QML Estimation of Dynamic Panel Data Models with Spatial Errors

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Z. Yang

Follow this and additional works at: http://ink.library.smu.edu.sg/soe_research
Part of the Economics Commons

Citation
Available at: http://ink.library.smu.edu.sg/soe_research/1490
QML Estimation of Dynamic Panel Data Models with Spatial Errors

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

October 9, 2012

Abstract

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


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

1 Introduction

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

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

†Corresponding Author: 90 Stamford Road, Singapore 178903. Phone: +65-6828-0852; Fax: +65-6828-0833. Email: zlyang@smu.edu.sg.
data (SPD) models.\textsuperscript{1} Adding a dynamic element into a SPD model further increases its flexibility, which has, since Anselin (2001), attracted the attention of many econometricians. The spatial dynamic panel data (SDPD) models can be broadly classified into two categories: one is that described in Anselin et al. (2008) where the dynamic and spatial effects both appear in the model in the forms of lags (in time and spatial) of the response variable, and the other allows the dynamic effect in the same manner but builds the spatial effects into the disturbance term. The former has been studied by Yu et al. (2007, 2008) and Yu and Lee (2007), and the latter by Elhorst (2005), Yang et al. (2006), Mutl (2006), Su and Yang (2007), and Lee and Yu (2010b). Lee and Yu (2010c) provide an excellent survey on the spatial panel data models (static and dynamic) and report some recent developments.

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

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

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

\textsuperscript{2}A panel with large $n$ and small $T$, called a short panel, remains the prevalent setting in the majority of empirical microeconometric research (Binder et al., 2005), and evidence from the standard dynamic panel data models shows that QML estimators are more efficient than GMM estimators (Hsiao et al., 2002; Binder et al., 2005).
large-\(n\) and large-\(T\) setting as considered in Yu et al. (2008). Third, spatial dependence is present only in the error term in our model whereas Yu et al. (2008) consider spatial lag model. It would be interesting to extend our work to the SDPD model with spatial lag, or with both spatial lag and spatial error.

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

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

## 2 Model Specification

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

\[
y_{it} = \rho y_{it-1} + x_{it}' \beta + z_{it}' \gamma + u_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{2.1}
\]

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

\[
u_{t} = \mu + \varepsilon_{t}, \tag{2.2}
\]

\[
\varepsilon_{t} = \lambda W_{n} \varepsilon_{t} + v_{t}, \tag{2.3}
\]

where \(\mu = (\mu_{1}, \cdots, \mu_{n})'\), \(\varepsilon_{t} = (\varepsilon_{1t}, \cdots, \varepsilon_{nt})'\), and \(v_{t} = (v_{1t}, \cdots, v_{nt})'\), with \(\mu\) representing the unobservable individual or space-specific effects, \(\varepsilon_{t}\) representing the spatially correlated errors, and \(v_{t}\) representing the random innovations that are assumed to be independent and identically distributed (iid) with mean zero and variance \(\sigma_{\varepsilon}^{2}\). The parameter \(\lambda\) is a spatial autoregressive coefficient and \(W_{n}\) is a known \(n \times n\) spatial weight matrix whose diagonal elements are zero.\(^{3}\)

Denoting \(y_{t} = (y_{1t}, \cdots, y_{nt})'\), \(x_{t} = (x_{1t}, \cdots, x_{nt})'\), and \(z = (z_{1}, \cdots, z_{n})'\), the model has the following reduced-form representation,

\[
y_{t} = \rho y_{t-1} + x_{t}' \beta + z_{t}' \gamma + u_{t}, \quad \text{with} \quad u_{t} = \mu + B_{n}^{-1} v_{t}, \quad t = 1, \ldots, T, \tag{2.4}
\]

\(^{3}\)It is worth mentioning that Eqs (2.1)-(2.3) allow spatial dependence to be present in the random disturbance term \(\varepsilon_{t}\) 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 in our model as is done by Kapoor et al. (2007) who consider GMM estimation of a static spatial panel model with random effects. Our theory can readily be modified to take into account of the latter case, and we conjecture that a specification test can also be developed to test for the two different specifications.
where $B_n = I_n - \lambda W_n$. The following specifications are essential for the subsequent developments.

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

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

3 The QML Estimators

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

3.1 QMLEs for the random effects model

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

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

4It is well known that when $T$ is fixed the likelihood function for a dynamic panel model depends on the assumptions on the initial observations (Hsiao, 2003). For example, if $|p| \geq 1$ or the process $\{x_t\}$ are not stationary, then it does not make sense to assume that the process generating the $y_t$ is the same prior to the periods of observations for $t = 1, \ldots, T$. In this case, it is reasonable to treat $y_0$ as exogenous. Otherwise, $y_0$ should be treated as endogenous. In general, $y_0$ is considered to be exogenous when it is reasonable to expect that $y_0$ varies “autonomously”, independently of the other variables in the model, otherwise it is considered as endogenous.
Conditional on the observed (exogenous) \( y_0 \), the distribution of \( y_1 \) can be easily derived, and hence the Gaussian quasi-likelihood function based on the observations \( y_1, y_2, \ldots, y_T \). Define \( Y = (y_1', \ldots, y_T')' \), \( Y_{\text{T}-1} = (y_0, \ldots, y_{\text{T}-1})' \), \( X = (x_1', \ldots, x_T')' \), \( Z = \nu_T \otimes z \), and \( v = (v_1', \ldots, v_T')' \). The SDPD model specified by (2.1)-(2.3) can be written in matrix form:

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

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

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

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

The distribution of \( u \) leads to the distribution of \( Y - \rho Y_{\text{T}-1} \), and hence the distribution of \( Y \) as the Jacobian of the transformation is one. Let \( \theta = (\beta', \gamma', \rho)' \), \( \delta = (\lambda, \phi_\mu)' \), and \( \psi = (\theta', \sigma_v^2, \delta)' \). Denoting \( u(\theta) = Y - \rho Y_{\text{T}-1} - X \beta - Z \gamma \), the quasi-log-likelihood function of \( \psi \) is

\[
L'(\psi) = -\frac{n_T}{2} \log(2\pi) - \frac{n_T}{2} \log(\sigma_v^2) - \frac{1}{2} \log |\Omega| - \frac{1}{2 \sigma_v^2} u(\theta)' \Omega^{-1} u(\theta). \tag{3.3}
\]

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

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

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

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

The QMLE \( \hat{\delta} = (\hat{\lambda}, \hat{\phi}_\mu)' \) of \( \delta \) 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 \( \sigma_v^2 \) is given by \( \hat{\sigma}_v^2 = \hat{\phi}_\mu \hat{\sigma}_v^2 \).

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

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

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

(3.6)

Letting $\eta_0$ and $\zeta_0$ be, respectively, the exogenous and endogenous components of $y_0$, we have

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

(3.7)

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

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

(3.8)

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

However, both $\{x_{-j}, j = 1, \ldots, m - 1\}$ for $m \geq 2$ and $y_{-m}$ for $m \geq 1$ in $\eta_m$ are unobserved, rendering that (3.7) cannot be used as a model for $\eta_0$. Some approximations are necessary. In this paper, we follow Bhargava and Sargan (1983) (see also Hsiao, 2003, p.76) and propose a model for the initial observations based on the following fundamental assumptions. Let $x \equiv (x_0, x_1, \ldots, x_T)$.

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

These assumptions lead immediately to the following model for $\eta_m$:

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

(3.9)

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

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

(3.10)

The ‘initial’ error vector $u_0$ is seen to contain three components: $\zeta$, $\mu \frac{1 - \rho^m}{1 - \rho}$, and $\sum_{j=0}^{m-1} \rho^j B^{-1} u_{-j}$, being, respectively, the prediction error from predicting the unobservables, the cumulative random effects up to the 0th period, and the ‘cumulative’ spatial effects and random shocks up to the 0th period. The term $z_m(\rho) \gamma = z \frac{1 - \rho^m}{1 - \rho}$ represents the cumulative impact of the time-invariant variables $z$ up to period 0 and needs not be predicted. However, the predictors for $\eta_m$ still include $z$, indicating that (i) the mean of $y_{-m}$ is allowed to be linearly related to $z$ and (ii) $\rho^m$ may not be small such that the effect of $y_{-m}$ on $\eta_m$ is not negligible. If $\rho^m$ is small which occurs when either $m$ is large or $\rho$ is small, the impact of $y_{-m}$
to \( \eta_m \) can be ignored, and the term \( z_\pi \) in (3.10) should be omitted. Some details about the cases with small \( \rho^m \) are given latter. For the cases where \( \rho^m \) is not negligible, one can easily show that, under strict exogeneity of \( x \) and \( z \), \( E(u_0) = 0 \),

\[
E(u_0u_0') = \sigma^2 \sigma^2 + \sigma^2 \sigma^2 m^2 \sigma^2 \sigma^2 m^2 (B'B)^{-1}, \quad \text{and } E(u_0u') = \sigma^2 a_m (I_T \otimes I_n),
\]

where \( a_m \equiv a_m (\rho) = \frac{1 - \rho^m}{1 - \rho^2} \) and \( b_m \equiv b_m (\rho) = \frac{1 - \rho^m}{1 - \rho^2} \). Let \( u^* = (u'_0, u')' \). Under the normality assumption for the original error components \( \mu \) and \( v \), and the ‘new’ prediction error \( \zeta \), we have \( u^* \sim N(0, \sigma^2 \Omega^*) \), where \( \Omega^* \) is \( n(T + 1) \times n(T + 1) \) and has the form:

\[
\Omega^* \equiv \Omega^*(\rho, \lambda, \phi, \phi) = \begin{pmatrix} \phi_c I_n + \phi_a a_m^2 m^2 \sigma^2 \sigma^2 m^2 (B'B)^{-1} + \phi_b m (B'B)^{-1} \sigma^2 \sigma^2 m^2 a_m (I_T \otimes I_n) & \phi_d \sigma^2 \sigma^2 m^2 a_m (I_T \otimes I_n) \\
\phi_d \sigma^2 \sigma^2 m^2 a_m (I_T \otimes I_n) & \Omega
\end{pmatrix}, \quad (3.11)
\]

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

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

\[
\mathcal{L}^\tau(\psi) = -\frac{n(T + 1)}{2} \log(2\pi) - \frac{n(T + 1)}{2} \log(\sigma^2) - \frac{1}{2} \log|\Omega^*| - \frac{1}{2 \sigma^2 v} u^*(\theta, \rho)' \Omega^{*-1} u^*(\theta, \rho), \quad (3.12)
\]

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

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

Maximizing (3.12) gives MLE of \( \psi \) if the error components are truly Gaussian and the QMLE otherwise. Similar to Case I, we work with the concentrated quasi-log-likelihood by concentrating out the parameters \( \theta \) and \( \sigma^2 \). The constrained QMLES of \( \theta \) and \( \sigma^2 \), given \( \delta \), are

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

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

\[
\mathcal{L}^\tau(\delta) = -\frac{n(T + 1)}{2} \log(2\pi) + 1 - \frac{n(T + 1)}{2} \log(\sigma^2) - \frac{1}{2} \log|\Omega^*|. \quad (3.14)
\]

Maximizing the concentrated quasi-log-likelihood (3.14) gives the QMLE of \( \delta \), denoted by \( \hat{\delta} = (\hat{\rho}, \hat{\lambda}, \hat{\phi}, \hat{\phi})' \). The QMLES of \( \theta \) and \( \sigma^2 \) are thus given by \( \hat{\theta}(\delta) \) and \( \hat{\sigma}_v^2(\delta) \), respectively, and those of \( \sigma^2_a \) and \( \sigma^2_c \) are given by \( \hat{\sigma}_a^2 = \hat{\phi}_a \hat{\sigma}_v^2 \) and \( \hat{\sigma}_c^2 = \hat{\phi}_c \hat{\sigma}_v^2 \), respectively.\footnote{Unlike the case of exogenous \( y_0 \), the dynamic parameter \( \rho \) now becomes a nonlinear parameter that has to be estimated, together with \( \lambda, \phi_\mu \), and \( \phi_c \), through a nonlinear optimization process.}

**Remark 1:** To utilize the information contained in the \( n \) endogenous initial observations \( y_0 \), we have introduced \( k = p(T + 1) + q + 1 \) additional parameters \( (\pi, \sigma^2) \) in the model (3.9). Besides the bias issue,
efficiency gain by utilizing additional n observations is reflected by n − k. Apparently, the condition n > k has to be satisfied in order for π and σ^2_ε to be identified. If both T and p are not so small (T = 9 and p = 10, say), one may consider to replace the regressors x in (3.9) by the most relevant ones (to the past), x_0 and x_1, say, or simply by \(x = (T + 1)^{-1} \sum_{t=0}^{T} x_t\). In this case \(k = 2p + q + 1,\) and \(p + q + 1,\) respectively.

**Remark II:** When \(y_0\) is exogenous, model (3.10) becomes \(y_0 = \tilde{x}_0 + u_0\), where \(u_0 \sim (0, \sigma^2_u I_n)\) and is independent of \(u\). In this case, we have \(\Omega^* = \text{diag} (\sigma^2_u I_n, \Omega)\). Model estimation may proceed by letting \(m = 0\) in (3.14), and the results are almost identical to those from maximizing (3.5). A special case of this is the one considered in Hsiao (2003, p.76, Case IIa) where \(y_{0t}^*\)'s are simply assumed to be iid independent of \(\mu_i\). If \(y_{0t}^*\)'s are allowed to be correlated with \(\mu_i\) (Case IIb, Hsiao, 2003, p.76), the model becomes a special case of endogenous \(y_0\) as considered above.

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

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

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

\[
\Omega^* = \begin{pmatrix}
(\phi_\zeta + \phi_\mu) I_n + (B' B)^{-1}, & \phi_\mu (v'_T \otimes I_n) \\
\phi_\mu (v'_T \otimes I_n), & \Omega
\end{pmatrix}
\]

In this case, \(\rho\) becomes a linear parameter again and the estimation can be simplified by putting \(\rho\) together with \(\beta, \gamma\) and \(\pi\) which can be concentrated out from the likelihood function. Now, denoting the response vector and the regressor matrix by:

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

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

(b) \(m = \infty\). When the process has reached stationarity \((m \to \infty\) and \(|\rho| < 1\), the model for the initial observations becomes \(y_0 = \sum_{j=0}^{\infty} \rho^j x_{-j} \beta + \frac{z_{2j}}{1 - \rho^2} + \frac{\tilde{z}_m}{1 - \rho} + \sum_{j=0}^{\infty} \rho^j B^{-1} v_{-j}\). As \(\eta_\infty = \sum_{j=0}^{\infty} \rho^j x_{-j} \beta\) involves only the time-varying predictors, its optimal predictors should be \((t_n, x)\). The estimation proceeds by letting \(z_m(\rho) = z_{\infty}(\rho) = \frac{1}{1 - \rho}, a_m = a_{\infty} = \frac{1}{1 - \rho}, b_m = b_{\infty} = \frac{1}{1 - \rho}, \tilde{x} = (t, x),\) and \(\pi = (\pi_1, \pi_2)'\).

(c) \(\rho = 0\). When the true value of the dynamic parameter is zero, the model becomes static with \(y_t = x_t \beta + z \gamma + \mu + B^{-1} v_t,\) \(t = 0, 1, \ldots, T\). At this point, the true values for all the added parameters, \(\pi\) and \(\sigma^2_\zeta,\) are automatically zero.
3.2 QMLEs for the fixed effects model

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

\[
\Delta y_t = \rho \Delta y_{t-1} + \Delta x_t \beta + \Delta u_t, \quad \Delta u_t = B^{-1} \Delta v_t, \quad t = 2, 3, \ldots, T. \tag{3.15}
\]

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

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

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

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

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

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

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

Clearly, the observations \( \Delta x_{1-j}, j = 1, \ldots, m - 1, m \geq 2 \), are not available, and the structure of \( E(\Delta y_{-m+1}), m \geq 1 \), is unknown. Hence \( \eta_m \) is completely unknown. Furthermore, as \( \eta_m \) is an \( n \times 1 \) vector, it cannot be treated as a free parameter vector to be estimated; otherwise the incidental parameters problem will be confronted again.\(^6\) Hsiao et al. (2002) remark that to get around this problem, the expected value of \( \eta_1 \), conditional on the observables, has to be a function of a finite number of

---

\(^6\)Unless the original model (2.4) does not contain time-varying variables as in Anderson and Hsiao (1981).
parameters, and that such a condition can hold provided that \( \{x_{it}\} \) are trend-stationary (with a common deterministic linear trend) or first-difference stationary processes. Letting \( \Delta x = (\Delta x_1, \cdots, \Delta x_T) \), we have the following fundamental assumptions.

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

Assumption F0(i) and Assumption F0(ii) lead immediately to a ‘predictive’ model for \( \eta_m \):

\[
\eta_m = \pi_1 x_n + \Delta x \pi_2 \equiv \Delta x \pi + \epsilon, \quad m = 0, 1, \cdots,
\]

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

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

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

\[
\Delta y_1 = \Delta x \pi + \Delta x_1 \beta + \epsilon + \Delta \zeta_1 \equiv \Delta x \pi + \Delta x_1 \beta + \Delta u_1,
\]

where \( \Delta u_1 = \epsilon + \Delta \zeta_1 = \epsilon - \rho^m e + \rho^m B^{-1} v_{m+1} + \sum_{j=0}^{m-1} \rho^j B^{-1} \Delta v_{1-j} \). Let \( \zeta = \epsilon - \rho^m e \). By assumption, the elements of \( \zeta \) are iid with variance \( \sigma^2_{\zeta} = \sigma^2_{\epsilon} + \sigma^2_e \rho^{2m} \).

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

\[
E(\Delta u_1 | \Delta u'_1) = \sigma^2_{\zeta} I_n + \sigma^2_e c_n (B'B)^{-1} = \sigma^2_{\epsilon} B^{-1} (\phi_e BB' + c_n I_n) B'^{-1}, \quad \text{and}
\]

\[
E(\Delta u_1 | \Delta u'_1) = -\sigma^2_{\zeta} (B'B)^{-1} \quad \text{for } t = 2, \quad \text{and for } t = 3, 4, \cdots, T,
\]

where \( c_n \equiv c_n(\rho) = \frac{-1}{1-\rho} - \frac{2(1-\rho)}{1-\rho} \) and \( \phi_e = \sigma^2_e/\sigma^2_{\epsilon} \). Note that \( c_0 = 1 \), \( c_\infty = \frac{-1}{1-\rho} \) and \( c_m(0) = 2 \).

Letting \( \Delta u = (\Delta u'_1, \Delta u'_2, \cdots, \Delta u'_T) \), we have \( \text{Var}(\Delta u) = \sigma^2_{\zeta} \Omega \), where

\[
\Omega^T \equiv \Omega \equiv \Phi \lambda \phi_\zeta = (I_T \otimes B^{-1}) \mathbb{H}(I_T \otimes B^{-1}),
\]

\[
E = \phi_\zeta BB' + c_n I_n, \quad \text{and } H_E \text{ is an } nT \times nT \text{ matrix defined as}
\]

\[
H_E = \left( \begin{array}{ccccccccc}
E & -I_n & 0 & \cdots & 0 & 0 & 0 \\
-I_n & 2I_n & -I_n & \cdots & 0 & 0 & 0 \\
0 & -I_n & 2I_n & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 2I_n & -I_n & 0 \\
0 & 0 & 0 & \cdots & -I_n & 2I_n & -I_n \\
0 & 0 & 0 & \cdots & 0 & -I_n & 2I_n \\
\end{array} \right).
\]
The expression for $\Omega^\dagger$ given in (3.22) greatly facilitates the calculation of the determinant and inverse of $\Omega^\dagger$ as seen in the subsequent subsection. Derivations of score and Hessian matrix requires the derivatives of $\Omega^\dagger$, which can be made much easier based on the following alternative expression

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

(3.24)

where $\ell_1$ is a $T \times T$ matrix with 1 in its top-left corner and zero elsewhere, and $h_{cm}$ is $h_s$ defined in Section 3.3 with $s$ replaced by $c_m$.

