QML Estimation of Dynamic Panel Data Models with Spatial Errors*

Liangjun Su and Zhenlin Yang†
School of Economics, Singapore Management University

October 9, 2012

Abstract

We propose quasi maximum likelihood (QML) estimation of dynamic panel models with spatial errors when the cross-sectional dimension n is large and the time dimension T is fixed. We consider both the random effects and fixed effects models and derive the limiting distributions of the QML estimators under different assumptions on the initial observations. We propose a residual-based bootstrap method for estimating the standard errors of the QML estimators. Monte Carlo simulation shows that both the QML estimators and the bootstrap standard errors perform well in finite samples under a correct assumption on initial observations, but may perform poorly when this assumption is not met.

Key Words: Bootstrap Standard Errors, Dynamic Panel, Fixed Effects, Random Effects, Spatial Error Dependence, Quasi Maximum Likelihood, Initial Observations.

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

1 Introduction

Recently, there has been a growing interest in the estimation of panel data models with cross-sectional or spatial dependence. See, among others, Anselin (1981), Elhorst (2003), Baltagi et al. (2003), Baltagi and Li (2004), Chen and Conley (2001), Pesaran (2004), Kapoor et al. (2007), Baltagi et al. (2007a, b), Mutl and Pfaffermayr (2008), and Lee and Yu (2010a) for an overview on the static spatial panel

*We thank Badi Baltagi, Harry Kelejian, Lung-fei Lee, Peter C. B. Phillips, Ingmar Prucha, and the seminar participants of Singapore Econometrics Study Group Meeting 2006, the 1st World Conference of the Spatial Econometric Association 2007, and the 18th International Panel Data Conference 2012 for their helpful comments. The early version of this paper was completed in 2007 when Su was with Beijing University undertaking research supported by the grant NSFC (70501001). He thanks the School of Economics, Singapore Management University (SMU) for the hospitality during his two-month visit in 2006, and the Wharton-SMU research center for supporting his visit. Zhenlin Yang gratefully acknowledges the research support from the Wharton-SMU research center.

†Corresponding Author: 90 Stamford Road, Singapore 178903. Phone: +65-6828-0852; Fax: +65-6828-0833. Email: zlyang@smu.edu.sg.

data (SPD) models.[1] Adding a dynamic element into a SPD model further increases its flexibility, which has, since Anselin (2001), attracted the attention of many econometricians. The spatial dynamic panel data (SDPD) models can be broadly classified into two categories: one is that described in Anselin et al. (2008) where the dynamic and spatial effects both appear in the model in the forms of lags (in time and spatial) of the response variable, and the other allows the dynamic effect in the same manner but builds the spatial effects into the disturbance term. The former has been studied by Yu et al. (2007, 2008) and Yu and Lee (2007), and the latter by Elhorst (2005), Yang et al. (2006), Mutl (2006), Su and Yang (2007), and Lee and Yu (2010b). Lee and Yu (2010c) provide an excellent survey on the spatial panel data models (static and dynamic) and report some recent developments.

In this paper, we consider the latter type of SDPD model, in particular, the *dynamic panel data model with spatial error*. We focus on the more traditional panel data where the cross-sectional dimension n is allowed to grow but the time dimension T is held fixed (usually small), and follow the quasi-maximum likelihood (QML) approach for model estimation.[2] Elhorst (2005) studies the maximum likelihood estimation (MLE) of this model with fixed effects, but the asymptotic properties of the estimators are not given. Mutl (2006) investigates this model using the method of three-step generalized method of moments (GMM). Yang et al. (2006) consider a more general model where the response is subject to an unknown transformation and estimate the model by MLE. There are two well-known problems inherent from short panel and QML estimation, namely the *assumptions on the initial values* and the *incidental parameters*, and these problems remain for the SDPD model that we consider. In the early version of this paper (Su and Yang, 2007), we derived the asymptotic properties of the quasi-maximum likelihood estimators (QMLEs) of this model under both the random and fixed effects specifications with initial observations treated as either exogenous or endogenous, but methods for estimating the standard errors of the QMLEs were not provided. The main difficulty lies in the estimation of the variance-covariance (VC) matrix of the score function, where the traditional methods based on sample analogues or analytical expressions fail due to the presence of a time lag and spatial errors. This difficulty is now overcome by a residual-based bootstrap method.

For over thirty years of spatial econometrics history, the asymptotic theory for the (Q)ML estimation of spatial models has been taken for granted until the influential paper by Lee (2004), which establishes systematically the desirable consistency and asymptotic normality results for the Gaussian QML estimates of a spatial autoregressive model. He demonstrates that the rate of convergence of the QML estimates may depend on some general features of the spatial weights matrix. More recently, Yu et al. (2008) extend the work of Lee (2004) to spatial dynamic panel data models with fixed effects allowing both T and n to be large. While our work is closely related to theirs, there are clear distinctions. First, unlike Yu et al. (2008) who consider only fixed effects model, we shall consider both random and fixed effects specifications of the individual effects. Second, we shall focus on the case of small T, and deal with the problems of *initial conditions* and *incidental parameters*. In contrast, neither problem arises under the

[1] For alternative approaches to model cross-sectional dependence, see Phillips and Sul (2003), Andrews (2005), Pesaran (2006), Bai (2009), Moon and Weidner (2010), Pesaran and Tosetti (2011), Su and Jin (2012), among others.

[2] A panel with large n and small T, called a short panel, remains the prevalent setting in the majority of empirical microeconometric research (Binder et al., 2005), and evidence from the standard dynamic panel data models shows that QML estimators are more efficient than GMM estimators (Hsiao et al., 2002; Binder et al., 2005).

large-n and large-T setting as considered in Yu et al. (2008). Third, spatial dependence is present only in the error term in our model whereas Yu et al. (2008) consider spatial lag model. It would be interesting to extend our work to the SDPD model with spatial lag, or with both spatial lag and spatial error.

The rest of the paper is organized as follows. Section 2 introduces the model. Section 3 presents the quasi maximum likelihood estimates. Section 4 derives the asymptotic properties of the QMLEs. Section 5 introduces the bootstrap method for standard error estimation. Section 6 presents Monte Carlo results for the finite sample performance of the QMLEs and their estimated standard errors. Section 7 concludes the paper. All the proofs are relegated to the appendix.

To proceed, we introduce some general notation and convention. For a positive integer k, let I_k denote a $k \times k$ identity matrix, ι_k a $k \times 1$ vector of ones, 0_k a $k \times 1$ vector of zeros, and $J_k = \iota_k \iota_k'$, where $'$ denotes transpose. Let $A \otimes B$ denotes the Kronecker product of two matrices A and B. Let $|\cdot|$, $\|\cdot\|$, and $\text{tr}(\cdot)$ denote, respectively, the determinant, the Frobenius norm, and the trace of a matrix. When B is a symmetric matrix, we use $\lambda_{\max}(B)$ and $\lambda_{\min}(B)$ denote its largest and smallest eigenvalues, respectively.

2 Model Specification

We consider the spatial dynamic panel data (SDPD) model of the form

$$y_{it} = \rho y_{i,t-1} + x_{it}'\beta + z_i'\gamma + u_{it}, \;\; i = 1, \cdots, n, \;\; t = 1, \cdots, T, \tag{2.1}$$

where the scalar parameter ρ ($|\rho| < 1$) characterizes the dynamic effect, x_{it} is a $p \times 1$ vector of time-varying exogenous variables, z_i is a $q \times 1$ vector of time-invariant exogenous variables that may include the constant term, dummy variables representing individuals' gender, race, etc., and β and γ are the usual regression coefficients. The disturbance vector $u_t = (u_{1t}, \cdots, u_{nt})'$ is assumed to exhibit both non-observable individual effects and spatially autocorrelated structure, i.e.,

$$u_t = \mu + \varepsilon_t, \tag{2.2}$$

$$\varepsilon_t = \lambda W_n \varepsilon_t + v_t, \tag{2.3}$$

where $\mu = (\mu_1, \cdots, \mu_n)'$, $\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_t)'$, and $v_t = (v_{1t}, \cdots, v_{nt})'$, with μ representing the unobservable individual or space-specific effects, ε_t representing the spatially correlated errors, and v_t representing the random innovations that are assumed to be independent and identically distributed (iid) with mean zero and variance σ_v^2. The parameter λ is a spatial autoregressive coefficient and W_n is a known $n \times n$ spatial weight matrix whose diagonal elements are zero.[3]

Denoting $y_t = (y_{1t}, \cdots, y_{nt})'$, $x_t = (x_{1t}, \cdots, x_{nt})'$, and $z = (z_1, \cdots, z_n)'$, the model has the following reduced-form representation,

$$y_t = \rho y_{t-1} + x_t\beta + z\gamma + u_t, \text{ with } u_t = \mu + B_n^{-1} v_t, \quad t = 1, \cdots, T, \tag{2.4}$$

[3]It is worth mentioning that Eqs (2.1)-(2.3) allow spatial dependence to be present in the random disturbance term ε_t but not in the individual effect μ. See Baltagi, Song and Koh (2003) and Baltagi and Li (2004) for the application of this type of models. Alternatively we can allow both ε_t and μ to follow a spatial autoregressive model in our model as is done by Kapoor et al. (2007) who consider GMM estimation of a static spatial panel model with random effects. Our theory can readily be modified to take into account of the latter case, and we conjecture that a specification test can also be developed to test for the two different specifications.

where $B_n = I_n - \lambda W_n$. The following specifications are essential for the subsequent developments.

We focus on **short panels** where $n \to \infty$ but T is fixed and typically small. Throughout the paper, the **initial observations** designated by y_0 are considered to be available, which can be either **exogenous** or **endogenous**; the individual or space-specific effects μ can be either 'random' or 'fixed', giving the so-called **random effects** and **fixed effects** models. To clarify, we adopt the view that the fundamental distinction between random effects and fixed effects models is not whether the unobserved individual-specific effects μ is random or fixed, but rather whether μ is uncorrelated or correlated with the observed regressors, and make it clear that μ is considered as a random vector in both models.

To give a unified presentation, we adopt a similar framework as Hsiao et al. (2002): (*i*) data collection starts from the 0th period; the process starts from the $-m$th period, i.e., m periods before the start of data collection, $m = 0, 1, \cdots$, and then evolves according to the model specified by (2.4); (*ii*) the starting position of the process y_{-m} is treated as exogenous; hence the exogenous variables (x_t, z) and the errors u_t start to have impact on the response from period $-m + 1$ onwards; (*iii*) all exogenous quantities (y_{-m}, x_t, z) are considered as random and inferences proceed by conditioning on them, and (*iv*) variances of elements of y_{-m} are constant. Thus, when $m = 0$, $y_0 = y_m$ is exogenous, when $m \geq 1$, y_0 becomes endogenous, and when $m = \infty$, the process has reached stationarity.

3 The QML Estimators

In this section we develop quasi maximum likelihood estimates (QMLE) based on Gaussian likelihood for the SDPD model with random effects as well as the SDPD model with fixed effects. For the former, we start with the case of exogenous y_0, and then generalize it to give a unified treatment on the initial values. For the latter, a unified treatment is given directly.[4]

3.1 QMLEs for the random effects model

As indicated above, the main feature of the random effects SDPD model is that the state-specific effect μ is assumed to be uncorrelated with the observed regressors. Furthermore, it is assumed that μ contains iid elements of mean zero and variance σ_μ^2, and is independent of v_t.

Case I: y_0 is exogenous ($m = 0$)**.** In case when y_0 is exogenous, it essentially contains no information with respect to the structural parameters in the system, and thus can be treated as fixed constants. In this case, x_0 is not needed, and the estimation of the system makes use of T periods of data ($t = 1, \cdots, T$). In case when y_0 is endogenously generated from the system (2.4), it contains useful information about the parameters in the model, and hence should be used in the model estimation, particularly when T is small and n is large (Bhargava and Sargan, 1983; Hsiao et al., 2002). In this case x_0 is needed for modelling y_0.

[4] It is well known that when T is fixed the likelihood function for a dynamic panel model depends on the assumptions on the initial observations (Hsiao, 2003). For example, if $|\rho| \geq 1$ or the process $\{x_t\}$ are not stationary, then it does not make sense to assume that the process generating the y_t is the same prior to the periods of observations for $t = 1, \cdots, T$. In this case, it is reasonable to treat y_0 as exogenous. Otherwise, y_0 should be treated as endogenous. In general, y_0 is considered to be exogenous when it is reasonable to expect that y_0 varies "autonomously", independently of the other variables in the model, otherwise it is considered as endogenous.

Conditional on the observed (exogenous) y_0, the distribution of y_1 can be easily derived, and hence the Gaussian quasi-likelihood function based on the observations $y_1, y_2, \cdots, y_T$. Define $Y = (y_1', \cdots, y_T')'$, $Y_{-1} = (y_0', \cdots, y_{T-1}')'$, $X = (x_1', \cdots, x_T')'$, $Z = \iota_T \otimes z$, and $v = (v_1', \cdots, v_T')'$. The SDPD model specified by (2.1)-(2.3) can be written in matrix form:

$$Y = \rho Y_{-1} + X\beta + Z\gamma + u, \text{ with } u = (\iota_T \otimes I_n)\mu + (I_T \otimes B^{-1})v. \tag{3.1}$$

Pretending μ and v follow normal distributions leads to $u \sim N(0, \sigma_v^2 \Omega)$, where

$$\Omega \equiv \Omega(\lambda, \phi_\mu) = \phi_\mu (J_T \otimes I_n) + I_T \otimes (B'B)^{-1}, \tag{3.2}$$

$\phi_\mu = \sigma_\mu^2/\sigma_v^2$, and $J_T = \iota_T \iota_T'$. Note that the dependence of B_n on n and λ is suppressed. The same notational convention is applied to other quantities such as Y, X, Ω, etc., unless confusion arises.

The distribution of u leads to the distribution of $Y - \rho Y_{-1}$, and hence the distribution of Y as the Jacobian of the transformation is one. Let $\theta = (\beta', \gamma', \rho)'$, $\delta = (\lambda, \phi_\mu)'$, and $\psi = (\theta', \sigma_v^2, \delta')'$. Denoting $u(\theta) = Y - \rho Y_{-1} - X\beta - Z\gamma$, the quasi-log-likelihood function of ψ is

$$\mathcal{L}^r(\psi) = -\frac{nT}{2}\log(2\pi) - \frac{nT}{2}\log(\sigma_v^2) - \frac{1}{2}\log|\Omega| - \frac{1}{2\sigma_v^2} u(\theta)'\Omega^{-1}u(\theta). \tag{3.3}$$

Maximizing (3.3) gives the maximum likelihood estimators (MLEs) of ψ if the error components are truly Gaussian and the quasi maximum likelihood estimators (QMLEs) otherwise. Computationally it is more convenient to work with the concentrated log-likelihood by concentrating out the parameters θ and σ_v^2. From (3.3), the constrained QMLEs of θ and σ_v^2, for a given δ, are,

$$\hat{\theta}(\delta) = (\tilde{X}'\Omega^{-1}\tilde{X})^{-1}\tilde{X}'\Omega^{-1}Y \quad \text{and} \quad \hat{\sigma}_v^2(\delta) = \frac{1}{nT}\tilde{u}(\delta)'\Omega^{-1}\tilde{u}(\delta), \tag{3.4}$$

respectively, where $\tilde{X} = (X, Z, Y_{-1})$ and $\tilde{u}(\delta) = Y - \tilde{X}\hat{\theta}(\delta)$. Substituting $\hat{\theta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ given in (3.4) back into (3.3) for θ and σ_v^2, we obtain the concentrated quasi-log-likelihood function of δ:

$$\mathcal{L}_c^r(\delta) = -\frac{nT}{2}[\log(2\pi) + 1] - \frac{nT}{2}\log[\hat{\sigma}_v^2(\delta)] - \frac{1}{2}\log|\Omega|. \tag{3.5}$$

The QMLE $\hat{\delta} = (\hat{\lambda}, \hat{\phi}_\mu)'$ of δ maximizes the concentrated log-likelihood (3.5). The QMLEs of θ and σ_v^2 are given by $\hat{\theta}(\hat{\delta})$ and $\hat{\sigma}_v^2(\hat{\delta})$, respectively. Further, the QMLE of σ_μ^2 is given by $\hat{\sigma}_\mu^2 = \hat{\phi}_\mu \hat{\sigma}_v^2$.

The QML estimation of the random effects SDPD model is seen to be very simple under exogenous y_0. The numerical maximization involves only two parameters, namely, the spatial parameter λ and the variance ratio ϕ_μ. The dynamic parameter ρ is estimated in the same way as the usual regression coefficients and its QMLE has an explicit expression given λ and ϕ_μ.

Case II: y_0 is endogenous ($m \geq 1$)**.** The log-likelihood function (3.3) is derived under the assumption that the initial observation y_0 is exogenously given. If this assumption is not satisfied, maximizing (3.2) generally produces biased estimators (see Bhargava and Sargan, 1983). On the other hand, if the initial observation y_0 is taken as endogenous in the sense that it is generated from the process specified by (2.4), which starts m periods before the 0th period, then y_0 contains useful information about the model parameters and hence should be utilized in the model estimation. We now present a unified set-up for a general m and then argue (see Remark II below) that by letting $m = 0$ it reduces to the case of exogenous y_0. By successive back substitution using (2.4), we have

$$y_0 = \rho^m y_{-m} + \sum_{j=0}^{m-1} \rho^j x_{-j}\beta + z\gamma\frac{1-\rho^m}{1-\rho} + \mu\frac{1-\rho^m}{1-\rho} + \sum_{j=0}^{m-1} \rho^j B^{-1} v_{-j}. \tag{3.6}$$

Letting η_0 and ζ_0 be, respectively, the exogenous and endogenous components of y_0, we have

$$\eta_0 = \rho^m y_{-m} + \sum_{j=0}^{m-1} \rho^j x_{-j}\beta + z\gamma\frac{1-\rho^m}{1-\rho} = \eta_m + x_0\beta + z_m(\rho)\gamma, \tag{3.7}$$

where $\eta_m = \rho^m y_{-m} + \sum_{j=1}^{m-1} \rho^j x_{-j}\beta$ and $z_m(\rho) = z\frac{1-\rho^m}{1-\rho}$; and

$$\zeta_0 = \mu\frac{1-\rho^m}{1-\rho} + \sum_{j=0}^{m-1} \rho^j B^{-1} v_{-j}, \tag{3.8}$$

where $\mathrm{E}(\zeta_0) = 0$ and $\mathrm{Var}(\zeta_0) = \sigma_\mu^2 \left(\frac{1-\rho^m}{1-\rho}\right)^2 I_n + \sigma_v^2 \frac{1-\rho^{2m}}{1-\rho^2}(B'B)^{-1}$. Clearly, both the mean and variance of y_0 are functions of the model parameters and hence y_0 is informative to model estimation. Treating y_0 as exogenous will lose such information and cause bias in model estimation.

However, both $\{x_{-j}, j = 1, \cdots, m-1\}$ for $m \geq 2$ and y_{-m} for $m \geq 1$ in η_m are unobserved, rendering that (3.7) cannot be used as a model for η_0. Some approximations are necessary. In this paper, we follow Bhargava and Sargan (1983) (see also Hsiao, 2003, p.76) and propose a model for the initial observations based on the following fundamental assumptions. Let $\mathbf{x} \equiv (x_0, x_1, \cdots, x_T)$.

Assumption R0: *(i) Conditional on the observables* $\mathbf{x}$ *and* z*, the optimal predictors for* $x_{-j}, j \geq 1$*, are* $\mathbf{x}$ *and the optimal predictors for* $E(y_{-m}), m \geq 1$*, are* $\mathbf{x}$ *and* z*; and (ii) The error resulted from predicting* η_m *using* $\mathbf{x}$ *and* z *is* ζ *such that* $\zeta \sim (0, \sigma_\zeta^2 I_n)$ *and is independent of* u*,* $\mathbf{x}$ *and* z*.*

These assumptions lead immediately to the following model for η_m:

$$\eta_m = \iota_n \pi_1 + \mathbf{x}\pi_2 + z\pi_3 + \zeta \equiv \tilde{\mathbf{x}}\pi + \zeta, \tag{3.9}$$

where $\tilde{\mathbf{x}} = (\iota_n, \mathbf{x}, z)$ and $\pi = (\pi_1, \pi_2', \pi_3')'$. Clearly, the variability of ζ comes from two sources: the variability of y_{-m} and the variability of the prediction error from predicting $\mathrm{E}(y_{-m})$ and $\sum_{j=1}^{m-1} \rho^j x_{-j}\beta$ by $\mathbf{x}$ and z. Hence, we have the following model for y_0 based on (3.6)-(3.9):

$$y_0 = \tilde{\mathbf{x}}\pi + x_0\beta + z_m(\rho)\gamma + u_0, \;\; u_0 = \zeta + \zeta_0. \tag{3.10}$$

The 'initial' error vector u_0 is seen to contain three components: ζ, $\mu\frac{1-\rho^m}{1-\rho}$, and $\sum_{j=0}^{m-1} \rho^j B^{-1} v_{-j}$, being, respectively, the prediction error from predicting the unobservables, the cumulative random effects up to the 0th period, and the 'cumulative' spatial effects and random shocks up to the 0th period. The term $z_m(\rho)\gamma = z\frac{1-\rho^m}{1-\rho}\gamma$ represents the cumulative impact of the time-invariant variables z up to period 0 and needs not be predicted. However, the predictors for η_m still include z, indicating that (i) the mean of y_{-m} is allowed to be linearly related to z and (ii) ρ^m may not be small such that the effect of y_{-m} on η_m is not negligible. If ρ^m is small which occurs when either m is large or ρ is small, the impact of y_{-m}

to η_m can be ignored, and the term $z\pi_3$ in (3.10) should be omitted. Some details about the cases with small ρ^m are given latter. For the cases where ρ^m is not negligible, one can easily show that, under strict exogeneity of $\mathbf{x}$ and z, $\mathrm{E}(u_0) = 0$,

$$\mathrm{E}(u_0 u_0') = \sigma_\zeta^2 I_n + \sigma_\mu^2 a_m^2 I_n + \sigma_v^2 b_m (B'B)^{-1}, \text{ and } \mathrm{E}(u_0 u') = \sigma_\mu^2 a_m (\iota_T' \otimes I_n),$$

where $a_m \equiv a_m(\rho) = \frac{1-\rho^m}{1-\rho}$ and $b_m \equiv b_m(\rho) = \frac{1-\rho^{2m}}{1-\rho^2}$. Let $u^* = (u_0', u')'$. Under the normality assumption for the original error components μ and v, and the 'new' prediction error ζ, we have $u^* \sim N(0, \sigma_v^2 \Omega^*)$, where Ω^* is $n(T+1) \times n(T+1)$ and has the form:

$$\Omega^* \equiv \Omega^*(\rho, \lambda, \phi_\mu, \phi_\zeta) = \begin{pmatrix} \phi_\zeta I_n + \phi_\mu a_m^2 I_n + b_m (B'B)^{-1} & \phi_\mu a_m (\iota_T' \otimes I_n) \\ \phi_\mu a_m (\iota_T \otimes I_n) & \Omega \end{pmatrix}, \tag{3.11}$$

$\phi_\zeta = \sigma_\zeta^2 / \sigma_v^2$, and Ω is given by (3.2). This leads to the joint distribution of $(y_0', (Y - \rho Y_{-1})')'$, and hence the joint distribution of $(y_0', Y')'$ or the likelihood function. Again, the arguments of Ω^* are frequently suppressed should no confusion arise.

Now let $\theta = (\beta', \gamma', \pi')'$, $\delta = (\rho, \lambda, \phi_\mu, \phi_\zeta)'$, and $\psi = (\theta', \sigma_v^2, \delta')'$. Based on (2.4) and (3.10), the Gaussian quasi-log-likelihood function of ψ has the form:

$$\mathcal{L}^{rr}(\psi) = -\frac{n(T+1)}{2}\log(2\pi) - \frac{n(T+1)}{2}\log(\sigma_v^2) - \frac{1}{2}\log|\Omega^*| - \frac{1}{2\sigma_v^2} u^*(\theta,\rho)' \Omega^{*-1} u^*(\theta,\rho), \tag{3.12}$$

where $u^*(\theta, \rho) = \{(y_0 - x_0\beta - z_m(\rho)\gamma - \tilde{\mathbf{x}}\pi)', (Y - \rho Y_{-1} - X\beta - Z\gamma)'\}' \equiv Y^* - X^*\theta$,

$$Y^* = Y^*(\rho) = \begin{pmatrix} y_0 \\ Y - \rho Y_{-1} \end{pmatrix} \quad \text{and} \quad X^* = X^*(\rho) = \begin{pmatrix} x_0 & z_m(\rho) & \tilde{\mathbf{x}} \\ X & Z & 0_{nT \times k} \end{pmatrix}.$$

Maximizing (3.12) gives MLE of ψ if the error components are truly Gaussian and the QMLE otherwise. Similar to **Case I**, we work with the concentrated quasi-log-likelihood by concentrating out the parameters θ and σ_v^2. The constrained QMLEs of θ and σ_v^2, given δ, are

$$\hat{\theta}(\delta) = (X^{*\prime}\Omega^{*-1}X^*)^{-1}X^{*\prime}\Omega^{*-1}Y^* \quad \text{and} \quad \hat{\sigma}_v^2(\delta) = \frac{1}{n(T+1)}\tilde{u}^*(\delta)'\Omega^{*-1}\tilde{u}^*(\delta), \tag{3.13}$$

where $\tilde{u}^*(\delta) = u^*(\hat{\theta}(\delta), \rho) = Y^* - X^*\hat{\theta}(\delta)$, and $\hat{\theta}(\delta) = (\hat{\beta}(\delta)', \hat{\gamma}(\delta)', \hat{\pi}(\delta)')'$. Substituting $\hat{\theta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ back into (3.12) for θ and σ_v^2, we obtain the concentrated quasi-log-likelihood function of δ:

$$\mathcal{L}_c^{rr}(\delta) = -\frac{n(T+1)}{2}[\log(2\pi) + 1] - \frac{n(T+1)}{2}\log\hat{\sigma}_v^2(\delta) - \frac{1}{2}\log|\Omega^*|. \tag{3.14}$$

Maximizing the concentrated quasi-log-likelihood (3.14) gives the QMLE of δ, denoted by $\hat{\delta} = (\widehat{\rho}, \hat{\lambda}, \hat{\phi}_\mu, \hat{\phi}_\zeta)'$. The QMLEs of θ and σ_v^2 are thus given by $\hat{\theta}(\hat{\delta})$ and $\hat{\sigma}_v^2(\hat{\delta})$, respectively, and these of σ_μ^2 and σ_ζ^2 are given by $\hat{\sigma}_\mu^2 = \hat{\phi}_\mu \hat{\sigma}_v^2$ and $\hat{\sigma}_\zeta^2 = \hat{\phi}_\zeta \hat{\sigma}_v^2$, respectively.[5]

Remark I: To utilize the information contained in the n endogenous initial observations y_0, we have introduced $k = p(T+1) + q + 1$ additional parameters (π, σ_ζ^2) in the model (3.9). Besides the bias issue,

[5] Unlike the case of exogenous y_0, the dynamic parameter ρ now becomes a nonlinear parameter that has to be estimated, together with λ, ϕ_μ and ϕ_ζ, through a nonlinear optimization process.

efficiency gain by utilizing additional n observations is reflected by $n - k$. Apparently, the condition $n > k$ has to be satisfied in order for π and σ_ζ^2 to be identified. If both T and p are not so small ($T = 9$ and $p = 10$, say), one may consider to replace the regressors $\mathbf{x}$ in (3.9) by the most relevant ones (to the past), x_0 and x_1, say, or simply by $\overline{x} = (T+1)^{-1}\sum_{t=0}^{T} x_t$. In this case $k = 2p + q + 1$, and $p + q + 1$, respectively.

Remark II: When y_0 is exogenous, model (3.10) becomes $y_0 = \tilde{\mathbf{x}}\pi + u_0$, where $u_0 \sim (0, \sigma_0^2 I_n)$ and is independent of u. In this case, we have $\Omega^* = diag(\sigma_0^2 I_n, \Omega)$. Model estimation may proceed by letting $m = 0$ in (3.14), and the results are almost identical to those from maximizing (3.5). A special case of this is the one considered in Hsiao (2003, p.76, Case IIa) where $y_{i0}'s$ are simply assumed to be iid independent of μ_i'. If $y_{i0}'s$ are allowed to be correlated with μ_i' (Case IIb, Hsiao, 2003, p.76), the model becomes a special case of endogenous y_0 as considered above.

Remark III: In general, m is unknown. In dealing with a dynamic panel model with fixed effects but without spatial dependence, Hsiao et al. (2002) recommend treating m or a function of it as a free parameter, which is estimated jointly with the other model parameters. However, we note that their approach requires $\rho \neq 0$, as when $\rho = 0$, m disappears from the model and hence cannot be identified. Elhorst (2005) recommends that an appropriate value of m should be chosen in advance. We concur with his view for two reasons: (i) an empirical study often tells roughly what the m value is (see, e.g., the application considered by Elhorst), and (ii) the estimation is often not sensitive to the choice of m unless it is very small ($m \leq 2$), and $|\rho|$ is close to 1, as evidenced by the Monte Carlo results given in Section 6.

While the results given above are under a rather general set-up, some special cases deserve detailed discussions, which are (a) $m = 1$, (b) $m = \infty$, and (c) $\rho = 0$.

(a) m=1. When the process starts just one period before the start of data collection, the model (3.10) becomes $y_0 = \rho y_{-1} + x_0\beta + z\gamma + \mu + B^{-1}v_0$, $z_m(\rho) = z$, and

$$\Omega^* = \begin{pmatrix} (\phi_\zeta + \phi_\mu)I_n + (B'B)^{-1}, & \phi_\mu(\iota_T' \otimes I_n) \\ \phi_\mu(\iota_T' \otimes I_n), & \Omega \end{pmatrix}.$$

In this case, ρ becomes a linear parameter again and the estimation can be simplified by putting ρ together with β, γ and π which can be concentrated out from the likelihood function. Now, denoting the response vector and the regressor matrix by:

$$\tilde{Y} = \begin{pmatrix} y_0 \\ Y \end{pmatrix} \quad \text{and} \quad \tilde{X} = \begin{pmatrix} x_0 & z & 0_{n\times 1} & \tilde{\mathbf{x}} \\ X & Z & Y_{-1} & 0_{nT\times k} \end{pmatrix},$$

the estimation proceeds with $\theta = (\beta', \gamma', \rho, \pi)'$ and $\delta = (\lambda, \phi_\mu, \phi_\zeta)'$.

(b) m=∞. When the process has reached stationarity ($m \to \infty$ and $|\rho| < 1$), the model for the initial observations becomes $y_0 = \sum_{j=0}^{\infty}\rho^j x_{-j}\beta + \frac{z\gamma}{1-\rho} + \frac{\mu}{1-\rho} + \sum_{j=0}^{\infty}\rho^j B^{-1}v_{-j}$. As $\eta_\infty = \sum_{j=0}^{\infty}\rho^j x_{-j}\beta$ involves only the time-varying regressors, its optimal predictors should be $(\iota_n, \mathbf{x})$. The estimation proceeds by letting $z_m(\rho) = z_\infty(\rho) = \frac{z}{1-\rho}$, $a_m = a_\infty = \frac{1}{1-\rho}$, $b_m = b_\infty = \frac{1}{1-\rho^2}$, $\tilde{\mathbf{x}} = (\iota, \mathbf{x})$, and $\pi = (\pi_1, \pi_2')'$.

(c) $\rho = 0$. When the true value of the dynamic parameter is zero, the model becomes static with $y_t = x_t\beta + z\gamma + \mu + B^{-1}v_t,\ t = 0, 1, \cdots, T$. At this point, the true values for all the added parameters, π and σ_ζ, are automatically zero.

3.2 QMLEs for the fixed effects model

In this section, we consider the QML estimation of the SDPD model with fixed effects, i.e., the vector of unobserved individual-specific effects μ in model (2.4) is allowed to correlate with the time-varying regressors x_t. Due to this unknown correlation, μ acts as if they are n free parameters, and with T fixed the model cannot be consistently estimated due to the incident parameter problem. Following the standard practice, we eliminate μ by first-differencing (2.4) to give

$$\Delta y_t = \rho \Delta y_{t-1} + \Delta x_t \beta + \Delta u_t, \;\; \Delta u_t = B^{-1} \Delta v_t, \;\; t = 2, 3, \cdots, T. \tag{3.15}$$

Clearly, (3.15) is not defined for $t = 1$ as Δy_1 depends on Δy_0 and the latter is not observed. Thus, even if y_0 (hence Δy_0) is exogenous, one cannot formulate the likelihood function by conditioning on Δy_0 as in the early case. To obtain the joint distribution of $\Delta y_1, \Delta y_2, \cdots, \Delta y_T$ or the transformed likelihood function for the remaining parameters based on (3.15), a proper approximation for Δy_1 needs to be made so that its marginal distribution can be obtained, whether y_0 is exogenous or endogenous. We present a unified treatment for the fixed effects model where the initial observations y_0 can be exogenous ($m = 0$) as well as endogenous ($m \geq 1$).

Under the general specifications given at the end of Section 2, continuous back substitution to the previous $m(\geq 1)$ periods leads to

$$\Delta y_1 = \rho^m \Delta y_{-m+1} + \sum_{j=0}^{m-1} \rho^j \Delta x_{1-j} \beta + \sum_{j=0}^{m-1} \rho^j B^{-1} \Delta v_{1-j}. \tag{3.16}$$

Note that (*i*) Δy_{-m+1} represents the changes after the process has made its first move, called the *initial endowment*; (*ii*) while the starting position y_{-m} is assumed exogenous, the initial endowment Δy_{-m+1} is endogenous, and (*iii*) when $m = 0$, $\Delta y_{-m+1} = \Delta y_1$, i.e., the initial endowment becomes the *observed initial difference*. The effect of the initial endowment decays as m increases. However, when m is small, their effect can be significant, and hence a proper approximation to it is important. In general, write $\Delta y_1 = \Delta \eta_1 + \Delta \zeta_1$, where $\Delta \eta_1$ and $\Delta \zeta_1$, the exogenous and endogenous components of Δy_1, are given as

$$\Delta \eta_1 = \rho^m E(\Delta y_{-m+1}) + \sum_{j=0}^{m-1} \rho^j \Delta x_{1-j} \beta \equiv \eta_m + \Delta x_1 \beta, \tag{3.17}$$

$$\Delta \zeta_1 = \rho^m [\Delta y_{-m+1} - E(\Delta y_{-m+1})] + \sum_{j=0}^{m-1} \rho^j B^{-1} \Delta v_{1-j}, \tag{3.18}$$

where $\eta_m = \rho^m E(\Delta y_{-m+1}) + \sum_{j=1}^{m-1} \rho^j \Delta x_{1-j} \beta$. Note that when $m = 0$, the summation terms in (3.17) and (3.18) should vanish, and as a result $\Delta \eta_1 = E(\Delta y_1)$ and $\Delta \zeta_1 = \Delta y_1 - E(\Delta y_1)$.

Clearly, the observations Δx_{1-j}, $j = 1, \cdots, m-1$, $m \geq 2$, are not available, and the structure of $E(\Delta y_{-m+1})$, $m \geq 1$, is unknown. Hence η_m is completely unknown. Furthermore, as η_m is an $n \times 1$ vector, it cannot be treated as a free parameter vector to be estimated; otherwise the incidental parameters problem will be confronted again.[6] Hsiao et al. (2002) remark that to get around this problem, the expected value of η_1, conditional on the observables, has to be a function of a finite number of

[6] Unless the original model (2.4) does not contain time-varying variables as in Anderson and Hsiao (1981).

parameters, and that such a condition can hold provided that $\{x_{it}\}$ are trend-stationary (with a common deterministic linear trend) or first-difference stationary processes. Letting $\Delta\mathbf{x} = (\Delta x_1, \cdots, \Delta x_T)$, we have the following fundamental assumptions.

Assumption F0: *(i) The optimal predictors for Δx_{1-j}, $j = 1, 2, \cdots$ and $E(\Delta y_{-m+1}), m = 0, 1, \cdots$, conditional on the observables, are $\Delta\mathbf{x}$; (ii) Collectively, the errors from using $\Delta\mathbf{x}$ to predict η_m is $\epsilon \sim (0, \sigma_\epsilon^2 I_n)$, and (iii) $y_{-m} = E(y_{-m}) + e$, where $e \sim (0, \sigma_e^2 I_n)$.*

Assumption F0(i) and Assumption F0(ii) lead immediately to a 'predictive' model for η_m:

$$\eta_m = \pi_1 \iota_n + \Delta\mathbf{x}\,\pi_2 \equiv \tilde{\Delta}\mathbf{x}\,\pi + \epsilon, \quad m = 0, 1, \cdots,$$

where $\tilde{\Delta}\mathbf{x} = (\iota_n, \Delta\mathbf{x})$ and $\pi = (\pi_1, \pi_2')'$. Thus, $\Delta\eta_1$ defined in (3.17) can be predicted by: $\Delta\eta_1 = \tilde{\Delta}\mathbf{x}\,\pi + \Delta x_1\beta + \epsilon$. The original theoretical model (2.1) and Assumption F0(iii) lead to

$$\Delta y_{-m+1} - E(\Delta y_{-m+1}) = B^{-1}v_{-m+1} - e, \quad m = 0, 1, \cdots,$$

which gives $\Delta\zeta_1 = -\rho^m e + \rho^m B^{-1}v_{-m+1} + \sum_{j=0}^{m-1}\rho^j B^{-1}\Delta v_{1-j}$ when $m \geq 1$, and $-e + B^{-1}v_1$ when $m = 0$. We thus have the following model for the observed initial difference,

$$\Delta y_1 = \tilde{\Delta}\mathbf{x}\pi + \Delta x_1\beta + \epsilon + \Delta\zeta_1 \equiv \tilde{\Delta}\mathbf{x}\pi + \Delta x_1\beta + \Delta\tilde{u}_1, \tag{3.19}$$

where $\Delta\tilde{u}_1 = \epsilon + \Delta\zeta_1 = \epsilon - \rho^m e + \rho^m B^{-1}v_{-m+1} + \sum_{j=0}^{m-1}\rho^j B^{-1}\Delta v_{1-j}$. Let $\zeta = \epsilon - \rho^m e$. By assumption, the elements of $\zeta = \epsilon - \rho^m e$ are iid with variance $\sigma_\zeta^2 = \sigma_\epsilon^2 + \sigma_e^2\rho^{2m}$.

