{V}_{nt}\big\},\ J_{\sigma^2\sigma^2} = -\tfrac{n(T-1)}{2\sigma_0^4} + \tfrac{1}{\sigma_0^6}\textstyle\sum_{t=1}^T \tilde{V}'_{nt}\tilde{V}_{nt}.
\end{aligned}
$$

Its expectation $I_{\mathtt{SL1}}(\theta_0)$ has the components:

$$
\begin{aligned}
&I_{\beta\beta} = \mathtt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}X'_{nt}X_{nt}\big\} - \big\{\tfrac{1}{T\sigma_0^2}X'_{nt}X_{ns}\big\},\ I_{\lambda\beta} = \mathtt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta'_{nt}X_{nt}\big\} - \big\{\tfrac{1}{T\sigma_0^2}\eta'_{nt}X_{ns}\big\},\\
&I_{\lambda\lambda} = \mathtt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}\eta'_{nt}\eta_{nt} + \tfrac{T-1}{T}\mathrm{tr}(G^s_{nt}G_{nt})\big\} - \big\{\tfrac{1}{T\sigma_0^2}\eta'_{nt}\eta_{ns}\big\},\ \ I_{\sigma^2\lambda} = \big\{\tfrac{T-1}{T\sigma_0^2}\mathrm{tr}(G_{nt})\big\},\\
&I_{\sigma^2\beta} = \mathbf{0},\ I_{\sigma^2\sigma^2} = \tfrac{n(T-1)}{2\sigma_0^4},\ \text{where } \eta_{nt} = G_{nt}(X_{nt}\beta_{t0} + c_n) \text{ and } G^s_{nt} = G_{nt} + G'_{nt}.
\end{aligned}
$$

The VC matrix $\Sigma_{\mathtt{SL1}}(\theta_0)$ takes an identical form as $\Sigma_{\mathtt{SL2}}(\theta_0)$ but with $\Pi_{1t} = \frac{1}{\sigma_0^2}Z^\circ_{Nt}X_{nt}$, $\Pi_{2t} = \frac{1}{\sigma_0^2}Z^\circ_{Nt}\eta_{nt0}$, $\Phi_t = \frac{1}{\sigma_0^2}Z_{Nt}G'_{nt0}Z^{\circ\prime}_{Nt}$, and $\Psi = \frac{1}{2\sigma^4}\sum_{t=1}^T Z^\circ_{Nt}Z^{\circ\prime}_{Nt}$.

B.3 Proofs of Theorems

The following matrix results are used in the proof: (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) $\gamma_{\min}(A)\text{tr}(B) \leqslant \text{tr}(AB) \leqslant \gamma_{\max}(A)\text{tr}(B)$ for symmetric matrix A and positive semidefinite (p.s.d.) matrix B; (iv) $\gamma_{\max}(A+B) \leqslant \gamma_{\max}(A) + \gamma_{\max}(B)$ for symmetric matrices A and B; and (v) $\gamma_{\max}(AB) \leqslant \gamma_{\max}(A)\gamma_{\max}(B)$ for p.s.d. matrices A and B. See, e.g, Bernstein (2009).

Proof of Theorem 3.1: From (3.10) and (3.23), we have

$$S^{\star c}(\delta) - \bar{S}^{\star c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}^2(\delta)} Y_N^{\circ\prime} B_N^{*\prime} \hat{V}_N^*(\delta) - \frac{1}{\bar{\sigma}^2(\delta)} \text{E}[Y_N^{\circ\prime} B_N^{*\prime} \bar{V}_N^*(\delta)], \\ \frac{1}{\hat{\sigma}^2(\delta)} \hat{V}_N^{*\prime}(\delta) H_N^*(\rho) \hat{V}_N^*(\delta) - \frac{1}{\bar{\sigma}^2(\delta)} \text{E}[\bar{V}_N^{*\prime}(\delta) H_N^*(\rho) \bar{V}_N^*(\delta)], \end{cases}$$

With Assumption F, consistency of $\hat{\delta}$ follows from:

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

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

(c) $\sup_{\delta \in \mathbf{\Delta}} \frac{1}{(n-1)(T-1)} \left| Y_N^{\circ\prime} B_N^{*\prime} \hat{V}_N^*(\delta) - \text{E}[Y_N^{\circ\prime} B_N^{*\prime} \bar{V}_N^*(\delta)] \right| = o_p(1)$,

(d) $\sup_{\delta \in \mathbf{\Delta}} \frac{1}{(n-1)(T-1)} \left| \hat{V}_N^{*\prime}(\delta) H_N^*(\rho) \hat{V}_N^*(\delta) - \text{E}[\bar{V}_N^{*\prime}(\delta) H_N^*(\rho) \bar{V}_N^*(\delta)] \right| = o_p(1)$,

Proof of (a). By $\bar{V}_N^*(\delta) = \mathbf{M}\Omega B_N^* A_N^* Y_N^* + \mathbf{P}\Omega B_N^* A_N^* \widetilde{Y}_N^*$ given in (3.25), and the orthogonality between the two projection matrices $\mathbf{M}$ and $\mathbf{P}$, we have,

$$\begin{aligned} \bar{\sigma}^2(\delta) = \; & \tfrac{1}{(n-1)(T-1)} \text{E}[\bar{V}_N^{*\prime}(\delta) \bar{V}_N^*(\delta)] = \tfrac{1}{(n-1)(T-1)} \text{tr}[\text{Var}(\Omega B_N^* A_N^* Y_N^*) \\ & + \tfrac{1}{(n-1)(T-1)} \text{E}(\Omega B_N^* A_N^* Y_N^*)' \mathbf{M} \text{E}(\Omega B_N^* A_N^* Y_N^*). \end{aligned}$$

As $\mathbf{M}$ is p.s.d., the second term is nonnegative uniformly in $\delta \in \mathbf{\Delta}$. The first term is $\frac{1}{(n-1)(T-1)} \text{tr}[\text{Var}(\Omega B_N^* A_N^* Y_N^*)] = \sigma_0^2 > c > 0$, uniformly in $\delta \in \mathbf{\Delta}$ by the assumption given in the theorem. It follows that $\inf_{\delta \in \mathbf{\Delta}} \bar{\sigma}^2(\delta) > c > 0$.

Proof of (b). Noting that $\hat{V}_N^*(\delta) = \mathbf{M}\Omega B_N^* A_N^* Y_N^*$, we have,

$$\hat{\sigma}^2(\delta) = \tfrac{1}{(n-1)(T-1)} \hat{V}_N^{*\prime}(\delta) \hat{V}_N^*(\delta) = \tfrac{1}{(n-1)(T-1)} (\Omega B_N^* A_N^* Y_N^*)' \mathbf{M} (\Omega B_N^* A_N^* Y_N^*).$$

It follows that, by denoting $Q_1 = \frac{1}{(n-1)(T-1)} (\Omega B_N^* A_N^* Y_N^*)' \mathbf{M} (\Omega B_N^* A_N^* Y_N^*)$ and $Q_2 = \frac{1}{(n-1)(T-1)} (\Omega B_N^* A_N^* \widetilde{Y}_N^*)' \mathbf{P} (\Omega B_N^* A_N^* \widetilde{Y}_N^*)$,

$$\hat{\sigma}^2(\delta) - \bar{\sigma}^2(\delta) = Q_1 - \text{E}Q_1 - \text{E}Q_2. \tag{B.3.1}$$

The results follows if $Q_1 - \mathrm{E}Q_1 \xrightarrow{p} 0$, and $\mathrm{E}Q_2 \longrightarrow 0$, uniformly in $\delta \in \mathbf{\Delta}$.

The uniform convergence of $Q_1 - \mathrm{E}Q_1$ to zero in probability follows from the pointwise convergence for each $\delta \in \mathbf{\Delta}$ and the stochastic equicontinuity of Q_1, according to Theorem 1 of Andrews (1992). By Lemma 3.1,

$$Q_1 = \frac{1}{(n-1)(T-1)} \left(\eta_N^{*\prime} \mathbf{M} \eta_N^* + \mathbb{V}_N' F_N' \Omega \mathbf{M} \Omega F_N \mathbb{V}_N + 2\eta_N^{*\prime} \mathbf{M} \Omega F_N \mathbb{V}_N \right),$$

where $\eta_N^* = \Omega B_N^* (X_N^* \boldsymbol{\beta} + C_N^*)$, where $C_N^* = l_T \otimes c_n^*$ and $F_N = I_T \otimes F_{n,n-1}'$. It gives $Q_1 - \mathrm{E}Q_1 = \sum_{\ell=1}^{2} (Q_{1,\ell} - \mathrm{E}Q_{1,\ell})$, where $Q_{1,\ell}, \ell = 1 \; and \; 2$, denote the two stochastic terms of Q_1, and $\mathrm{E}Q_{1,2} = 0$;

Thus, Q_1 is decomposed into terms: $\frac{1}{(n-1)(T-1)} \mathbb{V}_N' \mathrm{Z} \mathbb{V}_N$ and $\frac{1}{(n-1)(T-1)} \xi' \mathbb{V}_N$, where the matrix Z and the vector ξ are defined in terms of F_N, Ω, $\mathbf{M}$ and η_N^*. Note that η_N^* depend on true parameter values, whereas $\mathbf{M}$ depends on ρ.

To show $Q_{1,\ell}(\delta) - \mathrm{E}Q_{1,\ell}(\delta) \xrightarrow{p} 0$, for each $\delta \in \mathbf{\Delta}$, and all ℓ, the following results are used: (i) For the terms quadratic in $\mathbb{V}_N$, they can be written as $\frac{1}{(n-1)(T-1)} \sum_{t=1}^{T} \sum_{s=1}^{T} V_{nt}' \mathrm{Z}_{ts} V_{ns}$. The pointwise convergence of $\frac{1}{n-1} [V_{nt}' \mathrm{Z}_{ts} V_{ns} - \mathrm{E}(V_{nt}' \mathrm{Z}_{ts} V_{ns})]$ follows from Lemma B.1.5 (v), for each $t, s = 1, \ldots, T$; (ii) The pointwise convergence of $\frac{1}{(n-1)(T-1)} \xi' \mathbb{V}_N$ follows from Chebyshev inequality.

