Unified M-Estimation of Fixed-Effects Spatial Dynamic Models with Short Panels*

Zhenlin Yang

School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903

Email: zlyang@smu.edu.sg

January 16, 2016

Abstract

It is well known that quasi maximum likelihood (QML) estimation of dynamic panel data (DPD) models with short panels depends on the assumptions on the initial values, and a wrong treatment of them will result in inconsistency and serious bias. The same issues apply to spatial DPD (SDPD) models with short panels. In this paper, a unified M-estimation method is proposed for estimating the fixed-effects SDPD models containing three major types of spatial effects, namely spatial lag, spatial error and space-time lag. The method is free from the specification of the distribution of the initial observations and robust against nonnormality of the errors. Consistency and asymptotic normality of the proposed M-estimator are established. A martingale difference representation of the underlying estimating functions is developed, which leads to an initial-condition free estimate of the variance of the M-estimators. Monte Carlo results show that the proposed methods have excellent finite sample performance.

Key Words: Adjusted quasi score; Dynamic panels; Fixed effects; Initial-condition free estimation; Martingale difference; Spatial effects; Short panels.

JEL classifications: C10, C13, C21, C23, C15

1 Introduction

In the majority of empirical microeconometric research involving panel data, a panel with a large number of cross-sectional units and a small number of time periods, called a short panel, remains the prevalent setting (Hsiao et al., 2002; Binder et al., 2005), and evidence from the standard dynamic panel data models shows that maximum likelihood (ML) estimators are more efficient than GMM estimators (Hsiao et al., 2002; Binder et al., 2005; Bun and Caree, 2005; Gouriéroux, et al., 2010; Kruiniger, 2013). Hsiao (2003, Ch. 4) gives an excellent summary on dynamic panel data (DPD) models with random or fixed effects.

*I would like to thank Anil Bera, Ingmar Prucha, the participants of the VIII World Conference of the Spatial Econometrics Association, ETH Zurich, 2014, and the seminar participants of Shanghai University of Finance and Economics, May 8, 2015, for their helpful comments. The financial support from Singapore Management University under Grant C244/MSS12E007 is gratefully acknowledged.

In recent years, there has been a growing interest in the estimation of dynamic panel data models with cross-sectional or spatial dependence, arising from economic processes such as housing decisions, technology adoption, unemployment, welfare participation, price decisions, etc. The resulted models are referred to as spatial DPD (SDPD) models (Anselin, 2001; Anselin et al., 2008), where the spatial effects may appear in the model either in the form of spatial lag(s) of the response variable (Yu et al., 2008; Yu and Lee, 2010; Lee and Yu, 2010a; Korniotis, 2010; Elhorst, 2010), or in the form of spatial errors (Elhorst, 2005; Yang et al., 2006; Mutl, 2006; Su and Yang, 2015). Lee and Yu (2010b, 2015a) provide an excellent survey on the SDPD models. Most of the studies on the SDPD modes are either based on the GMM-type method or under a large panel set-up, except Elhorst (2010) and Su and Yang (2015) who consider the quasi-ML (QML) estimation of the SDPD model with short panels.

As ML estimators are more efficient than the GMM estimators, and the latter can perform poorly (Gouriéroux, et al., 2010), it is natural to expect that any ML-type estimation be more efficient than the corresponding GMM estimation. The main difficulty in using ML or QML method to estimate the DPD or SDPD models with short panels is the modeling of the initial observations of the response vector, say y_0, for the random effects model, or the initial differences, say Δy_1, for the fixed effects model. This is because y_0 may be exogenous in the sense that it varies autonomously, independent of other variables in the model; or endogenous in the sense that it is generated in the same way as the other values of the response vector y in the latter time periods. In case that y_0 is endogenous, it depends on the processes starting values and the past values of time-varying regressors, both of which are not observable, leading to incidental parameters. In the case of fixed effects model, Δy_1 is endogenous whether y_0 is exogenous or endogenous and this incidental parameters problem always exists. The traditional way of handling this problem is to predict these quantities using the observed values of the regressors (Anderson and Hisao, 1981, 1982; Bhargava and Sargan, 1983; Hsiao et al. 2002; Elhorst, 2010; Su and Yang, 2015).[1] However, the model for the initial differences involves the unknown process starting time. Also, its predictability typically requires that the time-varying regressors be stationary or trend-stationary. And, when there are many time-varying regressors in the model, modelling the initial difference may introduce 'too many' additional parameters, causing a efficiency decline. Most importantly, this linear projection method may not be applicable to an SDPD model with spatial lags (see the footnote at the end of Section 2 for some details). It is therefore highly desirable to have a general method that is free from the specification of the initial conditions.

In this paper, we propose a unified, initial-condition free approach to estimate the SDPD models with fixed effects, allowing all three major types of spatial dependence to be present in the model, namely, the spatial lag, space-time lag, and spatial error. The approach starts from the 'conditional' quasi-likelihood, with the initial differences being treated as if they are exogenous, and then makes corrections on the conditional quasi-score functions to give a set

[1]In case of fixed effects models, the incidental parameters problem also occurs in the model itself (the fixed effects), but this problem can be resolved by first-differencing or some kind of orthogonal transformations.

of unbiased estimating equations.[2] Solving these unbiased estimating equations (EFs) leads to estimators that are consistent and asymptotically normal. It turns out that the corrections or the adjustments on the conditional quasi scores are totally free from the specification of the distribution of the initial differences, resulting *initial-condition free* estimators for the SDPS model. The proposed estimator is simply referred to in this paper as the M-estimator according to Huber (1981) or van der Vaart (1998).[3]

For initial-condition free inferences, a martingale differences (MDs) representation of the EFs (being the adjusted quasi scores) is developed, and the average of the outer products of the MDs (OPMD) is shown to give a consistent and initial-condition free estimate of the variance covariance (VC) matrix of the EFs. This and the estimated Hessian matrix together give a consistent estimate of the VC matrix of the M-estimators, referred to as the OPMD-estimator in this paper. Monte Carlo results show that the proposed M-estimators of the model parameters and the OPMD-estimator of their VC matrix have excellent finite sample performance – robust against the way the initial observations being generated and nonnormality of the error distributions. Under a special submodel where only spatial error is present, the proposed methods are compared with the traditional method where the initial observations are modeled and the full quasi likelihood is used (Su and Yang, 2015). The results show that the two methods are comparable when the initial observations are correctly specified, but the proposed methods are more robust against misspecifications of the initial conditions. Our Monte Carlo results show that the proposed unified M-estimation method, for the FE-SDPD model with all three types of spatial effects, is not only valid when T is small, but also provides better estimators when T is not small, compared with the conditional quasi likelihood approach. The proposed OPMD method for the VC matrix estimation is valid only when T is small, but when T is large, a plug-in method based on the conditional variance of the adjusted quasi scores, treating the initial differences as exogenous, can be used.

The rest of the paper is organized as follows. Section 2 describes the general SDPD model, its submodels, and discusses the limitations of the method of modelling the initial conditions with the SDPD models. Section 3 introduces the unified M-estimation method for the general SDPD model with fixed effects, presents the asymptotic properties of the proposed M-estimators, and introduces the OPMD method for VC matrix estimation and proves its consistency. Section 4 considers several important submodels and discusses how the general results are simplified, and how the proposed methods compare with the existing QML approach. Section 5 presents Monte Carlo results. Section 6 concludes the paper.

---

[2] Clearly, this 'conditional' quasi likelihood can never be a correct likelihood as Δy_1 is always endogenous, and thus maximizing it may produce inconsistent estimators when the time dimension T is fixed and small. This is intuitively clear as the conditional likelihood ignores the information contained in Δy_1 which is a fixed proportion of the whole data $\Delta y_1, \Delta y_2, \ldots, \Delta y_T$. Beside, Δy_1 may contain some additional information about the model parameters that is accumulated from the past. In this sense, some form of modifications is necessary before this conditional likelihood approach can be followed for model estimation.

[3] The term 'M-estimator' was coined by Huber (1964) to mean maximum-likelihood type. It can be defined in either of the two ways: (a) as the solution of a maximization problem and (b) as the root of an estimating equation. Clearly, our estimation strategy falls into the category (b). van der Vaart (1998) also called the M-estimator defined in (b) a zero estimator. See also Huber (1981) and Newey and MacFadden (1994).

2 Spatial Dynamic Panel Data Models

Consider the spatial dynamic panel data (SDPD) model where the spatial effects appear in the model in the forms of spatial lag (SL), space-time lag (STL), and spatial error (SE):

$$\begin{aligned} y_t &= \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t\beta + Z\gamma + \mu + \alpha_t 1_n + u_t, \\ & u_t = \lambda_3 W_3 u_t + v_t, \quad t = 1, 2, \dots, T, \end{aligned} \tag{2.1}$$

where $y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$ and $v_t = (v_{1t}, v_{2t}, \dots, v_{nt})'$ are $n \times 1$ vectors of response values and idiosyncratic errors at time t, and $\{v_{it}\}$ are independent and identically distributed (iid) across i and t with mean zero and variance σ_v^2; the scalar parameter ρ characterizes the dynamic effect, λ_1 the spatial lag effect, λ_2 the space-time effect, and λ_3 the spatial error effect; $\{X_t\}$ are $n \times p$ matrices containing values of p time-varying exogenous variables, Z is an $n \times q$ matrix containing the values of q time-invariant exogenous variables that may include the constant term, dummy variables (e.g., individuals' gender and race), etc.; β and γ are the usual regression coefficients; $W_r, r = 1, 2, 3$ are the given $n \times n$ spatial weight matrices; and μ is an $n \times 1$ vector of unobserved individual-specific effects, $\{\alpha_t\}$ are the time-specific effects, and 1_n is an $n \times 1$ vector of ones.

Model (2.1) is fairly general. It embeds several important submodels popular in the literature. Thus, it is highly desirable to have a unified method of inference for this general model so that the method can easily be simplified to suit each special model of interest to a particular applied problem. On the other hand, Model (2.1) can be further extended to contain higher-order spatial lags in y_t, in y_{t-1}, as well as in u_t. See the end of Section 3.1 for further discussions. In this paper, we focus on Model (2.1) because it is general enough and further generalizations can be done at the expense of more tedious algebra. Each submodel has its own features and merits, and thus deserves some specific attention.

First, setting λ_1 and λ_2 to zero, Model (2.1) reduces to an SDPD model with only SE,[4] in the form of a spatial autoregressive (SAR) error process,

$$y_t = \rho y_{t-1} + X_t'\beta + Z\gamma + \mu + \alpha_t 1_n + u_t, \;\; u_t = \lambda_3 W_3 u_t + v_t, \;\; t = 1, 2, \dots, T. \tag{2.2}$$

Su and Yang (2015) provide formal asymptotic results for the quasi maximum likelihood (QML) estimation of Model (2.2) with short panels (T small), and random or fixed effects. To give a full quasi likelihood function, the initial observations y_0 or the initial differences Δy_0 are modeled under some fundamental assumptions adapted from Hsiao et al. (2002). These assumptions may not hold and thus the model for the initial observations or differences is subject to model misspecification. See the end of this section for a detailed discussion.

[4]This SE dependence structure was introduced by Anselin (1988). Subsequently, alternative or extended SE structures have been suggested, e.g., to replace the SAR process by a spatial moving average SMA process, to allow μ to be spatially correlated as well, etc. See Kapoor et al. (2007), Anselin et al. (2008), Lee and Yu (2012) and Baltagi et al. (2013). However, with short panels and fixed effects, μ must be eliminated by a data transformation to avoid incidental parameters problem, and after that only the SE structure in u_t is kept.

Setting λ_2 and λ_3 in (2.1) to zero gives an SDPD model with only SL,

$$y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + X_t'\beta + Z\gamma + \mu + \alpha_t 1_n + v_t, \quad t = 1, 2, \ldots, T. \tag{2.3}$$

A spatial model (not necessarily dynamic panel) with only SL effect may be more popular than that with only SE effect, as in the former, both the mean and variance of a spatial unit are directly affected by some other spatial units, whereas in the latter only the variance is so, resulting a model with so-called cross-section dependence.[5]

Setting λ_2 to zero, Model (2.1) reduces to an SDPD model with both SL and SE, also referred to as the SDPD model with SARAR effect in the literature,

$$y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + X_t'\beta + Z\gamma + \mu + \alpha_t \ell_n + u_t, \quad u_t = \lambda_3 W_3 u_t + v_t, \quad t = 1, 2, \ldots, T. \tag{2.4}$$

This model encompasses Models (2.2) and (2.3) has not been formally treated under the QML-type approach.[6] Čížk et al. (2014) considered GMM estimation of this model by extending the three steps approach of Kapoor et al. (2007) with large n and fixed T.

Setting λ_3 in (2.1) to zero gives an SDPD model with SL and STL,

$$y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t'\beta + Z\gamma + \mu + \alpha_t \ell_n + v_t, \quad t = 1, 2, \ldots, T. \tag{2.5}$$

Under fixed effects, Yu et al. (2008) presented formal asymptotic results for the QML estimation of Model (2.5) under large n and large T set-up, and Lee and Yu (2014) studied this model based on GMM approach where n is large and T can be large but small relative to n. The case of large n and fixed T for Model (2.5) has not been formally treated in the literature, in particular under the QML approach.

Setting ρ, λ_1 and λ_3 to zero in Model (2.1), we have a panel data model with only STL dependence, referred to as *pure space recursive* model in Anselin et al. (2008). Finally, when all the spatial parameters are set to zero, Model (2.1) reduces to the regular dynamic panel data (DPD) model, which has been extensively treated in the literature.

The current study focuses on the general Model (2.1) with fixed effects, large n and small T. Under this scenario, Z must be excluded from the model, and $\alpha_t 1_n$ can be merged into $X_t\beta$. Thus, the model under study takes the form:

$$y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t\beta + \mu + u_t, \quad u_t = \lambda_3 W_3 u_t + v_t, \tag{2.6}$$

$t = 1, 2, \ldots, T$. The QML approach requires the initial observations (differences) be specified (modeled) in order to construct the full likelihood function. The standard approach of modelling the initial observations or initial differences is through a linear projection onto the space

[5] An alternative way of modelling the cross-section dependence may be the factor model; see, e.g., Andrews (2005), Pesaran (2006), Bai (2009), and Pesaran and Tosetti (2011).

[6] Anselin et al. (2008, p. 647) point out that combinations of both spatially lagged dependent variables and spatially lagged error terms may lead to identification problems unless the covariate effects are non-zero. For detailed discussions on the identification of the SDPD models, see Elhorst (2012) and Lee and Yu (2015b).

of observed regressors (Hsiao et al., 2002). This approach has been successfully adapted by Su and Yang (2015) to give consistent estimations of Model (2.2) under both random effects and fixed-effects. However, this approach requires the following **initial conditions**:

(i) data collection starts from the $0th$ *period; the processes start from the* $-mth$ *period, i.e.,* m *periods before the start of data collection, where* $m = 0, 1, \ldots,$ *and then evolve according to the prescribed processes, i.e., one of the models described above; (ii) starting positions of the process* y_{-m} *are treated as exogenous; hence the exogenous variables* X_t *and the errors* u_t *start to have impact on the response from the period* $-m+1$ *onwards;* (iii) *all the exogenous quantities* (y_{-m}, X_t) *are considered as random and inferences proceed by conditioning on them, (iv)* $\{x_{it}, t = \ldots, -1, 0, 1, \ldots\}$ *are trend-stationary or first-difference stationary for all* $i = 1, \ldots, n,$ *and (v) the variances of* y_{-m} *are constant.*

Evidently, what happened in the 'past' is not observed, the process starting time or m is unknown, the processes movements may not be the same before and after the start of data collection, and the processes are observed only for a few periods. Hence, the above assumptions, in particular the later part of (i) and (iv), may not hold and the linear projection model for the initial observations may be misspecified.[7] Furthermore, even if these assumptions do hold, the linear projection method for modelling the initial observations/differences may not have a straightforward extension to SDPD models that contain `SL` and/or `STL` structures.[8] Alternative methods that are free from the initial conditions, or the methods without the need of explicitly modelling the initial differences, are therefore highly desirable.

3 Unified M-Estimation of Fixed-Effects SDPD Models

In this section, we present a unified framework for estimating the fixed effects (FE) SDPD models, where all three types of spatial dependence are allowed to present in the model and the time dimension T is allowed to be small and fixed. The basic idea of this unified approach is to first formulate the Gaussian likelihood function conditional on the initial differences Δy_1 as if they are exogenous, and then modify the resulted quasi score function to account for the ignorance of Δy_1 in this 'conditional' quasi Gaussian score. We shall start with short panels, i.e., panels with large n and small and fixed T, to show the exact cause of inconsistency of the estimators based on this conditional likelihood, and how the quasi scores be adjusted so

[7]Consider Model (2.2) after first-difference. Under the **initial conditions**, we obtain,

$$\Delta y_1 = \lambda_3^m \Delta y_{-m+1} + \textstyle\sum_{j=0}^{m-1} \lambda_3^j \Delta x_{-j+1}\beta + \sum_{j=0}^{m-1} \lambda_3^j B_3^{-(j+1)} \Delta v_{-j+1},$$

by successive backward substitutions. Clearly, the exogenous part $\Delta\eta_1$ of Δy_1 enjoys an approximately linear structure, which makes the linear project of $\Delta\eta_1$ onto the space of the observed $\Delta x_t, t = 1, \ldots, T,$ valid. However, if the specified initial conditions do not hold, this linear projection is in doubt.

[8]Consider simply the first-differenced Model (2.3). Under the **initial conditions**, we obtain,

$$\Delta y_1 = \lambda_1^m B_1^{-m} \Delta y_{-m+1} + \textstyle\sum_{j=0}^{m-1} \lambda_1^j B_1^{-(j+1)} \Delta x_{-j+1}\beta + \sum_{j=0}^{m-1} \lambda_1^j B_1^{-(j+1)} \Delta v_{-j+1},$$

by successive backward substitutions. Now, different from the SDPD model with `SE` only, the exogenous part $\Delta\eta_1$ of Δy_1 also contains spatial effect through B_1, and the linear structure is no longer there.

as to give consistent estimators. Then we argue that when T grows with n the proposed estimation strategy remains valid, and in fact it gives better estimators than those based on the conditional likelihood, which is the usual QML estimators under the large n and large T set up (see, e.g., Yu et al., 2008). We first present the results for the most general model, and then in Section 4 specialize them to each of several submodels to facilitate the practical applications and to compare with the existing QML estimation if available. Proofs of the theoretical results are lengthy, in particular the proofs of the theorems, and are put in Appendix B (for the lemmas) and Appendix C (for the theorems).

3.1 The M-estimation

To facilitate the introduction of the general theory and method, we differentiate the true value of a parameter vector from its general value by adding by adding a subscript '0', e.g., β_0 is the true value of β, and emphasize that Model (2.6) holds only under the true parameter values. Following the standard practice, we eliminate μ by first-differencing (2.6) to give,

$$\Delta y_t = \rho_0 \Delta y_{t-1} + \lambda_{10} W_1 \Delta y_t + \lambda_{20} W_2 \Delta y_{t-1} + \Delta X_t \beta_0 + \Delta u_t, \quad \Delta u_t = \lambda_{30} W_3 \Delta u_t + \Delta v_t, \tag{3.1}$$

for $t = 2, 3, \cdots, T$. The parameters left in Model (3.1) are $\psi_0 = \{\beta_0', \sigma_{v0}^2, \rho_0, \lambda_0'\}'$ where $\lambda_0 = (\lambda_{10}, \lambda_{20}, \lambda_{30})'$. Note that Δy_1 depends on both the initial observations y_0 and the first period observations y_1. Thus, even if y_0 is exogenous, y_1 and thus Δy_1 is not. However, we still formulate a likelihood function as if Δy_1 is exogenous, and then make corrections on the relevant elements of the score function. Let $\Delta Y = \{\Delta y_2', \ldots, \Delta y_T'\}'$, $\Delta Y_{-1} = \{\Delta y_1', \ldots, \Delta y_{T-1}'\}'$, $\Delta X = \{\Delta X_2', \ldots, \Delta X_T'\}'$, and $\Delta v = \{\Delta v_2', \ldots, \Delta v_T'\}'$. Let $\mathbf{W}_r = I_{T-1} \otimes W_r, r = 1, 2, 3,$ where $\otimes$ denotes the Kronecker product and I_m an $m \times m$ identity matrix. Define $B_r(\lambda_r) = I_n - \lambda_r W_r, r = 1, 3,$ and $B_2(\rho, \lambda_2) = \rho I_n + \lambda_2 W_2$. Model (3.1) can be written as:

$$\Delta Y = \rho_0 \Delta Y_{-1} + \lambda_{10} \mathbf{W}_1 \Delta Y + \lambda_{20} \mathbf{W}_2 \Delta Y_{-1} + \Delta X \beta_0 + \Delta u, \quad \Delta u = \lambda_{30} \mathbf{W}_3 \Delta u + \Delta v. \tag{3.2}$$

It is easy to see that

$$\text{Var}(\Delta u) = \sigma_{v0}^2 \big\{ C \otimes [B_3'(\lambda_{30}) B_3(\lambda_{30})]^{-1} \big\} \equiv \sigma_{v0}^2 \Omega(\lambda_{30}),$$

where C is a $(T-1) \times (T-1)$ constant matrix,

$$C = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}.$$

Under normality of v_t, the joint distribution of Δu can be easily obtained, which translates directly to the conditional joint distribution of ΔY. The quasi Gaussian loglikelihood of ψ

in terms of $\Delta y_2, \ldots, \Delta y_T$, as if Δy_1 is exogenous, has the form, ignoring the constant term:

$$\ell_{\texttt{STLE}}(\psi) = -\tfrac{n(T-1)}{2}\log(\sigma_v^2) - \tfrac{1}{2}\log|\Omega(\lambda_3)| + \log|\mathbf{B}_1(\lambda_1)| - \tfrac{1}{2\sigma_v^2}\Delta u(\theta)'\Omega(\lambda_3)^{-1}\Delta u(\theta), \quad (3.3)$$

where $\theta = (\beta', \rho, \lambda_1, \lambda_2)'$, $\Delta u(\theta) = \mathbf{B}_1(\lambda_1)\Delta Y - \mathbf{B}_2(\rho, \lambda_2)\Delta Y_{-1} - \Delta X\beta$, $\mathbf{B}_1(\lambda_1) = I_{T-1} \otimes B_1(\lambda_1)$, and $\mathbf{B}_2(\rho, \lambda_2) = I_{T-1} \otimes B_2(\rho, \lambda_2)$.

Let $\theta_1 = (\beta', \rho, \lambda_2)'$. Given λ_1 and λ_3, (3.3) is maximized at

$$\tilde{\theta}_1(\lambda_1, \lambda_3) = (\Delta\mathbb{X}'\Omega^{-1}\Delta\mathbb{X})^{-1}\Delta\mathbb{X}'\Omega\mathbf{B}_1(\lambda_1)\Delta Y, \quad (3.4)$$

$$\tilde{\sigma}_v^2(\lambda_1, \lambda_3) = \tfrac{1}{n(T-1)}\Delta\tilde{u}'(\lambda_1, \lambda_3)\Omega^{-1}\Delta\tilde{u}(\lambda_1, \lambda_3), \quad (3.5)$$

where $\tilde{u}(\lambda_1, \lambda_3) = \mathbf{B}_1(\lambda_1)\Delta Y - \Delta\mathbb{X}\tilde{\theta}(\lambda_1, \lambda_3)$ and $\Delta\mathbb{X} = (\Delta X, \Delta Y_{-1}, \mathbf{W}_2\Delta Y_{-1})$. Substituting $\tilde{\theta}_1(\lambda_1, \lambda_3)$ and $\tilde{\sigma}_v^2(\lambda_1, \lambda_3)$ back into (3.3) gives the concentrated conditional quasi loglikelihood of (λ_1, λ_3), ignoring the constant term,

$$\ell^c_{\texttt{STLE}}(\lambda_1, \lambda_3) = -\tfrac{n(T-1)}{2}\log[\tilde{\sigma}_v^2(\lambda_1, \lambda_3)] - \tfrac{1}{2}\log|\Omega(\lambda_3)| + \log|\mathbf{B}_1(\lambda_1)|. \quad (3.6)$$

Maximizing $\ell^c_{\texttt{STLE}}(\lambda_1, \lambda_3)$ gives the conditional QML (CQML) estimators $\tilde{\lambda}_1$ and $\tilde{\lambda}_3$ of λ_1 and λ_3, and thus the CQML estimators of θ_1 and σ_v^2 as $\tilde{\theta}_1 \equiv \tilde{\theta}_1(\tilde{\lambda}_1, \tilde{\lambda}_3)$ and $\tilde{\sigma}_v^2 \equiv \tilde{\sigma}_v^2(\tilde{\lambda}_1, \tilde{\lambda}_3)$.

Note that $\ell_{\texttt{STLE}}(\psi)$ is a quasi Gaussian loglikelihood both in the traditional sense that $\{v_{it}\}$ are not exactly Gaussian but Gaussian likelihood is used, and the sense that Δy_1 is not exogenous but is treated as exogenous. The latter causes inconsistency of the CQML estimators when T is small. We see from the results presented below that even T increases with n, the CQML estimators may encounter an asymptotic bias. We now introduce a method that not only gives a consistent estimator of the model parameters when T is small, but also eliminates the asymptotic bias when T is large. To simplify the notation, a parametric quantity (scalar, vector or matrix) evaluated at the general values of the parameters is denoted by dropping its arguments, e.g., $B_1 \equiv B_1(\lambda_1)$, $\mathbf{B}_1 \equiv \mathbf{B}_1(\lambda_1)$, $\Omega \equiv \Omega(\lambda_3)$, and similarly for B_r and $\mathbf{B}_r, r = 2, 3$; and that evaluated at the true values of the parameters is denoted by dropping its argument and then adding a subscript 0, e.g., $B_{10} \equiv B_1(\lambda_{10})$, $\Omega_0 \equiv \Omega(\lambda_{30})$. Let $\mathbf{C} = C \otimes I_n$. Denote $\Delta u \equiv \Delta u(\theta_0)$. The usual expectation, variance and covariance operators, 'E', 'Var' and 'Cov', correspond to the true parameter values.

Let $S_{\texttt{STLE}}(\psi) = \frac{\partial}{\partial\psi}\ell_{\texttt{STLE}}(\psi)$ be the conditional quasi score (CQS) function. We have

$$S_{\texttt{STLE}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta), \\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y - \mathrm{tr}(\mathbf{B}_1^{-1}\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}, \\ \frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1} \otimes A_3)\Delta u(\theta) - (T-1)\mathrm{tr}(G_3), \end{cases} \quad (3.7)$$

where $A_3 = W_3' B_3 + B_3' W_3$ and $G_3 = W_3 B_3^{-1}$.

Under mild conditions, maximizing the conditional loglikelihood $\ell_{\texttt{STLE}}(\psi)$ is equivalent to solving the estimating equation $S_{\texttt{STLE}}(\psi) = 0$. It is well known that the QML type estimation or an extremum type estimation is special case of M-estimation and that for a regular M-estimation problem, a necessary condition for the M-estimators to be consistent is that the probability limit of the estimating function (in this case, the averaged conditional quasi score) at the true parameter value is zero, i.e.,

$$\lim_{n\to\infty} \tfrac{1}{nT} S_{\texttt{STLE}}(\psi_0) \xrightarrow{p} 0,$$

see, e.g., van der Vaart (1998). However, as shown below this is not the case unless T also goes to infinity. Thus, the CQMLEs are not consistent unless $T \to \infty$. Further, even if T goes to infinity with n (proportionally), the CQMLEs encounter a bias of order $O(T^{-1})$, giving the so-called the asymptotic bias.[9] To overcome this major problem, and to avoid the stringent initial conditions and the difficulty in modelling the initial differences under the FE-SDPD models with SL and/or STL effects, we first derive $\mathrm{E}[S_{\texttt{STLE}}(\psi_0)]$, and then adjust the quasi scores $S_{\texttt{STLE}}(\psi)$ so that the adjusted quasi score (AQS) vector, say $S^*_{\texttt{STLE}}(\psi)$, is such that $\mathrm{plim}_{n\to\infty} \frac{1}{nT} S^*_{\texttt{STLE}}(\psi_0) = 0$.

In contrast with Hsiao et al . (2002), Elhorst (2010), and Su and Yang (2015), we only need to have very minimum knowledge about the processes in the past.

Assumption A: *Under Model* (2.1)*, (i) the processes started m periods before the start of data collection, the* 0*th period, and (ii) if $m \geq 1$, Δy_0 is independent of future errors $\{v_t, t \geq 1\}$; if $m = 0$, y_0 is independent of future errors $\{v_t, t \geq 1\}$.*

Assumption A implies that the proposed method does not impose the conditions that $\{y_s, s = -m, \dots, -1\}$ follow the same processes as $\{y_t, t = 0, 1, \dots, T\}$, and $\{x_{it}\}$ be trend-stationary or first-difference stationary. It has a much weaker requirement on the processes starting positions y_m. We have an important lemma based on the reduced form of (3.1):

$$\Delta y_t = \mathcal{B}_0 \Delta y_{t-1} + B_{10}^{-1} \Delta X_t \beta_0 + B_{10}^{-1} B_{30}^{-1} \Delta v_t, \;\; t = 2, \dots, T, \tag{3.8}$$

where $\mathcal{B} \equiv \mathcal{B}(\rho, \lambda_1, \lambda_2) = B_1^{-1}(\lambda_1) B_2(\rho, \lambda_2)$.

Lemma 3.1 *Suppose Assumption A holds. Assume further that, for $i = 1, \dots, n$ and $t = 0, 1, \dots, T$, (i) the idiosyncratic errors $\{v_{it}\}$ in Model* (2.1) *are iid across i and t with mean 0 and variance σ_{v0}^2, (ii) the time-varying regressors X_t are exogenous, and (iii) both B_{10}^{-1} and B_{30}^{-1} exist. We have*

$$\mathrm{E}(\Delta Y_{-1} \Delta v') = -\sigma_{v0}^2 \mathbf{D}_{-10} \mathbf{B}_{30}^{-1}, \tag{3.9}$$
$$\mathrm{E}(\Delta Y \Delta v') = -\sigma_{v0}^2 \mathbf{D}_0 \mathbf{B}_{30}^{-1}, \tag{3.10}$$

[9] To be exact, if $\frac{1}{nT}\mathrm{E}[S_{\texttt{STLE}}(\psi_0)] = O(\frac{1}{T})$, then $\frac{1}{\sqrt{nT}}\mathrm{E}[S_{\texttt{STLE}}(\psi_0)] = O((\frac{n}{T})^{\frac{1}{2}})$, implying $\mathrm{E}[\sqrt{nT}(\tilde{\psi} - \psi_0)] = O((\frac{n}{T})^{\frac{1}{2}})$. The latter says that $\sqrt{nT}(\tilde{\psi} - \psi_0)$ would converge to a non-centered normal if $\frac{n}{T} \to c > 0$. If $\frac{n}{T} \to 0$ (large T case), the asymptotic bias vanishes, but this would not be a case of interest to a spatial panel model.

where $\mathbf{D}_{-1} \equiv \mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)$ *and* $\mathbf{D} \equiv \mathbf{D}(\rho, \lambda_1, \lambda_2)$*, having the following expressions,*

$$\mathbf{D}_{-1} = \begin{pmatrix} I_n, & 0, & \dots & 0, & 0 \\ \mathcal{B} - 2I_n, & I_n, & \dots & 0, & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}^{T-4}(I_n - \mathcal{B})^2, & \mathcal{B}^{T-5}(I_n - \mathcal{B})^2, & \dots & \mathcal{B} - 2I_n, & I_n \end{pmatrix} \mathbf{B}_1^{-1},$$

$$\mathbf{D} = \begin{pmatrix} \mathcal{B} - 2I_n, & I_n, & \dots & 0 \\ (I_n - \mathcal{B})^2, & \mathcal{B} - 2I_n, & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{B}^{T-3}(I_n - \mathcal{B})^2, & \mathcal{B}^{T-4}(I_n - \mathcal{B})^2, & \dots & \mathcal{B} - 2I_n \end{pmatrix} \mathbf{B}_1^{-1}.$$

The results of Lemma 3.1 lead immediately to

$$\mathrm{E}(\Delta u'\Omega_0^{-1}\Delta Y_{-1}) = -\sigma_{v0}^2 \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}), \tag{3.11}$$

$$\mathrm{E}(\Delta u'\Omega_0^{-1}\mathbf{W}_1\Delta Y) = -\sigma_{v0}^2 \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_0\mathbf{W}_1), \tag{3.12}$$

$$\mathrm{E}(\Delta u'\Omega_0^{-1}\mathbf{W}_2\Delta Y_{-1}) = -\sigma_{v0}^2 \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}\mathbf{W}_2), \tag{3.13}$$

showing that the $(\rho, \lambda_1, \lambda_2)$ elements of $\mathrm{E}[S(\psi_0)]$ are not zero, typically of order $O(n)$.[10] Hence, $\mathrm{plim}_{n\to\infty}\frac{1}{nT}\frac{\partial}{\partial\rho}\ell_{\mathtt{STLE}}(\psi_0)$, $\mathrm{plim}_{n\to\infty}\frac{1}{nT}\frac{\partial}{\partial\lambda_1}\ell_{\mathtt{STLE}}(\psi_0)$, and $\mathrm{plim}_{n\to\infty}\frac{1}{nT}\frac{\partial}{\partial\lambda_2}\ell_{\mathtt{STLE}}(\psi_0)$ are all non-zero, suggesting that the conditional QMLEs of $(\rho, \lambda_1, \lambda_2)$, treating Δy_1 as exogenous, cannot be consistent in general.

