QML Estimation of Dynamic Panel Data Models with Spatial Errors

Liangjun Su† and Zhenlin Yang‡

February, 2007

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 individual effects and on the initial observations. Monte Carlo simulation shows that the estimators perform well in finite samples.

JEL classifications: C12, C14, C22, C5

Key Words: Dynamic Panel, Fixed Effects, Random Effects, Spatial Dependence, Quasi Maximum Likelihood

*Liangjun Su gratefully acknowledges the financial support from the NSFC (70501001). He also thanks the School of Economics and Social Sciences, Singapore Management University (SMU), for the hospitality during his two-month visit, and the Wharton-SMU research center, SMU, for supporting his visit. Zhenlin Yang gratefully acknowledges the research support from the Wharton-SMU research center, Singapore Management University.

†Guanghua School of Management, Peking University, Beijing 100871, China. Telephone: +86-10-6276-7444. Email address: lsu@gsm.pku.edu.cn.

‡School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road, Singapore 178903. Telephone: +65-0828-0852. Fax: +65-6828-0853. Email address: zlyang@smu.edu.sg.
1 Introduction

Recently there has been a growing interest in the estimation of panel data models with cross-sectional or spatial dependence. See Baltagi and Li (2004), Baltagi, Song and Koh (2003), Chen and Conley (2001), Elhorst (2003, 2005), Huang (2004), Kapoor, Kelejian and Prucha (2006), Persaran (2003, 2004), Phillips and Sul (2003), Yang, Li and Tse (2006), among others, for an overview. In this paper we focus on the quasi-maximum likelihood estimation of dynamic panel data models with spatial errors.

The history of spatial econometrics can be traced back at least to Cliff and Ord (1973). Since then, various methods have been proposed to estimate the spatial dependence models, including the method of maximum likelihood (ML) (Ord, 1975; Smirnov and Anselin, 2001), the method of moments (MM) (Kelejian and Prucha, 1998, 1999, 2006; Lin and Lee, 2006), and the method of quasi-maximum likelihood (QML) (Lee, 2004). A common feature of these methods is that they are all developed for estimations of estimate a cross-sectional model with no time dimension. Recently, Elhorst (2003, 2005) studies the ML estimation of (dynamic) panel data models with certain spatial dependence structure, but the asymptotic properties of the estimators are not given.

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, de Jong, and Lee (2006) extend the work of Lee (2004) to spatial dynamic panel data models with fixed effects allowing both the time dimension \(T\) and the cross-sectional dimension \(n\) large.

This paper concerns with the more traditional panel data model where \(n\) is allowed to grow but \(T\) is held fixed (usually small). As Binder, Hsiao and Pesaran (2005) remarked, this model remains the prevalent setting for the majority of empirical microeconomic research. Our work is distinct from that of Yu, de Jong, and Lee (2006) in several aspects. First, unlike Yu, de Jong, and Lee (2006) who consider only fixed effects model, we shall consider both random and fixed effects specification of the individual effects and highlight their differences and implications these differences have for estimation and inference. Second, when we keep \(T\) fixed, our estimation strategy is quite different from that in the large-\(n\) and large-\(T\) setting. In case of fixed effects model, we have to difference-out
the fixed effects whereas Yu, de Jong, and Lee (2006) need not do so. Third, spatial dependence is present only in the error term in our model whereas Yu, de Jong, and Lee (2006) considers spatial lag model. Consequently the two approaches complement each other. We conjecture we can extend our work to the general spatial autoregressive model with spatial error (SARAR).

The rest of the paper is organized as follows. In Section 2 we introduce our model specification. We propose the quasi maximum likelihood estimates in Section 3 and study their asymptotic properties in Section 4. In Section 5 we provide a small set of Monte Carlo experiments to evaluate the finite sample performance of our estimators. All proofs are relegated to the appendix.

To proceed, we introduce some notation and convention. Let $I_n$ denote an $n \times n$ identity matrix. Let $\iota_T$ denote a $T \times 1$ vector of ones and $J_T = \iota_T \iota_T'$, where prime denotes transposition throughout this paper. $\otimes$ denotes the Kronecker product. $|\cdot|$ denotes the absolute value of a scalar or determinant of a matrix.

2 Model Specification

We consider the model of the form

$$y_{it} = \rho y_{i,t-1} + x_{it}' \beta + z_i' \gamma + u_{it},$$

for $i = 1, ..., n$, $t = 1, ..., T$, where the scalar parameter $\rho$ with $|\rho| < 1$ 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 such as the constant term or dummies representing individuals’ gender or race, and the disturbance vector $u_t = (u_{it}, ..., u_{nt})'$ is assumed to exhibit both non-observable individual effects and spatially autocorrelated structure, i.e.,

$$u_t = \mu + \varepsilon_t,$$

$$\varepsilon_t = \lambda W_n \varepsilon_t + v_t,$$

where $\mu = (\mu_1, ..., \mu_n)'$, $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{nt})'$, and $v_t = (v_{1t}, ..., v_{nt})'$, with $\mu$ representing the unobservable individual effects which could be either random or fixed, $\varepsilon_t$ representing the spatially correlated errors, and $v_t$ representing the random innovations that are assumed to be independent and identically distributed (i.i.d.) with zero mean and variance $0, \sigma_v^2$. In the case where $\mu$ is random, its elements are assumed to be i.i.d. $(0, \sigma^2_\mu)$ and to be independent of $v_t$. In the case where $\mu$ is fixed, the time invariant regressors should be removed from the model due to multicollinearity between the
observed and unobserved individual-specific effects. The parameter $\lambda$ is a the spatial autoregressive coefficient and $W_n$ is a known $n \times n$ spatial weight matrix whose diagonal elements are zero. Following the literature in spatial econometrics, we assume that $I_n - \lambda W_n$ is nonsingular. We will also assume the observations on $(y_{it}, x'_{it}, z'_i)'$ are available at the initial period $t = 0$.

Let $B_n = B_n(\lambda) = I_n - \lambda W_n$. Frequently, we will suppress the dependence of $B_n$ and $W_n$ on $n$ and write $B$ and $W$ instead. We have $\varepsilon_t = B^{-1} v_t$. Let $y_t = (y_{1t}, ..., y_{nt})'$, and $x_t = (x_{1t}, ..., x_{nt})'$. Define $Y = (y'_1, ..., y'_T)$, $Y_{-1} = (y'_{0}, ..., y'_{T-1})'$, $X = (x'_1, ..., x'_T)'$, and $Z = \nu T \otimes z$, where $z = (z_1, ..., z_n)'$.

Using matrix notation, we can write the model specified by Eqs (2.1)-(2.3) as

$$Y = \rho Y_{-1} + X\beta + Z\gamma + u, \text{ with } u = (\nu T \otimes I_n)\mu + (I_T \otimes B^{-1}) v. \quad (2.4)$$

It is worth mentioning that Eqs (2.1)-(2.3) allow spatial dependence to be present in the random disturbance term $\varepsilon_t$ but not in the individual effect $\mu$. See Baltagi, Song and Koh (2003) and Baltagi and Li (2004) for the application of this type of models. Alternatively we can allow both $\varepsilon_t$ and $\mu$ to follow a spatial autoregressive model as is done by Kapoor, Kelejian and Prucha (2006) who consider GMM estimation of static spatial panel model with random effects. Our theory can readily be modified to take into account the latter case, and we conjecture that a specification test can also be developed to test for the two different specifications.

## 3 Quasi Maximum Likelihood Estimation

In this section we develop quasi maximum likelihood estimates (QMLEs) based on Gaussian likelihood for the models specified above.

### 3.1 QMLE for the Random Effects Model

For the random effects model, the covariance matrix of $u$ has the familiar form $E(uu') = \sigma_u^2 \Omega$, with

$$\Omega = \Omega (\phi_\mu, \lambda) = \phi_\mu (J_T \otimes I_n) + I_T \otimes (B'B)^{-1}, \quad (3.1)$$

where $\phi_\mu = \sigma_\mu^2 / \sigma_\nu^2$, $J_T = \nu T \nu_T'$, and we suppress the dependence of $\Omega$ on $n$. We frequently suppress the argument of $\Omega$ when no confusion can arise. It is well known that the likelihood function for a dynamic panel model depends on the assumptions on the initial observations (Hsiao, 2003).

If $|\rho| \geq 1$ or the processes generating the $x_{it}$ are not stationary, it does not make sense to assume
that the process generating the \( y_{it} \) is the same prior to the period of observations as for \( t = 1, \ldots, T \). For this reason, we consider two sets of assumptions about initial observations \( \{ y_{i0} \} \).

**Case I: \( y_{i0} \) is exogenous**

If \( y_{i0} \) is taken as exogenous, it rules out the plausible assumption that it is generated by the same process as generates \( y_{it}, t = 1, \ldots, T \). In this case, we can easily derive the likelihood function for the model (2.4) conditional on \( y_{i0} \).

Let \( \theta = (\beta', \gamma', \rho)' \), \( \delta = (\lambda, \phi, \mu)' \), and \( \varsigma = (\theta', \sigma_v^2, \delta')' \). The log likelihood function of (2.4) is

\[
L_r(\varsigma) = -\frac{nT}{2} \log(2\pi) - \frac{nT}{2} \log(\sigma_v^2) - \frac{1}{2} \log |\Omega| - \frac{1}{2\sigma_v^2} u(\theta)' \Omega^{-1} u(\theta) \tag{3.2}
\]

where \( u(\theta) = Y - \rho Y_{-1} - X \beta - Z \gamma \).

Maximizing (3.2) gives the QMLE of the model parameters based on the Gaussian likelihood. Computationally it is convenient to work with the concentrated log-likelihood by concentrating out the parameters \( \theta \) and \( \sigma_v^2 \). From (3.2), given \( \delta \), the QMLE of \( \theta \) is

\[
\tilde{\theta}(\delta) = \left[ \tilde{X}' \Omega^{-1} \tilde{X} \right]^{-1} \tilde{X}' \Omega^{-1} Y, \tag{3.3}
\]

and the QMLE of \( \sigma_v^2 \) is

\[
\tilde{\sigma}_v^2(\delta) = 1 \frac{nT}{2} \tilde{u}(\delta)' \Omega^{-1} \tilde{u}(\delta), \tag{3.4}
\]

where \( \tilde{X} = (X, Z, Y_{-1}) \), and \( \tilde{u}(\delta) = Y - \tilde{X} \tilde{\theta}(\delta) \). Substituting (3.3) and (3.4) back into (3.2) for \( \theta \) and \( \sigma_v^2 \), we obtain the concentrated log-likelihood function of \( \delta \) :

\[
L_c(\delta) = -\frac{nT}{2} \log(2\pi + 1) - \frac{nT}{2} \log [\tilde{\sigma}_v^2(\delta)] - \frac{1}{2} \log |\Omega|. \tag{3.5}
\]

The QMLE \( \tilde{\delta} = (\tilde{\phi}, \tilde{\lambda})' \) of \( \delta \) maximizes the concentrated log-likelihood (3.5). The QMLEs of \( \theta_1 \) and \( \sigma_v^2 \) are given by \( \tilde{\theta}_1(\tilde{\delta}) \) and \( \tilde{\sigma}_v^2(\tilde{\delta}) \), respectively. Further, the QMLE of \( \sigma_\mu^2 \) is given by \( \tilde{\sigma}_\mu^2 = \tilde{\phi}_\mu \tilde{\sigma}_v^2 \).

**Case II: \( y_{i0} \) is endogenous**

If \( y_{i0} \) is taken as endogenous, there are several approaches to treat initial observations. Assume that, possibly after some differencing, both \( y_{it} \) and \( x_{it} \) are stationary. In this case, the initial observations are determined by

\[
y_{i0} = \sum_{j=0}^{\infty} \rho^j x_{-j} \beta + \frac{z\gamma}{1 - \rho} + \frac{\mu}{1 - \rho} + \sum_{j=0}^{\infty} \rho^j B^{-1} v_{-j}. \tag{3.6}
\]
Since \( x_{-j}, j = 1, 2, \ldots \), are not observable, we cannot use \( x_{-j} \) in our estimation procedure. In this paper we follow Bhargava and Sargan (1983) (see also Hsiao, 2003, p.76) and assume that the initial observation \( y_0 \) can be approximated by

\[
y_0 = \pi_0 t_n + x \pi_1 + z \pi_2 + \epsilon \equiv \tilde{x}\pi + \epsilon, \tag{3.7}
\]

where \( x = (x_0, x_1, \ldots, x_T) \), \( \tilde{x} = (t_n, x, z) \), \( \pi = (\pi_0, \pi'_1, \pi'_2)' \), \( E(\epsilon|x, z) = 0 \), and the covariance structure of \( \epsilon \) is affected by the spatial weight matrix \( W \). Note that if \( z \) contains the constant term, then \( t_n \) should vanish in (3.7).

Under the stationarity assumption, (3.6) implies that \( y_0 = \tilde{y}_0 + \zeta_0 \), where \( \tilde{y}_0 \) is the systematic or exogenous part of \( y_0 \) and \( \zeta_0 \) is the endogenous part, namely,

\[
\tilde{y}_0 = \sum_{j=0}^{\infty} \rho^j x_{-j} \beta + \frac{z_0}{1-\rho} \quad \text{and} \quad \zeta_0 = \frac{\mu}{1-\rho} + \sum_{j=0}^{\infty} \rho^j B^{-1} v - j. \tag{3.8}
\]

(3.7) then follows by assuming that the optimal predictor of \( \tilde{y}_0 \) conditional on the observable \( x \) and \( z \) is \( \tilde{x}\pi : \tilde{y}_0 = \tilde{x}\pi + \zeta \), where \( \zeta = (\zeta_1, \ldots, \zeta_n)' \), and \( \zeta_0 \) are i.i.d. \((0, \sigma_{\zeta}^2)\) and they are independent of \( x_{it}, z_i, \mu_i \) and \( \varepsilon_{it} \).

By construction, \( \epsilon = \zeta + \zeta_0 \), and \( E(\epsilon_i) = 0 \). We can verify that under strict exogeneity of \( x_{it} \) and \( z_i \),

\[
E(\epsilon') = \sigma_{\zeta}^2 I_n + \frac{\sigma_\mu^2}{(1-\rho)^2} I_n + \frac{\sigma_\nu^2}{1-\rho^2} (B'B)^{-1} \tag{3.9}
\]

and

\[
E(\epsilon u') = \frac{\sigma_\mu^2}{1-\rho} I_T \otimes I_n \tag{3.10}
\]

We require that \( n > p(T + 1) + q + 1 \) for the identification of the parameters in (3.7). The is impossible if \( T \) is relatively large and \( p \neq 0 \). For this reason, the regressor in (3.7) is frequently replaced by \( \bar{x} \equiv (T+1)^{-1} \sum_{t=0}^{T} x_t \), and thus we have

\[
y_0 = \pi_0 t_n + \bar{x} \pi_1 + z \pi_2 + \epsilon \equiv \tilde{x}\pi + \epsilon, \tag{3.11}
\]

where now \( \tilde{x} = (t_n, \bar{x}, z) \), \( \pi = (\pi_0, \pi'_1, \pi'_2)' \), and the variance-covariance structure of \( \epsilon \) and \( u \) is the same as before except for the definition of \( \sigma_{\zeta}^2 \).

In contrast, Hsiao (2003, p.76, Case II) simply assumes that \( y_{0i} \) are random with a common mean \( \pi \) and write \( y_0 \) as

\[
y_0 = \epsilon_i + \mu_i \equiv \tilde{x}\pi + \epsilon, \tag{3.12}
\]
where now \( \bar{x} = x_n \), \( \epsilon = (\epsilon_1, \ldots, \epsilon_n)' \), \( \epsilon_i \) represents the deviation of initial individual endowments from the mean, and the variance-covariance structure of \( \epsilon \) and \( u \) is the same as before except for the definition of \( \sigma^2_\epsilon \).

In the following, \( \bar{x} \) can be any one of them defined in (3.7), (3.11) or (3.12). We will simply refer its dimension as \( n \times k \), where \( k \) varies from case to case.

Because the likelihood function (3.2) assumes that the initial observations are exogenously given, it generally produces biased estimators when this assumption is not satisfied (see Bhargava and Sargan, 1983). Under the assumption that the presample values are generated by the same process as the within-sample observations, we need to derive the joint distribution of \( y_T, \ldots, y_1, y_0 \) from (2.4) and (3.7), (3.11) or (3.12). Denoting by \( \sigma^2_\Omega^* \) the \( n (T + 1) \times n (T + 1) \) symmetric matrix of \( u^* = (\epsilon, u)' \), we see that \( \Omega^* \) has the form

\[
\sigma^2_\Omega^* = \left( \begin{array}{c}
\phi_\zeta I_n + \frac{\sigma_\mu^2}{(1 - \rho)^2} I_n + \frac{\sigma_\nu^2}{1 - \rho^2} \left( B'B \right)^{-1} \frac{\sigma_\mu^2}{1 - \rho^2} I_T \otimes I_n \\
\phi_\zeta I_n + \frac{\sigma_\nu^2}{(1 - \rho)^2} I_n + \frac{1}{1 - \rho^2} \left( B'B \right)^{-1} \frac{\sigma_\mu^2}{1 - \rho^2} I_T \otimes I_n \\
\phi_\zeta I_n + \frac{\sigma_\nu^2}{(1 - \rho)^2} I_n + \frac{\sigma_\mu^2}{1 - \rho^2} I_T \otimes I_n \\
\Omega
\end{array} \right)
\]

where \( \phi_\zeta = \sigma^2_\epsilon / \sigma^2_\nu \) and we frequently suppress the argument of \( \Omega^* \) when no confusion can arise.

Now let \( \theta = (\beta', \gamma', \pi')' \), \( \delta = (\rho, \lambda, \phi_\mu, \phi_\zeta)' \), and \( \zeta = (\theta', \sigma^2_\nu, \delta')' \). Note that \( \zeta \) is a \( (p + q + k + 5) \times 1 \) vector of unknown parameters. Based on (2.4) and (3.7), (3.11) or (3.12), and assuming Gaussian likelihood, the random effect QMLE estimator of \( \zeta \) is derived by maximizing the following log-likelihood function:

\[
\mathcal{L}^* (\zeta) = -\frac{n (T + 1)}{2} \log (2\pi) - \frac{n (T + 1)}{2} \log (\sigma^2_\epsilon) - \frac{1}{2} \log |\Omega^*| - \frac{1}{2} \sigma^2_\nu u^* (\theta)' \Omega^{-1} u^* (\theta)
\]

where \( u^* (\theta) = (y_0^* - \pi^* \beta^*, u (\beta, \gamma, \rho)')' \), and \( u (\beta, \gamma, \rho) = Y - \rho Y_{-1} - X \beta - Z \gamma \).

Maximizing (3.13) gives the quasi maximum likelihood estimates (QMLEs) of the model parameters based on the Gaussian likelihood. We will work with the concentrated log-likelihood by concentrating out the parameters \( \theta \) and \( \sigma^2_\nu \). From (3.13), given \( \delta = (\rho, \lambda, \phi_\mu, \phi_\zeta)' \), the QMLE of \( \theta \) is

\[
\hat{\theta} (\delta) = \left[ X^* \Omega^{-1} X^* \right]^{-1} X^* \Omega^{-1} Y^*,
\]

(3.14)
and the QMLE of \( \sigma_v^2 \) is

\[
\hat{\sigma}_v^2(\delta) = \frac{1}{nT} \tilde{u}^*(\delta)' \Omega_v^{-1} \tilde{u}^*(\delta),
\]

(3.15)

where

\[
Y^* = \begin{pmatrix} y_0 \\ Y - \rho Y_{-1} \end{pmatrix}, \quad X^* = \begin{pmatrix} 0_{n \times p} & 0_{n \times q} & \tilde{x} \\ X & Z & 0_{nT \times k} \end{pmatrix}, \quad \tilde{u}^*(\delta) = \begin{pmatrix} y_0 \tilde{x} \tilde{\pi} \delta \\ Y - X \tilde{\beta}(\delta) - Z \tilde{\gamma}(\delta) - \rho Y_{-1} \end{pmatrix},
\]

and \( \hat{\theta}(\delta) = (\hat{\beta}(\delta)', \hat{\gamma}(\delta)', \hat{\pi}(\delta)')' \). Substituting (3.14) and (3.15) back into (3.13) for \( \theta \) and \( \sigma_v^2 \), we obtain the concentrated log-likelihood function of \( \delta \):

\[
L^c(\delta) = -\frac{n(T+1)}{2} \log(2\pi) + 1 - \frac{n(T+1)}{2} \log \hat{\sigma}_v^2(\delta) - \frac{1}{2} \log |\Omega^*|.
\]

(3.16)

The QMLE \( \hat{\delta} = (\hat{\rho}, \hat{\lambda}, \hat{\sigma}_v, \hat{\phi}_y)' \) of \( \delta \) maximizes the concentrated log-likelihood (3.5). The QMLEs of \( \theta \) and \( \sigma_v^2 \) are given by \( \hat{\theta}(\hat{\delta}) \) and \( \hat{\sigma}_v^2(\hat{\delta}) \), respectively. Further, the QMLE of \( \sigma_v^2 \) and \( \sigma_y^2 \) are given by \( \hat{\sigma}_v^2 = \hat{\phi}_v \hat{\sigma}_v^2 \) by \( \hat{\sigma}_y^2 = \hat{\phi}_y \hat{\sigma}_y^2 \), respectively.