Let δ_1 and δ_2 be in $\mathbf{\Delta}$, We have by the mean value theorem that for all the $Q_{1,\ell}(\delta)$ terms:

$$Q_{1,\ell}(\delta_2) - Q_{1,\ell}(\delta_1) = \frac{\partial}{\partial \delta'} Q_{1,\ell}(\bar{\delta})(\delta_2 - \delta_1),$$

where $\bar{\delta}$ lies between δ_1 and δ_2 elementwise. The partial derivatives takes simple form, for $Q_{1,\ell}(\delta)$ that is linear or quadratic in λ_t, it is easy to show that $\sup_{\delta \in \mathbf{\Delta}} |\frac{\partial}{\partial \lambda_t} Q_{1,\ell}(\delta)| = O_p(1)$, for t=1,...,T. As for $\frac{\partial}{\partial \rho} Q_{1,\ell}(\delta)$, note that only the matrix $\mathbf{M}$ involves ρ. Some algebra is used for derivative:

$$\frac{d}{d\rho} \mathbf{M} = \mathbf{M} \Omega \mathbb{M}_N \Gamma B_N^{*\prime} \Omega + \Omega B_N^* \Gamma' \mathbb{M}_N' \Omega \mathbf{M}$$

where $\mathbb{M}_N = I_T \otimes M_n^*$ and $\Gamma = X_N^* (X_N^{*\prime} B_N^{*\prime} \Omega B_N^* X_N^*)^{-1} X_N^{*\prime}$. The results $\sup_{\delta \in \mathbf{\Delta}} |\frac{\partial}{\partial \rho} Q_{1,\ell}(\delta)| = O_p(1)$ can be easily proved for all the $Q_{1,\ell}(\delta)$ quantities. For example, for $Q_{1,1}(\delta)$, noting that $\gamma_{\max}(\mathbf{M}) = 1$,

$$\begin{aligned}
\sup_{\delta\in\mathbf{\Delta}} |\tfrac{\partial}{\partial\rho}Q_{1,1}(\delta)| &= \sup_{\delta\in\mathbf{\Delta}} |\tfrac{1}{(n-1)(T-1)}\tfrac{\partial}{\partial\rho}\mathbb{V}_N' F_N'\Omega\mathbf{M}\Omega F_N\mathbb{V}_N| \\
&= \sup_{\delta\in\mathbf{\Delta}} \tfrac{1}{(n-1)(T-1)}|\mathbb{V}_N' F_N'\Omega\mathbf{M}\Omega\mathbb{M}_N\Gamma B_N^{*\prime}\Omega F_N\mathbb{V}_N \\
&+ \mathbb{V}_N' F_N'\Omega B_N^{*}\Gamma'\mathbb{M}_N'\Omega\mathbf{M}\Omega F_N\mathbb{V}_N| \\
&\leqslant \sup_{\delta\in\mathbf{\Delta}} \tfrac{1}{(n-1)(T-1)}|\mathbb{V}_N' F_N'\mathbb{M}_N\Gamma B_N^{*\prime} F_N\mathbb{V}_N + \mathbb{V}_N' F_N' B_N^{*}\Gamma'\mathbb{M}_N' F_N\mathbb{V}_N| \\
&\leqslant 2\gamma_{\max}(\mathbb{M}_N)\gamma_{\max}(\Gamma)\gamma_{\max}(B_N^{*\prime})\tfrac{1}{(n-1)(T-1)}|\mathbb{V}_N' F_N' F_N\mathbb{V}_N| = O_p(1),
\end{aligned}$$

It follows that $Q_{1,\ell}(\delta)$ are stochastically equicontinuous. Hence, by Theorem 1 of Andrews (1992), $Q_{1,\ell}(\delta) - \mathrm{E}Q_{1,\ell}(\delta) \xrightarrow{p} 0$, uniformly in $\delta \in \mathbf{\Delta}$ for all ℓ. It follows that $Q_1(\delta) - \mathrm{E}Q_1(\delta) \xrightarrow{p} 0$, uniformly in $\delta \in \mathbf{\Delta}$.

It left to show that $\mathrm{E}Q_2(\delta) \to 0$, uniformly in $\delta \in \mathbf{\Delta}$:

$$\begin{aligned}
\mathrm{E}Q_2 = &\ \tfrac{1}{(n-1)(T-1)}\mathrm{tr}[\Omega B_N^* X_N^*(X_N^{*\prime}B_N^{*\prime}\Omega B_N^* X_N^*)^{-1}X_N^{*\prime}B_N^{*\prime}\Omega\mathrm{Var}(B_N^* A_N^* Y_N^*)] \\
\leqslant &\ \tfrac{1}{(n-1)(T-1)}\gamma_{\min}^{-1}(X_N^{*\prime}B_N^{*\prime}\Omega B_N^* X_N^*)\mathrm{tr}[X_N^{*\prime}B_N^{*\prime}\mathrm{Var}(B_N^* A_N^* Y_N^*)B_N^* X_N^*] \\
= &\ \tfrac{1}{(n-1)(T-1)}\gamma_{\min}^{-1}\big(\tfrac{X_N^{*\prime}B_N^{*\prime}\Omega B_N^* X_N^*}{(n-1)(T-1)}\big)\ \tfrac{1}{(n-1)(T-1)}\mathrm{tr}[X_N^{*\prime}B_N^{*\prime}\mathrm{Var}(B_N^* A_N^* Y_N^*)B_N^* X_N^*].
\end{aligned}$$

By Assumption C, we have, $0 < \underline{c}_x \leqslant \gamma_{\min}\big(\tfrac{X_N^{*\prime}B_N^{*\prime}\Omega B_N^* X_N^*}{(n-1)(T-1)}\big)$. It follows that

$$\begin{aligned}
\mathrm{E}Q_2 &\leqslant \tfrac{1}{(n-1)(T-1)}\underline{c}_x^{-1}\tfrac{1}{(n-1)(T-1)}\mathrm{tr}[X_N^{*\prime}B_N^{*\prime}\mathrm{Var}(B_N^* A_N^* Y_N^*)B_N^* X_N^*] \\
&\leqslant \tfrac{1}{(n-1)(T-1)}\underline{c}_x^{-1}\bar{c}_y\tfrac{1}{(n-1)(T-1)}\mathrm{tr}[X_N^{*\prime}B_N^{*\prime}B_N^* X_N^*], \ \ \text{by the assumption in Theorem 3.1} \\
&= O(n^{-1}), \ \ \text{by Assumption C.}
\end{aligned}$$

Hence, $\hat{\sigma}^2(\delta) - \bar{\sigma}^2(\delta) \xrightarrow{p} 0$, uniformly in $\delta \in \mathbf{\Delta}$, completing the proof of **(b)**.

Proofs of (c)-(d). By the expressions of $\hat{V}_N^*(\delta)$, $\bar{V}_N^*(\delta)$ and the Lemma 3.1, all the quantities inside $|\cdot|$ in (c)-(d) can all be expressed in the forms similar to (B.3.1). Thus, the proofs of (c)-(d) follow the proof of (b).

Proof of Theorem 3.2: By the mean value theorem, we have:

$$0 = \tfrac{1}{\sqrt{(n-1)(T-1)}}S^\star(\hat{\theta}) = \tfrac{1}{\sqrt{(n-1)(T-1)}}S^\star(\theta_0) + \big[\tfrac{1}{(n-1)(T-1)}\tfrac{\partial}{\partial\theta'}S^\star(\bar{\theta})\big]\sqrt{(n-1)(T-1)}(\hat{\theta}-\theta_0),$$

where $\bar{\theta}$ lies elementwise between $\hat{\theta}$ and θ_0. The result of the theorem follows if

(a) $\tfrac{1}{\sqrt{(n-1)(T-1)}}S^\star(\theta_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty}\Sigma^\circ(\theta_0)\big]$,

(b) $\frac{1}{(n-1)(T-1)}\left[\frac{\partial}{\partial\theta'}S^\star(\bar{\theta})-\frac{\partial}{\partial\theta'}S^\star(\theta_0)\right]\xrightarrow{p}0$, and

(c) $\frac{1}{(n-1)(T-1)}\left[\frac{\partial}{\partial\theta'}S^\star(\theta_0)-\mathrm{E}\big(\frac{\partial}{\partial\theta'}S^\star(\theta_0)\big)\right]\xrightarrow{p}0$.

Proof of (a). Recall the representation of $S^\star(\theta_0)$ given in (3.28):

$$S^*(\theta_0)=\begin{cases}\Pi_{1t}'\mathbb{V}_N, \quad t=1,\ldots,T,\\ \Pi_{2t}'\mathbb{V}_N+\mathbb{V}_N'\Phi_{1t}\mathbb{V}_N-\frac{T-1}{T}\mathrm{tr}(G_{nt0}^*), \quad t=1,\ldots,T,\\ \mathbb{V}_N'\Phi_2\mathbb{V}_N-(T-1)\mathrm{tr}(H_{n0}^*),\\ \mathbb{V}_N'\Psi\mathbb{V}_N-\frac{(n-1)(T-1)}{2\sigma_0^2},\end{cases} \tag{B.3.2}$$

As the elements of X_{nt} are non-stochastic and uniformly bounded (by Assumption C), and the row and column sums of B_n^* are also uniformly bounded in absolute values by Assumption E and Lemma B.1.1. It follows that the elements of Π_{1t} are uniformly bounded. By Assumption D and Lemma B.1.1(i), G_{nt} is uniformly bounded in both row and column sums. Then by Lemma A.4 of Lee and Yu (2010), we have $A_{nt}^{*-1}=F_{n,n-1}'A_{nt}^{-1}F_{n,n-1}$. Thus, G_{nt}^* is uniformly bounded in both row and column sums by Lemma B.1.1(iii), and the elements of $\eta_{nt}^*=G_{nt}^*(X_{nt}^*\beta_{t0}+c_n^*)$ are also uniformly bounded by Assumption C. It follows that the elements of Π_{2t} are uniformly bounded. Similarly, $B_n^{*-1}=F_{n,n-1}'B_n^{-1}F_{n,n-1}$, and therefore the elements of H_n^* is uniformly bounded in both row and column sums. With these and the definitions of Z_{Nt} and $Z_{Nt}^\diamond$, it is easy to show that Φ_{1t}, Φ_2 and Ψ are uniformly bounded in both row and column sums. Thus, under Assumptions A-F, the central limit theorem (CLT) of linear-quadratic (LQ) form of Kelejian and Prucha (2001) or its simplified version (under iid errors) given in Lemma B.1.7 can be applied to each element of $S^\star(\theta_0)$ to establish its asymptotic normality. Then, an application of Cramér-Wold device gives, $\frac{1}{\sqrt{N^*}}S^\star(\theta_0)\xrightarrow{D}N\big(0,\ \lim_{N^*\to\infty}\Sigma^\circ(\theta_0)\big)$, as $N^*\to\infty$.