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

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

(3.25)

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

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

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

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

(3.26)

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

$$L^f_1(\delta) = -\frac{nT}{2} \log(2\pi) + 1 - \frac{nT}{2} \log(2\pi) + 1 - \frac{1}{2} \log |\Omega^\dagger|.$$

(3.27)

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

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

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

Maximization of \( \mathcal{L}_c^T(\delta) \), \( \mathcal{L}_c^T(\delta) \) and \( \mathcal{L}_c(\delta) \) involves repeated evaluations of the inverse and determinants of the \( nT \times nT \) matrices \( \Omega \) and \( \Omega^\dagger \), and the \( (n+1) \times (n+1) \) 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 \( \Omega \) defined in (3.1):

\[
|\Omega| = |(B'B)^{-1} + \phi_\mu TI_n| \cdot |B|^{-2(T-1)},
\]

\[
\Omega^{-1} = T^{-1}J_T \otimes ((B'B)^{-1} + \phi_\mu TI_n)^{-1} + (I_T - T^{-1}J_T) \otimes (B'B).
\]

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

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

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

For the determinant and inverse of \( \Omega^* \) defined in (3.11), let \( \omega_{11} = \phi_\mu I_n + \phi_\alpha a^\mu_n I_n + b_m(B'B)^{-1}, \)
\( \omega_{21} = \omega_{12} = \phi_\mu a_m(T \otimes I_n) \), and \( D = \omega_{11} - \omega_{12} \Omega^{-1} \omega_{21} \). We have by using the formulas for a partitioned matrix (e.g., Magnus and Neudecker, 2002, p.106), \(|\Omega^*| = |\Omega| \cdot |D|\), and

\[
\Omega^{*-1} = \begin{pmatrix}
D^{-1} & -D^{-1}\omega_{12}\Omega^{-1} \\
-\Omega^{-1}\omega_{21}D^{-1} & \Omega^{-1} + \Omega^{-1}\omega_{21}D^{-1}\omega_{12}\Omega^{-1}
\end{pmatrix}.
\]

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

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

\[
|\Omega^\dagger| = |(I_T \otimes B^{-1})| \cdot |H_{E'}| \cdot |(I_T \otimes B'^{-1})| = |B|^{-2T} |H_{E'}|,
\]

\[
\Omega^\dagger^{-1} = (I_T \otimes B^{-1})^{-1}H_{E'}^{-1}(I_T \otimes B'^{-1})^{-1} = (I_T \otimes B')H_{E'}^{-1}(I_T \otimes B),
\]

where \( |H_{E'}| = |TE - (T-1)I_n| = \prod_{i=1}^{n}|T\phi_\zeta(1 - \lambda w_i)^2 + Tc_m - T + 1| \) as in (3.30), and

\[
H_{E'}^{-1} = (1 - T)(h_0^{-1} \otimes E'^{-1}) + (h_1^{-1} - (1 - T)h_0^{-1}) \otimes (E'^{-1}E),
\]

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

\[
h_s = \begin{pmatrix}
s & -1 & 0 & \cdots & 0 & 0 & 0 \\
-1 & 2 & -1 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & \cdots & -1 & 2 & -1 \\
0 & 0 & 0 & \cdots & 0 & -1 & 2
\end{pmatrix},
\]

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

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

General notation: (i) recall \( \psi = (\theta', \sigma^2_\epsilon, \delta')' \), where \( \theta \) and \( \sigma^2_\epsilon \) are the linear and scale parameters and can be concentrated out from the likelihood function, and \( \delta \) is the vector of nonlinear parameters left in the concentrated likelihood function. Let \( \psi_0 = (\theta_0', \sigma^2_{\epsilon 0}, \delta_0')' \) be the true parameter vector. Let \( \Psi \) be the parameter space of \( \psi \), and \( \Delta \) the space of \( \delta \). (ii) A parametric function, or vector, or matrix, evaluated at \( \psi_0 \), is denoted by adding a subscript \( 0 \), e.g., \( B_0 = B|_{\lambda = \lambda_0} \), and similarly for \( \Omega_0, \Omega^*_0, \Omega^*_0 \), etc. (iii) The common expectation and variance operators ‘\( E \)’ and ‘\( \text{Var} \)’ correspond to \( \psi \).

4.1 Generic assumptions

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

Assumption G1: (i) The available observations are: \( (y_{it}, x_{it}, z_i), i = 1, \cdots, n, \ t = 0, 1, \cdots, T, \) with \( T \geq 2 \) fixed and \( n \to \infty \); (ii) The disturbance vector \( u_t = (u_{1t}, \cdots, u_{nt})' \) exhibits both individual effects and spatially autocorrelated structure defined in (2.2) and (2.3) and \( v_{it} \) are iid for all \( i \) and \( t \) with \( \text{E}(v_{it}) = 0, \text{Var}(v_{it}) = \sigma^2_v, \text{and} \text{E}|v_{it}|^4 < \infty \) for some \( \epsilon_0 > 0 \); (iii) \( \{x_{it}, t = \cdots, -1, 0, 1, \cdots\} \) and \( \{z_i\} \) are strictly exogenous and independent across \( i \); (iv) \( |\rho| < 1 \) in (2.1); and (v) The true parameter \( \psi_0 \) lies in the interior of a convex compact set \( \Psi \).

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

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

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

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

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

### 4.2 Random effects model

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

**Assumption R:** (i) $\mu_i$’s are iid 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 $v_{ij}$ are mutually independent, and they are independent of $x_{ks}$ and $z_k$ for all $i, j, k, t, s$; (iii) All elements in $(x_{it}, z_i)$ have $4 + \epsilon_0$ moments for some $\epsilon_0 > 0$.

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

**Case I:** $y_0$ is exogenous. To derive the consistency of the QML estimators, we need to ensure that $\delta = (\lambda, \phi_0)'$ is identifiable. Then, the identifiability of other parameters follows. Following White (1994) and Lee (2004), define $L^*_{\omega}(\delta) = \max_{\theta, \sigma^2} E[L'(\theta, \sigma^2, \delta)]$, where we suppress the dependence of $L^*_{\omega}(\delta)$ on $n$. The optimal solution to $\max_{\theta, \sigma^2} E[L'(\theta, \sigma^2, \delta)]$ is given by

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

$$
\tilde{\sigma}^2(\delta) = \frac{1}{nT} E[u(\tilde{\theta}(\delta))'(\Omega^{-1}u(\tilde{\theta}(\delta))]. \quad (4.2)
$$
Consequently, we have
\[ \mathcal{L}_r^*(\delta) = -\frac{nT}{2} \log(2\pi) + 1 - \frac{nT}{2} \log(\hat{\sigma}_r^2(\delta)) - \frac{1}{2} \log(|\Omega|). \] (4.3)

We impose the following identification condition.

**Assumption R:** (iv) \[ \lim_{n \to \infty} \frac{1}{nT} \{ \log[\sigma_0^2(\delta)] - \log[\hat{\sigma}_r^2(\delta)] \} \neq 0 \] for any \( \delta \neq \delta_0 \), and \( \frac{1}{nT} \hat{X}'\hat{X} \) is positive definite almost surely for sufficiently large \( n \).

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

**Theorem 4.1** Under Assumptions G1, G2, and R(i)-(iv), if the initial observations \( y_0 \) are exogenously given, then \( \hat{\psi} \to^p \psi_0 \).

To derive the asymptotic distribution of \( \hat{\psi} \), we need to make a Taylor expansion of \( \frac{\partial}{\partial \psi} \mathcal{L}_r^*(\psi) = 0 \) at \( \psi_0 \), and then to check that the score function and Hessian matrix have proper asymptotic behavior. We report both the score and Hessian here to provide insights for the asymptotic theory and to facilitate the practical applications. First, the score function \( S(\psi) = \frac{\partial}{\partial \psi} \mathcal{L}_r^*(\psi) \) has the elements

\[ \frac{\partial \mathcal{L}_r^*(\psi)}{\partial \theta} = \frac{1}{\sigma^2} \hat{X}' \Omega^{-1} u(\theta), \]

\[ \frac{\partial \mathcal{L}_r^*(\psi)}{\partial \sigma^2} = \frac{1}{2\sigma^4} u(\theta)' \Omega^{-1} u(\theta) - \frac{nT}{2\sigma^2}, \]

\[ \frac{\partial \mathcal{L}_r^*(\psi)}{\partial \omega} = \frac{1}{2\sigma^2} u(\theta)' P_{\omega} u(\theta) - \frac{1}{2} \text{tr} (P_{\omega} \Omega), \quad \omega = \lambda, \phi, \mu, \]

where \( P_{\omega} = \Omega^{-1} \Omega_{\omega,\omega} \) and \( \Omega_{\omega,\omega} = \frac{\partial}{\partial \omega} \Omega(\delta) \) for \( \omega = \lambda, \phi, \mu \). One can easily verify that \( \Omega_\lambda = I_T \otimes A \) and \( \Omega_{\phi,\mu} = J_T \otimes I_n \), where \( A = \frac{\partial}{\partial \lambda} (B'B)^{-1} = (B'B)^{-1}(W'B + B'W)(B'B)^{-1} \). At \( \psi = \psi_0 \), the last three components of the score function are linear and quadratic functions of \( u \equiv u(\theta_0) \) and one can readily verify that their expectations are zero. The first component also has a zero expectation by Lemma B.6.

Note that the elements in \( u \) are not independent and that \( \hat{X} \) contains the lagged dependent variable, thus the standard results, such as the central limit theorem (CLT) for linear and quadratic forms in Kelejian and Prucha (2001) cannot be directly applied. For the last three components, we need to plug \( u = (I_T \otimes I_n) \mu + (I_T \otimes B_0)^{-1} v \) into \( \frac{\partial}{\partial \psi} \mathcal{L}_r^*(\psi_0) \) and apply the CLT to linear and quadratic functions of \( \mu \) and \( v \) separately. For the first component, a special care has to be given to \( Y_{-1} \) (see Lemmas B.6 and B.8 for details).

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

\[ \frac{\partial^2 \mathcal{L}_r^*(\psi)}{\partial \theta^2} = \frac{1}{\sigma^2} \hat{X}' \Omega^{-1} \hat{X}, \]

\[ \frac{\partial^2 \mathcal{L}_r^*(\psi)}{\partial \sigma^4} = \frac{1}{2\sigma^4} u(\theta)' P_{\sigma} u(\theta), \quad \omega = \lambda, \phi, \mu, \]

\[ \frac{\partial^2 \mathcal{L}_r^*(\psi)}{\partial \sigma^2 \partial \rho} = \frac{1}{2\sigma^2} u(\theta)' P_{\sigma} u(\theta) - \frac{1}{2} \text{tr} (P_{\sigma} \Omega), \quad \rho = \lambda, \phi, \mu, \]

where \( q_{\omega,\omega}(u) \equiv \frac{1}{2} \text{tr} (P_{\omega} \Omega_{\omega,\omega}) - \frac{1}{2\sigma^2} u'(2P_{\omega} \Omega_{\omega,\omega} - \Omega^{-1} \Omega_{\omega,\omega}) \Omega^{-1} u \) for \( \omega, \rho = \lambda, \phi, \mu \); and \( \Omega_{\omega,\omega} = \frac{\partial^2}{\partial \omega \partial \omega} \Omega(\delta) \) for \( \omega, \rho = \lambda, \phi, \mu \). It is easy to see that \( \Omega_{\lambda,\lambda} = I_T \otimes \hat{A} \) where \( \hat{A} = \frac{\partial}{\partial \lambda} A = 2(B'B)^{-1}([W'B + B'W]A - W'W) \), and all other \( \Omega_{\omega,\omega} \) matrices are \( 0_{nT \times nT} \).
Again, we see that most of the Hessian elements are quadratic forms of $u(\theta)$ whose asymptotic behavior is easy to study. Special care needs to be given to the elements involving $\tilde{X}$ (see Lemma B.7 for details). Let $\Gamma_{r,n}(\psi) = E\left[ \frac{\partial}{\partial \psi} \mathcal{L}'(\psi) \frac{\partial}{\partial \psi} \mathcal{L}''(\psi) \right]$ be the variance-covariance matrix of the score vector.\footnote{It is well known that for normally distributed individual-specific effects $\mu_i$ and error terms $v_{it}$, $\Gamma_{r,n}(\psi_{0}) = -E[H_{r,n}(\psi_{0})]$ under some mild conditions. We do not impose normality restriction in this paper.} See Appendix A for the expression of $\Gamma_{r,n}(\psi)$. We have the following theorem.

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

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

**Case II:** $y_{i0}$ is endogenous. In this case, define $\mathcal{L}_{r}^{\ast\ast}(\delta) = \max_{\theta, \sigma^2} E[\mathcal{L}''(\theta, \sigma^2, \delta)]$, where we suppress the dependence of $\mathcal{L}_{r}^{\ast\ast}(\delta)$ on $n$. The optimal solution to $\max_{\theta, \sigma^2} E[\mathcal{L}''(\theta, \sigma^2, \delta)]$ is now given by

\[
\tilde{\theta}(\delta) = \frac{1}{n(T+1)}E[u^{\ast}(\tilde{\theta}(\delta), \rho)\Omega^{-1}(\delta)u^{\ast}(\tilde{\theta}(\delta), \rho)].
\]

\[
\tilde{\sigma}^{2}(\delta) = \frac{1}{n(T+1)}E[u^{\ast}(\tilde{\theta}(\delta), \rho)\Omega^{-1}(\delta)u^{\ast}(\tilde{\theta}(\delta), \rho)].
\]

Consequently, we have

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

We make the following identification assumption.

**Assumption R:** (iv) $\lim_{n \to \infty} \frac{1}{2n(T+1)} \left[ \log |\sigma^{2}_{\epsilon}\Omega^{*}_{\ast} - \log |\tilde{\sigma}^{2}(\delta)\Omega^{*}(\delta)| \right] \neq 0$ for any $\delta \neq \delta_{0}$. Both $\frac{1}{n} \tilde{X}'\tilde{X}$ and $\frac{1}{n} \tilde{X}'(X, Z)'(X, Z) \tilde{X}$ are positive definite almost surely for sufficiently large $n$.

The following theorem establishes the consistency of QMLE for the random effects model with endogenous initial observations. Similarly, the key result is to show that $\frac{1}{n(T+1)}[\mathcal{L}_{r}^{\ast\ast}(\delta) - \mathcal{L}_{r}^{\ast\ast}(\delta_{0})]$ converges to zero uniformly in $\delta \in \Delta$, which is given in Appendix C.

**Theorem 4.3** Under Assumptions G1, G2, R0, R(i)-(iii) and R(iv'), if the initial observations $y_{i0}$ are endogenously given, then $\tilde{\psi} \xrightarrow{p} \psi_{0}$.

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

\[
\frac{\partial \mathcal{L}''(\psi)}{\partial \theta} = \frac{1}{\sigma^{2}_{\epsilon}} X^* \Omega^{-1} u^*(\theta, \rho),
\]

\[
\frac{\partial \mathcal{L}''(\psi)}{\partial \sigma_{\epsilon}} = \frac{1}{2\sigma^{2}_{\epsilon}} u^*(\theta, \rho)' \Omega^{-1} u^*(\theta, \rho) - \frac{n(T+1)}{2\sigma^{2}_{\epsilon}},
\]

\[
\frac{\partial \mathcal{L}''(\psi)}{\partial \mu_{i}} = -\frac{1}{\sigma^{2}_{\epsilon}} u^*(\theta, \rho)' \Omega^{-1} u^*(\theta, \rho) + \frac{1}{\sigma^{2}_{\epsilon}} \omega^*(\theta, \rho)' P_{\omega} u^*(\theta, \rho) - \frac{1}{2} \text{tr}(P_{\omega}^{*} \Omega^{*}),
\]

\[
\frac{\partial \mathcal{L}''(\psi)}{\partial \phi} = -\frac{1}{\sigma^{2}_{\epsilon}} u^*(\theta, \rho)' P_{\omega} u^*(\theta, \rho) - \frac{1}{2} \text{tr}(P_{\omega}^{*} \Omega^{*}) \text{ for } \omega = \lambda, \phi_{\mu}, \text{ and } \phi_{\epsilon}.
\]
where $u^*_p(\theta, \rho) = \frac{\partial}{\partial \rho} u^*(\theta, \rho)$, $P^*_\rho = \Omega^*-1 \Omega^*^{-1}$, $\Omega^*_\rho = \frac{\partial}{\partial \rho} \Omega^* (\delta)$ for $\omega = \rho, \lambda, \phi_\mu$, and $\phi_\zeta$ have the expressions

$$u^*_p(\theta, \rho) = -\left( \dot{a}_m Z \gamma \right) \left( Y_i - 1 \right), \quad \Omega^*_\rho = \begin{pmatrix} 2 \phi_\mu a_m \dot{a}_m I_n + \dot{b}_m (B'B)^{-1} \\ \phi_\mu \dot{a}_m (t' \otimes I_n) \end{pmatrix}, \quad \Omega^*_\rho = \begin{pmatrix} 0_n \times n_T \end{pmatrix}, \quad \Omega^*_\phi = \begin{pmatrix} m \end{pmatrix}, \quad \Omega^*_\phi = \begin{pmatrix} 1 \end{pmatrix},$$

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

$$\frac{\partial^2 \mathcal{L}^T(\psi)}{\partial \psi \partial \psi^T} = -\frac{1}{\sigma^2} \left( X' \Omega^*^{-1} X \right),$$

where $q^*_\omega(u^*) \equiv \frac{1}{2} \left( P^*_\rho \Omega^*_\rho - \Omega^*-1 \Omega^*_\rho \right)$ for $\omega = \rho, \lambda, \phi_\mu$, and $\phi_\zeta$. $X^*_\rho = \frac{\partial}{\partial \rho} X^*$, $u^*_p(\theta, \rho) = \frac{\partial^2}{\partial \rho \partial \psi} u^* (\theta, \rho)$, and $\Omega^*_\rho = \frac{\partial^2}{\partial \rho \partial \psi} \Omega^*$ for $\omega = \rho, \lambda, \phi_\mu$, and $\phi_\zeta$. The second-order partial derivatives of $u^*$ are

$$\Omega^*_p = \begin{pmatrix} 2 \phi_\mu (\dot{a}_m + \ddot{a}_m) I_n + \ddot{b}_m (B'B)^{-1} \\ \phi_\mu \dot{a}_m (t' \otimes I_n) \end{pmatrix}, \quad \Omega^*_\rho = \begin{pmatrix} 0_n \times n_T \end{pmatrix}, \quad \Omega^*_\phi = \begin{pmatrix} m \end{pmatrix}, \quad \Omega^*_\phi = \begin{pmatrix} 1 \end{pmatrix},$$

and all other $\Omega^*_\omega$ matrices are $0_n(T+1) \times n(T+1)$, where $\ddot{a}_m = \frac{\partial}{\partial \rho} \ddot{a}_m$ and $\ddot{b}_m = \frac{\partial}{\partial \rho} \ddot{b}_m$ and their exact expressions can be easily derived. Finally, $X^*_\rho$ has a sole non-zero element $\dot{a}_m$, and $u^*_p(\theta, \rho) = (-\dot{a}_m \gamma' z, 0_{1 \times nT} \gamma')$. Let $\Gamma_{rr,n}(\psi) = E[U^*_p | \mathbf{y}_n] = \frac{\partial^2}{\partial \psi \partial \psi^T} \mathcal{L}^T(\psi)$ be the variance-covariance matrix of the score vector with its detail given in Appendix A. We now state the asymptotic normality result.
Theorem 4.4 Under Assumptions G1, G2, R0, R(i)-(iii) and R(iv), if the initial observations are endogenously given, then \( \sqrt{n} (\hat{\psi} - \psi_0) \overset{d}{\rightarrow} \mathcal{N}(0, H_{rr}^{-1} \Gamma_{rr} H_{rr}^{-1}) \), where \( H_{rr} = \lim_{n \rightarrow \infty} \frac{1}{n(t+1)} E[H_{rr,n} (\psi_0)] \) and \( \Gamma_{rr} = \lim_{n \rightarrow \infty} \frac{1}{n(t+1)} \Gamma_{rr,n} (\psi_0) \), both assumed to exist, and \(-H_{rr}\) is assumed to be positive definite.

4.3 Fixed effects model

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

Assumption F: (i) The processes \( \{x_{it}, t = \cdots, -1, 0, 1, \cdots\} \) are trend-stationary or first-differencing stationary for all \( i = 1, \cdots, n \); (ii) All elements in \( (\Delta v_{1t}, \Delta x_{it}) \) have \( 4 + \epsilon_0 > 0 \); (iii) \( \frac{1}{\sqrt{n}} \Delta X' \Delta X \) is positive definite almost surely for sufficiently large \( n \).

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

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

\[
\hat{\delta}_*^2 (\delta) = \frac{1}{nT} E[\Delta u (\hat{\theta} (\delta), \rho) \Omega^{-1} \Delta u (\hat{\theta} (\delta), \rho)]. \quad (4.8)
\]

Consequently, we have

\[
\mathcal{L}_*^f (\delta) = -\frac{nT}{2} [\log (2\pi) + 1] - \frac{nT}{2} \log [\hat{\sigma}_*^2 (\delta)] - \frac{1}{2} \log |\Omega^\dagger|. \quad (4.9)
\]

The following identification condition is needed for our consistency result.

Assumption F: (iv) \( \lim_{n \rightarrow \infty} \frac{1}{nT} \left\{ \log |\sigma_{**}^2 (\Omega^\dagger) | - \log |\sigma_*^2 (\delta) \Omega^\dagger (\delta) | \right\} \neq 0 \) for any \( \delta \neq \delta_0 \).

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

Theorem 4.5 Under Assumptions G1, G2, F0, and F, we have \( \hat{\psi} \overset{P}{\rightarrow} \psi_0 \).