### 3.2 QMLE for the Fixed Effects Model

In this section, we consider the dynamic panel data model with fixed effects. In this case, we write the model in vector notation

\[
y_t = \rho y_{t-1} + x_t' \beta + z\gamma + \mu + B^{-1} v_t,
\]

(3.17)

where, for example, \( \mu = (\mu_1, ..., \mu_n)' \) denotes the fixed effects that may be correlated with the regressors \( x_t \) and \( z \), and the specification for other variables is the same as the random effects case.

Following the standard practice, we eliminate \( \mu \) by first-differencing (3.17), namely,

\[
\Delta y_t = \rho \Delta y_{t-1} + \Delta x_t' \beta + B^{-1} \Delta v_t.
\]

(3.18)

(3.18) is well defined for \( t = 2, 3, ..., T \) but not for \( t = 1 \) because observations on \( y_{i,-1} \) are not available. By continuous substitution, we can write \( \Delta y_1 \) as

\[
\Delta y_1 = \rho^n \Delta y_{-m+1} + \sum_{j=0}^{m-1} \rho^j \Delta x_{1-j} \beta + \sum_{j=0}^{m-1} \rho^j B^{-1} \Delta v_{1-j}.
\]

(3.19)

Like Hsiao et al. (2002), since the observations \( \Delta x_{1-j}, j = 1, 2, ... \) are not available, the conditional mean of \( \Delta y_1 \) given \( \Delta y_{-m+1} \) and \( \Delta x_{1-j}, j = 0, 1, 2, ... \) as defined by

\[
\eta_1 = E(\Delta y_1|\Delta y_{-m+1}, \Delta x_1, \Delta x_0, ...) = \rho^n \Delta y_{-m+1} + \sum_{j=0}^{m-1} \rho^j \Delta x_{1-j} \beta,
\]

(3.20)
is unknown even if we assume that $m$ is sufficiently large. Noting that $\eta_1$ is an $n \times 1$ vector, we will confront the incidental parameters problem if we treat $\eta_1$ as a free parameter to be estimated. As Hsiao et al. (2002) remark, to get around this problem, the expected value of $\eta_1$, conditional on the observables, has to be a function of a finite number of parameters, and such a condition can hold provided that $\{x_{it}\}$ are trend-stationary (with a common deterministic linear trend) or first-difference stationary processes. In this case, the expected value of $\Delta x_{i,1-\cdot}$, conditional on the $pT \times 1$ vector $\Delta x_i = (\Delta x'_{i1}, \ldots, \Delta x'_{iT})'$, is linear in $\Delta x_i$, i.e.,

$$E(\Delta x_{i,1-\cdot}|\Delta x_i) = \pi_{0j} + \pi_{1j}\Delta x_i,$$

(3.21)

where $\pi_{0j}$ and $\pi_{1j}$ don’t depend on $i$. Denote $\Delta x = (\Delta x_1, \ldots, \Delta x_n)'$, an $n \times pT$ matrix. Then under the assumption that $E(\Delta y_{i,-m+1}|\Delta x_{i1}, \Delta x_{i2}, \ldots, \Delta x_{iT})$ is the same across all individuals, we have

$$\Delta y_1 = \pi_0 t_n + \Delta x \pi_1 + e \equiv \Delta \bar{x} \pi + e,$$

(3.22)

where $e = (\eta_1 - E(\eta_1|\Delta x)) + \sum_{j=0}^{m-1} \rho^j B^{-1} \Delta v_{1-j}$ is an $n \times 1$ random vector with typical element $e_i$ ($i = 1, \ldots, n$); $\pi = (\pi_0, \pi_1)'$ is a $(pT+1) \times 1$ vector of parameters associated with the conditional mean of $\Delta y_1$; and $\Delta \bar{x} = (t_n, \Delta x)$. (3.21) and (3.22) are associated with the Bhargava and Sargan’s (1983) approximation for the 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 $x_{it}$,

$$E(e_i|\Delta x_i) = 0, E(ee') = \sigma^2 \mathbf{I}_n + \sigma^2 \mathbf{c}_m (B' B)^{-1} \equiv \sigma^2 \mathbf{c}_m (B' B)^{-1} (\phi_e \mathbf{B} \mathbf{B}' + \mathbf{c}_m \mathbf{I}_n) (B' B)^{-1},$$

(3.23)

and

$$E(e \Delta u_2') = -\sigma^2 \epsilon (B' B)^{-1}, E(e \Delta u'_t) = 0 \text{ for } t = 3, \ldots, T,$$

(3.24)

where $\Delta u_t = B^{-1} \Delta v_t$, $\sigma^2 = \text{Var}(\eta_1)$ is identical across $i$, $\mathbf{c}_m = 2 \left(1 + \rho^{2m-1}\right)/(1 + \rho)$, and $\phi_e = \sigma^2 \epsilon / \sigma^2 v$. Clearly, when $m \to \infty$, $c_{\infty} = 2/(1 + \rho)$, which is not a free parameter. We assume that $m$ is an unknown finite and identical across individuals so that we will estimate $c_m$ below. See Hsiao et al. (2002, p.110).

We require that $n > pT + 1$ for the identification of the parameters in (3.22). This becomes impossible if $T$ is relatively large in applications and $p > 1$. Like the random effects model, $\Delta \bar{x}$ in (3.22) can be chosen to be other variables, so that we have

$$\Delta y_1 = \pi_0 t_n + \Delta x \pi_1 + e \equiv \Delta \bar{x} \pi + e,$$

(3.25)
or
\[
\Delta y_1 = \pi \tau_n + e = \Delta \tilde{x} \pi + e, \quad (3.26)
\]

where \(\Delta \tilde{x} = (\tau_n, \Delta \pi)\) with \(\Delta \tau = T^{-1} \sum_{t=1}^{T} \Delta x_t\) in (3.25), \(\Delta \tilde{x} = \tau_n\) in (3.26), and in each case the variance-covariance structure of \(e\) and \((\Delta v_2, ..., \Delta v_T)\) are the same as above. In the following, we simply refer to the dimension of \(\pi\) to be \(k\).

Let \(E = \phi_e BB' + c_m I_n\). Then the covariance matrix of \(\Delta u \equiv (e', \Delta u'_2, ..., \Delta u'_T)\) is given by
\[
\text{Var}(\Delta u) = \sigma_v^2 (I_T \otimes B^{-1}) H_E (I_T \otimes B'^{-1}) \equiv \sigma_v^2 \Omega', \quad (3.27)
\]

where the \(nT \times nT\) matrix \(H_E\) is defined by
\[
H_E = \left( \begin{array}{cccccc}
E & -I_n & 0 & \cdots & 0 & 0 \\
-I_n & 2I_n & -I_n & \cdots & 0 & 0 \\
0 & -I_n & 2I_n & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 2I_n & -I_n \\
0 & 0 & 0 & \cdots & -I_n & 2I_n \\
0 & 0 & 0 & \cdots & 0 & -I_n & 2I_n \\
\end{array} \right), \quad (3.28)
\]

Maximizing (3.29) gives the Gaussian QMLE of the model parameters. We will work with the concentrated log-likelihood by concentrating out the parameters \(\theta\) and \(\sigma_v^2\). From (3.13), given \(\delta = (\lambda, c_m, \phi_e)'\), the QMLE of \(\theta\) is derived by maximizing the following log-likelihood function:
\[
L'(\varsigma) = -\frac{nT}{2} \log(2\pi) - \frac{nT}{2} \log(\sigma_v^2) - \frac{1}{2} \log |\Omega^\dagger| - \frac{1}{2\sigma_v^2} \Delta u(\theta)' \Omega^{-1} \Delta u(\theta), \quad (3.29)
\]

where
\[
\Delta u(\theta) = \begin{pmatrix}
\Delta y_1 - \Delta \tilde{x} \pi \\
\Delta y_2 - \rho \Delta y_1 - \Delta x_2 \beta \\
\vdots \\
\Delta y_T - \rho \Delta y_{T-1} - \Delta x_T \beta
\end{pmatrix}, \quad (3.30)
\]

Maximizing (3.29) gives the Gaussian QMLE of the model parameters. We will work with the concentrated log-likelihood by concentrating out the parameters \(\theta\) and \(\sigma_v^2\). From (3.13), given \(\delta = (\lambda, c_m, \phi_e)'\), the QMLE of \(\theta\) is
\[
\hat{\theta}(\delta) = [\Delta X' \Omega^{-1} \Delta X]^{-1} \Delta X' \Omega^{-1} \Delta Y, \quad (3.31)
\]
and the QMLE of $\sigma^2_v$ is
$$\hat{\sigma}^2_v(\delta) = \frac{1}{nT} \tilde{\Delta}u(\delta)' \Omega^{-1} \tilde{\Delta}u(\delta),$$
(3.32)
where
$$\Delta Y = \begin{pmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_T \end{pmatrix}, \quad \Delta X = \begin{pmatrix} 0_{n \times p} & 0_{n \times 1} & \Delta \tilde{x} \\ \Delta x_2 & \Delta y_1 & 0_{n \times k} \\ \vdots & \vdots & \vdots \\ \Delta x_T & \Delta y_{T-1} & 0_{n \times k} \end{pmatrix},$$
and $\tilde{\Delta}u(\delta)$ equals $\Delta u(\theta)$ with $\theta$ being replaced by $\tilde{\theta}(\delta)$.

Substituting (3.14) and (3.15) back into (3.13) for $\theta$ and $\sigma^2_v$, we obtain the concentrated log-likelihood function of $\delta$:
$$L^c_f(\delta) = -\frac{nT}{2} \left(\log(2\pi)+1\right) - \frac{nT}{2} \log \tilde{\sigma}^2_v(\delta) - \frac{1}{2} \log |\Omega^1|. \quad (3.33)$$
The QMLE $\tilde{\theta} = (\tilde{\lambda}, \tilde{\phi}_m, \tilde{\phi}_e)'$ of $\delta$ maximizes the concentrated log-likelihood (3.33). The QMLEs of $\theta$ and $\sigma^2_v$ are given by $\tilde{\theta}(\tilde{\delta})$ and $\tilde{\sigma}^2_v(\tilde{\delta})$, respectively. Further, the QMLE of $\sigma^2_v$ is given by $\tilde{\sigma}^2_v = \tilde{\phi}_e \tilde{\sigma}^2_v$.

### 3.3 Computational Issues

Maximization of $L^c(\delta), L^{c'}(\delta)$ and $L^f(\delta)$ involves repeated evaluations of the inverse and determinants of the $nT \times nT$ matrices $\Omega$ and $\Omega^*$, and the $n(T+1) \times n(T+1)$ matrix $\Omega^*$. 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 $\Omega$:
$$|\Omega| = \left|(B'B)^{-1} + \phi_p TI_n\right| \cdot |B|^{-2(T-1)},$$
(3.34)
$$\Omega^{-1} = T^{-1} J_T \otimes \left((B'B)^{-1} + \phi_p TI_n\right)^{-1} + (I_T - T^{-1} J_T) \otimes (B'B).$$
(3.35)
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. By Griffith (1988), calculations of the determinants can be further simplified:
$$|B| = \prod_{i=1}^n (1 - \lambda w_i), \quad \left|(B'B)^{-1} + \phi_p TI_n\right| = \prod_{i=1}^n \left(1 - \lambda w_i\right)^{-1} + \phi_p T,$$  
(3.36)
where $w_i's$ are the eigenvalues of $W$. The above simplifications are also used in Yang et al. (2006).
For the inverse of $\Omega^*$, using the formula for the inverse of a partitioned matrix (e.g., Magnus and Neudecker, 2002), we have

$$\Omega^{*-1} = \begin{pmatrix} D^{-1} & -D^{-1}\omega_{12}\Omega^{-1} \\ -\Omega^{-1}\omega_{21}D^{-1} & \Omega^{-1} + \Omega^{-1}\omega_{21}D^{-1}\omega_{12}\Omega^{-1} \end{pmatrix} \equiv \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}, \quad (3.37)$$

where $D = \omega_{11} - \omega_{12}\Omega^{-1}\omega_{21}$. For $|\Omega^*|$, we have

$$|\Omega^*| = |\Omega| |\omega_{11} - \omega_{12}\Omega^{-1}\omega_{21}|, \quad (3.38)$$

which involves the inverse and determinants of several $n \times n$ matrices by using (3.34) and (3.35).

For $\Omega^t$, by the properties of matrix operation,

$$|\Omega^t| = |(I_T \otimes B^{-1})| |H_E| |(I_T \otimes B'^{-1})| = |B|^{-2T} |H_E|,$$

$$\Omega^{t-1} = (I_T \otimes B'^{-1})^{-1} H_E^{-1} (I_T \otimes B^{-1})^{-1} = (I_T \otimes B') H_E^{-1} (I_T \otimes B),$$

where

$$|H_E| = |E^*|,$$

$$H_E^{-1} = (1 - T) (h_0^{-1} \otimes E^{*-1}) + (h_1^{-1} - (1 - T) h_0^{-1}) \otimes (E^{*-1} E), \quad (3.39)$$

$$E^* = TE - (T - 1) I_n,$$

and the $T \times T$ matrix $h_\theta$ ($\theta = 0, 1$) is defined as

$$h_\theta = \begin{pmatrix} \theta & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}. $$
Hsiao et al. (2002) give the determinant of $h_\theta$ as
$$|h_\theta| = 1 + T(\theta - 1),$$
and the inverse of $h_\theta$ as
$$h_\theta^{-1} = [1 + T(\theta - 1)]^{-1}
\begin{pmatrix}
T & T - 1 & T - 2 & \cdots & 3 & 2 & 1 \\
T - 1 & (T - 1) \theta & (T - 2) \theta & \cdots & 3 \theta & 2 \theta & \theta \\
T - 2 & (T - 2) \theta & (T - 2) (2 \theta - 1) & \cdots & 3 (2 \theta - 1) & 2 (2 \theta - 1) & (2 \theta - 1) \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
3 & 3 \theta & 3 (2 \theta - 1) & \cdots & 3 [(T - 3) \theta - (T - 4)] & 2 [(T - 3) \theta - (T - 4)] & (T - 3) \theta - (T - 4) \\
2 & 2 \theta & 2 (2 \theta - 1) & \cdots & 2 [(T - 3) \theta - (T - 4)] & 2 [(T - 2) \theta - (T - 3)] & (T - 2) \theta - (T - 3) \\
1 & \theta & (2 \theta - 1) & \cdots & (T - 3) \theta - (T - 4) & (T - 2) \theta - (T - 3) & (T - 1) \theta - (T - 2)
\end{pmatrix}.$$
Assumption G(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 G(ii) is standard in the literature. Assumption G(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, 2006; Lee, 2002, 2004; Lin and Lee, 2006; Robinson, 2006). One can relax the strict exogeneity condition in Assumption G(iii) like Hsiao et al. (2002) but this will complicate our analysis in case of spatially correlated errors. Assumption G(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 did Hsiao et al. (2002). Assumption G(v) is commonly assumed in the literature but deserves some further discussion: if we assume that the spatial weight matrix \( W \) is row normalized, it is well known that the largest eigenvalue (in absolute value) of \( W \) is 1, so that only \( \lambda \in (-1, 1) \) is required for the good behavior of \((I_n - \lambda W)^{-1}\). In this case Assumption G(v) requires that \( \lambda \) lie in a compact subset of \((-1, 1)\). Further, once we relax the assumption that \( W \) is row-normalized, the parameter space usually varies with the spatial units \( n \). For more detailed and excellent discussions, see Kelejian and Prucha (2006).

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

**Assumption W** (i) The elements \( w_{n,ij} \) of \( W_n \) are at most of order \( h_n^{-1} \), denoted by \( O(1/h_n) \), uniformly in all \( i,j \). As a normalization, \( W_{n,ii} = 0 \) for all \( i \). (ii) The ratio \( h_n/n \to 0 \) as \( n \) goes to infinity. (iii) The matrix \( B_n \) is nonsingular. (iv) The sequences of matrices \( \{W_n\} \) and \( \{B_n^{-1}\} \) are uniformly bounded in both row and column sums. (v) \( \{B_n^{-1}(\lambda)\} \) are uniformly bounded in either row or column sums, uniformly in \( \lambda \) in a compact parameter space \( \Lambda \).

Assumptions W(i)-(iv) parallel Assumptions 2-4 of Lee (2004). Like Lee (2004), Assumptions W(i)-(iv) provide the essential features of the weight matrix for the model. Assumption W(i) is always satisfied if \( \{h_n\} \) is a bounded sequence. We allow \( \{h_n\} \) to be divergent but at rate smaller than \( n \) as specified in Assumption W(ii). Assumption W(iii) guarantees that disturbance term is well defined. Kelejian and Prucha (1998, 1999, 2001) and Lee (2004) also assume Assumption 2(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 Hohnson (1985, p. 301), \( \limsup_n \| \lambda_0 W_n \| < 1 \) is sufficient to guarantee that \( B_n^{-1} \) is uniformly bounded in both row and column sums. By Lee (2002, Lemma A.3), Assumption W(iv) implies \( \{ B_n^{-1} (\lambda) \} \) are uniformly bounded in both row and column sums uniformly in a neighborhood of \( \lambda_0 \) as explicitly given in Assumption W(v).

To proceed, let \( \Omega_0, \Omega_0^* \) and \( \Omega_0^\dagger \) be respectively \( \Omega, \Omega^* \) and \( \Omega^\dagger \) evaluated at \( \delta_0 \).

4.2 Asymptotic Distributions for Random Effects Model

For the random effects model, we need to supplement the generic Assumption G with additional assumptions on the individual-specific effects \( \mu \), and the regressors. In particular, we make the following assumptions.

**Assumption R** (i) The individual effects \( \mu_i \) are i.i.d. with \( E(\mu_i) = 0 \), \( \text{Var}(\mu_i) = \sigma^2_\mu \), and \( E|\mu_i|^{4+\epsilon_0} < \infty \) for some \( \epsilon_0 > 0 \). (ii) \( \mu_i \) 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) is 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: \( y_{i0} \) is exogenous**

To derive the consistency of the QML estimator, we need the parameters of interest to be identifiable. For this purpose, define \( \mathcal{L}_{c}^{*} (\delta) = \max_{\theta, \sigma_v^2} E \left[ \mathcal{L}^{*} (\theta, \sigma_v^2, \delta) \right] \), where we suppress the dependence of \( \mathcal{L}_{c}^{*} (\delta) \) on \( n \). The optimal solution to \( \max_{\theta, \sigma_v^2} E \left[ \mathcal{L}^{*} (\theta, \sigma_v^2, \delta) \right] \) is given by

\[
\hat{\theta} (\delta) = \left[ E \left( \tilde{X}' \Omega^{-1} \tilde{X} \right) \right]^{-1} E \left( \tilde{X}' \Omega^{-1} Y \right) = \theta_0 \tag{4.1}
\]

and

\[
\hat{\sigma}_v^2 (\delta) = \frac{1}{nT} E \left[ u (\theta_0)' \Omega^{-1} u (\theta_0) \right] = \frac{\sigma^2_u}{nT} \text{tr} (\Omega^{-1} \Omega_0) \tag{4.2}
\]

where the second equality in (4.1) follows from the fact that \( Y = \tilde{X} \theta_0 + u \) and \( E(\tilde{X}' \Omega^{-1} u) = 0 \) by Lemma B.9. Consequently, we have

\[
\mathcal{L}_{c}^{*} (\delta) = -\frac{nT}{2} (\log(2\pi) + 1) - \frac{nT}{2} \log [\hat{\sigma}_v^2 (\delta)] - \frac{1}{2} \log |\Omega| \tag{4.3}
\]
We impose the following identification condition.

**Assumption R** (iv) $\hat{\mathcal{L}}^r(\delta) \equiv \lim_{n \to \infty} (nT)^{-1} L^r_e(\delta)$ is uniquely maximized at $\delta_0$.

The following theorem establishes the consistency of QMLE for the random effects model with exogenous initial observations.

**Theorem 4.1** Under Assumptions G, W, and R(i)-(iv), if the initial observations $y_{i0}$ are exogenously given, then $\tilde{\xi} \xrightarrow{p} \xi_0$.

