For Japanese Journal of Statistics and Data Science
February 2, 2020

Robust Estimation and Inference of Spatial Panel Data Models with Fixed Effects

Shew Fan Liu[1] and Zhenlin Yang[2]

Abstract

It is well established that the quasi maximum likelihood (QML) estimation of the spatial regression models is generally inconsistent under unknown cross-sectional heteroskedasticity (CH) and the CH-robust methods have been developed. The same issue remains for the spatial panel data (SPD) models but the similar studies based on QML approach do not seem to have been carried out. This paper focuses on the SPD model with fixed effects (FE). We argue that under unknown CH the QML estimator for the SPD-FE model is inconsistent in general, but there are 'special cases' where it may remain consistent although the exact conditions may not be possible to check, as in practice the type of CH is generally unknown. Thus, we introduce a new set of estimation and inference methods based on the *adjusted quasi scores* (AQS), which are fully robust against unknown CH. Consistency and asymptotic normality of the proposed AQS estimators are established. Robust standard error estimates are provided and their consistency is proved. To improve the finite sample performance, a set of AQS methods based on concentrated quasi scores is also introduced and its asymptotic properties examined. Extensive Monte Carlo results show that the new estimator outperforms the QML estimator even when the latter seems robust.

Key Words: Spatial dependence; Spatial panel data; Fixed effects; Unknown heteroskedasticity; Non-normality, AQS estimator; Robust standard error.

1. Introduction

Exploring how correlation in space extends to and interacts over time is a long standing question since the onset of the literature relating to spatial econometrics such as Anselin (1988). Spatial panel data (SPD) models have the versatility of allowing a location related dependence structure to be attached to the conventional panel model in terms of spatial dependence or spatial heterogeneity (Anselin et al., 2008). With a fast evolving literature (see surveys in Lee and Yu, 2010b, 2015), panel models with fixed effects (FE) and spatial or social interactions remain popular due to its wide practical applicability. Examples of recent empirical studies include Baltagi et al. (2016), Hsieh & Lee (2014), Kelejian & Piras (2016), and Millimet & Roy (2016). In this paper, we consider SPD models with FE and cross-sectional heteroskedasticity (CH) of unknown form, where a spatial autoregressive (SAR) process is built on both dependent variable and disturbance term, and introduce CH-robust estimation and inference methods.

Zhenlin Yang `Email:` zlyang@smu.edu.sg
[1]School of Economics, Singapore Management University, Singapore
[2]School of Economics, Singapore Management University, Singapore

SPD models with homoskedastic disturbances have been well studied (see, e.g., Baltagi et al. 2003, 2013; Baltagi and Yang 2013; Fingleton 2008; Kelejian and Prucha 2007; Lee & Yu 2010a, 2012; Robinson & Rossi 2015; and Yang et al. 2016). The literature on CH-robust estimation of cross-sectional spatial models is fairly comprehensive as well; See LeSage (1997) for Bayesian estimation; Badinger & Egger (2011), Lin & Lee (2010), and Kelejian & Prucha (2010) for GMM estimation; and Jin & Lee (2012), and Liu & Yang (2015) for QML based estimation. However, the study on SPD models with unknown CH has been limited to Moscone and Tosetti (2011) who extend the robust GMM estimation methods for a cross-sectional spatial model, given in Kelejian & Prucha (2010) and Lin & Lee (2010), to the SPD framework where they consider only spatial error dependence, and Badinger and Egger (2015) who consider CH-robust 2SLS estimation of a higher order spatial panel model by extending the methods in Kapoor et al. (2007), where the individual-specific effects are treated using the Mundlak (1978) approach. However, the individual-specific effects may correlate with time-varying regressors in an arbitrary manner and in this case they have to be treated as fixed parameters. Also, the 2SLS estimator may lack efficiency compared to a general GMM or an ML-based estimator, as it focuses only on the deterministic part of the model based on linear moments and ignores the reduced form model that incorporates the information contained in the disturbances.

Since ML methods provide the most efficient estimates, QML-based methods may also provide more efficient estimates compared with GMM and 2SLS methods, in particular the latter. Therefore, QML-based methods for FE-SPD models that are simple to implement and robust to unknown CH would be very useful. When the disturbances are **homoskedastic**, Lee and Yu (2010a) show that a direct QML estimation yields consistent estimators for all parameters in the FE-SPD model (including the FE parameters), when the number of spatial units (n) and time periods (T) are both large. When T is fixed, the QML estimators (QMLEs) for error variance and FEs are inconsistent. Upon transformation of the model to wipe out the FEs, QMLEs of all the structural parameters become consistent irrespective of the size of T. However, Lee and Yu (2010a) does not consider unknown CH. This paper aims to fill this gap in the literature.

In a cross-sectional SAR model with unknown CH, Lin and Lee (2010) show that the usual QMLE of the spatial parameter is inconsistent in general. A similar phenomenon is observed in the FE-SPD models. We argue that under unknown CH the QMLEs for the SPD-FE model are inconsistent in general, but there are 'special cases' where they may remain consistent although the exact conditions may not be possible to check, as in practice the type of CH is generally unknown. We therefore propose a new set of estimation and inference methods based on the *adjusted quasi scores* (AQS), fully robust against unknown CH. Consistency and asymptotic normality of the proposed AQS estimators (AQSEs) are established. To conduct CH-robust inferences, we propose an *outer-product-of-martingale-difference* (OPMD) method for estimating the variance-covariance matrix of the AQSEs, first under normality, and then generalized to allow for non-normality. Consistency of this OPMD-based estimate is also established. To capture the extra variability coming from the estimation of the regression coefficients and the average of error variance, a set of AQS methods based on *concentrated quasi score* is also proposed, which may offer finite-sample improvements. The AQS estimation is easy to implement and is effective in attaining consistency under unknown CH while limiting the compromise on efficiency of the usual QMLE. The AQS estimation and inference generally perform very well under CH, but the regular QML estimation and inference do not, even when they are valid under CH, as demonstrated by the extensive Monte Carlo results.

AQS estimators broadly fall into the umbrella of estimators known as M-estimators in the literature, which can be either a maxima of an objective function or a root of an estimating equation. The proposed robust estimator falls into the latter which is also known as the Z(ero)-estimator in van der Vaart (1998). Very interestingly, this idea finds its root in Neyman and Scott (1948) on *Modified Equations of Maximum Likelihood*, but it was only recently that the idea was picked up by Baltagi and Yang (2013) to give CH-robust LM tests for spatial dependence, Liu and Yang (2015) to give CH-robust estimation of spatial cross-sectional model, and Yang (2018) to give initial condition free estimation of spatial dynamic panel data (SDPD) models with FE, which is extended by Li and Yang (2020) to allow for unknown CH, Xu and Yang (2020) to give tests for temporal heterogeneity in FE-SPD models, and Li and Yang (2019) to give initial condition free estimation of SDPD models with correlated random effects.

The rest of the paper is organized as follows. Section 2 outlines the transformation-based QML estimation of the FE-SPD model and examines its robustness. Section 3 introduces the CH-robust AQS estimators for the SPD model with individual FE, presents asymptotic properties and introduces CH-robust inference methods. Section 4 extends the AQS methods to the SPD model with both individual and time FE. Section 5 presents the Monte Carlo results. Section 6 concludes the paper. All technical details are given in Appendix B.

2. QML Estimation of FE-SPD Model and its Robustness

The spatial panel data (SPD) model with individual and time specific fixed effects (FE), containing a spatial autoregressive (SAR) process in responses and a SAR process in errors, called the FE-SPD model in this paper, has the form:

$$Y_{nt} = \lambda_0 W_{1n} Y_{nt} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt}, \quad U_{nt} = \rho_0 W_{2n} U_{nt} + V_{nt}, \quad t = 1, \dots, T, \tag{2.1}$$

where $Y_{nt} = (y_{1t}, y_{2t}, \dots, y_{nt})'$ is an $n \times 1$ vector of observations on the responses, X_{nt} is an $n \times k$ matrix containing the values of k non-stochastic but time varying regressors, $V_{nt} = (v_{1t}, v_{2t}, \dots, v_{nt})'$ is the vector of idiosyncratic errors, β_0 is a $k \times 1$ vector of regression coefficients, λ_0 and ρ_0 are the spatial lag and error parameters, W_{1n} and W_{2n} are the respective $n \times n$ non-stochastic spatial weights matrices, $\mathbf{c}_{n0}$ is the $n \times 1$ time invariant vector of individual-specific FE, and $\{\alpha_{t0}\}$ are the time-specific FE with l_n being an $n \times 1$ vector of ones.

The fixed effects in a panel data model induce the so-called *incidental parameters problem* of Neyman and Scott (1948). The existence of unknown heteroskedasticity might induce another set of incidental parameters. The standard way of dealing with FE problem is to eliminate the FE by some transformation, such as first-difference, demean, and orthonormal transformation. However, there seems no standard solution to the problem of unknown heteroskedasticity.

In this section, we first outline the transformation-based QML estimator (QMLE) of the FE-SPD model, where the idiosyncratic errors $\{v_{it}\}$ are first set to be independent and identically distributed (iid) with mean 0 and variance σ_0^2, as in Lee and Yu (2010a). Then, we examine the properties of the QMLE when the errors are independent but with unknown cross-sectional heteroskedasticity (CH). We show that under unknown CH, the necessary conditions for QMLE to be consistent can be violated and therefore QMLE cannot be consistent in general.

Notation. Some notation and convention would be helpful in the theoretical developments. Let $\mathtt{tr}(\cdot)$, $|\cdot|$, and $\|\cdot\|$ be, respectively, the trace, determinant and Frobenius norm of a square matrix. The operator $\mathtt{diag}(\cdot)$ forms a diagonal matrix based on a vector or the diagonal elements of a square matrix, and $\mathtt{diagv}(\cdot)$ forms a column vector by the diagonal elements of a square matrix. Let $\theta_0 = (\beta_0', \sigma_0^2, \lambda_0, \rho_0)'$ be the **true** parameter vector and $\theta = (\beta', \sigma^2, \lambda, \rho)'$ be any value of it. The usual expectation, variance and covariance operators, $\mathrm{E}(\cdot)$, $\mathrm{Var}(\cdot)$, and $\mathrm{Cov}(\cdot)$, correspond to θ_0. However, for two non-stochastic vectors a and b of the same length, $\mathrm{Var}(a)$ denotes the sample variance of a, and $\mathrm{Cov}(a, b)$ the sample covariance between a and b.

2.1. The one-way FE-SPD model

Consider first the SPD model with only individual-specific FE (FE_1), i.e., dropping α_t from (2.1). For an identity matrix I_T and a vector of ones l_T, let $J_T = I_T - \frac{1}{T} l_T l_T'$, the *time demean* operator, which is idempotent with rank $T-1$, and thus has $T-1$ eigenvalues of 1 and one eigenvalue of 0. Let $F_{T,T-1}$ be the first $T-1$ eigenvectors of J_T corresponding to eigenvalue 1. The last eigenvector is $\frac{1}{\sqrt{T}} l_T$, orthogonal to $F_{T,T-1}$. Now, for an $n \times T$ matrix $[Z_{n1}, \ldots, Z_{nT}]$, defined $[Z_{n1}^*, \ldots, Z_{n,T-1}^*] = [Z_{n1}, \ldots, Z_{nT}] F_{T,T-1}$. We have the transformed FE_1-SPD model:

$$Y_{nt}^* = \lambda_0 W_{1n} Y_{nt}^* + X_{nt}^* \beta_0 + U_{nt}^*, \; U_{nt}^* = \rho_0 W_{2n} U_{nt}^* + V_{nt}^*, \;\; t = 1, \ldots, T-1, \tag{2.2}$$

where the individual-specific fixed effects $\mathbf{c}_{n0}$ are transformed away and the effective sample size post transformation is $N = n(T-1)$. Stack the transformed vectors to give $\mathbf{Y}_N = (Y_{n1}^{*\prime}, \ldots, Y_{n,T-1}^{*\prime})'$, similarly $\mathbf{U}_N$ and $\mathbf{V}_N$, $\mathbf{X}_{Nj} = (X_{jn,1}^{*\prime}, \ldots, X_{jn,T-1}^{*\prime})'$ for the jth regressor and $\mathbf{X}_N = [\mathbf{X}_{1N}, \ldots, \mathbf{X}_{kN}]$. Let $\mathbf{W}_{rN} = I_{T-1} \otimes W_{rn}, \; r = 1, 2$, where $\otimes$ denotes the kronecker product. Model (2.2) is written as,

$$\mathbf{Y}_N = \lambda_0 \mathbf{W}_{1N} \mathbf{Y}_N + \mathbf{X}_N \beta_0 + \mathbf{U}_N, \quad \mathbf{U}_N = \rho_0 \mathbf{W}_{2N} \mathbf{U}_N + \mathbf{V}_N. \tag{2.3}$$

The transformed errors, $\{v_{it}^*\}$, are iid $N(0, \sigma_0^2)$ if the original errors, $\{v_{it}\}$, are iid $N(0, \sigma_0^2)$, as $\mathbf{V}_N = (F_{T,T-1}' \otimes I_n)(V_{n1}', \ldots, V_{nT}')'$ and $\mathrm{E}(\mathbf{V}_N \mathbf{V}_N') = \sigma_0^2 (F_{T,T-1}' \otimes I_n)(F_{T,T-1} \otimes I_n) = \sigma_0^2 I_N$.

The (quasi) Gaussian loglikelihood of θ, *as if* $\{v_{it}^*\}$ are iid $N(0, \sigma_0^2)$, is,

$$\ell_N(\theta) = -\frac{N}{2} \ln(2\pi\sigma^2) + \ln|\mathbf{A}_{1N}(\lambda)| + \ln|\mathbf{A}_{2N}(\rho)| - \frac{1}{2\sigma^2} \mathbf{V}_N'(\beta, \delta) \mathbf{V}_N(\beta, \delta), \tag{2.4}$$

where $\mathbf{V}_N(\beta, \delta) = \mathbf{A}_{2N}(\rho)[\mathbf{A}_{1N}(\lambda)\mathbf{Y}_N - \mathbf{X}_N \beta]$, $\mathbf{A}_{1N}(\lambda) = I_N - \lambda \mathbf{W}_{1N}$, $\mathbf{A}_{2N}(\rho) = I_N - \rho \mathbf{W}_{2N}$, and $\delta = (\lambda, \rho)'$. Given δ, $\ell_N(\theta)$ is partially maximized at:

$$\tilde{\beta}_N(\delta) = [\mathbb{X}_N'(\rho)\mathbb{X}_N(\rho)]^{-1}\mathbb{X}_N'(\rho)\mathbb{Y}_N(\delta) \;\text{ and }\; \tilde{\sigma}_N^2(\delta) = \tfrac{1}{N}\mathbb{Y}_N'(\delta)\mathbb{M}_N(\rho)\mathbb{Y}_N(\delta), \tag{2.5}$$

where $\mathbb{Y}_N(\delta) = \mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{Y}_N$, $\mathbb{M}_N(\rho) = I_N - \mathbb{X}_N(\rho)[\mathbb{X}_N'(\rho)\mathbb{X}_N(\rho)]^{-1}\mathbb{X}_N'(\rho)$, and $\mathbb{X}_N(\rho) = \mathbf{A}_{2N}(\rho)\mathbf{X}_N$. The concentrated quasi Gaussian loglikelihood function of δ is, upon substitution:

$$\ell_N^c(\delta) = -\tfrac{N}{2}(\ln(2\pi) + 1) + \ln|\mathbf{A}_{1N}(\lambda)| + \ln|\mathbf{A}_{2N}(\rho)| - \tfrac{N}{2}\ln\tilde{\sigma}_N^2(\delta). \tag{2.6}$$

Maximizing (2.6) gives the unconstrained QMLE $\hat{\delta}_{\texttt{QML1}}$ of δ, and thus the unconstrained QMLEs of β and σ^2: $\hat{\beta}_{\texttt{QML1}} \equiv \tilde{\beta}_N(\hat{\delta}_{\texttt{QML1}})$ and $\hat{\sigma}^2_{\texttt{QML1}} \equiv \tilde{\sigma}^2_N(\hat{\delta}_{\texttt{QML1}})$. Under the assumptions that the errors are iid and some additional regularity conditions, Lee and Yu (2010a) show that $\hat{\theta}_{\texttt{QML1}} = (\hat{\beta}'_{\texttt{QML1}}, \hat{\delta}'_{\texttt{QML1}}, \hat{\sigma}^2_{\texttt{QML1}})'$ is $\sqrt{N}$-consistent and asymptotically normal.

To examine the **robustness** of the transformation-based QMLE of the FE$_1$-SPD model, consider the quasi score function derived from (2.4) under homoskedasticity assumption:

$$\frac{\partial}{\partial\theta}\ell_N(\theta) = \begin{cases} \frac{1}{\sigma^2}\mathbb{X}_N'(\rho)\mathbf{V}_N(\beta,\delta), \\ \frac{1}{2\sigma^4}\left[\mathbf{V}_N'(\beta,\delta)\mathbf{V}_N(\beta,\delta) - N\sigma^2\right], \\ \frac{1}{\sigma^2}\mathbf{V}_N'(\beta,\delta)\mathbf{A}_{2N}(\rho)\mathbf{W}_{1N}\mathbf{Y}_N - \mathrm{tr}(\mathbf{G}_{1N}(\lambda)), \\ \frac{1}{\sigma^2}\mathbf{V}_N'(\beta,\delta)\mathbf{G}_{2N}(\rho)\mathbf{V}_N(\beta,\delta) - \mathrm{tr}(\mathbf{G}_{2N}(\rho)), \end{cases} \tag{2.7}$$

where $\mathbf{G}_{1N}(\lambda) = \mathbf{W}_{1N}\mathbf{A}_{1N}^{-1}(\lambda)$ and $\mathbf{G}_{2N}(\rho) = \mathbf{W}_{2N}\mathbf{A}_{2N}^{-1}(\rho)$.

Suppose now the errors are independent but not identically distributed (inid), i.e., $v_{it} \sim inid(0, \sigma_0^2 h_i)$, where $\frac{1}{n}\sum_{i=1}^n h_i = 1$ and $h_i > 0$ so that σ_0^2 represents the average of $\mathrm{Var}(v_{it})$ over i for any t. A necessary condition for the consistency of an extremum estimator is that the probability limit of the average objective function at the true parameter is zero. In the present QMLE case, this surmounts to, $\mathrm{plim}_{N\to\infty}\frac{1}{N}\frac{\partial}{\partial\theta}\ell_N(\theta_0) = 0$ (Amemiya, 1985). It is evident that this condition is still satisfied by the β- and σ^2-components under unknown CH. However, it may not be always true for the λ- and ρ-components. Let $\mathbf{h}_n = (h_1, \ldots, h_n)'$, $H_n = \mathtt{diag}(\mathbf{h}_n)$ and $\mathbf{H}_N = I_{T-1} \otimes H_n$. Note that under CH, $\mathrm{Var}(\mathbf{V}_N) = \sigma_0^2\mathbf{H}_N$. It is easy to see that,

$$\begin{aligned} \frac{\partial}{\partial\lambda}\frac{1}{N}\ell_N(\theta_0) &= \frac{1}{N\sigma_0^2}\mathbf{V}_N'\bar{\mathbf{G}}_{1N}\mathbf{V}_N - \frac{1}{N}\mathtt{tr}(\mathbf{G}_{1N}) + o_p(1) \\ &= \frac{1}{N\sigma_0^2}\mathbf{V}_N'\left(\bar{\mathbf{G}}_{1N} - \frac{1}{N}\mathtt{tr}(\mathbf{G}_{1N})I_N\right)\mathbf{V}_N + o_p(1) \\ &= \frac{1}{N}\left(\mathtt{tr}(\mathbf{H}_N\bar{\mathbf{G}}_{1N}) - \frac{1}{N}\mathtt{tr}(\bar{\mathbf{G}}_{1N})\mathtt{tr}(\mathbf{H}_N)\right) + o_p(1) \\ &= \mathrm{Cov}(\bar{g}_{1n}, \mathbf{h}_n) + o_p(1), \\ \frac{\partial}{\partial\rho}\frac{1}{N}\ell_N(\theta_0) &= \frac{1}{N\sigma_0^2}\mathbf{V}_N'\mathbf{G}_{2N}\mathbf{V}_N - \frac{1}{N}\mathtt{tr}(\mathbf{G}_{2N}) + o_p(1) \\ &= \frac{1}{N\sigma_0^2}\mathbf{V}_N'\left(\mathbf{G}_{2N} - \frac{1}{N}\mathtt{tr}(\mathbf{G}_{2N})I_N\right)\mathbf{V}_N + o_p(1) \\ &= \frac{1}{N}\left(\mathtt{tr}(\mathbf{H}_N\mathbf{G}_{2N}) - \frac{1}{N}\mathtt{tr}(\mathbf{G}_{2N})\mathtt{tr}(\mathbf{H}_N)\right) + o_p(1) \\ &= \mathrm{Cov}(g_{2n}, \mathbf{h}_n) + o_p(1), \end{aligned}$$

where $\bar{\mathbf{G}}_{1N} = I_{T-1} \otimes \bar{G}_{1n}$, $\mathbf{G}_{2N} = I_{T-1} \otimes G_{2n}$, $\bar{g}_{1n} = \mathtt{diagv}(\bar{G}_{1n})$ and $g_{2n} = \mathtt{diagv}(G_{2n})$;

$\bar{G}_{1n} = A_{2n}G_{1n}A_{2n}^{-1}$, $G_{1n} = W_{1n}A_{1n}^{-1}$ and $G_{2n} = W_{2n}A_{2n}^{-1}$; $A_{1n} \equiv A_{1n}(\lambda_0) = I_n - \lambda_0 W_{1n}$ and $A_{2n} \equiv A_{2n}(\rho_0) = I_n - \rho_0 W_{2n}$. It follows that $\text{plim}_{N\to\infty}\frac{1}{N}\frac{\partial}{\partial\theta}\ell_N(\theta_0) = 0$ if,

$$\text{Cov}(\bar{g}_{1n}, \mathbf{h}_n) \to 0 \;\text{ and }\; \text{Cov}(g_{2n}, \mathbf{h}_n) \to 0. \tag{2.8}$$

Therefore, (2.8) constitutes two **necessary conditions** for QMLEs to remain consistent. Obviously, these conditions would hold if (i) $\text{Var}(\bar{g}_{1n}) \to 0$ and $\text{Var}(g_{2n}) \to 0$, where $\bar{g}_{1n}$ and g_{2n} relate to the spatial layouts, or (ii) CH, $\mathbf{h}_n$, arises due to reasons unrelated to spatial layouts. See Liu and Yang (2015, Sec. 2.2) for a detailed discussion. Assessing whether these conditions are satisfied in practice may not be feasible as CH is of unknown form, and thus makes the QML estimator rather unappealing when CH is suspected.

Furthermore, for $\hat{\delta}_{\texttt{QML1}}$ to be consistent under the unknown CH, **sufficient conditions** (van der Vaart 1998, Theorem 5.7) are much more than those in (2.8): $\sup_{\delta\in\Delta}\frac{1}{N}|\ell_N^c(\delta) - \bar{\ell}_N^c(\delta)| \xrightarrow{p} 0$ and $\sup_{\delta:d(\delta,\delta_0)\geq\varepsilon}\bar{\ell}_N^c(\delta) < \bar{\ell}_N^c(\delta_0)$ for every $\varepsilon > 0$, where $\bar{\ell}_N^c(\delta) = \max_{\beta,\sigma^2}\text{E}[\ell_N(\theta)]$ and $d(\delta,\delta_0)$ is a measure of distance between δ and δ_0. The latter condition, called the *identification uniqueness condition*, boils down to the following two conditions:

Condition I. *Either (a)* $\lim_{N\to\infty}\frac{1}{N}\{\mathbf{X}_N, \mathbf{G}_{1N}\mathbf{X}_N\beta_0\}'\mathbf{A}_{2N}'(\rho)\mathbf{A}_{2N}(\rho)\{\mathbf{X}_N, \mathbf{G}_{1N}\mathbf{X}_N\beta_0\}$ *is nonsingular* $\forall\rho$, *and* $\lim_{n\to\infty}\frac{1}{n}(\ln|\sigma_0^2 A_{2n}'^{-1}A_{2n}^{-1}| - \ln|\sigma_n^2(\lambda_0,\rho)A_{2n}'^{-1}(\rho)A_{2n}^{-1}(\rho)|) \neq 0$ *for* $\rho \neq \rho_0$,

or (b) $\lim_{n\to\infty}\frac{1}{n}(\ln|\sigma_0^2 D_n'^{-1}D_n^{-1}| - \ln|\sigma_n^2(\delta)D_n'^{-1}(\delta)D_n^{-1}(\delta)|) \neq 0$ *for* $\delta \neq \delta_0$,
where $D_n(\delta) = A_{2n}(\rho)A_{1n}(\lambda)$, $D_n \equiv D_n(\delta_0)$, *and* $\sigma_n^2(\delta) = \frac{1}{n}\sigma_0^2\texttt{tr}(H_n D_n'^{-1}D_n'(\delta)D_n(\delta)D_n^{-1})$;

Condition II. $\lim_{n\to\infty}\text{Cov}[\texttt{diagv}(A_n), \texttt{diagv}(H_n)] = 0$, *for* $A_n = \bar{G}_{1n}$, G_{2n}, $\bar{G}_{1n}'\bar{G}_{1n}$, $G_{2n}'G_{2n}$, $G_{2n}\bar{G}_{1n}$, $\bar{G}_{1n}'G_{2n}'G_{2n}\bar{G}_{1n}$, $G_{2n}'G_{2n}\bar{G}_{1n}$, $G_{2n}'\bar{G}_{1n}$, *and* $\bar{G}_{1n}'G_{2n}\bar{G}_{1n}$.

Condition I extends the identification uniqueness conditions of Lee and Yu (2010a) to allow for unknown CH and guarantees that $\limsup_{N\to\infty}\frac{1}{N}[\bar{\ell}_N^c(\delta) - \bar{\ell}_N^c(\delta_0)] \neq 0$, and Condition II extends those given in (2.8) and guarantees that it is less than or equal zero. Moreover, the uniform convergence, $\sup_{\delta\in\Delta}\frac{1}{N}|\ell_N^c(\delta) - \bar{\ell}_N^c(\delta)| \xrightarrow{p} 0$, also requires Condition II. Appendix B (the beginning part) provides some details on how these conditions arise.

Clearly, these conditions cannot be met in general, and even if some key conditions are met, e.g., those in (2.8), it is difficult to verify the remaining. One such a situation may be when spatial layouts are contiguity-based such as Rook, Queen, and group interactions, where the number of neighbors for each spatial unit 'does not vary much', or the unknown CH depends only on the exogenous regressors (see Liu and Yang 2015, and Monte Carlo results in this paper). However, practical applications often use spatial weight matrices constructed base on 'economic' or 'financial' distances, and in these cases even the necessary conditions (2.8) might be violated. Furthermore, this simple solution may not extend to the more general two-way FE model as discussed below. A more general approach is therefore called for.

2.2. The two-way FE-SPD model

When T is small, the above discussions extend in a straightforward manner to the SDP model with two-way FE (FE_2-SPD), by adding the time-specific FE, $\{\alpha_t\}_{t=1}^T$ into the model in the form of dummy variables. When T is large, however, $\{\alpha_t\}_{t=1}^T$ constitute another set of incidental parameters and it is customary to apply another transformation to eliminate them.

Let $F_{n,n-1}$ be the first $n-1$ eigenvectors of the *individual demean* operator, $J_n = I_n - \frac{1}{n} l_n l_n'$. Lee and Yu (2010a) show that this transformation is valid as long as the spatial weights matrices are row-normalized (i.e., each row sums to 1), since it ensures $J_n W_{rn} = J_n W_{rn} J_n$. By *Spectral Theorem*, $J_n = F_{n,n-1} F_{n,n-1}'$. Now, for an $n \times T$ matrix $[Z_{n1}, \cdots, Z_{nT}]$, define its transformed version as $[Z_{n1}^*, \ldots, Z_{n,T-1}^*] = F_{n,n-1}'[Z_{n1}, \ldots, Z_{nT}] F_{T,T-1}$. This gives the transformed variables (upon stacking): $\mathbf{Y}_N = (Y_{n1}^{*\prime}, \ldots, Y_{n,T-1}^{*\prime})'$, $\mathbf{U}_N = (U_{n1}^{*\prime}, \ldots, U_{n,T-1}^{*\prime})'$, $\mathbf{V}_N = (V_{n1}^{*\prime}, \ldots, V_{n,T-1}^{*\prime})'$, $\mathbf{X}_{jN} = (X_{jn,1}^{*\prime}, \ldots, X_{jn,T-1}^{*\prime})'$, for the jth regressor, $j = 1, \ldots, k$, and $\mathbf{X}_N = \{\mathbf{X}_{1N}, \ldots, \mathbf{X}_{kN}\}$. Define $\mathbf{W}_{rN} = I_{T-1} \otimes W_{rn}^*$, where $W_{rn}^* = F_{n,n-1}' W_{rn} F_{n,n-1}$. We have the following transformed FE$_2$-SPD model, identical in form to Model (2.3):

$$\mathbf{Y}_N = \lambda_0 \mathbf{W}_{1N} \mathbf{Y}_N + \mathbf{X}_N \beta_0 + \mathbf{U}_N, \quad \mathbf{U}_N = \rho_0 \mathbf{W}_{2N} \mathbf{U}_N + \mathbf{V}_N. \tag{2.9}$$

The effective sample size now becomes $N = (n-1) \times (T-1)$. It is easy to see that $\mathbf{V}_N = (F_{T,T-1}' \otimes F_{n,n-1}')(V_{n1}', \ldots, V_{nT}')'$. Then, $E(\mathbf{V}_N \mathbf{V}_N') = \sigma^2 (F_{T,T-1}' \otimes F_{n,n-1}')(F_{T,T-1} \otimes F_{n,n-1}) = \sigma^2 I_N$ under homoskedasticity. Hence, $\{v_{it}^*\}$ are iid $N(0, \sigma^2)$ if the original errors $\{v_{it}\}$ are iid $N(0, \sigma^2)$. Given the similarity between (2.9) and (2.3), QML estimation proceeds in the same way. When $\{v_{it}\}$ are iid but may not be normal, Lee and Yu (2010a) show that, under some regularity conditions, the resulting QMLE $\hat\theta_{\mathtt{QML2}} = (\hat\beta_{\mathtt{QML2}}', \hat\delta_{\mathtt{QML2}}', \hat\sigma_{\mathtt{QML2}}^2)'$ is $\sqrt{N}$-consistent and asymptotically normal. Finally, for simplifications in calculating the determinant terms in the concentrated loglikelihood functions, see Lee & Yu (2010a) and Griffith (1988).

