SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Asymptotics and Bootstrap for Transformed Panel Data Regressions

Liangjun Su and Zhenlin Yang

December 2008

Paper No. 03-2009

Asymptotics and Bootstrap for Transformed Panel Data Regressions[1]

Liangjun Su and Zhenlin Yang

School of Economics, Singapore Management University

email: ljsu@smu.edu.sg, zlyang@smu.edu.sg

December, 2008

Abstract

This paper investigates the asymptotic properties of quasi-maximum likelihood estimators for transformed random effects models where both the response and (some of) the covariates are subject to transformations for inducing normality, flexible functional form, homoscedasticity, and simple model structure. We develop a quasi maximum likelihood-type procedure for model estimation and inference. We prove the consistency and asymptotic normality of the parameter estimates, and propose a simple bootstrap procedure that leads to a robust estimate of the variance-covariance matrix. Monte Carlo results reveal that these estimates perform well in finite samples, and that the gains by using bootstrap procedure for inference can be enormous.

Key Words: Asymptotics; Bootstrap; Quasi-MLE; Transformed panels; Variance-covariance matrix estimate.

JEL Classification: C23, C15, C51

1 Introduction.

Panel data regression models with error components have been extensively treated in the literature, and almost all the standard econometrics text books on panel data models cover those topics (see, among the others, Baltagi, 2001; Arellano, 2003; Hsiao, 2003; Frees, 2004). However, the literature on transformed panel data regression models is rather sparse, and many issues of immediate theoretical and practical relevance, such as the properties of

[1]We would like to thank Peter Phillips and Myoung-Jae Lee for their helpful comments that have led to improvements in the paper. Thanks are also due to the seminar participants at the Department of Economics, Korea University (2007), and the conference participants at the 14th International Conference on Panel Data (2007), WISE, Xiamen University, China, for their comments and discussions. Zhenlin Yang gratefully acknowledges the support from a research grant (Grant number: C244/MSS7E005) from the Office of Research, Singapore Management University.

parameter estimates in terms of consistency, asypmtotic normality and robustness against heavy-tailed distributions; variance-covariance matrix estimation in the situations where transformation can only bring the data to near-normality, etc., have not been formally studied.[2] This is in a great contrast to the literature on transformed cross-sectional data, where almost all the standard econometrics text book cover this topic (e.g., Davidson and MacKinnon, 1993; Greene, 2000), and some of the popular commercial software, such as SAS and Matlab, have implemented the normal-transformation technique.

It is well known that the purposes of transforming the economic data are to induce (i) normality, (ii) flexible functional form, (iii) homoscedastic errors, and (iv) simple model structure. However, it is generally acknowledged that with a single transformation, it is difficult to reach all the four goals simultaneously, in particular, the normality. Nevertheless, it is still reasonable to believe that a normalizing transformation should be able to bring the data closer to being normally distributed (see, e.g., Hinkley (1975), Hernadze and Johnson (1980), Yeo and Johnson (2000), Yang and Tsui (2004), and Yang and Tse (2007)). Thus, in the framework of quasi-maximum likelihood estimation (QMLE) where one needs to choose a likelihood to approximate the true but unknown one, the normalizing transformation makes it more valid to use the popular Gaussian likelihood for model estimation.

In this paper, we concentrate on the transformed two-way random effects model,

$$\begin{aligned} h(Y_{it}, \lambda) &= \sum_{j=1}^{k_1} \beta_j X_{itj} + \sum_{j=k_1+1}^{k} \beta_j h(X_{itj}, \lambda) + u_{it}, \\ u_{it} &= \mu_i + \eta_t + v_{it}, \;\; i = 1, 2, \cdots, N, \;\; t = 1, 2, \cdots, T \end{aligned} \tag{1}$$

where $h(\cdot, \lambda)$ is a monotonic transformation (e.g., Box and Cox, 1964), known except the indexing parameter λ, called the *transformation parameter*, $X_{itj}, j = 1, \cdots, k_1$, are the exogenous variables containing a column of ones, dummy variables, etc., that do not need to be transformed, $X_{itj}, j = k_1 + 1, \cdots, k$, are the exogenous variables that need to be transformed, and $\{\mu_i\}, \{\eta_t\}$ and $\{v_{it}\}$ are error components assumed to be independent of each other with $\{\mu_i\}$ being independent and identically distributed (i.i.d.) of mean zero and variance σ_η^2, representing the time-invariant and individual-specific effects; $\{\eta_t\}$ i.i.d. of mean zero and variance σ_η^2, representing the individual-invariant and time-specific effects; and $\{v_{it}\}$ i.i.d. of mean zero and variance σ_v^2, representing the pure random errors. In

[2]Baltagi (1997) gives LM tests for linear and log-linear error components regression against Box-Cox alternatives. Abrevaya (1999) proposes a nonparametric estimation of a fixed-effects model with unknown transformation of the dependent variable. Giannakas *et al.* (2003) considers the choice of functional form in stochastic frontier model using panel data. Yang and Huang (2004) consider the maximum likelihood estimation of a transformed random effects model and proposed a simple computational devise for handling of large panels.

the following we will assume that $X_{itj}, j = 1, \cdots, k$, are all non-random regressors. Our analysis will hold with probability one if some elements of X_{itj} are generated randomly, and in this case, we can interpret our analysis as being conditional on $\{X_{itj}, j = 1, \cdots, k, i = 1, 2, \cdots, N,\ t = 1, 2, \cdots, T\}$.

Model (1) gives a useful extension of the standard random effects model by allowing the distribution of Y_{it} to be in a broad family (transformed normal family) not just normal or lognormal. It also allows easy testing of the traditional economic theories of lognormality for production function, firm-size distribution, income distribution, etc., as governed by the Cobb-Douglas production function and Gibrat's Law. Yang and Huang (2004) considered the maximum likelihood estimation (MLE) of Model (1) under Gaussian distributions and provided a simple method for handling the large panel data. Their Monte Carlo simulation results show that the finite sample performance of the MLE-based inference is excellent if the errors are normal or close to normal, but our Monte Carlo results show that it can be quite poor if the errors are fairly non-normal (e.g., there exist gross errors or outliers). Thus, there is a need for an alternative method for the MLE-based inference. Also, to the best of our knowledge there are so far no rigorous large sample theories for Model (1) for either the case of normal errors or the case of non-normal errors. Furthermore, for the cases where the error components follow nonnormal distributions, there are no available methods for estimating the variance-covariance matrix. The reason for the lack of these important results for the transformed two-way random effects panel model is, at least partially, due to the technical complications caused by the nonlinear response transformation and the cross-sectional and time wise dependence induced by the two-way error components, which render the standard large sample techniques not directly applicable.

This paper is organized as follows. Section 2 outlines the quasi-maximum likelihood estimation for the model. Section 3 presents the large sample results concerning the consistency and asymptotic normality of the QMLEs of model parameters, and their rates of convergence under different relative magnitudes of N and T. Section 4 introduces a bootstrap method for estimating the variance-covariance matrix which leads to robust inferences. Section 5 presents some Monte Carlo results concerning the finite sample behavior of the QMLEs and the bootstrap-based inference. Section 6 concludes the paper.

Some generic notation. Throughout the paper we adopt the following notation and conventions. The Euclidean norm of a matrix A is denoted by $\|A\| = [\text{tr}(AA')]^{1/2}$. When A is a square matrix, its smallest and largest eigenvalues are denoted, respectively, by $\gamma_{\min}(A)$ and $\gamma_{\max}(A)$. As usual, convergence in probability is denoted by $\xrightarrow{p}$ and convergence in distribution by $\xrightarrow{D}$. That both N and T approach to infinity concurrently is denoted by $N, T \to \infty$, and that either N or T or both approach to infinity is denoted by $N \cup T \to \infty$.

Partial derivatives of $h(Y_{it},\lambda)$ of various order are denoted by adding subscripts to h, e.g., $h_Y(Y_{it},\lambda)$ is the first-order partial derivative of h w.r.t. Y_{it}, $h_{Y\lambda}(Y_{it},\lambda)$ the partial derivative of h w.r.t. (Y_{it},λ), $h_{\lambda\lambda}(Y_{it},\lambda)$ the second-order partial derivative of h w.r.t. λ, etc.

2 Quasi Maximum Likelihood Estimation

Stacking the data according to $t = 1,\cdots,T$, for each of $i = 1,\cdots,N$, Model (1) can be compactly written in matrix form,

$$h(Y,\lambda) = X(\lambda)\beta + u, \quad \text{with} \quad u = Z_\mu\mu + Z_\eta\eta + v \tag{2}$$

where $Z_\mu = I_N \otimes 1_T$ and $Z_\eta = 1_N \otimes I_T$ with I_N being an $N \times N$ identity matrix, 1_N an N-vector of ones, and $\otimes$ the Kronecker product. Define $J_N = 1_N 1_N'$. The Gaussian log likelihood function after dropping the constant term takes the form

$$\ell(\psi) = -\frac{1}{2}\log|\Sigma| - \frac{1}{2}[h(Y,\lambda) - X(\lambda)\beta]'\Sigma^{-1}[h(Y,\lambda) - X(\lambda)\beta] + J(\lambda), \tag{3}$$

where $\psi = (\beta', \sigma_\mu^2, \sigma_\eta^2, \sigma_v^2, \lambda)'$, and $J(\lambda) = \sum_{i=1}^N \sum_{t=1}^T \log h_Y(Y_{it},\lambda)$ is the log Jacobian of the transformation, and Σ is the variance-covariance matrix of u which takes the form

$$\Sigma = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\eta^2(J_N \otimes I_T) + \sigma_v^2(I_N \otimes I_T).$$

When the error components μ, η and v are exactly normal, (3) gives the exact log likelihood and thus maximizing $\ell(\psi)$ gives the maximum likelihood estimator (MLE) of ψ. However, when one or more of the error components are not exactly normal, the $\ell(\psi)$ function defined by (3) is no longer the true likelihood function. Nevertheless, when $\ell(\psi)$ satisfies certain conditions, maximizing it still gives consistent estimators of model parameters, which are often termed as quasi-maximum likelihood estimator (QMLE). See, for example, White (1994). Furthermore, as pointed out in the introduction, the normalizing transformation makes it more valid to use Gaussian likelihood as an approximation to the true but unknown likelihood.

Yang and Huang (2004) pointed out that direct maximization of $\ell(\psi)$ may be impractical as the dimension of ψ may be high and calculation of $|\Sigma|$ and Σ^{-1} can be difficult if panels are large. They, following Baltagi and Li (1992) and others, considered a spectral decomposition: $\Omega = \frac{1}{\sigma_v^2}\Sigma = Q + \frac{1}{\theta_1}P_1 + \frac{1}{\theta_2}P_2 + \frac{1}{\theta_3}P_3$, where $Q = I_{NT} - \frac{1}{T}I_N \otimes J_T - \frac{1}{N}J_N \otimes I_T + \frac{1}{NT}J_{NT}$, $P_1 = \frac{1}{T}I_N \otimes J_T - \frac{1}{NT}J_{NT}$, $P_2 = \frac{1}{N}J_N \otimes I_T - \frac{1}{NT}J_{NT}$, $P_3 = \frac{1}{NT}J_{NT}$, $\theta_1 = 1/(T\phi_\mu + 1)$, $\theta_2 = 1/(N\phi_\eta + 1)$, and $\theta_3 = 1/(T\phi_\mu + N\phi_\eta + 1)$, $\phi_\mu = \sigma_\mu^2/\sigma_v^2$, and $\phi_\eta = \sigma_\eta^2/\sigma_v^2$. This leads to

$$\Omega^{-1} = Q + \theta_1 P_1 + \theta_2 P_2 + \theta_3 P_3, \quad \text{and} \quad |\Sigma|^{-1} = (\sigma_v^2)^{-NT}\theta_1^{N-1}\theta_2^{T-1}\theta_3. \tag{4}$$

In what follows, we adopt the following parameterization: $\psi = (\beta', \sigma_v^2, \phi')'$ with $\phi = (\phi_\mu, \phi_\eta, \lambda)'$. The log likelihood function in terms of this new parameterization thus becomes

$$\ell(\psi) = c(\phi_\mu, \phi_\eta) - \frac{NT}{2}\log(\sigma_v^2) - \frac{1}{2\sigma_v^2}[h(Y,\lambda) - X(\lambda)\beta]'\Omega^{-1}[h(Y,\lambda) - X(\lambda)\beta] + J(\lambda),$$

where $c(\phi_\mu, \phi_\eta) = \frac{N-1}{2}\log(\theta_1) + \frac{T-1}{2}\log(\theta_2) + \frac{1}{2}\log(\theta_3)$. The expressions θ_1, θ_2, and θ_3 defined above are often used for convenience.

It is easy to see that, for a given ϕ, $\ell(\psi)$ is partially maximized at

$$\hat{\beta}(\phi) = [X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda) \tag{5}$$

$$\hat{\sigma}_v^2(\phi) = \frac{1}{NT}[h(Y,\lambda) - X(\lambda)\hat{\beta}(\phi)]'\Omega^{-1}[h(Y,\lambda) - X(\lambda)\hat{\beta}(\phi)], \tag{6}$$

resulting the concentrated quasi log likelihood for ϕ as

$$\ell_{\max}(\phi) = c(\phi_\mu, \phi_\eta) - \frac{NT}{2}[1 + \log\hat{\sigma}_v^2(\phi)] + J(\lambda), \tag{7}$$

Maximizing $\ell_{\max}(\phi)$ gives the QMLE $\hat{\phi}$ of ϕ, and hence the QMLEs $\hat{\beta}(\hat{\phi})$ and $\hat{\sigma}_v^2(\hat{\phi})$ of β and σ_v^2, respectively. Yang and Huang (2004) further noted that maximization of (7) may still be computationally infeasible when panels become large, i.e., N and T become large, because the process involves repeated calculations of the $NT \times NT$ the matrices Q, P_1, P_2, and P_3. They provided a simple computational device that overcomes this difficulty.

3 Asymptotic Properties of the QMLE

As discussed in the introduction, large sample properties of Model (1) have not been formally considered in the literature when the errors are either normal or non-normal. In this section, we first treat the consistency of the QMLEs of the model parameters, and then the asymptotic normality where the different convergence rates of QMLEs are identified.

Let Λ, Φ and Ψ be, respectively, the parameter space for λ, ϕ and ψ; λ_0, ϕ_0 and ψ_0 be the true parameter values; "E" and "Var" be expectation and variance operators corresponding to the true parameter ψ_0.