Very interestingly, these results are derived under only a very minimum set of conditions given in Assumption A and in Lemma 3.1, they are free from the initial conditions usually set for short dynamic panels, and are independent of the way the past observations being generated, i.e., when the processes started and how the process evolved before the start of data collection. These provide a simple way to adjust the quasi scores so as to give a set of unbiased estimating functions. The adjusted quasi score (AQS) functions are:

$$S^*_{\mathtt{STLE}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta), \\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2), \\ \frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes A_3)\Delta u(\theta) - (T-1)\mathrm{tr}(G_3), \end{cases} \tag{3.14}$$

which lead to, as shown in Theorems 3.1 and 3.2, an estimator of ψ that not only is consistent but also has a centered asymptotic distribution, whether T is fixed or grows with n. The latter

[10] It is shown in Sec. 4.1 that when $\lambda_1 = \lambda_2 = 0$, $\mathrm{plim}\frac{1}{nT}\frac{\partial}{\partial\rho}\ell_{\mathtt{STLE}}(\psi_0) = \frac{1-\rho_0^T}{T^2(1-\rho_0)^2} - \frac{1}{T(1-\rho_0)}$. Thus, $\tilde{\rho}$ has a bias of order $O(\frac{1}{T})$. As all the matrices involved in (3.11)-(3.13) are uniformly bounded in row and column sums, this result would hold for the general model, and the bias in $\tilde{\rho}$ would spill over to the other CQMLEs.

implies that when T grows with n, the estimation based on the AQS functions eliminates the asymptotic bias incurred in the conditional QML approach. The 'adjustments' in the AQS functions have another interesting feature: they are independent of the SE structure, i.e., free from B_3. This means that the adjustments to the conditional quasi scores of (3.7) remain the same if the SAR error is replaced by SMA error, or the spatial errors are of higher order.[11] Comparing (3.14) with (3.7), we see that after the adjustments, both ρ and λ_2 parameters become non-linear in the sense that their estimation has to be done through a nonlinear root-finding process. Evidently, the adjustments recovered the 'neglected' information contained in the initial observations (by the conditional likelihood) about these parameters.

Solving $S^*_{\mathtt{STLE}}(\psi) = 0$ leads to the M-estimator $\hat{\psi}_{\mathtt{M}}$ of ψ. This root-finding process can be simplified by first solving the equations for β and σ_v^2, given $\delta = (\rho, \lambda')'$, resulting in the constrained M-estimators of β and σ_v^2 as

$$\begin{aligned}
\hat{\beta}(\delta) &= (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1}), && (3.15)\\
\hat{\sigma}_v^2(\delta) &= \tfrac{1}{n(T-1)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta), && (3.16)
\end{aligned}$$

where $\Delta\hat{u}(\delta) = \Delta u(\hat{\beta}(\delta), \rho, \lambda_1, \lambda_2)$. Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ back into the last four components of the AQS function in (3.14) gives the concentrated AQS functions:

$$S^{*c}_{\mathtt{STLE}}(\delta) = \begin{cases}
\frac{1}{\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}),\\
\frac{1}{\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1),\\
\frac{1}{\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2),\\
\frac{1}{2\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{u}(\delta)'(C^{-1}\otimes A_3)\Delta\hat{u}(\delta) - (T-1)\mathrm{tr}(G_3).
\end{cases} \quad (3.17)$$

Solving the resulted concentrated estimating equations, $S^{*c}_{\mathtt{STLE}}(\delta) = 0$, we obtain the unconstrained M-estimators $\hat{\delta}_{\mathtt{M}}$ of δ. The unconstrained M-estimators of β and σ_v^2 are thus $\hat{\beta}_{\mathtt{M}} \equiv \hat{\beta}(\hat{\delta}_{\mathtt{M}})$ and $\hat{\sigma}^2_{v,\mathtt{M}} \equiv \hat{\sigma}^2_v(\hat{\delta}_{\mathtt{M}})$. Denote $\hat{\psi}_{\mathtt{M}} = (\hat{\beta}'_{\mathtt{M}}, \hat{\sigma}^2_{v,\mathtt{M}}, \hat{\rho}_{\mathtt{M}}, \hat{\lambda}'_{\mathtt{M}})'$.

We end this subsection by noting that Model (3.1) and its M-estimation strategy can even be further extended to allow for higher-order spatial lags:

$$\begin{aligned}
\Delta y_t &= \rho\Delta y_{t-1} + \textstyle\sum_{j=1}^{k_1}\lambda_{1j}W_{1j}\Delta y_t + \sum_{j=1}^{k_2}\lambda_{2j}W_{2j}\Delta y_{t-1} + \Delta X_t\beta + \Delta u_t,\\
\Delta u_t &= \textstyle\sum_{j=1}^{k_3}\lambda_{3j}W_{3j}\Delta u_t + \Delta,
\end{aligned}$$

for $t = 2, 3, \cdots, T$. Now, let $\lambda_r = (\lambda_{rj}, j = 1, \ldots, k_1)'$, $r = 1, 2, 3$; $B_r(\lambda_r) = I_n - \sum_{j=1}^{k_r}\lambda_{rj}W_{rj}$, $r = 1, 3$, and $\mathcal{B}(\rho, \lambda_1, \lambda_2) = B_1^{-1}(\rho I_n + \sum_{j=1}^{k_2}\lambda_{2j}W_{2j})$. The M-estimation proceeds in the same manner. In this paper, we focus on Model (3.1) as it is general enough for most of the empirical applications and it leads to a set of inference theories that are fairly simple and yet can be extended to suit a larger model in a straightforward manner.

[11] This feature may also hold if the SE structure is replaced by the other forms of cross-section dependence induced by common factors; see, e.g., Andrews (2005), Pesaran (2006), Bai (2009), Pesaran and Tosetti (2011).

3.2 Asymptotic properties of the M-estimators

In this section we study the consistency and asymptotic normality of the proposed M-estimators for the FE-SDPD model with the general spatial dependence structure. To facilitate the discussions of the asymptotic properties of the proposed M-estimators, first recall, ψ_0 denotes the true value of the parameter vector ψ; and a parametric function at the true parameter value is differentiated from that at a general parameter value by adding a subscript '0', e.g., $B_1 \equiv B_1(\lambda_1)$ and $B_{10} \equiv B_1(\lambda_{10})$, $\Omega \equiv \Omega(\lambda_3)$ and $\Omega_0 \equiv \Omega(\lambda_{30})$, etc.; $\Delta u \equiv \Delta u(\theta_0)$; and the common expectation, variance and covariance operators 'E' 'Var' and 'Cov' correspond to the true parameter vector ψ_0. Second, some general notation and convention are as follows: (*i*) δ denotes the vector of parameters in the concentrated modified score function, and $\mathbf{\Delta}$ the space from which δ takes values; (*ii*) $\text{tr}(\cdot)$, $|\cdot|$ and $\|\cdot\|$ denote, respectively, the trace, determinant, and Frobenius norm of a matrix; (*iii*) $\gamma_{\max}(A)$ and $\gamma_{\min}(A)$ denote, respectively, the largest and smallest eigenvalues of a real symmetric matrix A; and (*iv*) $\text{diag}(a_k)$ forms a diagonal matrix using the elements $\{a_k\}$ and $\text{blkdiag}(A_k)$ forms a block-diagonal matrix using the matrices $\{A_k\}$.

Assumption B: *The innovations v_{it} are iid for all i and t with $E(v_{it}) = 0$, $Var(v_{it}) = \sigma_v^2$, and $E|v_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.*

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

Assumption D: *The time-varying regressors $\{X_t,\ t = 0, 1, \ldots, T\}$ are exogenous, their values are uniformly bounded, and $\lim_{n\to\infty} \frac{1}{nT}\Delta X'\Delta X$ exists and is nonsingular.*

Assumption E: *(i) For $r = 1, 2, 3$, the elements $w_{r,ij}$ of W_r are at most of order h_n^{-1}, uniformly in all i and j, and $w_{r,ii} = 0$ for all i; (ii) $h_n/n \to 0$ as $n \to \infty$; (iii) $\{W_r, r = 1, 2, 3\}$ and $\{B_{r0}^{-1}, r = 1, 3\}$ are uniformly bounded in both row and column sums; (iv) For $r = 1, 3$, $\{B_r^{-1}\}$ are uniformly bounded in either row or column sums, uniformly in λ_r in a compact parameter space $\mathbf{\Lambda}_r$, and $0 < \underline{c}_r \leq \inf_{\lambda_r\in\mathbf{\Lambda}_r} \gamma_{\min}(B_r'B_r) \leq \sup_{\lambda_r\in\mathbf{\Lambda}_r} \gamma_{\max}(B_r'B_r) \leq \bar{c}_r < \infty$.*

Assumption F: *For an $n\times n$ matrix Φ uniformly bounded in either row or column sums, with elements of uniform order h_n^{-1}, and an $n \times 1$ vector ϕ with elements of uniform order $h_n^{-1/2}$, (i) $\frac{h_n}{n}\Delta y_1'\Phi\Delta y_1 = O_p(1)$ and $\frac{h_n}{n}\Delta y_1'\Phi\Delta v_2 = O_p(1)$; (ii) $\frac{h_n}{n}(\Delta y_1 - \text{E}(\Delta y_1))'\phi = o_p(1)$; (iii) $\frac{h_n}{n}[\Delta y_1'\Phi\Delta y_1 - \text{E}(\Delta y_1'\Phi\Delta y_1)] = o_p(1)$, and (iv) $\frac{h_n}{n}[\Delta y_1'\Phi\Delta v_2 - \text{E}(\Delta y_1'\Phi\Delta v_2)] = o_p(1)$.*

Assumptions B-E are standard in the spatial econometrics literature (see, e.g., Lee (2004a), Lee and Yu (2010), Su and Yang (2015)). Assumption F imposes some fairly mild conditions on the initial differences Δy_1. These conditions clearly hold if the process starting position y_{-m} is exogenous with $m = 0$ or 1. They can also be proved for a general m if, in addition to Assumptions A, B, and D, it is further assumed that the processes start at exogenous positions y_{-m} at time $-m$, and then evolves according to (2.6).

Now, the consistency of the proposed M-estimators $\hat{\psi}_{\texttt{M}}$ lies with the consistency of $\hat{\delta}_{\texttt{M}}$, as under Assumptions D and E, the consistency of $\hat{\beta}_{\texttt{M}}$ and $\hat{\sigma}^2_{v,\texttt{M}}$ follows almost immediately that of $\hat{\delta}_{\texttt{M}}$. The concentrated estimating function (CEF) $S^{*c}_{\texttt{STLE}}(\delta)$ and its population counter part play a major role for the consistency of $\hat{\delta}_{\texttt{M}}$ for δ. We present theories and proofs for the most general model, and then discuss in Section 4 how the general theories reduce to a given submodel and how the assumptions are simplified.

Define $\bar{S}^*_{\mathtt{STLE}}(\psi) = \mathrm{E}[S^*_{\mathtt{STLE}}(\psi)]$, the population counter part of the joint estimating function JEF given in (3.14). Given δ, the population joint estimation equation (JEE) $\bar{S}^*_{\mathtt{STLE}}(\psi) = 0$ is partially solved, by working with its first two components, at

$$
\begin{aligned}
\bar{\beta}(\delta) &= (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}(\mathbf{B}_1\mathrm{E}\Delta Y - \mathbf{B}_2\mathrm{E}\Delta Y_{-1}), \qquad (3.18)\\
\bar{\sigma}^2_v(\delta) &= \tfrac{1}{n(T-1)}\mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\Delta\bar{u}(\delta)], \qquad (3.19)
\end{aligned}
$$

where $\Delta\bar{u}(\delta) = \Delta u(\theta)|_{\beta=\bar{\beta}(\delta)} = \mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1} - \Delta X\bar{\beta}(\delta)$. These lead to the population counter part of the CEF given in (3.17), upon substituting $\bar{\beta}(\delta)$ and $\bar{\sigma}^2_v(\delta)$ back into the δ-component of $\bar{S}^*_{\mathtt{STLE}}(\psi)$, as

$$
\bar{S}^{*c}_{\mathtt{STLE}}(\delta) = \begin{cases}
\frac{1}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\Delta Y_{-1}] + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}),\\
\frac{1}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y] + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1),\\
\frac{1}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y_{-1}] + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_1),\\
\frac{1}{2\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathrm{E}[\Delta\bar{u}(\delta)'(C^{-1}\otimes A_3)\Delta\bar{u}(\delta)] - (T-1)\mathrm{tr}(G_3).
\end{cases} \qquad (3.20)
$$

Note that more detailed expressions for $\bar{\sigma}^2_v(\delta)$ and thus $\bar{S}^{*c}_{\mathtt{STLE}}(\delta)$ can be obtained through the following very useful identity:

$$
\Delta\bar{u}^*(\delta) = \mathbf{M}(\mathbf{B}^*_1\Delta Y - \mathbf{B}^*_2\Delta Y_{-1}) + \mathbf{P}(\mathbf{B}^*_1\Delta Y^\circ - \mathbf{B}^*_2\Delta Y^\circ_{-1}), \qquad (3.21)
$$

where $\Delta\bar{u}^*(\delta) = \Omega^{-\frac{1}{2}}\Delta\bar{u}(\delta)$, $\mathbf{B}^*_r = \Omega^{-\frac{1}{2}}\mathbf{B}_r$, $\Delta Y^\circ = \Delta Y - \mathrm{E}(\Delta Y)$, $\Delta Y^\circ_{-1} = \Delta Y_{-1} - \mathrm{E}(\Delta Y_{-1})$, $\Omega^{\frac{1}{2}}$ is the square-root matrix of Ω, $\mathbf{M} = I_{n(T-1)} - \Omega^{-\frac{1}{2}}\Delta X(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-\frac{1}{2}}$, and $\mathbf{P} = I_{n(T-1)} - \mathbf{M}$. Also note that the quantities $\mathrm{E}(\Delta Y)$, $\mathrm{E}(\Delta Y_{-1})$, $\mathrm{Var}(\Delta Y)$, $\mathrm{Cov}(\Delta Y, \Delta Y_{-1})$, etc., involved in (3.18)-(3.20) are functions of ψ_0, but not ψ.

Clearly, the M-estimator $\hat{\delta}_{\mathtt{M}}$ of δ_0 is a zero of $S^{*c}_{\mathtt{STLE}}(\delta)$. It is easy to see that $\bar{S}^{*c}_{\mathtt{STLE}}(\delta_0) = 0$ trough $\bar{\beta}(\delta_0) = \beta_0$ and $\bar{\sigma}^2_v(\delta_0) = \sigma^2_{v0}$, i.e., δ_0 is a zero of $\bar{S}^{*c}_{\mathtt{STLE}}(\delta)$. Thus, by Theorem 5.9 of van der Vaart (1998), $\hat{\delta}_{\mathtt{M}}$ will be consistent for δ_0 if $\sup_{\delta\in\Delta}\frac{1}{\sqrt{n(T-1)}}\big\|S^{*c}_{\mathtt{STLE}}(\delta) - \bar{S}^{*c}_{\mathtt{STLE}}(\delta)\big\| \xrightarrow{p} 0$, and the following identification condition holds.

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

For a simpler model (see Section 4), the corresponding expression for $\bar{S}^{*c}_{\mathtt{M}}(\delta)$ can be easily obtained by dropping the relevant terms. The identification condition becomes simpler, allowing us to gain insights on the nature of such a condition. See Lee and Yu (2015b) and references therein for a detailed discussion on the identification of the SDPD models.

Theorem 3.1 *Suppose Assumptions A-G hold. Assume further that* (i) $\gamma_{\max}[\mathrm{Var}(\Delta Y)]$ *and* $\gamma_{\max}[\mathrm{Var}(\Delta Y_{-1})]$ *are bounded, and* (ii) $\inf_{\delta\in\Delta}\gamma_{\min}\big(\mathrm{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})\big) \geq \underline{c}_y > 0$.

We have, as $n \to \infty$, $\hat{\psi}_{\mathtt{M}} \xrightarrow{p} \psi_0$.

To derive the asymptotic distribution of $\hat{\psi}_{\mathtt{M}}$, we start with a Taylor expansion of the JEE $S^*_{\mathtt{STLE}}(\hat{\psi}_{\mathtt{M}}) = 0$ at ψ_0, and then verify that the AQS function $S^*_{\mathtt{STLE}}(\psi_0)$ is asymptotically normal and that the adjusted Hessian $\frac{\partial}{\partial \psi'} S^*_{\mathtt{STLE}}(\bar{\psi})$ has proper asymptotic behavior, for some $\bar{\psi}$ lying between $\hat{\psi}_{\mathtt{M}}$ and ψ_0 elementwise. To most of the static models or dynamic models where the initial conditions are specified, both problems are fairly standard in that the regular law of large numbers and central limit theorem (CLT) for linear-quadratic forms (e.g., Kelejian and Prucha, 2001) would be sufficient. In our approach, the initial conditions need not be specified, and thus an extended CLT for bilinear-quadratic form (given in Lemma A.5) is required for establishing the asymptotic normality of $S^*_{\mathtt{STLE}}(\psi_0)$. The following representations for ΔY and ΔY_{-1} in terms of $\Delta \mathbf{y}_1 = 1_{T-1} \otimes \Delta y_1$ and Δv are crucial.

Lemma 3.2 *Under the assumptions of Lemma 3.1, we have,*

$$\Delta Y = \mathbb{R}\, \Delta \mathbf{y}_1 + \boldsymbol{\eta} + \mathbb{S} \Delta v, \tag{3.22}$$

$$\Delta Y_{-1} = \mathbb{R}_{-1} \Delta \mathbf{y}_1 + \boldsymbol{\eta}_{-1} + \mathbb{S}_{-1} \Delta v, \tag{3.23}$$

where $\mathbb{R} = \text{blkdiag}(\mathcal{B}_0, \mathcal{B}_0^2, \ldots, \mathcal{B}_0^{T-1})$, $\mathbb{R}_{-1} = \text{blkdiag}(I_n, \mathcal{B}_0, \ldots, \mathcal{B}_0^{T-2})$, $\boldsymbol{\eta} = \mathbb{B}\mathbf{B}_{10}^{-1} \Delta X \beta_0$, $\boldsymbol{\eta}_{-1} = \mathbb{B}_{-1} \mathbf{B}_{10}^{-1} \Delta X \beta_0$, $\mathbb{S} = \mathbb{B}\mathbf{B}_{10}^{-1}\mathbf{B}_{30}^{-1}$, $\mathbb{S}_{-1} = \mathbb{B}_{-1}\mathbf{B}_{10}^{-1}\mathbf{B}_{30}^{-1}$,

$$\mathbb{B} = \begin{pmatrix} I_n & 0 & \ldots & 0 & 0 \\ \mathcal{B}_0 & I_n & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}_0^{T-2} & \mathcal{B}_0^{T-3} & \ldots & \mathcal{B}_0 & I_n \end{pmatrix}, \quad \textit{and} \quad \mathbb{B}_{-1} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ I_n & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}_0^{T-3} & \mathcal{B}_0^{T-4} & \ldots & I_n & 0 \end{pmatrix}.$$

The representations for ΔY and ΔY_{-1} given in Lemma 3.2 turn out to be very useful. They lead to a simple way for establishing the asymptotic normality of the AQS vector, and a simple way for estimating the variance-covariance (VC) matrix of it. Using these representations and $\Delta u = \mathbf{B}_{30}^{-1} \Delta v$, the AQS function at ψ_0 can be written as

$$S^*_{\mathtt{STLE}}(\psi_0) = \begin{cases} \Delta v' \Pi_1, \\ \Delta v' \Phi_1 \Delta v - \frac{n(T-1)}{2\sigma_{v0}^2}, \\ \Delta v' \Psi_1 \Delta \mathbf{y}_1 + \Delta v' \Pi_2 + \Delta v' \Phi_2 \Delta v + \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}), \\ \Delta v' \Psi_2 \Delta \mathbf{y}_1 + \Delta v' \Pi_3 + \Delta v' \Phi_3 \Delta v + \text{tr}(\mathbf{C}^{-1}\mathbf{D}_0 \mathbf{W}_1), \\ \Delta v' \Psi_3 \Delta \mathbf{y}_1 + \Delta v' \Pi_4 + \Delta v' \Phi_4 \Delta v + \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10} \mathbf{W}_2), \\ \Delta v' \Phi_5 \Delta v - (T-1)\text{tr}(G_{30}), \end{cases} \tag{3.24}$$

where $\Pi_1 = \frac{1}{\sigma_{v0}^2} \mathbb{C}_b \Delta X$, $\Pi_2 = \frac{1}{\sigma_{v0}^2} \mathbb{C}_b \boldsymbol{\eta}_{-1}$, $\Pi_3 = \frac{1}{\sigma_{v0}^2} \mathbb{C}_b \mathbf{W}_1 \boldsymbol{\eta}$, $\Pi_4 = \frac{1}{\sigma_{v0}^2} \mathbb{C}_b \mathbf{W}_2 \boldsymbol{\eta}_{-1}$, $\Phi_1 = \frac{1}{2\sigma_{v0}^4} (C^{-1} \otimes I_n)$, $\Phi_2 = \frac{1}{\sigma_{v0}^2} \mathbb{C}_b \mathbb{S}_{-1}$, $\Phi_3 = \frac{1}{\sigma_{v0}^2} \mathbb{C}_b \mathbf{W}_1 \mathbb{S}$, $\Phi_4 = \frac{1}{\sigma_{v0}^2} \mathbb{C}_b \mathbf{W}_2 \mathbb{S}_{-1}$, $\Phi_5 = \frac{1}{\sigma_{v0}^2} [C^{-1} \otimes (G_{30}' + G_{30})]$, $\Psi_1 = \frac{1}{\sigma_{v0}^2} \mathbb{C}_b \mathbb{R}_{-1}$, $\Psi_2 = \frac{1}{\sigma_{v0}^2} \mathbb{C}_b \mathbf{W}_1 \mathbb{R}$, $\Psi_3 = \frac{1}{\sigma_{v0}^2} \mathbb{C}_b \mathbf{W}_2 \mathbb{R}_{-1}$, and $\mathbb{C}_b = C^{-1} \otimes B_{30}$.

Theorem 3.2 *Assume Assumptions A-F hold. We have, as $n \to \infty$,*

$$\sqrt{n(T-1)}\big(\hat{\psi}_{\mathtt{M}} - \psi_0\big) \xrightarrow{D} n\big[0, \lim_{n\to\infty} \Sigma^{*-1}_{\mathtt{STLE}}(\psi_0)\Gamma^{*}_{\mathtt{STLE}}(\psi_0)\Sigma^{*-1}_{\mathtt{STLE}}(\psi_0)\big],$$

where $\Sigma^{*}_{\mathtt{STLE}}(\psi_0) = -\frac{1}{n(T-1)}\mathrm{E}[\frac{\partial}{\partial\psi'}S^{*}_{\mathtt{STLE}}(\psi_0)]$ *and* $\Gamma^{*}_{\mathtt{STLE}}(\psi_0) = \frac{1}{n(T-1)}\mathrm{Var}[S^{*}_{\mathtt{STLE}}(\psi_0)]$, *both assumed to exist and* $\Sigma^{*}_{\mathtt{STLE}}(\psi_0)$ *to be positive definite, for sufficiently large n.*

3.3 The OPMD estimation of robust VC matrix

The practical applications of the M-estimation of the FE-SDPD models, i.e., model inferences, depend on the availability of a consistent estimate of $\Sigma^{*-1}_{\mathtt{STLE}}(\psi_0)\Gamma^{*}_{\mathtt{STLE}}(\psi_0)\Sigma^{*-1}_{\mathtt{STLE}}(\psi_0)$, the VC matrix of $\hat{\psi}_{\mathtt{M}}$. As $\Sigma^{*}_{\mathtt{STLE}}(\psi_0)$ is the expected negative modified Hessian, its observed counter part immediately offers a consistent estimate of it, i.e.,

$$\Sigma^{*}_{\mathtt{STLE}}(\hat{\psi}_{\mathtt{M}}) = -\frac{1}{n(T-1)}\frac{\partial}{\partial\psi'}S^{*}_{\mathtt{STLE}}(\psi)\big|_{\psi=\hat{\psi}_{\mathtt{M}}}. \tag{3.25}$$

The detailed expression of $\frac{\partial}{\partial\psi'}S^{*}_{\mathtt{STLE}}(\psi)$ for the most general model is given in the proof of Theorem 3.2 in Appendix C, which can easily be simplified to various submodels by deleting relevant terms. The consistency of $\Sigma^{*}_{\mathtt{STLE}}(\hat{\psi}_{\mathtt{M}})$ is also proved in the proof of Theorem 3.2.

However, the traditional plug-in method of estimation of $\Gamma^{*}_{\mathtt{STLE}}(\psi_0)$, the VC matrix of the joint AQS function $S^{*}_{\mathtt{STLE}}(\psi_0)$, runs into a similar problems as the QML estimation of the model – initial differences Δy_1 need to be specified or modeled when T is fixed and small as seen from (3.24). To make the estimation of $\Gamma^{*}_{\mathtt{STLE}}(\psi_0)$ also free from the specification of initial conditions so that the inferences for the general FE-SDPD model is fully free from the initial conditions. We propose a martingale difference (M.D.) method, i.e., decompose the AQS function into the sum of a vector M.D. sequence so that the 'average' of the outer products of the elements of the M.D. sequence gives a consistent estimate of $\Gamma^{*}_{\mathtt{STLE}}(\psi_0)$.

From (3.24) we see that the AQS function $S^{*}_{\mathtt{STLE}}(\psi_0)$ contains three types of elements:

$$\Pi'\Delta v, \quad \Delta v'\Phi\Delta v, \quad \text{and } \Delta v'\Psi\Delta\mathbf{y}_1,$$

where Π, Φ and Ψ are nonstochastic matrices (depending on ψ_0) with Π being $n(T-1)\times p$ or $n(T-1)\times 1$, and Φ and Ψ being $n(T-1)\times n(T-1)$. Clearly, the closed form expressions for variances of $\Pi'\Delta v$ and $\Delta v'\Phi\Delta v$, and their covariance can readily be derived, so that a plug-in method may be used to estimate these variances and covariances. However, closed-form expression for the variance of $\Delta v'\Psi\Delta\mathbf{y}_1$ and its covariances with $\Pi'\Delta v$ and $\Delta v'\Phi\Delta v$ depend on the knowledge of the distribution of Δy_1, which is unavailable.

To give a unified method of estimating the variance of AQS function so that it is also free from the specifications of the initial conditions, we first write the AQS function as a sum of a vector M.D. array, taking the advantage that Δv_{it} are independent across i for each t and that T is small. We show in the following lemma that $\Phi'\Delta v$ can be written as a sum of n independent terms, and $\Delta v'\Phi\Delta v - \mathrm{E}(\Delta v'\Phi_{\Delta}v)$ can be written as the sum of a

M.D. array. Under the assumption that Δy_0 depends only on the current and past errors $(v_s, s \leq 0)$ but not the future errors $(v_t, t \geq 1)$, we show that $\Delta v'\Phi\Delta \mathbf{y}_1$ can also be written as the sum of a M.D. sequence. Note from (3.24) that $S^*_{\texttt{STLE}}(\psi_0)$ is a function of ψ_0, Δv and Δy_1, where ψ_0 is consistently estimated by $\hat{\psi}_{\texttt{M}}$, Δv is consistently estimated by $\widehat{\Delta} v$, and Δy_1 itself is observed. A new method for estimating the variance of the AQS function, namely the *outer-product-of-martingale-difference* (OPMD) method, arises.

For a square matrix A, let A^u, A^l and A^d be, respectively, its upper-triangular, lower-triangular, and diagonal matrix such that $A = A^u + A^l + A^d$. Denote by Π_t, Φ_{ts} and Ψ_{ts} the submatrices of Π, Φ and Ψ partitioned according to $t, s = 2, \ldots, T$. Define $\Psi_{t+} = \sum_{s=2}^{T} \Psi_{ts}$, $t = 2, \ldots, T$, $\Theta = \Psi_{2+}(B_{30}B_{10})^{-1}$, $\Delta y_1^{\circ} = B_{30}B_{10}\Delta y_1$, and $\Delta y_{1t}^* = \Psi_{t+}\Delta y_1$. Let $\{\mathcal{G}_{n,i}\}$ be the increasing sequence of σ-fields generated by $(v_{j1}, \ldots, v_{jT}, j = 1, \ldots, i), i = 1, \ldots, n$, $n \geq 1$. Let $\mathcal{F}_{n,0}$ be the σ-field generated by $(v_0, \Delta y_0)$, and define $\mathcal{F}_{n,i} = \mathcal{F}_{n,0} \otimes \mathcal{G}_{n,i}$. Clearly, $\mathcal{F}_{n,i-1} \subseteq \mathcal{F}_{n,i}$, i.e., $\{\mathcal{F}_{n,i}\}_{i=1}^n$ is an increasing sequence of σ-fields, for each $n \geq 1$.