Proof of (b). Denote $J(\theta)=-\frac{\partial}{\partial\theta'}S^\star(\theta)$, the negative Hessian matrix of $S^\star(\theta)$. It is easy to show that $\frac{1}{(n-1)(T-1)}J(\theta_0)=O_p(1)$ by Lemma B.1.1 and the model assumptions. $\hat{\theta}\xrightarrow{p}\theta_0$ implies $\bar{\theta}-\theta_0=o_p(1)$, thus $\frac{1}{(n-1)(T-1)}J(\bar{\theta})=O_p(1)$. As $\bar{\sigma}^2\xrightarrow{p}\sigma_0^2$, $\bar{\sigma}^{-r}=\sigma_0^{-r}+o_p(1), r=2,4,6$. Noting that σ^r appears in $J(\theta)$ multiplicatively, $\frac{1}{(n-1)(T-1)}J(\bar{\theta})=\frac{1}{(n-1)(T-1)}J(\bar{\beta},\sigma_0^2,\bar{\rho},\bar{\lambda})+o_p(1)$, i.e., replacing $\bar{\sigma}^2$ by σ_0^2 results in an asymptotically negligible error. The results of (b) follows if

$$\tfrac{1}{(n-1)(T-1)}\left[J(\bar{\boldsymbol{\beta}},\sigma_0^2,\bar{\rho},\bar{\boldsymbol{\lambda}})-J(\boldsymbol{\theta}_0)\right]\xrightarrow{p}0.$$

All the random elements of $J(\boldsymbol{\theta})$ are linear, bilinear, or quadratic in Y_{nt}^* or $\widetilde{V}_{nt}^*$, and linear or quadratic in $\boldsymbol{\beta},\rho$, and $\boldsymbol{\lambda}$. This means that all the corresponding elements in $\tfrac{1}{(n-1)(T-1)}\left[J(\bar{\boldsymbol{\beta}},\sigma_0^2,\bar{\rho},\bar{\boldsymbol{\lambda}})-J(\boldsymbol{\theta}_0)\right]$ are linear, bilinear, or quadratic in Y_{nt}^* or $\widetilde{V}_{nt}^*$, and linear, bilinear or quadratic in $\bar{\boldsymbol{\beta}}-\boldsymbol{\beta}_0,\bar{\rho}-\rho_0$, and $\bar{\boldsymbol{\lambda}}-\boldsymbol{\lambda}_0$, and thus are all $o_p(1)$ by the consistency of $\hat{\boldsymbol{\theta}}$, Lemma 3.1, Lemma B.1.1.

Besides the random elements, it also needs to show that all the 'trace' terms in $\tfrac{1}{(n-1)(T-1)}\left[J(\bar{\boldsymbol{\beta}},\sigma_0^2,\bar{\rho},\bar{\boldsymbol{\lambda}})-J(\boldsymbol{\theta}_0)\right]$ are $o_p(1)$, e.g., $\tfrac{1}{(n-1)(T-1)}[\mathrm{tr}(G_{nt}^{*2}(\bar{\lambda}_t))-\mathrm{tr}(G_{nt}^{*2}(\lambda_{t0}))]=o_p(1)$, for $J_{\lambda_t\lambda_t}$. Let λ_t^* be between $\bar{\lambda}_t$ and λ_{t0}. By the mean value theorem,

$$\tfrac{1}{(n-1)(T-1)}[\mathrm{tr}(G_{nt}^{*2}(\bar{\lambda}_t))-\mathrm{tr}(G_{nt}^{*2}(\lambda_{t0}))]=\tfrac{\bar{\lambda}_t-\lambda_{t0}}{(n-1)(T-1)}\mathrm{tr}(G_{nt}^{*2\ \lambda_t^*}),$$

where $G_{nt}^{*2\ \lambda_t^*}$ are the partial derivatives of G_{nt}^{*2} evaluated at λ_t^*. The elements in G_{nt}^{*2} are the multiplications of the matrices W_n^* and $A_{nt}^{*-1}(\lambda_t)$. Therefore, $G_{nt}^{*2\ \lambda_t^*}$ have elements being the multiplications of the matrices W_n^* and $A_{nt}^{*-1}(\lambda_t)$, and hence are uniformly bounded in a matrix norm, in the neighborhood of λ_{t0} by Lemmas B.1.1 and B.1.2. Therefore, $\tfrac{1}{(n-1)(T-1)}\mathrm{tr}(G_{nt}^{*2\ \lambda_t^*})=O_p(1)$, leading to (**b**).

Proof of (c). For the terms involving only $\widetilde{V}_{nt}^*$, the results follows Lemma B.1.5(v)-(vi), noticing $\widetilde{V}_{nt}^*=F_{n,n-1}'Z_{Nt}^{\circ\prime}\mathbb{V}_N$. For example,

$$\begin{aligned}
&J_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)-\mathrm{E}[J_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)]=\tfrac{1}{\sigma_0^6}[\textstyle\sum_{t=1}^T\widetilde{V}_{nt}^{*\prime}\widetilde{V}_{nt}^*-\mathrm{E}(\sum_{t=1}^T\widetilde{V}_{nt}^{*\prime}\widetilde{V}_{nt}^*)]\\
=\ &\tfrac{1}{\sigma_0^6}[\textstyle\sum_{t=1}^T\mathbb{V}_N'Z_{Nt}^{\circ}F_{n,n-1}F_{n,n-1}'Z_{Nt}^{\circ\prime}\mathbb{V}_N-\mathrm{E}(\sum_{t=1}^T\mathbb{V}_N'Z_{Nt}^{\circ}F_{n,n-1}F_{n,n-1}'Z_{Nt}^{\circ\prime}\mathbb{V}_N)]
\end{aligned}$$

which is easily seen that $Z_{Nt}^{\circ}F_{n,n-1}F_{n,n-1}'Z_{Nt}^{\circ\prime}$ is uniformly bounded in both row and column sums. Thus, Lemma B.1.5(v) leads to $\tfrac{1}{(n-1)(T-1)}\{J_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)-\mathrm{E}[J_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)]\}=o_p(1)$. By Lemma 3.1 all the terms involving Y_{nt}^* can be written as sums of the terms linear in $\mathbb{V}_N$. Thus, the results follow by repeatedly applying Lemma B.1.1, Lemma B.1.5.

Proof of Theorem 3.3: In the large panels, as n and T goes to infinity, $n-1$ is asymptotically equivalent to n, $T-1$ is asymptotically equivalent to T, and N^* is asymptotically equivalent to N. Therefore, the results of Theorem 3.3 is simply proceed by applying the Cramér-Wold device. Brief discussions are as followings.

The asymptotic normality of each element of $\boldsymbol{\theta}$ when T goes to infinity follows from the results of Theorem 3.2, with one more consideration of the adjusted normalizing factor. In Sec. 2.2, we have discussed that the normalizing factor should be adjusted to reflect the different rates of convergence of $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and σ^2. It is obvious that β_t and λ_t components of $S^{\star}(\boldsymbol{\theta})$ are $Op(\sqrt{n})$, but ρ and σ^2 component of $S^{\star}(\boldsymbol{\theta})$ is $Op(\sqrt{N})$, when both n and T approaches to infinity. Therefore, the results of Theorem 3.3 follows from results of Theorem 3.2 and Cramér-Wold device.

C Appendix to Chapter 4

C.1 Some Basic Lemmas

The following lemmas are essential for the derivations and proofs of theoretical results.

Lemma C.1.1 *(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 bounded. 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 bounded.

Lemma C.1.2 *(Lee, 2004, p.1918): For W_n and A_{nt} defined in Model (4.1), if $\|W_n\|$ and $\|A_{nt0}^{-1}\|$ are uniformly bounded, where $\|\cdot\|$ is a matrix norm, then $\|A_{nt}^{-1}\|$ is uniformly bounded in a neighborhood of λ_{t0}.*

Lemma C.1.3 *(Lee, 2004, p.1918): Let X_n be an $n \times p$ matrix. If the elements X_n are uniformly bounded and $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular, then $P_n = X_n(X_n'X_n)^{-1}X_n'$ and $M_n = I_n - P_n$ are uniformly bounded in both row and column sums.*

Lemma C.1.4 *(Lemma B.4, Yang, 2015a, extended): Let $\{A_n\}$ be a sequence of $n \times n$ matrices that are uniformly bounded in either row or column sums. Suppose that the elements $a_{n,ij}$ of A_n are $O(h_n^{-1})$ uniformly in all i and j. Let v_n be a random n-vector of inid elements satisfying Assumption A, and b_n a constant n-vector of elements of uniform order $O(h_n^{-1/2})$. Then*

(i) $\mathrm{E}(v_n' A_n v_n) = O(\frac{n}{h_n})$, *(ii)* $\mathrm{Var}(v_n' A_n v_n) = O(\frac{n}{h_n})$,

(iii) $\mathrm{Var}(v_n' A_n v_n + b_n' v_n) = O(\frac{n}{h_n})$, *(iv)* $v_n' A_n v_n = O_p(\frac{n}{h_n})$,

(v) $v_n' A_n v_n - \mathrm{E}(v_n' A_n v_n) = O_p((\frac{n}{h_n})^{\frac{1}{2}})$, *(vi)* $v_n' A_n b_n = O_p((\frac{n}{h_n})^{\frac{1}{2}})$,

the results (iii) and (vi) remain valid if b_n is a random n-vector independent of v_n such that $\{\mathrm{E}(b_{ni}^2)\}$ are of uniform order $O(h_n^{-1})$.