Robustness of QMLE of FE$_2$-SPD model. When T is also large, the results above for the FE$_1$-SPD model are invalid as $\{\alpha_t\}_{t=1}^T$ induce another set of incidental parameters. While the transformed FE$_2$-SPD model given in (2.9) takes an identical form as the transformed FE$_1$-SPD model given in (2.3), and the corresponding quantities also take the same forms as those given in equations (2.5), written in terms of the new transformed variables, the major difference is that in the presence of unknown CH the transformed errors in the FE$_2$-SPD model are no longer uncorrelated across i as seen below,

$$E(\mathbf{V}_N \mathbf{V}_N') = \sigma_0^2 (F_{T,T-1}' \otimes F_{n,n-1}')(I_T \otimes H_n)(F_{T,T-1} \otimes F_{n,n-1}) = \sigma_0^2 (I_T \otimes H_n^*),$$

where $H_n^* = (F_{n,n-1}' H_n F_{n,n-1})$, which no longer is a diagonal matrix. In addition, the two necessary conditions for the QMLEs of the FE$_2$-SPD model to be robust against CH become:

$$\tfrac{1}{n}(\mathtt{tr}(H_n^* \bar G_{1n}) - \tfrac{1}{n}\mathtt{tr}(\bar G_{1n})\mathtt{tr}(H_N^*)) \to 0 \ \text{ and } \ \tfrac{1}{n}(\mathtt{tr}(H_n^* G_{2n}) - \tfrac{1}{n}\mathtt{tr}(G_{2n})\mathtt{tr}(H_n^*)) \to 0,$$

which are even more difficult to verify and more unlikely to be satisfied in practical applications compared to (2.8). Therefore, it may not be of practical interest to pursue further in this direction. However, the study of this section sends a clear message: the standard QML estimation is not robust against unknown CH in general and effort should be diverted to the development of new estimation and inference methods that are generally robust against unknown CH.

3. Robust Estimation and Inference for FE$_1$-SPD Model

Note that CH of completely unknown form may induce another set of incidental parameters besides the fixed effects and this problem is even more profound for smaller T. In case of classical linear regression, it posts no problem in terms of point estimation, but does cause problem on standard error estimation which generates a series of works spurred by White (1980). In cases of spatial econometric models or models containing 'non-linear' structural parameters, it causes problems on both point estimation and inference. Developing a general method to solve these problems, for the SPD model with individual-specific FE, is the focus of this section.

3.1. The adjusted quasi score method

We propose an *adjusted quasi score* (AQS) method for estimating the common parameters in the FE-SPD model, by adjusting the joint quasi score function of θ. Following the notation of Sec. 2.1, the quasi score function, $\mathbf{S}_N(\theta) = \frac{\partial}{\partial\theta}\ell_N(\theta)$ given in (2.7), can be written at θ_0 as:

$$\mathbf{S}_N(\theta_0) = \begin{cases} \frac{1}{\sigma_0^2}\mathbb{X}_N'\mathbf{V}_N, \\ \frac{1}{2\sigma_0^4}(\mathbf{V}_N'\mathbf{V}_N - N\sigma_0^2), \\ \frac{1}{\sigma_0^2}\mathbf{V}_N'(\boldsymbol{\eta}_N + \bar{\mathbf{G}}_{1N}\mathbf{V}_N) - \mathrm{tr}(\bar{\mathbf{G}}_{1N}), \\ \frac{1}{\sigma_0^2}\mathbf{V}_N'\mathbf{G}_{2N}\mathbf{V}_N - \mathtt{tr}(\mathbf{G}_{2N}), \end{cases} \tag{3.1}$$

where $\boldsymbol{\eta}_N = \bar{\mathbf{G}}_{1N}\mathbb{X}_N\beta_0$. As evident from (3.1), the main cause of inconsistency of the QMLEs may be the score elements with respect to the spatial parameters, which fail to reach the desired probability limit of zero under CH. As such, one could naturally look at adjustments to these score components by brute force so that the resulting AQS functions become unbiased and have the desired probability limits under unknown CH. From (3.1), it is clear that these can be achieved by replacing $\bar{\mathbf{G}}_{1N}$ by $\bar{\mathbf{G}}_{1N}^\circ = \bar{\mathbf{G}}_{1N} - \mathtt{diag}(\bar{\mathbf{G}}_{1N})$ and $\mathbf{G}_{2N}$ by $\mathbf{G}_{2N}^\circ = \mathbf{G}_{2N} - \mathtt{diag}(\mathbf{G}_{2N})$:

$$\psi_N(\theta) = \begin{cases} \frac{1}{\sigma^2}\mathbb{X}_N'(\rho)\mathbf{V}_N(\beta,\delta), \\ \frac{1}{2\sigma^4}[\mathbf{V}_N'(\beta,\delta)\mathbf{V}_N(\beta,\delta) - N\sigma^2], \\ \frac{1}{\sigma^2}\mathbf{V}_N'(\beta,\delta)[\boldsymbol{\eta}_N(\beta,\delta) + \bar{\mathbf{G}}_{1N}^\circ(\delta)\mathbf{V}_N(\beta,\delta)], \\ \frac{1}{\sigma^2}\mathbf{V}_N'(\beta,\delta)\mathbf{G}_{2N}^\circ(\rho)\mathbf{V}_N(\beta,\delta), \end{cases} \tag{3.2}$$

to give an AQS function $\psi_N(\theta)$ with the desired property: $\mathrm{E}[\psi_N(\theta_0)] = 0$ and $\mathrm{plim}_{N\to\infty}\frac{1}{N}\psi_N(\theta_0) = 0$ under unknown CH. The AQS estimator (AQSE) of the structural parameters θ is thus

$$\hat{\theta}_{\mathtt{AQS1}} = \arg\{\psi_N(\theta) = 0\}. \tag{3.3}$$

The root-finding process can be simplified by first concentrating out β and σ^2 from $\psi_N(\theta)$ using $\tilde{\beta}_N(\delta)$ and $\tilde{\sigma}_N^2(\delta)$ given in (2.5) (the constrained QMLEs and AQSEs of β and σ^2 are the same), and then solving the concentrated AQS equations to give $\hat{\delta}_{\mathtt{AQS1}} = \arg\{\tilde{\psi}_N^c(\delta) = 0\}$, where

$$\tilde{\psi}_N^c(\delta) = \begin{cases} \frac{1}{\tilde{\sigma}_N^2(\delta)} \mathbf{V}_N'(\tilde{\beta}_N(\delta),\delta)[\boldsymbol{\eta}_N(\tilde{\beta}_N(\delta),\delta) + \bar{\mathbf{G}}_{1N}^{\circ}(\delta)\mathbf{V}_N(\tilde{\beta}_N(\delta),\delta)], \\ \frac{1}{\tilde{\sigma}_N^2(\delta)} \mathbf{V}_N'(\tilde{\beta}_N(\delta),\delta)\mathbf{G}_{2N}^{\circ}(\rho)\mathbf{V}_N(\tilde{\beta}_N(\delta),\delta). \end{cases} \tag{3.4}$$

Then, the AQS estimators of β and σ_0^2 are $\hat{\beta}_{\texttt{AQS1}} = \tilde{\beta}_N(\hat{\delta}_{\texttt{AQS1}})$ and $\hat{\sigma}^2_{\texttt{AQS1}} = \tilde{\sigma}^2_N(\hat{\delta}_{\texttt{AQS1}})$. The concentrated AQS vector $\tilde{\psi}_N^c(\delta)$ is also crucial in establishing the asymptotic properties of AQSE $\hat{\theta}_{\texttt{AQS1}}$, which are given below.

Asymptotic analyses. Our asymptotic analyses of the AQS estimator $\hat{\theta}_{\texttt{AQS1}}$ cover the cases where n is large and T is finite or large. The case of finite n and large T is of less interest as (i) individual FE and CH can be consistently estimated, and (ii) the spatial weights matrices can be estimated non-parametrically using the T observations for each cross section. Following is a set of generic assumptions for the asymptotic analyses of the FE-SPD models.

Assumption 1: The true spatial parameters δ_0 is in the interior of a compact set Δ.

Assumption 2: The errors $\{v_{it}\}$ are independent over $i = 1, \ldots, n$ and $t = 1, \ldots, T$, with mean 0 and variances $\sigma_0^2 h_i$ such that $\frac{1}{n}\sum_{i=1}^n h_i = 1$ and $h_i > 0, \forall i$, and $\mathrm{E}|v_{it}|^{4+\gamma} < c$ for some $\gamma > 0$ and constant c for all i and t.

Assumption 3: The elements of X_{nt} are non-stochastic and bounded, uniformly in i and t, and $\lim_{N\to\infty} \frac{1}{N}\mathbb{X}_N'\mathbb{X}_N$ exists and is non-singular.

Assumption 4: The spatial weights matrices W_{rn}, $r = 1, 2$, are uniformly bounded in absolute value in both row and column sums and are of zero diagonal elements.

Assumption 5: The matrices $\mathbf{A}_{rN}$ are non-singular and $\mathbf{A}_{rN}^{-1}$ are uniformly bounded in absolute value in both row and column sums. Further, $\mathbf{A}_{1N}^{-1}(\lambda)$ and $\mathbf{A}_{2N}^{-1}(\rho)$ are uniformly bounded in either row or column sums, uniformly in $\delta \in \Delta$.

Assumption 2 extends Lee and Yu (2010a) to allow for unknown CH. Assumptions 1 and 3-5 are as in Lee and Yu (2010a). Compactness of the parameter space Δ is needed due to the non-linearity of δ in the reduced form of the model (Lee and Yu, 2010a), and in the AQS function $\psi_N(\theta)$. Consistent estimation of δ requires that the difference between $\frac{1}{N}\tilde{\psi}_N^c(\delta)$ and its population counterpart converges in probability to zero, uniformly in $\delta \in \Delta$ and such a uniform convergence requires the compactness of Δ (Newey, 1991), as further explained below.

Let $\bar{\psi}_N(\theta) = \mathrm{E}[\psi_N(\theta)]$, the population counterpart of $\psi_N(\theta)$. Let $\bar{\psi}_N^c(\delta)$ be the population counterpart of $\tilde{\psi}_N^c(\delta)$ obtained by concentrating out β and σ^2 from the $\bar{\psi}_N(\theta) = 0$ (see the proof of Theorem 3.1 in Appendix B for details). By Theorem 5.9 of van der Vaart (1998), consistency of $\hat{\delta}_{\texttt{AQS1}}$ follows from (a) the uniform convergence: $\sup_{\delta\in\Delta} \frac{1}{N}\|\tilde{\psi}_N^c(\delta) - \bar{\psi}_N^c(\delta)\| \xrightarrow{p} 0$, and (b) the identification uniqueness condition: $\inf_{\delta: d(\delta,\delta_0)\geq\varepsilon} \frac{1}{N}\|\mathrm{E}[\bar{\psi}_N^c(\delta)]\| > 0 = \frac{1}{N}\|\mathrm{E}[\bar{\psi}_N^c(\delta_0)]\|$, for every $\varepsilon > 0$, where $d(\delta, \delta_0)$ is a measure of distance between δ and δ_0. The latter is satisfied by Assumption 6 given below. Let $\mathbf{D}_N(\delta) = \mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)$ and $\mathbf{D}_N \equiv \mathbf{D}_N(\delta_0)$.

Assumption 6: $\lim_{N\to\infty} \frac{1}{N} F_N(\delta) \neq 0$, $\forall \delta \neq \delta_0$, where, letting $\mathbf{f}_N = \mathbf{A}_{1N}^{-1}\mathbf{X}_N\beta_0$,

$$F_N(\delta) = \begin{cases} \mathbf{f}_N'\mathbf{D}_N'(\delta)\bar{\mathbf{G}}_{1N}^{\circ}(\delta)\mathbf{D}_N(\delta)\mathbf{f}_N + \sigma_0^2 \mathrm{tr}\big(\mathbf{H}_N\mathbf{D}_N'^{-1}\mathbf{D}_N'(\delta)\bar{\mathbf{G}}_{1N}^{\circ}(\delta)\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}\big), \\ \mathbf{f}_N'\mathbf{D}_N'(\delta)\mathbf{G}_{2N}^{\circ}(\delta)\mathbf{D}_N(\delta)\mathbf{f}_N + \sigma_0^2 \mathrm{tr}\big(\mathbf{H}_N\mathbf{D}_N'^{-1}\mathbf{D}_N'(\delta)\mathbf{G}_{2N}^{\circ}(\delta)\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}\big). \end{cases}$$

Once δ_0 is identified, the identification for β_0 and σ_0^2 follows from Assumptions 3-5. In contrast to δ, due to the linearity of β and σ^2 in the AQS function, the compactness of the parameter space of β and σ^2 is not needed.

The joint asymptotic normality of $\hat{\theta}_{\texttt{AQS1}}$ is established based on the fact that AQS function $\psi_N(\psi_0)$ can be written as linear-quadratic forms in the original error vector so that the central limit theorem (CLT) of linear-quadratic forms of Kelejian and Prucha (2001, 2010), or Lemma A.3, can be applied, and that the Hessian and the VC matrix of the AQS function possess desired properties. We have the following theorem with its proof given in Appendix B.

Theorem 3.1. *Under Assumptions 1-6, the AQSE $\hat{\theta}_{\texttt{AQS1}}$ is consistent and asymptotically normal, i.e., as $N \to \infty$, $\hat{\theta}_{\texttt{AQS1}} \xrightarrow{p} \theta_0$ and*

$$\sqrt{N}(\hat{\theta}_{\texttt{AQS1}} - \theta_0) \xrightarrow{D} N\big(\mathbf{0},\ \lim_{N\to\infty} \Phi_N^{-1}\Omega_N\Phi_N^{-1}\big),$$

where $\Phi_N = -\frac{1}{N}\mathrm{E}[\frac{\partial}{\partial\theta_0'}\psi_N(\theta_0)]$ and $\Omega_N = \frac{1}{N}\mathrm{E}[\psi_N(\theta_0)\psi_N'(\theta_0)]$, both are assumed to exist for large enough N and Φ_N is assumed to be positive definite for large enough N.

Robust inference. The robust inferences for ψ_0 depends on the availability of the robust estimators of Φ_N and Ω_N. The former can be consistently estimated by its sample analog $\widehat{\Phi}_{\texttt{AQS1}} = -\frac{1}{N}\frac{\partial}{\partial\theta'}\psi_N(\theta)|_{\theta=\hat{\theta}_{\texttt{AQS1}}}$, but the latter may contain second, third and fourth moments of v_{it} which vary across i in the presence of CH, making plug-in method infeasible. Here, we provide a simple remedy on the standard inference methods so that they remain valid even if there exists unknown CH. Similar to (3.1), we can write $\psi_N(\theta_0)$ as,

$$\psi_N(\theta_0) = \begin{cases} \frac{1}{\sigma_0^2}\mathbb{X}_N'\mathbf{V}_N, \\ \frac{1}{2\sigma_0^4}(\mathbf{V}_N'\mathbf{V}_N - N\sigma_0^2), \\ \frac{1}{\sigma_0^2}\mathbf{V}_N'(\boldsymbol{\eta}_N + \bar{\mathbf{G}}_{1N}^\circ\mathbf{V}_N), \\ \frac{1}{\sigma_0^2}\mathbf{V}_N'\mathbf{G}_{2N}^\circ\mathbf{V}_N, \end{cases} \tag{3.5}$$

As $\psi_N(\theta_0)$ contains linear-quadratic forms of $\mathbf{V}_N$, it can be decomposed into a sum of N uncorrelated terms (martingale differences) so that its variance can be estimated by the outer products of the summands (Baltagi and Yang, 2013). Given $\bar{\mathbf{G}}_{1N}^\circ$ has diagonal elements 0, the term $\mathbf{V}_N'\bar{\mathbf{G}}_{1N}^\circ\mathbf{V}_N$ in $\psi_N(\theta_0)$ can be written as,

$$\mathbf{V}_N'\bar{\mathbf{G}}_{1N}^\circ\mathbf{V}_N = \mathbf{V}_N'(\bar{\mathbf{G}}_{1N}^{\circ u} + \bar{\mathbf{G}}_{1N}^{\circ l})\mathbf{V}_N = \mathbf{V}_N'(\bar{\mathbf{G}}_{1N}^{\circ u\prime} + \bar{\mathbf{G}}_{1N}^{\circ l})\mathbf{V}_N = \mathbf{V}_N'\boldsymbol{\zeta}_{1N}^\circ,$$

where $\bar{\mathbf{G}}_{1N}^{\circ u}$ and $\bar{\mathbf{G}}_{1N}^{\circ l}$ are, respectively, the upper triangular and lower triangular matrices of $\bar{\mathbf{G}}_{1N}^\circ$, and $\boldsymbol{\zeta}_{1N}^\circ = (\bar{\mathbf{G}}_{1N}^{\circ u} + \bar{\mathbf{G}}_{1N}^{\circ l})\mathbf{V}_N$; similarly the term $\mathbf{V}_N'\mathbf{G}_{2N}^\circ\mathbf{V}_N$ is represented. Therefore, the AQS function can be written as $\psi_N(\theta_0) = \sum_{j=1}^N \mathbf{s}_{N,j}$, where,

$$\mathbf{s}_{N,j} = \begin{cases} \frac{1}{\sigma_0^2}\mathbb{X}_{N,j}\mathbf{v}_{N,j}, \\ \frac{1}{2\sigma_0^4}(\mathbf{v}_{N,j}^2 - \sigma_0^2 h_j^c), \\ \frac{1}{\sigma_0^2}\mathbf{v}_{N,j}(\boldsymbol{\eta}_{N,j} + \boldsymbol{\zeta}_{1N,j}^\circ), \\ \frac{1}{\sigma_0^2}\mathbf{v}_{N,j}\boldsymbol{\zeta}_{2N,j}^\circ, \end{cases} \tag{3.6}$$

where $\mathbb{X}'_{N,j}$ is the jth row of $\mathbb{X}_N$, $\mathbf{v}_{N,j}$ is the jth element of $\mathbf{V}_N$ and similarly are the other quantities defined, for $j = 1, \ldots, N$, a combined index for $i = 1, \ldots, n$ and $t = 1, \ldots, T-1$ with i being the faster running index, and $h^c_j = h_{it} = h_i$ for $t = 1, \ldots, T-1$.

The $\{\mathbf{s}_{N,j}, \mathcal{F}_{N,j}\}$ form a vector martingale difference (MD) sequence, with respect to the increasing sequence of σ-fields $\{\mathcal{F}_{N,j}\}$ generated by $\{\mathbf{v}_{N,1}, \ldots, \mathbf{v}_{N,j}\}$, **if** the elements of $\mathbf{V}_N$ are inid normal, which is the case when the original errors are inid normal. It follows that $\Omega_N = \frac{1}{N}\text{Var}[\psi_N(\theta_0)] = \frac{1}{N}\sum_{j=1}^N \text{E}(\mathbf{s}_{N,j}\mathbf{s}'_{N,j})$, and hence can be consistently estimated by,

$$\widehat{\Omega}_{\texttt{AQS1}} = \tfrac{1}{N}\textstyle\sum_{j=1}^N \hat{\mathbf{s}}_{N,j}\hat{\mathbf{s}}'_{N,j}, \tag{3.7}$$

which is termed as the *outer-product-of-martingale-difference* (OPMD) estimate as in Yang (2018), where $\hat{\mathbf{s}}_{N,j}$ are the estimates of $\mathbf{s}_{N,j}$ by plugging $\hat{\theta}_{\texttt{AQS1}}$ and $\hat{\mathbf{V}}_N$ into $\mathbf{s}_{N,j}$ for θ_0 and $\mathbf{V}_N$. Finally, since $\text{E}(v^{*2}_{it}) = \sigma_0^2 h_i$, it is natural to replace h_i in (3.6) by $\hat{h}_i = \frac{1}{(T-1)\hat{\sigma}^2_{\texttt{AQS1}}}\sum_{t=1}^{T-1}\hat{v}^{*2}_{it}$.

When the original errors are inid non-normal, the elements of $\mathbf{V}_N$ are independent across i and uncorrelated (but may not be independent) across t. Hence, there may exist higher-order dependence among the elements of $\mathbf{V}_N$ across t, i.e., between v^*_{it} and v^{*2}_{is} and v^{*2}_{it} and v^{*2}_{is} for $t \neq s$, implying that $\mathbf{s}_{N,it}$ and $\mathbf{s}_{N,is}$ may be correlated and that the above OPMD estimate of Ω_N may not be strictly valid. As $\mathbf{s}_{N,j}$ or $\mathbf{s}_{N,it}$ are uncorrelated across i for all t, we have

$$\begin{aligned}\text{Var}[\psi_N(\theta_0)] &= \text{Var}(\textstyle\sum_{i=1}^n\sum_{t=1}^{T-1}\mathbf{s}_{N,it}) = \sum_{i=1}^n \text{Var}(\sum_{t=1}^{T-1}\mathbf{s}_{N,it})\\ &= \textstyle\sum_{j=1}^N \text{E}(\mathbf{s}_{N,j}\mathbf{s}'_{N,j}) + 2\sum_{i=1}^n\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}\text{E}(\mathbf{s}_{N,it}\mathbf{s}'_{N,is}),\end{aligned}$$

where we freely switch between the single index j and the double indices (i,t) for convenience. This immediately suggests the following general estimator fully robust against non-normality:

$$\widehat{\Omega}^{\dagger}_{\texttt{AQS1}} = \tfrac{1}{N}\textstyle\sum_{j=1}^N \hat{\mathbf{s}}_{N,j}\hat{\mathbf{s}}'_{N,j} + \tfrac{1}{n}\sum_{i=1}^n \hat{\mathbf{r}}_{N,i}, \tag{3.8}$$

where $\hat{\mathbf{r}}_{N,i} = \frac{2}{T-1}\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}(\hat{\mathbf{s}}_{N,it}\hat{\mathbf{s}}'_{N,is})$. Its **sequential** limiting behavior is given below.

Theorem 3.2. *Under Assumptions 1-6, we have, (i)* $\widehat{\Phi}_{\texttt{AQS1}} - \Phi_N \xrightarrow{p} 0$ *as* $N \to \infty$*;*

$$(ii)\ \ \widehat{\Omega}^{\dagger}_{\texttt{AQS1}} - \Omega_N \xrightarrow{p} 0, \quad \textit{as } n \to \infty \textit{ first, and then followed by } T \to \infty;$$

and (iii) $\widehat{\Omega}_{\texttt{AQS1}} - \Omega_N \xrightarrow{p} 0$ *as* $N \to \infty$*, if the skewness and excess kurtosis of* v_{it} *are both zero.*

Intuitively, the higher-order dependence causing the additional term in (3.8) relative to (3.7) when the errors are non-normal may be asymptotically negligible due to the fact that $\{f_t\}$, the columns of $F_{T,T-1}$, are orthonormal. See the proof of Theorem 3.2 in Appendix B for details. Our Monte Carlo results (unreported for brevity) show that this is indeed the case. When T is small, the approach of Yang (2018) can be also followed, i.e., first sum the elements of $\psi_N(\theta_0)$ over t, and then decompose over i to give a true MD representation of $\psi_N(\theta_0)$ in n terms.

3.2. Finite sample improved AQS method

While it seems fairly easy to adjust the full score function (2.7) to attain a robust estimator, with a desired asymptotic performance, the finite sample performance is less than optimal by the fact that the full score function does not take into account the variability caused by estimating the other model parameters β and σ^2. As such modifying the concentrated quasi score functions are desirable to ensure both asymptotic as well as finite sample performance in the robust estimator, since the concentrated score function captures the variability coming from estimating β and σ^2. See Liu and Yang (2015) for more discussions.

The concentrated score function derived by taking the derivatives of the concentrated log-likelihood function (2.6) with respect to δ, or by concentrating (2.7), is,

$$S_N^c(\delta) = \begin{cases} \frac{1}{\tilde{\sigma}_N^2(\delta)} \mathbb{Y}_N'(\delta)\mathbb{M}_N(\rho)\big[\bar{\mathbf{G}}_{1N}(\delta) - \frac{1}{n}\mathtt{tr}(\mathbf{G}_{1N}(\lambda))I_N\big]\mathbb{Y}_N(\delta), \\ \frac{1}{\tilde{\sigma}_N^2(\delta)} \mathbb{Y}_N'(\delta)\mathbb{M}_N(\rho)\big[\bar{\mathbf{G}}_{2N}(\rho) - \frac{1}{n}\mathtt{tr}(\mathbf{G}_{2N}(\rho))I_N\big]\mathbb{Y}_N(\delta), \end{cases} \tag{3.9}$$

where $\bar{\mathbf{G}}_{1N}(\delta) = \mathbf{A}_{2N}(\rho)\mathbf{G}_{1N}(\lambda)\mathbf{A}_{2N}^{-1}(\rho)$ as defined above, and $\bar{\mathbf{G}}_{2N}(\rho) = \mathbf{G}_{2N}(\rho)\mathbb{M}_N(\rho)$.

Using $S_N^c(\delta)$, the regular QMLE is defined as, $\hat{\delta}_N = \arg\{S_N^c(\delta) = 0\}$. Clearly, the root-finding process is independent of $\tilde{\sigma}_N^2(\delta)$ as long as $\tilde{\sigma}_N^2(\delta)$ is bounded from below, away from 0 for δ in a neighborhood of δ_0. We therefore adjust the numerators of (3.9) so that the adjusted quantities have zero expectation at θ_0. Note, $\mathrm{E}(\mathbb{Y}_N'\mathbb{M}_N\bar{\mathbf{G}}_{rN}\mathbb{Y}_N) = \sigma_0^2\mathtt{tr}(\mathbf{H}_N\mathbb{M}_N\bar{\mathbf{G}}_{rN}) = \sigma_0^2\mathtt{tr}(\mathbf{H}_N\mathtt{diag}(\mathbb{M}_N\bar{\mathbf{G}}_{rN}))$. Hence, a possible way to go is to replace $\frac{1}{n}\mathtt{tr}(\mathbf{G}_{rN})$ of (3.9) with $\mathtt{diag}(\mathbb{M}_N\bar{\mathbf{G}}_{rN})$. However, this introduces an additional $\mathbb{M}_N$, i.e., $\mathrm{E}(\mathbb{Y}_N'\mathtt{diag}(\mathbb{M}_N\bar{\mathbf{G}}_{rN})\mathbb{Y}_N) = \sigma_0^2\mathtt{tr}(\mathbf{H}_N\mathbb{M}_N\mathtt{diag}(\mathbb{M}_N\bar{\mathbf{G}}_{rN}))$. To cancel out this effect, the final adjustment made is of the form $\mathtt{diag}(\mathbb{M}_N)^{-1}\mathtt{diag}(\mathbb{M}_N\bar{\mathbf{G}}_{rN})$. The final AQS function is simply,

$$\tilde{\psi}_N^*(\delta) = \begin{cases} \mathbb{Y}_N'(\delta)\mathbb{M}_N(\rho)\bar{\mathbf{G}}_{1N}^*(\delta)\mathbb{Y}_N(\delta), \\ \mathbb{Y}_N'(\delta)\mathbb{M}_N(\rho)\bar{\mathbf{G}}_{2N}^*(\rho)\mathbb{Y}_N(\delta), \end{cases} \tag{3.10}$$

where $\bar{\mathbf{G}}_{rN}^*(\delta) = \bar{\mathbf{G}}_{rN}(\delta) - \mathrm{diag}(\mathbb{M}_N(\rho))^{-1}\mathrm{diag}[\mathbb{M}_N(\rho)\bar{\mathbf{G}}_{rN}(\delta)],\ r = 1, 2.$

It can be seen that $\mathrm{E}[\tilde{\psi}_N^*(\delta_0)] = 0$ and $\mathrm{plim}_{N\to\infty}\frac{1}{N}\tilde{\psi}_N^*(\delta_0) = 0$, i.e., $\tilde{\psi}_N^*(\delta)$ gives a set of unbiased and consistent estimating functions, leading to an AQS estimator of δ, possibly consistent under unknown CH, asymptotically normal, and with a finite sample improved performance:

$$\hat{\delta}_{\mathtt{AQS1}}^* = \arg\{\tilde{\psi}_N^*(\delta) = 0\}. \tag{3.11}$$

Once the CH-robust estimator $\hat{\delta}_{\mathtt{AQS1}}^*$ is obtained, the CH-robust estimators for β and σ^2 follow from $\hat{\beta}_{\mathtt{AQS1}}^* \equiv \tilde{\beta}_N(\hat{\delta}_{\mathtt{AQS1}}^*)$ and $\hat{\sigma}_{\mathtt{AQS1}}^{*2} \equiv \tilde{\sigma}_N^2(\hat{\delta}_{\mathtt{AQS1}}^*)$. Denote $\hat{\theta}_{\mathtt{AQS1}}^* = (\hat{\beta}_{\mathtt{AQS1}}^{*\prime}, \hat{\sigma}_{\mathtt{AQS1}}^{*2}, \hat{\delta}_{\mathtt{AQS1}}^{*\prime})'$, called the AQS* estimator in this paper. Note that the AQS functions (3.10) do not depend on $\tilde{\sigma}_N^2(\delta)$.