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

\[
\begin{align*}
\frac{\partial S^f (\psi)}{\partial \psi} &= \frac{1}{\sigma^2} \Delta X' \Omega^{-1} \Delta u (\theta, \rho), \\
\frac{\partial S^f (\psi)}{\partial \sigma^2} &= \frac{1}{2 \sigma^2} \Delta u (\theta, \rho)' \Omega^{-1} \Delta u (\theta, \rho) - \frac{nT}{2 \sigma^2}, \\
\frac{\partial S^f (\psi)}{\partial \theta} &= \frac{1}{2 \sigma^2} \Delta u (\theta, \rho)' \Omega^{-1} \Delta u (\theta, \rho) + \frac{1}{2 \sigma^2} \Delta u (\theta, \rho)' P^\dagger_{\rho} \Delta u (\theta, \rho) - \frac{1}{2} \text{tr} (\Omega^{-1} \Omega^\dagger_{**}), \\
\frac{\partial S^f (\psi)}{\partial \rho} &= \frac{1}{2 \sigma^2} \Delta u (\theta, \rho)' P^\dagger_{\rho} \Delta u (\theta, \rho) - \frac{1}{2} \text{tr} (\Omega^{-1} \Omega^\dagger_{**}) \text{ for } \omega = \lambda, \phi_c,
\end{align*}
\]

where \( \Delta u (\theta, \rho) = \frac{\partial}{\partial \rho} \Delta u (\theta, \rho) = - (0_{n \times 1}, \Delta y_{1}', \cdots, \Delta y_{T-1}')' \) and \( \Omega^\dagger_{**} = \frac{\partial}{\partial \rho} \Omega^\dagger (\delta) \) and \( P^\dagger_{\rho} = \Omega^{-1} \Omega^\dagger_{**} \Omega^{-1} \) for \( \omega = \rho, \lambda, \text{and } \phi_c \). From (3.24), it is easy to see that \( \Omega^\dagger_\rho = h_{c_m} \otimes (B' B)^{-1}, \Omega^\dagger_{\phi_c} = h_{c_m} \otimes A, \text{ and } \Omega^\dagger_\phi_\phi = \lambda_1 \otimes I_4 \), where \( c_m = \frac{\partial}{\partial \rho} c_m (\rho) \). Again, one can readily verify that \( E[\frac{\partial}{\partial \psi} \mathcal{L}_*^f (\psi_0)] = 0 \). The asymptotic normality of the score is given in Lemma B.15. The asymptotic normality of \( \hat{\psi} \) thus follows.
if the Hessian matrix, \( H_{f,n}(\psi) = \frac{\partial^2 L_f(\psi)}{\partial \psi \partial \psi'} \), given below possesses the desired stochastic convergence property.

\[
\frac{\partial^2 L_f(\psi)}{\partial \theta \partial \rho} = -\frac{1}{\sigma^2} \Delta X^\top \Omega^{-1} \Delta X,
\]

\[
\frac{\partial^2 L_f(\psi)}{\partial \theta \partial \sigma} = -\frac{1}{\sigma^2} \Delta X^\top \Omega^{-1} \Delta u(\theta, \rho),
\]

\[
\frac{\partial^2 L_f(\psi)}{\partial \theta \partial \phi} = \frac{1}{\sigma^2} \Delta X^\top \Omega^{-1} \Delta u(\theta, \rho) - \frac{1}{\sigma^2} \Delta X^\top P^\top \rho \Delta u(\theta, \rho),
\]

\[
\frac{\partial^2 L_f(\psi)}{\partial \rho \partial \omega} = -\frac{1}{\sigma^2} \Delta X^\top \Omega^{-1} \Delta u(\theta, \rho), \text{ for } \omega = \lambda, \phi, \zeta.
\]

\[
\frac{\partial^2 L_f(\psi)}{\partial \rho \partial \omega,} = -\frac{1}{\sigma^2} \Delta u(\theta, \rho) \Omega^{-1} \Delta u(\theta, \rho) + \frac{\omega}{\sigma^2} \Delta u(\theta, \rho) \rho \Delta u(\theta, \rho), \text{ for } \omega = \lambda, \phi, \zeta,
\]

\[
\frac{\partial^2 L_f(\psi)}{\partial \theta \partial \phi} = \frac{1}{\sigma^2} \Delta u(\theta, \rho) \Omega^{-1} \Delta u(\theta, \rho) + \frac{\omega}{\sigma^2} \Delta u(\theta, \rho) \rho \Delta u(\theta, \rho), \text{ for } \omega = \lambda, \phi, \zeta.
\]

\[
\frac{\partial^2 L_f(\psi)}{\partial \theta \partial \omega} = \frac{1}{\sigma^2} \Delta u(\theta, \rho) \Omega^{-1} \Delta u(\theta, \rho), \text{ for } \omega = \lambda, \phi, \zeta.
\]

where \( \Delta(\omega) \equiv \frac{1}{\sigma^2} \text{tr}(P_{\omega}^\top \Omega^{-1} \Delta u - \Omega^{-1} \Omega_{\omega} - \Omega^{-1} \Omega_{\omega} \Omega^{-1} \Delta u) \) for \( \omega = \rho, \lambda, \phi, \zeta \). The second derivatives \( \Omega_{\omega} \) of \( \Omega \) are: \( \Omega_{\omega,} = h_{\omega} \otimes (B' B)^{-1} \) where \( h_{\omega} = \frac{\partial}{\partial \omega} \hat{c}_m, \Omega_{\omega,} = h_{\omega} \otimes A, \Omega = h_{\omega} \otimes \hat{A}, \) and the remaining are all zero matrices.

Let \( \Gamma_{f,n}(\psi) = E[\frac{\partial^2 L_f(\psi)}{\partial \psi \partial \psi'}] \). (See Appendix A for some details.) We now state the asymptotic normality result.

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

## 5 Bootstrap Estimate of the Variance-Covariance Matrix

From Theorems 4.2, 4.4 and 4.6, we see that the asymptotic variance-covariance (VC) matrices of the QMLEs of the three models considered are, respectively, \( H_{r}^{-1} \Gamma_r H_{r}^{-1}, H_{rr}^{-1} \Gamma_{rr} H_{rr}^{-1}, \) and \( H_{f}^{-1} \Gamma_f H_{f}^{-1} \).

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

---

8This is not a problem for the exact likelihood inference (Elhorst, 2005, Yang et al., 2006) as in this case the VC matrix of the score function equals the negative expected Hessian. Hence, the asymptotic VC matrices of the MLEs in the three models considered reduce to \( -H_{r}^{-1}, -H_{rr}^{-1} \) and \( -H_{f}^{-1} \), respectively, of which sample analogues exist.
In this section, we introduce a residual-based bootstrap method for estimating the variance of the scores, with the bootstrap draws made on the joint empirical distribution function (EDF) of the transformed vectors of residuals. While the general principle for our bootstrap method is the same for all the three models considered above, different structures of the residuals and the score functions render them a separate consideration.

5.1 Random effects model with exogenous initial values

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

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

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

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

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

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

$$y_{1b}^b = \hat{\rho}_y y_0 + x_1 \hat{\beta} + z_1 \hat{\gamma} + \Sigma^{\frac{1}{2}}(\hat{\lambda}, \hat{\phi}, \mu) \hat{r}_{1b},$$

$$y_{tb}^b = \hat{\rho}_y y_{t-1} + x_t \hat{\beta} + z_t \hat{\gamma} + \Sigma^{\frac{1}{2}}(\hat{\lambda}, \hat{\phi}, \mu) \hat{r}_{tb}, \quad t = 2, 3, \cdots, T.$$

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

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

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

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

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

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

which can be made to an arbitrary level of accuracy by choosing an arbitrarily large \( M \). Note that \( u_t = \sigma \Sigma \hat{\psi}(\lambda_0, \phi_{0}) \psi_t \). The step (i) above is equivalent to draw random sample \( r^m_t \) from the distribution \( F \) of \( r_t \), the \( i \) element of \( r_t \), and compute \( u^m_t = \sigma \Sigma \hat{\psi}(\lambda_0, \phi_{0}) \psi_t \).

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

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

In the real world, \( F \), or the distributions of \( \mu_t \) and \( v_t \) are also unknown. However, we note that the only difference between \( \hat{\Gamma}_{n,r}^b \) given in (5.1) and \( \hat{\Gamma}_{n,r}^m \) given in (5.3) is that \( r^m_t \) for the former is from the EDF \( \hat{F}_{n,t} \), but \( r^b_t \) for the latter is drawn from the true distribution \( F \). The bootstrap DGP that mimics the real-world DGP must be \( y^b_t = \rho y_{t-1} + x_t \hat{\beta} + z_t \hat{\gamma} + u^b_t \), and \( y^b_t = \rho y_{t-1} + x_t \hat{\beta} + z_t \hat{\gamma} + u^b_t, t = 2, \ldots, T \). Thus, if \( \hat{F}_{n,t} \) provides a consistent estimate for the true but unknown distribution \( F \), which is typically the case as \( \hat{\psi} \) is consistent for \( \psi_0 \), then \( \hat{\Gamma}_{n,r}^b \) and \( \hat{\Gamma}_{n,r}^m \) are asymptotically equivalent. The extra variability caused by replacing \( F \) by \( \hat{F}_{n,t} \) is of the same order as that from replacing \( \psi_0 \) by \( \hat{\psi} \). This justifies the validity of the proposed bootstrap procedure.

### 5.2 Random effects model with endogenous initial values

When the initial observations \( y_0 \) are endogenously given, the disturbance vector now becomes \( (u_0, u_1, u_2, \ldots, u_T) \) such that \( \text{Var}(u_0) = \sigma^2 \omega_{11} \) and \( \text{Var}(u_t) = \sigma^2 \Sigma(\lambda, \phi_t), t = 1, 2, \ldots, T \), where \( \omega_{11} \) is defined above (3.31) and \( \Sigma(\lambda, \phi_t) \) is defined in Section 5.1. Define the transformed residuals: \( r_0 = \omega_{11}^{-1/2} u_0 \), and \( r_t = \Sigma^{-1/2}(\lambda, \phi_t) u_t, t = 1, \ldots, T \), where \( \omega_{11}^{-1} \) is a square-root matrix of \( \omega_{11} \). Now, denote the QML estimates of the transformed residuals as \( \{r^b_0, r^b_1, \ldots, r^b_T\} \), and the EDF of the centered \( \hat{r}_t \) by \( \hat{F}_{n,t} \), \( t = 0, 1, \ldots, T \). Draw a random sample of size \( n \) each from \( F_{n,t} \), to give bootstrap residuals \( \{\hat{r}^b_0, \hat{r}^b_1, \ldots, \hat{r}^b_T\} \). The bootstrap values for the response variables are thus generated according to

\[
\hat{y}^b_0 = \hat{\beta} \hat{x}_0^b + \omega^b \hat{x}_0^b, \quad \text{and} \quad y^b_t = \rho y_{t-1}^b + x_t \hat{\beta} + z_t \hat{\gamma} + \Sigma^b(\hat{\lambda}, \hat{\phi}_t) \hat{r}^b_t, t = 1, 2, \ldots, T.
\]

The rest is analogous to those described in Section 5.1, including the justifications for the validity of this bootstrap procedure.
5.3 Fixed effects model with endogenous initial values

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

\[
\Delta y_t^b = \Delta \tilde{x}_t^b + \omega^{\frac{1}{2}} r_0^b, \quad \text{and} \quad y_t^b = \rho \Delta y_{t-1}^b + \Delta x_t^b \hat{\beta} + \sqrt{2} \hat{B}^{-1} \hat{r}_t^b, t = 2, 3, \ldots, T.
\]

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

6 Finite Sample Properties of the QMLEs

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

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

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

\(^{9}\)The detail is: \(x_t = \mu_x + g t_n + \zeta_t, (1 - \phi_1 L) \zeta_t = \varepsilon_t + \phi_2 \varepsilon_{t-1}, \varepsilon_t \sim N(0, \sigma_\varepsilon^2 I_n), \mu_x = e + \frac{1}{n-1} \sum_{t=2}^T \varepsilon_{t-1}\) and \(e \sim N(0, \sigma_e^2)\). Let \(\theta = (g, \phi_1, \phi_2, \sigma_1, \sigma_2)\). Alternatively, the elements of \(x_t\) can be randomly generated from \(N(0, 4)\).
**Random effects model.** Table 1 reports the Monte Carlo mean and rmse for the random effects model when the data are generated according to either \( m = 0 \) or \( m = 6 \), but the model is estimated under \( m = 0, 6, \) and 200. The results show clearly that a correct treatment on the initial values leads to excellent estimation results in general, but a wrong treatment may give totally misleading results.

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

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

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

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

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

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

7 Conclusion

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

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

\[
\begin{align*}
\Gamma_{r,\theta} &= \frac{1}{\sigma_\theta^2} \sigma_\theta \Rightarrow E(\dot{X}^T \Omega_0^{-1} \dot{X}), \\
\Gamma_{r,\sigma_\theta^2} &= \frac{1}{\sigma_\theta^2} g(\Omega_0^{-1}), \\
\Gamma_{r,\varpi} &= \frac{1}{\sigma_{\varpi}^2} \sigma_{\varpi} \Rightarrow E(\dot{X}^T \Omega_0^{-1} \dot{u} \dot{u}^T P_{\varpi} u), \\
\Gamma_{r,\varpi^2} &= \frac{1}{\sigma_{\varpi}^2} g(\Omega_0^{-1}, P_{\varpi}), \\
\Gamma_{r,\omega} &= \frac{1}{\sigma_\omega^2} \sigma_\omega \Rightarrow E(\dot{X}^T \Omega_0^{-1} \dot{u} \dot{u}^T P_{\omega} u), \\
\Gamma_{r,\omega^2} &= \frac{1}{\sigma_\omega^2} g(\Omega_0^{-1}, P_{\omega}),
\end{align*}
\]

where \( g(A, B) \equiv \frac{1}{4 \sigma_\omega^2} E(\dot{u}^T \Omega_0^{-1} \dot{u}^T P_{\omega} \dot{u}) - \frac{1}{4} \text{tr}(A \Omega_0) \text{tr}(B \Omega_0) \), and \( P_{\omega} \) is defined below Theorem 4.1. The explicit form of \( g \) can be obtained from Lemma B.4(1). The other elements do not possess explicit forms due to the complications caused by \( Y_{-1} \).

The elements of the information matrix for the random effects model with endogenous \( y_0 \), \( \Gamma_{r,n}(\psi) \equiv E(\begin{bmatrix} \frac{\partial}{\partial \psi} L'(\psi) \frac{\partial}{\partial \psi} L'(\psi) \end{bmatrix}) \), are, for \( \omega, \varpi = \lambda, \phi_\mu \), or \( \phi_\zeta \):

\[
\begin{align*}
\Gamma_{r,\theta} &= \frac{1}{\sigma_\theta^2} \sigma_\theta \Rightarrow E(\dot{X}^T \Omega_0^{-1} \dot{X}), \\
\Gamma_{r,\sigma_\theta^2} &= \frac{1}{\sigma_\theta^2} f_1^\dagger(\Omega_0^{-1}), \\
\Gamma_{r,\varpi} &= \frac{1}{\sigma_{\varpi}^2} \sigma_{\varpi} \Rightarrow E(\dot{X}^T \Omega_0^{-1} \dot{u} \dot{u}^T P_{\varpi} u), \\
\Gamma_{r,\varpi^2} &= \frac{1}{\sigma_{\varpi}^2} g(\Omega_0^{-1}, P_{\varpi}), \\
\Gamma_{r,\omega} &= \frac{1}{\sigma_\omega^2} \sigma_\omega \Rightarrow E(\dot{X}^T \Omega_0^{-1} \dot{u} \dot{u}^T P_{\omega} u), \\
\Gamma_{r,\omega^2} &= \frac{1}{\sigma_\omega^2} g(\Omega_0^{-1}, P_{\omega}),
\end{align*}
\]

The elements of the information matrix for the fixed effects model with endogenous \( y_0 \), \( \Gamma_{f,n}(\psi) \equiv E(\begin{bmatrix} \frac{\partial}{\partial \psi} L'(\psi) \frac{\partial}{\partial \psi} L'(\psi) \end{bmatrix}) \), are, for \( \omega, \varpi = \lambda, \phi_\zeta \):

\[
\begin{align*}
\Gamma_{f,\theta} &= \frac{1}{\sigma_\theta^2} \sigma_\theta \Rightarrow E(\dot{X}^T \Omega_0^{-1} \dot{X}), \\
\Gamma_{f,\sigma_\theta^2} &= \frac{1}{\sigma_\theta^2} f_1^\dagger(\Omega_0^{-1}), \\
\Gamma_{f,\varpi} &= \frac{1}{\sigma_{\varpi}^2} \sigma_{\varpi} \Rightarrow E(\dot{X}^T \Omega_0^{-1} \dot{u} \dot{u}^T P_{\varpi} u), \\
\Gamma_{f,\varpi^2} &= \frac{1}{\sigma_{\varpi}^2} g(\Omega_0^{-1}, P_{\varpi}), \\
\Gamma_{f,\omega} &= \frac{1}{\sigma_\omega^2} \sigma_\omega \Rightarrow E(\dot{X}^T \Omega_0^{-1} \dot{u} \dot{u}^T P_{\omega} u), \\
\Gamma_{f,\omega^2} &= \frac{1}{\sigma_\omega^2} g(\Omega_0^{-1}, P_{\omega}),
\end{align*}
\]

Appendix B: Some Useful Lemmas

We introduce some fundamental lemmas (existing and new) that are used in the proofs of the main results. For any random variable \( a \) with a zero mean and a finite fourth moment, let \( \kappa_a \equiv E(a^4) - 25 \).
Lemma B.1 Let $P_n$ and $Q_n$ be two $n \times n$ matrices that are uniformly bounded in both row and column sums. Let $R_n$ be a conformable matrix whose elements are uniformly $O(o_n)$ for a certain sequence $o_n$. Then we have: (1) $P_nQ_n$ is also uniformly bounded in both row and column sums; (2) any $(i, j)$ elements $P_{n,ij}$ of $P_n$ are uniformly bounded in $i$ and $j$ and $tr(P_n) = O(n)$; (3) the elements of $P_nR_n$ and $R_nP_n$ are uniformly $O(o_n)$.

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

Lemma B.2 (1) $B'B, (B'B)^{-1}, \Omega, \Omega^{-1}, \Omega^*, \Omega^{*-1}, \Omega^T, \Omega^{*-1}, A$, and $\hat{A}$ are all uniformly bounded in both row and column sums.

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

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

Lemma B.3 Let $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$ be two independent iid sequences with zero means and fourth moments. Let $\sigma_a^4 = E(a_i^4), \sigma_b^4 = E(b_i^4)$. Let $q_n$ and $p_n$ be $n \times n$ nonstochastic matrices. Then

(1) $E[(a'q_n a)(a'p_n a)] = \kappa_n \sum_{i=1}^n q_{n,ii}p_{n,ii} + \sigma_a^2 tr(q_n)tr(p_n) + tr(q_n(p_n + p_n'))$.

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

(3) $E[(a'q_n b)(a'p_n b)] = \sigma_a^2 \sigma_b^2 tr(q_n p_n')$.

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

Proof. To show (1), write $E[(a'q_n a)(a'p_n a)] = E(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ik}q_{n,ij}a_{lk}p_{n,kl})$. Noting that $E(a_{ik}a_{lk}a_{ik}a_{lk})$ 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

$$E[(a'q_n a)(a'p_n a)] = E(a') \sum_{i=1}^n q_{n,ii}p_{n,ii} + \sigma_a^2 \sum_{i=1}^n \sum_{j \neq i}^n (q_{n,ij}p_{n,ij} + q_{n,ji}p_{n,ji})$$

$$= \kappa_n \sum_{i=1}^n q_{n,ii}p_{n,ii} + \sigma_a^2 \sum_{i=1}^n q_{n,ij}p_{n,ji} + \sigma_a^2 \sum_{i=1}^n q_{n,ji}p_{n,ij} + \sigma_a^2 \sum_{i=1}^n \sum_{j \neq i}^n q_{n,ij}p_{n,ji}.$$ 

The result (2) follows from the independence between $a'q_n a$ and $b'p_n b$. For (3), $E[(a'q_n b)(a'p_n b)] = E(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij}b_{kl}q_{n,ij}a_{kl}p_{n,kl}) = E(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij}^2 b_{kl}^2 q_{n,ij}p_{n,ij}) = \sigma_a^2 \sigma_b^2 tr(q_n p_n').$
Suppose that $G$ of other claims is similar. First, let
\[ \lim_{n \to \infty} = G_{q_n,1} + G_{p_n,1} + G_{p_n,2} + G_{q_n,2}. \]
Similarly, we can prove the other claims.

Next, write $E[(u', s_n u)(a't_n u)]$ and Lemma B.3, we have
\[ E[(u', s_n u)(a't_n u)] = \kappa \sum_{i=1}^{n} G_{q_n,1} + \kappa \sum_{i=1}^{n} G_{p_n,1} + \kappa \sum_{i=1}^{n} G_{p_n,2} + G_{q_n,2}. \]

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

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

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

Similarly, we can prove the other claims.