To derive the asymptotic distribution of $\xi$, we need to make a Taylor expansion of $\partial \mathcal{L}^r(\xi) / \partial \xi = 0$ at $\xi_0$. The first order derivatives of $\mathcal{L}^r(\xi)$ are

$$
\frac{\partial \mathcal{L}^r(\xi)}{\partial \theta} = \frac{1}{\sigma_v^2} \tilde{X} \Omega^{-1} u(\theta),
$$

$$
\frac{\partial \mathcal{L}^r(\xi)}{\partial \sigma_v^2} = \frac{1}{2\sigma_v^4} u(\theta)' \Omega^{-1} u(\theta) - \frac{nT}{2\sigma_v^2},
$$

$$
\frac{\partial \mathcal{L}^r(\xi)}{\partial \lambda} = \frac{1}{2\sigma_v^2} u(\theta)' \Omega^{-1} (J_T \otimes A) \Omega^{-1} u(\theta) - \frac{1}{2} tr \left( \Omega^{-1} (J_T \otimes A) \right),
$$

$$
\frac{\partial \mathcal{L}^r(\xi)}{\partial \phi_\mu} = \frac{1}{2\sigma_v^2} u(\theta)' \Omega^{-1} (J_T \otimes I_n) \Omega^{-1} u(\theta) - \frac{1}{2} tr \left( \Omega^{-1} (J_T \otimes I_n) \right),
$$

where $A = \partial (B' B)^{-1} / \partial \lambda = (B' B)^{-1} (W' B + B' W) (B' B)^{-1}$. At $\xi = \xi_0$, these are linear and quadratic functions of $u \equiv u(\theta_0)$. Note that the elements in $u$ are not independent and the components in $\tilde{X}$ are random, thus the central limit theorem (CLT) for linear and quadratic forms in Kelejian and Prucha (2001) cannot be directly applied to $u$. We need to plug $u = (i_T \otimes I_n) \mu + (I_T \otimes B^{-1}) v$ into $\partial \mathcal{L}^r(\xi_0) / \partial \xi$ and apply the CLT to linear and quadratic functions of $\mu$ and $v$ separately.

Let $H^r_n(\xi) = \partial^2 \mathcal{L}^r(\xi) / (\partial \xi \partial \xi')$ and $\Gamma^r_n(\xi) = E(\partial \mathcal{L}^r(\xi) / \partial \xi) \partial \mathcal{L}^r(\xi) / (\partial \xi')$. Then $H^r_n(\xi)$ is the Hessian matrix and $\Gamma^r_n(\xi_0)$ is the variance matrix of $\partial \mathcal{L}^r(\xi_0) / \partial \xi$, both of which are reported in Appendix A. Using some of the lemmas in Appendix B (Lemmas B.3-B.6 in particular), we can verify that for normally distributed individual-specific effects $\mu_i$ and error terms $v_{it}$, $\Gamma^r_n(\xi_0) = -E \left( H^r_n(\xi_0) \right)$.

**Theorem 4.2** Under Assumptions G, W, and R(i)-(iv), if the initial observations $y_{i0}$ are exogenously given, then $\sqrt{nT} (\xi - \xi_0) \xrightarrow{d} N(0, H^r \Gamma_r H^r^{-1})$, where $H_r = \lim_{n \to \infty} (nT)^{-1} H^r_n(\xi_0)$ and $\Gamma_r = \lim_{n \to \infty} (nT)^{-1} \Gamma^r_n(\xi_0)$.
As in Lee (2004), the asymptotic results in Theorem 4.2 are valid regardless of whether the sequence \( \{h_n\} \) is bounded or divergent. Even though not reported here, the matrices \( \Gamma_r \) and \( H_r \) can be simplified if \( h_n \to \infty \) as \( n \to \infty \). When both \( \mu_i \) and \( \nu_i \) are normally distributed, the asymptotic variance-covariance matrix reduces to \( H_r^{-1} \).

In practice, we need to estimate the variance-covariance matrix. By the lemmas in Appendix B, we can show that \( \hat{H}_{r,n} \equiv (nT)^{-1} H_n (\zeta) \) is a consistent estimate of \( H_r \). The consistent estimate of \( \Gamma_r \) can be obtained by its sample analogue. To conserve space, we omit the details here. Similar remarks hold for the other cases.

Case II: \( y_{i0} \) is endogenous

In this case, define \( \mathcal{L}_{r*} (\delta) = \max_{\theta, \sigma^2} E \left[ \mathcal{L}_r (\theta, \sigma^2, \delta) \right] \), where we suppress the dependence of \( \mathcal{L}_{r*} (\delta) \) on \( n \). The optimal solution to \( \max_{\theta, \sigma^2} E \left[ \mathcal{L}_r (\theta, \sigma^2, \delta) \right] \) is now given by

\[
\tilde{\theta} (\delta) = \left[ E \left( X^* \Omega^{-1} (\delta) X^* \right) \right]^{-1} E \left( X^* \Omega^{-1} (\delta) Y^* (\rho) \right) = \theta_0 \tag{4.4}
\]

and

\[
\tilde{\sigma}_{*}^2 (\delta) = \frac{1}{n(T+1)} E \left[ u^* (\theta_0)^T \Omega^{-1} (\delta) u^* (\theta_0) \right] = \frac{\sigma^2_{*0}}{n(T+1)} \text{tr} \left( \Omega^{-1} (\delta) \Omega_0^* \right). \tag{4.5}
\]

Let \( Y_{*1} = (0_{1 \times n}, Y_{*1})' \). The second equality in (4.4) follows from the fact that \( Y^* (\rho) = X^* \theta_0 + u^* (\theta_0) + (\rho_0 - \rho) Y_{*1} \), and \( E \left( X^* \Omega^{-1} (\delta) u^* (\theta_0) + (\rho_0 - \rho) Y_{*1} \right) = 0 \) by arguments similar to those used in the proof of Lemma B.13. Consequently, we have

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

We make the following identification assumption.

**Assumption R** (iv*) \( \tilde{\mathcal{L}}_{r*} (\delta) \equiv \lim_{n \to \infty} (nT)^{-1} \mathcal{L}_{r*} (\delta) \) is uniquely maximized at \( \delta_0 \).

The following theorem establishes the consistency of QMLE for the random effects model with endogenous initial observations.

**Theorem 4.3** Under Assumptions G, W, R(i)-(iii) and R(iv*), if the initial observations \( y_{i0} \) are endogenously given, then \( \hat{\zeta} \to^P \hat{\zeta}_0 \).

17
The score functions are given below:

\[
\frac{\partial L^r}{\partial \theta} = \frac{1}{\sigma^2} X^* \Omega^{r-1} u^* (\theta),
\]

\[
\frac{\partial L^r}{\partial \sigma^2} = \frac{1}{2\sigma^2} u^* (\theta)' \Omega^{r-1} u^* (\theta) - \frac{n(T + 1)}{2\sigma^2},
\]

\[
\frac{\partial L^r}{\partial \rho} = \frac{1}{2\sigma^2} u^* (\theta)' \Omega^{r-1} (I_T \otimes A) \Omega^{r-1} u^* (\theta) - \frac{1}{2} \text{tr} \left[ \Omega^{r-1} (I_T \otimes A) \right],
\]

\[
\frac{\partial L^r}{\partial \lambda} = \frac{1}{2\sigma^2} u^* (\theta)' \Omega^{r-1} (J_T^* \otimes I_n) \Omega^{r-1} u^* (\theta) - \frac{1}{2} \text{tr} \left[ \Omega^{r-1} (J_T^* \otimes I_n) \right],
\]

\[
\frac{\partial L^r}{\partial \phi} = \frac{1}{2\sigma^2} u^* (\theta)' \Omega^{r-1} (K_T^* \otimes I_n) \Omega^{r-1} u^* (\theta) - \frac{1}{2} \text{tr} \left[ \Omega^{r-1} (K_T^* \otimes I_n) \right],
\]

where \( \Omega^*_r = \partial \Omega^* / \partial \rho \).

\[
I_T^* = \begin{pmatrix} \frac{1}{\sigma^2} & 0_{1 \times T} \\ 0_{T \times 1} & I_T \end{pmatrix}, \quad J_T^* = \begin{pmatrix} 0 & \iota_T' \\ \iota_T & J_T \end{pmatrix}, \quad \text{and} \quad K_T^* = \begin{pmatrix} 1 & 0_{1 \times T} \\ 0_{T \times 1} & 0_{T \times T} \end{pmatrix}.
\]

Let \( H''_{n}(\zeta) = (\partial^2 L^r (\zeta) / (\partial \delta^2)) \) and \( \Gamma''_{n}(\zeta) = E(\partial L^r (\zeta) / (\partial \delta) \partial L^r (\zeta) / (\partial \delta^2)) \). We report the Hessian matrix \( H''_{n}(\zeta) \) and \( \Gamma''_{n}(\zeta_0) \) in Appendix A. We now state the asymptotic normality result.

**Theorem 4.4** Under Assumptions G, W, R(i)-(iii) and R(iv), if the initial observations \( y_{t0} \) satisfy (3.7), (3.11) or (3.12), then \( \sqrt{nT}(\zeta - \zeta_0) \to N(0, H_{rr}^{-1} \Gamma_r H_{rr}^{-1}) \), where \( H_{rr} = \lim_{n \to \infty} (nT)^{-1} H''_{n}(\zeta_0) \) and \( \Gamma_r = \lim_{n \to \infty} (nT)^{-1} \Gamma''_{n}(\zeta_0) \).

### 4.3 Asymptotic Distribution for the Fixed Effects Model

For the fixed effect model, we need to supplement the general Assumption G 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 \( i = 1, \ldots, n \). (ii) All elements in \( \{\Delta v_{it}, \Delta x_{it}\} \) have \( 4 + \epsilon_0 \) moments for some \( \epsilon_0 > 0 \).

Define \( L^r(\delta) = \max_{\theta, \sigma^2} E \left[ L^r (\theta, \sigma^2, \delta) \right] \), where we suppress the dependence of \( L^r(\delta) \) on \( n \).

The optimal solution to \( \max_{\theta, \sigma^2} E \left[ L^r (\theta, \sigma^2, \delta) \right] \) is now given by

\[
\tilde{\theta}(\delta) = \left[ E(\Delta X' \Omega^{\frac{1}{2}} \Delta X) \right]^{-1} E(\Delta X' \Omega^{\frac{1}{2}} \Delta Y) = \theta_0 \quad (4.7)
\]
and
\[ \tilde{\sigma}_c^2 (\delta) = \frac{1}{nT} E \left[ \Delta u (\theta_0)' \Omega^{1-1} \Delta u (\theta_0) \right] = \frac{\sigma^2}{nT} \text{tr} \left( \Omega^{1-1} \Omega_0^b \right), \] (4.8)
where the second equality in (4.7) follows from the fact that \( \Delta Y = \Delta X \theta_0 + \Delta u (\theta_0) \) and \( E (\Delta X' \Omega^{1-1} \Delta u (\theta_0)) = 0 \) by Lemma B.16. Consequently, we have
\[ \mathcal{L}_{c^*}^* (\delta) = -\frac{nT}{2} \log(2\pi) - \frac{nT}{2} \log [\tilde{\sigma}_c^2 (\delta)] - \frac{1}{2} \log |\Omega^f|. \] (4.9)

The following identification condition is needed for our consistency result.

**Assumption F** (iii) \( \hat{\mathcal{L}}^f (\delta) \equiv \lim_{n \to \infty} (nT)^{-1} \mathcal{L}_{c^*}^* (\delta) \) is uniquely maximized at \( \delta_0 \).

**Theorem 4.5** Under Assumptions G, F, and W, if the initial observations satisfy (3.22), (3.25), or (3.26), then \( \tilde{\gamma} \overset{p}{\to} \zeta_0 \).

The first order derivatives of \( \mathcal{L}^f (\zeta) \) are
\[
\begin{align*}
\frac{\partial \mathcal{L}^f (\zeta)}{\partial \theta} &= \frac{1}{\sigma^2} \Delta X' \Omega^{1-1} \Delta u (\theta), \\
\frac{\partial \mathcal{L}^f (\zeta)}{\partial \sigma^2} &= \frac{1}{2\sigma^2} \Delta u (\theta)' \Omega^{1-1} \Delta u (\theta) - \frac{nT}{2\sigma^2}, \\
\frac{\partial \mathcal{L}^f (\zeta)}{\partial \lambda} &= \frac{1}{2\sigma^2} \Delta u (\theta)' \Omega^{1-1} \Omega_\lambda^{1-1} \Delta u (\theta) - \frac{1}{2} \text{tr} \left[ \Omega^{1-1} \Omega_\lambda^1 \right], \\
\frac{\partial \mathcal{L}^f (\zeta)}{\partial c_m} &= \frac{1}{2\sigma^2} \Delta u (\theta)' \Omega^{1-1} \Omega_m^{1-1} \Delta u (\theta) - \frac{1}{2} \text{tr} \left[ \Omega^{1-1} \Omega_m^1 \right], \\
\frac{\partial \mathcal{L}^f (\zeta)}{\partial \phi_c} &= \frac{1}{2\sigma^2} \Delta u (\theta)' \Omega^{1-1} \Omega_{\phi_c}^{1-1} \Delta u (\theta) - \frac{1}{2} \text{tr} \left[ \Omega^{1-1} \Omega_{\phi_c}^1 \right],
\end{align*}
\]
where
\[
\begin{align*}
\Omega_\lambda^1 &\equiv \frac{\partial \Omega^1}{\partial \lambda} = (I_T \otimes B^{-1}W B^{-1}) H_E (I_T \otimes B^{-1}) + (I_T \otimes B^{-1}) H_E \left( I_T \otimes (B^{-1}W B^{-1})' \right) + (I_T \otimes B^{-1}) H_{E,\lambda} (I_T \otimes B^{-1}), \\
\Omega_m^1 &\equiv \frac{\partial \Omega^1}{\partial c_m} = (I_T \otimes B^{-1}) H_{E,c_m} (I_T \otimes B^{-1}), \\
\Omega_{\phi_c}^1 &\equiv \frac{\partial \Omega^1}{\partial \phi_c} = (I_T \otimes B^{-1}) H_{E,\phi_c} (I_T \otimes B^{-1}),
\end{align*}
\]
and \( H_{E,\lambda}, H_{E,c_m} \) and \( H_{E,\phi_c} \) are the derivatives of \( H_E \) with respect to \( \lambda, c_m \) and \( \phi_c \), respectively.

Let \( H_n^f (\zeta) = \partial^2 \mathcal{L}^f (\zeta) / (\partial \zeta \partial \zeta' \zeta) \) and \( \Gamma_n^f (\zeta) = E (\partial \mathcal{L}^f (\zeta) / (\partial \zeta) \partial \mathcal{L}^f (\zeta) / (\partial \zeta')) \). As before we report the Hessian matrix \( H_n^f (\zeta_0) \) and \( \Gamma_n^f (\zeta_0) \) in Appendix A. We now state the asymptotic normality result.

**Theorem 4.6** Under Assumptions G, F and W, if the initial observations satisfy (3.22), (3.25), or (3.26), then \( \sqrt{nT} (\zeta - \zeta_0) \overset{d}{\to} N \left( 0, H_f^{-1} \Gamma_f H_f^{-1} \right) \), where \( H_f = \lim_{n \to \infty} (nT)^{-1} H_n^f (\zeta_0) \) and \( \Gamma_r = \lim_{n \to \infty} (nT)^{-1} \Gamma_n^f (\zeta_0) \).
5 Finite Sample Properties of the QMLEs

In this section, we investigate the performance of the QMLEs when sample sizes are finite. In particular, we investigate under the random effects model the consequences of treating the initial observations as endogenous when they are in fact exogenous, and vice versa. In the case of fixed effects model, we pay attention to whether there is a difference in parameter estimates when the fixed effects are uncorrelated with the regressor(s) versus when they are correlated with the regressor(s).

We use the following data generating process (DGP):

\[
\begin{align*}
y_t &= \rho y_{t-1} + \beta_0 \xi_t + x_t \beta_1 + x_t \gamma + u_t \\
u_t &= \mu + \varepsilon_t \\
\varepsilon_t &= \lambda W_n \varepsilon_t + \nu_t
\end{align*}
\]

where \(y_t, y_{t-1}, x_t,\) and \(z\) are all \(n \times 1\) vectors. The elements of \(x_t\) are randomly generated from \(N(0, 4)\), and the elements of \(z\) are randomly generated from Bernoulli(0.5). The spatial weight matrix is generated according to Rook 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 \(\beta_0 = 5, \beta_1 = 1, \gamma = 1, \sigma_\mu = 0.5, \sigma_v = 0.5, \rho \in \{0.25, 0.50, 0.75\},\) and \(\lambda \in \{0.25, 0.50, 0.75\}.\)

Each set of Monte Carlo results (corresponding to a combination of the \(\rho\) and \(\lambda\) values) is based on 1000 samples.

Table 1 summarizes the results for the random effects model and Table 2 reports the results for the fixed effects model. From the upper panel of Table 1, we see that when \(y_0\) is exogenous and is correctly treated as exogenous, the QMLEs perform quite well. However, when \(y_0\) is exogenous but treated incorrectly as endogenous, the QMLEs can be quite biased with large RMSEs, in particular in the cases that \(\rho\) and \(\lambda\) are large. When \(y_0\) is endogenous and is correctly treated as endogenous, the QMLEs also perform quite well. When \(y_0\) is endogenous but treated incorrectly as exogenous (lower panel of Table 1), the QMLEs perform reasonably well, except for case of \(\gamma\) and \(\sigma_\mu\) where the QMLEs are slightly biased. From Table 2, we see that QMLEs under the fixed effects perform generally well, whether the fixed effects are correlated with the regressor or not.

REFERENCES


Appendix

A The Hessian and Information Matrices

In this appendix, we give the Hessian and information matrices for each scenario. For any random variable \( a \) with zero mean and fourth moments, let \( \kappa_a = E(a^4) - 3E(a^2) \).