Note that when $m = 0$, $\Delta u_1 = \epsilon - e + B^{-1}v_1$. The approximation (3.19) is associated with Bhargava and Sargan's (1983) approximation for the standard dynamic random effects model with endogenous initial observations. See Ridder and Wansbeek (1990) and Blundell and Smith (1991) for a similar approach. By construction, we can verify that under strict exogeneity of x_{it}, i.e., $E(\zeta_i|\Delta x_{i,1}, \cdots, \Delta x_{i,T}) = 0$, and independence between ζ and $\{\Delta v_{1-j}, j = 0, 1, \cdots, m-1\}$,

$$E(\Delta\tilde{u}_1\Delta\tilde{u}_1') = \sigma_\zeta^2 I_n + \sigma_v^2 c_m (B'B)^{-1} = \sigma_v^2 B^{-1}(\phi_\zeta BB' + c_m I_n)B'^{-1}, \text{ and} \tag{3.20}$$

$$E(\Delta\tilde{u}_1\Delta u_t') = -\sigma_v^2(B'B)^{-1} \text{ for } t = 2, \text{ and } 0 \text{ for } t = 3, 4, \cdots, T, \tag{3.21}$$

where $c_m \equiv c_m(\rho) = \frac{2}{1+\rho} - \frac{\rho^{2m}(1-\rho)}{1+\rho}$ and $\phi_\zeta = \sigma_\zeta^2/\sigma_v^2$. Note that $c_0 = 1$, $c_\infty = \frac{2}{1+\rho}$ and $c_m(0) = 2$.

Letting $\Delta u = (\Delta\tilde{u}_1', \Delta u_2', \cdots, \Delta u_T')$, we have $\text{Var}(\Delta u) = \sigma_v^2\Omega^\dagger$, where

$$\Omega^\dagger \equiv \Omega^\dagger(\rho, \lambda, \phi_\zeta) = (I_T \otimes B^{-1})H_E(I_T \otimes B'^{-1}), \tag{3.22}$$

$E = \phi_\zeta BB' + c_m I_n$, and H_E is an $nT \times nT$ matrix defined as

$$H_E = \begin{pmatrix} E & -I_n & 0 & \cdots & 0 & 0 & 0 \\ -I_n & 2I_n & -I_n & \cdots & 0 & 0 & 0 \\ 0 & -I_n & 2I_n & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2I_n & -I_n & 0 \\ 0 & 0 & 0 & \cdots & -I_n & 2I_n & -I_n \\ 0 & 0 & 0 & \cdots & 0 & -I_n & 2I_n \end{pmatrix}. \tag{3.23}$$

The expression for $\Omega^\dagger$ given in (3.22) greatly facilitates the calculation of the determinant and inverse of $\Omega^\dagger$ as seen in the subsequent subsection. Derivations of score and Hessian matrix requires the derivatives of $\Omega^\dagger$, which can be made much easier based on the following alternative expression

$$\Omega^\dagger = \phi_\zeta(\ell_1 \otimes I_n) + h_{c_m} \otimes (B'B)^{-1}, \tag{3.24}$$

where ℓ_1 is a $T \times T$ matrix with 1 in its top-left corner and zero elsewhere, and h_{c_m} is h_s defined in Section 3.3 with s replaced by c_m.

In the following, we simply refer to the dimension of π to be k. Now let $\theta = (\beta', \pi')'$, $\delta = (\rho, \lambda, \phi_\zeta)'$, and $\psi = (\theta', \sigma_v^2, \delta')'$. Note that ψ is a $(p+k+4) \times 1$ vector of unknown parameters. Based on (3.15) and (3.19), the Gaussian quasi-log-likelihood of ψ has the form:

$$\mathcal{L}^f(\psi) = -\frac{nT}{2}\log(2\pi) - \frac{nT}{2}\log(\sigma_v^2) - \frac{1}{2}\log\left|\Omega^\dagger\right| - \frac{1}{2\sigma_v^2}\Delta u(\theta,\rho)'\Omega^{\dagger-1}\Delta u(\theta,\rho), \tag{3.25}$$

where $\Delta u(\theta,\rho) = \Delta Y^\dagger(\rho) - \Delta X^\dagger \theta$,

$$\Delta Y^\dagger(\rho) = \begin{pmatrix} \Delta y_1 \\ \Delta y_2 - \rho\Delta y_1 \\ \vdots \\ \Delta y_T - \rho\Delta y_{T-1} \end{pmatrix}, \text{ and } \Delta X^\dagger = \begin{pmatrix} \Delta x_1 & \tilde{\Delta}\mathbf{x} \\ \Delta x_2 & 0_{n\times k} \\ \vdots & \vdots \\ \Delta x_T & 0_{n\times k} \end{pmatrix}.$$

Maximizing (3.25) gives the Gaussian MLE or QMLE of ψ. First, given $\delta = (\rho, \lambda, \phi_\zeta)'$, the constrained MLEs or QMLEs of θ and σ_v^2 are, respectively,

$$\hat{\theta}(\delta) = (\Delta X^{\dagger\prime}\Omega^{\dagger-1}\Delta X^\dagger)^{-1}\Delta X^{\dagger\prime}\Omega^{\dagger-1}\Delta Y^\dagger(\rho) \quad \text{and} \quad \hat{\sigma}_v^2(\delta) = \frac{1}{nT}\widetilde{\Delta} u(\delta)'\Omega^{\dagger-1}\widetilde{\Delta} u(\delta), \tag{3.26}$$

where $\widetilde{\Delta} u(\delta)$ equals $\Delta u(\theta,\rho)$ with θ being replaced by $\hat{\theta}(\delta)$. Substituting $\hat{\theta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ back into (3.25) for θ and σ_v^2, we obtain the concentrated quasi-log-likelihood function of δ:

$$\mathcal{L}_c^f(\delta) = -\frac{nT}{2}[\log(2\pi)+1] - \frac{nT}{2}\log\hat{\sigma}_v^2(\delta) - \frac{1}{2}\log|\Omega^\dagger|. \tag{3.27}$$

The QMLE $\delta = (\hat{\rho}, \hat{\lambda}, \hat{\phi}_\zeta)'$ of δ maximizes the concentrated quasi-log-likelihood (3.27). The QMLEs of θ and σ_v^2 are given by $\hat{\theta}(\hat{\delta})$ and $\hat{\sigma}_v^2(\hat{\delta})$, respectively. Further, the QMLE of σ_ζ^2 are given by $\hat{\sigma}_\zeta^2 = \hat{\phi}_\zeta\hat{\sigma}_v^2$.

Remark IV: We require that $n > pT+1$ for the identification of the parameters in (3.19). When this is too demanding, it can be addressed in the same manner as in the random effects model by choosing variables $\Delta\tilde{\mathbf{x}}$ with a smaller dimension. For example, replacing $\Delta\mathbf{x}$ in (3.19) by $\overline{\Delta x} = T^{-1}\sum_{t=1}^T \Delta x_t$ gives $\Delta\tilde{\mathbf{x}} = (\iota_n, \overline{\Delta x})$, and dropping $\Delta\mathbf{x}$ in (3.19) gives $\tilde{\Delta}\mathbf{x} = \iota_n$. In each case, the variance-covariance structure of Δu remains the same.

Remark V: Hsiao et al. (2002, p.110), in dealing with a dynamic panle data model without spatial effect, recommend treating $c_m(\rho)$ as a free parameter to be estimated together with other model parameters. This essentially requires that $\rho \neq 0$ and m be an unknown number. Note that $c_m(0) = 2$ and $c_\infty(\rho) = 2/(1+\rho)$, which become either a constant or a pure function of ρ. Our set-up allows both $\rho = 0$ and $m = \infty$ so that a test for the existence of dynamics can be carried out and a stationary model can be fit. As in the case of the random effects model, we again treat m as known, chosen in advance based on the given data (see Remark III given in section 3.2).

3.3 Some computational notes

Maximization of $\mathcal{L}_c^r(\delta)$, $\mathcal{L}_c^{rr}(\delta)$ and $\mathcal{L}_c^f(\delta)$ involves repeated evaluations of the inverse and determinants of the $nT \times nT$ matrices Ω and $\Omega^\dagger$, and the $n(T+1) \times n(T+1)$ matrix Ω^*. This can be a great burden when n or T or both are large. By Magnus (1982, p.242), the following identities can be used to simplify the calculation involving Ω defined in (3.1):

$$\begin{aligned} |\Omega| &= |(B'B)^{-1} + \phi_\mu T I_n| \cdot |B|^{-2(T-1)}, & (3.28)\\ \Omega^{-1} &= T^{-1} J_T \otimes \left((B'B)^{-1} + \phi_\mu T I_n\right)^{-1} + (I_T - T^{-1} J_T) \otimes (B'B). & (3.29)\end{aligned}$$

The above formulae reduce the calculations of the inverse and determinant of an $nT \times nT$ matrix to the calculations of those of several $n \times n$ matrices, where the key element is the $n \times n$ matrix B. By Griffith (1988), calculations of the determinants can be further simplified as:

$$|B| = \prod_{i=1}^{n}(1-\lambda w_i), \text{ and } |(B'B)^{-1} + \phi_\mu T I_n| = \prod_{i=1}^{n}[(1-\lambda w_i)^{-2} + \phi_\mu T], \tag{3.30}$$

where $w_i's$ are the eigenvalues of W. The above simplifications are also used in Yang et al. (2006).

For the determinant and inverse of Ω^* defined in (3.11), let $\omega_{11} = \phi_\zeta I_n + \phi_\mu a_m^2 I_n + b_m (B'B)^{-1}$, $\omega_{21} = \omega_{12}' = \phi_\mu a_m (\iota_T \otimes I_n)$, and $D = \omega_{11} - \omega_{12}\Omega^{-1}\omega_{21}$. We have by using the formulas for a partitioned matrix (e.g., Magnus and Neudecker, 2002, p.106), $|\Omega^*| = |\Omega| \cdot |D|$, and

$$\Omega^{*-1} = \begin{pmatrix} D^{-1} & -D^{-1}\omega_{12}\Omega^{-1} \\ -\Omega^{-1}\omega_{21}D^{-1} & \Omega^{-1} + \Omega^{-1}\omega_{21}D^{-1}\omega_{12}\Omega^{-1} \end{pmatrix}. \tag{3.31}$$

Thus, the calculations of the determinant and inverse of the $n(T+1) \times n(T+1)$ matrix Ω^* are reduced to the calculations of those of the $n \times n$ matrix D, and those of Ω given in (3.28) and (3.29).

For the determinant and inverse of $\Omega^\dagger$ defined in (3.22), by the properties of matrix operation,

$$\begin{aligned} |\Omega^\dagger| &= |(I_T \otimes B^{-1})| \cdot |H_E| \cdot |(I_T \otimes B'^{-1})| = |B|^{-2T}|H_E|, \\ \Omega^{\dagger -1} &= (I_T \otimes B'^{-1})^{-1} H_E^{-1} (I_T \otimes B^{-1})^{-1} = (I_T \otimes B') H_E^{-1} (I_T \otimes B), \end{aligned}$$

where $|H_E| = |TE - (T-1)I_n = \prod_{i=1}^n [T\phi_\zeta(1-\lambda w_i)^2 + Tc_m - T + 1]$ as in (3.30), and

$$H_E^{-1} = (1-T)(h_0^{-1} \otimes E^{*-1}) + (h_1^{-1} - (1-T)h_0^{-1}) \otimes (E^{*-1}E), \tag{3.32}$$

where $E^* = TE - (T-1)I_n$, and the $T \times T$ matrices $h_s, s = 0, 1$, are

$$h_s = \begin{pmatrix} s & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix},$$

as in Hsiao et al. (2002, Appendix B), who also give $|h_s| = 1 + T(s-1)$ and the expression for h_s^{-1}.

4 Asymptotic Properties of the QMLEs

In this section we study the consistency and asymptotic normality of the proposed QML estimators for the dynamic panel data models with spatial errors. We first state and discuss a set of generic assumptions applicable to all three scenarios discussed in Section 3. Then we proceed with each specific scenario where, under some additional assumptions, the key asymptotic results are presented. To facilitate the presentation, some general notation (old and new) is given.

General notation: (*i*) recall $\psi = (\theta', \sigma_v^2, \delta')'$, where θ and σ_v^2 are the linear and scale parameters and can be concentrated out from the likelihood function, and δ is the vector of nonlinear parameters left in the concentrated likelihood function. Let $\psi_0 = (\theta_0', \sigma_{v0}^2, \delta_0')'$ be the true parameter vector. Let $\mathbf{\Psi}$ be the parameter space of ψ, and $\mathbf{\Delta}$ the space of δ. (*ii*) A parametric function, or vector, or matrix, evaluated at ψ_0, is denoted by adding a subscript $_0$, e.g., $B_0 = B|_{\lambda=\lambda_0}$, and similarly for Ω_0, Ω_0^*, $\Omega_0^\dagger$, etc. (*iii*) The common expectation and variance operators 'E' and 'Var' correspond to ψ_0.

4.1 Generic assumptions

To provide a rigorous analysis of the QMLEs, we need to assume different sets of conditions based on different model specifications. Nevertheless, for both the random and fixed effects specifications we first make the following generic assumptions.

Assumption G1: *(i) The available observations are: $(y_{it}, x_{it}, z_i), i = 1, \cdots, n,\ t = 0, 1, \cdots, T$, with $T \geq 2$ fixed and $n \to \infty$; (ii) The disturbance vector $u_t = (u_{1t}, \cdots, u_{nt})'$ exhibits both individual effects and spatially autocorrelated structure defined in (2.2) and (2.3) and 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 $\epsilon_0 > 0$; (iii) $\{x_{it}, t = \cdots, -1, 0, 1, \cdots\}$ and $\{z_i\}$ are strictly exogenous and independent across i; (iv) $|\rho| < 1$ in (2.1); and (v) The true parameter ψ_0 lies in the interior of a convex compact set $\mathbf{\Psi}$.*

Assumption G1(i) corresponds to traditional panel data models with large n and small T. One can consider extending the QMLE procedure to panels with large n and large T; see, for example, Phillips and Sul (2003). Assumption G1(ii) is standard in the literature. Assumption G1(iii) is not as strong as it appears in the spatial econometrics literature, since in most spatial analysis regressors are treated as nonstochastic fixed constants (e.g., Anselin, 1988; Kelejian and Prucha, 1998, 1999, 2010; Lee, 2002, 2004; Lin and Lee, 2010; Robinson, 2010; Su and Jin, 2010; Su, 2012). One can relax the strict exogeneity condition in Assumption G1(iii) like Hsiao et al. (2002) but this will complicate our analysis in case of spatially correlated errors. Assumption G1(iv) can be relaxed for the case of random effects with exogenous initial observations without any change of the derivation. It can also be relaxed for the fixed effects model with some modification of the derivation as in Hsiao et al. (2002). Assumption G1(v) is commonly assumed in the literature but deserves some further discussion.

For QML estimation, it is required that λ lie within a certain space so as to guarantee the positiveness of the determinant of $I_n - \lambda W$ and hence the existence of $(I_n - \lambda W)^{-1}$. If the eigenvalues of the spatial weight matrix W are real, then such a space would be $(1/w_{\min}, 1/w_{\max})$ where $w_{\min}$ and $w_{\max}$ are, respectively, the smallest and the largest eigenvalues of W; if, further, W is row normalized, then $w_{\max} = 1$ and $1/w_{\min} < -1$, and the parameter space of λ becomes $(1/w_{\min}, 1)$ (see Anselin, 1988). In general, the eigenvalues of W may not be all real and in this case Kelejian and Prucha (2010) suggest the parameter space be $(-1/\tau_n, 1/\tau_n)$ where τ_n is the spectral radius of W, giving a parameter space dependent upon the number of spatial units. This parameter space can be converted to $(-1, 1)$ if one works with $\tau_n^{-1}W$. In this case Assumption G1(v) requires that λ lies in a compact subset of $(-1, 1)$.

For the spatial weight matrix, we make the following assumptions.

Assumption G2: *(i) The elements w_{ij} of W are at most of order h_n^{-1}, denoted by $O(h_n^{-1})$, uniformly in all i and j. As a normalization, $w_{ii} = 0$ for all i; (ii) The ratio $h_n/n \to 0$ as n goes to infinity; (iii) The matrix B_0 is nonsingular; (iv) The sequences of matrices $\{W\}$ and $\{B_0^{-1}\}$ are uniformly bounded in both row and column sums; (v) $\{B^{-1}\}$ are uniformly bounded in either row or column sums, uniformly in λ in a compact parameter space $\mathbf{\Lambda}$, and $\underline{c}_\lambda \le \inf_{\lambda\in\mathbf{\Lambda}}\lambda_{\max}(B) \le \sup_{\lambda\in\mathbf{\Lambda}} \lambda_{\max}(B) \le \bar{c}_\lambda < \infty$.*

Assumptions G2(i)-(iv) parallel Assumptions 2-4 of Lee (2004). Like Lee (2004), Assumptions G2(i)-(iv) provide the essential features of the weight matrix for the model. Assumption G2(ii) is always satisfied if $\{h_n\}$ is a bounded sequence. We allow $\{h_n\}$ to be divergent but at a rate smaller than n as in Lee (2004). Assumption G2(iii) guarantees that the disturbance term is well defined. Kelejian and Prucha (1998, 1999, 2001) and Lee (2004) also assume Assumption G2(iv) which limits the spatial correlation to some degree but facilitates the study of the asymptotic properties of the spatial parameter estimators. By Horn and Johnson (1985, p. 301), $\text{limsup}_n\|\lambda_0 W\| < 1$ is sufficient to guarantee that B_0^{-1} is uniformly bounded in both row and column sums. By Lee (2002, Lemma A.3), Assumption G2(iv) implies $\{B^{-1}\}$ are uniformly bounded in both row and column sums uniformly in a neighborhood of λ_0. Assumption G2(v) is stronger than Assumption G2(iv) and is required in establishing the consistency results.

4.2 Random effects model

We now present detailed asymptotic results for the SDPD model with random effects. Beside the generic assumptions given earlier, some additional assumptions specific for this model are necessary.

Assumption R: *(i) μ_i's are iid with $E(\mu_i) = 0$, $Var(\mu_i) = \sigma_\mu^2$, and $E|\mu_i|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$; (ii) μ_i and v_{jt} are mutually independent, and they are independent of x_{ks} and z_k for all i, j, k, t, s; (iii) All elements in (x_{it}, z_i) have $4 + \epsilon_0$ moments for some $\epsilon_0 > 0$.*

Assumption R(i) and the first part of Assumption R(ii) are standard in the random effects panel data literature. The second part of Assumption R(ii) is for convenience. Alternatively we can treat the regressors as nonstochastic matrix.

Case I: y_0 is exogenous. To derive the consistency of the QML estimators, we need to ensure that $\delta = (\lambda, \phi_\mu)'$ is identifiable. Then, the identifiability of other parameters follows. Following White (1994) and Lee (2004), define $\mathcal{L}_c^{r*}(\delta) = \max_{\theta,\sigma_v^2} E[\mathcal{L}^r(\theta, \sigma_v^2, \delta)]$, where we suppress the dependence of $\mathcal{L}_c^{r*}(\delta)$ on n. The optimal solution to $\max_{\theta,\sigma_v^2} E[\mathcal{L}^r(\theta, \sigma_v^2, \delta)]$ is given by

$$\tilde{\theta}(\delta) = [E(\tilde{X}'\Omega^{-1}\tilde{X})]^{-1}E(\tilde{X}'\Omega^{-1}Y) \text{ and} \tag{4.1}$$

$$\tilde{\sigma}_v^2(\delta) = \frac{1}{nT}E[u(\tilde{\theta}(\delta))'\Omega^{-1}u(\tilde{\theta}(\delta))]. \tag{4.2}$$

Consequently, we have

$$\mathcal{L}_c^{r*}(\delta) = -\frac{nT}{2}\left[\log(2\pi)+1\right] - \frac{nT}{2}\log[\tilde{\sigma}_v^2(\delta)] - \frac{1}{2}\log|\Omega|. \tag{4.3}$$

We impose the following identification condition.

Assumption R: *(iv)* $\lim_{n\to\infty}\frac{1}{2nT}\left\{\log|\sigma_{v0}^2\Omega_0| - \log|\tilde{\sigma}_v^2(\delta)\,\Omega(\delta)|\right\} \neq 0$ for any $\delta \neq \delta_0$, *and* $\frac{1}{nT}\tilde{X}'\tilde{X}$ *is positive definite almost surely for sufficiently large* n.

The first part of Assumption R(*iv*) parallels Assumption 9 in Lee (2004). It is a global identification condition related to the uniqueness of the variance-covariance matrix of u. With this and the uniform convergence of $\frac{1}{nT}[\mathcal{L}_c^r(\delta) - \mathcal{L}_c^{r*}(\delta)]$ to zero on $\mathbf{\Delta}$ proved in the Appendix C, the consistency of $\hat{\delta}$ follows. The consistency of $\hat{\theta}$ and $\hat{\sigma}_v^2$ follows from that of $\hat{\delta}$ and the second part of Assumption R(iv).

Theorem 4.1 *Under Assumptions G1, G2, and R(i)-(iv), if the initial observations y_{i0} are exogenously given, then* $\hat{\psi} \xrightarrow{p} \psi_0$.

To derive the asymptotic distribution of $\hat{\psi}$, we need to make a Taylor expansion of $\frac{\partial}{\partial\psi}\mathcal{L}^r(\hat{\psi}) = 0$ at ψ_0, and then to check that the score function and Hessian matrix have proper asymptotic behavior. We report both the score and Hessian here to provide insights for the asymptotic theory and to facilitate the practical applications. First, the score function $S(\psi) = \frac{\partial}{\partial\psi}\mathcal{L}^r(\psi)$ has the elements

$$\begin{aligned}
\frac{\partial\mathcal{L}^r(\psi)}{\partial\theta} &= \frac{1}{\sigma_v^2}\tilde{X}'\Omega^{-1}u(\theta),\\
\frac{\partial\mathcal{L}^r(\psi)}{\partial\sigma_v^2} &= \frac{1}{2\sigma_v^4}u(\theta)'\Omega^{-1}u(\theta) - \frac{nT}{2\sigma_v^2},\\
\frac{\partial\mathcal{L}^r(\psi)}{\partial\omega} &= \frac{1}{2\sigma_v^2}u(\theta)'P_\omega u(\theta) - \frac{1}{2}\mathrm{tr}\left(P_\omega\Omega\right),\ \ \omega = \lambda, \phi_\mu,
\end{aligned}$$

where $P_\omega = \Omega^{-1}\Omega_\omega\Omega^{-1}$ and $\Omega_\omega = \frac{\partial}{\partial\omega}\Omega(\delta)$ for $\omega = \lambda, \phi_\mu$. One can easily verify that $\Omega_\lambda = I_T \otimes A$ and $\Omega_{\phi_\mu} = J_T \otimes I_n$ where $A = \frac{\partial}{\partial\lambda}(B'B)^{-1} = (B'B)^{-1}(W'B + B'W)(B'B)^{-1}$. At $\psi = \psi_0$, the last three components of the score function are linear and quadratic functions of $u \equiv u(\theta_0)$ and one can readily verify that their expectations are zero. The first component also has a zero expectation by Lemma B.6. Note that the elements in u are not independent and that $\tilde{X}$ contains the lagged dependent variable, thus the standard results, such as the central limit theorem (CLT) for linear and quadratic forms in Kelejian and Prucha (2001) cannot be directly applied. For the last three components, we need to plug $u = (\iota_T \otimes I_n)\mu + (I_T \otimes B_0^{-1})v$ into $\frac{\partial}{\partial\psi}\mathcal{L}^r(\psi_0)$ and apply the CLT to linear and quadratic functions of μ and v separately. For the first component, a special care has to be given to Y_{-1} (see Lemmas B.6 and B.8 for details).

The Hessian matrix $H_{r,n}(\psi) \equiv \frac{\partial^2}{\partial\psi\partial\psi'}\mathcal{L}^r(\psi)$ has the elements

$$\begin{aligned}
&\frac{\partial^2\mathcal{L}^r(\psi)}{\partial\theta\partial\theta'} = -\frac{1}{\sigma_v^2}\tilde{X}'\Omega^{-1}\tilde{X}, && \frac{\partial^2\mathcal{L}^r(\psi)}{\partial\theta\partial\sigma_v^2} = -\frac{1}{\sigma_v^4}\tilde{X}'\Omega^{-1}u(\theta),\\
&\frac{\partial^2\mathcal{L}^r(\psi)}{\partial\theta\partial\omega} = -\frac{1}{\sigma_v^2}\tilde{X}'P_\omega u(\theta),\ \ \omega = \lambda, \phi_\mu, && \frac{\partial^2\mathcal{L}^r(\psi)}{\partial\sigma_v^2\partial\sigma_v^2} = -\frac{1}{\sigma_v^6}u(\theta)'\Omega^{-1}u(\theta) + \frac{nT}{2\sigma_v^4},\\
&\frac{\partial^2\mathcal{L}^r(\psi)}{\partial\sigma_v^2\partial\omega} = -\frac{1}{2\sigma_v^4}u(\theta)'P_\omega u(\theta),\ \ \omega = \lambda, \phi_\mu, && \frac{\partial^2\mathcal{L}^r(\psi)}{\partial\omega\partial\varpi} = q_{\omega\varpi}[u(\theta)],\ \ \text{for } \omega, \varpi = \lambda, \phi_\mu,
\end{aligned}$$

where $q_{\omega\varpi}(u) \equiv \frac{1}{2}\mathrm{tr}(P_\varpi\Omega_\omega - \Omega^{-1}\Omega_{\omega\varpi}) - \frac{1}{2\sigma_v^2}u'(2P_\varpi\Omega_\omega - \Omega^{-1}\Omega_{\omega\varpi})\Omega^{-1}u$ for $\omega, \varpi = \lambda, \phi_\mu$; and $\Omega_{\omega\varpi} = \frac{\partial^2}{\partial\omega\partial\varpi}\Omega(\delta)$ for $\omega, \varpi = \lambda, \phi_\mu$. It is easy to see that $\Omega_{\lambda\lambda} = I_T \otimes \dot{A}$ where $\dot{A} = \frac{\partial}{\partial\lambda}A = 2(B'B)^{-1}[(W'B + B'W)A - W'W]$, and all other $\Omega_{\omega\varpi}$ matrices are $0_{nT\times nT}$.

Again, we see that most of the Hessian elements are quadratic forms of $u(\theta)$ whose asymptotic behavior is easy to study. Special care needs to be given to the elements involving $\tilde{X}$ (see Lemma B.7 for details). Let $\Gamma_{r,n}(\psi) = E[\frac{\partial}{\partial\psi}\mathcal{L}^r(\psi)\frac{\partial}{\partial\psi'}\mathcal{L}^r(\psi)]$ be the variance-covariance matrix of the score vector.[7] See Appendix A for the expression of $\Gamma_{r,n}(\psi)$. We have the following theorem.

Theorem 4.2 *Under Assumptions G1, G2, and R(i)-(iv), if the initial observations y_{i0} are exogenously given, then $\sqrt{nT}(\hat{\psi} - \psi_0) \xrightarrow{d} N(0, H_r^{-1}\Gamma_r H_r^{-1})$, where $H_r = \lim_{n\to\infty} \frac{1}{nT}E\left[H_{r,n}(\psi_0)\right]$ and $\Gamma_r = \lim_{n\to\infty} \frac{1}{nT}\Gamma_{r,n}(\psi_0)$, both assumed to exist, and $(-H_r)$ is assumed to be positive definite.*

As in Lee (2004), the asymptotic results in Theorem 4.2 is valid regardless of whether the sequence $\{h_n\}$ is bounded or divergent. The matrices Γ_r and H_r can be simplified if $h_n \to \infty$ as $n \to \infty$. When both μ_i and v_{it} are normally distributed, the asymptotic variance-covariance matrix reduces to $-H_r^{-1}$.

Case II: y_0 is endogenous. In this case, define $\mathcal{L}_c^{rr*}(\delta) = \max_{\theta,\sigma_v^2} E[\mathcal{L}^{rr}(\theta, \sigma_v^2, \delta)]$, where we suppress the dependence of $\mathcal{L}_c^{rr*}(\delta)$ on n. The optimal solution to $\max_{\theta,\sigma_v^2} E[\mathcal{L}^{rr}(\theta, \sigma_v^2, \delta)]$ is now given by

$$\tilde{\theta}(\delta) = [E(X^{*\prime}\Omega^{*-1}(\delta)X^*)]^{-1}E\left[X^{*\prime}\Omega^{*-1}(\delta)Y^*(\rho)\right] \text{ and} \tag{4.4}$$

$$\tilde{\sigma}_v^2(\delta) = \frac{1}{n(T+1)}E[u^*(\tilde{\theta}(\delta), \rho)'\Omega^{*-1}(\delta)u^*(\tilde{\theta}(\delta), \rho)]. \tag{4.5}$$

Consequently, we have

$$\mathcal{L}_c^{rr*}(\delta) = -\frac{n(T+1)}{2}[\log(2\pi) + 1] - \frac{n(T+1)}{2}\log\tilde{\sigma}_v^2(\delta) - \frac{1}{2}\log|\Omega^*|. \tag{4.6}$$

We make the following identification assumption.

Assumption R: (iv^*) $\lim_{n\to\infty} \frac{1}{2n(T+1)}\left\{\log|\sigma_{v0}^2\Omega_0^*| - \log|\tilde{\sigma}_v^2(\delta)\Omega^*(\delta)|\right\} \neq 0$ for any $\delta \neq \delta_0$. *Both $\frac{1}{n}\tilde{\mathbf{x}}'\tilde{\mathbf{x}}$ and $\frac{1}{nT}(X, Z)'(X, Z)$ are positive definite almost surely for sufficiently large n.*

The following theorem establishes the consistency of QMLE for the random effects model with endogenous initial observations. Similarly, the key result is to show that $\frac{1}{n(T+1)}[\mathcal{L}_c^{rr}(\delta) - \mathcal{L}_c^{rr*}(\delta)]$ coverges to zero uniformly in $\delta \in \mathbf{\Delta}$, which is given in Appendix C.

Theorem 4.3 *Under Assumptions G1, G2, R0, R(i)-(iii) and R(iv*), if the initial observations y_{i0} are endogenously given, then $\hat{\psi} \xrightarrow{p} \psi_0$.*

Again, to derive the asymptotic distribution of $\hat{\psi}$, one starts with a Taylor expansion of the score function, $S^{rr}(\psi) = \frac{\partial}{\partial\psi}\mathcal{L}^{rr}(\psi)$, of which the elements are given below:

$$\begin{aligned}
\frac{\partial\mathcal{L}^{rr}(\psi)}{\partial\theta} &= \frac{1}{\sigma_v^2}X^{*\prime}\Omega^{*-1}u^*(\theta,\rho), \\
\frac{\partial\mathcal{L}^{rr}(\psi)}{\partial\sigma_v^2} &= \frac{1}{2\sigma_v^4}u^*(\theta,\rho)'\Omega^{*-1}u^*(\theta,\rho) - \frac{n(T+1)}{2\sigma_v^2}, \\
\frac{\partial\mathcal{L}^{rr}(\psi)}{\partial\rho} &= -\frac{1}{\sigma_v^2}u_\rho^*(\theta,\rho)'\Omega^{*-1}u^*(\theta,\rho) + \frac{1}{2\sigma_v^2}u^*(\theta,\rho)'P_\rho^*u^*(\theta,\rho) - \frac{1}{2}\text{tr}(P_\rho^*\Omega^*), \\
\frac{\partial\mathcal{L}^{rr}(\psi)}{\partial\omega} &= \frac{1}{2\sigma_v^2}u^*(\theta,\rho)'P_\omega^*u^*(\theta,\rho) - \frac{1}{2}\text{tr}(P_\omega^*\Omega^*) \text{ for } \omega = \lambda,\ \phi_\mu,\ \text{and } \phi_\zeta,
\end{aligned}$$

[7] It is well known that for normally distributed individual-specific effects μ_i and error terms v_{it}, $\Gamma_{r,n}(\psi_0) = -E[H_{r,n}(\psi_0)]$ under some mild conditions. We do not impose normality restriction in this paper.

where $u^*_\rho(\theta,\rho) = \frac{\partial}{\partial\rho}u^*(\theta,\rho)$, $P^*_\omega = \Omega^{*-1}\Omega^*_\omega\Omega^{*-1}$, $\Omega^*_\omega = \frac{\partial}{\partial\omega}\Omega^*(\delta)$ for $\omega = \rho, \lambda, \phi_\mu$, and ϕ_ζ have the expressions

$$
\begin{aligned}
u^*_\rho(\theta,\rho) &= -\begin{pmatrix} \dot{a}_m Z\gamma \\ Y_{-1} \end{pmatrix},\ \Omega^*_\rho = \begin{pmatrix} 2\phi_\mu a_m \dot{a}_m I_n + \dot{b}_m (B'B)^{-1} & \phi_\mu \dot{a}_m(\iota'\otimes I_n) \\ \phi_\mu \dot{a}_m(\iota\otimes I_n) & 0_{nT\times nT} \end{pmatrix}, \\
\Omega^*_\lambda &= \begin{pmatrix} b_m & 0'_T \\ 0_T & I_T \end{pmatrix}\otimes A,\ \Omega^*_{\phi_\mu} = \begin{pmatrix} a_m^2 & a_m\iota'_T \\ a_m\iota_T & J_T \end{pmatrix}\otimes I_n,\ \text{and}\ \Omega^*_{\phi_\zeta} = \begin{pmatrix} 1 & 0'_T \\ 0_T & 0_{T\times T} \end{pmatrix}\otimes I_n,
\end{aligned}
$$

where $\dot{a}_m = \frac{d}{d\rho}a_m(\rho)$ and $\dot{b}_m = \frac{d}{d\rho}b_m(\rho)$, and their expressions can easily be obtained. One can readily verify that $E[\frac{\partial}{\partial\psi}\mathcal{L}^{rr}(\psi_0)] = 0$. The asymptotic normality of the score is given in Lemma B.13. The asymptotic normality of the QMLE thus follows if the Hessian matrix, $H_{rr,n}(\psi) \equiv \frac{\partial^2}{\partial\psi\partial\psi'}\mathcal{L}^{rr}(\psi)$, given below possesses the desired stochastic convergence property.

$$
\begin{aligned}
\frac{\partial^2\mathcal{L}^{rr}(\psi)}{\partial\theta\partial\theta'} &= -\frac{1}{\sigma_v^2}X^{*\prime}\Omega^{*-1}X^*, \\
\frac{\partial^2\mathcal{L}^{rr}(\psi)}{\partial\theta\partial\sigma_v^2} &= -\frac{1}{\sigma_v^4}X^{*\prime}\Omega^{*-1}u^*(\theta,\rho), \\
\frac{\partial^2\mathcal{L}^{rr}(\psi)}{\partial\theta\partial\rho} &= \frac{1}{\sigma_v^2}X^{*\prime}_\rho\Omega^{*-1}u^*(\theta,\rho) + \frac{1}{\sigma_v^2}X^{*\prime}\Omega^{*-1}u^*_\rho(\theta,\rho) - \frac{1}{\sigma_v^2}X^{*\prime}P^*_\rho u^*(\theta,\rho), \\
\frac{\partial^2\mathcal{L}^{rr}(\psi)}{\partial\theta\partial\omega} &= -\frac{1}{\sigma_v^2}X^{*\prime}P^*_\omega u^*(\theta,\rho), \quad \text{for } \omega = \lambda, \phi_\mu, \text{ and } \phi_\zeta, \\
\frac{\partial^2\mathcal{L}^{rr}(\psi)}{\partial\sigma_v^2\partial\sigma_v^2} &= -\frac{1}{\sigma_v^6}u^*(\theta,\rho)'\Omega^{*-1}u^*(\theta,\rho) + \frac{n(T+1)}{2\sigma_v^4}, \\
\frac{\partial^2\mathcal{L}^{rr}(\psi)}{\partial\sigma_v^2\partial\rho} &= \frac{1}{\sigma_v^4}u^*_\rho(\theta,\rho)'\Omega^{*-1}u^*(\theta,\rho) - \frac{1}{2\sigma_v^4}u^*(\theta)'P^*_\rho u^*(\theta,\rho), \\
\frac{\partial^2\mathcal{L}^{rr}(\psi)}{\partial\sigma_v^2\partial\omega} &= -\frac{1}{2\sigma_v^4}u^*(\theta,\rho)'P^*_\omega u^*(\theta,\rho), \quad \text{for } \omega = \lambda, \phi_\mu, \text{ and } \phi_\zeta, \\
\frac{\partial^2\mathcal{L}^{rr}(\psi)}{\partial\rho\partial\rho} &= -\frac{1}{\sigma_v^2}u^*_{\rho\rho}(\theta,\rho)'\Omega^{*-1}u^*(\theta,\rho) - \frac{1}{\sigma_v^2}u^*_\rho(\theta,\rho)'\Omega^{*-1}u^*_\rho(\theta,\rho) + \frac{2}{\sigma_v^2}u^*_\rho(\theta,\rho)'P^*_\rho u^*(\theta,\rho) + q^*_{\rho\rho}[u^*(\theta,\rho)], \\
\frac{\partial^2\mathcal{L}^{rr}(\psi)}{\partial\rho\partial\omega} &= \frac{1}{\sigma_v^2}u^*_\rho(\theta,\rho)'P^*_\omega u^*(\theta,\rho) + q^*_{\rho\omega}[u^*(\theta,\rho)], \quad \text{for } \omega = \lambda, \phi_\mu, \text{ and } \phi_\zeta, \\
\frac{\partial^2\mathcal{L}^{rr}(\psi)}{\partial\omega\partial\varpi} &= q^*_{\omega\varpi}[u^*(\theta,\rho)], \quad \text{for } \omega, \varpi = \lambda, \phi_\mu, \text{ and } \phi_\zeta.
\end{aligned}
$$

where $q^*_{\omega\varpi}(u^*) \equiv \frac{1}{2}\text{tr}(P^*_\varpi\Omega^*_\omega - \Omega^{*-1}\Omega^*_{\omega\varpi}) - \frac{1}{2\sigma_v^2}u^{*\prime}(2P^*_\varpi\Omega^*_\omega - \Omega^{*-1}\Omega^*_{\omega\varpi})\Omega^{*-1}u^*$ for $\omega, \varpi = \rho, \lambda, \phi_\mu$, and ϕ_ζ, $X^*_\rho = \frac{\partial}{\partial\rho}X^*$, $u^*_{\rho\rho}(\theta,\rho) = \frac{\partial^2}{\partial\rho^2}u^*(\theta,\rho)$, and $\Omega^*_{\rho\omega} = \frac{\partial^2}{\partial\rho\partial\omega}\Omega^*$ for $\omega = \rho, \lambda, \phi_\mu$, and ϕ_ζ. The second-order partial derivatives of Ω^* are

$$
\begin{aligned}
&\Omega^*_{\rho\rho} = \begin{pmatrix} 2\phi_\mu(\dot{a}_m^2 + \ddot{a}_m)I_n + \ddot{b}_m(B'B)^{-1}, & \phi_\mu\ddot{a}_m(\iota'\otimes I_n) \\ \phi_\mu\ddot{a}_m(\iota\otimes I_n) & 0_{nT\times nT} \end{pmatrix},\quad \Omega^*_{\rho\lambda} = \begin{pmatrix} \dot{b}_m A, & 0_{n\times nT} \\ 0_{nT\times n} & 0_{nT\times nT} \end{pmatrix}, \\
&\Omega^*_{\rho\phi_\mu} = \begin{pmatrix} 2a_m\dot{a}_m I_n, & \dot{a}_m(\iota'\otimes I_n) \\ \dot{a}_m(\iota\otimes I_n) & 0_{nT\times nT} \end{pmatrix},\quad \Omega^*_{\lambda\lambda} = \begin{pmatrix} b_m & 0 \\ 0 & I_T \end{pmatrix}\otimes\dot{A},
\end{aligned}
$$

and all other $\Omega^*_{\omega\varpi}$ matrices are $0_{n(T+1)\times n(T+1)}$, where $\ddot{a}_m = \frac{\partial}{\partial\rho}\dot{a}_m$ and $\ddot{b} = \frac{\partial}{\partial\rho}\dot{b}_m$ and their exact expressions can be easily derived. Finally, X^*_ρ has a sole non-zero element $\dot{a}_m z$, and $u^*_{\rho\rho}(\theta,\rho) = (-\ddot{a}_m\gamma'z',\ 0_{1\times nT})'$.

