

# Asymptotics and Bootstrap for Random-Effects Panel Data Transformation Models\*

Liangjun Su and Zhenlin Yang<sup>†</sup>

*School of Economics, Singapore Management University*

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

October, 2015

## Abstract

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

**Key Words:** Asymptotics; Error components bootstrap; Quasi-MLE; Transformed panels; Random-effects; Robust VC matrix estimation.

**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 the parameter estimates in terms of consistency, asymptotic normality and robustness against heavy-tailed

---

\*We would like to thank Peter C. B. Phillips, Myoung-Jae Lee, two anonymous referees, and the Editor Esfandiar Maasoumi for their helpful comments that have led to improvements in the paper. Su gratefully acknowledges the support from an academic research fund (Grant number: MOE2012-T2-2-021) from the Ministry of Education, Singapore. Yang gratefully acknowledges the support from a research grant (Grant number: C244/MSS7E005) from the Office of Research, Singapore Management University.

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

distributions, variance-covariance matrix estimation in the situations where transformation can only bring the data to near-normality, etc., have not been formally studied, in particular for the models with random-effects.<sup>1</sup>

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

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

where  $h(\cdot, \lambda)$  and  $g(\cdot, \rho_j)$  are the monotonic transformations (e.g., Box and Cox, 1964), known except the indexing parameters  $\lambda$  and  $\{\rho_j\}$ , called the *transformation parameters*,  $X_{itj}$ ,  $j = 1, \dots, k_1$ , are the exogenous variables containing a column of ones, dummy variables, etc., that do not need to be transformed,  $X_{itj}$ ,  $j = k_1+1, \dots, k$ , are the exogenous variables that need to be transformed, and  $\{\mu_i\}$ ,  $\{\eta_t\}$  and  $\{v_{it}\}$  are the error components representing, respectively, the individual-specific effects, the time-specific effects and the pure random errors, assumed to be independent of each other, and each containing elements that are independent and identically distributed (i.i.d.) of means zero and variances  $\sigma_\mu^2$ ,  $\sigma_\eta^2$  and  $\sigma_v^2$ , respectively. In the following we will assume that the regressors  $X_{itj}$ ,  $j = 1, \dots, k$ , are either non-stochastic or stochastic but independent of the errors. In the latter case, our analysis can be interpreted as being conditional on the stochastic regressors.

Yang and Huang (2011) considered the maximum likelihood estimation (MLE) of Model (1) under Gaussian distributions with  $g = h$  and  $\rho_j = \lambda$ , and provided a simple method for handling the large panel data. As indicated in Yang and Huang (2011), Model (1) gives a useful extension of the standard error components model by allowing the distribution of  $Y_{it}$  to be in a broad 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. Interesting examples for which Model (1), or an extended version of it, can be useful include the *public capital productivity* (Baltagi, 2001, Ch. 3) and the *wage distribution of U.S. male workers* (Polacheck and Yoon, 1996), where strong evidence was found by Yang and Huang (2011) for a power transformation rather than linear or log-linear form. See, e.g., Baltagi (2001, Ch. 2 and Ch. 3) and Cameron and Trivedi (2005) for more examples.

The Monte Carlo results of Yang and Huang (2011) show that the finite sample performance

---

<sup>1</sup>While the fixed-effects models have the attraction of allowing one to use panel data to establish causation under weaker assumptions, they do suffer from several practical weaknesses for being unable to estimate the effects of time-invariant regressors, imprecise in estimating the effects in time-varying regressors of which the variation in time-dimension is small, etc. For these reasons economists also use random-effects models in particular when causation is clear (Cameron and Trivedi, 2005, Ch. 21). Panel data transformation models with random-effects are typically treated parametrically, see, e.g., Baltagi (1997), Giannakas *et al.* (2003), and Yang and Huang (2011); whereas those with fixed-effects are typically estimated semi-parametrically, see, e.g., Abrevaya (1999a, 2000), and Chen (2002, 2010).

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 nonnormal (e.g., there exist gross errors or outliers).<sup>2</sup> Thus, there is a need for an alternative method to 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 nonnormal 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, making the standard large sample techniques not directly applicable.<sup>3</sup> This paper fills in these gaps. Model (1) can be further extended to include heteroskedasticity in  $\{\mu_i\}$  and serial correlation in  $\{\eta_t\}$ . It can also be simplified by letting  $g = h$  and  $\rho_j = \lambda$ , or by dropping the time-specific effects  $\{\eta_t\}$ .

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 the 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 convention. The Euclidean norm of a matrix  $A$  is denoted by  $\|A\| = [\text{tr}(AA')]^{1/2}$ . When  $A$  is a real symmetric 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 infinity concurrently is denoted by  $N, T \rightarrow \infty$ , and that either  $N$  or  $T$  or both approach infinity is denoted by  $N \cup T \rightarrow \infty$ . 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.  $(Y_{it}, \lambda)$ ,  $h_{\lambda\lambda}(Y_{it}, \lambda)$  the second-order partial derivative of  $h$  w.r.t.  $\lambda$ , etc.

---

<sup>2</sup>Transformation aims to induce (i) normality, (ii) flexible functional form, (iii) homoskedastic errors, and (iv) simple model structure (Box and Cox, 1964). 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), and Yang and Tse (2007, 2008)). In this sense, the use of quasi-maximum likelihood method provides an additional protection when the exact normality is not achieved.

<sup>3</sup>In contrast, the transformed cross-sectional model does not suffer from the dependence problem. Almost all the standard econometrics/statistics textbooks cover this topic (e.g., Davidson and MacKinnon, 1993; Greene, 2000; Draper and Smith, 1998; Cook and Weisberg, 1999), and some popular commercial softwares, such as SAS and Matlab, have implemented the normal-transformation technique.

## 2 Quasi Maximum Likelihood Estimation

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

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

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 quasi-Gaussian loglikelihood function after dropping the constant term takes the form

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

where  $\psi = (\beta', \sigma_\mu^2, \sigma_\eta^2, \sigma_v^2, \lambda, \rho)'$ , 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  $\Sigma = \text{Var}(u) = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\eta^2(J_N \otimes I_T) + \sigma_v^2(I_N \otimes I_T)$ .

When the error components  $\mu$ ,  $\eta$  and  $v$  are exactly normal, (3) gives the exact loglikelihood and thus maximizing  $\ell(\psi)$  gives the maximum likelihood estimator (MLE) of  $\psi$ . 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 loglikelihood function. Nevertheless, when  $\ell(\psi)$  satisfies certain conditions, maximizing it still gives consistent estimators of model parameters, which are often termed as QML estimator (QMLE). See, e.g., White (1994). Furthermore, as pointed out in the introduction, one of the aims of transformation is to bring the data to near-normality. In case that the exact normality is not achieved, the QML method provides an extra protection against the ‘left-over’ nonnormality.

Yang and Huang (2011) pointed out that direct maximization of  $\ell(\psi)$  may be impractical as the dimension of  $\psi$  may be high and calculation of  $|\Sigma|$  and  $\Sigma^{-1}$  can be difficult if panels are large. Following Baltagi and Li (1992) and others, they 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

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

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

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

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  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  defined above are often used for convenience.

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

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

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

resulting the concentrated quasi loglikelihood for  $\phi$  as

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

Maximizing  $\ell_{\max}(\phi)$  gives the QMLE  $\hat{\phi}$  of  $\phi$ , and hence the QMLEs  $\hat{\beta}(\hat{\phi})$  and  $\hat{\sigma}_v^2(\hat{\phi})$  of  $\beta$  and  $\sigma_v^2$ , respectively. Yang and Huang (2011) 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

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. To ease the exposition, formal results are proved without loss of generality under the simpler model with  $g = h$ ,  $\rho_j = \lambda$ , and  $\phi = (\phi_\mu, \phi_\eta, \lambda)'$ . Let  $\Lambda$ ,  $\Phi$  and  $\Psi$  be, respectively, the parameter space for  $\lambda$ ,  $\phi$  and  $\psi$ ;  $\lambda_0$ ,  $\phi_0$  and  $\psi_0$  be the true parameter values; “E” and “Var” be expectation and variance operators corresponding to the true parameter  $\psi_0$ .