Lemma 3.3 *Consider Model* (3.1)*, and suppose the assumptions of Lemma 3.1 hold. Consider the general* Π *which is* $n(T-1) \times p$*, and let* Π_{it} *be the ith row of* Π_t*. Define*

$$\begin{aligned}
g_{1i} &= \textstyle\sum_{t=2}^{T} \Pi'_{it}\Delta v_{it}, && (3.26)\\
g_{2i} &= \textstyle\sum_{t=2}^{T} (\Delta v_{it}\Delta\xi_{it} + \Delta v_{it}\Delta v_{it}^* - \sigma_{v0}^2 d_{it}), && (3.27)\\
g_{3i} &= \Delta v_{2i}\Delta\zeta_i + \Theta_{ii}(\Delta v_{2i}\Delta y_{1i}^{\circ} + \sigma_{v0}^2) + \textstyle\sum_{t=3}^{T} \Delta v_{it}\Delta y_{1it}^*, && (3.28)
\end{aligned}$$

where for (3.27)*,* $\xi_t = \sum_{s=2}^{T}(\Phi_{st}^{u\prime} + \Phi_{ts}^l)\Delta v_s$*,* $\Delta v_t^* = \sum_{s=2}^{T} \Phi_{ts}^d \Delta v_s$*, and* $\{d_{it}\}$ *are the diagonal elements of* $\mathbf{C}\Phi$*; for* (3.28)*,* $\{\Delta\zeta_i\} = \Delta\zeta = (\Theta^u + \Theta^l)\Delta y_1^{\circ}$*, and* $\text{diag}\{\Theta_{ii}\} = \Theta^d$*. Then,*

$$\begin{aligned}
\Pi'\Delta v &= \textstyle\sum_{i=1}^n g_{1i},\\
\Delta v'\Phi\Delta v - \mathrm{E}(\Delta v'\Phi\Delta v) &= \textstyle\sum_{i=1}^n g_{2i},\\
\Delta v'\Psi\Delta \mathbf{y}_1 - \mathrm{E}(\Delta v'\Psi\Delta \mathbf{y}_1) &= \textstyle\sum_{i=1}^n g_{3i},
\end{aligned}$$

and $\{(g'_{1i},\ g_{2i},\ g_{3i})', \mathcal{F}_{n,i}\}_{i=1}^n$ *form a vector M.D. sequence.*

Now, following Lemma 3.3, for each $\Pi_r, r = 1, 2, 3, 4$, defined in (3.24), define g_{1ri} according to (3.26); for each $\Phi_r, r = 1, \ldots, 5$, define g_{2ri} according to (3.27); and for each $\Psi_r, r = 1, 2, 3$, define g_{3ri} according to (3.28). Define

$$g_i = (g'_{11i},\ g_{21i},\ g_{31i} + g_{12i} + g_{22i},\ g_{32i} + g_{13i} + g_{23i},\ g_{33i} + g_{14i} + g_{24i},\ g_{25i})'.$$

Then, $S^*_{\texttt{SELE}}(\psi_0) = \sum_{i=1}^n g_i$, and $\{g_i, \mathcal{F}_{n,i}\}$ form a vector M.D. sequence. It follows that $\mathrm{Var}[S^*_{\texttt{SELE}}(\psi_0)] = \sum_{i=1}^n \mathrm{E}(g_i g'_i)$. The 'average' of the outer products of the estimated $g'_i s$, i.e.,

$$\widehat{\Gamma}^*_{\texttt{STLE}} = \tfrac{1}{n(T-1)} \textstyle\sum_{i=1}^n \hat{g}_i\hat{g}'_i, \tag{3.29}$$

thus gives a consistent estimator of the variance of $\Gamma^*_{\texttt{STLE}}(\psi_0)$, where $\hat{g}_i$ is obtained by replacing ψ_0 in g_i by $\hat{\psi}_{\texttt{M}}$ and Δv in g_i by its observed counterpart $\widehat{\Delta} v$. Noting that Δy_1 is observed, we have the following theorem.

Theorem 3.3 *Under the assumptions of Theorem* (3.1)*, we have, as* $n \to \infty$*,*

$$\widehat{\Gamma}^*_{\mathtt{STLE}} - \Gamma^*_{\mathtt{STLE}}(\psi_0) = \frac{1}{n(T-1)} \sum_{i=1}^{n} \left[\hat{g}_i \hat{g}_i' - \mathrm{E}(g_i g_i')\right] \xrightarrow{p} 0,$$

and hence, $\Sigma^{*-1}_{\mathtt{STLE}}(\hat{\psi}_{\mathtt{M}})\widehat{\Gamma}^*_{\mathtt{STLE}}\Sigma^{*-1}_{\mathtt{STLE}}(\hat{\psi}_{\mathtt{M}}) - \Sigma^{*-1}_{\mathtt{STLE}}(\psi_0)\Gamma^*_{\mathtt{STLE}}(\psi_0)\Sigma^{*-1}_{\mathtt{STLE}}(\psi_0) \xrightarrow{p} 0.$

The estimator $\Sigma^{*-1}_{\mathtt{STLE}}(\hat{\psi}_{\mathtt{M}})\widehat{\Gamma}^*_{\mathtt{STLE}}\Sigma^{*-1}_{\mathtt{STLE}}(\hat{\psi}_{\mathtt{M}})$ of the VC matrix of $\hat{\psi}_{\mathtt{M}}$ is referred to as the OPMD estimator in this paper, to reflect the fact that $\widehat{\Gamma}^*_{\mathtt{STLE}}$ is obtained from the **o**uter **p**roducts of the elements of a **m**artingale **d**ifference (OPMD) sequence.[12]

4 Unified Estimation and Inference for Some Submodels

Certain submodels deserve some special attention. We concentrate on the submodels that contain spatial dependence, namely, the FE-DPD model with only SE dependence, the FE-DPD model with only SL dependence, the FE-DPD model with both SL and STL dependence, and the FE-DPD model with both SL and SE dependence. We are particularly interested in comparing our approach with the standard small T or large T approaches, to demonstrate that when T is small our approach provides results that are comparable with the standard full QML approach when the initial model is correctly specified. However, our approach provides results that are more robust against misspecification of the initial model than does the full QML approach. When T is large, our approach provides results that are less biased compared with the conditional QML approach.

4.1 The FE-SDPD model with SE effect

Setting $\lambda_1 = \lambda_2 = 0$, Model (3.1) reduces to an FE-SDPD with only SE dependence of a SAR form, which has been rigorously treated in Su and Yang (2015) based on a full QML approach where the initial differences are modeled. It would be certainly interesting to see how the proposed approach compares with this full QML approach. The conditional quasi Gaussian loglikelihood (3.3) simplifies to:

$$\ell_{\mathtt{SE}}(\psi) = -\tfrac{n(T-1)}{2}\log(\sigma_v^2) - \tfrac{1}{2}\log|\Omega| - \tfrac{1}{2\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta), \qquad (4.1)$$

where $\psi = \{\beta', \sigma_v^2, \rho, \lambda_3\}'$ and $\theta = (\beta', \rho)'$ and $u(\theta) = \Delta Y - \rho\Delta Y_{-1} - \Delta X\beta$. Given λ_3, $\ell_{\mathtt{SE}}(\psi)$ is maximized at $\tilde{\theta}(\lambda_3) = (\Delta\mathbb{X}'\Omega^{-1}\Delta\mathbb{X})^{-1}\Delta\mathbb{X}'\Omega\Delta Y$ and $\tilde{\sigma}_v^2(\lambda_3) = \frac{1}{n(T-1)}\Delta\tilde{u}'(\lambda_3)\Omega^{-1}\Delta\tilde{u}(\lambda_3)$,

[12]Practical implementations of the OPMD estimator of the VC matrix of the AQS vector can be greatly facilitated by the vector and matrix representation of the quantities defined in Lemma 3.3. For example, to compute $g_1 = \{g_{1i}\}$, let π_k be the kth column of Π. Reshape π_k and Δv into $n \times (T-1)$ matrices $\boldsymbol{\pi}_k$ and $\Delta\mathbf{v}$. Then g_1 equals the vector of row sums of $\boldsymbol{\pi}_k \odot \Delta\mathbf{v}$, where $\odot$ denotes the Hadamard product.

Moreover, the partial derivatives of $\mathbf{D}$ and $\mathbf{D}_{-1}$ defined in Lemma 3.1 are needed for the evaluation of the Hessian matrix, which can be algebraically tedious if T is not so small. In this case, these partial derivatives can be replaced by the numerical partial derivatives without losing much of the accuracy of the OPMD estimator.

where $\Delta\tilde{u}(\lambda_3) = \Delta Y - \Delta\mathbb{X}\,\tilde{\theta}(\lambda_3)$, and $\Delta\mathbb{X} = (\Delta X, \Delta Y_{-1})$. Substituting $\tilde{\theta}(\lambda_3)$ and $\tilde{\sigma}_v^2(\lambda_3)$ back into $\ell_{\texttt{SE}}(\psi)$ gives the concentrated quasi loglikelihood function of λ_3,

$$\ell_{\texttt{SE}}^c(\lambda_3) = -\tfrac{n(T-1)}{2}\log(\tilde{\sigma}_v^2(\lambda_3)) - \tfrac{1}{2}\log|\Omega|. \tag{4.2}$$

Maximizing $\ell_{\texttt{SE}}^c(\lambda_3)$ gives the CQMLE $\tilde{\lambda}_3$ of λ_3, and thus the CQMLEs $\tilde{\theta} \equiv \tilde{\theta}(\tilde{\lambda}_3)$ and $\tilde{\sigma}_v^2 \equiv \tilde{\sigma}_v^2(\tilde{\lambda}_3)$ of β and σ_v^2, respectively.

Now, $S_{\texttt{SE}}(\psi) = \frac{\partial}{\partial\psi}\ell_{\texttt{SE}}(\psi)$ has elements: $\frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta)$, $\frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}$, $\frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1}$, $\frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes A_3)\Delta u(\theta) - (T-1)\mathrm{tr}(G_3)$. Only the ρ-element of $\mathrm{E}[S_{\texttt{SE}}(\psi_0)]$ is non-zero, noting that $\mathbf{D}_{-1}$ in Lemma 3.1 reduces to $D(\rho)\otimes I_n$,

$$\sigma_{v0}^{-2}\mathrm{E}(\Delta u'\Omega^{-1}Y_{-1}) = -n\mathrm{tr}[C^{-1}D(\rho_0)], \tag{4.3}$$

where the $(T-1)\times(T-1)$ matrix $D(\rho)$ has the following expression:

$$D(\rho) = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \rho-2 & 1 & \cdots & 0 & 0 \\ (1-\rho)^2 & \rho-2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{T-5}(1-\rho)^2 & \rho^{T-6}(1-\rho)^2 & \cdots & 1 & 0 \\ \rho^{T-4}(1-\rho)^2 & \rho^{T-5}(1-\rho)^2 & \cdots & \rho-2 & 1 \end{pmatrix}.$$

It is easy to see that, when $|\rho|<1$, $\mathrm{tr}[C^{-1}D(\rho)] = \frac{1}{1-\rho} - \frac{1-\rho^T}{T(1-\rho)^2}$, a result that has appeared in the literature of non-spatial dynamic panel data models (e.g., Nickell, 1981; Lancaster, 2002; and Alvarez and Arellano, 2004), and was derived from different angles.

The result suggests that the ρ-element of the conditional quasi score function is such that $\mathrm{plim}_{n\to\infty}\frac{1}{nT\sigma_v^2}\Delta u'\Omega^{-1}\Delta Y_{-1} \neq 0$, unless T also approaches ∞. A necessary condition for consistency is violated, and hence the conditional QMLE of ρ is inconsistent when T is fixed. This result also suggests that even under the large n and large T set up, the conditional QMLE of ρ would incur a bias of order $O(T^{-1})$ as shown in Hahn and Kuersteiner (2002) for the regular DPD model. With (3.6) and the fact that other score elements have zero expectation, the adjusted quasi score becomes

$$S_{\texttt{SE}}^*(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta), \\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1} + n\mathrm{tr}(C^{-1}D(\rho)), \\ \frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes A_3)\Delta u(\theta) - (T-1)\mathrm{tr}(G_3). \end{cases} \tag{4.4}$$

Solving $S_{\texttt{SE}}^*(\psi) = 0$ leads to the M-estimator $\hat{\psi}_{\texttt{M}}$ of ψ. This root-finding process can be simplified by first solving the equations for β and σ_v^2, given $\delta = (\rho, \lambda_3)'$, resulting in the constrained M-estimators of β and σ_v^2 as $\hat{\beta}(\delta) = (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\Delta Y(\rho)$ and $\hat{\sigma}_v^2(\delta) = \frac{1}{n(T-1)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta)$, where $\Delta Y(\rho) = \Delta Y - \rho\Delta Y_{-1}$ and $\Delta\hat{u}(\delta) = \Delta u(\hat{\beta}(\delta), \rho)$.

Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ into the last two components of the AQS function in (4.4) gives the concentrated AQS functions:

$$S_{\texttt{SE}}^{*c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}_{v,\texttt{M}}^2(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} + n\text{tr}(C^{-1}D(\rho)), \\ \frac{1}{2\hat{\sigma}_{v,\texttt{M}}^2(\delta)}\Delta\hat{u}(\delta)'(C^{-1}\otimes A_3)\Delta\hat{u}(\delta) - (T-1)\text{tr}(G_3). \end{cases} \tag{4.5}$$

Solving the resulted concentrated estimating equations, $S_{\texttt{SE}}^{*c}(\delta) = 0$, we obtain the unconstrained M-estimators $\hat{\delta}_{\texttt{M}} = (\hat{\rho}_{\texttt{M}}, \hat{\lambda}_{3,\texttt{M}})'$ of δ. The unconstrained M-estimators of β and σ_v^2 are thus $\hat{\beta}_{\texttt{M}} \equiv \hat{\beta}(\hat{\delta}_{\texttt{M}})$ and $\hat{\sigma}_{v,\texttt{M}}^2 \equiv \hat{\sigma}_v^2(\hat{\delta}_{\texttt{M}})$.

Compared with the full QML estimation of Su and Yang (2015), the proposed M-estimation, though slightly less efficient, is much simpler as it is free from the specification of the initial conditions, and is thus robust against misspecifications of initial conditions. In contrast, the full QML estimation requires that the process starting time m is known a priori and that the processes evolve in the same manner before and after the data collection. Our Monte Carlo results and those in Su and Yang (2015) confirm these points.

4.2 The FE-SDPD model with SL effect

Setting $\lambda_2 = \lambda_3 = 0$, Model (3.1) reduces to a FE-SDPD model with only SL dependence. Now, $\psi = (\beta', \sigma_v^2, \rho, \lambda_1)'$. The conditional quasi Gaussian loglikelihood of ψ reduces to:

$$\ell_{\texttt{SL}}(\psi) = -\tfrac{n(T-1)}{2}\log(\sigma_v^2) + \log|\mathbf{B}_1| - \tfrac{1}{2}\log|\mathbf{C}| - \tfrac{1}{2\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta), \tag{4.6}$$

where $\theta = (\theta_1', \lambda_1)'$, $\theta_1 = (\beta', \rho)'$, and $v(\theta) = \mathbf{B}_1\Delta Y - \rho\Delta Y_{-1} - \Delta X\beta$. Given λ_1, $\ell_{\texttt{SL}}(\psi)$ is maximized at $\tilde{\theta}_1(\lambda_1) = (\Delta\mathbb{X}'\mathbf{C}^{-1}\Delta\mathbb{X})^{-1}\Delta\mathbb{X}'\mathbf{C}^{-1}\mathbf{B}_1\Delta Y$ and $\tilde{\sigma}_v^2(\lambda_1) = \frac{1}{n(T-1)}\Delta\tilde{v}'(\lambda_1)\mathbf{C}^{-1}\Delta\tilde{v}(\lambda_1)$, where $\Delta\tilde{v}(\lambda_1) = \mathbf{B}_1\Delta Y - \Delta\mathbb{X}\,\tilde{\theta}(\lambda_1)$, and $\Delta\mathbb{X} = (\Delta X, \Delta Y_{-1})$. Substituting $\tilde{\theta}_1(\lambda_1)$ and $\tilde{\sigma}_v^2(\lambda_1)$ back into $\ell_{\texttt{SL}}(\psi)$ gives the concentrated conditional loglikelihood function of λ_1,

$$\ell_{\texttt{SL}}^c(\lambda_1) = \log|\mathbf{B}_1| - \tfrac{n(T-1)}{2}\log(\tilde{\sigma}_v^2(\lambda_1)) - \tfrac{1}{2}\log|\mathbf{C}|. \tag{4.7}$$

Maximizing $\ell_{\texttt{SL}}^c(\lambda_1)$ gives the CQMLE $\tilde{\lambda}_1$ of λ_1, and thus the CQMLEs $\tilde{\theta} \equiv \tilde{\theta}(\hat{\lambda}_1)$ and $\tilde{\sigma}_v^2 \equiv \tilde{\sigma}_v^2(\tilde{\lambda}_1)$ of θ and σ_v^2, respectively.

The CQS function $S_{\texttt{SL}}(\psi)$ has elements: $\frac{1}{\sigma_v^2}\Delta X'\mathbf{C}^{-1}\Delta v(\theta)$, $\frac{1}{2\sigma_v^4}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta) - \frac{n(T-1)}{2\sigma_v^2}$, $\frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta Y_{-1}$, $\frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y - \text{tr}(\mathbf{B}_1^{-1}\mathbf{W}_1)$. The expectations of the first two components of $S_{\texttt{SL}}(\psi_0)$ are zero, but these of the last two are not as by Lemma 3.1,

$$\begin{aligned} \text{E}(\Delta v'\mathbf{C}^{-1}\Delta Y_{-1}) &= -\sigma_{v0}^2\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}), \quad \text{and} \qquad (4.8)\\ \text{E}(\Delta v'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y) &= -\sigma_{v0}^2\text{tr}(\mathbf{C}^{-1}\mathbf{D}_0\mathbf{W}_1), \qquad (4.9) \end{aligned}$$

where $\mathbf{D}_{-1}$ and $\mathbf{D}$ are given in Section 3.1 with $\mathcal{B}$ simplifies to $\rho B_1(\lambda_1)^{-1}$. These show that the last two elements of $\text{plim}_{n\to\infty}\frac{1}{n}S_{\texttt{SL}}(\psi_0)$ are not zero, showing that the CQMLEs of the SL model are inconsistent. Even when T grows with n, it can be shown that the CQMLE of ρ

has a bias of order $O(T^{-1})$ instead of the desired order $O((nT)^{-1})$. Some modifications are thus necessary whether T is fixed or not. The adjusted quasi score function is,

$$S^*_{\text{SL}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\mathbf{C}^{-1}\Delta v(\theta), \\ \frac{1}{2\sigma_v^4}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta Y_{-1} + \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y + \text{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1). \end{cases} \tag{4.10}$$

The M-estimator for the FE-SDPD-SLD model is thus defined as $\hat{\psi}_{\text{M}} = \arg\{S^*_{\text{SL}}(\psi) = 0\}$. The root-finding process can be simplified by first solving the equations for β and σ_v^2, given $\delta = (\rho, \lambda_1)'$, leading to the constrained M-estimators $\hat{\beta}(\delta) = (\Delta X'\mathbf{C}^{-1}\Delta X)^{-1}\Delta X'\mathbf{C}^{-1}\Delta Y(\delta)$ and $\hat{\sigma}_v^2(\delta) = \frac{1}{n(T-1)}\Delta\tilde{v}(\delta)'\mathbf{C}^{-1}\Delta\tilde{v}(\delta)$ for β and σ_v^2, where $\Delta Y(\delta) = \mathbf{B}_1\Delta Y - \rho\Delta Y_{-1}$ and $\Delta\hat{v}(\delta) = \Delta v(\hat{\beta}(\delta), \delta)$. Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ into the last two components of (4.10) gives the concentrated AQS function of δ:

$$S^{*c}_{\text{SL}}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{v,\text{M}}(\delta)}\Delta\hat{v}(\delta)'\mathbf{C}^{-1}\Delta Y_{-1} + \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\hat{\sigma}^2_{v,\text{M}}(\delta)}\Delta\hat{v}(\delta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y + \text{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1). \end{cases} \tag{4.11}$$

Solving the concentrated equations, $S^{*c}_{\text{SL}}(\delta) = 0$, gives the unconstrained M-estimator $\hat{\delta}_{\text{M}}$ of δ. The unconstrained M-estimators of β and σ_v^2 are thus $\hat{\beta}_{\text{M}} \equiv \hat{\beta}(\hat{\delta}_{\text{M}})$ and $\hat{\sigma}^2_{v,\text{M}} \equiv \hat{\sigma}_v^2(\hat{\delta}_{\text{M}})$.

4.3 The FE-SDPD model with SL and STL effects

Setting $\lambda_3 = 0$, Model (3.1) reduces to a FE-SDPD model with SL and STL dependence. Now, $\psi = (\beta', \sigma_v^2, \rho, \lambda_1, \lambda_2)'$. The conditional quasi Gaussian loglikelihood of ψ reduces to:

$$\ell_{\text{STL}}(\psi) = -\tfrac{n(T-1)}{2}\log(\sigma_v^2) + \log|\mathbf{B}_1| - \tfrac{1}{2}\log|\mathbf{C}| - \tfrac{1}{2\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta), \tag{4.12}$$

where $\theta = (\beta', \rho, \lambda_1, \lambda_2)'$, $\theta_1 = (\beta', \rho, \lambda_2)'$, and $v(\theta) = \mathbf{B}_1\Delta Y - (\rho I_n + \lambda_2\mathbf{W}_2)\Delta Y_{-1} - \Delta X\beta$. Given λ_1, $\ell_{\text{STL}}(\psi)$ is maximized at $\tilde{\theta}_1(\lambda_1) = (\Delta\mathbb{X}'\mathbf{C}^{-1}\Delta\mathbb{X})^{-1}\Delta\mathbb{X}'\mathbf{C}^{-1}\mathbf{B}_1\Delta Y$ and $\tilde{\sigma}_v^2(\lambda_1) = \frac{1}{n(T-1)}\Delta\tilde{v}'(\lambda_1)\mathbf{C}^{-1}\Delta\tilde{v}(\lambda_1)$, where $\Delta\mathbb{X} = (\Delta X, \Delta Y_{-1}, \mathbf{W}_2\Delta Y_{-1})$ and $\Delta\tilde{v}(\lambda_1) = \mathbf{B}_1\Delta Y - \Delta\mathbb{X}\,\tilde{\theta}(\lambda_1)$. Thus, the concentrated conditional quasi loglikelihood function of λ_1 is,

$$\ell^c_{\text{STL}}(\lambda_1) = +\log|\mathbf{B}_1| - \tfrac{n(T-1)}{2}\log(\tilde{\sigma}_v^2(\lambda_1)) - \tfrac{1}{2}\log|\mathbf{C}|. \tag{4.13}$$

Maximizing $\ell^c_{\text{STL}}(\lambda_1)$ gives the CQMLE $\tilde{\lambda}_1$, and thus the CQMLEs $\tilde{\theta} \equiv \tilde{\theta}(\hat{\lambda}_1)$ and $\tilde{\sigma}_v^2 \equiv \tilde{\sigma}_v^2(\tilde{\lambda}_1)$.

The CQS function $S_{\text{STL}}(\psi)$ contains elements: $\frac{1}{\sigma_v^2}\Delta X'\mathbf{C}^{-1}\Delta v(\theta)$, $\frac{1}{2\sigma_v^4}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta) - \frac{n(T-1)}{2\sigma_v^2}$, $\frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta Y_{-1}$, $\frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y - \text{tr}(\mathbf{B}_1^{-1}\mathbf{W}_1)$, $\frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_2\Delta Y_{-1}$. It is easy to see that the first two components of $\text{E}[S_{\text{SL}}(\psi_0)]$ are zero, but these of the last three components are not, and are obtained from (3.11)-(3.13). Thus a necessary condition for the consistency of parameter estimators is violated, suggesting that with T fixed the conditional

QMLEs of the FE-DPD-SLD model are inconsistent. Even when T grows with n, it can be shown that the conditional QMLE of ρ has a bias of order $O(T^{-1})$ instead of the desired order $O((nT)^{-1})$. Some modifications are thus necessary whether T is fixed or not, and the adjusted quasi score function is,

$$S^*_{\mathtt{STL}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\mathbf{C}^{-1}\Delta v(\theta), \\ \frac{1}{2\sigma_v^4}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2). \end{cases} \tag{4.14}$$

The M-estimator for the FE-SDPD-SLD model is thus defined as $\hat{\psi}_{\mathtt{SL}} = \arg\{S^*_{\mathtt{M}}(\psi) = 0\}$. The root-finding process can be simplified by first solving the equations for β and σ_v^2, given $\delta = (\rho, \lambda_1, \lambda_2)'$, leading to the constrained M-estimators $\hat{\beta}(\delta) = (\Delta X'\mathbf{C}^{-1}\Delta X)^{-1}\Delta X'\mathbf{C}^{-1}\Delta Y(\delta)$ and $\hat{\sigma}_v^2(\delta) = \frac{1}{n(T-1)}\Delta\hat{v}(\delta)'\mathbf{C}^{-1}\Delta\hat{v}(\delta)$, where $\Delta Y(\delta) = \mathbf{B}_1\Delta Y - (\rho I_n + \lambda_2\mathbf{W}_2)\Delta Y_{-1}$ and $\Delta\hat{v}(\delta) = \Delta v(\hat{\beta}(\delta), \delta)$. Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ into the last two components of (4.14) gives the concentrated AQS function of δ:

$$S^{*c}_{\mathtt{STL}}(\delta) = \begin{cases} \frac{1}{\tilde{\sigma}^2_{v,\mathtt{M}}(\delta)}\tilde{v}(\delta)'\mathbf{C}^{-1}\Delta\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\tilde{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\tilde{v}(\delta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{W}_1\mathbf{C}^{-1}\mathbf{D}), \\ \frac{1}{\tilde{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\tilde{v}(\delta)'\mathbf{C}^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2). \end{cases} \tag{4.15}$$

Solving the concentrated equations, $S^{*c}_{\mathtt{SL}}(\delta) = 0$, gives the M-estimator $\hat{\delta}_{\mathtt{M}}$ of δ. The M-estimators of β and σ_v^2 are, thus, $\hat{\beta}_{\mathtt{M}} \equiv \hat{\beta}(\hat{\delta}_{\mathtt{M}})$ and $\hat{\sigma}^2_{v,\mathtt{M}} \equiv \hat{\sigma}_v^2(\hat{\delta}_{\mathtt{M}})$.

4.4 The FE-SDPD model with SL and SE effects

Setting $\lambda_2 = 0$ in Model (3.1) yields an FE-SDPD model with both SL and SE dependence. The conditional quasi Gaussian loglikelihood of $\psi = (\beta', \sigma_v^2, \rho, \lambda_1, \lambda_3)'$ reduces to,

$$\ell_{\mathtt{SLE}}(\psi) = -\tfrac{n(T-1)}{2}\log(\sigma_v^2) + \log|\mathbf{B}_1| - \tfrac{1}{2}\log|\Omega| - \tfrac{1}{2\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta), \tag{4.16}$$

where $\theta = (\beta', \rho, \lambda_1)'$ and $\Delta u(\theta) = \mathbf{B}_1\Delta Y - \rho\Delta Y_{-1} - \Delta X\beta$. Given $\lambda = (\lambda_1, \lambda_3)'$, $\ell_{\mathtt{SLE}}(\psi)$ is maximized at $\tilde{\theta}(\lambda) = (\Delta\mathbb{X}'\Omega^{-1}\Delta\mathbb{X})^{-1}\Delta\mathbb{X}'\Omega^{-1}\mathbf{B}_1\Delta Y$ and $\tilde{\sigma}_v^2(\lambda) = \frac{1}{n(T-1)}\Delta\tilde{u}'(\lambda)\Omega^{-1}\Delta\tilde{u}(\lambda)$, where $\Delta\tilde{u}(\lambda) = \mathbf{B}_1\Delta Y - \Delta\mathbb{X}\,\tilde{\theta}(\lambda)$, and $\Delta\mathbb{X} = (\Delta X, \Delta Y_{-1})$. Substituting $\tilde{\theta}(\lambda)$ and $\tilde{\sigma}_v^2(\lambda)$ back into $\ell_{\mathtt{SLE}}(\psi)$ gives the concentrated loglikelihood function of λ,

$$\ell^c_{\mathtt{SLE}}(\lambda) = \log|\mathbf{B}_1| - \tfrac{n(T-1)}{2}\log(\tilde{\sigma}_v^2(\lambda)) - \tfrac{1}{2}\log|\Omega|. \tag{4.17}$$

Maximizing $\ell^c_{\mathtt{SLE}}(\lambda)$ gives the CQMLE $\tilde{\lambda}$, and thus the CQMLEs $\tilde{\theta} \equiv \tilde{\theta}(\hat{\lambda})$ and $\tilde{\sigma}_v^2 \equiv \tilde{\sigma}_v^2(\tilde{\lambda})$.

The CQS function $S_{\mathtt{SLE}}(\psi)$ has the components: $\frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta)$, $\frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) -$

$\frac{n(T-1)}{2\sigma_v^2}$, $\frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1}$, $\frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y - \text{tr}(\mathbf{B}_1^{-1}\mathbf{W}_1)$, $\frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes A_3)\Delta u(\theta) - (T-1)\text{tr}(G_3)$. The β, σ_v^2 and λ_3 components of $\text{E}[S_{\texttt{SLE}}(\psi_0)]$ are zero, but the ρ and λ_1 components are not as seen from Lemma 3.1: $\text{E}(\Delta u'\Omega^{-1}\Delta Y_{-1}) = -\sigma_{v0}^2\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10})$ and $\text{E}(\Delta u'\Omega^{-1}\mathbf{W}_1\Delta Y) = -\sigma_{v0}^2\text{tr}(\mathbf{C}^{-1}\mathbf{D}_0\mathbf{W}_1)$, which are of identical forms as those for the SLD model. The results show that the CQMLEs are not consistent unless T also approaches infinity. To achieve consistency, the conditional quasi score function should be modified as:

$$S_{\texttt{SLE}}^*(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta), \\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1} + \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \text{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes A_3)\Delta u(\theta) - (T-1)\text{tr}(G_3). \end{cases} \tag{4.18}$$

The M-estimator of the FE-DPD-SLE model is defined as $\hat{\psi}_{\texttt{M}} = \arg\{S_{\texttt{SLE}}^*(\psi) = 0\}$. The root-finding process can be simplified by first solving the equations for β and σ_v^2, resulting in the constrained M-estimators $\hat{\beta}(\delta) = (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\Delta Y(\rho, \lambda_1)$ and $\hat{\sigma}_v^2(\delta) = \frac{1}{n(T-1)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta)$, given $\delta = (\rho, \lambda_1, \lambda_3)'$, where $\Delta Y(\rho, \lambda_1) = \mathbf{B}_1\Delta Y - \rho\Delta Y_{-1}$ and $\Delta\hat{u}(\delta) = \Delta u(\hat{\beta}(\delta), \rho, \lambda_1)$. Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_{v,\texttt{M}}^2(\delta)$ into the last three components of $S_{\texttt{SLE}}^*(\psi)$ gives the concentrated AQS function:

$$S_{\texttt{SLE}}^{*c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}_{v,\texttt{M}}^2(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} + \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\hat{\sigma}_{v,\texttt{M}}^2(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \text{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{2\hat{\sigma}_{v,\texttt{M}}^2(\delta)}\Delta\hat{u}(\delta)'(C^{-1}\otimes A_3)\Delta\hat{u}(\delta) - (T-1)\text{tr}(G_3). \end{cases} \tag{4.19}$$

Solving the concentrated equations, $S_{\texttt{SLE}}^{*c}(\delta) = 0$, gives the M-estimator $\hat{\delta}_{\texttt{M}}$ of δ. The M-estimators of β and σ_v^2 are thus $\hat{\beta}_{\texttt{M}} \equiv \hat{\beta}(\hat{\delta}_{\texttt{M}})$ and $\hat{\sigma}_{v,\texttt{M}}^2 \equiv \hat{\sigma}_v^2(\hat{\delta}_{\texttt{M}})$.

