Econometrica Supplementary Material

SUPPLEMENT TO "FIXED-EFFECTS DYNAMIC PANEL MODELS, A FACTOR ANALYTICAL METHOD"

(*Econometrica*, Vol. 81, No. 1, January 2013, 285–314)

BY JUSHAN BAI

THIS SUPPLEMENT PROVIDES the technical proofs and additional related results that were omitted due to space constraint.

S.1. SEMIPARAMETRIC EFFICIENCY BOUND AND PROOF OF PROPOSITION 1

The proof follows closely that of Hahn and Kuersteiner (2002). The analysis here focuses on insight instead of rigor. Under normality of u_{it} and under the fixed-effects assumption that η_i are constants, the likelihood function is

$$\ell(\theta) = -\frac{N}{2}\sum_{t=1}^{T}\log\sigma_t^2 - \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_t^2}(y_{it}-\eta_i-\rho y_{it-1})^2.$$

Let $\eta = (\eta_1, \eta_2, \ldots, \eta_N)$ and $\psi = (\psi_1, \psi_2, \ldots, \psi_T)$ with $\psi_t = \frac{1}{\sigma_t^2}$, and similarly, let $\widetilde{\eta} = (\widetilde{\eta}_1, \widetilde{\eta}_2, \ldots, \widetilde{\eta}_N)$ and $\widetilde{\psi} = (\widetilde{\psi}_1, \ldots, \widetilde{\psi}_T)$. We further put $\theta = (\rho, \eta, \psi)$ and $\widetilde{\theta} = (\widetilde{\rho}, \widetilde{\eta}, \widetilde{\psi})$. Consider the local likelihood ratio $\ell(\theta + (NT)^{-1/2}\widetilde{\theta}) - \ell(\theta)$. It is not difficult to show that, when θ is the true parameter,

$$\ell\bigl(\theta + (NT)^{-1/2}\widetilde{\theta}\bigr) - \ell(\theta) = \Delta_{NT}(\widetilde{\theta}) - \frac{1}{2}E\bigl[\Delta_{NT}(\widetilde{\theta})\bigr]^2 + o_p(1), \tag{S.1}$$

where

$$\begin{aligned}\Delta_{NT}(\widetilde{\theta}) = &-\frac{1}{2\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\psi}_t\bigl(\varepsilon_{it}^2 - \sigma_t^2\bigr)\\ &+ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\bigl(\widetilde{\eta}_i + \widetilde{\rho}y_{i,t-1}^*\bigr)\psi_t\varepsilon_{it},\end{aligned}$$

and $y_{it}^* = \eta_i/(1-\rho) + \varepsilon_{it} + \rho\varepsilon_{i,t-1} + \cdots + \rho^{t-1}\varepsilon_{i1}$. In fact, assuming θ is the true parameter, $\ell(\theta) = \frac{N}{2}\sum_{t=1}^{T}\log(\psi_t) - \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2\psi_t^2$, and

$$\begin{aligned}\ell\bigl(\theta + (NT)^{-1/2}\widetilde{\theta}\bigr) = &\frac{N}{2}\sum_{t=1}^{T}\log\left(\psi_t + \frac{1}{\sqrt{NT}}\widetilde{\psi}_t\right)\\ &- \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{it} - \frac{1}{\sqrt{NT}}\widetilde{\eta}_i - \frac{1}{\sqrt{NT}}\widetilde{\rho}y_{i,t-1}\right)^2\\ &\times\left(\psi_t + \frac{1}{\sqrt{NT}}\widetilde{\psi}_t\right)^2.\end{aligned}$$

 DOI: 10.3982/ECTA9409

Expanding $\log(\psi_t + \frac{1}{\sqrt{NT}}\widetilde{\psi}_t) = \log\psi_t + \frac{1}{\sqrt{NT}}\sigma_t^2\widetilde{\psi}_t - \frac{1}{2NT}\sigma_t^4\widetilde{\psi}^2 + O(1/(NT)^{3/2})$, we can rewrite

$$\ell\big(\theta + (NT)^{-1/2}\widetilde{\theta}\big) = \ell(\theta) + \Delta_{NT}(\widetilde{\theta}) - \frac{1}{4}\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{\psi}_t^2}{\psi_t^2} - \frac{1}{2NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(\widetilde{\alpha}_i + \widetilde{\rho}y_{i,t-1}^*\big)^2\psi_t^2 + o_p(1).$$

Note that we have replaced $y_{i,t-1}$ by $y_{i,t-1}^*$. This replacement only contributes an $o_p(1)$ term to the likelihood ratio under large T. Next, it is easy to see that

$$\frac{1}{4}\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{\psi}_t^2}{\psi_t^2} + \frac{1}{2NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(\widetilde{\alpha}_i + \widetilde{\rho}y_{i,t-1}^*\big)^2\psi_t^2 = \frac{1}{2}E\big[\Delta_{NT}(\widetilde{\theta})\big]^2 + o_p(1).$$