#### 3.1 Consistency

Let  $\bar{\ell}(\psi)$  be the expected loglikelihood, i.e.,  $\bar{\ell}(\psi) \equiv E[\ell(\psi)] = -\frac{NT}{2} \log(\sigma_v^2) + c(\phi_\mu, \phi_\eta) - \frac{1}{2\sigma_v^2} E \{ [h(Y, \lambda) - X(\lambda)\beta]'\Omega^{-1}[h(Y, \lambda) - X(\lambda)\beta] \} + E[J(\lambda)]$ . Given  $\phi$ ,  $\bar{\ell}(\psi)$  is maximized at

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

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

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

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

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  $\mu, \eta$ , 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  $\sigma_{v 0}^2$  respectively for  $\mu, \eta$ , and  $v$ .*

**Assumption C2:**  *$\Phi$  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  $\Phi$ .*

**Assumption C3:** *The elements of  $X(\lambda)$  are uniformly bounded, uniformly in  $\lambda \in \Lambda$ ; and  $\lim_{N, T \rightarrow \infty} \frac{1}{NT} [X'(\lambda)X(\lambda)]$  exists and is nonsingular, uniformly in  $\lambda \in \Lambda$ .*

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

**Assumption C5:** *Let  $\bar{h}_i = \frac{1}{T} \sum_{t=1}^T h(Y_{it}, \lambda)$  and  $\bar{h}_t = \frac{1}{N} \sum_{i=1}^N h(Y_{it}, \lambda)$ . As  $N \cup T \rightarrow \infty$ ,*

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

(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  $\lambda \in \Lambda$ , and*

(iii) *the row sums of the absolute values of  $\text{Var}[h(Y, \lambda)]$  are  $o(NT)$  uniformly in  $\lambda \in \Lambda$ .*

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

(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  $j = k_1 + 1, \dots, k$ ,*

(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  $\bar{\beta}(\phi)$  uniformly in  $\phi \in \Phi$ , as under Assumption C3, the limit of  $\frac{1}{NT} [X'(\lambda)\Omega^{-1}X(\lambda)]$  exists and is nonsingular uniformly in  $\phi \in \Phi$ .<sup>4</sup>

Assumption C4 ensures the uniform boundedness of  $\bar{\sigma}^2(\phi)$  and  $\frac{1}{NT} E[J(\lambda)]$ , and thus the uniform boundedness of  $\frac{1}{NT} \bar{\ell}_{\max}(\phi)$ . Assumption C5 says 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), and the dependence among  $\{h(Y_{it}, \lambda)\}$  is of smaller order of magnitude than the information available. Assumption C6 says that those sequences are well behaved uniformly for  $\lambda$  in the compact set  $\Lambda$ , which are essential for us to apply the weak uniform law of large numbers (ULLN). Alternatively, one can require those sequences satisfy certain Lipschitz condition, as specified in, say, Andrews (1987, 1992), Pötscher and Prucha (1989), and Davidson (1994, Chapter 21).

<sup>4</sup>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 eigenvalues of a projection matrix are either 0 or 1. We have by (4),  $\gamma_{\max}(\Omega^{-1}) \leq [\gamma_{\max}(Q) + \sum_{j=1}^3 \theta_j \gamma_{\max}(P_j)] \leq 4$ , because  $0 < \theta_j \leq 1$  for  $j = 1, 2, 3$ , and  $Q, P_1, P_2$ , and  $P_3$  are projection matrices. It follows that  $\gamma_{\min}(\Omega^{-1})X'(\lambda)X(\lambda) \leq X'(\lambda)\Omega^{-1}X(\lambda) \leq \gamma_{\max}(\Omega^{-1})X'(\lambda)X(\lambda) \leq 4X'(\lambda)X(\lambda)$ , where  $\gamma_{\min}(\Omega^{-1})$  is strictly positive as  $\Omega$  is positive definite. Here,  $A \geq B$  means  $A - B$  is positive semidefinite, and  $A \leq B$  is defined similarly.

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  $\lambda$  up to any order. We have the following consistency result.

**Theorem 1:** *Assume Assumptions C1-C6 hold. Assume further that (a)  $h(Y_{it}, \lambda)$  is monotonically increasing in  $Y_{it}$ , and (b)  $\bar{\ell}_{\max}(\phi)$  has a unique global maximum at  $\phi^*$  such that  $\phi^* \rightarrow \phi_0$  as  $N, T \rightarrow \infty$ . Then,  $\hat{\psi} \xrightarrow{p} \psi_0$ , as  $N, T \rightarrow \infty$ .*

The proof is relegated to Appendix. The identification uniqueness condition ( $\bar{\ell}_{\max}(\phi)$  has a unique global maximum at  $\phi_0$ ) 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 loglikelihood function  $\ell(\psi)$ . Let  $J_\lambda(\lambda)$ ,  $J_{\lambda\lambda}(\lambda)$  and  $J_{\lambda\lambda\lambda}(\lambda)$  be the first three partial derivatives of  $J(\lambda)$  w.r.t.  $\lambda$  (the  $\lambda$ -derivatives), and  $X_\lambda(\lambda)$ ,  $X_{\lambda\lambda}(\lambda)$  and  $X_{\lambda\lambda\lambda}(\lambda)$  the first three  $\lambda$ -derivatives of  $X(\lambda)$ . The  $G(\psi)$  function has the elements:

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

where  $u \equiv u(\beta, \lambda) = h(Y, \lambda) - X(\lambda)\beta$ ,  $u_\lambda = \frac{\partial}{\partial \lambda} u(\beta, \lambda)$ ,  $A_\mu \equiv -\frac{\partial}{\partial \phi_\mu} \Omega^{-1} = T(\theta_1^2 P_1 + \theta_3^2 P_3)$ , and  $A_\eta \equiv -\frac{\partial}{\partial \phi_\eta} \Omega^{-1} = N(\theta_2^2 P_2 + \theta_3^2 P_3)$ . The detailed expression of  $H(\psi)$  is given in Appendix. For the asymptotic normality, we need some further assumptions.

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

**Assumption N2:**  $\psi_0$  is an interior point of  $\Psi$ .

**Assumption N3:**  $\frac{1}{\sqrt{NT}} E[G_\lambda(\psi_0)] = o(1)$ .

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

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

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

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

**Assumption N5:** As  $N \cup T \rightarrow \infty$ ,

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

(ii)  $\frac{1}{NT}\{h'_\lambda(Y, \lambda_0)h(Y, \lambda_0) - E[h'_\lambda(Y, \lambda_0)h(Y, \lambda_0)]\} = o_p(1)$ , 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)  $\frac{1}{NT}\{J_{\lambda\lambda}(\lambda_0) - E[J_{\lambda\lambda}(\lambda_0)]\} = o_p(1)$ .

Assumptions N1-N3 are standard for quasi maximum likelihood inference. Under Assumption C1, the first four components of  $G(\psi_0)$  have mean zero, and hence the requirement,  $\frac{1}{\sqrt{NT}}E[G_\lambda(\psi_0)] = o(1)$ , becomes essential for the limiting distribution of the normalized gradient to be centered at zero. Intuitively, such a requirement is more likely to be met when the errors are more symmetric due to the nonlinearity of the  $G_\lambda(\psi_0)$  function. This intuition is indeed supported by the results presented in Footnote 6 corresponding to the Box-Cox power transformation. In fact, when both  $N$  and  $T$  are large it is more crucial that  $v_{it}$  is symmetrically distributed, or otherwise  $\sigma_{v_0}$  needs to be ‘small’ for the assumption to be met (see the discussions below Assumption N6). See also Hinkley (1975) and Yang (1999) for some related discussions and useful results. This assumption is also related to the Assumption (b) stated in Theorem 1, for the unique global maximum point of  $\bar{\ell}_{\max}(\phi)$  to converge to  $\phi_0$ . Assumptions N4 and N5 spell out conditions on the transformation function and its 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  $\lambda$ , and  $J_{\lambda\lambda}(\lambda) = J_{\lambda\lambda\lambda}(\lambda) = 0$ .

One of the key steps 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  $\phi_\eta$  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 by Kelejian and Prucha (2001). However, the component of  $G(\psi_0)$  corresponding to  $\lambda$  involves the nonlinear function  $h$  and its partial derivatives:

$$G_\lambda(\psi_0) = \sum_i \sum_t \frac{h_{Y\lambda}(Y_{it}, \lambda_0)}{h_Y(Y_{it}, \lambda_0)} - \frac{1}{\sigma_{v_0}^2} u_{0\lambda} \Omega^{-1} u_0,$$

where  $u_0 \equiv u(\beta_0, \lambda_0)$ , and  $u_{0\lambda} = u_\lambda(\beta_0, \lambda_0)$ . Moreover, the two-way error components  $\mu$  and  $\eta$  induce dependence along both the cross-sectional and time-wise directions. These render the standard limiting theorems inapplicable<sup>5</sup> and hence some high-level condition needs to be imposed. Some heuristic arguments for its plausibility follow.

**Assumption N6:**  $\frac{1}{\sqrt{NT}}G_\lambda(\psi_0) \xrightarrow{D} N(0, \tau^2)$ , and any linear combination of  $G_\lambda(\psi_0)$  with

---

<sup>5</sup>This is in contrast to the fixed-effects panel transformation models (Abrevaya, 1999a, 2000; Chen, 2002, 2010), or the cross-sectional transformation models (Abrevaya, 1999b; Shin, 2008; Honoré and Hu, 2010).

the other elements of  $G(\psi_0)$  is also asymptotically normal.