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

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

We now present lemmas needed in the proofs of the main theorems. For ease of exposition, we assume that both $x_{it}$ and $z_{ij}$ are scalar random variables ($p = 1, q = 1$) in this Appendix. For the proofs of
Theorems 2 and 4 for the SDPD model with random effects, the following presentations are essential. By
continuous back substitutions, we have for \( t = 0, 1, 2, \cdots \),

\[
y_t = X_t \beta_0 + c_{p_0,t} z \gamma_0 + c_{p_0,t} \mu + V_t + Y_{0,t}, \tag{B.1}
\]
where for fixed \( y_0 \),

\[
X_t = \sum_{j=0}^{t-1} \rho_j x_{t-j}, \quad V_t = \sum_{j=0}^{t-1} \rho_j B_0^{-1} v_{t-j}, \quad Y_{0,t} = \rho_0 y_0 \text{ and } c_{p_0,t} = (1 - \rho^t)/(1 - \rho);
\]
and for endogenous \( y_0 \),

\[
X_t = \sum_{j=0}^{t+m-1} \rho_j x_{t-j}, \quad V_t = \sum_{j=0}^{t+m-1} \rho_j B_0^{-1} v_{t-j}, \quad Y_{0,t} = \rho_0^m y_{-m}, \quad \text{and } c_{p_0,t} = (1 - \rho^{t+m})/(1 - \rho).
\]
Now, define \( Y_0 = (Y_{0,0}', Y_{0,1}', \cdots, Y_{0,T-1}') \). Then

\[
Y_{-1} = X_{(-1)} \beta_0 + (l_{p_0} \otimes I_n) z \gamma_0 + (l_{p_0} \otimes I_n) \mu + V_{(-1)} + Y_0, \tag{B.2}
\]
where \( X_{(-1)} = (0, X_1', \cdots, X_{T-1}')', \quad V_{(-1)} = (0, V_1', \cdots, V_{T-1}')', \quad \text{and } l_\rho = (0, c_{p_1,1}, \cdots, c_{p,T-1})' \) when \( y_0 \) is fixed, and \( X_{(-1)} = (X_0, X_1', \cdots, X_{T-1}')', \quad V_{(-1)} = (V_0', V_1', \cdots, V_{T-1}')', \quad \text{and } l_\rho = (c_{p,T}, \cdots, c_{p,T-1})' \) when \( y_0 \) is endogenous. Notice that when \( y_0 \) is exogenous, \( Y_{-1} \) can also be expressed as

\[
Y_{-1} = A_x X' \beta_0 + (l_{p_0} \otimes I_n) z \gamma_0 + (l_{p_0} \otimes I_n) \mu + A_v v + Y_0, \tag{B.3}
\]
where \( A_x = J'_{p_0} \otimes I_n \) and \( A_v = J'_{p_0} \otimes B_0^{-1} \) with

\[
J_\rho = \begin{pmatrix}
0 & 1 & \rho & \cdots & \rho^{T-2} \\
0 & 0 & 1 & \cdots & \rho^{T-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
0 & 0 & 0 & \cdots & 0
\end{pmatrix}. \tag{B.4}
\]

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

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

**Proof.** Note that \( \tilde{X} = (X, Z, Y_{-1}) \). By the strict exogeneity of \( X \) and \( Z \), we can readily show that both \( X' \Omega_0^{-1} u \) and \( Z' \Omega_0^{-1} u \) have expectations zero. We are left to show \( E(Y_{-1}' \Omega_0^{-1} u) = 0 \). By (B.3),

\[
E(Y_{-1}' \Omega_0^{-1} u) = E[\mu'(l_{p_0} \otimes I_n) \Omega_0^{-1} u] + E[v' A'_0 \Omega_0^{-1} u].
\]
Using \( u = (v T \otimes I_n) \mu + (I_T \otimes B_0^{-1}) \) and (3.29), we have

\[
E[\mu'(l_{p_0} \otimes I_n) \Omega_0^{-1} u] = E[\mu'(l_{p_0} \otimes I_n) \Omega_0^{-1} (v T \otimes I_n) \mu] = \phi_{p_0} \sigma_{v0}^2 \text{tr}\{[\Omega_0^{-1} (v T l_{p_0}) \otimes I_n]\} = \phi_{p_0} \sigma_{v0}^2 \text{tr}\{(J_T J_{p_0}) \otimes [(B_0' B_0)^{-1} + \phi_{p_0} T I_n]^{-1}\},
\]
and

\[
E[v' A'_0 \Omega_0^{-1} u] = E[v' A'_0 \Omega_0^{-1} (I_T \otimes B_0^{-1}) u] = \sigma_{v0}^2 \text{tr}\{[\Omega_0^{-1} (I_T \otimes B_0^{-1})(J_{p_0} \otimes (B_0' B_0)^{-1})]\} = \sigma_{v0}^2 \text{tr}\{(T^{-1} J_T J_{p_0}) \otimes [(B_0' B_0)^{-1} + \phi_{p_0} T I_n]^{-1}(B_0' B_0)^{-1}\} + \sigma_{v0}^2 \text{tr}\{(J_{p_0} - T^{-1} J_T J_{p_0}) \otimes I_n\},
\]
where we have used the fact that \( E(v v' A'_0) = J_{p_0} \otimes B_0^{-1} \). It follows that \( E(Y_{-1}' \Omega_0^{-1} u) = \sigma_{v0}^2 \text{tr}(J_{p_0} \otimes I_n) = \sigma_{v0}^2 \text{tr}(J_{p_0}) \text{tr}(I_n) = 0. \)
Lemma B.7 Under the assumptions of Theorem 4.2, \( \frac{1}{n} \left\{ \frac{\partial^2 \mathcal{L}(\phi_0)}{\partial \psi_0^2} - E \left[ \frac{\partial^2 \mathcal{L}(\phi_0)}{\partial \psi_0^2} \right] \right\} = o_p(1). \)

Proof. By the expressions of the Hessian matrix \( \frac{\partial^2 \mathcal{L}(\phi_0)}{\partial \psi_0^2} \) in Section 4.2, it suffices to prove (i) \( n^{-1}[\bar{X}'\bar{\Omega}_0^{-1}\bar{X} - E(\bar{X}'\bar{\Omega}_0^{-1}\bar{X})] = o_p(1) \); (ii) \( n^{-1}[\bar{X}'R_u - E(\bar{X}'R_u)] = o_p(1) \) for \( R = \bar{\Omega}_0^{-1} \) and \( P_{cu} \) with \( \omega = \lambda \) and \( \phi_\mu \); (iii) \( n^{-1}[u'R_u - \sigma^2_{e_0}\text{tr}(R_{00})] = o_p(1) \) for \( R = \bar{\Omega}_0^{-1} \) and \( P_{cu} \) with \( \omega = \lambda \) and \( \phi_\mu \); and (iv) \( n^{-1}[q_{cu}(u) - E(q_{cu}(u))] = o_p(1) \) for \( \omega, \bar{\omega} = \lambda \) and \( \phi_\mu \).

Let \( \Omega_{\omega\omega_0} = \frac{\partial^2}{\partial \omega^2} \Omega(\delta_0) \) for \( \omega, \bar{\omega} = \lambda \) and \( \phi_\mu \). Noting that \( \Omega_{\omega_0}^{-1}, \Omega_{\omega_00}, P_{cu}, \Omega_{\omega_0\omega_0} \) are uniformly bounded in both row and column sums by Lemmas B.1-B.2 and \( q_{\omega\omega}(u) \) is quadratic in \( u \), we can readily show that (iii)-(iv) hold by straightforward moment calculations, Chebyshev inequality, and Lemma B.4. For example, to show (iii), first note that \( E(u'R_u) = \sigma^2_{e_0}\text{tr}(R_{00}) \). By Lemma B.4,

\[
\text{Var}(n^{-1}u'R_u) = n^{-2}\{E(u'R_u\bar{u}'R_u) - [E(u'R_u)]^2\} = n^{-2}\kappa_n \sum_{i=1}^n G_{R_{11}}^2 + n^{-2}\kappa_n \sum_{i=1}^n G_{R_{22}}^2 + 2n^{-2}\sigma^2_{e_0}\text{tr}(R_{00}R_{00}) = O(n^{-1}),
\]

where the last equality follows from the fact that \( G_{R_{11}}^2, G_{R_{22}}^2, \) and \( R_{00}R_{00} \) are all uniformly bounded in both row and column sums. Then (iii) follows by Chebyshev inequality.

To prove (i), let \( R = \bar{\Omega}_0^{-1} \). Noting that \( \bar{X} = (X, Z, Y_1) \), it is easy to show that the terms not involving \( Y_{-1} \), such as \( n^{-1}X'R_X, n^{-1}X'R_Z \), and \( n^{-1}Z'R_Z \) converge in probability to their expectations. For the terms involving \( Y_{-1} \), we first have by (B.3),

\[
n^{-1}Y_{-1}'R_{Y_{-1}} = n^{-1}[A_2X'b_0 + (l_{p_0} \otimes I_n)z_{\gamma_0}]R[\bar{A}_2X'b_0 + (l_{p_0} \otimes I_n)z_{\gamma_0}]
+n^{-1}[l_{p_0} \otimes I_n]v + A_v v'R[(l_{p_0} \otimes I_n)\mu + A_v v]
+n^{-1}Y_{0}'R_{Y_{0}} + 2n^{-1}[A_2X'b_0 + (l_{p_0} \otimes I_n)z_{\gamma_0}]R[(l_{p_0} \otimes I_n)\mu + A_v v]
+2n^{-1}[A_2X'b_0 + (l_{p_0} \otimes I_n)z_{\gamma_0}]R_{Y_{0}} + 2n^{-1}[(l_{p_0} \otimes I_n)\mu + A_v v]'R_{Y_{0}}
= \sum_{i=1}^6 A_{ni}, \quad \text{say}.
\]

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

\[
\text{Var}(A_{n6}) = 4n^{-2}\{E[u'(l_{p_0} \otimes I_n)R_{Y_{0}}Y_{0}'R'(l_{p_0} \otimes I_n)\mu] + E(v' A_v R_{Y_{0}}Y_{0}'R'A_vv')\}
= 4n^{-2}\{\sigma^2_{e_0}\text{tr}(R_{Y_{0}}Y_{0}'R'(l_{p_0} \otimes I_n)\mu) + \sigma^2_{e_0}\text{tr}(A_v R_{Y_{0}}Y_{0}'R'A_v')\} = O(n^{-1}),
\]

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

Lemma B.8 Under the assumptions of Theorem 4.2, \( \frac{1}{\sqrt{n}} \frac{\partial \mathcal{L}(\phi_0)}{\partial \psi_0} \overset{d}{\longrightarrow} N(0(\Gamma_r)) \) as \( \sqrt{n} \to \infty \).

Proof. The key step of the proof is to show that \( \frac{1}{\sqrt{n}} \bar{X}'\bar{\Omega}_0^{-1}u \overset{d}{\longrightarrow} N(0(\Gamma_{r,11})) \) where \( \Gamma_{r,11} = \lim_{n \to \infty} (nT)^{-1}X_0\bar{\Omega}_0^{-1} \bar{X} \). By Cramér-Wold device, it suffices to show that for any \( c = (c_1', c_2', c_3') \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R} \) with \( ||c|| = 1 \), \( (nT)^{-1/2}c' \bar{X}'\bar{\Omega}_0^{-1} \bar{X} \overset{d}{\longrightarrow} N(0, c'\Gamma_{r,11}c) \). Using (B.3) and \( u = (v \otimes I_n)\mu + (I_T \otimes B_{-1})v \), we have \( c' \bar{X}'\bar{\Omega}_0^{-1}u = c_1 X_0^{-1}u + c_2 Z_0^{-1}u + c_3 Y_{-1}^{-1}u = \sum_{i=1}^3 T_{ni} \), where

\[
T_{n1} = [c_1' X + c_2' Z + c_3' \beta_0' X_0' + c_3' \gamma_0' z_{\gamma_0} (l_{p_0} \otimes I_n) + c_3' Y_0' I_0^{-1}(I_T \otimes I_n) \mu + c_3' \mu'(l_{p_0} \otimes I_n) \Omega_0^{-1}(I_T \otimes I_n) \mu, T_{n2} = c_1' X + c_2' Z + c_3' \beta_0' X_0' + c_3' \gamma_0' z_{\gamma_0} (l_{p_0} \otimes I_n) + c_3' Y_0' \Omega_0^{-1}(I_T \otimes B_{-1})v + c_3' v' A_0' \Omega_0^{-1}(I_T \otimes B_{-1})v, \]
\[
T_{n3} = c_3' \mu' (l_{p_0} \otimes I_n) \Omega_0^{-1}(I_T \otimes B_{-1}) + (I_T \otimes I_n) \Omega_0^{-1} A_v].
\]

29
It is easy to verify that $E(T_{n1}) = 0$, $E(T_{n2}) = c_3 \phi_m \sigma_0^2 \text{tr}[^{1} \nu_{v''o} \otimes I_n]$, and thus $E(T_{n3}) = -E(T_{n1})$ by Lemma B.6. Also, we can verify that $\text{Cov}(T_{ni}, T_{nj}) = 0$ for $i \neq j$. It suffices to show that each $T_{ni}$ (after appropriately centered for $T_{n1}$ and $T_{n2}$) is asymptotically normal with mean zero.

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

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

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

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

implying that $(nT)^{-1/2} \sqrt{\text{Var}(\Omega_0^{-1} u)} \xrightarrow{d} N(0, \Gamma_{r,11})$ because we can readily show that $(nT)^{-1/2} \sum_{i=1}^{3} \text{Var}(\Omega_0^{-1} u_i) = o_p(1)$.

Noticing that each component of $\partial L^*(\psi_0)/\partial v$ can be written as linear and quadratic functions of $\mu$ or $v$, the rest of the proof proceeds by following the above steps closely.

**Lemmas B.9-B.13 are used in the proof of Theorem 4.4.** for the SDPD model with random effects and endogenous $y_n$. Let $R_{ts}$ be an $n \times n$ symmetric and positive semidefinite (p.s.d.) nonstochastic square matrix for $t, s = 0, 1, \cdots, T-1$. Assume that $R_{ts}$ are uniformly bounded in both row and column sums. Recall for this case, $\chi_t = \sum_j x_j x_t - j$ and $\psi_t = \sum_j x_j x_t - j$.

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

1. $E(\psi_t' R_{ts} \psi_s) = \sigma_0^2 \text{tr}(B_0^{-1} R_{ts} B_0^{-1}) \sum_{i=\max(0, t-s)}^{t+s-1} \rho_0^{i-t+2s}$,
2. $E(\chi_t' R_{ts} \chi_s) = \text{tr}(\sum_{j=0}^{s+t-m-1} \sum_{k=0}^{t+s-1} \rho_0^{j+k} R_{ts} E(x_{s-j} x_{t-k}))$,
3. $E(\psi_t' R_{ts} \psi_s) = 0$.

**Proof.** Let $P_j = \rho_0^j B_0^{-1}$. Then $\psi_t = \sum_{j=0}^{t+s-1} P_j v_{t-j}$. Noting that $E(v_i' D v_s) = \sigma_0^2 \text{tr}(D)$ for any nonstochastic conformable matrix $D$ if $t = s$ and 0 otherwise, we have

$$
E(\psi_t' R_{ts} \psi_s) = \sum_{i=0}^{s+t-m-1} \sum_{k=0}^{t+s-1} E(v_i' P_i' R_{ts} P_j v_s) = \sum_{i=\max(0, t-s)}^{t+s-1} \sigma_0^2 \text{tr}(B_0^{-1} R_{ts} B_0^{-1}) \rho_0^{i-t+2s}.
$$

Next, noting that $\chi_t = \sum_{j=0}^{t+s-1} \rho_0^j x_t - j$, we have

$$
E(\chi_t' R_{ts} \chi_s) = \sum_{j=0}^{s+t-m-1} \sum_{k=0}^{t+s-1} \rho_0^{j+k} E(x_{t-k} R_{ts} x_{s-j}) = \text{tr}(\sum_{j=0}^{s+t-m-1} \sum_{k=0}^{t+s-1} \rho_0^{j+k} R_{ts} E(x_{s-j} x_{t-k})).
$$

Lastly, $E(\psi_t' R_{ts} \psi_s) = \sum_{j=0}^{s+t-m-1} \sum_{k=0}^{t+s-1} \rho_0^{j+k} E(x_{t-k} R_{ts} B_0^{-1} v_{s-j}) = 0$.■

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

1. $\text{Cov}(\psi_t' R_{ts} \psi_s, \psi_g' R_{gh} \psi_h) = \rho_{sg,h} \{ \sum_{i=0}^{n} \beta_{[s,i]} \beta_{[g,h,i]} + 2 \sigma_0^2 \text{tr}(B_0^{-1} R_{ts} B_0^{-1} R_{gh} B_0^{-1}) \}

+ \rho_{sg,h,2} \sigma_0^2 \text{tr}(B_0^{-1} R_{ts} B_0^{-1} R_{gh} B_0^{-1}) \rho_{gh,h} B_0^{-1} + \rho_{sg,h,3} \sigma_0^2 \text{tr}(B_0^{-1} R_{ts} B_0^{-1} R_{gh} B_0^{-1}) \rho_{gh,h} B_0^{-1} \}$,
2. $\text{Cov}(\chi_t' R_{ts} \chi_s, \chi_g' R_{gh} \chi_h) = \sigma_0^2 \text{tr}(\sum_{j=0}^{s+t-m-1} \sum_{k=0}^{t+s-1} \sum_{h=0}^{n} \sum_{k=0}^{n} \sum_{j=\max(0, s-k)}^{t+s-1} \rho_0^{j+k+h+s+2j} R_{ts} \times (B_0^{-1} R_{gh} E(x_{s-h} x_{t-k})))$,
3. $\text{Cov}(\chi_t' R_{ts} \chi_s, \chi_g' R_{gh} \chi_h) = O(n)$.
where \( \mathcal{B}_{ts,ii} \) denotes the \((i,i)\)th element of \( \mathcal{B}_{ts} \equiv B_0^{-1} R_{ts} B_0^{-1} \), \( \rho_{sgh} = \sum_{t=t-s-t}^{t=m-1} \rho_0^{h-t-s-t} \), \( \rho_{sgh,1} = \sum_{t=t-s-t}^{t=m-1} \rho_0^{h-t-s-t} \), and \( \rho_{sgh,2} = \sum_{t=t-s-t}^{t=m-1} \rho_0^{h-t-s-t} \sum_{j=0}^{m-1} \rho_0^{h-s-j} \) \((j \neq i + s - t) \).

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

\[
E[(v'_i R_1 v_s) (v'_g R_2 v_h)] = \begin{cases} 
\kappa_i \sum_{t=0}^{m-1} \rho_0^{t-s+m-1} \sum_{s=0}^{m-1} \sum_{j=0}^{s-m-1} (i+j+s) & \text{if } t = s = g = h \\
\sigma_{v_0}^4 \text{tr}(R_1) \text{tr}(R_2) & \text{if } t = s \neq g = h \\
\sigma_{v_0}^4 \text{tr}(R_1 R_2) & \text{if } t = g \neq s = h \\
\sigma_{v_0}^4 \text{tr}(R_1 R_2)' & \text{if } t = h \neq s = g \\
0 & \text{otherwise}
\end{cases}
\]

Consequently,

\[
E(V'_t R_s V'_t R'_s R_{gh} V_h) = E[(V'_t R_t V_s (X'_t R_{gh} V_s)')'] = \begin{cases} 
\kappa_i \sum_{t=0}^{m-1} \rho_0^{t+s+m-1} \sum_{s=0}^{m-1} \sum_{j=0}^{s+m-1} \rho_0^{i+j+k} E[x'_i - R_t B_0^{-1} v_s - j x'_g - k R_{gh} B_0^{-1} v_h - t]' \\
= \sum_{t=0}^{m-1} \sum_{s=0}^{m-1} \sum_{j=0}^{m-1} \rho_0^{i+j+k} \sum_{t=0}^{m-1} \sum_{s=0}^{m-1} \rho_0^{j+k+s-t} R_{ts} B_0^{-1} R_{gh} B_0^{-1} + C(h) \end{cases}
\]

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

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

(1) \((nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} [V'_t R_t V_s - E(V'_t R_t V_s)] \to P 0,

(2) \((nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} X'_t R_t V_s \to P 0,

(3) \((nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} X'_t R_t X_s - E(X'_t R_t X_s) \to P 0.

**Proof.** By Lemmas B.1, B.2, B.9, and B.10, we can show that \((nT)^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} E(V'_t R_t V_s) = O(1)\), and \(\text{Var}(n^{-1} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} V'_t R_t V_s) = n^{-2} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \text{Var}(V'_t R_t V_s) = O(n^{-2})\). Then (1) follows from Chebyshev inequality. For (2), we have \(E[1] \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \sum_{0}^{T-1} X'_t R_t V_s = 0,\)

31
where the last equality follows because (i) $x_{it}$ are independent across $i$ with second moments uniformly bounded in $i$, (ii) $R_t(B_0^g B_0^r)^{-1} R_{gh}^r$ are uniformly bounded in both row and column sums by Lemmas B.1-B.2, and (iii) elements of $R_t(B_0^g B_0^r)^{-1} R_{gh}^r E(x_{g-k}'x_{i-t}^r)$ are uniformly bounded by the same lemmas. Hence the conclusion follows from Chebyshev inequality. (3) follows from Lemma B.10 and Chebyshev inequality.