A.1 Random Effects Model with Fixed Initial Observations

Recall \( u(\theta) = Y - \bar{X}\theta = Y - \rho Y_{-1} - X\beta - Z\gamma \), \( A = (B'B)^{-1} (W'W + B'W)(B'B)^{-1} \), and \( A \equiv \partial A/\partial \lambda = 2(B'B)^{-1} [W'W + B'W] A - W'W \). To conserve space, let \( P_1 = \Omega^{-1} (I_T \otimes A) \Omega^{-1} \), \( P_2 = \Omega^{-1} (I_T \otimes I_n) \Omega^{-1} \), \( P_3 = \Omega^{-1} (I_T \otimes \bar{A}) \Omega^{-1} \), and \( P_4 = P_1 P_2 P_3 - P_3 \). Then the Hessian matrix

\[
H_{r,n} (\varsigma) = \partial^2 \mathcal{L}^\tau (\varsigma) / (\partial \varsigma \partial \varsigma')
\]

has elements

\[
\begin{align*}
\frac{\partial^2 \mathcal{L}^\tau (\varsigma)}{\partial \theta \partial \mu} &= -\frac{1}{\sigma^2} \bar{X}' \Omega^{-1} \bar{X}, \\
\frac{\partial^2 \mathcal{L}^\tau (\varsigma)}{\partial \sigma^2 \partial \mu} &= \frac{1}{\sigma^2} \bar{X}' P_1 u (\theta_1), \\
\frac{\partial^2 \mathcal{L}^\tau (\varsigma)}{\partial \sigma^2 \partial \sigma} &= -\frac{1}{\sigma^4} \bar{X}' P_1 u (\theta_1), \\
\frac{\partial^2 \mathcal{L}^\tau (\varsigma)}{\partial \sigma \partial \mu} &= 0, \\
\frac{\partial^2 \mathcal{L}^\tau (\varsigma)}{\partial \sigma \partial \sigma} &= -\frac{1}{\sigma^2} \bar{X}' P_1 u (\theta_1), \\
\frac{\partial^2 \mathcal{L}^\tau (\varsigma)}{\partial \mu \partial \mu} &= -\frac{1}{\sigma^2} \bar{X}' P_1 u (\theta_1).
\end{align*}
\]

For an \( nT \times nT \) matrix \( q_n \), let \( G_{q_{n,1}} = (I_T \otimes I_n) q_n (I_T \otimes I_n) \), \( G_{q_{n,2}} = (I_T \otimes B^{-1}) q_n (I_T \otimes B^{-1}) \), and \( G_{q_{n,3}} = (I_T \otimes I_n) q_n (I_T \otimes B^{-1}) \). The elements of \( \Gamma_{r,n} (\varsigma_0) \equiv E \{ [\partial \mathcal{L}^\tau (\varsigma_0) / \partial \varsigma] [\partial \mathcal{L}^\tau (\varsigma_0) / \partial \varsigma'] \} \) are given by:

\[
\begin{align*}
\Gamma_{r,\theta \theta} &= \frac{1}{\sigma^2} E(\bar{X}' \Omega^{-1} \bar{X}), \\
\Gamma_{r,\theta \sigma^2} &= \frac{1}{\sigma^4} E(\bar{X}' \Omega^{-1} \bar{X}' \Omega^{-1} u), \\
\Gamma_{r,\theta \lambda} &= \frac{1}{\sigma^2} E(\bar{X}' \Omega^{-1} u' \Omega^{-1} (I_T \otimes A) \Omega^{-1} u), \\
\Gamma_{r,\theta \phi_n} &= \frac{1}{\sigma^2} E(\bar{X}' \Omega^{-1} u' \Omega^{-1} (J_T \otimes I_n) \Omega^{-1} u), \\
\Gamma_{r,\theta \sigma^2 \phi_n} &= \frac{1}{2 \sigma^2} E(\bar{X}' \Omega^{-1} u' \Omega^{-1} (J_T \otimes I_n) \Omega^{-1} u) \\
&= \frac{1}{2 \sigma^2} \left\{ \kappa_\mu \sum_{i=1}^n G_{\Omega^{-1},1ii} + \kappa_\sigma \sum_{i=1}^n G_{\Omega^{-1},2ii} \right\}, \\
\Gamma_{r,\theta \lambda \phi_n} &= \frac{1}{2 \sigma^2} E(\bar{X}' \Omega^{-1} (P_1 \Omega) + \frac{1}{\sigma^2} \left\{ \kappa_\mu \sum_{i=1}^n G_{\Omega^{-1},1ii} G_{P_{1},1ii} + \kappa_\nu \sum_{i=1}^n G_{\Omega^{-1},2ii} G_{P_{2},2ii} \right\}, \\
\Gamma_{r,\theta \lambda \phi_n} &= \frac{1}{2 \sigma^2} \left\{ \kappa_\mu \sum_{i=1}^n G_{\Omega^{-1},1ii} G_{P_{1},1ii} + \kappa_\nu \sum_{i=1}^n G_{\Omega^{-1},2ii} G_{P_{2},2ii} \right\}, \\
\Gamma_{r,\phi_n} &= \frac{1}{2 \sigma^2} \left\{ \kappa_\mu \sum_{i=1}^n G_{P_{1},1ii} + \kappa_\nu \sum_{i=1}^n G_{P_{2},2ii} \right\}.
\end{align*}
\]

The expression for \( \Gamma_{r,\theta \sigma^2 \phi_n} \) and \( \Gamma_{r,\theta \lambda \phi_n} \) can be expressed out by using Lemma B.18.
A.2 Random Effects Model with Endogenous Initial Observations

To conserve space, let $Y_{I-1} = (0_{1 \times n}, Y'_{I-1})'$, $P_0^* = \Omega^{-1} \Omega_{\rho}^\ast \Omega^{-1}$, $P_1^* = \Omega^{-1} (J_1^* \otimes A) \Omega^{-1}$, $P_2^* = \Omega^{-1} (J_2^* \otimes I_n) \Omega^{-1}$, $P_3^* = \Omega_{\phi \ast}^{-1} (K_2^* \otimes I_n) \Omega^{-1}$, $P_4^* = \Omega^{-1} (I_T \otimes A) \Omega^{-1}$, and $P_5^* = P_1^* P_3^* - P_4^* P_2^*$.

Write $u^* = u^* (\Theta)$. Then the Hessian matrix $H_{rr, \Theta} (\cdot) = \partial^2 L_{rr} (\cdot) / (\partial \Theta \partial \Theta')$ has elements

\[
\frac{\partial^2 L_{rr} (\cdot)}{\partial \Theta \partial \Theta'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} X^\ast, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \Theta \partial \rho} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} u^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \Theta \partial \lambda} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \rho \partial \rho} = -\frac{1}{\sigma} X^\ast P_1^* u^*, 
\]

\[
\frac{\partial^2 L_{rr} (\cdot)}{\partial \rho \partial \lambda} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \rho \partial \lambda'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q'^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \lambda \partial \lambda'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \lambda \partial \lambda'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q'^*, 
\]

\[
\frac{\partial^2 L_{rr} (\cdot)}{\partial \rho \partial \phi} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \rho \partial \phi'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q'^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \phi \partial \phi'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \phi \partial \phi'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q'^*, 
\]

\[
\frac{\partial^2 L_{rr} (\cdot)}{\partial \rho \partial \phi} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \rho \partial \phi'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q'^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \phi \partial \phi'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \phi \partial \phi'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q'^*, 
\]

\[
\frac{\partial^2 L_{rr} (\cdot)}{\partial \rho \partial \phi} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \rho \partial \phi'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q'^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \phi \partial \phi'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q^*, \quad \frac{\partial^2 L_{rr} (\cdot)}{\partial \phi \partial \phi'} = -\frac{1}{\sigma} X^\ast \Omega_{\phi \ast}^{-1} q'^*, 
\]

where e.g., $\Omega_{\rho \lambda} = \partial \Theta^\ast / \partial \rho \partial \lambda$, and $\Omega_{\rho \rho \lambda} = \partial^2 \Theta^\ast / \partial \rho \partial \lambda \partial \lambda$.

$\Gamma_{rr, \Theta} (\cdot) = E [\{ \partial^2 L_{rr} (\cdot) / \partial \Theta \} \{ \partial^2 L_{rr} (\cdot) / \partial \Theta' \}]$ has elements

\[
\Gamma_{rr, \Theta} = \frac{1}{\sigma} E (X^\ast \Omega_{\phi \ast}^{-1} X^\ast), 
\]

\[
\Gamma_{rr, \Theta, \rho} = \frac{1}{\sigma} E (X^\ast \Omega_{\phi \ast}^{-1} u^* \langle \ast \rangle \Omega_{\phi \ast}^{-1} u^*), 
\]

\[
\Gamma_{rr, \Theta, \phi} = \frac{1}{\sigma} E (X^\ast \Omega_{\phi \ast}^{-1} u^* \langle \ast \rangle \Omega_{\phi \ast}^{-1} u^* - \frac{1}{2} tr (\Omega_{\phi \ast}^{-1} \Omega_{\phi \ast}^{-1} u^*)), 
\]

\[
\Gamma_{rr, \Theta, \phi, \rho} = \frac{1}{\sigma} E (X^\ast \Omega_{\phi \ast}^{-1} u^* \langle \ast \rangle \Omega_{\phi \ast}^{-1} u^* - \frac{1}{2} tr (\Omega_{\phi \ast}^{-1} \Omega_{\phi \ast}^{-1} u^*)), 
\]

\[
\Gamma_{rr, \Theta, \phi, \phi} = \frac{1}{\sigma} E (X^\ast \Omega_{\phi \ast}^{-1} u^* \langle \ast \rangle \Omega_{\phi \ast}^{-1} u^* - \frac{1}{2} tr (\Omega_{\phi \ast}^{-1} \Omega_{\phi \ast}^{-1} u^*)), 
\]

\[
\Gamma_{rr, \Theta, \phi, \phi, \rho} = \frac{1}{\sigma} E (X^\ast \Omega_{\phi \ast}^{-1} u^* \langle \ast \rangle \Omega_{\phi \ast}^{-1} u^* - \frac{1}{2} tr (\Omega_{\phi \ast}^{-1} \Omega_{\phi \ast}^{-1} u^*)), 
\]

\[
\Gamma_{rr, \Theta, \phi, \phi, \phi} = \frac{1}{\sigma} E (X^\ast \Omega_{\phi \ast}^{-1} u^* \langle \ast \rangle \Omega_{\phi \ast}^{-1} u^* - \frac{1}{2} tr (\Omega_{\phi \ast}^{-1} \Omega_{\phi \ast}^{-1} u^*)), 
\]
\[ \Gamma_{rr,\sigma^2_\lambda} = \frac{1}{4\sigma^2} E \left( u^{1T} \Omega^{\lambda-1} u^{1}\right) - \frac{n(T+1)}{4\sigma^2} \text{tr} \left( P_1^* \Omega^* \right), \]
\[ \Gamma_{rr,\sigma^2_\phi_0} = \frac{1}{\sigma^2_0} E \left( u^{1T} \Omega^{\phi_0-1} u^{1}\right) - \frac{n(T+1)}{4\sigma^2} \text{tr} \left( P_2^* \Omega^* \right), \]
\[ \Gamma_{rr,\sigma^2_\phi_\zeta} = \frac{1}{\sigma^2_\zeta} E \left( u^{1T} \Omega^{\phi_\zeta-1} u^{1}\right) - \frac{n(T+1)}{4\sigma^2} \text{tr} \left( P_3^* \Omega^* \right), \]
\[ \Gamma_{rr,\rho} = \frac{1}{\sigma^2_\rho} E \left( \frac{1}{2} \frac{Y^{1T} \Omega^{\phi_\zeta-1} u^{1} + \frac{1}{2\sigma^2_\zeta} u^{T} P_0^* u^{*} - \frac{1}{2} \text{tr} \left( P_0^* \Omega^* \right)}{2} \right) \left( \frac{Y^{1T} \Omega^{\phi_0-1} u^{1} + \frac{1}{2\sigma^2_0} u^{T} P_0^* u^{*} - \frac{1}{2} \text{tr} \left( P_0^* \Omega^* \right)}{2} \right), \]
\[ \Gamma_{rr,\lambda} = \frac{1}{\sigma^2_\lambda} E \left( \frac{1}{2} \frac{Y^{1T} \Omega^{\phi_0-1} u^{1} + \frac{1}{2\sigma^2_0} u^{T} P_0^* u^{*} - \frac{1}{2} \text{tr} \left( P_0^* \Omega^* \right)}{2} \right) \left( \frac{Y^{1T} \Omega^{\phi_\zeta-1} u^{1} + \frac{1}{2\sigma^2_\zeta} u^{T} P_0^* u^{*} - \frac{1}{2} \text{tr} \left( P_0^* \Omega^* \right)}{2} \right). \]
\[ A.3 \text{ Fixed Effects Model} \]

To conserve space, let \( P_\alpha^1 = \Omega^1 \Omega\alpha^1 \Omega^{-1} \), for \( a = \lambda, c_m, \phi_\epsilon \), where \( \Omega^1_a = \partial \Omega^1 / \partial a \). Let \( \Omega_{aa} = \partial^2 \Omega^1 / (\partial \alpha \partial \alpha) \). Write \( \Delta u = \Delta u (\theta) \). Then the Hessian matrix \( H_{f,n} (s) = \partial^2 L^f (s) / (\partial \zeta \partial \zeta') \) has elements

\[ \frac{\partial^2 L^f (s)}{\partial \zeta \partial \alpha} = -\frac{1}{\sigma^2_\alpha} \Delta X' \Omega^{\lambda-1} \Delta X, \quad \frac{\partial^2 L^f (s)}{\partial \alpha \partial \zeta} = -\frac{1}{\sigma^2_\zeta} \Delta X' \Omega^{\phi_\epsilon-1} \Delta X, \]
\[ \frac{\partial^2 L^f (s)}{\partial \alpha \partial \alpha} = -\frac{1}{\sigma^2_\alpha} \Delta X' P_{\alpha} \Delta u, \quad \frac{\partial^2 L^f (s)}{\partial \zeta \partial \zeta} = -\frac{1}{\sigma^2_\zeta} \Delta X' P_{\zeta} \Delta u, \]
\[ \frac{\partial^2 L^f (s)}{\partial \alpha \partial \zeta} = -\frac{1}{\sigma^2_\zeta} \Delta X' P_{\phi_\epsilon} \Delta u, \quad \frac{\partial^2 L^f (s)}{\partial \zeta \partial \alpha} = -\frac{1}{\sigma^2_\alpha} \Delta u P_{\alpha} \Delta u, \]
\[ \frac{\partial^2 L^f (s)}{\partial \alpha \partial \alpha} = -\frac{1}{\sigma^2_\alpha} \Delta u P_{\phi_\epsilon} \Delta u, \]
\[ \frac{\partial^2 L^f (s)}{\partial \alpha \partial \alpha} = \frac{1}{2 \sigma^2_\alpha} \text{tr} \left( P_0^* \Omega^{\phi_\zeta-1} \Delta u + \frac{1}{2\sigma^2_\zeta} \Delta u P_0^* \Omega^{\phi_0-1} \Delta u \right), \]
\[ \frac{\partial^2 L^f (s)}{\partial \zeta \partial \alpha} = \frac{1}{2 \sigma^2_\alpha} \text{tr} \left( P_0^* \Omega^{\phi_0-1} \Delta u + \frac{1}{2\sigma^2_0} \Delta u P_0^* \Omega^{\phi_\zeta-1} \Delta u \right), \]
\[ \frac{\partial^2 L^f (s)}{\partial \alpha \partial \alpha} = \frac{1}{2 \sigma^2_\alpha} \text{tr} \left( P_0^* \Omega^{\phi_\epsilon-1} \Delta u + \frac{1}{2\sigma^2_\epsilon} \Delta u P_0^* \Omega^{\phi_0-1} \Delta u \right), \]
\[ \frac{\partial^2 L^f (s)}{\partial \zeta \partial \alpha} = \frac{1}{2 \sigma^2_\alpha} \text{tr} \left( P_0^* \Omega^{\phi_0-1} \Delta u + \frac{1}{2\sigma^2_0} \Delta u P_0^* \Omega^{\phi_\epsilon-1} \Delta u \right), \]
\[ \Gamma_{f,n} (s) = E \left[ \left\{ \partial^2 L^f (s) / \partial \alpha \partial \alpha \right\} \left\{ \partial^2 L^f (s) / \partial \alpha \partial \alpha \right\} \right], \]
\[ \Gamma_{f,\theta \theta} = \frac{1}{\sigma^2_\theta} E (\Delta X' \Omega^{\lambda-1} \Delta X), \]
\[ \Gamma_{f,\theta} = \frac{1}{2\sigma^2} E(\Delta X'\Omega^{-1} \Delta u' \Omega^{-1} \Delta u), \]
\[ \Gamma_{f,\alpha} = \frac{1}{2\sigma^2} E(\Delta X'\Omega^{-1} \Delta u' \Omega^{-1} \Omega^t_{\alpha} \Omega^{-1} \Delta u), \]
\[ \Gamma_{f,\theta B} = \frac{1}{2\sigma^2} E(\Delta X'\Omega^{-1} \Delta u' \Omega^{-1} \Omega^t_{\theta} \Omega^{-1} \Delta u), \]
\[ \Gamma_{f,\theta \phi} = \frac{1}{2\sigma^2} E(\Delta X'\Omega^{-1} \Delta u' \Omega^{-1} \Omega^t_{\phi} \Omega^{-1} \Delta u), \]
\[ \Gamma_{f,\theta \sigma \theta} = \frac{1}{2\sigma^2} E((\Delta u' \Omega^{-1} \Delta u)^2) - \frac{n^2}{4\sigma^4}. \]

By the three evident facts and Lemma B.1, we have the following lemma.

**Lemma B.1** $B' B, (B' B)^{-1}$, $\Omega$, $\Omega^{-1}$, $\Omega^*$, $\Omega^*$, $\Omega^t$, $\Omega^{-1}$, $A$, and $\hat{A}$ are all uniformly bounded in both row and column sums, where recall $A = (B' B)^{-1} (W' B + B' W) (B' B)^{-1}$ and $\hat{A} = 2 (B' B)^{-1} [(W' B + B' W) A - W' W]$. 

By the three evident facts and Lemma B.1, we have the following lemma.

**B Some Useful Lemmas**

We prove some lemmas that are used in the proof of the main theorems in the text. For the ease of notation, we assume that both $x_{it}$ and $z_i$ are scalar random variables ($p = 1, q = 1$) in this Appendix.

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 certain sequence $o_n$. Frequently we will use three evident facts (see, e.g., Koekoek and Prucha, 1999, Lee, 2002): (1) $P_n Q_n$ is also uniformly bounded in both row and column sums; (2) any $(i, j)$ elements $P_n$ are uniformly bounded in $i$ and $j$ and $tr(P_n) = O(n)$; (3) the elements of $P_n R_n$ and $R_n P_n$ are uniformly $O(o_n)$.

Noting that both $W$ and $B^{-1}$ are all uniformly bounded in both row and column sums under our assumption. It is easy to apply the first evident fact to prove the following lemma.

**Lemma B.1** $B' B, (B' B)^{-1}$, $\Omega$, $\Omega^{-1}$, $\Omega^*$, $\Omega^*$, $\Omega^t$, $\Omega^{-1}$, $A$, and $\hat{A}$ are all uniformly bounded in both row and column sums, where recall $A = (B' B)^{-1} (W' B + B' W) (B' B)^{-1}$ and $\hat{A} = 2 (B' B)^{-1} [(W' B + B' W) A - W' W]$.
Lemma B.2 1) $tr(D_1\Omega D_2)/n = O(1)$ for $D_1, D_2 = \Omega^{-1}, \Omega^{-1}(I_T \otimes A)\Omega^{-1}, \Omega^{-1}(J_T \otimes I_n)\Omega^{-1}$, $\Omega^{-1}(I_T \otimes A')$. The same conclusion holds when $\Omega$ is replaced by $\Omega^* \text{ or } \Omega^!$, and $D_1$ and $D_2$ are replaced by their analogs corresponding to the case of $\Omega^* \text{ or } \Omega^!$.

2) $tr(B'^{-1}R B^{-1})/n = O(1)$ 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 i.i.d., sequences with zero means and fourth moments. Let $\sigma_a^2 = E(a_1^2), \sigma_b^2 = E(b_1^2)$. Let $\kappa_a = E(a_1^4) - 3\sigma_a^4$, and $\kappa_b = E(b_1^4) - 3\sigma_b^4$. Let $q_n, p_n$ be $n \times n$ nonstochastic matrices. Then

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

2) $E\left(a'q_n a\right) \left(b'p_n b\right) = \sigma_a^2 \sigma_b^2 tr\left(q_n p_n\right)$.

3) $E\left(a'q_n a\right) \left(a'p_n a\right) = \sigma_a^4 tr\left(q_n p_n\right)$.

Proof. To show 1), write $E\left(a'q_n a\right) \left(a'q_n a\right) = E\left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_ia_j q_{n,ij} a_k q_{n,kl}\right)$. Noting that $E(a_i a_j a_k a_l)$ 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

$$E\left(a'q_n a\right) \left(a'q_n a\right) = E\left(\sum_{i=1}^n q_{n,ii}p_{n,ii} + \sigma_a^4 \sum_{i=1}^n \sum_{j=1}^n q_{n,ij}p_{n,ij} + q_{n,ij}p_{n,ji}\right) \left(\sum_{i=1}^n q_{n,ii}p_{n,ii} + \sigma_a^4 \sum_{i=1}^n \sum_{j=1}^n q_{n,ij}p_{n,ij} + q_{n,ij}p_{n,ji}\right)$$

The result in 2) is trivial since $E\left(a'q_n a\right) \left(b'p_n b\right) = E\left(a'q_n a\right) E\left(b'p_n b\right) = \sigma_a^2 \sigma_b^2 tr\left(q_n p_n\right)$. Now, $E\left(a'q_n b\right) \left(a'p_n b\right) = E\left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_i b_j q_{n,ij} a_k b_l p_{n,kl}\right) = E\left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_i^2 b_j^2 q_{n,ij} p_{n,kl}\right) = \sigma_a^2 \sigma_b^2 tr\left(q_n p_n\right)$.

Lemma B.4 Recall $u = (\kappa T \otimes I_n)\mu + (I_T \otimes B^{-1})v$. Let $a = \xi + \mu/(1 - \rho) + \sum_{j=0}^\infty \rho^j B^{-1}v - j$, where $\xi, \mu, \text{ and } v$ are defined in the text. In particular, $\xi_i's$ are i.i.d., and independent of $\mu$ 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, $q_n, p_n, \text{ and } r_n$ are symmetric. Then

1) $E\left(u'q_n u\right) \left(u'p_n u\right) = \kappa_u \sum_{i=1}^n G_{q_n,i} G_{p_n,i} + \kappa_v \sum_{i=1}^n G_{q_n,2i} G_{p_n,2i} + \sigma_v^2 \left[tr\left(q_n \Omega\right) tr\left(p_n \Omega\right) + 2tr\left(q_n \Omega p_n \Omega\right)\right]$. 28
2) \( E \left( u' q_n u (a' r_n a) \right) = \frac{\kappa_\mu}{(1-\rho)^2} \sum_{i=1}^{\infty} G_{q_n,1ii}(r_n a_i) + \sigma_\nu^4 \{ \text{tr} (r_n \omega_{11}) \text{tr} (q_n \Omega) + 2 \text{tr} (\omega_{12} q_n \omega_{21} p_n) \} \).