Let $\Gamma_{rr,n}(\psi) = E[\frac{\partial}{\partial\psi}\mathcal{L}^{rr}(\psi)\frac{\partial}{\partial\psi'}\mathcal{L}^{rr}(\psi)]$ be the variance-covariance matrix of the score vector with its detail given in Appendix A. We now state the asymptotic normality result.

Theorem 4.4 *Under Assumptions G1, G2, R0, R(i)-(iii) and R(iv*), if the initial observations are endogenously given, then* $\sqrt{nT}(\hat{\psi}-\psi_0) \xrightarrow{d} N(0, H_{rr}^{-1}\Gamma_{rr}H_{rr}^{-1})$*, where* $H_{rr} = \lim_{n\to\infty} \frac{1}{n(T+1)} E\left[H_{rr,n}(\psi_0)\right]$ *and* $\Gamma_{rr} = \lim_{n\to\infty} \frac{1}{n(T+1)} \Gamma_{rr,n}(\psi_0)$*, both assumed to exist, and* $(-H_{rr})$ *is assumed to be positive definite.*

4.3 Fixed effects model

For the fixed effects model, we need to supplement the generic assumptions, Assumptions G1 and G2, made above with the following assumption on the regressors.

Assumption F: *(i) The processes* $\{x_{it}, t = \cdots, -1, 0, 1, \cdots\}$ *are trend-stationary or first-differencing stationary for all* $i = 1, \cdots, n$*; (ii) All elements in* $(\Delta v_{it}, \Delta x_{it})$ *have* $4 + \epsilon_0$ *moments for some* $\epsilon_0 > 0$*; (iii)* $\frac{1}{nT}\Delta X^{\dagger\prime}\Delta X^{\dagger}$ *is positive definite almost surely for sufficiently large* n*.*

Define $\mathcal{L}_c^{f*}(\delta) = \max_{\theta,\sigma_v^2} E[\mathcal{L}^f(\theta, \sigma_v^2, \delta)]$, where we suppress the dependence of $\mathcal{L}_c^{f*}(\delta)$ on n. Let $\Delta Y = (0_{1\times n}, \Delta y_1', ..., \Delta y_{T-1}')'$. The optimal solution to $\max_{\theta,\sigma_v^2} E[\mathcal{L}^f(\theta, \sigma_v^2, \delta)]$ is now given by

$$\tilde{\theta}(\delta) = \left\{E\left[\left(\Delta X^{\dagger}\right)' \Omega^{\dagger-1}\Delta X^{\dagger}\right]\right\}^{-1} E\left[\left(\Delta X^{\dagger}\right)' \Omega^{\dagger-1}\Delta Y^{\dagger}(\rho)\right] \text{ and} \tag{4.7}$$

$$\tilde{\sigma}_v^2(\delta) = \frac{1}{nT} E[\Delta u(\tilde{\theta}(\delta), \rho)'\Omega^{\dagger-1}\Delta u(\tilde{\theta}(\delta), \rho)]. \tag{4.8}$$

Consequently, we have

$$\mathcal{L}_c^{f*}(\delta) = -\frac{nT}{2}\left[\log(2\pi) + 1\right] - \frac{nT}{2}\log[\tilde{\sigma}_v^2(\delta)] - \frac{1}{2}\log|\Omega^{\dagger}|. \tag{4.9}$$

The following identification condition is needed for our consistency result.

Assumption F: *(iv)* $\lim_{n\to\infty} \frac{1}{2nT}\left\{\log|\sigma_{v0}^2\Omega_0^{\dagger}| - \log|\tilde{\sigma}_v^2(\delta)\,\Omega^{\dagger}(\delta)\,|\right\} \neq 0$ for any $\delta \neq \delta_0$.

With this identification condition, the consistency of $\hat{\delta}$ follows if $\frac{1}{nT}[\mathcal{L}_c^f(\delta) - \mathcal{L}_c^{f*}(\delta)]$ converges to zero uniformly on $\boldsymbol{\Delta}$. The consistency of $\hat{\theta}$ and $\hat{\sigma}_v^2$ then follows from the consistency of $\hat{\delta}$ and the identification condition given in Assumption F(iii). We have the following theorem.

Theorem 4.5 *Under Assumptions G1, G2, F0, and F, we have* $\hat{\psi} \xrightarrow{p} \psi_0$*.*

To derive the asymptotic distribution of $\hat{\psi}$, one needs the score function $S^f(\psi) = \frac{\partial}{\partial\psi}\mathcal{L}^f(\psi)$:

$$\begin{aligned}
\frac{\partial\mathcal{L}^f(\psi)}{\partial\theta} &= \frac{1}{\sigma_v^2}\Delta X^{\dagger\prime}\Omega^{\dagger-1}\Delta u(\theta,\rho),\\
\frac{\partial\mathcal{L}^f(\psi)}{\partial\sigma_v^2} &= \frac{1}{2\sigma_v^4}\Delta u(\theta,\rho)'\Omega^{\dagger-1}\Delta u(\theta,\rho) - \frac{nT}{2\sigma_v^2},\\
\frac{\partial\mathcal{L}^f(\psi)}{\partial\rho} &= -\frac{1}{\sigma_v^2}\Delta u_\rho(\theta,\rho)'\Omega^{\dagger-1}\Delta u(\theta,\rho) + \frac{1}{2\sigma_v^2}\Delta u(\theta,\rho)'P_\rho^{\dagger}\Delta u(\theta,\rho) - \frac{1}{2}\text{tr}(\Omega^{\dagger-1}\Omega_\rho^{\dagger}),\\
\frac{\partial\mathcal{L}^f(\psi)}{\partial\omega} &= \frac{1}{2\sigma_v^2}\Delta u(\theta,\rho)'P_\omega^{\dagger}\Delta u(\theta,\rho) - \frac{1}{2}\text{tr}(\Omega^{\dagger-1}\Omega_\omega^{\dagger}) \text{ for } \omega = \lambda,\ \phi_\zeta,
\end{aligned}$$

where $\Delta u_\rho(\theta,\rho) = \frac{\partial}{\partial\rho}\Delta u(\theta,\rho) = -(0_{n\times1}', \Delta y_1', \cdots, \Delta y_{T-1}')'$, and $\Omega_\omega^{\dagger} = \frac{\partial}{\partial\omega}\Omega^{\dagger}(\delta)$ and $P_\omega^{\dagger} = \Omega^{\dagger-1}\Omega_\omega^{\dagger}\Omega^{\dagger-1}$ for $\omega = \rho, \lambda$, and ϕ_ζ. From (3.24), it is easy to see that $\Omega_\rho^{\dagger} = h_{\dot{c}_m} \otimes (B'B)^{-1}$, $\Omega_\lambda^{\dagger} = h_{c_m} \otimes A$, and $\Omega_{\phi_\zeta}^{\dagger} = \ell_1 \otimes I_n$, where $\dot{c}_m = \frac{\partial}{\partial\rho}c_m(\rho)$. Again, one can readily verify that $E[\frac{\partial}{\partial\psi}\mathcal{L}^f(\psi_0)] = 0$. The asymptotic normality of the score is given in Lemma B.15. The asymptotic normality of $\hat{\psi}$ thus follows if the Hessian matrix, $H_{f,n}(\psi) \equiv \frac{\partial^2}{\partial\psi\partial\psi'}\mathcal{L}^f(\psi)$, given below possesses the desired stochastic convergence property.

$$
\begin{aligned}
\frac{\partial^2\mathcal{L}^f(\psi)}{\partial\theta\partial\theta'} &= -\frac{1}{\sigma_v^2}\Delta X^{\dagger\prime}\Omega^{\dagger-1}\Delta X^{\dagger}, \\
\frac{\partial^2\mathcal{L}^f(\psi)}{\partial\theta\partial\sigma_v^2} &= -\frac{1}{\sigma_v^4}\Delta X^{\dagger\prime}\Omega^{\dagger-1}\Delta u(\theta,\rho), \\
\frac{\partial^2\mathcal{L}^f(\psi)}{\partial\theta\partial\rho} &= \frac{1}{\sigma_v^2}\Delta X^{\dagger\prime}\Omega^{\dagger-1}\Delta u_\rho(\theta,\rho) - \frac{1}{\sigma_v^2}\Delta X^{\dagger\prime}P_\rho^{\dagger}\Delta u(\theta,\rho), \\
\frac{\partial^2\mathcal{L}^f(\psi)}{\partial\theta\partial\omega} &= -\frac{1}{\sigma_v^2}\Delta X^{\dagger\prime}P_\omega^{\dagger}\Delta u(\theta,\rho), \text{ for } \omega=\lambda,\ \phi_\zeta, \\
\frac{\partial^2\mathcal{L}^f(\psi)}{\partial\sigma_v^2\partial\sigma_v^2} &= -\frac{1}{\sigma_v^6}\Delta u(\theta,\rho)'\Omega^{\dagger-1}\Delta u(\theta,\rho) + \frac{nT}{2\sigma_v^4}, \\
\frac{\partial^2\mathcal{L}^f(\psi)}{\partial\sigma_v^2\partial\rho} &= \frac{1}{\sigma_v^4}\Delta u_\rho(\theta,\rho)'\Omega^{\dagger-1}\Delta u(\theta,\rho) - \frac{1}{2\sigma_v^4}\Delta u(\theta,\rho)'P_\rho^{\dagger}\Delta u(\theta,\rho), \\
\frac{\partial^2\mathcal{L}^f(\psi)}{\partial\sigma_v^2\partial\omega} &= -\frac{1}{2\sigma_v^4}\Delta u(\theta,\rho)'P_\omega^{\dagger}\Delta u(\theta,\rho), \text{ for } \omega=\lambda,\ \phi_\zeta, \\
\frac{\partial^2\mathcal{L}^f(\psi)}{\partial\rho\partial\rho} &= -\frac{1}{\sigma_v^2}\Delta u_\rho(\theta,\rho)'\Omega^{\dagger-1}\Delta u_\rho(\theta,\rho) + \frac{2}{\sigma_v^2}\Delta u_\rho(\theta,\rho)'P_\rho^{\dagger}\Delta u(\theta,\rho) + q_{\rho\rho}^{\dagger}[\Delta u(\theta,\rho)], \\
\frac{\partial^2\mathcal{L}^f(\psi)}{\partial\rho\partial\omega} &= \frac{1}{\sigma_v^2}\Delta u_\rho(\theta,\rho)'P_\omega^{\dagger}\Delta u(\theta,\rho) + q_{\rho\omega}^{\dagger}[\Delta u(\theta,\rho)], \text{ for } \omega=\lambda,\ \phi_\zeta, \\
\frac{\partial^2\mathcal{L}^f(\psi)}{\partial\omega\partial\varpi} &= q_{\omega\varpi}^{\dagger}[\Delta u(\theta,\rho)], \text{ for } \omega,\varpi=\lambda,\ \phi_\zeta,
\end{aligned}
$$

where $q_{\omega\varpi}^{\dagger}(\Delta u) \equiv \frac{1}{2}\text{tr}(P_\varpi^{\dagger}\Omega_\omega^{\dagger} - \Omega^{\dagger-1}\Omega_{\omega\varpi}^{\dagger}) - \frac{1}{2\sigma_v^2}\Delta u'(2P_\varpi^{\dagger}\Omega_\omega^{\dagger} - \Omega^{\dagger-1}\Omega_{\omega\varpi}^{\dagger})\Omega^{\dagger-1}\Delta u$ for $\omega,\ \varpi=\rho,\ \lambda$, and ϕ_ζ. The second derivatives $\Omega_{\omega\varpi}$ of Ω are: $\Omega_{\rho\rho} = h_{\ddot{c}_m}\otimes(B'B)^{-1}$ where $\ddot{c}_m = \frac{\partial}{\partial\rho}\dot{c}_m$, $\Omega_{\rho\lambda} = h_{\dot{c}_m}\otimes A$, $\Omega_{\lambda\lambda} = h_{c_m}\otimes\dot{A}$, and the remaining are all zero matrices.

Let $\Gamma_{f,n}(\psi) = E[\frac{\partial}{\partial\psi}\mathcal{L}^f(\psi)\frac{\partial}{\partial\psi'}\mathcal{L}^f(\psi)]$. (See Appendix A for some details.) We now state the asymptotic normality result.

Theorem 4.6 *Under Assumptions G1, G2, F0 and F, we have* $\sqrt{nT}(\hat{\psi}-\psi_0) \xrightarrow{d} N(0, H_f^{-1}\Gamma_f H_f^{-1})$, *where* $H_f = \lim_{n\to\infty}\frac{1}{nT}E[H_{f,n}(\psi_0)]$ *and* $\Gamma_r = \lim_{n\to\infty}\frac{1}{nT}\Gamma_{f,n}(\psi_0)$, *both assumed to exist, and* $(-H_f)$ *is assumed to be positive definite.*

5 Bootstrap Estimate of the Variance-Covariance Matrix

From Theorems 4.2, 4.4 and 4.6, we see that the asymptotic variance-covariance (VC) matrices of the QMLEs of the three models considered are, respectively, $H_r^{-1}\Gamma_r H_r^{-1}$, $H_{rr}^{-1}\Gamma_{rr}H_{rr}^{-1}$, and $H_f^{-1}\Gamma_f H_f^{-1}$. Practical applications of the asymptotic normality theory depend upon the availability of a consistent estimator of the asymptotic VC matrix. Obviously, the Hessian matrices evaluated at the QMLEs provide consistent estimators for H_r, H_{rr}, and H_f, i.e., $\hat{H}_r \equiv \frac{1}{nT}H_{r,n}(\hat{\psi})$, $\hat{H}_{rr} \equiv \frac{1}{n(T+1)}H_{rr,n}(\hat{\psi})$, and $\hat{H}_f \equiv \frac{1}{nT}H_{f,n}(\hat{\psi})$. The formal proofs of the consistency of these estimators can be found in the proofs of Theorems 4.2, 4.4, and 4.6, respectively. However, consistent estimators for Γ_r, Γ_{rr}, and Γ_f, the VC matrices of the scores (normalized), are not readily available due to the presence of the lagged dependent variable in the regressors. The basic problem is that the explicit expressions for $\Gamma_{r,n}(\psi_0), \Gamma_{rr,n}(\psi_0)$, and $\Gamma_{f,n}(\psi_0)$ are not readily available, and hence the usual plug-in method cannot be applied.[8] Thus, an alternative method is desired.

[8] This is not a problem for the exact likelihood inference (Elhorst, 2005, Yang et al. 2006) as in this case the VC matrix of the score function equals the negative expected Hessian. Hence, the asymptotic VC matrices of the MLEs in the three models considered reduce to $-H_r^{-1}, -H_{rr}^{-1}$ and $-H_f^{-1}$, respectively, of which sample analogues exist.

In this section, we introduce a residual-based bootstrap method for estimating the variance of the scores, with the bootstrap draws made on the joint empirical distribution function (EDF) of the n transformed vectors of residuals. While the general principle for our bootstrap method is the same for all the three models considered above, different structures of the residuals and the score functions render them a separate consideration.

5.1 Random effects model with exogenous initial values

Write the model as: $y_t = \rho y_{t-1} + x_t\beta + z\gamma + u_t,\ u_t = \mu + B^{-1}v_t,\ t = 1, 2, \cdots, T$, now viewed as a real-world data generating process (DGP). We have, $\text{Var}(u_t) = \sigma_v^2(\phi_\mu I_n + (B'B)^{-1}) \equiv \sigma_v^2\Sigma(\lambda, \phi_\mu)$. Define the *transformed residuals* (t-residuals):

$$r_t = \Sigma^{-\frac{1}{2}}(\lambda, \phi_\mu)u_t, \quad t = 1, \cdots, T,$$

where $\Sigma^{\frac{1}{2}}(\lambda, \phi_\mu)$ is a square-root matrix of $\Sigma(\lambda, \phi_\mu)$. Then, $\text{E}(r_t) = 0$ and $\text{Var}(r_t) = I_n$. Thus, the elements of r_t are uncorrelated, which are iid if μ and v_t are normal satisfying the conditions given in Assumptions G1 and R. As our asymptotics depend only on n, these uncorrelated residuals lay out the theoretical foundation for a residual-based bootstrap method. Let $\hat{r}_t$ be the QML estimate of r_t, and $\hat{\mathcal{F}}_{n,t}$ be the empirical distribution function (EDF) of the centered $\hat{r}_t$, for $t = 1, 2, \cdots, T$. Let $S(Y_{-1}, u, \psi_0)$ be the score function given below Theorem 4.1, written in terms of the lagged response Y_{-1}, the disturbance vector u and the true parameter vector ψ_0. The bootstrap procedure for estimating $\Gamma_{n,r}(\psi_0)$ is as follows.

1. Compute the QMLE $\hat{\psi}$ and obtain the QML residuals $\{\hat{r}_t, t = 1, 2, \cdots, T\}$. For each t, center $\hat{r}_t$ to obtain $\hat{\mathcal{F}}_{n,t}$.

2. Draw a random sample of size n from each $\hat{\mathcal{F}}_{n,t}$, $t = 1, 2, \cdots, T$, to give T samples of bootstrap residuals $\{\hat{r}_1^b, \cdots, \hat{r}_T^b\}$.

3. Conditional on y_0, x_t, z, and the QMLE $\hat{\psi}$, generate the bootstrap data according to

$$\begin{aligned} y_1^b &= \hat{\rho}y_0 + x_1\hat{\beta} + z\hat{\gamma} + \Sigma^{\frac{1}{2}}(\hat{\lambda}, \hat{\phi}_\mu)\hat{r}_1^b, \\ y_t^b &= \hat{\rho}y_{t-1}^b + x_t\hat{\beta} + z\hat{\gamma} + \Sigma^{\frac{1}{2}}(\hat{\lambda}, \hat{\phi}_\mu)\hat{r}_t^b, \quad t = 2, 3, \cdots, T. \end{aligned}$$

   The bootstrapped values of u and Y_{-1} are given by $u^b = vec[\Sigma^{\frac{1}{2}}(\hat{\lambda}, \hat{\phi}_\mu)(\hat{r}_1^b, \cdots, \hat{r}_T^b)]$ and$Y_{-1}^b = vec(y_0, y_1^b, \cdots, y_{t-1}^b)$, respectively.

4. Compute $S(Y_{-1}^b, u^b, \hat{\psi})$, where $S(Y_{-1}, u, \psi_0)$ is the score function.

5. Repeat steps 2-4 B times, and the bootstrap estimate of $\Gamma_{n,r}(\psi_0)$ is given by

$$\hat{\Gamma}_{n,r}^b = \frac{1}{B}\sum_{b=1}^{B}\Big(S(Y_{-1}^b, u^b, \hat{\psi})S(Y_{-1}^b, u^b, \hat{\psi})'\Big) - \frac{1}{B}\sum_{b=1}^{B}S(Y_{-1}^b, u^b, \hat{\psi}) \cdot \frac{1}{B}\sum_{b=1}^{B}S(Y_{-1}^b, u^b, \hat{\psi})'. \quad (5.1)$$

A justification for the validity of the above bootstrap procedure goes as follows. First, note that the score function can be written as $S(Y_{-1}, u, \psi)$, viewed as a function of random components and parameters. Note that $u_t = \mu + B^{-1}v_t, t = 1, \cdots, T$. If ψ_0 and the distributions of μ_i and v_{it} were all known, then to compute the value of $\Gamma_{n,r}(\psi_0)$, one can simply use the Monte Carlo method: (i) generate Monte Carlo samples μ^m and $v_t^m, t = 1, \cdots, T$, to give a Monte Carlo value u^m, (ii) compute the Monte Carlo value Y_{-1}^m based on u^m, $\mathbf{x}$ and z, through the real-world DGP, (iii) compute a Monte Carlo value $S^m(\psi_0) = S(Y_{-1}^m, u^m, \psi_0)$ for the score function, and (iv) repeat (i)-(iii) M times to give a Monte Carlo approximation to the value of $\Gamma_{n,r}(\psi_0)$ as

$$\Gamma_{n,r}^m(\psi_0) \approx \frac{1}{M}\sum_{m=1}^{B} S^m(\psi_0)S^m(\psi_0)' - \frac{1}{M}\sum_{m=1}^{M} S^m(\psi_0) \cdot \frac{1}{M}\sum_{m=1}^{M} S^m(\psi_0)', \tag{5.2}$$

which can be made to an arbitrary level of accuracy by choosing an arbitrarily large M. Note that $u_t = \sigma_{v0}\Sigma^{\frac{1}{2}}(\lambda_0, \phi_{\mu 0})r_t$. The step (i) above is equivalent to draw random sample r_t^m from the distribution $\mathcal{F}$ of r_{it}, the i element of r_t, and compute $u_t^m = \sigma_{v0}\Sigma^{\frac{1}{2}}(\lambda_0, \phi_{\mu 0})r_t^m$.

However, in the real world, ψ_0 is unknown. In this case, it is clear that a Monte Carlo estimate of $\Gamma_{n,r}(\psi_0)$ can be obtained by plugging $\hat{\psi}$ into (5.2),

$$\hat{\Gamma}_{n,r}^m = \left(\frac{1}{M}\sum_{m=1}^{B} S^m(\hat{\psi})S^m(\hat{\psi})' - \frac{1}{M}\sum_{m=1}^{M} S^m(\hat{\psi}) \cdot \frac{1}{M}\sum_{m=1}^{M} S^m(\hat{\psi})'\right). \tag{5.3}$$

In the real world, $\mathcal{F}$, or the distributions of μ_i and v_{it} are also unknown. However, we note that the only difference between $\hat{\Gamma}_{n,r}^b$ given in (5.1) and $\hat{\Gamma}_{n,r}^m$ given in (5.3) is that r_t^b for the former is from the EDF $\hat{\mathcal{F}}_{n,t}$, but r_t^m for the latter is drawn from the true distribution $\mathcal{F}$. The bootstrap DGP that mimics the real-world DGP must be $y_1^b = \hat{\rho}y_0 + x_1\hat{\beta} + z\hat{\gamma} + u_1^b$, and $y_t^b = \hat{\rho}y_{t-1} + x_t\hat{\beta} + z\hat{\gamma} + u_t^b, t = 2, \cdots, T$. Thus, if $\hat{\mathcal{F}}_{n,t}$ provides a consistent estimate for the true but unknown distribution $\mathcal{F}$, which is typically the case as $\hat{\psi}$ is consistent for ψ_0, then $\hat{\Gamma}_{n,r}^b$ and $\hat{\Gamma}_{n,r}^m$ are asymptotically equivalent. The extra variability caused by replacing $\mathcal{F}$ by $\hat{\mathcal{F}}_{n,t}$ is of the same order as that from replacing ψ_0 by $\hat{\psi}$. This justifies the validity of the proposed bootstrap procedure.

5.2 Random effects model with endogenous initial values

When the initial observations y_0 are endogenously given, the disturbance vector now becomes $(u_0, u_1, u_2, \cdots, u_T)$ such that $\mathrm{Var}(u_0) = \sigma_v^2\omega_{11}$ and $\mathrm{Var}(u_t) = \sigma_v^2\Sigma(\lambda, \phi_\mu), t = 1, 2, \cdots, T$, where ω_{11} is defined above (3.31) and $\Sigma(\lambda, \phi_\mu)$ is defined in Section 5.1. Define the *transformed residuals*: $r_0 = \omega_{11}^{-\frac{1}{2}}u_0$, and $r_t = \Sigma^{-\frac{1}{2}}(\lambda, \phi_\mu)u_t,\ t = 1, \cdots, T$, where $\omega_{11}^{\frac{1}{2}}$ is a square-root matrix of ω_{11}. Now, denote the QML estimates of the transformed residuals as $\{\hat{r}_0, \hat{r}_1, \cdots, \hat{r}_T\}$, and the EDF of the centered $\hat{r}_t$ by $\hat{\mathcal{F}}_{n,t}, t = 0, 1, \cdots, T$. Draw a random sample of size n each from $\mathcal{F}_{n,t}$, to give bootstrap residuals $\{\hat{r}_0, \hat{r}_1^b, \cdots, \hat{r}_T^b\}$. The bootstrap values for the response variables are thus generated according to

$$y_0^b = \tilde{\mathbf{x}}\hat{\pi} + \hat{\omega}^{\frac{1}{2}}\hat{r}_0^b, \ \text{ and } \ y_t^b = \hat{\rho}y_{t-1}^b + x_t\hat{\beta} + z\hat{\gamma} + \Sigma^{\frac{1}{2}}(\hat{\lambda}, \hat{\phi}_\mu)\hat{r}_t^b,\ t = 1, 2, \cdots, T.$$

The rest is analogous to those described in Section 5.1, including the justifications for the validity of this bootstrap procedure.

5.3 Fixed effects model with endogenous initial values

When the individual effects are treated as fixed, and the initial differences are modelled by (3.19), the disturbance vector becomes after first-differencing: $(\Delta\tilde{u}_1, \Delta u_2, \cdots, \Delta u_T)$, where $\Delta\tilde{u}_1$ is defined in (3.19) and $\Delta u_t = B^{-1}v_t$ as in (3.15) such that $\text{Var}(\Delta\tilde{u}_1) = \sigma_v^2(\phi_\zeta I_n + c_m(B'B)^{-1}) \equiv \sigma_v^2\omega$ and $\text{Var}(u_t) = 2\sigma_v^2(B'B)^{-1}$, $t = 2, \cdots, T$. Define the *transformed residuals*: $r_1 = \omega^{-\frac{1}{2}}\Delta\tilde{u}_1$ and $r_t = \frac{1}{\sqrt{2}}Bu_t$, $t = 2, \cdots, T$, where $\omega^{\frac{1}{2}}$ is square-root matrix of ω. Denote the QML estimates of the transformed residuals as $\{\hat{r}_1, \hat{r}_2, \cdots, \hat{r}_T\}$, and the EDF of the centered $\hat{r}_t$ by $\hat{\mathcal{F}}_{n,t}$, $t = 1, \cdots, T$. Draw a random sample of size n from $\mathcal{F}_{n,t}$, $t = 1, \cdots, T$, to give bootstrap residuals $\{\hat{r}_1, \hat{r}_2^b, \cdots, \hat{r}_T^b\}$. The bootstrap values for the response variables are thus generated according to

$$\Delta y_1^b = \Delta\tilde{\mathbf{x}}\hat{\pi} + \hat{\omega}^{\frac{1}{2}}\hat{r}_0^b, \text{ and } y_t^b = \hat{\rho}\Delta y_{t-1}^b + \Delta x_t\hat{\beta} + \sqrt{2}\hat{B}^{-1}\hat{r}_t^b,\ t = 2, 3, \cdots, T.$$

The rest is analogous to those described in Section 5.1, including the justifications for the validity of this bootstrap procedure.

6 Finite Sample Properties of the QMLEs

Monte Carlo experiments are carried out to investigate the performance of the QMLEs in finite samples and that of the bootstrapped estimates of the standard errors. In the former case, we investigate the consequences of treating the initial observations as endogenous when they are in fact exogenous, and vice versa. In the latter case we study the performance of standard error estimates based on only the Hessian, or only the bootstrapped variance of the score, or both, when errors are normal or nonnormal. We use the following data generating process (DGP):

$$\begin{aligned} y_t &= \rho y_{t-1} + \beta_0 \iota_n + x_t\beta_1 + z\gamma + u_t \\ u_t &= \mu + \varepsilon_t \\ \varepsilon_t &= \lambda W_n \varepsilon_t + v_t \end{aligned}$$

where y_t, y_{t-1}, x_t, and z are all $n \times 1$ vectors. The elements of x_t are generated in a similar fashion as in Hsiao et al. (2002),[9] and the elements of z are randomly generated from $Bernoulli(0.5)$. The spatial weight matrix is generated according to Rook or Queen contiguity, by randomly allocating the n spatial units on a lattice of $k \times m$ $(\geq n)$ squares, finding the neighbors for each unit, and then row normalizing. We choose $\beta_0 = 5, \beta_1 = 1, \gamma = 1, \sigma_\mu = 1, \sigma_v = 1$, a set of values for ρ ranging from -0.9 to 0.9, a set of values for λ in a similar range, $T = 3$ or 7, and $n = 50$ or 100. Each set of Monte Carlo results (corresponding to a combination of the ρ and λ values) is based on 1000 samples. For bootstrapping standard errors, the number of bootstrap samples is chosen to be $B = 999 + \lfloor n^{0.75} \rfloor$ where $\lfloor \cdot \rfloor$ denotes the integer part of $\cdot$. Due to space constraints, only a subset of results are reported. The error (v_t) distributions can be (i) normal, (ii) normal mixture (10% $N(0, 4)$ and 90% $N(0, 1)$), or (iii) centered $\chi^2(5)$ or $\chi^2(3)$. For the case of random effects model, μ and v_t are generated from the same distribution.

[9] 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 $e \sim N(0, \sigma_2^2)$. Let $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2)$. Alternatively, the elements of x_t can be randomly generated from $N(0, 4)$.

Random effects model. Table 1 reports the Monte Carlo mean and rmse for the random effects model when the data are generated according to either $m = 0$ or $m = 6$, but the model is estimated under $m = 0, 6$, and 200. The results show clearly that a correct treatment on the initial values leads to excellent estimation results in general, but a wrong treatment may give totally misleading results.

Some details are as follows. When the true m value is 0, i.e., y_0 is exogenous, estimating the model as if $m = 6$ or 200 can give very poor results when ρ is large. When ρ is not large or when ρ is negative (not reported for brevity), the estimates under a wrong m value improve but are still far from being satisfactory. In contrast, when the true m value is 6 but are treated as either 0 or 200, the resulted estimates are in general quite close to the true estimates except for the case of $m = 0$ under a large and positive ρ. This shows that the model estimates are not sensitive to the exact choice of m when y_0 is endogenous and is treated as endogenous. Comparing the results of Table 1a and 1b, we see that non-normality does not deteriorate the results of a wrong treatment of the initial values in terms of mean, but it does in terms of rmse. We note that, when the true m value is 0 but is treated as 6 or 200, the poor performance of the estimates when ρ is large and positive may be attributed to the fact that the quantities $z_m(\rho)$ and $a_m(\rho)$, given below (3.7) and above (3.11), have $1 - \rho$ as their denominators.

Table 2 reports the standard errors of the estimates based on (1) only the bootstrapped variance of the score (`seSCb`), (2) only the Hessian matrix (`seHS`), and (3) both the bootstrapped variance of the score and the Hessian (`seHSb`). The results show that when errors are normal, all three methods give averaged standard errors very close to the corresponding Monte Carlo SDs; but when errors are non-normal, only the `seHSb` method gives standard errors close to the corresponding Monte Carlo SDs; see in particular the standard errors of ϕ_μ and σ_v^2. More results corresponding to other choices of the spatial weight matrices, and other values of ρ and λ are available from the authors upon request.

Fixed effects model. The fixed effects μ are generated according to either $\frac{1}{T}\sum_{t=1}^{T} x_t + e$ or e, where e is generated in the same way as μ in the random effects model. The reported results correspond to the former. Table 3 reports the Monte Carlo mean and rmse for the fixed effects model when the data are generated according to either $m = 0$ or $m = 6$, but the model is estimated under $m = 0, 6$, and 200. The results show again that a correct treatment on the initial values leads to excellent estimation results in general, and that a wrong treatment on the initial values may lead to misleading results though to a much lesser degree as compared with the case of random effects model. When results corresponding to uncorrelated fixed effects (unreported for brevity) show that whether the individual effects are correlated with the regressors or not does not affect the performance of the fixed-effects QMLEs.

Some details are as follows. When the true m value is 0, i.e., y_0 is exogenous, estimates of the model parameters as if $m = 6$ or 200 can be poor when ρ is negative and large. When ρ is not large or when ρ is positive (not reported for brevity), the estimates under a wrong m are quite satisfactory. This shows that the model estimates are less sensitive to the treatment on y_0 when it is endogenous. Comparing the results of Table 3a and 3b, we see that non-normality does not deteriorate the results of a wrong treatment of the initial values in terms of mean, but it does in terms of rmse.

Contrary to the case of random effects model, when the true m value is 0 but is treated as 6 or 200 the estimates of the fixed effects model are poor when ρ is large but negative. This may be attributed to the quantity $c_m(\rho)$ defined below (3.21) which has $1 + \rho$ as its denominator. Comparing the results for the fixed effects model with those for the random effects model, it seems that the fixed effects model is less sensitive to the treatment of the initial values.

Table 4 reports `seSCb`, `seHS`, and `seHSb` along with the Monte Carlo SDs for comparison. The results show that when errors are normal, all three methods give averaged standard errors very close to the corresponding Monte Carlo SDs; but when errors are non-normal, the standard errors of $\hat{\sigma}_v^2$ from the `seHSb` method are much closer to the corresponding Monte Carlo SDs than those from the other two methods. More results corresponding to other choices of the spatial weight matrices, and other values of ρ and λ are available from the authors upon request.

7 Conclusion

The asymptotic properties of the quasi maximum likelihood estimators of dynamic panel models with spatial errors are studied in detail under the framework that the cross-sectional dimension n is large and the time dimension T is fixed, a typical framework for microeconomics data. Both the random effects and fixed effects models are considered, and the assumptions on the initial values and their impact on the subsequent analyses are given a special attention. The difficulty in implementing the robust standard error estimates (due to the lack of analytical expressions for the variance of the score function) is overcome by a simple residual-based bootstrap method. Monte Carlo simulation shows that both the QML estimators and the bootstrap standard errors perform well in finite samples under a correct assumption on initial observations, but the QMLEs can perform poorly when this assumption is not met.

Appendix A: Information Matrices

The elements of the information matrix for the random effects model with exogenous y_0, $\Gamma_{r,n}(\psi_0) \equiv E[\frac{\partial}{\partial\psi}\mathcal{L}^r(\psi_0)\frac{\partial}{\partial\psi'}\mathcal{L}^r(\psi_0)]$, are, for $\omega, \varpi = \lambda, \phi_\mu$:

$$\Gamma_{r,\theta\theta} = \tfrac{1}{\sigma_{v0}^2}E(\tilde{X}'\Omega_0^{-1}\tilde{X}), \qquad \Gamma_{r,\theta\sigma_v^2} = \tfrac{1}{2\sigma_{v0}^6}E(\tilde{X}'\Omega_0^{-1}uu'\Omega_0^{-1}u),$$
$$\Gamma_{r,\theta\omega} = \tfrac{1}{2\sigma_{v0}^4}E(\tilde{X}'\Omega_0^{-1}uu'P_{\omega 0}u), \qquad \Gamma_{r,\sigma_v^2\sigma_v^2} = \tfrac{1}{\sigma_{v0}^4}g(\Omega_0^{-1}, \Omega_0^{-1}),$$
$$\Gamma_{r,\sigma_v^2\omega} = \tfrac{1}{\sigma_{v0}^2}g(\Omega_0^{-1}, P_{\omega 0}), \qquad \Gamma_{r,\omega\varpi} = g(P_{\omega 0}, P_{\varpi 0}),$$

where $g(A, B) \equiv \frac{1}{4\sigma_{v0}^4}E(u'Auu'Bu) - \frac{1}{4}\text{tr}(A\Omega_0)\text{tr}(B\Omega_0)$, and P_ω is defined below Theorem 4.1. The explicit form of g can be obtained from Lemma B.4(1). The other elements do not possess explicit forms due to the complications caused by Y_{-1}.