The **asymptotic properties** of the AQS* estimators $\hat{\delta}_{\mathtt{AQS1}}^*$ and $\hat{\beta}_{\mathtt{AQS1}}^*$ are studied under the same set of regularity conditions. In particular, $\bar{\mathbf{G}}_{1N}^*(\delta)$ is asymptotically equivalent to $\bar{\mathbf{G}}_{1N}^\circ(\delta)$, $\bar{\mathbf{G}}_{2N}^*(\delta)$ is asymptotically equivalent to $\mathbf{G}_{2N}^\circ(\delta)$, and $\bar{\sigma}_N^2(\delta)$ is bounded from below away from 0, uniformly in $\delta \in \Delta$ as shown in the proof of Theorem 3.1 and so is $\tilde{\sigma}_N^2(\delta)$ for large enough N. Thus, it is valid to work with the numerators of (3.9), and the identification uniqueness condition for δ_0, given in Assumption 6, remains.

To establish asymptotic normality of $\hat{\delta}^*_{\texttt{AQS1}}$, note that,

$$\tilde{\psi}^*_N(\delta_0) = \begin{cases} \mathbf{V}'_N \mathbf{B}_{1N}\mathbf{V}_N + \mathbf{c}'_{1N}\mathbf{V}_N, \\ \mathbf{V}'_N \mathbf{B}_{2N}\mathbf{V}_N + \mathbf{c}'_{2N}\mathbf{V}_N, \end{cases} \tag{3.12}$$

where $\mathbf{B}_{rN} = \mathbb{M}_N \bar{\mathbf{G}}^*_{rN}$ and $\mathbf{c}_{rN} = \mathbb{M}_N \bar{\mathbf{G}}^*_{rN} \mathbb{X}_N \beta_0, r = 1, 2$. Clearly, $\texttt{diag}(\mathbf{B}_{rN}) = \mathbf{0}_{N\times N}$ by construction. The AQS function $\tilde{\psi}^*_N(\delta_0)$ can be further rewritten as a linear-quadratic form of the original disturbances, $\{v_{it}\}$, and therefore its asymptotic normality can be established by the CLT for linear-quadratic forms of Kelejian and Prucha (2001) or its multivariate version of Kelejian and Prucha (2010) extended in Lemma A.3. This together with the proper asymptotic behavior of the Hessian and VC matrices of $\tilde{\psi}^*_N(\delta_0)$ lead to the asymptotic normality of $\hat{\delta}^*_{\texttt{AQS1}}$.

Theorem 3.3. *Under Assumptions 1-6, the AQS* estimator $\hat{\delta}^*_{\texttt{AQS1}}$ is consistent and asymptotically normal, i.e., as $N \to \infty$, $\hat{\delta}^*_{\texttt{AQS1}} \xrightarrow{p} \delta_0$ and*

$$\sqrt{N}(\hat{\delta}^*_{\texttt{AQS1}} - \delta_0) \xrightarrow{D} N\big(\mathbf{0},\ \lim_{N\to\infty} \Phi^{*-1}_N \Omega^*_N \Phi^{*-1}_N\big),$$

where $\Omega^*_N = \frac{1}{N}\mathrm{Var}[\tilde{\psi}^*_N(\delta_0)]$ and $\Phi^*_N = -\frac{1}{N}\mathrm{E}[\frac{\partial}{\partial \delta'_0}\tilde{\psi}^*_N(\delta_0)]$, both are assumed to exist, and Φ^*_N is further assumed to be positive definite for large enough N.

Finally, for the AQS* estimator $\hat{\beta}^*_{\texttt{AQS1}} = \tilde{\beta}_N(\hat{\delta}_{\texttt{AQS1}})$, we have by a Taylor expansion:

$$\begin{aligned} \hat{\beta}^*_{\texttt{AQS1}} - \beta_0 &= \tilde{\beta}_N(\delta_0) - \beta_0 + [\tfrac{\partial}{\partial \delta'_0}\hat{\beta}_N(\delta_0)](\hat{\delta}_{\texttt{AQS1}} - \delta_0) + O_p(\tfrac{1}{N}) \\ &= (\mathbb{X}'_N\mathbb{X}_N)^{-1}\mathbb{X}'_N\mathbf{V}_N + \mathrm{E}[\tfrac{\partial}{\partial \delta'_0}\hat{\beta}_N(\delta_0)]\Phi^{-1}_N \tilde{\psi}^*_N(\delta_0) + O_p(\tfrac{1}{N}) \\ &= (\mathbb{X}'_N\mathbb{X}_N)^{-1}[\mathbb{X}'_N\mathbf{V}_N + \Pi_N \tilde{\psi}^*_N(\delta_0)] + O_p(\tfrac{1}{N}), \end{aligned}$$

where $\Pi_N = \mathrm{E}[\frac{\partial}{\partial \delta'_0}\hat{\beta}_N(\delta_0)]\Phi^{-1}_N$, and $\mathrm{E}[\frac{\partial}{\partial \delta'_0}\hat{\beta}_N(\delta_0)] = -\big[(\mathbb{X}'_N\mathbb{X}_N)^{-1}\mathbb{X}'_N\bar{\mathbf{G}}_{1N}\mathbb{X}_N\beta_0,\ \mathbf{0}_{k\times 1}\big]$ because

$$\tfrac{\partial}{\partial \delta'_0}\hat{\beta}_N(\delta_0)] = \big[(\mathbb{X}'_N\mathbb{X}_N)^{-1}\mathbb{X}'_N\mathbf{A}_{2N}\mathbf{W}_{1N}\mathbf{Y}_N,\ (\mathbb{X}'_N\mathbb{X}_N)^{-1}\mathbb{X}'_N(\mathbf{G}'_{1N} + \mathbf{G}_{1N})\mathbb{M}_N\mathbb{Y}_N\big].$$

These lead to the asymptotic distribution of $\hat{\beta}^*_{\texttt{AQS1}}$.

Theorem 3.4. *Under Assumptions 1-6, the AQS* estimator $\hat{\beta}^*_{\texttt{AQS1}}$ is consistent and asymptotically normal, i.e., as $N \to \infty$, $\hat{\beta}^*_{\texttt{AQS1}} \xrightarrow{p} \beta_0$, and*

$$\sqrt{N}(\hat{\beta}^*_{\texttt{AQS1}} - \beta_0) \xrightarrow{D} N\big[0,\ \lim_{N\to\infty}(\tfrac{1}{N}\mathbb{X}'_N\mathbb{X}_N)^{-1}\Sigma_N(\tfrac{1}{N}\mathbb{X}'_N\mathbb{X}_N)^{-1}\big],$$

where $\Sigma_N = \frac{1}{N}\mathrm{Var}(\mathbb{X}'_N\mathbf{V}_N + \Pi_N\tilde{\psi}^*_N)$.

The **robust inferences** for δ and β are carried out in a similar manner as in Sec. 3.1. First, to conduct robust inference for δ, Φ^*_N is consistently and robustly estimated by its sample analog, $\widehat{\Phi}^*_{\texttt{AQS1}} = -\frac{\partial}{\partial \delta'_0}\tilde{\psi}^*_N(\delta_0)|_{\delta_0 = \hat{\delta}^*_{\texttt{AQS1}}}$. Then, the linear-quadratic forms of the elements of $\tilde{\psi}^*_N(\delta_0)$ leads to an OPMD estimate of the VC matrix Ω^*_N of $\tilde{\psi}^*_N(\delta_0)$, similar to Sec. 3.1.

Write $\mathbf{B}_{rN} = \mathbf{B}^u_{rN} + \mathbf{B}^l_{rN}$. Define $\zeta_{rN} = (\mathbf{B}^{u\prime}_{rN} + \mathbf{B}^l_{rN})\mathbf{V}_N$, and let $\mathbf{s}^*_{N,j} = (\zeta_{1N,j} + c_{1N,j},\ \zeta_{2N,j} + c_{2N,j})'$, $r = 1, 2, j = 1, \ldots, N$. It follows that $\tilde{\psi}^*_N(\delta_0) = \sum_{j=1}^N \mathbf{v}_{N,j}\mathbf{s}^*_{N,j}$. If the elements $\{\mathbf{v}_{N,j}\}$ of the transformed error vector $\mathbf{V}_N$ are independent, which is the case if the original errors are independent normal, then it can be shown that $\{\mathbf{v}_{N,j}\mathbf{s}^*_{N,j}\}$ form a sequence of martingale differences and hence are uncorrelated. It follows that $\Omega^*_N = \frac{1}{N}\mathrm{Var}[\tilde{\psi}^*_N(\delta_0)] = \frac{1}{N}\sum_{j=1}^N \mathrm{E}[\mathbf{v}^2_{N,j}\mathbf{s}^*_{N,j}\mathbf{s}^{*\prime}_{N,j}]$. Therefore, a heteroskedasticity robust estimator of Ω^*_N is given as

$$\widehat{\Omega}^*_{\texttt{AQS1}} = \tfrac{1}{N}\textstyle\sum_{j=1}^N \hat{\mathbf{v}}^2_{N,j}\hat{\mathbf{s}}^*_{N,j}\hat{\mathbf{s}}^{*\prime}_{N,j}, \tag{3.13}$$

where $\hat{\mathbf{v}}_{N,j}$ and $\hat{\mathbf{s}}^*_{N,j}$ are, respectively, the estimates of $\mathbf{v}_{N,j}$ and $\mathbf{s}^*_{N,j}$, based on $\hat{\theta}^*_{\texttt{AQS1}}$.

Now suppose the disturbances are not Gaussian. In this case, $\{\mathbf{v}_{N,j}\mathbf{s}^*_{N,j}\}$ are no longer strictly uncorrelated and hence the OPMD estimator given in (3.13) may not be a valid estimator of the VC matrix of the AQS function. However, similar to Theorem 3.2, an extended estimator fully robust against non-normality and unknown heteroskedasticity is given as follows:

$$\widehat{\Omega}^{*\dagger}_{\texttt{AQS1}} = \tfrac{1}{N}\textstyle\sum_{j=1}^N \hat{\mathbf{v}}^2_{N,j}\hat{\mathbf{s}}^*_{N,j}\hat{\mathbf{s}}^{*\prime}_{N,j} + \tfrac{1}{n}\sum_{i=1}^n \hat{\mathbf{r}}^*_{N,i}, \tag{3.14}$$

where $\hat{\mathbf{r}}^*_{N,i} = \frac{2}{T-1}\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}\hat{v}^*_{it}\hat{v}^*_{is}\hat{\mathbf{s}}^*_{N,it}\hat{\mathbf{s}}^{*\prime}_{N,is}$.

Finally, to estimate Σ_N in Theorem 3.4 for inference on β, based on the MD decomposition for $\tilde{\psi}^*_N(\delta_0)$ given above under normality, we obtain $\Sigma_N = \frac{1}{N}\sum_{j=1}^N \mathrm{E}[\mathbf{v}^2_{N,j}(\mathbb{X}_{N,j} + \Pi_{N,j}\mathbf{s}_{N,j})(\mathbb{X}_{N,j} + \Pi_{N,j}\mathbf{s}_{N,j})']$. It follows that a heteroskedasticity robust estimator of Σ_N is,

$$\widehat{\Sigma}_{\texttt{AQS1}} = \tfrac{1}{N}\textstyle\sum_{j=1}^N \hat{\mathbf{v}}^2_{N,j}(\mathbb{X}_{N,j} + \widehat{\Pi}_{N,j}\hat{\mathbf{s}}^*_{N,j})(\mathbb{X}_{N,j} + \widehat{\Pi}_{N,j}\hat{\mathbf{s}}^*_{N,j})'. \tag{3.15}$$

Under non-normality, one immediately obtain a fully robust estimator:

$$\widehat{\Sigma}^{\dagger}_{\texttt{AQS1}} = \widehat{\Sigma}_{\texttt{AQS1}} + \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{\mathbf{q}}^*_{N,i} \tag{3.16}$$

where $\hat{\mathbf{q}}^*_{N,i} = \frac{2}{T-1}\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}(\widehat{\Pi}_{N,it}\hat{v}^*_{it}\hat{v}^*_{is}\hat{\mathbf{s}}^*_{N,it}\hat{\mathbf{s}}^{*\prime}_{N,is}\widehat{\Pi}'_{N,is} + 2\mathbb{X}_{N,it}\hat{v}^*_{it}\hat{v}^*_{is}\hat{\mathbf{s}}^{*\prime}_{N,is}\widehat{\Pi}'_{N,is})$.

Theorem 3.5. *Under Assumptions 1-6, we have, (i)* $\widehat{\Phi}^*_{\texttt{AQS1}} - \Phi^*_N \xrightarrow{p} 0$ *as* $N \to \infty$*;*

$$(ii)\ \ \widehat{\Omega}^{*\dagger}_{\texttt{AQS1}} - \Omega^*_N \xrightarrow{p} 0 \ \ \textit{and} \ \ \widehat{\Sigma}^{\dagger}_{\texttt{AQS1}} - \Sigma_N \xrightarrow{p} 0, \ \textit{as } n \to \infty \textit{ first and then } T \to \infty;$$

and (iii) $\widehat{\Omega}^*_{\texttt{AQS1}} - \Omega^*_N \xrightarrow{p} 0$ *and* $\widehat{\Sigma}_{\texttt{AQS1}} - \Sigma_N \xrightarrow{p} 0$ *as* $N \to \infty$*, if the skewness and excess kurtosis of* v_{it} *are zero.*

Similar to the arguments given below Theorem 3.2 and its proof in Appendix B, the additional terms in $\widehat{\Omega}^{*\dagger}_{\texttt{AQS1}}$ and $\widehat{\Sigma}^{\dagger}_{\texttt{AQS1}}$ do not play much a role due to the fact that the columns of $F_{T,T-1}$ are orthonormal. Finally, similar steps lead to the asymptotic results for the AQS* estimator $\hat{\sigma}^{*2}_{\texttt{AQS1}}$, and the CH-robust inference method for inference on σ^2. As this is not a case of major interest, and the methods based on the joint AQS function have already covered this case, we do not present details to conserve space.

4. Robust Estimation and Inference for FE$_2$-SPD Model

The AQS estimation. The AQS estimation method for the FE$_1$-SPD model introduced above may be extendible to the FE$_2$-SPD model. As the FE$_2$-SPD model (2.9) takes an identical form as the FE$_1$-SPD model (2.3), the likelihood and quasi score functions remain in the same form as well. These motivate that for the robust estimation and inference for the FE$_2$-SPD Model, the same form of the AQS function (3.2) of the FE$_1$-SPD model may be used. However, this AQS function may not achieve the desired property for the FE$_2$-SPD model as $\frac{1}{N}\mathrm{E}(\mathbf{V}_N'\bar{\mathbf{G}}_{rN}^{\circ}\mathbf{V}_N) = \frac{1}{n}\mathtt{tr}(\bar{G}_{rn}^{\circ}H_n^{\circ}) = \frac{1}{n}\mathtt{tr}(\bar{G}_{rn}^{\circ}F_{n,n-1}'H_nF_{n,n-1}) \neq 0$, $r = 1, 2$, due to the difference in transformed errors $\mathbf{V}_N$ for the SPD-2F model, which are correlated across i under CH. This may pose a potential problem in terms of attaining consistency for the AQSE even after making the adjustments as those for FE$_1$-SPD model. However, it is easy to see that,

$$\frac{1}{n}\mathtt{tr}(\bar{G}_{rn}^{\circ}F_{n,n-1}'H_nF_{n,n-1}) = \frac{1}{n}\sum_{i=1}^{n-1}\sum_{j\neq i}^{n-1}\bar{G}_{rn,ij}^{\circ}f_j'H_nf_i \ \equiv \ k_{rn},$$

where f_i denotes the ith column of $F_{n,n-1}$ and $\bar{G}_{rn,ij}^{\circ}$ denotes the ijth element of $\bar{G}_{rn}^{\circ}$. As $f_j'f_i = 0$, for $j \neq i$, we have $k_{rn} = 0$ if $H_n = cI_n$, Also, note that $f_i'f_i = 1$. Therefore, it is reasonable to **assume**

$$k_{rn} \to 0, \text{ as } n \to \infty, r = 1, 2.$$

We are unable to provide simpler justifications for its validity, but instead we have performed extensive Monte Carlo experiments and the results show clearly that this is indeed the case. Thus, under this condition the AQS method inherited from the FE$_1$-SPD model remains asymptotically valid for the FE$_2$-SPD model. Therefore, we proceed using the same AQS function (3.2) to give an AQS estimator, denoted as $\hat{\theta}_{\mathtt{AQS2}}$, of the structural parameters θ in the FE$_2$-SPD model, and do not pursue rigorous asymptotic theories in this paper.

Based on the AQS estimator $\hat{\theta}_{\mathtt{AQS2}}$, robust inference for θ can be carried out in a similar manner. In particular, the asymptotic variance of $\hat{\theta}_{\mathtt{AQS2}}$ is $\Phi_N^{-1}\Omega_N\Phi_N^{-1}$, where Φ_N and Ω_N are defined in the same way and estimated in the identical manners as those for the FE$_1$-SPA model. But again, we do not pursue the rigorous theoretical work in this occasion.

Finite sample improved AQS estimation. Similar to the considerations given in Sec. 3.2, the finite sample improved AQS estimation strategy for the FE$_1$-SPD model may be extended directly to the FE$_2$-SPD model using the newly defined quantities for the transformed FE$_2$-SPD model given in Sec. 2.2., due to the fact that the two transformed models and the corresponding quasi score functions are identical in forms. However, unlike the case of FE$_1$-SPD model, Var$(\mathbf{V}_N)$ is no longer diagonal under CH. Therefore, for the AQS function $\tilde{\psi}_N^*(\delta)$ given in (3.10) to be applicable to the FE$_2$-SPD model, it requires additional minor conditions which can be seen to be asymptotically equivalent to the conditions given in Sec. 3.1 for AQS estimation of FE$_2$-SPD model: $k_{rn} \to 0$ as $n \to \infty$, $r = 1, 2$. The resulting finite sample improved AQS estimators are denoted by $\hat{\theta}_{\mathtt{AQS2}}^* = (\hat{\beta}_{\mathtt{AQS2}}^{*\prime}, \hat{\sigma}_{\mathtt{AQS2}}^{*2}, \hat{\delta}_{\mathtt{AQS2}}^{*\prime})'$.

The results given in Sections 3 and 4 show that the AQS estimators $\hat{\theta}_{\mathtt{AQS1}}$, $\hat{\theta}_{\mathtt{AQS2}}$, $\hat{\theta}_{\mathtt{AQS1}}^*$, and $\hat{\theta}_{\mathtt{AQS2}}^*$ are computationally as simple as the original QML estimators $\hat{\theta}_{\mathtt{QML1}}$ and $\hat{\theta}_{\mathtt{QML2}}$, while being generally consistent under unknown CH and preserving the nature of being robust against non-normality. Monte Carlo results given in the following section confirm the excellent performance of these estimators, in particular the pair of finite sample improved AQSEs $\hat{\theta}_{\mathtt{AQS1}}^*$ and $\hat{\theta}_{\mathtt{AQS2}}^*$.

5. Monte Carlo Study

Extensive Monte Carlo experiments were conducted to investigate the finite sample performance of the original QMLE $\hat{\delta}_N$ and the proposed AQSEs $\hat{\delta}_{\texttt{AQS1}}$ and $\hat{\delta}^*_{\texttt{AQS1}}$, and their impacts on the estimators of β_0 and σ_0^2, with respect to changes in the sample size, spatial layouts, error distributions and the model parameters when the disturbances are heteroskedastic. We consider cases where the original QMLEs may be robust against CH and the cases they are not. The simulations are carried out based on the following data generation process (DGP):

$$Y_{nt} = \lambda_0 W_n Y_{nt} + X_{1,nt}\beta_1 + X_{2,nt}\beta_2 + \mathbf{c}_n + U_{nt}, \quad U_{nt} = \rho_0 W_n U_{nt} + V_{nt}, \quad t = 1, 2, 3,$$

where $X_{1,nt}$ and $X_{2,nt}$ are the two fixed regressors, and $V_{nt} = \sigma H_n e_{nt}$. The regression coefficients β is set to $(1,1)'$, σ is set to 1, λ and ρ takes values from $\{-0.5, -0.25, 0, 0.25, 0.5\}$, n take values from $\{50, 100, 250, 500\}$ and T is initially set to be 3. The ways of generating the values for (X_{1n}, X_{2n}), the spatial weights matrix W_n, the CH measure H_n, and the idiosyncratic errors e_{nt} are described below. Each set of Monte Carlo results is based on $5,000$ Monte Carlo samples.

Spatial Weights Matrix: We use three different spatial layouts: (i) `Circular Neighbors`, (ii) `Group Interaction` and (iii) `Queen Contiguity`. In (i), neighbors occur in the positions immediately ahead and behind a particular spatial unit. For example, for the ith spatial unit with 6 neighbors, the ith row of W_n matrix has non-zero elements in the positions: $i-3, i-2, i-1, i+1, i+2$, and $i+3$. The weights matrix we consider has 2, 4, 6, 8 and 10 neighbors with equal proportion. In (ii), neighbors occur in groups where each group member is spatially related to one another resulting in a symmetric W_n matrix. In (iii), neighbors could occur in the eight cardinal and ordinal positions of each unit. To ensure the CH does not fade as n increases (so that the regular QMLE is 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 or fixing the number of neighbors behind and ahead in the Circular Neighbors scheme. The degree of spatial dependence is naturally bounded in the Queen Contiguity weights matrix. To analyze the performance of the original QMLE when it is likely to be robust against CH, we use Queen Contiguity scheme and the `balanced Circular Neighbors` scheme where all spatial units have 6 peers each.

Heteroskedasticity: For the `unbalanced Circular Neighbor` scheme, $h_{n,i}$ is generated as the ratio of the total number of neighbors to the average number of neighbors for each i while for the Group Interaction scheme $h_{n,i}$ is generated as the ratio of the group size to mean group size. For the balanced Circular Neighbor and the Queen Contiguity schemes, we generate CH as $h_{n,i} = n[\sum_{i=1}^n (|X_{1n,i}| + |X_{2n,i}|)]^{-1}(|X_{1n,i}| + |X_{2n,i}|)$.

Regressors: The regressors are generated according to `REG1:` $\{x_{1,it}, x_{2,it}\} \overset{iid}{\sim} N(0,1)/\sqrt{2}$. For the Group Interaction scheme, the regressors can also be generated according to `REG2:` $\{x_{1,it,r}, x_{2,it,r}\} \overset{iid}{\sim} (2z_r + z_{it,r})/\sqrt{10}$, where $(z_r, z_{it,r}) \overset{iid}{\sim} N(0,1)$, for the ith observation in the rth group, to give a case of non-iid regressors taking into account the impact of group sizes on the regressors. Both schemes give a signal-to-noise ratio of 1 when $\beta_1 = \beta_2 = \sigma = 1$.

Error Distribution: To generate the e_{nt} component of the disturbance term, three DGPs are considered: `DGP1:` $\{e_{n,it}\}$ are iid standard normal, `DGP2:` $\{e_{n,it}\}$ are iid standardized normal mixture with 10% of values from $N(0,4)$ and the remaining from $N(0,1)$ and `DGP3:` $\{e_{n,it}\}$ iid standardized log-normal with parameters 0 and 1. Thus, the error distribution from `DGP2` is leptokurtic, and that of `DGP3` is both skewed and leptokurtic.

Tables 1-3 (a,b,c) summarize partial results for the QML and AQS* estimation of δ (the worst and the best among the three estimators), where in each table, the Monte Carlo means, root mean square errors (rmse) and the standard deviations (sd) of the estimators are reported. To investigate the finite sample performance of the proposed OPMD-based robust standard error estimators, we also report the averaged standard errors ($\widehat{\text{sd}}$) of the AQS* estimator (AQSE*) based on $\widehat{\Omega}^*_{\texttt{AQS1}}$ in Theorem 3.5. Table 4 (a,b) gives empirical sizes of the t tests of H_0: $\beta_1 = \beta_2$ under the Group Interaction scheme, using the QML and AQS* estimators, respectively. The main observations made from the Monte Carlo results are summarized as follows:

(i) For the case where QMLE is likely to be consistent such as in Queen contiguity given in Tables 1a-1c, both estimators perform equally well, consistency of both the estimators is clearly shown, and the consistency of the OPMD-based standard error estimate for the AQSE* is also clearly demonstrated.

(ii) For the cases where the original QMLE is inconsistent as in Tables 2-3, AQSE* provides a useful consistent alternative with significantly less bias and with little or no impact on the efficiency. The inconsistency of the QMLE and the consistency (robustness) of the AQSE* are clearly demonstrated by the Monte Carlo results.

(iii) The OPMD-based estimates of the robust standard errors of λ_0 and ρ_0 perform well with their values very close to their Monte Carlo counterparts in general.

(iv) As the theory suggests, the QMLEs for the covariate effects are less affected by CH. The AQSE* for the covariate effects (unreported for brevity) performs well as well.

(v) The t-statistics based on the AQSE* outperform the ones based on the QMLE in terms of size. The AQSE*-based test is oversized but not severe, and with the increase of sample size, its empirical sizes quickly converge to their nominal levels. In contrast, the QMLE-based test is more severely oversized when sample size is not large, its empirical sizes depend strongly on the values of the spatial parameters, keep decreasing as sample size increases, and as sample size becomes large it becomes significantly undersized (see the lower parts of Tables 4a and 4b).

(vi) The cases with larger T were also investigated. The results (unreported for brevity) show that the AQSE* for δ and the OPMD-based estimate for the standard errors continue to perform well, irrespective of whether the errors are normal or non-normal. These conclusions support the discussions below Theorem 3.5.

Extensive Monte Carlo experiments were also conducted for the estimators based on the joint AQS function, corresponding to the results of Theorems 3.1 and 3.2. The results generally support the theories, in particular, the AQSE performs not as well as AQSE* although generally consistent. The results (not reported for conserving space) with $\widehat{\Omega}^{\dagger}_{\mathtt{AQS1}}$ or $\widehat{\Omega}^{*\dagger}_{\mathtt{AQS1}}$ in VC matrix estimation do not show significant difference from those using $\widehat{\Omega}_{\mathtt{AQS1}}$ or $\widehat{\Omega}^{*}_{\mathtt{AQS1}}$. Furthermore, Monte Carlo experiments were conducted as well for the QMLE, AQSE and AQSE* estimators of FE_2-SPD model, and the results (available from the authors upon requests) show similar patterns, showing that the assumption on quantities $k_{rn}, r = 1, 2$, defined in Section 4 and the related discussions are valid. Therefore, the methods developed for the FE_1-SPD model can be directly applied to the FE_2-SPD model, although rigorous theories are yet to be developed.

6. Conclusion

In this paper we consider the problem of cross-sectional heteroskedasticity (CH) and non-normality of the disturbances in a fixed effects spatial panel data (FE-SPD) model with spatial autoregressive dependent variable and disturbances. CH in particular causes the traditional QML estimator to be inconsistent in general, and for this we proposed the *adjusted quasi score* (AQS) methods, based on joint AQS or concentrated AQS functions, giving AQS and AQS* estimators that are generally robust against unknown CH. For CH-robust inferences, we proposed an *outer-product-of-martingale-differences* (OPMD) method to estimate the variance of the AQS or AQS* functions, which together with the Hessian matrices of the AQS or AQS* functions give robust estimator of the variance-covariance (VC) matrix of the AQS or AQS* estimators. Monte Carlo results reveal excellent performance of the proposed methods.

Motivated by the pioneering research in the cross-sectional spatial econometric literature, we also give some formal arguments that the traditional QMLE of the FE-SPD model can be consistent under CH of certain 'types'. However, the conditions under which the QMLE is robust against unknown CH are difficult to verify, and even if the conditions were satisfied under some CH structures, these CH structures may not suit the practical applications well. Therefore, the proposed set of fully robust AQS-estimation method and OPMD-inference method, which are computationally as simple as the QML methods, are recommended for practical applications.

The studies given in this paper on SPD models with one-way fixed effects or two-way additive fixed effects shed much light on the AQS strategy for robust estimation of structural parameters in the model, and the corresponding OPMD strategy on the VC matrix estimation for robust inferences, for future studies on more general models or different models. For example, in cases where the spatial weights matrices changes with time so that the transformation method cannot be applied, the AQS method may be able to provide a solution. In a situation where the two-way fixed effects are *interactive*, the AQS method may be able to provide an alternative, and perhaps simpler method to estimate the model. A more difficult issue remains on the estimation of the VC matrix. It would be interesting to pursue these issues in future research.