This verifies (S.1). Rewrite

$$\Delta_{NT}(\widetilde{\theta}) = \widetilde{\rho}\Delta_{NT,1} + \Delta_{NT,2}(\widetilde{\eta},\widetilde{\psi}) + \Delta_{NT,3}(\widetilde{\eta},\widetilde{\psi}),$$

where

$$\Delta_{NT,1} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\eta_i}{1-\rho}\psi_t\varepsilon_{it} + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it-1}\psi_t\varepsilon_{it},$$

$$\Delta_{NT,2}(\widetilde{\eta},\widetilde{\psi}) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\eta}_i\psi_t\varepsilon_{it},$$

$$\Delta_{NT,3}(\widetilde{\eta},\widetilde{\psi}) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\psi}_t\big(\varepsilon_{it}^2 - \sigma_t^2\big),$$

and $w_{it-1} = \varepsilon_{i,t-1} + \rho\varepsilon_{i,t-2} + \cdots + \rho^{t-2}\varepsilon_{i1}$. The efficiency bound is $1/E(\bar{\Delta}_1^2)$, where $\bar{\Delta}_1$ is the residual in the projection of Δ_1 on the linear space spanned by $\Delta_2(\widetilde{\eta},\widetilde{\psi})$ and $\Delta_3(\widetilde{\eta},\widetilde{\psi})$, where Δ_k $(k = 1, 2, 3)$ is the limit of $\Delta_{NT,k}$; see Theorem 6 of Hahn and Kuersteiner (2002). In the limit, $\widetilde{\eta}$ and $\widetilde{\psi}$ are elements of an infinite dimensional Banach space. For insight, let us examine the finite sample projection. Under normality, $\Delta_{NT,3}$ is uncorrelated with (and is asymptotically independent of) $\Delta_{NT,k}$ $(k = 1, 2)$. Thus, to minimize the variance of the projection residual, the optimal choice of $\widetilde{\psi}_t$ is $\widetilde{\psi}_t = 0$ for all t. It remains to consider the optimal projection of $\Delta_{NT,1}$ on $\Delta_{NT,2}$ alone. The first term of $\Delta_{NT,1}$ is "perfectly" correlated with $\Delta_{NT,2}$, and the second term is uncorrelated with $\Delta_{NT,2}$. Thus the optimal projection is to set $\widetilde{\eta}_i = -\eta_i/(1-\rho)$, leaving the projection residual as

$$\bar{\Delta}_{NT,1} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} w_{it-1} \psi_t \varepsilon_{it}.$$

The limit distribution of the above is $N(0, \gamma)$, where γ is defined in the main text. Thus the semiparametric efficiency bound is $1/\gamma$. *Q.E.D.*

S.2. FIXED-T CONSISTENCY

Fixed-T consistency follows from existing literature on factor analysis. However, a simple direct proof (main sketch) is given here. Assume $\pi_N \to \pi$, so that $\theta_N^0 \to \theta^0$. By the law of large numbers, $S_N \xrightarrow{p} \Sigma(\theta^0)$. Uniformly over the compact set described in the text, as $N \to \infty$,

$$Q_N(\theta) \xrightarrow{p} Q(\theta) = \log\left|\Sigma(\theta)\right| + \operatorname{tr}\left[\Sigma\left(\theta^0\right)\Sigma^{-1}(\theta)\right].$$

It is well known that $Q(\theta)$ is minimized when $\Sigma(\theta) = \Sigma(\theta^0)$ (e.g., Arnold (1981, p. 460)). Since the factor structure is identifiable, $\Sigma(\theta) = \Sigma(\theta^0)$ if and only if $\theta = \theta^0$. It follows that the limiting function $Q(\theta)$ is uniquely minimized at θ^0. Since the objective function $Q_N(\theta)$ is also continuous in θ, by classical results on consistency (e.g., Amemiya (1985), Newey and McFadden (1994)), we have $\widehat{\theta} \xrightarrow{p} \theta^0$. Without assuming π_N^0 converging to π^0, the argument is modified as follows. We can show that

$$Q_N(\theta) = Q_N^*(\theta) + o_p(1),$$

where $o_p(1)$ is uniform over the compact set, as defined earlier; $Q_N^*(\theta)$ has the same form as $Q(\theta)$ but with $\Sigma(\theta^0)$ replaced by $\Sigma(\theta_N^0)$, so $Q_N^*(\theta)$ depends on N via θ_N^0 only. But $Q_N^*(\theta)$ is uniquely minimized at $\theta = \theta_N^0$ by the same reasoning used for $Q(\theta)$. This implies that $\widehat{\theta} = \theta_N^0 + o_p(1)$.

After consistently estimating ρ, the time-effects parameter δ is easily recovered as $\widehat{\delta} = \widehat{\Gamma}^{-1}\bar{y}$, since $\mu = \Gamma\delta$ and μ is estimated by $\bar{y}$. The estimator $\widehat{\delta}$ is $\sqrt{N}$ consistent for $\delta + 1_T\bar{\eta}$ (a shift in δ). We may impose $\bar{\eta} = 0$ or $\sum_{t=1}^{T} \delta_t = 0$. Such a restriction is necessary since we cannot separately identify the sample mean of δ_t and of η_i. With $\bar{\eta} = 0$, $\widehat{\delta}$ is consistent for δ. With $\sum_{t=1}^{T} \delta_t = 0$, we obtain an estimate of $\bar{\eta}$ as $\widehat{\bar{\eta}} = 1_T'\widehat{\delta}/T$. Then $\widehat{\delta} - 1_T\widehat{\bar{\eta}}$ is consistent for δ. Under either restriction, a consistent estimate for δ is available.