5 Monte Carlo Results

Monte Carlo experiments are carried out to investigate (i) the finite sample performance of the M-estimators of three FE-SDPD models considered in this paper, and (ii) the finite sample performance of the proposed OPMD estimates of the robust standard errors. We use the following three data generating processes (DGPs):

$\texttt{SE}:\quad y_t = \rho y_{t-1} + \beta_0\iota_n + X_t\beta_1 + Z\gamma + \mu + u_t,\ \ u_t = \lambda_3 W_3 u_t + v_t,$

$\texttt{SL}:\quad y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \beta_0\iota_n + X_t\beta_1 + Z\gamma + \mu + v_t,$

$\texttt{SLE}:\quad y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \beta_0\iota_n + X_t\beta_1 + Z\gamma + \mu + u_t,\ \ u_t = \lambda_3 W_3 u_t + v_t,$

$\texttt{STLE}: y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + \beta_0\iota_n + X_t\beta_1 + Z\gamma + \mu + v_t,$

$\texttt{STLE}: y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + \beta_0\iota_n + X_t\beta_1 + Z\gamma + \mu + u_t,\ u_t = \lambda_2 W_3 u_t + v_t.$

The elements of X_t are generated in a similar fashion as in Hsiao et al. (2002),[13] and the elements of Z are randomly generated from $Bernoulli(0.5)$. The spatial weight matrices are generated according to Rook or Queen contiguity, or group interaction schemes.[14] We choose $\beta_0 = 5, \beta_1 = 1, \gamma = 1, \sigma_\mu = 1, \sigma_v = 1$ or 2, a set of values for ρ ranging from -0.9 to 0.9, a set of values for λ_1 and/or λ_2 in a similar range, $T = 3$ or 6, and $N = 50$, 100, and 200. Each set of Monte Carlo results, corresponding to a combination of the values of $(n, T, m, \rho, \lambda' s, \sigma_v)$, is based on 2000 samples. The error (v_t) distributions can be (i) normal, (ii) normal mixture (10% $N(0, 4)$ and 90% $N(0, 1)$), or (iii) chi-square with 3 degrees of freedom.[15] The fixed effects μ are generated according to either $\frac{1}{T}\sum_{t=1}^{T} X_t + e$ or e, where $e \sim (0, I_N)$, resulting in the fixed effects that are either correlated or uncorrelated with the regressors.

The reported results are the Monte Carlo mean and Monte Carlo standard deviation (sd) for the conditional QMLEs (CQMLE) and the proposed M-estimators. The standard errors (ses): $\widetilde{se}$ calculated based on $\widehat{\Gamma}_{\mathtt{M}}^{*-1}$, $\widehat{se}$ calculated based on $\Sigma_{\mathtt{M}}^{*-1}(\hat{\psi}_{\mathtt{M}})$ and the robust ses $\widehat{rse}$ calculated based on $\Sigma_{\mathtt{M}}^{*-1}(\hat{\psi}_{\mathtt{M}})\widehat{\Gamma}_{\mathtt{M}}^{*}\Sigma_{\mathtt{M}}^{*-1}(\hat{\psi}_{\mathtt{M}})$, are also reported for the proposed M-estimators, for Model `M = SE, SL, SLE, STL` and `STLE`. Due to the space constraints, only a subset of results, corresponding to the case of correlated fixed effects, are reported.

Table 1a presents empirical means and sds for the `SE` model and Table 1b presents the corresponding ses for the M-estimators. For this model, the full QMLE (FQMLE) is available from Su and Yang (2015) and thus is included in the Monte Carlo experiments for comparison purpose. The results show an excellent performance of the proposed M-estimators, and an excellent performance of the proposed OPMD estimate of the robust standard error (rse). The M-estimator of the dynamic parameter is nearly unbiased, as is the FQMLE, whereas the CQMLE can be quite biased and as n increases it does not show a sign of convergence. As expected, the M-estimator is slightly less efficient than the FQMLE, but when T is increased from 3 to 7, the difference becomes negligible. However, the FQMLE depends on m value and a wrong specification of it may result in poor estimate when ρ is negative and large (see Su and Yang, 2015). In contrast, the m value does not have any significant impact on either CQMLE or the M-estimator. The OPMD estimates of the rses are very close to the corresponding Monte Carlo sds. In contrast, the non-robust se of $\hat{\sigma}_v^2$ can be quite different from the corresponding Monte Carlo sd when the errors are nonnormal. When T is increased from 3 to 7, the CQMLE of ρ improves significantly. Both FQMLE and M-estimator of the spatial parameter λ_3 show some bias (the CQMLE is more biased). This is perhaps due to the intrinsic nature of the QML-type estimation of the spatial effects.[16] The $\sqrt{n}$-consistency of the FQMLE and M-estimator is clearly demonstrated by the Monte Carlo sds.

[13]The detail is: $X_t = \mu_x + gt1_n + \zeta_t$, $(1 - \phi_1 L)\zeta_t = \varepsilon_t + \phi_2\varepsilon_{t-1}$, $\varepsilon_t \sim N(0, \sigma_1^2 I_n)$, $\mu_x = e + \frac{1}{T+m+1}\sum_{t=-m}^{T}\varepsilon_t$, and $e \sim N(0, \sigma_2^2 I_n)$. Let $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2)$. Alternatively, X_t can be randomly generated from $N(0, \sigma_1^2 I_n)$. The σ_1^2 is the key parameter that controls the variability of the regressors, and thus the signal-to-noise ratio.

[14]The Rook and Queen schemes are standard. For group interaction, we first generate $k = n^\alpha$ groups of sizes $n_g \sim U(.5\bar{n}, 1.5\bar{n})$, $g = 1, \cdots, k$, where $0 < \alpha < 1$ and $\bar{n} = n/k$, and then adjust n_g so that $\sum_{g=1}^{k} n_g = N$. The reported results correspond to $\alpha = 0.5$. See Yang (2015) for details in generating these spatial layouts.

[15]In both (ii) and (iii), the generated errors are standardized to have mean zero and variance σ_v^2.

[16]See Yang (2015) for a detailed examination on the bias of estimating a spatial lag model and the general methodology on finite sample bias corrections for nonlinear estimators.

Table 2a presents empirical means and sds for the SL model and Table 2b the corresponding ses for the M-estimators. The results again show excellent performance of the proposed M-estimation strategy, which offers dramatic improvements over the conditional QML estimation method when T is small. The results also show that the proposed OPMD estimates of standard errors also perform excellently. As discussed in Section 2, a spatial model with SL effects may be more popular due to the fact that it is able to capture the neighborhood or spatial effects on both the mean and variance levels, and hence it is important to have simple and reliable estimation and inference methods for the SDPD models with SL effects.

Tables 3a and 3b present results for the model SLE model, i.e., the FE-DPD model with both SL and SE effects. Similar observations as in the two simpler models can be made, except that the estimators for the spatial parameters, in particular the spatial error parameter, are more biased, and that the OPMD se estimates of them are slightly less accurate.

Tables 4a and 4b present results for the STL model that incorporates both SL and STL effects into the FE-DPD model. Our Monte Carlo results show that the numerical stability in the estimation of the STL effect requires a larger signal-to-noise ratio (the σ_1/σ_v value given at the bottom of Table 4a). With a larger signal-to-noise ratio, however, the CQMLE performs better, though it is still inconsistent and clearly outperformed by the proposed M-estimator.

Tables 5a and 5b present results for the STLE model with all the three spatial effects. The results show that the proposed M-estimators of the model parameters, and the proposed OPMD-based standard error estimates of the M-estimators perform very well. Similar to the STL model, the estimation of this model requires a larger signal-to-noise ratio than a model without the STL effect. Finally, for all the models, the nonnormality can have a significant effect on the se of the error variance σ_v^2, and hence it is important to use the OPMD-based robust standard errors in statistical inferences when the normality of the errors is in doubt. More Monte Carlo results and Matlab codes are available from the author upon request.

6 Conclusion

We introduce a general strategy (M-estimation) for estimating a fixed-effects dynamic panel data (DPD) model with three major forms of spatial effects: the spatial lag, space-time lag, and spatial error, based on short panels. The proposed M-estimation method is simple as it is based on the adjusted quasi score functions, and is robust in the sense that it is free from the specification of the initial conditions and allowing errors to be nonnormal. A initial condition free method for estimating the robust standard errors of the M-estimators is also given. These together lead to a complete set of inference methods for the fixed-effects spatial DPD models that are free from the specification of the initial conditions, and robust against error distributions. The simplicity and generality of the proposed methods render them to be very attractive to the practitioners.

Appendix A: Some Basic Lemmas

The following lemmas are essential for the proofs of the main results in this paper.

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

(i) the sequence $\{A_nB_n\}$ are uniformly bounded in both row and column sums,
(ii) the elements of A_n are uniformly bounded and $tr(A_n) = O(n)$, and
(iii) the elements of A_nC_n and C_nA_n are uniformly $O(h_n^{-1})$.

Lemma A.2 *(Lee, 2004a, p.1918): For W_1 and B_1 defined in Model (3.1), if $\|W_1\|$ and $\|B_{10}^{-1}\|$ are uniformly bounded, where $\|\cdot\|$ is a matrix norm, then $\|B_1^{-1}\|$ is uniformly bounded in a neighborhood of λ_{10}.*

Lemma A.3 *(Lee, 2004a, p.1918): Let X_n be an $n \times p$ matrix. If the elements X_n are uniformly bounded and $\lim_{n\to\infty} \frac{1}{n}X_n'X_n$ exists and is nonsingular, then $P_n = X_n(X_n'X_n)^{-1}X_n'$ and $M_n = I_n - P_n$ are uniformly bounded in both row and column sums.*

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

(i) $\mathrm{E}(v_n'A_nv_n) = O(\frac{n}{h_n})$, *(ii)* $\mathrm{Var}(v_n'A_nv_n) = O(\frac{n}{h_n})$,
(iii) $\mathrm{Var}(v_n'A_nv_n + b_n'v_n) = O(\frac{n}{h_n})$, *(iv)* $v_n'A_nv_n = O_p(\frac{n}{h_n})$,
(v) $v_n'A_nv_n - \mathrm{E}(v_n'A_nv_n) = O_p((\frac{n}{h_n})^{\frac{1}{2}})$, *(vi)* $v_n'A_nb_n = O_p((\frac{n}{h_n})^{\frac{1}{2}})$,

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

Lemma A.5 *(**Central Limit Theorem for bilinear quadratic forms**). Let $\{\Phi_n\}$ be a sequence of $n \times n$ matrices with row and column sums uniformly bounded, and elements of uniform order $O(h_n^{-1})$. Let $v_n = (v_1, \cdots, v_n)'$ be a random vector of iid elements with mean zero, variance σ_v^2, and finite $(4+2\epsilon_0)$th moment for some $\epsilon_0 > 0$. Let $b_n = \{b_{ni}\}$ be a sequence of $n \times 1$ random vectors such that (i) $\{\mathrm{E}(b_{ni}^2)\}$ are of uniform order $O(h_n^{-1})$, (ii) $sup_i E|b_{ni}|^{2+\epsilon_0} < \infty$, (iii) $\frac{h_n}{n}\sum_{i=1}^n[\phi_{n,ii}(b_{ni} - \mathrm{E}b_{ni})] = o_p(1)$ where $\{\phi_{n,ii}\}$ are the diagonal elements of Φ_n, and (iv) $\frac{h_n}{n}\sum_{i=1}^n[b_{ni} - \mathrm{E}(b_{ni}^2)] = o_p(1)$. Define the bilinear-quadratic form:*

$$Q_n = b_n'v_n + v_n'\Phi_nv_n - \sigma_v^2 tr(\Phi_n),$$

and let $\sigma_{Q_n}^2$ be the variance of Q_n. If $lim_{n\to\infty} h_n^{1+2/\epsilon_0}/n = 0$ and $\{\frac{h_n}{n}\sigma_{Q_n}^2\}$ are bounded away from zero, then $Q_n/\sigma_{Q_n} \xrightarrow{d} N(0,1)$.

Proof of Lemma A.5: The proof proceeds by assuming (W.L.O.G.) Φ_n being symmetric, with elements denoted by $\phi_{n,ij}$. Write $Q_n = \sum_{i=1}^n [b_{ni}v_i + v_i\xi_{ni} + \phi_{n,ii}(v_i^2 - \sigma_v^2)] \equiv \sum_{i=1}^n Y_{ni}$, where $\xi_{ni} = 2\sum_{j=1}^{i-1}\phi_{n,ij}v_j$. Consider the σ-fields $\mathcal{G}_i = \sigma(v_1, \ldots, v_i)$ generated by $(v_1, \ldots, v_i), i = 1, \ldots, n$. By construction, $\mathcal{G}_{i-1} \subseteq \mathcal{G}_i$. Define the σ-field $\mathcal{F}_{n0}$ generated by b_n. By independence between b_n and v_n, $\mathcal{F}_{ni} = \mathcal{F}_{n0} \times \mathcal{G}_i$ is the σ-field generated by $(b_n, v_1, \ldots, v_i)$. Clearly, Y_{ni} is $\mathcal{F}_{ni}$-measurable and ξ_{ni} is $\mathcal{F}_{n,i-1}$-measurable. It follows that $\mathrm{E}(Y_{ni}|\mathcal{F}_{n,i-1}) = b_{ni}\mathrm{E}(v_i) + \mathrm{E}(v_i)\xi_{ni} + \phi_{n,ii}\mathrm{E}(v_i^2 - \sigma_v^2) = 0$, and hence $\{Y_{ni}, \mathcal{F}_{ni},\ 1 \le i \le n\}$ forms a martingale difference (M.D.) array, and $\sigma_{Q_n}^2 = \sum_{i=1}^n E(Y_{ni}^2)$. Define $Z_{ni} = Y_{ni}/\sigma_{Q_n}$. Then, $\{Z_{ni}, \mathcal{F}_{ni},\ 1 \le i \le n\}$ also forms a M.D. array. The proof of the lemma thus amounts to verify the conditions for the central limit theorem (CLT) for M.D. arrays, e.g., the condition (A.1) or (A.3) and condition (A.2) of Theorem A.1 in Kelejian and Prucha (2001):

$$\begin{aligned}
&(a) \quad \textstyle\sum_{i=1}^n \mathrm{E}[\mathrm{E}(|Z_{ni}|^{2+\delta}|\mathcal{F}_{n,i-1})] \longrightarrow 0, \quad \text{for some } \delta > 0;\\
&(b) \quad \textstyle\sum_{i=1}^n \mathrm{E}(Z_{ni}^2|\mathcal{F}_{n,i-1}) \xrightarrow{p} 1.
\end{aligned}$$

The details for the proof of (a) follow closely that of Theorem 1 of Kelejian and Prucha (2001), where the quantities $|b_{ni}|$, b_{ni}^2, $|b_{ni}|^q$ are replaced by their expectations, and references are made to the proof of Lemma A.13 of Lee (2004b) to take care of the case when h_n is unbounded.

To prove (b), we have $\sum_{i=1}^n \mathrm{E}[Z_{ni}^2|\mathcal{F}_{n,i-1}] - 1 = \sigma_{Q_n}^{-2}\sum_{i=1}^n[\mathrm{E}(Y_{ni}^2|\mathcal{F}_{n,i-1}) - \mathrm{E}(Y_{ni}^2)]$, and

$$\begin{aligned}
& \tfrac{h_n}{n}\textstyle\sum_{i=1}^n[\mathrm{E}(Y_{ni}^2|\mathcal{F}_{n,i-1}) - \mathrm{E}(Y_{ni}^2)]\\
= \;& \sigma_v^2\tfrac{h_n}{n}\textstyle\sum_{i=1}^n(\xi_{ni}^2 - \tau_{ni}^2) + 2\sigma_v^2\tfrac{h_n}{n}\sum_{i=1}^n(b_{ni}\xi_{ni}) + 2\mu_3\tfrac{h_n}{n}\sum_{i=1}^n(\phi_{n,ii}\xi_{ni})\\
& +2\mu_3\tfrac{h_n}{n}\textstyle\sum_{i=1}^n[\phi_{n,ii}(b_{ni} - \mathrm{E}b_{ni})] + \sigma_v^2\tfrac{h_n}{n}\sum_{i=1}^n[b_{ni}^2 - \mathrm{E}(b_{ni}^2)]\\
\equiv \;& \sigma_v^2 Q_1 + 2\sigma_v^2 Q_2 + 2\mu_3 Q_3 + 2\mu_3 Q_4 + \sigma_v^2 Q_5,
\end{aligned}$$

where $\tau_{ni}^2 = \mathrm{Var}(\xi_{ni}) = 4\sigma_v^2\sum_{j=1}^{i-1}\phi_{n,ij}^2$. We have

$$Q_1 = \tfrac{h_n}{n}\textstyle\sum_{i=1}^n(\xi_{ni}^2 - \tau_{ni}^2) = 4\tfrac{h_n}{n}\sum_{j=1}^{n-1}a_{nj}(v_j^2 - \sigma_v^2) + 8\tfrac{h_n}{n}\sum_{j=1}^{n-1}v_j\varepsilon_{nj}.$$

where $a_{nj} = \sum_{i=j+1}^n \phi_{n,ij}^2$, $\varepsilon_{nj} = \sum_{k=1}^{j-1}c_{n,ik}v_k$, and $c_{n,ik} = \sum_{i=j+1}^n \phi_{n,ij}\phi_{n,ik}$. Clearly, both $\{(v_j^2 - \sigma_v^2), \mathcal{G}_i\}$ and $\{v_j\varepsilon_{nj}, \mathcal{G}_i\}$ are M.D. arrays, and hence their convergence in probability to zero is proved by applying the weak law of large numbers (WLLN) for M.D. arrays of Davidson (1994, p. 299). It follows that $Q_1 \xrightarrow{p} 0$.

By applying Chebyshev inequality, we show that $Q_2 \xrightarrow{p} 0$. Now, it is easy to see that $Q_3 = \frac{h_n}{n}\sum_{j=1}^{n-1}d_{n,j}v_j$ where $d_{n,j} = \sum_{i=j+1}^n \phi_{n,ii}\phi_{ij}$. Thus, the convergence of Q_3 is proved by applying Chebyshev inequality. Finally, by Assumptions (iii) and (iv) stated in Lemma A.5, both Q_4 and Q_5 converge to zero in probability. This completes the proof Lemma A.5 (details, in particular the proof of (a), are available from the author upon request). ■

Appendix B: Proofs of Lemmas 3.1-3.3

Proof of Lemma 3.1: Using $\Delta y_t = \mathcal{B}_0 \Delta y_{t-1} + B_{10}^{-1}\Delta X_t + B_{10}^{-1}B_{30}^{-1}\Delta v_t, \quad t = 2, \ldots, T$ given in (3.8), we have under Assumption A: if $m \geq 1$, then

$$\mathrm{E}(\Delta y_1 \Delta v_2') = B_{10}^{-1}B_{30}^{-1}\mathrm{E}(\Delta v_1 \Delta v_2') = -\sigma_{v0}^2 B_{10}^{-1}B_{30}^{-1};$$

if $m = 0$, then $\mathrm{E}(\Delta y_1 \Delta v_2') = B_{10}^{-1}B_{30}^{-1}\mathrm{E}(y_1 \Delta v_2') = B_{10}^{-1}B_{30}^{-1}\mathrm{E}(v_1 \Delta v_2') = -\sigma_{v0}^2 B_{10}^{-1}B_{30}^{-1}$.

Now, for $t \geq 2$, we have, $\mathrm{E}(\Delta y_t \Delta v_{t+1}') = B_{10}^{-1}B_{30}^{-1}\mathrm{E}(v_t \Delta v_{t+1}') = -\sigma_{v0}^2 B_{10}^{-1}B_{30}^{-1}$,
$\mathrm{E}(\Delta y_t \Delta v_t') = \mathcal{B}_0\mathrm{E}(\Delta y_{t-1}\Delta v_t') + B_{10}^{-1}B_{30}^{-1}\mathrm{E}(\Delta v_t \Delta v_t') = \sigma_{v0}^2(2I_n - \mathcal{B}_0)B_{10}^{-1}B_{30}^{-1}$, and
$\mathrm{E}(\Delta y_{t+1}\Delta v_t') = \mathcal{B}_0\mathrm{E}(\Delta y_t \Delta v_t') + B_{10}^{-1}B_{30}^{-1}\mathrm{E}(\Delta v_{t+1}\Delta v_t') = -\sigma_{v0}^2(I_n - \mathcal{B}_0)^2 B_{10}^{-1}B_{30}^{-1}$.
For $t \geq 3$, we have, $\mathrm{E}(\Delta y_t \Delta v_2') = -\sigma_{v0}^2\mathcal{B}_0^{t-3}(I_n - \mathcal{B}_0)^2 B_{10}^{-1}B_{30}^{-1}$,
For $t \geq 4$, we have, $\mathrm{E}(\Delta y_t \Delta v_3') = -\sigma_{v0}^2\mathcal{B}_0^{t-4}(I_n - \mathcal{B}_0)^2 B_{10}^{-1}B_{30}^{-1}$, etc.

Summarize above, we obtain results of Lemma (3.1). ■

Proof of Lemma 3.2: By (3.8), continuous substitution gives, for $t = 2, \ldots, T$,

$$\begin{aligned}
\Delta y_t &= \mathcal{B}_0^{t-1}\Delta y_1 + \mathcal{B}_0^{t-2}B_{10}^{-1}\Delta X_2\beta_0 + \ldots + B_{10}^{-1}\Delta X_t\beta_0 \\
&\quad + \mathcal{B}_0^{t-2}B_{10}^{-1}B_{30}^{-1}\Delta v_2 + \ldots + B_{10}^{-1}B_{30}^{-1}\Delta v_t \\
&= \mathcal{B}_0^{t-1}\Delta y_1 + \{\mathcal{B}_0^{t-2}, \mathcal{B}_0^{t-3}, \ldots, I_n, 0, \ldots, 0\}\mathbf{B}_{10}^{-1}\Delta X\beta_0 \\
&\quad + \{\mathcal{B}_0^{t-2}, \mathcal{B}_0^{t-3}, \ldots, I_n, 0, \ldots, 0\}\mathbf{B}_{10}^{-1}\mathbf{B}_{30}^{-1}\Delta v.
\end{aligned}$$

The results of Lemma 3.2 thus follow. ■

Proof of Lemma 3.3: First, for the terms linear in Δv, we have,

$$\begin{aligned}
\Pi'\Delta v &= \textstyle\sum_{t=2}^T \Pi_t'\Delta v_t \\
&= \textstyle\sum_{t=2}^T \sum_{i=1}^n \Pi_{it}'\Delta v_{it} \\
&= \textstyle\sum_{i=1}^n \sum_{t=2}^T \Pi_{it}'\Delta v_{it} \\
&= \textstyle\sum_{i=1}^n g_{1i}.
\end{aligned}$$

Clearly, $\{g_{1i}\}$ are independent with mean zero, and thus form a vector M.D. sequence. Now, for the terms quadratic in Δv, we have,

$$\begin{aligned}
\mathrm{E}(\Delta v'\Phi\Delta v) &= \sigma_{v0}^2\mathrm{tr}[(C \otimes I_n)\Phi] \\
&= \sigma_{v0}^2 \textstyle\sum_{t=2}^T \sum_{s=2}^T \mathrm{tr}(c_{ts}\Phi_{st}) \\
&= \sigma_{v0}^2 \textstyle\sum_{i=1}^n \sum_{t=2}^T \sum_{s=2}^T (c_{ts}\Phi_{ii,st}) \\
&= \textstyle\sum_{i=1}^n \sum_{t=2}^T d_{it},
\end{aligned}$$

where $\{c_{ts}, t, s = 2, \ldots, T\}$ are the elements of the matrix C given in Section 3.1, $\{\Phi_{ii,ts}, i =$

$1,\ldots,n\}$ are the diagonal elements of Φ_{ts}, and $d_{it} = \sigma_{v0}^2 \sum_{s=2}^T (c_{ts}\Phi_{ii,st})$; and

$$\begin{aligned}
\Delta v'\Phi\Delta v - \mathrm{E}(\Delta v'\Phi\Delta v) &= \textstyle\sum_{t=2}^T \sum_{s=2}^T \Delta v_t'\Phi_{ts}\Delta v_s - \sum_{i=1}^n \sum_{t=2}^T d_{it} \\
&= \textstyle\sum_{t=2}^T \sum_{s=2}^T \Delta v_t'(\Phi_{ts}^u + \Phi_{ts}^l + \Phi_{ts}^d)\Delta v_s - \sum_{i=1}^n \sum_{t=2}^T d_{it} \\
&= \textstyle\sum_{t=2}^T \sum_{s=2}^T \left[\Delta v_s'\Phi_{ts}^{u\prime}\Delta v_t + \Delta v_t'(\Phi_{ts}^l + \Phi_{ts}^d)\Delta v_s\right] - \sum_{i=1}^n \sum_{t=2}^T d_{it} \\
&= \textstyle\sum_{t=2}^T \Delta v_t'\Delta\xi_t + \sum_{t=2}^T \Delta v_t'\Delta v_t^* - \sum_{i=1}^n \sum_{t=2}^T d_{it}, \\
&= \textstyle\sum_{i=1}^n \sum_{t=2}^T \left(\Delta v_{it}\Delta\xi_{it} + \Delta v_{it}\Delta v_{it}^* - d_{it}\right) \\
&\equiv \textstyle\sum_{i=1}^n g_{2i},
\end{aligned}$$

where $\Delta\xi_t = \sum_{s=2}^T (\Phi_{st}^{u\prime} + \Phi_{ts}^l)\Delta v_s$, and $\Delta v_t^* = \sum_{s=2}^T \Phi_{ts}^d \Delta v_s$. Noting that $\Delta\xi_{it}$ is $\mathcal{G}_{n,i-1}$-measurable, it is easy to see that $\mathrm{E}(g_{2i}|\mathcal{G}_{n,i-1}) = 0$. Thus, $\{g_{2i}, \mathcal{G}_{n,i}\}$ form a M.D. sequence.

Finally, for the terms bilinear in Δv and $\Delta\mathbf{y}_1 = 1_{T-1} \otimes \Delta y_1$, we have,

$$\begin{aligned}
\Delta v'\Psi\mathbf{y}_1 &= \textstyle\sum_{t=2}^T \sum_{s=2}^T \Delta v_t'\Psi_{ts}\Delta y_1 \\
&= \textstyle\sum_{t=2}^T \Delta v_t'\big(\sum_{s=2}^T \Psi_{ts}\big)\Delta y_1 \\
&= \textstyle\sum_{t=2}^T \Delta v_t'\Psi_{t+}\Delta y_1 \\
&= \textstyle\Delta v_2'\Psi_{2+}\Delta y_1 + \sum_{t=3}^T \Delta v_t'\Psi_{t+}\Delta y_1 \\
&= \textstyle\Delta v_2'\Theta\Delta y_1^\circ + \sum_{t=3}^T \Delta v_t'\Delta y_{1t}^*,
\end{aligned}$$

where $\Delta y_1^\circ = B_{30}B_{10}\Delta y_1$ and $\Delta y_{1t}^* = \Psi_{t+}\Delta y_1$. The second term equals $\sum_{i=1}^n(\sum_{t=3}^T \Delta v_{it}\Delta y_{1ti}^*)$, which is the sum of n uncorrelated terms of mean zero, due to the fact that Δy_1 is independent of $\Delta v_t, t \geq 3$. The term $\Delta v_2'\Theta\Delta y_1^\circ$ needs some special attention. Noting that

$$\Delta y_1^\circ = B_{30}B_{10}\Delta y_1 = B_{30}B_{20}\Delta y_0 + B_{30}\Delta x_1\beta_0 + \Delta v_1, \tag{B.1}$$

and as Δy_0 is independent of $v_t, t \geq 1$ by Assumption A, $\mathrm{E}(\Delta v_2'\Theta\Delta y_1^\circ) = -\sigma_{v0}^2\mathrm{tr}(\Theta)$, and

$$\begin{aligned}
\Delta v_2'\Theta\Delta y_1^\circ - \mathrm{E}(\Delta v_2'\Theta\Delta y_1^\circ) &= \Delta v_2'(\Theta^u + \Theta^l + \Theta^d)\Delta y_1^\circ + \sigma_{v0}^2\mathrm{tr}(\Theta) \\
&= \Delta v_2'(\Theta^u + \Theta^l)\Delta y_1^\circ + \Delta v_2'\Theta^d\Delta y_1^\circ + \sigma_{v0}^2\mathrm{tr}(\Theta) \\
&= \textstyle\sum_{i=1}^n \Delta v_{2i}\Delta\zeta_i + \sum_{i=1}^n \Theta_{ii}(\Delta v_{2i}\Delta y_{1i}^\circ + \sigma_{v0}^2),
\end{aligned}$$

where $\{\Delta\zeta_i\} = \Delta\zeta = (\Theta^u + \Theta^l)\Delta y_1^\circ$. As $\Delta\zeta_i$ is measurable w.r.t. $\mathcal{F}_{n,i-1}$ and $\{\Delta v_{1,i+1}, \ldots \Delta v_{1,n}\}$, the first term is the sum of a M.D. sequence. The second term is easily seen to be the sum of n uncorrelated terms by (B.1). It follows that $\Delta v'\Psi\Delta\mathbf{y}_1 - \mathrm{E}(\Delta v'\Psi\Delta\mathbf{y}_1) = \sum_{i=1}^n g_{3i}$, where

$$g_{3i} = \Delta v_{2i}\Delta\zeta_i + \Theta_{ii}(\Delta v_{2i}\Delta y_{1i}^\circ + \sigma_{v0}^2) + \textstyle\sum_{t=3}^T \Delta v_{it}\Delta y_{1it}^*.$$

It is easy to see that $\mathrm{E}(g_{3i}|\mathcal{F}_{n,i-1}) = 0$. Hence, $\{g_{3i}, \mathcal{F}_{n,i}\}$ form a M.D. sequence. Finally, it is easy to verify that $\mathrm{E}[(g_{1i}', g_{2i}, g_{3i})|\mathcal{F}_{n,i-1}] = 0$. Hence, $\{(g_{1i}', g_{2i}, g_{3i})', \mathcal{F}_{n,i}\}$ form a vector M.D. sequence. ■

Appendix C: Proofs of Theorems 3.1-3.3

In proving the theorems, the following matrix results are used: (i) the eigenvalues of a projection matrix are either 0 or 1; (ii) the eigenvalues of a positive definite (p.d.) matrix are strictly positive; (iii) $\gamma_{\min}(A)\text{tr}(B) \le \text{tr}(AB) \le \gamma_{\max}(A)\text{tr}(B)$ for symmetric matrix A and positive semidefinite (p.s.d.) matrix B; (iv) $\gamma_{\max}(A+B) \le \gamma_{\max}(A) + \gamma_{\max}(B)$ for symmetric matrices A and B; and (v) $\gamma_{\max}(AB) \le \gamma_{\max}(A)\gamma_{\max}(B)$ for p.s.d. matrices A and B. See, e.g, Bernstein (2009).

Proof of Theorem 3.1: From (3.17) and (3.20), we have

$$S^{*c}_{\texttt{STLE}}(\delta) - \bar{S}^{*c}_{\texttt{STLE}}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{v,\texttt{M}}(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} - \frac{1}{\bar{\sigma}^2_{v,\texttt{M}}(\delta)}\text{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\Delta Y_{-1}], \\ \frac{1}{\hat{\sigma}^2_{v,\texttt{M}}(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y - \frac{1}{\bar{\sigma}^2_{v,\texttt{M}}(\delta)}\text{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y], \\ \frac{1}{\hat{\sigma}^2_{v,\texttt{M}}(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} - \frac{1}{\bar{\sigma}^2_{v,\texttt{M}}(\delta)}\text{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}], \\ \frac{1}{\hat{\sigma}^2_{v,\texttt{M}}(\delta)}\Delta\hat{u}(\delta)'\Upsilon\Delta\hat{u}(\delta) - \frac{1}{\bar{\sigma}^2_{v,\texttt{M}}(\delta)}\text{E}[\Delta\bar{u}(\delta)'\Upsilon\Delta\bar{u}(\delta)], \end{cases}$$

where $\Upsilon = \frac{1}{2}(C^{-1} \otimes A_3)$. With Assumption G, consistency of $\hat{\delta}_{\texttt{M}}$ follows from:

- (**a**) $\inf_{\delta\in\Delta}\bar{\sigma}^2_{v,\texttt{M}}(\delta)$ is bounded away from zero,
- (**b**) $\sup_{\delta\in\Delta}\big|\hat{\sigma}^2_{v,\texttt{M}}(\delta) - \bar{\sigma}^2_{v,\texttt{M}}(\delta)\big| = o_p(1)$,
- (**c**) $\sup_{\delta\in\Delta}\frac{1}{n(T-1)}\big|\Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} - \text{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\Delta Y_{-1}]\big| = o_p(1)$,
- (**d**) $\sup_{\delta\in\Delta}\frac{1}{n(T-1)}\big|\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y - \text{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y]\big| = o_p(1)$,
- (**e**) $\sup_{\delta\in\Delta}\frac{1}{n(T-1)}\big|\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} - \text{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}]\big| = o_p(1)$,
- (**f**) $\sup_{\delta\in\Delta}\frac{1}{n(T-1)}\big|\Delta\hat{u}(\delta)'\Upsilon\Delta\hat{u}(\delta) - \text{E}[\Delta\bar{u}(\delta)'\Upsilon\Delta\bar{u}(\delta)]\big| = o_p(1)$.