Appendix A: Some Useful Lemmas

Following lemmas extend the selected lemmas from Lee (2004), Yu et al. (2008), Lin & Lee (2010), and Kelejian & Prucha (2010), which are essential in proving our main results.

Lemma A.1: *For $\mathbb{X}_N(\rho)$ defined in Sec. 2, under Assumptions 1, 3 and 4, the projection matrices, $\mathbb{P}_N(\rho) = \mathbb{X}_N(\rho)[\mathbb{X}'_N(\rho)\mathbb{X}_N(\rho)]^{-1}\mathbb{X}'_N(\rho)$ and $\mathbb{M}_N(\rho) = \mathbf{I}_N - \mathbb{P}_N(\rho)$ and are uniformly bounded in both row and column sums, for each ρ in its compact parameter space.*

Lemma A.2: *Let $\mathbf{A}_N$ and $\mathbf{B}_N$ be $N \times N$ matrices, uniformly bounded in both row and column sums, and $\mathbb{M}_N(\rho)$ be defined in Lemma A.1. Then, we have,*

(i) *the elements of A_N are uniformly bounded,*

(ii) $\mathtt{tr}(\mathbf{A}_N^m) = O(N)$ *for $m \geq 1$,*

(iii) $\mathtt{tr}(\mathbf{A}'_N\mathbf{A}_N) = O(N)$,

(iv) $\mathtt{tr}((\mathbb{M}_N(\rho)\mathbf{A}_N)^m) = \mathtt{tr}(\mathbf{A}_N^m) + O(1)$ *for $m \geq 1$ and each ρ,*

(v) $\mathtt{tr}((\mathbf{A}'_N\mathbb{M}_N(\rho)\mathbf{A}_N)^m) = \mathtt{tr}((\mathbf{A}'_N\mathbf{A}_N)^m) + O(1)$ *for $m \geq 1$ and each ρ,*

(vi) $\mathbf{A}_N\mathbf{B}_N$ *is uniformly bounded in both row and column sums.*

Lemma A.3: *Let $\mathbf{A}_N$ be an $N \times N$ matrix of uniformly bounded column sums, $\mathbf{C}_N$ be an $N \times k$ matrix ($k < N$) of uniformly bounded elements, and $\mathbb{V}_N$ be an $N \times 1$ random vector of independent elements with zero mean, and uniformly bounded third absolute moments. Then,*

(i) $\frac{1}{\sqrt{N}}\mathbf{C}'_N\mathbf{A}_N\mathbb{V} = O_p(1)$ *and* $\frac{1}{N}\mathbf{C}'_N\mathbf{A}_N\mathbb{V} = o_p(1)$,

(ii) $\frac{1}{\sqrt{N}}\mathbf{C}'_N\mathbf{A}_N\mathbb{V} \xrightarrow{D} N(0, \lim_{N\to\infty}\frac{1}{N}\mathbf{C}'_N\mathbf{A}_N\mathbf{H}_N\mathbf{A}'_N\mathbf{C}_N)$, *where $\mathbf{H}_N = \mathrm{Var}(\mathbb{V})$ and the 'limit' is assumed to exist and to be positive definite.*

Lemma A.4 (Moments and Limiting Distribution for Linear Quadratic forms): *Let $\mathbf{B}_{rN}$ be $N \times N$ matrices of uniformly bounded row and column sums, and $\mathbf{c}_{rN}$ be $N \times 1$ vectors with elements c_{ri} satisfying $\sup_N \frac{1}{N}\sum_{i=1}^N |c_{ri}|^{2+\epsilon} < \infty$ for some $\epsilon > 0$. Let $\mathbb{V}_N$ be an $N \times 1$ random vector with elements: $\{v_i\} \sim inid(0, \sigma_0^2 h_i)$, where $h_i > 0$ such that $\frac{1}{N}\sum_{i=1}^N h_i = 1$, and $E|v_i|^{4+\epsilon} < c < \infty$ for all i, for some $\epsilon > 0$ and constant c. Consider the linear-quadratic forms: $\mathbf{Q}_{rN} = \mathbb{V}'_N\mathbf{B}_{rN}\mathbb{V}_N + \mathbf{c}'_{rN}\mathbb{V}_N$, $r = 1, 2$. Denote the diagonal elements of $\mathbf{B}_{rN}$ by $b_{r,ii}$. Let s_i and κ_i be, respectively, the measures of skewness and excess kurtosis of v_i. We have,*

(i) $\mathrm{E}(\mathbf{Q}_{rN}) = \sigma_0^2\mathtt{tr}(\mathbf{H}_N\mathbf{B}_{rN})$, *where* $\mathbf{H}_N = \mathtt{diag}(h_1, \ldots, h_N)$,

(ii) $\mathrm{Var}(\mathbf{Q}_{rN}) = \sigma_0^4\mathtt{tr}[\mathbf{H}_N\mathbf{B}_{rN}(\mathbf{H}_N\mathbf{B}_{rN} + \mathbf{B}'_{rN}\mathbf{H}_N)] + \sigma_0^2\mathbf{c}'_{rN}\mathbf{H}_N\mathbf{c}_{rN}$
$\qquad + \sum_{i=1}^N(\sigma_0^4 b_{r,ii}^2 h_i^2\kappa_i + 2\sigma_0^3 b_{r,ii}c_{ri}h_i^{3/2}s_i)$,

(iii) $\mathrm{Cov}(\mathbf{Q}_{1N}, \mathbf{Q}_{2N}) = 2\sigma_0^4\mathtt{tr}(\mathbf{B}_{1N}\mathbf{H}_N\mathbf{B}_{2N}\mathbf{H}_N) + \sigma_0^2\mathbf{c}'_{1N}\mathbf{H}_N\mathbf{c}_{2N}$
$\qquad + \sum_{i=1}^N\left[\sigma_0^4 b_{1,ii}b_{2,ii}h_i^2\kappa_i + \sigma_0^3(b_{1,ii}c_{2i} + b_{2,ii}c_{1i})h_i^{3/2}s_i\right]$,

(iv) $\mathrm{E}(\mathbf{Q}_{rN}) = O(N)$, $\quad Var(\mathbf{Q}_{rN}) = O(N)$, *and* $\mathbf{Q}_{rN} = O_p(N)$,

(v) $\frac{1}{N}\mathbf{Q}_{rN} - \frac{1}{N}\mathrm{E}(\mathbf{Q}_{rN}) = O_p\big(N^{-\frac{1}{2}}\big)$,

(vi) $\frac{\mathbf{Q}_{rN} - \mathrm{E}(\mathbf{Q}_{rN})}{\sqrt{\mathrm{Var}(\mathbf{Q}_{rN})}} \xrightarrow{D} N(0, 1)$, *and for* $\mathbf{Q}_N = (\mathbf{Q}_{1N}, \mathbf{Q}_{2N})'$,

(vii) $\Sigma_N^{-1/2}(\mathbf{Q}_N - \mathrm{E}(\mathbf{Q}_N)) \xrightarrow{D} N(\mathbf{0}, I_2)$, *where* $\Sigma_N = \mathrm{Var}(\mathbf{Q}_N)$, *and* $\Sigma_N^{1/2}\Sigma_N^{1/2} = \Sigma_N$.

Appendix B: Proofs of Theorems

More on the Robustness of QMLE. We continue on the discussion at the end of Sec. 2.1 to give some more useful details on the nature of **Condition I** and **Condition II**.

First, given δ, $\bar{\ell}_N(\theta) = \mathrm{E}[\ell_N(\theta)]$ is partially maximized at

$$\bar{\beta}_N(\delta) = [\mathbb{X}'_N(\rho)\mathbb{X}_N(\rho)]^{-1}\mathbb{X}'_N(\rho)\mathbf{D}_N(\delta)\mathbf{f}_N, \tag{B-1}$$

$$\bar{\sigma}^2_N(\delta) = \tfrac{1}{N}\mathbf{f}'_N\mathbf{D}'_N(\delta)\mathbb{M}_N(\rho)\mathbf{D}_N(\delta)\mathbf{f}_N + \tfrac{\sigma_0^2}{N}\mathtt{tr}[\mathbf{H}_N\mathbf{D}'^{-1}_N\mathbf{D}'_N(\delta)\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}], \tag{B-2}$$

giving the population counterpart of $\ell^c_N(\delta)$ (see (2.6)) upon substitution:

$$\bar{\ell}^c_N(\delta) = \max_{\beta,\sigma^2}\mathrm{E}[\ell_N(\theta)] = -\tfrac{N}{2}\ln(2\pi+1) + \ln|\mathbf{D}_N(\delta)| - \tfrac{N}{2}\ln(\bar{\sigma}^2_N(\delta)), \tag{B-3}$$

recalling $\mathbf{D}_N(\delta) = I_{T-1}\otimes D_n(\delta), \mathbf{D}_N = \mathbf{D}_N(\delta_0)$ and $\mathbf{f}_N = \mathbf{A}^{-1}_{1N}\mathbf{X}_N\beta_0$. We have $\bar{\sigma}^2_N(\delta_0) = \sigma_0^2$, and $\bar{\sigma}^2_N(\delta) = \sigma^2_n(\delta)\big[1 + \frac{1}{N\sigma^2_n(\delta)}\mathbf{f}'_N\mathbf{D}'_N(\delta)\mathbb{M}_N(\rho)\mathbf{D}_N(\delta)\mathbf{f}_N\big] \equiv \sigma^2_n(\delta)\mu_N(\delta)$. Thus,

$$\bar{\ell}^c_N(\delta) - \bar{\ell}^c_N(\delta_0) = \ln|\mathbf{D}_N(\delta)| - \ln|\mathbf{D}_N| - \tfrac{N}{2}(\ln(\sigma^2_n(\delta)) - \ln(\sigma_0^2)) - \tfrac{N}{2}\ln(\mu_N(\delta)).$$

It can be shown that $\sigma^2_n(\delta)$ (which is the 2nd part of (B-3)) is bounded from below away from 0 (see the proof of 3.1). By the first part of **Condition I**(a), $\frac{1}{N}\mathbf{f}'_N\mathbf{D}'_N(\delta)\mathbb{M}_N(\rho)\mathbf{D}_N(\delta)\mathbf{f}_N > 0$, and thus $\mu_N(\delta) > 1$ for $\lambda \neq \lambda_0$ given any ρ. Now, given λ_0, $\lim_{N\to\infty}\frac{1}{N}[\bar{\ell}^c_N(\lambda_0,\rho) - \bar{\ell}^c_N(\delta_0)] \neq 0$ for $\rho \neq \rho_0$ by the second part of **Condition I**(a). Hence, δ_0 is identified if further: $\lim_{N\to\infty}\frac{1}{N}[\bar{\ell}^c_N(\lambda_0,\rho) - \bar{\ell}^c_N(\delta_0)] \leq 0$ for $\rho \neq \rho_0$, which is a special case of the following.

When **Condition I**(a) fails, $\bar{\ell}^c_N(\delta) - \bar{\ell}^c_N(\delta_0) \neq 0$ $\forall\delta \neq \delta_0$ by **Condition I**(b). To ensure $\bar{\ell}^c_N(\delta) < \bar{\ell}^c_N(\delta_0)$ $\forall\delta \neq \delta_0$, one needs additional conditions so that $\bar{\ell}^c_N(\delta) \leq \bar{\ell}^c_N(\delta_0)$ $\forall\delta \neq \delta_0$. Note that $p_N(\theta_0) = \exp[\ell_N(\theta_0)]$ is the *quasi* joint pdf of $\mathbf{Y}_N$ under $\mathbf{V}_N \sim N(0,\sigma^2 I_N)$. Let $p^0_N(\theta_0)$ be the *true* joint pdf of $\mathbf{Y}_N$ under $\mathbf{V}_N \sim (0,\sigma^2\mathbf{H}_N)$. Let E^q denote the expectation with respect to $p_N(\delta_0)$, to differentiate from the usual notation E that corresponds to $p^0_N(\theta_0)$. Write

$$\mathbf{D}_N(\delta)\mathbf{Y}_N = \mathbf{D}_N(\delta)\mathbf{f}_N + \mathbf{B}_N(\delta)\mathbf{V}_N, \text{ and } \mathbf{V}_N(\beta,\delta) = \mathbf{B}_N(\delta)\mathbf{V}_N + \mathbf{b}_N(\beta,\delta),$$

where $\mathbf{B}_N(\delta) = \mathbf{D}_N(\delta)\mathbf{D}_N^{-1}$ and $\mathbf{b}_N(\beta,\delta) = \mathbf{D}_N(\delta)\mathbf{f}_N - \mathbf{A}_{2N}(\rho)\mathbf{X}_N\beta$. Then, for $\ell_N(\theta)$ in (2.4),

$$\mathrm{E}^q[\ell_N(\theta_0)] = \mathrm{E}[\ell_N(\theta_0)] = -\tfrac{N}{2}\ln(2\pi\sigma^2) + \ln|\mathbf{D}_N| - \tfrac{N}{2}, \text{ as } \tfrac{1}{N}\mathtt{tr}(\mathbf{H}_N) = 1, \text{ and}$$
$$\mathrm{E}^q[\ell_N(\theta)] = -\tfrac{N}{2}\ln(2\pi\sigma^2) + \ln|\mathbf{D}_N(\delta)| - \tfrac{1}{2\sigma^2}[\sigma_0^2\mathtt{tr}(\mathbf{B}'_N(\delta)\mathbf{B}_N(\delta)) + \mathbf{b}'_N(\beta,\delta)\mathbf{b}_N(\beta,\delta)],$$
$$\mathrm{E}[\ell_N(\theta)] = -\tfrac{N}{2}\ln(2\pi\sigma^2) + \ln|\mathbf{D}_N(\delta)| - \tfrac{1}{2\sigma^2}[\sigma_0^2\mathtt{tr}(\mathbf{H}_N\mathbf{B}'_N(\delta)\mathbf{B}_N(\delta)) + \mathbf{b}'_N(\beta,\delta)\mathbf{b}_N(\beta,\delta)].$$

By Jensen's inequality, $\mathrm{E}^q\big[\ln\big(\frac{p_N(\theta)}{p_N(\theta_0)}\big)\big] \leq \ln\mathrm{E}^q\big(\frac{p_N(\theta)}{p_N(\theta_0)}\big) = 0$. **If**, $\mathrm{E}[\ell_N(\theta)] - \mathrm{E}^q[\ell_N(\theta)] = o(N)$, then $\mathrm{E}[\ln p_N(\theta)] \leq \mathrm{E}[\ln p_N(\theta_0)]$, for large enough N. Thus, $\bar{\ell}^c_N(\delta) = \max_{\beta,\sigma^2}\mathrm{E}[\ln p_N(\theta)] \leq \max_{\beta,\sigma^2}\mathrm{E}[\ln p_N(\theta_0)] = \bar{\ell}^c_N(\delta_0), \forall\delta \neq \delta_0$, and N large enough. Clearly,

$$\mathrm{E}[\ell_N(\theta)] - \mathrm{E}^q[\ell_N(\theta)] = \tfrac{\sigma_0^2}{2\sigma^2}[\mathtt{tr}(\mathbf{B}'_N(\delta)\mathbf{B}_N(\delta)) - \mathtt{tr}(\mathbf{H}_N\mathbf{B}'_N(\delta)\mathbf{B}_N(\delta))].$$

Using $\mathbf{A}_{1N}(\lambda) = \mathbf{A}_{1N} + (\lambda_0 - \lambda)\mathbf{W}_{1N}$ and $\mathbf{A}_{2N}(\rho) = \mathbf{A}_{2N} + (\rho_0 - \rho)\mathbf{W}_{2N}$, we have

$$\mathbf{B}_N(\delta) = I_N + (\rho_0 - \rho)\mathbf{G}_{2N} + (\lambda_0 - \lambda)\bar{\mathbf{G}}_{1N} + (\lambda_0 - \lambda)(\rho_0 - \rho)\mathbf{G}_{2N}\bar{\mathbf{G}}_{1N}. \tag{B-4}$$

Using (B-4) it is easy to see that **Condition II** ensures $\mathrm{E}[\ell_N(\theta)]-\mathrm{E}^q[\ell_N(\theta)]=o(N)$. Therefore, if **Condition I** and **Condition II** are met, $\sup_{\delta:d(\delta,\delta_0)>\varepsilon}\bar{\ell}_N^c(\delta)<\bar{\ell}_N^c(\delta_0)$ for every $\varepsilon>0$, i.e., δ_0 are uniquely identified by the QML estimation. Finally, it can be seen that the uniform convergence, $\sup_{\delta\in\Delta}\frac{1}{N}|\ell_N^c(\delta)-\bar{\ell}_N^c(\delta)|\xrightarrow{p}0$, also requires **Condition II**.

Proof of Theorem 3.1: Proof of consistency. Let $\bar{\psi}_N(\theta)=\mathrm{E}[\psi_N(\theta)]$, the population counterpart of the joint estimating function $\psi_N(\theta)$ given in (3.2). Given δ, $\bar{\psi}_N(\theta)$ is partially solved at $\bar{\beta}_N(\delta)$ and $\bar{\sigma}_N^2(\delta)$, given in (B-1) and (B-2). Plugging $\bar{\beta}_N(\delta)$ and $\bar{\sigma}_N^2(\delta)$ back into the λ- and ρ-components $\bar{\psi}_N(\theta)$, we get the population counterpart of $\tilde{\psi}_N^c(\delta)$:

$$\bar{\psi}_N^c(\delta)=\begin{cases}\frac{1}{\bar{\sigma}_N^2(\delta)}\mathrm{E}\big\{\mathbf{V}(\bar{\beta}_N(\delta),\delta)'[\boldsymbol{\eta}_N(\bar{\beta}_N(\delta),\delta)+\bar{\mathbf{G}}_{1N}^{\circ}(\delta)\mathbf{V}(\bar{\beta}_N(\delta),\delta)]\big\}\\ \frac{1}{\bar{\sigma}_N^2(\delta)}\mathrm{E}\big\{\mathbf{V}(\bar{\beta}_N(\delta),\delta)'\mathbf{G}_{2N}^{\circ}(\rho)\mathbf{V}(\bar{\beta}_N(\delta),\delta)\big\}.\end{cases}$$

where $\mathbf{V}(\bar{\beta}_N(\delta),\delta)=\mathbb{Y}_N(\delta)-\mathbb{X}_N(\rho)\bar{\beta}_N(\delta)$. Working on the numerators of $\bar{\psi}_N^c(\delta)$ and dropping the terms of smaller order, we arrive at $F_N(\delta)$ given in Assumption 6, which shows that the identification uniqueness condition of Theorem 5.9 of van der Vaart (1998) holds, i.e., for every $\epsilon>0$, $\inf_{\delta:d(\delta,\delta_0)\geq\epsilon}\frac{1}{N}\|\bar{\psi}_N^c(\delta)\|>0=\frac{1}{N}\|\bar{\psi}_N^c(\delta_0)\|$, provided that $\bar{\sigma}_N^2(\delta)$ is bounded from below away from zero. Then, $\hat{\delta}_{\mathtt{AQS1}}$ is consistent if the uniform convergence condition of Theorem 5.9 of van der Vaart (1998) holds, i.e., $\sup_{\delta\in\Delta}\frac{1}{N}\|\tilde{\psi}_N^c(\delta)-\bar{\psi}_N^c(\delta)\|=o_p(1)$. These amount to show

- **(a)** $\bar{\sigma}_N^2(\delta)$ is bounded from below away from zero;
- **(b)** $\sup_{\delta\in\Delta}|\tilde{\sigma}_N^2(\delta)-\bar{\sigma}_N^2(\delta)|=o_p(1)$, uniformly in $\delta\in\Delta$;
- **(c)** $\sup_{\delta\in\Delta}\frac{1}{N}\big|\mathbf{V}(\tilde{\beta}_N(\delta),\delta)'\boldsymbol{\eta}_N(\tilde{\beta}_N(\delta),\delta)-\mathrm{E}[\mathbf{V}(\bar{\beta}_N(\delta),\delta)]'\boldsymbol{\eta}_N(\bar{\beta}_N(\delta),\delta)\big|=o_p(1)$;
- **(d)** $\sup_{\delta\in\Delta}\frac{1}{N}\big|\mathbf{V}(\tilde{\beta}_N(\delta),\delta)'\bar{\mathbf{G}}_{1N}^{\circ}(\delta)\mathbf{V}(\tilde{\beta}_N(\delta),\delta)-\mathrm{E}[\mathbf{V}(\bar{\beta}_N(\delta),\delta)'\bar{\mathbf{G}}_{1N}^{\circ}(\delta)\mathbf{V}(\bar{\beta}_N(\delta),\delta)]\big|=o_p(1)$;
- **(e)** $\sup_{\delta\in\Delta}\frac{1}{N}\big|\mathbf{V}(\tilde{\beta}_N(\delta),\delta)'\mathbf{G}_{2N}^{\circ}(\delta)\mathbf{V}(\tilde{\beta}_N(\delta),\delta)-\mathrm{E}[\mathbf{V}(\bar{\beta}_N(\delta),\delta)'\mathbf{G}_{2N}^{\circ}(\delta)\mathbf{V}(\bar{\beta}_N(\delta),\delta)]\big|=o_p(1)$;

where $\mathbf{V}(\tilde{\beta}_N(\delta),\delta)=\mathbb{Y}_N(\delta)-\mathbb{X}_N(\rho)\tilde{\beta}_N(\delta)=\mathbb{M}_N(\rho)\mathbb{Y}_N(\delta)$, following the notation defined between (2.4) and (2.6). Similarly, $\mathbf{V}(\bar{\beta}_N(\delta),\delta)=\mathbb{Y}_N(\delta)-[\mathbf{I}_N-\mathbb{M}_N(\rho)]\mathrm{E}[\mathbb{Y}_N(\delta)]$.

For condition (a), from (B-2), it is obvious that the first term of $\bar{\sigma}_N^2(\delta)$ is nonnegative. It suffices to show that the second term, which is $\sigma_n^2(\delta)$ defined in **Condition I**, is uniformly bounded from below away from zero. Consider the model with $\beta_0=0$ and $H_n=I_n$. We have the loglikelihood: $\ell_N^*(\theta)=-\frac{N}{2}\ln(2\pi\sigma^2)+\ln|\mathbf{D}_N(\delta)|-\frac{1}{2\sigma^2}\mathbb{Y}_N'(\delta)\mathbb{Y}_N(\delta)$ and $\bar{\ell}_N^*(\delta)=\max_{\sigma^2}\mathrm{E}[\ell_N^*(\theta)]=const.-\frac{N}{2}\ln(\sigma_{\mathrm{o}n}^2(\delta))+\ln|\mathbf{D}_N(\delta)|$, where $\sigma_{\mathrm{o}n}^2(\delta)=\frac{\sigma_0^2}{n}\mathtt{tr}[D_n'^{-1}D_n'(\delta)D_n(\delta)D_n^{-1}]$. As $D_n'^{-1}D_n'(\delta)D_n(\delta)D_n^{-1}$ is positive semidefinite (p.s.d.), $\sigma_{\mathrm{o}n}^2(\delta)\geq0$. By Jensen's inequality, $\bar{\ell}_N^*(\delta)\leq\max_{\sigma^2}\mathrm{E}[\ell_N^*(\theta_0)]=\bar{\ell}_N^*(\delta_0)$, implying $-\ln(\sigma_n^2(\delta))\leq-\ln(\sigma_0^2)+\frac{2}{N}\ln|\mathbf{D}_N|-\frac{2}{N}\ln|\mathbf{D}_N(\delta)|=O(1)$ by Lemma A.2 and the fact that σ_0^2 is bounded away from 0. Thus, $-\ln(\sigma_{\mathrm{o}n}^2(\delta))$ is bounded from above, implying $\sigma_{\mathrm{o}n}^2(\delta)\neq0$. Therefore, $\sigma_{\mathrm{o}n}^2(\delta)$ is bounded from below away from 0. Finally, $\sigma_n^2(\delta)=\frac{\sigma_0^2}{n}\mathtt{tr}[H_nD_n'^{-1}D_n'(\delta)D_n(\delta)D_n^{-1}]\geq\min(h_i)\sigma_{\mathrm{o}n}^2(\delta)\geq c>0$, as $D_n'^{-1}D_n'(\delta)D_n(\delta)D_n^{-1}$ is p.s.d., and H_n is a diagonal matrix with strictly positive elements.

For condition (b), using $\mathbb{Y}_N(\delta)=\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}(\mathbf{A}_{2N}\mathbf{X}_N\beta_0+(F_{T,T-1}'\otimes I_n)\mathbb{V}_{nT})$, where $\mathbb{V}_{nT}$

is the $nT \times 1$ vector of original errors, we can write $\tilde{\sigma}_N^2(\delta) = \frac{1}{N}\mathbb{Y}'_N(\delta)\mathbb{M}_N(\rho)\mathbb{Y}_N(\delta)$ as

$$\begin{aligned}\tilde{\sigma}_N^2(\delta) = & \tfrac{1}{N}\mathbf{f}'_N\mathbf{D}'_N(\delta)\mathbb{M}_N(\rho)\mathbf{D}_N(\delta)\mathbf{f}_N + \tfrac{2}{N}\mathbf{f}'_N\mathbf{D}'_N(\delta)\mathbb{M}_N(\rho)\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}(F'_{T,T-1}\otimes I_n)\mathbb{V}_{nT} \\ & + \tfrac{1}{N}\mathbb{V}'_{nT}(F_{T,T-1}\otimes I_n)\mathbf{D}'^{-1}_N\mathbf{D}'_N(\delta)\mathbb{M}_N(\rho)\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}(F'_{T,T-1}\otimes I_n)\mathbb{V}_{nT},\end{aligned}$$

giving $\tilde{\sigma}_N^2(\delta)-\bar{\sigma}_N^2(\delta) = Q_1+Q_2-\sigma_N^2(\delta)$, where $Q_1(\delta) = \frac{2}{N}\mathbf{f}'_N\mathbf{D}'_N(\delta)\mathbb{M}_N(\rho)\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}(F'_{T,T-1}\otimes I_n)\mathbb{V}_{nT}$, and $Q_2(\delta) = \frac{1}{N}\mathbb{V}'_{nT}(F_{T,T-1}\otimes I_n)\mathbf{D}'^{-1}_N\mathbf{D}'_N(\delta)\mathbb{M}_N(\rho)\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}(F'_{T,T-1}\otimes I_n)\mathbb{V}_{nT}$.

For $Q_1(\delta)$, it is easy to see that, under Assumptions 3-5 and by Lemma A.2, the elements of $\mathbf{f}'_N\mathbf{D}'_N(\delta)\mathbb{M}_N(\rho)\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}(F'_{T,T-1}\otimes I_n)$ are uniformly bounded for each $\delta\in\Delta$, the pointwise convergence, $Q_1(\delta)\xrightarrow{p}0$, therefore follows from Lemma A.3. For $Q_2(\delta)$, under Assumptions 3-5 and by Lemma A.2(v), $\mathtt{tr}[\mathbf{D}'^{-1}_N\mathbf{D}'_N(\delta)\mathbb{M}_N(\rho)\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}] = \mathtt{tr}[\mathbf{D}'^{-1}_N\mathbf{D}'_N(\delta)\mathbf{D}_N(\delta)\mathbf{D}_N^{-1}]+O(1)$. It follows that, by Lemma A.4(v), $Q_2(\delta)-\sigma_N^2(\delta)\xrightarrow{p}0$, for each $\delta\in\Delta$.

To show that $Q_r(\delta), r=1,2$, are stochastically equicontinuous, let δ_1 and δ_2 be two points in Δ. We have by the mean value theorem:

$$Q_r(\delta_2)-Q_r(\delta_1) = \tfrac{\partial}{\partial\delta'}Q_r(\bar{\delta})(\delta_2-\delta_1),\ r=1,2,$$

where $\bar{\delta}$ lies between δ_1 and δ_2 elementwise. It is easy to show that $\sup_{\delta\in\Delta}|\frac{\partial}{\partial\lambda}Q_r(\delta)| = O_p(1)$, by Assumptions 1, 3, 4, and 5, and Lemma A.2, as $Q_r(\delta)$ are linear or quadratic in λ by the expression $\mathbf{D}_N(\delta)\mathbf{D}_N^{-1} = I_N+(\rho_0-\rho)\mathbf{G}_{2N}+(\lambda_0-\lambda)\bar{\mathbf{G}}_{1N}+(\lambda_0-\lambda)(\rho_0-\rho)\mathbf{G}_{2N}\bar{\mathbf{G}}_{1N}$. Now, ρ appears in $Q_r(\delta)$ nonlinearly only through $\mathbb{M}_N(\rho)$. It is easy to show that $\frac{\partial}{\partial\rho}\mathbb{M}_N(\rho)$ is uniformly bounded in both row and column sums by Lemma A.2, uniformly in ρ in its compact space, and that $\sup_{\delta\in\Delta}|\frac{\partial}{\partial\rho}Q_r(\delta)| = O_p(1)$. Therefore, $Q_r(\delta), r=1,2$, are stochastically equicontinuous. The pointwise convergence and stochastic equicontinuity imply that $Q_r(\delta)-\mathrm{E}[Q_r(\delta)]\xrightarrow{p}0$, uniformly in $\delta\in\Delta, r=1,2$, leading to condition (b) (Newey, 1991).