Lemma C.1.5 *(CLT for Linear-Quadratic Forms, Kelejian and Prucha, 2001). Let* A_n *be* $n \times n$ *matrices and* a_n *be the vector of diagonal elements of* A_n*, Let* v_n *be an* $n \times 1$ *ramdom vector satisfying Assumption A. Let* c_n *be an* $n \times 1$ *random vector, independent of* v_n*. Assume (i)* A_n *is bounded uniformly in row and column sums, (ii)* $n^{-1} \sum_{i=1}^{n} |c_{n,i}^{2+\iota_1}| < \infty, \iota_1 > 0$ *and (iii)* $E|v_{n,i}^{4+\iota_2}| < \infty, \iota_2 > 0$*. Let* $R_n = \mathtt{diag}(r_{n,1}, \ldots, r_{n,n})$*. Define the bilinear-quadratic form:*

$$Q_n = v_n' A_n v_n + c_n' v_n - \sigma^2 \mathrm{tr}(R_n A_n),$$

and let $\sigma_{Q_n}^2$ *be the variance of* Q_n*. Then* $Q_n / \sigma_{Q_n} \xrightarrow{d} N(0, 1)$*.*

C.2 Hessian and Expected Hessian Matrices

Notation. For $t, s = 1, \ldots, T$, $\texttt{blkdiag}\{A_t\}$ forms a block-diagonal matrix by placing A_t diagonally, $\{A_t\}$ forms a matrix by stacking A_t horizontally, and $\{B_{ts}\}$ forms a matrix by the component matrices B_{ts}. The negative Hessian $J(\boldsymbol{\theta}_0)$ and expected negative Hessian $I(\boldsymbol{\theta}_0)$ of the AQS function, are both partitioned according to the slope parameters $\boldsymbol{\beta}$, the spatial lag parameters $\boldsymbol{\lambda}$, and the error variance σ^2, with the sub-matrices denoted by, e.g., $I_{\boldsymbol{\beta\beta}}$, $I_{\boldsymbol{\beta\lambda}}$, $J_{\boldsymbol{\beta\beta}}$, $J_{\boldsymbol{\beta\lambda}}$. Furthermore, $\texttt{diag}(\cdot)$ forms a diagonal matrix and $\texttt{diagv}(\cdot)$ a column vector, based on the diagonal elements of a square matrix.

Parametric quantities, e.g., $A_n(\lambda_{t0})$ and $B_n(\rho_0)$, evaluated at the true parameters are denoted as A_{nt} and B_n. For a matrix A_n, denote $A_n^s = A_n + A_n'$. The bold **0** represents generically a vector or a matrix of zeros, to distinguish from the scalar 0.

Letting $\eta_{nt} = G_{nt}(X_{nt}\beta_t + c_n)$ and $g_{nt} = \texttt{diagv}(G_{nt})$, the negative Hessian matrix, $J_{\texttt{SL1}}(\boldsymbol{\theta}_0)$, has the components:

$$
\begin{aligned}
J_{\boldsymbol{\beta\beta}} &= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2} X_{nt}' X_{nt}\big\} - \big\{\tfrac{1}{T\sigma_0^2} X_{nt}' X_{ns}\big\}, \\
J_{\boldsymbol{\beta\lambda}} &= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2} X_{nt}'(W_n Y_{nt})\big\} - \big\{\tfrac{1}{T\sigma_0^2} X_{nt}'(W_n Y_{ns})\big\}, \\
J_{\boldsymbol{\lambda\beta}} &= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}[(W_n Y_{nt}) - 2\texttt{diag}(G_{nt})\widetilde{V}_{nt}]' X_{nt}\big\} \\
&\quad - \big\{\tfrac{1}{T\sigma_0^2}[(W_n Y_{nt}) - 2\texttt{diag}(G_{nt})\widetilde{V}_{nt}]' X_{ns}\big\}, \\
J_{\boldsymbol{\lambda\lambda}} &= \texttt{blkdiag}\big\{\tfrac{1}{\sigma_0^2}(W_n Y_{nt})'[(W_n Y_{nt}) - 2\texttt{diag}(G_{nt})\widetilde{V}_{nt}] + \tfrac{1}{\sigma_0^2}\widetilde{V}_{nt}'\texttt{diag}(G_{nt}^2)\tilde{V}_{nt}\big\} \\
&\quad - \big\{\tfrac{1}{T\sigma_0^2}(W_n Y_{ns})'[(W_n Y_{nt}) - 2\texttt{diag}(G_{nt})\widetilde{V}_{nt}]\big\}, \\
J_{\sigma^2\boldsymbol{\beta}} &= \big\{\tfrac{1}{\sigma_0^4}\widetilde{V}_{nt}' X_{nt}\big\}, \\
J_{\sigma^2\boldsymbol{\lambda}} &= \big\{\tfrac{1}{\sigma_0^4}(W_n Y_{nt})'\widetilde{V}_{nt}\big\}, \\
J_{\boldsymbol{\lambda}\sigma^2} &= \big\{\tfrac{1}{\sigma_0^4}\widetilde{V}_{nt}'[(W_n Y_{nt}) - \texttt{diag}(G_{nt})\tilde{V}_{nt}]\big\}, \\
J_{\sigma^2\sigma^2} &= -\tfrac{n(T-1)}{2\sigma_0^4} + \tfrac{1}{\sigma_0^6}\textstyle\sum_{t=1}^T \widetilde{V}_{nt}'\widetilde{V}_{nt}.
\end{aligned}
$$

The expected negative Hessian matrix, $I_{\mathtt{SL1}}(\theta_0)$, has the components:

$$
\begin{aligned}
I_{\beta\beta} &= \mathtt{blkdiag}\left\{\tfrac{1}{\sigma_0^2}X'_{nt}X_{nt}\right\} - \left\{\tfrac{1}{T\sigma_0^2}X'_{nt}X_{ns}\right\}, \\
I_{\lambda\beta} &= \mathtt{blkdiag}\left\{\tfrac{1}{\sigma_0^2}\eta'_{nt}X_{nt}\right\} - \left\{\tfrac{1}{T\sigma_0^2}\eta'_{nt}X_{ns}\right\}, \\
I_{\lambda\lambda} &= \mathtt{blkdiag}\left\{\tfrac{1}{\sigma_0^2}\eta'_{nt}\eta_{nt} + \tfrac{T-1}{T}\mathrm{tr}[R_n G'_{nt}G_{nt} + R_n\mathtt{diag}(G_{nt}^2)] - \tfrac{2(T-2)}{T}\mathrm{tr}[R_n\mathtt{diag}(G_{nt})G_{nt}]\right\} \\
&\quad - \left\{\tfrac{1}{T\sigma_0^2}\eta'_{nt}\eta_{ns} + \tfrac{2}{T^2}\mathrm{tr}[R_n\mathtt{diag}(G_{nt})G_{ns}]\right\}, \\
I_{\sigma^2\lambda} &= \left\{\tfrac{T-1}{T\sigma_0^2}\mathrm{tr}(R_n G_{nt})\right\}, \\
I_{\lambda\sigma^2} &= \left\{\tfrac{T-1}{T\sigma_0^2}\mathrm{tr}(R_n G_{nt}^{\circ})\right\} \\
I_{\sigma^2\beta} &= \mathbf{0}, \\
I_{\sigma^2\sigma^2} &= \tfrac{n(T-1)}{2\sigma_0^4}.
\end{aligned}
$$

C.3 Proofs of Theorems

In the proofs, the following matrix results are useful: (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) $\gamma_{\min}(A)\mathrm{tr}(B) \leqslant \mathrm{tr}(AB) \leqslant \gamma_{\max}(A)\mathrm{tr}(B)$ for symmetric matrix A and positive semidefinite (p.s.d.) matrix B; (iv) $\gamma_{\max}(A+B) \leqslant \gamma_{\max}(A) + \gamma_{\max}(B)$ for symmetric matrices A and B; and (v) $\gamma_{\max}(AB) \leqslant \gamma_{\max}(A)\gamma_{\max}(B)$ for p.s.d. matrices A and B. See, e.g, Bernstein (2009).

proof of Theorem 4.1: From (4.10) and (4.15), we have $S^{\star c}(\boldsymbol{\lambda}) - \bar{S}^{\star c}(\boldsymbol{\lambda})$ equals to

$$\tfrac{1}{\hat{\sigma}^2(\boldsymbol{\lambda})}\hat{\eta}_N'\hat{V}_N(\boldsymbol{\lambda}) - \tfrac{1}{\bar{\sigma}^2(\boldsymbol{\lambda})}\mathrm{E}[\bar{\eta}_N'\bar{V}_N(\boldsymbol{\lambda})] + \tfrac{1}{\hat{\sigma}^2(\boldsymbol{\lambda})}\hat{V}_N^{\circ\prime}(\boldsymbol{\lambda})G_N^{\circ}(\boldsymbol{\lambda})\hat{V}_N(\boldsymbol{\lambda}) - \tfrac{1}{\bar{\sigma}^2(\boldsymbol{\lambda})}\mathrm{E}[\bar{V}_N^{\circ\prime}(\boldsymbol{\lambda})G_N^{\circ}(\boldsymbol{\lambda})\bar{V}_N(\boldsymbol{\lambda})]$$

With Assumption E, consistency of $\hat{\boldsymbol{\lambda}}$ follows from:

(**a**) $\inf_{\boldsymbol{\lambda}\in\boldsymbol{\lambda}}\bar{\sigma}^2(\boldsymbol{\lambda})$ is bounded away from zero,

(**b**) $\sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\big|\hat{\sigma}^2(\boldsymbol{\lambda}) - \bar{\sigma}^2(\boldsymbol{\lambda})\big| = o_p(1)$,

(**c**) $\sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\frac{1}{n(T-1)}\big|\hat{\eta}_N'\hat{V}_N(\boldsymbol{\lambda}) - \mathrm{E}[\bar{\eta}_N'\bar{V}_N(\boldsymbol{\lambda})]\big| = o_p(1)$,

(**d**) $\sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\frac{1}{n(T-1)}\big|\hat{V}_N^{\circ\prime}(\boldsymbol{\lambda})G_N^{\circ}(\boldsymbol{\lambda})\hat{V}_N(\boldsymbol{\lambda}) - \mathrm{E}[\bar{V}_N^{\circ\prime}(\boldsymbol{\lambda})G_N^{\circ}(\boldsymbol{\lambda})\bar{V}_N(\boldsymbol{\lambda})]\big| = o_p(1)$,

Proof of (a). The identity (4.14), $\bar{V}_N(\boldsymbol{\lambda}) = \mathbf{M}\Omega A_N Y_N + \mathbf{P}\Omega A_N \widetilde{Y}_N$, is useful in obtaining the expressions for $\bar{\sigma}^2(\boldsymbol{\lambda})$. By the orthogonality between the two projection matrices $\mathbf{M}$ and $\mathbf{P}$, we have,

$$\begin{aligned}\bar{\sigma}^2(\boldsymbol{\lambda}) = \; & \tfrac{1}{n(T-1)}\mathrm{E}[\bar{V}_N'(\boldsymbol{\lambda})\bar{V}_N(\boldsymbol{\lambda})] = \tfrac{1}{n(T-1)}\mathrm{tr}[\mathrm{Var}(\Omega A_N Y_N)] \\ & + \tfrac{1}{n(T-1)}\mathrm{E}(\Omega A_N Y_N)'\mathbf{M}\mathrm{E}(\Omega A_N Y_N).\end{aligned}$$

As $\mathbf{M}$ is p.s.d., the second term is nonnegative uniformly in $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$. The first term is $\frac{1}{n(T-1)}\mathrm{tr}[\mathrm{Var}(\Omega A_N Y_N)] = \sigma_0^2 > c > 0$, uniformly in $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$ by the assumption in the theorem. It follows that $\inf_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\bar{\sigma}^2(\boldsymbol{\lambda}) > c > 0$.