The elements of the information matrix for the random effects model with endogenous y_0, $\Gamma_{rr,n}(\psi_0) \equiv E[\frac{\partial}{\partial\psi}\mathcal{L}^{rr}(\psi_0)\frac{\partial}{\partial\psi'}\mathcal{L}^{rr}(\psi_0)]$, are, for ω and $\varpi = \lambda, \phi_\mu$, or ϕ_ζ:

$$\Gamma_{rr,\theta\theta} = \tfrac{1}{\sigma_{v0}^2}E(X^{*\prime}\Omega_0^{*-1}X^*), \qquad \Gamma_{rr,\theta\sigma_v^2} = \tfrac{1}{\sigma_{v0}^2}f_1^*(\Omega_0^{*-1}),$$
$$\Gamma_{rr,\theta\rho} = f_1^*(P_{\rho 0}^*) - f_2^*(\Omega_0^{*-1}), \qquad \Gamma_{rr,\theta\omega} = g_1^*(P_{\omega 0}^*),$$
$$\Gamma_{rr,\sigma_v^2\sigma_v^2} = \tfrac{1}{\sigma_{v0}^4}g_1^*(\Omega_0^{*-1}, \Omega_0^{*-1}), \qquad \Gamma_{rr,\sigma_v^2\rho} = \tfrac{1}{\sigma_{v0}^2}[g_1^*(P_{\rho 0}^*, \Omega_0^{*-1}) - g_2^*(\Omega_0^{*-1}, \Omega_0^{*-1})],$$
$$\Gamma_{rr,\sigma_v^2\omega} = \tfrac{1}{\sigma_{v0}^2}g_1^*(\Omega_0^{*-1}, P_{\omega 0}^*), \qquad \Gamma_{rr,\rho\rho} = \tfrac{1}{\sigma_{v0}^4}E[(u_\rho^{*\prime}\Omega_0^{*-1}u^*)^2] + g_1^*(P_{\rho 0}^*, P_{\rho 0}^*) - 2g_2^*(\Omega_0^{*-1}, P_{\rho 0}^*),$$
$$\Gamma_{rr,\rho\omega} = g_1^*(P_{\rho 0}^*, P_{\omega 0}^*) - g_2^*(\Omega_0^{*-1}, P_{\omega 0}^*), \qquad \Gamma_{rr,\omega\varpi} = g_1^*(P_{\omega 0}^*, P_{\varpi 0}^*),$$

where $f_1^*(A) \equiv \frac{1}{2\sigma_{v0}^4}E(X^{*\prime}\Omega_0^{*-1}u^*u^{*\prime}Au^*)$, $f_2^*(A) \equiv \frac{1}{\sigma_{v0}^4}E(X^{*\prime}\Omega_0^{*-1}u^*u_\rho^{*\prime}Au^*)$, P_ω^* is defined below Theorem 4.3, $g_1^*(A, B) \equiv \frac{1}{4\sigma_v^4}E(u^{*\prime}Au^*u^{*\prime}Bu^*) - \frac{1}{4}\text{tr}(A\Omega_0^*)\text{tr}(B\Omega_0^*)$, and $g_2^*(A, B) \equiv \frac{1}{4\sigma_{v0}^4}E(u_\rho^{*\prime}Au^*u^{*\prime}Bu^*)$. As X^* is exogenous, the explicit forms of f_1^* and g_1^* can be obtained from Lemma B.4. The functions f_2^* and g_2^*, however, do not possess explicit expressions due to the complications caused by u_ρ^*.

The elements of the information matrix for the fixed effects model with exogenous or endogenous y_0, $\Gamma_{f,n}(\psi_0) = E[\frac{\partial}{\partial\psi}\mathcal{L}^f(\psi_0)\frac{\partial}{\partial\psi'}\mathcal{L}^f(\psi_0)]$, are, for $\omega, \varpi = \lambda, \phi_\zeta$:

$$\Gamma_{rr,\theta\theta} = \tfrac{1}{\sigma_{v0}^2}E(\Delta X^{\dagger\prime}\Omega_0^{\dagger-1}\Delta X^\dagger), \qquad \Gamma_{rr,\theta\sigma_v^2} = \tfrac{1}{\sigma_{v0}^2}f_1^\dagger(\Omega_0^{\dagger-1}),$$
$$\Gamma_{rr,\theta\rho} = f_1^\dagger(P_{\rho 0}^\dagger) - f_2^\dagger(\Omega_0^{\dagger-1}), \qquad \Gamma_{rr,\theta\omega} = f_1^\dagger(P_{\omega 0}^\dagger),$$
$$\Gamma_{rr,\sigma_v^2\sigma_v^2} = \tfrac{1}{\sigma_{v0}^4}g_1^\dagger(\Omega_0^{\dagger-1}, \Omega_0^{\dagger-1}), \qquad \Gamma_{rr,\sigma_v^2\rho} = \tfrac{1}{\sigma_{v0}^2}[g_1^\dagger(P_{\rho 0}^\dagger, \Omega_0^{\dagger-1}) - g_2^\dagger(\Omega_0^{\dagger-1}, \Omega_0^{\dagger-1})],$$
$$\Gamma_{rr,\sigma_v^2\omega} = \tfrac{1}{\sigma_{v0}^2}g_1^\dagger(\Omega_0^{\dagger-1}, P_{\omega 0}^\dagger), \qquad \Gamma_{rr,\rho\rho} = \tfrac{1}{\sigma_{v0}^4}E[(\Delta u_\rho^{\dagger\prime}\Omega_0^{\dagger-1}\Delta u^\dagger)^2] + g_1^\dagger(P_{\rho 0}^\dagger, P_{\rho 0}^\dagger) - 2g_2^\dagger(\Omega_0^{\dagger-1}, P_{\rho 0}^\dagger),$$
$$\Gamma_{rr,\rho\omega} = g_1^\dagger(P_{\rho 0}^\dagger, P_{\omega 0}^\dagger) - g_2^\dagger(\Omega_0^{\dagger-1}, P_{\omega 0}^\dagger), \qquad \Gamma_{rr,\omega\varpi} = g_1^\dagger(P_{\omega 0}^\dagger, P_{\varpi 0}^\dagger),$$

where $f_1^\dagger(A) \equiv \frac{1}{2\sigma_{v0}^4}E(\Delta X^{\dagger\prime}\Omega_0^{\dagger-1}\Delta u^\dagger\Delta u^{\dagger\prime}A\Delta u^\dagger)$, $f_2^\dagger(A) \equiv \frac{1}{\sigma_{v0}^4}E(\Delta X^{\dagger\prime}\Omega_0^{\dagger-1}\Delta u^\dagger\Delta u_\rho^{\dagger\prime}A\Delta u^\dagger)$, $g_1^\dagger(A, B) \equiv \frac{1}{4\sigma_v^4}E(\Delta u^{\dagger\prime}A\Delta u^\dagger\Delta u^{\dagger\prime}B\Delta u^\dagger) - \frac{1}{4}\text{tr}(A\Omega_0^\dagger)\text{tr}(B\Omega_0^\dagger)$, $g_2^\dagger(A, B) \equiv \frac{1}{4\sigma_{v0}^4}E(\Delta u_\rho^{\dagger\prime}A\Delta u^\dagger\Delta u^{\dagger\prime}B\Delta u^\dagger)$, and $P_\omega^\dagger$ is defined below Theorem 4.5. As $\Delta X^\dagger$ is exogenous, the explicit forms of $f_1^\dagger$ and $g_1^\dagger$ can be obtained from Lemma B.4. The functions $f_2^\dagger$ and $g_2^\dagger$, however, do not possess explicit expressions as they involve $\Delta u_\rho^\dagger$.

Appendix B: Some Useful Lemmas

We introduce some fundamental lemmas (existing and new) that are used in the proofs of the main results. For any random variable a with a zero mean and a finite fourth moment, let $\kappa_a \equiv E(a^4) -$

$3[E(a^2)]^2$. The first one is from Kelejian and Prucha (1999) and Lee (2002).

Lemma B.1 *Let P_n and Q_n be two $n \times n$ matrices that are uniformly bounded in both row and column sums. Let R_n be a conformable matrix whose elements are uniformly $O(o_n)$ for a certain sequence o_n. Then we have: (1) P_nQ_n is also uniformly bounded in both row and column sums; (2) any (i,j) elements $P_{n,ij}$ of P_n are uniformly bounded in i and j and $tr(P_n) = O(n)$; (3) the elements of P_nR_n and R_nP_n are uniformly $O(o_n)$.*

Noting that both W and B^{-1} are all uniformly bounded in both row and column sums under our assumptions, and recalling $A = (B'B)^{-1}(W'B + B'W)(B'B)^{-1}$ and $\dot{A} = 2(B'B)^{-1}[(W'B + B'W)A - W'W]$, it is easy to apply the above results to prove the following lemma.

Lemma B.2 *(1) $B'B, (B'B)^{-1}, \Omega, \Omega^{-1}, \Omega^*, \Omega^{*-1}, \Omega^\dagger, \Omega^{\dagger-1}, A$, and $\dot{A}$ are all uniformly bounded in both row and column sums.*

(2) $tr(D_1\Omega D_2)/n = O(1)$ for $D_1, D_2 = \Omega^{-1}, \Omega^{-1}(I_T \otimes A)\Omega^{-1}, \Omega^{-1}(J_T \otimes I_n)\Omega^{-1}$, and $\Omega^{-1}(I_T \otimes \dot{A})$. The same conclusion holds when Ω is replaced by Ω^* or $\Omega^\dagger$, and D_1 and D_2 are replaced by their analogs corresponding to the case of Ω^* or $\Omega^\dagger$.

(3) $tr(B'^{-1}RB^{-1})/n = O(1)$ where R is an $n \times n$ nonstochastic matrix that is uniformly bounded in both row and column sums.

Lemma B.3 *Let $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$ be two independent iid sequences with zero means and fourth moments. Let $\sigma_a^2 = E(a_1^2)$, $\sigma_b^2 = E(b_1^2)$. Let q_n and p_n be $n \times n$ nonstochastic matrices. Then*

(1) $E[(a'q_na)(a'p_na)] = \kappa_a \sum_{i=1} q_{n,ii}p_{n,ii} + \sigma_a^4[tr(q_n)tr(p_n) + tr(q_n(p_n + p_n'))]$,

(2) $E[(a'q_na)(b'p_nb)] = \sigma_a^2\sigma_b^2 tr(q_n)tr(p_n)$,

(3) $E[(a'q_nb)(a'p_nb)] = \sigma_a^2\sigma_b^2 tr(q_np_n')$,

where, e.g., $q_{n,ij}$ denotes the (i,j)th element of q_n.

Proof. To show (1), write $E[(a'q_na)(a'p_na)] = E(\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n a_ia_jq_{n,ij}a_ka_lp_{n,kl})$. Noting that $E(a_ia_ja_ka_l)$ will not vanish only when $i = j = k = l$, $(i = j) \neq (k = l)$, $(i = k) \neq (j = l)$, and $(i = l) \neq (j = k)$, we have

$$\begin{aligned} E[(a'q_na)(a'p_na)] &= E(a_1^4)\textstyle\sum_{i=1}^n q_{n,ii}p_{n,ii} + \sigma_a^4 \sum_{i=1}^n\sum_{j\neq i}^n(q_{n,ii}p_{n,jj} + q_{n,ij}p_{n,ij} + q_{n,ij}p_{n,ji}) \\ &= \kappa_a \textstyle\sum_{i=1}^n q_{n,ii}p_{n,ii} + \sigma_a^4\sum_{i=1}^n\sum_{j=1}^n(q_{n,ii}p_{n,jj} + q_{n,ij}p_{n,ij} + q_{n,ij}p_{n,ji}) \\ &= \kappa_a \textstyle\sum_{i=1}^n q_{n,ii}p_{n,ii} + \sigma_a^4[tr(q_n)tr(p_n) + tr(q_n(p_n + p_n'))]. \end{aligned}$$

The result (2) follows from the independence between $a'q_na$ and $b'p_nb$. For (3), $E[(a'q_nb)(a'p_nb)] = E(\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n a_ib_jq_{n,ij}a_kb_lp_{n,kl}) = E(\sum_{i=1}^n\sum_{j=1}^n a_i^2b_j^2q_{n,ij}p_{n,ij}) = \sigma_a^2\sigma_b^2 tr(q_np_n')$. ■

Lemma B.4 *Recall $u = (\iota_T \otimes I_n)\mu + (I_T \otimes B_0^{-1})v$. Let $a = \zeta + \mu(1-\rho_0^m)/(1-\rho_0) + \sum_{j=0}^{m-1}\rho_0^jB_0^{-1}v_{-j}$, where ζ, μ, and v are defined in the text. In particular, ζ_i's are iid and independent of μ and v. Let q_n, p_n, r_n, s_n, t_n be $nT \times nT$, $nT \times nT$, $n \times n$, $n \times nT$ and $n \times nT$ nonstochastic matrices, respectively. Further, q_n, p_n, and r_n are symmetric. Then*

$$\begin{array}{rcl}
\text{(1) } E[(u'q_nu)(u'p_nu)] &=& \kappa_\mu \sum_{i=1}^n G_{q_n,1ii}G_{p_n,1ii} + \kappa_v \sum_{i=1}^n G_{q_n,2ii}G_{p_n,2ii} \\
&& +\sigma_v^4[tr(q_n\Omega_0)tr(p_n\Omega_0) + 2tr(q_n\Omega_0 p_n\Omega_0)], \\
\text{(2) } E[(u'q_nu)(a'r_na)] &=& \frac{\kappa_\mu(1-\rho_0^m)^2}{(1-\rho_0)^2} \sum_{i=1}^n G_{q_n,1ii}r_{n,ii} + \sigma_v^4[tr(r_n\omega_{11})tr(q_n\Omega_0) + 2tr(\omega_{12}q_n\omega_{21}p_n)], \\
\text{(3) } E[(a's_nu)(a't_nu)] &=& \frac{\kappa_\mu(1-\rho_0^m)^2}{(1-\rho_0)^2} \sum_{i=1}^n (s_n(\iota_T\otimes I_n))_{ii}(t_n(\iota_T\otimes I_n))_{ii} \\
&& +\sigma_v^4[tr(s_n\omega_{21})tr(t_n\omega_{21}) + tr(s_n\omega_{21}t_n\omega_{21}) + tr(s_n\Omega_0 t_n'\omega_{11})], \\
\text{(4) } E[(u'q_nu)(u's_n'a)] &=& \frac{\kappa_\mu(1-\rho_0^m)}{1-\rho_0} \sum_{i=1}^n G_{q_n,1ii}((\iota_T'\otimes I_n)s_n')_{ii} + \sigma_v^4[tr(q_n\Omega_0)tr(s_n'\omega_{12}) + 2tr(\Omega_0 s_n'\omega_{12}q_n)], \\
\text{(5) } E[(a'r_na)(a's_nu)] &=& \frac{\kappa_\mu(1-\rho_0^m)^3}{(1-\rho_0)^3} \sum_{i=1}^n r_{n,ii}(s_n(\iota_T\otimes I_n))_{ii} + \sigma_v^4[(r_n\omega_{11})tr(s_n\omega_{21}) + 2tr(r_n\omega_{11}s_n\omega_{21})],
\end{array}$$

where $G_{q_n,1} \equiv (\iota_T'\otimes I_n)q_n(\iota_T\otimes I_n)$, $G_{q_n,2} \equiv (I_T\otimes B_0'^{-1})q_n(I_T\otimes B_0^{-1})$, *and, e.g.,* $G_{q_n,1ij}$ *denotes the* (i,j)*th element of* $G_{q_n,1}$.

Proof. We only sketch the proof of (1) and (2) since it mainly follows from Lemma B.3 and the proof of other claims is similar. First, let $G_{q_n,3} \equiv (\iota_T'\otimes I_n)q_n(I_T\otimes B_0^{-1})$. Then by the independence of μ and v and Lemma B.3, we have

$$\begin{array}{rcl}
E[(u'q_nu)(u'p_nu)] &=& E(\mu'G_{q_n,1}\mu\mu'G_{p_n,1}\mu + v'G_{q_n,2}vv'G_{p_n,2}v + \mu'G_{q_n,1}\mu v'G_{p_n,2}v \\
&& +v'G_{q_n,2}v\mu'G_{p_n,1}\mu + 2\mu'G_{q_n,3}v\mu'G_{p_n,3}v + 2v'G_{q_n,3}'\mu v'G_{p_n,3}\mu) \\
&=& \kappa_\mu \sum_{i=1}^n G_{q_n,1ii}G_{p_n,1ii} + \kappa_v \sum_{i=1}^n G_{q_n,2ii}G_{p_n,2ii} \\
&& +\sigma_v^4[\mathrm{tr}(q_n\Omega_0)\mathrm{tr}(p_n\Omega_0)+2\mathrm{tr}(q_n\Omega_0p_n\Omega_0)].
\end{array}$$

Next, write $a = b + B_0^{-1}c$, where $b = \zeta + \mu(1-\rho_0^m)/(1-\rho_0)$ and $c = \sum_{j=0}^{m-1}\rho_0^j v_{-j}$. Then b and c are iid and mutually independent. It follows that

$$\begin{array}{rcl}
E[(u'q_nu)(a'r_na)] &=& E(\mu'G_{q_n,1}\mu b'r_nb + v'G_{q_n,2}vc'B_0'^{-1}r_nB_0^{-1}c + \mu'G_{q_n,1}\mu c'B_0'^{-1}r_nB_0^{-1}c \\
&& +v'G_{q_n,2}vb'r_nb) \\
&=& \frac{\kappa_\mu(1-\rho_0^m)^2}{(1-\rho_0)^2} \sum_{i=1}^n G_{q_n,1ii}r_{n,ii} + \sigma_v^4[\mathrm{tr}(r_n\omega_{11})\mathrm{tr}(q_n\Omega_0) + 2\mathrm{tr}(\omega_{12}q_n\omega_{21}p_n)].
\end{array}$$

Similarly, we can prove the other claims. ■

Lemma B.5 *Suppose that* $\{P_{1n}\}$ *and* $\{P_{2n}\}$ *are sequences of matrices with row and column sums uniformly bounded. Let* $a = (a_1, \cdots, a_n)'$*, where* a_i*'s are independent random variables such that* $sup_i E|a_i|^{2+\epsilon_0}$ $< \infty$ *for some* $\epsilon_0 > 0$*. Let* $b = (b_1, \cdots, b_n)'$*, where* $b_i's$ *are iid with mean zero and* $(4+2\epsilon_0)$*th finite moments, and* $\{b_i\}$ *is independent of* $\{a_i\}$*. Let* $\sigma_{Q_n}^2$ *be the variance of* $Q_n = a'P_{1n}b + b'P_{2n}b - \sigma_v^2 tr(P_{2n})$*. Assume that the elements of* P_{1n}*,* P_{2n} *are of uniform order* $O(1/\sqrt{h_n})$ *and* $O(1/h_n)$*, respectively. If* $lim_{n\to\infty} h_n^{1+2/\epsilon_0}/n = 0$*, then* $Q_n/\sigma_{Q_n} \xrightarrow{d} N(0,1)$.

Proof. Note that Q_n is a linear-quadratic form of b as in Theorem 1 of Kelejian and Prucha (2001). The difference is that the coefficient $a'P_{1n}$ of the linear term is random. The proof proceeds by modifying that of Theorem 1 in Kelejian and Prucha (2001) or Lemma A.13 of Lee (2002). ■

We now present lemmas needed in the proofs of the main theorems. For ease of exposition, we assume that both x_{it} and z_i are scalar random variables ($p = 1$, $q = 1$) in this Appendix. For the proofs of

Theorems 2 and 4 for the SDPD model with random effects, the following presentations are essential. By continuous back substitutions, we have for $t = 0, 1, 2, \cdots$,

$$y_t = \mathbb{X}_t\beta_0 + c_{\rho_0,t}z\gamma_0 + c_{\rho_0,t}\mu + \mathbb{V}_t + \mathbb{Y}_{0,t}, \tag{B.1}$$

where for fixed y_0, $\mathbb{X}_t = \sum_{j=0}^{t-1}\rho_0^j x_{t-j}$, $\mathbb{V}_t = \sum_{j=0}^{t-1}\rho_0^j B_0^{-1}v_{t-j}$, $\mathbb{Y}_{0,t} = \rho_0^t y_0$ and $c_{\rho,t} = (1-\rho^t)/(1-\rho)$; and for endogenous y_0, $\mathbb{X}_t = \sum_{j=0}^{t+m-1}\rho_0^j x_{t-j}$, $\mathbb{V}_t = \sum_{j=0}^{t+m-1}\rho_0^j B_0^{-1}v_{t-j}$, $\mathbb{Y}_{0,t} = \rho_0^{t+m}y_{-m}$, and $c_{\rho,t} = (1-\rho^{t+m})/(1-\rho)$. Now, define $\mathbb{Y}_0 = (\mathbb{Y}_{0,0}', \mathbb{Y}_{0,1}', \cdots, \mathbb{Y}_{0,T-1}')'$. Then

$$Y_{-1} = \mathbb{X}_{(-1)}\beta_0 + (l_{\rho_0}\otimes I_n)z\gamma_0 + (l_{\rho_0}\otimes I_n)\mu + \mathbb{V}_{(-1)} + \mathbb{Y}_0, \tag{B.2}$$

where $\mathbb{X}_{(-1)} = (0, \mathbb{X}_1', \cdots, \mathbb{X}_{T-1}')'$, $\mathbb{V}_{(-1)} = (0, \mathbb{V}_1', \cdots, \mathbb{V}_{T-1}')'$, and $l_\rho = (0, c_{\rho,1}, \cdots, c_{\rho,T-1})'$ when y_0 is fixed, and $\mathbb{X}_{(-1)} = (\mathbb{X}_0', \mathbb{X}_1', \cdots, \mathbb{X}_{T-1}')'$, $\mathbb{V}_{(-1)} = (\mathbb{V}_0', \mathbb{V}_1', \cdots, \mathbb{V}_{T-1}')')$, and $l_\rho = (c_{\rho,0}, c_{\rho,1}, \cdots, c_{\rho,T-1})'$ when y_0 is endogenous. Notice that when y_0 is exogenous, Y_{-1} can also be expressed as

$$Y_{-1} = A_x X'\beta_0 + (l_{\rho_0}\otimes I_n)z\gamma_0 + (l_{\rho_0}\otimes I_n)\mu + A_v v + \mathbb{Y}_0, \tag{B.3}$$

where $A_x = \mathcal{J}_{\rho_0}'\otimes I_n$ and $A_v = \mathcal{J}_{\rho_0}'\otimes B_0^{-1}$ with

$$\mathcal{J}_\rho = \begin{pmatrix} 0 & 1 & \rho & \cdots & \rho^{T-2} \\ 0 & 0 & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{B.4}$$

Lemmas B.6-B.8 given below are used in the proof of Theorem 4.2.

Lemma B.6 *Under the assumptions of Theorem 4.2,* $E(\tilde{X}'\Omega_0^{-1}u) = 0$.

Proof. Note that $\tilde{X} = (X, Z, Y_{-1})$. By the strict exogeneity of X and Z, we can readily show that both $X'\Omega_0^{-1}u$ and $Z'\Omega_0^{-1}u$ have expectations zero. We are left to show $E(Y_{-1}'\Omega_0^{-1}u) = 0$. By (B.3), $E(Y_{-1}'\Omega_0^{-1}u) = E\left[\mu'(l_{\rho_0}'\otimes I_n)\Omega_0^{-1}u\right] + E\left[v'A_v'\Omega_0^{-1}u\right]$. Using $u = (\iota_T\otimes I_n)\mu + (I_T\otimes B_0^{-1})$ and (3.29), we have

$$\begin{aligned} E\left[\mu'(l_{\rho_0}'\otimes I_n)\Omega_0^{-1}u\right] &= E[\mu'(l_{\rho_0}'\otimes I_n)\Omega_0^{-1}(\iota_T\otimes I_n)\mu] = \phi_{\mu 0}\sigma_{v0}^2\text{tr}[\Omega_0^{-1}((\iota_T l_{\rho_0}')\otimes I_n)] \\ &= \phi_{\mu 0}\sigma_{v0}^2\text{tr}\left\{(J_T\mathcal{J}_{\rho_0})\otimes[(B_0'B_0)^{-1} + \phi_{\mu 0}TI_n]^{-1}\right\}, \end{aligned}$$

and

$$\begin{aligned} E\left[v'A_v'\Omega_0^{-1}u\right] &= E[v'A_v'\Omega_0^{-1}(I_T\otimes B_0^{-1})v] \\ &= \sigma_{v0}^2\text{tr}[\Omega_0^{-1}(I_T\otimes B_0^{-1})(\mathcal{J}_{\rho_0}\otimes B_0'^{-1})] = \sigma_{v0}^2\text{tr}[\Omega_0^{-1}(\mathcal{J}_{\rho_0}\otimes(B_0'B_0)^{-1})] \\ &= \sigma_{v0}^2\text{tr}\left\{(T^{-1}J_T\mathcal{J}_{\rho_0})\otimes[(B_0'B_0)^{-1} + \phi_{\mu 0}TI_n]^{-1}(B_0'B_0)^{-1}\right\} + \sigma_{v0}^2\text{tr}\left[(\mathcal{J}_{\rho_0} - T^{-1}J_T\mathcal{J}_{\rho_0})\otimes I_n\right], \end{aligned}$$

where we have used the fact that $E(vv'A_v') = \mathcal{J}_{\rho_0}\otimes B_0'^{-1}$. It follows that $E(Y_{-1}'\Omega_0^{-1}u) = \sigma_{v0}^2\text{tr}(\mathcal{J}_{\rho_0}\otimes I_n) = \sigma_{v0}^2\text{tr}(\mathcal{J}_{\rho_0})\text{tr}(I_n) = 0$. ■

Lemma B.7 *Under the assumptions of Theorem 4.2,* $\frac{1}{nT}\left\{\frac{\partial \mathcal{L}^r(\psi_0)}{\partial\psi\partial\psi'} - E\left[\frac{\partial \mathcal{L}^r(\psi_0)}{\partial\psi\partial\psi'}\right]\right\} = o_P(1).$

Proof. By the expressions of the Hessian matrix $\frac{\partial \mathcal{L}^r(\psi_0)}{\partial\psi\partial\psi'}$ in Section 4.2, it suffices to prove (i) $n^{-1}[\tilde{X}'\Omega_0^{-1}\tilde{X} - E(\tilde{X}'\Omega_0^{-1}\tilde{X})] = o_P(1)$; (ii) $n^{-1}[\tilde{X}'Ru - E(\tilde{X}'Ru)] = o_P(1)$ for $R = \Omega_0^{-1}$ and $P_{\omega 0}$ with $\omega = \lambda$ and ϕ_μ; (iii) $n^{-1}[u'Ru - \sigma_{v0}^2 \mathrm{tr}(R\Omega_0)] = o_P(1)$ for $R = \Omega_0^{-1}$ and $P_{\omega 0}$ with $\omega = \lambda$ and ϕ_μ; and (iv) $n^{-1}[q_{\omega\bar{\omega}}(u) - E(q_{\omega\bar{\omega}}(u))] = o_P(1)$ for ω, $\bar{\omega} = \lambda$ and ϕ_μ.

Let $\Omega_{\omega\bar{\omega}0} = \frac{\partial^2}{\partial\omega\partial\bar{\omega}}\Omega(\delta_0)$ for $\omega, \bar{\omega} = \lambda$ and ϕ_μ. Noting that Ω_0^{-1}, $\Omega_{\omega 0}$, $P_{\omega 0}$, and $\Omega_{\omega\bar{\omega}0}$ with $\omega, \bar{\omega} = \lambda$ and ϕ_μ are uniformly bounded in both row and column sums by Lemmas B.1-B.2 and $q_{\omega\bar{\omega}}(u)$ is quadratic in u, we can readily show that (iii)-(iv) hold by straightforward moment calculations, Chebyshev inequality, and Lemma B.4. For example, to show (iii), first note that $E(u'Ru) = \sigma_{v0}^2\mathrm{tr}(R\Omega_0)$. By Lemma B.4,

$$\begin{aligned}\mathrm{Var}(n^{-1}u'Ru) &= n^{-2}\{E(u'Ruu'Ru) - [E(u'Ru)]^2\} \\ &= n^{-2}\kappa_\mu \textstyle\sum_{i=1}^n G_{R,1ii}^2 + n^{-2}\kappa_v \sum_{i=1}^n G_{R,2ii}^2 + 2n^{-2}\sigma_{v0}^4 \mathrm{tr}(R\Omega_0 R\Omega_0) = O(n^{-1}),\end{aligned}$$

where the last equality follows from the fact that $G_{R,1}^2$, $G_{R,2}^2$, and $R\Omega_0 R\Omega_0$ are all uniformly bounded in both row and column sums. Then (iii) follows by Chebyshev inequality.

To prove (i), let $R = \Omega_0^{-1}$. Noticing that $\tilde{X} = (X, Z, Y_{-1})$, it is easy to show that the terms not involving Y_{-1}, such as $n^{-1}X'RX$, $n^{-1}X'RZ$, and $n^{-1}Z'RZ$ converge in probability to their expectations. For the terms involving Y_{-1}, we first have by (B.3),

$$\begin{aligned}n^{-1}Y_{-1}'RY_{-1} &= n^{-1}[A_x X'\beta_0 + (l_{\rho_0}\otimes I_n)z\gamma_0]'R[A_x X'\beta_0 + (l_{\rho_0}\otimes I_n)z\gamma_0] \\ &\quad + n^{-1}[(l_{\rho_0}\otimes I_n)\mu + A_v v]'R[(l_{\rho_0}\otimes I_n)\mu + A_v v] \\ &\quad + n^{-1}\mathbb{Y}_0'R\mathbb{Y}_0 + 2n^{-1}[A_x X'\beta_0 + (l_{\rho_0}\otimes I_n)z\gamma_0]'R[(l_{\rho_0}\otimes I_n)\mu + A_v v] \\ &\quad + 2n^{-1}[A_x X'\beta_0 + (l_{\rho_0}\otimes I_n)z\gamma_0]'R\mathbb{Y}_0 + 2n^{-1}[(l_{\rho_0}\otimes I_n)\mu + A_v v]'R\mathbb{Y}_0 \\ &\equiv \textstyle\sum_{i=1}^6 A_{ni}, \text{ say.}\end{aligned}$$

It suffices to show that each A_{ni} $(i = 1, \cdots, 6)$ converges in probability to its expectations. Take A_{n6} as an example. $E(A_{n6}) = 0$ because $\mathbb{Y}_0$ is kept fixed here. For the second moment,

$$\begin{aligned}\mathrm{Var}(A_{n6}) &= 4n^{-2}\{E[\mu'(l_{\rho_0}'\otimes I_n)R\mathbb{Y}_0\mathbb{Y}_0'R'(l_{\rho_0}\otimes I_n)\mu] + E(v'A_v R\mathbb{Y}_0\mathbb{Y}_0'R'A_v v)\} \\ &= 4n^{-2}\{\sigma_{\mu 0}^2\mathrm{tr}[R\mathbb{Y}_0\mathbb{Y}_0'R'(l_{\rho_0}l_{\rho_0}'\otimes I_n)] + \sigma_{v0}^2\mathrm{tr}(A_v R\mathbb{Y}_0\mathbb{Y}_0'R'A_v)\} = O(n^{-1}),\end{aligned}$$

where the last equality follows from the fact that both matrices in the two trace operators are uniformly bounded in both row and column sums. Similarly, we can show that $n^{-1}X'RY_{-1}$ and $n^{-1}Z'RY_{-1}$ converge to their expectations in probability, and thus (i) follows. Analogously, we can show (ii). ■

Lemma B.8 *Under the assumptions of Theorem of 4.2,* $\frac{1}{\sqrt{nT}}\frac{\partial \mathcal{L}^r(\psi_0)}{\partial\psi} \xrightarrow{d} N(0, \Gamma_r)$.

Proof. The key step of the proof is to show that $\frac{1}{\sqrt{nT}}\tilde{X}'\Omega_0^{-1}u \xrightarrow{d} N(0, \Gamma_{r,11})$ where $\Gamma_{r,11} = \mathrm{plim}_{n\to\infty}(nT)^{-1}\tilde{X}'\Omega_0^{-1}\tilde{X}$. By Cramér-Wold device, it suffices to show that for any $c = (c_1', c_2', c_3)' \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}$ with $\|c\| = 1$, $(nT)^{-1/2}c'\tilde{X}'\Omega_0^{-1}u \xrightarrow{d} N(0, c'\Gamma_{r,11}c)$. Using (B.3) and $u = (\iota_T\otimes I_n)\mu + (I_T\otimes B_0^{-1})v$, we have $c'\tilde{X}'\Omega_0^{-1}u = c_1'X\Omega_0^{-1}u + c_2'Z\Omega_0^{-1}u + c_3 Y_{-1}\Omega_0^{-1}u = \sum_{i=1}^3 T_{ni}$, where

$$\begin{aligned}T_{n1} &= [c_1'X + c_2'Z + c_3\beta_0'XA_x' + c_3\gamma_0'z(l_{\rho_0}'\otimes I_n) + c_3\mathbb{Y}_0']\Omega_0^{-1}(\iota_T\otimes I_n)\mu + c_3\mu'(l_{\rho_0}'\otimes I_n)\Omega_0^{-1}(\iota_T\otimes I_n)\mu, \\ T_{n2} &= [c_1'X + c_2'Z + c_3\beta_0'XA_x' + c_3\gamma_0'z(l_{\rho_0}'\otimes I_n) + c_3\mathbb{Y}_0']\Omega^{-1}(I_T\otimes B_0^{-1})v + c_3 v'A_v'\Omega_0^{-1}(I_T\otimes B_0^{-1})v, \\ T_{n3} &= c_3\mu'[(l_{\rho_0}'\otimes I_n)\Omega_0^{-1}(I_T\otimes B_0^{-1}) + (\iota_T'\otimes I_n)\Omega_0^{-1}A_v]v.\end{aligned}$$

It is easy to verify that $E(T_{n3})=0$, $E(T_{n1})=c_3\phi_{\mu 0}\sigma_{v0}^2\text{tr}[\Omega_0^{-1}(\iota_T l'_{\rho_0}\otimes I_n)]$, and thus $E(T_{n2})=-E(T_{n1})$ by Lemma B.6. Also, we can verify that $\text{Cov}(T_{ni},T_{nj})=0$ for $i\neq j$. It suffices to show that each T_{ni} (after appropriately centered for T_{n1} and T_{n2}) is asymptotically normal with mean zero.

Note that T_{n1} and T_{n2} are linear and quadratic functions of μ and v, respectively. For T_{n3}, it is a special case of Lemma B.5 since it can be regarded as a linear function of either μ or v, with μ and v independent of each other. So we can apply Lemma B.5 to T_{ni} to obtain

$$\{T_{ni}-E(T_{ni})\}/\sqrt{\text{Var}(T_{ni})}\xrightarrow{d}N(0,1)\text{ for }i=1,2,3.$$

Now by the independence of T_{n1} and T_{n2}, and the asymptotic independence of T_{n3} with T_{n1} and T_{n2}, we have

$$\frac{1}{\sqrt{nT}}c'\tilde{X}'\Omega_0^{-1}u=\frac{1}{\sqrt{nT}}\sum_{i=1}^3 T_{ni}\xrightarrow{d}N(0,\lim_{n\to\infty}(nT)^{-1}\sum_{i=1}^3\text{Var}(T_{ni})),$$

implying that $(nT)^{-1/2}\tilde{X}'\Omega_0^{-1}u\xrightarrow{d}N(0,\Gamma_{r,11})$ because we can readily show that $(nT)^{-1}[\tilde{X}'\Omega_0^{-1}\tilde{X}-\text{Var}(\tilde{X}'\Omega_0^{-1}u)]=o_P(1)$.

Noticing that each component of $\partial\mathcal{L}^r(\psi_0)/\partial\psi$ can be written as linear and quadratic functions of μ or v, the rest of the proof proceeds by following the above steps closely. ■

Lemmas B.9-B.13 are used in the proof of Theorem 4.4, for the SDPD model with random effects and endogenous y_0. Let R_{ts} be an $n\times n$ symmetric and positive semidefinite (p.s.d.) nonstochastic square matrix for $t,s=0,1,\cdots,T-1$. Assume that R_{ts} are uniformly bounded in both row and column sums. Recall for this case, $\mathbb{X}_t=\sum_{j=0}^{t+m-1}\rho_0^j x_{t-j}$ and $\mathbb{V}_t=\sum_{j=0}^{t+m-1}\rho_0^j B_0^{-1}v_{t-j}$.

Lemma B.9 *Suppose that the conditions in Theorem 4.4 are satisfied. Then*
(1) $E(\mathbb{V}_t'R_{ts}\mathbb{V}_s)=\sigma_v^2 tr(B_0'^{-1}R_{ts}B_0^{-1})\sum_{i=\max(0,t-s)}^{t+m-1}\rho_0^{s-t+2i}$,
(2) $E(\mathbb{X}_t'R_{ts}\mathbb{X}_s)=tr(\sum_{j=0}^{s+m-1}\sum_{k=0}^{t+m-1}\rho_0^{j+k}R_{ts}E(x_{s-j}x'_{t-k}))$,
(3) $E(\mathbb{X}_t'R_{ts}\mathbb{V}_s)=0$.