For condition (c), we have $\boldsymbol{\eta}_N(\tilde{\beta}_N(\delta),\delta) = \bar{\mathbf{G}}_{1N}(\delta)\mathbb{P}_N(\rho)\mathbb{Y}_N(\delta)$ and $\boldsymbol{\eta}_N(\bar{\beta}_N(\delta),\delta) = \bar{\mathbf{G}}_{1N}(\delta)\mathbb{P}_N(\rho)\mathrm{E}[\mathbb{Y}_N(\delta)]$. With $\mathbf{V}(\tilde{\beta}_N(\delta),\delta) = \mathbb{M}_N(\rho)\mathbb{Y}_N(\delta)$ and $\mathbb{Y}_N(\delta) = \mathbf{D}_N(\delta)\mathbf{D}_N^{-1}(\mathbf{A}_{2N}\mathbf{X}_N\beta_0+(F'_{T,T-1}\otimes I_n)\mathbb{V}_{nT})$, we see that $\frac{1}{N}\{\mathbf{V}(\tilde{\beta}_N(\delta),\delta)'\boldsymbol{\eta}_N(\tilde{\beta}_N(\delta),\delta)-\mathrm{E}[\mathbf{V}(\bar{\beta}_N(\delta),\delta)]'\boldsymbol{\eta}_N(\bar{\beta}_N(\delta),\delta)\}$ is of the linear-quadratic form: $\mathbb{V}'_{nT}\mathbb{A}_{nT}(\delta)\mathbb{V}_{nT}+\mathbf{c}'_{nT}(\delta)\mathbb{V}_{nT}$, for suitably defined matrix $\mathbb{A}_{nT}(\delta)$ and vector $\mathbf{c}_{nT}(\delta)$. Its pointwise convergence follows from Lemma A.4(v), and uniform convergence is proved in a similar way as that for (b), based on the theorem of Newey (1991).

For condition (d), again with the expressions for $\mathbf{V}(\tilde{\beta}_N(\delta),\delta)$ and $\mathbb{Y}_N(\delta)$, we can write $\frac{1}{N}\{\mathbf{V}(\tilde{\beta}_N(\delta),\delta)'\bar{\mathbf{G}}^{\circ}_{1N}(\delta)\mathbf{V}(\tilde{\beta}_N(\delta),\delta)-\mathrm{E}[\mathbf{V}(\bar{\beta}_N(\delta),\delta)'\bar{\mathbf{G}}^{\circ}_{1N}(\delta)\mathbf{V}(\bar{\beta}_N(\delta),\delta)]\}$ as a linear-quadratic form in $\mathbb{V}_{nT}$, and the proof of uniform convergence proceeds similarly.

For condition (e), similar to the proof of (d).

Proof of asymptotic normality. First note that $\mathrm{tr}(H_n)=n$. By the mean value theorem,

$$\sqrt{N}(\hat{\theta}_{\mathtt{AQS1}}-\theta_0) = -\big[\tfrac{1}{N}\tfrac{\partial}{\partial\theta'}\psi_N(\tilde{\theta})\big]^{-1}\tfrac{1}{\sqrt{N}}\psi_N(\theta_0),$$

where $\tilde{\theta}$ lies element-wise between $\hat{\theta}_{\mathtt{AQS1}}$ and θ_0. It mounts to show that,

(i) $\frac{1}{\sqrt{N}}\psi_N(\theta_0)\xrightarrow{D}N(0,\ \lim_{N\to\infty}\Omega_N)$, where $\Omega_N = \frac{1}{N}\mathrm{Var}[\psi_N(\theta_0)]$

(*ii*) $\frac{1}{N}\left[\frac{\partial}{\partial\theta'}\psi_N(\tilde{\theta})-\frac{\partial}{\partial\theta'}\psi_N(\theta_0)\right]=o_p(1)$, and

(*iii*) $\frac{1}{N}\left[\frac{\partial}{\partial\theta'}\psi_N(\theta_0)-\mathrm{E}(\frac{\partial}{\partial\theta'}\psi_N(\theta_0))\right]=o_p(1)$.

As argued above Theorem 3.1, the components of $\psi_N(\theta_0)$ are linear or linear-quadratic forms in the original error vector $\mathbb{V}_{nT}$ since $\mathbf{V}_N=(f'_{T,T-1}\otimes I_n)\mathbb{V}_{nT}$. Assumptions 1-5 ensures that every fixed linear combination of $\frac{1}{\sqrt{N}}\psi_N(\theta_0)$ satisfies the conditions of the central limit theorem (CLT) for linear-quadratic (LQ) forms of Kelejian and Prucha (2001) and hence is asymptotically normal. Therefore, Cramér-Wold device leads to $\frac{1}{\sqrt{N}}\psi_N(\theta_0)\xrightarrow{D}N(0,\lim_{N\to\infty}\Omega_N)$.

For condition (*ii*)**:** letting $\mathcal{H}_N(\theta)=-\frac{1}{N}\frac{\partial}{\partial\theta'}\psi_N(\theta)$ and denoting $A_n^s=A_n+A_n'$ for a matrix A_n, we have the expression for $N\sigma^2\mathcal{H}_N(\theta)$:

$$\begin{pmatrix} \mathbb{X}'_N(\rho)\mathbb{X}_N(\rho), & \frac{1}{\sigma^2}\mathbb{X}'_N(\rho)\mathbf{V}_N(\beta,\delta), & \mathbb{X}'_N(\rho)\mathbf{\Pi}_{1N}, & \mathbb{X}'_N(\rho)\mathbf{G}^s_{2N}(\rho)\mathbf{V}_N(\beta,\delta) \\ \frac{1}{\sigma^2}\mathbf{V}'_N(\beta,\delta)\mathbb{X}_N(\rho), & \frac{1}{\sigma^4}\|\mathbf{V}_N(\beta,\delta)\|^2-\frac{N}{2\sigma^2}, & \frac{1}{\sigma^2}\mathbf{V}'_N(\beta,\delta)\mathbf{\Pi}_{1N}, & \frac{1}{\sigma^2}\mathbf{V}'_N(\beta,\delta)\mathbf{\Pi}_{2N} \\ \mathbf{\Pi}^{\circ\prime}_{1N}\mathbb{X}_N(\rho), & \frac{1}{\sigma^2}\mathbf{V}'_N(\beta,\delta)\mathbf{\Pi}^\circ_{1N}, & \mathcal{H}_{33}(\beta,\delta), & \mathcal{H}_{34}(\beta,\delta) \\ \mathbf{V}'_N(\beta,\delta)\mathbf{G}^{\circ s}_{2N}(\rho)\mathbb{X}_N(\rho), & \frac{1}{\sigma^2}\mathbf{V}'_N(\beta,\delta)\mathbf{\Pi}^\circ_{2N}, & \mathcal{H}_{43}(\beta,\delta), & \mathcal{H}_{44}(\beta,\delta) \end{pmatrix},$$

where $\mathcal{H}_{33}(\beta,\delta)=\mathbf{\Pi}^{\circ\prime}_{1N}\mathbf{\Pi}^\circ_{1N}+\mathbf{V}'_N(\beta,\delta)\dot{\mathbf{\Pi}}^\circ_{1N}$, $\mathcal{H}_{43}(\beta,\delta)=\mathbf{\Pi}^{\circ\prime}_{1N}\mathbf{\Pi}^\circ_{2N}+\mathbf{V}'_N(\beta,\delta)\mathbf{G}^{\circ s}_{2N}(\rho)\mathbf{V}_N(\beta,\delta)=\mathcal{H}'_{34}(\beta,\delta)$, $\mathcal{H}_{44}(\beta,\delta)=\mathbf{V}'_N(\beta,\delta)\mathbf{G}^{\circ s}_{2N}(\rho)\mathbf{V}_N(\beta,\delta)$, $\mathbf{\Pi}_{1N}=\boldsymbol{\eta}_N(\beta,\delta)+\bar{\mathbf{G}}_{1N}(\delta)\mathbf{V}_N(\beta,\delta)$, $\dot{\mathbf{\Pi}}_{1N}=\frac{\partial}{\partial\lambda}\mathbf{\Pi}_{1N}$, $\mathbf{\Pi}_{2N}=\mathbf{G}_{2N}(\rho)\mathbf{V}_N(\beta,\delta)$, $\mathbf{\Pi}^\circ_{1N}=\boldsymbol{\eta}_N(\beta,\delta)+\bar{\mathbf{G}}^\circ_{1N}(\delta)\mathbf{V}_N(\beta,\delta)$, and $\mathbf{\Pi}^\circ_{2N}=\mathbf{G}^\circ_{2N}(\rho)\mathbf{V}_N(\beta,\delta)$.

By Assumptions 3-5, Lemma A.2-A.3, and the following facts: $\tilde{\theta}-\theta_0=o_p(1)$, $\mathbf{V}_N(\tilde{\beta},\tilde{\delta})=\mathbf{A}_{2N}\mathbf{X}_N(\beta_0-\tilde{\beta})+(\lambda_0-\tilde{\lambda})\mathbf{A}_{2N}\mathbf{W}_{1N}\mathbf{Y}_N+(\rho_0-\tilde{\rho})\mathbf{W}_{2N}\mathbf{A}_{1N}\mathbf{Y}_N+(\lambda_0-\tilde{\lambda})(\rho_0-\tilde{\rho})\mathbf{W}_{2N}\mathbf{W}_{1N}\mathbf{Y}_N-(\rho_0-\tilde{\rho})\mathbf{W}_{2N}\mathbf{X}_N\tilde{\beta}+\mathbf{V}_N$, $\frac{1}{N}\mathbf{V}'_N(\tilde{\beta},\tilde{\delta})\mathbf{V}_N(\tilde{\beta},\tilde{\delta})=\frac{1}{N}\mathbf{V}'_N\mathbf{V}_N+o_p(1)$, and the $\boldsymbol{\eta}_N$ and $\mathbf{G}$-quantities are all smooth functions of β and δ, it is straightforward but tedious to show that each term in $\mathcal{H}_N(\tilde{\theta})-\mathcal{H}_N(\theta_0)$ is $o_p(1)$. We thus omit the details.

For condition (*iii*)**,** recall $\Phi_N=\mathrm{E}[\mathcal{H}_N(\theta_0)]$. We have

$$\Phi_N=\frac{1}{N\sigma_0^2}\begin{pmatrix} \mathbb{X}'_N\mathbb{X}_N, & \sim, & \sim, & \sim \\ 0, & \frac{N}{2\sigma_0^2}, & \mathtt{tr}(\mathbf{H}_N\bar{\mathbf{G}}_{1N}), & \mathtt{tr}(\mathbf{H}_N\mathbf{G}_{2N}) \\ \boldsymbol{\eta}'_N\mathbb{X}_N & 0, & \boldsymbol{\eta}'_N\boldsymbol{\eta}_N+\sigma_0^2\mathtt{tr}(\mathbf{H}_N\bar{\mathbf{G}}^{\circ s}_{1N}\bar{\mathbf{G}}^\circ_{1N}), & \sim \\ 0, & 0, & \sigma_0^2\mathtt{tr}(\mathbf{H}_N\mathbf{G}^{\circ s}_{2N}\bar{\mathbf{G}}^\circ_{1N}), & \sigma_0^2\mathtt{tr}(\mathbf{H}_N\mathbf{G}^{\circ s}_{2N}\mathbf{G}^\circ_{2N}) \end{pmatrix}.$$

By Lemma A.4 and $\mathbf{V}_N=(F'_{T,T-1}\otimes I_n)\mathbb{V}_{nT}$, we have, $\mathrm{Var}[\frac{1}{N}(\mathbf{V}'_N\mathbf{B}_N\mathbf{V}_N+c'_N\mathbf{V}_N)]=o(1)$ for any $N\times N$ matrix $\mathbf{B}_N$ and $N\times 1$ vector c_N satisfying the conditions of Lemma A.4. By these results and Chebyshev inequality, we can show that all the terms in $\mathcal{H}_N(\theta_0)-\Phi_N$ are $o_p(1)$. ■

Proof of Theorem 3.2: The result $\widehat{\Phi}_{\mathtt{AQS1}}-\Phi_N\xrightarrow{p}0$ follows from the results (*ii*) and (*iii*) in the proof of asymptotic normality part of the proof of Theorem 3.1. This result holds irrespective of whether the errors are normal or non-normal, and T is small or large.

To show $\widehat{\Omega}^\dagger_{\mathtt{AQS1}}-\Omega_N\xrightarrow{p}0$, we first prove the following general result:

$$\frac{1}{N}\sum_{j=1}^N[\hat{\mathbf{s}}_{N,j}\hat{\mathbf{s}}'_{N,j}-\mathrm{E}(\mathbf{s}_{N,j}\mathbf{s}'_{N,j})]\xrightarrow{p}0. \tag{B-5}$$

Under normality, $\Omega_N=\frac{1}{N}\sum_{j=1}^N\mathrm{E}(\mathbf{s}_{N,j}\mathbf{s}'_{N,j})$ and therefore (B-5) already gives the desired result.

The proof of (B-5) is relatively simple, as in this case the transformed errors $\mathbf{v}_j$ are inid normal, and hence $\{\mathbf{s}_{N,j}, \mathcal{F}_{N,j}\}$ form an MD sequence. See the proof of Theorem 3.5 for details.

Under non-normality, the proof of (B-5) is not trivial, and therefore for the proof of this theorem we concentrate on the case of non-normal errors. First, we prove (B-5) by showing $\frac{1}{N}\sum_{j=1}^{N}(\hat{\mathbf{s}}_{N,j}\hat{\mathbf{s}}_{N,j}' - \mathbf{s}_{N,j}\mathbf{s}_{N,j}') \xrightarrow{p} 0$, and $\frac{1}{N}\sum_{j=1}^{N}[\mathbf{s}_{N,j}\mathbf{s}_{N,j}' - \mathrm{E}(\mathbf{s}_{N,j}\mathbf{s}_{N,j}')] \xrightarrow{p} 0$. The proof of the former is trivial by applying the mean value theorem, due to the consistency of the parameter estimates. We focus on the proof of the latter result. To facilitate the proofs, we freely switching between the single index j for the combined unit and time, and the double indices (i,t) for unit i and time t. Recall v_{it} are the original errors and v_{it}^* are the transformed errors, and v_t^* is the $n\times 1$ vector of transformed errors for period t. As $\mathbf{s}_{N,j}$ or $\mathbf{s}_{N,it}$ contains only two types of quantities: $\Pi_{N,j}\mathbf{v}_{N,j}$ and $\mathbf{v}_{N,j}\boldsymbol{\zeta}_{N,j}^{\circ}$ or $\Pi_{it}\mathbf{v}_{it}^*$ and $\mathbf{v}_{it}^*\boldsymbol{\zeta}_{it}^{\circ}$, it suffices to show

(*a*) $\frac{1}{N}\sum_{j=1}^{N}[\Pi_{N,j}\Pi_{N,j}'(\mathbf{v}_{N,j}^2 - \mathrm{E}\mathbf{v}_{N,j}^2)] \xrightarrow{p} 0$,
(*b*) $\frac{1}{N}\sum_{j=1}^{N}[\Pi_{N,j}(\mathbf{v}_{N,j}^2\boldsymbol{\zeta}_{N,j}^{\circ} - \mathrm{E}(\mathbf{v}_{N,j}^2\boldsymbol{\zeta}_{N,j}^{\circ}))] \xrightarrow{p} 0$, and
(*c*) $\frac{1}{N}\sum_{j=1}^{N}[(\mathbf{v}_{N,j}\boldsymbol{\zeta}_{N,j}^{\circ})^2 - \mathrm{E}((\mathbf{v}_{N,j}\boldsymbol{\zeta}_{N,j}^{\circ})^2))] \xrightarrow{p} 0$.

To show (*a*), we have $\frac{1}{N}\sum_{j=1}^{N}[\Pi_{N,j}\Pi_{N,j}'(\mathbf{v}_{N,j}^2 - \mathrm{E}(\mathbf{v}_{N,j}^2))] = \frac{1}{T-1}\sum_{t=1}^{T-1}\{\frac{1}{n}\sum_{i=1}^{n}[\Pi_{it}\Pi_{it}'(v_{it}^{*2} - \mathrm{E}(v_{it}^{*2}))]\} \equiv \frac{1}{T-1}\sum_{t=1}^{T-1}P_{nt}$. For each t, v_{it}^* are independent over i, and thus $\{v_{it}^{*2} - \mathrm{E}(v_{it}^{*2})\}$ form an MD sequence. The weak law of large numbers (WLLN) for MD arrays of Davidson (1994, p.299) leads to $P_{nt} \xrightarrow{p} 0$. Thus, $\frac{1}{T-1}\sum_{t=1}^{T-1}P_{nt} \xrightarrow{p} 0$, as $n\to\infty$ and then $T\to\infty$.

To show (*b*), note that $\boldsymbol{\zeta}_N^{\circ} = \mathbf{B}_N\mathbf{V}_N$ by definition given above (3.6), where $\mathbf{B}_N$ is a strictly lower triangular matrix. Decompose $\boldsymbol{\zeta}_N^{\circ}$ into $\{\boldsymbol{\zeta}_t^{\circ}\}$ and $\mathbf{B}_N$ into $\{\mathbf{B}_{ts}\}$, $t,s = 1,\ldots,T-2$. Note that $\mathbf{B}_{ts}$ is a zero matrix if $s>t$, a strictly lower triangular matrix if $s=t$ and a full $n\times n$ matrix if $s<t$. We have, $\frac{1}{N}\sum_{j=1}^{N}[\Pi_{N,j}(\mathbf{v}_{N,j}^2\boldsymbol{\zeta}_{N,j}^{\circ} - \mathrm{E}(\mathbf{v}_{N,j}^2\boldsymbol{\zeta}_{N,j}^{\circ}))] = \frac{1}{T-1}\sum_{t=1}^{T-1}\{\frac{1}{n}\sum_{i=1}^{n}[\Pi_{it}(v_{it}^{*2}\zeta_{it}^{\circ} - \mathrm{E}(v_{it}^{*2}\zeta_{it}^{\circ}))]\} \equiv \frac{1}{T-1}\sum_{t=1}^{T-1}Q_{nt}$. We shall show that for each t, $Q_{nt} \xrightarrow{p} 0$. First, we have,

$$Q_{n1} = \frac{1}{n}\sum_{i=1}^{n}[\Pi_{i1}((v_{i1}^{*2} - \sigma_0^2 h_i)\boldsymbol{\zeta}_{i1}^{\circ} + \sigma_0^2 h_i\boldsymbol{\zeta}_{i1}^{\circ} - \mathrm{E}(v_{i1}^{*2}\boldsymbol{\zeta}_{i1}^{\circ}))] = Q_{n1}^a + Q_{n1}^b.$$

Let $\mathcal{G}_{n,i}$ be the increasing σ-field generated by $(v_{1\cdot},\ldots,v_{i\cdot})$, where $v_{i\cdot}$ is the $T\times 1$ vector of the original idiosyncratic errors corresponding to the ith spatial unit. As $\boldsymbol{\zeta}_{i1}^{\circ}$ is $\mathcal{G}_{n,i-1}$-measurable, $\mathrm{E}[(v_{i1}^{*2} - \sigma_0^2 h_i)\boldsymbol{\zeta}_{i1}^{\circ}|\mathcal{G}_{n,i-1}] = 0$. Thus, $Q_{n1}^a = \frac{1}{n}\sum_{i=1}^{n}\Pi_{i1}(v_{i1}^{*2} - \sigma_0^2 h_i)\boldsymbol{\zeta}_{i1}^{\circ}$ is the sum of an MD array. By the WLLN for MD arrays of (Davidson, 1994, p.299), $Q_{n1}^a \xrightarrow{p} 0$. Now, as $\mathrm{E}(v_{i1}^{*2}\boldsymbol{\zeta}_{i1}^{\circ})) = 0$, $Q_{n1}^b = \frac{\sigma_0^2}{n}\sum_{i=1}^{n}\Pi_{i1}h_i\boldsymbol{\zeta}_{i1}^{\circ}$. Then, $Q_{n1}^b = \frac{\sigma_0^2}{n}\Pi_1' H_n\boldsymbol{\zeta}_1^{\circ} = \frac{\sigma_0^2}{n}\Pi_1' H_n\mathbf{B}_{11}v_1^* \xrightarrow{p} 0$, by Assumptions 2-5, Lemma A.2 and Chebyshev's WLLN (Serfling, 1980, p.27). Therefore, $Q_{n1} \xrightarrow{p} 0$.

Next, to show $Q_{n2} \xrightarrow{p} 0$, first note that $\boldsymbol{\zeta}_2^{\circ} = \mathbf{B}_{21}v_1^* + \mathbf{B}_{22}v_2^* = (\mathbf{B}_{21}^u + \mathbf{B}_{21}^l + \mathbf{B}_{21}^d)v_1^* + \mathbf{B}_{22}v_2^* = \boldsymbol{\zeta}_{2,1}^{\circ u} + \boldsymbol{\zeta}_{2,1}^{\circ l} + \boldsymbol{\zeta}_{2,1}^{\circ d} + \boldsymbol{\zeta}_{2,2}^{\circ}$. We have, $Q_{n2} = \frac{1}{n}\sum_{i=1}^{n}[\Pi_{i2}(v_{i2}^{*2}\boldsymbol{\zeta}_{i2}^{\circ} - \mathrm{E}(v_{i2}^{*2}\boldsymbol{\zeta}_{i2}^{\circ}))] = \sum_{r=1}^{4}Q_{n2}^{(r)}$, where

$$Q_{n2}^{(1)} = \frac{1}{n}\sum_{i=1}^{n}[\Pi_{i1}(v_{i2}^{*2} - \sigma_0^2 h_i)\boldsymbol{\zeta}_{i2,1}^{\circ u}] + \frac{\sigma_0^2}{n}\sum_{i=1}^{n}h_i\boldsymbol{\zeta}_{i2,1}^{\circ u},$$
$$Q_{n2}^{(2)} = \frac{1}{n}\sum_{i=1}^{n}[\Pi_{i1}(v_{i2}^{*2} - \sigma_0^2 h_i)\boldsymbol{\zeta}_{i2,1}^{\circ l}] + \frac{\sigma_0^2}{n}\sum_{i=1}^{n}h_i\boldsymbol{\zeta}_{i2,1}^{\circ l},$$
$$Q_{n2}^{(3)} = \frac{1}{n}\sum_{i=1}^{n}[\Pi_{i1}(v_{i2}^{*2}\boldsymbol{\zeta}_{i2,1}^{\circ d} - \mathrm{E}(v_{i2}^{*2}\boldsymbol{\zeta}_{i2,1}^{\circ d})].$$
$$Q_{n2}^{(4)} = \frac{1}{n}\sum_{i=1}^{n}[\Pi_{i1}(v_{i2}^{*2} - \sigma_0^2 h_i)\boldsymbol{\zeta}_{i2,2}^{\circ}] + \frac{\sigma_0^2}{n}\sum_{i=1}^{n}h_i\boldsymbol{\zeta}_{i2,2}^{\circ},$$

The first terms of $Q_{n2}^{(1)}$ and $Q_{n2}^{(4)}$ are like Q_{n1}^a and their second terms are like Q_{n1}^b; thus $Q_{n2}^{(2)} =$

$o_p(1)$ and $Q_{n2}^{(4)} = o_p(1)$. As $\boldsymbol{\zeta}_{i2,1}^{\circ u}$ is $\bar{\mathcal{G}}_{n,i+1}$-measurable, where $\bar{\mathcal{G}}_{n,i}$ is a decreasing σ-field generated by $(v_{i\cdot}, \dots, v_{n\cdot})$, $\frac{1}{n}\sum_{i=1}^{n}[\Pi_{i1}(v_{i2}^{*2} - \sigma_0^2 h_i)\boldsymbol{\zeta}_{i2,1}^{\circ u}]$ is the sum of an MD sequence, shown to be $o_p(1)$. That $\frac{\sigma_0^2}{n}\sum_{i=1}^{n} h_i \boldsymbol{\zeta}_{i2,1}^{\circ u}$ is $o_p(1)$ follows from some similar arguments for Q_{n1}^b. Thus, $Q_{n2}^{(2)} = o_p(1)$. Finally, as $v_{i2}^{*2}\boldsymbol{\zeta}_{i2,1}^{\circ d}$ is measurable w.r.t. $v_{i\cdot}$ and thus are independent. An application of WLLN for MD arrays shows that $Q_{n2}^{(3)} = o_p(1)$. Therefore, $Q_{n2} = o_p(1)$. The proof of $Q_{nt} \xrightarrow{p} 0$ for $t \geq 3$ follows similar arguments as those for Q_{n2}, although more tedious.

To show (c), we have $\frac{1}{N}\sum_{j=1}^{N}[(\mathbf{v}_{N,j}\boldsymbol{\zeta}_{N,j}^{\circ})^2 - \mathrm{E}((\mathbf{v}_{N,j}\boldsymbol{\zeta}_{N,j}^{\circ})^2))] = \frac{1}{T-1}\sum_{t=1}^{T-1}\left\{\frac{1}{n}\sum_{i=1}^{n}[(v_{it}^{*}\boldsymbol{\zeta}_{it}^{\circ})^2 - \mathrm{E}((v_{it}^{*}\boldsymbol{\zeta}_{it}^{\circ})^2)]\right\} = \frac{1}{T-1}\sum_{t=1}^{T-1} R_{nt}$. Thus, the result follows if each R_{nt} is $o_p(1)$.

First, we have $R_{n1} = \frac{1}{n}\sum_{i=1}^{n}(v_{i1}^{*2} - \sigma_0^2 h_i)\boldsymbol{\zeta}_{i1}^{\circ 2} + \frac{\sigma_0^2}{n}\sum_{i=1}^{n} h_i(\boldsymbol{\zeta}_{i1}^{\circ 2} - \mathrm{E}(\boldsymbol{\zeta}_{i1}^{\circ 2}))$. Obviously, the first term of R_{n1} is the sum of an MD sequence, which can easily be shown to be $o_p(1)$ by applying the WLLN for MD arrays. For the second term, note that $\boldsymbol{\zeta}_{i1}^{\circ} = \sum_{k=1}^{i-1} b_{11,ik}v_{1k}^{*}$, where $b_{11,ik}$ is the (i,k)th element of $\mathbf{B}_{11}$. Thus, $\boldsymbol{\zeta}_{i1}^{\circ 2} = \sum_{k=1}^{i-1} b_{11,ik}^2 v_{1k}^{*2} + 2\sum_{k=1}^{i-1}\sum_{l=1}^{k-1} b_{11,ik}v_{1k}^{*}b_{11,il}v_{1l}^{*}$. Then,

$$
\begin{aligned}
&\tfrac{\sigma_0^2}{n}\textstyle\sum_{i=1}^{n} h_i(\boldsymbol{\zeta}_{i1}^{\circ 2} - \mathrm{E}(\boldsymbol{\zeta}_{i1}^{\circ 2}))\\
&= \tfrac{\sigma_0^2}{n}\textstyle\sum_{i=1}^{n} h_i[\sum_{k=1}^{i-1} b_{11,ik}^2(v_{1k}^{*2} - \mathrm{E}(v_{1k}^{*2}))] + \tfrac{2\sigma_0^2}{n}\sum_{i=1}^{n} h_i[\sum_{k=1}^{i-1}\sum_{l=1}^{k-1} b_{11,ik}b_{11,il}v_{1k}^{*}v_{1l}^{*}]\\
&= \tfrac{\sigma_0^2}{n}\textstyle\sum_{k=1}^{n-1}(\sum_{i=k+1}^{n} h_i b_{11,ik}^2)(v_{1k}^{*2} - \mathrm{E}(v_{1k}^{*2})) + \tfrac{2\sigma_0^2}{n}\sum_{k=1}^{n-1}\xi_k^{*}v_{1k}^{*},
\end{aligned}
$$

where $\xi_k^{*} = \sum_{l=1}^{k-1}(\sum_{i=k+1}^{n} h_i b_{11,ik}b_{11,il})v_{1l}^{*}$, and the last equality is obtained by switching the orders of summations. Both terms are sums of MD sequences as ξ_k^{*} is $\mathcal{G}_{n,k-1}$-measurable, which are shown to be $o_p(1)$ by applying the WLLN for MD sequences. Therefore, $R_{n1} = o_p(1)$.

Next, we have $R_{n2} = \frac{1}{n}\sum_{i=1}^{n}(v_{i2}^{*2} - \sigma_0^2 h_i)\boldsymbol{\zeta}_{i2}^{\circ 2} + \frac{\sigma_0^2}{n}\sum_{i=1}^{n} h_i(\boldsymbol{\zeta}_{i2}^{\circ 2} - \mathrm{E}(\boldsymbol{\zeta}_{i2}^{\circ 2}))$. Applying the decomposition $\boldsymbol{\zeta}_2^{\circ} = \boldsymbol{\zeta}_{2,1}^{\circ u} + \boldsymbol{\zeta}_{2,1}^{\circ l} + \boldsymbol{\zeta}_{2,1}^{\circ d} + \boldsymbol{\zeta}_{2,2}^{\circ}$ as used in proving $Q_{n2} \xrightarrow{p} 0$, we are able to decompose R_{n2} into a sum of a finite number of terms, of which each is $o_p(1)$, and hence R_{n2} itself is $o_p(1)$. The detail for this and these for $R_{nt}, t \geq 3$, are very tedious and hence are omitted.

It remains to show that $\frac{2}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}[\mathbf{s}_{N,it}\mathbf{s}_{N,is}' - \mathrm{E}(\mathbf{s}_{N,it}\mathbf{s}_{N,is}')] \xrightarrow{p} 0$, and that $\frac{2}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}[\hat{\mathbf{s}}_{N,it}\hat{\mathbf{s}}_{N,is}' - \mathbf{s}_{N,it}\mathbf{s}_{N,is}'] \xrightarrow{p} 0$. The latter is straightforward by applying the mean value theorem, and the former can proved a long the same line of the proof above.