3.1 Consistency

Let $\bar{\ell}(\psi)$ be the expected log likelihood, i.e., $\bar{\ell}(\psi) \equiv \mathrm{E}[\ell(\psi)] = -\frac{NT}{2}\log(\sigma_v^2) + c(\phi_\mu, \phi_\eta) - \frac{1}{2\sigma_v^2}\mathrm{E}\left\{[h(Y,\lambda) - X(\lambda)\beta]'\Omega^{-1}[h(Y,\lambda) - X(\lambda)\beta]\right\} + \mathrm{E}[J(\lambda)]$. It is easy to show that, for a given ϕ, $\bar{\ell}(\psi)$ is maximized at

$$\bar{\beta}(\phi) = [X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}\mathrm{E}[h(Y,\lambda)] \tag{8}$$

$$\bar{\sigma}_v^2(\phi) = \frac{1}{NT}\mathrm{E}\left\{[h(Y,\lambda) - X(\lambda)\bar{\beta}(\phi)]'\Omega^{-1}[h(Y,\lambda) - X(\lambda)\bar{\beta}(\phi)]\right\}. \tag{9}$$

Thus, the partially maximized $\bar{\ell}(\psi)$ takes the form

$$\bar{\ell}_{\max}(\phi) = c(\phi_\mu, \phi_\eta) - \frac{NT}{2}[1 + \log \bar{\sigma}_v^2(\phi)] + \mathrm{E}[J(\lambda)]. \tag{10}$$

According to White (1994, Theorem 3.4), the uniform convergence of $\frac{1}{NT}[\ell_{\max}(\phi) - \bar{\ell}_{\max}(\phi)]$ to zero is the focal point for the consistency of the QMLE $\hat{\phi}$. Once the consistency of $\hat{\phi}$ is established, the consistency of $\hat{\beta}(\hat{\phi})$ and $\hat{\sigma}_v^2(\hat{\phi})$ follows immediately, although some standard conditions on the regressors are necessary. We now list a set of sufficient conditions for the consistency of the QMLE.

Assumption C1: *The error components μ, η, and v are independent of each other, and each contains i.i.d. elements with a zero mean and a constant variance denoted by $\sigma_{\mu 0}^2, \sigma_{\eta 0}^2$, and σ_{v0}^2 respectively for μ, η, and v.*

Assumption C2: *Φ is convex and compact. $\phi_\mu = \sigma_\mu^2/\sigma_v^2$ and $\phi_\eta = \sigma_\eta^2/\sigma_v^2$ are bounded away from 0 in Φ.*

Assumption C3: $\lim_{N,T\to\infty} \frac{1}{NT}[X'(\lambda)\Omega^{-1}X(\lambda)]$ *exists and is nonsingular, uniformly in $\phi \in \Phi$.*

Assumption C4: $E[h^2(Y_{it}, \lambda)] < \Delta_1 < \infty$ *and* $E|\log h_Y(Y_{it}, \lambda)| < \Delta_2 < \infty$, *for all $i = 1, \cdots, N$, $t = 1, \cdots, T$, and $\lambda \in \Lambda$.*

Assumption C5: *As $N \cup T \to \infty$,*

(i) $\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T [h^2(Y_{it}, \lambda) - E(h^2(Y_{it}, \lambda))] \xrightarrow{p} 0$, *for each $\lambda \in \Lambda$, and*

(ii) $\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T [\log h_Y(Y_{it}, \lambda) - E(\log h_Y(Y_{it}, \lambda))] \xrightarrow{p} 0$, *for each $\lambda \in \Lambda$.*

Assumption C6: *The partial derivatives $\{h_\lambda(X_{itj}, \lambda), j = k_1 + 1, \cdots, k\}$, $h_\lambda(Y_{it}, \lambda)$ and $h_{Y\lambda}(Y_{it}, \lambda)$ exist such that as $N \cup T \to \infty$,*

(i) $\sup_{\lambda\in\Lambda} \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T h_\lambda^2(X_{itj}, \lambda) = O(1)$ *for $j = k_1 + 1, \cdots, k$,*

(ii) $\sup_{\lambda\in\Lambda} \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T h^2(Y_{it}, \lambda) = O_p(1)$,

(iii) $\sup_{\lambda\in\Lambda} \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T h_\lambda^2(Y_{it}, \lambda) = O_p(1)$, *and*

(iv) $\sup_{\lambda\in\Lambda} \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T h_{Y\lambda}(Y_{it}, \lambda)/h_Y(Y_{it}, \lambda) = O_p(1)$.

Assumptions C1-C2 are standard in quasi maximum likelihood estimation. Assumption C3 guarantees the existence of $\bar{\beta}(\phi)$ uniform in $\phi \in \Phi$. This assumption is weak and can in fact be ensured by a simpler version:

$\lim_{N,T\to\infty} \frac{1}{NT}[X'(\lambda)X(\lambda)]$ *exists and is nonsingular, uniformly in $\lambda \in \Lambda$.*

This can be seen by the following matrix results: (i) for any two real symmetric matrices A and B, $\gamma_{\max}(A+B) \leq \gamma_{\max}(A) + \gamma_{\max}(B)$, and (ii) the largest eigenvalue of a projection matrix is less than or equal to one. We have by (4), $\gamma_{\max}(\Omega^{-1}) \leq [\gamma_{\max}(Q) + \sum_{j=1}^3 \theta_j \gamma_{\max}(P_j)] \leq$

4, where the last inequality follows from the facts that $0 < \theta_j \leq 1$ for $j = 1, 2, 3$, and that Q, P_1, P_2, and P_3 are projection matrices. Thus, the existence of the limit in Assumption C3 follows from

$$X'(\lambda)\Omega^{-1}X(\lambda) \leq \gamma_{\max}(\Omega^{-1})X'(\lambda)X(\lambda) \leq 4X'(\lambda)X(\lambda),$$

and the nonsingularity of the limiting matrix follows from

$$X'(\lambda)\Omega^{-1}X(\lambda) \geq \gamma_{\min}(\Omega^{-1})X'(\lambda)X(\lambda),$$

where $\gamma_{\min}(\Omega^{-1})$ is strictly positive as Ω^{-1} is positive definite. Here, $A \geq B$ means $A - B$ is positive semidefinte for two matrices A and B of the same order, and $A \leq B$ is defined similarly.

Assumption C4 ensures the uniform boundedness of $\bar{\sigma}^2(\phi)$ and $\frac{1}{NT}\mathrm{E}[J(\lambda)]$, and thus the uniform boundedness of $\frac{1}{NT}\bar{\ell}_{\max}(\phi)$. Assumption C5 says that the sequences of random variables $\{h^2(Y_{it}, \lambda)\}$ and $\{\log h_Y(Y_{it}, \lambda)\}$ satisfy a pointwise weak law of large numbers (LLN). Assumption C6 says that these two sequences are well behaved uniformly for λ in the compact set Λ, which are essential for us to apply the weak uniform law of large numbers (ULLN). Alternatively, one can require that the two sequences satisfy certain Lipschitz condition, as specified in, say, Andrews (1987, 1992), Pötscher and Prucha (1989), and Davidson (1994, Chapter 21). The smoothness condition in Assumption C6 is not restrictive as the transformation functions $h(Y_{it}, \lambda)$ applied in practice, such as the Box-Cox power transformation (Box and Cox, 1964), and more recently the power transformations by Yeo and Johnson (2000) and the dual-power transformation by Yang (2006), typically possess continuous partial derivatives in Y_{it} and λ up to any order. We have the following consistency result.

Theorem 1: *Suppose the data generating process is given by Model (1). Assume Assumptions C1-C6 hold. Assume further that (a) $h(Y_{it}, \lambda)$ is monotonic increasing in Y_{it}, and (b) $\bar{\ell}_{\max}(\phi)$ has a unique global maximum at ϕ_0. Then, $\hat{\psi} \xrightarrow{p} \psi_0$, as $N, T \to \infty$.*

The proof is relegated to Appendix. The identification uniqueness condition ($\bar{\ell}_{\max}(\phi)$ has a unique global maximum at ϕ_0) stated in Theorem 1 may be proved directly with some additional minor regularity conditions. Some details on the order of convergence of $\hat{\psi}$ with respect to the relative magnitudes of N and T are given in the next subsection.

3.2 Asymptotic normality

Let $G(\psi) = \partial\ell(\psi)/\partial\psi$ and $H(\psi) = \partial^2\ell(\psi)/(\partial\psi\partial\psi')$ be, respectively, the gradient and the Hessian of the log likelihood function $\ell(\psi)$. Their detailed expressions are given in Appendix. For the asymptotic normality, we need some further assumptions.

Assumption N1: $\mathrm{E}|\mu_i|^{4+\epsilon_1} < \infty$, $\mathrm{E}|\eta_t|^{4+\epsilon_2} < \infty$, *and* $\mathrm{E}|v_{it}|^{4+\epsilon_3} < \infty$, *for some* ϵ_1, ϵ_2 *and* $\epsilon_3 > 0$, *all* $i = 1, \cdots, N$; $t = 1, \cdots, T$

Assumption N2: ψ_0 *is an interior point of* Ψ.

Assumption N3: $\mathrm{E}[G(\psi_0)] = 0$.

Assumption N4: $X(\lambda)$ *and* $h(Y, \lambda)$ *are third order differentiable w.r.t.* λ *such that for* $N_\epsilon(\lambda_0) = \{\lambda \in \Lambda : |\lambda - \lambda_0| \le \epsilon\}$, *and as* $N \cup T \to \infty$,

(i) $\sup_{\lambda \in N_\epsilon(\lambda_0)} \frac{1}{NT} \|X^*(\lambda)\|^2 = O(1)$, *where* $X^*(\lambda) = X(\lambda), X_\lambda(\lambda), X_{\lambda\lambda}(\lambda)$, *or* $X_{\lambda\lambda\lambda}(\lambda)$.

(ii) $\sup_{\lambda \in N_\epsilon(\lambda_0)} \frac{1}{NT} \|h^*(Y, \lambda)\|^2 = O_p(1)$, *where* $h^*(Y, \lambda) = h(Y, \lambda), h_\lambda(Y, \lambda), h_{\lambda\lambda}(Y, \lambda)$, *or* $h_{\lambda\lambda\lambda}(Y, \lambda)$.

(iii) $\sup_{\lambda \in N_\epsilon(\lambda_0)} \frac{1}{NT} |J^*(\lambda)| = O_p(1)$, *where* $J^*(\lambda) = J(\lambda), J_\lambda(\lambda), J_{\lambda\lambda}(\lambda)$, *or* $J_{\lambda\lambda\lambda}(\lambda)$.

Assumption N5: As $N \cup T \to \infty$,

(i) $\frac{1}{NT} X'(\lambda_0)[h_\lambda(Y, \lambda_0) - \mathrm{E}(h_\lambda(Y, \lambda_0))] = o_p(1)$ *and the same is true when* $h_\lambda(Y, \lambda_0)$ *is replaced by* $h_{\lambda\lambda}(Y, \lambda_0)$.

(ii) $\frac{1}{NT}\{h'_\lambda(Y, \lambda_0) h(Y, \lambda_0) - \mathrm{E}[h'_\lambda(Y, \lambda_0) h(Y, \lambda_0)]\} = o_p(1)$, *and the same is true when* $h(Y, \lambda_0)$ *is replaced by* $h_\lambda(Y, \lambda_0)$ *or* $h_\lambda(Y, \lambda_0)$ *is replaced by* $h_{\lambda\lambda}(Y, \lambda_0)$.

(iii) $\frac{1}{NT}\{J_{\lambda\lambda}(\lambda_0) - \mathrm{E}[J_{\lambda\lambda}(\lambda_0)]\} = o_p(1)$.

Assumptions N1-N3 are standard for quasi maximum likelihood inference. Under Assumption C1, the first four components of $G(\psi_0)$ automatically satisfies Assumption N3. For the last component $G_\lambda(\psi_0)$ of $G(\psi_0)$, the requirement $\mathrm{E}[G_\lambda(\psi_0)] = 0$ is tricky, but is likely to be true if the error distributions are symmetric. See Hinkley (1975) and Yang (1999) for some discussions and useful results. This assumption is related to the Assumption (b) stated in Theorem 1. Assumptions N4 and N5 spell out conditions on the transformation function and it partial derivatives to ensure the existence of the information matrix and convergence in probability of various quantities. In particular, Assumptions N4(iii) and N5(iii) set out conditions on the derivatives of the Jacobian term. They are not restrictive as in the special case of Box-Cox power transformation, $J_\lambda(\lambda)$ is free of λ, and $J_{\lambda\lambda}(\lambda) = J_{\lambda\lambda\lambda}(\lambda) = 0$.

One of the key step in proving the asymptotic normality of the QMLE $\hat{\psi}$ is to show that the gradient function $G(\psi_0)$ after being suitably normalized is asymptotic normal. The asymptotic normality of the components of $G(\psi_0)$ corresponding to $\beta, \sigma_v^2, \phi_\mu$ and ϕ_η can be proved using the central limit theorem (CLT) for linear-quadratic forms of error components given in Lemma A3 in Appendix, which adapts the CLT for linear-quadratic forms of i.i.d. errors by Kelejian and Prucha (2001). However, the component of $G(\psi_0)$ corresponding to λ involves the nonlinear function h and its partial derivatives. Moreover, the two-way error components μ and η induce dependence along both the cross-sectional and time-wise directions. These render the standard limiting theorems not applicable and hence some high-level condition needs to be imposed. Define

$$g_{it} \equiv g_{it}(\psi_0) = \frac{h_{Y\lambda}(Y_{it}, \lambda)}{h_Y(Y_{it}, \lambda)} + c_1 u_{\lambda,it} u_{it} + c_2 u_{it}^2 + c_3 u_{it} + c_4 \mu_i^2 + c_5 \eta_t^2 + c_6 \mu_i + c_7 \eta_t + c_8,$$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$, where c_1-c_8 are constants depending on ψ_0, $\{u_{it}\}$ are the elements of $u = h(Y, \lambda_0) - X(\lambda_0)\beta_0$, and $u_{\lambda,it} = \partial u_{it}/\partial \lambda_0$.

Assumption N6: $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} g_{it} \xrightarrow{D} N(0, \tau^2)$ *as* $N, T \to \infty$.

It is extremely difficult, if possible at all, to specify explicitly detailed conditions on g_{it} so that a version of CLT can apply. Given the highly nonlinear dependence of g_{it} on the non-identically distributed dependent data, no generic CLT for dependent sequence (as in McLeish (1975)) is applicable. Alternatively, one can directly assume the score function $G(\psi_0)$ to be asymptotically normal. Now, letting $C = \text{diag}\{I_{k+1}, \sqrt{T}, \sqrt{N}, 1\}$, we have the following theorem.

Theorem 2: *Given Assumptions C1-C6 and Assumptions N1-N6, we have*

$$\sqrt{NT}C^{-1}(\hat{\psi} - \psi_0) \xrightarrow{D} N\left(0, I^{-1}(\psi_0)K(\psi_0)I^{-1}(\psi_0)\right), \quad \textit{as } N, T \to \infty.$$

where $I(\psi_0) = -\lim_{N,T\to\infty} \frac{1}{NT} C\text{E}[H(\psi_0)]C$ *and* $K(\psi_0) = \lim_{N,T\to\infty} \frac{1}{NT} C\text{E}[G(\psi_0)G'(\psi_0)]C$, *both assumed to exist with* $I(\psi_0)$ *being positive definite. Furthermore, if* μ_i*'s,* η_t*'s and* v_{it}*'s are all normally distributed, then* $\sqrt{NT}C^{-1}(\hat{\psi} - \psi_0) \xrightarrow{D} N(0, I^{-1}(\psi_0))$*, as* $N, T \to \infty$.