Proof. Let $P_j\equiv\rho_0^jB_0^{-1}$. Then $\mathbb{V}_t=\sum_{j=0}^{t+m-1}P_jv_{t-j}$. Noting that $E(v_t'Dv_s)=\sigma_{v0}^2\text{tr}(D)$ for any nonstochastic conformable matrix D if $t=s$ and 0 otherwise, we have

$$\begin{aligned}E(\mathbb{V}_t'R_{ts}\mathbb{V}_s)&=\textstyle\sum_{i=0}^{t+m-1}\sum_{j=0}^{s+m-1}E(v'_{t-i}P_i'R_{ts}P_jv_{s-j})=\sum_{i=\max(0,t-s)}^{t+m-1}E(v'_{t-i}P_i'R_{ts}P_{s-t+i}v_{t-i})\\&=\textstyle\sigma_{v0}^2\text{tr}(\sum_{i=\max(0,t-s)}^{t+m-1}P_i'R_{ts}P_{s-t+i})=\sigma_{v0}^2\text{tr}(B_0'^{-1}R_{ts}B_0^{-1})\sum_{i=\max(0,t-s)}^{t+m-1}\rho_0^{s-t+2i}.\end{aligned}$$

Next, noting that $\mathbb{X}_t=\sum_{j=0}^{t+m-1}\rho_0^jx_{t-j}$, we have

$$E(\mathbb{X}_t'R_{ts}\mathbb{X}_s)=\sum_{j=0}^{s+m-1}\sum_{k=0}^{t+m-1}\rho_0^{j+k}E(x'_{t-k}R_{ts}x_{s-j})=\text{tr}(\sum_{j=0}^{s+m-1}\sum_{k=0}^{t+m-1}\rho_0^{j+k}R_{ts}E(x_{s-j}x'_{t-k})).$$

Lastly, $E(\mathbb{X}_t'R_{ts}\mathbb{V}_s)=\sum_{j=0}^{s+m-1}\sum_{k=0}^{t+m-1}\rho_0^{j+k}E(x'_{t-k}R_{ts}B_0^{-1}v_{s-j})=0$. ■

Lemma B.10 *Suppose that the conditions in Theorem 4.4 are satisfied. Then*

$$\begin{aligned}(1)\ Cov(\mathbb{V}_t'R_{ts}\mathbb{V}_s,\mathbb{V}_g'R_{gh}\mathbb{V}_h)&=\rho_{tsgh,1}\{\kappa_v\textstyle\sum_{i=1}^n\overline{B}_{ts,ii}\overline{B}_{gh,ii}+2\sigma_{v0}^4tr[\overline{B}_{ts}(\overline{B}_{gh}+\overline{B}'_{gh})]\}\\&\quad+\rho_{tsgh,2}\sigma_{v0}^4tr[B_0'^{-1}R_{ts}(B_0'B_0)^{-1}R_{gh}B_0^{-1}]\\&\quad+\rho_{tsgh,3}\sigma_{v0}^4tr[B_0'^{-1}R_{ts}(B_0'B_0)^{-1}R'_{gh}B_0^{-1}],\\(2)\ Cov(\mathbb{X}_t'R_{ts}\mathbb{V}_s,\mathbb{X}_g'R_{gh}\mathbb{V}_h)&=\sigma_{v0}^2tr[\textstyle\sum_{i=0}^{t+m-1}\sum_{k=0}^{g+m-1}\sum_{j=\max(0,s-h)}^{s+m-1}\rho_0^{i+k+h-s+2j}R_{ts}\\&\quad\times(B_0'B_0)^{-1}R'_{gh}E(x'_{g-k}x_{t-i})],\\(3)\ Cov(\mathbb{X}_t'R_{ts}\mathbb{X}_s,\mathbb{X}_g'R_{gh}\mathbb{X}_h)&=O(n),\end{aligned}$$

where $\overline{B}_{ts,ii}$ *denotes the* (i,i)*th element of* $\overline{B}_{ts} \equiv B_0'^{-1} R_{ts} B_0^{-1}$, $\rho_{tsgh,1} = \sum_{j=\max(0,t-s,t-g,t-h)}^{t+m-1} \rho_0^{(s+g+h-3t+4j)}$, $\rho_{tsgh,2} = \sum_{i=\max(0,t-g)}^{t+m-1} \rho_0^{g-t+2i} \sum_{j=\max(0,s-h)}^{s+m-1} \rho_0^{h-s+2j} 1(j \neq i+s-t)$, *and* $\rho_{tsgh,3} = \sum_{i=\max(0,t-h)}^{t+m-1} \rho_0^{h-t+2i} \sum_{j=\max(0,s-g)}^{s+m-1} \rho_0^{g-s+2j} 1(j \neq i+s-t)$.

Proof. Let R_1 and R_2 be arbitrary $n \times n$ nonstochastic matrices. We can show that

$$E[(v_t' R_1 v_s)(v_g' R_2 v_h)] = \begin{cases} \kappa_v \sum_{i=1}^n R_{1,ii} R_{2,ii} + \sigma_{v0}^4 \{\text{tr}(R_1)\text{tr}(R_2) + \text{tr}[R_1(R_2 + R_2')]\} & \text{if } t = s = g = h \\ \sigma_{v0}^4 \text{tr}(R_1)\text{tr}(R_2) & \text{if } t = s \neq g = h \\ \sigma_{v0}^4 \text{tr}(R_1 R_2) & \text{if } t = g \neq s = h \\ \sigma_{v0}^4 \text{tr}(R_1 R_2') & \text{if } t = h \neq s = g \\ 0 & \text{otherwise} \end{cases}.$$

Consequently,

$$\begin{aligned} & E(\mathbb{V}_t' R_{ts} \mathbb{V}_s \mathbb{V}_g' R_{gh} \mathbb{V}_h) \\ = \; & E(\textstyle\sum_{i=0}^{t+m-1} \sum_{j=0}^{s+m-1} \sum_{k=0}^{g+m-1} \sum_{l=0}^{h+m-1} \rho_0^{i+j+k+l} v_{t-i}' B_0'^{-1} R_{ts} B_0^{-1} v_{s-j} v_{g-k}' B_0'^{-1} R_{gh} B_0^{-1} v_{h-l}) \\ = \; & \textstyle\sum_{j=\max(0,t-s,t-g,t-h)}^{t+m-1} \rho_0^{(s+g+h-3t+4j)} \{\kappa_v \sum_{i=1}^n (B_0'^{-1} R_{ts} B_0^{-1})_{ii} (B_0'^{-1} R_{gh} B_0^{-1})_{ii} \\ & + \sigma_{v0}^4 [\text{tr}(B_0'^{-1} R_{ts} B_0^{-1}) \text{tr}(B_0'^{-1} R_{gh} B_0^{-1}) + 2\text{tr}(B_0'^{-1} R_{ts} B_0^{-1} (B_0'^{-1} R_{gh} B_0^{-1} + B_0'^{-1} R_{gh}' B_0^{-1}))]\} \\ & + \sigma_{v0}^4 \textstyle\sum_{i=\max(0,t-s)}^{t+m-1} \rho_0^{s-t+2i} \text{tr}(B_0'^{-1} R_{ts} B_0^{-1}) \sum_{j=\max(0,g-h)}^{g+m-1} \rho_0^{h-g+2j} \text{tr}(B_0'^{-1} R_{gh} B_0^{-1}) 1(j \neq i+g-t) \\ & + \textstyle\sum_{i=\max(0,t-g)}^{t+m-1} \rho_0^{g-t+2i} \sum_{j=\max(0,s-h)}^{s+m-1} \rho_0^{h-s+2j} \sigma_{v0}^4 \text{tr}(B_0'^{-1} R_{ts} (B_0' B_0)^{-1} R_{gh} B_0^{-1}) 1(j \neq i+s-t) \\ & + \textstyle\sum_{i=\max(0,t-h)}^{t+m-1} \rho_0^{h-t+2i} \sum_{j=\max(0,s-g)}^{s+m-1} \rho_0^{g-s+2j} \sigma_{v0}^4 \text{tr}(B_0'^{-1} R_{ts} (B_0' B_0)^{-1} R_{gh}' B_0^{-1}) 1(j \neq i+s-t). \end{aligned}$$

Then (1) follows by Lemma B.9. For (2), we have

$$\begin{aligned} \text{Cov}(\mathbb{X}_t' R_{ts} \mathbb{V}_s, \mathbb{X}_g' R_{gh} \mathbb{V}_h) &= E(\mathbb{X}_t' R_{ts} \mathbb{V}_s (\mathbb{X}_t' R_{gh} \mathbb{V}_s)') \\ &= \sum_{i=0}^{t+m-1} \sum_{j=0}^{s+m-1} \sum_{k=0}^{g+m-1} \sum_{l=0}^{h+m-1} \rho_0^{i+j+k+l} E[x_{t-i}' R_{ts} B_0^{-1} v_{s-j} (x_{g-k}' R_{gh} B_0^{-1} v_{h-l})'] \\ &= \sigma_{v0}^2 \text{tr} \left[\sum_{i=0}^{t+m-1} \sum_{k=0}^{g+m-1} \sum_{j=\max(0,s-h)}^{s+m-1} \rho_0^{i+k+h-s+2j} R_{ts} (B_0' B_0)^{-1} R_{gh}' E(x_{g-k}' x_{t-i}) \right]. \end{aligned}$$

The expression for $\text{Cov}(\mathbb{X}_t' R_{ts} \mathbb{X}_t, \mathbb{X}_g' R_{gh} \mathbb{X}_h)$ is quite complicated, but we can use Lemmas B.1-B.2 to show it is of order $O(n)$, which suffices for our purpose. ■

Lemma B.11 *Suppose that the conditions in Theorem 4.4 are satisfied. Then*

(1) $(nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} [\mathbb{V}_t' R_{ts} \mathbb{V}_s - E(\mathbb{V}_t' R_{ts} \mathbb{V}_s)] \xrightarrow{p} 0$,

(2) $(nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \mathbb{X}_t' R_{ts} \mathbb{V}_s \xrightarrow{p} 0$,

(3) $(nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} [\mathbb{X}_t' R_{ts} \mathbb{X}_s - E(\mathbb{X}_t' R_{ts} \mathbb{X}_s)] \xrightarrow{p} 0$.

Proof. By Lemmas B.1, B.2, B.9, and B.10, we can show that $(nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} E(\mathbb{V}_t' R_{ts} \mathbb{V}_s) = O(1)$, and $\text{Var}(n^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \mathbb{V}_t' R_{ts} \mathbb{V}_s) = n^{-2} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \sum_{g=0}^{T-1} \sum_{h=0}^{T-1} \text{Cov}(\mathbb{V}_t' R_{ts} \mathbb{V}_s, \mathbb{V}_g' R_{gh} \mathbb{V}_h) = O(n^{-1})$. Then (1) follows from Chebyshev inequality. For (2), we have $E[\frac{1}{nT} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \mathbb{X}_t' R_{ts} \mathbb{V}_s] = 0$, and

$$\begin{aligned}\mathrm{Var}\left(n^{-1}\sum_{t=0}^{T-1}\sum_{s=0}^{T-1}\mathbb{X}_t' R_{ts}\mathbb{V}_s\right) &= n^{-2}\sum_{t=0}^{T-1}\sum_{s=0}^{T-1}\sum_{g=0}^{T-1}\sum_{h=0}^{T-1}\mathrm{Cov}(\mathbb{X}_t' R_{ts}\mathbb{V}_s, \mathbb{X}_g' R_{ts}\mathbb{V}_h)\\ &= n^{-2}\sum_{t=0}^{T-1}\sum_{s=0}^{T-1}\sum_{g=0}^{T-1}\sum_{h=0}^{T-1}\sigma_{v0}^2\sum_{i=0}^{t+m-1}\sum_{k=0}^{g+m-1}\sum_{j=\max(0,s-h)}^{s+m-1}\mathrm{tr}[\rho_0^{i+k+h-s+2j}R_{ts}\\ &\quad\times(B_0'B_0)^{-1}R_{gh}'E(x_{g-k}x_{t-i}')]\\ &= O(n^{-1}),\end{aligned}$$

where the last equality follows because (i) x_{it} are independent across i with second moments uniformly bounded in i, (ii) $R_{ts}(B_0'B_0)^{-1}R_{gh}'$ are uniformly bounded in both row and column sums by Lemmas B.1-B.2, and (iii) elements of $R_{ts}(B_0'B_0)^{-1}R_{gh}'E(x_{g-k}x_{t-i}')$ are uniformly bounded by the same lemmas. Hence the conclusion follows from Chebyshev inequality. (3) follows from Lemma B.10 and Chebyshev inequality. ■

Lemma B.12 *Under the assumptions of Theorem 4.4,* $\frac{1}{n(T+1)}\left\{\frac{\partial\mathcal{L}^{rr}(\psi_0)}{\partial\psi\partial\psi'} - E\left[\frac{\partial\mathcal{L}^{rr}(\psi_0)}{\partial\psi\partial\psi'}\right]\right\} = o_P(1)$.

Proof. Let $u^* = u^*(\theta_0,\rho_0)$ and $u_\rho^* = u_\rho^*(\theta_0,\rho_0) = \frac{\partial}{\partial\rho}u^*(\theta_0,\rho_0)$. Noting that $E(X^{*\prime}Ru^*) = 0$ for any $n(T+1)\times n(T+1)$ nonstochastic matrix R and X_ρ^* is free of ρ, by the expressions of the Hessian matrix $\frac{\partial\mathcal{L}^{rr}(\psi_0)}{\partial\psi\partial\psi'}$ in Section 4.2, it suffices to prove

(i) $n^{-1}\left[X^{*\prime}\Omega_0^{*-1}X^* - E\left(X^{*\prime}\Omega_0^{*-1}X^*\right)\right] = o_P(1)$;
(ii) $n^{-1}X^{*\prime}Ru^* = o_P(1)$ for $R = \Omega_0^{*-1}$ and $P_{\omega0}^*$ with $\omega = \rho$, λ, ϕ_μ, and ϕ_ζ;
(iii) $n^{-1}\left[u^{*\prime}Ru^* - E\left(u^{*\prime}Ru^*\right)\right] = o_P(1)$ for $R = \Omega_0^{*-1}$ and $P_{\omega0}^*$ with $\omega = \rho$, λ, ϕ_μ, and ϕ_ζ;
(iv) $n^{-1}\left[X_\rho^{*\prime}\Omega_0^{*-1}u^* - E\left(X_\rho^{*\prime}\Omega_0^{*-1}u^*\right)\right] = o_P(1)$;
(v) $n^{-1}\left[X^{*\prime}\Omega_0^{*-1}u_\rho^* - E\left(X^{*\prime}\Omega_0^{*-1}u_\rho^*\right)\right] = o_P(1)$;
(vi) $n^{-1}\left[u_\rho^{*\prime}Ru^* - E\left(u_\rho^{*\prime}Ru^*\right)\right] = o_P(1)$ for $R = \Omega_0^{*-1}$ and $P_{\omega0}^*$ with $\omega = \rho$, λ, ϕ_μ, and ϕ_ζ;
(vii) $n^{-1}\left[u_{\rho\rho}^{*\prime}\Omega_0^{*-1}u^* - E\left(u_{\rho\rho}^{*\prime}\Omega_0^{*-1}u^*\right)\right] = o_P(1)$;
$(viii)$ $n^{-1}\left[u_\rho^{*\prime}\Omega_0^{*-1}u_\rho^* - E\left(u_\rho^{*\prime}\Omega_0^{*-1}u_\rho^*\right)\right] = o_P(1)$;
(ix) $n^{-1}\left[q_{\omega\bar\omega}^*(u^*) - E\left(q_{\omega\bar\omega}^*(u^*)\right)\right] = o_P(1)$ for $\omega, \bar\omega = \rho$, λ, ϕ_μ, and ϕ_ζ.

Let $\Omega_{\omega\bar\omega0}^* = \frac{\partial^2}{\partial\omega\partial\bar\omega}\Omega^*(\delta_0)$ for ρ, λ, ϕ_μ, and ϕ_ζ. Noting that Ω_0^{*-1}, $\Omega_{\omega0}^*$, $P_{\omega0}^*$ and $\Omega_{\omega\bar\omega0}^*$ with $\omega,\bar\omega = \rho$, λ, ϕ_μ, and ϕ_ζ are uniformly bounded in both row and column sums and $q_{\omega\bar\omega}^*(u^*)$ is quadratic in u^*, we can readily show that (i)-(iv) and (ix) hold by straightforward moment calculations and Chebyshev inequality. Noting that $u_\rho^* = -\begin{pmatrix}\dot a_{m0}z\gamma_0\\ Y_{-1}\end{pmatrix}$ and $u_{\rho\rho}^* = -\begin{pmatrix}\ddot a_{m0}z\gamma_0\\ 0_{nT\times1}\end{pmatrix}$ with $\dot a_{m0} = \frac{d}{d\rho}a_m(\rho_0)$ and $\ddot a_{m0} = \frac{d^2}{d\rho^2}a_m(\rho_0)$, we can readily prove (v)-(vii) by Chebyshev inequality. In fact, $E\left(u_{\rho\rho}^{*\prime}\Omega_0^{*-1}u^*\right) = 0$ in (vii).

We are left to prove $(viii)$. Write $\Omega_0^{*-1} = \begin{pmatrix} \omega_*^{11} & \omega_*^{12} \\ \omega_*^{12\prime} & \omega_*^{22} \end{pmatrix}$ where $\omega_*^{11}, \omega_*^{12}$, and ω_*^{22} are $n \times n$, $n \times nT$, and $nT \times nT$ matrices, respectively.

$$
\begin{aligned}
n^{-1} u_\rho^{*\prime} \Omega_0^{*-1} u_\rho^* &= n^{-1} \begin{pmatrix} \dot{a}_{m0} z \gamma_0 \\ Y_{-1} \end{pmatrix}' \begin{pmatrix} \omega_*^{11} & \omega_*^{12} \\ \omega_*^{12\prime} & \omega_*^{22} \end{pmatrix} \begin{pmatrix} \dot{a}_{m0} z \gamma_0 \\ Y_{-1} \end{pmatrix} \\
&= n^{-1} \left((\dot{a}_{m0})^2 \gamma_0' z' \omega_*^{11} z \gamma_0 + 2 \dot{a}_{m0} \gamma_0' z' \omega_*^{12} Y_{-1} + Y_{-1}' \omega_*^{22} Y_{-1} \right).
\end{aligned}
$$

To show the convergence of $n^{-1} u_\rho^{*\prime} \Omega_0^{*-1} u_\rho^*$ to its expectation, it suffices show each term in the last expression converges to its expectation. We only show $n^{-1}[Y_{-1}' \omega_*^{22} Y_{-1} - E\left(Y_{-1}' \omega_*^{22} Y_{-1}\right)] = o_P(1)$ since the proof that $n^{-1}[(\dot{a}_{m0})^2 \gamma_0' z' \omega_*^{11} z \gamma_0 \ - E\left((\dot{a}_{m0})^2 \gamma_0' z' \omega_*^{11} z \gamma_0\right)] = o_P(1)$ and that $n^{-1}[\dot{a}_{m0} \gamma_0' z' \omega_*^{12} Y_{-1} - \left(\dot{a}_{m0} \gamma_0' z' \omega_*^{12} Y_{-1}\right)] = o_P(1)$ is similar and simpler. By (B.2)

$$
\begin{aligned}
n^{-1} Y_{-1}' \omega_*^{22} Y_{-1} &= n^{-1} \left(\mathbb{X}_{(-1)} \beta_0 + (l_{\rho_0} \otimes I_n) z \gamma_0 + (l_{\rho_0} \otimes I_n) \mu + \mathbb{V}_{(-1)} + \mathbb{Y}_0 \right)' \omega_*^{22} \\
&\quad \times \left(\mathbb{X}_{(-1)} \beta_0 + (l_{\rho_0} \otimes I_n) z \gamma_0 + (l_{\rho_0} \otimes I_n) \mu + \mathbb{V}_{(-1)} + \mathbb{Y}_0 \right).
\end{aligned}
$$

After expressing out the right hand side of the last expression, it has 25 terms, most of which can easily be shown to converge to their respective expectations. The exceptions are terms involving $\mathbb{X}_{(-1)}$ and $\mathbb{V}_{(-1)}$, namely: $n^{-1} \beta_0' \mathbb{X}_{(-1)}' \omega_*^{22} \mathbb{X}_{(-1)} \beta_0$, $n^{-1} \beta_0' \mathbb{V}_{(-1)}' \omega_*^{22} \mathbb{V}_{(-1)}$, $n^{-1} \beta_0' \mathbb{X}_{(-1)}' \omega_*^{22} \mathbb{V}_{(-1)}$, $n^{-1} \beta_0' \mathbb{X}_{(-1)}' \omega_*^{22} (l_{\rho_0} \otimes I_n) z \gamma_0$, $n^{-1} \beta_0' \mathbb{X}_{(-1)}' \omega_*^{22} (l_{\rho_0} \otimes I_n) \mu$, $n^{-1} \mathbb{V}_{(-1)}' \omega_*^{22} (l_{\rho_0} \otimes I_n) z \gamma_0$, $n^{-1} \mathbb{V}_{(-1)}' \omega_*^{22} (l_{\rho_0} \otimes I_n) \mu$, $n^{-1} \beta_0' \mathbb{X}_{(-1)}' \omega_*^{22} \mathbb{Y}_0$, and $n^{-1} \mathbb{V}_{(-1)}' \omega_*^{22} \mathbb{Y}_0$. The first three terms converge in probability to their expectations by Lemma B.11. We can show the other terms converge in probability to their expectations by similar arguments to those used in proving Lemmas B.9-B.11. ■

Lemma B.13 *Under the assumptions of Theorem 4.4,* $\frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \psi} \xrightarrow{d} N(0, \Gamma_{rr})$.

Proof. By Cramér-Wold device, it suffices to show that for any $c = (c_1', c_2, c_3, c_4, c_5, c_6)' \in \mathbb{R}^{p+q+k} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ with $\|c\| = 1$, $S_n^* \equiv \frac{1}{\sqrt{nT}} c' \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \psi} \xrightarrow{d} N(0, c' \Gamma_{rr} c)$. Using the expression for elements of $\frac{\partial \mathcal{L}^{rr}(\psi)}{\partial \psi}$ defined in Section 4.2, we can readily obtain

$$
\begin{aligned}
S_n^* &= \frac{1}{\sqrt{nT}} \left[c_1' \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \theta'} + c_2 \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \sigma_v^2} + c_3 \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \rho} + c_4 \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \lambda} + c_5 \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \phi_\mu} + c_6 \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \phi_{\phi_\varsigma}} \right] \\
&= \frac{1}{\sqrt{nT}} \left\{ \frac{1}{\sigma_{v0}^2} c_1' X^{*\prime} \Omega_0^{*-1} u^* - \frac{c_3}{\sigma_{v0}^2} u_\rho^{*\prime} \Omega_0^{*-1} u^* + \frac{c_2}{2\sigma_{v0}^2} \left[\sigma_{v0}^{-2} u^{*\prime} \Omega_0^{*-1} u^* - n(T+1) \right] \right. \\
&\quad + \frac{c_3}{2\sigma_{v0}^2} \left[u^{*\prime} P_{\rho 0}^* u^* - \sigma_{v0}^2 \mathrm{tr}(P_{\rho 0}^* \Omega_0^*) \right] + \frac{c_4}{2\sigma_{v0}^2} \left[u^{*\prime} P_{\lambda 0}^* u^* - \sigma_{v0}^2 \mathrm{tr}(P_{\lambda 0}^* \Omega_0^*) \right] \\
&\quad \left. + \frac{c_5}{2\sigma_{v0}^2} \left[u^{*\prime} P_{\phi_\mu 0}^* u^* - \sigma_{v0}^2 \mathrm{tr}(P_{\phi_\mu 0}^* \Omega_0^*) \right] + \frac{c_6}{2\sigma_{v0}^2} \left[u^{*\prime} P_{\phi_\varsigma 0}^* u^* - \sigma_{v0}^2 \mathrm{tr}(\Omega_{\phi_\varsigma 0}^* \Omega_0^*) \right] \right\} \\
&= S_{n1}^* + S_{n2}^* + \left[S_{n3}^* - \mathrm{E}\left(S_{n3}^*\right) \right]
\end{aligned}
$$

where $S_{n1}^{*}=\frac{1}{\sqrt{nT}}\frac{1}{\sigma_{v0}^{2}}c_{1}^{\prime}X^{*\prime}\Omega_{0}^{*-1}u^{*}$, $S_{n2}^{*}=\frac{-1}{\sqrt{nT}}\frac{c_{3}}{\sigma_{v0}^{2}}u_{\rho}^{*\prime}\Omega_{0}^{*-1}u^{*}$, $S_{n3}^{*}=\frac{1}{\sqrt{nT}}\frac{1}{2\sigma_{v0}^{2}}u^{*\prime}\bar{\Omega}_{0}^{*}u^{*}$ and $\bar{\Omega}_{0}^{*}=\frac{c_{2}}{\sigma_{v0}^{2}}\Omega_{0}^{*-1}$ $+c_{3}P_{\rho0}^{*}+c_{4}P_{\lambda0}^{*}+c_{5}P_{\phi_{\mu}0}^{*}+c_{6}P_{\phi_{\varsigma}0}^{*}$. Note that

$$
\begin{aligned}
S_{n1}^{*} &= \frac{1}{\sqrt{nT}}\frac{1}{\sigma_{v0}^{2}}c_{1}^{\prime}X^{*\prime}\begin{pmatrix}\omega_{*}^{11} & \omega_{*}^{12}\\ \omega_{*}^{21} & \omega_{*}^{22}\end{pmatrix}\begin{pmatrix}\zeta+a_{m0}\mu+\sum_{j=0}^{m-1}\rho_{0}^{j}B_{0}^{-1}v_{-j}\\ (\iota_{T}\otimes I_{n})\mu+(I_{T}\otimes B_{0}^{-1})v\end{pmatrix}\\
&= \frac{1}{\sqrt{nT}}\frac{1}{\sigma_{v0}^{2}}c_{1}^{\prime}X^{*\prime}\begin{pmatrix}\omega_{*}^{11}\\ \omega_{*}^{21}\end{pmatrix}\zeta+\frac{1}{\sqrt{nT}}\frac{1}{\sigma_{v0}^{2}}c_{1}^{\prime}X^{*\prime}\begin{pmatrix}\omega_{*}^{11}a_{m0}+\omega_{*}^{12}(\iota_{T}\otimes I_{n})\\ \omega_{*}^{21}a_{m0}+\omega_{*}^{22}(\iota_{T}\otimes I_{n})\end{pmatrix}\mu\\
&\quad +\frac{1}{\sqrt{nT}}\frac{1}{\sigma_{v0}^{2}}c_{1}^{\prime}X^{*\prime}\begin{pmatrix}\omega_{*}^{12}\\ \omega_{*}^{22}\end{pmatrix}(I_{T}\otimes B_{0}^{-1})v+\frac{1}{\sqrt{nT}}\frac{1}{\sigma_{v0}^{2}}c_{1}^{\prime}X^{*\prime}\begin{pmatrix}\omega_{*}^{11}\\ \omega_{*}^{21}\end{pmatrix}\sum_{j=0}^{m-1}\rho_{0}^{j}B_{0}^{-1}v_{-j}\\
&\equiv S_{n1,1}^{*}+S_{n1,2}^{*}+S_{n1,3}^{*}+S_{n1,4}^{*},\ \text{say},
\end{aligned}
$$

where $S_{n1,1}^{*}, S_{n1,2}^{*}$, $S_{n1,3}^{*}$, and $S_{n1,4}^{*}$ are linear in ζ, μ, v and v_{-j}'s, respectively. Similarly

$$
\begin{aligned}
S_{n3}^{*} &= \frac{1}{\sqrt{nT}}\frac{1}{2\sigma_{v0}^{2}}\Big\{\zeta^{\prime}\bar{\omega}_{*}^{11}\zeta+\mu^{\prime}\left[a_{m0}\bar{\omega}_{*}^{11}+(\iota_{T}^{\prime}\otimes I_{n})\bar{\omega}_{*}^{22}(\iota_{T}\otimes I_{n})+2a_{m0}(\iota_{T}^{\prime}\otimes I_{n})\bar{\omega}_{*}^{21}\right]\mu\\
&\quad +v^{\prime}(I_{T}\otimes B^{\prime-1})\bar{\omega}_{*}^{22}(I_{T}\otimes B_{0}^{-1})v+\left(\sum_{j=0}^{m-1}\rho_{0}^{j}B_{0}^{-1}v_{-j}\right)^{\prime}\bar{\omega}_{*}^{11}\left(\sum_{j=0}^{m-1}\rho_{0}^{j}B_{0}^{-1}v_{-j}\right)\\
&\quad +2\left[\left(a_{m0}^{2}\mu+\sum_{j=0}^{m-1}\rho_{0}^{j}B_{0}^{-1}v_{-j}\right)+(\iota_{T}\otimes I_{n})\mu+(I_{T}\otimes B_{0}^{-1})v\right]^{\prime}\bar{\omega}_{*}^{21}\zeta\\
&\quad +2a_{m0}\mu^{\prime}\bar{\omega}_{*}^{11}\sum_{j=0}^{m-1}\rho_{0}^{j}B_{0}^{-1}v_{-j}+2\mu^{\prime}(\iota_{T}^{\prime}\otimes I_{n})\bar{\omega}_{*}^{22}(I_{T}\otimes B_{0}^{-1})v\\
&\quad +2v^{\prime}(I_{T}\otimes B_{0}^{\prime-1})\bar{\omega}_{*}^{21}\left(a_{m0}\mu+\sum_{j=0}^{m-1}\rho_{0}^{j}B_{0}^{-1}v_{-j}\right)+2\mu^{\prime}(\iota_{T}^{\prime}\otimes I_{n})\bar{\omega}_{*}^{21}\sum_{j=0}^{m-1}\rho_{0}^{j}B_{0}^{-1}v_{-j}\Big\}.
\end{aligned}
$$

where $\bar{\Omega}_{0}^{*-1}=\begin{pmatrix}\bar{\omega}_{*}^{11} & \bar{\omega}_{*}^{21\prime}\\ \bar{\omega}_{*}^{21} & \bar{\omega}_{*}^{22}\end{pmatrix}$ with $\bar{\omega}_{*}^{11}$, $\bar{\omega}_{*}^{12}$, and $\bar{\omega}_{*}^{22}$ being $n\times n$, $nT\times n$, and $nT\times nT$ matrices. Apparently, S_{n3}^{*} can be written as the summation of five asymptotically independent terms, i.e., $S_{n3}^{*}=\sum_{j=1}^{5}S_{n3,j}^{*}$, where $S_{n3,1}^{*}$, $S_{n3,2}^{*}, S_{n3,3}^{*}$, and $S_{n3,4}^{*}$ are quadratic functions of ζ, μ, v, and v_{-j}'s, respectively, and $S_{n3,5}^{*}$ is the summation of terms that are bilinear in any two of ζ, μ, v, and v_{-j}'s. Analogous to the proof of Lemma B.8, we can use $u_{\rho}^{*}=-\left(\dot{a}_{m0}\left(z\gamma_{0}\right)^{\prime},Y_{-1}^{\prime}\right)$ and the expression of Y_{-1} in (B.2) to write $S_{n2}^{*}=\sum_{j=1}^{5}S_{n2,j}^{*}$, where $S_{n2,1}^{*}$, $S_{n2,2}^{*}$, and $S_{n2,3}^{*}$ are quadratic functions of μ, v, and v_{-j}'s, respectively, $S_{n2,4}^{*}$ is a bilinear function that contains summation of terms which are linear in any two of ζ, μ, v, and v_{-j}'s, and $S_{n2,5}^{*}$ is the summation of terms that are linear in one of ς, μ, v, and v_{-j}'s. Consequently, we can write $S_{n}^{*}=\sum_{j=1}^{6}s_{nj}^{*}$, where $s_{n1}^{*},...,s_{n4}^{*}$ are quadratic functions of ζ, μ, v, and v_{-j}'s, respectively, s_{n5}^{*} is a summation of terms tat are bilinear in any two of ζ, μ, v, and v_{-j}'s, and s_{n6}^{*} is summation of terms that are linear in ζ, μ, v, and v_{-j}'s. By the mutual independence of ζ, μ, v, and v_{-j}'s and their zero mean property, these six terms are either independent or asymptotically independent. By Lemma B.5,

$$\{s_{nj}^{*}-\mathrm{E}(s_{nj}^{*})\}/\sqrt{\mathrm{Var}(s_{nj}^{*})}\xrightarrow{d}N(0,1).$$

It follows that $S_{n}^{*}\xrightarrow{d}N(0,\lim_{n\to\infty}\sum_{j=1}^{6}\mathrm{Var}(s_{nj}^{*}))$, implying that $S_{n}^{*}\xrightarrow{d}N(0,c^{\prime}\Gamma_{rr}c)$. ■

Lemmas B.14-B.15 are used in the proof of Theorem 4.6 for the fixed effects model.

Lemma B.14 *Under the assumptions of Theorem 4.6,* $\frac{1}{nT}\left\{\frac{\partial\mathcal{L}^{f}(\psi_{0})}{\partial\psi\partial\psi^{\prime}}-E\left[\frac{\partial\mathcal{L}^{f}(\psi_{0})}{\partial\psi\partial\psi^{\prime}}\right]\right\}=o_{P}(1)$.

Proof. Noting that $E\left(\Delta X^{\dagger\prime} R \Delta u\right)=0$ for any $nT \times nT$ nonstochastic matrix R, by the expressions of the Hessian matrix $\frac{\partial \mathcal{L}^f(\psi_0)}{\partial \psi \partial \psi'}$ in Section 4.3, it suffices to prove

(*i*) $n^{-1}[\Delta X^{\dagger\prime} \Omega_0^{\dagger-1} \Delta X^{\dagger} - E(\Delta X^{\dagger\prime} \Omega_0^{\dagger-1} \Delta X^{\dagger})] = o_P(1)$;
(*ii*) $n^{-1} \Delta X^{\dagger\prime} R \Delta u = o_P(1)$ for $R = \Omega_0^{\dagger-1}$ and $P_{\omega 0}^{\dagger}$ with $\omega = \rho,\ \lambda,$and ϕ_ζ;
(*iii*) $n^{-1}[\Delta u' R \Delta u - \sigma_{v0}^2 \operatorname{tr}(R \Omega_0^{\dagger})] = o_P(1)$ for $R = \Omega_0^{\dagger-1}$ and $P_{\omega 0}^{\dagger}$ with $\omega = \rho,\ \lambda,$ and ϕ_ζ;
(*iv*) $n^{-1}[\Delta X^{\dagger\prime} \Omega_0^{\dagger-1} \Delta u_\rho - E(\Delta X^{\dagger\prime} \Omega_0^{\dagger-1} \Delta u_\rho)] = o_P(1)$;
(*v*) $n^{-1}\left[\Delta u_\rho' R \Delta u - E\left(\Delta u_\rho' R \Delta u\right)\right] = o_P(1)$ for $R = \Omega_0^{\dagger-1}$ and $P_{\omega 0}^{\dagger}$ with $\omega = \rho,\ \lambda,$ and ϕ_ζ;
(*vi*) $n^{-1}[\Delta u_\rho' \Omega_0^{\dagger-1} \Delta u_\rho - E\left(\Delta u_\rho' \Omega_0^{\dagger-1} \Delta u_\rho\right)] = o_P(1)$;
(*vii*) $n^{-1}[q_{\omega\bar{\omega}}^{\dagger}(\Delta u) - E(q_{\omega\bar{\omega}}^{\dagger}(\Delta u))] = o_P(1)$ for $\omega,\ \bar{\omega} = \rho,\ \lambda,$ and ϕ_ζ.

Let $\Omega_{\omega\bar{\omega}0}^{\dagger} = \frac{\partial^2}{\partial \omega \partial \bar{\omega}} \Omega^{\dagger}(\delta_0)$ for $\rho,\ \lambda,$ and ϕ_ζ. Noting that $\Omega_0^{\dagger-1}$, $\Omega_{\omega 0}^{\dagger}$, $P_{\omega 0}^{\dagger}$ and $\Omega_{\omega\bar{\omega}0}^{\dagger}$ with $\omega,\ \bar{\omega} = \rho,\ \lambda,$ and ϕ_ζ are uniformly bounded in both row and column sums and $q_{\omega\bar{\omega}}^{\dagger}(\Delta u)$ is quadratic in Δu, we can show that (*i*)-(*vii*) hold by straightforward moment calculations and Chebyshev inequality. Below we only demonstrate the proof of (*iii*) and (*vi*) since the proof of the other claims is similar or simpler.

Since $E(\Delta u' R \Delta u) = \sigma_{v0}^2 \operatorname{tr}(R \Omega_0^{\dagger})$, by Chebyshev inequality (*iii*) follows provided $\operatorname{Var}(n^{-1} \Delta u' R \Delta u) = o(1)$. Let $\Delta v_{(0)} = B_0 \zeta + \rho_0^m v_{-m+1} + \sum_{j=0}^{m-1} \rho_0^j \Delta v_{1-j}$, $\Delta v_{(1)} = (\Delta v_2', \ldots . \Delta v_T')'$, and $\Delta v = (\Delta v_{(0)}', \Delta v_{(1)}')'$. Then $\Delta u = (I_n \otimes B_0^{-1}) \Delta v$ and $\Delta u' R \Delta u = \Delta v' (I_n \otimes B_0^{'-1}) R (I_n \otimes B_0^{-1}) \Delta v = \Delta v' \widetilde{R} \Delta v$, where $\widetilde{R} \equiv (I_n \otimes B_0^{'-1}) R (I_n \otimes B_0^{-1})$. Now, write

$$R = \begin{pmatrix} \underset{n\times n}{R_{00}} & \underset{n\times n(T-1)}{R_{01}} \\ \underset{n(T-1)\times n}{R_{10}} & \underset{n(T-1)\times n(T-1)}{R_{11}} \end{pmatrix}$$

and partition $\widetilde{R}$ similarly. Let C be a $(T-1) \times T$ matrix with $C_{ij} = -1$ if $i = j$, $C_{ij} = 1$ if $j = i+1$, and $C_{ij} = 0$ otherwise. Then $\Delta v_{(1)} = (C \otimes I_n) v$, where $v = (v_1', \cdots, v_T')'$. So

$$\begin{aligned} \Delta v' \widetilde{R} \Delta v &= \Delta v_{(0)}' \widetilde{R}_{00} \Delta v_{(0)} + \Delta v_{(1)}' \widetilde{R}_{11} \Delta v_{(1)} + \Delta v_{(0)}' (R_{01} + R_{10}') \Delta v_{(1)} \\ &= \Delta v_{(0)}' \widetilde{R}_{00} \Delta v_{(0)} + v' (C' \otimes I_n) \widetilde{R}_{11} (C \otimes I_n) v + \Delta v_{(0)}' (R_{01} + R_{10}') (C \otimes I_n) v \end{aligned}$$

Then by Cauchy-Schwarz inequality

$$\operatorname{Var}(\Delta u' R \Delta u) \leq 3\operatorname{Var}(\Delta v_{(0)}' \widetilde{R}_{00} \Delta v_{(0)}) + 3\operatorname{Var}(v' (C' \otimes I_n) \widetilde{R}_{11} (C \otimes I_n) v) + 3\operatorname{Var}(\Delta v_{(0)}' (R_{01} + R_{10}') (C \otimes I_n) v).$$

Write $\Delta v_{(0)} = B_0 \zeta + v_1 + \rho_0^{m-1}(\rho_0 - 1) v_{-m+1} + \sum_{j=0}^{m-2} \rho_0^j (\rho_0 - 1) v_{-j}$. Since $B_0' \widetilde{R}_{00} B_0$ is uniformly bounded in both row and column sums, by Lemma B.3(1)

$$\operatorname{Var}(\zeta' B_0' \widetilde{R}_{00} B_0 \zeta) = \kappa_\zeta \sum_{i=1}^{n} [(B_0' \widetilde{R}_{00} B_0)_{ii}]^2 + \sigma_{\zeta 0}^4 \operatorname{tr}(B_0' \widetilde{R}_{00} B_0 B_0' (\widetilde{R}_{00} + \widetilde{R}_{00}') B_0) = O(n).$$

Similarly, we can show that $\operatorname{Var}(v_1' \widetilde{R}_{00} v_1) = O(n)$, $\operatorname{Var}(v_{-m+1}' \widetilde{R}_{00} v_{-m+1}) = O(n)$, and $\operatorname{Var}(\sum_{j=0}^{m-2} \rho_0^j v_{-j}' \widetilde{R}_{00} \times \sum_{j=0}^{m-2} \rho_0^j v_{-j}) = O(n)$. It follows from Cauchy-Schwarz inequality that $\operatorname{Var}(\Delta v_{(0)}' \widetilde{R}_{00} \Delta v_{(0)}) = O(n)$. By the same token, we can show that $\operatorname{Var}(v' (C' \otimes I_n) \widetilde{R}_{11} (C \otimes I_n) v) = O(n)$, and $\operatorname{Var}(\Delta v_{(0)}' (R_{01} + R_{10}') (C \otimes I_n) v) = O(n)$. This completes the proof of (*iii*).