Proof of (b). Noting that $\hat{V}_N(\boldsymbol{\lambda}) = \mathbf{M}\Omega A_N Y_N$, we have,

$$\hat{\sigma}^2(\boldsymbol{\lambda}) = \tfrac{1}{n(T-1)}\hat{V}_N'(\boldsymbol{\lambda})\hat{V}_N(\boldsymbol{\lambda}) = \tfrac{1}{n(T-1)}(\Omega A_N Y_N)'\mathbf{M}(\Omega A_N Y_N).$$

By denoting $Q_1 = \frac{1}{n(T-1)}(\Omega A_N Y_N)'\mathbf{M}(\Omega A_N Y_N)$ and $Q_2 = \frac{1}{n(T-1)}(\Omega A_N \widetilde{Y}_N)'\mathbf{P}(\Omega A_N \widetilde{Y}_N)$,

$$\hat{\sigma}^2(\boldsymbol{\lambda}) - \bar{\sigma}^2(\boldsymbol{\lambda}) = Q_1 - \mathrm{E}Q_1 - \mathrm{E}Q_2. \tag{C.3.1}$$

The results follows if $Q_1 - \mathrm{E}Q_1 \xrightarrow{p} 0$, and $\mathrm{E}Q_2 \longrightarrow 0$, uniformly in $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$.

According to Theorem 1 of Andrews (1992), the uniform convergence of $Q_1 - \mathrm{E}Q_1$ to zero in probability follows from the pointwise convergence for each $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$ and the stochastic equicontinuity of Q_1. Q_1 can be written in the form of $\mathbb{V}_N$, where $\mathbb{V}_N = (V_{n1}', \ldots, V_{nT}')'$ is the vector of original errors with elements $\{v_{it}\}$ satisfying Assumption A.

$$Q_1 = \frac{1}{n(T-1)}\left(\eta_N^{*\prime}\mathbf{M}\eta_N^* + \mathbb{V}_N'\Omega\mathbf{M}\Omega\mathbb{V}_N + 2\eta_N^{*\prime}\mathbf{M}\Omega\mathbb{V}_N\right),$$

where $\eta_N^* = \Omega(X_N\boldsymbol{\beta} + C_N)$ and $C_N = l_T \otimes c_n$. Denote $Q_1 - \mathrm{E}Q_1 = \sum_{\ell=1}^{2}(Q_{1,\ell} - \mathrm{E}Q_{1,\ell})$, where $Q_{1,\ell}, \ell = 1 \; and \; 2$ are the two stochastic terms of Q_1, and $\mathrm{E}Q_{1,2} = 0$.

The above decomposition contains terms in the form: $\frac{1}{n(T-1)}\mathbb{V}_N'\mathrm{Z}\mathbb{V}_N$ and $\frac{1}{n(T-1)}\xi'\mathbb{V}_N$, where the matrix Z and the vector ξ are defined in terms of Ω, $\mathbf{M}$ and η_N^*. Note that η_N^* depends on true parameter values.

For the term quadratic in $\mathbb{V}_N$, it can be written as $\frac{1}{n(T-1)}\sum_{t=1}^{T}\sum_{s=1}^{T}V_{nt}'\mathrm{Z}_{ts}V_{ns}$. The pointwise convergence of $\frac{1}{n}[V_{nt}'\mathrm{Z}_{ts}V_{ns} - \mathrm{E}(V_{nt}'\mathrm{Z}_{ts}V_{ns})]$ follows from Lemma C.1.4(v), for each $t, s = 1, \ldots, T$. The pointwise convergence of $\frac{1}{n(T-1)}\xi'\mathbb{V}_N$ follows from Chebyshev inequality. Thus, it follows that $Q_{1,\ell}(\boldsymbol{\lambda}) - \mathrm{E}Q_{1,\ell}(\boldsymbol{\lambda}) \xrightarrow{p} 0$, for each $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$, and all ℓ.

Let $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ be in $\boldsymbol{\Lambda}$, We have by the mean value theorem that for all the $Q_{1,\ell}(\boldsymbol{\lambda})$ terms:

$$Q_{1,\ell}(\boldsymbol{\lambda}_2) - Q_{1,\ell}(\boldsymbol{\lambda}_1) = \frac{\partial}{\partial\boldsymbol{\lambda}'}Q_{1,\ell}(\bar{\boldsymbol{\lambda}})(\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1),$$

where $\bar{\boldsymbol{\lambda}}$ lies between $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ elementwise. The partial derivatives take simple form, for $Q_{1,\ell}(\boldsymbol{\lambda})$ that is linear or quadratic in λ_t, it is easy to show that $\sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}|\frac{\partial}{\partial\lambda_t}Q_{1,\ell}(\boldsymbol{\lambda})| = O_p(1)$, for t=1,...,T. Therefore, it follows that $Q_{1,\ell}(\boldsymbol{\lambda})$ are stochastically equicontinuous. Hence, by Theorem 1 of Andrews (1992), $Q_{1,\ell}(\boldsymbol{\lambda}) - \mathrm{E}Q_{1,\ell}(\boldsymbol{\lambda}) \xrightarrow{p} 0$, uniformly in $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$ for all ℓ. It follows that $Q_1(\boldsymbol{\lambda}) - \mathrm{E}Q_1(\boldsymbol{\lambda}) \xrightarrow{p} 0$, uniformly in $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$.

To show that $\mathrm{E}Q_2(\boldsymbol{\lambda}) \to 0$, uniformly in $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$. We have,

$$
\begin{aligned}
\mathrm{E}Q_2 = \;& \tfrac{1}{n(T-1)}\mathrm{tr}[\Omega X_N(X_N'\Omega X_N)^{-1}X_N'\Omega \mathrm{Var}(A_N Y_N)] \\
&\leqslant \tfrac{1}{n(T-1)}\gamma_{\min}^{-1}(X_N'\Omega X_N)\mathrm{tr}[X_N'\mathrm{Var}(A_N Y_N)X_N] \\
&= \tfrac{1}{n(T-1)}\gamma_{\min}^{-1}\big(\tfrac{X_N'\Omega X_N}{n(T-1)}\big)\ \tfrac{1}{n(T-1)}\mathrm{tr}[X_N'\mathrm{Var}(A_N Y_N)X_N].
\end{aligned}
$$

By Assumption C, we have, $0 < \underline{c}_x \leqslant \gamma_{\min}\big(\frac{X_N'\Omega X_N}{n(T-1)}\big)$. It follows that

$$
\begin{aligned}
\mathrm{E}Q_2 \;&\leqslant\; \tfrac{1}{n(T-1)}\underline{c}_x^{-1}\tfrac{1}{n(T-1)}\mathrm{tr}[X_N'\mathrm{Var}(A_N Y_N)X_N] \\
&\leqslant\; \tfrac{1}{n(T-1)}\underline{c}_x^{-1}\bar{c}_y\ \tfrac{1}{n(T-1)}\mathrm{tr}[X_N'X_N], \quad \text{by the assumption in Theorem 4.1} \\
&=\; O(n^{-1}), \quad \text{by the assumption C}
\end{aligned}
$$

Hence, $\hat{\sigma}^2(\boldsymbol{\lambda}) - \bar{\sigma}^2(\boldsymbol{\lambda}) \xrightarrow{p} 0$, uniformly in $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$, completing the proof of **(b)**.

Proof of (c)-(d). By the expressions of $\hat{V}_N(\boldsymbol{\lambda})$, $\bar{V}_N(\boldsymbol{\lambda})$ and Lemma 4.1, all the quantities inside $|\cdot|$ in (c)-(d) can all be expressed in the forms similar to (C.3.1). Thus, the proofs of (c)-(d) follow the proof of (b).

Proof of Theorem 4.2: By the mean value theorem,

$$
0 = \tfrac{1}{\sqrt{n(T-1)}}S^\star(\hat{\boldsymbol{\theta}}) = \tfrac{1}{\sqrt{n(T-1)}}S^\star(\boldsymbol{\theta}_0) + \big[\tfrac{1}{n(T-1)}\tfrac{\partial}{\partial\boldsymbol{\theta}'}S^\star(\bar{\boldsymbol{\theta}})\big]\sqrt{n(T-1)}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),
$$

where $\bar{\boldsymbol{\theta}}$ lies elementwise between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. The theorem follows if

(a) $\frac{1}{\sqrt{n(T-1)}}S^\star(\boldsymbol{\theta}_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty}\Sigma^\circ(\boldsymbol{\theta}_0)\big]$,
(b) $\frac{1}{n(T-1)}\big[\frac{\partial}{\partial\boldsymbol{\theta}'}S^\star(\bar{\boldsymbol{\theta}}) - \frac{\partial}{\partial\boldsymbol{\theta}'}S^\star(\boldsymbol{\theta}_0)\big] \xrightarrow{p} 0$, and
(c) $\frac{1}{n(T-1)}\big[\frac{\partial}{\partial\boldsymbol{\theta}'}S^\star(\boldsymbol{\theta}_0) - \mathrm{E}\big(\frac{\partial}{\partial\boldsymbol{\theta}'}S^\star(\boldsymbol{\theta}_0)\big)\big] \xrightarrow{p} 0$.