The proof of Theorem 2 is given in Appendix. From Theorem 2, we see that the involvement of the C matrix clearly spells out the rate of convergence for the parameter estimates. The behavior of the QMLEs is different under following different scenarios:

(a) $N, T \to \infty$ such that $N/T \to c$, a positive finite constant;
(b) $N, T \to \infty$ such that $N/T \to \infty$;
(c) $N, T \to \infty$ such that $N/T \to 0$;
(d) $N \to \infty$, T is fixed;
(e) $T \to \infty$, N is fixed;

Under these scenarios, the asymptotic behavior of the QMLEs are as follows

(i) $\hat{\beta}, \hat{\sigma}_v^2$ and $\hat{\lambda}$ are $\sqrt{NT}$-consistent under (a)-(e);
(ii) ϕ_μ or (σ_μ^2) is $\sqrt{N}$-consistent under (a)-(d), but is inconsistent under (e).
(iii) ϕ_η or (σ_η^2) is $\sqrt{T}$-consistent under (a)-(c) and (e), but is inconsistent under (d).

Thus, the QMLEs $\hat{\beta}, \hat{\sigma}_v^2$ and $\hat{\lambda}$ are consistent when either N or T of both approach to infinity. In the case where both approach to infinity, they are $\sqrt{NT}$-consistent irrespective of the relative magnitude of N and T. When N approaches infinity but T is fixed, $\hat{\sigma}_\mu^2$ is consistent but $\hat{\sigma}_\eta^2$ is inconsistent. This is because there is no sufficient variations in η_t no matter how large N is. Similarly, when T goes to infinity but N is fixed, $\hat{\sigma}_\eta^2$ is consistent but $\hat{\sigma}_\mu^2$ is inconsistent. See Hsiao (2003, p. 41) for a discussion on a random effects model without functional form transformation.

The result of Theorem 2 provides theoretical base for statistical inferences for the transformed random effects models. Practical application of this result involves the estimation of $I(\psi_0)$ and $K(\psi_0)$. The former can be consistently estimated by $\widehat{I(\psi_0)} = -\frac{1}{NT}CH(\hat{\psi})C$, but for the latter, there are no readily available methods. This is because (i) Var$[G(\psi_0)$ does not have an explicit expression for the transformed panel models, (ii) $G(\psi_0)$ cannot be written as summation of NT independent terms, nor in the form of a U- or V-statistic. Thus, traditional methods of estimating Var$[G(\psi_0)]$ are not applicable.

4 Bootstrap Estimate of Variance-Covariance Matrix

As mentioned above, the difficulty in estimating the variance-covariance matrix of $\hat{\psi}$ is due to the lack of analytical expression for $K(\psi_0)$ or due to the fact that $G(\psi_0)$ does not have the desirable structure. Thus, we turn to the bootstrap method. The bootstrap procedure given below is inspired by the idea of *transformation based bootstrap (TBB)* put fourth by Lahiri (2003, p. 40), which generalized the idea of Hurvivh and Zeger (1987). The central idea can simply be stated as follows. If (a) a statistic is a function of a dependent sequence, (ii) this sequence can be transformed through a one-to-one transformation to a sequence that is approximately independent, and (c) the statistic can be expressed (at least to a close approximation) in terms of this new sequence, then the distribution of this statistic can be obtained by bootstrapping the new sequence in the usual way.

The bootstrap procedure is called the *Error Components Bootstrap* as it directly bootstraps on the estimated error components obtained by decomposing the estimated error vector $\hat{u} = h(Y, \hat{\lambda}) - X(\hat{\lambda})\hat{\beta}$. These estimated error components clearly contain approximately independent elements. The procedure is summarized as follows.

1. Reshape $\hat{u}$ into an $N \times T$ matrix denoted by $\hat{U}$. Decompose $\hat{u}$ into three components:
   - $\hat{\mu} = N \times 1$ vector of row means of $\hat{U}$
   - $\hat{\eta} = T \times 1$ vector of column means of $\hat{U}$
   - $\hat{v} = \hat{u} - \hat{\mu} \otimes 1_T - 1_N \otimes \hat{\eta}$
2. Resample in the usual way $\hat{\mu}, \hat{\eta}$ and $\hat{v}$ respectively to give $\hat{\mu}^*, \hat{\eta}^*$ and $\hat{v}^*$, and thus

$$\hat{u}^* = \hat{\mu}^* \otimes 1_T + 1_N \otimes \hat{\eta}^* + \hat{v}^*$$

3. Compute $G(\psi_0)$ using $\hat{u}^*$ and $\hat{\psi}$, denoted as $G^*(\hat{\psi})$

4. Repeat 1-4 B times to give $G_1^*(\hat{\psi}), G_2^*(\hat{\psi}), \cdots, G_B^*(\hat{\psi})$. The bootstrap estimate of $\mathrm{E}[G(\psi_0)G'(\psi_0)]$ is then given as

$$\frac{1}{B-1}\sum_{i=1}^{B}[G_i^*(\hat{\psi}) - \mu_G^*][G_i^*(\hat{\psi}) - \mu_G^*]',$$

   where $\mu_G^* = \frac{1}{B}\sum_{i=1}^{B} G_i^*(\hat{\psi})$. This gives a bootstrap estimate of $K(\psi_0)$, which together with $\widehat{I(\psi_0)}$ gives an estimate of the VC matrix of the QMLE $\hat{\psi}$.

Some details in calculating $G^*(\hat{\psi})$ is given as follows. From Appendix (proof of Theorem 2), we see that the first four elements of $G(\psi_0)$ are all explicit functions of u_0 and the true parameters no matter what transformation function is adopted. Their bootstrapped values can thus be obtained by plugging $\hat{u}^*$ and $\hat{\psi}$ in these functions for u_0 and ψ_0, respectively. Calculating bootstrapped values of the last element of $G^*(\hat{\psi})$, i.e., $G_\lambda(\psi_0) = J_\lambda(\lambda) - \frac{1}{\sigma_v^2}u_\lambda'\Omega^{-1}u$, requires some algebra which is transformation specific.

For the Box-Cox power transformation, the transformation used in our Monte Carlo simulation, we have $J_\lambda(\lambda_0) = \sum_{i=1}^{N}\sum_{t=1}^{T}\log Y_{it}$, and $h_\lambda(Y_{iy}, \lambda_0) = \lambda_0^{-1}[1+\lambda_0 h(Y_{it}, \lambda_0)]\log Y_{it} - \lambda_0^{-1}h(Y_{it}, \lambda_0)$ when $\lambda_0 \neq 0$; $\frac{1}{2}(\log Y_{it})^2$ when $\lambda_0 = 0$. Since $h(Y_{it}, \lambda_0) = x_{it}'(\lambda_0)\beta_0 + u_{0,it}$, $\log Y_{it} = \frac{1}{\lambda}\log[1 + \lambda_0 h(Y_{it}, \lambda_0)]$, and $u_\lambda = h_\lambda(Y_{iy}, \lambda_0) - X_\lambda(\lambda_0)\beta_0$, the gradient $G_\lambda(\psi_0)$ can also be expressed analytically in terms of u_0 and ψ_0. Thus, the bootstrapped values of $G_\lambda(\psi_0)$ can again be obtained by plugging $\hat{u}^*$ and $\hat{\psi}$ in $G_\lambda(\psi_0)$ for u_0 and ψ_0. For other transformations, one could go through the same process, although the expressions may be more complicated than those of Box-Cox power transformation.

The advantage of the proposed bootstrap procedure is that it is computationally feasible even for large panels. This is because it bootstraps the score function only, for given QMLEs of parameters, based on resampling the estimated error components, thus avoiding the numerical optimization as in obtaining the parameter estimates. Section 5 presents Monte Carlo results for the performance of this bootstrap procedure for VC estimation in the forms of confidence intervals for model parameters.

5 Monte Carlo Results

Monte Carlo experiments we conducted serve two purposes: one is for checking the convergence rates of the QMLEs under different scenarios concerning the relative magnitude of N and T discussed in Section 3, and the other is for investigating the finite sample performance of the bootstrap estimate of VC matrix when used in confidence interval construction. The data generating process (DGP) used in the Monte Carlo experiments is as follows.

$$h(Y,\lambda) = \beta_0 + \beta_1 X_1 + \beta_2 h(X_2,\lambda) + Z_\mu \mu + Z_\eta \eta + v$$

where h is the Box-Cox power transformation with $\lambda = 0.1$, X_1 is generated from $U(0,5)$, X_2 from $exp[N(0,1)]$, $\beta = (20,5,1)'$, $\sigma_\mu = \sigma_\eta = 0.5$, and $\sigma_v = 1.0$.

To generate error components $\{\mu_i\}$, $\{\eta_t\}$ and $\{v_{it}\}$, we consider three distributions: (i) normal, (ii) normal-mixture, and (iii) normal-gamma mixture, all standardized to have zero mean and unit variance. The standardized normal-mixture random variates are generated according to

$$W_i = ((1-\xi_i)Z_i + \xi_i \tau Z_i)/(1-p+p\tau^2)^{0.5},$$

where ξ is a Bernoulli random variable with probability of success p and Z_i is standard normal independent of ξ. The parameter p in this case also represents the proportion of mixing the two normal populations. In our experiments, we choose $p = 0.05$ or 0.10, meaning that 95% or 90% of the random variates are generated from the standard normal and the remaining 5% or 10% are from another normal population with standard deviation τ. We choose $\tau = 5$ or 10 to simulate the situations where there are gross errors in the data. Similarly, the standardized normal-gamma mixture random variates are generated according to

$$W_i = ((1-\xi_i)Z_i + \xi_i(V_i - \alpha))/(1-p+p\alpha)^{0.5},$$

where V_i is a gamma random variable with scale parameter 1 and shape parameter α, and is independent of Z_i and ξ_i. The other quantities are the same as in the definition of normal-mixture. We choose $p = 0.05$ or 0.10, and $\alpha = 4$ or 9.

Note that the normal-mixture gives a non-normal distribution that is still symmetric like normal distribution but leptokurtic, whereas the normal-gamma mixture gives a non-normal distribution that is both skewed and leptokurtic. As we discussed in the introduction, one of the main purposes of a response transformation is to induce normality of the data. We argued that while exact normality may be impractical, the transformed observations can be close to normal or at least more symmetrically distributed, which makes the use of Gaussian likelihood more valid in the QMLE process. This means that there could still be 'mild' departure from normality for the error distributions in the forms of excess kurtosis or skewness or both. As symmetry can pretty much be achieved by transformation, it is thus more interesting to see the behaviors of the QMLEs and bootstrap VC matrix estimation in the case of excess kurtosis, i.e., the case of normal-mixture. Nevertheless, we still include the normal-gamma mixture case to see what happens when the transformed data is still 'far' from being normal in the sense that there is still a certain degree of skewness left after the so called 'normalizing' transformation.

5.1 Convergence of the QMLEs

Table 1 presents Monte Carlo results for the finite sample performance of the QMLEs of the model parameters ψ, where DGP 1 corresponds to the case that all errors are normal, DGP 2 corresponds to the case that μ and η are normal but v follows a normal-mixture distribution with $p = 0.05$ and $\tau = 5$, and DGP 3 corresponds to the case that μ and η are normal but v follows a normal-gamma mixture with $p = 0.05$ and $\alpha = 4$. The results corresponding to each combination of values of N, T and DGP are based on 10,000 samples.

Tables 1a-1d correspond to the cases where N and T increase concurrently with the same or different speeds. The results clearly show that as N and T get larger, the bias and the root mean squared error (rmse) get smaller. If N is relatively larger than T, then the bias and rmse of ϕ_η are larger than those of ϕ_μ and vise versa. If T is fixed as in Table 1e, then the bias and rmse of ϕ_η, in particular the former, do not go down as N increases. Similarly, if N is fixed as in Table 1f, then the bias and rmse of ϕ_μ do not go down as T increases.

The results corresponding to DGP 2 and DGP 3 do not differ much from those corresponding to DGP 1 as far as the general observations are concerned. However, introducing the nonnormality does make the rmse larger, especially in the case of DGP 2. As discussed in Section 3, the consistency of $\hat{\lambda}$ may require that the errors in the transformed model be near symmetric. The results in Table 1 do indicate that the bias of $\hat{\lambda}$ is indeed smaller in the case of symmetric nonnormal errors (DGP 2) than in the case of asymmetric nonnormal errors (DGP 3). However, the magnitude of bias is still quite small. Monte Carlo experiments are repeated under other parameter (ψ) values as well. The results (not reported for brevity) show similar patterns and lead to the same conclusions.

5.2 Performance of the bootstrap estimate of VC matrix

This subsection concerns on the finite sample performance of the bootstrap estimates of the variance-covariance (VC) matrix of the QMLEs in terms of the coverage probability of confidence interval (CI) for each parameter in the model. The same DGPs as in Section 5.1 are used with some changes on the parameter values in the mixture distributions. The results are summarized in Table 5.2. Due to the fact that bootstrap procedure is computationally more demanding, we use 5,000 samples for each Monte Carlo experiment instead of 10,000 as in Section 5.1. The number of bootstrap samples is chosen to be 300.

From the results we see that in case of symmetric non-normal errors, the bootstrap procedure leads to confidence intervals (`Boot`) with coverage probabilities generally quite close to their nominal level even though the mixture distributions considered in the Monte Carlo experiment are quite different from normal distribution. The Hessian-based confidence intervals (Hess) can perform quite poorly with coverage probabilities significantly below the nominal level.

In case of skewed errors (DGP 3), the performance of bootstrap confidence intervals are not as good as in the case of symmetric but nonnormal error (DGP 2), but still significantly better than the confidence intervals based on Hessian matrix. One point to note is that the inferences based on both methods do not improve when N and T increase. The cause of this may be the biasedness of some parameter estimates (e.g., $\hat{\lambda}$), inherited from the skewness of the error distribution. However, the amount of skewness generated by DGP 3 is incompatible with the transformtion model as response transformation can typically achieve near-symmetry as discussed earlier. Another interesting point to note is that the CIs for $\hat{\phi}_\mu$ and $\hat{\phi}_\eta$ based on both methods perform equally well in all situations. One possible reason for this may be that the error components μ and η are generated from normal for all the reported Monte Carlo results. This shows that whether the pure error v is normal or non-normal does not affect the performance of the inference for ϕ_μ and ϕ_η.