It is extremely difficult, if possible at all, to specify explicitly detailed conditions so that a version of CLT can apply to  $G_\lambda(\psi_0)$ . Given the highly nonlinear dependence of  $r(Y_{it}, \lambda_0) = \frac{h_{Y\lambda}(Y_{it}, \lambda_0)}{h_Y(Y_{it}, \lambda_0)}$  and  $h_\lambda(Y_{it}, \lambda_0)$  on the non-identically distributed dependent data, no generic CLT for dependent sequences (as in McLeish (1975)) is applicable. The following heuristic arguments help understand the nature of Assumption N6. Denote  $h(Y_{it}, \lambda_0) = \mu_{it} + u_{it}$  where  $\mu_{it}$  are the elements of  $X(\lambda_0)\beta_0$ . As  $h$  is strictly monotonic in  $Y_{it}$ , one can write  $r(Y_{it}, \lambda_0) = r(\mu_{it} + u_{it})$  and  $h_\lambda(Y_{it}, \lambda_0) = h_\lambda(\mu_{it} + u_{it})$ . If  $\sigma_{\mu 0}$ ,  $\sigma_{\eta 0}$  and  $\sigma_{v 0}$  are all small in the sense that as  $N, T \rightarrow \infty$ , they approach zero, then we have  $r(\mu_{it} + u_{it}) \approx r(\mu_{it}) + r_\mu(\mu_{it})u_{it}$  and  $h_\lambda(\mu_{it} + u_{it}) \approx h_\lambda(\mu_{it}) + h_{\lambda\mu}(\mu_{it})u_{it}$  (Bickel and Doksum, 1981; Yang, 1999).<sup>6</sup> Hence,  $G_\lambda(\psi_0)$  becomes a linear-quadratic form in  $u$ , as the other elements of  $G(\psi_0)$ , and Lemma A3 leads to Assumption N6. For generality, however, the small- $\sigma$  condition is not imposed.

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*

$$\sqrt{NT}C^{-1}(\hat{\psi} - \psi_0) \xrightarrow{D} N(0, I^{-1}(\psi_0)K(\psi_0)I^{-1}(\psi_0)), \text{ as } N, T \rightarrow \infty,$$

where  $I(\psi_0) = -\lim_{N, T \rightarrow \infty} \frac{1}{NT}CE[H(\psi_0)]C$  and  $K(\psi_0) = \lim_{N, T \rightarrow \infty} \frac{1}{NT}CE[G(\psi_0)G'(\psi_0)]C$ , both assumed to exist with  $I(\psi_0)$  being positive definite. Furthermore, if  $\mu_i$ 's,  $\eta_t$ '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  $N, T \rightarrow \infty$ .

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 the following different scenarios:

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

---

<sup>6</sup>For the Box-Cox power transformation:  $h(y, \lambda) = \frac{1}{\lambda}(y^\lambda - 1)$  if  $\lambda \neq 0$ ;  $\log y$  if  $\lambda = 0$ , we have  $r(Y_{it}, \lambda_0) = \log Y_{it} = \frac{1}{\lambda_0} \log[1 + \lambda_0 h(Y_{it}, \lambda_0)]$ , and  $h_\lambda(Y_{it}, \lambda_0) = \lambda_0^{-1}[1 + \lambda_0 h(Y_{it}, \lambda_0)] \log Y_{it} - \lambda_0^{-1} h(Y_{it}, \lambda_0)$  when  $\lambda_0 \neq 0$ ;  $\frac{1}{2}(\log Y_{it})^2$  when  $\lambda_0 = 0$ , which are clearly analytical functions of  $h(Y_{it}, \lambda_0) = \mu_{it} + u_{it}$ . As the Box-Cox power transformation has a bounded range when  $\lambda \neq 0$ , the small- $\sigma$  approximation makes it more compatible with the near-normality assumption. Under this approximation,  $\log Y_{it} \approx \frac{1}{\lambda_0} \log(1 + \lambda_0 \mu_{it}) + \theta_{it} u_{it}$ , and

$$\frac{1}{NT}E[G_\lambda(\psi_0)] \approx -\frac{1}{2}\sigma_{v0}\gamma_v\bar{\theta} + \frac{1}{2T}(\sigma_{v0}\gamma_v - \frac{T\sigma_{\mu 0}^3\gamma_\mu + \sigma_{v0}^3\gamma_v}{T\sigma_{\mu 0}^2 + \sigma_{v0}^2})\bar{\theta} + \frac{1}{2N}(\sigma_{v0}\gamma_v - \frac{N\sigma_{\eta 0}^3\gamma_\eta + \sigma_{v0}^3\gamma_v}{N\sigma_{\eta 0}^2 + \sigma_{v0}^2})\bar{\theta} + O(\frac{1}{NT}),$$

where  $\theta_{it} = (1 + \lambda_0 \mu_{it})^{-1}$ ,  $\bar{\theta} = \frac{1}{NT} \sum_i \sum_t \theta_{it}$ , and  $\gamma_\mu, \gamma_\eta$  and  $\gamma_v$  are the measures of skewness of  $\mu_i, \eta_t$ , and  $v_{it}$ , respectively. Thus, for Assumption N3 to be satisfied it is necessary that  $\sigma_{v0} = o((NT)^{-1/2})$  if  $\gamma_v \neq 0$ ,  $\sigma_{\mu 0} = o((N/T)^{-1/2})$  if  $\gamma_\mu \neq 0$ , and  $\sigma_{\eta 0} = o((T/N)^{-1/2})$  if  $\gamma_\eta \neq 0$ .

- (ii)  $\phi_\mu$  (or  $\sigma_\mu^2$ ) is  $\sqrt{N}$ -consistent under (a)-(d), but is inconsistent under (e).
- (iii)  $\phi_\eta$  (or  $\sigma_\eta^2$ ) 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  $T$  or both approach infinity. In the case where both pass 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  $\eta_t$  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  $I(\hat{\psi}) = -\frac{1}{NT}CH(\hat{\psi})C$ , but for the latter, there are no readily available methods. This is because (i)  $\text{Var}[G(\psi_0)]$  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  $\text{Var}[G(\psi_0)]$  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 forth by Lahiri (2003, p. 40), which generalizes the idea of Hurvich and Zeger (1987). The central idea can simply be stated as follows. If (a) a statistic is a function of a dependent sequence, (b) 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 (ECB)* as it directly bootstraps on the estimated error components obtained by decomposing the estimated error vector  $\hat{u} = h(Y, \hat{\lambda}) - X(\hat{\lambda})\hat{\beta}$ . Each of these estimated error components contains asymptotically independent elements, which sets up the theoretical base for the bootstrap method. Furthermore and very importantly,  $G(\psi_0)$  can be written as an analytical function of  $u_0$  and  $\psi_0$  (see details below). 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 steps 1-3  $B$  times to give  $G_1^*(\hat{\psi}), G_2^*(\hat{\psi}), \dots, G_B^*(\hat{\psi})$ . The bootstrap estimate of  $\text{Var}[G(\psi_0)] = E[G(\psi_0)G'(\psi_0)]$  is then given as

$$\widehat{\text{Var}}[G(\psi_0)] = \frac{1}{B-1} \sum_{b=1}^B [G_b^*(\hat{\psi}) - \mu_G^*][G_b^*(\hat{\psi}) - \mu_G^*]',$$

$$\text{where } \mu_G^* = \frac{1}{B} \sum_{b=1}^B G_b^*(\hat{\psi}).$$

This gives a bootstrap estimate of  $K(\psi_0)$ ,  $\widehat{K}(\psi_0) = \frac{1}{NT} C \widehat{\text{Var}}[G(\psi_0)] C$ , which together with  $I(\hat{\psi})$  gives an estimate of the robust VC matrix of the QMLE  $\hat{\psi}$ .

Some details in calculating  $G^*(\hat{\psi})$  are given as follows. From the expression of  $G(\psi_0)$  given in Section 3.2, we see that the first four elements of  $G(\psi_0)$  are all explicit functions of  $u_0$  and the true parameters  $\psi_0$  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  $\psi_0$ , respectively. Calculating bootstrapped values of the last element of  $G^*(\hat{\psi})$ , i.e.,  $G_\lambda(\psi_0) = J_\lambda(\lambda) - \frac{1}{\sigma_\lambda^2} u'_{0\lambda} \Omega^{-1} u_0$ , 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  $h_\lambda(Y_{it}, \lambda_0) = \lambda_0^{-1} [1 + \lambda_0 h(Y_{it}, \lambda_0)] \log Y_{it} - \lambda_0^{-1} h(Y_{it}, \lambda_0)$  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  $u_{0\lambda} = h_\lambda(Y, \lambda_0) - X_\lambda(\lambda_0) \beta_0$ , the gradient  $G_\lambda(\psi_0)$  can also be expressed analytically in terms of  $u_0$  and  $\psi_0$ . 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  $\psi_0$ . For other transformations, one could go through the same process, although the expressions may be more complicated than those of Box-Cox power transformation.

For practical applications, it is essential to establish the validity of the proposed bootstrap procedure. Let  $F_\mu, F_\eta$  and  $F_v$  be, respectively, the true distributions of  $\mu_i, \eta_t$  and  $v_{it}$ . Let  $\hat{F}_\mu$  be the empirical distribution function (EDF) of  $\hat{\mu}$ ,  $\hat{F}_\eta$  the EDF of  $\hat{\eta}$  and  $\hat{F}_v$  the EDF of  $\hat{v}$ . Let  $E^*$  and  $\text{Var}^*$  be the mean and variance operators corresponding to  $(\hat{F}_\mu, \hat{F}_\eta, \hat{F}_v)$ . The theoretical bootstrap estimate of  $\text{Var}[G(\psi_0)]$  is thus,

$$\text{Var}^*[G^*(\hat{\psi})] = E^*[G^*(\hat{\psi})G^*(\hat{\psi})'] - E^*[G^*(\hat{\psi})]E^*[G^*(\hat{\psi})'], \quad (11)$$

which is approximated by the feasible version  $\widehat{\text{Var}}[G(\psi_0)]$  given above. Evidently, this approximation can be made arbitrarily accurate by choosing an arbitrarily large  $B$ .