Proof of (a). Elements in the AQS function that are in the form of $\widetilde{V}_N$ can be written in terms of the original error $\mathbb{V}_N$. Thus, We represent $S^\star(\boldsymbol{\theta}_0)$ in terms of $\mathbb{V}_N$. Let z_t be a $T\times 1$ vector of element 1 in the tth position and 0 elsewhere, and define $Z_{Nt} = z_t \otimes I_n$, $\bar{Z}_N = \frac{1}{T}(l_T \otimes I_n)$, and $Z_{Nt}^\circ = Z_{Nt} - \bar{Z}_N$. Thus, $V_{nt} = Z_{Nt}'\mathbb{V}_N$ and $\widetilde{V}_{nt} = V_{nt} - \overline{V}_n = Z_{Nt}^{\circ\prime}\mathbb{V}_N$. The AQS function $S^\star(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$ takes the form:

$$
S^\star(\boldsymbol{\theta}_0) = \begin{cases} \Pi_{1t}^{\diamond\prime}\mathbb{V}_N, \ t = 1,\ldots,T, \\ \Pi_{2t}^{\diamond\prime}\mathbb{V}_N + \mathbb{V}_N'\Phi_{1t}^\diamond\mathbb{V}_N - \frac{T-1}{T}\mathrm{tr}(G_{nt0}), \ t = 1,\ldots,T, \\ \mathbb{V}_N'\Phi_2^\diamond\mathbb{V}_N - \frac{n(T-1)}{2\sigma_0^2}, \end{cases} \tag{C.3.2}
$$

where $\Pi_{1t}^{\diamond} = \frac{1}{\sigma_0^2} Z_{Nt}^{\circ} X_{nt}$, $\Pi_{2t}^{\diamond} = \frac{1}{\sigma_0^2} Z_{Nt}^{\circ} \tilde{\eta}_{nt0}$, $\Phi_{1t}^{\diamond} = \frac{1}{\sigma_0^2} Z_{Nt}^{\circ} G_{nt0} Z_{Nt}^{\circ\prime}$, and $\Phi_2^{\diamond} = \frac{1}{2\sigma_0^4} \sum_{t=1}^{T}$ $Z_{Nt}^{\circ} Z_{Nt}^{\circ\prime}$.

As the elements of X_{nt} are non-stochastic and uniformly bounded (by Assumption C), it is easy to see that the elements of Π_{1t} are uniformly bounded. By Assumption D and Lemma C.1.1(i), G_{nt} is uniformly bounded in both row and column sums. The elements of $\tilde{\eta}_{nt} = G_{nt}(X_{nt}\beta_{t0} + \tilde{c}_n)$ are uniformly bounded by Assumption C. It follows that the elements of Π_{2t} are uniformly bounded. With these and the definition of Z_{Nt} and Z_{Nt}°, it is easy to see that Φ_{1t} and Φ_2 are uniformly bounded in both row and column sums. Thus, under Assumptions A-E, the central limit theorem (CLT) of linear-quadratic (LQ) form of Kelejian and Prucha (2001) or its simplified version given in Lemma C.1.5 can be applied to the elements of $S^{\star}(\theta_0)$ to establish the asymptotic normality. Then, an application of Cramér-Wold device under a finit T gives, as $N^* \to \infty$, $\frac{1}{\sqrt{N^*}} S^{\star}(\theta_0) \xrightarrow{D} N\big(0, \lim_{N^* \to \infty} \Sigma^{\circ}(\theta_0)\big)$.

Proof of (b). Denote $J(\theta)$ as the negative Hessian matrix of $S^{\star}(\theta)$, that is $J(\theta) = -\frac{\partial}{\partial \theta'} S^{\star}(\theta)$. It is easy to show that $\frac{1}{n(T-1)} J(\theta_0) = O_p(1)$ by Lemma C.1.1 and the model assumptions. $\hat{\theta} \xrightarrow{p} \theta_0$ by the consistency, which implies $\bar{\theta} - \theta_0 = o_p(1)$, thus $\frac{1}{n(T-1)} J(\bar{\theta}) = O_p(1)$. As $\bar{\sigma}^2 \xrightarrow{p} \sigma_0^2$, $\bar{\sigma}^{-r} = \sigma_0^{-r} + o_p(1), r = 2, 4, 6$. Noting that σ^r appears in $J(\theta)$ multiplicatively, $\frac{1}{n(T-1)} J(\bar{\theta}) = \frac{1}{n(T-1)} J(\bar{\beta}, \sigma_0^2, \bar{\lambda}) + o_p(1)$, i.e., replacing $\bar{\sigma}^2$ by σ_0^2 results in an asymptotically negligible error. The results of $(\mathbf{b})$ follows if

$$\frac{1}{n(T-1)} \big[J(\bar{\beta}, \sigma_0^2, \bar{\lambda}) - J(\theta_0) \big] \xrightarrow{p} 0. \tag{C.3.3}$$

As all the random elements of $J(\theta)$ are linear, bilinear, or quadratic in Y_{nt} or $\widetilde{V}_{nt}$, and linear or quadratic in β and λ. This means that all the corresponding elements in $\frac{1}{n(T-1)} \big[J(\bar{\beta}, \sigma_0^2, \bar{\lambda}) - J(\theta_0) \big]$ are linear, bilinear, or quadratic in Y_{nt} or $\widetilde{V}_{nt}$, and linear, bilinear or quadratic in $\bar{\beta} - \beta_0$ and $\bar{\lambda} - \lambda_0$, and thus are all $o_p(1)$ by the consistency of $\hat{\theta}$ and Lemma C.1.1.

Proof of (c). For the terms involving $\widetilde{V}_{nt}$, it can be written in the form of $\mathbb{V}_N$, where $\mathbb{V}_N = (V_{n1}', \ldots, V_{nT}')'$ is the vector of original errors satisfying Assumption A. the results follows Lemma C.1.4(v)-(vi). Let $\bar{Z}_N = \frac{1}{T}(l_T \otimes I_n)$, and $Z_{Nt}^{\circ} = Z_{Nt} - \bar{Z}_N$, noticing

$\widetilde{V}_{nt} = Z^{\circ\prime}_{Nt}\mathbb{V}_N$. For example,

$$J_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0) - \mathrm{E}[J_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)] = \tfrac{1}{\sigma_0^6}[\textstyle\sum_{t=1}^T \widetilde{V}'_{nt}\widetilde{V}_{nt} - \mathrm{E}(\sum_{t=1}^T \widetilde{V}'_{nt}\widetilde{V}_{nt})]$$
$$= \tfrac{1}{\sigma_0^6}[\textstyle\sum_{t=1}^T \mathbb{V}'_N Z^{\circ}_{Nt} Z^{\circ\prime}_{Nt}\mathbb{V}_N - \mathrm{E}(\sum_{t=1}^T \mathbb{V}'_N Z^{\circ}_{Nt} Z^{\circ\prime}_{Nt}\mathbb{V}_N)]$$

which is easily seen that $Z^{\circ}_{Nt}Z^{\circ\prime}_{Nt}$ is uniformly bounded in both row and column sums. Thus, $\frac{1}{n(T-1)}\{J_{\sigma^2\sigma^2}(\theta_0) - \mathrm{E}[J_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)]\} = o_p(1)$ by Lemma C.1.4.

Similarly, all the terms involving Y_{nt} can be written as sums of the terms linear in terms of $\mathbb{V}_N$. Thus, the results follow by repeatedly applying Lemma C.1.1 and Lemma C.1.4.

Proof of Theorem 4.3: In the large panels, as n and T goes to infinity, $n-1$ is asymptotically equivalent to n, $T-1$ is asymptotically equivalent to T, and N^* is asymptotically equivalent to N. Therefore, the results of Theorem 4.3 is simply proceed by applying the Cramér-Wold device. Brief discussions are as followings.

The asymptotic normality of each element of θ when T goes to infinity follows from the results of Theorem 4.2, with one more consideration of the adjusted normalizing factor. In Sec. 4.2, we have discussed that the normalizing factor should be adjusted to reflect the different rates of convergence of $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and σ^2. It is obvious that β_t and λ_t components of $S^\star(\theta)$ are $Op(\sqrt{n})$, but σ^2 component of $S^\star(\theta)$ is $Op(\sqrt{N})$, when both n and T approaches to infinity. Therefore, the results of Theorem 4.3 follows from results of Theorem 4.2 and Cramér-Wold device.

Proof of Theorem 4.4: The following *dot* notation is introduced in the proof: (a) for an $nT \times 1$ vector $\widetilde{V}_N$ with elements $\{\widetilde{V}_{it}\}$ double indexed by $i = 1, \ldots, n$ for each $t = 1, \ldots, T$, $\{\widetilde{V}_{\cdot t}\}$ is the subvector that contains all the elements with the same t, and $\{\widetilde{V}_{i\cdot}\}$ is the subvector that picks up the elements with the same i; (b) for an $nT \times nT$ matrix Φ with elements $\{\Phi_{it,ls},\ i, l = 1, \ldots, n;\ t, s = 1, \ldots, T\}$, where it is the double index for the rows and ls the double index for the columns, $\Phi_{\cdot t,\cdot s}$ is the $n \times n$ submatrix corresponding to the (t, s) periods, $\Phi_{i\cdot,l\cdot}$ the $T \times T$ submatrix corresponding to the (i, l) units, $\Phi_{it,l\cdot}$ the $T \times 1$ subvector that picks up the element from the itth row corresponding to $s = 1, \ldots, T$.

Firstly, the result $I^{\circ}(\hat{\theta}) - I^{\circ}(\boldsymbol{\theta}_0) \xrightarrow{p} 0$ is implied by the result **(b)** in the proof of Theorem 4.2. Secondly, the result $\frac{1}{n(T-1)} \sum_{i=1}^{n} [\hat{g}_i \hat{g}_i' - \mathrm{E}(g_i g_i')] \xrightarrow{p} 0$ follows if

(a) $\frac{1}{n(T-1)} \sum_{i=1}^{n} (\hat{g}_i \hat{g}_i' - g_i g_i') \xrightarrow{p} 0$

(b) $\frac{1}{n(T-1)} \sum_{i=1}^{n} [g_i g_i' - \mathrm{E}(g_i g_i')] \xrightarrow{p} 0.$

Proof of (a). By applying the mean value theorem, the proof is straightforward.

Proof of (b). As in Lemma 4.2, the elements of $S^{\star}(\theta_0)$ are mixtures of terms of the forms $\Pi' \widetilde{V}_N = \sum_{i=1}^{n} g_{\pi i}$ and $\widetilde{V}_N' \Phi \widetilde{V}_N - \mathrm{E}(\widetilde{V}_N' \Phi \widetilde{V}_N) = \sum_{i=1}^{n} g_{\Phi i}$, it suffices to show that

$$\frac{1}{n(T-1)} \sum_{i=1}^{n} [g_{ki} g_{ri}' - \mathrm{E}(g_{ki} g_{ri}')] = o_p(1), \;\; k, r = \Pi, \Phi.$$

Notations defined in Lemma 4.2 can be written in the form of *vector dot*. $g_{\Pi i} = \Pi_{i\cdot}' \widetilde{V}_{i\cdot}$ and $g_{\Phi i} = \widetilde{V}_{i\cdot}' \xi_{i\cdot} + \widetilde{V}_{i\cdot}' \widetilde{V}_{i\cdot}^{*} - 1_T' d_{i\cdot}$. Note that by Assumptions C, D and Lemma C.1.1 that the elements of all the Π's and Φ's, defined in (4.17), are uniformly bounded. The proofs proceed by applying the weak law of large numbers (WLLN) for M.D. arrays, see, e.g., Davidson (1994, p. 299).