Proof of (a). By $\Delta\bar{u}^*(\delta) = \mathbf{M}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1}) + \mathbf{P}(\mathbf{B}_1^*\Delta Y^\circ - \mathbf{B}_2^*\Delta Y^\circ_{-1})$ given in (3.21), and the fact that the two projection matrices $\mathbf{M}$ and $\mathbf{P}$ are orthogonal to each other, we have,

$$\begin{aligned} \bar{\sigma}^2_{v,\texttt{M}}(\delta) &= \tfrac{1}{n(T-1)}\text{E}[\Delta\bar{u}^{*\prime}(\delta)\Delta\bar{u}^*(\delta)] \\ &= \tfrac{1}{n(T-1)}\text{E}[(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})'\mathbf{M}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})] \\ &\quad + \tfrac{1}{n(T-1)}\text{E}[(\mathbf{B}_1^*\Delta Y^\circ - \mathbf{B}_2^*\Delta Y^\circ_{-1})'\mathbf{P}(\mathbf{B}_1^*\Delta Y^\circ - \mathbf{B}_2^*\Delta Y^\circ_{-1})] \\ &= \tfrac{1}{n(T-1)}\text{tr}[\text{Var}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})] \\ &\quad + \tfrac{1}{n(T-1)}(\mathbf{B}_1^*\text{E}\Delta Y - \mathbf{B}_2^*\text{E}\Delta Y_{-1})'\mathbf{M}(\mathbf{B}_1^*\text{E}\Delta Y - \mathbf{B}_2^*\text{E}\Delta Y_{-1}). \end{aligned}$$

As $\mathbf{M}$ is p.s.d., the second term is nonnegative uniformly in $\delta \in \Delta$. The first term is $\frac{1}{n(T-1)}\text{tr}[\Omega^{-1}\text{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})] \ge \frac{1}{n(T-1)}\gamma_{\min}(C^{-1})\gamma_{\min}(B_3'B_3)\text{tr}[\text{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})] > c > 0$, uniformly in $\delta \in \Delta$, by the definition of the matrix C, Assumption E(iv) and the assumption given in the theorem. It follows that $\inf_{\delta\in\Delta}\bar{\sigma}^2_{v,\texttt{M}}(\delta) > c > 0$.

Proof of (b). Noting that $\Delta\hat{u}^*(\delta) = \mathbf{M}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})$, we have,

$$\hat{\sigma}^2_{v,\mathtt{M}}(\delta) = \tfrac{1}{n(T-1)}\Delta\hat{u}^{*\prime}(\delta)\Delta\hat{u}^*(\delta) = \tfrac{1}{n(T-1)}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})'\mathbf{M}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1}).$$

It follows that

$$\begin{aligned}
\hat{\sigma}^2_{v,\mathtt{M}}(\delta) - \bar{\sigma}^2_{v,\mathtt{M}}(\delta) &= \tfrac{1}{n(T-1)}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})'\mathbf{M}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1}) \\
&\quad -\tfrac{1}{n(T-1)}\mathrm{E}[(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})'\mathbf{M}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})] \\
&\quad -\tfrac{1}{n(T-1)}\mathrm{E}[(\mathbf{B}_1^*\Delta Y^\circ - \mathbf{B}_2^*\Delta Y_{-1}^\circ)'\mathbf{P}(\mathbf{B}_1^*\Delta Y^\circ - \mathbf{B}_2^*\Delta Y_{-1}^\circ)] \\
&= \tfrac{1}{n(T-1)}[\Delta Y'\mathbf{B}_1^{*\prime}\mathbf{M}\mathbf{B}_1^*\Delta Y - \mathrm{E}(\Delta Y'\mathbf{B}_1^{*\prime}\mathbf{M}\mathbf{B}_1^*\Delta Y)] \\
&\quad +\tfrac{1}{n(T-1)}[\Delta Y_{-1}'\mathbf{B}_2^{*\prime}\mathbf{M}\mathbf{B}_2^*\Delta Y_{-1} - \mathrm{E}(\Delta Y_{-1}'\mathbf{B}_2^{*\prime}\mathbf{M}\mathbf{B}_2^*\Delta Y_{-1})] \\
&\quad -\tfrac{2}{n(T-1)}[\Delta Y'\mathbf{B}_1^{*\prime}\mathbf{M}\mathbf{B}_2^*\Delta Y_{-1} - \mathrm{E}(\Delta Y'\mathbf{B}_1^{*\prime}\mathbf{M}\mathbf{B}_2^*\Delta Y_{-1})] \\
&\quad -\tfrac{1}{n(T-1)}\mathrm{E}[(\mathbf{B}_1^*\Delta Y^\circ - \mathbf{B}_2^*\Delta Y_{-1}^\circ)'\mathbf{P}(\mathbf{B}_1^*\Delta Y^\circ - \mathbf{B}_2^*\Delta Y_{-1}^\circ)] \\
&\equiv (Q_1 - \mathrm{E}Q_1) + (Q_2 - \mathrm{E}Q_2) - (Q_3 - \mathrm{E}Q_3) - \mathrm{E}Q_4.
\end{aligned}$$

The results follows if $Q_j - \mathrm{E}Q_j \xrightarrow{p} 0, j = 1, 2, 3$, and $\mathrm{E}Q_4 \longrightarrow 0$, uniformly in $\delta \in \Delta$.

First, by $\Delta Y = \mathbb{R}\ \Delta\mathbf{y}_1 + \boldsymbol{\eta} + \mathbb{S}\Delta v$ in Lemma 3.2, and letting $\mathbf{M}^* = \Omega^{-\frac{1}{2}}\mathbf{M}\Omega^{-\frac{1}{2}}$, we have,

$$\begin{aligned}
Q_1 &= \tfrac{1}{n(T-1)}\Delta Y'\mathbf{B}_1^{*\prime}\mathbf{M}\mathbf{B}_1^*\Delta Y \\
&= \tfrac{1}{n(T-1)}\big(\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_1\mathbb{R}\Delta\mathbf{y}_1 + \boldsymbol{\eta}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_1\boldsymbol{\eta} + \Delta v'\mathbb{S}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_1\mathbb{S}\Delta v \\
&\quad +2\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_1\boldsymbol{\eta} + \Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_1\mathbb{S}\Delta v + 2\boldsymbol{\eta}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_1\mathbb{S}\Delta v\big),
\end{aligned}$$

which leads to $Q_1 - \mathrm{E}Q_1 = \sum_{\ell=1}^5(Q_{1,\ell} - \mathrm{E}Q_{1,\ell})$, where $Q_{1,\ell}, \ell = 1, \ldots, 5$, denote the five stochastic terms of Q_1, and $\mathrm{E}Q_{1,5} = 0$.

Second, by $\Delta Y_{-1} = \mathbb{R}_{-1}\Delta\Delta\mathbf{y}_1 + \boldsymbol{\eta}_{-1} + \mathbb{S}_{-1}\Delta v$ given in Lemma 3.2, we have,

$$\begin{aligned}
Q_2 &= \tfrac{1}{n(T-1)}\Delta Y_{-1}'\mathbf{B}_2^{*\prime}\mathbf{M}\mathbf{B}_2^*\Delta Y_{-1} \\
&= \tfrac{1}{n(T-1)}\big(\Delta\mathbf{y}_1'\mathbb{R}_{-1}'\mathbf{B}_2'\mathbf{M}^*\mathbf{B}_2\mathbb{R}_{-1}\Delta\mathbf{y}_1 + \boldsymbol{\eta}_{-1}'\mathbf{B}_2'\mathbf{M}^*\mathbf{B}_2\boldsymbol{\eta}_{-1} + \Delta v'\mathbb{S}_{-1}'\mathbf{B}_2'\mathbf{M}^*\mathbf{B}_2\mathbb{S}_{-1}\Delta v \\
&\quad +2\Delta\mathbf{y}_1'\mathbb{R}_{-1}'\mathbf{B}_2'\mathbf{M}^*\mathbf{B}_2\boldsymbol{\eta}_{-1} + 2\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_2'\mathbf{M}^*\mathbf{B}_2\mathbb{S}_{-1}\Delta v + 2\boldsymbol{\eta}_{-1}'\mathbf{B}_2'\mathbf{M}^*\mathbf{B}_2\mathbb{S}_{-1}\Delta v\big),
\end{aligned}$$

leading to $Q_2 - \mathrm{E}Q_2 = \sum_{\ell=1}^5(Q_{2,\ell} - \mathrm{E}Q_{2,\ell})$, where $Q_{2,\ell}, \ell = 1, \ldots, 5$, denote the five stochastic terms of Q_2, and $\mathrm{E}Q_{2,5} = 0$.

Third, by both identities given in Lemma 3.2, we have

$$\begin{aligned}
Q_3 &= \tfrac{1}{n(T-1)}\Delta Y'\mathbf{B}_1^{*\prime}\mathbf{M}\mathbf{B}_2^*\Delta Y_{-1} \\
&= \tfrac{1}{n(T-1)}\big(\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_2\mathbb{R}_{-1}\Delta\mathbf{y}_1 + \boldsymbol{\eta}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_2\boldsymbol{\eta}_{-1} + \Delta v'\mathbb{S}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_2\mathbb{S}_{-1}\Delta v \\
&\quad +\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_2\boldsymbol{\eta}_{-1} + \boldsymbol{\eta}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_2\mathbb{R}_{-1}\Delta\mathbf{y}_1 + \Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_2\mathbb{S}_{-1}\Delta v \\
&\quad +\Delta v'\mathbb{S}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_2\mathbb{R}_{-1}\Delta\mathbf{y}_1' + \boldsymbol{\eta}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_2\mathbb{S}_{-1}\Delta v + \Delta v'\mathbb{S}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_2\boldsymbol{\eta}_{-1}\big),
\end{aligned}$$

leading to $Q_3 - \mathrm{E}Q_3 = \sum_{\ell=1}^{8}(Q_{3,\ell} - \mathrm{E}Q_{3,\ell})$, where $Q_{3,\ell}$, $\ell = 1,\ldots,8$, denote the eight random terms in Q_3 and the last two terms have expectations zero.

Thus, Q_k, $k = 1,2,3$, are decomposed into sums of terms of the forms: $\frac{1}{n(T-1)}\Delta\mathbf{y}_1'\Phi\Delta\mathbf{y}_1$, $\frac{1}{n(T-1)}\Delta v'\Pi\Delta v$, $\frac{1}{n(T-1)}\Delta\mathbf{y}_1'\Psi\Delta v$, $\frac{1}{n(T-1)}\Delta\mathbf{y}_1'\phi$, and $\frac{1}{n(T-1)}\Delta v'\psi$, where the matrices Φ, Π and ξ, and the vectors ϕ and ξ are defined in terms of $\mathbb{R}$, $\mathbb{R}_{-1}$, $\mathbb{S}$, $\mathbb{S}_{-1}$, $\boldsymbol{\eta}$, $\boldsymbol{\eta}_{-1}$, $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{M}^*$. Note that $\mathbb{R}$, $\mathbb{R}_{-1}$, $\mathbb{S}$, $\mathbb{S}_{-1}$, $\boldsymbol{\eta}$ and $\boldsymbol{\eta}_{-1}$ depend on true parameter values, whereas $\mathbf{B}_1$ depends on λ_1, $\mathbf{B}_2$ depends on ρ and λ_2, and $\mathbf{M}^*$ depends on λ_3.

For the terms quadratic in $\Delta\mathbf{y}_1$, they can be written as $\frac{1}{n}\Delta y_1'\Phi_{++}(\delta)\Delta y_1$ where $\Phi_{++}(\delta) = \sum_t\sum_s \Phi_{t,s}(\delta)$. It can easily be seen by Lemma A.1 and Lemma A.3 that for each $\delta \in \Delta$, $\Phi_{t,s}(\delta)$ are uniformly bounded in either row or column sums. The pointwise convergence of $\frac{1}{n}[\Delta y_1'\Phi_{++}(\delta)\Delta y_1 - \mathrm{E}(\Delta y_1'\Phi_{++}(\delta)\Delta y_1)]$ thus follows from Assumption F(ii). For the terms quadratic in Δv, they can be written as $\frac{1}{n(T-1)}\sum_{t=1}^{T}\sum_{s=1}^{T} v_t'\Pi_{ts}v_s$. The pointwise convergence of $\frac{1}{n}[v_t'\Pi_{ts}v_s - \mathrm{E}(v_t'\Pi_{ts}v_s)]$ follows from Lemma A.4 (v), for each $t,s = 1,\ldots,T$. The pointwise convergence of $\frac{1}{n(T-1)}[\Delta\mathbf{y}_1'\Psi\Delta v - \mathrm{E}(\Delta\mathbf{y}_1'\Psi\Delta v)]$ follows by writing $\Delta\mathbf{y}_1'\Psi\Delta v = \sum_s \Delta y_1\Psi_{+s}\Delta v_s$ and then applying Lemma A.4 (vii) and Assumption F(iii). The pointwise convergence of $\frac{1}{n(T-1)}[\Delta\mathbf{y}_1'\phi - \mathrm{E}(\Delta\mathbf{y}_1'\phi)$ follows from Assumption F(ii), and of $\frac{1}{n(T-1)}\Delta v'\psi$ from Chebyshev inequality. Thus, $Q_{k,\ell}(\delta) - \mathrm{E}Q_{k,\ell}(\delta) \xrightarrow{p} 0$, for each $\delta \in \Delta$, and all k and ℓ.

Now, for all the $Q_{k,\ell}(\delta)$ terms, let δ_1 and δ_2 be in $\boldsymbol{\Delta}$. We have by the mean value theorem:

$$Q_{k,\ell}(\delta_2) - Q_{k,\ell}(\delta_1) = \tfrac{\partial}{\partial\delta'}Q_{k,\ell}(\bar{\delta})(\delta_2 - \delta_1),$$

where $\bar{\delta}$ lies between δ_1 and δ_2 elementwise. Note that $Q_{k,\ell}(\delta)$ is linear or quadratic in ρ, λ_1 and λ_2, and thus the corresponding partial derivatives takes simple form. It is easy to show that $\sup_{\delta\in\boldsymbol{\Delta}} |\frac{\partial}{\partial\omega}Q_{k,\ell}(\delta)| = O_p(1)$, for $\omega = \rho, \lambda_1, \lambda_2$. For $\frac{\partial}{\partial\lambda_3}Q_{k,\ell}(\delta)$, note that only the matrix M^* involves λ_3. Some algebra leads to the following simple expression for its derivative:

$$\tfrac{d}{d\lambda_3}M^* = \mathbf{M}^*\Omega\dot{\Omega}^{-1}\Omega\mathbf{M}^*,$$

where $\dot{\Omega}^{-1} = \frac{d}{d\lambda_3}\Omega^{-1} = C^{-1}\otimes A_3$. Thus, the results $\sup_{\delta\in\boldsymbol{\Delta}} |\frac{\partial}{\partial\lambda_3}Q_{k,\ell}(\delta)| = O_p(1)$ can be easily proved for all the $Q_{k,\ell}(\delta)$ quantities. For example, for $Q_{1,1}(\delta)$, noting that $\gamma_{\max}(\mathbf{M}) = 1$,

$$\begin{aligned}
\sup_{\delta\in\boldsymbol{\Delta}} |Q_{1,1}(\delta)| &= \sup_{\delta\in\boldsymbol{\Delta}} |\tfrac{1}{n(T-1)}\tfrac{\partial}{\partial\lambda_3}\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_1\mathbb{R}\Delta\mathbf{y}_1| \\
&= \sup_{\delta\in\boldsymbol{\Delta}} \tfrac{1}{n(T-1)}|\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\Omega\dot{\Omega}^{-1}\Omega\mathbf{M}^*\mathbf{B}_1\mathbb{R}\Delta\mathbf{y}_1| \\
&\leq \sup_{\delta\in\boldsymbol{\Delta}} \tfrac{1}{n(T-1)}|\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\dot{\Omega}^{-1}\mathbf{B}_1\mathbb{R}\Delta\mathbf{y}_1| \\
&\leq \gamma_{\max}(\dot{\Omega}^{-1})\gamma_{\max}(\mathbf{B}_1'\mathbf{B}_1)\tfrac{1}{n(T-1)}|\Delta\mathbf{y}_1'\mathbb{R}'\mathbb{R}\Delta\mathbf{y}_1| \\
&= O(1)\times O(1)\times O_p(1) = O_p(1), \text{ by Assumption F(i).}
\end{aligned}$$

It follows that $Q_{k,\ell}(\delta)$ are stochastic equicontinuous, and by Theorem 1 of Andrews (1992) $Q_{k,\ell}(\delta) - \mathrm{E}Q_{k,\ell}(\delta) \xrightarrow{p} 0$, uniformly in $\delta \in \boldsymbol{\Delta}$. Thus, $Q_k(\delta) - \mathrm{E}Q_k(\delta) \xrightarrow{p} 0$, uniformly in $\delta \in \Delta$, $k = 1,2,3$. It left to show that $\mathrm{E}Q_4(\delta) \to 0$, uniformly in $\delta \in \Delta$. We have

$$
\begin{aligned}
\mathrm{E}Q_4 &= \tfrac{1}{n(T-1)}\mathrm{E}[(\mathbf{B}_1^*\Delta Y^\circ - \mathbf{B}_2^*\Delta Y_{-1}^\circ)'\mathbf{P}(\mathbf{B}_1^*\Delta Y^\circ - \mathbf{B}_2^*\Delta Y_{-1}^\circ)] \\
&= \tfrac{1}{n(T-1)}\mathrm{tr}[\Omega^{-1}\Delta X(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\mathrm{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})] \\
&\le \tfrac{1}{n(T-1)}\gamma_{\max}(\Omega^{-2})\gamma_{\min}^{-1}(\Delta X'\Omega^{-1}\Delta X)\ \mathrm{tr}[\Delta X'\mathrm{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})\Delta X] \\
&= \tfrac{1}{n(T-1)}\gamma_{\max}(\Omega^{-2})\gamma_{\min}^{-1}\big(\tfrac{\Delta X'\Omega^{-1}\Delta X}{n(T-1)}\big)\ \tfrac{1}{n(T-1)}\mathrm{tr}[\Delta X'\mathrm{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})\Delta X].
\end{aligned}
$$

As $\Omega^{-1} = C^{-1} \otimes B_3'B_3$, we have by the matrix C defined at the beginning of Section 3.1 and Assumption E(iv), $0 < \underline{c}_w \le \inf_{\lambda_3\in\Lambda_3}\gamma_{\min}(\Omega^{-1}) \le \sup_{\lambda_3\in\Lambda_3}\gamma_{\min}(\Omega^{-1}) \le \bar{c}_w < \infty$. By Assumption D, we have, $0 < \underline{c}_x \le \inf_{\lambda_3\in\Lambda_3}\gamma_{\min}(\Omega^{-1})\gamma_{\min}\big(\frac{\Delta X'\Delta X}{n(T-1)}\big) \le \gamma_{\min}\big(\frac{\Delta X'\Omega^{-1})\Delta X}{n(T-1)}\big) \le \gamma_{\max}\big(\frac{\Delta X'\Omega^{-1})\Delta X}{n(T-1)}\big) \le \sup_{\lambda_3\in\Lambda_3}\gamma_{\max}(\Omega^{-1})\gamma_{\max}\big(\frac{\Delta X'\Delta X}{n(T-1)}\big) \le \bar{c}_x < \infty$. It follows that

$$
\begin{aligned}
\mathrm{E}Q_4 &\le \tfrac{1}{n(T-1)}\bar{c}_w^2\underline{c}_x\tfrac{1}{n(T-1)}\mathrm{tr}[\Delta X'\mathrm{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})\Delta X] \\
&\le \tfrac{1}{n(T-1)}\bar{c}_w^2\underline{c}_x\bar{c}_y\ \tfrac{1}{n(T-1)}\mathrm{tr}[\Delta X'\Delta X], \text{ by the assumption in Theorem 3.1} \\
&= O(n^{-1}), \text{ by Assumption D.}
\end{aligned}
$$

Hence, $\hat{\sigma}_{v,\mathtt{M}}^2(\delta) - \bar{\sigma}_{v,\mathtt{M}}^2(\delta) \xrightarrow{p} 0$, uniformly in $\delta \in \Delta$, completing the proof of **(b)**.

Proofs of (c)-(f). By the expressions of $\Delta\hat{u}(\delta)$ and $\Delta\bar{u}(\delta)$ given earlier, we have,

$$
\begin{aligned}
&\Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} - \mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\Delta Y_{-1}] \\
=\ &[\Delta Y'\mathbf{B}_1'\mathbf{M}^*\Delta Y_{-1} - \mathrm{E}(\Delta Y'\mathbf{B}_1'\mathbf{M}^*\Delta Y_{-1})] - [\Delta Y_{-1}'\mathbf{B}_2'\mathbf{M}^*\Delta Y_{-1} - \mathrm{E}(\Delta Y_{-1}'\mathbf{B}_2'\mathbf{M}^*\Delta Y_{-1})] \\
&-\mathrm{E}(\Delta Y^{\circ\prime}\mathbf{B}_1'\mathbf{P}^*\Delta Y_{-1}^\circ) + \mathrm{E}(\Delta Y_{-1}^{\circ\prime}\mathbf{B}_2'\mathbf{P}^*\Delta Y_{-1}^\circ);
\end{aligned}
$$

$$
\begin{aligned}
&\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y - \mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y] \\
=\ &\Delta Y'\mathbf{B}_1'\mathbf{M}^*\mathbf{W}_1\Delta Y - \mathrm{E}(\Delta Y'\mathbf{B}_1'\mathbf{M}^*\mathbf{W}_1\Delta Y) - \Delta Y_{-1}'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\Delta Y \\
&-\mathrm{E}(\Delta Y_{-1}'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\Delta Y] - \mathrm{E}(\Delta Y^{\circ\prime}\mathbf{B}_1'\mathbf{P}^*\mathbf{W}_1\Delta Y^\circ) + \mathrm{E}(\Delta Y_{-1}^{\circ\prime}\mathbf{B}_2'\mathbf{P}^*\mathbf{W}_1\Delta Y^\circ);
\end{aligned}
$$

$$
\begin{aligned}
&\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} - \mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}] \\
=\ &\Delta Y'\mathbf{B}_1'\mathbf{M}^*\mathbf{W}_2\Delta Y_{-1} - \mathrm{E}(\Delta Y'\mathbf{B}_1'\mathbf{M}^*\mathbf{W}_2\Delta Y_{-1}) - \Delta Y_{-1}'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_2\Delta Y_{-1} \\
&-\mathrm{E}(\Delta Y_{-1}'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_2\Delta Y_{-1}] - \mathrm{E}(\Delta Y^{\circ\prime}\mathbf{B}_1'\mathbf{P}^*\mathbf{W}_2\Delta Y_{-1}^\circ) + \mathrm{E}(\Delta Y_{-1}^{\circ\prime}\mathbf{B}_2'\mathbf{P}^*\mathbf{W}_2\Delta Y_{-1}^\circ);\ \text{and}
\end{aligned}
$$

$$
\begin{aligned}
&\Delta\hat{u}(\delta)'\Upsilon\Delta\hat{u}(\delta) - \mathrm{E}[\Delta\bar{u}(\delta)' \\
=\ &\Delta Y'\mathbf{B}_1'\mathbf{M}^{\circ\prime}\Upsilon\mathbf{M}^\circ\mathbf{B}_1\Delta Y - \mathrm{E}(\Delta Y'\mathbf{B}_1'\mathbf{M}^{\circ\prime}\Upsilon\mathbf{M}^\circ\mathbf{B}_1\Delta Y) \\
&+\Delta Y_{-1}'\mathbf{B}_2'\mathbf{M}^{\circ\prime}\Upsilon\mathbf{M}^\circ\mathbf{B}_2\Delta Y_{-1} - \mathrm{E}(\Delta Y_{-1}'\mathbf{B}_2'\mathbf{M}^{\circ\prime}\Upsilon\mathbf{M}^\circ\mathbf{B}_2\Delta Y_{-1}) \\
&-2\Delta Y'\mathbf{B}_1'\mathbf{M}^{\circ\prime}\Upsilon\mathbf{M}^\circ\mathbf{B}_2\Delta Y_{-1} + 2\mathrm{E}(\Delta Y'\mathbf{B}_1'\mathbf{M}^{\circ\prime}\Upsilon\mathbf{M}^\circ\mathbf{B}_2\Delta Y_{-1}) \\
&+2\mathrm{E}[(\mathbf{B}_1\Delta Y^\circ - \mathbf{B}_2\Delta Y_{-1}^\circ)'\mathbf{M}^{\circ\prime}\Upsilon\mathbf{P}^\circ(\mathbf{B}_1\Delta Y^\circ - \mathbf{B}_2\Delta Y_{-1}^\circ)] \\
&+2\mathrm{E}[(\mathbf{B}_1\Delta Y^\circ - \mathbf{B}_2\Delta Y_{-1}^\circ)'\mathbf{P}^{\circ\prime}\Upsilon\mathbf{P}^\circ(\mathbf{B}_1\Delta Y^\circ - \mathbf{B}_2\Delta Y_{-1}^\circ)],
\end{aligned}
$$

where $\mathbf{M}^\circ = \mathbf{M}\Omega^{-\frac{1}{2}}$. By Lemma 3.2, all the quantities involving ΔY and ΔY_{-1} can be decomposed into sums of quadratic, bilinear and linear forms in $\Delta\mathbf{y}_1$ and/or Δv, and all the quantities involving ΔY° and ΔY_{-1}° can be handled in a similar manner as for Q_4 in **(b)**. The rest of the proof proceeds in a similar manner as for the proof of **(b)**. ∎

Proof of Theorem 3.2: We have by the mean value theorem,

$$0 = \tfrac{1}{\sqrt{n(T-1)}} S^*_{\mathtt{STLE}}(\hat{\psi}_{\mathtt{STLE}}) = \tfrac{1}{\sqrt{n(T-1)}} S^*_{\mathtt{STLE}}(\psi_0) + \big[\tfrac{1}{n(T-1)} \tfrac{\partial}{\partial \psi'} S^*_{\mathtt{STLE}}(\bar{\psi})\big] \sqrt{n(T-1)}(\hat{\psi}_{\mathtt{M}} - \psi_0),$$

where $\bar{\psi}$ lies elementwise between $\hat{\psi}_{\mathtt{M}}$ and ψ_0. The result of the theorem follows if

- **(a)** $\frac{1}{\sqrt{n(T-1)}} S^*_{\mathtt{STLE}}(\psi_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty} \Gamma^*_{\mathtt{STLE}}(\psi_0)\big]$,
- **(b)** $\frac{1}{n(T-1)} \big[\frac{\partial}{\partial \psi'} S^*_{\mathtt{STLE}}(\bar{\psi}) - \frac{\partial}{\partial \psi'} S^*_{\mathtt{STLE}}(\psi_0)\big] \xrightarrow{p} 0$, and
- **(c)** $\frac{1}{n(T-1)} \big[\frac{\partial}{\partial \psi'} S^*_{\mathtt{STLE}}(\psi_0) - \mathrm{E}\big(\frac{\partial}{\partial \psi'} S^*_{\mathtt{STLE}}(\psi_0)\big)\big] \xrightarrow{p} 0$.

Proof of (a). From (3.24), we see that $S^*_{\mathtt{STLE}}(\psi_0)$ consists of three types of elements: $\Pi'\Delta v$, $\Delta v'\Phi\Delta v$ and $\Delta v'\Psi\Delta \mathbf{y}_1$, which can be written as

$$\Pi'\Delta v = \textstyle\sum_{t=1}^T \Pi_t^{*\prime} v_t, \;\; \Delta v \Phi \Delta v = \sum_{t=1}^T \sum_{s=1}^T v_t' \Phi^*_{ts} v_s, \;\text{ and }\; \Delta v'\Psi\Delta\mathbf{y}_1 = \sum_{t=1}^T v_t' \Psi_t^* \Delta y_1,$$

where Π_t^*, Φ_{ts}^* and Ψ_t^* are formed by the elements of the partitioned Π, Φ and Ψ, respectively. By (2.1), $y_1 = B_{10}^{-1} B_{20} y_0 + \eta_1 + B_{10}^{-1} B_{30}^{-1} v_1$, leading to $\sum_{t=1}^T v_t' \Psi_t^* \Delta y_1 = \sum_{t=1}^T v_t' \Psi_t^{**} y_0 + \sum_{t=1}^T v_t' \Psi_t^{*+} v_1$, for suitably defined non-stochastic quantities η_1, Ψ_t^{**} and Ψ_t^{*+}. These show that, for every non-zero $(p+5)\times 1$ vector of constants c, $c' S^*_{\mathtt{STLE}}(\psi_0)$ can be expressed as

$$c' S^*_{\mathtt{STLE}}(\psi_0) = \sum_{t=1}^T \sum_{s=1}^T v_t' A_{ts} v_s + \sum_{t=1}^T v_t' B_t v_1 + \sum_{t=1}^T v_t' g(y_0),$$

for suitably defined non-stochastic matrices A_{ts} and B_t, and the function $g(y_0)$ linear in y_0. As, $\{y_0, v_1, \ldots, v_T\}$ are independent, the asymptotic normality of $\frac{1}{\sqrt{n(T-1)}} c' S^*_{\mathtt{STLE}}(\psi_0)$ follows from Lemma A.5. Finally, the Cramér-Wold devise leads to the joint asymptotic normality.