**Corollary 1:** *Under the assumptions of Theorems 2 and 3, we have*

$$\frac{1}{NT}C\{\text{Var}^*[G^*(\hat{\psi})] - \text{Var}[G(\psi_0)]\}C \xrightarrow{p} 0, \text{ as } N, T \rightarrow \infty,$$

*and hence the proposed ECB procedure is asymptotically valid.*

The advantage of the proposed ECB procedure is that it is computationally feasible even for large panels. This is because it bootstraps the score function only by resampling the estimated error components, conditional on the QMLEs of the parameters, thus the numerical optimization as in the process of obtaining the parameter estimates is avoided. The Monte Carlo results presented in Section 5 show that this procedure performs well when used to construct robust 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  $\hat{K}(\psi_0)$  when used in constructing robust confidence intervals (CIs) for the elements of  $\psi_0$ , in comparison with the CIs based on Hessian only.

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  $\exp[N(0, 1)]$ ,  $\beta = (20, 5, 1)'$ ,  $\sigma_\mu^2 = \sigma_\eta^2 = 0.25$ , and  $\sigma_v^2 = 1.0$ .

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  $\xi_i$  is a Bernoulli random variable with probability of success  $p$  and  $Z_i$  is standard normal independent of  $\xi_i$ . The parameter  $p$  here 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  $\tau$ . 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  $\alpha$ , and is independent of  $Z_i$  and  $\xi_i$ . 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 nonnormal distribution that is still symmetric like normal distribution but leptokurtic, whereas the normal-gamma mixture gives a nonnormal distribution that is both skewed and leptokurtic.<sup>7</sup> As discussed in the introduction, one of the main purposes of a response transformation is to induce normality of the observations. We argued that while exact normality may be impractical, the transformed observations can be close to normal or at least more symmetrically distributed. In case that the exact normality is not achieved, the QML method provides an extra protection. 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  $\psi$ , in terms of **bias**(%) ( $(\text{bias}/\text{true parameter value}) \times 100\%$ ) and **rmse** (root mean squared error), with DGP 1 corresponding to the case that all errors are normal, DGP 2 the case that  $\mu$  and  $\eta$  are normal but  $v$  follows a normal-mixture distribution with  $p = 0.05$  and  $\tau = 5$ , and DGP 3 the case that  $\mu$  and  $\eta$  are normal but  $v$  follows a normal-gamma mixture with  $p = 0.05$  and  $\alpha = 4$ . The results corresponding to each combination of the 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 rmse get smaller. If  $N$  is relatively larger than  $T$ , then the bias and rmse of  $\phi_\eta$  are generally larger than those of  $\phi_\mu$  and vice versa. If  $T$  is fixed as in Table 1e, then the bias and rmse of  $\phi_\eta$ , 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  $\phi_\mu$  do not go down as  $T$  increases. More specifically, the results of Tables 1a-1f show that the rmse of a  $\sqrt{NT}$ -consistent estimator at  $(N_2, T_2)$  approximately equals that at  $(N_1, T_1)$  multiplied by  $\sqrt{N_1 T_1 / N_2 T_2}$ ; the rmse of a  $\sqrt{N}$ -consistent estimator at

---

<sup>7</sup>Using the fact that  $\xi_i^j(1-\xi_i)^k = 0, i = 1, 2, \dots, j = 1, 2, \dots$ , one can easily show that the excess kurtosis for the normal-mixture random variable is  $\kappa = \frac{3(1-p+p\tau^4)}{(1-p+p\tau^2)^2} - 3$ , and that the skewness and excess kurtosis for the normal-gamma mixture random variable are, respectively,  $\gamma = \frac{2p\alpha}{(1-p+p\alpha)^{3/2}}$  and  $\kappa = \frac{3(1-p+p\alpha(2+\alpha))}{(1-p+p\alpha)^2} - 3$ . Thus for the four sets of  $(p, \tau)$  values considered for the normal-mixture,  $\kappa = (13.45, 16.96, 22.27, 39.45)$ , and for the four sets of  $(p, \alpha)$  values considered,  $\gamma = (.3243, .5397, .5433, .7454,)$  and  $\kappa = (1.88, 2.86, 6.03, 7.00)$ .

$N_2$  approximately equals that at  $N_1$  multiplied by  $\sqrt{N_1/N_2}$ ; and the rmse of a  $\sqrt{T}$ -consistent estimator at  $T_2$  approximately equals that at  $T_1$  multiplied by  $\sqrt{T_1/T_2}$ . Further, the rmse of  $\hat{\phi}_\mu$  approximately equals that of  $\hat{\phi}_\eta$  multiplied by  $\sqrt{T/N}$  (note  $\phi_{\mu 0} = \phi_{\eta 0} = .25$ ).

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 Subsection 3.2 below Assumption N5, it may be easier to achieve consistency and asymptotic normality of  $\hat{\lambda}$  when the error distributions in the transformed model are closer to symmetry. 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. The results of Table 1 also show that the rmse under DGP 2 is generally larger than that under DGP 3. This is because DGP 2 has a much larger excess kurtosis than DGP3 (16.96 vs 1.88 as in Footnote 7). This larger kurtosis also causes larger biases of the other estimators (beside  $\hat{\lambda}$ ) when  $NT$  is not large. Monte Carlo experiments are repeated under other parameter ( $\psi$ ) 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

To investigate the finite sample performance of the bootstrap estimate of  $\text{Var}[G(\psi_0)]$ , we look at the coverage probability of the confidence interval (CI) for each parameter in the model, constructed based on the robust VC matrix estimate  $I^{-1}(\hat{\psi})\hat{K}(\psi_0)I^{-1}(\hat{\psi})$  (RCI), and compare it with the CI constructed based on  $I(\hat{\psi})$  only (HCI). Both CIs use standard normal critical values, and are asymptotically valid when the errors are normal. When the errors are nonnormal, however, only RCI can be asymptotically valid. The same DGPs as in Section 5.1 are used with some changes on the parameter values in the mixture distributions. 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. Partial results for 95% CIs are summarized in Table 5.2.

From the results we see that under normal errors, the two CIs perform equally well. Under nonnormal but symmetric errors, the bootstrap procedure leads to robust CIs with coverage probabilities generally quite close to their nominal levels even though the mixture distributions considered in the Monte Carlo experiment are quite different from normal distribution (see Footnote 7). In contrast, the Hessian-based CIs can perform quite poorly with coverage probabilities significantly below the nominal levels. The results (unreported for brevity) for the 90% and 99% CIs show the same pattern. Furthermore, the results (reported and unreported but available from the authors upon request) show that as  $N \cup T$  increases, the coverage probabilities of the robust CIs get closer to their nominal levels, and the empirical distributions of the robust  $t$ -statistics get closer to the standard normal distribution. However, the same

are not true for the Hessian-based CIs and  $t$ -statistics unless the errors are normal.

When the errors are skewed (DGP 3), the performance of RCIs may not be as good as in the case of symmetric but nonnormal error (DGP 2), but can still be significantly better than HCIs. One point to note is that when the error standard deviations are not small, e.g.,  $(\sigma_{v0}, \sigma_{\mu0}, \sigma_{\eta0}) = (1, .5, .5)$ , the inferences based on both methods do not improve much when  $N$  and  $T$  increase, suggesting that Assumption N3 may have been violated (see Footnote 6). However, the amount of skewness generated by DGP 3 (see Footnote 7) is incompatible with the transformation model as a response transformation can typically achieve near-symmetry as discussed earlier. Interestingly, the additional Monte Carlo results (unreported for brevity) show that by simply reducing  $\sigma_{v0}$  by half, the RCIs perform reasonably well, and so are the HCIs except the HCI for  $\sigma_{v0}^2$ ; further reducing the error standard deviations to  $(.5, .25, .25)$  the RCIs perform very well and the HCI for  $\sigma_{v0}^2$  continues to perform poorly. Another interesting point to note is that the CIs for  $\phi_{\mu}$  and  $\phi_{\eta}$  based on both methods perform equally well in all situations. One possible reason for this may be that the error components  $\mu$  and  $\eta$  are generated from normal for all the reported Monte Carlo results. This shows that whether the pure error  $v_{it}$  is normal or not does not affect the performance of the inference for  $\phi_{\mu}$  and  $\phi_{\eta}$ .

## 6 Discussions

Asymptotic properties of the QMLEs of the transformed panel model with two-way random-effects are studied and an error components bootstrap (ECB) method is introduced for estimating the robust VC matrix. 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 ECB method works well when model assumptions are met, and in this case the resulted CIs are robust against departures from normality, in particular in the form of excess kurtosis.