As $g_{\pi i} = \Pi_{i\cdot}' \widetilde{V}_{i\cdot}$, we have $\frac{1}{n(T-1)} \sum_{i=1}^{n} [g_{\pi i} g_{\pi i}' - \mathrm{E}(g_{\pi i} g_{\pi i}')] = \frac{1}{n(T-1)} \sum_{i=1}^{n} \Pi_{i\cdot}' (\widetilde{V}_{i\cdot} \widetilde{V}_{i\cdot}' - \frac{T-1}{T} \sigma_0^2 r_{n,i} I_T) \Pi_{i\cdot}$
$\equiv \frac{1}{n(T-1)} \sum_{i=1}^{n} U_{n,i}$. Without loss of generality, assume U_{ni} is a scalar, it is easy to see that $\{U_{n,i}\}$ are independent, thus form a M.D. array. By Assumption A and the property that the elements of $\Pi_{i\cdot}$ are uniformly bounded, it is easy to show that $\mathrm{E}|U_{n,i}|^{1+\epsilon} \leqslant K_u < \infty$, for $\epsilon > 0$. Thus, $\{U_{n,i}\}$ are uniformly integrable and $\frac{1}{n(T-1)} \sum_{i=1}^{n} U_{n,i} \xrightarrow{p} 0$ by applying the WLLN for M.D. arrays of Davidson.

As $g_{\Phi i} = \widetilde{V}_{i\cdot}' \xi_{i\cdot} + \widetilde{V}_{i\cdot}' \widetilde{V}_{i\cdot}^{*} - 1_T' d_{i\cdot}$, the expression of $\frac{1}{n(T-1)} \sum_{i=1}^{n} [g_{\Phi i}^2 - \mathrm{E}(g_{\Phi i}^2)]$ is more complicated as more terms are involved in, we simplify it to five terms, that is:

$$\begin{aligned}
& \frac{1}{n(T-1)} \sum_{i=1}^{n} [g_{\Phi i}^2 - \mathrm{E}(g_{\Phi i}^2)] \equiv \sum_{r=1}^{5} H_r \\
= \; & \frac{1}{n(T-1)} \sum_{i=1}^{n} [(\widetilde{V}_{i\cdot}' \xi_{i\cdot})^2 - \mathrm{E}((\widetilde{V}_{i\cdot}' \xi_{i\cdot})^2)] \\
& + \frac{1}{n(T-1)} \sum_{i=1}^{n} [(\widetilde{V}_{i\cdot}' \widetilde{V}_{i\cdot}^{*})^2 - \mathrm{E}((\widetilde{V}_{i\cdot}' \widetilde{V}_{i\cdot}^{*})^2)] \\
& + \frac{2}{n(T-1)} \sum_{i=1}^{n} (\widetilde{V}_{i\cdot}' \xi_{i\cdot})(\widetilde{V}_{i\cdot}' \widetilde{V}_{i\cdot}^{*}) - \frac{2}{n(T-1)} \sum_{i=1}^{n} (1_T' d_{i\cdot})(\widetilde{V}_{i\cdot}' \xi_{i\cdot}) \\
& - \frac{2}{n(T-1)} \sum_{i=1}^{n} [(1_T' d_{i\cdot})(\widetilde{V}_{i\cdot}' \widetilde{V}_{i\cdot}^{*} - \mathrm{E}(\widetilde{V}_{i\cdot}' \widetilde{V}_{i\cdot}^{*}))]
\end{aligned}$$

Now, we have $H_1 = \frac{1}{n(T-1)}\sum_{i=1}^{n}[\xi_{i\cdot}'(\widetilde{V}_{i\cdot}\widetilde{V}_{i\cdot}'-\frac{T-1}{T}\sigma_0^2 r_{n,i}I_T)\xi_{i\cdot}]+\frac{\sigma_0^2}{n(T-1)}\sum_{i=1}^{n}[\xi_{i\cdot}'\frac{T-1}{T}r_{n,i}I_T\xi_{i\cdot}-\mathrm{E}(\xi_{i\cdot}'\frac{T-1}{T}$

$r_{n,i}I_T\xi_{i\cdot})]$. For the first term, let $V_{n,i} = \xi_{i\cdot}'(\widetilde{V}_{i\cdot}\widetilde{V}_{i\cdot}' - \frac{T-1}{T}\sigma_0^2 r_{ni}I_T)\xi_{i\cdot}$. As $\xi_{i\cdot}$ is $\mathcal{F}_{n,i-1}$-measurable, $\mathrm{E}(V_{n,i}|\mathcal{F}_{n,i-1}) = 0$. Thus, $\{V_{n,i}, \mathcal{F}_{n,i}\}$ form a M.D. array. It is easy to see that $\mathrm{E}|V_{n,i}^{1+\epsilon}| \leqslant K_v < \infty$, for some $\epsilon > 0$. Thus, $\{V_{n,i}\}$ is uniformly integrable. Again, conditions of the WLLN for M.D. arrays of Davidson are satisfied, thus, $\frac{1}{n(T-1)}\sum_{i=1}^{n}V_{n,i} \xrightarrow{p} 0$.

For the second term of H_1, note that $\xi_{i\cdot}'\frac{T-1}{T}r_{ni}I_T\xi_{i\cdot} = \sum_t\sum_s \xi_{it}'\frac{T-1}{T}r_{ni}I_{ts}\xi_{is}$, where $\{I_{ts}\} = I_T$. In Lemma 4.2, $\xi_t = \sum_{s=1}^{T}(\Phi_{st}^{u\prime} + \Phi_{ts}^{\ell})\widetilde{V}_s$. We have,

$$\xi_{it} = \sum_{s=1}^{T}\sum_{l=1}^{i-1}(\Phi_{ls,it} + \Phi_{it,ls})\widetilde{V}_{ls} = \sum_{l=1}^{i-1}\sum_{s=1}^{T}(\Phi_{ls,it} + \Phi_{it,ls})\widetilde{V}_{ls} = \sum_{l=1}^{i-1}\phi_{ilt}'\widetilde{V}_{l\cdot},$$

where $\phi_{ilt} = (\Phi_{l\cdot,it}+\Phi_{it,l\cdot})$. Thus, $(\xi_{it})^2-\mathrm{E}[(\xi_{it})^2] = \sum_{l=1}^{i-1}[\phi_{ilt}'(\widetilde{V}_{l\cdot}\widetilde{V}_{l\cdot}'-\frac{T-1}{T}\sigma_0^2 r_{n,i}J_T)\phi_{ilt}]+2\sum_{l=1}^{i-1}\sum_{k=1}^{l-1}\widetilde{V}_{l\cdot}'\phi_{ilt}\phi_{ikt}'\widetilde{V}_{k\cdot}$, where $J_T = I_T - \frac{1}{T}1_T1_T'$. It follows that

$$\begin{aligned}
&\frac{1}{n(T-1)}\sum_{i=1}^{n}\{(\xi_{it})^2 - \mathrm{E}[(\xi_{it})^2]\} \\
= \;& \frac{1}{n(T-1)}\sum_{l=1}^{n-1}\Big\{\sum_{i=l+1}^{n}[\phi_{ilt}'(\widetilde{V}_{l\cdot}\widetilde{V}_{l\cdot}' - \frac{T-1}{T}\sigma_0^2 r_{n,i}J_T)\phi_{ilt}]\Big\} \\
&+2\frac{1}{n(T-1)}\sum_{l=1}^{n-1}\widetilde{V}_{l\cdot}'\Big\{\sum_{i=l+1}^{n}\sum_{k=1}^{l-1}\phi_{ilt}\phi_{ikt}'\widetilde{V}_{k\cdot}\Big\}.
\end{aligned}$$

Clearly, the first term is the 'average' of $n-1$ independent terms, as the second term in the curling brackets is $\mathcal{F}_{n,j-1}$-measurable, therefore it is the 'average' of a M.D. array. Conditions of Theorem 19.7 of Davidson (1994) are easily verified, and hence $\frac{1}{n(T-1)}\sum_{i=1}^{n}\{(\xi_{it})^2-\mathrm{E}[(\xi_{it})^2]\} = o_p(1)$. Similarly, we can show that $\frac{1}{n(T-1)}\sum_{i=1}^{n}\{\xi_{it}\xi_{is}-\mathrm{E}[(\xi_{it}\xi_{is})]\} = o_p(1)$ for $s \neq t$. Since $r_{n,i}$ is uniformly bounded, therefore

$$\frac{\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^{n}[\xi_{i\cdot}'\tfrac{T-1}{T}r_{ni}I_T\xi_{i\cdot} - \mathrm{E}(\xi_{i\cdot}'\tfrac{T-1}{T}r_{ni}I_T\xi_{i\cdot})] = o_p(1), \text{ and } H_1 = o_p(1).$$

The proofs for H_3 and H_4 are similar as the proof for the second term of H_1. The proofs for H_2 and H_5 are similar to the proof of the first part of H_1, as they each involves a sum of n independent terms.

Subsequently, the cross-product term $\frac{1}{n(T-1)}\sum_{i=1}^{n}[g_{\pi i}g_{\Phi i} - \mathrm{E}(g_{\pi i}g_{\Phi i})]$ can be decomposed in a similar manner, and the convergence of each of the decomposed terms can be proved in a similar way.

Proof of Theorem 4.5: Comparing with the proof of theorem 4.4, the proof of showing $\widehat{\Sigma} - \Sigma(\boldsymbol{\theta}_0) \xrightarrow{p} 0$ when T is large is more complicated due to the involvement of a new term that capture the dependence among the elements of $\widetilde{V}_N$ across t. It follows if

(a) $\frac{1}{N}\sum_{j=1}^{N}[\hat{s}_{N,j}\hat{s}'_{N,j} - E(s_{N,j}s'_{N,j})] \xrightarrow{p} 0,$

(b) $\frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[\hat{s}_{N,it}\hat{s}'_{N,is} - \mathrm{E}(s_{N,it}s'_{N,is})] \xrightarrow{p} 0,$

Proof of (a). The proof is similar as the proof of Theorem 4.4. Without loss of generality, we express the terms on the scalar level and work on it. See the proof of Theorem 4.4 for details.