Finally, we offer a discussion on the magnitude of the additional term in $\widehat{\Omega}_{\mathtt{AQS1}}^{\dagger}$. It is asymptotically equivalent to $\frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}\mathrm{E}(\mathbf{s}_{N,it}\mathbf{s}_{N,is}')$. Denote the elements of $\mathrm{E}(\mathbf{s}_{N,it}\mathbf{s}_{N,is}')$ by $\Upsilon_{i,pq}$, where $p, q = 1, 2, 3, 4$, corresponding to β, σ^2, λ, and ρ, respectively. Let f_t be the tth column of $F_{T,T-1}$ and $v_{i\cdot}$ be the $T \times 1$vector of idiosyncratic errors of the spatial unit ith. We have $v_{it}^{*} = f_t' v_{i\cdot}$. It is easy to see that $\Upsilon_{i,11} = 0$. By Lemma A.4 (iii) and the homoskedasticity of v_{it}^{*} across t given i, we have $\Upsilon_{i,22} = \frac{1}{4\sigma_0^8}\mathrm{Cov}(v_{it}^{*2}, v_{is}^{*2}) = \frac{1}{4\sigma_0^4}h_i^2\kappa_i(f_t^2)'f_s^2$, $\Upsilon_{i,33} = \frac{1}{\sigma_0^4}\mathrm{Cov}[\mathbf{v}_{N,it}(\boldsymbol{\eta}_{N,it} + \boldsymbol{\zeta}_{N,is}), \mathbf{v}_{N,is}(\boldsymbol{\eta}_{N,is} + \boldsymbol{\zeta}_{N,is})] = \frac{1}{\sigma_0^4}\mathrm{E}(v_{it}^{*2}v_{is}^{*}b_{ts,ii}\boldsymbol{\eta}_{is}) = \frac{1}{\sigma_0^4}\gamma_i f_t' f_s^2 b_{ts,ii}\boldsymbol{\eta}_{is}$, and $\Upsilon_{i,44} = 0$, where $\{b_{ts,ii}\}$ are diagonal elements of $\mathbf{B}_{ts}$, $\boldsymbol{\eta}_{is}$ are the (i,s)th element of $\boldsymbol{\eta}_N$, and γ_i and κ_i are the measures of skewness and excess kurtosis of v_{it}. Thus, $\frac{1}{N}\sum_{i=1}^{n}\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}\mathrm{Cov}(\mathbf{s}_{N,it}, \mathbf{s}_{N,is}) = o(1)$, if

(a) $\frac{1}{N}\sum_{i=1}^{n} h_i^2\kappa_i\sum_{t=2}^{T-1}\sum_{s=1}^{t-1}(f_t^2)'f_s^2 = o(1)$, and
(b) $\frac{1}{N}\sum_{i=1}^{n}\gamma_i\sum_{t=2}^{T-1}\sum_{s=1}^{t-1} f_t' f_s^2 b_{ts,ii}\boldsymbol{\eta}_{is} = o(1)$,

as for the other terms with $p \neq q$, we have $|\Upsilon_{i,pq}| \leq |\Upsilon_{i,pp}||\Upsilon_{i,qq}|$. ■

Proof of Theorem 3.3: Proof of consistency. Let $\bar{\psi}^*_N(\delta) = \mathrm{E}(\tilde{\psi}^*_N(\delta))$. By Theorem 5.9 of van der Vaart (1998), consistency of $\hat{\delta}^*_{\texttt{AQS1}}$ follows from (a) $\sup_{\delta\in\Delta}\frac{1}{N}\|\tilde{\psi}^*_N(\delta) - \bar{\psi}^*(\delta)\| = o_p(1)$ and (b) for every $\varepsilon > 0$, $\inf_{\delta: d(\delta,\delta_0)\geq\epsilon}\frac{1}{N}\|\bar{\psi}^*(\delta)\| > 0 = \frac{1}{N}\|\bar{\psi}^*(\delta_0)\|$. Write the two components of the AQS function $\tilde{\psi}^*_N(\delta)$ as $R_{rN}(\delta) = T_{rN}(\delta) - S_{rN}(\delta), r = 1,2$, where $T_{rN}(\delta) = \mathbb{Y}'_N(\delta)\mathbb{M}_N(\rho)\bar{\mathbf{G}}_{rN}(\delta)\mathbb{Y}_N(\delta)$ and $S_{rN}(\delta) = \mathbb{Y}'_N(\delta)\mathbb{M}_N(\rho)\mathrm{diag}[\mathbb{M}_N(\rho)]^{-1}\mathrm{diag}[\mathbb{M}_N(\rho)\bar{\mathbf{G}}_{rN}(\delta)]\mathbb{Y}_N(\delta)$.

For condition (a), with $\mathbb{Y}_N(\delta) = \mathbf{D}_N(\delta)\mathbf{D}_N^{-1}(\mathbf{A}_{2N}\mathbf{X}_N\beta_0 + (F'_{T,T-1}\otimes I_n)\mathbb{V}_{nT})$, we see that $T_{rN}(\delta)$ and $S_{rN}(\delta)$ are all linear-quadratic in $\mathbb{V}_{nT}$. Therefore, for each δ, the pointwise convergence to zero of $\frac{1}{N}[T_{rN}(\delta) - \mathrm{E}(T_{rN}(\delta))]$ and $\frac{1}{N}[S_{rN}(\delta) - \mathrm{E}(S_{rN}(\delta))]$ for $r = 1, 2$, can easily be established along the lines of the proof for Theorem 3.1. For stochastic equicontinuity of the two types of quantities, note that $\mathbf{D}_N(\delta)\mathbf{D}_N^{-1} = I_N + (\rho_0-\rho)\mathbf{G}_{2N} + (\lambda_0-\lambda)\bar{\mathbf{G}}_{1N} + (\lambda_0-\lambda)(\rho_0-\rho)\mathbf{G}_{2N}\bar{\mathbf{G}}_{1N}$, and the partial derivatives $\frac{\partial}{\partial\lambda}\bar{\mathbf{G}}_{1N}(\delta)$, $\frac{\partial}{\partial\rho}\bar{\mathbf{G}}_{1N}(\delta)$, $\frac{\partial}{\partial\rho}\bar{\mathbf{G}}_{2N}(\rho)$, and $\frac{\partial}{\partial\rho}\mathbb{M}_N(\rho)$ are all uniformly bounded in row and column sums, uniformly in $\delta\in\Delta$ by Lemma A.2. Therefore, $T_{rN}(\delta)$ and $S_{rN}(\delta)$ are stochastically equicontinuous. The pointwise convergence and stochastic equicontinuity lead to the uniform convergence results: $\sup_{\delta\in\Delta}\frac{1}{N}|T_{rN}(\delta) - \mathrm{E}(T_{rN}(\delta))| = o_p(1)$ and $\sup_{\delta\in\Delta}\frac{1}{N}|S_{rN}(\delta) - \mathrm{E}(S_{rN}(\delta))| = o_p(1)$ for $r = 1, 2$, under Assumptions 1-6 and using the theorem of Newey (1991). Thus, $\frac{1}{N}[R_{rN}(\delta) - \mathrm{E}[R_{rN}(\delta)]] = o_p(1)$.

For condition (b), first, we have $\mathrm{E}[R_{rN}(\delta_0)] = 0$. By Assumption 6 and Lemma A.2, $\mathrm{E}[R_{rN}(\delta)] \neq 0$, for any $\delta\neq\delta_0$. It follows that the conditions of Theorem 5.9 of van der Vaart (1998) hold, and thus the consistency of $\hat{\delta}^*_{\texttt{AQS1}}$ follows.

Proof of asymptotic normality. To establish the asymptotic normality of $\hat{\delta}^*_{\texttt{AQS1}}$, we have, by the mean value theorem,

$$0 = \frac{1}{\sqrt{N}}\tilde{\psi}^*_N(\hat{\delta}^*_{\texttt{AQS1}}) = \frac{1}{\sqrt{N}}\tilde{\psi}^*_N(\delta_0) + \frac{1}{N}\frac{\partial}{\partial\delta'}\tilde{\psi}^*_N(\bar{\delta}_N)\sqrt{N}(\hat{\delta}^*_{\texttt{AQS1}} - \delta_0), \tag{B-6}$$

where $\bar{\delta}_N$ lies between $\hat{\delta}^*_{\texttt{AQS1}}$ and δ_0 elementwise. It suffices to show that

(i) $\frac{1}{\sqrt{N}}\tilde{\psi}^*_N(\delta_0) \xrightarrow{D} N(0, \lim_{N\to\infty}\Omega^*_N)$,

(ii) $\frac{1}{N}\left[\frac{\partial}{\partial\delta'}\tilde{\psi}^*_N(\bar{\delta}_N) - \frac{\partial}{\partial\delta'}\tilde{\psi}^*_N(\delta_0)\right] = o_p(1)$, and

(iii) $\frac{1}{N}\left[\frac{\partial}{\partial\delta'}\tilde{\psi}^*_N(\delta_0) - \mathrm{E}(\frac{\partial}{\partial\delta'}\tilde{\psi}^*_N(\delta_0))\right] = o_p(1)$.

To prove (i), note $\tilde{\psi}^*_N(\delta_0)$ can be written in LQ forms in original errors, the CLT for LQ forms of Kelejian and Prucha (2001) leads to the result.

To prove (ii), let $\mathcal{H}^*_N(\delta) = -\frac{\partial}{\partial\delta'}\tilde{\psi}^*_N(\delta) = [\mathcal{H}^*_{N,11}(\delta), \mathcal{H}^*_{N,12}(\delta);\ \mathcal{H}^*_{N,21}(\delta), \mathcal{H}^*_{N,22}(\delta)]$, where,

$$\mathcal{H}^*_{N,11}(\delta) = \mathbb{Y}'_N(\delta)[\dot{\mathbf{B}}^*_{11N}(\delta) + \bar{\mathbf{G}}'_{1N}(\lambda)\mathbf{B}^*_{1N}(\delta) + \mathbf{B}^*_{1N}(\delta)\bar{\mathbf{G}}_{1N}(\lambda)]\mathbb{Y}_N(\delta),$$
$$\mathcal{H}^*_{N,12}(\delta) = \mathbb{Y}'_N(\delta)[\dot{\mathbf{B}}^*_{12N}(\delta) + \mathbf{G}'_{2N}(\lambda)\mathbf{B}^*_{1N}(\delta) + \mathbf{B}^*_{1N}(\delta)\mathbf{G}_{2N}(\lambda) + \dot{\mathbb{M}}_N(\rho)\bar{\mathbf{G}}^*_{1N}(\delta)]\mathbb{Y}_N(\delta),$$
$$\mathcal{H}^*_{N,21}(\delta) = \mathbb{Y}'_N(\delta)[\bar{\mathbf{G}}'_{1N}(\delta)\mathbf{B}^*_{2N}(\rho) + \mathbf{B}^*_{2N}(\rho)\bar{\mathbf{G}}_{1N}(\delta)]\mathbb{Y}_N(\delta),$$
$$\mathcal{H}^*_{N,22}(\delta) = \mathbb{Y}'_N(\delta)[\dot{\mathbf{B}}^*_{22N}(\delta) + \mathbf{G}'_{2N}(\lambda)\mathbf{B}^*_{2N}(\rho) + \mathbf{B}^*_{2N}(\rho)\mathbf{G}_{2N}(\lambda) + \dot{\mathbb{M}}_N(\rho)\bar{\mathbf{G}}^*_{2N}(\delta)]\mathbb{Y}_N(\delta),$$

where $\mathbf{B}^*_{rN}(\delta) = \mathbb{M}_N(\rho)\bar{\mathbf{G}}^*_{rN}(\delta)$, $\dot{\mathbf{B}}^*_{rsN}(\delta) = \mathbb{M}_N(\rho)\dot{\bar{\mathbf{G}}}^*_{rs,N}(\delta)$, $\dot{\bar{\mathbf{G}}}^*_{r1,N}(\delta)$ is the partial derivative of $\bar{\mathbf{G}}^*_{rN}(\delta)$ $(r = 1, 2)$, w.r.t. λ and ρ $(s = 1, 2)$, $\dot{\mathbb{M}}_N(\rho)$ is the derivative of $\mathbb{M}_N(\rho)$ w.r.t. ρ,

$\dot{\bar{\mathbf{G}}}^*_{11,N}(\delta) = \bar{\mathbf{G}}^2_{1N}(\delta) - \text{diag}[\mathbb{M}_N(\rho)]^{-1}\text{diag}[\mathbf{B}_{1N}(\rho)\bar{\mathbf{G}}_{1N}(\delta)]$,
$\dot{\bar{\mathbf{G}}}^*_{12,N}(\delta) = \bar{\mathbf{G}}_{1N}(\delta)\mathbf{G}_{2N}(\rho) - \mathbf{G}_{2N}(\rho)\bar{\mathbf{G}}_{1N}(\delta) + \text{diag}[\mathbb{M}_N(\rho)]^{-2}\text{diag}[\dot{\mathbb{M}}_N(\rho)]\text{diag}[\mathbf{B}_{2N}(\rho)]$
$\quad +\text{diag}[\mathbb{M}_N(\rho)]^{-1}\text{diag}[\mathbb{M}_N(\rho)\mathbf{G}_{2N}(\rho)\bar{\mathbf{G}}_{1N}(\delta) - \mathbf{B}_{1N}\mathbf{G}_{2N}(\rho) - \dot{\mathbb{M}}_N(\rho)\bar{\mathbf{G}}_{1N}(\delta)]$,
$\dot{\bar{\mathbf{G}}}^*_{22,N}(\delta) = \mathbf{G}_{2N}(\rho)\bar{\mathbf{G}}_{2N}(\rho) + \mathbf{G}_{2N}(\rho)\dot{\mathbb{M}}_N(\rho) + \text{diag}[\mathbb{M}_N(\rho)]^{-2}\text{diag}[\dot{\mathbb{M}}_N(\rho)]\text{diag}[\mathbf{B}_{2N}(\rho)]$
$\quad -\text{diag}[\mathbb{M}_N(\rho)]^{-1}\text{diag}[\mathbb{M}_N(\rho)\mathbf{G}_{2N}(\rho)\bar{\mathbf{G}}_{2N}(\rho) + \mathbb{M}_N(\rho)\mathbf{G}_{2N}(\rho)\dot{\mathbb{M}}_N(\rho) + \dot{\mathbb{M}}_N(\rho)\bar{\mathbf{G}}_{2N}(\rho)]$
$\dot{\mathbb{M}}_N(\rho) = \mathbb{M}_N(\rho)\mathbf{G}_{2N}(\rho)\mathbf{P}_N(\rho) + \mathbf{P}_N(\rho)\mathbf{G}'_{2N}(\rho)\mathbb{M}_N(\rho)$,
$\mathbf{B}_{rN}(\delta) = \mathbb{M}_N(\rho)\bar{\mathbf{G}}_{rN}(\delta)$, for $r = 1, 2$.

By Assumptions 4, 5 and continuous mapping theorem (CMT), $\bar{\mathbf{G}}^\circ_{rN}(\bar{\delta}_N) = \bar{\mathbf{G}}^*_{rN} + o_p(1)$ and $\dot{\bar{\mathbf{G}}}^*_{rN}(\bar{\delta}_N) = \dot{\bar{\mathbf{G}}}^*_{rN} + o_p(1)$ for $r = 1, 2$. Thus using a Taylor expansion, terms of the sort $\mathbb{Q}_{1N}(\bar{\delta}) = \frac{1}{N}\mathbb{Y}'_n(\bar{\delta})Q_{1N}(\bar{\delta})\mathbb{Y}_N(\bar{\delta})$ can be written as, $\mathbb{Q}_{1N} + (\bar{\delta} - \delta_0)'\frac{\partial}{\partial\delta}\mathbb{Q}_{1N}$. Together with the CMT, Lemma A.2, Assumptions 3-5 and some tedious algebra, we have $\mathbb{Q}_{1N}(\bar{\delta}) = \mathbb{Q}_{1N} + o_p(1)$. Collecting these results we have $\frac{\partial}{\partial\delta'}\tilde{\psi}^*_N(\bar{\delta}_N) - \frac{\partial}{\partial\delta'}\tilde{\psi}^*_N = o_p(1)$.

To prove (iii), the negative of the expected Hessian, Φ^*_N, is given as:

$$\Phi^*_N = \frac{1}{N}\begin{pmatrix} \sigma_0^2\text{tr}(\mathbf{H}_N\phi_{11,N}) + \beta_0'\mathbb{X}_N'\phi_{11,N}\mathbb{X}_N\beta, & \sigma_0^2\text{tr}(\mathbf{H}_N\phi_{12,N}) + \beta_0'\mathbb{X}_N'\phi_{12,N}\mathbb{X}_N\beta \\ \sigma_0^2\text{tr}(\mathbf{H}_N\phi_{21,N}) + \beta_0'\mathbb{X}_N'\phi_{21,N}\mathbb{X}_N\beta, & \sigma_0^2\text{tr}(\mathbf{H}_N\phi_{22,N}) + \beta_0'\mathbb{X}_N'\phi_{22,N}\mathbb{X}_N\beta \end{pmatrix},$$

where $\phi_{11,N} = \dot{\mathbf{B}}^*_{11N} + \bar{\mathbf{G}}'_{1N}\mathbf{B}^*_{1N} + \mathbf{B}^*_{1N}\bar{\mathbf{G}}_{1N}$,
$\phi_{12,N} = \dot{\mathbf{B}}^*_{12N} + \mathbf{G}'_{2N}\mathbf{B}^*_{1N} + \mathbf{B}^*_{1N}\mathbf{G}_{2N} + \dot{\mathbb{M}}_N\bar{\mathbf{G}}^*_{1N}$,
$\phi_{21,N} = \bar{\mathbf{G}}'_{1N}\mathbf{B}^*_{2N} + \mathbf{B}^*_{2N}\bar{\mathbf{G}}_{1N}$ and
$\phi_{22,N} = \dot{\mathbf{B}}^*_{22N} + \mathbf{G}'_{2N}\mathbf{B}^*_{2N} + \mathbf{B}^*_{2N}\mathbf{G}_{2N} + \dot{\mathbb{M}}_N\bar{\mathbf{G}}^*_{2N}$.
The result of (iii) follows by showing $\mathcal{H}_{N,rs} - \Phi^*_{N,rs} = o_p(1), r, s = 1, 2$.
With (B-6), and (i)-(iii), the asymptotic normality follows. ∎

Proof of Theorem 3.4: The proof is straightforward following the derivations above Theorem 3.4 in the main text and the proof of Theorem 3.3, and thus is omitted. ∎

Proof of Theorem 3.5: Similarly, the result $\widehat{\Phi}^*_{\texttt{AQS1}} - \Phi^*_N \xrightarrow{p} 0$ follows from the results (ii) and (iii) in the proof of asymptotic normality part of the proof of Theorem 3.3. This result holds irrespective of whether the errors are normal or non-normal, and T is small or large.

To prove the consistency of the OPMD-type estimators of Ω^*_N and Σ_N, we focus on the case of normal errors for this theorem, i.e., the estimators $\widehat{\Omega}^*_{\texttt{AQS1}}$ and $\widehat{\Sigma}_{\texttt{AQS1}}$. The case of non-normal errors can be proved in a similar manner as the proof of Theorem 3.2. It amounts to show that $\frac{1}{N}\sum_{j=1}^N[\hat{\mathbf{v}}^2_{N,j}\hat{\mathbf{s}}^*_{N,j}\hat{\mathbf{s}}^{*\prime}_{N,j} - \text{E}(\mathbf{v}^2_{N,j}\mathbf{s}^*_{N,j}\mathbf{s}^{*\prime}_{N,j})] = o_p(1)$, where $\mathbf{s}^*_{N,j} = (\zeta_{rN,j} + c_{rN,j})'_{r=1,2} \equiv (\mathbf{s}^*_{1N,j}, \mathbf{s}^*_{2N,j})'$ and $\hat{\mathbf{v}}_{N,j}$ and $\hat{\mathbf{s}}^*_{N,j}$ are estimates based on $\hat{\theta}^*_{\texttt{AQS1}}$. The result holds if
$\frac{1}{N}\sum_{j=1}^N[\hat{\mathbf{v}}^2_{N,j}\hat{\mathbf{s}}^*_{N,j}\hat{\mathbf{s}}^{*\prime}_{N,j} - \mathbf{v}^2_{N,j}\mathbf{s}^*_{N,j}\mathbf{s}^{*\prime}_{N,j}] = o_p(1)$ and
$\frac{1}{N}\sum_{j=1}^N[\mathbf{v}^2_{N,j}\mathbf{s}^*_{N,j}\mathbf{s}^{*\prime}_{N,j} - \text{E}(\mathbf{v}^2_{N,j}\mathbf{s}^*_{N,j}\mathbf{s}^{*\prime}_{N,j})] = o_p(1)$.
The former is straightforward by using the mean value theorem, and therefore we focus on the

latter. Denote $\{\Delta_{r,s}\}_{r,s=1,2} = \frac{1}{N}\sum_{j=1}^{N}[\mathbf{v}_{N,j}^2\mathbf{s}_{N,j}^*\mathbf{s}_{N,j}^{*\prime} - \mathrm{E}(\mathbf{v}_{N,j}^2\mathbf{s}_{N,j}^*\mathbf{s}_{N,j}^{*\prime})]$. We have, for $r,s = 1,2$,

$$\begin{aligned}\Delta_{r,s} &= \tfrac{1}{N}\textstyle\sum_{j=1}^{N}[\mathbf{s}_{rN,j}^*\mathbf{s}_{sN,j}^*(\mathbf{v}_{N,j}^2 - \mathrm{E}(\mathbf{v}_{N,j}^2)] \\ &\quad + \tfrac{1}{N}\textstyle\sum_{j=1}^{N}[\mathrm{E}(\mathbf{v}_{N,j}^2)(\zeta_{rN,j}\zeta_{sN,j} - \mathrm{E}(\zeta_{rN,j}\zeta_{sN,j}))] \\ &\quad + \tfrac{1}{N}\textstyle\sum_{j=1}^{N}(c_{rN,j}\zeta_{sN,j} + c_{sN,j}\zeta_{rN,j})\mathrm{E}(\mathbf{v}_{N,j}^2) \equiv \sum_{k=1}^{3}T_{kN}.\end{aligned}$$

As $\mathbf{s}_{1N,j}^*$ and $\mathbf{s}_{2N,j}^*$ are $\mathcal{F}_{N,j-1}$-measurable, where $\mathcal{F}_{N,j}$ is an increasing σ-field generated by $\{\mathbf{v}_{N,1},\ldots,\mathbf{v}_{N,j}\}$, $\mathrm{E}[\mathbf{s}_{rN,j}^*\mathbf{s}_{sN,j}^*(\mathbf{v}_{N,j}^2 - \mathrm{E}(\mathbf{v}_{N,j}^2)|\mathcal{F}_{N,j-1}] = 0$, and thus T_{1N} is the sum of an MD sequence. The conditions of Theorem 19.7 of Davidson (1994) (the WLLN for MD sequences) can be easily verified under Assumptions 1-5, and hence $T_{1N} = o_p(1)$.

For T_{2N}, note that $\zeta_{rN,j} = \sum_{k=1}^{j-1}b_{rN,jk}\mathbf{v}_{N,k}$, where $b_{rN,jk}$ is the (j,k)th element of $\mathbf{B}_{rN}^{u\prime} + \mathbf{B}_{rN}^{l}$. Hence, $\mathrm{E}(\zeta_{rN,j}\zeta_{sN,j}) = \sum_{k=1}^{j-1}b_{rN,jk}b_{sN,jk}\mathrm{E}(\mathbf{v}_{N,k}^2) \equiv d_{rsN,j}$ and,

$$\begin{aligned}T_{2N} &= \tfrac{1}{N}\textstyle\sum_{j=1}^{N}\mathrm{E}(\mathbf{v}_{N,j}^2)(\zeta_{rN,j}\zeta_{sN,j} - d_{rsN,j}) \\ &= \tfrac{1}{N}\textstyle\sum_{j=1}^{N}\mathrm{E}(\mathbf{v}_{N,j}^2)\sum_{k=1}^{j-1}b_{rN,jk}b_{sN,jk}(\mathbf{v}_{N,k}^2 - \mathrm{E}(\mathbf{v}_{N,k}^2)) \\ &\quad + \tfrac{2}{N}\textstyle\sum_{j=1}^{N}\mathrm{E}(\mathbf{v}_{N,j}^2)\sum_{k=1}^{j-1}b_{rN,jk}\mathbf{v}_{N,k}\sum_{l=1}^{k-1}b_{sN,jl}\mathbf{v}_{N,l} \\ &= \tfrac{1}{N}\textstyle\sum_{j=1}^{N-1}\phi_{rsN,j}(\mathbf{v}_{N,j}^2 - \mathrm{E}(\mathbf{v}_{N,j}^2)) + \tfrac{1}{N}\sum_{j=1}^{N-1}\varphi_{rsN,j}\mathbf{v}_{N,j},\end{aligned}$$

where, by switching the order of summations, we have $\phi_{rsN,j} = \frac{1}{N}\sum_{k=j+1}^{N}b_{rN,kj}b_{sN,kj}\mathrm{E}(\mathbf{v}_{N,k}^2)$, $\varphi_{rsN,j} = \sum_{k=1}^{j-1}\xi_{rsN,jk}\mathbf{v}_{N,k}$ and $\xi_{rsN,jk} = 2\sum_{l=j+1}^{N}b_{rN,lj}b_{sN,lk}\mathrm{E}(\mathbf{v}_{N,l}^2)$. Thus T_{2N} is the sum of two MD sequences and the WLLN for MD sequences implies $T_{2N} \xrightarrow{p} 0$. The last term T_{3N} is simpler than T_{3N}. Thus, similar but simpler arguments show $T_{3N} \xrightarrow{p} 0$.

For the case of non-normal errors, refer to the proof of Theorem 3.2 for details.

A similar line of arguments can be used to show $\hat{\Sigma}_{\mathtt{AQS1}} - \Sigma_N = o_p(1)$. ■

Acknowledgments. We thank the two referees for their helpful comments. Zhenlin Yang gratefully acknowledges the financial support from Singapore Management University under Grant C244/MSS16E003.

References

Amemiya, T., 1985. *Advanced Econometrics.* Cambridge, Massachusetts: Harvard University Press.

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

Anselin, L., Le Gallo, J., Jayet, J., 2008. Spatial panel econometrics. In: Mátyás, L., Sevestre, P. (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice.* Springer-Verlag, Berlin Heidelberg, pp. 625-660.

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* 10:1, 11-35.

Baltagi, B., Egger, P., Pfaffermayr, M., 2013. A generalised spatial panel data model with random effects. *Econometric Reviews* 32, 650-685.

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

Baltagi, B., Song, S. H., Kon, W., 2003. Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 117, 123-150.

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

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

Elhorst, J. P., Lacombe, D. J., Piras, G., 2012. On model specification and parameter space definitions in higher order spatial econometrics models. *Regional Science and Urban Economics* 42, 211-220.

Fingleton, B., 2008. A generalised method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. *Spatial Economic Analysis* 3, 27-44.

Griffith, D. A., 1988. *Advanced Spatial Statistics.* Kluwer, Dordrecht, The Netherlands.

Hsieh, C. S., Lee, L. F., 2014. A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics* 31, 301-319.

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., Piras, G., 2016. An extension of the J-test to spatial panel data framework. *Journal of Applied Econometrics* 31, 387-402.

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., 2007. HAC estimation in a spatial framework. *Journal of Econometrics* 140, 131-154.

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.

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

Lee, L. F., Yu, J., 2010a. Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects. *Journal of Econometrics* 154, 165-185.

Lee, L. F., Yu, J., 2010b. Some recent developments in spatial panel data models. *Regional Science and Urban Economics* 40, 255-271.

Lee, L. F., Yu, J., 2012. Spatial panels: random components vs. fixed effects. *International Economic Review* 53, 1369-1412.

Lee, L. F., Yu, J., 2015. Spatial panel data models. In: Baltagi, B. H. (Eds.), The Oxford Handbook of Panel Data. *Oxford University Press*, Oxford. 363-401.

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

LeSage, J., Pace, R. K., 2009. *Introduction to spatial econometrics.* New York: CRC Press.

Li, L. Y., Yang, Z. L., 2019. Spatial Dynamic Panel Data Models with Correlated Random Effects. *Journal of Econometrics.* Conditionally accepted.

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*, forthcoming.

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

Liu, S. F., Yang, Z. L., 2015. Modified QML Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity and Non-normality. *Regional Science and Urban Economics* 52, 50-70.

Millimet, D. L., Roy, J., 2016. Empirical tests of the pollution haven hypothesis when environmental regulations is endogenous. *Journal of Applied Econometrics* 31, 652-677.

Moscone, F., Tosetti, E., 2011. GMM estimation of spatial panels with fixed effects and unknown heteroskedasticity. *Regional Science and Urban Economics* 41 487-497.

Mundlak, Y., 1978. On the pooling of time series and cross section data. *Econometrica* 46, 69-85.

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

Newey, W. K., 1991. Uniform Convergence in Probability and Stochastic Equicontinuity. *Econometrica* 59, 1161-1167.

Robinson, P. M., Rossi, F., 2015. Refinements in maximum likelihood inference on spatial autocorrelation in panel data. *Journal of Econometrics* 189, 447-456.

Serfling, R. J., 1980. *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, Inc.

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

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817-838.