**Lemma B.12** Under the assumptions of Theorem 4.4, \( \frac{1}{n(T+1)} \left\{ \frac{\partial^2 \mathcal{L}^*(\psi)}{\partial \psi \partial \psi} - E \left[ \frac{\partial^2 \mathcal{L}^*(\psi)}{\partial \psi \partial \psi} \right] \right\} = o_p(1). \)

**Proof.** Let $u^* = u^*(\theta_0, \rho_0)$ and $u^*_p = u^*_p(\theta_0, \rho_0) = \frac{\partial}{\partial \rho} u^*(\theta_0, \rho_0)$. Noting that $E(X^*RU^*) = 0$ for any $n(T+1) \times n(T+1)$ nonstochastic matrix $R$ and $X^*_p$ is free of $\rho$, by the expressions of the Hessian matrix $\frac{\partial^2 \mathcal{L}^*(\psi)}{\partial \psi \partial \psi}$ in Section 4.2, it suffices to prove

(i) $n^{-1} \left[ X^* \Omega^{-1} \Omega_0^{-1} X^* - E \left( X^* \Omega_0^{-1} X^* \right) \right] = o_p(1);$

(ii) $n^{-1} X^* R u^* = o_p(1)$ for $R = \Omega_0^{-1}$ and $P_{\psi 0}^*$ with $\omega = \rho, \lambda, \phi_\mu,$ and $\phi_\xi$;

(iii) $n^{-1} [u^* R u^* - E(u^* R u^*)] = o_p(1)$ for $R = \Omega_0^{-1}$ and $P_{\psi 0}^*$ with $\omega = \rho, \lambda, \phi_\mu,$ and $\phi_\xi$;

(iv) $n^{-1} [X^* \Omega_0^{-1} u^* - E(X^* \Omega_0^{-1} u^*)] = o_p(1);$

(v) $n^{-1} [X^* \Omega_0^{-1} u^*_p - E(X^* \Omega_0^{-1} u^*_p)] = o_p(1);$

(vi) $n^{-1} [u^*_p R u^* - E(u^*_p R u^*)] = o_p(1)$ for $R = \Omega_0^{-1}$ and $P_{\psi 0}^*$ with $\omega = \rho, \lambda, \phi_\mu,$ and $\phi_\xi$;

(vii) $n^{-1} [u^*_p R u^* - E(u^*_p R u^*)] = o_p(1);$

(viii) $n^{-1} [u^*_p \Omega_0^{-1} u^*_p - E(u^*_p \Omega_0^{-1} u^*_p)] = o_p(1);$

(ix) $n^{-1} [\omega_{\psi} u^* - E(\omega_{\psi} u^*)] = o_p(1)$ for $\omega, \bar{\omega} = \rho, \lambda, \phi_\mu,$ and $\phi_\xi$.

Let $\Omega_0^{-1} = \frac{\partial^2}{\partial \rho \partial \rho} \Omega^*(\delta_0)$ for $\rho, \lambda, \phi_\mu,$ and $\phi_\xi$. Noting that $\Omega_0^{-1}, \Omega_0^*, P_{\psi 0}^*$ and $\Omega_0^* \omega_{\psi 0}$ with $\omega, \bar{\omega} = \rho, \lambda, \phi_\mu,$ and $\phi_\xi$ are uniformly bounded in both row and column sums and $q_{\psi 0}(u^*)$ is quadratic in $u^*$, we can readily show that (i)-(iv) and (ix) hold by straightforward moment calculations and Chebyshev inequality. Noting that $u^*_p = \left( \begin{array}{c} \bar{a}_{m0}\gamma_0 \\ Y_{-1} \end{array} \right)$ and $u^*_p = \left( \begin{array}{c} \bar{a}_{m0}\gamma_0 \\ 0_{nT \times 1} \end{array} \right)$ with $\bar{a}_{m0} = \frac{\partial}{\partial m} a_m (\rho_0)$ and $\bar{m0} = \frac{\partial}{\partial m} a_m (\rho_0)$, we can readily prove (v)-(vii) by Chebyshev inequality. In fact, $E(u^*_p \Omega_0^{-1} u^*) = 0$ in (vii).
Lemma B.13. Under the assumptions of Theorem 4.4, \(X\) and arguments to those used in proving Lemmas B.9-B.11.

\[
\frac{1}{nT} \sum_{t=1}^{nT} \frac{\partial \mathcal{L}^{(\psi)}(\psi)}{\partial \psi} \xrightarrow{d} N(0, \Gamma_{rr}).
\]

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

\[
S_n^* = \frac{1}{\sqrt{nT}} \left[ c_1 \frac{\partial \mathcal{L}^{(\psi)}(\psi)}{\partial \psi} + c_2 \frac{\partial \mathcal{L}^{(\psi)}(\psi)}{\partial \psi} + c_3 \frac{\partial \mathcal{L}^{(\psi)}(\psi)}{\partial \psi} + c_4 \frac{\partial \mathcal{L}^{(\psi)}(\psi)}{\partial \psi} + c_5 \frac{\partial \mathcal{L}^{(\psi)}(\psi)}{\partial \psi} + c_6 \frac{\partial \mathcal{L}^{(\psi)}(\psi)}{\partial \psi} \right]
\]

\[
= \frac{1}{\sqrt{nT}} \left[ \frac{1}{\sigma c_0} \left( c_1 X^* \Omega^{-1} u - \frac{c_3}{\sigma c_0} \frac{\partial \mathcal{L}^{(\psi)}(\psi)}{\partial \psi} \right) + \frac{c_4}{\sigma c_0} \left( \frac{\partial \mathcal{L}^{(\psi)}(\psi)}{\partial \psi} \right) \right]
\]

\[
= S_{n1}^* + S_{n2}^* + S_{n3}^* - E(S_{n3}^*)
\]

33
It follows that

\[ + \bar{\Omega} = \frac{-\hat{\sigma}_{\tilde{w}0}}{\sigma_{w0}} \Omega_0^{-1} + c_3 P_{\rho 0} + c_4 P_{\rho \omega 0} + c_5 P_{v \omega 0} + c_6 P_{\phi 0} \]  

Note that

\[ S_{n1}^* = \frac{1}{\sqrt{nT}} \frac{1}{\sigma v_0} c_1'(X^v \Omega_0^{-1} u^*)_n, S_{n2}^* = \frac{1}{\sqrt{nT}} \frac{1}{\sigma v_0} c_3 u^*_m \Omega_0^{-1} u^*_n, S_{n3}^* = \frac{1}{\sqrt{nT}} \frac{1}{\sigma v_0} u^* \Theta_0^* u^*_n \] 

where \( S_{n1}, S_{n2}, S_{n3} \) are linear in \( \zeta, \mu, v \) and \( v_{-j}'s \), respectively. Similarly

\[ S_{n3}^* = \frac{1}{\sqrt{nT}} \frac{1}{2 \sigma v_0} \left\{ \bar{\omega}_1 \zeta + \mu' [a_{m0} \zeta_1 + (\nu' \otimes I_n) \hat{\omega}_2] + 2a_{m0} (\nu' \otimes I_n) \hat{\omega}_2 \right\} + \nu' (I_T \otimes B_0^{-1}) \hat{\omega}_2 v + \left( \sum_{j=0}^{m-1} \rho_j B_0^{-1} v_{-j} \right) \bar{\omega}_1 + \sum_{j=0}^{m-1} \rho_j B_0^{-1} v_{-j} \right\} \]

where \( \bar{\omega}_1 = \left( \begin{array}{c} \bar{\omega}_1 \\ \bar{\omega}_2 \\ \bar{\omega}_3 \end{array} \right) \) with \( \bar{\omega}_1, \bar{\omega}_2, \bar{\omega}_3 \) being \( n \times n, nT \times n \), and \( nT \times nT \) matrices.

Apparently, \( S_{n3}^* \) can be written as the summation of five asymptotically independent terms, i.e., \( S_{n3}^* = \sum_{j=1}^{5} S_{n3,j}^* \), where \( S_{n3,1}^*, S_{n3,2}^*, S_{n3,3}^*, \) and \( S_{n3,4}^* \) are quadratic functions of \( \zeta, \mu, v, \) and \( v_{-j}'s \), respectively, and \( S_{n3,5}^* \) is the summation of terms that are bilinear in any two of \( \zeta, \mu, v, \) and \( v_{-j}'s \). Analogous to the proof of Lemma B.8, we can use \( u_5^* = - (\hat{a}_{m0} (z_0)^', Y_{-1}') \) and the expression of \( Y_{-1} \) in (B.2) to write \( S_{n2}^* = \sum_{j=1}^{6} S_{n2,j}^* \), where \( S_{n2,1}^*, S_{n2,2}^*, \) and \( S_{n2,3}^* \) are quadratic functions of \( \mu, v, \) and \( v_{-j}'s \), respectively, \( S_{n2,4}^* \) is a bilinear function that contains summation of terms which are linear in any two of \( \zeta, \mu, v, \) and \( v_{-j}'s \), and \( S_{n2,5}^* \) is the summation of terms that are linear in one of \( \zeta, \mu, v, \) and \( v_{-j}'s \). Consequently, we can write \( S_n^* = \sum_{j=1}^{6} s_{n,j}^* \), where \( s_{n,1}, \ldots, s_{n,6}^* \) are quadratic functions of \( \zeta, \mu, v, \) and \( v_{-j}'s \), respectively, \( s_{n,5}^* \) is a summation of terms that are linear in any two of \( \zeta, \mu, v, \) and \( v_{-j}'s \), and \( s_{n,6}^* \) is summation of terms that are linear in \( \zeta, \mu, v, \) and \( v_{-j}'s \). By the mutual independence of \( \zeta, \mu, v, \) and \( v_{-j}'s \) and their zero mean property, these six terms are either independent or asymptotically independent. By Lemma B.5,

\[ \{ s_{n,j}^* - E(s_{n,j}^*) \}/\sqrt{\text{Var}(s_{n,j}^*)} \xrightarrow{d} N(0,1). \]

It follows that \( S_n^* \xrightarrow{d} N(0, \lim_{n \to \infty} \sum_{j=1}^{6} \text{Var}(s_{n,j}^*)) \), implying that \( S_n^* \xrightarrow{d} N(0, cT_{\rho} \cdot c) \)

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

**Lemma B.14** Under the assumptions of Theorem 4.6, \( \frac{1}{nT} \left\{ \frac{\partial c'(w_0)}{\partial (w_0)} - E \left[ \frac{\partial c'(w_0)}{\partial (w_0)} \right] \right\} \) is \( o_p(1) \).
Proof. Noting that $E(\Delta X'\Delta u) = 0$ for any $nT \times nT$ nonstochastic matrix $R$, by the expressions of the Hessian matrix $\frac{\partial^2 f(v_\theta)}{\partial \theta \partial \phi}$ in Section 4.3, it suffices to prove

(i) $n^{-1}[\Delta X'\Omega_{n}^{1/2} \Delta X] \equiv o_p$ for $R = \Omega_{n}^{1/2}$ and $P_{\nu(x)}$ with $\omega = \rho$, $\lambda$, and $\phi$.

(ii) $n^{-1}[\Delta u'\Delta u - \sigma_{\nu(x)}(\Omega_{n}^{1/2} \Delta u)] \equiv o_p(1)$ for $R = \Omega_{n}^{1/2}$ and $P_{\nu(x)}$ with $\omega = \rho$, $\lambda$, and $\phi$.

(iii) $n^{-1}[\Delta u'\Delta u - \sigma_{\nu(x)}^2(\Omega_{n}^{1/2} \Delta u)] \equiv o_p(1)$ for $R = \Omega_{n}^{1/2}$ and $P_{\nu(x)}$ with $\omega = \rho$, $\lambda$, and $\phi$.

(iv) $n^{-1}[\Delta X'\Omega_{n}^{1/2} \Delta u - E(\Delta X'\Omega_{n}^{1/2} \Delta u)] \equiv o_p(1)$ for $R = \Omega_{n}^{1/2}$ and $P_{\nu(x)}$ with $\omega = \rho$, $\lambda$, and $\phi$.

(v) $n^{-1}[\Delta u'\Delta u - E(\Delta u'\Delta u)] \equiv o_p(1)$ for $R = \Omega_{n}^{1/2}$ and $P_{\nu(x)}$ with $\omega = \rho$, $\lambda$, and $\phi$.

(vi) $n^{-1}[\Delta u'\Omega_{n}^{1/2} \Delta u - E(\Delta u'\Omega_{n}^{1/2} \Delta u)] \equiv o_p(1)$ for $R = \Omega_{n}^{1/2}$ and $P_{\nu(x)}$ with $\omega = \rho$, $\lambda$, and $\phi$.

(vii) $n^{-1}[\Omega_{n}^{1/2}(\Delta u) - E(\Omega_{n}^{1/2}(\Delta u))] \equiv o_p(1)$ for $R = \Omega_{n}^{1/2}$ and $P_{\nu(x)}$ with $\omega = \rho$, $\lambda$, and $\phi$.

Let $\Omega_{n}^{1/2} = \frac{\partial^2 f(v_\theta)}{\partial \theta \partial \phi}$. Noting that $\Omega_{n}^{1/2}$, $\Omega_{n}^{1/2}$, $P_{\nu(x)}$ with $\omega = \rho$, $\lambda$, and $\phi$ are uniformly bounded in both row and column sums and $\Omega_{n}^{1/2}(\Delta u)$ is quadratic in $\Delta u$, we can show that (i)-(vii) hold by straightforward moment calculations and Chebyshev inequality. Below we only demonstrate the proof of (ii) and (vii) since the proof of the other claims is similar or simpler.

Since $E(\Delta u'\Delta u) = \sigma_{\nu(x)}^2(\Omega_{n}^{1/2})$, by Chebyshev inequality (ii) follows provided $\text{Var}(n^{-1}\Delta u'\Delta u) = o(1)$. Let $\Delta v(0) = B_0c + \rho_0 \sum_{j=0}^{m-1} \rho_0^j \Delta v_{-j}$, $\Delta v(1) = (\Delta v_1', \ldots, \Delta v_{-m}'\Delta v_{-m+1}')$, and $\Delta v = (\Delta v_{(0)}', \Delta v_{(1)}')$. Then $\Delta u = (I_n \otimes B_0^{-1}) \Delta v$ and $\Delta u'\Delta u = \Delta v'(I_n \otimes B_0^{-1}) R(I_n \otimes B_0^{-1}) \Delta v = \Delta v' \tilde{R} \Delta v$, where $\tilde{R} = (I_n \otimes B_0^{-1}) R(I_n \otimes B_0^{-1})$. Now, write

$$
\tilde{R} = \begin{pmatrix}
  R_{00} & R_{01} \\
  R_{10} & R_{11}
\end{pmatrix}
$$

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

$$
\nabla v' \tilde{R} \nabla v = \nabla v'(0) \tilde{R} \nabla v(0) + \nabla v'(1) \tilde{R} \nabla v(1) + \nabla v'(0)(R_{01} + R_{10}) \nabla v(1)
$$

$$
= \nabla v'(0) \tilde{R} \nabla v(0) + v'(C \otimes I_n) \tilde{R} \nabla v(1) + \nabla v'(0)(R_{01} + R_{10})(C \otimes I_n) v
$$

Then by Cauchy-Schwarz inequality

$$
\text{Var}(\Delta u'\Delta u) \leq 3\text{Var}(\nabla v'(0) \tilde{R} \nabla v(0)) + 3\text{Var}(v'(C \otimes I_n) \tilde{R} \nabla v(1)) + 3\text{Var}(\nabla v'(0)(R_{01} + R_{10})(C \otimes I_n) v).
$$

Write $\nabla v(0) = B_0c + v_1 + \rho_0 \sum_{j=0}^{m-1} \rho_0^j (\rho_0 - 1) v_{-j}$ and $\sum_{j=0}^{m-1} \rho_0^j (\rho_0 - 1) v_{-j}$. Since $B_0^T \tilde{R} B_0$ is uniformly bounded in both row and column sums, by Lemma B.3(1)

$$
\text{Var}(\nabla v'(0) \tilde{R} \nabla v(0)) = \kappa \sum_{i=1}^{n} \sigma_{\nu(x)}^2(\tilde{R} B_0^T B_0 \nabla v_i(0))^2 + \sigma_{\nu(x)}^2(\tilde{R} B_0^T B_0 \nabla v_i(0))^2 + \sigma_{\nu(x)}^2(\tilde{R} B_0^T B_0 \nabla v_i(0))^2 = O(n).
$$

Similarly, we can show that $\text{Var}(v_1 \tilde{R} B_0 v_1) = O(n)$, $\text{Var}(v_{-m+1} \tilde{R} B_0 v_{-m+1}) = O(n)$, and $\text{Var}(\sum_{j=0}^{m-2} \rho_0^j v_{-j} \tilde{R} B_0 v_{-j}) = O(n)$. It follows from Cauchy-Schwarz inequality that $\text{Var}(\nabla v'(0) \tilde{R} \nabla v(0)) = O(n)$. By the same token, we can show that $\text{Var}(v'(C \otimes I_n) \tilde{R} \nabla v(1)) = O(n)$, and $\text{Var}(\nabla v'(0)(R_{01} + R_{10})(C \otimes I_n) v) = O(n)$. This completes the proof of (ii).
Now, we show (vi). Let \( \Delta Y^* = (0_{1 \times n}, \Delta y_0', \ldots, \Delta y_{T-1}') \). Then \( \Delta u_\rho = -\Delta Y^* \). Let \( k_\rho = (0, 1, \rho, \ldots, \rho^{T-2})' \), \( X = (0_{1 \times n}, 0_{1 \times n}, (\Delta x_2^{b_0}), \ldots, \sum_{j=0}^{T-3} \rho_0^j (\Delta x_{T-1-j}^{b_0})') \), and \( V = (0_{1 \times n}, 0_{1 \times n}, (\Delta v_2)', \ldots, \sum_{j=0}^{T-3} \rho_0^j (\Delta v_{T-1-j})') \). Since \( \Delta y_t = \Delta x_\pi + \Delta x_{j_0} + \Delta u_1 + \)and

\[
\Delta y_t = \rho_0^{t-1} \Delta y_1 + \sum_{j=0}^{t-2} \rho_0^j \Delta x_{j_0} + \sum_{j=0}^{t-2} \rho_0^j B_0^{-1} \Delta v_{t-j} \quad \text{for } t = 2, 3, \ldots, \tag{B.5}
\]

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

\[
\text{Var} \left( \Delta u_\rho \Omega_0^{-1} \Delta u_\rho \right) \leq 3 \text{Var} \left( (k_\rho \otimes \Delta y_1) \Omega_0^{-1} (k_\rho \otimes \Delta y_1) \right) + 3 \text{Var} \left( \chi' \Omega_0^{-1} \chi \right) + 3 \text{Var} \left( \chi' (I_T \otimes B_0^{-1}) \Omega_0^{-1} (I_T \otimes B_0^{-1}) \chi \right).
\]

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

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

**Proof.** By Cramér-Wold device, it suffices to show that for any \( c = (c_1, c_2, c_3, c_4, c_5)' \in \mathbb{R}^{p+k} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \) with \( ||c|| = 1 \), \( S_n^c = \frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}(\psi)}{\partial \psi} ) \xrightarrow{d} N(0, c' \Gamma_f c) \). Recall \( \Delta u = \Delta u(\theta_0, \rho_0) \). Let \( \Delta u_\rho = -(0_{1 \times n}, \Delta y_0', \ldots, \Delta y_{T-1}') \), and \( P_{0 \rho}^j = P_{0 \rho}^j (\delta_0) \) for \( \omega = \rho, \lambda, \) and \( \phi_\zeta \). Using the expression for elements of \( \frac{\partial \mathcal{L}(\psi)}{\partial \psi} \) defined in Section 4.3, we can readily obtain

\[
S_n^c = \frac{1}{\sqrt{nT}} \left[ c_1 \frac{\partial \mathcal{L}(\psi)}{\partial \theta} \right]
\]

\[
= \frac{1}{\sqrt{nT}} \left[ c_1 \frac{\Delta X' \Omega_0^{-1} \Delta u}{\sigma_{\theta}^2} + c_2 \frac{\partial \mathcal{L}(\psi)}{\partial \sigma_{\theta}^2} + c_3 \frac{\partial \mathcal{L}(\psi)}{\partial \rho} + c_4 \frac{\partial \mathcal{L}(\psi)}{\partial \lambda} + c_5 \frac{\partial \mathcal{L}(\psi)}{\partial \phi_\zeta} \right]
\]

where \( S_{n1}^c = \frac{1}{\sqrt{nT}} \frac{\Delta X' \Omega_0^{-1} \Delta u}{\sigma_{\theta}^2} \), \( S_{n2}^c = \frac{1}{\sqrt{nT}} \frac{\Delta X' \Omega_0^{-1} \Delta u}{\sigma_{\theta}^2} \), \( S_{n3}^c = \frac{1}{\sqrt{nT}} \frac{\Delta X' \Omega_0^{-1} \Delta u}{\sigma_{\theta}^2} \), and \( S_{n4}^c = \frac{1}{\sqrt{nT}} \frac{\Delta X' \Omega_0^{-1} \Delta u}{\sigma_{\theta}^2} \). Analogous to the proof of Lemma B.13, one can write \( S_n^c = S_{n1}^c + S_{n2}^c + S_{n3}^c + S_{n4}^c \), where \( S_{n1}, S_{n2}, S_{n3}, S_{n4} \) are quadratic functions of \( \zeta, v, \) and \( v_{-j}'s \), respectively, \( S_{n5}^c \) is a sum of terms that are quadratic in \( \zeta, v, \) and \( v_{-j}'s \). By the mutual independence of \( \zeta, v, \) and \( v_{-j}'s \) and their zero mean property, these five terms are either independent or asymptotically independent. By Lemma B.5,

\[
\text{Var} (S_{n5}^c) = 0, \quad \text{implies that } S_n^c \xrightarrow{d} N(0, \Omega_{\Gamma_f})
\]