Proof of (b). We prove it by showing $\frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[\hat{s}_{N,it}\hat{s}'_{N,is} - s_{N,it}s'_{N,is}] \xrightarrow{p} 0$, and $\frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[s_{N,it}s'_{N,is} - \mathrm{E}(s_{N,it}s'_{N,is})] \xrightarrow{p} 0$. Due to the consistency of the parameter estimates, the proof of the former is straightforward by applying the mean value theorem. We focus on the proof of the later result. There is a free switch between the index j for the combined unit i and time t for convenience. It suffices to show that

$$\frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[g_{kit}^{\diamond}g_{ris}^{\diamond\prime} - \mathrm{E}(g_{kit}^{\diamond}g_{ris}^{\diamond\prime})] = o_p(1),\ \ k, r = \Pi, \Phi.$$

The proofs proceed similarly as the proof of Theorem 4.4, the weak law of large numbers (WLLN) for M.D. arrays, see, e.g., Davidson(1994, p.299) are widely applied.

First, with $g_{\pi it}^{\diamond} = \Pi'_{N,j}\widetilde{V}_{N,j}$, we have

$$\begin{aligned}
&\frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[g_{\pi it}^{\diamond}g_{\pi is}^{\diamond\prime} - \mathrm{E}(g_{\pi it}^{\diamond}g_{\pi is}^{\diamond\prime})] \\
= \ &\frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[\Pi_{N,it}\Pi'_{N,is}(\widetilde{V}_{N,it}\widetilde{V}_{N,is} - E(\widetilde{V}_{N,it}\widetilde{V}_{N,is}))] \\
= \ &\frac{1}{n}\sum_{i=1}^{n}\{\frac{2}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[\Pi_{N,it}\Pi'_{N,is}(\widetilde{V}_{N,it}\widetilde{V}_{N,is} - E(\widetilde{V}_{N,it}\widetilde{V}_{N,is}))]\} \equiv \frac{1}{n}\sum_{i=1}^{n}P_{n,i}.
\end{aligned}$$

For each t and s, $\widetilde{V}_{N,it}$ and $\widetilde{V}_{N,is}$ are independent over i, and thus $\{P_{n,i}\}$ form an M.D. array. Applying the weak law of large number (WLLN) for MD arrays of Davidson (1994 p.299) leads to $\frac{1}{n}\sum_{i=1}^{n}P_{n,i} \xrightarrow{p} 0$, as $n \to \infty$ and then $T \to \infty$.

Second, with $g^{\diamond}_{\Phi it} = \widetilde{V}_{N,j}\xi_{N,j} + \widetilde{V}_{N,j}\widetilde{V}^{*}_{N,j} - d_{N,j}$, we have

$$
\begin{aligned}
& \tfrac{2}{N}\textstyle\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[g^{\diamond}_{\Phi it}g^{\diamond}_{\Phi is} - \mathrm{E}(g^{\diamond}_{\Phi it}g^{\diamond}_{\Phi is})] \equiv \textstyle\sum_{r=1}^{5} H_r \\
= \; & \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\{\tfrac{2}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[(\widetilde{V}_{N,it}\xi_{N,it})(\widetilde{V}_{N,is}\xi_{N,is}) - \mathrm{E}((\widetilde{V}_{N,it}\xi_{N,it})(\widetilde{V}_{N,is}\xi_{N,is}))]\} \\
& +\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\{\tfrac{2}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[(\widetilde{V}_{N,it}\widetilde{V}^{*}_{N,it})(\widetilde{V}_{N,is}\widetilde{V}^{*}_{N,is}) - \mathrm{E}((\widetilde{V}_{N,it}\widetilde{V}^{*}_{N,it})(\widetilde{V}_{N,is}\widetilde{V}^{*}_{N,is}))]\} \\
& +\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\{\tfrac{4}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[(\widetilde{V}_{N,it}\xi_{N,it})(\widetilde{V}_{N,is}\widetilde{V}^{*}_{N,is})] \\
& -\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\{\tfrac{4}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1} d_{N,it}(\widetilde{V}_{N,is}\widetilde{V}^{*}_{N,is} - \mathrm{E}(\widetilde{V}_{N,is}\widetilde{V}^{*}_{N,is})\} \\
& -\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\{\tfrac{4}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1} d_{N,it}(\widetilde{V}_{N,is}\xi_{N,is})\}
\end{aligned}
$$

We have $H_1 = \frac{1}{n}\sum_{i=1}^{n}\{\frac{2}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[(\widetilde{V}_{N,it}\xi_{N,it})(\widetilde{V}_{N,is}\xi_{N,is}) - \mathrm{E}((\widetilde{V}_{N,it}\xi_{N,it})(\widetilde{V}_{N,is}\xi_{N,is}))]\} = \frac{1}{n}\sum_{i=1}^{n} D_{ni}$. Thus, the result follows if D_{ni} form a M.D. array.

First, we have $D_{ni} = \{\frac{2}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[(\widetilde{V}_{N,it}\xi_{N,it})(\widetilde{V}_{N,is}\xi_{N,is}) - \mathrm{E}((\widetilde{V}_{N,it}\xi_{N,it})(\widetilde{V}_{N,is}\xi_{N,is}))]\}$ $= \{\frac{2}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[(\widetilde{V}_{N,it}\widetilde{V}_{N,is} - \frac{-1}{T}\sigma_0^2 r_{ni})\xi_{N,it}\xi_{N,is}] + \frac{-\sigma_0^2}{T}\{\frac{2}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[r_{ni}(\xi_{N,it}\xi_{N,is} - \mathrm{E}(\xi_{N,it}\xi_{N,is}))]\} = D_{n1,i} + D_{n2,i}$. As $\xi_{i\cdot}$ is $\mathcal{F}_{n,i-1}$-measurable, $\mathrm{E}(D_{n1,i}|\mathcal{F}_{n,i-1}) = 0$. Thus, $\{D_{n1,i}, \mathcal{F}_{n,i}\}$ form a M.D. array. It is easy to see that $\mathrm{E}|D_{n1,i}^{1+\epsilon}| \leqslant K_D < \infty$, for some $\epsilon > 0$. Thus, $\{D_{n1,i}\}$ is uniformly integrable. Again, conditions of the WLLN for M.D. arrays of Davidson are satisfied, thus, $\frac{1}{n}\sum_{i=1}^{n} D_{n1,i} \xrightarrow{p} 0$.

For the second term $D_{n2,i}$, note that in Lemma 4.2, $\xi_t = \sum_{s=1}^{T}(\Phi^{u\prime}_{st} + \Phi^{\ell}_{ts})\widetilde{V}_s$. We have,

$$
\xi_{it} = \textstyle\sum_{s=1}^{T}\sum_{l=1}^{i-1}(\Phi_{ls,it} + \Phi_{it,ls})\widetilde{V}_{ls} = \sum_{l=1}^{i-1}\sum_{s=1}^{T}(\Phi_{ls,it} + \Phi_{it,ls})\widetilde{V}_{ls} = \sum_{l=1}^{i-1}\phi'_{ilt}\widetilde{V}_{l\cdot},
$$

where $\phi_{ilt} = (\Phi_{l\cdot,it} + \Phi_{it,l\cdot})$. Thus, $\xi_{it}\xi_{is} - \mathrm{E}[\xi_{it}\xi_{is}] = \sum_{l=1}^{i-1}[\phi'_{ilt}(\widetilde{V}_{l\cdot}\widetilde{V}'_{l\cdot} - \sigma_0^2 r_{ni}J_T)\phi_{ils}] + \sum_{l=1}^{i-1}\sum_{k=1}^{l-1}$
$\widetilde{V}'_{l\cdot}\phi_{ilt}\phi'_{iks}\widetilde{V}_{k\cdot} + \sum_{l=1}^{i-1}\sum_{k=1}^{l-1}\widetilde{V}'_{l\cdot}\phi_{ikt}\phi'_{ils}\widetilde{V}_{k\cdot}$, where $J_T = I_T - \frac{1}{T}1_T 1'_T$ It follows that

$$
\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{i=1}^{n}[\{\xi_{it}\xi_{is} - \mathrm{E}[\xi_{it}\xi_{is}]\} \\
= \; & \tfrac{1}{n}\textstyle\sum_{l=1}^{n-1}\big\{\sum_{i=l+1}^{n}[\phi'_{ilt}(\widetilde{V}_{l\cdot}\widetilde{V}'_{l\cdot} - \tfrac{T-1}{T}\sigma_0^2 r_{ni}J_T)\phi_{ilt}]\big\} \\
& +\tfrac{1}{n}\textstyle\sum_{l=1}^{n-1}\widetilde{V}'_{l\cdot}\big\{\sum_{i=l+1}^{n}\sum_{k=1}^{l-1}(\phi_{ilt}\phi'_{iks}\widetilde{V}_{k\cdot} + \phi_{ikt}\phi'_{ils}\widetilde{V}_{k\cdot})\big\}.
\end{aligned}
$$

Clearly, the first term is the 'average' of $n-1$ independent terms, as the second term in the curling brackets is $\mathcal{F}_{n,j-1}$-measurable, therefore it is the 'average' of a M.D. array. Conditions of Theorem 19.7 of Davidson (1994) are easily verified, and hence

$\frac{1}{n}\sum_{i=1}^{n}\{\xi_{it}\xi_{is}-\mathrm{E}[(\xi_{it}\xi_{is})]\}=o_p(1)$ for $s\neq t$. Since r_{ni} is uniformly bounded, therefore

$$\frac{2\sigma_0^2}{N}\sum_{i=1}^{n}\{\frac{-1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[r_{ni}(\xi_{N,it}\xi_{N,is}-\mathrm{E}(\xi_{N,it}\xi_{N,is}))]\}=o_p(1), \text{ and } H_1=o_p(1).$$

The proofs for H_3 and H_5 are similar as the proof for the second term of H_1. The proofs for H_2 and H_4 are similar to the proof of the first part of H_1, as they each involves a sum of n independent terms.

Subsequently, the cross-product term $\frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s=1}^{t-1}[g^{\diamond}_{\Pi it}g^{\diamond\prime}_{\Phi is}-\mathrm{E}(g^{\diamond}_{\Pi it}g^{\diamond\prime}_{\Phi is})]$ can be decomposed in a similar manner, and the proofs of convergence of each of the decomposed terms are similar.