6 Discussions

Asymptotic properties of the QMLEs of the transformed panel model with two-way random effects are studies and a bootstrap method for estimating the variance-covariance matrix is introduced. Typically, a consistent estimate of the model requires both N and T to be large. When N is large but T is fixed, only the variance of the time-specific random effects cannot be consistently estimated; when T is large but N is fixed, only the variance of the individual-specific random effects cannot be consistently estimated. The error component bootstrap (ECB) estimate of the VC matrix works well when errors are not normal but still symmetrically distributed, and in this case it shows enormous improvements over the case where the VC matrix is estimated by the Hessian matrix.

Appendix: Proofs of the Theorems

Denote $u \equiv u(\beta, \lambda) = H(Y, \lambda) - X(\lambda)\beta$, $u_0 = u(\beta_0, \lambda_0)$, $u_\lambda \equiv u_\lambda(\beta, \lambda) = \partial u(\beta, \lambda)/\partial\lambda$, and $u_{\lambda\lambda} \equiv u_{\lambda\lambda}(\beta, \lambda) = \partial^2 u(\beta, \lambda)/\partial\lambda^2$. Recall $\Omega^{-1} = Q + \theta_1 P_1 + \theta_2 P_2 + \theta_3 P_3$, where $\theta_1 = 1/(T\phi_\mu + 1)$, $\theta_2 = 1/(N\phi_\eta + 1)$, and $\theta_3 = 1/(T\phi_\mu + N\phi_\eta + 1)$. The proof of Theorem 1 needs the following two lemmas.

Lemma A1: *Under the assumptions of Theorem 1, the quantity $\bar{\sigma}_v^2(\phi)$ defined in 9 is bounded below from zero as $N, T \to \infty$, uniformly in $\phi \in \Phi$.*

Proof: Recall $\bar{\ell}(\psi) \equiv \mathrm{E}[\ell(\psi)] = -\frac{NT}{2}\log(\sigma_v^2) + c(\phi_\mu, \phi_\eta) - \frac{1}{2\sigma_v^2}\mathrm{E}[u'(\beta, \lambda)\Omega^{-1}u(\beta, \lambda)] + \mathrm{E}[J(\lambda)]$. This gives $\bar{\ell}(\psi_0) = -\frac{NT}{2}\log(\sigma_{v0}^2) + c(\phi_{\mu 0}, \phi_{\eta 0}) - \frac{NT}{2} + \mathrm{E}[J(\lambda_0)]$. By information inequality (see, e.g., White (1994, p.9)), we have

$$\mathrm{E}[\ell(\psi)] \leq \mathrm{E}[\ell(\psi_0)], \forall \psi \in \Psi.$$

It follows that $\bar{\ell}_{\max}(\phi) = \max_{\beta, \sigma_v^2} \mathrm{E}[\ell(\psi)] \leq \mathrm{E}[\ell(\psi_0)], \forall \phi \in \Phi$. That is,

$$c(\phi_\mu, \phi_\eta) - \frac{NT}{2}[1 + \log \bar{\sigma}_v^2(\phi)] + \mathrm{E}[J(\lambda)] \leq c(\phi_{\mu 0}, \phi_{\eta 0}) - \frac{NT}{2}[1 + \log(\sigma_{v0}^2)] + \mathrm{E}[J(\lambda_0)],$$

or equivalently, $\log \bar{\sigma}_v^2(\phi) \geq \frac{1}{NT}[c(\phi_\mu, \phi_\eta) - c(\phi_{\mu 0}, \phi_{\eta 0})] + \log \sigma_{v0}^2 + \frac{2}{NT}[\mathrm{E}J(\lambda) - \mathrm{E}J(\lambda_0)]$. Assumption C4 guarantees that $\frac{2}{NT}[\mathrm{E}J(\lambda) - \mathrm{E}J(\lambda_0)]$ is bounded for all N and T, uniformly in $\lambda \in \Lambda$. It follows that $\bar{\sigma}_v^2(\phi)$ is bounded away from zero uniformly in $\phi \in \Phi$. *Q.E.D.*

Lemma A2: *Under the assumptions of Theorem 1, $|\hat{\sigma}_v^2(\phi) - \bar{\sigma}_v^2(\phi)| \xrightarrow{p} 0$ as $N, T \to \infty$, uniformly in $\phi \in \Phi$.*

Proof:[3] By (6) and (9), we have,

$$\begin{aligned} & \hat{\sigma}_v^2(\phi) - \bar{\sigma}_v^2(\phi) \\ = & \frac{1}{NT}\{h'(Y, \lambda)\Omega^{-1}h(Y, \lambda) - \mathrm{E}[h'(Y, \lambda)\Omega^{-1}h(Y, \lambda)]\} \\ & -\frac{1}{NT}\{h'(Y, \lambda)\Omega^{-\frac{1}{2}}P^*(\phi)\Omega^{-\frac{1}{2}}h(Y, \lambda) - \mathrm{E}[h'(Y, \lambda)\Omega^{-\frac{1}{2}}P^*(\phi)\Omega^{-\frac{1}{2}}h(Y, \lambda)]\}, \quad (11) \end{aligned}$$

where $P^*(\phi) = \Omega^{-\frac{1}{2}}X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-\frac{1}{2}}$, a projection matrix, and $\Omega^{-\frac{1}{2}}$ is the symmetric square root of Ω^{-1}. We prove the lemma in two steps.

Step 1. To show $\sup_{\phi \in \Phi} |Q_1(\phi) - \mathrm{E}[Q_1(\phi)]| = o_p(1)$, where $Q_1(\phi) = \frac{1}{NT}h'(Y, \lambda)\Omega^{-1}h(Y, \lambda)$. Note that

$$Q_1(\phi) \leq \frac{\gamma_{\max}(\Omega^{-1})}{NT}\|h(Y, \lambda)\|^2 \leq \frac{4}{NT}\|h(Y, \lambda)\|^2$$

[3]In proving Lemma A2, the following matrix results are repeatedly used: (i) the largest eigenvalue of a projection matrix is less than and equal to 1; (ii) $\gamma_{\min}(A)\mathrm{tr}B \leq \mathrm{tr}(AB) \leq \gamma_{\max}(A)\mathrm{tr}B$ for symmetric matrix A and positive semidefinte (p.s.d.) matrix B, (ii) $\gamma_{\max}(A+B) \leq \gamma_{\max}(A) + \gamma_{\max}(B)$ for symmetric matrices A and B; and (iv) $\gamma_{\max}(AB) \leq \gamma_{\max}(A)\gamma_{\max}(B)$ for p.s.d. matrices A and B.

where the last inequality follows from the fact that $\sup_{\phi\in\Phi}\gamma_{\max}(\Omega^{-1}) \leq 4$ (see the remark after Assumptions C1-C6). The pointwise convergence follows from Assumptions C4-C5(i) and the dominated convergence theorem. We are left to show the stochastic equicontinuity of $Q_1(\phi)$. By the mean value theorem,

$$Q_1(\phi) - Q_1(\tilde{\phi}) = Q_{1\lambda}(\phi^*)(\lambda - \tilde{\lambda}) + Q_{1\phi_\mu}(\phi^*)(\phi_\mu - \tilde{\phi}_\mu) + Q_{1\phi_\eta}(\phi^*)(\phi_\eta - \tilde{\phi}_\eta)$$

where $\phi^* \equiv (\phi_\mu^*, \phi_\eta^*, \lambda^*)'$ lies between ϕ and $\tilde{\phi}$,

$$\begin{aligned}
Q_{1\lambda}(\phi) &= \frac{2}{NT} h'_\lambda(Y,\lambda)\Omega^{-1}h(Y,\lambda), \\
Q_{1\phi_\mu}(\phi) &= -\frac{1}{NT} h'(Y,\lambda)(T\theta_1^2 P_1 + T\theta_3^2 P_3)h(Y,\lambda), \text{ and} \\
Q_{1\phi_\eta}(\phi) &= -\frac{1}{NT} h'(Y,\lambda)(N\theta_2^2 P_2 + N\theta_3^2 P_3)h(Y,\lambda).
\end{aligned}$$

By Lemma 1 of Andrews (1992), it suffices to show that $\sup_{\phi\in\Phi}|Q_{1\xi}(\phi)| = O_p(1)$ for $\xi = \lambda, \phi_\mu$, and ϕ_η. First, by Cauchy-Schwarz inequality and Assumption C6(ii) and C6(iii),

$$\begin{aligned}
\sup_{\phi\in\Phi}|Q_{1\lambda}(\phi)| &\leq \sup_{\phi\in\Phi}\frac{2}{NT}\{h'_\lambda(Y,\lambda)\Omega^{-1}h_\lambda(Y,\lambda)\}^{1/2}\{h'(Y,\lambda)\Omega^{-1}h(Y,\lambda)\}^{\frac{1}{2}} \\
&\leq 2\{\sup_{\phi\in\Phi}\gamma_{\max}(\Omega^{-1})\}\sup_{\lambda\in\Lambda}\frac{1}{\sqrt{NT}}\|h_\lambda(Y,\lambda)\|\sup_{\lambda\in\Lambda}\frac{1}{\sqrt{NT}}\|h(Y,\lambda)\| \\
&\leq 8\cdot O_p(1)\cdot O_p(1) = O_p(1).
\end{aligned}$$

Now, by Assumption C2, the positive constants $T\theta_1^2$, $N\theta_2^2$, $T\theta_3^2$ and $N\theta_3^2$ are such that $T\theta_3^2$ and $N\theta_3^2$ are $o(1)$ if $N \cup T \to \infty$; $T\theta_1^2$ is free of N, which is $O(1)$ if T is fixed and $o(1)$ if T grows; and $N\theta_2^2$ is free of T, which is $O(1)$ if N is fixed and $o(1)$ if N grows. In any case, they are bounded uniformly by a constant $\overline{c}$, say. We have

$$\begin{aligned}
\sup_{\phi\in\Phi}|Q_{1\phi_\mu}(\phi)| &\leq \sup_{\phi\in\Phi}\frac{1}{NT}h'(Y,\lambda)(T\theta_1^2P_1 + T\theta_3^2P_3)h(Y,\lambda) \\
&\leq \sup_{\phi\in\Phi}(T\theta_1^2\gamma_{\max}(P_1) + T\theta_3^2\gamma_{\max}(P_3))\frac{1}{NT}\|h(Y,\lambda)\|^2 \\
&\leq 2\overline{c}\sup_{\lambda\in\Lambda}\frac{1}{NT}\|h(Y,\lambda)\|^2 = O_p(1) \text{ by Assumption C6(ii),}
\end{aligned}$$

and similarly

$$\begin{aligned}
\sup_{\phi\in\Phi}|Q_{1\phi_\eta}(\phi)| &\leq \sup_{\phi\in\Phi}\frac{1}{NT}h'(Y,\lambda)(N\theta_2^2P_2 + N\theta_3^2P_3)h(Y,\lambda) \\
&\leq \sup_{\phi\in\Phi}(N\theta_2^2\gamma_{\max}(P_2) + N\theta_3^2\gamma_{\max}(P_3))\frac{1}{NT}\|h(Y,\lambda)\|^2 \\
&\leq 2\overline{c}\sup_{\lambda\in\Lambda}\frac{1}{NT}\|h(Y,\lambda)\|^2 = O_p(1) \text{ by Assumption C6(ii).}
\end{aligned}$$

Step 2. To show $\sup_{\phi\in\Phi}|Q_2(\phi)-\mathrm{E}[Q_2(\phi)]|=o_p(1)$, where $Q_2(\phi)=\frac{1}{NT}h'(Y,\lambda)\Omega^{-\frac{1}{2}}P^*(\phi)\Omega^{-\frac{1}{2}}h(Y,\lambda)$. Noting that $Q_2(\phi)\le\gamma_{\max}(P^*(\phi))\frac{1}{NT}h'(Y,\lambda)\Omega^{-1}h(Y,\lambda)\le Q_1(\phi)$, the pointwise convergence follows from Step 1 and the dominated convergence theorem. We now show the stochastic equicontinuity of $Q_2(\phi)$. By the mean value theorem,

$$Q_2(\phi)-Q_2(\tilde\phi)=Q_{2\lambda}(\phi^{**})(\lambda-\tilde\lambda)+Q_{2\phi_\mu}(\phi^{**})(\phi_\mu-\tilde\phi_\mu)+Q_{2\phi_\eta}(\phi^{**})(\phi_\eta-\tilde\phi_\eta),$$

where $\phi^{**}\equiv(\phi_\mu^{**},\phi_\eta^{**},\lambda^{**})'$ lies between ϕ and $\tilde\phi$,

$$\begin{aligned}
Q_{2\lambda}(\phi) &= \frac{2}{NT}h'_\lambda(Y,\lambda)\Omega^{-1}X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda)\\
&\quad+\frac{2}{NT}h'(Y,\lambda)\Omega^{-1}X_\lambda(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda)\\
&\quad-\frac{2}{NT}h'(Y,\lambda)\Omega^{-1}X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'_\lambda(\lambda)\Omega^{-1}X(\lambda)\\
&\quad\times[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda)\\
&\equiv 2Q_{2\lambda,1}(\phi)+2Q_{2\lambda,2}(\phi)-2Q_{2\lambda,3}(\phi),\\
Q_{2\phi_\mu}(\phi) &= -\frac{2}{NT}h'(Y,\lambda)(T\theta_1^2P_1+T\theta_3^2P_3)X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda)\\
&\quad+\frac{1}{NT}h'(Y,\lambda)\Omega^{-1}X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}(T\theta_1^2P_1+T\theta_3^2P_3)\\
&\quad\times[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda),\ \textit{and}\\
Q_{2\phi_\eta}(\phi) &= -\frac{2}{NT}h'(Y,\lambda)(N\theta_2^2P_2+N\theta_3^2P_3)X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda)\\
&\quad+\frac{1}{NT}h'(Y,\lambda)\Omega^{-1}X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}(N\theta_2^2P_2+N\theta_3^2P_3)\\
&\quad\times[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda).
\end{aligned}$$

By the Cauchy-Schwarz inequality, and Assumption C6(ii)-(iii),

$$\begin{aligned}
&\sup_{\phi\in\Phi}|Q_{2\lambda,1}(\phi)|\\
=\;&\sup_{\phi\in\Phi}\frac{1}{NT}\left|h'_\lambda(Y,\lambda)\Omega^{-\frac{1}{2}}P^*(\phi)\Omega^{-\frac{1}{2}}h(Y,\lambda)\right|\\
\le\;&\sup_{\phi\in\Phi}\frac{1}{NT}\left\{h'_\lambda(Y,\lambda)\Omega^{-\frac{1}{2}}P^*(\phi)\Omega^{-\frac{1}{2}}h_\lambda(Y,\lambda)\right\}^{\frac{1}{2}}\left\{h'(Y,\lambda)\Omega^{-\frac{1}{2}}P^*(\phi)\Omega^{-\frac{1}{2}}h(Y,\lambda)\right\}^{\frac{1}{2}}\\
\le\;&\sup_{\phi\in\Phi}\gamma_{\max}(P^*(\phi))\frac{1}{NT}\left\{h'_\lambda(Y,\lambda)\Omega^{-1}h_\lambda(Y,\lambda)\right\}^{\frac{1}{2}}\left\{h'(Y,\lambda)\Omega^{-1}h(Y,\lambda)\right\}^{\frac{1}{2}}\\
\le\;&\sup_{\phi\in\Phi}\gamma_{\max}(\Omega^{-1})\frac{1}{NT}\|h_\lambda(Y,\lambda)\|\|h(Y,\lambda)\|\\
\le\;&4\sup_{\lambda\in\Lambda}\frac{1}{\sqrt{NT}}\|h_\lambda(Y,\lambda)\|\sup_{\lambda\in\Lambda}\frac{1}{\sqrt{NT}}\|h(Y,\lambda)\|=O_p(1).
\end{aligned}\tag{12}$$