It follows that \( S_n^c \xrightarrow{d} N(0, \lim_{m \to \infty} \sum_{j=1}^{5} \text{Var}(s_{n5}^c)) \), implying that \( S_n^c \xrightarrow{d} N(0, \Omega_{\Gamma_f}) \)
Appendix C: Proofs of the Theorems

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

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

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

By (4.1) and the fact that $Y = \bar{X}\theta + u$, $\bar{u}(\delta)$ is defined in (3.2). Combining (C.2)-(C.3) yields
\[
\hat{\delta}^2(\delta) = \frac{1}{nT}E\left\{(u - \bar{X}\theta^* (\delta))'\Omega^{-1}(u - \bar{X}\theta^* (\delta))\right\}
\]
\[
= \frac{1}{nT}E\left\{u'(\Omega^{-1}u) + \frac{1}{nT}\theta^* (\delta)'E(\bar{X}'\Omega^{-1}\bar{X})\theta^* (\delta) - \frac{2}{nT}\theta^* (\delta)'E(\bar{X}'\Omega^{-1}u)\right\}
\]
\[
= \frac{\sigma^2_\bar{u}}{nT}\text{tr} (\Omega^{-1}\Omega_0) - \frac{1}{nT}\left\{E(\bar{X}'\Omega^{-1}u)^' [E(\bar{X}'\Omega^{-1}\bar{X})]^{-1}E(\bar{X}'\Omega^{-1}u)\right\},
\] (C.3)

where recall $\Omega_0 \equiv (\Omega(\delta_0)$ and $\Omega(\delta)$ is defined in (3.2). Combining (C.2)-(C.3) yields
\[
\hat{\delta}^2(\delta) - \delta^2(\delta) = \frac{1}{nT}[u'(\Omega^{-1}u - \sigma^2_\bar{u}\text{tr} (\Omega^{-1}\Omega_0))] - \frac{1}{nT}u'(\Omega^{-1}2\Omega^{-1/2}P\Omega^{-1/2}u)
\]
\[
+ \frac{1}{nT}\left\{E(\bar{X}'\Omega^{-1}u)^' [E(\bar{X}'\Omega^{-1}\bar{X})]^{-1}E(\bar{X}'\Omega^{-1}u)\right\}
\]
\[
= \frac{1}{nT}\left\{\text{tr}(\Omega^{-1}(uu' - \sigma^2_\bar{u}\Omega_0))\right\}
\]
\[
- \left\{Q_{\bar{x}u}(\delta)' Q_{\bar{x}u}(\delta)^{-1}Q_{\bar{x}u}(\delta) - \left\{E[Q_{\bar{x}u}(\delta)]\right\}' \left\{E[Q_{\bar{x}u}(\delta)]\right\}^{-1}E[Q_{\bar{x}u}(\delta)]\right\}
\]
\[
\equiv \Pi_{n1}(\delta) - \Pi_{n2}(\delta),
\]

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

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

\[
\inf_{\delta \in \Delta} \lambda_{\min}(\Omega(\delta)) \geq \inf_{\delta \in \Delta} \lambda_{\min} (\phi_{\mu}(I_{nT} \otimes I_{n_u}) + I_{T} \otimes [B(\lambda)B(\lambda)]^{-1})
\]
\[
\geq \inf_{\lambda \in \Lambda} \lambda_{\min} (I_{T} \otimes [B(\lambda)B(\lambda)]^{-1}) \geq \inf_{\lambda \in \Lambda} (\inf_{\lambda \in \Lambda} (B(\lambda)B(\lambda)]^{-1})
\]
\[
\geq \inf_{\lambda \in \Lambda} \{\lambda_{\min}(B(\lambda))\}^2 = \{\sup_{\lambda \in \Lambda} \lambda_{\max}(B(\lambda))\}^2 \geq c_\lambda^{-2} > 0
\] (C.4)
by Cauchy-Schwarz inequality,
\[ |\Pi_n(\delta) - \Pi_{n1}(\delta)| = \frac{1}{nT} \text{tr} \{ \Omega(\delta)^{-1} [\Omega(\delta) - \Omega(\delta)] \Omega(\delta)^{-1} (uu' - \sigma^2_v\Omega_0) \} \]
\[ \leq \frac{1}{nT} \| \text{tr} \{ \Omega(\delta)^{-1} [\Omega(\delta) - \Omega(\delta)] \Omega(\delta)^{-1} \} \|^{1/2} \| uu' - \sigma^2_v\Omega_0 \| \]
\[ \leq \frac{1}{\sqrt{nT}} \| \Omega(\delta) - \Omega(\delta) \| \| uu' - \sigma^2_v\Omega_0 \| \cdot \]

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

For \( \Pi_n(\delta) \), we decompose it as follows
\[ \Pi_n(\delta) = \{ Q_{ux}(\delta) - E[Q_{ux}(\delta)] \}' Q_{xx}(\delta)^{-1} Q_{ux}(\delta) \]
\[ + \{ E[Q_{ux}(\delta)] \}' Q_{xx}(\delta)^{-1} \{ E[Q_{ux}(\delta)] - Q_{xx}(\delta) \} \{ E[Q_{ux}(\delta)] \}^{-1} Q_{ux}(\delta) \]
\[ + \{ E[Q_{ux}(\delta)] \}' \{ E[Q_{xx}(\delta)] \}^{-1} \{ Q_{ux}(\delta) - E[Q_{ux}(\delta)] \} \]
\[ \equiv \Pi_{n2,1}(\delta) + \Pi_{n2,2}(\delta) + \Pi_{n2,3}(\delta), \text{ say.} \]

By Assumption G1(v), \( \sup \phi_{\mu} \leq c_\phi \) for some \( c_\phi < \infty \). Noting that by G2(v)
\[ \sup_{\delta \in \Delta} \lambda_{\max}(\Omega(\delta)) \leq \sup_{\delta \in \Delta} \lambda_{\max} \left\{ \phi_{\mu} (J_T \otimes I_n) + I_T \otimes [B(\lambda)'B(\lambda)]^{-1} \right\} \]
\[ \leq \sup_{\phi_{\mu}} \left\{ \phi_{\mu} \lambda_{\max}(J_T \otimes I_n) + \lambda_{\max} \left\{ [B(\lambda)'B(\lambda)]^{-1} \right\} \right\} \]
\[ \leq c_\phi T + \left\{ \inf_{\lambda \in \Lambda} \lambda_{\min}(B(\lambda)) \right\}^{-2} \leq c_\phi T + \frac{1}{\Delta^2} < \infty, \] \hfill (C.5)

we have
\[ \inf_{\delta \in \Delta} \lambda_{\min}(Q_{xx}(\delta)) \geq \left[ \sup_{\delta \in \Delta} \lambda_{\max}(\Omega(\delta)) \right]^{-1} \lambda_{\min} \left( \frac{1}{nT} \hat{X}'\hat{X} \right) \]
\[ \geq \left( c_\phi T + \frac{1}{\Delta^2} \right)^{-1} \lambda_{\min} \left( \frac{1}{nT} \hat{X}'\hat{X} \right). \]

This implies that \( \sup_{\delta \in \Delta} \| Q_{xx}(\delta)^{-1} \| = O_P(1) \) by Assumption R(iv). It is straightforward to show that \( Q_{ux}(\delta) - E[Q_{ux}(\delta)] = O_P(1) \) uniformly in \( \delta \) by Chebyshev inequality and the arguments for stochastic equicontinuity. In addition, \( E[Q_{ux}(\delta)] = O(1) \) uniformly in \( \delta \). So \( Q_{ux}(\delta) = O_P(1) \) uniformly in \( \delta \). Consequently,
\[ |\Pi_{n2,1}(\delta)| \leq \| Q_{ux}(\delta) - E[Q_{ux}(\delta)] \| \| Q_{xx}(\delta)^{-1} \| \| Q_{ux}(\delta) \| \]
\[ = O_P(1) O_P(1) O_P(1) \text{ uniformly in } \delta. \]

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

To show (ii), we can define an auxiliary process \( \{ U_{nT} \} \) such that (3.1) is now satisfied with \( u \) replaced by \( U_{nT} \) and \( U_{nT} \sim N(0, \sigma^2_v\Omega) \) with \( \Omega = \Omega(\delta) \) and is independent of \( (X, Z) \). If \( u \) is normally distributed,
one just sets $U_{nT} = u$. The true value of $(\theta, \sigma^2_\omega, \delta)$ is given by $(\theta_0, \sigma_{\omega_0}^2, \delta_0)$. Now the quasi-log-likelihood function $\mathcal{L}_T^\ast(\psi)$ in (3.3) becomes the exact log-likelihood function. By the principle of maximum likelihood and Jensen inequality, one can readily show that $\mathcal{L}_T^\ast(\delta) \leq \mathcal{L}_T^\ast(\delta_0)$ for any $\delta \in \Delta$. Observing that $\sigma^2_\omega(\delta) = \frac{\partial^2}{\partial \omega^2} \text{tr}(\Omega^{-1}_0 \Omega_0) = \sigma^2_{\omega_0}$ by (C.3) and Lemma B.6, we have

$$\frac{1}{nT} \left[ \mathcal{L}_T^\ast(\delta) - \mathcal{L}_T^\ast(\delta_0) \right] = \frac{1}{2nT} \left\{ \log |\Omega_0| - \log |\Omega(\delta)| \right\} + \frac{1}{2} \left\{ \log \left[ \sigma^2_\omega(\delta_0) \right] - \log \left[ \sigma^2_\omega(\delta) \right] \right\}$$

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

**Proof of Theorem 4.2.** By Taylor series expansion,

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

where elements of $\tilde{\psi} = (\bar{\theta}, \sigma^2_\omega, \lambda)'$ lie in the segment joining the corresponding elements of $\psi$ and $\psi_0$ and $\bar{\delta} = (\bar{\lambda}, \bar{\delta}_0)'$. Thus

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

By Theorem 4.1, $\tilde{\psi} \xrightarrow{p} \psi_0$. Consequently, $\tilde{\psi} \xrightarrow{d} \psi_0$, and it suffices to show that: (i) $\frac{1}{nT} \frac{\partial^2 \mathcal{L}_T(\psi)}{\partial \omega \partial \omega'} = o_P(1)$, (ii) $\frac{1}{nT} \frac{\partial^2 \mathcal{L}_T(\psi_0)}{\partial \omega \partial \omega'} \xrightarrow{d} H_\tau$, and (iii) $\frac{1}{nT} \frac{\partial \mathcal{L}_T(\psi_0)}{\partial \omega} \xrightarrow{d} N(0, \Sigma)$. (ii) and (iii) follow from Lemmas B.7 and B.8, respectively. We are left to show (i).

With the expression of $\frac{\partial^2}{\partial \omega \partial \omega'} \mathcal{L}_T(\psi)$ given in Section 4.2, it suffices to show that $\frac{1}{nT} \frac{\partial^2 \mathcal{L}_T(\psi_0)}{\partial \omega \partial \omega'} = o_P(1)$ for $\omega$, $\varpi = \theta$, $\sigma^2_\omega$, $\lambda$, and $\phi_\mu$. We do this only for the cases of $(\omega, \varpi) = (\theta, \theta)$, $(\theta, \sigma^2_\omega)$, and $(\sigma^2_\omega, \sigma^2_\omega)$ as the other cases can be shown analogously. First, write

$$- \frac{1}{nT} \left[ \frac{\partial^2 \mathcal{L}_T(\psi)}{\partial \theta \partial \theta'} + \frac{\partial^2 \mathcal{L}_T(\psi_0)}{\partial \theta \partial \theta'} \right] = \left( \frac{1}{\sigma^2_\omega} \frac{1}{\sigma^2_{\omega_0}} \right) \frac{X' \Omega^{-1}(\delta)^{-1} \tilde{X}}{nT} + \frac{1}{nT \sigma^2_{\omega_0}} \tilde{X}' \Omega^{-1} - \Omega^{-1}_0 \tilde{X}.$$

Noting that $\sigma^2_\omega - \sigma^2_{\omega_0} = o_P(1)$ by Theorem 4.1 and $(nT)^{-1} \tilde{X}' \Omega^{-1}(\delta)^{-1} \tilde{X} = O_P(1)$, the first term on the right hand side of the last expression is $o_P(1)$. For the second term, we first show that

$$\lambda_{\max}(\Omega_0 - \Omega(\delta)) = O_P(\|\tilde{\delta} - \delta_0\|).$$

To see this, write $\Omega_0 - \Omega(\delta) = (\phi_{\mu_0} - \bar{\phi}_{\mu})(J_T \otimes I_n) + r_n(\lambda)$, where $r_n(\lambda) = I_T \otimes \{ [B(\lambda)']^2 B(\lambda)]^{-1} - [B(\lambda)'B(\lambda)]^{-1} \}$ is a symmetric matrix. By the repeated use of the fact that

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

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

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

39
where the last equality follows from Assumption G2 and the fact that
\[
\lambda_{\max}[B(\bar{\lambda})'B(\bar{\lambda}) - B(\lambda_0)'B(\lambda_0)] = \lambda_{\max}[(\lambda_0 - \bar{\lambda})(W' + W) + (\bar{\lambda}^2 - \lambda_0^2)W'W] \\
\leq |\bar{\lambda} - \lambda_0|\lambda_{\max}(W' + W) + (\bar{\lambda}^2 - \lambda_0^2)\lambda_{\max}(W'W) \\
= O_P(\bar{\lambda} - \lambda_0)
\]
under Assumption G2. Noting that \(\lambda_{\max}(J_F \otimes I_n) = T\), we can apply the fact that
\[
\lambda_{\max}(A + C) \leq \lambda_{\max}(A) + \lambda_{\max}(C)
\]  
(C.9)
to obtain \(\lambda_{\max}([\Omega_0 - \Omega(\bar{\delta})] \leq T\|\phi_{\mu_0} - \phi_p\| + \lambda_{\max}(r_n(\bar{\lambda})) = O_P(\|\bar{\delta} - \delta_0\|)\). Thus (C.7) follows. Let \(c\) be an arbitrary column vector in \(\mathbb{R}^{p+q+1}\). Then by Cauchy-Schwarz inequality, (C.4), and (C.7)
\[
\frac{1}{n}[c'\bar{X}'[\Omega(\bar{\delta})^{-1} - \Omega_0^{-1}]}\bar{X}c] \\
= \frac{1}{n}[c'\bar{X}'\Omega(\bar{\delta})^{-1}[\Omega_0 - \Omega(\bar{\delta})]\Omega_0^{-1}]}\bar{X}c] \\
\leq \frac{1}{n}[c'\bar{X}'\Omega(\bar{\delta})^{-1}][\Omega_0 - \Omega(\bar{\delta})][\Omega_0 - \Omega(\bar{\delta})\Omega(\bar{\delta})^{-1}]}\bar{X}c]^{1/2}[c'\bar{X}'\Omega_0^{-1}]}\bar{X}c]^{1/2} \\
\leq \lambda_{\max}[\Omega_0 - \Omega(\bar{\delta})][\lambda_{\max}(\Omega(\bar{\delta}))^{-1}][\lambda_{\max}(\Omega_0)]^{-1}\frac{1}{n}\|\bar{X}c\|^2 = O_P(\|\delta - \delta_0\|) = o_P(1).
\]  
(T.10)
It follows that the second term on the right hand side of (C.6) is \(o_P(1)\). Consequently, \(\frac{1}{nT}\frac{\partial^2 C_c(\psi)}{\partial \theta \partial \sigma_v} = o_P(1)\).

Next, we consider \(-\frac{1}{nT}\frac{\partial^2 C_c(\psi)}{\partial \theta \partial \sigma_v} + \frac{1}{nT}\frac{\partial^2 C_c(\psi)}{\partial \sigma_v \partial \sigma_v}\). This term is equal to
\[
\frac{1}{nT}\bar{X}'\Omega(\bar{\delta})^{-1}u(\bar{\theta}) - \frac{1}{nT}\bar{X}'\Omega_0^{-1}u(\bar{\theta}) \\
= \left(\frac{1}{\sigma_v^2} - \frac{1}{\sigma_v^2}\right)\frac{\bar{X}'\Omega(\bar{\delta})^{-1}}{nT}u(\bar{\theta}) + \frac{1}{\sigma_v^2}\frac{\bar{X}'\Omega_0^{-1}}{nT}u(\bar{\theta}) + \frac{1}{\sigma_v^2}\frac{\bar{X}'\Omega_0^{-1}}{nT}u(\bar{\theta}) - u(\bar{\theta})
\]
Using \(u(\bar{\theta}) = Y - \bar{X}\theta = u + \bar{X}(\theta_0 - \bar{\theta})\), we can readily show that \(\frac{1}{nT}\bar{X}'\Omega(\bar{\delta})^{-1}u(\bar{\theta}) = O_P(1)\), which implies that the first term in the last expression is \(o_P(1)\) by Theorem 4.1. The second term is \(o_P(1)\) by arguments analogous to those used above. The third term is \(\sigma_v^2(nT)^{-1}\bar{X}'\Omega(\bar{\delta})^{-1}\bar{X}(\theta_0 - \bar{\theta}) = O_P(1)\|\theta_0 - \bar{\theta}\| = o_P(1)\). It follows that \(\frac{1}{nT}\frac{\partial^2 C_c(\psi)}{\partial \theta \partial \sigma_v} - \frac{1}{nT}\frac{\partial^2 C_c(\psi)}{\partial \sigma_v \partial \sigma_v} = o_P(1)\). Now, write
\[
-\frac{1}{nT}\left[\frac{\partial^2 C_c(\psi)}{\partial \theta \partial \sigma_v} - \frac{\partial^2 C_c(\psi)}{\partial \sigma_v \partial \sigma_v}\right] = \left(\frac{1}{\sigma_v^2}\frac{u(\bar{\theta})}'\Omega(\bar{\delta})^{-1}u(\bar{\theta}) - \frac{1}{\sigma_v^2}\frac{u(\bar{\theta})'}{\Omega_0^{-1}}u(\bar{\theta}) + \frac{1}{2}\right)\left(\frac{1}{\sigma_v^2} - \frac{1}{\sigma_v^2}\right).
\]
Clearly, the second term is \(o_P(1)\) by Theorem 4.1. We can use the decomposition \(u(\bar{\theta}) = u + \bar{X}(\theta_0 - \bar{\theta})\) and the consistency of \(\bar{\psi}\) to show that the first term is also \(o_P(1)\). This completes the proof. ■

Proof of Theorem 4.3

Let \(T_1 = T + 1\). As in the proof of Theorem 4.1, we prove the theorem by showing that (i) \(\frac{1}{nT_1}[\bar{L}^{cr}(\delta) - \bar{L}^{cr}(\epsilon)] \overset{P}{\to} 0\) uniformly in \(\epsilon \in \Delta\), and (ii) \(\lim_{\epsilon \to 0} \max_{\epsilon \in \Delta} \sup_{x \in N_x(\epsilon)} \frac{1}{nT_1}[\bar{L}^{cr}(\delta) - \bar{L}^{cr}(\epsilon)] < 0\) for any \(\epsilon > 0\). The proof of (ii) is almost identical to that of (i) in the proof of Theorem 4.1 and thus omitted.