Proof of (b). The Hessian matrix, $H^*_{\mathtt{STLE}}(\psi) = \frac{\partial}{\partial \psi'} S^*_{\mathtt{STLE}}(\psi)$, has the elements:

$$\begin{aligned}
H^*_{\beta\beta} &= -\tfrac{1}{\sigma_v^2} \Delta X' \Omega^{-1} \Delta X, & H^*_{\sigma_v^2\sigma_v^2} &= -\tfrac{1}{\sigma_v^6} \Delta u(\theta)' \Omega^{-1} \Delta u(\theta) + \tfrac{n(T-1)}{2\sigma_v^4},\\
H^*_{\beta\sigma_v^2} &= -\tfrac{1}{\sigma_v^4} \Delta X' \Omega^{-1} \Delta u(\theta), & H^*_{\sigma_v^2\lambda_2} &= -\tfrac{1}{\sigma_v^4} \Delta Y_{-1}' \mathbf{W}_2' \Omega^{-1} \Delta u(\theta),\\
H^*_{\beta\rho} &= -\tfrac{1}{\sigma_v^2} \Delta X' \Omega^{-1} \Delta Y_{-1}, & H^*_{\sigma_v^2\lambda_3} &= \tfrac{1}{2\sigma_v^4} \Delta u(\theta)' \dot{\Omega}^{-1} \Delta u(\theta),\\
H^*_{\beta\lambda_1} &= -\tfrac{1}{\sigma_v^2} \Delta X' \Omega^{-1} \mathbf{W}_1 \Delta Y, & H^*_{\rho\rho} &= -\tfrac{1}{\sigma_v^2} \Delta Y_{-1}' \Omega^{-1} \Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\rho}),\\
H^*_{\beta\lambda_2} &= -\tfrac{1}{\sigma_v^2} \Delta X' \Omega^{-1} \mathbf{W}_2 \Delta Y_{-1}, & H^*_{\rho\lambda_1} &= -\tfrac{1}{\sigma_v^2} \Delta Y_{-1}' \Omega^{-1} \mathbf{W}_1 \Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\lambda_1}),\\
H^*_{\beta\lambda_3} &= \tfrac{1}{\sigma_v^2} \Delta X' \dot{\Omega}^{-1} \Delta u(\theta), & H^*_{\rho\lambda_2} &= -\tfrac{1}{\sigma_v^2} \Delta Y_{-1}' \Omega^{-1} \mathbf{W}_2 \Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\lambda_2}),\\
H^*_{\sigma_v^2\rho} &= -\tfrac{1}{\sigma_v^4} \Delta Y_{-1}' \Omega^{-1} \Delta u(\theta), & H^*_{\lambda_1\lambda_1} &= -\tfrac{1}{\sigma_v^2} \Delta Y' \mathbf{W}_1' \Omega^{-1} \mathbf{W}_1 \Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}\mathbf{W}_1),\\
H^*_{\sigma_v^2\lambda_1} &= -\tfrac{1}{\sigma_v^4} \Delta Y' \mathbf{W}_1' \Omega^{-1} \Delta u(\theta), & H^*_{\lambda_1\lambda_2} &= -\tfrac{1}{\sigma_v^2} \Delta Y' \mathbf{W}_1' \Omega^{-1} \mathbf{W}_2 \Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_2}\mathbf{W}_1),\\
H^*_{\rho\lambda_3} &= \tfrac{1}{\sigma_v^2} \Delta Y_{-1}' \dot{\Omega}^{-1} \Delta u(\theta), & H^*_{\lambda_2\lambda_2} &= -\tfrac{1}{\sigma_v^2} \Delta Y_{-1}' \mathbf{W}_2' \Omega^{-1} \mathbf{W}_2 \Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\lambda_2}\mathbf{W}_2),\\
H^*_{\lambda_1\lambda_3} &= \tfrac{1}{\sigma_v^2} \Delta Y' \mathbf{W}_1' \dot{\Omega}^{-1} \Delta u(\theta), & H^*_{\lambda_3\lambda_3} &= -\tfrac{1}{\sigma_v^2} \Delta u(\theta)' [C^{-1} \otimes (W_3' W_3)] \Delta u(\theta) - \mathrm{tr}(\mathbf{G}_3^2).\\
H^*_{\lambda_2\lambda_3} &= \tfrac{1}{\sigma_v^2} \Delta Y_{-1}' \mathbf{W}_2' \dot{\Omega}^{-1} \Delta u(\theta), & &
\end{aligned}$$

where $\dot{\Omega}^{-1} = \frac{\partial}{\partial \lambda_3} \Omega^{-1}$, $\mathbf{D}_{-1,\omega} = \frac{\partial}{\partial \omega} \mathbf{D}_{-1}$ and $\mathbf{D}_\omega = \frac{\partial}{\partial \omega} \mathbf{D}$, $\omega = \rho, \lambda_1, \lambda_2$, and $\mathbf{G}_3 = \mathbf{W}_3 \mathbf{B}_3^{-1}$.

It is easy to show that $\frac{1}{n(T-1)} H^*_{\mathtt{STLE}}(\psi_0) = O_p(1)$ by Lemma A.1 and the model assumptions. Thus, $\frac{1}{n(T-1)} H^*_{\mathtt{STLE}}(\bar{\psi}) = O_p(1)$ because $\bar{\psi} - \psi_0 = o_p(1)$ which is implied by $\hat{\psi}_{\mathtt{M}} \xrightarrow{p} \psi_0$. As $\bar{\sigma}^2 \xrightarrow{p} \sigma_0^2$, $\bar{\sigma}^{-r} = \sigma_0^{-r} + o_p(1), r = 2, 4, 6$. Noting that σ^r appears in $H^*_{\mathtt{STLE}}(\psi)$ multiplicatively, $\frac{1}{n(T-1)} H^*_{\mathtt{STLE}}(\bar{\psi}) = \frac{1}{n(T-1)} H^*_{\mathtt{STLE}}(\bar{\beta}, \sigma_0^2, \bar{\rho}, \bar{\lambda}) + o_p(1)$, i.e., replacing $\bar{\sigma}^2$ by σ_{v0}^2 results in an asymptotically negligible error. The proof of (**b**) is thus equivalent to the proof of

$$\tfrac{1}{n(T-1)} \left[H^*_{\mathtt{STLE}}(\bar{\beta}, \sigma_0^2, \bar{\rho}, \bar{\lambda}) - H^*_{\mathtt{STLE}}(\psi_0) \right] \xrightarrow{p} 0.$$

From $\Delta u(\theta) = \Delta u - (\lambda_1 - \lambda_{10}) \mathbf{W}_1 \Delta Y - (\rho - \rho_0) \Delta Y_{-1} - (\lambda_2 - \lambda_{20}) \mathbf{W}_2 \Delta Y_{-1} - \Delta X(\beta - \beta_0)$, $\Omega^{-1}(\lambda_3) - \Omega^{-1}(\lambda_{30}) = (\lambda_3^2 - \lambda_{30}^2) C^{-1} \otimes (W_3' W_3) - (\lambda_3 - \lambda_{30}) C^{-1} \otimes (W_3' + W_3)$, and $\dot{\Omega}^{-1} = C^{-1} \otimes (W_3' B_3 + B_3' W_3)$, we see that all the random elements of $H^*_{\mathtt{STLE}}(\psi)$ are linear, bilinear, or quadratic in ΔY, ΔY_{-1} or Δu, and linear or quadratic in β, ρ, and λ. This means that all the corresponding elements in $\frac{1}{n(T-1)} \left[H^*_{\mathtt{STLE}}(\bar{\beta}, \sigma_0^2, \bar{\rho}, \bar{\lambda}) - H^*_{\mathtt{STLE}}(\psi_0) \right]$ are linear, bilinear, or quadratic in ΔY, ΔY_{-1} or Δu, and linear, bilinear or quadratic in $\bar{\beta} - \beta_0, \bar{\rho} - \rho_0$, and $\bar{\lambda} - \lambda_0$, and thus are all $o_p(1)$ by the consistency of $\hat{\psi}_{\mathtt{M}}$, Lemma 3.2, Lemma A.1 and Assumption F. This can be easily seen for all the terms linear in ΔY or ΔY_{-1}, quadratic in ΔY or ΔY_{-1}, or bilinear in ΔY and ΔY_{-1}. For example, for the term corresponding to $H^*_{\lambda_1 \lambda_1}$, we have, by the consistency of $\hat{\lambda}_{\mathtt{M}}$, Lemma 3.2, Lemma A.1, and Assumption F,

$$\begin{aligned}
& \tfrac{1}{n(T-1)} [-\tfrac{1}{\sigma_v^2} \Delta Y' \mathbf{W}_1' \Omega^{-1}(\bar{\lambda}_3) \mathbf{W}_1 \Delta Y + \tfrac{1}{\sigma_v^2} \Delta Y' \mathbf{W}_1' \Omega_0^{-1}(\lambda_{30}) \mathbf{W}_1 \Delta Y] \\
= \; & \tfrac{1}{n(T-1)} \tfrac{1}{\sigma_{v0}^2} \Delta Y' \mathbf{W}_1' [-\Omega^{-1}(\bar{\lambda}_3) + \Omega_0^{-1}(\lambda_{30})] \mathbf{W}_1 \Delta Y] \\
= \; & (\bar{\lambda}_3 - \lambda_{30}) \tfrac{1}{n(T-1)\sigma_{v0}^2} \Delta Y' \mathbf{W}_1' [C^{-1} \otimes (W_3' + W_3)] \mathbf{W}_1 \Delta Y \\
& -(\bar{\lambda}_3^2 - \lambda_{30}^2) \tfrac{1}{n(T-1)\sigma_{v0}^2} \Delta Y' \mathbf{W}_1' [C^{-1} \otimes (W_3' W_3)] \mathbf{W}_1 \Delta Y \\
= \; & o_p(1) \times O_p(1) - o_p(1) \times O_p(1) = o_p(1).
\end{aligned}$$

Now, all the remaining terms involve $\Delta u(\theta)$. We have, for example,

$$\begin{aligned}
& H^*_{\sigma_v^2 \lambda_1}(\bar{\beta}, \sigma_0^2, \bar{\rho}, \bar{\lambda}) - H^*_{\sigma_v^2 \lambda_1}(\psi_0) \\
= \; & -\tfrac{1}{\sigma_{v0}^4} \Delta Y' \mathbf{W}_1' [\Omega^{-1}(\bar{\lambda}_3) \Delta u(\bar{\theta}) - \Omega_0^{-1} \Delta u] \\
= \; & -\tfrac{1}{\sigma_{v0}^4} \Delta Y' \mathbf{W}_1' \big\{ [\Omega_0^{-1} + (\bar{\lambda}_3^2 - \lambda_{30}^2) C^{-1} \otimes (W_3' W_3) - (\bar{\lambda}_3 - \lambda_{30}) C^{-1} \otimes (W_3' + W_3)] \\
& \times [\Delta u - (\bar{\lambda}_1 - \lambda_{10}) \mathbf{W}_1 \Delta Y - (\bar{\rho} - \rho_0) \Delta Y_{-1} - (\bar{\lambda}_2 - \lambda_{20}) \mathbf{W}_2 \Delta Y_{-1} - \Delta X(\bar{\beta} - \beta_0] - \Omega_0^{-1} \Delta u \big\},
\end{aligned}$$

from which one sees clearly that it is linear, bilinear or quadratic in ΔY, ΔY_{-1}, or Δu, and linear, bilinear or quadratic in $\bar{\beta} - \beta_0, \bar{\rho} - \rho_0$, and $\bar{\lambda} - \lambda_0$. The proof of

$$\tfrac{1}{n(T-1)} [H^*_{\sigma_v^2 \lambda_1}(\bar{\beta}, \sigma_0^2, \bar{\rho}, \bar{\lambda}) - H^*_{\sigma_v^2 \lambda_1}(\psi_0)] = o_p(1)$$

boils down to show that the quantities $\frac{1}{n(T-1)} \Delta Y' \mathbf{W}_1' \Omega_0^{-1} \mathbf{W}_1 \Delta Y$, $\frac{1}{n(T-1)} \Delta Y' \mathbf{W}_1' \Omega_0^{-1} \Delta Y_{-1}$, $\frac{1}{n(T-1)} \Delta Y' \mathbf{W}_1' \Omega_0^{-1} \Delta X$, etc., are all $O_p(1)$, which can be done easily by Lemma 3.2, Lemma A.1 and Assumption F. The proofs for the other terms involving $\Delta u(\theta)$ proceed in the same manner. It left to show that

(a) $\frac{1}{n(T-1)}[\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\rho}(\bar{\rho},\bar{\lambda}_1,\bar{\lambda}_2)) - \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\rho}(\rho_0,\lambda_{10},\lambda_{20}))] = o_p(1)$
(b) $\frac{1}{n(T-1)}[\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\lambda_1}(\bar{\rho},\bar{\lambda}_1,\bar{\lambda}_2)) - \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\lambda_1}(\rho_0,\lambda_{10},\lambda_{20}))] = o_p(1)$
(c) $\frac{1}{n(T-1)}[\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\lambda_2}(\bar{\rho},\bar{\lambda}_1,\bar{\lambda}_2)) - \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\lambda_2}(\rho_0,\lambda_{10},\lambda_{20}))] = o_p(1)$
(d) $\frac{1}{n(T-1)}[\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}(\bar{\rho},\bar{\lambda}_1,\bar{\lambda}_2)\mathbf{W}_1) - \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}(\rho_0,\lambda_{10},\lambda_{20})\mathbf{W}_1)] = o_p(1)$
(e) $\frac{1}{n(T-1)}[\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_2}(\bar{\rho},\bar{\lambda}_1,\bar{\lambda}_2)\mathbf{W}_1) - \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_2}(\rho_0,\lambda_{10},\lambda_{20})\mathbf{W}_1)] = o_p(1)$
(f) $\frac{1}{n(T-1)}[\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\lambda_2}(\bar{\rho},\bar{\lambda}_1,\bar{\lambda}_2)\mathbf{W}_2)) - \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1,\lambda_2}(\rho_0,\lambda_{10},\lambda_{20})\mathbf{W}_2))] = o_p(1)$
(g) $\frac{1}{n(T-1)}[\text{tr}(\mathbf{G}(\bar{\lambda}_3)^2) - \text{tr}(\mathbf{G}(\lambda_{30})^2)] = o_p(1)$.

It is easy to show the last result. By the mean value theorem, $\text{tr}(\mathbf{G}(\bar{\lambda}_3)^2) - \text{tr}(\mathbf{G}(\lambda_{30})^2) = 2(\bar{\lambda}_3 - \lambda_{30})\text{tr}(\mathbf{G}(\lambda_3^*)^3)$, where λ_3^* lies between $\bar{\lambda}_3$ and λ_{30}. By Lemmas A.1 and A.2, the elements of $\mathbf{G}(\lambda_3^*)^3$ is uniformly bounded. Thus, $\frac{1}{n(T-1)}\text{tr}(\mathbf{G}(\lambda_3^*)^3) = O_p(1)$, leading to (g). The proofs of (a)-(f) are similar, and some details are given for the most complicate case (d). Let $(\rho^*, \lambda_1^*, \lambda_2^*)$ be between $(\bar{\rho},\bar{\lambda}_1,\bar{\lambda}_2)$ and $(\rho_0,\lambda_{10},\lambda_{20})$. By the mean value theorem,

$$\begin{aligned}&\tfrac{1}{n(T-1)}[\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}(\bar{\rho},\bar{\lambda}_1,\bar{\lambda}_2)\mathbf{W}_1) - \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}(\rho_0,\lambda_{10},\lambda_{20})\mathbf{W}_1)]\\ =\ &\tfrac{\bar{\rho}-\rho_0}{n(T-1)}\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}^{\rho^*}\mathbf{W}_1) + \tfrac{\bar{\lambda}_1-\lambda_{10}}{n(T-1)}\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}^{\lambda_1^*}\mathbf{W}_1) + \tfrac{\bar{\lambda}_2-\lambda_{20}}{n(T-1)}\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}^{\lambda_2^*}\mathbf{W}_1),\end{aligned}$$

where $\mathbf{D}_{\lambda_1}^{\rho^*}$, $\mathbf{D}_{\lambda_1}^{\lambda_1^*}$ and $\mathbf{D}_{\lambda_1}^{\lambda_2^*}$ are the partial derivatives of $\mathbf{D}_{\lambda_1}$ evaluated at $(\rho^*, \lambda_1^*, \lambda_2^*)$. Consider, W.L.O.G., $T = 3$. Recall $B_1 = I_n - \lambda W_1$, $B_2 = \rho I_n + \lambda_2 W_2$ and $\mathcal{B} = B_2^{-1}B_2$. We have

$$\mathbf{D}(\rho,\lambda_1,\lambda_2) = \begin{pmatrix} B_1^{-1}B_2B_1^{-1}, & B_1^{-1} \\ (I_n - B_1^{-1}B_2)^2B_1^{-1}, & B_1^{-1}B_2B_1^{-1}\end{pmatrix}.$$

This shows that the elements of $\mathbf{D}_{\lambda_1}$ are the multiplications of the matrices W_1, B_1^{-1} and B_2. Subsequently, $\mathbf{D}_{\lambda_1}^{\rho}$, $\mathbf{D}_{\lambda_1}^{\lambda_1}$ and $\mathbf{D}_{\lambda_1}^{\lambda_2}$ have elements being the multiplications of the matrices W_1, W_2, $B_1^{-1}(\lambda_1)$, and $B_2(\rho,\lambda_2)$, and hence are uniformly bounded in the neighborhood of $(\rho_0,\lambda_{10},\lambda_{20})$ by Lemmas A.1 and A.2. Therefore, $\frac{1}{n(T-1)}\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}^{\rho^*}\mathbf{W}_1) = O_p(1)$, $\frac{1}{n(T-1)}\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}^{\lambda_1^*}\mathbf{W}_1) = O_p(1)$, and $\frac{1}{n(T-1)}\text{tr}(\mathbf{C}^{-1}\mathbf{D}_{\lambda_1}^{\lambda_2^*}\mathbf{W}_1) = O_p(1)$, leading to (d).

Proof of (c). First, for the terms involving only Δu (linear or quadratic), the results follows Lemma A.4(v)-(vi), noticing $\Delta u = \mathbf{B}_{30}^{-1}\mathbf{F}v$ where $\mathbf{F}v = \Delta v$. For example,

$$H^*_{\sigma_v^2\lambda_3}(\psi_0) - \text{E}[H^*_{\sigma_v^2\lambda_3}(\psi_0)] = \tfrac{1}{2\sigma_{v0}^4}[\Delta u'\dot{\Omega}_0^{-1}\Delta u - \text{E}(\Delta u'\dot{\Omega}_0^{-1}\Delta u)] = \tfrac{1}{2\sigma_{v0}^4}[v'\mathbf{A}v - \text{E}(v'\mathbf{A}v)],$$

where $\mathbf{A} = \mathbf{F}'\mathbf{B}_{30}'^{-1}\dot{\Omega}_0^{-1}\mathbf{B}_{30}^{-1}\mathbf{F}$, which is easily seen to be uniformly bounded in both row and column sums. Thus, Lemma A.4(v) leads to $\frac{1}{n(T-1)}\{H^*_{\sigma_v^2\lambda_3}(\psi_0) - \text{E}[H^*_{\sigma_v^2\lambda_3}(\psi_0)]\} = o_p(1)$.

Second, by Lemma 3.2 all the terms involving ΔY and ΔY_{-1} can be written as sums of the terms linear in $\Delta\mathbf{y}$, quadratic in $\Delta\mathbf{y}$, bilinear in $\Delta\mathbf{y}$ and Δv, or quadratic in Δv. Thus, the results follow by repeatedly applying Lemma A.1, Lemma A.4, and Assumption F. ■

Proof of Theorem 3.3: First, the result $\Sigma^*_{\mathtt{STLE}}(\hat{\psi}_{\mathtt{M}}) - \Sigma^*_{\mathtt{STLE}}(\psi_0) \xrightarrow{p} 0$ is implied by the result **(b)** in the proof of Theorem 3.2. The result $\frac{1}{n(T-1)}\sum_{i=1}^n[\hat{g}_i\hat{g}_i' - \mathrm{E}(g_i g_i')] \xrightarrow{p} 0$ follows from $\frac{1}{n(T-1)}\sum_{i=1}^n[\hat{g}_i\hat{g}_i' - g_i g_i'] \xrightarrow{p} 0$ and $\frac{1}{n(T-1)}\sum_{i=1}^n[g_i g_i' - \mathrm{E}(g_i g_i')] \xrightarrow{p} 0$. The proof of the former is straightforward by applying the mean value theorem. We focus on the proof of the latter result. As the elements of $S^*_{\mathtt{STLE}}(\psi_0)$ are mixtures of terms of the forms $\Pi'\Delta v = \sum_{i=1}^n g_{1i}$, $\Delta v'\Phi\Delta v = \sum_{i=1}^n g_{2i}$ and $\Delta v'\Psi\Delta\mathbf{y}_1 = \sum_{i=1}^n g_{2i}$, it suffices to show that

$$\frac{1}{n(T-1)}\sum_{i=1}^n[g_{ki}g_{ri}' - \mathrm{E}(g_{ki}g_{ri}')] = o_p(1),\ k,r = 1,2,3.$$

To facilitate the proof, the following *dot* notation is convenient: (a) for an $n(T-1)\times 1$ vector Δv with elements $\{\Delta v_{it}\}$ double indexed by $i = 1,\ldots,n$ for each $t = 2,\ldots,T$, $\Delta v_{\cdot t}$ is the subvector that contains all the elements with the same t, and $\Delta v_{i\cdot}$ is the subvector that picks up the elements with the same i; (b) for an $n(T-1)\times n(T-1)$ matrix Φ with elements $\{\Phi_{it,js},\ i,j = 1,\ldots,n;\ t,s = 2,\ldots,T\}$, where it is the double index for the rows and js the double index for the columns, $\Phi_{\cdot t,\cdot s}$ is the $n\times n$ submatrix corresponding to the (t,s) periods, $\Phi_{i\cdot,j\cdot}$ the $(T-1)\times(T-1)$ submatrix corresponding to the (i,j) units, $\Phi_{it,j\cdot}$ the $(T-1)\times 1$ subvector that picks up the element from the itth row corresponding to $s = 2,\ldots,T$.

With the vector dot notation, the $g_{ri}, r = 1,2,3$, defined in Lemma 3.3 can be written as $g_{1i} = \Pi_{i\cdot}'\Delta v_{i\cdot}$, $g_{2i} = \Delta v_{i\cdot}'\Delta\xi_{i\cdot} + \Delta v_{i\cdot}'\Delta v_{i\cdot}^* - 1_{T-1}'d_{i\cdot}$, and $g_{3i} = \Delta v_{2i}\Delta\zeta_i + \Theta_{ii}(\Delta v_{2i}\Delta y_{1i}^\circ + \sigma_{v0}^2) + \Delta v_{i-}'\Delta y_{1i-}^*$ where '$-$' plays the same role as '.' but corresponds to $t = 3,\ldots,T$. Note that under Assumptions D and E, one can easily see by Lemma A.1 that the elements of all the Π's, Φ's,, and Ψ's, defined in (3.24), are uniformly bounded. The proofs proceed by applying the weak law of large numbers (WLLN) for M.D. arrays, see, e.g., Davidson (1994, p. 299).

First, with $g_{1i} = \Pi_{i\cdot}'\Delta v_{i\cdot}$, $\frac{1}{n(T-1)}\sum_{i=1}^n g_{1i}g_{1i}' - \mathrm{E}(g_{1i}g_{1i}')] = \frac{1}{n(T-1)}\sum_{i=1}^n \Pi_{i\cdot}'(\Delta v_{i\cdot}\Delta v_{i\cdot}' - \sigma_{v0}^2 C)\Pi_{i\cdot} \equiv \frac{1}{n(T-1)}\sum_{i=1}^n U_{n,i}$, where C is defined below (3.2). Without loss of generality, assume U_{ni} is a scalar, as if not we can work on each element of it. Clearly, $\{U_{n,i}\}$ are independent, thus form a M.D. array. By Assumption B and using the fact that the elements of $\Pi_{i\cdot}$ are uniformly bounded, it is easy to show that $\mathrm{E}|U_{n,i}|^{1+\epsilon} \le K_u < \infty$, for $\epsilon > 0$. Thus, $\{U_{n,i}\}$ are uniformly integrable. With the constant coefficients $\frac{1}{n(T-1)}$ the other two conditions of WLLN for M.D. arrays of Davidson are satisfied. Thus, $\frac{1}{n(T-1)}\sum_{i=1}^n U_{n,i} \xrightarrow{p} 0$.

Second, with $g_{2i} = \Delta v_{i\cdot}'\Delta\xi_{i\cdot} + \Delta v_{i\cdot}'\Delta v_{i\cdot}^* - 1_{T-1}'d_{i\cdot}$, we have,

$$\begin{aligned}
& \tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[g_{2i}^2 - \mathrm{E}(g_{2i}^2)] \\
= \ & \tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(\Delta v_{i\cdot}'\Delta\xi_{i\cdot})^2 - \mathrm{E}((\Delta v_{i\cdot}'\Delta\xi_{i\cdot})^2)] \\
& +\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(\Delta v_{i\cdot}'\Delta v_{i\cdot}^*)^2 - \mathrm{E}((\Delta v_{i\cdot}'\Delta v_{i\cdot}^*)^2)] \\
& +\tfrac{2}{n(T-1)}\textstyle\sum_{i=1}^n(\Delta v_{i\cdot}'\Delta\xi_{i\cdot})(\Delta v_{i\cdot}'\Delta v_{i\cdot}^*) - \tfrac{2}{n(T-1)}\sum_{i=1}^n(1_{T-1}'d_{i\cdot})(\Delta v_{i\cdot}'\Delta\xi_{i\cdot}) \\
& -\tfrac{2}{n(T-1)}\textstyle\sum_{i=1}^n[(1_{T-1}'d_{i\cdot})(\Delta v_{i\cdot}'\Delta v_{i\cdot}^* - \mathrm{E}(\Delta v_{i\cdot}'\Delta v_{i\cdot}^*))] \equiv \sum_{r=1}^5 H_r.
\end{aligned}$$

Now, $H_1 = \frac{1}{n(T-1)}\sum_{i=1}^n[\Delta\xi_{i\cdot}'(\Delta v_{i\cdot}\Delta v_{i\cdot}' - \sigma_{vo}^2 C)\Delta\xi_{i\cdot}] + \frac{\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n[\Delta\xi_{i\cdot}'C\Delta\xi_{i\cdot} - \mathrm{E}(\Delta\xi_{i\cdot}'C\Delta\xi_{i\cdot})]$. For the first term, let $V_{n,i} = \Delta\xi_{i\cdot}'(\Delta v_{i\cdot}\Delta v_{i\cdot}' - \sigma_{vo}^2 C)\Delta\xi_{i\cdot}$. As $\Delta\xi_{i\cdot}$ is $\mathcal{G}_{n,i-1}$-measurable, $\mathrm{E}(V_{n,i}|\mathcal{G}_{n,i-1}) = 0$. Thus, $\{V_{n,i},\mathcal{G}_{n,i}\}$ form a M.D. array. It is easy to see that $\mathrm{E}|V_{n,i}^{1+\epsilon}| \le$

$K_v < \infty$, for some $\epsilon > 0$. Thus, $\{V_{n,i}\}$ is uniformly integrable. The other two conditions of the WLLN for M.D. arrays of Davidson are satisfied. Thus, $\frac{1}{n(T-1)}\sum_{i=1}^n V_{n,i} \xrightarrow{p} 0$.

For the second term of H_1, recall $\xi_t = \sum_{s=2}^T(\Phi_{ts}^{u\prime} + \Phi_{ts}^{\ell})\Delta v_s$. We have,

$$\Delta\xi_{it} = \sum_{s=2}^T\sum_{j=1}^{i-1}(\Phi_{jt,is} + \Phi_{it,js})\Delta v_{js} = \sum_{j=1}^{i-1}\sum_{s=2}^T(\Phi_{jt,is} + \Phi_{it,js})\Delta v_{js} = \sum_{j=1}^{i-1}\phi_{ijt}\Delta v_{j\cdot},$$

where $\phi_{ijt} = (\Phi_{ji,t\cdot} + \Phi_{ij,t\cdot})$. Thus, $(\Delta\xi_{it})^2 - \mathrm{E}[(\Delta\xi_{it})^2] = \sum_{j=1}^{i-1}[\phi_{ijt\cdot}'(\Delta v_{j\cdot}\Delta v_{j\cdot}' - \sigma_{v0}^2 C)\phi_{ijt\cdot}] + 2\sum_{j=1}^{i-1}\sum_{k=1}^{j-1}\Delta v_{j\cdot}'\phi_{ijt\cdot}\phi_{ikt\cdot}'\Delta v_{k\cdot}$. It follows that

$$\begin{aligned}
& \tfrac{1}{n(T-1)}\sum_{i=1}^n\{(\Delta\xi_{it})^2 - \mathrm{E}[(\Delta\xi_{it})^2]\} \\
= \;& \tfrac{1}{n(T-1)}\sum_{i=1}^n\sum_{j=1}^{i-1}[\phi_{ijt\cdot}'(\Delta v_{j\cdot}\Delta v_{j\cdot}' - \sigma_{v0}^2 C)\phi_{ijt\cdot}] \\
& +2\tfrac{1}{n(T-1)}\sum_{i=1}^n\sum_{j=1}^{i-1}\sum_{k=1}^{j-1}\Delta v_{j\cdot}'\phi_{ijt\cdot}\phi_{ikt\cdot}'\Delta v_{k\cdot} \\
= \;& \tfrac{1}{n(T-1)}\sum_{j=1}^{n-1}\big\{\sum_{i=j+1}^n[\phi_{ijt\cdot}'(\Delta v_{j\cdot}\Delta v_{j\cdot}' - \sigma_{v0}^2 C)\phi_{ijt\cdot}]\big\} \\
& +2\tfrac{1}{n(T-1)}\sum_{j=1}^{n-1}\Delta v_{j\cdot}'\big\{\sum_{i=j+1}^n\sum_{k=1}^{j-1}\phi_{ijt\cdot}\phi_{ikt\cdot}'\Delta v_{k\cdot}\big\}.
\end{aligned}$$

Clearly, the first term is the 'average' of $n-1$ independent terms, and the second is the 'average' of a M.D. array as the term in the curling brackets is $G_{n,j-1}$-measurable. Conditions of Theorem 19.7 of Davidson (1994) are easily verified, and hence $\frac{1}{n(T-1)}\sum_{i=1}^n\{(\Delta\xi_{it})^2 - \mathrm{E}[(\Delta\xi_{it})^2]\} = o_p(1)$. Similarly, one shows that $\frac{1}{n(T-1)}\sum_{i=1}^n\{\Delta\xi_{it}\Delta\xi_{is} - \mathrm{E}[(\Delta\xi_{it}\Delta\xi_{is})]\} = o_p(1)$ for $s \neq t$. Thus, $\frac{\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n[\Delta\xi_{i\cdot}'C\Delta\xi_{i\cdot} - \mathrm{E}(\Delta\xi_{i\cdot}'C\Delta\xi_{i\cdot})] = o_p(1)$, and $H_1 = o_p(1)$.

The proofs for H_3 and H_4 can be done in a similar manner as the proof for the second term of H_1. The proofs for H_2 and H_5 are similar to the proof of the first part of H_1, as they each involves a sum of n independent terms.