Similarly,

$$\begin{aligned}
\sup_{\phi\in\Phi}|Q_{2\lambda,2}(\phi)| &= \sup_{\phi\in\Phi}\frac{1}{NT}\left|h'(Y,\lambda)\Omega^{-1}X_\lambda(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda)\right| \\
&\le \sup_{\phi\in\Phi}\left\{\frac{1}{NT}h'(Y,\lambda)\Omega^{-1}X_\lambda(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'_\lambda(\lambda)\Omega^{-1}h(Y,\lambda)\right\}^{\frac{1}{2}} \\
&\quad\times\sup_{\phi\in\Phi}\left\{\frac{1}{NT}h'(Y,\lambda)\Omega^{-1}X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda)\right\}^{\frac{1}{2}}.
\end{aligned}$$

Let $Q_{XX}(\phi)\equiv\frac{1}{NT}X'(\lambda)\Omega^{-1}X(\lambda)$. The first term on the right hand side (r.h.s.) is

$$\begin{aligned}
&\le \sup_{\phi\in\Phi}\left[\gamma_{\min}\left(Q_{XX}(\phi)\right)\right]^{-\frac{1}{2}}\sup_{\phi\in\Phi}\frac{1}{NT}\left\{h'(Y,\lambda)\Omega^{-1}X_\lambda(\lambda)X'_\lambda(\lambda)\Omega^{-1}h(Y,\lambda)\right\}^{\frac{1}{2}} \\
&\le \sup_{\phi\in\Phi}[\gamma_{\min}(Q_{XX}(\phi))]^{-1}\sup_{\lambda\in\Lambda}\frac{1}{\sqrt{NT}}\|X_\lambda(\lambda)\|\sup_{\phi\in\Phi}\gamma_{\max}(\Omega^{-1})\sup_{\lambda\in\Lambda}\frac{1}{\sqrt{NT}}\|h'(Y,\lambda)\| \\
&= O_p(1)O_p(1)O(1)O_p(1)=O_p(1),\ \ \text{by Assumptions C3 and C6(i),}
\end{aligned}$$

and $\sup_{\phi\in\Phi}\gamma_{\max}(\Omega^{-1})\le 4$. The second term on the r.h.s. is

$$\begin{aligned}
&\sup_{\phi\in\Phi}\left\{\frac{1}{NT}h'(Y,\lambda)\Omega^{-\frac{1}{2}}P^*(\phi)\Omega^{-\frac{1}{2}}h(Y,\lambda)\right\}^{\frac{1}{2}} \\
&\le \left\{\sup_{\phi\in\Phi}\gamma_{\max}(\Omega^{-\frac{1}{2}}P^*(\phi)\Omega^{-\frac{1}{2}})\right\}^{1/2}\sup_{\lambda\in\Lambda}\frac{1}{\sqrt{NT}}\|h'(Y,\lambda)\| \\
&\le 2\sup_{\lambda\in\Lambda}\frac{1}{\sqrt{NT}}\|h'(Y,\lambda)\|=O_p(1),\ \ \text{by Assumption C6(i).}
\end{aligned}$$

Consequently,

$$\sup_{\phi\in\Phi}|Q_{2\lambda,2}(\phi)|=O_p(1). \tag{13}$$

Next,

$$\begin{aligned}
\sup_{\phi\in\Phi}|Q_{2\lambda,3}(\phi)| &\le \sup_{\phi\in\Phi}\|h'(Y,\lambda)\Omega^{-1}X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}\|\ \sup_{\lambda\in\Lambda}\frac{1}{\sqrt{NT}}\|X'_\lambda(\lambda)\| \\
&\quad\times\sup_{\phi\in\Phi}\frac{1}{\sqrt{NT}}\|\Omega^{-1}X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}X'(\lambda)\Omega^{-1}h(Y,\lambda)\|.
\end{aligned}$$

The first term on the r.h.s. is

$$\begin{aligned}
&\sup_{\phi\in\Phi}\|h'(Y,\lambda)\Omega^{-1}X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}\| \\
&\le \sup_{\phi\in\Phi}\left\{\|h'(Y,\lambda)\Omega^{-1}X(\lambda)[X'(\lambda)\Omega^{-1}X(\lambda)]^{-2}X'(\lambda)\Omega^{-1}h(Y,\lambda)\|\right\}^{\frac{1}{2}} \\
&\le \sup_{\phi\in\Phi}[\gamma_{\min}(Q_{XX}(\phi))]^{-\frac{1}{2}}\frac{1}{\sqrt{NT}}\left\{\left\|h'(Y,\lambda)\Omega^{-\frac{1}{2}}P^*(\phi)\Omega^{-\frac{1}{2}}h(Y,\lambda)\right\|\right\}^{\frac{1}{2}} \\
&\le \sup_{\phi\in\Phi}[\gamma_{\min}(Q_{XX}(\phi))]^{-\frac{1}{2}}\sup_{\phi\in\Phi}\left\{\gamma_{\max}(\Omega^{-1})\right\}^{\frac{1}{2}}\frac{1}{\sqrt{NT}}\|h(Y,\lambda)\| \\
&= O_p(1)O(1)O_p(1)=O_p(1)\ \ \text{by Assumptions C3 and C6(i),}
\end{aligned}$$

the middle term is $O(1)$ by Assumption C6(i), and the third term on r.h.s. is

$$\begin{aligned}
& \sup_{\phi\in\Phi} \frac{1}{\sqrt{NT}} \left\| \Omega^{-\frac{1}{2}} P^*(\phi) \Omega^{-\frac{1}{2}} h(Y,\lambda) \right\| \\
= \;& \sup_{\phi\in\Phi} \frac{1}{\sqrt{NT}} \left\{ h'(Y,\lambda) \Omega^{-\frac{1}{2}} P^*(\phi) \Omega^{-1} P^*(\phi) \Omega^{-\frac{1}{2}} h(Y,\lambda) \right\}^{\frac{1}{2}} \\
\leq \;& \left\{ \gamma_{\max}\left(\Omega^{-\frac{1}{2}} P^*(\phi) \Omega^{-1} P^*(\phi) \Omega^{-\frac{1}{2}} \right) \right\}^{\frac{1}{2}} \sup_{\lambda\in\Lambda} \frac{1}{\sqrt{NT}} \|h(Y,\lambda)\| \\
\leq \;& 4 \sup_{\lambda\in\Lambda} \frac{1}{\sqrt{NT}} \|h(Y,\lambda)\| = O_p(1) \;\text{ by Assumption C6(i),}
\end{aligned}$$

where we repeatedly use the fact that $\gamma_{\max}(AB) \leq \gamma_{\max}(A)\gamma_{\max}(B)$ for positive semidefinite matrices A and B. Consequently, we have

$$\sup_{\phi\in\Phi} |Q_{2\lambda,3}(\phi)| = O_p(1). \tag{14}$$

By the triangle inequality, combining (12)-(14) yields $\sup_{\phi\in\Phi} |Q_{2\lambda}(\phi)| = O_p(1)$.

Analogously, we can show $\sup_{\phi\in\Phi} |Q_{2\phi_\mu}(\phi)| = O_p(1)$ and $\sup_{\phi\in\Phi} |Q_{2\phi_\eta}(\phi)| = O_p(1)$. This completes the proof of the lemma. *Q.E.D.*

Proof Theorem 1: Since $\ell_{\max}(\phi)$ has identifiably unique maximizer ϕ_0, following White (1994, Theorem 3.4), the proof of the consistency of $\hat{\phi}$ amounts to show the uniform convergence

$$\sup_{\phi\in\Phi} \frac{1}{NT} \left| \ell_{\max}(\phi) - \bar{\ell}_{\max}(\phi) \right| \xrightarrow{p} 0, \;\text{ as }\; N, T \longrightarrow \infty. \tag{15}$$

From (7) and (10), we have $\frac{1}{NT}[\ell_{\max}(\phi) - \bar{\ell}_{\max}(\phi)] = -\frac{1}{2}[\log \hat{\sigma}_v^2(\phi) - \log \bar{\sigma}_v^2(\phi)] + \frac{1}{NT}\{J(\lambda) - \mathrm{E}[J(\lambda)]\}$. By a Taylor expansion of $\log \hat{\sigma}_v^2(\phi)$ at $\bar{\sigma}_v^2(\phi)$, we obtain

$$|\log \hat{\sigma}_v^2(\phi) - \log \bar{\sigma}_v^2(\phi)| = |\hat{\sigma}_v^2(\phi) - \bar{\sigma}_v^2(\phi)| / \tilde{\sigma}_v^2(\phi),$$

where $\tilde{\sigma}_v^2(\phi)$ lies between $\hat{\sigma}_v^2(\phi)$ and $\bar{\sigma}_v^2(\phi)$. Lemma A1 shows that $\tilde{\sigma}_v^2(\phi)$ is bounded below from zero uniformly in $\phi \in \Phi$, and Lemma A2 shows that $|\hat{\sigma}_v^2(\phi) - \bar{\sigma}_v^2(\phi)| \xrightarrow{p} 0$, uniformly in $\phi \in \Phi$. Hence, $|\log \hat{\sigma}_v^2(\phi) - \log \bar{\sigma}_v^2(\phi)| \xrightarrow{p} 0$, uniformly in $\phi \in \Phi$. Now, Assumptions C5(ii) and C6(ii) ensure that $\frac{1}{NT}\{J(\lambda) - \mathrm{E}[J(\lambda)]\} \xrightarrow{p} 0$ uniformly in $\phi \in \Phi$. The result (A.1) thus follows. Finally, the consistency of $\hat{\beta}(\hat{\phi})$ and $\hat{\sigma}_v^2(\hat{\phi})$ follows from the consistency of $\hat{\phi}$ and Assumption C3. *Q.E.D.*

The proof of Theorem 2 requires Lemma A.3, which essentially gives a central limit theorem (CLT) for linear-quadratic forms of error components $u = Z_\mu \mu + Z_\eta \eta + v$ defied in (2). Let $m_{\mu 0}^{(k)}, m_{\eta 0}^{(k)}$ and $m_{v0}^{(k)}$ be, respectively, the kth moment of μ_i, η_t and $v_{it}, k = 1, 2, 3, 4$.

Let $\kappa_{\mu 0} = m_{\mu 0}^{(k)}/\sigma_{\mu 0}^4 - 3$, $\kappa_{\eta 0} = m_{\eta 0}^{(k)}/\sigma_{\eta 0}^4 - 3$ and $\kappa_{v0} = m_{v0}^{(k)}/\sigma_{v0}^4 - 3$. Denote $n = NT$. Let $W = \{w_{jk}\}$ be an $n \times n$ symmetric matrix and $b = \{b_j\}$ be an $n \times 1$ column vector. Consider the following linear-quadratic form:

$$q = u'Wu + b'u.$$

Lemma A3. *Assume (i)* $\mathrm{E}|\mu_i|^{4+\epsilon_1} < \infty$, $\mathrm{E}|\eta_t|^{4+\epsilon_2} < \infty$, *and* $\mathrm{E}|v_{it}|^{4+\epsilon_3} < \infty$, *for some* ϵ_1, ϵ_2 *and* $\epsilon_3 > 0$, *and for all* $i = 1, \cdots, N$ *and* $t = 1, \cdots, T$; *(ii)* $\sup_{1 \le k \le n} \sum_{j=1}^n |w_{jk}| < \infty$; *and (iii)* $\sup_n n^{-1} \sum_{j=1}^n |b_j|^{2+\epsilon} < \infty$ *for some* $\epsilon > 0$. *Then, we have, as* $n \longrightarrow \infty$,

$$\frac{q - \mu_q}{\sigma_q} \xrightarrow{D} N(0, 1),$$

where $\mu_q = \mathrm{E}(q) = \sigma_{\mu 0}^2 tr(W_\mu) + \sigma_\eta^2 tr(W_\eta) + \sigma_v^2 tr(W)$, *and*

$$\begin{aligned}\sigma_q^2 = {} & \sigma_{\mu 0}^4(\kappa_{\mu 0} \textstyle\sum w_{\mu,ii}^2 + 2tr(W_\mu^2)) + \sigma_{\eta 0}^4(\kappa_{\eta 0} \sum w_{\eta,ii}^2 + 2tr(W_\eta^2)) + \sigma_{v0}^4(\kappa_{v0} \sum w_{ii}^2 + 2tr(W^2)) \\ & + 4\sigma_{\mu 0}^2\sigma_{\eta 0}^2 tr(Z_\mu' W Z_\eta Z_\eta' W Z_\mu) + 4\sigma_{\mu 0}^2\sigma_{v0}^2 tr(Z_\mu' W^2 Z_\mu) + 4\sigma_{\eta 0}^2\sigma_{v0}^2 tr(Z_\eta' W^2 Z_\eta), \\ & + 2m_{\mu 0}^{(3)} \textstyle\sum w_{\mu,ii} b_{\mu,i} + 2m_{\eta 0}^{(3)} \sum w_{\eta,ii} b_{\eta,i} + 2m_{v0}^{(3)} \sum w_{ii} b_i + \sigma_{v0}^2 b'\Omega b,\end{aligned}$$

where $W_\mu = Z_\mu' W Z_\mu$, $W_\eta = Z_\eta' W Z_\eta$, $b_\mu = Z_\mu' b$, *and* $b_\eta = Z_\eta' b$.