Xu, Y. H, Yang, Z. L., 2020. Specification tests for temporal heterogeneity in spatial panel data models with fixed effects. *Regional Science and Urban Economics*, forthcoming.

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

Yang, Z. L., 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.

Yu, J., de Jong, R., Lee, L. F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146, 118-134.

Table 1a. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, FE_1-SPD Model

Case when the regular QMLE is consistent under heteroskedasticity

$T = 3, \beta = (1, 1)', \sigma = 1$, `Queen Contiguity, REG-1, DGP 1`

n	λ	ρ	QMLE-λ	AQSE*-λ	QMLE-ρ	AQSE*-ρ
50	.50	.50	.474[.202](.200)	.490[.209](.207){.190}	.452[.239](.234)	.449[.244](.238){.234}
		.25	.462[.190](.186)	.470[.195](.191){.180}	.225[.266](.265)	.221[.268](.267){.266}
		.00	.468[.166](.163)	.470[.168](.165){.158}	-.017[.275](.274)	-.021[.273](.272){.279}
		-.25	.469[.150](.147)	.472[.151](.148){.149}	-.257[.271](.271)	-.258[.267](.267){.271}
		-.50	.472[.138](.135)	.476[.138](.136){.129}	-.501[.271](.271)	-.500[.267](.267){.270}
	-.50	.50	-.469[.225](.223)	-.480[.226](.225){.215}	.450[.211](.205)	.450[.211](.205){.192}
		.25	-.475[.222](.221)	-.480[.224](.223){.220}	.196[.252](.246)	.194[.252](.245){.239}
		.00	-.484[.222](.221)	-.485[.223](.223){.219}	-.049[.277](.273)	-.001[.275](.271){.268}
		-.25	-.487[.218](.218)	-.486[.220](.220){.217}	-.288[.286](.284)	-.274[.284](.281){.281}
		-.50	-.489[.219](.219)	-.490[.221](.221){.221}	-.532[.288](.287)	-.521[.285](.284){.280}
100	.50	.50	.472[.169](.167)	.470[.169](.166){.151}	.485[.179](.178)	.490[.177](.177){.172}
		.25	.474[.144](.142)	.474[.143](.140){.150}	.244[.194](.194)	.250[.191](.191){.200}
		.00	.481[.119](.118)	.481[.118](.117){.118}	-.005[.196](.196)	-.003[.192](.192){.195}
		-.25	.486[.099](.097)	.490[.098](.097){.093}	-.253[.193](.193)	-.249[.190](.190){.192}
		-.50	.487[.087](.086)	.490[.087](.086){.083}	-.504[.186](.185)	-.498[.183](.183){.185}
	-.50	.50	-.486[.181](.181)	-.485[.180](.179){.174}	.474[.151](.149)	.469[.151](.148){.148}
		.25	-.495[.174](.174)	-.500[.172](.172){.169}	.228[.181](.180)	.230[.179](.177){.177}
		.00	-.494[.173](.173)	-.493[.171](.171){.170}	-.022[.202](.201)	-.023[.199](.197){.196}
		-.25	-.501[.169](.169)	-.500[.167](.167){.162}	-.263[.212](.212)	-.261[.208](.208){.208}
		-.50	-.501[.169](.169)	-.500[.167](.167){.160}	-.510[.216](.216)	-.504[.211](.211){.214}
250	.50	.50	.486[.118](.118)	.490[.121](.120){.119}	.489[.128](.127)	.490[.130](.130){.128}
		.25	.486[.098](.097)	.488[.099](.098){.096}	.248[.134](.134)	.250[.135](.135){.133}
		.00	.487[.081](.080)	.490[.081](.080){.078}	.001[.135](.135)	.000[.134](.134){.132}
		-.25	.490[.068](.068)	.500[.068](.068){.066}	-.247[.128](.128)	-.250[.128](.128){.127}
		-.50	.493[.059](.059)	.500[.059](.059){.058}	-.500[.122](.122)	-.500[.121](.121){.121}
	-.50	.50	-.486[.127](.127)	-.491[.128](.127){.127}	.481[.100](.098)	.484[.099](.098){.096}
		.25	-.490[.126](.126)	-.493[.126](.126){.126}	.233[.122](.121)	.240[.122](.121){.121}
		.00	-.493[.125](.125)	-.500[.126](.125){.124}	-.014[.141](.140)	-.013[.141](.140){.140}
		-.25	-.497[.123](.123)	-.497[.123](.123){.121}	-.260[.149](.148)	-.258[.148](.148){.146}
		-.50	-.500[.118](.118)	-.500[.118](.118){.118}	-.505[.148](.148)	-.502[.147](.147){.146}
500	.50	.50	.492[.082](.082)	.500[.083](.083){.083}	.497[.089](.089)	.497[.089](.089){.088}
		.25	.494[.066](.066)	.495[.066](.066){.064}	.250[.095](.095)	.250[.095](.095){.092}
		.00	.496[.052](.052)	.500[.052](.052){.052}	-.001[.093](.093)	.000[.093](.093){.092}
		-.25	.497[.045](.045)	.500[.045](.045){.045}	-.251[.088](.088)	-.250[.088](.088){.088}
		-.50	.497[.041](.041)	.500[.041](.041){.040}	-.501[.086](.086)	-.500[.086](.086){.085}
	-.50	.50	-.497[.086](.086)	-.500[.086](.086){.086}	.494[.065](.065)	.500[.065](.065){.065}
		.25	-.498[.087](.087)	-.500[.087](.087){.086}	.244[.085](.085)	.243[.085](.085){.083}
		.00	-.499[.085](.085)	-.499[.084](.084){.084}	-.004[.096](.096)	-.001[.096](.096){.094}
		-.25	-.502[.082](.082)	-.500[.082](.082){.082}	-.252[.102](.102)	-.252[.101](.101){.101}
		-.50	-.502[.081](.081)	-.501[.080](.080){.080}	-.502[.101](.101)	-.500[.100](.100){.101}

Table 1b. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, FE_1-SPD Model
Case when the regular QMLE is consistent under heteroskedasticity
$T = 3, \beta = (1, 1)', \sigma = 1$, `Queen Contiguity, REG-1, DGP 2`

n	λ	ρ	QMLE-λ	AQSE*-λ	QMLE-ρ	AQSE*-ρ
50	.50	.50	.475[.201](.200)	.472[.208](.206){.220}	.451[.239](.234)	.450[.243](.237){.237}
		.25	.467[.183](.180)	.467[.187](.184){.173}	.227[.255](.254)	.230[.256](.255){.258}
		.00	.469[.165](.162)	.470[.167](.164){.160}	-.016[.268](.267)	-.012[.266](.265){.265}
		-.25	.469[.152](.148)	.480[.152](.149){.140}	-.255[.268](.268)	-.255[.264](.264){.260}
		-.50	.471[.143](.140)	.480[.143](.140){.145}	-.503[.269](.269)	-.500[.264](.264){.259}
	-.50	.50	-.469[.225](.223)	-.480[.226](.224){.217}	.448[.211](.205)	.447[.211](.204){.196}
		.25	-.481[.223](.222)	-.484[.224](.223){.210}	.201[.251](.246)	.200[.249](.244){.246}
		.00	-.487[.217](.216)	-.487[.218](.217){.210}	-.041[.274](.271)	-.042[.271](.268){.265}
		-.25	-.494[.216](.216)	-.492[.218](.217){.200}	-.279[.282](.281)	-.272[.279](.277){.277}
		-.50	-.499[.216](.216)	-.495[.216](.216){.210}	-.516[.283](.283)	-.512[.278](.278){.274}
100	.50	.50	.473[.167](.165)	.473[.165](.163){.148}	.483[.177](.176)	.482[.174](.173){.169}
		.25	.473[.144](.141)	.480[.140](.138){.133}	.246[.193](.193)	.250[.189](.189){.189}
		.00	.479[.123](.121)	.480[.121](.119){.110}	-.001[.199](.199)	.000[.194](.194){.191}
		-.25	.487[.101](.100)	.487[.100](.099){.092}	-.252[.192](.192)	-.248[.188](.188){.187}
		-.50	.487[.091](.090)	.487[.091](.090){.090}	-.501[.185](.185)	-.495[.182](.182){.182}
	-.50	.50	-.488[.181](.181)	-.486[.179](.179){.169}	.476[.151](.149)	.480[.150](.147){.143}
		.25	-.494[.177](.177)	-.500[.174](.174){.165}	.226[.183](.181)	.223[.180](.178){.173}
		.00	-.499[.174](.174)	-.497[.171](.171){.160}	-.015[.201](.201)	-.012[.197](.196){.192}
		-.25	-.498[.173](.173)	-.497[.171](.170){.159}	-.264[.213](.213)	-.262[.208](.208){.199}
		-.50	-.503[.169](.169)	-.500[.167](.167){.157}	-.506[.214](.214)	-.501[.209](.209){.200}
250	.50	.50	.485[.119](.118)	.484[.122](.121){.119}	.493[.128](.128)	.500[.130](.130){.127}
		.25	.485[.099](.098)	.486[.100](.099){.095}	.251[.132](.132)	.250[.133](.133){.132}
		.00	.489[.080](.079)	.499[.080](.079){.076}	.001[.132](.132)	.000[.132](.132){.130}
		-.25	.491[.066](.065)	.493[.066](.065){.065}	-.248[.126](.126)	-.250[.125](.125){.125}
		-.50	.492[.060](.059)	.500[.060](.059){.058}	-.498[.124](.124)	-.499[.124](.124){.120}
	-.50	.50	-.490[.127](.126)	-.500[.127](.127){.125}	.485[.097](.096)	.490[.097](.096){.094}
		.25	-.491[.130](.130)	-.500[.130](.130){.126}	.233[.125](.124)	.240[.125](.124){.120}
		.00	-.498[.126](.126)	-.499[.126](.126){.123}	-.011[.140](.139)	-.010[.139](.139){.136}
		-.25	-.498[.123](.123)	-.498[.123](.123){.120}	-.261[.149](.149)	-.254[.149](.148){.143}
		-.50	-.502[.118](.118)	-.500[.118](.118){.117}	-.507[.147](.147)	-.504[.146](.146){.144}
500	.50	.50	.493[.082](.082)	.500[.083](.082){.080}	.496[.089](.089)	.496[.089](.089){.088}
		.25	.494[.066](.065)	.495[.066](.065){.064}	.251[.093](.093)	.250[.093](.093){.092}
		.00	.497[.053](.053)	.500[.053](.053){.052}	-.003[.093](.093)	-.002[.092](.092){.091}
		-.25	.496[.046](.046)	.500[.046](.046){.045}	-.251[.090](.090)	-.250[.090](.090){.089}
		-.50	.498[.040](.040)	.500[.040](.040){.040}	-.503[.085](.085)	-.500[.084](.084){.084}
	-.50	.50	-.497[.087](.087)	-.497[.087](.086){.086}	.494[.065](.065)	.493[.066](.065){.065}
		.25	-.500[.087](.087)	-.499[.087](.087){.086}	.246[.084](.084)	.250[.083](.083){.083}
		.00	-.500[.084](.084)	-.499[.084](.084){.084}	-.004[.094](.094)	-.005[.093](.093){.093}
		-.25	-.499[.085](.084)	-.498[.084](.084){.082}	-.255[.103](.103)	-.252[.102](.102){.100}
		-.50	-.502[.082](.082)	-.501[.081](.081){.080}	-.502[.104](.104)	-.500[.103](.103){.101}

Table 1c. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, FE$_1$-SPD Model

Case when the regular QMLE is consistent under heteroskedasticity

$T = 3, \beta = (1, 1)', \sigma = 1$, `Queen Contiguity, REG-1, DGP 3`

n	λ	ρ	QMLE-λ	AQSE*-λ	QMLE-ρ	AQSE*-ρ
50	.50	.50	.475[.194](.193)	.480[.200](.198){.195}	.456[.228](.223)	.453[.231](.226){.221}
		.25	.466[.182](.179)	.470[.187](.184){.188}	.228[.250](.249)	.230[.251](.249){.252}
		.00	.466[.172](.168)	.468[.173](.169){.153}	-.009[.265](.264)	-.011[.261](.261){.263}
		-.25	.471[.149](.146)	.473[.149](.146){.140}	-.256[.265](.265)	-.255[.260](.260){.263}
		-.50	.475[.140](.138)	.477[.140](.138){.130}	-.500[.258](.258)	-.495[.253](.253){.252}
	-.50	.50	-.467[.222](.220)	-.480[.221](.219){.200}	.448[.209](.203)	.450[.208](.201){.190}
		.25	-.477[.222](.221)	-.480[.223](.221){.199}	.201[.242](.237)	.199[.241](.236){.234}
		.00	-.487[.214](.214)	-.490[.214](.213){.199}	-.036[.268](.265)	-.038[.264](.261){.259}
		-.25	-.491[.209](.209)	-.490[.209](.208){.198}	-.285[.273](.270)	-.250[.268](.266){.269}
		-.50	-.498[.214](.214)	-.500[.213](.213){.197}	-.519[.280](.280)	-.515[.274](.274){.270}
100	.50	.50	.478[.162](.160)	.480[.158](.156){.144}	.484[.170](.170)	.482[.168](.167){.164}
		.25	.475[.145](.143)	.480[.140](.138){.137}	.244[.189](.189)	.250[.184](.184){.184}
		.00	.480[.124](.123)	.480[.122](.120){.107}	.001[.189](.189)	.002[.184](.184){.185}
		-.25	.486[.104](.103)	.490[.103](.101){.090}	-.254[.187](.187)	-.249[.182](.182){.179}
		-.50	.487[.090](.089)	.486[.091](.089){.084}	-.499[.180](.180)	-.491[.176](.176){.177}
	-.50	.50	-.491[.180](.179)	-.490[.174](.173){.160}	.476[.150](.148)	.480[.146](.143){.145}
		.25	-.493[.176](.176)	-.490[.171](.171){.155}	.226[.180](.178)	.230[.175](.173){.174}
		.00	-.496[.173](.173)	-.500[.167](.167){.155}	-.019[.198](.197)	-.021[.191](.190){.187}
		-.25	-.500[.171](.171)	-.498[.164](.164){.150}	-.260[.214](.213)	-.259[.203](.203){.194}
		-.50	-.501[.170](.170)	-.500[.164](.164){.150}	-.509[.215](.215)	-.500[.205](.205){.199}
250	.50	.50	.489[.118](.118)	.490[.119](.119){.120}	.489[.127](.126)	.490[.128](.127){.129}
		.25	.485[.102](.100)	.486[.102](.101){.100}	.248[.137](.137)	.250[.137](.137){.137}
		.00	.487[.082](.081)	.489[.082](.082){.080}	.003[.133](.133)	.001[.133](.133){.130}
		-.25	.493[.064](.064)	.495[.064](.063){.063}	-.250[.125](.125)	-.250[.123](.123){.120}
		-.50	.493[.058](.058)	.496[.058](.057){.056}	-.500[.120](.120)	-.500[.118](.118){.114}
	-.50	.50	-.488[.130](.130)	-.491[.128](.127){.128}	.483[.101](.099)	.485[.099](.098){.098}
		.25	-.491[.131](.131)	-.500[.129](.129){.124}	.233[.127](.126)	.235[.125](.124){.120}
		.00	-.501[.128](.128)	-.500[.126](.126){.120}	-.010[.142](.141)	-.010[.140](.140){.140}
		-.25	-.495[.123](.123)	-.500[.122](.122){.117}	-.262[.147](.147)	-.261[.146](.145){.140}
		-.50	-.502[.123](.123)	-.501[.121](.121){.120}	-.504[.153](.153)	-.501[.150](.150){.149}
500	.50	.50	.496[.082](.082)	.500[.081](.081){.078}	.494[.089](.089)	.494[.088](.088){.086}
		.25	.493[.065](.064)	.494[.064](.064){.063}	.251[.092](.092)	.251[.091](.091){.090}
		.00	.496[.053](.053)	.500[.053](.053){.051}	-.003[.092](.092)	-.002[.092](.092){.089}
		-.25	.497[.045](.045)	.497[.044](.044){.044}	-.251[.088](.088)	-.250[.087](.087){.086}
		-.50	.498[.041](.040)	.498[.040](.040){.040}	-.501[.086](.086)	-.499[.085](.085){.082}
	-.50	.50	-.498[.088](.088)	-.500[.087](.087){.084}	.495[.067](.067)	.494[.066](.065){.063}
		.25	-.498[.087](.087)	-.500[.086](.086){.084}	.243[.085](.084)	.242[.084](.083){.080}
		.00	-.500[.087](.087)	-.499[.085](.085){.082}	-.004[.096](.096)	-.006[.095](.095){.092}
		-.25	-.503[.084](.084)	-.500[.082](.082){.080}	-.250[.102](.102)	-.250[.100](.100){.098}
		-.50	-.499[.084](.084)	-.500[.081](.081){.080}	-.503[.104](.104)	-.500[.101](.101){.100}

Table 2a. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, FE$_1$-SPD Model

Case when the regular QMLE is inconsistent under heteroskedasticity

$T = 3, \beta = (1,1)', \sigma = 1$, `Circular Neighbors, REG-1, DGP 1`

n	λ	ρ	QMLE-λ	AQSE*-λ	QMLE-ρ	AQSE*-ρ
50	.50	.50	.486[.124](.123)	.485[.165](.164){.208}	.422[.181](.164)	.444[.220](.213){.218}
		.25	.451[.123](.112)	.476[.144](.142){.144}	.229[.172](.171)	.213[.236](.233){.239}
		.00	.435[.123](.104)	.480[.126](.124){.127}	.043[.179](.174)	-.026[.241](.240){.229}
		-.25	.418[.129](.100)	.480[.116](.114){.115}	-.142[.198](.166)	-.267[.233](.232){.232}
		-.50	.405[.137](.099)	.479[.112](.110){.115}	-.321[.241](.161)	-.493[.219](.219){.226}
	-.50	.50	-.390[.152](.104)	-.481[.117](.115){.117}	.368[.199](.149)	.457[.163](.157){.157}
		.25	-.401[.143](.103)	-.480[.126](.124){.121}	.127[.208](.167)	.202[.207](.201){.201}
		.00	-.421[.128](.100)	-.480[.137](.136){.134}	-.078[.182](.165)	-.047[.233](.228){.207}
		-.25	-.443[.117](.102)	-.478[.151](.149){.171}	-.258[.152](.152)	-.288[.237](.234){.379}
		-.50	-.478[.106](.104)	-.485[.161](.160){.155}	-.426[.156](.137)	-.523[.226](.225){.282}
100	.50	.50	.485[.096](.095)	.490[.133](.132){.136}	.447[.129](.117)	.481[.154](.153){.153}
		.25	.459[.093](.083)	.483[.106](.105){.109}	.245[.123](.123)	.237[.163](.163){.162}
		.00	.443[.095](.076)	.486[.088](.087){.086}	.053[.134](.123)	-.005[.165](.165){.165}
		-.25	.435[.095](.069)	.490[.075](.074){.073}	-.142[.161](.120)	-.258[.161](.161){.161}
		-.50	.428[.097](.065)	.491[.068](.067){.072}	-.332[.202](.112)	-.495[.148](.148){.101}
	-.50	.50	-.359[.166](.088)	-.487[.101](.100){.100}	.364[.174](.108)	.477[.114](.112){.112}
		.25	-.381[.144](.082)	-.487[.105](.105){.105}	.121[.175](.118)	.220[.149](.146){.146}
		.00	-.409[.120](.079)	-.489[.110](.110){.107}	-.081[.144](.118)	-.029[.171](.168){.168}
		-.25	-.441[.095](.075)	-.493[.113](.113){.114}	-.257[.108](.108)	-.269[.175](.174){.174}
		-.50	-.479[.077](.074)	-.498[.119](.119){.120}	-.421[.125](.097)	-.504[.168](.168){.162}
250	.50	.50	.490[.059](.058)	.491[.086](.086){.083}	.458[.082](.071)	.494[.099](.099){.100}
		.25	.461[.065](.052)	.495[.067](.066){.066}	.255[.078](.077)	.242[.108](.108){.108}
		.00	.441[.076](.048)	.495[.055](.055){.055}	.066[.102](.077)	-.003[.107](.107){.107}
		-.25	.427[.086](.045)	.495[.050](.049){.050}	-.124[.148](.076)	-.251[.105](.105){.105}
		-.50	.418[.093](.043)	.496[.046](.046){.047}	-.318[.195](.070)	-.497[.093](.093){.093}
	-.50	.50	-.370[.141](.053)	-.495[.057](.057){.060}	.374[.143](.068)	.491[.067](.066){.066}
		.25	-.384[.127](.051)	-.497[.061](.061){.061}	.129[.143](.075)	.239[.088](.088){.088}
		.00	-.407[.105](.048)	-.497[.066](.065){.065}	-.078[.107](.073)	-.009[.103](.103){.103}
		-.25	-.436[.080](.047)	-.495[.073](.073){.073}	-.259[.067](.066)	-.258[.111](.110){.111}
		-.50	-.476[.053](.048)	-.497[.084](.084){.084}	-.422[.099](.060)	-.502[.113](.113){.113}
500	.50	.50	.492[.039](.038)	.497[.054](.054){.054}	.460[.063](.048)	.497[.066](.066){.066}
		.25	.464[.050](.034)	.498[.043](.043){.043}	.257[.053](.053)	.246[.072](.072){.072}
		.00	.445[.064](.033)	.498[.038](.038){.038}	.064[.084](.054)	-.003[.076](.076){.076}
		-.25	.430[.076](.031)	.498[.034](.034){.034}	-.125[.136](.053)	-.252[.074](.074){.074}
		-.50	.419[.086](.029)	.497[.032](.032){.032}	-.319[.187](.049)	-.499[.066](.066){.070}
	-.50	.50	-.377[.129](.036)	-.497[.039](.039){.040}	.380[.129](.048)	.494[.047](.046){.046}
		.25	-.389[.116](.035)	-.498[.041](.041){.041}	.136[.126](.052)	.245[.060](.060){.060}
		.00	-.409[.096](.033)	-.497[.043](.043){.043}	-.074[.090](.051)	-.005[.070](.070){.070}
		-.25	-.438[.070](.033)	-.496[.049](.049){.049}	-.258[.049](.048)	-.257[.079](.078){.078}
		-.50	-.477[.040](.033)	-.498[.057](.057){.060}	-.422[.088](.042)	-.502[.078](.078){.080}

Table 2b. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, FE_1-SPD Model

Case when the regular QMLE is inconsistent under heteroskedasticity

$T = 3, \beta = (1,1)', \sigma = 1$, `Circular Neighbors, REG-1, DGP 2`

n	λ	ρ	QMLE-λ	AQSE*-λ	QMLE-ρ	AQSE*-ρ
50	.50	.50	.483[.125](.124)	.482[.163](.162){.163}	.432[.177](.163)	.454[.215](.210){.211}
		.25	.456[.119](.111)	.479[.142](.140){.174}	.230[.171](.169)	.216[.231](.229){.228}
		.00	.438[.122](.105)	.482[.126](.124){.123}	.038[.177](.173)	-.029[.241](.240){.240}
		-.25	.420[.132](.105)	.479[.119](.117){.122}	-.143[.202](.172)	-.264[.237](.237){.232}
		-.50	.406[.139](.102)	.479[.111](.109){.107}	-.328[.238](.165)	-.500[.218](.218){.276}
	-.50	.50	-.396[.151](.110)	-.484[.117](.116){.116}	.376[.196](.152)	.461[.163](.158){.158}
		.25	-.406[.144](.109)	-.481[.127](.126){.121}	.135[.203](.167)	.207[.202](.198){.204}
		.00	-.420[.130](.103)	-.476[.135](.133){.131}	-.076[.180](.164)	-.048[.230](.224){.227}
		-.25	-.445[.118](.104)	-.480[.151](.150){.183}	-.257[.151](.151)	-.290[.236](.233){.262}
		-.50	-.475[.110](.107)	-.483[.164](.164){.103}	-.425[.161](.143)	-.523[.230](.229){.297}
100	.50	.50	.486[.095](.094)	.484[.130](.129){.122}	.445[.128](.116)	.477[.151](.149){.115}
		.25	.461[.092](.083)	.485[.105](.104){.102}	.240[.125](.125)	.232[.165](.164){.201}
		.00	.446[.094](.077)	.488[.088](.087){.087}	.048[.133](.125)	-.011[.167](.167){.167}
		-.25	.434[.098](.072)	.487[.076](.075){.075}	-.139[.165](.122)	-.249[.160](.160){.159}
		-.50	.430[.097](.067)	.492[.067](.067){.067}	-.338[.200](.117)	-.502[.144](.144){.144}
	-.50	.50	-.363[.167](.096)	-.488[.103](.102){.102}	.365[.177](.114)	.476[.117](.115){.115}
		.25	-.384[.144](.086)	-.487[.105](.104){.104}	.126[.173](.120)	.220[.147](.144){.144}
		.00	-.411[.120](.081)	-.490[.108](.108){.108}	-.075[.139](.117)	-.024[.166](.164){.160}
		-.25	-.441[.098](.078)	-.491[.117](.117){.117}	-.257[.109](.108)	-.271[.177](.176){.176}
		-.50	-.479[.078](.075)	-.497[.120](.120){.124}	-.420[.126](.098)	-.504[.166](.166){.160}
250	.50	.50	.490[.059](.058)	.491[.086](.085){.103}	.456[.084](.072)	.491[.100](.100){.104}
		.25	.460[.067](.054)	.493[.068](.068){.068}	.256[.078](.078)	.244[.108](.108){.108}
		.00	.441[.077](.049)	.495[.056](.056){.056}	.064[.102](.079)	-.005[.109](.109){.109}
		-.25	.427[.087](.048)	.495[.050](.050){.050}	-.124[.148](.078)	-.252[.104](.104){.104}
		-.50	.419[.093](.046)	.496[.046](.046){.046}	-.320[.195](.075)	-.499[.093](.093){.093}
	-.50	.50	-.371[.142](.059)	-.496[.058](.058){.060}	.375[.144](.070)	.490[.067](.066){.066}
		.25	-.383[.129](.055)	-.494[.063](.063){.063}	.129[.144](.078)	.238[.089](.088){.088}
		.00	-.405[.107](.050)	-.494[.066](.065){.065}	-.080[.108](.073)	-.013[.103](.102){.102}
		-.25	-.435[.081](.048)	-.493[.073](.073){.074}	-.258[.068](.067)	-.259[.112](.112){.112}
		-.50	-.477[.053](.048)	-.499[.083](.083){.083}	-.422[.099](.060)	-.501[.109](.109){.107}
500	.50	.50	.491[.040](.039)	.496[.055](.055){.055}	.460[.063](.050)	.497[.067](.067){.067}
		.25	.464[.051](.036)	.498[.044](.044){.044}	.256[.053](.053)	.246[.073](.073){.073}
		.00	.445[.064](.033)	.499[.038](.037){.037}	.064[.084](.055)	-.004[.075](.075){.075}
		-.25	.431[.077](.033)	.498[.035](.035){.035}	-.124[.137](.055)	-.250[.074](.074){.074}
		-.50	.420[.086](.031)	.498[.032](.032){.032}	-.320[.188](.054)	-.500[.066](.066){.070}
	-.50	.50	-.378[.129](.039)	-.498[.039](.039){.040}	.382[.128](.050)	.496[.047](.046){.046}
		.25	-.390[.116](.037)	-.498[.041](.041){.041}	.136[.126](.054)	.245[.061](.061){.061}
		.00	-.411[.095](.035)	-.499[.044](.044){.044}	-.072[.088](.051)	-.003[.070](.070){.070}
		-.25	-.438[.070](.034)	-.497[.050](.050){.050}	-.256[.048](.048)	-.254[.078](.078){.078}
		-.50	-.477[.040](.033)	-.498[.057](.057){.060}	-.423[.088](.043)	-.502[.077](.077){.080}