Third, with $g_{3i} = \Delta v_{2i}\Delta\zeta_i + \Theta_{ii}(\Delta v_{2i}\Delta y_{1i}^\circ + \sigma_{v0}^2) + \Delta v_{i-}'\Delta y_{1i-}^*$, we obtain,

$$\begin{aligned}
& \tfrac{1}{n(T-1)}\sum_{i=1}^n[g_{3i}^2 - \mathrm{E}(g_{3i}^2)] \\
= \;& \tfrac{1}{n(T-1)}\sum_{i=1}^n[(\Delta v_{2i}^2 - 2\sigma_{v0}^2)\Delta\zeta_i^2] + \tfrac{2\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n[\Delta\zeta_i^2 - \mathrm{E}(\Delta\zeta_i^2)] \\
& +\tfrac{1}{n(T-1)}\sum_{i=1}^n\Theta_{ii}^2[(\Delta v_{2i}\Delta y_{1i}^\circ)^2 - \mathrm{E}((\Delta v_{2i}\Delta y_{1i}^\circ)^2)] \\
& +\tfrac{2\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n\Theta_{ii}^2[\Delta v_{2i}\Delta y_{1i}^\circ - \mathrm{E}(\Delta v_{2i}\Delta y_{1i}^\circ)] \\
& +\tfrac{1}{n(T-1)}\sum_{i=1}^n[(\Delta v_{i-}'\Delta y_{1i-}^*)^2 - \mathrm{E}((\Delta v_{i-}'\Delta y_{1i-}^*)^2)] \\
& +\tfrac{2}{n(T-1)}\sum_{i=1}^n\Theta_{ii}[\Delta v_{2i}^2\Delta\zeta_i\Delta y_{1i}^\circ - \mathrm{E}(\Delta v_{2i}^2\Delta\zeta_i\Delta y_{1i}^\circ)] + \tfrac{2\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n\Theta_{ii}\Delta v_{2i}\Delta\zeta_i \\
& +\tfrac{2}{n(T-1)}\sum_{i=1}^n[\Delta v_{2i}\Delta\zeta_i(\Delta v_{i-}'\Delta y_{1i-}^*) - \mathrm{E}(\Delta v_{2i}\Delta\zeta_i(\Delta v_{i-}'\Delta y_{1i-}^*))] \\
& +\tfrac{2}{n(T-1)}\sum_{i=1}^n\Theta_{ii}[(\Delta v_{2i}\Delta y_{1i}^\circ)(\Delta v_{i-}'\Delta y_{1i-}^*) - \mathrm{E}((\Delta v_{2i}\Delta y_{1i}^\circ)(\Delta v_{i-}'\Delta y_{1i-}^*))] \\
& +\tfrac{2\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n\Theta_{ii}[\Delta v_{i-}'\Delta y_{1i-}^* - \mathrm{E}(\Delta v_{i-}'\Delta y_{1i-}^*)] \equiv \sum_{r=1}^{10}Q_r.
\end{aligned}$$

As $\Delta\zeta_i^2$ is $\mathcal{F}_{n,i-1}$-measurable, Q_1 is the average of a M.D. array and its convergence follows from WLLN for M.D. array, and the convergence of Q_7 immediately follows. For Q_2, note that $\Delta\zeta = (\Theta^{u,\prime} + \Theta^{\ell})\Delta y_1^\circ = (\Theta^{u\prime} + \Theta^{\ell})B_{30}B_{10}\Delta y_1$. It follows that $Q_2 = \frac{2\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n(\Delta y_1'A\Delta y_1 - \mathrm{E}(\Delta y_1'A\Delta y_1)] = o_p(1)$ by Assumption F, where $A = ((\Theta^{u\prime} + \Theta^{\ell})B_{30}B_{10})'(\Theta^{u\prime} + \Theta^{\ell})B_{30}B_{10}$ is easily seen to be uniformly bounded in both row and column sums. Writing $\Delta y_1^\circ =$

$B_{30}B_{10}\Delta y_0+B_{30}\Delta x_1\beta_0+\Delta v_1 \equiv g(y_0,v_0)+v_1$, the convergence of Q_3, Q_4 and Q_6 can be easily proved though tedious. The results for Q_5 and Q_{10} are proved by the independence between $\Delta v'_{i-}$ and Δy^*_{1i-} are independent, $\Delta y^*_{1t} = \Phi_{t+}\Delta y_1$, and Assumption F. Finally, the results for Q_8 and Q_9 can be proved by further writing $\Delta y^*_{1t} = \Phi_{t+}\Delta y_1 = \Phi_{t+}(B_{30}B_{10})^{-1}\Delta y^\circ_1 \equiv q(\Delta y_0, v_0) + \Phi_{t+}(B_{30}B_{10})^{-1}v_1$.

Subsequently, for the cross-product terms, we have,

$$\begin{aligned}
&\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[g_{1i}g_{2i} - \mathrm{E}(g_{1i}g_{2i})]\\
=\;&\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[\Pi'_{i\cdot}(\Delta v_{i\cdot}\Delta v'_{i\cdot} - \sigma^2_{v0}C)\Delta\xi_{i\cdot}] + \tfrac{\sigma^2_{v0}}{n(T-1)}\textstyle\sum_{i=1}^n(\Pi'_{i\cdot}C\Delta\xi_{i\cdot})\\
&+\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n\Pi'_{i\cdot}[\Delta v_{i\cdot}\Delta v'_{i\cdot}\Delta v^*_{i\cdot} - \mathrm{E}(\Delta v_{i\cdot}\Delta v'_{i\cdot}\Delta v^*_{i\cdot})] + \tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(1'_{T-1}d_{i\cdot})\Pi'_{i\cdot}\Delta v_{i\cdot}].
\end{aligned}$$

$$\begin{aligned}
&\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[g_{1i}g_{3i} - \mathrm{E}(g_{1i}g_{3i})]\\
=\;&\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n\Pi'_{i\cdot}[\Delta v_{i\cdot}\Delta v_{2i}\Delta\zeta_i - \mathrm{E}(\Delta v_{i\cdot}\Delta v_{2i}\Delta\zeta_i)]\\
&+\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n\Theta_{ii}\Pi'_{i\cdot}[\Delta v_{i\cdot}(\Delta v_{2i}\Delta y^\circ_{1i} + \sigma^2_{v0}) - \mathrm{E}(\Delta v_{i\cdot}(\Delta v_{2i}\Delta y^\circ_{1i} + \sigma^2_{v0}))]\\
&+\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n\Pi'_{i\cdot}[\Delta v_{i\cdot}\Delta v'_{i-}\Delta y^*_{1i-} - \mathrm{E}(\Delta v_{i\cdot}\Delta v'_{i-}\Delta y^*_{1i-})].
\end{aligned}$$

$$\begin{aligned}
&\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[g_{2i}g_{3i} - \mathrm{E}(g_{2i}g_{3i})]\\
=\;&\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(\Delta v'_{i\cdot}\Delta\xi_i)(\Delta v_{2i}\Delta\zeta_i) - \mathrm{E}((\Delta v'_{i\cdot}\Delta\xi_i)(\Delta v_{2i}\Delta\zeta_i))]\\
&+\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n\Theta_{ii}[(\Delta v'_{i\cdot}\Delta\xi_i)(\Delta v_{2i}\Delta y^\circ_{1i} + \sigma^2_{v0}) - \mathrm{E}((\Delta v'_{i\cdot}\Delta\xi_i)(\Delta v_{2i}\Delta y^\circ_{1i} + \sigma^2_{v0}))]\\
&+\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(\Delta v'_{i\cdot}\Delta\xi_i)(\Delta v'_{i-}\Delta y^*_{1i-}) - \mathrm{E}((\Delta v'_{i\cdot}\Delta\xi_i)(\Delta v'_{i-}\Delta y^*_{1i-}))]\\
&+\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(\Delta v'_{i\cdot}\Delta v^*_{i\cdot})(\Delta v_{2i}\Delta\zeta_i) - \mathrm{E}((\Delta v'_{i\cdot}\Delta v^*_{i\cdot})(\Delta v_{2i}\Delta\zeta_i))]\\
&+\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(\Delta v'_{i\cdot}\Delta v^*_{i\cdot})(\Delta v_{2i}\Delta y^\circ_{1i} + \sigma^2_{v0}) - \mathrm{E}((\Delta v'_{i\cdot}\Delta v^*_{i\cdot})(\Delta v_{2i}\Delta y^\circ_{1i} + \sigma^2_{v0}))]\\
&+\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(\Delta v'_{i\cdot}\Delta v^*_{i\cdot})(\Delta v'_{i-}\Delta y^*_{1i-}) - \mathrm{E}((\Delta v'_{i\cdot}\Delta v^*_{i\cdot})(\Delta v'_{i-}\Delta y^*_{1i-}))]\\
&+\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(1'_{T-1}d_{i\cdot})\Delta v_{2i}\Delta\zeta_i] + \tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(1'_{T-1}d_{i\cdot})\Theta_{ii}(\Delta v_{2i}\Delta y^\circ_{1i} + \sigma^2_{v0})]\\
&+\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[(1'_{T-1}d_{i\cdot})(\Delta v'_{i-}\Delta y^*_{1i-} - \mathrm{E}(\Delta v'_{i-}\Delta y^*_{1i-}))]
\end{aligned}$$

The convergence of each of the terms above can be proved in a similarly manner as these terms appear in similar forms as the terms appeared in the H_r and Q_r. ■

References

[1] Alvarez, J., Arellano, M., 2004. Robust likelihood estimation of dynamic panel data models. Working Paper #0421, CEMFI.

[2] Anderson, T. W., Hsiao, C., 1981. Estimation of dynamic models with error components. *Journal of American Statistical Association* 76, 598-606.

[3] Anderson, T. W., Hsiao, C., 1982. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18, 47-82.

[4] Andrews, D. W. K., 1992. Generic uniform convergence. *Econometric Theory* 8, 241-257.

[5] Andrews, D. W. K., 2005. Cross-section regression with common shocks. Econometrica 73, 1551-1585.

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

[7] Anselin, L., 2001. Spatial econometrics. In: Baltagi, B. H. (Eds.), A companion to theoretical econometrics. Blackwell Publishers Ltd., Massachusetts, pp. 310-330.

[8] 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.

[9] Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

[10] Baltagi, B. H., Egger, P., Pfaffermayr, M., 2013. A generalized spatial panel model with random effects. Econometric Reviews 32, 650-685.

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

[12] Bhargava, A., Sargan, J. D., 1983. Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51, 1635-1659.

[13] Binder, M., Hsiao, C., Pesaran, M. H., 2005. Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21, 795-837.

[14] Bun, M.J., Carree, M.A., 2005. Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics* 23, 200-210.

[15] Čížk, P., Jacobs, J. P. A. M., Ligthart J., Vrijburg, H., 2014. GMM estimation of fixed effects dynamic panel data models with spatial lag and spatial errors. CentER Discussion Paper No. 2014-003, Tilburg University.

[16] Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, Oxford.

[17] Elhorst, J. P., 2005. Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis* 37, 85-106.

[18] Elhorst, J. P., 2010. Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics* 40, 272-282.

[19] Elhorst, J. P., 2012. Dynamic spatial panels: models, methods, and applications. *Journal of Geographical Systems* 14, 5-28.

[20] Gouriéroux, C. Phillips, P. C. B, Yu, J., 2010. Indirect inference for dynamic panel models. *Journal of Econometrics* 157, 68-77.

[21] Hahn, J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic model with fixed effects when both n and T are large. *Econometrica* 70, 1639-1657.

[22] Hsiao, C., Pesaran, M. H., Tahmiscioglu, A. K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109, 107-150.

[23] Hsiao, C., 2003. *Analysis of Panel Data. 2nd edition.* Cambridge University Press.

[24] Huber, P. J., 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73-101.

[25] Huber, P. J., 1981. *Robust Statistics.* Wiley, New York.

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

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

[28] 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.

[29] Korniotis, G. M., 2010. Estimating panel models with internal and external habit formation. *Journal of Business and Economic Statistics* 28, 145-158.

[30] Kruiniger, H., 2013. Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions. *Journal of Econometrics* 173, 175-188.

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

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

[33] Lee, L. F., 2004b. A supplement to 'Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models'. Working paper, Department of Economics, Ohio State University.

[34] Lee, L. F., Yu, J., 2010a. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory* 26, 564-597.

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

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

[37] Lee, L. F., Yu, J., 2014. Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics* 180, 174-197.

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

[39] Lee, L. F., Yu, J., 2015b. Identification of spatial Durbin panel models. *Journal of Applied Econometrics* DOI: 10.1002/jae.2450.

[40] Lancaster, T, 2002. Orthogonal parameters and panel data. *Review of Economic Studies* 69, 647-666.

[41] Mutl, J., 2006. Dynamic panel data models with spatially correlated disturbances. *PhD Thesis*, University of Maryland, College Park.

[42] Newey, W., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: McFadden, D., Engle, R. F. (Eds.), *Handbook of Econometrics Vol. IV*. Elsevier, North Holland, pp. 2111-2245.

[43] Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica* 49, 1417-1426.

[44] Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967-1012.

[45] Pesaran, M. H., Tosetti, E., 2011. Large panels with common factors and spatial correlation. Journal of Econometrics 161, 182-202.

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

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

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

[49] Yang, Z. (2015). A general method for third-order bias and variance correction on a nonlinear estimator. *Journal of Econometrics* 186, 178-200.

[50] 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.

[51] Yu, J., Lee, L. F., 2010. Estimation of unit root spatial dynamic panel data models. *Econometric Theory* 26, 1332-1362.

Table 1a. Empirical Mean(sd) of CQMLE, FQMLE and M-Estimator, SE Model, $T = 3, m = 5$

		$n=50$			$n=200$		
dgp	ψ	CQMLE	FQMLE	M-Est	CQMLE	FQMLE	M-Est
1	1	1.0152(.096)	1.0017(.100)	1.0015(.100)	1.0109(.050)	1.0021(.052)	1.0020(.053)
	1	0.9154(.135)	0.9678(.148)	0.9719(.154)	0.9080(.065)	0.9960(.079)	0.9962(.080)
	.5	0.3605(.055)	0.4995(.065)	0.5015(.066)	0.2869(.033)	0.5009(.043)	0.5013(.044)
	.5	0.4702(.107)	0.4761(.093)	0.4793(.105)	0.4775(.073)	0.4877(.060)	0.4907(.070)
2	1	1.0142(.098)	1.0007(.102)	1.0002(.102)	1.0099(.050)	1.0015(.053)	1.0014(.053)
	1	0.9176(.266)	0.9662(.284)	0.9785(.307)	0.9045(.128)	0.9920(.152)	0.9935(.155)
	.5	0.3610(.066)	0.4975(.069)	0.5023(.078)	0.2876(.041)	0.5002(.047)	0.5018(.052)
	.5	0.4701(.106)	0.4770(.092)	0.4803(.104)	0.4741(.075)	0.4844(.063)	0.4883(.072)
3	1	1.0133(.099)	1.0001(.103)	0.9997(.103)	1.0090(.047)	1.0003(.049)	1.0003(.049)
	1	0.9192(.198)	0.9678(.212)	0.9771(.227)	0.9060(.099)	0.9938(.119)	0.9947(.121)
	.5	0.3585(.059)	0.4953(.066)	0.4992(.071)	0.2881(.036)	0.5018(.046)	0.5029(.048)
	.5	0.4681(.110)	0.4736(.093)	0.4786(.106)	0.4741(.075)	0.4852(.062)	0.4884(.073)
1	1	1.0525(.100)	1.0035(.103)	1.0012(.104)	1.0517(.052)	1.0009(.053)	0.9999(.053)
	1	0.9204(.138)	0.9255(.126)	0.9702(.154)	0.9313(.069)	0.9712(.066)	0.9915(.078)
	0	-0.1524(.065)	-0.0036(.074)	0.0032(.078)	-0.1825(.035)	-0.0032(.042)	0.0005(.043)
	.5	0.4731(.106)	0.4848(.085)	0.4807(.105)	0.4820(.072)	0.4897(.059)	0.4881(.070)
2	1	1.0528(.099)	1.0042(.102)	1.0006(.104)	1.0479(.053)	0.9979(.055)	0.9962(.055)
	1	0.9230(.265)	0.9032(.241)	0.9764(.299)	0.9327(.133)	0.9596(.129)	0.9940(.150)
	0	-0.1529(.071)	-0.0091(.076)	0.0022(.086)	-0.1821(.039)	-0.0042(.043)	0.0018(.047)
	.5	0.4741(.103)	0.4880(.086)	0.4805(.102)	0.4806(.073)	0.4917(.059)	0.4873(.072)
3	1	1.0515(.100)	1.0021(.103)	0.9990(.104)	1.0497(.053)	0.9998(.054)	0.9985(.054)
	1	0.9250(.200)	0.9194(.185)	0.9767(.224)	0.9319(.102)	0.9661(.100)	0.9924(.115)
	0	-0.1543(.068)	-0.0077(.076)	0.0014(.083)	-0.1831(.037)	-0.0045(.043)	0.0001(.045)
	.5	0.4740(.107)	0.4855(.088)	0.4811(.105)	0.4834(.072)	0.4929(.057)	0.4906(.070)
1	1	1.0484(.103)	0.9987(.104)	1.0015(.104)	1.0418(.053)	0.9989(.054)	0.9997(.054)
	1	0.9504(.139)	0.9552(.135)	0.9764(.147)	0.9641(.072)	0.9849(.071)	0.9909(.075)
	-.5	-0.6034(.059)	-0.4915(.067)	-0.4978(.070)	-0.6070(.030)	-0.4970(.035)	-0.4988(.036)
	.5	0.4798(.108)	0.4830(.090)	0.4815(.108)	0.4889(.072)	0.4880(.060)	0.4905(.072)
2	1	1.0494(.102)	0.9980(.102)	1.0024(.103)	1.0419(.054)	0.9981(.054)	0.9997(.054)
	1	0.9380(.271)	0.9261(.250)	0.9642(.284)	0.9661(.138)	0.9775(.131)	0.9933(.145)
	-.5	-0.6028(.065)	-0.4885(.070)	-0.4981(.075)	-0.6072(.032)	-0.4949(.035)	-0.4989(.037)
	.5	0.4792(.104)	0.4831(.090)	0.4822(.102)	0.4849(.073)	0.4889(.060)	0.4862(.073)
3	1	1.0481(.105)	0.9971(.106)	1.0009(.106)	1.0409(.054)	0.9975(.054)	0.9989(.054)
	1	0.9388(.195)	0.9340(.182)	0.9647(.205)	0.9658(.103)	0.9808(.099)	0.9928(.108)
	-.5	-0.6059(.061)	-0.4924(.068)	-0.5005(.072)	-0.6092(.030)	-0.4974(.034)	-0.5008(.035)
	.5	0.4744(.108)	0.4790(.091)	0.4764(.108)	0.4841(.073)	0.4871(.059)	0.4856(.072)

Note: Par = $\psi = (\beta, \sigma_v^2, \rho, \lambda_2)'$; dgp=1 (normal), 2 (normal mixture), and 3 (chi-square).
X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 1, .5)$, as in Footnote 11.
W_3 is generated according to Group Interaction scheme as in Footnote 12.

Table 1a. Cont'd, $T = 7$

		$n=50$			$n=100$		
dgp	ψ	CQMLE	FQMLE	M-Est	CQMLE	FQMLE	M-Est
1	1	1.0248(.044)	1.0015(.044)	1.0013(.044)	1.0231(.033)	1.0018(.033)	1.0017(.033)
	1	0.9771(.081)	0.9888(.083)	0.9893(.083)	0.9821(.059)	0.9949(.060)	0.9956(.061)
	.5	0.4456(.028)	0.4987(.029)	0.4990(.029)	0.4407(.021)	0.4990(.022)	0.4994(.022)
	.5	0.4928(.057)	0.4920(.055)	0.4947(.056)	0.4931(.047)	0.4904(.044)	0.4953(.046)
2	1	1.0247(.045)	1.0012(.045)	1.0010(.045)	1.0232(.033)	1.0020(.033)	1.0019(.033)
	1	0.9776(.183)	0.9887(.186)	0.9899(.187)	0.9806(.129)	0.9931(.132)	0.9942(.133)
	.5	0.4461(.028)	0.4988(.029)	0.4992(.029)	0.4412(.022)	0.4991(.022)	0.4996(.022)
	.5	0.4919(.058)	0.4914(.055)	0.4941(.057)	0.4906(.048)	0.4882(.046)	0.4928(.047)
3	1	1.0250(.044)	1.0015(.044)	1.0013(.044)	1.0214(.033)	1.0003(.033)	1.0002(.033)
	1	0.9751(.130)	0.9863(.133)	0.9872(.134)	0.9779(.095)	0.9908(.097)	0.9915(.097)
	.5	0.4458(.028)	0.4986(.029)	0.4990(.029)	0.4413(.020)	0.4996(.021)	0.5000(.021)
	.5	0.4903(.057)	0.4896(.056)	0.4923(.057)	0.4919(.048)	0.4898(.045)	0.4940(.047)
1	1	1.0342(.048)	1.0017(.048)	1.0009(.048)	1.0360(.035)	1.0021(.035)	1.0016(.036)
	1	0.9840(.083)	0.9660(.077)	0.9922(.084)	0.9873(.059)	0.9807(.056)	0.9961(.060)
	0	-0.0603(.036)	-0.0020(.037)	-0.0005(.038)	-0.0632(.026)	-0.0012(.027)	-0.0003(.028)
	.5	0.4932(.055)	0.4953(.052)	0.4945(.055)	0.4932(.047)	0.4946(.044)	0.4941(.047)
2	1	1.0345(.047)	1.0024(.047)	1.0013(.047)	1.0341(.034)	1.0006(.034)	0.9999(.034)
	1	0.9860(.182)	0.9575(.170)	0.9943(.185)	0.9805(.133)	0.9666(.126)	0.9893(.135)
	0	-0.0603(.037)	-0.0028(.038)	-0.0008(.039)	-0.0631(.027)	-0.0018(.027)	-0.0005(.028)
	.5	0.4920(.057)	0.4959(.054)	0.4931(.057)	0.4917(.048)	0.4943(.045)	0.4928(.047)
3	1	1.0327(.046)	1.0004(.046)	0.9994(.047)	1.0343(.035)	1.0005(.035)	0.9999(.035)
	1	0.9882(.134)	0.9642(.125)	0.9966(.135)	0.9863(.093)	0.9759(.088)	0.9951(.094)
	0	-0.0606(.038)	-0.0025(.039)	-0.0007(.039)	-0.0637(.027)	-0.0018(.028)	-0.0007(.029)
	.5	0.4933(.056)	0.4958(.052)	0.4942(.056)	0.4930(.047)	0.4953(.044)	0.4941(.046)
1	1	1.0244(.047)	0.9977(.047)	1.0002(.047)	1.0251(.034)	0.9988(.034)	1.0003(.034)
	1	0.9822(.082)	0.9705(.078)	0.9867(.082)	0.9906(.058)	0.9859(.056)	0.9952(.059)
	-.5	-0.5454(.037)	-0.4945(.038)	-0.4992(.039)	-0.5472(.027)	-0.4973(.028)	-0.5002(.029)
	.5	0.4919(.057)	0.4920(.054)	0.4923(.057)	0.4959(.047)	0.4952(.044)	0.4963(.047)
2	1	1.0262(.047)	0.9983(.047)	1.0019(.047)	1.0241(.036)	0.9969(.036)	0.9992(.036)
	1	0.9859(.177)	0.9655(.165)	0.9906(.179)	0.9945(.128)	0.9834(.121)	0.9992(.129)
	-.5	-0.5460(.037)	-0.4931(.039)	-0.5000(.040)	-0.5466(.027)	-0.4950(.029)	-0.4995(.029)
	.5	0.4903(.059)	0.4924(.055)	0.4904(.059)	0.4944(.046)	0.4955(.043)	0.4947(.046)
3	1	1.0251(.047)	0.9977(.047)	1.0008(.047)	1.0246(.035)	0.9978(.035)	0.9998(.035)
	1	0.9791(.132)	0.9630(.123)	0.9837(.133)	0.9892(.095)	0.9816(.090)	0.9939(.096)
	-.5	-0.5462(.036)	-0.4943(.039)	-0.5002(.039)	-0.5468(.027)	-0.4962(.029)	-0.4998(.029)
	.5	0.4924(.055)	0.4932(.051)	0.4927(.055)	0.4956(.046)	0.4952(.043)	0.4958(.046)

Table 1b. Empirical sd and average of estimated standard errors of M-Estimator
`SE` Model, $T = 3, m = 5$, Parameter configurations as in Table 1a.

dgp	ψ	$n=50$ sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	$n=100$ sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	$n=200$ sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.100	.112	.099	.096	.071	.073	.070	.069	.053	.053	.051	.051
	1	.154	.165	.150	.146	.113	.114	.110	.109	.080	.081	.079	.080
	.5	.066	.068	.064	.065	.059	.054	.054	.056	.044	.040	.042	.044
	.5	.105	.111	.099	.096	.083	.086	.081	.080	.070	.070	.068	.068
2	1	.102	.124	.099	.093	.071	.078	.069	.068	.053	.055	.051	.050
	1	.307	.117	.152	.263	.209	.076	.110	.198	.155	.050	.079	.147
	.5	.078	.071	.064	.070	.065	.053	.054	.063	.052	.037	.042	.051
	.5	.104	.126	.099	.090	.089	.095	.082	.078	.072	.074	.068	.067
3	1	.103	.117	.099	.095	.070	.075	.069	.069	.049	.053	.051	.051
	1	.227	.133	.151	.203	.162	.089	.110	.153	.121	.061	.079	.113
	.5	.071	.070	.064	.066	.062	.053	.054	.060	.048	.039	.042	.047
	.5	.106	.118	.099	.093	.088	.091	.082	.079	.073	.072	.068	.067
1	1	.104	.113	.102	.100	.072	.074	.071	.071	.053	.054	.052	.052
	1	.154	.165	.149	.144	.111	.112	.107	.106	.078	.078	.076	.076
	.0	.078	.081	.075	.075	.056	.057	.055	.055	.043	.042	.042	.042
	.5	.105	.111	.099	.094	.087	.086	.082	.081	.070	.071	.069	.068
2	1	.104	.126	.103	.131	.074	.078	.071	.070	.055	.056	.052	.052
	1	.299	.117	.157	.568	.211	.073	.107	.196	.150	.048	.076	.143
	.0	.086	.086	.077	.173	.065	.058	.055	.060	.047	.041	.042	.046
	.5	.102	.126	.099	.094	.087	.094	.082	.078	.072	.075	.069	.067
3	1	.104	.120	.102	.099	.073	.076	.071	.071	.054	.054	.052	.052
	1	.224	.132	.150	.203	.156	.086	.107	.149	.115	.058	.076	.109
	.0	.083	.084	.075	.077	.059	.057	.055	.057	.045	.042	.042	.044
	.5	.105	.117	.099	.093	.084	.091	.082	.079	.070	.072	.068	.067
1	1	.104	.115	.104	.101	.072	.075	.072	.071	.054	.054	.053	.052
	1	.147	.161	.145	.140	.103	.108	.103	.101	.075	.075	.073	.072
	-.5	.070	.077	.069	.067	.048	.051	.048	.047	.036	.036	.035	.035
	.5	.108	.111	.099	.095	.086	.087	.082	.081	.072	.070	.068	.068
2	1	.103	.126	.103	.099	.072	.081	.072	.071	.054	.056	.053	.052
	1	.284	.107	.143	.255	.215	.067	.104	.196	.145	.043	.073	.142
	-.5	.075	.085	.069	.067	.051	.054	.048	.049	.037	.037	.035	.036
	.5	.102	.127	.099	.089	.084	.095	.082	.077	.073	.075	.069	.067
3	1	.106	.120	.103	.100	.074	.077	.072	.071	.054	.055	.053	.053
	1	.205	.124	.143	.194	.151	.081	.103	.147	.108	.054	.073	.106
	-.5	.072	.080	.069	.068	.049	.052	.048	.048	.035	.037	.035	.035
	.5	.108	.118	.100	.094	.088	.092	.083	.079	.072	.073	.069	.067

Table 1b. Cont'd, $T = 7$

		$n = 50$				$n = 100$				$n = 200$			
dgp	ψ	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.044	.047	.044	.043	.033	.034	.032	.032	.025	.026	.025	.025
	1	.083	.090	.084	.082	.061	.062	.059	.058	.042	.043	.042	.042
	.5	.029	.031	.028	.028	.022	.022	.021	.021	.016	.016	.016	.015
	.5	.056	.060	.055	.054	.046	.048	.046	.045	.039	.040	.039	.039
2	1	.045	.050	.044	.043	.033	.035	.032	.032	.025	.026	.025	.025
	1	.187	.050	.084	.172	.133	.032	.059	.127	.093	.021	.042	.092
	.5	.029	.032	.028	.028	.022	.023	.021	.021	.017	.016	.016	.016
	.5	.057	.066	.055	.052	.047	.051	.046	.045	.040	.041	.039	.038
3	1	.044	.049	.044	.043	.033	.034	.032	.032	.026	.026	.025	.025
	1	.134	.062	.083	.128	.097	.041	.059	.092	.069	.028	.042	.066
	.5	.029	.031	.028	.028	.021	.022	.021	.021	.017	.016	.016	.016
	.5	.057	.063	.056	.053	.047	.049	.046	.046	.039	.040	.039	.038
1	1	.048	.051	.047	.047	.036	.036	.035	.035	.027	.027	.027	.027
	1	.084	.090	.083	.082	.060	.061	.059	.058	.040	.042	.042	.042
	0	.038	.042	.039	.038	.028	.029	.028	.028	.021	.021	.021	.021
	.5	.055	.061	.055	.054	.047	.048	.046	.046	.040	.040	.039	.039
2	1	.047	.054	.047	.046	.034	.038	.035	.034	.027	.028	.027	.027
	1	.185	.050	.084	.174	.135	.031	.058	.126	.095	.021	.042	.092
	0	.039	.045	.039	.037	.028	.030	.028	.028	.021	.022	.021	.021
	.5	.057	.066	.055	.052	.047	.051	.046	.045	.040	.041	.039	.038
3	1	.047	.052	.047	.047	.035	.037	.035	.035	.027	.028	.027	.027
	1	.135	.062	.084	.128	.094	.040	.059	.093	.068	.027	.042	.066
	0	.039	.043	.039	.038	.029	.030	.028	.028	.021	.021	.021	.021
	.5	.056	.063	.055	.052	.046	.049	.046	.045	.039	.040	.039	.039
1	1	.047	.051	.047	.046	.034	.036	.035	.035	.027	.027	.027	.027
	1	.082	.089	.083	.081	.059	.061	.059	.058	.041	.042	.041	.041
	-.5	.039	.043	.040	.040	.029	.030	.029	.028	.021	.021	.021	.021
	.5	.057	.061	.056	.054	.047	.048	.046	.045	.039	.040	.039	.039
2	1	.047	.054	.047	.046	.036	.038	.035	.035	.027	.028	.027	.026
	1	.179	.048	.083	.174	.129	.031	.059	.127	.095	.020	.041	.091
	-.5	.040	.048	.040	.038	.029	.032	.029	.028	.021	.022	.021	.020
	.5	.059	.067	.056	.052	.046	.051	.046	.045	.038	.041	.039	.038
3	1	.047	.052	.047	.046	.035	.037	.035	.034	.026	.027	.027	.027
	1	.133	.061	.082	.125	.096	.040	.059	.093	.067	.027	.041	.066
	-.5	.039	.045	.040	.039	.029	.031	.029	.028	.021	.021	.021	.021
	.5	.055	.063	.056	.053	.046	.049	.046	.045	.039	.040	.039	.039

Table 2a. Empirical Mean(sd) of CQMLE and M-Estimator, `SL` Model, $T = 3, m = 5$

		$n = 50$		$n = 100$		$n = 200$	
dgp	ψ	`CQMLE`	`M-Est`	`CQMLE`	`M-Est`	`CQMLE`	`M-Est`
1	1	1.0160(.045)	0.9999(.047)	0.9881(.039)	1.0001(.040)	0.9993(.025)	0.9997(.026)
	1	0.9384(.135)	0.9665(.144)	0.9510(.094)	0.9850(.101)	0.9659(.069)	0.9942(.074)
	.5	0.4365(.038)	0.4991(.042)	0.4186(.030)	0.4985(.034)	0.4331(.020)	0.5005(.022)
	.2	0.2327(.064)	0.1975(.068)	0.2078(.069)	0.1971(.074)	0.1986(.043)	0.1983(.046)
2	1	1.0162(.046)	1.0004(.047)	0.9877(.038)	0.9996(.039)	0.9988(.025)	0.9992(.025)
	1	0.9415(.268)	0.9713(.285)	0.9494(.193)	0.9844(.208)	0.9622(.138)	0.9907(.146)
	.5	0.4370(.043)	0.4997(.045)	0.4200(.033)	0.4998(.036)	0.4330(.022)	0.5000(.022)
	.2	0.2310(.064)	0.1960(.068)	0.2067(.070)	0.1965(.075)	0.1979(.043)	0.1977(.046)
3	1	1.0148(.044)	0.9988(.046)	0.9884(.040)	1.0004(.041)	0.9984(.025)	0.9988(.026)
	1	0.9461(.207)	0.9755(.220)	0.9516(.148)	0.9861(.159)	0.9627(.104)	0.9909(.110)
	.5	0.4388(.041)	0.5021(.044)	0.4209(.033)	0.5011(.036)	0.4334(.020)	0.5003(.021)
	.2	0.2294(.060)	0.1938(.064)	0.2058(.068)	0.1954(.074)	0.1985(.042)	0.1978(.045)
1	1	1.0289(.048)	0.9990(.049)	1.0153(.039)	0.9999(.039)	1.0167(.026)	1.0000(.026)
	1	0.9452(.138)	0.9712(.146)	0.9527(.094)	0.9814(.100)	0.9696(.069)	0.9927(.073)
	0	-0.0752(.046)	0.0001(.051)	-0.0876(.034)	-0.0019(.038)	-0.0681(.022)	0.0006(.023)
	.2	0.2012(.093)	0.1891(.094)	0.1980(.074)	0.1942(.075)	0.1995(.046)	0.1986(.046)
2	1	1.0283(.047)	0.9982(.048)	1.0160(.039)	1.0005(.040)	1.0157(.026)	0.9991(.026)
	1	0.9538(.277)	0.9814(.292)	0.9527(.194)	0.9823(.205)	0.9668(.141)	0.9901(.148)
	0	-0.0746(.049)	0.0008(.052)	-0.0845(.038)	0.0011(.041)	-0.0678(.024)	0.0006(.025)
	.2	0.2034(.088)	0.1919(.089)	0.1982(.074)	0.1947(.076)	0.1988(.045)	0.1981(.046)
3	1	1.0271(.048)	0.9969(.050)	1.0172(.040)	1.0015(.041)	1.0169(.026)	1.0002(.026)
	1	0.9487(.201)	0.9753(.212)	0.9591(.142)	0.9885(.150)	0.9696(.104)	0.9928(.109)
	0	-0.0745(.049)	0.0007(.055)	-0.0869(.037)	-0.0008(.040)	-0.0691(.022)	-0.0003(.024)
	.2	0.1994(.092)	0.1870(.094)	0.1974(.076)	0.1934(.078)	0.1993(.047)	0.1984(.047)
1	1	1.0205(.047)	0.9975(.048)	1.0209(.041)	0.9989(.042)	1.0222(.027)	1.0005(.028)
	1	0.9613(.140)	0.9749(.144)	0.9703(.097)	0.9864(.100)	0.9809(.069)	0.9936(.071)
	-.5	-0.5505(.045)	-0.4955(.049)	-0.5646(.034)	-0.5004(.037)	-0.5521(.022)	-0.4992(.024)
	.2	0.1886(.090)	0.1913(.092)	0.1947(.068)	0.1942(.068)	0.1896(.045)	0.1975(.046)
2	1	1.0224(.049)	0.9997(.050)	1.0228(.041)	1.0008(.041)	1.0210(.027)	0.9994(.027)
	1	0.9533(.270)	0.9673(.277)	0.9684(.199)	0.9848(.205)	0.9778(.140)	0.9905(.144)
	-.5	-0.5515(.048)	-0.4969(.051)	-0.5614(.035)	-0.4974(.038)	-0.5526(.023)	-0.4998(.024)
	.2	0.1843(.086)	0.1865(.088)	0.1948(.067)	0.1945(.068)	0.1888(.044)	0.1971(.044)
3	1	1.0222(.049)	0.9991(.050)	1.0227(.040)	1.0009(.041)	1.0209(.027)	0.9992(.028)
	1	0.9589(.214)	0.9726(.220)	0.9693(.145)	0.9854(.149)	0.9809(.106)	0.9936(.109)
	-.5	-0.5527(.046)	-0.4979(.049)	-0.5630(.034)	-0.4992(.037)	-0.5522(.023)	-0.4993(.024)
	.2	0.1894(.088)	0.1915(.090)	0.1946(.068)	0.1947(.069)	0.1912(.044)	0.1991(.044)

Note: Par = $\psi = (\beta, \sigma_v^2, \rho, \lambda_2)'$; dgp=1 (normal), 2 (normal mixture), and 3 (chi-square).
X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$, as in Footnote 11.
W_1 is generated according to Queen Contiguity scheme.