By (3.14) and (4.6), \(\frac{1}{nT_1}[\bar{L}^{cr}(\delta) - \bar{L}^{cr}(\epsilon)] = \ln \bar{\delta}_e^2(\delta) - \ln \bar{\delta}_e^2(\epsilon)\). To show (i), it suffices to show
\[
\bar{\delta}_e^2(\delta) - \bar{\delta}_e^2(\epsilon) = o_P(1)\]  
uniformly on \(\Delta\).  
(C.11)
By the definition of $\bar{u}^*(\delta)$ below (3.13), we have $\bar{u}^*(\delta) = Y^*(\rho) - X^*(X^*\Omega^{-1}X^*)^{-1}X^*\Omega^{-1}Y^*(\rho) = \Omega^{*1/2}M^*\Omega^{*1/2}X^*$, where $M^* = I_{nT_1} - \Omega^{*-1/2}X^*(X^*\Omega^{*1/2}X^*)^{-1}X^*\Omega^{*1/2}X^*$ is a projection matrix. Observe that $Y^*(\rho) = Y^*(\rho_0) + Y^*(\rho) - Y^*(\rho_0) = X^*\theta_0 + u^* + (\rho_0 - \rho) Y^*_{-1}$ where $Y^*_{-1} = (0_{1\times n}, Y^*_{-1})$. This, in conjunction with the fact that $M^*\Omega^{*1/2}X^* = 0$, implies that

$$\tilde{\sigma}^2_v(\delta) = \frac{1}{nT_1} \bar{u}^*(\delta)'\Omega^{-1}\bar{u}^*(\delta) = \frac{1}{nT_1} Y^*(\rho)'\Omega^{-1} \Omega^{*1/2}M^*\Omega^{*1/2}Y^*(\rho)$$

$$= \frac{1}{nT_1} \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right]'\Omega^{-1/2}M^*\Omega^{-1/2} \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right]. \quad (C.12)$$

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

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

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

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

By (4.5),

$$\tilde{\sigma}^2_v(\delta) = \frac{1}{nT_1} E \left\{ \left[ X^*\theta^*(\delta) + u^* + (\rho_0 - \rho) Y^*_{-1} \right]'\Omega^{-1} \left[ X^*\theta^*(\delta) + u^* + (\rho_0 - \rho) Y^*_{-1} \right] \right\}$$

$$= \frac{1}{nT_1} E \left\{ \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right]'\Omega^{-1} \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right] \right\} + \frac{1}{nT_1} \theta^*(\delta)' E \left( X^*\Omega^{-1}X^* \right) \theta^*(\delta) + \frac{2(\rho_0 - \rho)}{nT_1} \theta^*(\delta)' E \left( X^*\Omega^{-1}Y^*_{-1} \right)$$

$$= \frac{1}{nT_1} E \left\{ \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right]'\Omega^{-1} \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right] \right\} + \frac{(\rho_0 - \rho)}{nT_1} \theta^*(\delta)' E \left( X^*\Omega^{-1/2}Y^*_{-1} \right). \quad (C.13)$$

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

$$\tilde{\sigma}^2_v(\delta) - \tilde{\sigma}^2_v(\delta) = \frac{1}{nT_1} \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right]'\Omega^{-1/2}M^*\Omega^{*1/2} \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right] - \tilde{\sigma}^2_v(\delta)$$

$$= \frac{1}{nT_1} \left\{ \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right]'\Omega^{-1/2}M^*\Omega^{*1/2} \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right] - E \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right]'\Omega^{-1/2}M^*\Omega^{*1/2} \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right] \right\}$$

$$+ Q_{xx}(\delta)' Q_{xx}(\delta) = \frac{2(\rho_0 - \rho)}{nT_1} \theta^*(\delta)' E \left( X^*\Omega^{-1}Y^*_{-1} \right) + (\rho_0 - \rho)^2 \left\{ Q_{zx}(\delta)' Q_{zx}(\delta)' Q_{yy}(\delta)' \right\}$$

$$= 1 \left( n_{T_1} \right) \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right]'\Omega^{-1/2}M^*\Omega^{*1/2} \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right] - E \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right]'\Omega^{-1/2}M^*\Omega^{*1/2} \left[ u^* + (\rho_0 - \rho) Y^*_{-1} \right] \right\}$$

$$\equiv \Pi_{n1}(\delta) + \Pi_{n2}(\delta) + 2(\rho_0 - \rho) \Pi_{n3}(\delta) + (\rho_0 - \rho)^2 \Pi_{n4}(\delta),$$

where $Q_{xx}(\delta) = \frac{1}{nT_1} X^*\Omega^{-1}X^*$, $Q_{zx}(\delta) = \frac{1}{nT_1} X^*\Omega^{-1}Y^*_1$, and $Q_{zx}(\delta) = \frac{1}{nT_1} X^*\Omega^{-1}Y^*_1$. We prove (i) by showing that $\Pi_{n1}(\delta) = \alpha_p(1)$ uniformly in $\delta$ for $s = 1, 2, 3, 4$.

We can decompose $\Pi_{n1}(\delta)$ as follows

$$\Pi_{n1}(\delta) = \frac{1}{nT_1} \left[ u^* \Omega^{-1/2} - E \left( u^* \Omega^{-1/2} u^* \right) \right] + \frac{(\rho_0 - \rho)^2}{nT_1} \left[ Y^*_1 \Omega^{-1/2} Y^*_1 - E \left( Y^*_1 \Omega^{-1/2} Y^*_1 \right) \right]$$

$$+ 2(\rho_0 - \rho) \left[ \left( u^* \Omega^{-1/2} Y^*_1 - E \left( u^* \Omega^{-1/2} Y^*_1 \right) \right) \right]$$

$$\equiv \Pi_{n1,1}(\delta) + \Pi_{n1,2}(\delta) + \Pi_{n1,3}(\delta),$$

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

$$|\Pi_{n,1}^*(\delta) - \Pi_{n,1}^*(\bar{\delta})| = \left| \frac{1}{nT_1} \text{tr} \left\{ \Omega^* (\delta)^{-1} [\Omega^* (\delta) - \Omega^*(\bar{\delta})] \Omega^* (\delta)^{-1} [u^* u'^* - E(u^* u'^*)] \right\} \right| \leq \frac{1}{nT_1} \left( \text{tr}(\Omega^* (\delta)^{-1} [\Omega^* (\delta) - \Omega^*(\bar{\delta})] \Omega^* (\delta)^{-2} [\Omega^* (\delta) - \Omega^*(\bar{\delta})] \Omega^* (\delta)^{-1}) \right)^{1/2} \times \|u^* u'^* - E(u^* u'^*)\| \leq \frac{1}{\sqrt{nT_1}} \|\Omega^* (\delta)^{-1}\| \|\Omega^* (\delta)\| \frac{1}{\sqrt{nT_1}} \|u^* u'^* - E(u^* u'^*)\| \cdot$$

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

To show $\Pi_{n,2}^*(\delta) = o_P(1)$ uniformly in $\delta$, we first argue that $\Omega^* (\delta)$ is positive definite uniformly in $\delta$, i.e.,

$$\inf_{\delta \in \Delta} \lambda_{\min} (\Omega^* (\delta)) \geq c^* \text{ for some } c^* > 0.$$

Let $u^* = (a_m, \mu, u'^*)$ and $\Omega^* (\delta) = \begin{pmatrix} \phi_\mu a_m^2 I_n & \phi_\mu a_m (\nu'^* \otimes I_n) \\ \phi_\mu a_m (\nu'^* \otimes I_n) & \Omega \end{pmatrix}$. Noting that $\bar{\Omega}^* (\delta) = E(\bar{u}^* \bar{u}'^*)$, it is positive semidefinite uniformly in $\delta$. By Theorem 8.4.11 in Bernstein (2005) and (C.4), $\lambda_{\min} (\phi_\nu I_n + b_m (B'B)^{-1}) \geq \phi_\nu + b_m \lambda_{\min} ((B'B)^{-1}) \geq \phi_\nu + b_m \delta^2 > 0$ uniformly in $\delta$ as $\phi_\nu$ is positive and bounded away from 0 and $b_m > 0$, implying that $\phi_\nu I_n + b_m (B'B)^{-1}$ is positive definite uniformly in $\delta$. Noting $\Omega^* (\delta)$ is equal to $\bar{\Omega}^* (\delta)$ with its upper-left $(n, n)$-submatrix added by a uniformly positive definite matrix $\phi_\nu I_n + b_m (B'B)^{-1}$, we can apply Fact 8.9.19 in Bernstein (2005) to conclude that $\Omega^* (\delta)$ is positive definite uniformly in $\delta$. Similarly, we can readily show that

$$\sup_{\delta \in \Delta} \lambda_{\max} (\Omega^* (\delta)) \leq \sup_{\delta \in \Delta} \lambda_{\max} (\bar{\Omega}^* (\delta)) + \sup_{\delta \in \Delta} (\phi_\nu I_n + b_m (B'B)^{-1})) \leq \sup_{\delta \in \Delta} \lambda_{\max} (\Omega^* (\delta)) + \sup_{\delta \in \Delta} (\phi_\nu + b_m \lambda_{\min} ((B'B)^{-1})) \leq c^*, \text{ say.}$$

Next, write

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

Noting that $A_1 (\rho)$ is a block triangular matrix. Its eigenvalues are given by those of $\frac{1}{nT_1} \begin{pmatrix} X' X & X' Z \\ Z' X & Z' Z \end{pmatrix}$ and those of $\frac{1}{nT_1} \tilde{x}' \tilde{x}$. By Assumption R* (iv), the minimum of these eigenvalues are bounded away from 0, say by $\omega_{\text{x}}$, uniformly in $\rho$. Similarly, the minimum eigenvalues of $A_2 (\rho) = 0$ uniformly in $\rho$. It follows

42
that \( \inf_\rho \lambda_{\min}\left(\frac{1}{nT_1}X'\rho X^* \right) \geq \inf_\rho [\lambda_{\min}(A_1(\rho)) + \lambda_{\min}(A_2(\rho))] \geq \Delta > 0 \). Consequently,

\[
\inf_{\delta \in \Delta} \lambda_{\min}(Q_{xx}^*(\delta)) = \inf_{\delta \in \Delta} \lambda_{\min}\left(\frac{1}{nT_1}X'\rho^{-1}X^* \right) \geq c^{r-1} \inf_\rho \lambda_{\min}\left(\frac{1}{nT_1}X'\rho X^* \right) \geq c^{r-1} \Delta > 0. \tag{C.14}
\]

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

\[
\|\Pi_{n2}^*(\delta)\| \leq \left[ \inf_{\delta \in \Delta} \lambda_{\min}(Q_{xx}^*(\delta)) \right]^{-1} \|Q_{xx}^*(\delta)\|^2 = o_P(1) \text{ uniformly in } \delta.
\]

For \( \Pi_{n3}^*(\delta) \), we have \( \Pi_{n3}^*(\delta) \leq \|Q_{xx}^*(\delta)\| \left\|Q_{xx}^*(\delta)^{-1}\right\| \left\|Q_{xy_{s-1}}^*(\delta)\right\| = o_P(1) \text{ uniformly in } \delta \) as one can readily show that \( Q_{xy_{s-1}}^*(\delta) = O_P(1) \) uniformly in \( \delta \).

For \( \Pi_{n4}^*(\delta) \), we have

\[
\Pi_{n4}^*(\delta) = \left\{ Q_{xy_{s-1}}^*(\delta) - E[Q_{xy_{s-1}}^*(\delta)] \right\}'Q_{xx}^*(\delta)^{-1}Q_{xy_{s-1}}^*(\delta)
+ E[Q_{xy_{s-1}}^*(\delta)]'Q_{xx}^*(\delta)^{-1}\{E[Q_{xx}^*(\delta)] - Q_{xx}^*(\delta)\} \{E[Q_{xx}^*(\delta)]\}^{-1}Q_{xy_{s-1}}^*(\delta)
+ E[Q_{xy_{s-1}}^*(\delta)]'E[Q_{xx}^*](\delta) \left\{ Q_{xy_{s-1}}^*(\delta) - E[Q_{xy_{s-1}}^*(\delta)] \right\}.
\]

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

**Proof of Theorem 4.4**

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

**Proof of Theorem 4.5**

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

**Proof of Theorem 4.6**

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


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

<table>
<thead>
<tr>
<th>ψ</th>
<th>m = 0</th>
<th>m = 6</th>
<th>m = 200</th>
<th>m = 0</th>
<th>m = 6</th>
<th>m = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 50, T = 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>5.0266[0.334]</td>
<td>4.9604[0.338]</td>
<td>5.0030[0.328]</td>
<td>4.5591[0.378]</td>
<td>4.9940[0.411]</td>
<td>5.0988[0.411]</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0011[0.040]</td>
<td>0.9917[0.045]</td>
<td>0.9981[0.045]</td>
<td>0.9626[0.041]</td>
<td>0.9980[0.040]</td>
<td>1.0057[0.039]</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9951[0.345]</td>
<td>0.9852[0.350]</td>
<td>0.9927[0.352]</td>
<td>0.7418[0.365]</td>
<td>0.9384[0.391]</td>
<td>0.9790[0.395]</td>
</tr>
<tr>
<td>0.5</td>
<td>0.7991[0.023]</td>
<td>0.8071[0.024]</td>
<td>0.8018[0.022]</td>
<td>0.8238[0.015]</td>
<td>0.8015[0.017]</td>
<td>0.7963[0.016]</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4827[0.099]</td>
<td>0.3023[0.115]</td>
<td>0.2686[0.114]</td>
<td>0.4732[0.101]</td>
<td>0.4886[0.098]</td>
<td>0.4868[0.098]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9681[0.147]</td>
<td>0.1469[0.116]</td>
<td>0.0214[0.055]</td>
<td>0.8648[0.145]</td>
<td>0.9528[0.158]</td>
<td>0.9280[0.161]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9834[0.072]</td>
<td>1.2563[0.087]</td>
<td>1.2805[0.088]</td>
<td>1.0056[0.076]</td>
<td>0.9880[0.073]</td>
<td>1.0019[0.076]</td>
</tr>
<tr>
<td>n = 100, T = 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>4.9785[0.357]</td>
<td>4.9683[0.400]</td>
<td>4.9719[0.400]</td>
<td>4.7922[0.353]</td>
<td>5.0164[0.352]</td>
<td>5.0162[0.352]</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0003[0.040]</td>
<td>0.9964[0.045]</td>
<td>0.9967[0.045]</td>
<td>0.9780[0.041]</td>
<td>0.9981[0.039]</td>
<td>0.9981[0.039]</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9937[0.324]</td>
<td>1.0022[0.328]</td>
<td>1.0028[0.328]</td>
<td>0.8910[0.352]</td>
<td>0.9374[0.360]</td>
<td>0.9376[0.361]</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4015[0.034]</td>
<td>0.4025[0.044]</td>
<td>0.4019[0.044]</td>
<td>0.4271[0.032]</td>
<td>0.4090[0.030]</td>
<td>0.4090[0.030]</td>
</tr>
<tr>
<td>0.5</td>
<td>0.7991[0.103]</td>
<td>0.3694[0.141]</td>
<td>0.3690[0.142]</td>
<td>0.4765[0.104]</td>
<td>0.4912[0.093]</td>
<td>0.4912[0.093]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9690[0.146]</td>
<td>0.6380[0.229]</td>
<td>0.6364[0.231]</td>
<td>0.9141[0.155]</td>
<td>0.9725[0.148]</td>
<td>0.9712[0.149]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9838[0.074]</td>
<td>1.1272[0.137]</td>
<td>1.1280[0.138]</td>
<td>1.0056[0.080]</td>
<td>0.9960[0.074]</td>
<td>0.9964[0.074]</td>
</tr>
</tbody>
</table>

Note: \( \psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \sigma_\mu, \sigma_\nu)' \). Parameters values for generating \( x_t: \theta_x = (0.01, 0.5, 0.5, 2.1) \) (see Footnote 7).
Table 1b. Monte Carlo Mean [RMSE] for the QMLEs, Random Effects Model with Normal Mixture

<table>
<thead>
<tr>
<th>ψ</th>
<th>true m = 0</th>
<th>m = 6</th>
<th>m = 200</th>
<th>true m = 6</th>
<th>m = 6</th>
<th>m = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 50, T = 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>5.0194[0.342]</td>
<td>4.9734[0.356]</td>
<td>5.0140[0.340]</td>
<td>4.5754[0.416]</td>
<td>4.9935[0.429]</td>
<td>5.0941[0.430]</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0005[0.039]</td>
<td>0.9948[0.047]</td>
<td>1.0006[0.047]</td>
<td>0.9656[0.041]</td>
<td>0.9984[0.039]</td>
<td>1.0057[0.039]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9874[0.335]</td>
<td>0.9778[0.339]</td>
<td>0.9858[0.340]</td>
<td>0.7650[0.383]</td>
<td>0.9558[0.405]</td>
<td>0.9981[0.410]</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7992[0.022]</td>
<td>0.8047[0.024]</td>
<td>0.7998[0.022]</td>
<td>0.8225[0.017]</td>
<td>0.8011[0.016]</td>
<td>0.7960[0.016]</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4788[0.100]</td>
<td>0.2652[0.130]</td>
<td>0.2489[0.129]</td>
<td>0.4766[0.099]</td>
<td>0.4916[0.097]</td>
<td>0.4902[0.096]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9544[0.249]</td>
<td>0.1551[0.120]</td>
<td>0.0283[0.061]</td>
<td>0.8470[0.228]</td>
<td>0.9330[0.259]</td>
<td>0.9101[0.260]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9792[0.145]</td>
<td>1.2519[0.163]</td>
<td>1.2776[0.167]</td>
<td>0.9984[0.147]</td>
<td>0.9821[0.143]</td>
<td>0.9954[0.147]</td>
</tr>
<tr>
<td>n = 100, T = 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>5.0179[0.343]</td>
<td>5.0602[0.344]</td>
<td>5.0602[0.344]</td>
<td>4.9083[0.343]</td>
<td>5.0085[0.339]</td>
<td>5.0085[0.339]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9990[0.044]</td>
<td>1.0016[0.044]</td>
<td>1.0016[0.044]</td>
<td>0.9884[0.040]</td>
<td>1.0000[0.038]</td>
<td>1.0000[0.038]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9981[0.333]</td>
<td>1.0061[0.333]</td>
<td>1.0067[0.333]</td>
<td>0.9917[0.336]</td>
<td>0.9928[0.339]</td>
<td>0.9928[0.339]</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.0009[0.043]</td>
<td>-0.0094[0.043]</td>
<td>-0.0094[0.043]</td>
<td>0.0917[0.045]</td>
<td>-0.0017[0.042]</td>
<td>-0.0017[0.042]</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4822[0.097]</td>
<td>0.4484[0.096]</td>
<td>0.4484[0.096]</td>
<td>0.4808[0.100]</td>
<td>0.4926[0.089]</td>
<td>0.4926[0.089]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9469[0.259]</td>
<td>0.8501[0.259]</td>
<td>0.8501[0.259]</td>
<td>0.9081[0.247]</td>
<td>0.9435[0.246]</td>
<td>0.9434[0.246]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9784[0.144]</td>
<td>1.1470[0.183]</td>
<td>1.1485[0.184]</td>
<td>0.9895[0.148]</td>
<td>0.9770[0.138]</td>
<td>0.9774[0.138]</td>
</tr>
</tbody>
</table>