In certain economics applications, it may be more appropriate to apply a one-way random-effects model. The results in the paper can easily be simplified to suit this purpose. For the one-way random-effects model with only individual-specific effects, (i) in the derivations in Section 2, set  $\phi_{\eta} = 0$  and  $\theta_3 = \theta_1$ , (ii) in the gradient and Hessian expressions given in Section 3.2 and in the proof of Theorem 2, do (i) and drop the components corresponding to  $\phi_{\eta}$ , and (iii) in the bootstrap procedure presented in Section 4, drop the terms involving  $\hat{\eta}$  and  $\hat{\eta}^*$ . Similarly, for the one-way random-effects model with only time-specific effects, set  $\phi_{\mu} = 0$  and  $\theta_3 = \theta_2$ , and drop all the terms corresponding to  $\mu$  in the gradient and Hessian expressions, and in the bootstrap procedure. Clearly for the former the asymptotics only require  $N$  to be large, and for the latter the asymptotics only require  $T$  to be large.

## Appendix: Proofs of the Theorems

Recall  $u \equiv u(\beta, \lambda) = h(Y, \lambda) - X(\lambda)\beta$ ,  $u_\lambda = \frac{\partial}{\partial \lambda} u(\beta, \lambda)$ , and  $u_{\lambda\lambda} = \frac{\partial^2}{\partial \lambda^2} u(\beta, \lambda)$  with their values at  $(\beta_0, \lambda_0)$  denoted as  $u_0$ ,  $u_{0\lambda}$  and  $u_{0\lambda\lambda}$ . 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 away from zero as  $N, T \rightarrow \infty$ , uniformly in  $\phi \in \Phi$ .*

**Proof:** By Assumption (b) stated in Theorem 1,  $\bar{\ell}_{\max}(\phi) \leq \bar{\ell}_{\max}(\phi_0) + \epsilon$ ,  $\forall \phi \in \Phi$ , i.e.,

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

or  $\log \bar{\sigma}_v^2(\phi) \geq \frac{1}{NT}[c(\phi_\mu, \phi_\eta) - c(\phi_{\mu 0}, \phi_{\eta 0}) - \epsilon] + \log \sigma_{v0}^2 + \frac{2}{NT}[\mathbb{E}J(\lambda) - \mathbb{E}J(\lambda_0)]$ , for small  $\epsilon$  ( $> 0$ ). Assumption C4 guarantees that  $\frac{2}{NT}[\mathbb{E}J(\lambda) - \mathbb{E}J(\lambda_0)]$  is bounded, uniformly in  $\lambda \in \Lambda$ . It follows that  $\bar{\sigma}_v^2(\phi)$  is bounded away from zero as  $N, T \rightarrow \infty$ , uniformly in  $\phi \in \Phi$ .  $\blacksquare$

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

**Proof:** The proof goes in three steps.<sup>8</sup> Let  $P^*(\phi) = \Omega^{-\frac{1}{2}} X(\lambda) [X'(\lambda) \Omega^{-1} X(\lambda)]^{-1} X'(\lambda) \Omega^{-\frac{1}{2}}$  where  $\Omega^{-\frac{1}{2}}$  is the symmetric square root of  $\Omega^{-1}$ , we have by (6) and (9),

$$\hat{\sigma}_v^2(\phi) - \bar{\sigma}_v^2(\phi) = \{Q_1(\phi) - \mathbb{E}[Q_1(\phi)]\} - \{Q_2(\phi) - \mathbb{E}[Q_2(\phi)]\} + \mathbb{E}[Q_3(\phi)], \quad (\text{A-1})$$

where  $Q_1(\phi) = \frac{1}{NT} h'(Y, \lambda) \Omega^{-1} h(Y, \lambda)$ ,  $Q_2(\phi) = -\frac{1}{NT} h'(Y, \lambda) \Omega^{-\frac{1}{2}} P^*(\phi) \Omega^{-\frac{1}{2}} h(Y, \lambda)$ , and  $Q_3(\phi) = \frac{1}{NT} \{h'(Y, \lambda) - \mathbb{E}h'(Y, \lambda)\} \Omega^{-\frac{1}{2}} P^*(\phi) \Omega^{-\frac{1}{2}} [h(Y, \lambda) - \mathbb{E}h(Y, \lambda)]$ .

**Step 1.** We show  $\sup_{\phi \in \Phi} |Q_1(\phi) - \mathbb{E}[Q_1(\phi)]| = o_p(1)$ . Using  $\Omega^{-1} = Q + \theta_1 P_1 + \theta_2 P_2 + \theta_3 P_3$  given by (4), we can prove  $Q_1(\phi) - \mathbb{E}[Q_1(\phi)] = o_p(1)$  for each  $\phi \in \Phi$  by showing that  $\frac{1}{NT} \{h'(Y, \lambda) Q h(Y, \lambda) - \mathbb{E}[h'(Y, \lambda) Q h(Y, \lambda)]\} = o_p(1)$  for each  $\lambda \in \Lambda$  and  $\frac{1}{NT} \{h'(Y, \lambda) P_k h(Y, \lambda) - \mathbb{E}[h'(Y, \lambda) P_k h(Y, \lambda)]\} = o_p(1)$  for each  $\lambda \in \Lambda$ ,  $k = 1, 2, 3$ . Using the expressions for  $Q$  and  $P_k$ 's, it is trivial to show that these pointwise convergence results hold under Assumptions C4 and C5(i). To show the stochastic equicontinuity of  $Q_1(\phi)$ , we have 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  $\phi$  and  $\tilde{\phi}$  elementwise,  $Q_{1\lambda}(\phi) = \frac{2}{NT} h'_\lambda(Y, \lambda) \Omega^{-1} h(Y, \lambda)$ ,  $Q_{1\phi_\mu}(\phi) = -\frac{1}{N} h'(Y, \lambda) (\theta_1^2 P_1 + \theta_3^2 P_3) h(Y, \lambda)$ , and  $Q_{1\phi_\eta}(\phi) = -\frac{1}{T} h'(Y, \lambda) (\theta_2^2 P_2 + \theta_3^2 P_3) h(Y, \lambda)$ .

<sup>8</sup>In proving Lemma A2, the following matrix results are repeatedly used: (i) the eigenvalues of a projection matrix are either 0 or 1; (ii)  $\gamma_{\min}(A) \text{tr} B \leq \text{tr}(AB) \leq \gamma_{\max}(A) \text{tr} B$  for symmetric matrix  $A$  and positive semidefinite (p.s.d.) matrix  $B$ ; (iii)  $\gamma_{\max}(A+B) \leq \gamma_{\max}(A) + \gamma_{\max}(B)$  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  $B$ .

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  $\phi_\eta$ . First, by Cauchy-Schwarz inequality and Assumption C6(ii) and C6(iii),

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

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  $N \cup T \rightarrow \infty$ ;  $T\theta_1^2$  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  $\bar{c}$ , say. We have by Assumption C6(ii),

$$\begin{aligned} \sup_{\phi \in \Phi} |Q_{1\phi_\mu}(\phi)| &\leq \sup_{\phi \in \Phi} \frac{1}{NT} h'(Y, \lambda) (T\theta_1^2 P_1 + T\theta_3^2 P_3) h(Y, \lambda) \\ &\leq \sup_{\phi \in \Phi} (T\theta_1^2 \gamma_{\max}(P_1) + T\theta_3^2 \gamma_{\max}(P_3)) \frac{1}{NT} \|h(Y, \lambda)\|^2 \\ &\leq 2\bar{c} \sup_{\lambda \in \Lambda} \frac{1}{NT} \|h(Y, \lambda)\|^2 = O_p(1), \text{ and} \\ \sup_{\phi \in \Phi} |Q_{1\phi_\eta}(\phi)| &\leq \sup_{\phi \in \Phi} \frac{1}{NT} h'(Y, \lambda) (N\theta_2^2 P_2 + N\theta_3^2 P_3) h(Y, \lambda) \\ &\leq \sup_{\phi \in \Phi} (N\theta_2^2 \gamma_{\max}(P_2) + N\theta_3^2 \gamma_{\max}(P_3)) \frac{1}{NT} \|h(Y, \lambda)\|^2 \\ &\leq 2\bar{c} \sup_{\lambda \in \Lambda} \frac{1}{NT} \|h(Y, \lambda)\|^2 = O_p(1). \end{aligned}$$

**Step 2.** We now show  $\sup_{\phi \in \Phi} |Q_2(\phi) - E[Q_2(\phi)]| = o_p(1)$ . As  $P^*(\phi)$  is a projection matrix,  $Q_2(\phi) \leq \gamma_{\max}(P^*(\phi)) \frac{1}{NT} h'(Y, \lambda)\Omega^{-1}h(Y, \lambda) \leq Q_1(\phi)$ . 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  $\phi$  and  $\tilde{\phi}$  elementwise, and for  $A(\phi) = [X'(\lambda)\Omega^{-1}X(\lambda)]^{-1}$ ,