Proof: Since $u'Wu = (Z_\mu\mu + Z_\eta\eta + v)'W(Z_\mu\mu + Z_\eta\eta + v) = \mu' W_\mu \mu + \eta' W_\eta \eta + v'Wv + 2\mu' Z_\mu' W Z_\eta \eta + 2\mu' Z_\mu' W v + 2\eta' Z_\eta' W v$, we have $\mathrm{E}(q) = \sigma_{\mu 0}^2 tr(W_\mu) + \sigma_{\eta 0}^2 tr(W_\eta) + \sigma_{v0}^2 tr(W)$. Noting that the six terms in the expansion of $u'Wu$ are mutually uncorrelated, we have

$$\begin{aligned}\mathrm{Var}(u'Wu) &= \mathrm{Var}(\mu' W_\mu \mu) + \mathrm{Var}(\eta' W_\eta \eta) + \mathrm{Var}(v'Wv) \\ &\quad + 4\mathrm{Var}(\mu' Z_\mu' W Z_\eta \eta) + 4\mathrm{Var}(\mu' Z_\mu' W v) + 4\mathrm{Var}(\eta' Z_\eta' W v)\end{aligned}$$

It is easy to show that

$$\begin{aligned}\mathrm{Var}(\mu' W_\mu \mu) &= \sigma_{\mu 0}^4 \left(\kappa_{\mu 0} \textstyle\sum w_{\mu,ii}^2 + 2tr(W_\mu^2)\right), \\ \mathrm{Var}(\mu' Z_\mu W Z_\eta \eta) &= \sigma_{\mu 0}^2 \sigma_{\eta 0}^2 tr(Z_\mu' W Z_\eta Z_\eta' W Z_\mu).\end{aligned}$$

The former leads to the expressions for $\mathrm{Var}(\eta' W_\eta \eta)$ and $\mathrm{Var}(v'Wv)$, and the latter leads to the expressions for $\mathrm{Var}(\mu' Z_\mu' W v)$ and $\mathrm{Var}(\eta' Z_\eta' W v)$. Finally,

$$\mathrm{Cov}(u'Wu, b'u) = m_{\mu 0}^{(3)} \sum w_{\mu,ii} b_{\mu,i} + m_{\eta 0}^{(3)} \sum w_{\eta,ii} b_{\eta,i} + m_{v0}^{(3)} \sum w_{ii} b_i,$$

where we note the number of items in each summation is, respectively, N, T, and NT. Putting all together gives the expression for σ_q^2 =Var(q).

For the asymptotic normality of q, we note that $q = u'Wu + b'u = (\mu' W_\mu \mu + b_\mu' \mu) + (\eta' W_\eta \eta + b_\eta' \eta) + (v'Wv + b'v) + 2\mu' Z_\mu' W Z_\eta \eta + 2\mu' Z_\mu' W v + 2\eta' Z_\eta' W v$. The asymptotic normality of the first three bracketed terms follow from the CLT for linear-quadratic forms of vector of i.i.d. elements given in Kelejian and Prucha (2001). The asymptotic normality of the last three terms can easily be proved using the fact that the two random vectors involved in each term are independent. *Q.E.D.*

Proof of Theorem 2: Let $G^{\dagger}(\psi) = CG(\psi)$, where $G(\psi) \equiv \partial\ell(\psi)/\partial\psi$ is the gradient function containing the following elements.

$$
\begin{aligned}
G_{\beta}(\psi) &= \tfrac{1}{\sigma_v^2}X'(\lambda)\Omega^{-1}u \\
G_{\sigma_v^2}(\psi) &= \tfrac{1}{2\sigma_v^4}u'\Omega^{-1}u - \tfrac{NT}{2\sigma_v^2} \\
G_{\phi_\mu}(\psi) &= \tfrac{1}{2\sigma_v^2}u'A_\mu u - \tfrac{1}{2}T(N-1)\theta_1 - \tfrac{1}{2}T\theta_3 \\
G_{\phi_\eta}(\psi) &= \tfrac{1}{2\sigma_v^2}u'A_\eta u - \tfrac{1}{2}N(T-1)\theta_2 - \tfrac{1}{2}N\theta_3 \\
G_{\lambda}(\psi) &= J_\lambda(\lambda) - \tfrac{1}{\sigma_v^2}u_\lambda'\Omega^{-1}u
\end{aligned}
$$

where $A_\mu \equiv -\frac{\partial}{\partial\phi_\mu}\Omega^{-1} = T(\theta_1^2P_1+\theta_3^2P_3)$, and $A_\eta \equiv -\frac{\partial}{\partial\phi_\eta}\Omega^{-1} = N(\theta_2^2P_2+\theta_3^2P_3)$. The proof of the theorem starts from a Taylor expansion of $G^{\dagger}(\hat{\psi})$ around ψ_0:

$$
0 = \frac{1}{\sqrt{NT}}G^{\dagger}(\hat{\psi}) = \frac{1}{\sqrt{NT}}G^{\dagger}(\psi_0) + \left(\frac{1}{NT}CH(\bar{\psi})C\right)\sqrt{NT}C^{-1}(\hat{\psi}-\psi_0)
$$

where $\bar{\psi}$ lies between $\hat{\psi}$ and ψ_0, and $H(\psi)$ is the Hessian matrix containing the following elements

$$
\begin{aligned}
H_{\beta\beta} &= -\tfrac{1}{\sigma_v^2}X'(\lambda)\Omega^{-1}X(\lambda) \\
H_{\beta\sigma_v^2} &= -\tfrac{1}{\sigma_v^4}X'(\lambda)\Omega^{-1}u \\
H_{\beta\phi_\mu} &= -\tfrac{1}{\sigma_v^2}X'(\lambda)A_\mu u \\
H_{\beta\phi_\eta} &= -\tfrac{1}{\sigma_v^2}X'(\lambda)A_\eta u \\
H_{\beta\lambda} &= \tfrac{1}{\sigma_v^2}[X_\lambda'(\lambda)\Omega^{-1}u + X'(\lambda)\Omega^{-1}u_\lambda] \\
H_{\sigma_v^2\sigma_v^2} &= \tfrac{NT}{2\sigma_v^4} - \tfrac{1}{\sigma_v^6}u'\Omega^{-1}u \\
H_{\sigma_v^2\phi_\mu} &= -\tfrac{1}{2\sigma_v^4}u'A_\mu u \\
H_{\sigma_v^2\phi_\eta} &= -\tfrac{1}{2\sigma_v^4}u'A_\eta u \\
H_{\sigma_v^2\lambda} &= \tfrac{1}{\sigma_v^4}u_\lambda'\Omega^{-1}u \\
H_{\phi_\mu\phi_\mu} &= \tfrac{1}{2}T^2((N-1)\theta_1^2+\theta_3^2) - \tfrac{1}{2\sigma_v^2}u'A_{\mu\mu}u \\
H_{\phi_\mu\phi_\eta} &= \tfrac{1}{2}NT\theta_3^2 - \tfrac{1}{2\sigma_v^2}u'A_{\mu\eta}u \\
H_{\phi_\mu\lambda} &= -\tfrac{1}{\sigma_v^2}u_\lambda'A_\mu u \\
H_{\phi_\eta\phi_\eta} &= \tfrac{1}{2}N^2((T-1)\theta_2^2+\theta_3^2) - \tfrac{1}{2\sigma_v^2}u'A_{\eta\eta}u \\
H_{\phi_\eta\lambda} &= -\tfrac{1}{\sigma_v^2}u_\lambda'A_\eta u \\
H_{\lambda\lambda} &= -\tfrac{1}{\sigma_v^2}(u_{\lambda\lambda}'\Omega^{-1}u + u_\lambda'\Omega^{-1}u_\lambda) + J_{\lambda\lambda}(\lambda).
\end{aligned}
$$

where $A_{\mu\mu} = \frac{\partial^2}{\partial\phi_\mu^2}\Omega^{-1} = 2T^2(\theta_1^3P_1+\theta_3^3P_3)$, $A_{\mu\eta} = \frac{\partial^2}{\partial\phi_\mu\partial\phi_\eta}\Omega^{-1} = 2NT\theta_3^3P_3$, and $A_{\eta\eta} = \frac{\partial^2}{\partial\phi_\eta^2}\Omega^{-1} = 2N^2(\theta_2^3P_2+\theta_3^3P_3)$.

The result of the theorem follows from the following three results:

$$\begin{array}{ll}(i) & \frac{1}{\sqrt{NT}}CG(\psi_0) \xrightarrow{D} N(0, K(\psi_0)) \\ (ii) & \frac{1}{NT}C\{H(\bar{\psi}) - H(\psi_0)\}C = o_p(1) \\ (iii) & \frac{1}{NT}C\{H(\psi_0) - \mathrm{E}[H(\psi_0)]\}C = o_p(1).\end{array}$$

For (i), the joint asymptotic normality of the first four elements of $G(\psi_0)$ follows from Lemma A3 and the Cramér-Wold device. The last element of $G(\psi_0)$ can be written as

$$\begin{aligned}G_\lambda(\psi_0) &= \sum_i \sum_t \frac{h_{Y\lambda}(Y_{it}, \lambda_0)}{h_Y(Y_{it}, \lambda_0)} - \frac{1}{\sigma_{v0}^2} \sum_i \sum_t (u_{\lambda,it} u_{it} + (\theta_1 - 1)\bar{u}_{\lambda i} \bar{u}_i \\ &\quad + (\theta_2 - 1)\bar{u}_{\lambda t} \bar{u}_t + (1 - \theta_1 - \theta_2 + \theta_3)\bar{\bar{u}}_\lambda \bar{\bar{u}})\end{aligned}$$

where $\bar{u}_{i\cdot}$ and $\bar{u}_{\lambda i\cdot}$ are the averages over t for a given i of $\{u_{it}(\beta_0, \lambda_0)\}$ and $\{u_{\lambda,it}(\beta_0, \lambda_0)\}$, respectively; $\bar{u}_{\cdot t}$ and $\bar{u}_{\lambda\cdot t}$ are the averages over i for a given t, and $\bar{\bar{u}}$ and $\bar{\bar{u}}_\lambda$ are the overall averages. It is easy to see that $\bar{u}_{i\cdot} = \mu_i + O_p(T^{-\frac{1}{2}})$, $\bar{u}_{\cdot t} = \eta_t + O_p(N^{-1/2})$, and $\bar{\bar{u}} = O_p(T^{-1/2}) + O_p(N^{-1/2}) + O_p((NT)^{-1/2})$. Since $\bar{u}_{\lambda i\cdot}$, $\bar{u}_{\lambda\cdot t}$, and $\bar{\bar{u}}_\lambda$ are all $O_p(1)$ by Assumption N5, it follows that the linear combinations of $G_\lambda(\psi_0)$ with other elements of $G(\psi_0)$ can be written in the form of $\sum_i \sum_t g_{it}$ where g_{it} is defined before Assumption N6. Thus, Assumption N6 and Cramér-Wold device lead to the joint asymptotic normality of $G(\psi_0)$.

What is left is to show that the normalizing factor should be adjusted by the matrix C to reflect the different rates of convergence of the components of $\hat{\psi}$. This amounts to show that $G_\beta(\psi_0)$, $G_{\sigma_v^2}(\psi_0)$ and $G_\lambda(\psi_0)$ are all $O_p(\sqrt{NT})$, but $G_{\phi_\mu}(\psi_0) = O_p(\sqrt{N})$ and $G_{\phi_\eta}(\psi_0) = O_p(\sqrt{T})$. The first three results are trivial. To prove the latter two, note that

$$\begin{aligned}u'P_1 u &= NT(s_\mu^2 + s_{v1}^2 + s_{\mu v}^2) \\ u'P_2 u &= NT(s_\eta^2 + s_{v2}^2 + s_{\eta v}^2) \\ u'P_3 u &= NT(\bar{\mu} + \bar{\eta} + \bar{\bar{v}})^2\end{aligned}$$

where $s_\mu^2 = \frac{1}{N}\sum_{i=1}^N (\mu_i - \bar{\mu})^2$, $s_{v1}^2 = \frac{1}{N}\sum_{i=1}^N (\bar{v}_{i\cdot} - \bar{\bar{v}})^2$, and $s_{\mu v}^2 = \frac{1}{N}\sum_{i=1}^N [\mu_i(\bar{v}_{i\cdot} - \bar{\bar{v}})]$; $s_\eta^2 = \frac{1}{T}\sum_{t=1}^T (\eta_i - \bar{\eta})^2$, $s_{v2}^2 = \frac{1}{T}\sum_{t=1}^T (\bar{v}_{\cdot t} - \bar{\bar{v}})^2$, and $s_{\eta v}^2 = \frac{1}{T}\sum_{t=1}^T [\eta_t(\bar{v}_{\cdot t} - \bar{\bar{v}})]$; $\bar{\mu} = \frac{1}{N}\sum_{i=1}^N \mu_i$, $\bar{\eta} = \frac{1}{T}\sum_{t=1}^T \eta_i$, $\bar{v}_{i\cdot} = \frac{1}{T}\sum_{t=1}^T v_{it}$, $\bar{v}_{\cdot t} = \frac{1}{N}\sum_{i=1}^N v_{it}$, and $\bar{\bar{v}} = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T v_{it}$. These give

$$\begin{aligned}G_{\phi_\mu}(\psi_0) &= \frac{1}{2\sigma_{v0}^2} u' A_\mu u - \frac{1}{2}T(N-1)\theta_1 - \frac{1}{2}T\theta_3 \\ &= \frac{T}{2\sigma_{v0}^2}(\theta_1^2 u' P_1 u + \theta_3^2 u' P_3 u) - \frac{1}{2}T(N-1)\theta_1 - \frac{1}{2}T\theta_3 \\ &= \frac{NT^2(s_\mu^2 + s_{v1}^2 + s_{\mu v}^2)}{2\sigma_{v0}^2(T\phi_{\mu 0} + 1)^2} + \frac{NT^2(\bar{\mu} + \bar{\eta} + \bar{\bar{v}})^2}{2\sigma_{v0}^2(T\phi_{\mu 0} + N\phi_{\eta 0} + 1)^2} \\ &\quad - \frac{(N-1)T}{2(T\phi_{\mu 0} + 1)} - \frac{T}{2(T\phi_{\mu 0} + N\phi_{\eta 0} + 1)}\end{aligned}$$

Under Assumption C1, we have, as $N \to \infty$, $\bar{\mu} \xrightarrow{p} 0$, $\bar{\bar{v}} \xrightarrow{p} 0$, $\bar{\eta} \xrightarrow{p} 0$ if $T \to \infty$ as well, but otherwise does not converge, $s_\mu^2 \xrightarrow{p} \sigma_{\mu 0}^2$, $s_{v1}^2 \xrightarrow{p} \sigma_{v0}^2/T$, $s_{\mu v}^2 \xrightarrow{p} 0$. These give for N large

$$G_{\phi_\mu}(\psi_0) \approx \frac{\sqrt{N}}{2(\phi_{\mu 0} + 1/T)} \sqrt{N} \left(\frac{s_\mu^2 + s_{v1}^2}{\sigma_{\mu 0}^2 + \sigma_{v0}^2/T} - 1 \right)$$

Clearly, the term $\sqrt{N}((s_\mu^2 + s_{v1}^2)/(\sigma_{\mu 0}^2 + \sigma_{v0}^2/T) - 1)$ is $O_p(1)$ as N approaches infinity irrespective of whether T being fixed or approaching to infinity, and hence $G_{\phi_\mu}(\psi_0) = O_p(\sqrt{N})$ as N approaches to infinity irrespective of whether T being fixed or approaching to infinity. Similarly, one can show that as $T \longrightarrow \infty$,

$$G_{\phi_\eta}(\psi_0) \approx \frac{\sqrt{T}}{2(\phi_{\eta 0} + 1/N)} \sqrt{T} \left(\frac{s_\eta^2 + s_{v2}^2}{\sigma_{\eta 0}^2 + \sigma_{v0}^2/N} - 1 \right)$$

showing that it is $O_p(\sqrt{T})$, irrespective of weather N being fixed or approaching to infinity.