Now, we show (vi). Let $\Delta Y^* = (0_{1\times n}, \Delta y_1', \cdots, \Delta y_{T-1}')'$. Then $\Delta u_\rho = -\Delta Y^*$. Let $k_\rho = (0, 1, \rho, \cdots, \rho^{T-2})'$, $\mathcal{X} = (0_{1\times n}, 0_{1\times n}, (\Delta x_2\beta_0)', \cdots, \sum_{j=0}^{T-3} \rho_0^j(\Delta x_{T-1-j}\beta_0)')$, and $\mathcal{V} = (0_{1\times n},\, 0_{1\times n}, (\Delta v_2)', \cdots, \sum_{j=0}^{T-3} \rho_0^j(\Delta v_{T-1-j})')$. Since $\Delta y_1 = \tilde{\Delta}x\pi_0 + \Delta x_1\beta_0 + \tilde{\Delta}u_1$ and

$$\Delta y_t = \rho_0^{t-1}\Delta y_1 + \sum_{j=0}^{t-2}\rho_0^j\Delta x_{t-j}\beta_0 + \sum_{j=0}^{t-2}\rho_0^j B_0^{-1}\Delta v_{t-j} \text{ for } t = 2, 3, \cdots, \tag{B.5}$$

we have $\Delta Y^* = k_{\rho_0} \otimes \Delta y_1 + \mathcal{X} + (I_T \otimes B_0^{-1})\mathcal{V}$. It follows that

$$\begin{aligned}\operatorname{Var}\left(\Delta u_\rho'\Omega_0^{\dagger-1}\Delta u_\rho\right) \quad \leq \quad & 3\operatorname{Var}\left(\left(k_{\rho_0}'\otimes\Delta y_1\right)\Omega_0^{\dagger-1}\left(k_{\rho_0}\otimes\Delta y_1\right)\right) + 3\operatorname{Var}\left(\mathcal{X}'\Omega_0^{\dagger-1}\mathcal{X}\right)\\ & +3\operatorname{Var}\left(\mathcal{V}'(I_T\otimes B_0'^{-1})\Omega_0^{\dagger-1}(I_T\otimes B_0^{-1})\mathcal{V}\right)\end{aligned}$$

We can show that each term on the right hand side of the last expression is $O(n)$. Then (vi) follows by Chebyshev inequality. ■

Lemma B.15 *Suppose that the conditions in Theorem 4.6 are satisfied. Then* $\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}^f(\psi_0)}{\partial\psi} \xrightarrow{d} N(0, \Gamma_f)$.

Proof. By Cramér-Wold device, it suffices to show that for any $c = (c_1', c_2, c_3, c_4, c_5)' \in \mathbb{R}^{p+k} \times \mathbb{R}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}$ with $\|c\| = 1$, $S_n^\dagger \equiv \frac{1}{\sqrt{nT}}c'\frac{\partial\mathcal{L}^f(\psi_0)}{\partial\psi} \xrightarrow{d} N(0, c'\Gamma_f c)$. Recall $\Delta u = \Delta u(\theta_0, \rho_0)$. Let $\Delta u_\rho = -(0_{n\times1}', \Delta y_1', \cdots, \Delta y_{T-1}')'$, and $P_{\omega0}^\dagger = P_\omega^\dagger(\delta_0)$ for $\omega = \rho, \lambda$, and ϕ_ζ. Using the expression for elements of $\frac{\partial\mathcal{L}^f(\psi)}{\partial\psi}$ defined in Section 4.3, we can readily obtain

$$\begin{aligned}S_n^\dagger \quad = \quad & \frac{1}{\sqrt{nT}}\left[c_1'\frac{\partial\mathcal{L}^f(\psi_0)}{\partial\theta'} + c_2\frac{\partial\mathcal{L}^f(\psi_0)}{\partial\sigma_v^2} + c_3\frac{\partial\mathcal{L}^f(\psi_0)}{\partial\rho} + c_4\frac{\partial\mathcal{L}^f(\psi_0)}{\partial\lambda} + c_5\frac{\partial\mathcal{L}^f(\psi_0)}{\partial\phi_{\phi_\zeta}}\right]\\ = \quad & \frac{1}{\sqrt{nT}}\left\{\frac{1}{\sigma_{v0}^2}c_1'\Delta X^{\dagger\prime}\Omega_0^{\dagger-1}\Delta u - \frac{c_3}{\sigma_{v0}^2}\Delta u_\rho'\Omega_0^{\dagger-1}\Delta u\right.\\ & +\frac{c_2}{2\sigma_{v0}^2}\left[\frac{1}{2\sigma_{v0}^2}\Delta u'\Omega_0^{\dagger-1}\Delta u - nT\right] + \frac{c_3}{2\sigma_{v0}^2}\left[\Delta u'P_{\rho0}^\dagger\Delta u - \sigma_{v0}^2\operatorname{tr}(P_{\rho0}^\dagger\Omega_0^\dagger)\right]\\ & \left.+\frac{c_4}{2\sigma_{v0}^2}\left[\Delta u'P_{\lambda0}^\dagger\Delta u - \sigma_{v0}^2\operatorname{tr}(P_{\lambda0}^\dagger\Omega_0^\dagger)\right] + \frac{c_5}{2\sigma_{v0}^2}\left[\Delta u'P_{\phi_\zeta0}^\dagger\Delta u - \sigma_{v0}^2\operatorname{tr}(P_{\phi_\zeta0}^\dagger\Omega_0^\dagger)\right]\right\}\\ = \quad & S_{n1}^\dagger + S_{n2}^\dagger + \left[S_{n3}^\dagger - \mathrm{E}\left(S_{n3}^\dagger\right)\right]\end{aligned}$$

where $S_{n1}^\dagger = \frac{1}{\sqrt{nT}}\frac{1}{\sigma_{v0}^2}c_1'\Delta X^{\dagger\prime}\Omega_0^{\dagger-1}\Delta u$, $S_{n2}^\dagger = \frac{-1}{\sqrt{nT}}\frac{c_3}{\sigma_{v0}^2}\Delta u_\rho'\Omega_0^{\dagger-1}\Delta u$, $S_{n3}^\dagger = \frac{1}{\sqrt{nT}}\frac{1}{2\sigma_{v0}^2}\Delta u'\bar{\Omega}_0^\dagger\Delta u$ and $\bar{\Omega}_0^\dagger = \frac{c_2}{\sigma_{v0}^2}\Omega_0^{\dagger-1} + c_3P_{\rho0}^\dagger + c_4P_{\lambda0}^\dagger + c_5P_{\phi_\zeta0}^\dagger$. Analogous to the proof of Lemma B.13, one can write $S_n^\dagger = \sum_{j=1}^5 s_{nj}^\dagger$, where $s_{n1}^\dagger, ..., s_{n3}^\dagger$ are quadratic functions of ζ, v, and v_{-j}'s, respectively, $s_{n4}^\dagger$ is a summation of terms tat are bilinear in any two of ζ, v, and v_{-j}'s, and $s_{n5}^\dagger$ is summation of terms that are linear in ς, v, and v_{-j}'s. By the mutual independence of ζ, v, and v_{-j}'s and their zero mean property, these five terms are either independent or asymptotically independent. By Lemma B.5,

$$\{s_{nj}^\dagger - \mathrm{E}(s_{nj}^\dagger)\}/\sqrt{\operatorname{Var}(s_{nj}^\dagger)} \xrightarrow{d} N(0,1).$$

It follows that $S_n^\dagger \xrightarrow{d} N(0, \lim_{n\to\infty}\sum_{j=1}^5\operatorname{Var}(s_{nj}^\dagger))$, implying that $S_n^\dagger \xrightarrow{d} N(0, c'\Gamma_{rr}c)$. ■

Appendix C: Proofs of the Theorems

Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be, respectively, the smallest and the largest eigenvalues of the matrix A.

Proof of Theorem 4.1. By Theorem 3.4 of White (1994), it suffices to show that: (i) $\frac{1}{nT}[\mathcal{L}_c^{r*}(\delta) - \mathcal{L}_c^r(\delta)] \xrightarrow{p} 0$ uniformly in $\delta \in \mathbf{\Delta}$, and (ii) $\limsup_{n\to\infty} \max_{\delta \in N_\epsilon^c(\delta_0)} \frac{1}{nT}[\mathcal{L}_c^{r*}(\delta) - \mathcal{L}_c^{r*}(\delta_0)] < 0$ for any $\epsilon > 0$, where $N_\epsilon^c(\delta_0)$ is the complement of an open neighborhood of δ_0 on $\mathbf{\Delta}$ of radius ϵ. By (3.5) and (4.3), $\frac{2}{nT}[\mathcal{L}_c^{r*}(\delta) - \mathcal{L}_c^r(\delta)] = -\ln \tilde{\sigma}_v^2(\delta) + \ln \hat{\sigma}_v^2(\delta)$. To show (i), it is sufficient to show

$$\hat{\sigma}_v^2(\delta) - \tilde{\sigma}_v^2(\delta) = o_P(1) \text{ uniformly in } \delta \in \mathbf{\Delta}. \tag{C.1}$$

By the definition of $\tilde{u}(\delta)$ below (3.4), we have $\tilde{u}(\delta) = Y - \tilde{X}(\tilde{X}'\Omega^{-1}\tilde{X})^{-1}\tilde{X}'\Omega^{-1}Y = \Omega^{1/2}M\Omega^{-1/2}Y$ where $M = I_{nT} - \Omega^{-1/2}\tilde{X}(\tilde{X}'\Omega^{-1}\tilde{X})^{-1}\tilde{X}'\Omega^{-1/2}$ is a projection matrix. This, in conjunction with the fact that $M\Omega^{-1/2}\tilde{X} = 0$, implies that

$$\hat{\sigma}_v^2(\delta) = \frac{1}{nT}\tilde{u}(\delta)'\Omega^{-1}\tilde{u}(\delta) = \frac{1}{nT}Y'\Omega^{-1/2}M\Omega^{-1/2}Y = \frac{1}{nT}u'\Omega^{-1/2}M\Omega^{-1/2}u. \tag{C.2}$$

By (4.1) and the fact that $Y = \tilde{X}\theta_0 + u$, $\tilde{\theta}(\delta) = \theta_0 + \theta^*(\delta)$ where $\theta^*(\delta) = [E(\tilde{X}'\Omega^{-1}\tilde{X})]^{-1}E(\tilde{X}'\Omega^{-1}u)$. Then $u(\tilde{\theta}(\delta)) = Y - \tilde{X}\tilde{\theta}(\delta) = u - \tilde{X}\theta^*(\delta)$. By (4.2) and using the expression for $\theta^*(\delta)$, we have

$$\begin{aligned}
\tilde{\sigma}_v^2(\delta) &= \frac{1}{nT}E\left\{\left[u - \tilde{X}\theta^*(\delta)\right]'\Omega^{-1}\left[u - \tilde{X}\theta^*(\delta)\right]\right\} \\
&= \frac{1}{nT}E\left(u'\Omega^{-1}u\right) + \frac{1}{nT}\theta^*(\delta)'E(\tilde{X}'\Omega^{-1}\tilde{X})\theta^*(\delta) - \frac{2}{nT}\theta^*(\delta)'E(\tilde{X}'\Omega^{-1}u) \\
&= \frac{\sigma_{v0}^2}{nT}\operatorname{tr}\left(\Omega^{-1}\Omega_0\right) - \frac{1}{nT}[E(\tilde{X}'\Omega^{-1}u)]'[E(\tilde{X}'\Omega^{-1}\tilde{X})]^{-1}E(\tilde{X}'\Omega^{-1}u),
\end{aligned} \tag{C.3}$$

where recall $\Omega_0 \equiv \Omega(\delta_0)$ and $\Omega(\delta)$ is defined in (3.2). Combining (C.2)-(C.3) yields

$$\begin{aligned}
\hat{\sigma}_v^2(\delta) - \tilde{\sigma}_v^2(\delta) &= \frac{1}{nT}[u'\Omega^{-1}u - \sigma_{v0}^2\operatorname{tr}\left(\Omega^{-1}\Omega_0\right)] - \frac{1}{nT}u'\Omega^{-1/2}P\Omega^{-1/2}u \\
&\quad + \frac{1}{nT}[E(\tilde{X}'\Omega^{-1}u)]'[E(\tilde{X}'\Omega^{-1}\tilde{X})]^{-1}E(\tilde{X}'\Omega^{-1}u) \\
&= \frac{1}{nT}\operatorname{tr}[\Omega^{-1}(uu' - \sigma_{v0}^2\Omega_0)] \\
&\quad - \left\{Q_{xu}(\delta)'Q_{xx}(\delta)^{-1}Q_{xu}(\delta) - \{E[Q_{xu}(\delta)]\}'\{E[Q_{xx}(\delta)]\}^{-1}E[Q_{xu}(\delta)]\right\} \\
&\equiv \Pi_{n1}(\delta) - \Pi_{n2}(\delta), \text{ say,}
\end{aligned}$$

where $P = \Omega^{-1/2}\tilde{X}(\tilde{X}'\Omega^{-1}\tilde{X})^{-1}\tilde{X}'\Omega^{-1/2}$, $Q_{xx}(\delta) = \frac{1}{nT}\tilde{X}'\Omega^{-1}\tilde{X}$, and $Q_{xu}(\delta) = \frac{1}{nT}\tilde{X}'\Omega^{-1}u$.

For $\Pi_{n1}(\delta)$, we can show that $E[\Pi_{n1}(\delta)] = 0$ and $E[\Pi_{n1}(\delta)]^2 = O(n^{-1})$ as in the proof of Lemma B.7. So the pointwise convergence of $\Pi_{n1}(\delta)$ to 0 follows by Chebyshev inequality. The uniform convergence holds if we can show that $\Pi_{n1}(\delta)$ is stochastic equicontinuous. To achieve this, we first show that $\inf_{\delta\in\mathbf{\Delta}} \lambda_{\min}(\Omega(\delta))$ is bounded away from 0:

$$\begin{aligned}
\inf_{\delta\in\mathbf{\Delta}} \lambda_{\min}(\Omega(\delta)) &\geq \inf_{\delta\in\mathbf{\Delta}} \lambda_{\min}\{\phi_\mu(J_T \otimes I_n) + I_T \otimes [B(\lambda)'B(\lambda)]^{-1}\} \\
&\geq \inf_{\lambda\in\mathbf{\Lambda}} \lambda_{\min}(I_T \otimes [B(\lambda)'B(\lambda)]^{-1}) \geq \inf_{\lambda\in\mathbf{\Lambda}} \lambda_{\min}([B(\lambda)'B(\lambda)]^{-1}) \\
&\geq \inf_{\lambda\in\mathbf{\Lambda}} \{\lambda_{\min}[B(\lambda)^{-1}]\}^2 = \{\sup_{\lambda\in\mathbf{\Lambda}} \lambda_{\max}[B(\lambda)]\}^{-2} \geq \bar{c}_\lambda^{-2} > 0
\end{aligned} \tag{C.4}$$

by Facts 8.16.20 and B.14.20 in Bernstein (2005) and Assumption G2(v). Now, let δ, $\bar{\delta} \in \mathbf{\Delta}$. By Cauchy-Schwarz inequality,

$$\begin{aligned}
|\Pi_{n1}(\delta) - \Pi_{n1}(\bar{\delta})| &= \left| \frac{1}{nT} \text{tr}\{\Omega(\delta)^{-1}[\Omega(\delta) - \Omega(\bar{\delta})]\Omega(\bar{\delta})^{-1}(uu' - \sigma_{v0}^2 \Omega_0)\} \right| \\
&\leq \frac{1}{nT}[\text{tr}\{\Omega(\delta)^{-1}[\Omega(\delta) - \Omega(\bar{\delta})]\Omega(\bar{\delta})^{-2}[\Omega(\delta) - \Omega(\bar{\delta})]\Omega(\delta)^{-1}\}]^{1/2} \left\| uu' - \sigma_{v0}^2 \Omega_0 \right\| \\
&\leq [\lambda_{\min}(\Omega(\bar{\delta}))]^{-2} \frac{1}{\sqrt{nT}} \left\| \Omega(\delta) - \Omega(\bar{\delta}) \right\| \frac{1}{\sqrt{nT}} \left\| uu' - \sigma_{v0}^2 \Omega_0 \right\|.
\end{aligned}$$

Straightforward moment calculations and Chebyshev inequality lead to $\left\| uu' - \sigma_{v0}^2 \Omega_0 \right\| / \sqrt{nT} = O_P(1)$. In addition, $\left\| \Omega(\delta) - \Omega(\bar{\delta}) \right\| / \sqrt{nT} \to 0$ as $\left\| \delta - \bar{\delta} \right\| \to 0$. Thus, $\{\Pi_{n1}(\delta)\}$ is stochastically equicontinuous by Theorem 21.10 in Davidson (1994).

For $\Pi_{n2}(\delta)$, we decompose it as follows

$$\begin{aligned}
\Pi_{n2}(\delta) &= \{Q_{xu}(\delta) - E[Q_{xu}(\delta)]\}' Q_{xx}(\delta)^{-1} Q_{xu}(\delta) \\
&\quad + \{E[Q_{xu}(\delta)]\}' Q_{xx}(\delta)^{-1} \{E[Q_{xx}(\delta)] - Q_{xx}(\delta)\} \{E[Q_{xx}(\delta)]\}^{-1} Q_{xu}(\delta) \\
&\quad + \{E[Q_{xu}(\delta)]\}' \{E[Q_{xx}(\delta)]\}^{-1} \{Q_{xu}(\delta) - E[Q_{xu}(\delta)]\} \\
&\equiv \Pi_{n2,1}(\delta) + \Pi_{n2,2}(\delta) + \Pi_{n2,3}(\delta), \text{ say.}
\end{aligned}$$

By Assumption G1(v), $\sup |\phi_\mu| \leq c_\phi$ for some $c_\phi < \infty$. Noting that by G2(v)

$$\begin{aligned}
\sup_{\delta \in \mathbf{\Delta}} \lambda_{\max}(\Omega(\delta)) &\leq \sup_{\delta \in \mathbf{\Delta}} \lambda_{\max}\{\phi_\mu(J_T \otimes I_n) + I_T \otimes [B(\lambda)'B(\lambda)]^{-1}\} \\
&\leq \sup_{\phi_\mu} \{\phi_\mu \lambda_{\max}(J_T \otimes I_n) + \lambda_{\max}\{[B(\lambda)'B(\lambda)]^{-1}\}\} \\
&\leq c_\phi T + \{\inf_{\lambda \in \mathbf{\Lambda}} \lambda_{\min}[B(\lambda)]\}^{-2} \leq c_\phi T + \underline{c}_\lambda^{-2} < \infty,
\end{aligned} \tag{C.5}$$

we have

$$\begin{aligned}
\inf_{\delta \in \mathbf{\Delta}} \lambda_{\min}(Q_{xx}(\delta)) &\geq \left[\sup_{\delta \in \mathbf{\Delta}} \lambda_{\max}(\Omega(\delta)) \right]^{-1} \lambda_{\min}\left(\frac{1}{nT} \tilde{X}'\tilde{X} \right) \\
&\geq \left(c_\phi T + \underline{c}_\lambda^{-2} \right)^{-1} \lambda_{\min}\left(\frac{1}{nT} \tilde{X}'\tilde{X} \right).
\end{aligned}$$

This implies that $\sup_{\delta \in \mathbf{\Delta}} \left\| Q_{xx}(\delta)^{-1} \right\| = O_P(1)$ by Assumption R(iv). It is straightforward to show that $Q_{xu}(\delta) - E[Q_{xu}(\delta)] = o_P(1)$ uniformly in δ by Chebyshev inequality and the arguments for stochastic equicontinuity. In addition, $E[Q_{xu}(\delta)] = O(1)$ uniformly in δ. So $Q_{xu}(\delta) = O_P(1)$ uniformly in δ. Consequently,

$$\begin{aligned}
|\Pi_{n2,1}(\delta)| &\leq \| Q_{xu}(\delta) - E[Q_{xu}(\delta)] \| \left\| Q_{xx}(\delta)^{-1} \right\| \| Q_{xu}(\delta) \| \\
&= o_P(1) O_P(1) O_P(1) = o_P(1) \text{ uniformly in } \delta.
\end{aligned}$$

By the same token, we can show that $\Pi_{n2,s}(\delta) = o_P(1)$ uniformly in δ for $s = 2, 3$. It follows that $\Pi_{n2}(\delta) = o_P(1)$ uniformly in δ. Hence $\sup_{\delta \in \mathbf{\Delta}} |\hat{\sigma}_v^2(\delta) - \tilde{\sigma}_v^2(\delta)| = o_P(1)$ as desired.

To show (ii), we can define an auxiliary process $\{U_{nT}\}$ such that (3.1) is now satisfied with u replaced by U_{nT} and $U_{nT} \sim N(0, \sigma_v^2 \Omega)$ with $\Omega = \Omega(\delta)$ and is independent of (X, Z). [If u is normally distributed, one just sets $U_{nT}=u$.] The true value of $(\theta,\sigma_v^2,\delta)$ is given by $(\theta_0,\sigma_{v0}^2,\delta_0)$. Now the quasi-log-likelihood function $\mathcal{L}^r(\psi)$ in (3.3) becomes the exact log-likelihood function. By the principle of maximum likelihood and Jensen inequality, one can readily show that $\mathcal{L}_c^{r*}(\delta)\leq\mathcal{L}_c^{r*}(\delta_0)$ for any $\delta\in\mathbf{\Delta}$. Observing that $\tilde{\sigma}_v^2(\delta_0)=\frac{\sigma_{v0}^2}{nT}\text{tr}(\Omega_0^{-1}\Omega_0)=\sigma_{v0}^2$ by (C.3) and Lemma B.6, we have

$$\begin{aligned}\frac{1}{nT}[\mathcal{L}_c^{r*}(\delta)-\mathcal{L}_c^{r*}(\delta_0)] &= \frac{1}{2nT}\{\log|\Omega_0|-\log|\Omega(\delta)|\}+\frac{1}{2}\{\log[\tilde{\sigma}_v^2(\delta_0)]-\log[\tilde{\sigma}_v^2(\delta)]\}\\ &= \frac{1}{2nT}\{\log|\sigma_{v0}^2\Omega_0|-\log|\tilde{\sigma}_v^2(\delta)\Omega(\delta)|\}.\end{aligned}$$

Then (ii) follows from Assumption R(iv). This completes the proof of the theorem. ■

Proof of Theorem 4.2. By Taylor series expansion,

$$0=\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}^r(\hat{\psi})}{\partial\psi}=\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}^r(\psi_0)}{\partial\psi}+\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\bar{\psi})}{\partial\psi\partial\psi'}\sqrt{nT}(\hat{\psi}-\psi_0),$$

where elements of $\bar{\psi}=(\bar{\theta}',\bar{\sigma}_v^2,\bar{\delta})'$ lie in the segment joining the corresponding elements of $\hat{\psi}$ and ψ_0 and $\bar{\delta}=(\bar{\lambda},\bar{\phi}_\mu)'$. Thus

$$\sqrt{nT}(\hat{\psi}-\psi_0)=-\left[\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\bar{\psi})}{\partial\psi\partial\psi'}\right]^{-1}\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}^r(\psi_0)}{\partial\psi}.$$

By Theorem 4.1, $\hat{\psi}\xrightarrow{p}\psi_0$. Consequently, $\bar{\psi}\xrightarrow{p}\psi_0$, and it suffices to show that: (i) $\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\bar{\psi})}{\partial\psi\partial\psi'}-\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\psi_0)}{\partial\psi\partial\psi'}=o_P(1)$, (ii) $\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\psi_0)}{\partial\psi\partial\psi'}\xrightarrow{p}H_r$, and (iii) $\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}^r(\psi_0)}{\partial\psi}\xrightarrow{d}N(0,\Gamma_r)$. (ii) and (iii) follow from Lemmas B.7 and B.8, respectively. We are left to show (i).

With the expression of $\frac{\partial^2}{\partial\psi\partial\psi'}\mathcal{L}^r(\psi)$ given in Section 4.2, it suffices to show that $\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\bar{\psi})}{\partial\omega\partial\varpi'}-\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\psi_0}{\partial\omega\partial\varpi'}=o_P(1)$ for ω, $\varpi=\theta$, σ_v^2, λ, and ϕ_μ. We do this only for the cases of $(\omega,\varpi)=(\theta,\theta)$, (θ,σ_v^2), and (σ_v^2,σ_v^2) as the other cases can be shown analogously. First, write

$$-\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}^r(\bar{\psi})}{\partial\theta\partial\theta'}-\frac{\partial^2\mathcal{L}^r(\psi_0)}{\partial\theta\partial\theta'}\right]=\left(\frac{1}{\bar{\sigma}_v^2}-\frac{1}{\sigma_{v0}^2}\right)\frac{\tilde{X}'\Omega(\bar{\delta})^{-1}\tilde{X}}{nT}+\frac{1}{nT\sigma_{v0}^2}\tilde{X}'[\Omega(\bar{\delta})^{-1}-\Omega_0^{-1}]\tilde{X}. \tag{C.6}$$

Noting that $\bar{\sigma}_v^2-\sigma_{v0}^2=o_P(1)$ by Theorem 4.1 and $(nT)^{-1}\tilde{X}'\Omega(\bar{\delta})^{-1}\tilde{X}=O_P(1)$, the first term on the right hand side of the last expression is $o_P(1)$. For the second term, we first show that

$$\lambda_{\max}[\Omega_0-\Omega(\bar{\delta})]=O_p(\|\bar{\delta}-\delta_0\|). \tag{C.7}$$

To see this, write $\Omega_0-\Omega(\bar{\delta})=(\phi_{\mu0}-\bar{\phi}_\mu)(J_T\otimes I_n)+r_n(\bar{\lambda})$, where $r_n(\lambda)=I_T\otimes\{[B(\lambda_0)'B(\lambda_0)]^{-1}-[B(\lambda)'B(\lambda)]^{-1}\}$ is a symmetric matrix. By the repeated use of the fact that

$$\lambda_{\max}(A\otimes C)\leq\lambda_{\max}(A)\lambda_{\max}(C) \tag{C.8}$$

for any two real symmetric matrices [see, e.g., Fact 8.16.20 of Bernstein (2005)], we have

$$\begin{aligned}\lambda_{\max}[r_n(\bar{\lambda})] &\leq \lambda_{\max}\{[B(\lambda_0)'B(\lambda_0)]^{-1}-[B(\bar{\lambda})'B(\bar{\lambda})]^{-1}\}\\ &= \lambda_{\max}([B(\lambda_0)'B(\lambda_0)]^{-1}[B(\bar{\lambda})'B(\bar{\lambda})-B(\lambda_0)'B(\lambda_0)][B(\bar{\lambda})'B(\bar{\lambda})]^{-1})\\ &\leq \{\inf_{\lambda\in\mathbf{\Lambda}}\lambda_{\min}[B(\lambda)'B(\lambda)]\}^{-2}\lambda_{\max}[B(\bar{\lambda})'B(\bar{\lambda})-B(\lambda_0)'B(\lambda_0)]=O_P(\bar{\lambda}-\lambda_0)\end{aligned}$$

where the last equality follows from Assumption G2 and the fact that

$$\begin{aligned}\lambda_{\max}[B(\bar{\lambda})'B(\bar{\lambda})-B(\lambda_0)'B(\lambda_0)] &= \lambda_{\max}[(\lambda_0-\bar{\lambda})(W'+W)+(\bar{\lambda}^2-\lambda_0^2)W'W] \\ &\leq |\bar{\lambda}-\lambda_0|\lambda_{\max}(W'+W)+(\bar{\lambda}^2-\lambda_0^2)\lambda_{\max}(W'W) \\ &= O_P(\bar{\lambda}-\lambda_0)\end{aligned}$$

under Assumption G2. Noting that $\lambda_{\max}(J_T\otimes I_n)=T$, we can apply the fact that

$$\lambda_{\max}(A+C)\leq\lambda_{\max}(A)+\lambda_{\max}(C) \tag{C.9}$$

to obtain $\lambda_{\max}[\Omega_0-\Omega(\bar{\delta})]\leq T|\phi_{\mu0}-\bar{\phi}_\mu|+\lambda_{\max}(r_n(\bar{\lambda}))=O_p(\|\bar{\delta}-\delta_0\|)$. Thus (C.7) follows. Let c be an arbitrary column vector in $\mathbb{R}^{p+q+1}$. Then by Cauchy-Schwarz inequality, (C.4), and (C.7)

$$\begin{aligned}&\frac{1}{n}|c'\tilde{X}'[\Omega(\bar{\delta})^{-1}-\Omega_0^{-1}]\tilde{X}c| \\ =\;&\frac{1}{n}|c'\tilde{X}'\Omega(\bar{\delta})^{-1}[\Omega_0-\Omega(\bar{\delta})]\Omega_0^{-1}\tilde{X}c| \\ \leq\;&\frac{1}{n}\{c'\tilde{X}'\Omega(\bar{\delta})^{-1}[\Omega_0-\Omega(\bar{\delta})][\Omega_0-\Omega(\bar{\delta})]\Omega(\bar{\delta})^{-1}\tilde{X}c\}^{1/2}[c'\tilde{X}'\Omega_0^{-1}\Omega_0^{-1}\tilde{X}c]^{1/2} \\ \leq\;&\lambda_{\max}[\Omega_0-\Omega(\bar{\delta})][\lambda_{\min}(\Omega(\bar{\delta}))]^{-1}[\lambda_{\min}(\Omega_0)]^{-1}\frac{1}{n}\left\|\tilde{X}c\right\|^2=O_P(\|\bar{\delta}-\delta_0\|)=o_P(1).\end{aligned} \tag{C.10}$$

It follows that the second term on the right hand side of (C.6) is $o_P(1)$. Consequently, $\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\bar{\psi})}{\partial\theta\partial\theta'}-\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\psi_0)}{\partial\theta\partial\theta'}=o_P(1)$.

Next we consider $-\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\bar{\psi})}{\partial\theta\partial\sigma_v^2}+\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\psi_0)}{\partial\theta\partial\sigma_v^2}$. This term is equal to

$$\begin{aligned}&\frac{1}{nT\tilde{\sigma}_v^4}\tilde{X}'\Omega(\bar{\delta})^{-1}u(\bar{\theta})-\frac{1}{nT\sigma_{v0}^4}\tilde{X}'\Omega_0^{-1}u \\ =\;&\left(\frac{1}{\bar{\sigma}_v^4}-\frac{1}{\sigma_{v0}^4}\right)\frac{\tilde{X}'\Omega(\bar{\delta})^{-1}u(\bar{\theta})}{nT}+\frac{1}{\sigma_{v0}^4}\frac{\tilde{X}'[\Omega(\bar{\delta})^{-1}-\Omega_0^{-1}]u(\bar{\theta})}{nT}+\frac{1}{\sigma_{v0}^4}\frac{\tilde{X}'\Omega_0^{-1}[u(\bar{\theta})-u]}{nT}.\end{aligned}$$

Using $u(\bar{\theta})=Y-\tilde{X}\bar{\theta}=u+\tilde{X}(\theta_0-\bar{\theta})$, we can readily show that $\frac{1}{nT}\tilde{X}'\Omega(\bar{\delta})^{-1}u(\bar{\theta})=O_P(1)$, which implies that the first term in the last expression is $o_P(1)$ by Theorem 4.1. The second term is $o_P(1)$ by arguments analogous to those used above. The third term is $\sigma_{v0}^{-4}(nT)^{-1}\tilde{X}'\Omega(\bar{\delta})^{-1}\tilde{X}(\theta_0-\bar{\theta})=O_P(1)||\theta_0-\bar{\theta}||=o_P(1)$. It follows that $\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\bar{\psi})}{\partial\theta\partial\sigma_v^2}-\frac{1}{nT}\frac{\partial^2\mathcal{L}^r(\psi_0)}{\partial\theta\partial\sigma_v^2}=o_P(1)$. Now, write

$$-\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}^r(\bar{\psi})}{\partial\sigma_v^2\partial\sigma_v^2}-\frac{\partial^2\mathcal{L}^r(\psi_0)}{\partial\sigma_v^2\partial\sigma_v^2}\right]=\left(\frac{1}{\bar{\sigma}_v^6}u(\bar{\theta})'\Omega(\bar{\delta})^{-1}u(\bar{\theta})-\frac{1}{\sigma_v^2}u'\Omega_0^{-1}u\right)+\frac{1}{2}\left(\frac{1}{\sigma_{v0}^4}-\frac{1}{\bar{\sigma}_v^4}\right).$$

Clearly, the second term is $o_P(1)$ by Theorem 4.1. We can use the decomposition $u(\bar{\theta})=u+\tilde{X}(\theta_0-\bar{\theta})$ and the consistency of $\bar{\psi}$ to show the first term is also $o_P(1)$. This completes the proof. ■

Proof of Theorem 4.3

Let $T_1=T+1$. As in the proof of Theorem 4.1, we prove the theorem by showing that (i) $\frac{1}{nT_1}[\mathcal{L}_c^{rr*}(\delta)-\mathcal{L}_c^{rr}(\delta)]\xrightarrow{p}0$ uniformly in $\delta\in\mathbf{\Delta}$, and (ii) $\limsup_{n\to\infty}\max_{\rho\in N_\epsilon^c(\rho_0)}\frac{1}{nT_1}[\mathcal{L}_c^{rr*}(\delta)-\mathcal{L}_c^{rr*}(\delta_0)]<0$ for any $\epsilon>0$. The proof of (ii) is almost identical to that of (ii) in the proof of Theorem 4.1 and thus omitted.