Note: ψ = (γ0, β, γ1, ρ, λ, σμ, σε)′. Parameters values for generating x_t: θ_p = (.01, .5, .5, 2, 1) (see Footnote 7).
<table>
<thead>
<tr>
<th>n</th>
<th>ψ</th>
<th>Mean</th>
<th>SD</th>
<th>seSCb</th>
<th>seHS</th>
<th>seHSb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T + 1 = 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5.0</td>
<td>5.0155 0.3595 0.3257 0.3428 0.3759</td>
<td>5.0040 0.2736 0.2436 0.2695 0.3149</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0003 0.0422 0.0373 0.0403 0.0443</td>
<td>0.9999 0.0229 0.0203 0.0222 0.0246</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9949 0.3462 0.3321 0.3291 0.3288</td>
<td>0.9996 0.3017 0.2981 0.2978 0.2988</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.4987 0.0332 0.0312 0.0321 0.0342</td>
<td>0.4995 0.0150 0.0140 0.0149 0.0162</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.4775 0.1035 0.1037 0.1003 0.1104</td>
<td>0.4973 0.0608 0.0632 0.0588 0.0631</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9998 0.3622 0.3885 0.3543 0.3692</td>
<td>0.9734 0.2657 0.2727 0.2543 0.2621</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9775 0.1441 0.1416 0.1455 0.1686</td>
<td>0.9883 0.0822 0.0821 0.0837 0.0981</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>5.0</td>
<td>5.0021 0.2634 0.2421 0.2571 0.2797</td>
<td>5.0014 0.1860 0.1591 0.1806 0.2145</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0000 0.0287 0.0270 0.0285 0.0305</td>
<td>1.0000 0.0155 0.0148 0.0160 0.0175</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9949 0.2412 0.2360 0.2350 0.2351</td>
<td>1.0109 0.2168 0.2141 0.2161 0.2190</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.5000 0.0223 0.0211 0.0216 0.0226</td>
<td>0.4999 0.0105 0.0098 0.0105 0.0113</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.4896 0.0726 0.0750 0.0715 0.0766</td>
<td>0.4976 0.0398 0.0466 0.0425 0.0444</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0040 0.2540 0.2636 0.2495 0.2589</td>
<td>0.9866 0.1889 0.1871 0.1815 0.1885</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9899 0.1027 0.0964 0.1038 0.1227</td>
<td>0.9966 0.0602 0.0560 0.0596 0.0710</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>ψ</th>
<th>Mean</th>
<th>SD</th>
<th>seSCb</th>
<th>seHS</th>
<th>seHSb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T + 1 = 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5.0</td>
<td>5.0105 0.3450 0.3340 0.3389 0.3735</td>
<td>4.9986 0.2828 0.2555 0.2685 0.3100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0005 0.0394 0.0368 0.0398 0.0441</td>
<td>1.0001 0.0208 0.0190 0.0205 0.0224</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9972 0.3300 0.3244 0.3215 0.3220</td>
<td>1.0029 0.3045 0.2977 0.2945 0.2928</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.4997 0.0331 0.0308 0.0316 0.0345</td>
<td>0.4998 0.0159 0.0143 0.0149 0.0161</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.4887 0.1011 0.0984 0.0985 0.1178</td>
<td>0.4928 0.0575 0.0584 0.0590 0.0719</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0376 0.6779 0.3104 0.3636 0.5621</td>
<td>1.0135 0.5932 0.1917 0.2625 0.4643</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9813 0.2916 0.0897 0.1464 0.2867</td>
<td>0.9964 0.1770 0.0413 0.0844 0.1923</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>5.0</td>
<td>5.0098 0.2541 0.2313 0.2420 0.2676</td>
<td>4.9899 0.1900 0.1671 0.1842 0.2175</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0002 0.0293 0.0272 0.0290 0.0316</td>
<td>0.9997 0.0154 0.0139 0.0151 0.0164</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9842 0.2397 0.2344 0.2310 0.2290</td>
<td>1.0070 0.2189 0.2115 0.2151 0.2197</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.5004 0.0240 0.0208 0.0218 0.0236</td>
<td>0.5002 0.0106 0.0101 0.0106 0.0114</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.4900 0.0696 0.0730 0.0713 0.0834</td>
<td>0.4972 0.0421 0.0440 0.0425 0.0502</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0239 0.4462 0.1898 0.2532 0.4188</td>
<td>1.0078 0.3683 0.1162 0.1850 0.3578</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9927 0.2081 0.0569 0.1042 0.2177</td>
<td>0.9901 0.1289 0.0265 0.0592 0.1416</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: ψ = (γ₀, β, γ₁, ρ, λ, φ, σₚ²). Parameters values for generating x₁: θᵣ = (0.1, 5, 5, 2, 1) (see Footnote 7).
<table>
<thead>
<tr>
<th>( n )</th>
<th>( \psi )</th>
<th>( \text{Mean} )</th>
<th>( \text{SD} )</th>
<th>( \text{seSCb} )</th>
<th>( \text{seHS} )</th>
<th>( \text{seHSb} )</th>
<th>( T + 1 = 4 )</th>
<th>( \text{Mean} )</th>
<th>( \text{SD} )</th>
<th>( \text{seSCb} )</th>
<th>( \text{seHS} )</th>
<th>( \text{seHSb} )</th>
<th>( T + 1 = 8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5.0</td>
<td>5.0006</td>
<td>0.3692</td>
<td>0.3683</td>
<td>0.3677</td>
<td>0.3947</td>
<td>5.0104</td>
<td>0.2857</td>
<td>0.2931</td>
<td>0.2770</td>
<td>0.3033</td>
<td>0.4991</td>
<td>0.0155</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9999</td>
<td>0.0371</td>
<td>0.0364</td>
<td>0.0378</td>
<td>0.0408</td>
<td>1.0014</td>
<td>0.0247</td>
<td>0.0253</td>
<td>0.0251</td>
<td>0.0264</td>
<td>0.9489</td>
<td>0.3510</td>
<td>0.3637</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5014</td>
<td>0.0275</td>
<td>0.0289</td>
<td>0.0277</td>
<td>0.0281</td>
<td>0.4990</td>
<td>0.0151</td>
<td>0.0206</td>
<td>0.0153</td>
<td>0.0121</td>
<td>0.4972</td>
<td>0.0907</td>
<td>0.0953</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9905</td>
<td>0.3505</td>
<td>0.3737</td>
<td>0.3424</td>
<td>0.3635</td>
<td>0.9678</td>
<td>0.2583</td>
<td>0.2832</td>
<td>0.2534</td>
<td>0.2584</td>
<td>0.9805</td>
<td>0.1439</td>
<td>0.1381</td>
</tr>
<tr>
<td>100</td>
<td>5.0</td>
<td>5.0276</td>
<td>0.2902</td>
<td>0.2687</td>
<td>0.2739</td>
<td>0.2910</td>
<td>5.0036</td>
<td>0.2046</td>
<td>0.2037</td>
<td>0.1966</td>
<td>0.2126</td>
<td>1.0017</td>
<td>0.0297</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0203</td>
<td>0.2406</td>
<td>0.2402</td>
<td>0.2351</td>
<td>0.2331</td>
<td>0.9996</td>
<td>0.2197</td>
<td>0.2158</td>
<td>0.2128</td>
<td>0.2130</td>
<td>0.4973</td>
<td>0.0212</td>
<td>0.0209</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4989</td>
<td>0.0691</td>
<td>0.0714</td>
<td>0.0676</td>
<td>0.0718</td>
<td>0.4966</td>
<td>0.0412</td>
<td>0.0451</td>
<td>0.0414</td>
<td>0.0436</td>
<td>1.0103</td>
<td>0.2634</td>
<td>0.2666</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9879</td>
<td>0.1020</td>
<td>0.0946</td>
<td>0.1015</td>
<td>0.1203</td>
<td>0.9948</td>
<td>0.0579</td>
<td>0.0559</td>
<td>0.0594</td>
<td>0.0710</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

- Normal Errors
- Normal Mixture Errors
- Chi-Square Errors, df=5

Note: \( \psi = (\gamma_0, \beta, \gamma_1, \phi, \lambda, \varphi, \sigma^2_\varphi) \). Parameters values for generating \( \mathbf{x}_t \): \( \theta_x = (0.1, 5, 5, 2, 1) \) (see Footnote 7).
<table>
<thead>
<tr>
<th>( n )</th>
<th>( \psi )</th>
<th>( T+1=4 )</th>
<th>( T+1=8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal Errors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5.0</td>
<td>Mean</td>
<td>5.0283</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.2098</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seSCb</td>
<td>0.0250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seHS</td>
<td>0.2059</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seHSb</td>
<td>0.2244</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.0122</td>
<td>0.3683</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3803</td>
<td>0.0250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3677</td>
<td>0.0230</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4082</td>
<td>0.0250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9898</td>
<td>0.0250</td>
</tr>
<tr>
<td>100</td>
<td>5.0</td>
<td>Mean</td>
<td>5.0121</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.0316</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seSCb</td>
<td>0.0230</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seHS</td>
<td>0.0177</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seHSb</td>
<td>0.0176</td>
</tr>
<tr>
<td>Normal Mixture Errors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5.0</td>
<td>Mean</td>
<td>0.5012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.0412</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seSCb</td>
<td>0.0385</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seHS</td>
<td>0.0316</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seHSb</td>
<td>0.0316</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0001</td>
<td>0.0230</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9853</td>
<td>0.0230</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9895</td>
<td>0.0177</td>
</tr>
<tr>
<td>100</td>
<td>5.0</td>
<td>Mean</td>
<td>0.5018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.0208</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seSCb</td>
<td>0.0208</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seHS</td>
<td>0.0316</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seHSb</td>
<td>0.0316</td>
</tr>
<tr>
<td>Chi-Square Errors, df=5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5.0</td>
<td>Mean</td>
<td>5.0403</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.0040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seSCb</td>
<td>0.0386</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seHS</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seHSb</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0009</td>
<td>0.0423</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9893</td>
<td>0.0162</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4933</td>
<td>0.0586</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5003</td>
<td>0.0408</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0146</td>
<td>0.4149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9863</td>
<td>0.0162</td>
</tr>
</tbody>
</table>

**Note:** \( \psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \phi_0, \sigma^2_f) \). Parameters values for generating \( x_t \): \( \theta_x = (0.1, 5, 5, 2, 1) \) (see Footnote 7).
### Table 3a. Monte Carlo Mean [RMSE] for the QMLEs, Fixed Effects Model, Normal Errors

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$m = 0$</th>
<th>$m = 6$</th>
<th>$m = 200$</th>
<th>$m = 0$</th>
<th>$m = 6$</th>
<th>$m = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 50, T = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9957[.090]</td>
<td>0.9702[.088]</td>
<td>0.9589[.087]</td>
<td>1.0006[.127]</td>
<td>0.9983[.126]</td>
<td>0.9891[.125]</td>
</tr>
<tr>
<td>-0.9</td>
<td>-0.8960[.045]</td>
<td>-0.8390[.038]</td>
<td>-0.8139[.029]</td>
<td>-0.8976[.037]</td>
<td>-0.8934[.034]</td>
<td>-0.8744[.026]</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4764[.105]</td>
<td>0.4471[.100]</td>
<td>0.4584[.100]</td>
<td>0.4912[.104]</td>
<td>0.4889[.088]</td>
<td>0.4837[.088]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9775[.141]</td>
<td>0.8568[.113]</td>
<td>0.8747[.116]</td>
<td>0.9934[.132]</td>
<td>0.9632[.131]</td>
<td>0.9521[.131]</td>
</tr>
<tr>
<td>$n = 100, T = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9991[.090]</td>
<td>0.9990[.090]</td>
<td>0.9990[.090]</td>
<td>0.9994[.139]</td>
<td>1.0012[.136]</td>
<td>1.0012[.136]</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0004[.065]</td>
<td>0.0004[.055]</td>
<td>-0.0004[.055]</td>
<td>0.0280[.103]</td>
<td>-0.0059[.087]</td>
<td>-0.0059[.087]</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4925[.100]</td>
<td>0.4780[.097]</td>
<td>0.4780[.097]</td>
<td>0.5281[.101]</td>
<td>0.4903[.089]</td>
<td>0.4903[.089]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9673[.149]</td>
<td>0.9619[.147]</td>
<td>0.9619[.147]</td>
<td>0.9973[.174]</td>
<td>0.9703[.156]</td>
<td>0.9702[.156]</td>
</tr>
<tr>
<td></td>
<td>1.0035[.089]</td>
<td>1.0037[.089]</td>
<td>1.0037[.089]</td>
<td>0.9977[.133]</td>
<td>0.9976[.133]</td>
<td>0.9976[.133]</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8991[.025]</td>
<td>0.8993[.025]</td>
<td>0.8993[.025]</td>
<td>0.9004[.044]</td>
<td>0.9002[.044]</td>
<td>0.9002[.044]</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4704[.112]</td>
<td>0.4695[.112]</td>
<td>0.4692[.112]</td>
<td>0.4862[.104]</td>
<td>0.4859[.103]</td>
<td>0.4858[.103]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9682[.149]</td>
<td>0.9682[.149]</td>
<td>0.9681[.149]</td>
<td>0.9803[.151]</td>
<td>0.9803[.151]</td>
<td>0.9803[.151]</td>
</tr>
</tbody>
</table>

**Note:** $\psi = (\beta, \rho, \lambda, \sigma_e)'$. Parameters values for generating $x_t$: $\theta_x = (.01, .5, .5, 1, 5)$ (see Footnote 7).
<table>
<thead>
<tr>
<th>$\psi$</th>
<th>true $m=0$</th>
<th>true $m=6$</th>
<th>true $m=200$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m=0$</td>
<td>$m=6$</td>
<td>$m=200$</td>
</tr>
<tr>
<td>$n = 50, T = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0021[0.092]</td>
<td>0.9906[0.091]</td>
<td>0.9826[0.090]</td>
</tr>
<tr>
<td>-0.9</td>
<td>-0.8987[0.041]</td>
<td>-0.8648[0.040]</td>
<td>-0.8416[0.033]</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4862[103]</td>
<td>0.4035[0.098]</td>
<td>0.4113[0.097]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9822[300]</td>
<td>0.9147[252]</td>
<td>0.9238[262]</td>
</tr>
<tr>
<td>$n = 100, T = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0026[0.091]</td>
<td>1.0013[0.091]</td>
<td>1.0013[0.091]</td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.5009[0.050]</td>
<td>-0.4969[0.049]</td>
<td>-0.4969[0.049]</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4894[103]</td>
<td>0.4415[103]</td>
<td>0.4415[103]</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9862[285]</td>
<td>0.9687[278]</td>
<td>0.9687[278]</td>
</tr>
</tbody>
</table>

Note: $\psi = (\beta, \rho, \lambda, \sigma_v)'$. Parameters values for generating $x_t$: $\theta_x = (0.01, 0.5, 0.5, 1, 0.5)$ (see Footnote 7).
### Table 4a. Monte Carlo Mean and SD, and Bootstrap Standard Errors, $m = 0$

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

Note: $\psi = (\beta, \rho, \lambda, \sigma^2_v)^T$. Parameters values for generating $x_t$: $\theta = (1.5, 5, 5, 5, 1)$ (see Footnote 7).
<table>
<thead>
<tr>
<th>$n$</th>
<th>$\psi$</th>
<th>Mean</th>
<th>SD</th>
<th>seSCb</th>
<th>seHS</th>
<th>seHSb</th>
<th>Mean</th>
<th>SD</th>
<th>seSCb</th>
<th>seHS</th>
<th>seHSb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$T = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$T = 7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.0</td>
<td>1.0000 0.0182 0.0189 0.0184 0.0183</td>
<td>1.0004 0.0095 0.0098 0.0096 0.0117</td>
<td>50</td>
<td>1.0</td>
<td>0.5010 0.0198 0.0188 0.0190 0.0229</td>
<td>0.5001 0.0070 0.0073 0.0070 0.0089</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5000 0.1037 0.0999 0.1016 0.1058</td>
<td>0.4956 0.0603 0.0565 0.0594 0.0633</td>
<td></td>
<td></td>
<td>0.9744 0.1450 0.1602 0.1427 0.1358</td>
<td>0.9914 0.0814 0.0907 0.0836 0.0809</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>0.9998 0.0150 0.0151 0.0149 0.0148</td>
<td>0.9999 0.0064 0.0068 0.0066 0.0075</td>
<td>0.4992 0.0108 0.0117 0.0112 0.0121</td>
<td>0.5000 0.0052 0.0051 0.0051 0.0060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4954 0.0701 0.0735 0.0728 0.0730</td>
<td>0.4991 0.0433 0.0418 0.0425 0.0437</td>
<td>0.9805 0.1040 0.1082 0.1013 0.0990</td>
<td>0.9916 0.0638 0.0619 0.0591 0.0581</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal Mixture Errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.0</td>
<td>1.0004 0.0186 0.0187 0.0180 0.0179</td>
<td>0.9996 0.0093 0.0098 0.0095 0.0117</td>
<td>0.4999 0.0196 0.0185 0.0187 0.0235</td>
<td>0.4999 0.0067 0.0073 0.0069 0.0089</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4993 0.1029 0.0978 0.1019 0.1090</td>
<td>0.4977 0.0572 0.0537 0.0592 0.0662</td>
<td>0.9558 0.2840 0.0986 0.1400 0.2405</td>
<td>0.9857 0.1572 0.0471 0.0832 0.1677</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>0.9993 0.0156 0.0151 0.0149 0.0149</td>
<td>1.0000 0.0067 0.0067 0.0066 0.0074</td>
<td>0.4997 0.0119 0.0117 0.0112 0.0128</td>
<td>0.4998 0.0049 0.0051 0.0051 0.0060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4948 0.0719 0.0726 0.0729 0.0741</td>
<td>0.4976 0.0438 0.0407 0.0426 0.0451</td>
<td>0.9906 0.2015 0.0647 0.1024 0.1908</td>
<td>0.9897 0.1301 0.0371 0.0590 0.1243</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal Mixture Errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.0</td>
<td>0.9991 0.0187 0.0189 0.0183 0.0182</td>
<td>1.0001 0.0100 0.0099 0.0096 0.0118</td>
<td>0.4994 0.0195 0.0186 0.0189 0.0232</td>
<td>0.4998 0.0072 0.0074 0.0070 0.0089</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4958 0.0998 0.0997 0.1022 0.1071</td>
<td>0.4981 0.0569 0.0552 0.0593 0.0646</td>
<td>0.9691 0.2161 0.1221 0.1418 0.1884</td>
<td>0.9995 0.1353 0.0615 0.0844 0.1269</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>1.0007 0.0146 0.0151 0.0149 0.0148</td>
<td>1.0000 0.0067 0.0068 0.0066 0.0075</td>
<td>0.4999 0.0115 0.0117 0.0112 0.0124</td>
<td>0.4998 0.0049 0.0051 0.0051 0.0060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4919 0.0704 0.0734 0.0732 0.0740</td>
<td>0.4977 0.0425 0.0414 0.0426 0.0443</td>
<td>0.9811 0.1476 0.0803 0.1014 0.1418</td>
<td>0.9959 0.0955 0.0415 0.0594 0.0912</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $\psi = (\beta, \rho, \lambda, \sigma^2_v)'$. Parameters values for generating $x_t$: $\theta_x = (1.1, 5, 5, 5, 1)$ (see Footnote 7)
### Table 4c. Monte Carlo Mean and SD, and Bootstrap Standard Errors, \(m = 200\)

<table>
<thead>
<tr>
<th>(n)</th>
<th>(\psi)</th>
<th>(T = 3)</th>
<th>(T = 7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>Mean</td>
<td>SD</td>
<td>seSCb</td>
</tr>
<tr>
<td>50</td>
<td>1.0</td>
<td>1.0004</td>
<td>0.0210</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4999</td>
<td>0.0197</td>
<td>0.0199</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4866</td>
<td>0.0974</td>
<td>0.1011</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9624</td>
<td>0.1422</td>
<td>0.1573</td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>1.0001</td>
<td>0.0139</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5001</td>
<td>0.0117</td>
<td>0.0116</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4977</td>
<td>0.0736</td>
<td>0.0726</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9886</td>
<td>0.1064</td>
<td>0.1091</td>
</tr>
</tbody>
</table>

**Normal Errors**

<table>
<thead>
<tr>
<th>(n)</th>
<th>(\psi)</th>
<th>Mean</th>
<th>SD</th>
<th>seSCb</th>
<th>seHS</th>
<th>seHSb</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>1.0</td>
<td>0.4999</td>
<td>0.0197</td>
<td>0.0199</td>
<td>0.0197</td>
<td>0.0231</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4866</td>
<td>0.0974</td>
<td>0.1011</td>
<td>0.1009</td>
<td>0.1027</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9624</td>
<td>0.1422</td>
<td>0.1573</td>
<td>0.1406</td>
<td>0.1349</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>1.0001</td>
<td>0.0139</td>
<td>0.0140</td>
<td>0.0138</td>
<td>0.0154</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5001</td>
<td>0.0117</td>
<td>0.0116</td>
<td>0.0116</td>
<td>0.0144</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.4977</td>
<td>0.0736</td>
<td>0.0726</td>
<td>0.0745</td>
<td>0.0775</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.9886</td>
<td>0.1064</td>
<td>0.1091</td>
<td>0.1019</td>
<td>0.0993</td>
<td></td>
</tr>
</tbody>
</table>

**Normal Mixture Errors**

<table>
<thead>
<tr>
<th>(n)</th>
<th>Mean</th>
<th>SD</th>
<th>seSCb</th>
<th>seHS</th>
<th>seHSb</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.4999</td>
<td>0.0197</td>
<td>0.0199</td>
<td>0.0197</td>
<td>0.0231</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4866</td>
<td>0.0974</td>
<td>0.1011</td>
<td>0.1009</td>
<td>0.1027</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9624</td>
<td>0.1422</td>
<td>0.1573</td>
<td>0.1406</td>
<td>0.1349</td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>1.0001</td>
<td>0.0139</td>
<td>0.0140</td>
<td>0.0138</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5001</td>
<td>0.0117</td>
<td>0.0116</td>
<td>0.0116</td>
<td>0.0144</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4977</td>
<td>0.0736</td>
<td>0.0726</td>
<td>0.0745</td>
<td>0.0775</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9886</td>
<td>0.1064</td>
<td>0.1091</td>
<td>0.1019</td>
<td>0.0993</td>
</tr>
</tbody>
</table>

**Chi-Square, df=3**

<table>
<thead>
<tr>
<th>(n)</th>
<th>Mean</th>
<th>SD</th>
<th>seSCb</th>
<th>seHS</th>
<th>seHSb</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.4999</td>
<td>0.0197</td>
<td>0.0199</td>
<td>0.0197</td>
<td>0.0231</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4866</td>
<td>0.0974</td>
<td>0.1011</td>
<td>0.1009</td>
<td>0.1027</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9624</td>
<td>0.1422</td>
<td>0.1573</td>
<td>0.1406</td>
<td>0.1349</td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>1.0001</td>
<td>0.0139</td>
<td>0.0140</td>
<td>0.0138</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5001</td>
<td>0.0117</td>
<td>0.0116</td>
<td>0.0116</td>
<td>0.0144</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4977</td>
<td>0.0736</td>
<td>0.0726</td>
<td>0.0745</td>
<td>0.0775</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9886</td>
<td>0.1064</td>
<td>0.1091</td>
<td>0.1019</td>
<td>0.0993</td>
</tr>
</tbody>
</table>

**Note:** \(\psi = (\beta, \rho, \lambda, \sigma^2_v)\). Parameters values for generating \(x_t\): \(\theta_x = (1.0, 5.5, 5.5, 1)\) (see Footnote 7)