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

For the first part of  $Q_{2\lambda}(\phi)$ , by Cauchy-Schwarz inequality and Assumption C6(ii)-(iii),

$$\begin{aligned}
\sup_{\phi \in \Phi} |Q_{2\lambda,1}(\phi)| &= \sup_{\phi \in \Phi} \frac{1}{NT} |h'_\lambda(Y, \lambda) \Omega^{-\frac{1}{2}} P^*(\phi) \Omega^{-\frac{1}{2}} h(Y, \lambda)| \\
&\leq \sup_{\phi \in \Phi} \frac{1}{\sqrt{NT}} \{h'_\lambda(Y, \lambda) \Omega^{-\frac{1}{2}} P^*(\phi) \Omega^{-\frac{1}{2}} h_\lambda(Y, \lambda)\}^{\frac{1}{2}} Q_2(\phi)^{\frac{1}{2}} \\
&\leq \sup_{\phi \in \Phi} \gamma_{\max}(P^*(\phi)) \frac{1}{\sqrt{NT}} \{h'_\lambda(Y, \lambda) \Omega^{-1} h_\lambda(Y, \lambda)\}^{\frac{1}{2}} Q_1(\phi)^{\frac{1}{2}} \\
&\leq \sup_{\phi \in \Phi} \gamma_{\max}(\Omega^{-1}) \frac{1}{NT} \|h_\lambda(Y, \lambda)\| \|h(Y, \lambda)\| \\
&\leq 4 \sup_{\lambda \in \Lambda} \frac{1}{\sqrt{NT}} \|h_\lambda(Y, \lambda)\| \sup_{\lambda \in \Lambda} \frac{1}{\sqrt{NT}} \|h(Y, \lambda)\| = O_p(1).
\end{aligned}$$

Similarly, for the second component of  $Q_{2\lambda}(\phi)$ ,

$$\begin{aligned}
\sup_{\phi \in \Phi} |Q_{2\lambda,2}(\phi)| &= \sup_{\phi \in \Phi} \frac{1}{NT} |h'(Y, \lambda) \Omega^{-1} X_\lambda(\lambda) A(\phi) X'_\lambda(\lambda) \Omega^{-1} h(Y, \lambda)| \\
&\leq \sup_{\phi \in \Phi} \frac{1}{\sqrt{NT}} \{h'(Y, \lambda) \Omega^{-1} X_\lambda(\lambda) A(\phi) X'_\lambda(\lambda) \Omega^{-1} h(Y, \lambda)\}^{\frac{1}{2}} \cdot \sup_{\phi \in \Phi} Q_2(\phi)^{\frac{1}{2}}.
\end{aligned}$$

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

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

Note that  $Q_1(\phi) \leq \frac{\gamma_{\max}(\Omega^{-1})}{NT} \|h(Y, \lambda)\|^2 \leq \frac{4}{NT} \|h(Y, \lambda)\|^2$  as  $\sup_{\phi \in \Phi} \gamma_{\max}(\Omega^{-1}) \leq 4$  (see Footnote 4). By Assumption C6(ii),  $\sup_{\phi \in \Phi} Q_2(\phi)^{\frac{1}{2}} \leq \sup_{\phi \in \Phi} Q_1(\phi)^{\frac{1}{2}} \leq 2 \sup_{\lambda \in \Lambda} \frac{1}{\sqrt{NT}} \|h'(Y, \lambda)\| = O_p(1)$ . Consequently,  $\sup_{\phi \in \Phi} |Q_{2\lambda,2}(\phi)| = O_p(1)$ . For the third component of  $Q_{2\lambda}(\phi)$ ,

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

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

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

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

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

where we repeatedly use the fact that  $\gamma_{\max}(AB) \leq \gamma_{\max}(A)\gamma_{\max}(B)$  for p.s.d. matrices  $A$  and  $B$ . Consequently, we have  $\sup_{\phi \in \Phi} |Q_{2\lambda,3}(\phi)| = O_p(1)$ . Finally, the triangle inequality leads to:  $\sup_{\phi \in \Phi} |Q_{2\lambda}(\phi)| \leq 2 \sup_{\phi \in \Phi} |Q_{2\lambda,1}(\phi)| + 2 \sup_{\phi \in \Phi} |Q_{2\lambda,2}(\phi)| + 2 \sup_{\phi \in \Phi} |Q_{2\lambda,3}(\phi)| = O_p(1)$ .

Analogously, we can show  $\sup_{\phi \in \Phi} |Q_{2\phi_\mu}(\phi)| = O_p(1)$  and  $\sup_{\phi \in \Phi} |Q_{2\phi_\eta}(\phi)| = O_p(1)$ .

**Step 3.** Finally, we show  $\sup_{\phi \in \Phi} |EQ_3(\phi)| = o(1)$ . By Assumption C3,

$$\begin{aligned}
& \sup_{\phi \in \Phi} E[Q_3(\phi)] \\
&= \sup_{\phi \in \Phi} \frac{1}{NT} E\{[h'(Y, \lambda) - Eh'(Y, \lambda)] \Omega^{-\frac{1}{2}} P^*(\phi) \Omega^{-\frac{1}{2}} [h(Y, \lambda) - Eh(Y, \lambda)]\} \\
&= \sup_{\phi \in \Phi} \frac{1}{(NT)^2} E\{[h'(Y, \lambda) - Eh'(Y, \lambda)] \Omega^{-1} X(\lambda) Q_{XX}(\phi) X'(\lambda) \Omega^{-1} [h(Y, \lambda) - Eh(Y, \lambda)]\} \\
&\leq [\gamma_{\min}(Q_{XX}(\phi))]^{-1} \sup_{\lambda \in \Lambda} \frac{1}{(NT)^2} E\|X'(\lambda) \Omega^{-1} [h(Y, \lambda) - Eh(Y, \lambda)]\|^2 \\
&= [\gamma_{\min}(Q_{XX}(\phi))]^{-1} [\gamma_{\max}(\Omega^{-1})]^2 \sup_{\lambda \in \Lambda} \frac{1}{(NT)^2} \text{tr}\{X'(\lambda) \text{Var}[h(Y, \lambda)] X(\lambda)\} \\
&= O(1)O(1)o(1) = o(1), \text{ by Assumption C5(iii).}
\end{aligned}$$

This completes the proof of the lemma.  $\blacksquare$

**Proof Theorem 1:** Since  $\ell_{\max}(\phi)$  has a unique global maximum at  $\phi^*$  such that  $\phi^* \rightarrow \phi_0$  as  $N, T \rightarrow \infty$ , the proof of the consistency of  $\hat{\phi}$  amounts to show the uniform convergence

$$\sup_{\phi \in \Phi} \frac{1}{NT} |\ell_{\max}(\phi) - \bar{\ell}_{\max}(\phi)| \xrightarrow{p} 0, \text{ as } N, T \rightarrow \infty, \quad (\text{A-2})$$

(White, 1994, Theorem 3.4). From (7) and (10),  $\frac{1}{NT} [\ell_{\max}(\phi) - \bar{\ell}_{\max}(\phi)] = -\frac{1}{2} [\log \hat{\sigma}_v^2(\phi) - \log \bar{\sigma}_v^2(\phi)] + \frac{1}{NT} \{J(\lambda) - E[J(\lambda)]\}$ . By a Taylor expansion of  $\log \hat{\sigma}_v^2(\phi)$  at  $\bar{\sigma}_v^2(\phi)$ ,

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

where  $\bar{\sigma}_v^2(\phi)$  lies between  $\hat{\sigma}_v^2(\phi)$  and  $\bar{\sigma}_v^2(\phi)$ . Lemma A1 shows that  $\bar{\sigma}_v^2(\phi)$  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, Assumption C5(ii) leads to the pointwise convergence of  $\frac{1}{NT} \{J(\lambda) - E[J(\lambda)]\}$  to zero in probability, and Assumption C6(iv) leads to the stochastic equicontinuity of  $\frac{1}{NT} \{J(\lambda)\}$ . Thus, by Lemma 1 of Andrews (1992),  $\frac{1}{NT} \{J(\lambda) - E[J(\lambda)]\} \xrightarrow{p} 0$  uniformly in  $\phi \in \Phi$ . The

result (A-2) thus follows. Finally, the consistency of  $\hat{\beta}(\hat{\phi})$  and  $\hat{\sigma}_v^2(\hat{\phi})$  follows from that of  $\hat{\phi}$  and Assumption C3. ■

The proof of Theorem 2 requires Lemma A.3, which essentially gives a central limit theorem (CLT) for linear-quadratic forms of error components  $u = Z_\mu\mu + Z_\eta\eta + v$  defined in (2). Let  $m_{\mu 0}^{(k)}, m_{\eta 0}^{(k)}$  and  $m_{v 0}^{(k)}$  be, respectively, the  $k$ th moment of  $\mu_i, \eta_t$  and  $v_{it}, k = 1, 2, 3, 4$ . Let  $\kappa_{\mu 0} = m_{\mu 0}^{(4)}/\sigma_{\mu 0}^4 - 3$ ,  $\kappa_{\eta 0} = m_{\eta 0}^{(4)}/\sigma_{\eta 0}^4 - 3$  and  $\kappa_{v 0} = m_{v 0}^{(4)}/\sigma_{v 0}^4 - 3$ . 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)  $E|\mu_i|^{4+\epsilon_1} < \infty$ ,  $E|\eta_t|^{4+\epsilon_2} < \infty$ , and  $E|v_{it}|^{4+\epsilon_3} < \infty$ , for some  $\epsilon_1, \epsilon_2$  and  $\epsilon_3 > 0$ , and for all  $i = 1, \dots, N$  and  $t = 1, \dots, T$ ; (ii)  $\sup_{1 \leq k \leq n} \sum_{j=1}^n |w_{jk}| < \infty$ ; and (iii)  $\sup_n n^{-1} \sum_{j=1}^n |b_j|^{2+\epsilon} < \infty$  for some  $\epsilon > 0$ . Then, we have, as  $n \rightarrow \infty$ ,*

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

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

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

where  $W_\mu = Z'_\mu W Z_\mu$ ,  $W_\eta = Z'_\eta W Z_\eta$ ,  $b_\mu = Z'_\mu b$ ,  $b_\eta = Z'_\eta b$ ,  $b_{\mu, i}$  denotes the  $i$ th element of  $b_\mu$  and similarly  $b_{\eta, i}$ .