By (3.14) and (4.6), $\frac{2}{nT_1}[\mathcal{L}_c^{rr*}(\delta)-\mathcal{L}_c^{rr}(\delta)]=\ln\hat{\sigma}_v^2(\delta)-\ln\tilde{\sigma}_v^2(\delta)$. To show (i), it suffices to show

$$\hat{\sigma}_v^2(\delta)-\tilde{\sigma}_v^2(\delta)=o_P(1)\text{ uniformly on }\mathbf{\Delta}. \tag{C.11}$$

By the definition of $\tilde{u}^*(\delta)$ below (3.13), we have $\tilde{u}^*(\delta) = Y^*(\rho) - X^*(X^{*\prime}\Omega^{*-1}X^*)^{-1}X^{*\prime}\Omega^{*-1}Y^*(\rho) = \Omega^{*1/2}M^*\Omega^{*-1/2}Y^*(\rho)$ where $M^* = I_{nT_1} - \Omega^{*-1/2}X^*(X^{*\prime}\Omega^{*-1}X^*)^{-1}X^{*\prime}\Omega^{*-1/2}$ is a projection matrix. Observe that $Y^*(\rho) = Y^*(\rho_0) + [Y^*(\rho) - Y^*(\rho_0)] = X^*\theta_0 + u^* + (\rho_0 - \rho)Y^*_{-1}$ where $Y^*_{-1} = (0_{1\times n}, Y'_{-1})'$. This, in conjunction with the fact that $M^*\Omega^{*-1/2}X^* = 0$, implies that

$$\begin{aligned}\hat{\sigma}^2_v(\delta) &= \frac{1}{nT_1}\tilde{u}^*(\delta)'\Omega^{*-1}\tilde{u}^*(\delta) = \frac{1}{nT_1}Y^*(\rho)'\,\Omega^{*-1/2}M^*\Omega^{*-1/2}Y^*(\rho)\\ &= \frac{1}{nT_1}\left[u^* + (\rho_0-\rho)Y^*_{-1}\right]'\Omega^{*-1/2}M^*\Omega^{*-1/2}\left[u^* + (\rho_0-\rho)Y^*_{-1}\right]. \qquad \text{(C.12)}\end{aligned}$$

By (4.4) and the above expression for $Y^*(\rho)$, we have

$$\tilde{\theta}(\delta) = \left[E\left(X^{*\prime}\Omega^{*-1}X^*\right)\right]^{-1}E\left[X^{*\prime}\Omega^{*-1}Y^*(\rho)\right] = \theta_0 - \theta^*(\delta),$$

where $\theta^*(\delta) = (\rho - \rho_0)\left[E\left(X^{*\prime}\Omega^{*-1}X^*\right)\right]^{-1}E\left(X^{*\prime}\Omega^{*-1}Y^*_{-1}\right)$. Then by the definition of $u^*(\theta,\rho)$ after (3.12),

$$u^*(\tilde{\theta}(\delta),\rho) = Y^*(\rho) - X^*\tilde{\theta}(\delta) = X^*\theta^*(\delta) + u^* + (\rho_0-\rho)Y^*_{-1}.$$

By (4.5),

$$\begin{aligned}\tilde{\sigma}^2_v(\delta) &= \frac{1}{nT_1}E\left\{\left[X^*\theta^*(\delta) + u^* + (\rho_0-\rho)Y^*_{-1}\right]'\Omega^{*-1}\left[X^*\theta^*(\delta) + u^* + (\rho_0-\rho)Y^*_{-1}\right]\right\}\\ &= \frac{1}{nT_1}E\left\{\left[u^* + (\rho_0-\rho)Y^*_{-1}\right]'\Omega^{*-1}\left[u^* + (\rho_0-\rho)Y^*_{-1}\right]\right\}\\ &\quad + \frac{1}{nT_1}\theta^*(\delta)'E\left(X^{*\prime}\Omega^{*-1}X^*\right)\theta^*(\delta) + \frac{2(\rho_0-\rho)}{nT_1}\theta^*(\delta)'E\left(X^{*\prime}\Omega^{*-1}Y^*_{-1}\right)\\ &= \frac{1}{nT_1}E\left\{\left[u^* + (\rho_0-\rho)Y^*_{-1}\right]'\Omega^{*-1}\left[u^* + (\rho_0-\rho)Y^*_{-1}\right]\right\}\\ &\quad + \frac{(\rho_0-\rho)}{nT_1}\theta^*(\delta)'E\left(X^{*\prime}\Omega^{*-1}Y^*_{-1}\right). \qquad \text{(C.13)}\end{aligned}$$

Using (C.12)-(C.13) and $\Omega^{*-1/2}M^*\Omega^{*-1/2} = \Omega^{*-1} - \Omega^{*-1}X^*(X^{*\prime}\Omega^{*-1}X^*)^{-1}X^{*\prime}\Omega^{*-1}$, we have

$$\begin{aligned}&\hat{\sigma}^2_v(\delta) - \tilde{\sigma}^2_v(\delta)\\ =\ &\frac{1}{nT_1}\left[u^* + (\rho_0-\rho)Y^*_{-1}\right]'\Omega^{*-1/2}M^*\Omega^{*-1/2}\left[u^* + (\rho_0-\rho)Y\right]^*_{-1} - \tilde{\sigma}^2_v(\delta)\\ =\ &\frac{1}{nT_1}\left\{\left[u^* + (\rho_0-\rho)Y^*_{-1}\right]'\Omega^{*-1}\left[u^* + (\rho_0-\rho)Y^*_{-1}\right] - E\left[u^* + (\rho_0-\rho)Y^*_{-1}\right]'\Omega^{*-1}\left[u^* + (\rho_0-\rho)Y^*_{-1}\right]\right\}\\ &+Q^*_{xu}(\delta)'Q^*_{xx}(\delta)^{-1}Q^*_{xu}(\delta)' + 2(\rho_0-\rho)Q^*_{xu}(\delta)'Q^*_{xx}(\delta)^{-1}Q^*_{xy_{-1}}(\delta)\\ &+(\rho_0-\rho)^2\left\{Q^*_{xy_{-1}}(\delta)'Q^*_{xx}(\delta)^{-1}Q^*_{xy_{-1}}(\delta) - E[Q^*_{xy_{-1}}(\delta)']\left\{E\left[Q^*_{xx}(\delta)\right]\right\}^{-1}E[Q^*_{xy_{-1}}(\delta)]\right\}\\ \equiv\ &\Pi^*_{n1}(\delta) + \Pi^*_{n2}(\delta) + 2(\rho_0-\rho)\Pi^*_{n3}(\delta) + (\rho_0-\rho)^2\Pi^*_{n4}(\delta),\ \text{say,}\end{aligned}$$

where $Q^*_{xx}(\delta) = \frac{1}{nT_1}X^{*\prime}\Omega^{*-1}X^*$, $Q^*_{xu}(\delta) = \frac{1}{nT_1}X^{*\prime}\Omega^{*-1}u^*$, and $Q^*_{xy_{-1}}(\delta) = \frac{1}{nT_1}X^{*\prime}\Omega^{*-1}Y^*_{-1}$. We prove (i) by showing that $\Pi^*_{ns}(\delta) = o_P(1)$ uniformly in δ for $s = 1, 2, 3$, and 4.

We can decompose $\Pi^*_{n1}(\delta)$ as follows

$$\begin{aligned}\Pi^*_{n1}(\delta) &= \frac{1}{nT_1}\left[u^{*\prime}\Omega^{*-1}u^* - E\left(u^{*\prime}\Omega^{*-1}u^*\right)\right] + \frac{(\rho_0-\rho)^2}{nT_1}\left[Y^{*\prime}_{-1}\Omega^{*-1}Y^*_{-1} - E\left(Y^{*\prime}_{-1}\Omega^{*-1}Y^*_{-1}\right)\right]\\ &\quad + \frac{2(\rho_0-\rho)}{nT_1}\left[u^{*\prime}\Omega^{*-1}Y^*_{-1} - E\left(u^{*\prime}\Omega^{*-1}Y^*_{-1}\right)\right]\\ &\equiv \Pi^*_{n1,1}(\delta) + \Pi^*_{n1,2}(\delta) + \Pi^*_{n1,3}(\delta),\ \text{say.}\end{aligned}$$

For $\Pi^*_{n1,1}(\delta)$, we can show that $E[\Pi^*_{n1,1}(\delta)] = 0$ and $E[\Pi^*_{n1,1}(\delta)]^2 = O(n^{-1})$ as in the proof of Lemma B.7. So the pointwise convergence of $\Pi^*_{n1,1}(\delta)$ to 0 follows by Chebyshev inequality. The uniform convergence holds if we can show that $\Pi^*_{n1,1}(\delta)$ is stochastic equicontinuous. Let $\delta,\ \bar{\delta} \in \mathbf{\Delta}$. By Cauchy-Schwarz inequality,

$$\begin{aligned}
|\Pi^*_{n1,1}(\delta) - \Pi^*_{n1,1}(\bar{\delta})| &= \left| \frac{1}{nT_1} \operatorname{tr} \left\{ \Omega^*(\delta)^{-1} \left[\Omega^*(\bar{\delta}) - \Omega^*(\delta) \right] \Omega^*(\bar{\delta})^{-1} \left[u^* u^{*\prime} - E(u^* u^{*\prime}) \right] \right\} \right| \\
&\leq \frac{1}{nT_1} [\operatorname{tr}\{\Omega^*(\delta)^{-1} \left[\Omega^*(\bar{\delta}) - \Omega^*(\delta) \right] \Omega^*(\bar{\delta})^{-2} \left[\Omega^*(\bar{\delta}) - \Omega^*(\delta) \right] \Omega^*(\delta)^{-1}\}]^{1/2} \\
&\quad \times \| u^* u^{*\prime} - E(u^* u^{*\prime}) \| \\
&\leq [\lambda_{\min}(\Omega^*(\bar{\delta}))]^{-2} \frac{1}{\sqrt{nT_1}} \left\| \Omega^*(\bar{\delta}) - \Omega^*(\delta) \right\| \frac{1}{\sqrt{nT_1}} \| u^* u^{*\prime} - E(u^* u^{*\prime}) \| .
\end{aligned}$$

Straightforward moment calculations and Chebyshev inequality lead to $\| u^* u^{*\prime} - E(u^* u^{*\prime}) \| / \sqrt{nT_1} = O_P(1)$. In addition, $\left\| \Omega^*(\bar{\delta}) - \Omega^*(\delta) \right\| / \sqrt{nT_1} \to 0$ as $\left\| \delta - \bar{\delta} \right\| \to 0$. Thus, $\{\Pi^*_{n1,1}(\delta)\}$ is stochastically equicontinuous by Theorem 21.10 in Davidson (1994). Consequently, $\Pi^*_{n1,1}(\delta) = o_P(1)$ uniformly in δ. By the same token, $\Pi^*_{n1,s}(\delta) = o_P(1)$ uniformly in δ for $s = 2, 3$. It follows that $\Pi^*_{n1}(\delta) = o_P(1)$ uniformly in δ.

To show $\Pi^*_{n2}(\delta) = o_P(1)$ uniformly in δ, we first argue that $\Omega^*(\delta)$ is positive definite uniformly in δ, i.e., $\inf_{\delta \in \mathbf{\Delta}} \lambda_{\min}(\Omega^*(\delta)) \geq \underline{c}^*$ for some $\underline{c}^* > 0$. Let $\bar{u}^* = (a_m \mu', u')'$ and $\bar{\Omega}^*(\delta) = \begin{pmatrix} \phi_\mu a_m^2 I_n & \phi_\mu a_m (\iota_T' \otimes I_n) \\ \phi_\mu a_m (\iota_T \otimes I_n) & \Omega \end{pmatrix}$. Noting that $\bar{\Omega}^*(\delta) = E(\bar{u}^* \bar{u}^{*\prime})$, it is positive semidefinite uniformly in δ. By Theorem 8.4.11 in Bernstein (2005) and (C.4), $\lambda_{\min}\left(\phi_\zeta I_n + b_m (B'B)^{-1}\right) \geq \phi_\zeta + b_m \lambda_{\min}\left((B'B)^{-1}\right) \geq \phi_\zeta + b_m \bar{c}_\lambda^{-2} > 0$ uniformly in δ as ϕ_ζ is positive and bounded away from 0 and $b_m > 0$, implying that $\phi_\zeta I_n + b_m (B'B)^{-1}$ is positive definite uniformly in δ. Noting $\Omega^*(\delta)$ is equal to $\bar{\Omega}^*(\delta)$ with its upper-left (n, n)-submatrix added by a uniformly positive definite matrix $\phi_\zeta I_n + b_m (B'B)^{-1}$, we can apply Fact 8.9.19 in Bernstein (2005) to conclude that $\Omega^*(\delta)$ is positive definite uniformly in δ. Similarly, we can readily show that

$$\begin{aligned}
\sup_{\delta \in \mathbf{\Delta}} \lambda_{\max}(\Omega^*(\delta)) &\leq \sup_{\delta \in \mathbf{\Delta}} \lambda_{\max}(\bar{\Omega}^*(\delta)) + \sup_{\delta \in \mathbf{\Delta}} \lambda_{\max}(\phi_\zeta I_n + b_m (B'B)^{-1})) \\
&\leq \sup_{\delta \in \mathbf{\Delta}} \lambda_{\max}(\bar{\Omega}^*(\delta)) + \sup_{\delta \in \mathbf{\Delta}} \left(\phi_\zeta + b_m \lambda_{\min}\left((B'B)^{-1}\right) \right) \\
&\leq \bar{c}^*, \text{ say.}
\end{aligned}$$

Next, write

$$\begin{aligned}
\frac{1}{nT_1} X^{*\prime} X^* &= \frac{1}{nT_1} \begin{pmatrix} X'X & X'Z & 0_{p \times k} \\ Z'X & Z'Z & 0_{q \times k} \\ \tilde{\mathbf{x}}' x_0 & \tilde{\mathbf{x}}' z_m(\rho) & \tilde{\mathbf{x}}' \tilde{\mathbf{x}} \end{pmatrix} + \frac{1}{nT_1} \begin{pmatrix} x_0' x_0 & x_0' z_m(\rho) & x_0' \tilde{\mathbf{x}} \\ z_m(\rho)' x_0 & z_m(\rho)' z_m(\rho) & z_m(\rho) \tilde{\mathbf{x}} \\ 0_{k \times p} & 0_{k \times q} & 0_{k \times k} \end{pmatrix} \\
&\equiv A_1(\rho) + A_2(\rho), \text{ say.}
\end{aligned}$$

Noting that $A_1(\rho)$ is a block triangular matrix. Its eigenvalues are given by those of $\frac{1}{nT_1} \begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z \end{pmatrix}$ and those of $\frac{1}{nT_1} \tilde{\mathbf{x}}' \tilde{\mathbf{x}}$. By Assumption R* (iv), the minimum of these eigenvalues are bounded away from 0, say by $\underline{c}_{xx}$, uniformly in ρ. Similarly, the minimum eigenvalues of $A_2(\rho)$ is 0 uniformly in ρ. It follows that $\inf_\rho \lambda_{\min}\left(\frac{1}{nT_1}X^{*\prime}X^*\right) \geq \inf_\rho\left[\lambda_{\min}\left(A_1(\rho)\right)+\lambda_{\min}\left(A_2(\rho)\right)\right] \geq \underline{c}_{xx} > 0$. Consequently,

$$\inf_{\delta\in\mathbf{\Delta}} \lambda_{\min}\left(Q^*_{xx}(\delta)\right) = \inf_{\delta\in\mathbf{\Delta}} \lambda_{\min}\left(\frac{1}{nT_1}X^{*\prime}\Omega^{*-1}X^*\right) \geq \bar{c}^{*-1}\inf_\rho \lambda_{\min}\left(\frac{1}{nT_1}X^{*\prime}X^*\right) \geq \bar{c}^{*-1}\underline{c}_{xx} > 0. \quad \text{(C.14)}$$

Next, noting that $\mathrm{E}[Q^*_{xu}(\delta)] = 0$ and $\mathrm{Var}(Q^*_{xu}(\delta)) = O(n^{-1})$, we have $Q^*_{xu}(\delta) = o_P(1)$ by Chebyshev inequality. In addition, it is straightforward to show that $Q^*_{xu}(\delta)$ is stochastic equicontinuous. So $Q^*_{xu}(\delta) = o_P(1)$ uniformly in δ. We have

$$|\Pi^*_{n2}(\delta)| \leq \left[\inf_{\delta\in\mathbf{\Delta}} \lambda_{\min}\left(Q^*_{xx}(\delta)\right)\right]^{-1} \left\|Q^*_{xu}(\delta)\right\|^2 = o_P(1) \text{ uniformly in } \delta.$$

For $\Pi^*_{n3}(\delta)$, we have $\Pi^*_{n3}(\delta) \leq \left\|Q^*_{xu}(\delta)\right\| \left\|Q^*_{xx}(\delta)^{-1}\right\| \left\|Q^*_{xy_{-1}}(\delta)\right\| = o_P(1)$ uniformly in δ as one can readily show that $Q^*_{xy_{-1}}(\delta) = O_P(1)$ uniformly in δ.

For $\Pi^*_{n4}(\delta)$, we have

$$\begin{aligned}
\Pi^*_{n4}(\delta) &= \left\{Q^*_{xy_{-1}}(\delta) - E[Q^*_{xy_{-1}}(\delta)]\right\}' Q^*_{xx}(\delta)^{-1} Q^*_{xy_{-1}}(\delta) \\
&\quad + E[Q^*_{xy_{-1}}(\delta)]' Q^*_{xx}(\delta)^{-1}\left\{E[Q^*_{xx}(\delta)] - Q^*_{xx}(\delta)\right\}\left\{E[Q^*_{xx}(\delta)]\right\}^{-1} Q^*_{xy_{-1}}(\delta) \\
&\quad + E[Q^*_{xy_{-1}}(\delta)]' E\left[Q^*_{xx}(\delta)\right]\left\{Q^*_{xy_{-1}}(\delta) - E[Q^*_{xy_{-1}}(\delta)]\right\} \\
&\equiv \Pi^*_{n4,1}(\delta) + \Pi^*_{n4,2}(\delta) + \Pi^*_{n4,3}(\delta), \text{ say.}
\end{aligned}$$

We can readily show that $Q^*_{xy_{-1}}(\delta) - E[Q^*_{xy_{-1}}(\delta)] = o_P(1)$ uniformly in δ by Chebyshev inequality and the arguments of stochastic equicontinuity. This, in conjunction with (C.14) and the fact that $Q^*_{xy_{-1}}(\delta) = O_P(1)$ uniformly in δ, implies that $\Pi^*_{n4,1}(\delta) = o_P(1)$ uniformly in δ. Similarly, we can show that $\Pi^*_{n4,s}(\delta) = o_P(1)$ uniformly in δ for $s = 2, 3$. Thus $\Pi^*_{n4}(\delta) = o_P(1)$ uniformly in δ. This completes the proof of (i). ■

Proof of Theorem 4.4

The proof is analogous to that of Theorem 4.2, but follows mainly from Lemmas B.12-B.13. ■

Proof of Theorem 4.5

The proof is almost identical to that of Theorem 4.3 and thus omitted. ■

Proof of Theorem 4.6

The proof is analogous to that of Theorem 4.2, but follows mainly from Lemmas B.14-B.15. ■

REFERENCES

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

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

Anselin, L., Le Gallo, J., Jayet, J., 2008. Spatial panel econometrics. In L. Mátyás & P. Sevestre (eds), *Fundamentals and Recent Developments in Theory and Practice*, pp. 625-660. Springer, Berlin Heidelberg.

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

Baltagi, B. H., Li, D., 2004. Prediction in the panel data models with spatial correlation. In L. Anselin, R. Florax, and S. J. Rey (eds), *Advances in Spatial Econometrics*, pp. 283-296. Springer-Verlag: Berlin.

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

Baltagi, B. H., Egger, P., Pfaffermayr, M., 2007a. A generalized spatial panel model with random effects. *Working Paper*, Syracuse University.

Baltagi, B. H., Song, S. K., Jung, B. C., Koh, W., 2007b. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* 140, 5-51.

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

Bernstein, D. S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory.* Princeton University Press, Princeton.

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.

Blundell, R., Smith, R. J., 1991. Conditions initiales et estimation efficace dans les modèles dynamiques sur données de panel. *Annales d'Économie et de Statistique* 20/21, 109-123.

Chen, X., Conley,T. G., 2001. A new semiparametric spatial model for panel time series. *Journal of Econometrics* 105, 59-83.

Cliff, A. D., Ord, J. K., 1973. *Spatial Autocorrelation.* London: Pion Ltd.

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

Elhorst, J. P., 2003. Specification and estimation of spatial panel data models. International *Regional Science Review* 26, 244-268.

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

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

Horn, R., Johnson, C., 1985. *Matrix Analysis.* New York. John Wiley & Sons.

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.

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

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

Kelejian, H. H., Prucha, I. R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics* 17, 377-398.

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

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

Kelejian, H. H., Prucha, I. R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53-67.

Lee, L. F., 2002. Asymptotic Distribution of Quasi-maximum Likelihood Estimators for Spatial Autoregressive Models, *Mimeo,* Dept. of Economics, Ohio State University.

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

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

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

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

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

Magnus, J. R., 1982. Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. *Journal of Econometrics* 19, 239-285.

Magnus, J. R., Neudecker, H., 2002 *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons.

Moon, H. & M. Weidner (2010) Dynamic linear panel regression models with interactive fixed effects. Manuscript. *Working paper*, University of Southern California.

Mutl, J., 2006. Dynamic panel data models with spatially correlated disturbances. Ph.D. Thesis, University of Maryland, College Park.

Mutl, J., Pfaffermayr, M., 2008. The spatial random effects and the spatial fixed effects model: the Hausman test in a Cliff and Ord panel model. Manuscript, Institute for Advanced Studies, Vienna.

Nerlove, M., 2002. Likelihood inference for dynamic panel models. In *Essays in Panel Data Econometrics*, M. Nerlove (eds), pp. 307-348. Cambridge University Press.

Ord, J. K., 1975. Estimation Methods for Models of Spatial Interaction, *Journal of the American Statistical Association* 70, 120-126.

Pesaran, M. H., 2004. General diagnostic tests for cross section dependence in panels. *Working Paper No. 1229*, University of Cambridge.

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

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

Phillips, P. C. B., Sul, D., 2003. Dynamic panel estimation and homogeneity testing under cross-section dependence. *Econometrics Journal* 6, 217-259.

Ridder, G., Wansbeek, T., 1990. Dynamic models for panel data. In F. van der Ploeg (ed.), *Advanced Lectures in Quantitative Economics*, pp. 557-582. London: Academic Press.

Robinson, P. M., 2010. Efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics* 157, 6-17.

Smirnov, O., Anselin, L., 2001. Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. *Computational Statistics and Data Analysis* 35, 301-319.

Su, L., 2012. A semiparametric GMM estimation of spatial autoregressive models. *Journal of Econometrics* 167, 543-560.

Su, L., Jin, S., 2010. Profile quasi-maximum likelihood estimation of spatial autoregressive models. *Journal of Econometrics* 157, 18-33.

Su, L., Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics* 169, 34-47.

Su, L., Yang, Z., 2007. QML estimation of dynamic panel data models with spatial errors. *Working Paper*, Singapore Management University.

White, H., 1994. *Estimation, Inference and Specification Analysis.* Econometric Society Monographs No. 22, Cambridge University Press.

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

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

Table 1a. Monte Carlo Mean[RMSE] for the QMLEs, Random Effects Model with Normal Errors

	true $m=0$			true $m=6$		
ψ	$m=0$	$m=6$	$m=200$	$m=0$	$m=6$	$m=200$
	$n=50, T=3$					
5.0	5.0266[0.334]	4.9604[0.338]	5.0030[0.328]	4.5591[0.378]	4.9940[0.411]	5.0988[0.411]
1.0	1.0011[0.040]	0.9917[0.045]	0.9981[0.045]	0.9626[0.041]	0.9980[0.040]	1.0057[0.039]
1.0	0.9951[0.345]	0.9852[0.350]	0.9927[0.352]	0.7418[0.365]	0.9384[0.391]	0.9790[0.395]
0.8	0.7991[0.023]	0.8071[0.024]	0.8018[0.022]	0.8238[0.015]	0.8015[0.017]	0.7963[0.016]
0.5	0.4827[0.099]	0.3023[0.115]	0.2868[0.114]	0.4732[0.101]	0.4886[0.098]	0.4868[0.098]
1.0	0.9681[0.147]	0.1469[0.116]	0.0214[0.055]	0.8648[0.145]	0.9528[0.158]	0.9280[0.161]
1.0	0.9834[0.072]	1.2563[0.087]	1.2805[0.088]	1.0056[0.076]	0.9880[0.073]	1.0019[0.076]
5.0	4.9785[0.357]	4.9683[0.400]	4.9719[0.400]	4.7922[0.353]	5.0164[0.352]	5.0162[0.352]
1.0	1.0003[0.040]	0.9964[0.045]	0.9967[0.045]	0.9780[0.041]	0.9981[0.039]	0.9981[0.039]
1.0	0.9937[0.323]	1.0022[0.328]	1.0028[0.328]	0.8910[0.352]	0.9374[0.360]	0.9370[0.361]
0.4	0.4015[0.034]	0.4025[0.044]	0.4019[0.044]	0.4271[0.032]	0.4009[0.030]	0.4009[0.030]
0.5	0.4799[0.103]	0.3694[0.141]	0.3690[0.142]	0.4765[0.104]	0.4912[0.093]	0.4911[0.093]
1.0	0.9609[0.146]	0.6380[0.229]	0.6364[0.231]	0.9141[0.155]	0.9725[0.148]	0.9712[0.149]
1.0	0.9838[0.074]	1.1272[0.137]	1.1280[0.138]	1.0056[0.080]	0.9960[0.074]	0.9964[0.074]
5.0	5.0096[0.337]	4.9719[0.352]	4.9719[0.352]	4.9061[0.328]	5.0103[0.328]	5.0103[0.328]
1.0	0.9987[0.040]	0.9947[0.042]	0.9947[0.042]	0.9872[0.040]	0.9991[0.039]	0.9991[0.039]
1.0	0.9944[0.336]	0.9805[0.337]	0.9805[0.337]	0.9481[0.356]	0.9897[0.361]	0.9897[0.361]
0.0	-0.0014[0.041]	0.0069[0.047]	0.0069[0.047]	0.0199[0.043]	-0.0021[0.042]	-0.0021[0.042]
0.5	0.4783[0.106]	0.3977[0.114]	0.3977[0.114]	0.4815[0.102]	0.4929[0.091]	0.4929[0.091]
1.0	0.9659[0.151]	0.7313[0.178]	0.7313[0.178]	0.9342[0.157]	0.9691[0.148]	0.9691[0.148]
1.0	0.9808[0.076]	1.0741[0.102]	1.0741[0.102]	0.9945[0.079]	0.9624[0.066]	0.9624[0.066]
	$n=100, T=3$					
5.0	4.9921[0.252]	4.9129[0.258]	4.9423[0.248]	4.5604[0.270]	5.0174[0.299]	5.1460[0.300]
1.0	0.9995[0.029]	0.9892[0.034]	0.9932[0.033]	0.9655[0.029]	0.9997[0.029]	1.0090[0.029]
1.0	1.0019[0.243]	0.9822[0.242]	0.9916[0.242]	0.9112[0.227]	1.0126[0.240]	1.0414[0.244]
0.8	0.8003[0.017]	0.8092[0.018]	0.8058[0.016]	0.8200[0.009]	0.7993[0.010]	0.7935[0.010]
0.5	0.4852[0.074]	0.2674[0.086]	0.2500[0.085]	0.4857[0.068]	0.4872[0.067]	0.4865[0.067]
1.0	0.9788[0.101]	0.1828[0.094]	0.0279[0.056]	0.9083[0.101]	0.9806[0.115]	0.9719[0.120]
1.0	0.9941[0.052]	1.2885[0.062]	1.3150[0.060]	1.0075[0.053]	0.9940[0.052]	1.0025[0.053]
5.0	4.9941[0.247]	4.9271[0.305]	4.9318[0.306]	4.7258[0.277]	4.9982[0.273]	4.9982[0.273]
1.0	0.9991[0.031]	0.9899[0.040]	0.9904[0.040]	0.9730[0.031]	1.0012[0.030]	1.0012[0.030]
1.0	1.0055[0.242]	0.9888[0.245]	0.9897[0.245]	0.9384[0.240]	1.0127[0.250]	1.0128[0.250]
0.4	0.4004[0.025]	0.4104[0.037]	0.4098[0.037]	0.4316[0.023]	0.3996[0.022]	0.3996[0.022]
0.5	0.4916[0.069]	0.3706[0.099]	0.3701[0.100]	0.4885[0.074]	0.4859[0.069]	0.4858[0.069]
1.0	0.9885[0.103]	0.6050[0.175]	0.6033[0.177]	0.9141[0.104]	0.9808[0.101]	0.9798[0.101]
1.0	0.9926[0.053]	1.1742[0.118]	1.1752[0.118]	1.0120[0.054]	0.9948[0.051]	0.9951[0.051]
5.0	5.0098[0.265]	5.0200[0.271]	5.0200[0.271]	4.8775[0.257]	5.0054[0.254]	5.0054[0.254]
1.0	1.0011[0.032]	1.0023[0.033]	1.0023[0.033]	0.9845[0.032]	0.9997[0.030]	0.9997[0.030]
1.0	0.9923[0.232]	0.9930[0.233]	0.9930[0.233]	0.9819[0.240]	1.0086[0.244]	1.0086[0.244]
0.0	0.0000[0.031]	-0.0021[0.033]	-0.0021[0.033]	0.0236[0.033]	-0.0010[0.031]	-0.0010[0.031]
0.5	0.4860[0.069]	0.4257[0.073]	0.4258[0.073]	0.4866[0.072]	0.4942[0.063]	0.4942[0.063]
1.0	0.9771[0.107]	0.8260[0.117]	0.8261[0.117]	0.9505[0.109]	0.9851[0.101]	0.9851[0.101]
1.0	0.9957[0.054]	1.0535[0.068]	1.0535[0.068]	1.0015[0.054]	0.9778[0.045]	0.9778[0.045]

Note: $\psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \sigma_\mu, \sigma_v)'$. Parameters values for generating x_t: $\theta_x = (.01, .5, .5, 2, 1)$ (see Footnote 7).

Table 1b. Monte Carlo Mean[RMSE] for the QMLEs, Random Effects Model with Normal Mixture

	true $m = 0$			true $m = 6$		
ψ	$m = 0$	$m = 6$	$m = 200$	$m = 0$	$m = 6$	$m = 200$
	$n = 50, T = 3$					
5.0	5.0194[0.342]	4.9734[0.350]	5.0140[0.340]	4.5754[0.416]	4.9935[0.429]	5.0941[0.430]
1.0	1.0005[0.039]	0.9948[0.047]	1.0006[0.047]	0.9656[0.041]	0.9984[0.039]	1.0057[0.039]
1.0	0.9874[0.335]	0.9778[0.339]	0.9858[0.340]	0.7650[0.383]	0.9558[0.405]	0.9981[0.410]
0.8	0.7992[0.022]	0.8047[0.024]	0.7998[0.022]	0.8225[0.017]	0.8011[0.016]	0.7960[0.016]
0.5	0.4788[0.100]	0.2652[0.130]	0.2489[0.129]	0.4766[0.099]	0.4916[0.097]	0.4902[0.096]
1.0	0.9544[0.249]	0.1551[0.120]	0.0283[0.061]	0.8470[0.228]	0.9330[0.259]	0.9101[0.260]
1.0	0.9792[0.145]	1.2519[0.163]	1.2776[0.167]	0.9984[0.147]	0.9821[0.143]	0.9954[0.147]
5.0	4.9914[0.340]	4.9151[0.373]	4.9190[0.374]	4.8085[0.368]	5.0216[0.361]	5.0215[0.361]
1.0	0.9990[0.042]	0.9887[0.047]	0.9891[0.047]	0.9814[0.040]	1.0002[0.038]	1.0002[0.038]
1.0	1.0152[0.332]	1.0061[0.333]	1.0067[0.333]	0.8921[0.357]	0.9384[0.361]	0.9381[0.361]
0.4	0.4003[0.033]	0.4120[0.041]	0.4114[0.041]	0.4265[0.033]	0.4016[0.030]	0.4016[0.030]
0.5	0.4784[0.099]	0.3775[0.115]	0.3770[0.116]	0.4804[0.097]	0.4914[0.090]	0.4913[0.090]
1.0	0.9488[0.256]	0.5328[0.299]	0.5307[0.302]	0.8779[0.250]	0.9387[0.249]	0.9375[0.249]
1.0	0.9799[0.144]	1.1476[0.183]	1.1485[0.184]	0.9895[0.148]	0.9770[0.138]	0.9774[0.138]
5.0	5.0179[0.343]	5.0602[0.344]	5.0602[0.344]	4.9083[0.343]	5.0085[0.339]	5.0085[0.339]
1.0	0.9990[0.044]	1.0016[0.044]	1.0016[0.044]	0.9884[0.040]	1.0000[0.038]	1.0000[0.038]
1.0	0.9981[0.343]	1.0043[0.344]	1.0043[0.344]	0.9497[0.346]	0.9928[0.349]	0.9929[0.349]
0.0	-0.0009[0.043]	-0.0094[0.043]	-0.0094[0.043]	0.0197[0.045]	-0.0017[0.042]	-0.0017[0.042]
0.5	0.4822[0.097]	0.4484[0.096]	0.4484[0.096]	0.4808[0.100]	0.4926[0.089]	0.4926[0.089]
1.0	0.9469[0.259]	0.8501[0.259]	0.8500[0.259]	0.9081[0.247]	0.9435[0.246]	0.9434[0.246]
1.0	0.9784[0.144]	1.0170[0.162]	1.0170[0.162]	0.9871[0.145]	0.9475[0.124]	0.9475[0.124]
	$n = 100, T = 3$					
5.0	4.9975[0.265]	4.9224[0.276]	4.9695[0.262]	4.6100[0.278]	5.0438[0.335]	5.1446[0.290]
1.0	1.0003[0.029]	0.9916[0.034]	0.9974[0.033]	0.9662[0.029]	1.0024[0.029]	1.0118[0.029]
1.0	1.0089[0.239]	0.9960[0.239]	1.0040[0.240]	0.9023[0.226]	0.9941[0.242]	1.0155[0.245]
0.8	0.8005[0.017]	0.8086[0.019]	0.8035[0.016]	0.8197[0.010]	0.7981[0.013]	0.7931[0.010]
0.5	0.4880[0.072]	0.2826[0.083]	0.2658[0.084]	0.4787[0.072]	0.4749[0.072]	0.4735[0.072]
1.0	0.9621[0.180]	0.1625[0.098]	0.0201[0.048]	0.8933[0.157]	0.9873[0.248]	0.9648[0.190]
1.0	0.9945[0.107]	1.2741[0.115]	1.2990[0.118]	1.0052[0.107]	0.9896[0.104]	0.9969[0.105]
5.0	4.9962[0.258]	4.8481[0.297]	4.8535[0.298]	4.7778[0.262]	5.0177[0.259]	5.0181[0.259]
1.0	1.0009[0.031]	0.9813[0.038]	0.9820[0.038]	0.9755[0.032]	1.0003[0.030]	1.0003[0.030]
1.0	1.0026[0.239]	0.9616[0.240]	0.9630[0.240]	0.9453[0.225]	0.9933[0.231]	0.9934[0.231]
0.4	0.4002[0.026]	0.4229[0.034]	0.4221[0.035]	0.4277[0.023]	0.3989[0.022]	0.3989[0.022]
0.5	0.4878[0.073]	0.3309[0.089]	0.3308[0.090]	0.4867[0.072]	0.4825[0.069]	0.4824[0.069]
1.0	0.9746[0.183]	0.4723[0.195]	0.4706[0.197]	0.9108[0.178]	0.9695[0.188]	0.9687[0.188]
1.0	0.9943[0.103]	1.1997[0.125]	1.2001[0.126]	1.0052[0.100]	0.9887[0.096]	0.9890[0.096]
5.0	4.9946[0.270]	5.0102[0.279]	5.0103[0.279]	4.9119[0.266]	5.0339[0.264]	5.0339[0.264]
1.0	0.9998[0.032]	0.9996[0.034]	0.9996[0.034]	0.9865[0.032]	1.0016[0.031]	1.0016[0.031]
1.0	1.0004[0.249]	0.9802[0.249]	0.9802[0.249]	0.9565[0.238]	0.9816[0.242]	0.9816[0.242]
0.0	0.0001[0.033]	-0.0008[0.036]	-0.0008[0.036]	0.0208[0.033]	-0.0032[0.031]	-0.0032[0.031]
0.5	0.4877[0.071]	0.4050[0.090]	0.4050[0.090]	0.4912[0.072]	0.5024[0.062]	0.5024[0.062]
1.0	0.9638[0.186]	0.8049[0.194]	0.8050[0.194]	0.9518[0.182]	0.9871[0.182]	0.9872[0.182]
1.0	0.9864[0.105]	1.0428[0.128]	1.0427[0.128]	0.9942[0.108]	0.9641[0.092]	0.9641[0.092]

Note: $\psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \sigma_\mu, \sigma_v)'$. Parameters values for generating x_t: $\theta_x = (.01, .5, .5, 2, 1)$ (see Footnote 7).

Table 2a. Monte Carlo Mean and SD, and Bootstrap Standard Errors, $m = 0$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
n	ψ			$T+1=4$					$T+1=8$		
		`Normal Errors`									
50	5.0	5.0155	0.3595	0.3257	0.3428	0.3759	5.0040	0.2736	0.2436	0.2695	0.3149
	1.0	1.0003	0.0422	0.0373	0.0403	0.0443	0.9999	0.0229	0.0203	0.0222	0.0246
	1.0	0.9949	0.3462	0.3321	0.3291	0.3288	0.9996	0.3017	0.2981	0.2978	0.2988
	0.5	0.4987	0.0332	0.0312	0.0321	0.0342	0.4995	0.0150	0.0140	0.0149	0.0162
	0.5	0.4775	0.1035	0.1037	0.1003	0.1104	0.4973	0.0608	0.0632	0.0588	0.0631
	1.0	0.9998	0.3622	0.3885	0.3543	0.3692	0.9734	0.2657	0.2727	0.2543	0.2621
	1.0	0.9775	0.1441	0.1416	0.1455	0.1686	0.9883	0.0822	0.0821	0.0837	0.0981
100	5.0	5.0021	0.2634	0.2421	0.2571	0.2797	5.0014	0.1860	0.1591	0.1806	0.2145
	1.0	1.0000	0.0287	0.0270	0.0285	0.0305	1.0000	0.0155	0.0148	0.0160	0.0175
	1.0	0.9949	0.2412	0.2360	0.2350	0.2351	1.0109	0.2168	0.2141	0.2161	0.2190
	0.5	0.5000	0.0223	0.0211	0.0216	0.0226	0.4999	0.0105	0.0098	0.0105	0.0113
	0.5	0.4896	0.0726	0.0750	0.0715	0.0766	0.4976	0.0398	0.0466	0.0425	0.0444
	1.0	1.0040	0.2540	0.2636	0.2495	0.2589	0.9866	0.1889	0.1871	0.1815	0.1885
	1.0	0.9899	0.1027	0.0964	0.1038	0.1227	0.9966	0.0602	0.0560	0.0596	0.0710
		`Normal Mixture Errors`									
50	5.0	5.0105	0.3450	0.3340	0.3389	0.3735	4.9986	0.2828	0.2555	0.2685	0.3100
	1.0	1.0005	0.0394	0.0368	0.0398	0.0441	1.0001	0.0208	0.0190	0.0205	0.0224
	1.0	0.9972	0.3300	0.3244	0.3215	0.3220	1.0029	0.3045	0.2977	0.2945	0.2928
	0.5	0.4997	0.0331	0.0308	0.0316	0.0345	0.4998	0.0159	0.0143	0.0149	0.0161
	0.5	0.4887	0.1011	0.0984	0.0985	0.1178	0.4928	0.0575	0.0584	0.0590	0.0719
	1.0	1.0376	0.6779	0.3104	0.3636	0.5621	1.0135	0.5932	0.1917	0.2625	0.4643
	1.0	0.9813	0.2916	0.0897	0.1464	0.2867	0.9964	0.1770	0.0413	0.0844	0.1923
100	5.0	5.0098	0.2541	0.2313	0.2420	0.2676	4.9899	0.1900	0.1671	0.1842	0.2175
	1.0	1.0002	0.0293	0.0272	0.0290	0.0316	0.9997	0.0154	0.0139	0.0151	0.0164
	1.0	0.9842	0.2397	0.2344	0.2310	0.2290	1.0070	0.2189	0.2115	0.2151	0.2197
	0.5	0.5004	0.0240	0.0208	0.0218	0.0236	0.5002	0.0106	0.0101	0.0106	0.0114
	0.5	0.4900	0.0696	0.0730	0.0713	0.0834	0.4972	0.0421	0.0440	0.0425	0.0502
	1.0	1.0239	0.4462	0.1898	0.2532	0.4188	1.0078	0.3683	0.1162	0.1850	0.3578
	1.0	0.9927	0.2081	0.0569	0.1042	0.2177	0.9901	0.1289	0.0265	0.0592	0.1416
		`Chi-Square Errors, df=5`									
50	5.0	4.9959	0.3544	0.3414	0.3420	0.3756	5.0178	0.3216	0.3135	0.3190	0.3535
	1.0	0.9994	0.0408	0.0373	0.0403	0.0443	1.0006	0.0236	0.0220	0.0231	0.0246
	1.0	0.9942	0.3366	0.3318	0.3287	0.3285	0.9943	0.3363	0.3330	0.3286	0.3258
	0.5	0.5017	0.0334	0.0307	0.0320	0.0350	0.4982	0.0154	0.0148	0.0153	0.0163
	0.5	0.4758	0.1012	0.1026	0.1005	0.1133	0.4959	0.0582	0.0615	0.0588	0.0651
	1.0	1.0195	0.4533	0.3659	0.3601	0.4293	0.9649	0.3417	0.2488	0.2527	0.3186
	1.0	0.9806	0.1876	0.1208	0.1460	0.2072	0.9895	0.1166	0.0631	0.0838	0.1273
100	5.0	4.9997	0.2478	0.2430	0.2455	0.2691	4.9919	0.1903	0.1788	0.1885	0.2209
	1.0	0.9997	0.0286	0.0262	0.0282	0.0308	0.9993	0.0156	0.0143	0.0155	0.0169
	1.0	0.9981	0.2343	0.2359	0.2352	0.2357	1.0062	0.2157	0.2116	0.2126	0.2143
	0.5	0.5002	0.0216	0.0204	0.0214	0.0229	0.5002	0.0110	0.0104	0.0110	0.0118
	0.5	0.4889	0.0673	0.0744	0.0716	0.0787	0.4974	0.0426	0.0455	0.0425	0.0458
	1.0	1.0103	0.3043	0.2381	0.2501	0.3066	0.9824	0.2466	0.1653	0.1810	0.2397
	1.0	0.9917	0.1391	0.0799	0.1040	0.1536	0.9946	0.0838	0.0421	0.0595	0.0934

Note: $\psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \phi_\mu, \sigma_v^2)'$. Parameters values for generating x_t: $\theta_x = (.01, .5, .5, 2, 1)$ (see Footnote 7).