3) \( E \left( a' s_n u (a' t_n a) \right) = \frac{\kappa_s}{(1-\rho)^2} \sum_{i=1}^{\infty} (s_n (r_T \otimes I_n))_{ii} (t_n (r_T \otimes I_n))_{ii} + \sigma_\nu^4 \{ \text{tr} (s_n \omega_{21}) \text{tr} (t_n \omega_{21}) + \text{tr} (s_n \omega_{21} t_n \omega_{21}) + \text{tr} (s_n \Omega a_n \omega_{21}) \} \).

4) \( E \left( u' q_n u (u' a_n a) \right) = \frac{\kappa_u}{(1-\rho)^2} \sum_{i=1}^{\infty} G_{q_n,1ii}(s_n (r_T \otimes I_n))_{ii} + \sigma_\nu^4 \{ \text{tr} (s_n \omega_{11}) \text{tr} (q_n \omega_{21}) + 2 \text{tr} (r_n \omega_{11} s_n \omega_{21}) \} \).

5) \( E \left( a' r_n a (a' s_n a) \right) = \frac{\kappa_a}{(1-\rho)^2} \sum_{i=1}^{\infty} G_{q_n,2ii}(s_n (r_T \otimes I_n))_{ii} + \sigma_\nu^4 \{ \text{tr} (r_n \omega_{11}) \text{tr} (s_n \omega_{21}) + 2 \text{tr} (r_n \omega_{11} s_n \omega_{21}) \} \),

where, e.g., \( G_{q_n,1} = (r_T \otimes I_n) q_n (r_T \otimes I_n) \), and \( G_{q_n,2} = (I_T \otimes B^{-1}) q_n (I_T \otimes B^{-1}) \).

**Proof.** We only sketch the proof of 1) and 2) since it mainly follows from Lemma 2.3 and the other proofs are similar. First, let \( G_{q_n,3} = (r_T \otimes I_n) q_n (I_T \otimes B^{-1}) \). Then by the independence of \( \mu \) and \( v \) and Lemma 2.3, we have

\[
E \left( u' q_n u (a' p_n u) \right) = E \{ \mu' G_{q_n,1} \mu' G_{p_n,1} + v' G_{q_n,2} v' G_{p_n,2} + \mu' G_{q_n,3} \mu' G_{p_n,3} + v' G_{q_n,2} v' G_{p_n,1} \mu \}
+ 2 \mu' G_{q_n,3} \mu' G_{p_n,3} + 2 v' G_{q_n,3} \mu' G_{p_n,3} \mu
= \kappa_\mu \sum_{i=1}^{\infty} G_{q_n,1ii} G_{p_n,1ii} + \kappa_v \sum_{i=1}^{\infty} G_{q_n,2ii} G_{p_n,2ii} + \sigma_\nu^4 \{ \text{tr} (q_n \Omega) \text{tr} (p_n \Omega) + 2 \text{tr} (q_n \Omega p_n \Omega) \}.
\]

Next, write \( a = b + B^{-1} c \), where \( b = \xi + \mu / (1 - \rho) \) and \( c = \sum_{j=0}^{\infty} \rho^j v_{-j} \). Then \( b \) and \( c \) are i.i.d. and mutually independent. So

\[
E \left( u' q_n u (a' r_n a) \right) = E \{ \mu' G_{q_n,1} \mu' r_n b + v' G_{q_n,2} v' B^{-1} r_n B^{-1} b_c + \mu' G_{q_n,1} \mu' B^{-1} r_n B^{-1} c + v' G_{q_n,2} v' r_n b \}
= \frac{\kappa_\mu}{(1-\rho)^2} \sum_{i=1}^{\infty} G_{q_n,1ii} G_{r_n,1ii} + \sigma_\nu^4 \{ \text{tr} (r_n \omega_{11}) \text{tr} (q_n \omega_{21}) + 2 \text{tr} (\omega_{12} q_n \omega_{21} p_n) \}.
\]

Similarly, we can prove the other claims. □

To proceed, we define some notation. By continuous back substitution, we have for \( t = 0, 1, 2, ..., \)

\[
y_t = \mathcal{X}_t y_0 + c_{p,t} \gamma_0 + c_{p,t} \mu + \mathcal{V}_t + \mathcal{Y}_{0,t},
\]

where \( \mathcal{X}_t = \sum_{j=0}^{t-1} \rho_0^j x_{t-j}, \mathcal{V}_t = \sum_{j=0}^{t-1} \rho_0^j B^{-1} v_{t-j}, \mathcal{Y}_{0,t} = \rho_0^t y_0 \) and \( c_{p,t} = (1 - \rho_0^t) / (1 - \rho_0) \) for the case of random effect with fixed initial observations (Case I), and \( \mathcal{X}_t = \sum_{j=0}^{\infty} \rho_0^j x_{t-j}, \mathcal{V}_t = \sum_{j=0}^{\infty} \rho_0^j B^{-1} v_{t-j}, \mathcal{Y}_{0,t} = 0, \) and \( c_{p,t} = 1 / (1 - \rho_0) \equiv c_p \) for the case of random effect with endogenous initial observations (Case II). Now, define \( \mathcal{V}_0 = (\mathcal{V}_0', \mathcal{V}_{0,1}', ..., \mathcal{V}_{0,T-1}') \). Then

\[
Y_{-1} = \mathcal{X}_{(-1)} \beta_0 + (l_p \otimes I_n) \gamma_0 + (l_p \otimes I_n) \mu + \mathcal{V}_{(-1)} + \mathcal{Y}_0,
\]

(2.2)
where

\[
X_{(-1)} = \begin{pmatrix}
0 \\
X_1 \\
\vdots \\
X_{T-1}
\end{pmatrix}, \quad V_{(-1)} = \begin{pmatrix}
0 \\
V_1 \\
\vdots \\
V_{T-1}
\end{pmatrix}, \quad \text{and } l_\rho = \begin{pmatrix}
0 \\
c_{\rho,1} \\
\vdots \\
c_{\rho,T-1}
\end{pmatrix}
\]  
(B.3)

for Case I and

\[
X_{(-1)} = \begin{pmatrix}
X_0 \\
X_1 \\
\vdots \\
X_{T-1}
\end{pmatrix}, \quad V_{(-1)} = \begin{pmatrix}
V_0 \\
V_1 \\
\vdots \\
V_{T-1}
\end{pmatrix}, \quad \text{and } l_\rho = c_\mu T
\]  
(B.4)

for Case II. Notice that in Case I, \( Y_{-1} \) can also be expressed as

\[
Y_{-1} = A_x X' \beta_0 + (I_P \otimes I_n) z \gamma_0 + (I_P \otimes I_n) \mu + A_v v + Y_0,
\]  
(B.5)

where \( A_x = C_\rho \otimes I_n \) and \( A_v = C_\rho \otimes B^{-1} \) with

\[
C_\rho = \begin{pmatrix}
0 & 0 & 0 & \cdots & 0 & 0 \\
1 & 0 & 0 & \cdots & 0 & 0 \\
\rho_0 & 1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\rho_0^{T-2} & \rho_0^{T-3} & \rho_0^{T-4} & \cdots & 1 & 0
\end{pmatrix}
\]

The expression in (B.5) will facilitate our analysis in several places. The following four lemmas (Lemmas B.5-B.8) are used in the proof of Theorem 4.2.

**Lemma B.5** For \( D_1, D_2 = \Omega^{-1}, \Omega^{-1} (I_T \otimes A) \Omega^{-1} \), or \( \Omega^{-1} (J_T \otimes I_n) \Omega^{-1} \),

1) \( n^{-1} \left[ u' D_1 \Omega D_2 u - \sigma_u^2 tr (D_1 \Omega D_2 \Omega) \right] = o_p(1), \)

2) \( n^{-1} \left[ \tilde{X}' D_1 \Omega D_2 \tilde{X} - E \left( \tilde{X}' D_1 \Omega D_2 \tilde{X} \right) \right] = o_p(1). \)

**Proof.** Let \( R = D_1 \Omega D_2 \). Note that \( R \) is uniformly bounded in both row and column sums. To show 1), first note that \( E (u' Ru) = \sigma_u^2 tr (R \Omega) \). By Lemma B.4,

\[
\text{Var} (n^{-1} u' Ru) = n^{-2} \left[ E (u' Ru' Ru) - (E (u' Ru))^2 \right] = n^{-2} \kappa \sum_{i=1}^{n} G_{R,1ii}^2 + n^{-2} \kappa \sum_{i=1}^{n} G_{R,2ii}^2 + 2n^{-2} \sigma_u^4 tr (R \Omega R \Omega) = O (n^{-1})
\]
where the last equality follows from the fact that $G_{R,1}^2$, $G_{R,2}^2$, and $R\Omega R\Omega$ are all uniformly bounded in both row and column sums. Then 1) follows by the Chebyshev inequality.

Next, noticing that $\bar{X} = (X, Z, Y_{-1})$, it is easy to show that terms like $n^{-1}X'RX$, $n^{-1}X'RZ$, and $n^{-1}Z'RZ$ converge to their expectation. For terms involving $Y_{-1}$, we need to use (B.5) to obtain

$$n^{-1}Y^{-1}_{-1}RY = n^{-1}[Ax'\beta_0 + (l_p \otimes I_n) z\gamma_0] R [Ax'\beta_0 + (l_p \otimes I_n) z\gamma_0] + n^{-1}[(l_p \otimes I_n) \mu + A_v v] R [(l_p \otimes I_n) \mu + A_v v] + n^{-1}Y^{-1}_0 RY_0 + 2n^{-1}[Ax'\beta_0 + (l_p \otimes I_n) z\gamma_0] R [Ax'\beta_0 + (l_p \otimes I_n) z\gamma_0] + 2n^{-1}[(l_p \otimes I_n) \mu + A_v v] RY_0 + 2n^{-1}[(l_p \otimes I_n) \mu + A_v v] RY_0$$

$$\equiv \sum_{i=1}^{r} A_{mi}.$$  

It suffices to show that each $A_{mi}$ ($i = 1, ..., 6$) converges to its expectation. Take $A_{n6}$ as an example. $E(A_{n6}) = 0$ because $\gamma_0$ is kept fixed here. For the second moment,

$$\text{Var} (A_{n6}) = 4n^{-2}\left\{ E \left( \mu' (l_p \otimes I_n) R\gamma_0 \gamma_0' R' (l_p \otimes I_n) \mu \right) + E \left( v' A_v R\gamma_0 \gamma_0' R' A_v v \right) \right\}$$

$$= 4n^{-2} \left\{ \sigma^2 \text{tr} (R\gamma_0 \gamma_0' R' (l_p' \otimes I_n)) + \sigma^2 \text{tr} (A_v R\gamma_0 \gamma_0' R' A_v) \right\} = O(n^{-1}),$$

where the last equality follows from the fact that both matrices in the trace operator 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 expectation in probability.

**Lemma B.6** Let $R$ be an $nT \times nT$ nonstochastic matrix that is uniformly bounded in both row and column sums, e.g., $I_n T$, $\Omega^{-1}(I_T \otimes A)$, or $\Omega^{-1}(I_T \otimes I_n)$. Then $n^{-1}\bar{X}'\Omega^{-1}u \xrightarrow{} (0_{1 \times p}, 0_{1 \times q}, \lim_{n \to \infty} \sigma^2 \text{tr} (R (J_p \otimes I_n)) / n)$, where

$$J_p = \begin{pmatrix} 0 & 1 & \rho & \cdots & \rho^{T-2} \\ 0 & 0 & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$  

In particular, when $R = I_{nT}$, $E(n^{-1}\bar{X}'\Omega^{-1}u) = 0$ and $n^{-1}\bar{X}'\Omega^{-1}u \xrightarrow{} 0(p+q+1) \times 1$.  

31
Proof. Note that $\bar{X} = (X, Z, Y_{-1})$. By the strict exogeneity of $X$ and $Z$, we can readily show that both $X'R\Omega^{-1}u$ and $Z'R\Omega^{-1}u$ have expectation zero, $n^{-1}X'R\Omega^{-1}u \overset{p}{\to} 0$ and $n^{-1}Z'R\Omega^{-1}u \overset{p}{\to} 0$.

We are left to show

$$n^{-1} \{Y_{-1}'R\Omega^{-1}u - \sigma_v^2 \text{tr} \left( R(J_p \otimes I_n) \right) \} \overset{p}{\to} 0,$$

and $E \{Y_{-1}'R\Omega^{-1}u \} = \sigma_v^2 \text{tr} \left( R(J_p \otimes I_n) \right)$.

Using (B.5), we have

$$Y_{-1}'R\Omega^{-1}u = \mu' \left( I_p' \otimes I_n \right) R\Omega^{-1}u + \nu' A_v' R\Omega^{-1}u + \beta_0 \sigma_v^2 \text{tr} \left( R\Omega^{-1} \left( I_p \otimes I_n \right) \right)$$

and

$$= \sum_{j=1}^5 A_{nj}.$$

It is easy to show that $n^{-1} \sigma_v^2 \text{tr} \left( R(J_p \otimes I_n) \right) = o_p(1)$ for $j = 3, 4, 5$. So we will show that $A_{n1} + A_{n2} = \sigma_v^2 \text{tr} \left( R(J_p \otimes I_n) \right) + o_p(n)$. Noting that $u = (\nu T \otimes I_n) \mu + (I_T \otimes B^{-1}) v$, we have

$$E \left( A_{n1} \right) = E \left[ \nu' A_v' R\Omega^{-1} (I_T \otimes B^{-1}) v \right] = \sigma_v^2 \text{tr} \left[ R\Omega^{-1} (I_T \otimes B^{-1}) (J_p \otimes B^{-1}) \right]$$

and

$$E \left( A_{n2} \right) = E \left[ \nu' A_v' R\Omega^{-1} (I_T \otimes B^{-1}) \nu \right] = \sigma_v^2 \text{tr} \left[ R\Omega^{-1} \left( J_p \otimes (B'B)^{-1} \right) \right].$$

where we have used (3.35) and the fact that $E (\nu' A_v') = J_p \otimes B^{-1}$. Hence

$$E \left( A_{n1} + A_{n2} \right) = \sigma_v^2 \text{tr} \left( R(J_p \otimes I_n) \right).$$

We can show that $E (A_{nj}^2) = O(n)$ for $j = 1, 2$. It follows from the Chebyshev inequality that

$$n^{-1} (A_{n1} + A_{n2}) = \sigma_v^2 \text{tr} \left( R(J_p \otimes I_n) \right) / n + o_p(1).$$

When $R = I_n$, $tr \left( R(J_p \otimes I_n) \right) = tr \left( J_p \right) tr \left( I_n \right) = 0$ and we can also verify that $E \left( A_{ni} \right) = 0$ for $i = 3, 4, 5$. This completes the proof. $\blacksquare$

Lemma B.7 Suppose that $\{P_{1n}\}$ and $\{P_{2n}\}$ are sequences of matrices with row and column sums uniformly bounded. Let $a = (a_1, ..., a_n)'$, where $a_i$ is a random variable and $\sup E |a_i|^{2+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$. Let $b = (b_1, ..., b_n)'$, where $b_i's$ are i.i.d. with mean zero and $(4 + 2\epsilon_0)$th finite moments,
and \( \{b_i\} \) is independent of \( \{a_i\} \). Let \( \sigma^2_{Q_n} \) be the variance of \( Q_n = a'P_{1n}b + b'P_{2n}b - \sigma^2_{\text{tr}}(P_{2n}) \). Assume that the elements of \( P_{1n}, P_{2n} \) are of uniform order \( O(1/\sqrt{n}) \) 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 \( b \) like that in Theorem 1 of Kelejian and Prucha (2001). The only difference is that the coefficient \( (a'P_{1n}) \) of the linear part of \( Q_n \) is random here. We can prove the lemma by modifying the proof of Theorem 1 in Kelejian and Prucha (2001) or Lemma A.13 of Lee (2002).

**Lemma B.8** Suppose that the conditions in Theorem 4.2 are satisfied. Then

1. \( \frac{1}{\sqrt{n}} \tilde{X} \Gamma^{-1} u \xrightarrow{d} N(0, \Gamma_{r,11}) \), where \( \Gamma_{r,11} = \rho \lim_{n \to \infty} (nT)^{-1} \tilde{X} \Gamma^{-1} \tilde{X} \).
2. \( \frac{1}{\sqrt{n}} \frac{\partial c(c_n)}{\partial c} \xrightarrow{d} N(0, \Gamma_r) \).

**Proof.** For 1), by the Cramer-Wold device, it suffices to show that for any \( c = (c'_1, c'_2, c'_3)' \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R} \) such that \( ||c|| = 1 \), \( (nT)^{-1/2} \tilde{X} \Gamma^{-1} u \xrightarrow{d} N(0, \epsilon \Gamma_{r,11} c) \). Using (B.5) and the expression \( u = (\epsilon_T \otimes I_n) \mu + (I_T \otimes B^{-1}) v, \) we have

\[
\begin{align*}
c' \tilde{X} \Gamma^{-1} u &= c'_1 X \Gamma^{-1} u + c'_2 Z \Gamma^{-1} u + c'_3 Y \Gamma^{-1} u \\
&= c'_1 X \Gamma^{-1} (\epsilon_T \otimes I_n) \mu + c'_1 X \Gamma^{-1} (I_T \otimes B^{-1}) v \\
&\quad + c'_2 Z \Gamma^{-1} (\epsilon_T \otimes I_n) \mu + c'_2 Z \Gamma^{-1} (I_T \otimes B^{-1}) v \\
&\quad + c_3 \beta'_0 X A'_0 \Gamma^{-1} (\epsilon_T \otimes I_n) \mu + c_3 \beta'_0 X A'_0 \Gamma^{-1} (I_T \otimes B^{-1}) v \\
&\quad + c_3 \gamma'_0 z (l'_p \otimes I_n) \Gamma^{-1} (\epsilon_T \otimes I_n) \mu + c_3 \gamma'_0 z (l'_p \otimes I_n) \Gamma^{-1} (I_T \otimes B^{-1}) v \\
&\quad + c_3 \gamma'_0 z (l'_p \otimes I_n) \Gamma^{-1} (\epsilon_T \otimes I_n) \mu + c_3 \gamma'_0 z (l'_p \otimes I_n) \Gamma^{-1} (I_T \otimes B^{-1}) v \\
&\quad + c_3 \gamma'_0 z (l'_p \otimes I_n) \Gamma^{-1} (\epsilon_T \otimes I_n) \mu + c_3 \gamma'_0 z (l'_p \otimes I_n) \Gamma^{-1} (I_T \otimes B^{-1}) v \\
&\quad + c_3 \epsilon' A'_0 \Gamma^{-1} (\epsilon_T \otimes I_n) \mu + c_3 \epsilon' A'_0 \Gamma^{-1} (I_T \otimes B^{-1}) v \\
&= \sum_{i=1}^{3} T_{ni},
\end{align*}
\]

where

\[
T_{n1} = [c'_1 X + c'_2 Z + c_3 \beta'_0 X A'_0 + c_3 \gamma'_0 z (l'_p \otimes I_n) + c_3 \gamma'_0 z (l'_p \otimes I_n) \Gamma^{-1} (\epsilon_T \otimes I_n) \mu \\
+ c_3 \epsilon' A'_0 \otimes (\epsilon_T \otimes I_n) \Gamma^{-1} v]
\]
\[ T_{n2} = \left[ c_1X + c_2Z + c_3\beta'X\Omega^{-1}\left( I_T \otimes B^{-1}\right) \right] v + c_3v'\Omega^{-1}(I_T \otimes B^{-1})v, \]

\[ T_{n3} = c_3\mu'\left( I_n' \otimes I_n \right) \Omega^{-1}(I_T \otimes B^{-1})v + c_3v'\Omega^{-1}(I_T \otimes B^{-1})\mu \]

It is easy to verify that \( E(T_{n3}) = 0, E(T_{n1}) = c_3\phi_0\sigma^2\text{tr} \left( \Omega^{-1}(I_T l' \otimes I_n) \right), \) 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 normally distributed with mean zero.