Table 2c. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, FE_1-SPD Model

Case when the regular QMLE is inconsistent under heteroskedasticity

$T = 3, \beta = (1,1)', \sigma = 1$, `Circular Neighbors, REG-1, DGP 3`

n	λ	ρ	QMLE-λ	AQSE*-λ	QMLE-ρ	AQSE*-ρ
50	.50	.50	.480[.129](.128)	.480[.160](.158){.154}	.435[.175](.163)	.459[.204](.200){.208}
		.25	.461[.126](.120)	.483[.139](.138){.139}	.224[.177](.175)	.212[.227](.224){.224}
		.00	.439[.131](.116)	.476[.132](.130){.130}	.035[.181](.178)	-.020[.238](.237){.237}
		-.25	.429[.132](.111)	.478[.124](.122){.125}	-.155[.203](.180)	-.261[.237](.237){.231}
		-.50	.416[.138](.109)	.478[.124](.122){.183}	-.334[.245](.180)	-.496[.234](.234){.266}
	-.50	.50	-.401[.159](.124)	-.480[.128](.126){.124}	.382[.197](.158)	.460[.168](.163){.169}
		.25	-.413[.145](.116)	-.481[.126](.125){.125}	.140[.201](.168)	.202[.201](.195){.192}
		.00	-.428[.135](.114)	-.479[.137](.135){.135}	-.064[.181](.169)	-.042[.223](.219){.219}
		-.25	-.448[.121](.110)	-.481[.146](.145){.147}	-.252[.162](.162)	-.287[.235](.232){.221}
		-.50	-.478[.112](.110)	-.489[.159](.158){.195}	-.419[.171](.151)	-.520[.230](.229){.210}
100	.50	.50	.486[.097](.096)	.486[.124](.123){.121}	.446[.129](.117)	.476[.146](.144){.140}
		.25	.461[.096](.087)	.482[.107](.105){.106}	.243[.125](.125)	.239[.162](.162){.162}
		.00	.451[.093](.079)	.488[.089](.089){.106}	.041[.132](.126)	-.012[.165](.164){.201}
		-.25	.436[.103](.081)	.485[.082](.081){.085}	-.145[.168](.131)	-.249[.161](.161){.160}
		-.50	.433[.102](.077)	.489[.082](.081){.080}	-.341[.208](.134)	-.497[.157](.157){.120}
	-.50	.50	-.369[.172](.111)	-.483[.110](.109){.110}	.369[.178](.121)	.472[.120](.117){.114}
		.25	-.393[.147](.100)	-.487[.106](.105){.105}	.136[.170](.126)	.223[.145](.143){.146}
		.00	-.417[.124](.092)	-.490[.111](.110){.109}	-.069[.138](.120)	-.025[.163](.161){.165}
		-.25	-.446[.102](.087)	-.494[.112](.111){.110}	-.249[.117](.117)	-.265[.170](.169){.169}
		-.50	-.476[.088](.085)	-.493[.123](.123){.121}	-.422[.133](.108)	-.512[.169](.169){.170}
250	.50	.50	.488[.063](.062)	.490[.086](.086){.083}	.457[.086](.074)	.492[.099](.098){.100}
		.25	.462[.067](.055)	.494[.067](.067){.067}	.255[.077](.076)	.245[.103](.103){.103}
		.00	.444[.078](.054)	.495[.056](.056){.056}	.060[.102](.082)	-.005[.106](.106){.106}
		-.25	.431[.088](.055)	.496[.053](.053){.050}	-.132[.148](.089)	-.254[.107](.107){.107}
		-.50	.420[.097](.055)	.495[.050](.050){.049}	-.324[.201](.096)	-.499[.097](.097){.097}
	-.50	.50	-.374[.147](.076)	-.494[.063](.063){.060}	.380[.145](.082)	.491[.070](.069){.069}
		.25	-.391[.128](.067)	-.495[.060](.060){.060}	.138[.140](.083)	.240[.086](.086){.086}
		.00	-.410[.109](.062)	-.495[.066](.065){.065}	-.073[.107](.079)	-.011[.101](.100){.099}
		-.25	-.440[.084](.059)	-.496[.075](.075){.075}	-.256[.072](.072)	-.260[.110](.110){.110}
		-.50	-.476[.059](.053)	-.497[.082](.082){.085}	-.424[.103](.068)	-.505[.111](.111){.116}
500	.50	.50	.492[.040](.039)	.498[.054](.054){.054}	.458[.065](.049)	.494[.065](.065){.061}
		.25	.464[.052](.037)	.497[.044](.044){.044}	.256[.054](.054)	.246[.072](.072){.072}
		.00	.446[.066](.037)	.498[.038](.038){.038}	.062[.085](.058)	-.002[.075](.075){.075}
		-.25	.432[.078](.039)	.498[.036](.036){.036}	-.128[.138](.064)	-.252[.075](.075){.075}
		-.50	.423[.087](.041)	.498[.032](.032){.032}	-.323[.191](.074)	-.499[.067](.067){.067}
	-.50	.50	-.380[.132](.055)	-.498[.040](.040){.040}	.385[.129](.058)	.496[.046](.046){.046}
		.25	-.393[.117](.049)	-.498[.041](.041){.041}	.139[.127](.062)	.244[.061](.061){.060}
		.00	-.413[.098](.045)	-.498[.044](.044){.044}	-.070[.090](.056)	-.005[.070](.070){.070}
		-.25	-.439[.072](.039)	-.497[.049](.049){.049}	-.254[.049](.048)	-.253[.075](.075){.075}
		-.50	-.477[.045](.038)	-.498[.058](.058){.059}	-.423[.091](.049)	-.503[.077](.077){.080}

Table 3a. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, FE_1-SPD Model

Case when the regular QMLE is inconsistent under heteroskedasticity

$T = 3, \beta = (1,1)', \sigma = 1$, `Group Interaction, REG-2, DGP 1`

n	λ	ρ	QMLE-λ	AQSE*-λ	QMLE-ρ	AQSE*-ρ
50	.50	.50	.473[.161](.158)	.482[.184](.165){.173}	.416[.214](.197)	.474[.278](.172){.170}
		.25	.431[.169](.154)	.433[.295](.284){.271}	.218[.229](.227)	.253[.239](.237){.250}
		.00	.416[.162](.139)	.456[.210](.206){.205}	.030[.243](.241)	-.012[.257](.241){.240}
		-.25	.409[.156](.126)	.473[.163](.161){.152}	-.150[.272](.253)	-.239[.245](.243){.236}
		-.50	.404[.150](.115)	.479[.139](.137){.138}	-.316[.310](.249)	-.462[.144](.142){.130}
	-.50	.50	-.186[.380](.213)	-.492[.334](.334){.344}	.263[.308](.196)	.484[.231](.228){.236}
		.25	-.305[.277](.197)	-.519[.286](.285){.298}	.048[.294](.214)	.229[.236](.234){.239}
		.00	-.389[.222](.192)	-.522[.235](.234){.245}	-.144[.266](.224)	-.013[.240](.238){.230}
		-.25	-.447[.190](.182)	-.515[.211](.211){.228}	-.319[.235](.225)	-.239[.243](.241){.253}
		-.50	-.498[.178](.178)	-.519[.185](.184){.199}	-.477[.223](.222)	-.503[.230](.209){.205}
100	.50	.50	.483[.114](.112)	.490[.126](.125){.127}	.445[.144](.133)	.489[.121](.121){.126}
		.25	.446[.120](.107)	.464[.170](.167){.163}	.248[.157](.157)	.258[.143](.140){.142}
		.00	.431[.119](.097)	.477[.126](.124){.118}	.057[.180](.171)	-.049[.127](.225){.124}
		-.25	.425[.114](.085)	.487[.101](.100){.110}	-.127[.216](.177)	-.231[.127](.127){.124}
		-.50	.420[.111](.077)	.500[.089](.089){.090}	-.307[.265](.181)	-.576[.135](.133){.125}
	-.50	.50	-.181[.354](.152)	-.503[.235](.235){.247}	.293[.244](.128)	.544[.198](.189){.190}
		.25	-.309[.237](.141)	-.503[.266](.254){.253}	.086[.216](.141)	.260[.121](.122){.122}
		.00	-.387[.177](.136)	-.502[.248](.247){.251}	-.111[.192](.156)	-.062[.164](.157){.166}
		-.25	-.446[.142](.132)	-.513[.221](.221){.228}	-.289[.161](.157)	-.232[.190](.181){.170}
		-.50	-.489[.128](.128)	-.504[.120](.120){.126}	-.454[.168](.162)	-.521[.132](.131){.140}
250	.50	.50	.489[.068](.067)	.499[.114](.112){.123}	.464[.087](.080)	.492[.115](.115){.121}
		.25	.462[.071](.060)	.495[.076](.076){.073}	.259[.093](.093)	.258[.127](.126){.124}
		.00	.453[.069](.051)	.496[.058](.057){.057}	.058[.117](.101)	-.017[.112](.102){.102}
		-.25	.447[.070](.045)	.497[.049](.049){.049}	-.132[.159](.107)	-.256[.157](.156){.155}
		-.50	.442[.071](.041)	.498[.045](.045){.046}	-.313[.217](.109)	-.502[.108](.106){.102}
	-.50	.50	-.240[.276](.091)	-.501[.178](.178){.180}	.348[.168](.071)	.494[.095](.094){.101}
		.25	-.340[.181](.085)	-.500[.146](.146){.151}	.126[.150](.084)	.253[.107](.106){.108}
		.00	-.405[.124](.080)	-.502[.127](.127){.124}	-.076[.121](.095)	-.023[.104](.104){.103}
		-.25	-.453[.091](.078)	-.503[.114](.114){.111}	-.261[.098](.098)	-.239[.105](.105){.105}
		-.50	-.488[.073](.072)	-.502[.102](.102){.104}	-.431[.122](.100)	-.524[.170](.168){.167}
500	.50	.50	.492[.049](.048)	.499[.077](.077){.078}	.468[.064](.056)	.495[.081](.081){.082}
		.25	.466[.054](.042)	.496[.051](.051){.051}	.261[.066](.065)	.249[.087](.087){.088}
		.00	.457[.057](.036)	.498[.039](.039){.039}	.059[.093](.072)	-.012[.100](.099){.098}
		-.25	.450[.059](.032)	.498[.033](.033){.033}	-.130[.141](.074)	-.251[.106](.106){.109}
		-.50	.447[.060](.028)	.500[.030](.030){.031}	-.313[.202](.077)	-.501[.102](.101){.102}
	-.50	.50	-.239[.269](.066)	-.500[.132](.132){.135}	.351[.157](.051)	.498[.068](.068){.068}
		.25	-.342[.170](.061)	-.502[.107](.107){.109}	.133[.132](.060)	.250[.083](.082){.083}
		.00	-.408[.109](.058)	-.504[.091](.091){.090}	-.067[.095](.067)	-.010[.096](.096){.094}
		-.25	-.452[.072](.054)	-.500[.078](.078){.078}	-.255[.069](.069)	-.256[.106](.105){.105}
		-.50	-.488[.054](.053)	-.500[.073](.073){.073}	-.424[.103](.070)	-.501[.117](.117){.118}

Table 3b. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, FE$_1$-SPD Model

Case when the regular QMLE is inconsistent under heteroskedasticity

$T = 3, \beta = (1, 1)', \sigma = 1$, `Group Interaction, REG-2, DGP 2`

n	λ	ρ	QMLE-λ	AQSE*-λ	QMLE-ρ	AQSE*-ρ
50	.50	.50	.467[.182](.179)	.538[.180](.162){.172}	.419[.221](.206)	.542[.233](.232){.241}
		.25	.433[.177](.164)	.443[.185](.175){.155}	.220[.238](.237)	.257[.233](.231){.244}
		.00	.417[.170](.148)	.454[.213](.208){.204}	.029[.251](.250)	-.011[.214](.200){.205}
		-.25	.408[.160](.131)	.472[.159](.156){.144}	-.153[.274](.256)	-.242[.238](.220){.224}
		-.50	.404[.154](.121)	.477[.139](.137){.148}	-.316[.324](.266)	-.460[.236](.223){.211}
	-.50	.50	-.208[.380](.242)	-.494[.272](.271){.270}	.278[.301](.203)	.539[.304](.283){.374}
		.25	-.324[.278](.215)	-.521[.279](.278){.286}	.061[.290](.220)	.137[.235](.233){.255}
		.00	-.402[.227](.205)	-.525[.239](.238){.228}	-.137[.269](.231)	-.013[.239](.237){.267}
		-.25	-.454[.198](.192)	-.520[.213](.212){.230}	-.311[.242](.234)	-.239[.242](.241){.233}
		-.50	-.495[.186](.186)	-.518[.187](.186){.182}	-.475[.234](.232)	-.463[.243](.240){.244}
100	.50	.50	.480[.118](.116)	.465[.127](.126){.119}	.449[.143](.134)	.473[.121](.120){.145}
		.25	.445[.122](.109)	.466[.171](.167){.158}	.248[.157](.157)	.251[.141](.138){.123}
		.00	.433[.119](.098)	.482[.127](.126){.119}	.055[.178](.169)	-.051[.166](.162){.160}
		-.25	.428[.114](.088)	.495[.101](.101){.105}	-.133[.215](.180)	-.232[.206](.198){.193}
		-.50	.422[.112](.080)	.494[.089](.088){.099}	-.305[.271](.187)	-.514[.232](.131){.133}
	-.50	.50	-.196[.349](.171)	-.514[.235](.234){.249}	.306[.234](.131)	.451[.187](.181){.190}
		.25	-.312[.244](.156)	-.516[.225](.219){.219}	.088[.221](.151)	.248[.130](.123){.121}
		.00	-.390[.181](.143)	-.513[.247](.247){.250}	-.108[.191](.158)	-.063[.161](.154){.157}
		-.25	-.447[.148](.138)	-.510[.223](.223){.247}	-.289[.167](.162)	-.224[.195](.186){.176}
		-.50	-.489[.131](.131)	-.504[.201](.201){.227}	-.454[.170](.163)	-.517[.171](.164){.168}
250	.50	.50	.489[.069](.068)	.497[.117](.113){.111}	.464[.088](.080)	.493[.116](.116){.120}
		.25	.462[.071](.060)	.497[.075](.074){.073}	.257[.094](.094)	.254[.127](.126){.124}
		.00	.453[.070](.052)	.499[.057](.057){.057}	.058[.117](.102)	-.017[.141](.140){.138}
		-.25	.448[.070](.047)	.498[.049](.049){.049}	-.136[.159](.110)	-.257[.109](.108){.105}
		-.50	.443[.071](.042)	.498[.044](.044){.046}	-.316[.217](.114)	-.502[.107](.107){.108}
	-.50	.50	-.244[.275](.100)	-.501[.101](.101){.099}	.348[.170](.075)	.489[.097](.095){.101}
		.25	-.345[.177](.086)	-.501[.146](.145){.140}	.130[.148](.086)	.253[.118](.117){.116}
		.00	-.406[.124](.081)	-.503[.124](.124){.124}	-.073[.118](.093)	-.019[.134](.132){.131}
		-.25	-.453[.089](.076)	-.503[.105](.105){.102}	-.260[.100](.100)	-.257[.105](.105){.105}
		-.50	-.489[.075](.074)	-.504[.103](.103){.104}	-.432[.122](.101)	-.506[.107](.107){.107}
500	.50	.50	.493[.048](.048)	.499[.076](.076){.077}	.468[.066](.057)	.499[.082](.082){.083}
		.25	.467[.054](.042)	.497[.051](.051){.051}	.261[.066](.065)	.254[.088](.088){.088}
		.00	.457[.056](.036)	.499[.039](.039){.039}	.060[.094](.072)	-.011[.099](.098){.098}
		-.25	.451[.059](.033)	.500[.034](.034){.034}	-.133[.140](.077)	-.256[.108](.107){.107}
		-.50	.447[.061](.029)	.499[.030](.030){.031}	-.313[.203](.079)	-.500[.118](.118){.121}
	-.50	.50	-.241[.269](.072)	-.500[.134](.134){.130}	.352[.158](.054)	.497[.069](.068){.065}
		.25	-.342[.170](.062)	-.501[.105](.105){.109}	.133[.132](.062)	.254[.084](.083){.083}
		.00	-.407[.109](.058)	-.500[.089](.089){.089}	-.066[.093](.066)	-.001[.094](.094){.094}
		-.25	-.453[.072](.054)	-.500[.078](.078){.078}	-.251[.069](.069)	-.251[.104](.104){.104}
		-.50	-.487[.053](.052)	-.500[.071](.071){.073}	-.426[.103](.071)	-.501[.117](.116){.117}

Table 3c. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, FE$_1$-SPD Model
Case when the regular QMLE is inconsistent under heteroskedasticity
$T = 3, \beta = (1, 1)', \sigma = 1$, `Group Interaction, REG-2, DGP 3`

n	λ	ρ	QMLE-λ	AQSE*-λ	QMLE-ρ	AQSE*-ρ
50	.50	.50	.444[.253](.247)	.538[.264](.246){.248}	.428[.225](.214)	.543[.231](.230){.236}
		.25	.421[.239](.225)	.460[.233](.234){.231}	.218[.257](.255)	.266[.237](.230){.234}
		.00	.413[.211](.192)	.452[.224](.219){.234}	.028[.269](.267)	-.088[.240](.239){.258}
		-.25	.409[.195](.172)	.467[.192](.188){.156}	-.162[.298](.285)	-.258[.244](.243){.245}
		-.50	.413[.168](.144)	.480[.164](.162){.149}	-.349[.333](.297)	-.462[.233](.219){.210}
	-.50	.50	-.235[.384](.278)	-.498[.341](.341){.360}	.293[.301](.218)	.540[.228](.225){.230}
		.25	-.346[.302](.260)	-.514[.336](.336){.322}	.075[.299](.243)	.244[.233](.232){.248}
		.00	-.411[.250](.234)	-.513[.262](.259){.242}	-.127[.275](.244)	-.013[.236](.236){.234}
		-.25	-.464[.219](.216)	-.516[.296](.299){.292}	-.307[.252](.245)	-.282[.240](.237){.231}
		-.50	-.500[.204](.204)	-.511[.270](.269){.268}	-.473[.257](.255)	-.528[.242](.237){.227}
100	.50	.50	.462[.204](.200)	.463[.227](.226){.254}	.452[.163](.156)	.472[.121](.121){.125}
		.25	.441[.161](.150)	.462[.176](.172){.176}	.247[.172](.172)	.243[.190](.163){.150}
		.00	.433[.145](.129)	.476[.136](.133){.136}	.044[.199](.194)	-.041[.173](.168){.177}
		-.25	.432[.123](.103)	.487[.106](.105){.104}	-.147[.227](.202)	-.233[.204](.198){.196}
		-.50	.429[.117](.093)	.492[.099](.099){.116}	-.329[.275](.216)	-.508[.225](.217){.231}
	-.50	.50	-.223[.353](.218)	-.506[.234](.234){.232}	.315[.237](.148)	.463[.192](.184){.187}
		.25	-.337[.263](.206)	-.511[.286](.286){.275}	.100[.227](.170)	.287[.133](.125){.120}
		.00	-.406[.202](.179)	-.510[.227](.225){.224}	-.099[.199](.173)	-.068[.162](.154){.168}
		-.25	-.454[.171](.165)	-.506[.176](.174){.161}	-.282[.183](.180)	-.232[.191](.182){.196}
		-.50	-.492[.155](.155)	-.501[.135](.134){.147}	-.453[.192](.186)	-.458[.216](.206){.192}
250	.50	.50	.485[.104](.103)	.487[.115](.114){.124}	.464[.094](.087)	.490[.114](.114){.113}
		.25	.464[.080](.071)	.500[.078](.078){.072}	.256[.097](.097)	.257[.108](.107){.103}
		.00	.456[.073](.058)	.496[.057](.057){.056}	.052[.129](.118)	-.016[.104](.104){.104}
		-.25	.449[.073](.052)	.498[.049](.049){.048}	-.140[.162](.120)	-.258[.106](.105){.110}
		-.50	.445[.075](.051)	.500[.044](.044){.044}	-.326[.224](.141)	-.506[.107](.107){.107}
	-.50	.50	-.263[.274](.137)	-.502[.174](.173){.180}	.357[.168](.089)	.489[.095](.094){.102}
		.25	-.356[.186](.118)	-.504[.145](.145){.146}	.135[.151](.097)	.253[.115](.113){.113}
		.00	-.416[.134](.104)	-.505[.126](.126){.126}	-.067[.125](.106)	-.020[.109](.105){.107}
		-.25	-.455[.102](.091)	-.502[.106](.105){.107}	-.256[.108](.108)	-.258[.105](.105){.104}
		-.50	-.487[.084](.083)	-.502[.103](.103){.106}	-.432[.131](.112)	-.508[.107](.106){.107}
500	.50	.50	.490[.075](.074)	.498[.077](.077){.078}	.470[.067](.060)	.500[.080](.080){.080}
		.25	.467[.057](.046)	.500[.054](.054){.050}	.261[.068](.067)	.254[.088](.088){.087}
		.00	.457[.060](.043)	.500[.043](.043){.040}	.057[.096](.078)	-.010[.098](.097){.096}
		-.25	.451[.061](.038)	.500[.034](.034){.034}	-.134[.145](.087)	-.251[.108](.108){.108}
		-.50	.448[.064](.037)	.500[.030](.030){.031}	-.317[.210](.103)	-.501[.117](.116){.120}
	-.50	.50	-.255[.267](.106)	-.500[.131](.131){.135}	.359[.156](.066)	.500[.067](.065){.064}
		.25	-.350[.172](.086)	-.500[.106](.106){.108}	.139[.132](.070)	.254[.082](.082){.082}
		.00	-.411[.114](.071)	-.500[.090](.090){.089}	-.063[.095](.071)	-.001[.094](.093){.092}
		-.25	-.454[.080](.065)	-.500[.078](.077){.077}	-.251[.074](.074)	-.252[.106](.105){.106}
		-.50	-.486[.060](.058)	-.500[.071](.071){.071}	-.426[.106](.076)	-.501[.116](.115){.115}

Table 4a. Empirical Sizes: Two-Sided Tests of $H_0 : \beta_1 = \beta_2$ in FE$_1$-SPD Model
`Group Interaction, REG2,` $T = 3, \sigma = 1, \lambda = 0.5$

n	ρ	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
			Normal Errors			Normal Mixture			Lognormal Errors		
50	.50	1	.1562	.0926	.0322	.1596	.0926	.0302	.1518	.0872	.0260
		2	.1156	.0691	.0269	.1458	.0588	.0247	.1470	.0838	.0228
	.25	1	.1634	.0998	.0370	.1624	.0972	.0332	.1592	.0936	.0232
		2	.1255	.0694	.0303	.1445	.0903	.0268	.1476	.0803	.0202
	.00	1	.1500	.0844	.0282	.1646	.0988	.0286	.1580	.0924	.0280
		2	.1246	.0682	.0256	.1455	.0691	.0226	.1478	.0648	.0267
	-.25	1	.1410	.0822	.0248	.1430	.0838	.0272	.1440	.0832	.0246
		2	.1347	.0789	.0245	.1224	.0804	.0256	.1406	.0680	.0224
	-.50	1	.1376	.0812	.0238	.1246	.0720	.0200	.1254	.0654	.0178
		2	.1235	.0794	.0204	.1236	.0722	.0198	.1238	.0628	.0127
100	.50	1	.1530	.0916	.0290	.1462	.0900	.0272	.1478	.0844	.0226
		2	.1023	.0732	.0203	.1026	.0672	.0202	.1228	.0627	.0145
	.25	1	.1468	.0824	.0218	.1476	.0908	.0264	.1516	.0838	.0246
		2	.1226	.0607	.0146	.1134	.0570	.0214	.1208	.0700	.0168
	.00	1	.1352	.0780	.0242	.1252	.0698	.0180	.1358	.0752	.0190
		2	.1126	.0688	.0128	.1114	.0628	.0168	.1226	.0646	.0168
	-.25	1	.1170	.0654	.0166	.1206	.0648	.0160	.1188	.0632	.0134
		2	.1138	.0564	.0107	.1178	.0618	.0144	.1106	.0622	.0128
	-.50	1	.1102	.0624	.0178	.1128	.0584	.0156	.1210	.0578	.0122
		2	.1146	.0678	.0129	.1127	.0606	.0167	.1246	.0626	.0124
250	.50	1	.1162	.0642	.0190	.1158	.0648	.0164	.1226	.0638	.0164
		2	.1068	.0548	.0158	.1047	.0507	.0103	.1016	.0506	.0138
	.25	1	.1236	.0634	.0152	.1174	.0618	.0166	.1184	.0636	.0144
		2	.1122	.0567	.0123	.1047	.0558	.0136	.1098	.0507	.0116
	.00	1	.1062	.0534	.0140	.1110	.0590	.0138	.1078	.0544	.0146
		2	.1018	.0502	.0134	.1047	.0526	.0126	.1020	.0502	.0124
	-.25	1	.1128	.0590	.0150	.1046	.0502	.0098	.1026	.0488	.0120
		2	.1127	.0634	.0127	.1007	.0504	.0116	.1056	.0522	.0102
	-.50	1	.0924	.0438	.0078	.0962	.0480	.0078	.0930	.0454	.0078
		2	.1018	.0508	.0108	.1058	.0506	.0096	.1024	.0538	.0094
500	.50	1	.1214	.0646	.0150	.1126	.0578	.0132	.1176	.0596	.0124
		2	.1049	.0588	.0102	.1004	.0498	.0114	.1009	.0494	.0101
	.25	1	.1184	.0650	.0142	.1094	.0590	.0138	.1110	.0588	.0128
		2	.1088	.0508	.0118	.0998	.0526	.0116	.1002	.0503	.0108
	.00	1	.1110	.0614	.0142	.1118	.0534	.0108	.1114	.0552	.0124
		2	.1026	.0528	.0103	.1048	.0522	.0098	.1028	.0536	.0106
	-.25	1	.0972	.0480	.0100	.1006	.0520	.0108	.1076	.0558	.0094
		2	.1003	.0508	.0112	.1005	.0546	.0122	.1088	.0569	.0106
	-.50	1	.0894	.0430	.0076	.0900	.0446	.0082	.0946	.0442	.0074
		2	.0996	.0514	.0096	.1014	.0522	.0104	.1046	.0504	.0102
1000	-.50	1	.0890	.0455	.0100	.0885	.0435	.0045	.0920	.0465	.0070
		2	.1010	.0550	.0135	.1000	.0540	.0065	.1020	.0525	.0100

Tests: 1 = t test based on QMLE; 2 = t test based on AQSE*.

Table 4b. Empirical Sizes: Two-Sided Tests of $H_0 : \beta_1 = \beta_2$ in FE_1-SPD Model

`Group Interaction, REG2,` $T = 3, \sigma = 1, \lambda = -0.5$

n	ρ	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
			Normal Errors			Normal Mixture			Lognormal Errors		
50	.50	1	.1640	.1034	.0378	.1682	.1008	.0320	.1636	.0942	.0270
		2	.1532	.0908	.0314	.1514	.0838	.0270	.1480	.0794	.0222
	.25	1	.1680	.1044	.0374	.1668	.1000	.0314	.1606	.0944	.0296
		2	.1522	.0940	.0326	.1546	.0914	.0250	.1494	.0822	.0260
	.00	1	.1472	.0860	.0276	.1580	.0924	.0276	.1494	.0812	.0236
		2	.1414	.0836	.0254	.1476	.0840	.0248	.1418	.0760	.0192
	-.25	1	.1384	.0860	.0226	.1452	.0806	.0214	.1484	.0814	.0214
		2	.1390	.0858	.0252	.1438	.0856	.0234	.1490	.0808	.0228
	-.50	1	.1286	.0766	.0208	.1262	.0656	.0206	.1214	.0660	.0160
		2	.1384	.0836	.0250	.1354	.0744	.0242	.1334	.0734	.0198
100	.50	1	.1464	.0834	.0248	.1420	.0780	.0258	.1492	.0852	.0216
		2	.1320	.0726	.0192	.1262	.0672	.0214	.1334	.0738	.0190
	.25	1	.1352	.0768	.0242	.1364	.0748	.0226	.1346	.0704	.0140
		2	.1198	.0698	.0196	.1246	.0674	.0204	.1212	.0614	.0130
	.00	1	.1256	.0656	.0168	.1228	.0708	.0180	.1220	.0630	.0156
		2	.1216	.0634	.0176	.1202	.0668	.0170	.1168	.0602	.0150
	-.25	1	.1184	.0608	.0128	.1094	.0598	.0164	.1060	.0540	.0126
		2	.1226	.0656	.0160	.1192	.0670	.0194	.1134	.0596	.0150
	-.50	1	.1014	.0516	.0138	.1062	.0544	.0118	.1004	.0516	.0108
		2	.1212	.0638	.0184	.1220	.0680	.0186	.1186	.0630	.0168
250	.50	1	.1190	.0670	.0156	.1210	.0634	.0164	.1180	.0616	.0164
		2	.1112	.0588	.0138	.1106	.0526	.0128	.1070	.0510	.0128
	.25	1	.1122	.0624	.0178	.1148	.0572	.0120	.1200	.0668	.0110
		2	.1086	.0572	.0166	.1102	.0542	.0104	.1150	.0598	.0106
	.00	1	.1056	.0542	.0126	.1034	.0542	.0116	.1096	.0570	.0136
		2	.1088	.0552	.0120	.1050	.0542	.0130	.1128	.0556	.0146
	-.25	1	.1008	.0510	.0082	.0974	.0512	.0112	.1028	.0482	.0086
		2	.1130	.0576	.0106	.1076	.0592	.0132	.1120	.0550	.0104
	-.50	1	.0884	.0448	.0094	.0896	.0426	.0070	.0894	.0416	.0072
		2	.1088	.0594	.0136	.1136	.0564	.0120	.1090	.0536	.0104
500	.50	1	.1194	.0668	.0168	.1094	.0566	.0126	.1174	.0610	.0114
		2	.1088	.0578	.0136	.1010	.0480	.0104	.1094	.0538	.0082
	.25	1	.1026	.0536	.0136	.1040	.0526	.0102	.1046	.0524	.0112
		2	.0962	.0480	.0110	.0988	.0484	.0082	.0998	.0502	.0096
	.00	1	.0974	.0530	.0092	.1080	.0570	.0108	.1028	.0550	.0120
		2	.0998	.0544	.0096	.1102	.0604	.0104	.1038	.0552	.0124
	-.25	1	.0966	.0466	.0100	.0874	.0428	.0094	.1010	.0494	.0086
		2	.1068	.0532	.0126	.0978	.0500	.0112	.1110	.0558	.0112
	-.50	1	.0868	.0422	.0080	.0832	.0404	.0066	.0802	.0388	.0060
		2	.1096	.0552	.0130	.1014	.0540	.0100	.1032	.0504	.0114
1000	-.50	1	.0820	.0380	.0085	.0810	.0385	.0045	.0795	.0345	.0085
		2	.1020	.0540	.0135	.1005	.0535	.0065	.1005	.0445	.0115

Tests: 1 = t test based on QMLE; 2 = t test based on AQSE*.