To show (ii) $\frac{1}{NT} CH(\bar{\psi})C - \frac{1}{NT} CH(\psi_0)C = o_p(1)$, Note that $\hat{\psi} \xrightarrow{p} 0$ implies $\bar{\psi} \xrightarrow{p} 0$. All parameters or their one-to-one functions, except λ, appear in $H(\psi)$ additively or multiplicatively. The parameter λ appears in $H(\psi)$ through either continuous non-stochastic functions $X(\lambda)$ and its derivatives up to second order, or stochastic functions $h(Y, \lambda)$ and its partial derivatives up to third order. Hence, it suffices to show the following:

(a) $\frac{1}{NT}[X'(\bar{\lambda})WX(\bar{\lambda}) - X'(\lambda_0)WX(\lambda_0)] = o_p(1)$, for $W = I_{NT}, P_1, P_2, P_3$, with the same being true when $X(\lambda)$ is replaced by its first and second derivatives with respect to λ;

(b) $\frac{1}{NT}[h'(Y, \bar{\lambda})WX(\bar{\lambda}) - h'(Y, \lambda_0)WX(\lambda_0)] = o_p(1)$, for $W = I_{NT}, P_1, P_2, P_3$, with the same being true when $X(\lambda)$ or $h(Y, \lambda)$ is replaced by its first and second derivatives with respect to λ;

(c) $\frac{1}{NT}[h'(Y, \bar{\lambda})Wh(Y, \bar{\lambda}) - h'(Y, \lambda_0)Wh(Y, \lambda_0)] = o_p(1)$, for $W = I_{NT}, P_1, P_2, P_3$, with the same being true when $h(Y, \lambda)$ is replaced by its first and second derivatives with respect to λ; and

(d) $\frac{1}{NT}[J_{\lambda\lambda}(\bar{\lambda}) - J_{\lambda\lambda}(\lambda_0)] = o_p(1)$.

To show (a), by the mean value theorem,

$$\frac{1}{NT}[X'(\bar{\lambda})WX(\bar{\lambda}) - X'(\lambda_0)WX(\lambda_0)] = \frac{2}{NT} X_\lambda'(\tilde{\lambda})WX(\tilde{\lambda})(\bar{\lambda} - \lambda_0)$$

where $\tilde{\lambda}$ lies between $\bar{\lambda}$ and λ_0. Let ι_i denotes a $k \times 1$ vector with 1 in the ith place and 0 elsewhere. Then by the fact that W is a projection matrix, the Cauchy-Schwarz inequality, and Assumption N4

$$
\begin{aligned}
& \frac{1}{NT}\left|\iota_i' X_\lambda'(\tilde{\lambda}) W X(\tilde{\lambda})\iota_j\right| \\
\leq\ & \left\{\frac{1}{NT}\iota_i' X_\lambda'(\tilde{\lambda}) W X_\lambda'(\tilde{\lambda})\iota_i\right\}^{1/2}\left\{\frac{1}{NT}\iota_j' X'(\tilde{\lambda}) W X(\tilde{\lambda})\iota_j\right\}^{\frac{1}{2}} \\
\leq\ & \gamma_{\max}(W)\frac{1}{\sqrt{NT}}\left\|\iota_i' X_\lambda'(\tilde{\lambda})\right\|\frac{1}{\sqrt{NT}}\left\|\iota_j' X'(\tilde{\lambda})\right\| \\
\leq\ & \sup_{\lambda\in N_\epsilon(\lambda_0)}\frac{1}{\sqrt{NT}}\|X_\lambda(\lambda)\|\sup_{\lambda\in N_\epsilon(\lambda_0)}\frac{1}{\sqrt{NT}}\|X(\lambda)\| \\
=\ & O(1)O(1)=O(1).
\end{aligned}
$$

It follows that $\frac{1}{NT}[X'(\bar{\lambda})WX(\bar{\lambda})-X'(\lambda_0)WX(\lambda_0)]=o_p(1)$ as $\bar{\lambda}-\lambda_0=o_p(1)$, and thus the first part of (a) follows. Noting that $\frac{1}{\sqrt{NT}}\|X_\lambda(\lambda)\|$, $\frac{1}{\sqrt{NT}}\|X_{\lambda\lambda}(\lambda)\|$, and $\frac{1}{\sqrt{NT}}\|X_{\lambda\lambda\lambda}(\lambda)\|$ are $O(1)$ uniformly in the ϵ-neighborhood of λ_0 by Assumption N4, the other parts of (a) follow by similar arguments. Analogously, one can show that the results (b)-(d) follow.

Finally, to show (iii) $\frac{1}{NT}C\left\{H(\psi_0)-\mathrm{E}[H(\psi_0)]\right\}C=o_p(1)$, it is straightforward to handle the terms which are linear or quadratic forms of u_0, i.e., $\frac{1}{NT}X_*'(\lambda_0)Wu_0=o_p(1)$ and $\frac{1}{NT}[u_0'Wu_0-\mathrm{E}(u_0'Wu_0)]=o_p(1)$, for $W=I_{NT},P_1,P_2,P_3$, and $X_*'(\lambda_0)=X(\lambda_0),X_\lambda(\lambda_0)$, and $X_{\lambda\lambda}(\lambda_0)$. For other items, Assumption N5 implies that

(a) $\frac{1}{NT}X'(\lambda_0)W[h_\lambda(Y,\lambda_0)-\mathrm{E}(h_\lambda(Y,\lambda_0))]=o_p(1)$, for $W=I_{NT}P_1,P_2,P_3$, with the same being true when $h_\lambda(Y,\lambda_0)$ is replaced by $h_{\lambda\lambda}(Y,\lambda_0)$;

(b) $\frac{1}{NT}\left\{h_\lambda'(Y,\lambda_0)Wh(Y,\lambda_0)-\mathrm{E}[h_\lambda'(Y,\lambda_0)Wh(Y,\lambda_0)]\right\}=o_p(1)$, for $W=I_{NT},P_1,P_2,P_3$, with the same being true when $h(Y,\lambda_0)$ is replaced by $h_\lambda(Y,\lambda_0)$ or $h_\lambda(Y,\lambda_0)$ is replaced by $h_{\lambda\lambda}(Y,\lambda_0)$.

Finally, Assumption 5N(iii) states $\frac{1}{NT}\left\{J_{\lambda\lambda}(\lambda_0)-\mathrm{E}[J_{\lambda\lambda}(\lambda_0)]\right\}=o_p(1)$. This completes the proof of (iii) and thus the proof of Theorem 2.

References

[1] Abrevaya, J. (1999). Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics*, **93**, 203-228.

[2] Andrews, D. W. K. (1987). Consistency in nonlinear econometric models: a generic uniform law of large numbers. *Econometrica*, **55**, 1465-1471.

[3] Andrews, D. W. K. (1992). Generic Uniform Convergence. *Econometric Theory*, **8**, 241-257.

[4] Arellano, M. (2003). *Panel Data Econometrics.* Oxford: Oxford University Press.

[5] Baltagi, B. H. (1997). Testing linear and log-linear error components regression against Box-Cox alternatives. *Statistics and Probability Letters*, **33**, 63-68.

[6] Baltagi, B. H. (2001). *Econometric Analysis of Panel Data.* New York: John Wiley & Sons, Ltd.

[7] Baltagi, B. H. and Li, Q. (1992). A monotonic property for iterative GLS in the two-way random effects model. *Journal of Econometrics*, **53**, 45-51.

[8] Box, G.E.P. and Cox, D.R. (1964). An Analysis of Transformations (with discussion). *Journal of the Royal Statistical Society*, **B 26**, 211-46.

[9] Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford: Oxford University Press.

[10] Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics.* Oxford: Oxford University Press.

[11] Frees, E. W. (2004). *Longitudinal and Panel Data.* Cambridge: Cambridge University Press.

[12] Giannakas, K., Tran, K. C., and Tzouvelekas, V. (2003). On the choice of functional form in stochastic frontier modeling. *Empirical Economics*, **28**, 75-100.

[13] Greene, W. H. (2000). *Econometric Analysis, 4th ed.* Singapore: Prentice-Hall Pte Ltd.

[14] Hernandze, F. and Johnson, R. A. (1980). The large-sample behavior of transformations to normality. *Journal of The American Statistical Association* **75**, 855-861.

[15] Hinkley, D. V. (1975). On power transformations to symmetry. *Biometrika*, **62**, 101-111.

[16] Hsiao, C. (2003). *Analysis of Panel Data.* Cambridge: Cambridge University Press.

[17] Hurvich, C. M. and Zeger, S. (1987). Frequency domain bootstrap methods for time series. Working Paper, Statistics and Operations Research, New York University, New York.

[18] Yeo, I. K. and Johnson, R.A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, **87**, 954-959.

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

[20] Lahiri, S. N. (2003). *Resampling Methods for Dependent Data.* New York: Springer-Verlag.

[21] McLeish, D. L. (1975). A maximal inequality and dependent strong laws. *Annals of probability* **3**, 829-839.

[22] Pötscher, B. M. and Prucha, I. R. (1989). A uniform law of large numbers for dependent and heterogeneous data process. *Econometrica* **57**, 675-683.

[23] White, H. (1994). *Estimation, Inference and Specification Analysis.* New York: Cambridge University Press.

[24] Yang, Z. L. (1999). Estimating a transformation and its effect on Box-Cox T-ratio. *Test*, **8**, 167-190.

[25] Yang, Z. L. and Huang, J. H. (2004). A transformed random effects model with applications. *Working Paper*, School of Economics, Singapore Management University, Singapore.

[26] Yang, Z. L. and Tsui, A. K. (2004). Analytically calibrated Box-Cox percentile limits for duration and event-time models. *Insurance: Mathematics and Economics* **35**, 649-677.

[27] Yang, Z. L. (2006). A modified family of power transformations. *Economics Letters* **92**, 14-19.

[28] Yang, Z. L. and Tse, Y. K. (2007). A corrected plug-in method for quantile interval construction through a transformed regression. *Journal of Business and Economic Statistics* **25**, 356-376.

Table 1a. Monte Carlo Results for Bias and RMSE: $T = Ceiling(N/3)$

(N, T)		(10, 4)		(30, 10)		(90, 30)		(200, 67)	
DGP	ψ	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE
1	20	1.4850	4.2025	0.1518	1.1387	0.0182	0.3640	0.0019	0.1734
	5	15.4692	3.4779	0.9954	0.6977	0.1178	0.2130	0.0211	0.0975
	1	16.4202	0.7732	1.0627	0.1556	0.1114	0.0463	0.0219	0.0211
	1	63.3115	2.5958	2.7896	0.2986	0.2773	0.0893	0.0727	0.0406
	.25	16.8834	0.3256	0.5225	0.0975	0.0573	0.0430	-0.0609	0.0267
	.25	-6.5971	0.2874	-7.1974	0.1282	-2.6440	0.0680	-1.2370	0.0437
	.10	-0.5324	0.0376	0.0250	0.0096	0.0188	0.0030	0.0020	0.0014
2	20	1.1845	4.6417	0.4475	1.6176	0.2634	0.5492	0.2531	0.2679
	5	21.1295	5.1422	2.3531	1.0404	0.7842	0.3376	0.6521	0.1619
	1	23.2728	1.1711	2.4264	0.2220	0.7495	0.0698	0.6495	0.0332
	1	191.1191	15.3145	8.0796	0.6139	1.8845	0.1637	1.4040	0.0764
	.25	59.6726	0.4670	6.5192	0.1224	0.2663	0.0476	0.1467	0.0283
	.25	26.9171	0.4165	-1.1966	0.1484	-2.4037	0.0702	-0.9441	0.0460
	.10	-1.4870	0.0433	0.2615	0.0137	0.3927	0.0046	0.4178	0.0022
3	20	1.1752	3.9537	0.6544	1.1921	-0.6135	0.4035	-0.5213	0.2097
	5	13.0648	3.2485	2.2002	0.7352	-1.4375	0.2344	-1.2778	0.1194
	1	14.0831	0.7210	2.3692	0.1650	-1.4595	0.0498	-1.2505	0.0251
	1	32.9719	2.1937	5.7527	0.3273	-3.4043	0.0962	-2.3534	0.0475
	.25	34.7490	0.3634	0.2941	0.1001	0.6344	0.0439	-0.2369	0.0267
	.25	9.2933	0.3244	-6.9805	0.1283	-2.0470	0.0677	-1.5639	0.0438
	.10	-0.7973	0.0353	0.8214	0.0099	-1.0804	0.0033	-0.9076	0.0017

Table 1b. Monte Carlo Results for Bias and RMSE: $N = Ceiling(T/3)$

(N, T)		(4, 10)		(10, 30)		(30, 90)		(67, 200)	
DGP	ψ	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE
1	20	1.6626	3.5220	0.1212	1.1225	0.0045	0.3700	0.0110	0.1729
	5	10.0607	2.5084	0.8350	0.6959	0.0648	0.2185	0.0238	0.0974
	1	8.4445	0.5141	0.6747	0.1443	0.0534	0.0473	0.0197	0.0211
	1	31.3067	1.4900	2.4561	0.3014	0.1868	0.0918	0.0693	0.0407
	.25	-7.4761	0.2938	-6.2675	0.1277	-2.8286	0.0670	-0.9196	0.0445
	.25	16.1751	0.3416	0.2472	0.0978	-0.3819	0.0434	-0.1904	0.0266
	.10	0.4069	0.0294	-0.0705	0.0095	-0.0195	0.0030	0.0036	0.0014
2	20	1.5019	3.8552	0.4219	1.5741	0.2936	0.5645	0.2745	0.2606
	5	12.0462	3.1864	2.2684	1.0372	0.8794	0.3499	0.7002	0.1573
	1	10.4869	0.6289	2.0535	0.2051	0.8561	0.0712	0.6784	0.0323
	1	70.3777	6.8206	8.0102	0.6051	2.1802	0.1699	1.4127	0.0744
	.25	23.8538	0.4217	-1.8562	0.1453	-1.7470	0.0705	-1.0862	0.0444
	.25	51.9917	0.4556	6.6628	0.1229	0.3339	0.0475	0.1305	0.0282
	.10	-0.2634	0.0331	0.2200	0.0135	0.4430	0.0048	0.4527	0.0022
3	20	0.5578	3.4725	-0.1199	1.1146	-0.5424	0.3945	-0.3817	0.1948
	5	7.2228	2.4724	0.2162	0.6831	-1.2801	0.2303	-0.9401	0.1103
	1	5.6161	0.4817	0.1704	0.1409	-1.2560	0.0491	-0.9247	0.0235
	1	16.6609	1.4810	1.0719	0.2971	-3.5683	0.0974	-2.4932	0.0476
	.25	1.9333	0.3197	-7.9895	0.1259	-1.3736	0.0682	-0.8216	0.0441
	.25	30.5226	0.3702	0.7213	0.0999	1.1827	0.0444	0.5264	0.0269
	.10	-1.4481	0.0293	-0.4884	0.0095	-0.9676	0.0033	-0.6692	0.0016