Note that \( T_{n1} \) and \( T_{n2} \) are linear and quadratic functions of \( \mu \) and \( v \), respectively. For \( T_{n3} \), it is a special case of Lemma B.7 since it can be regarded as a linear function of either \( \mu \) or \( v \) and \( \mu \) and \( v \) are independent of each other by assumption. So we can apply Lemma B.10 to \( T_{ni} \) \( (i = 1, 2, 3) \) to obtain

\[ \{T_{ni} - E(T_{ni})\} / \sqrt{\text{Var}(T_{ni})} \xrightarrow{d} N(0, 1). \]

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

\[ \frac{1}{\sqrt{nT}}X'\Omega^{-1}u = \frac{1}{\sqrt{nT}}\sum_{i=1}^{3}T_{ni} \xrightarrow{d} N\left( 0, \lim_{n \to \infty} (nT)^{-1}\sum_{i=1}^{3}\text{Var}(T_{ni}) \right), \]

implying that \( \frac{1}{\sqrt{nT}}X'\Omega^{-1}u \xrightarrow{d} N(0, I_{v_1}) \) because we can readily show that \( (nT)^{-1}(X'\Omega^{-1}X - \text{Var}(X'\Omega^{-1}u)) = o_p(1) \).

For 2), noticing that each component of \( \partial L^\tau(\zeta_0) / \partial \varsigma \) can be written as linear and quadratic functions of \( \mu \) or \( v \), the proof can be done by following that of 1) closely.

The next four lemmas are used in the proof of Theorem 4.4 for the random effects model with endogenous initial observations (Case II). Let \( R_{ts} \) be an \( n \times n \) symmetric and positive semidefinite (psd) nonstochastic square matrix for \( t, s = 0, 1, ..., T - 1 \). Assume that \( R_{ts} \) are uniformly bounded in both row and column sums. Recall for Case II \( X_t = \sum_{j=0}^{\infty} \rho_0 x_{t-j}, \) and \( V_t = \sum_{j=0}^{\infty} \rho_0 B^{-1}v_{t-j}. \)

**Lemma B.9** Suppose that the conditions in Theorem 4.4 are satisfied. Then

1) \( E(\psi'_t R_{ts} \psi_s) = \sigma^2 \text{tr} \left( B^{-1}R_{ts}B^{-1} \right) \sum_{i=\max(0,t-s)}^{\infty} \rho^{s+t+2i}, \)
2) \( E(X'_t R_{ts} X_s) = \text{tr} \left( \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} R_{ts} E \left( x_{s-j} x'_{t-k} \right) \right), \)
3) \( E(X'_t R_{ts} V_s) = 0. \)

34
Proof. Let $P_j = \rho^j B^{-1}$. Then $\mathbb{V}_t = \sum_{j=0}^{\infty} P_j v_{t-j}$. Noting that $E(v'_t D v_s) = \sigma^2_t tr(D)$ if $t = s$ and 0 otherwise, we have

$$E(\mathbb{V}_t' R_t s \mathbb{V}_s) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} E(\mathbb{V}_t' P'_i R_t s P_j v_{t-j}) = \sum_{i=\max(0,t-s)}^{\infty} E(\mathbb{V}_t' P'_i R_t s P_{s-t+i} v_{t-i})$$

$$= \sigma^2_t tr \left( \sum_{i=\max(0,t-s)}^{\infty} P'_i R_t s P_{s-t+i} \right) = \sigma^2_t tr \left( B'^{-1} R_t s B^{-1} \right) \sum_{i=\max(0,t-s)}^{\infty} \rho^{s-t+2i}.$$ 

Next, noting that $\mathbb{X}_t = \sum_{j=0}^{\infty} \rho^j x_{t-j}$, we have

$$E(\mathbb{X}_t' R_t s \mathbb{X}_s) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} E(x'_{t-k} R_t s x_{s-j}) = tr \left( \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} R_t s E(x_{s-j} x'_{t-k}) \right).$$

Now,

$$E(\mathbb{X}_t' R_t s \mathbb{X}_s) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} E(x'_{t-k} R_t s B^{-1} v_{s-j}) = 0.$$

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

1) $\text{Cov}(\mathbb{V}_t' R_t s \mathbb{V}_s, \mathbb{V}_s' R_{gh} \mathbb{V}_h) = \rho_{tsgh,1} \left\{ \kappa \sum_{i=1}^{n} (B'^{-1} R_t s B^{-1})_{ii} (B'^{-1} R_{gh} B^{-1})_{ii} + 2\sigma^4_t tr \left( B'^{-1} R_t s B^{-1} \left( B'^{-1} R_{gh} B^{-1} + B'^{-1} R_{gh} B^{-1} \right) \right) \right\}$

$\quad + \rho_{tsgh,2} \sigma^4_t tr \left( B'^{-1} R_t s (B'B)^{-1} R_{gh} B^{-1} \right) + \rho_{tsgh,3} \sigma^4_t tr \left( B'^{-1} R_t s (B'B)^{-1} R_{gh} B^{-1} \right).$

2) $\text{Cov}(\mathbb{X}_t' R_t s \mathbb{X}_s, \mathbb{X}_s' R_{gh} \mathbb{X}_h) = \sigma^2_t tr \left( \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{j=\max(0,s-t-h)}^{\min(m,s+h)} \rho^{j+k+h-s+2j} R_t s (B'B)^{-1} R_{gh} E(x'_{s-j} x_{t-k}) \right).$

3) $\text{Cov}(\mathbb{X}_t' R_t s \mathbb{X}_t, \mathbb{X}_s' R_{gh} \mathbb{X}_h) = O(n),$ where

$$\rho_{tsgh,1} = \sum_{j=\max(0,t-s-g-h-t)}^{\infty} \rho^{s+g+h+3t+4j}, \quad \rho_{tsgh,2} = \sum_{i=\max(0,t-g)}^{\infty} \rho^{g-t+2i} \sum_{j=\max(0,s-h)}^{\infty} \rho^{h-s+2j} (j \neq i + s - t), \quad \text{and} \quad \rho_{tsgh,3} = \sum_{i=\max(0,t-h)}^{\infty} \rho^{h-t+2i} \sum_{j=\max(0,s-g)}^{\infty} \rho^{g-s+2j} (j \neq i + s - t).$$

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

$$E \left( \left[ (v'_t R_1 v_s) (v'_g R_2 v_h) \right] \right) = \begin{cases} \kappa i \sum_{i=1}^{n} R_{1,i} R_{2,ii} + \sigma^4_t [tr(R_1) tr(R_2) + tr(R_1 (R_2 + R_2'))] & \text{if } t = s = g = h \\ \sigma^4_t tr(R_1) tr(R_2) & \text{if } t = s \neq g = h \\ \sigma^4_t tr(R_1 R_2) & \text{if } t = g = s = h \\ \sigma^4_t tr(R_1 R'_2) & \text{if } t = h \neq s = g \\ 0 & \text{otherwise} \end{cases}$$

35
Consequently,

\[
E \left[ \Psi'_t R_{ts} \Psi_s \Psi'_g R_{gh} \Psi_h \right] \\
= E \left[ \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} \left( R_{ts} R_{gh} B_t - 1 \right) v_{s-j} \Psi'_g B_{t-k} R_{gh} B_{t-l} \right] \\
= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} \kappa \sum_{i=1}^{\infty} \left( B_{t-i} R_{ts} B_{t-i} \right) \left( B_{t-i} R_{gh} B_{t-i} \right) \\
+ \sigma_v^4 \left( \rho^{j+k} \left( \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} \left( R_{ts} R_{gh} B_t - 1 \right) v_{s-j} \right) \left( B_{t-k} R_{gh} B_{t-l} \right) \right) \\
+ \sigma_v^4 \left( \rho^{j+k} \left( \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} \left( R_{ts} R_{gh} B_t - 1 \right) v_{s-j} \right) \left( B_{t-k} R_{gh} B_{t-l} \right) \right) \\
+ \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} \left( R_{ts} R_{gh} B_t - 1 \right) \left( B_{t-k} R_{gh} B_{t-l} \right) \\
+ \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} \left( R_{ts} R_{gh} B_t - 1 \right) \left( B_{t-k} R_{gh} B_{t-l} \right) \\
+ \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} \left( R_{ts} R_{gh} B_t - 1 \right) \left( B_{t-k} R_{gh} B_{t-l} \right) \\
\end{array}
\]

\[
\text{Then 1) follows by Lemma B.9.}
\]

For 2), we have

\[
\text{Cov} (\Psi'_t R_{ts} \Psi_s, \Psi'_g R_{gh} \Psi_h) = E \left[ \Psi'_t R_{ts} \Psi_s \text{Cov} (\Psi'_t R_{ts} \Psi_s) \right] \\
= E \left[ \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} \left( R_{ts} R_{gh} B_t - 1 \right) v_{s-j} \left( B_{t-k} R_{gh} B_{t-l} \right) \right] \\
= \sigma_v^4 \left( \rho^{j+k} \left( \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \rho^{j+k} \left( R_{ts} R_{gh} B_t - 1 \right) v_{s-j} \right) \left( B_{t-k} R_{gh} B_{t-l} \right) \right) \\
\]

The expression for \(\text{Cov}(\Psi'_t R_{ts} \Psi_s, \Psi'_g R_{gh} \Psi_h)\) is quite complicated, but we can use the three evident facts to show it is of order \(O(n)\), which suffices for our purpose. ■

**Lemma B.11**
1) \((nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \Psi'_t R_{ts} \Psi_s \rightarrow E \left[ (nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \Psi'_t R_{ts} \Psi_s \right] \rightarrow 0,

2) \((nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \Psi'_t R_{ts} \Psi_s \rightarrow 0,

3) \((nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \Psi'_t R_{ts} \Psi_s \rightarrow E \left[ (nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \Psi'_t R_{ts} \Psi_s \right] \rightarrow 0.\]

**Proof.** By the three evident facts and Lemmas B.2, B.9, and B.10, we can show that \(E[(nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \Psi'_t R_{ts} \Psi_s] = O(1)\), and

\[
\text{Var} \left( n^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \Psi'_t R_{ts} \Psi_s \right) = n^{-2} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \sum_{g=0}^{T-1} \sum_{h=0}^{T-1} \text{Cov} (\Psi'_t R_{ts} \Psi_s, \Psi'_g R_{gh} \Psi_h) = O(n^{-1}).
\]
Then 1) follows by the Chebyshev’s inequality. For 2), we have \( E\{(nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \mathcal{X}^t R_{ts} \mathcal{V}_s \} = 0 \), and

\[
\begin{align*}
\text{Var}\left(n^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \mathcal{X}^t R_{ts} \mathcal{V}_s \right) &= n^{-2} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \sum_{g=0}^{T-1} \sum_{h=0}^{T-1} \text{Cov}\left( \mathcal{X}^t R_{ts} \mathcal{V}_s, \mathcal{X}_g R_{ts} \mathcal{V}_h \right) \\
&= n^{-2} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \sum_{g=0}^{T-1} \sum_{h=0}^{T-1} \sigma_t^2 tr \left[ \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \sum_{j=\max(0,s-h)}^{\infty} \rho^{i+k+h-s+j+2i} R_{tg} E \left( (x_{g-k} x_{i-1}^t) \right) \right] \\
&= O(n^{-1}),
\end{align*}
\]

where the last equality follows because (i) \( x_{it} \) are i.i.d. across \( i \) with second moments uniformly bounded in \( i \), (ii) \( R_{ts} (B'B)^{-1} R_{gh} \) are uniformly bounded in both row and column sums by B.1, and (iii) elements of \( R_{ts} (B'B)^{-1} R_{gh} E \left( (x_{g-k} x_{i-1}^t) \right) \) are uniformly bounded by the third evident fact.

Hence the conclusion follows by the Chebyshev inequality.

3) follows by Lemma B.10 and the Chebyshev inequality.

**Lemma B.12** For \( D_1, D_2 = \Omega^{s-1}, \Omega^{s-1}(J_2^* \otimes A) \Omega^{s-1}, \Omega^{s-1}(J_2^* \otimes I_n) \Omega^{s-1} \) or \( \Omega^{s-1}(K_y^* \otimes I_n) \Omega^{s-1} \),

1) \( n^{-1} \left[ u^* D_1 \Omega^* D_2 u - \sigma^2 tr (D_1 \Omega^* D_2 \Omega^*) \right] = o_p(1) \),

2) \( n^{-1} \left[ X^* D_1 \Omega^* D_2 X - E(X^* D_1 \Omega^* D_2 X) \right] = o_p(1) \).

**Proof.** Let \( R = D_1 \Omega^* D_2 \). Note that \( R \) is uniformly bounded in both row and column sums.

To show 1), first note that \( E(u^* R u) = \sigma^2 tr (R \Omega^*) \). Now write \( R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \). Let \( u_0 = y_0 - \tilde{x} \pi_0 \) and \( u = Y - \rho_0 Y_1 - X \beta_0 - Z \gamma_0 \). Then,

\[
\begin{align*}
\text{Var}(u^* R u) &= \text{Var}(u_0 R_{11} u_0 + u^T R_{22} u + u_0 (R_{12} + R_{21}^T) u) \\
&= \text{Var}(u_0 R_{11} u_0) + \text{Var}(u^T R_{22} u) + \text{Var}(u_0 (R_{12} + R_{21}^T) u) + 2\text{Cov}(u_0 R_{11} u_0, u^T R_{22} u) \\
&\quad + 2\text{Cov}(u_0 R_{11} u_0, u_0 (R_{12} + R_{21}^T) u) + 2\text{Cov}(u^T R_{22} u, u_0 (R_{12} + R_{21}^T) u).
\end{align*}
\]

By Lemmas B.2-B.4, we can show that each term on the right hand side is \( O(n) \). Consequently, 1) follows by the Chebyshev inequality.

Next, \( X^* R X^* = \begin{pmatrix} \tilde{X}^T R_{22} \tilde{X} & \tilde{X}^T R_{21} \tilde{x} \\ \tilde{x}^T R_{12} \tilde{X} & \tilde{x}^T R_{11} \tilde{x} \end{pmatrix} \). It is easy to show \( n^{-1} \tilde{x}^T R_{11} \tilde{x} \) converges in probability to its expectation. To show \( n^{-1} \tilde{X}^T R_{22} \tilde{X}, n^{-1} \tilde{X}^T R_{21} \tilde{x}, \) and \( n^{-1} \tilde{x}^T R_{12} \tilde{X} \) converge in probability to

37
their expectations, the major difficulty lies in the appearance of $Y_{-1}$ in $\tilde{X}$ (recall $\tilde{X} = (X, Z, Y_{-1})$). We show below that $n^{-1}(Y'_1 R_{22} Y_1 - E \{ Y'_1 R_{22} Y_1 \}) \overset{p}{\to} 0$ and the other cases can be proved similarly. To this goal, we need to use (B.2) (with $Y_0 = 0_{nT \times 1}$) to obtain

$$
n^{-1} Y'_1 R_{22} Y_1 = n^{-1} (X_{(-1)} \beta_0 + (l_\rho \otimes I_n) z \gamma_0 + (l_\rho \otimes I_n) \mu + \mathcal{V}_{(-1)})' R_{22} \times (X_{(-1)} \beta_0 + (l_\rho \otimes I_n) z \gamma_0 + (l_\rho \otimes I_n) \mu + \mathcal{V}_{(-1)}).
$$

After expressing out the right hand side of the last expression, it has 16 terms, most of which can easily be shown to converge to their respective expectations. The exceptions are terms involving $X_{(-1)}$ or $\mathcal{V}_{(-1)}$, namely: $n^{-1} \beta'_0 X'_{(-1)} R_{22} X_{(-1)} \beta_0$, $n^{-1} \beta'_0 V'_{(-1)} R_{22} V_{(-1)}$, $n^{-1} \beta'_0 \mathcal{X}'_{(-1)} R_{22} \mathcal{V}_{(-1)}$, $n^{-1} \beta'_0 \mathcal{X}'_{(-1)} R_{22} (l_\rho \otimes I_n) z \gamma_0$, $n^{-1} \beta'_0 \mathcal{X}'_{(-1)} R_{22} (l_\rho \otimes I_n) \mu$, $n^{-1} \mathcal{V}'_{(-1)} R_{22} (l_\rho \otimes I_n) z \gamma_0$, and $n^{-1} \mathcal{V}'_{(-1)} R_{22} (l_\rho \otimes I_n) \mu$. 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** Let $R$ be an $n(T+1) \times n(T+1)$ nonstochastic matrix that is uniformly bounded in both row and column sums, e.g., $l_n(1+T)$, $\Omega^{-1} (J_1 \otimes A)$, $\Omega^{-1} (J_2 \otimes I_n)$, or $\Omega^{-1} (K_2 \otimes I_n)$. Then $n^{-1} X^* \Omega^{-1} u^* \overset{p}{\to} 0_{(p+q+1) \times 1}$ and $E(X^* \Omega^{-1} u^*) = 0$.

**Proof.** Recall

$$
X^* = \left( \begin{array}{cccc}
0_{n \times p} & 0_{n \times q} & 0_{n \times 1} & \tilde{X} \\
0_{n \times 1} & 0_{n \times 1} & X & Z \\
0_{nT \times k} & 0_{nT \times k} & Y_{-1} & 0_{nT \times k}
\end{array} \right),
\tilde{u}^* = \left( \begin{array}{c}
y_0 - \tilde{x} \pi_0 \\
y - X \beta_0 - Z \gamma_0 - Y_{-1} \rho_0
\end{array} \right).
$$

By the strict exogeneity of $X$ and $Z$, we can readily show that $[0_{p \times n} X'] R \Omega^{-1} u^*, [0_{q \times n} Z'] R \Omega^{-1} u^*$ and $[\tilde{X}' 0_{n \times 1}] R \Omega^{-1} u^*$ have expectation zero, $n^{-1} [0_{p \times n} X'] R \Omega^{-1} u^* \overset{p}{\to} 0$, $n^{-1} [0_{q \times n} Z'] R \Omega^{-1} u^* \overset{p}{\to} 0$, and $n^{-1} [\tilde{X}' 0_{k \times n}] R \Omega^{-1} u^* \overset{p}{\to} 0$. Let

$$
Y'_{-1} = \left( \begin{array}{c}
0_{n \times 1} \\
0_{n \times 1} \\
0_{n \times 1}
\end{array} \right),
X^* = \left( \begin{array}{c}
X_{(-1)} \\
X_{(-1)} \\
X_{(-1)}
\end{array} \right),
\mathcal{V}^* = \left( \begin{array}{c}
0_{n \times 1} \\
0_{n \times 1} \\
0_{n \times 1}
\end{array} \right).
$$
Then we are left to show

\[ n^{-1} \left\{ Y_{s-n} R \Omega_s^{-1} u^* - \sigma_v^2 tr \left( R \left( (l^*_\rho) \right) \otimes I_n + J^*_\rho \otimes (B'B)^{-1} \right) \right\} \rightarrow 0, \]

and

\[ E \left( Y_{s-n} R \Omega_s^{-1} u^* \right) = \sigma_v^2 tr \left( R \left( (l^*_\rho) \right) \otimes I_n + J^*_\rho \otimes (B'B)^{-1} \right). \]

Using (B.2), \( Y_{s-n} = X_{(s-n)} \beta_0 + (l_\rho \otimes I_n) z \gamma_0 + (l_\rho \otimes I_n) \mu + \mathbb{V}_{(s-n)} \), we have

\[ Y_{s-n} R \Omega_s^{-1} u^* = \mu' (l^*_\rho \otimes I_n) R \Omega_s^{-1} u^* + \mathbb{V}' R \Omega_s^{-1} u^* \]

\[ + \beta_0 X^* R \Omega_s^{-1} u^* + \gamma_0 z (l^*_\rho \otimes I_n) R \Omega_s^{-1} u^* = n \sum_{j=1}^4 A_{n j}. \]

It is easy to show that \( E(A_{n j}) = 0 \) and \( n^{-1} A_{n j} = o_p(1) \) for \( j = 3, 4 \). So we will show that

\[ A_{n 1} + A_{n 2} = \sigma_v^2 tr \left( R \left( (l^*_\rho) \otimes I_n + J^*_\rho \otimes (B'B)^{-1} \right) \right) + o_p(n). \]

Let

\[ v^* = \left( \sum_{j=0}^\infty \rho^j B^{-1} u_{-j} \right), \]

noting that \( u^* = (u_0, u')' \), where \( u_0 = \xi + \mu / (1 - \rho) + \sum_{j=0}^\infty \rho^j B^{-1} u_{-j} \), and \( u = (l_\rho \otimes I_n) \mu + (l_\rho \otimes B^{-1} v) \), we have

\[ E(A_{n 1}) = E \left[ \mu' (l^*_\rho \otimes I_n) R \Omega_s^{-1} (l^*_\rho \otimes I_n) \mu \right] = \sigma_v^2 tr \left[ R \Omega_s^{-1} \left( (l^*_\rho) \otimes I_n \right) \right] \]

and

\[ E(A_{n 2}) = E \left[ \mathbb{V}' R \Omega_s^{-1} v^* \right] = \sigma_v^2 tr \left[ R \Omega_s^{-1} \left( J^*_\rho \otimes (B'B)^{-1} \right) \right], \]

where we have used (3.35) and the fact that \( E(v^* \mathbb{V}') = J^*_\rho \otimes (B'B)^{-1} \). Hence

\[ E(A_{n 1} + A_{n 2}) = \sigma_v^2 tr \left( R \Omega_s^{-1} \left( (l^*_\rho) \otimes I_n + J^*_\rho \otimes (B'B)^{-1} \right) \right). \]