Table 2b. Empirical sd and average of estimated standard errors of M-Estimator
SL Model, $T = 3, m = 5$, Parameter configurations as in Table 2a.

		$n = 50$				$n = 100$				$n = 200$			
dgp	ψ	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.047	.050	.045	.043	.040	.041	.039	.038	.026	.026	.026	.025
	1	.144	.156	.141	.137	.101	.108	.102	.101	.074	.074	.073	.072
	.5	.042	.044	.041	.040	.034	.034	.033	.033	.022	.021	.021	.021
	.2	.068	.066	.065	.068	.074	.065	.070	.079	.046	.040	.043	.047
2	1	.047	.055	.045	.043	.039	.044	.039	.038	.025	.027	.025	.025
	1	.285	.104	.142	.252	.208	.066	.102	.192	.146	.042	.072	.140
	.5	.045	.047	.040	.041	.036	.035	.033	.034	.022	.021	.021	.022
	.2	.068	.072	.064	.068	.075	.070	.070	.078	.046	.042	.043	.047
3	1	.046	.052	.045	.044	.041	.042	.039	.038	.026	.027	.025	.025
	1	.220	.121	.142	.195	.159	.080	.103	.146	.110	.053	.072	.106
	.5	.044	.044	.041	.042	.036	.033	.033	.035	.021	.021	.021	.022
	.2	.064	.070	.065	.068	.074	.068	.070	.078	.045	.041	.043	.047
1	1	.049	.053	.048	.047	.039	.041	.039	.038	.026	.027	.026	.026
	1	.146	.156	.141	.137	.100	.107	.101	.100	.073	.074	.072	.072
	.0	.051	.054	.049	.048	.038	.039	.038	.038	.023	.024	.024	.024
	.2	.094	.100	.085	.076	.075	.079	.068	.061	.046	.047	.043	.040
2	1	.048	.059	.048	.046	.040	.044	.039	.038	.026	.028	.026	.026
	1	.292	.103	.143	.256	.205	.065	.102	.193	.148	.041	.072	.141
	.0	.052	.058	.049	.049	.041	.041	.038	.039	.025	.024	.024	.024
	.2	.089	.111	.084	.073	.076	.085	.068	.060	.046	.049	.043	.039
3	1	.050	.056	.048	.047	.041	.042	.039	.039	.026	.027	.026	.026
	1	.212	.121	.142	.195	.150	.079	.102	.147	.109	.053	.072	.106
	.0	.055	.055	.049	.050	.040	.039	.038	.039	.024	.025	.024	.023
	.2	.094	.105	.085	.076	.078	.082	.068	.060	.047	.048	.043	.039
1	1	.048	.054	.049	.047	.042	.043	.041	.040	.028	.028	.027	.027
	1	.144	.155	.140	.136	.100	.106	.101	.099	.071	.073	.071	.071
	-.5	.049	.054	.048	.047	.037	.039	.037	.037	.024	.025	.024	.023
	.2	.092	.094	.081	.076	.068	.068	.062	.060	.046	.044	.042	.040
2	1	.050	.059	.048	.046	.041	.046	.041	.040	.027	.029	.027	.027
	1	.277	.098	.139	.252	.205	.062	.100	.192	.144	.040	.071	.140
	-.5	.051	.059	.048	.047	.038	.041	.037	.037	.024	.025	.024	.024
	.2	.088	.104	.081	.073	.068	.072	.062	.058	.044	.046	.041	.039
3	1	.050	.056	.048	.047	.041	.044	.041	.040	.028	.028	.027	.027
	1	.220	.118	.140	.193	.149	.078	.100	.143	.109	.052	.071	.105
	-.5	.049	.055	.048	.047	.037	.040	.037	.037	.024	.025	.024	.024
	.2	.090	.099	.081	.074	.069	.070	.062	.059	.044	.045	.041	.040

Table 3a. Empirical Mean(sd) of CQMLE and M-Estimator, SLE Model, $T = 3, m = 5$

		$n = 50$		$n = 100$		$n = 200$	
dgp	ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
1	1	0.9955(.055)	0.9980(.056)	1.0091(.037)	0.9993(.038)	1.0038(.025)	1.0008(.025)
	1	0.9264(.134)	0.9589(.144)	0.9500(.099)	0.9766(.105)	0.9573(.068)	0.9851(.072)
	.5	0.4224(.043)	0.4991(.048)	0.4366(.028)	0.4994(.030)	0.4336(.020)	0.5000(.022)
	.2	0.1869(.108)	0.1882(.109)	0.2125(.078)	0.1968(.079)	0.2094(.057)	0.1972(.058)
	.2	0.1320(.210)	0.1421(.193)	0.1318(.164)	0.1510(.151)	0.1561(.128)	0.1678(.120)
2	1	0.9951(.056)	0.9976(.058)	1.0081(.035)	0.9984(.036)	1.0021(.024)	0.9991(.025)
	1	0.9260(.263)	0.9603(.283)	0.9544(.190)	0.9819(.201)	0.9587(.137)	0.9870(.145)
	.5	0.4224(.046)	0.4989(.048)	0.4356(.030)	0.4985(.031)	0.4336(.022)	0.5001(.023)
	.2	0.1811(.128)	0.1831(.126)	0.2087(.081)	0.1933(.081)	0.2073(.057)	0.1955(.058)
	.2	0.1357(.204)	0.1461(.185)	0.1323(.170)	0.1514(.157)	0.1567(.130)	0.1681(.121)
3	1	0.9969(.057)	0.9993(.058)	1.0085(.035)	0.9986(.036)	1.0029(.025)	0.9999(.025)
	1	0.9214(.196)	0.9540(.209)	0.9499(.142)	0.9767(.150)	0.9606(.099)	0.9887(.105)
	.5	0.4249(.044)	0.5007(.048)	0.4372(.028)	0.5002(.031)	0.4332(.021)	0.4997(.023)
	.2	0.1913(.108)	0.1912(.109)	0.2077(.078)	0.1923(.079)	0.2102(.057)	0.1977(.058)
	.2	0.1221(.210)	0.1352(.192)	0.1364(.161)	0.1551(.149)	0.1498(.129)	0.1607(.122)
1	1	1.0247(.053)	0.9971(.055)	1.0205(.038)	0.9973(.039)	1.0202(.026)	0.9991(.027)
	1	0.9409(.140)	0.9550(.145)	0.9605(.097)	0.9726(.100)	0.9769(.068)	0.9902(.070)
	-.5	-0.5629(.048)	-0.4978(.054)	-0.5548(.031)	-0.5001(.034)	-0.5546(.023)	-0.4996(.025)
	.2	0.1920(.119)	0.1878(.135)	0.1750(.108)	0.1784(.125)	0.1763(.076)	0.1894(.085)
	.2	0.1292(.211)	0.1272(.227)	0.1722(.172)	0.1618(.191)	0.1822(.139)	0.1635(.153)
2	1	1.0260(.054)	0.9984(.055)	1.0207(.039)	0.9973(.040)	1.0204(.027)	0.9996(.027)
	1	0.9456(.275)	0.9603(.283)	0.9573(.199)	0.9694(.204)	0.9731(.142)	0.9865(.145)
	-.5	-0.5643(.053)	-0.4994(.057)	-0.5530(.034)	-0.4984(.036)	-0.5542(.023)	-0.4995(.025)
	.2	0.1868(.119)	0.1819(.135)	0.1790(.106)	0.1802(.130)	0.1788(.077)	0.1922(.085)
	.2	0.1390(.211)	0.1381(.228)	0.1666(.171)	0.1592(.192)	0.1854(.131)	0.1663(.145)
3	1	1.0242(.052)	0.9968(.054)	1.0210(.040)	0.9976(.041)	1.0214(.027)	1.0003(.027)
	1	0.9398(.210)	0.9542(.216)	0.9665(.146)	0.9789(.150)	0.9754(.105)	0.9886(.108)
	-.5	-0.5617(.049)	-0.4964(.054)	-0.5536(.034)	-0.4985(.036)	-0.5541(.023)	-0.4993(.024)
	.2	0.1877(.117)	0.1838(.132)	0.1778(.106)	0.1805(.125)	0.1788(.078)	0.1922(.087)
	.2	0.1321(.209)	0.1297(.226)	0.1686(.174)	0.1595(.193)	0.1846(.136)	0.1660(.150)

Note: Par = $\psi = (\beta, \sigma_v^2, \rho, \lambda_2)'$; dgp=1 (normal), 2 (normal mixture), and 3 (chi-square).
X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$, as in Footnote 11.
Spatial weights: Group Interaction for both SE and SL effects; see Footnote 12.

Table 3b. Empirical sd and average of estimated standard errors of M-Estimator
SLE Model, $T = 3, m = 5$, Parameter configurations as in Table 3a.

		$n = 50$				$n = 100$				$n = 200$			
dgp	ψ	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.056	.064	.056	.055	.038	.038	.036	.036	.025	.025	.025	.024
	1	.144	.159	.141	.138	.105	.107	.101	.100	.072	.074	.072	.072
	.5	.048	.049	.045	.045	.030	.030	.029	.029	.022	.021	.021	.021
	.2	.109	.087	.102	.132	.079	.068	.074	.086	.058	.051	.056	.064
	.2	.193	.190	.173	.179	.151	.149	.141	.143	.120	.120	.117	.118
2	1	.058	.070	.056	.055	.036	.041	.036	.036	.025	.027	.025	.024
	1	.283	.111	.142	.248	.201	.067	.101	.189	.145	.043	.072	.140
	.5	.048	.053	.045	.046	.031	.031	.029	.030	.023	.021	.021	.022
	.2	.126	.098	.103	.135	.081	.074	.075	.087	.058	.053	.057	.065
	.2	.185	.213	.172	.174	.157	.162	.142	.141	.121	.126	.117	.116
3	1	.058	.066	.056	.055	.036	.040	.036	.036	.025	.026	.025	.024
	1	.209	.126	.141	.189	.150	.080	.101	.143	.105	.054	.072	.105
	.5	.048	.050	.045	.046	.031	.030	.029	.030	.023	.021	.021	.022
	.2	.109	.089	.100	.129	.079	.071	.075	.086	.058	.052	.056	.064
	.2	.192	.203	.173	.175	.149	.155	.141	.140	.122	.123	.118	.118
1	1	.055	.059	.053	.051	.039	.041	.039	.039	.027	.028	.027	.027
	1	.145	.156	.139	.135	.100	.106	.099	.098	.070	.074	.071	.071
	-.5	.054	.058	.053	.052	.034	.036	.034	.034	.025	.025	.024	.024
	.2	.135	.122	.110	.109	.125	.099	.103	.113	.085	.076	.076	.079
	.2	.227	.227	.201	.199	.191	.176	.171	.177	.153	.143	.140	.141
2	1	.055	.066	.052	.051	.040	.044	.039	.039	.027	.029	.027	.027
	1	.283	.108	.140	.247	.204	.064	.099	.187	.145	.041	.071	.139
	-.5	.057	.065	.053	.052	.036	.039	.034	.034	.025	.026	.024	.025
	.2	.135	.138	.111	.110	.130	.106	.103	.116	.085	.079	.075	.077
	.2	.228	.257	.200	.192	.192	.192	.172	.177	.145	.151	.139	.138
3	1	.054	.061	.052	.051	.041	.043	.039	.039	.027	.028	.027	.027
	1	.216	.123	.139	.189	.150	.079	.100	.144	.108	.052	.071	.105
	-.5	.054	.061	.053	.053	.036	.037	.034	.034	.024	.025	.024	.024
	.2	.132	.127	.110	.111	.125	.103	.103	.112	.087	.077	.076	.078
	.2	.226	.241	.201	.195	.193	.185	.172	.177	.150	.145	.139	.140

Table 4a. Empirical Mean(sd) of CQMLE and M-Estimator, STL Model, $T = 3, m = 5$

		$n = 50$		$n = 100$		$n = 200$	
dgp	ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
1	1	0.9997(.019)	1.0001(.019)	1.0002(.017)	1.0003(.017)	0.9982(.010)	0.9998(.010)
	1	0.9548(.137)	0.9584(.138)	0.9729(.100)	0.9774(.101)	0.9845(.069)	0.9901(.070)
	.5	0.4910(.016)	0.5002(.016)	0.4886(.012)	0.4998(.012)	0.4864(.009)	0.4999(.009)
	.2	0.1951(.042)	0.1969(.042)	0.1997(.033)	0.1981(.033)	0.2012(.022)	0.1992(.023)
	.2	0.2110(.041)	0.2025(.041)	0.2093(.030)	0.2017(.030)	0.2095(.020)	0.2011(.020)
2	1	1.0003(.020)	1.0007(.020)	0.9998(.016)	1.0000(.016)	0.9983(.011)	0.9999(.011)
	1	0.9594(.277)	0.9632(.280)	0.9726(.201)	0.9773(.203)	0.9807(.136)	0.9863(.138)
	.5	0.4908(.015)	0.5001(.015)	0.4890(.012)	0.5002(.012)	0.4867(.009)	0.5002(.009)
	.2	0.1948(.043)	0.1966(.043)	0.1990(.033)	0.1974(.033)	0.2014(.023)	0.1994(.023)
	.2	0.2111(.041)	0.2027(.041)	0.2096(.031)	0.2020(.031)	0.2086(.020)	0.2002(.020)
3	1	0.9997(.019)	1.0002(.019)	1.0004(.016)	1.0006(.016)	0.9983(.011)	0.9999(.011)
	1	0.9605(.218)	0.9643(.219)	0.9760(.151)	0.9806(.153)	0.9819(.106)	0.9875(.107)
	.5	0.4906(.016)	0.4999(.016)	0.4892(.012)	0.5005(.012)	0.4859(.010)	0.4994(.010)
	.2	0.1960(.043)	0.1978(.043)	0.2006(.034)	0.1990(.034)	0.2011(.023)	0.1991(.023)
	.2	0.2103(.042)	0.2018(.042)	0.2084(.030)	0.2008(.030)	0.2100(.021)	0.2016(.021)
1	1	1.0058(.022)	1.0002(.022)	1.0032(.017)	0.9994(.017)	1.0053(.011)	1.0002(.011)
	1	0.9580(.138)	0.9612(.139)	0.9779(.100)	0.9813(.101)	0.9833(.069)	0.9872(.070)
	-.5	-0.5129(.023)	-0.4997(.023)	-0.5128(.016)	-0.4996(.017)	-0.5156(.013)	-0.5000(.013)
	.2	0.1951(.049)	0.1958(.049)	0.2005(.037)	0.1972(.037)	0.1992(.025)	0.1989(.025)
	.2	0.2013(.051)	0.2005(.052)	0.2059(.034)	0.2009(.034)	0.2025(.026)	0.1993(.027)
2	1	1.0056(.022)	1.0001(.022)	1.0036(.018)	0.9999(.018)	1.0054(.011)	1.0002(.011)
	1	0.9530(.281)	0.9564(.283)	0.9823(.203)	0.9859(.205)	0.9894(.144)	0.9934(.145)
	-.5	-0.5134(.023)	-0.5003(.024)	-0.5136(.017)	-0.5004(.017)	-0.5157(.012)	-0.5000(.013)
	.2	0.1938(.050)	0.1944(.050)	0.2022(.037)	0.1989(.037)	0.1998(.024)	0.1995(.024)
	.2	0.2010(.050)	0.2003(.051)	0.2059(.035)	0.2009(.035)	0.2036(.027)	0.2005(.027)
3	1	1.0062(.022)	1.0006(.022)	1.0033(.017)	0.9996(.017)	1.0050(.011)	0.9998(.011)
	1	0.9520(.212)	0.9553(.213)	0.9744(.153)	0.9779(.154)	0.9874(.105)	0.9914(.106)
	-.5	-0.5124(.024)	-0.4992(.024)	-0.5130(.017)	-0.4999(.017)	-0.5157(.012)	-0.4999(.013)
	.2	0.1954(.049)	0.1960(.049)	0.2011(.037)	0.1978(.037)	0.2001(.024)	0.1998(.024)
	.2	0.2023(.050)	0.2015(.050)	0.2063(.035)	0.2014(.035)	0.2029(.026)	0.1997(.027)

Note: Par = $\psi = (\beta, \sigma_v^2, \rho, \lambda_2)'$; dgp=1 (normal), 2 (normal mixture), and 3 (chi-square).
X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 5, 1)$, as in Footnote 11.
Spatial weights: Queen Contiguity for both SL and ST effects.

Table 4b. Empirical sd and average of estimated standard errors of M-Estimator STL Model, $T = 3, m = 5$, Parameter configurations as in Table 4a.

		$n = 50$				$n = 100$				$n = 200$			
dgp	ψ	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.019	.021	.019	.018	.017	.017	.016	.016	.010	.011	.011	.010
	1	.138	.153	.136	.132	.101	.105	.098	.097	.070	.073	.070	.070
	.5	.016	.017	.015	.016	.012	.012	.012	.013	.009	.009	.009	.010
	.2	.042	.046	.045	.048	.033	.035	.035	.036	.023	.023	.023	.024
	.2	.041	.042	.042	.046	.030	.031	.032	.037	.020	.019	.020	.023
2	1	.020	.024	.019	.018	.016	.018	.016	.016	.011	.011	.011	.010
	1	.280	.099	.137	.248	.203	.061	.098	.188	.138	.039	.070	.139
	.5	.015	.019	.015	.015	.012	.013	.012	.013	.009	.010	.009	.010
	.2	.043	.051	.045	.048	.033	.037	.034	.035	.023	.024	.023	.023
	.2	.041	.047	.042	.045	.031	.033	.032	.036	.020	.020	.020	.023
3	1	.019	.022	.019	.018	.016	.018	.016	.016	.011	.011	.011	.010
	1	.219	.118	.137	.189	.153	.077	.099	.142	.107	.051	.070	.103
	.5	.016	.018	.015	.015	.012	.013	.012	.013	.010	.009	.009	.010
	.2	.043	.048	.045	.049	.034	.036	.035	.036	.023	.023	.023	.023
	.2	.042	.044	.042	.046	.030	.032	.032	.037	.021	.020	.020	.023
1	1	.022	.024	.021	.021	.017	.018	.017	.017	.011	.012	.011	.011
	1	.139	.154	.136	.133	.101	.105	.099	.097	.070	.073	.070	.070
	-.5	.023	.026	.023	.023	.017	.017	.017	.016	.013	.013	.013	.013
	.2	.049	.052	.050	.054	.037	.038	.037	.037	.025	.025	.024	.025
	.2	.052	.048	.053	.064	.034	.032	.035	.040	.027	.023	.027	.033
2	1	.022	.027	.021	.020	.018	.019	.017	.017	.011	.012	.011	.011
	1	.283	.099	.136	.245	.205	.062	.099	.190	.145	.040	.071	.140
	-.5	.024	.029	.023	.022	.017	.019	.016	.016	.013	.014	.013	.013
	.2	.050	.057	.049	.053	.037	.041	.037	.036	.024	.026	.024	.025
	.2	.051	.054	.052	.063	.035	.034	.035	.040	.027	.024	.027	.033
3	1	.022	.025	.021	.021	.017	.018	.017	.017	.011	.012	.011	.011
	1	.213	.117	.136	.186	.154	.076	.098	.142	.106	.051	.070	.104
	-.5	.024	.027	.023	.023	.017	.018	.016	.016	.013	.013	.013	.013
	.2	.049	.053	.050	.053	.037	.040	.037	.036	.024	.025	.024	.025
	.2	.050	.051	.052	.063	.035	.032	.035	.040	.027	.023	.027	.033

Table 5a. Empirical Mean(sd) of CQMLE and M-Estimator, STLE Model, $T = 3, m = 5$

		$n = 50$		$n = 100$		$n = 200$	
dgp	ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
1	1	1.0025(.024)	1.0004(.024)	1.0030(.014)	0.9996(.014)	1.0019(.010)	0.9999(.010)
	1	0.9406(.135)	0.9462(.137)	0.9665(.099)	0.9718(.100)	0.9841(.070)	0.9885(.071)
	.3	0.2855(.020)	0.3009(.021)	0.2861(.014)	0.3003(.014)	0.2886(.009)	0.3001(.009)
	.2	0.1961(.060)	0.1978(.061)	0.1968(.032)	0.1994(.032)	0.1968(.026)	0.1990(.026)
	.2	0.2054(.051)	0.2004(.053)	0.2051(.033)	0.1995(.034)	0.2024(.023)	0.2013(.024)
	.2	0.1454(.186)	0.1438(.187)	0.1751(.120)	0.1725(.121)	0.1897(.087)	0.1882(.087)
2	1	1.0019(.024)	0.9998(.024)	1.0035(.014)	1.0001(.014)	1.0018(.011)	0.9998(.011)
	1	0.9468(.280)	0.9528(.284)	0.9708(.194)	0.9763(.196)	0.9788(.145)	0.9832(.146)
	.3	0.2840(.021)	0.2994(.021)	0.2853(.014)	0.2996(.014)	0.2883(.009)	0.2997(.009)
	.2	0.1995(.059)	0.2012(.060)	0.1967(.031)	0.1993(.031)	0.1977(.026)	0.2000(.026)
	.2	0.2035(.052)	0.1986(.053)	0.2053(.033)	0.1998(.034)	0.2012(.024)	0.2001(.024)
	.2	0.1556(.179)	0.1538(.180)	0.1775(.121)	0.1749(.122)	0.1910(.088)	0.1895(.088)
3	1	1.0027(.024)	1.0006(.024)	1.0035(.014)	1.0000(.014)	1.0020(.011)	1.0000(.011)
	1	0.9499(.209)	0.9557(.212)	0.9697(.151)	0.9752(.153)	0.9826(.105)	0.9870(.106)
	.3	0.2844(.020)	0.2999(.021)	0.2857(.014)	0.2999(.014)	0.2887(.009)	0.3001(.009)
	.2	0.1970(.062)	0.1987(.063)	0.1971(.031)	0.1997(.032)	0.1971(.026)	0.1993(.027)
	.2	0.2062(.051)	0.2012(.053)	0.2057(.033)	0.2002(.034)	0.2018(.024)	0.2007(.024)
	.2	0.1478(.180)	0.1460(.182)	0.1786(.120)	0.1759(.122)	0.1870(.089)	0.1854(.090)
1	1	1.0050(.025)	0.9999(.026)	1.0066(.015)	0.9997(.015)	1.0052(.012)	1.0001(.012)
	1	0.9425(.141)	0.9474(.142)	0.9698(.100)	0.9746(.101)	0.9822(.070)	0.9866(.071)
	-.3	-0.3191(.027)	-0.3004(.027)	-0.3170(.018)	-0.2997(.018)	-0.3157(.012)	-0.3002(.012)
	.2	0.1988(.066)	0.1995(.066)	0.2008(.032)	0.1996(.032)	0.1975(.026)	0.1995(.026)
	.2	0.2044(.057)	0.1997(.058)	0.1989(.041)	0.2015(.042)	0.1961(.028)	0.2000(.029)
	.2	0.1487(.187)	0.1478(.190)	0.1757(.123)	0.1781(.125)	0.1884(.087)	0.1874(.088)
2	1	1.0050(.026)	0.9999(.026)	1.0062(.016)	0.9993(.016)	1.0047(.011)	0.9996(.011)
	1	0.9489(.277)	0.9541(.280)	0.9703(.200)	0.9752(.202)	0.9799(.145)	0.9844(.146)
	-.3	-0.3185(.027)	-0.2997(.027)	-0.3170(.018)	-0.2997(.018)	-0.3156(.012)	-0.3000(.012)
	.2	0.1989(.066)	0.1996(.066)	0.2013(.032)	0.2003(.033)	0.1979(.026)	0.2000(.026)
	.2	0.2040(.057)	0.1992(.058)	0.1957(.041)	0.1979(.042)	0.1962(.028)	0.2002(.029)
	.2	0.1448(.181)	0.1438(.184)	0.1693(.121)	0.1714(.123)	0.1895(.087)	0.1886(.088)
3	1	1.0052(.024)	1.0002(.025)	1.0070(.016)	1.0002(.016)	1.0050(.012)	0.9999(.012)
	1	0.9379(.203)	0.9429(.205)	0.9705(.148)	0.9753(.149)	0.9804(.108)	0.9847(.109)
	-.3	-0.3185(.026)	-0.2999(.026)	-0.3174(.017)	-0.3001(.018)	-0.3153(.011)	-0.2998(.012)
	.2	0.1980(.066)	0.1987(.067)	0.1998(.034)	0.1989(.034)	0.1977(.026)	0.1998(.026)
	.2	0.2041(.056)	0.1995(.057)	0.1969(.042)	0.1991(.043)	0.1975(.030)	0.2014(.030)
	.2	0.1509(.182)	0.1502(.184)	0.1691(.123)	0.1711(.125)	0.1880(.086)	0.1870(.087)

Note: Par = $\psi = (\beta, \sigma_v^2, \rho, \lambda_2)'$; dgp=1 (normal), 2 (normal mixture), and 3 (chi-square).
X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 5, 1)$, as in Footnote 11.
Spatial weights: Queen Contiguity for all the SL, ST and SE effects.

Table 5b. Empirical sd and average of estimated standard errors of M-Estimator STLE Model, $T = 3, m = 5$, Parameter configurations as in Table 5a.

		$n = 50$				$n = 100$				$n = 200$			
dgp	ψ	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.024	.027	.024	.024	.014	.015	.014	.014	.010	.011	.011	.011
	1	.137	.155	.135	.132	.100	.106	.098	.096	.071	.073	.070	.070
	.3	.021	.023	.020	.020	.014	.015	.014	.014	.009	.009	.009	.009
	.2	.061	.065	.061	.064	.032	.033	.034	.036	.026	.026	.027	.029
	.2	.053	.048	.053	.070	.034	.029	.035	.046	.024	.018	.024	.036
	.2	.187	.202	.177	.179	.121	.130	.121	.123	.087	.090	.087	.088
2	1	.024	.031	.024	.023	.014	.017	.014	.014	.011	.012	.011	.011
	1	.284	.106	.136	.243	.196	.063	.099	.188	.146	.041	.070	.138
	.3	.021	.026	.020	.020	.014	.016	.014	.014	.009	.009	.009	.009
	.2	.060	.074	.061	.063	.031	.037	.034	.035	.026	.028	.027	.029
	.2	.053	.054	.053	.068	.034	.031	.035	.046	.024	.019	.024	.035
	.2	.180	.233	.176	.171	.122	.144	.121	.118	.088	.096	.087	.086
3	1	.024	.029	.024	.024	.014	.016	.014	.014	.011	.011	.011	.011
	1	.212	.122	.137	.190	.153	.078	.098	.141	.106	.052	.070	.104
	.3	.021	.024	.020	.020	.014	.015	.014	.014	.009	.009	.009	.009
	.2	.063	.069	.061	.064	.032	.035	.034	.036	.027	.027	.027	.029
	.2	.053	.050	.053	.071	.034	.030	.035	.046	.024	.018	.024	.035
	.2	.182	.216	.177	.177	.122	.136	.121	.120	.090	.092	.087	.088
1	1	.026	.028	.025	.025	.015	.017	.016	.016	.012	.012	.011	.011
	1	.142	.155	.135	.132	.101	.105	.098	.097	.071	.073	.070	.070
	-.3	.027	.029	.026	.026	.018	.019	.018	.018	.012	.012	.012	.012
	.2	.066	.069	.065	.067	.032	.034	.035	.039	.026	.027	.028	.029
	.2	.058	.049	.054	.068	.042	.033	.043	.062	.029	.022	.029	.040
	.2	.190	.204	.179	.182	.125	.131	.123	.126	.088	.090	.088	.090
2	1	.026	.032	.025	.024	.016	.018	.016	.015	.011	.012	.011	.011
	1	.280	.107	.136	.243	.202	.063	.098	.188	.146	.041	.070	.138
	-.3	.027	.033	.026	.025	.018	.020	.018	.018	.012	.013	.012	.012
	.2	.066	.078	.064	.066	.033	.037	.035	.038	.026	.028	.027	.029
	.2	.058	.055	.054	.066	.042	.036	.043	.060	.029	.023	.029	.040
	.2	.184	.239	.180	.176	.123	.145	.123	.123	.088	.096	.088	.087
3	1	.025	.030	.025	.024	.016	.017	.016	.015	.012	.012	.011	.011
	1	.205	.121	.135	.185	.149	.078	.098	.141	.109	.052	.070	.103
	-.3	.026	.031	.026	.025	.018	.019	.018	.018	.012	.012	.012	.012
	.2	.067	.073	.065	.067	.034	.035	.035	.039	.026	.027	.027	.029
	.2	.057	.052	.054	.067	.043	.034	.043	.061	.030	.022	.029	.040
	.2	.184	.219	.179	.178	.125	.137	.123	.125	.087	.093	.088	.089