Table 1c. Monte Carlo Results for Bias and RMSE: $T = Ceiling(N^{2/3})$

(N,T)		(10, 5)		(30, 10)		(90, 21)		(200, 35)	
DGP	ψ	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE
1	20	0.6714	3.3944	0.1004	1.1168	0.0145	0.4312	0.0168	0.2370
	5	8.0998	2.4804	0.8047	0.6947	0.0994	0.2543	0.0579	0.1358
	1	9.1545	0.5718	0.8373	0.1484	0.0868	0.0551	0.0520	0.0290
	1	30.1281	1.4476	2.5120	0.3007	0.3006	0.1068	0.1385	0.0566
	.25	11.9192	0.2711	0.4978	0.0983	-0.0399	0.0459	0.0816	0.0286
	.25	-2.3470	0.2603	-6.0608	0.1279	-3.9286	0.0800	-2.3244	0.0615
	.10	-1.1210	0.0299	-0.0933	0.0095	-0.0162	0.0035	0.0144	0.0019
2	20	1.1472	3.9027	0.4735	1.6043	0.2578	0.6505	0.2455	0.3636
	5	14.2339	3.6941	2.5092	1.0660	0.8052	0.4011	0.6723	0.2210
	1	16.1274	0.8367	2.3411	0.2161	0.7838	0.0824	0.6600	0.0449
	1	97.4621	7.0443	8.6327	0.6275	1.9647	0.1989	1.4691	0.1040
	.25	39.7601	0.3674	6.1964	0.1201	1.2364	0.0517	0.1877	0.0311
	.25	22.7846	0.3527	-1.1836	0.1496	-2.9973	0.0842	-2.1562	0.0617
	.10	-0.5890	0.0356	0.3148	0.0138	0.3402	0.0055	0.3992	0.0030
3	20	5.2050	3.7566	-0.1560	1.1518	-0.4588	0.4618	-0.3992	0.2610
	5	21.9689	3.0336	0.2115	0.7048	-1.0756	0.2690	-0.9699	0.1489
	1	23.7486	0.7016	0.2117	0.1482	-1.0567	0.0575	-0.9586	0.0315
	1	76.8265	2.6040	-2.4828	0.2953	-2.1617	0.1112	-2.1341	0.0614
	.25	11.5621	0.2701	4.2328	0.1034	0.1849	0.0469	0.1804	0.0289
	.25	-2.6763	0.2631	-3.3042	0.1315	-4.0713	0.0813	-1.9501	0.0600
	.10	6.8458	0.0312	-0.5480	0.0098	-0.8493	0.0038	-0.7074	0.0021

Table 1d. Monte Carlo Results for Bias and RMSE: $N = Ceiling(T^{2/3})$

(N,T)		(5, 10)		(10, 30)		(21, 90)		(35, 200)	
DGP	ψ	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE
1	20	0.9086	2.9491	0.1771	1.0901	-0.0200	0.4507	0.0066	0.2342
	5	6.5530	2.0498	0.9423	0.6634	0.0284	0.2655	0.0469	0.1328
	1	5.9321	0.4541	0.9262	0.1454	0.0084	0.0571	0.0394	0.0291
	1	21.1822	1.1090	2.5036	0.2854	0.1470	0.1119	0.1083	0.0557
	.25	-6.6775	0.2615	-7.5289	0.1265	-4.0155	0.0806	-2.5084	0.0607
	.25	4.0118	0.2699	-0.1795	0.0971	-0.0217	0.0451	0.0496	0.0281
	.10	-0.2126	0.0251	0.0524	0.0091	-0.0795	0.0037	0.0079	0.0018
2	20	1.3733	3.4228	0.2871	1.4986	0.2388	0.6870	0.2433	0.3522
	5	10.4854	2.7818	1.6693	0.9516	0.8098	0.4254	0.6463	0.2139
	1	9.9401	0.5962	1.7204	0.1985	0.7852	0.0869	0.6374	0.0443
	1	55.8167	3.8679	5.7291	0.5497	1.9901	0.2046	1.5275	0.1021
	.25	19.8571	0.3558	-1.9789	0.1465	-3.3260	0.0829	-2.1208	0.0615
	.25	35.3587	0.3674	6.4728	0.1216	0.9734	0.0519	0.0544	0.0307
	.10	0.1624	0.0299	-0.0061	0.0127	0.3165	0.0058	0.3859	0.0029
3	20	0.2111	2.8711	0.0570	1.1180	-0.0357	0.4693	-0.4332	0.2603
	5	4.5863	1.9592	0.6030	0.6724	0.0096	0.2755	-1.0502	0.1480
	1	4.1735	0.4378	0.5002	0.1459	0.0412	0.0582	-1.0358	0.0317
	1	6.9426	0.9532	4.4785	0.3091	-1.3457	0.1140	-1.9337	0.0605
	.25	2.0814	0.2794	-10.0904	0.1253	-2.6570	0.0810	-2.3653	0.0605
	.25	14.4077	0.2909	-2.2718	0.0982	1.4432	0.0467	-0.1125	0.0285
	.10	-1.3220	0.0246	-0.1968	0.0093	-0.1020	0.0039	-0.7615	0.0021

Table 1e. Monte Carlo Results for Bias and RMSE: $T = 6$

(N, T)		(10, 6)		(30, 6)		(90, 6)		(200, 6)	
DGP	ψ	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE
1	20	0.5176	2.5188	0.1920	1.4818	0.0699	0.8520	0.0414	0.5945
	5	4.3508	1.6859	1.4480	0.9370	0.5334	0.5147	0.1935	0.3449
	1	4.4221	0.3974	1.6185	0.2067	0.4656	0.1125	0.1584	0.0739
	1	12.4808	0.8504	4.5732	0.4162	1.4810	0.2192	0.5530	0.1458
	.25	6.6293	0.2346	2.0671	0.1241	0.9901	0.0702	0.5473	0.0457
	.25	-4.8258	0.2287	-13.5153	0.1620	-15.5476	0.1458	-15.8451	0.1422
	.10	-0.4027	0.0214	-0.1371	0.0126	0.0032	0.0071	-0.0287	0.0048
2	20	0.8390	2.9618	0.5264	2.0239	0.2808	1.2295	0.2923	0.8783
	5	6.9884	2.2412	3.4373	1.3936	1.4094	0.7854	1.1367	0.5474
	1	7.8780	0.5155	3.5880	0.2968	1.4024	0.1623	1.0758	0.1115
	1	34.4753	2.5459	13.8585	0.9604	4.4340	0.4148	3.0105	0.2756
	.25	29.8492	0.3138	11.7330	0.1567	4.1343	0.0848	1.9740	0.0555
	.25	19.6292	0.3135	-4.0819	0.1956	-12.5116	0.1534	-15.2801	0.1457
	.10	-0.2556	0.0256	0.1047	0.0174	0.1649	0.0105	0.3879	0.0074
3	20	-0.1263	2.3799	0.0686	1.4659	-1.0046	0.9072	-0.9829	0.6624
	5	2.4054	1.5447	1.1418	0.9208	-2.1759	0.5324	-2.2858	0.3789
	1	2.4613	0.3618	1.1868	0.2057	-2.1201	0.1144	-2.2234	0.0800
	1	-2.4846	0.6659	3.7739	0.4198	-1.1348	0.2230	-1.4319	0.1515
	.25	15.5504	0.2486	1.9605	0.1258	-2.2069	0.0698	-2.3133	0.0467
	.25	4.8760	0.2470	-13.3253	0.1637	-17.1744	0.1430	-17.9169	0.1399
	.10	-1.3612	0.0204	-0.3407	0.0125	-1.9236	0.0077	-1.7920	0.0055

Table 1f. Monte Carlo Results for Bias and RMSE: $N = 6$

(N, T)		(6, 10)		(6, 30)		(6, 90)		(6, 200)	
DGP	ψ	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE
1	20	0.6010	2.5238	0.2521	1.5170	0.0873	0.8517	0.0351	0.5931
	5	4.5205	1.6876	1.6386	0.9616	0.5002	0.5173	0.1921	0.3430
	1	4.6810	0.3965	1.7997	0.2134	0.4451	0.1121	0.1510	0.0727
	1	13.1904	0.8543	5.1963	0.4276	1.4016	0.2186	0.5562	0.1454
	.25	-4.5043	0.2280	-12.9882	0.1609	-15.2029	0.1443	-16.1954	0.1424
	.25	2.1429	0.2275	1.4080	0.1228	1.0595	0.0690	0.5436	0.0461
	.10	-0.2702	0.0214	-0.0708	0.0130	-0.0196	0.0071	-0.0278	0.0047
2	20	0.7508	2.9542	0.6402	2.0800	0.3092	1.2393	0.3050	0.8668
	5	6.6834	2.1923	3.8620	1.4425	1.4862	0.7915	1.1148	0.5368
	1	7.4786	0.5041	4.1143	0.3096	1.4786	0.1626	1.0406	0.1089
	1	32.3731	2.3497	15.2978	1.0088	4.5079	0.4162	2.9287	0.2717
	.25	19.4406	0.3194	-4.4418	0.1918	-12.4435	0.1544	-14.4436	0.1449
	.25	30.2216	0.3201	11.9166	0.1577	4.3162	0.0843	2.2850	0.0560
	.10	-0.4123	0.0256	0.2626	0.0179	0.2051	0.0106	0.3868	0.0073
3	20	-0.1466	2.3923	0.0867	1.5013	-1.0186	0.9097	-0.9925	0.6549
	5	2.4299	1.5598	1.2234	0.9491	-2.2174	0.5348	-2.2764	0.3783
	1	2.3144	0.3604	1.2935	0.2103	-2.1578	0.1155	-2.2034	0.0795
	1	-1.8531	0.6859	4.1209	0.4329	-1.2705	0.2229	-1.4183	0.1513
	.25	3.8823	0.2450	-12.1404	0.1663	-17.3942	0.1429	-18.7594	0.1400
	.25	13.7751	0.2491	2.2762	0.1269	-1.8197	0.0700	-2.4326	0.0467
	.10	-1.3970	0.0205	-0.3495	0.0129	-1.9545	0.0077	-1.7853	0.0054

Table 2. Empirical Coverage Probabilities for 95% Confidence Intervals

(N, T)	(25,25)		(50, 25)		(50, 50)		(100, 50)		(100, 100)	
ψ	Hess	Boot	Hess	Boot	Hess	Boot	Hess	Boot	Hess	Boot
					Normal Errors					
20	.9468	.9382	.9456	.9354	.9454	.9406	.9502	.9450	.9490	.9482
5	.9480	.9366	.9438	.9352	.9496	.9426	.9510	.9444	.9512	.9466
1	.9490	.9404	.9442	.9330	.9472	.9386	.9534	.9512	.9502	.9462
1	.9390	.9276	.9372	.9294	.9486	.9404	.9482	.9418	.9512	.9496
.25	.8984	.9128	.9352	.9484	.9266	.9288	.9324	.9434	.9370	.9352
.25	.8988	.9144	.8938	.8868	.9216	.9260	.9238	.9148	.9416	.9376
.10	.9472	.9370	.9476	.9370	.9484	.9402	.9516	.9472	.9502	.9466
					Normal-Mixture, $p = .05, \tau = 5$					
20	.8322	.9200	.8054	.9210	.8114	.9332	.8000	.9318	.7966	.9446
5	.8260	.9146	.7994	.9200	.7986	.9366	.7848	.9390	.7702	.9434
1	.8402	.9164	.8200	.9206	.8212	.9362	.8060	.9334	.7924	.9436
1	.7542	.8828	.7388	.9068	.7500	.9260	.7294	.9360	.7168	.9384
.25	.8780	.9050	.9112	.9410	.9116	.9270	.9282	.9498	.9298	.9368
.25	.8924	.9206	.8860	.8928	.9240	.9356	.9142	.9124	.9292	.9366
.10	.8228	.9140	.7950	.9172	.7970	.9332	.7834	.9354	.7678	.9412
					Normal-Mixture, $p = .10, \tau = 5$					
20	.8402	.9206	.8382	.9316	.8410	.9362	.8368	.9408	.8470	.9442
5	.8360	.9166	.8348	.9250	.8246	.9318	.8184	.9412	.8242	.9458
1	.8488	.9186	.8512	.9320	.8472	.9352	.8410	.9378	.8468	.9438
1	.7804	.8922	.7838	.9156	.7640	.9270	.7760	.9390	.7720	.9470
.25	.8930	.9168	.9126	.9388	.9160	.9274	.9302	.9494	.9260	.9322
.25	.8946	.9174	.8932	.8972	.9180	.9292	.9168	.9208	.9326	.9400
.10	.8362	.9190	.8326	.9240	.8234	.9334	.8152	.9398	.8238	.9446
					Normal-Mixture, $p = .05, \tau = 10$					
20	.6904	.8916	.6770	.9096	.6638	.9244	.6612	.9316	.6474	.9352
5	.6782	.8940	.6608	.9092	.6444	.9256	.6358	.9318	.6070	.9378
1	.7056	.8988	.6928	.9140	.6756	.9270	.6664	.9342	.6492	.9396
1	.5902	.8464	.5778	.8840	.5750	.9186	.5642	.9316	.5578	.9456
.25	.8700	.9130	.8756	.9370	.8958	.9262	.8912	.9326	.9220	.9398
.25	.8616	.9068	.8802	.8994	.9044	.9346	.9080	.9188	.9212	.9404
.10	.6760	.8896	.6606	.9108	.6418	.9222	.6332	.9242	.6058	.9340
					Normal-Gamma Mixture, $p = .05, \alpha = 9$					
20	.8992	.9226	.8864	.9216	.9162	.9494	.8394	.8976	.8612	.9146
5	.8958	.9132	.8812	.9084	.9208	.9546	.8270	.8900	.8510	.9076
1	.9008	.9146	.8868	.9148	.9250	.9520	.8436	.8958	.8610	.9118
1	.8704	.8910	.8652	.8956	.9052	.9402	.8532	.9008	.8894	.9310
.25	.8920	.9110	.9186	.9406	.9212	.9254	.9232	.9392	.9260	.9274
.25	.8958	.9126	.9008	.8986	.9232	.9296	.9138	.9066	.9150	.9184
.10	.9016	.9214	.8876	.9214	.9194	.9534	.8378	.9010	.8524	.9152