We can show that \( E(A_{n j}^2) = o(n) \) for \( j = 1, 2 \). The first conclusion then follows from the Chebyshev inequality. When \( R = I_{n(t+1)} \), tedious calculation shows that \( tr \left( \Omega_s^{-1} \left( (l^*_\rho) \otimes I_n + J^*_\rho \otimes (B'B)^{-1} \right) \right) = 0 \) and we can also verify that \( E(A_{n 1}) = 0 \) for \( i = 3, 4 \). This completes the proof. □

**Lemma B.14** Suppose that the conditions in Theorem 4.4 are satisfied. Then

1. \( \frac{1}{\sqrt{n}} X_s \Omega_s^{-1} u^* \xrightarrow{d} N(0, \Gamma_{rr,1}) \), where \( \Gamma_{rr,1} = p \lim_{n \to \infty} (nT)^{-1} X_s \Omega_s^{-1} X_s \).
2. \( \frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}^*_{rr}(\phi)}{\partial \phi} \xrightarrow{d} N(0, \Gamma_{rr}). \)
Lemma B.15

Proof. For 1), by the Cramer-Wold device, it suffices to show that for any \( c = (c_1', c_2') \in \mathbb{R}^{p+q+1} \times \mathbb{R}^k \) such that \( \|c\| = 1 \), \((nT)^{-1/2} c' X^* \Omega^{-1} u^* \xrightarrow{d} N(0, c'T_{rr,11}c)\). Write

\[
T_n \equiv (nT)^{-1/2} c' X^* \Omega^{-1} u^* = (nT)^{-1/2} \left[ c_1' \tilde{X}' \omega_1^1 u_0 + c_1' \tilde{X}' \omega_2^2 u + c_2' \tilde{X}' \omega_1^1 u_0 + c_2' \tilde{X}' \omega_1^2 u \right].
\]

Analogous to the proof of Lemma B.8, we can write \( T_n \) as the summation of six asymptotically independent terms, namely, \( T_n = \sum_{i=1}^6 T_{ni} \), where \( T_{ni} \)'s are linear and quadratic functions of \( \mu \), linear function of \( \mu \) and \( v \), linear function of \( \mu \) and \( \tilde{C} \), linear function of \( v \) and \( \tilde{C} \) respectively. Further

\[
\{T_{ni} - E(T_{ni})\}/\sqrt{\text{Var}(T_{ni})} \xrightarrow{d} N(0,1),
\]

and \( E\left(\sum_{i=1}^6 T_{ni}\right) = 0 \). Now by the asymptotic independence of \( T_{ni} \)'s, we have

\[
T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^6 T_{ni} \xrightarrow{d} N\left(0, \lim_{n \to \infty} (nT)^{-1} \sum_{i=1}^6 \text{Var}(T_{ni})\right),
\]

implying that \((nT)^{-1/2} X^* \Omega^{-1} u^* \xrightarrow{d} N(0, \Gamma_{rr,11})\) because we can show that \((nT)^{-1} (X^* \Omega^{-1} X^* - \text{Var}(X^* \Omega^{-1} u^*)) = o_p(1)\). The proof of 2) is similar and thus omitted. \( \Box \)

The next three lemmas are used in the proof of Theorem 4.6 for the fixed effects model.

Lemma B.15 For \( D_1, D_2 = \Omega^{i-1}, \Omega^{i-1} \tilde{\Omega}^{i-1}, \Omega^{i-1} \Omega_m^{i-1}, \Omega^{i-1} \Omega_{\hat{\sigma}}^{i-1} \) or \( \Omega^{i-1} \Omega (\hat{\sigma})^{i-1} \),

1) \( n^{-1} [\Delta u'D_1\Omega^i D_2\Delta u - \sigma^2 tr \left(D_1\Omega^i D_2\Omega^i\right)] = o_p(1) \),

2) \( n^{-1} [\Delta X'D_1\Omega^i D_2\Delta X - E(\Delta X'D_1\Omega^i D_2\Delta X)] = o_p(1) \).

Proof. Let \( R = D_1\Omega^i D_2 \). Note that \( R \) is uniformly bounded in both row and column sums. Since \( E(\Delta u'R\Delta u) = \sigma^2 tr \left(R\Omega^i\right) \), by the Chebyshev inequality 1) follows provided \( \text{Var}(n^{-1} \Delta u'R\Delta u) = o(1) \). Let \( \Delta v(0) = B\zeta + \sum_{j=0}^{n-1} \rho^j \Delta v_{1-j} \), \( \Delta v(1) = (\Delta v_0', ..., \Delta v_T')' \), and \( \Delta v = \left(\Delta v(0), \Delta v(1)\right)' \). Then \( \Delta u'R\Delta u = \Delta v' \left(I_n \otimes B^{-1}\right) R \left(I_n \otimes B^{-1}\right) \Delta v = \Delta v' \tilde{R} \Delta v \), where \( \tilde{R} \equiv \left(I_n \otimes B^{-1}\right) R \left(I_n \otimes B^{-1}\right) \).

Now write \( R = \begin{pmatrix} R_{00} & R_{01} \\ R_{10} & R_{11} \end{pmatrix} \) and partition \( \tilde{R} \) similarly. Let \( C \) be a \((T-1) \times T\) matrix with \( C_{ij} = -1 \) if \( i = j \), \( C_{ij} = 1 \) if \( i = j + 1 \), and \( C_{ij} = 0 \) otherwise. Then \( \Delta v(1) = (C \otimes I_n) v \), where \( v = (v_1', ..., v_T')' \). So

\[
\Delta v' \tilde{R} \Delta v = \Delta v'(0) \tilde{R}_{00} \Delta v(0) + \Delta v'(1) \tilde{R}_{11} \Delta v(1) + \Delta v'(0) (R_{01} + R_{10}) \Delta v(1)
\]

\[
\Delta v'(0) \tilde{R}_{00} \Delta v(0) + v'(C' \otimes I_n) \tilde{R}_{11} (C \otimes I_n) v + \Delta v'(0) (R_{01} + R_{10}) (C \otimes I_n) v
\]

40
Then,
\[
\begin{align*}
\text{Var}(\Delta u' R \Delta u) &= \text{Var}\left(\Delta v'_{(0)} \tilde{R}_{00} \Delta v_{(0)}\right) + \text{Var}\left(v' (C' \otimes I_n) \tilde{R}_{11} (C \otimes I_n) v\right) + \text{Var}\left(\Delta v'_{(0)} (R_{01} + R'_{10}) (C \otimes I_n) v\right) \\
&\quad + 2 \text{Cov}\left(\Delta v'_{(0)} \tilde{R}_{00} \Delta v_{(0)}, v' (C' \otimes I_n) \tilde{R}_{11} (C \otimes I_n) v\right) \\
&\quad + 2 \text{Cov}\left(\Delta v'_{(0)} \tilde{R}_{00} \Delta v_{(0)}, \Delta v'_{(0)} (R_{01} + R'_{10}) (C \otimes I_n) v\right) \\
&\quad + 2 \text{Cov}\left(v' (C' \otimes I_n) \tilde{R}_{11} (C \otimes I_n) v, \Delta v'_{(0)} (R_{01} + R'_{10}) (C \otimes I_n) v\right).
\end{align*}
\]

By the Cauchy-Schwartz inequality it suffices to show that the first three terms on the right hand side are \(O(n)\) since then \(\text{Var}(n^{-1} \Delta u' R \Delta u) = O(n^{-1}) = o(1)\). Write \(\Delta v_{(0)} = B\zeta + v_1 - \rho^{m-1} v_{m+1} + \sum_{j=0}^{m-2} \rho^j (\rho - 1) v_j\). Since \(B' \tilde{R}_{00} B\) is uniformly bounded in both row and column sums,
\[
\text{Var}\left(\zeta' B' \tilde{R}_{00} B \zeta\right) = \kappa_n \sum_{i=1}^n \left[\left(B' \tilde{R}_{00} B\right)_{ii}\right]^2 + \sigma_n^2 \text{tr}\left(B' \tilde{R}_{00} B B' \left(\tilde{R}_{00} + \tilde{R}_{00}\right) B\right) = O(n).
\]

Similarly \(\text{Var}\left(v_1' \tilde{R}_{00} v_1\right) = O(n), \text{Var}\left(\left(\rho^{m-1} v_{m+1}\right)' \tilde{R}_{00} (\rho^{m-1} v_{m+1})\right) = O(n), \text{and Var}(\sum_{j=0}^{m-2} \rho^j (\rho - 1) v_j) = O(n).\) It follows by the Cauchy-Schwartz inequality that \(\text{Var}\left(\Delta v'_{(0)} \tilde{R}_{00} \Delta v_{(0)}\right) = O(n).\) By the same token, we can show that \(\text{Var}\left(v' (C' \otimes I_n) \tilde{R}_{11} (C \otimes I_n) v\right) = O(n),\) and \(\text{Var}\left(\Delta v'_{(0)} (R_{01} + R'_{10}) (C \otimes I_n) v\right) = O(n).\) This completes the proof of 1). The proof of 2) is similar and thus omitted. \(\blacksquare\)

**Lemma B.16** Let \(R\) be an \(nT \times nT\) nonstochastic matrix that is uniformly bounded in both row and column sums, e.g., \(I_n(T+1), \Omega^{-1}, \Omega^{-1} \Omega_1, \Omega^{-1} \Omega_{1,\infty}\) or \(\Omega^{-1} \Omega_{p,\infty}\). Then \(n^{-1} \Delta X' \Omega^{-1} \Delta u \xrightarrow{p} (0'_{p \times 1}, 0'_{q \times 1}, \lim_{n \to \infty} \sigma_n^2 \text{tr}\left\{R \Omega^{-1} \left[C_1 \otimes (B'B)^{-1} + C_2 \otimes I_n\right]\right\}/n, 0')\), where \(C_1\) is a \(T \times T\) matrix defined by
\[
C_1 = \begin{pmatrix}
0 & \frac{2}{T^2} \rho - 1 & \frac{2}{T^2} \rho - 1 & \cdots & \frac{2}{T^2} \rho - 1 & \rho \left(\frac{2}{T^2} \rho - 1\right) & \rho^3 \left(\frac{2}{T^2} \rho - 1\right) \\
0 & -1 & 2 - \rho & 2 - \rho - \rho^2 & \cdots & \rho^5 \left(2 \rho - 1 - \rho^2\right) & \rho^6 \left(2 \rho - 1 - \rho^2\right) \\
0 & 0 & -1 & 2 - \rho & \cdots & \rho^5 \left(2 \rho - 1 - \rho^2\right) & \rho^4 \left(2 \rho - 1 - \rho^2\right) \\
0 & 0 & 0 & 0 & \cdots & -1 & 2 - \rho \\
0 & 0 & 0 & 0 & \cdots & 0 & -1
\end{pmatrix}
\]
and \(C_2\) is a \(T \times T\) matrix whose first row is \((0, \phi_e, \rho \phi_e, \cdots, \rho^{T-2} \phi_e)\) and all other row elements are zero. In particular, when \(R = I_{nT}\), we have \(n^{-1} \Delta X' \Omega^{-1} \Delta u \xrightarrow{p} 0_{(p+k+1) \times 1}\) and \(E(\Delta X' \Omega^{-1} \Delta u) = 0\).
Proof. Let $\Delta X^* = (0_{p \times n}, \Delta x_2, ..., \Delta x_T)'$, $\Delta Y^* = \left( 0_{1 \times n}, \Delta y'_1, ..., \Delta y'_{T-1} \right)'$, $\Delta \tilde{x}^* = (\Delta \tilde{x}', 0_{n \times n(T-1)})'$. Then $\Delta X = (\Delta X^*, \Delta Y^*, \Delta \tilde{x}^*)$. By the notation in the proof of Lemma B.15, $\Delta u = (I_n \otimes B^{-1}) \Delta v$.

By the strict exogeneity of $X$ and $Z$, we can readily show that $\Delta X^* R \Omega^{-1} \Delta u$, and $\Delta \tilde{x}^* R \Omega^{-1} \Delta u$ have expectation zero, $n^{-1} \Delta X^* R \Omega^{-1} \Delta u \overset{P}{\rightarrow} 0$, and $n^{-1} \Delta \tilde{x}^* R \Omega^{-1} \Delta u \overset{P}{\rightarrow} 0$. We are left to show

$$n^{-1} \left\{ \Delta Y^* R \Omega^{-1} \Delta u - \sigma^2_t \text{tr} \left\{ R \Omega^{-1} \left[ C_1 \otimes (B'B)^{-1} + C_2 \otimes I_n \right] \right\} \right\} \overset{P}{\rightarrow} 0,$$

and $E (\Delta Y^* R \Omega^{-1} \Delta u) = \sigma^2_t \text{tr} \left\{ R \Omega^{-1} \left[ C_1 \otimes (B'B)^{-1} + C_2 \otimes I_n \right] \right\}$.

Let $k_p = (0, 1, \rho, ..., \rho^{T-2})'$, $\mathcal{X} = \left( 0_{1 \times n}, 0_{1 \times n}, (\Delta x_2 \beta_0)', ..., \sum_{j=0}^{T-3} \rho^j (\Delta x_{T-1-j} \beta_0)' \right)$, and $\mathcal{V} = \left( 0_{1 \times n}, 0_{1 \times n}, (\Delta v_2)', ..., \sum_{j=0}^{T-3} \rho^j (\Delta v_{T-1-j})' \right)$. Since $\Delta y_1 = \Delta \tilde{x} + \epsilon$ and

$$\Delta y_t = \rho^{t-1} \Delta y_1 + \sum_{j=0}^{t-2} \rho^j \Delta x_{t-j} + \sum_{j=0}^{t-2} \rho^j B^{-1} \Delta v_{t-j} \text{ for } t = 2, 3, ..., \quad (B.6)$$

we have

$$\Delta Y^* = k_p \otimes \Delta y_1 + \mathcal{X} + (I_T \otimes B^{-1}) \mathcal{V}.$$ 

So

$$\Delta Y^* R \Omega^{-1} \Delta u = \mathcal{X}' R \Omega^{-1} \Delta u + (k_p \otimes \Delta y_1') R \Omega^{-1} \Delta u + \mathcal{V}' (I_T \otimes B^{-1}) R \Omega^{-1} \Delta u \equiv \sum_{j=1}^3 A_{n_j}.$$ 

It is easy to show that $E (A_{n_1}) = 0$ and $n^{-1} A_{n_1} = o_p(1)$. Now after some tedious calculations we can show that

$$E (A_{n_2} + A_{n_3}) = E \left\{ R \Omega^{-1} \Delta u \left[ (k_p \otimes \Delta y_1') + \mathcal{V}' (I_T \otimes B^{-1}) \right] \right\} = \sigma^2_t \text{tr} \left\{ R \Omega^{-1} \left[ C_1 \otimes (B'B)^{-1} + C_2 \otimes I_n \right] \right\}.$$ 

and that $E (A_{n_j}^2) = O (n)$ for $j = 2, 3$. The first conclusion then follows from the Chebyshev inequality.

When $R = I_{n_T}$, write

$$E (A_{n_2} + A_{n_3}) = \sigma^2_t \text{tr} \left( \Omega^{-1} \left( C_1 \otimes (B'B)^{-1} \right) \right) + \sigma^2_t \text{tr} \left( \Omega^{-1} (C_2 \otimes I_n) \right) \equiv \overline{A}_{n_2} + \overline{A}_{n_3}.$$ 

Using (3.39) and the explicit expressions for $h_{0,-1}$ and $h_{1,-1}$, we can show that

$$\overline{A}_{n_2} = -c \phi \sigma^2 \text{tr} \left( E^{*-1} BB' \right) \text{ and } \overline{A}_{n_3} = -c \phi \sigma^2 \text{tr} \left( E^{*-1} BB' \right),$$

where $c = \sum_{j=1}^{T-1} (T-j) \rho^{j-1}$. Consequently, $E (A_{n_2} + A_{n_3}) = 0$. This completes the proof. \hfill \blacksquare
Lemma B.17 Suppose that the conditions in Theorem 4.6 are satisfied. Then

1) \( \frac{1}{\sqrt{nT}} \Delta X' \Omega^{-1} \Delta u \xrightarrow{d} N(0, \Gamma_{f,11}) \), where \( \Gamma_{f,11} = \text{plim}_{n \to \infty} (nT)^{-1} \Delta X' \Omega^{-1} \Delta X \).

2) \( \frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}(\omega)}{\partial \omega} \xrightarrow{d} N(0, \Gamma_f) \).

Proof. For 1), by the Cramer-Wold device, it suffices to show that for any \( c = (c_1', c_2, c_3') \in \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^k \) such that \( \|c\| = 1 \), \( (nT)^{-1/2} \Delta X' \Omega^{-1} \Delta u \xrightarrow{d} N(0, c \Gamma_{f,11} c) \). As in the proof of Lemma B.16, let \( \Delta X^* = (0_{p \times n}, \Delta x_2', ..., \Delta x_k') \), \( \Delta Y^* = (0_{n \times n}, \Delta y_1', ..., \Delta y_{T-1}') \), \( \Delta \hat{x}^* = (\Delta \hat{x}', 0_{k \times n(T-1)}) \). Then

\[
T_n \equiv \Delta X' \Omega^{-1} \Delta u = (c_1' \Delta X^* + c_2' \Delta Y^* + c_3' \Delta x^*) \Omega^{-1} \Delta u.
\]

By the proof of Lemma B.16,

\[
\Delta Y^* = X + (k_\rho \otimes \Delta y_1) + (I_T \otimes B^{-1}) V.
\]

Let \( V^* = (v_{-m+1}', v_{-m+2}', ..., v_T') \). We can write \( \Delta y_1 = \Delta \hat{x} + \zeta + \sum_{j=0}^{m-1} \rho^j B_1 \Delta v_{-j} = \Delta \hat{x} + \zeta + C_1 V^* \), \( V = C_2 V^* \) and \( \Delta u = C_4 \zeta + C_3 V^* \) for some matrices \( C_i \), \( i = 1, 2, 3, 4 \). Consequently,

\[
\Delta Y^* = X + k_\rho \otimes \zeta + C_5 V^*.
\]

where \( X' = X + k_\rho \otimes (\Delta \hat{x}) \), and \( C_5 = (k_{\rho,1} C_1', ..., k_{\rho,T} C_1')' + (I_T \otimes B^{-1}) C_2 \). As a result, we can write

\[
T_n = (c_1' \Delta X^* + c_2' (X' + k_\rho \otimes \zeta + C_5 V^*) + c_3' \Delta x^*) \Omega^{-1} (C_4 \zeta + C_3 V^*)
\]

\[
= (a_1' \zeta + \zeta' A_1 \zeta) + (a_2' V^* + V^* A_2 V^*) + \zeta' A_3 V^*.
\]

\[
= T_{n1} + T_{n2} + T_{n3}.
\]

where \( a_i \) and \( A_i \) are vectors or matrices that involve \( \Delta X^* \), \( X' \), \( \Delta x^* \), and the nonstochastic matrices \( C_j \) \( j = 3, 4, 5 \). By Lemma B.7,

\[
\{T_{ni} - E(T_{ni})\}/\sqrt{\text{Var}(T_{ni})} \xrightarrow{d} N(0, 1) \text{,}
\]

By the facts that \( E\left(\sum_{i=1}^{3} T_{ni}\right) = 0 \) and that \( T_{ni}, i = 1, 2, 3 \), are asymptotically independent, we have

\[
\frac{1}{\sqrt{nT}} T_n = \frac{1}{\sqrt{nT}} \sum_{i=1}^{3} T_{ni} \xrightarrow{d} N \left(0, \lim_{n \to \infty} (nT)^{-1} \sum_{i=1}^{3} \text{Var}(T_{ni})\right),
\]

implying that \( (nT)^{-1/2} \Delta X' \Omega^{-1} \Delta u \xrightarrow{d} N(0, \Gamma_{f,11}) \) because we can show that \( (nT)^{-1} (\Delta X' \Omega^{-1} \Delta X - \text{Var}(\Delta X' \Omega^{-1} \Delta u)) = o_p(1) \). The proof of 2) is similar and thus omitted. ■
The next three Lemmas are used in simplifying the covariance-covariance matrix of the score function.