Table 2b. Monte Carlo Mean and SD, and Bootstrap Standard Errors, $m = 6$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
n	ψ	$T+1=4$					$T+1=8$				
		`Normal Errors`									
50	5.0	5.0006	0.3692	0.3683	0.3677	0.3947	5.0104	0.2857	0.2931	0.2770	0.3033
	1.0	0.9989	0.0371	0.0364	0.0378	0.0408	1.0014	0.0247	0.0253	0.0251	0.0264
	1.0	0.9489	0.3510	0.3637	0.3626	0.3732	0.9917	0.3106	0.3047	0.2986	0.3001
	0.5	0.5014	0.0275	0.0289	0.0277	0.0281	0.4990	0.0151	0.0206	0.0153	0.0121
	0.5	0.4972	0.0907	0.0953	0.0906	0.1004	0.4832	0.0601	0.0616	0.0583	0.0637
	1.0	0.9905	0.3505	0.3737	0.3424	0.3635	0.9678	0.2583	0.2832	0.2534	0.2584
	1.0	0.9805	0.1439	0.1381	0.1425	0.1687	0.9900	0.0872	0.0828	0.0835	0.0989
100	5.0	5.0276	0.2902	0.2687	0.2739	0.2910	5.0036	0.2046	0.2037	0.1966	0.2126
	1.0	1.0017	0.0297	0.0285	0.0296	0.0314	1.0005	0.0163	0.0163	0.0163	0.0170
	1.0	1.0203	0.2406	0.2402	0.2351	0.2331	0.9996	0.2197	0.2158	0.2128	0.2130
	0.5	0.4973	0.0212	0.0209	0.0203	0.0203	0.4997	0.0109	0.0140	0.0112	0.0094
	0.5	0.4898	0.0681	0.0714	0.0676	0.0718	0.4966	0.0412	0.0451	0.0414	0.0436
	1.0	1.0103	0.2643	0.2666	0.2537	0.2649	0.9836	0.1796	0.1915	0.1816	0.1877
	1.0	0.9879	0.1020	0.0946	0.1015	0.1203	0.9948	0.0579	0.0559	0.0594	0.0710
		`Normal Mixture Errors`									
50	5.0	5.0188	0.3582	0.3763	0.3684	0.4236	5.0123	0.2804	0.3036	0.2777	0.3024
	1.0	1.0003	0.0383	0.0364	0.0378	0.0434	1.0013	0.0259	0.0252	0.0250	0.0263
	1.0	0.9170	0.3839	0.3591	0.3579	0.3835	0.9963	0.2960	0.3064	0.2996	0.3004
	0.5	0.5010	0.0282	0.0287	0.0281	0.0324	0.4991	0.0155	0.0205	0.0152	0.0121
	0.5	0.4941	0.0903	0.0922	0.0907	0.1096	0.4856	0.0567	0.0571	0.0581	0.0732
	1.0	1.0256	0.6788	0.3003	0.3543	0.5729	1.0370	0.5664	0.2124	0.2691	0.4816
	1.0	0.9938	0.2765	0.0843	0.1461	0.3087	0.9911	0.1791	0.0416	0.0836	0.1925
100	5.0	5.0199	0.2863	0.2722	0.2734	0.2941	4.9971	0.1975	0.2075	0.1960	0.2116
	1.0	1.0014	0.0295	0.0283	0.0294	0.0316	1.0003	0.0161	0.0163	0.0162	0.0170
	1.0	1.0066	0.2531	0.2387	0.2336	0.2319	1.0082	0.2109	0.2147	0.2116	0.2116
	0.5	0.4983	0.0206	0.0207	0.0202	0.0208	0.4997	0.0113	0.0139	0.0111	0.0094
	0.5	0.4905	0.0672	0.0695	0.0675	0.0795	0.4969	0.0397	0.0428	0.0415	0.0496
	1.0	1.0475	0.4597	0.2037	0.2626	0.4341	1.0091	0.4092	0.1281	0.1855	0.3568
	1.0	0.9837	0.2014	0.0537	0.1014	0.2178	0.9943	0.1302	0.0270	0.0593	0.1416
		`Chi-Square Errors, df=5`									
50	5.0	5.0165	0.3750	0.3859	0.3697	0.3991	5.0351	0.2870	0.3065	0.2770	0.3015
	1.0	0.9984	0.0383	0.0365	0.0378	0.0411	1.0013	0.0255	0.0251	0.0250	0.0263
	1.0	0.9227	0.3595	0.3633	0.3621	0.3754	0.9583	0.3014	0.3049	0.2985	0.2996
	0.5	0.5008	0.0277	0.0289	0.0278	0.0288	0.4992	0.0148	0.0205	0.0152	0.0120
	0.5	0.5031	0.0877	0.0938	0.0900	0.1028	0.4849	0.0584	0.0601	0.0582	0.0662
	1.0	0.9992	0.4431	0.3510	0.3446	0.4179	0.9925	0.3520	0.2700	0.2590	0.3251
	1.0	0.9906	0.1940	0.1202	0.1441	0.2107	0.9833	0.1181	0.0638	0.0829	0.1281
100	5.0	5.0307	0.2801	0.2807	0.2744	0.2908	5.0081	0.1999	0.2133	0.1967	0.2119
	1.0	1.0016	0.0296	0.0285	0.0296	0.0315	1.0004	0.0169	0.0163	0.0163	0.0170
	1.0	1.0172	0.2419	0.2405	0.2358	0.2343	0.9989	0.2137	0.2157	0.2128	0.2130
	0.5	0.4969	0.0203	0.0208	0.0203	0.0205	0.4996	0.0112	0.0140	0.0112	0.0094
	0.5	0.4888	0.0689	0.0709	0.0677	0.0741	0.4960	0.0426	0.0443	0.0415	0.0452
	1.0	1.0304	0.3157	0.2479	0.2584	0.3169	0.9949	0.2548	0.1757	0.1833	0.2396
	1.0	0.9867	0.1323	0.0791	0.1015	0.1512	0.9932	0.0810	0.0430	0.0593	0.0931

Note: $\psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \phi_\mu, \sigma_v^2)'$. Parameters values for generating x_t: $\theta_x = (.01, .5, .5, 2, 1)$ (see Footnote 7).

Table 2c. Monte Carlo Mean and SD, and Bootstrap Standard Errors, $m = 200$

n	ψ	Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
		$T+1=4$					$T+1=8$				
		`Normal Errors`									
50	5.0	5.0283	0.3738	0.3745	0.3731	0.3958	5.0117	0.2852	0.2966	0.2834	0.3117
	1.0	1.0012	0.0392	0.0387	0.0397	0.0423	1.0003	0.0250	0.0248	0.0237	0.0243
	1.0	0.9720	0.3411	0.3339	0.3321	0.3369	1.0028	0.3041	0.3033	0.3046	0.3130
	0.5	0.4970	0.0275	0.0280	0.0265	0.0263	0.4993	0.0157	0.0217	0.0162	0.0129
	0.5	0.4778	0.0907	0.0981	0.0934	0.1017	0.4922	0.0599	0.0611	0.0575	0.0627
	1.0	1.0255	0.3967	0.3912	0.3602	0.3833	0.9842	0.2643	0.2863	0.2576	0.2646
	1.0	0.9742	0.1484	0.1380	0.1424	0.1685	0.9898	0.0817	0.0825	0.0835	0.0991
100	5.0	5.0121	0.2733	0.2740	0.2727	0.2849	5.0113	0.2131	0.2116	0.2059	0.2254
	1.0	1.0001	0.0305	0.0287	0.0298	0.0316	1.0006	0.0177	0.0176	0.0176	0.0185
	1.0	1.0020	0.2423	0.2421	0.2418	0.2448	0.9853	0.2247	0.2155	0.2137	0.2149
	0.5	0.4988	0.0213	0.0218	0.0205	0.0199	0.5000	0.0120	0.0150	0.0117	0.0095
	0.5	0.4963	0.0663	0.0707	0.0667	0.0707	0.4989	0.0408	0.0452	0.0417	0.0438
	1.0	1.0026	0.2702	0.2679	0.2535	0.2638	0.9747	0.1845	0.1934	0.1813	0.1854
	1.0	0.9865	0.1024	0.0938	0.1015	0.1212	0.9985	0.0605	0.0564	0.0597	0.0711
		`Normal Mixture Errors`									
50	5.0	5.0122	0.3683	0.3803	0.3677	0.4082	5.0039	0.2902	0.3019	0.2799	0.3079
	1.0	0.9986	0.0412	0.0385	0.0395	0.0437	1.0001	0.0238	0.0247	0.0235	0.0241
	1.0	0.9767	0.3368	0.3274	0.3248	0.3312	1.0178	0.3164	0.2979	0.2987	0.3066
	0.5	0.4993	0.0263	0.0275	0.0263	0.0285	0.4995	0.0161	0.0214	0.0160	0.0130
	0.5	0.4707	0.0960	0.0948	0.0938	0.1130	0.4945	0.0585	0.0566	0.0573	0.0711
	1.0	1.0508	0.7028	0.3138	0.3660	0.5834	1.0052	0.5478	0.2101	0.2621	0.4621
	1.0	0.9808	0.2897	0.0855	0.1438	0.2965	0.9855	0.1855	0.0417	0.0832	0.1900
100	5.0	4.9976	0.2751	0.2757	0.2705	0.2861	5.0239	0.2076	0.2165	0.2058	0.2248
	1.0	1.0018	0.0304	0.0286	0.0296	0.0316	1.0000	0.0178	0.0176	0.0176	0.0185
	1.0	0.9985	0.2392	0.2392	0.2390	0.2422	0.9823	0.2159	0.2151	0.2127	0.2136
	0.5	0.5004	0.0208	0.0216	0.0204	0.0203	0.4992	0.0118	0.0150	0.0117	0.0096
	0.5	0.4933	0.0670	0.0690	0.0669	0.0781	0.5003	0.0408	0.0429	0.0416	0.0495
	1.0	1.0146	0.4514	0.2034	0.2555	0.4149	0.9902	0.3572	0.1302	0.1840	0.3490
	1.0	0.9863	0.1955	0.0547	0.1017	0.2159	1.0014	0.1294	0.0272	0.0599	0.1440
		`Chi-Square Errors, df=5`									
50	5.0	5.0403	0.3978	0.3932	0.3732	0.3927	5.0213	0.2890	0.3071	0.2811	0.3075
	1.0	0.9996	0.0405	0.0386	0.0396	0.0423	1.0007	0.0238	0.0247	0.0236	0.0242
	1.0	0.9744	0.3420	0.3345	0.3317	0.3358	1.0090	0.3283	0.2997	0.3014	0.3098
	0.5	0.4972	0.0264	0.0280	0.0264	0.0263	0.4983	0.0162	0.0216	0.0161	0.0128
	0.5	0.4766	0.0912	0.0976	0.0935	0.1041	0.4931	0.0586	0.0595	0.0574	0.0648
	1.0	1.0448	0.4633	0.3701	0.3627	0.4375	0.9824	0.3657	0.2678	0.2568	0.3208
	1.0	0.9703	0.1867	0.1194	0.1414	0.2023	0.9853	0.1162	0.0651	0.0831	0.1257
100	5.0	4.9983	0.2807	0.2860	0.2728	0.2836	5.0051	0.2098	0.2210	0.2059	0.2244
	1.0	1.0023	0.0299	0.0287	0.0298	0.0316	1.0001	0.0178	0.0176	0.0176	0.0185
	1.0	1.0055	0.2416	0.2425	0.2418	0.2443	0.9941	0.2150	0.2161	0.2139	0.2147
	0.5	0.4996	0.0212	0.0218	0.0205	0.0200	0.4998	0.0119	0.0150	0.0117	0.0095
	0.5	0.4995	0.0647	0.0700	0.0666	0.0725	0.4989	0.0400	0.0444	0.0417	0.0454
	1.0	1.0081	0.3351	0.2480	0.2542	0.3083	0.9862	0.2441	0.1769	0.1835	0.2393
	1.0	0.9921	0.1389	0.0798	0.1021	0.1514	0.9965	0.0805	0.0429	0.0596	0.0942

Note: $\psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \phi_\mu, \sigma_v^2)'$. Parameters values for generating x_t: $\theta_x = (.01, .5, .5, 2, 1)$ (see Footnote 7).

Table 3a. Monte Carlo Mean[RMSE] for the QMLEs, Fixed Effects Model, Normal Errors

	true $m=0$			true $m=6$		
ψ	$m=0$	$m=6$	$m=200$	$m=0$	$m=6$	$m=200$
	$n=50, T=3$					
1.0	0.9957[.090]	0.9702[.088]	0.9589[.087]	1.0006[.127]	0.9983[.126]	0.9891[.125]
-0.9	-0.8966[.045]	-0.8390[.038]	-0.8139[.029]	-0.8976[.037]	-0.8934[.034]	-0.8744[.026]
0.5	0.4764[.105]	0.4471[.100]	0.4584[.100]	0.4912[.104]	0.4889[.088]	0.4837[.088]
1.0	0.9775[.141]	0.8568[.113]	0.8747[.116]	0.9934[.132]	0.9632[.131]	0.9521[.131]
1.0	0.9989[.089]	0.9969[.089]	0.9969[.089]	0.9934[.135]	0.9926[.133]	0.9926[.133]
-0.5	-0.4996[.048]	-0.4926[.048]	-0.4925[.048]	-0.4943[.074]	-0.4924[.068]	-0.4923[.068]
0.5	0.4852[.102]	0.4092[.117]	0.4091[.117]	0.5149[.114]	0.4893[.095]	0.4893[.095]
1.0	0.9662[.142]	0.9493[.142]	0.9493[.142]	0.9734[.153]	0.9410[.136]	0.9410[.136]
1.0	0.9991[.090]	0.9990[.090]	0.9990[.090]	0.9904[.139]	1.0012[.136]	1.0012[.136]
0.0	0.0004[.055]	-0.0004[.055]	-0.0004[.055]	0.0280[.103]	-0.0059[.087]	-0.0059[.087]
0.5	0.4925[.100]	0.4780[.097]	0.4780[.097]	0.5281[.101]	0.4903[.089]	0.4903[.089]
1.0	0.9673[.149]	0.9619[.147]	0.9619[.147]	1.0134[.176]	0.9340[.130]	0.9340[.130]
1.0	0.9988[.095]	0.9989[.095]	0.9988[.095]	1.0031[.135]	1.0049[.134]	1.0050[.134]
0.5	0.4976[.040]	0.4977[.040]	0.4977[.040]	0.5155[.096]	0.4983[.089]	0.4982[.089]
0.5	0.4772[.108]	0.4675[.107]	0.4675[.107]	0.5081[.102]	0.4826[.098]	0.4826[.098]
1.0	0.9610[.144]	0.9586[.144]	0.9586[.144]	0.9973[.174]	0.9703[.156]	0.9702[.156]
1.0	1.0035[.089]	1.0037[.089]	1.0037[.089]	0.9977[.133]	0.9976[.133]	0.9976[.133]
0.9	0.8991[.025]	0.8993[.025]	0.8993[.025]	0.9004[.044]	0.9002[.044]	0.9002[.044]
0.5	0.4704[.112]	0.4695[.112]	0.4692[.112]	0.4862[.104]	0.4859[.103]	0.4858[.103]
1.0	0.9682[.149]	0.9682[.149]	0.9681[.149]	0.9803[.151]	0.9803[.151]	0.9803[.151]
	$n=100, T=3$					
1.0	1.0025[.074]	0.9882[.074]	0.9750[.073]	0.9986[.071]	0.9985[.071]	0.9935[.071]
-0.9	-0.8996[.026]	-0.8753[.023]	-0.8528[.017]	-0.8996[.026]	-0.8994[.024]	-0.8858[.019]
0.5	0.4937[.077]	0.3917[.075]	0.4014[.073]	0.5001[.076]	0.4876[.068]	0.4753[.068]
1.0	0.9848[.104]	0.9411[.089]	0.9410[.091]	1.0177[.093]	0.9847[.102]	0.9765[.098]
1.0	0.9972[.075]	0.9951[.075]	0.9950[.075]	0.9994[.071]	1.0007[.070]	1.0006[.070]
-0.5	-0.5026[.038]	-0.4977[.037]	-0.4976[.037]	-0.4951[.050]	-0.4983[.047]	-0.4983[.047]
0.5	0.4892[.076]	0.4289[.078]	0.4289[.078]	0.5302[.081]	0.4977[.065]	0.4977[.065]
1.0	0.9790[.107]	0.9696[.106]	0.9696[.106]	0.9984[.107]	0.9792[.098]	0.9792[.098]
1.0	0.9992[.076]	0.9997[.075]	0.9997[.075]	0.9941[.072]	1.0022[.071]	1.0022[.071]
0.0	0.0022[.041]	0.0011[.041]	0.0011[.041]	0.0223[.064]	-0.0072[.056]	-0.0072[.056]
0.5	0.4989[.073]	0.4848[.068]	0.4848[.068]	0.5472[.075]	0.4977[.063]	0.4977[.063]
1.0	0.9944[.106]	0.9916[.105]	0.9916[.105]	1.0225[.119]	0.9584[.091]	0.9584[.091]
1.0	0.9989[.075]	0.9989[.075]	0.9989[.075]	0.9997[.069]	1.0001[.069]	1.0001[.069]
0.5	0.5014[.031]	0.5012[.030]	0.5012[.030]	0.5188[.062]	0.5036[.057]	0.5036[.057]
0.5	0.5001[.077]	0.4969[.076]	0.4969[.076]	0.5193[.070]	0.4957[.067]	0.4957[.067]
1.0	0.9829[.106]	0.9827[.106]	0.9827[.106]	1.0224[.122]	1.0056[.113]	1.0056[.113]
1.0	0.9952[.071]	0.9952[.071]	0.9952[.071]	0.9990[.068]	0.9991[.068]	0.9991[.068]
0.9	0.9003[.021]	0.9001[.021]	0.9002[.021]	0.9018[.028]	0.9020[.028]	0.9020[.028]
0.5	0.4952[.077]	0.4954[.077]	0.4954[.077]	0.4864[.076]	0.4857[.075]	0.4855[.075]
1.0	0.9844[.108]	0.9843[.108]	0.9843[.108]	0.9834[.104]	0.9836[.104]	0.9835[.104]

Note: $\psi = (\beta, \rho, \lambda, \sigma_v)'$. Parameters values for generating x_t: $\theta_x = (.01, .5, .5, 1, .5)$ (see Footnote 7).

Table 3b. Monte Carlo Mean[RMSE] for the QMLEs, Fixed Effects Model, Normal Mixture

	true $m=0$			true $m=6$		
ψ	$m=0$	$m=6$	$m=200$	$m=0$	$m=6$	$m=200$
	$n=50, T=3$					
1.0	1.0021[.092]	0.9906[.091]	0.9826[.090]	0.9981[.126]	0.9980[.125]	0.9954[.125]
-0.9	-0.8987[.041]	-0.8648[.040]	-0.8416[.033]	-0.8956[.038]	-0.8924[.038]	-0.8770[.033]
0.5	0.4862[.103]	0.4035[.098]	0.4113[.097]	0.4829[.105]	0.4850[.092]	0.4770[.091]
1.0	0.9822[.300]	0.9147[.252]	0.9238[.262]	1.0121[.264]	0.9540[.268]	0.9473[.271]
1.0	1.0026[.091]	1.0013[.091]	1.0013[.091]	0.9923[.128]	0.9905[.127]	0.9905[.127]
-0.5	-0.5009[.050]	-0.4969[.049]	-0.4969[.049]	-0.4926[.079]	-0.4881[.072]	-0.4880[.072]
0.5	0.4894[.103]	0.4415[.103]	0.4415[.103]	0.5164[.103]	0.4934[.089]	0.4934[.089]
1.0	0.9802[.285]	0.9687[.278]	0.9687[.278]	0.9807[.291]	0.9301[.247]	0.9301[.247]
1.0	0.9986[.089]	0.9986[.089]	0.9986[.089]	0.9936[.139]	1.0045[.134]	1.0045[.134]
0.0	0.0017[.062]	0.0005[.062]	0.0005[.062]	0.0254[.106]	-0.0110[.091]	-0.0110[.091]
0.5	0.4917[.102]	0.4733[.098]	0.4733[.098]	0.5371[.099]	0.5045[.088]	0.5045[.088]
1.0	0.9761[.305]	0.9731[.302]	0.9731[.302]	1.0100[.309]	0.9057[.235]	0.9057[.235]
1.0	1.0004[.090]	1.0004[.090]	1.0004[.090]	1.0033[.129]	1.0051[.128]	1.0051[.128]
0.5	0.5001[.041]	0.5000[.041]	0.5000[.041]	0.5068[.100]	0.4911[.094]	0.4911[.094]
0.5	0.4826[.105]	0.4761[.104]	0.4761[.104]	0.5054[.097]	0.4809[.094]	0.4808[.094]
1.0	0.9865[.303]	0.9844[.301]	0.9844[.301]	0.9824[.313]	0.9551[.287]	0.9550[.286]
1.0	0.9968[.094]	0.9970[.094]	0.9970[.094]	0.9971[.128]	0.9970[.128]	0.9970[.128]
0.9	0.8991[.026]	0.8993[.026]	0.8993[.026]	0.9006[.049]	0.9004[.049]	0.9004[.049]
0.5	0.4797[.107]	0.4789[.107]	0.4786[.107]	0.4884[.106]	0.4881[.105]	0.4880[.105]
1.0	0.9760[.279]	0.9760[.279]	0.9759[.279]	0.9649[.285]	0.9648[.284]	0.9649[.284]
	$n=100, T=3$					
1.0	0.9986[.076]	0.9712[.075]	0.9564[.074]	1.0022[.072]	1.0028[.072]	0.9979[.072]
-0.9	-0.9005[.030]	-0.8549[.029]	-0.8303[.023]	-0.8964[.026]	-0.8972[.025]	-0.8853[.021]
0.5	0.4909[.078]	0.4299[.071]	0.4398[.072]	0.4938[.074]	0.4864[.068]	0.4744[.068]
1.0	0.9833[.205]	0.8850[.164]	0.8978[.173]	1.0367[.177]	0.9845[.200]	0.9779[.198]
1.0	0.9976[.074]	0.9964[.074]	0.9964[.074]	0.9971[.073]	0.9971[.072]	0.9971[.072]
-0.5	-0.4987[.039]	-0.4963[.039]	-0.4963[.039]	-0.4922[.055]	-0.4926[.052]	-0.4925[.052]
0.5	0.5002[.080]	0.4672[.074]	0.4672[.074]	0.5262[.076]	0.4967[.062]	0.4967[.062]
1.0	0.9862[.204]	0.9742[.200]	0.9742[.200]	0.9994[.219]	0.9641[.188]	0.9641[.188]
1.0	1.0016[.077]	1.0017[.077]	1.0017[.077]	0.9930[.073]	1.0011[.072]	1.0011[.072]
0.0	-0.0014[.038]	-0.0015[.038]	-0.0015[.038]	0.0229[.067]	-0.0072[.059]	-0.0072[.059]
0.5	0.4921[.073]	0.4694[.071]	0.4694[.071]	0.5428[.074]	0.4998[.064]	0.4998[.064]
1.0	0.9892[.208]	0.9864[.207]	0.9864[.207]	1.0143[.224]	0.9344[.175]	0.9344[.175]
1.0	1.0003[.074]	1.0005[.074]	1.0005[.074]	1.0005[.070]	1.0010[.069]	1.0010[.069]
0.5	0.5012[.033]	0.5005[.032]	0.5005[.032]	0.5201[.067]	0.5050[.062]	0.5050[.062]
0.5	0.5131[.076]	0.5162[.073]	0.5162[.073]	0.5174[.067]	0.4941[.063]	0.4941[.063]
1.0	0.9912[.218]	0.9912[.218]	0.9912[.218]	1.0245[.222]	1.0047[.204]	1.0047[.204]
1.0	1.0019[.073]	1.0019[.073]	1.0019[.073]	0.9976[.076]	0.9977[.076]	0.9977[.076]
0.9	0.9005[.021]	0.9003[.021]	0.9003[.021]	0.9011[.028]	0.9013[.028]	0.9013[.028]
0.5	0.4976[.079]	0.4979[.079]	0.4980[.079]	0.4853[.076]	0.4846[.075]	0.4843[.075]
1.0	0.9816[.205]	0.9814[.204]	0.9815[.204]	0.9801[.202]	0.9803[.202]	0.9802[.202]

Note: $\psi = (\beta, \rho, \lambda, \sigma_v)'$. Parameters values for generating x_t: $\theta_x = (.01, .5, .5, 1, .5)$ (see Footnote 7).

Table 4a. Monte Carlo Mean and SD, and Bootstrap Standard Errors, $m = 0$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
n	ψ			$T = 3$					$T = 7$		
		Normal Errors									
50	1.0	0.9986	0.0971	0.1001	0.0981	0.0982	1.0003	0.0559	0.0545	0.0532	0.0549
	0.5	0.4988	0.0348	0.0380	0.0326	0.0437	0.4995	0.0241	0.0259	0.0241	0.0363
	0.5	0.4888	0.1055	0.1016	0.1044	0.1127	0.4917	0.0612	0.0571	0.0597	0.0639
	1.0	0.9650	0.1489	0.1713	0.1411	0.1339	0.9861	0.0806	0.0990	0.0841	0.0794
100	1.0	1.0024	0.0720	0.0744	0.0737	0.0790	1.0005	0.0340	0.0343	0.0337	0.0342
	0.5	0.5012	0.0266	0.0288	0.0273	0.0417	0.5005	0.0167	0.0173	0.0170	0.0266
	0.5	0.4922	0.0759	0.0742	0.0749	0.0782	0.4986	0.0408	0.0419	0.0428	0.0443
	1.0	0.9889	0.1044	0.1219	0.1022	0.0980	0.9948	0.0592	0.0673	0.0600	0.0576
		Normal Mixture Errors									
50	1.0	0.9979	0.0967	0.0996	0.0971	0.0973	1.0016	0.0530	0.0550	0.0533	0.0563
	0.5	0.4976	0.0338	0.0385	0.0320	0.0461	0.4994	0.0252	0.0278	0.0249	0.0408
	0.5	0.4847	0.1017	0.1001	0.1046	0.1153	0.4953	0.0585	0.0542	0.0595	0.0671
	1.0	0.9586	0.2841	0.1207	0.1401	0.2372	0.9881	0.1855	0.0637	0.0844	0.1610
100	1.0	1.0027	0.0733	0.0742	0.0733	0.0791	0.9971	0.0328	0.0342	0.0336	0.0341
	0.5	0.5000	0.0269	0.0287	0.0262	0.0431	0.4994	0.0168	0.0173	0.0169	0.0275
	0.5	0.4933	0.0718	0.0731	0.0748	0.0794	0.4995	0.0435	0.0406	0.0428	0.0457
	1.0	0.9860	0.2121	0.0833	0.1019	0.1860	0.9894	0.1291	0.0408	0.0596	0.1198
		Chi-Square, df=3									
50	1.0	0.9942	0.1022	0.1001	0.0983	0.0995	1.0034	0.0544	0.0549	0.0534	0.0557
	0.5	0.4999	0.0361	0.0376	0.0333	0.0471	0.4991	0.0251	0.0265	0.0242	0.0369
	0.5	0.4785	0.1046	0.1015	0.1060	0.1171	0.4966	0.0588	0.0554	0.0595	0.0654
	1.0	0.9646	0.2141	0.1377	0.1409	0.1860	0.9908	0.1365	0.0741	0.0845	0.1218
100	1.0	1.0012	0.0734	0.0744	0.0737	0.0792	1.0010	0.0328	0.0344	0.0338	0.0345
	0.5	0.4999	0.0312	0.0290	0.0284	0.0487	0.5003	0.0175	0.0168	0.0169	0.0263
	0.5	0.4935	0.0771	0.0735	0.0755	0.0804	0.4976	0.0441	0.0414	0.0428	0.0449
	1.0	0.9918	0.1604	0.0971	0.1024	0.1425	0.9962	0.0971	0.0486	0.0600	0.0897

Note: $\psi = (\beta, \rho, \lambda, \sigma_v^2)'$. Parameters values for generating x_t: $\theta_x = (.1, .5, .5, 5, 1)$ (see Footnote 7).

Table 4b. Monte Carlo Mean and SD, and Bootstrap Standard Errors, $m = 6$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
n	ψ			$T = 3$					$T = 7$		
		Normal Errors									
50	1.0	1.0000	0.0182	0.0189	0.0184	0.0183	1.0004	0.0095	0.0098	0.0096	0.0117
	0.5	0.5010	0.0198	0.0188	0.0190	0.0229	0.5001	0.0070	0.0073	0.0070	0.0089
	0.5	0.5000	0.1037	0.0999	0.1016	0.1058	0.4956	0.0603	0.0565	0.0594	0.0633
	1.0	0.9744	0.1450	0.1602	0.1427	0.1358	0.9914	0.0814	0.0907	0.0836	0.0809
100	1.0	0.9998	0.0150	0.0151	0.0149	0.0148	0.9999	0.0064	0.0068	0.0066	0.0075
	0.5	0.4992	0.0108	0.0117	0.0112	0.0121	0.5000	0.0052	0.0051	0.0051	0.0060
	0.5	0.4954	0.0701	0.0735	0.0728	0.0730	0.4991	0.0433	0.0418	0.0425	0.0437
	1.0	0.9805	0.1040	0.1082	0.1013	0.0990	0.9916	0.0638	0.0619	0.0591	0.0581
		Normal Mixture Errors									
50	1.0	1.0004	0.0186	0.0187	0.0180	0.0179	0.9996	0.0093	0.0098	0.0095	0.0117
	0.5	0.4999	0.0196	0.0185	0.0187	0.0235	0.4999	0.0067	0.0073	0.0069	0.0089
	0.5	0.4993	0.1029	0.0978	0.1019	0.1090	0.4977	0.0572	0.0537	0.0592	0.0662
	1.0	0.9558	0.2840	0.0986	0.1400	0.2405	0.9857	0.1872	0.0471	0.0832	0.1677
100	1.0	0.9993	0.0156	0.0151	0.0149	0.0149	1.0000	0.0067	0.0067	0.0066	0.0074
	0.5	0.4997	0.0119	0.0117	0.0112	0.0128	0.4998	0.0049	0.0051	0.0051	0.0060
	0.5	0.4948	0.0719	0.0726	0.0729	0.0741	0.4976	0.0438	0.0407	0.0426	0.0451
	1.0	0.9906	0.2015	0.0647	0.1024	0.1908	0.9897	0.1301	0.0317	0.0590	0.1243
		Chi-Square, df=3									
50	1.0	0.9991	0.0187	0.0189	0.0183	0.0182	1.0001	0.0100	0.0099	0.0096	0.0118
	0.5	0.4994	0.0195	0.0186	0.0189	0.0232	0.4998	0.0072	0.0074	0.0070	0.0089
	0.5	0.4958	0.0998	0.0997	0.1022	0.1071	0.4981	0.0569	0.0552	0.0593	0.0646
	1.0	0.9691	0.2161	0.1221	0.1418	0.1884	0.9995	0.1353	0.0615	0.0844	0.1269
100	1.0	1.0007	0.0146	0.0151	0.0149	0.0148	1.0000	0.0067	0.0068	0.0066	0.0075
	0.5	0.4999	0.0115	0.0117	0.0112	0.0124	0.4998	0.0049	0.0051	0.0051	0.0060
	0.5	0.4919	0.0704	0.0734	0.0732	0.0740	0.4977	0.0425	0.0414	0.0426	0.0443
	1.0	0.9811	0.1476	0.0803	0.1014	0.1418	0.9959	0.0955	0.0415	0.0594	0.0912

Note: $\psi = (\beta, \rho, \lambda, \sigma_v^2)'$. Parameters values for generating x_t: $\theta_x = (.1, .5, .5, 5, 1)$ (see Footnote 7)

Table 4c. Monte Carlo Mean and SD, and Bootstrap Standard Errors, $m = 200$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
n	ψ			$T = 3$					$T = 7$		
		Normal Errors									
50	1.0	1.0004	0.0210	0.0213	0.0208	0.0210	1.0000	0.0097	0.0096	0.0093	0.0100
	0.5	0.4999	0.0197	0.0199	0.0197	0.0231	0.5000	0.0070	0.0072	0.0069	0.0081
	0.5	0.4866	0.0974	0.1011	0.1009	0.1027	0.4991	0.0626	0.0562	0.0588	0.0622
	1.0	0.9624	0.1422	0.1573	0.1406	0.1349	0.9909	0.0881	0.0914	0.0837	0.0800
100	1.0	1.0001	0.0139	0.0140	0.0138	0.0154	0.9990	0.0337	0.0339	0.0333	0.0358
	0.5	0.5001	0.0117	0.0117	0.0116	0.0144	0.4986	0.0201	0.0195	0.0206	0.0370
	0.5	0.4977	0.0736	0.0726	0.0745	0.0775	0.4991	0.0409	0.0397	0.0409	0.0430
	1.0	0.9886	0.1064	0.1091	0.1019	0.0993	0.9938	0.0585	0.0673	0.0601	0.0582
		Normal Mixture Errors									
50	1.0	1.0005	0.0208	0.0213	0.0207	0.0210	0.9996	0.0092	0.0095	0.0092	0.0100
	0.5	0.4999	0.0204	0.0200	0.0196	0.0244	0.4997	0.0069	0.0072	0.0069	0.0082
	0.5	0.4796	0.1010	0.0994	0.1017	0.1064	0.5014	0.0566	0.0534	0.0586	0.0653
	1.0	0.9685	0.2847	0.1000	0.1414	0.2444	0.9937	0.1837	0.0474	0.0840	0.1685
100	1.0	1.0001	0.0138	0.0139	0.0137	0.0153	0.9994	0.0328	0.0339	0.0333	0.0360
	0.5	0.5000	0.0117	0.0117	0.0115	0.0148	0.5006	0.0209	0.0194	0.0205	0.0403
	0.5	0.4988	0.0743	0.0714	0.0743	0.0785	0.4967	0.0408	0.0387	0.0410	0.0445
	1.0	0.9835	0.2065	0.0642	0.1013	0.1879	0.9933	0.1339	0.0430	0.0600	0.1200
		Chi-Square, df=3									
50	1.0	1.0002	0.0214	0.0213	0.0208	0.0211	1.0000	0.0094	0.0096	0.0093	0.0099
	0.5	0.4995	0.0203	0.0199	0.0197	0.0238	0.5001	0.0069	0.0072	0.0070	0.0081
	0.5	0.4835	0.1009	0.1003	0.1014	0.1048	0.4990	0.0549	0.0550	0.0587	0.0634
	1.0	0.9662	0.2116	0.1220	0.1411	0.1879	0.9944	0.1367	0.0614	0.0840	0.1255
100	1.0	1.0002	0.0144	0.0139	0.0137	0.0153	1.0009	0.0335	0.0338	0.0333	0.0359
	0.5	0.5005	0.0113	0.0117	0.0115	0.0145	0.4999	0.0207	0.0193	0.0205	0.0375
	0.5	0.4987	0.0732	0.0721	0.0744	0.0780	0.5004	0.0407	0.0392	0.0408	0.0435
	1.0	0.9807	0.1505	0.0796	0.1011	0.1432	0.9922	0.0961	0.0508	0.0600	0.0894

Note: $\psi = (\beta, \rho, \lambda, \sigma_v^2)'$. Parameters values for generating x_t: $\theta_x = (.1, .5, .5, 5, 1)$ (see Footnote 7)