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

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

It is easy to show that  $\text{Var}(\mu'W_\mu\mu) = \sigma_{\mu 0}^4 (\kappa_{\mu 0} \sum w_{\mu, ii}^2 + 2\text{tr}(W_\mu^2))$ , and  $\text{Var}(\mu'Z'_\mu W Z_\eta\eta) = \sigma_{\mu 0}^2 \sigma_{\eta 0}^2 \text{tr}(Z'_\mu W Z_\eta Z'_\eta W Z_\mu)$ . The former leads to the expressions for  $\text{Var}(\eta'W_\eta\eta)$  and  $\text{Var}(v'Wv)$ , and the latter leads to the expressions for  $\text{Var}(\mu'Z'_\mu W v)$  and  $\text{Var}(\eta'Z'_\eta W v)$ . Finally,

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

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

For the asymptotic normality of  $q$ , we note that  $q = u'Wu + b'u = (\mu'W_\mu\mu + b'_\mu\mu) + (\eta'W_\eta\eta + b'_\eta\eta) + (v'Wv + b'_v v) + 2\mu'Z'_\mu W Z_\eta\eta + 2\mu'Z'_\mu W v + 2\eta'Z'_\eta W v$ . 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. ■

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

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

where  $A_\mu = T(\theta_1^2 P_1 + \theta_3^2 P_3)$ , and  $A_\eta = N(\theta_2^2 P_2 + \theta_3^2 P_3)$ . The proof of the theorem starts from a Taylor expansion of  $G^\dagger(\hat{\psi})$  around  $\psi_0$ :

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

where  $\bar{\psi}$  lies between  $\hat{\psi}$  and  $\psi_0$  elementwise, and the Hessian matrix  $H(\psi)$  has the elements:

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

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

$$= 2N^2(\theta_2^3 P_2 + \theta_3^3 P_3).$$

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

- (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)  $\frac{1}{NT}C\{H(\psi_0) - E[H(\psi_0)]\}C = o_p(1).$

For (i), the joint asymptotic normality of the first four elements of  $G^\dagger(\psi_0)$  follows from Lemma A3 and Cramér-Wold device. Assumption N6 and Cramér-Wold device lead to the joint asymptotic normality of all elements 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  $G_\beta(\psi_0)$ ,  $G_{\sigma_2^2}(\psi_0)$  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

$$u'P_1u = NT(s_\mu^2 + s_{v_1}^2 + s_{\mu v}^2), \quad u'P_2u = NT(s_\eta^2 + s_{v_2}^2 + s_{\eta v}^2), \quad \text{and } u'P_3u = NT(\bar{\mu} + \bar{\eta} + \bar{v})^2,$$

where  $s_\mu^2 = \frac{1}{N} \sum_{i=1}^N (\mu_i - \bar{\mu})^2$ ,  $s_{v_1}^2 = \frac{1}{N} \sum_{i=1}^N (\bar{v}_i - \bar{v})^2$ , and  $s_{\mu v}^2 = \frac{1}{N} \sum_{i=1}^N \mu_i (\bar{v}_i - \bar{v})$ ;  $s_\eta^2 = \frac{1}{T} \sum_{t=1}^T (\eta_t - \bar{\eta})^2$ ,  $s_{v_2}^2 = \frac{1}{T} \sum_{t=1}^T (\bar{v}_t - \bar{v})^2$ , and  $s_{\eta v}^2 = \frac{1}{T} \sum_{t=1}^T [\eta_t (\bar{v}_t - \bar{v})]$ ;  $\bar{\mu} = \frac{1}{N} \sum_{i=1}^N \mu_i$ ,  $\bar{\eta} = \frac{1}{T} \sum_{t=1}^T \eta_t$ ,  $\bar{v}_i = \frac{1}{T} \sum_{t=1}^T v_{it}$ ,  $\bar{v}_t = \frac{1}{N} \sum_{i=1}^N v_{it}$ , and  $\bar{v} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T v_{it}$ . These give

$$\begin{aligned} G_{\phi_\mu}(\psi_0) &= \frac{1}{2\sigma_{v_0}^2} u' A_\mu u - \frac{1}{2} T(N-1)\theta_1 - \frac{1}{2} T\theta_3 \\ &= \frac{T}{2\sigma_{v_0}^2} (\theta_1^2 u' P_1 u + \theta_3^2 u' P_3 u) - \frac{1}{2} T(N-1)\theta_1 - \frac{1}{2} T\theta_3 \\ &= \frac{NT^2(s_\mu^2 + s_{v_1}^2 + s_{\mu v}^2)}{2\sigma_{v_0}^2(T\phi_{\mu 0} + 1)^2} + \frac{NT^2(\bar{\mu} + \bar{\eta} + \bar{v})^2}{2\sigma_{v_0}^2(T\phi_{\mu 0} + N\phi_{\eta 0} + 1)^2} \\ &\quad - \frac{(N-1)T}{2(T\phi_{\mu 0} + 1)} - \frac{T}{2(T\phi_{\mu 0} + N\phi_{\eta 0} + 1)} \end{aligned}$$

Under Assumption C1, we have, as  $N \rightarrow \infty$ ,  $\bar{\mu} \xrightarrow{p} 0$ ,  $\bar{v} \xrightarrow{p} 0$ ,  $\bar{\eta} \xrightarrow{p} 0$  if  $T \rightarrow \infty$  as well, but otherwise does not converge,  $s_\mu^2 \xrightarrow{p} \sigma_{\mu 0}^2$ ,  $s_{v_1}^2 \xrightarrow{p} \sigma_{v_0}^2/T$ ,  $s_{\mu v}^2 \xrightarrow{p} 0$ . These give for  $N$  large

$$G_{\phi_\mu}(\psi_0) \approx \frac{\sqrt{N}}{2(\phi_{\mu 0} + 1/T)} \sqrt{N} \left( \frac{s_\mu^2 + s_{v_1}^2}{\sigma_{\mu 0}^2 + \sigma_{v_0}^2/T} - 1 \right)$$

Clearly, the term  $\sqrt{N}((s_\mu^2 + s_{v_1}^2)/(\sigma_{\mu 0}^2 + \sigma_{v_0}^2/T) - 1)$  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  $T \rightarrow \infty$ ,

$$G_{\phi_\eta}(\psi_0) \approx \frac{\sqrt{T}}{2(\phi_{\eta_0} + 1/N)} \sqrt{T} \left( \frac{s_\eta^2 + s_{v2}^2}{\sigma_{\eta_0}^2 + \sigma_{v0}^2/N} - 1 \right)$$

showing that it is  $O_p(\sqrt{T})$ , irrespective of whether  $N$  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  $\lambda$ , appear in  $H(\psi)$  additively or multiplicatively. The parameter  $\lambda$  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 order derivatives w.r.t  $\lambda$ ;

(b)  $\frac{1}{NT}[h'(Y, \bar{\lambda})WX(\bar{\lambda}) - h'(Y, \lambda_0)WX(\lambda_0)] = o_p(1)$ , for  $W = I_{NT}, P_1, P_2, P_3$ , and the same holds if  $X(\lambda)$  or  $h(Y, \lambda)$  is replaced by its first and second order derivatives w.r.t.  $\lambda$ ;

(c)  $\frac{1}{NT}[h'(Y, \bar{\lambda})Wh(Y, \bar{\lambda}) - h'(Y, \lambda_0)Wh(Y, \lambda_0)] = o_p(1)$ , for  $W = I_{NT}, P_1, P_2, P_3$ , and the same is true when  $h(Y, \lambda)$  is replaced by its first and second order derivatives w.r.t.  $\lambda$ ; 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  $\lambda_0$ . Let  $\iota_i$  denotes a  $k \times 1$  vector with 1 in the  $i$ th place and 0 elsewhere. Then by the fact that  $W$  is a projection matrix, Cauchy-Schwarz inequality, and Assumption N4

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

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  $\epsilon$ -neighborhood of  $\lambda_0$  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):  $\frac{1}{NT}C\{H(\psi_0) - E[H(\psi_0)]\}C = o_p(1)$ , 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  $\frac{1}{NT}[u'_0Wu_0 - E(u'_0Wu_0)] = o_p(1)$ , 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)  $\frac{1}{NT}X'(\lambda_0)W[h_\lambda(Y, \lambda_0) - E(h_\lambda(Y, \lambda_0))] = o_p(1)$ , 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)  $\frac{1}{NT}\{h'_\lambda(Y, \lambda_0)Wh(Y, \lambda_0) - E[h'_\lambda(Y, \lambda_0)Wh(Y, \lambda_0)]\} = o_p(1)$ , for  $W = I_{NT}, P_1, P_2, P_3$ , and the same holds if  $h(Y, \lambda_0)$  is replaced by  $h_\lambda(Y, \lambda_0)$ , or  $h_\lambda(Y, \lambda_0)$  by  $h_{\lambda\lambda}(Y, \lambda_0)$ .