**Lemma B.18**  Suppose that the conditions in Theorem 4.2 are satisfied. Then

\[
E \left( \tilde{X}' \tilde{q}_n \tilde{u}' p_n u \right) = \begin{pmatrix}
E \left( \mu_1^T \right) X' g_{q_n,1} \text{diag} \left( G_{p_n,1} \right) + E \left( v_1^T \right) X' g_{q_n,2} \text{diag} \left( G_{p_n,2} \right) \\
E \left( \mu_2^T \right) Z' g_{q_n,1} \text{diag} \left( G_{p_n,1} \right) + E \left( v_1^T \right) Z' g_{q_n,2} \text{diag} \left( G_{p_n,2} \right) \\
E \left( Y_1' \right) q_n \tilde{u}' p_n u \\
E \left( Y_2' \right) q_n \tilde{u}' p_n u
\end{pmatrix},
\]

where \( E \left( Y_1' q_n \tilde{u}' p_n u \right) = \sigma_2^2 E \left( Y_1' \Omega p_n u \right) + E \left( \mu_2^T \right) \Omega' g_{q_n,1} \text{diag} \left( G_{p_n,1} \right) + E \left( v_1^T \right) \Omega' g_{q_n,2} \text{diag} \left( G_{p_n,2} \right) + \kappa_\mu \sum_{i=1}^n \left( \left( I_{p} \otimes I_n \right) q_n \left( \mu \otimes I_n \right) \right)_{ii} G_{p_n,1} + \kappa_\mu \sum_{i=1}^n \left( \left( \mu \otimes B^{-1} \right) \right)_{ii} G_{p_n,2}, \)
\( Q_n = E \left[ A \tilde{X}' \beta_0 + \left( \mu \otimes I_n \right) \gamma_0 + \Omega_0 \right], \) \( g_{q_n,1} = q_n \left( \mu \otimes I_n \right), \) \( g_{q_n,2} = q_n \left( \mu \otimes B^{-1} \right), \) and \( A, \mu, \gamma_0 \) and \( I_n \) are defined in (B.2) and (B.5).

**Proof.** The proof is similar to that of Lemma B.4.  

**Lemma B.19**  Let \( q_n \) and \( p_n \) be \( n (T + 1) \times n (T + 1) \) symmetric nonstochastic matrix. Suppose that the conditions in Theorem 4.4 are satisfied. Then

1) \( E \left( X' q_n u^* u^* p_n u^* \right) = E \left( \mu_1^T \right) X' q_n d_3 \text{diag} \left( d_3' p_n d_3 \right) + E \left( \mu_2^T \right) X' q_n d_3 \text{diag} \left( d_3' p_n d_3 \right), \)

2) \( E \left( u^* q_n u^* u^* p_n u^* \right) = \sigma_3^2 \text{tr} \left( \Omega' q_n \right) \text{tr} \left( \Omega' p_n \right) + \kappa_\mu \sum_{i=1}^n \left( d_1' q_n d_1 \right)_{ii} + \kappa_\mu \sum_{i=1}^n \left( d_2' q_n d_2 \right)_{ii} \times \left( d_2' p_n d_2 \right)_{ii} + \kappa_\mu \sum_{i=1}^n \left( d_3' q_n d_3 \right)_{ii} + \kappa_\mu \sum_{i=1}^n \left( d_3' p_n d_3 \right)_{ii} \times \left( d_3' d_3 \right)_{ii}, \)

where \( d_i' \) and \( a_2 \) are defined in (B.8).

**Proof.** Write

\[
u^* = \left( \begin{array}{c}
\zeta + \frac{\mu}{1 - \rho} + B^{-1} a_2 \\
\left( \mu \otimes I_n \right) \mu + \left( \mu \otimes B^{-1} \right) v
\end{array} \right) = d_1 \mu + d_2 \zeta + d_3 a_2 + d_4 v,
\]

where

\[
d_1 = \left( \begin{array}{c}
\frac{\mu}{1 - \rho} \\
\mu \otimes I_n
\end{array} \right) \otimes I_n, \quad d_2 = \left( \begin{array}{c}
1 \\
0_{T \times 1}
\end{array} \right) \otimes I_n, \quad d_3 = \left( \begin{array}{c}
1 \\
0_{T \times 1}
\end{array} \right) \otimes B^{-1},
\]

\[
d_4 = \left( \begin{array}{c}
0_{1 \times 1} \quad 0_{1 \times T} \\
0_{T \times 1} \quad I_T
\end{array} \right) \otimes B^{-1}, \quad \text{and} \quad a_2 = \sum_{j=0}^{\infty} \rho^j v_{-j}.
\]

Using the fact that \( \mu, \zeta, a_2, \) and \( v \) are mutually independent, we can readily show 1) and 3). For
example, for 3), we can apply Lemma B.3 to each term in (B.7) to get

\[ E(u^*q_nu^*u^*p_nu^*) \]

\[ = \sigma_{\mu}^4 [\text{tr} (d'_1q_n d_1) \text{tr} (d'_1p_n d_1)] + \sigma_{\mu}^4 \sum_{i=1}^n (d'_1q_n d_1)_{ii} (d'_1p_n d_1)_{ii} \]

\[ + \sigma_{v}^4 [\text{tr} (d'_2q_n d_2) \text{tr} (d'_2p_n d_2)] + \sigma_{v}^4 \sum_{i=1}^n (d'_2q_n d_2)_{ii} (d'_2p_n d_2)_{ii} \]

\[ = \sigma_{\mu}^4 [\text{tr} (\Omega^*q_n) \text{tr} (\Omega^*p_n) + \kappa_{\mu} \sum_{i=1}^n (d'_1q_n d_1)_{ii} (d'_1p_n d_1)_{ii} + \kappa_{\nu} \sum_{i=1}^n (d'_2q_n d_2)_{ii} (d'_2p_n d_2)_{ii} \]

\[ + \kappa_{\nu} \sum_{i=1}^n (d'_3q_n d_3)_{ii} (d'_3p_n d_3)_{ii} + \kappa_{v} \sum_{i=1}^n (d'_4q_n d_4)_{ii} (d'_4p_n d_4)_{ii} \].

\[ \]  

**Lemma B.20** Let \( q_n \) and \( p_n \) be \( nT \times nT \) symmetric nonstochastic matrix. Suppose that the conditions in Theorem 4.6 are satisfied. Write \( \Delta X = (\Delta X^*, \Delta Y^*, \Delta w^*) \) as in the proof of Lemma B.16. Then

1) \[ E(\Delta X^*q_n \Delta w u^*p_n \Delta u) \]

\[ = \left( E\left(\tilde{v}_1\right) \Delta X^*q_n^{(1)} \text{tr} (p_{11}) + E\left(\tilde{v}_1\right) \Delta X^*q_n^{(2)} \text{tr} (B'^{-1} p_{11} B^{-1}) + E\left(\tilde{v}_1\right) \Delta X^*q_n^{(1)} \text{tr} (d'^*p_n d) \right) \]

\[ + \left( E\left(\tilde{v}_1\right) \Delta X^*q_n^{(2)} \text{tr} (B'^{-1} p_{11} B^{-1}) + E\left(\tilde{v}_1\right) \Delta X^*q_n^{(1)} \text{tr} (d'^*p_n d) \right) \]

2) \[ E(\Delta u^*q_n \Delta w u^*p_n \Delta u) = \sigma_{\mu}^4 \text{tr} (\Omega^*q_n) \text{tr} (\Omega^*p_n) + \kappa_{\mu} \sum_{i=1}^n q_{11,ii} p_{11,ii} + \kappa_{\nu} \sum_{i=1}^n \left( B'^{-1} q_{11,ii} (B'^{-1} B)^{-1} \right)_{ii} \]

\[ \]  

where \( E(\Delta X^*q_n \Delta w u^*p_n \Delta u) = \sigma_{\mu}^2 E(\Delta X^*q_n \Omega^*p_n \Delta u) + E(\mu_{\mu}^2) Q_{n}^{(1)} q_{n,i}^{(-1)} \text{tr} (G_{p,n}) \]

\[ + \kappa_{\nu} \sum_{i=1}^n \left( \tilde{v}_1 \right) \text{tr} (p_{11} B^{-1}) \]

\[ + \kappa_{\nu} \sum_{i=1}^n \left( \tilde{v}_1 \right) \text{tr} (p_{11} B^{-1}) \]

\[ \]

\[ \tilde{v}_i = \sum_{j=0}^{n} (I^{p}_j \otimes I_n) q_n, p_{11} = (0_{n \times n(T-1)}, 0_{n \times n(T-1)}, 0_{n \times n(T-1)}, 0_{n \times n(T-1)}), q_{11} = (0_{n \times n(T-1)}, 0_{n \times n(T-1)}), q_{21} = (0_{n \times n(T-1)}, 0_{n \times n(T-1)}), \]

\[ \text{and } C \text{ is defined in the proof of Lemma B.15.} \]

**Proof.** The proof is similar to that of Lemma B.19 and thus omitted. \( \square \)
C Proofs of the Theorems

Proof of Theorem 4.1

By Theorem 3.4 of White (1994), it suffices to show that

\[(nT)^{-1} [\mathcal{L}^*_c (\delta) - \mathcal{L}^*_c (\delta_0)] \overset{p}{\to} 0 \text{ uniformly in } \delta \in \Lambda, \tag{C.1} \]

and

\[\limsup_{n \to \infty} \max_{\rho \in \mathcal{N}_c^* (\rho_0)} (nT)^{-1} [\mathcal{L}^*_c (\delta) - \mathcal{L}^*_c (\delta_0)] < 0 \text{ for any } \epsilon > 0, \tag{C.2} \]

where \( \mathcal{N}_c^* (\rho_0) \) is the complement of an open neighborhood of \( \rho_0 \) on \( \Lambda \) of diameter \( \epsilon \). By (3.5) and (4.3), \( (nT)^{-1} [\mathcal{L}^*_c (\delta) - \mathcal{L}^*_c (\delta_0)] = -\frac{1}{nT} \left( \ln \hat{\sigma}^2_v (\delta) - \ln \hat{\sigma}^2_v (\delta) \right) \). To show (C.1), it is sufficient to show

\[\hat{\sigma}^2_v (\delta) - \hat{\sigma}^2_v (\delta) = o_p (1) \text{ uniformly on } \Lambda. \tag{C.3} \]

By (3.4), we can write

\[\hat{\sigma}^2_v (\delta) = \frac{1}{nT} \bar{u} (\delta)' \Omega^{-1} \bar{u} (\delta) = \frac{1}{nT} Y' \Omega^{-1/2} M \Omega^{-1/2} Y, \]

where \( M = U_n = \Omega^{-1/2} \tilde{X} (\tilde{X}' \Omega^{-1} \tilde{X})^{-1/2} \tilde{X} \Omega^{-1/2} \). Noting that \( M \Omega^{-1/2} \tilde{X} = 0 \), we have \( \hat{\sigma}^2_v (\delta) = (nT)^{-1} u' \Omega^{-1/2} M \Omega^{-1/2} u \). Then by (4.2), we have

\[\hat{\sigma}^2_v (\delta) - \hat{\sigma}^2_v (\delta) = \frac{1}{nT} \left[ u' \Omega^{-1/2} M \Omega^{-1/2} u - \sigma^2_u tr \left( \Omega^{-1/2} M \Omega^{-1/2} \Omega_0 \right) \right] \]

\[+ \frac{\sigma^2_u}{nT} \left[ tr \left( \Omega^{-1/2} M \Omega^{-1/2} \Omega_0 \right) - tr \left( \Omega^{-1/2} \Omega_0 \right) \right] \equiv T_{n1} + T_{n2}. \]

We can readily show that both \( T_{n1} \) and \( T_{n2} \) are \( o_p (1) \) uniformly on \( \Lambda \).

To show (C.2), we follow Lee (2002) and define an auxiliary process: \( Y_{nT} = U_{nT} \), where \( U_{nT} \sim N (0, \sigma^2_v \Omega) \) with \( \Omega = \Omega (\delta) \). The true parameter value is given by \( (\sigma^2_v, \delta_0) \). Let \( \Omega_0 = \Omega (\delta_0) \). The log likelihood function of the auxiliary process is \( \mathcal{L}_a^* (\delta, \sigma^2_v) = -\frac{1}{2} n T \log (2\pi) - \frac{n T}{2} \log (\sigma^2_v) - \frac{1}{2} \log |\Omega| - \frac{1}{2 \sigma^2_v} U_{nT}' \Omega^{-1} U_{nT} \). We can verify that \( \mathcal{L}_a^* (\delta) = \max_{\sigma^2_v} E_a \mathcal{L}_a^* (\delta, \sigma^2_v) \), where \( E_a \) is the expectation under the auxiliary process. By the Jensen inequality, \( \mathcal{L}_a^* (\delta) \leq E_a \mathcal{L}_a^* (\delta_0, \sigma^2_v) = \mathcal{L}_a^* (\delta_0) \) for all \( \delta \). Suppose the identification uniqueness condition in (C.2) is not satisfied, then there would exist a sequence \( \delta_n \in \Lambda \) such that \( \delta_n \to \delta^* \neq \delta_0 \), and \( \lim_{n \to \infty} (nT)^{-1} [\mathcal{L}_a^* (\delta^*) - \mathcal{L}_a^* (\delta_0)] = 0 \). The latter would contradict to Assumption R(iv).

This completes the proof of the theorem. \( \blacksquare \)
Proof of Theorem 4.2

By the Taylor series expansion

\[ 0 = \frac{1}{\sqrt{nT}} \frac{\partial L_r^r}{\partial \varsigma} = \frac{1}{\sqrt{nT}} \frac{\partial L_r^r (\varsigma_0)}{\partial \varsigma} + \frac{1}{nT} \frac{\partial^2 L_r^r (\tilde{\varsigma})}{\partial \varsigma \partial \varsigma^r} \sqrt{nT} (\tilde{\varsigma} - \varsigma_0), \]

where elements of \( \tilde{\varsigma} \) lie in the segment joining the corresponding elements of \( \tilde{\varsigma} \) and \( \varsigma_0 \). Thus

\[ \sqrt{nT} (\tilde{\varsigma} - \varsigma_0) = - \left[ \frac{1}{nT} \frac{\partial^2 L_r^r (\tilde{\varsigma})}{\partial \varsigma \partial \varsigma^r} \right]^{-1} \frac{1}{\sqrt{nT}} \frac{\partial L_r^r (\varsigma_0)}{\partial \varsigma}. \]

By Theorem 4.1, \( \tilde{\varsigma} \xrightarrow{p} \varsigma_0 \). Consequently, \( \tilde{\varsigma} \xrightarrow{d} \varsigma_0 \), and it suffices to show that

\[ \frac{1}{nT} \frac{\partial^2 L_r^r (\tilde{\varsigma})}{\partial \varsigma \partial \varsigma^r} \xrightarrow{p} \Sigma^r, \quad (C.4) \]

and

\[ \frac{1}{\sqrt{nT}} \frac{\partial L_r^r (\varsigma_0)}{\partial \varsigma} \xrightarrow{d} N(0, \Sigma^r + \Lambda^r). \quad (C.6) \]

As \( \tilde{\varsigma} \xrightarrow{p} \varsigma_0 \), it follows that \( \tilde{\Omega} = \Omega \left( \tilde{\phi}, \tilde{\lambda} \right) \xrightarrow{p} \Omega_0 \). By Lemmas B.2, B.5 and B.6, and the fact that \( u \left( \tilde{\theta} \right) = Y - \tilde{X} \tilde{\theta} = u + \tilde{X} \left( \tilde{\theta}_0 - \tilde{\theta} \right) \), (C.4) holds for each of its component, e.g., \( \partial^2 L_r^r (\tilde{\varsigma}) / \partial \theta \partial \theta^r \). The result in (C.5) follows from Lemmas B.5 and B.6. (C.6) is proved in Lemma B.8.

Proof of Theorem 4.3

The proof is almost identical to that of Theorem 4.1 and thus omitted.

Proof of Theorem 4.4

The proof is analogous to that of Theorem 4.2 and now follows mainly from Lemmas B.12-B.14.

Proof of Theorem 4.5

The proof is almost identical to that of Theorem 4.1 and thus omitted.

Proof of Theorem 4.6

The proof is analogous to that of Theorem 4.2 and now follows mainly from Lemmas B.15-B.17.
Table 1. Monte Carlo Mean and RMSE for the QMLEs
Random Effects Model with Normal Errors, n = 50, T = 3

<table>
<thead>
<tr>
<th>r</th>
<th>Mean Rmse</th>
<th>Mean Rmse</th>
<th>Mean Rmse</th>
<th>Mean Rmse</th>
<th>Mean Rmse</th>
<th>Mean Rmse</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>0.731</td>
<td>0.070</td>
<td>0.720</td>
<td>0.074</td>
<td>0.494</td>
<td>0.038</td>
</tr>
<tr>
<td>0.51</td>
<td>0.732</td>
<td>0.071</td>
<td>0.721</td>
<td>0.072</td>
<td>0.494</td>
<td>0.038</td>
</tr>
<tr>
<td>0.52</td>
<td>0.733</td>
<td>0.072</td>
<td>0.722</td>
<td>0.073</td>
<td>0.494</td>
<td>0.038</td>
</tr>
<tr>
<td>0.53</td>
<td>0.734</td>
<td>0.073</td>
<td>0.723</td>
<td>0.074</td>
<td>0.494</td>
<td>0.038</td>
</tr>
<tr>
<td>0.54</td>
<td>0.735</td>
<td>0.074</td>
<td>0.724</td>
<td>0.075</td>
<td>0.494</td>
<td>0.038</td>
</tr>
<tr>
<td>0.55</td>
<td>0.736</td>
<td>0.075</td>
<td>0.725</td>
<td>0.076</td>
<td>0.494</td>
<td>0.038</td>
</tr>
<tr>
<td>0.56</td>
<td>0.737</td>
<td>0.076</td>
<td>0.726</td>
<td>0.077</td>
<td>0.494</td>
<td>0.038</td>
</tr>
<tr>
<td>0.57</td>
<td>0.738</td>
<td>0.077</td>
<td>0.727</td>
<td>0.078</td>
<td>0.494</td>
<td>0.038</td>
</tr>
<tr>
<td>0.58</td>
<td>0.739</td>
<td>0.078</td>
<td>0.728</td>
<td>0.079</td>
<td>0.494</td>
<td>0.038</td>
</tr>
<tr>
<td>0.59</td>
<td>0.740</td>
<td>0.079</td>
<td>0.729</td>
<td>0.080</td>
<td>0.494</td>
<td>0.038</td>
</tr>
<tr>
<td>0.60</td>
<td>0.741</td>
<td>0.080</td>
<td>0.730</td>
<td>0.081</td>
<td>0.494</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Note: Under each r value, first two columns correspond to estimates treating y0 as exogenous, whereas the last two columns correspond to estimates treating y0 as endogenous. Under each r value, the seven rows correspond to, respectively, $\beta_0 (= 5), \beta_1 (= 1), \gamma (= 1), r, \lambda, \sigma_\epsilon (= .5) \text{ and } \sigma_\gamma (= .5)$. 

48
### Table 2. Monte Carlo Mean and RMSE for the QMLEs

<table>
<thead>
<tr>
<th></th>
<th>$\lambda = .25$</th>
<th>$\lambda = .50$</th>
<th>$\lambda = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>RMSE</td>
<td>Mean</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.236</td>
<td>0.028</td>
<td>0.238</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.248</td>
<td>0.124</td>
<td>0.484</td>
</tr>
<tr>
<td>$\sigma_v$</td>
<td>0.489</td>
<td>0.038</td>
<td>0.489</td>
</tr>
</tbody>
</table>

The fixed effects are randomly generated

### Table 2. Monte Carlo Mean and RMSE for the QMLEs

<table>
<thead>
<tr>
<th></th>
<th>$\lambda = .25$</th>
<th>$\lambda = .50$</th>
<th>$\lambda = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>RMSE</td>
<td>Mean</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.481</td>
<td>0.033</td>
<td>0.484</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.246</td>
<td>0.123</td>
<td>0.488</td>
</tr>
<tr>
<td>$\sigma_v$</td>
<td>0.490</td>
<td>0.037</td>
<td>0.491</td>
</tr>
</tbody>
</table>

The fixed effects are average (over $T$) of the $X$ values

### Table 2. Monte Carlo Mean and RMSE for the QMLEs

<table>
<thead>
<tr>
<th></th>
<th>$\lambda = .25$</th>
<th>$\lambda = .50$</th>
<th>$\lambda = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>RMSE</td>
<td>Mean</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.719</td>
<td>0.042</td>
<td>0.722</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.239</td>
<td>0.125</td>
<td>0.490</td>
</tr>
<tr>
<td>$\sigma_v$</td>
<td>0.488</td>
<td>0.038</td>
<td>0.489</td>
</tr>
</tbody>
</table>

Note: $\beta_0 = 5$, $\beta_1 = 1$, $\gamma = 1$, $\sigma_\mu = .5$ and $\sigma_v = .5$. 

49