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

**Proof Corollary 1:** Let  $S(u_0, \psi_0) = \frac{1}{\sqrt{NT}}CG(\psi_0)$  be the normalized gradient function, viewed as a vector-valued function of the true error vector  $u_0$  and the true but unknown parameter vector  $\psi_0$ . Clearly, the bootstrap analogue of  $S(u_0, \psi_0)$  is  $S(\hat{u}^*, \hat{\psi})$ . It suffices to show that  $E^*[S(\hat{u}^*, \hat{\psi})] = o_p(1)$  and  $E^*[S(\hat{u}^*, \hat{\psi})S(\hat{u}^*, \hat{\psi})'] - E[S(u_0, \psi_0)S(u_0, \psi_0)'] = o_p(1)$ . Denote  $F_0 = (F_{\mu 0}, F_{\eta 0}, F_{v 0})$  and  $\hat{F} = (\hat{F}_{\hat{\mu}}, \hat{F}_{\hat{\eta}}, \hat{F}_{\hat{v}})$ . We have, as  $\hat{\psi}$  is consistent for  $\psi_0$ ,

$$\begin{aligned} E^*[S(\hat{u}^*, \hat{\psi})] &= \int S(u, \hat{\psi})d\hat{F}(u) = \int S(u, \psi_0)d\hat{F}(u) + o_p(1), \\ E^*[S(\hat{u}^*, \hat{\psi})S(\hat{u}^*, \hat{\psi})'] &= \int S(u, \hat{\psi})S(u, \hat{\psi})'d\hat{F}(u) = \int S(u, \psi_0)S(u, \psi_0)'d\hat{F}(u) + o_p(1), \end{aligned}$$

which are compared, respectively, with  $E[S(u_0, \psi_0)] = \int S(u, \psi_0)dF_0(u) = o_p(1)$  by Assumption N3, and with  $E[S(u_0, \psi_0)S(u_0, \psi_0)'] = \int S(u, \psi_0)S(u, \psi_0)'dF_0(u)$ . The results thus follow if  $\hat{F}_{\hat{\mu}}(u) - F_{\mu 0}(u) = o_p(1)$  for each continuity point  $u$  of  $F_{\mu 0}$ ,  $\hat{F}_{\hat{\eta}}(u) - F_{\eta 0}(u) = o_p(1)$  for each continuity point  $u$  of  $F_{\eta 0}$ , and  $\hat{F}_{\hat{v}}(u) - F_{v 0}(u) = o_p(1)$  for each continuity point  $u$  of  $F_{v 0}$ . To prove the first case, note that  $\hat{u} = h(Y, \hat{\lambda}) - X(\hat{\lambda})\hat{\beta}$ ,

$$\hat{\mu} = \frac{1}{T}(I_n \otimes 1'_T)\hat{u} = \frac{1}{T}(I_n \otimes 1'_T)u_0 + \frac{1}{T}(I_n \otimes 1'_T)u_{0\alpha'}(\hat{\alpha} - \alpha_0) = \mu + o_p(1),$$

where  $\alpha = (\beta', \lambda)'$  and  $u_{0\alpha'} = \frac{\partial}{\partial \alpha'}u(\alpha_0)$ , implying that  $\hat{F}_{\hat{\mu}}(u) - F_{\mu 0}(u) = o_p(1)$  for each continuity point  $u$  of  $F_{\mu 0}$ . Details on this and the proofs of the results corresponding to  $\eta$  and  $v$  (which can be proved in a similar fashion) are available from the authors.

## References

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

- [2] Abrevaya, J. (1999b). Rank estimation of transformation model with observed truncation. *Econometrics Journal* **2**, 292-305.
- [3] Abrevaya, J. (2000). Rank estimation of a generalized fixed-effects regression model. *Journal of Econometrics*, **95**, 1-23.
- [4] Andrews, D. W. K. (1987). Consistency in nonlinear econometric models: a generic uniform law of large numbers. *Econometrica*, **55**, 1465-1471.
- [5] Andrews, D. W. K. (1992). Generic Uniform Convergence. *Econometric Theory*, **8**, 241-257.
- [6] Arellano, M. (2003). *Panel Data Econometrics*. Oxford: Oxford University Press.
- [7] Baltagi, B. H. (1997). Testing linear and log-linear error components regression against Box-Cox alternatives. *Statistics and Probability Letters*, **33**, 63-68.
- [8] Baltagi, B. H. (2001). *Econometric Analysis of Panel Data*. New York: John Wiley & Sons, Ltd.
- [9] 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.
- [10] 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.
- [11] Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, **76**, 296-311.
- [12] Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.
- [13] Chen, S. (2002). Rank estimation of transformation models. *Econometrica*, **70**, 1683-1697.
- [14] Chen, S. (2010). Root- $N$ -consistent estimation of fixed-effect panel data transformation models. *Journal of Econometrics*, **159**, 222-234.
- [15] Cook R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York: John Wiley & Sons, Ltd.
- [16] Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford: Oxford University Press.
- [17] Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.
- [18] Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*. New York: John Wiley & Sons, Ltd.
- [19] Frees, E. W. (2004). *Longitudinal and Panel Data*. Cambridge: Cambridge University Press.
- [20] Giannakas, K., Tran, K. C., and Tzouvelekas, V. (2003). On the choice of functional form in stochastic frontier modeling. *Empirical Economics*, **28**, 75-100.
- [21] Greene, W. H. (2000). *Econometric Analysis, 4th ed.* Singapore: Prentice-Hall Pte Ltd.

- [22] Hernandze, F. and Johnson, R. A. (1980). The large-sample behavior of transformations to normality. *Journal of The American Statistical Association* **75**, 855-861.
- [23] Hinkley, D. V. (1975). On power transformations to symmetry. *Biometrika*, **62**, 101-111.
- [24] Honoré B. E. and Hu, L. (2010). Estimation of a transformation model with truncation, interval observation and time-varying covariates. *Econometrics Journal* **13**, 127-144.
- [25] Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge: Cambridge University Press.
- [26] Hurvich, C. M. and Zeger, S. (1987). Frequency domain bootstrap methods for time series. Working Paper, Statistics and Operations Research, New York University.
- [27] Yeo, I. K. and Johnson, R.A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, **87**, 954-959.
- [28] 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.
- [29] Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. New York: Springer-Verlag.
- [30] McLeish, D. L. (1975). A maximal inequality and dependent strong laws. *Annals of probability* **3**, 829-839.
- [31] Polachek, S. W. and Yoon, B. J. (1996). Panel estimates of a two-tiered earning frontier. *Journal of Applied Econometrics* **11**, 169-178.
- [32] Pötscher, B. M. and Prucha, I. R. (1989). A uniform law of large numbers for dependent and heterogeneous data process. *Econometrica* **57**, 675-683.
- [33] Shin, Y. (2008). Semiparametric estimation of the Box-Cox transformation model. *Econometrics Journal* **11**, 517-537.
- [34] White, H. (1994). *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.
- [35] Yang, Z. L. (1999). Estimating a transformation and its effect on Box-Cox  $T$ -ratio. *Test*, **8**, 167-190.
- [36] Yang, Z. L. and Huang, J. H. (2011). A transformed random effects model with applications. *Applied Stochastic Models in Business and Industry*, **27**, 222-234.
- [37] Yang, Z. L. (2006). A modified family of power transformations. *Economics Letters* **92**, 14-19.
- [38] 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.
- [39] Yang, Z. L. and Tse, Y. K. (2008). Generalized LM tests for functional form and heteroscedasticity. *Econometrics Journal* **11**, 349-376.

**Table 1a.** Monte Carlo Results for bias and rmse:  $T = \text{Ceiling}(N/3)$ 

$(N, T)$		(10, 4)		(30, 10)		(90, 30)		(200, 67)	
DGP	$\psi$	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:  $N = \text{Ceiling}(T/3)$ 

$(N, T)$		(4, 10)		(10, 30)		(30, 90)		(67, 200)	
DGP	$\psi$	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:  $T = \text{Ceiling}(N^{2/3})$ 

$(N, T)$		(10, 5)		(30, 10)		(90, 21)		(200, 35)	
DGP $\psi$		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
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
	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:  $N = \text{Ceiling}(T^{2/3})$ 

$(N, T)$		(5, 10)		(10, 30)		(21, 90)		(35, 200)	
DGP $\psi$		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
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
	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	$\psi$	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
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
	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	$\psi$	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
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
	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)	
$\psi$	HCI	RCI	HCI	RCI	HCI	RCI	HCI	RCI	HCI	RCI
	<u>Normal Errors</u>									
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
	<u>Normal-Mixture, <math>p = .05, \tau = 5</math></u>									
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
	<u>Normal-Mixture, <math>p = .10, \tau = 5</math></u>									
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
	<u>Normal-Mixture, <math>p = .05, \tau = 10</math></u>									
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
	<u>Normal-Gamma Mixture, <math>p = .05, \alpha = 9</math></u>									
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