S.3. ARBITRARY INITIAL CONDITIONS

In the main text, we assume $y_{i0} = 0$ for notational simplicity. If $y_{i0} \neq 0$, let $y_{it}^{\dagger} = y_{it} - y_{i0}$ for $t \geq 0$; then $y_{i0}^{\dagger} = 0$. The model for $y_{it}^{\dagger}$ has a newly defined individual heterogeneity. Consistency and asymptotic normality still hold with $y_{it}^{\dagger}$. But using $y_{it}^{\dagger}$ amounts to using differenced data; see Alvarez and Arellano (2004). This implies that the factor approach is also applicable for differenced data. However, if the initial observation is available, it should be used in estimation. We now consider general initial conditions, not necessarily being drawn from a stationary distribution and allowed to depend on the fixed effects. By repeated substitution, we have

$$y_i = \Gamma\delta + \rho\Gamma e_1 y_{i0} + \Gamma 1_T \eta_i + \Gamma u_i, \tag{S.2}$$

where $e_1 = (1, 0, \dots, 0)'$ is $T \times 1$ so that $\rho\Gamma e_1 = (\rho, \rho^2, \dots, \rho^T)'$. Consider the projection (or viewing y_{i0} is generated in this way)

$$y_{i0} = \delta_0 + \phi\eta_i + u_{i0},$$

where $E(u_{i0}) = 0$ and $\operatorname{var}(u_{i0}) = \sigma_0^2$. Substituting y_{i0} into (S.2), and stacking y_{i0} and y_i, we have

$$\begin{bmatrix} y_{i0} \\ y_i \end{bmatrix} = \begin{bmatrix} \delta_0 \\ \Gamma\delta \end{bmatrix} + \begin{bmatrix} \phi \\ \rho\Gamma e_1\phi + \Gamma 1_T \end{bmatrix} \eta_i + \begin{bmatrix} 1 & 0 \\ \rho\Gamma e_1 & \Gamma \end{bmatrix} \begin{bmatrix} u_{i0} \\ u_i \end{bmatrix},$$

or more compactly,

$$y_i^+ = \delta^+ + \Lambda^+\eta_i + \Gamma^+ u_i^+,$$

where $y_i^+ = (y_{i0}, y_i')'$, $\delta^+ = (\delta_0, \delta'\Gamma')'$, $u_i^+ = (u_{i0}, u_i')'$, the matrix Γ^+ has exactly the same form as Γ, but of dimension $(T+1) \times (T+1)$, and the factor loading vector Λ^+ is

$$\Lambda^+ = \begin{bmatrix} \phi \\ \phi\rho + 1 \\ \phi\rho^2 + 1 + \rho \\ \vdots \\ \phi\rho^T + 1 + \rho + \cdots + \rho^{T-1} \end{bmatrix},$$

which can also be written as $\Lambda^+ = \Gamma^+(\phi, 1_T')'$. Let

$$\theta_N = \big(\rho, \pi_N, \phi, \sigma_0^2, \sigma_1^2, \dots, \sigma_T^2\big).$$

Define

$$S_N = \frac{1}{n}\sum_{i=1}^{N}\big(y_i^+ - \bar{y}^+\big)\big(y_i^+ - \bar{y}^+\big)';$$

then

$$E(S_N) = \Sigma(\theta) = \Gamma^+\big[\big(\phi, 1_T'\big)'\big(\phi, 1_T'\big)\pi_N + \Phi^+\big]\Gamma^{+\prime},$$

where $\Phi^+ = \operatorname{diag}(\sigma_0^2, \sigma_1^2, \dots, \sigma_T^2)$. This is again an identifiable factor structure for $T \geq 2$ (three periods of data). We estimate θ by minimizing the loss function in (3) with the newly definded S_N and $\Sigma(\theta)$. Consistency and asymptotical normality remain valid.

S.4. CROSS-SECTIONAL HETEROSKEDASTICITY

Cross-sectional heteroskedasticity is easily incorporated into the factor approach. Suppose that $E(u_{it}^2) = \sigma_{it}^2$. In this case, σ_t^2 is replaced by $\bar{\sigma}_{Nt}^2 = \frac{1}{n}\sum_{i=1}^N \sigma_{it}^2$, the average variance over the cross sections. The Φ matrix is replaced by

$$\Phi_N = \operatorname{diag}\big(\bar{\sigma}_{N1}^2, \ldots, \bar{\sigma}_{NT}^2\big)$$

(a $T \times T$ diagonal matrix). We have $E(S_N) = \Gamma(1_T 1_T' \pi_N + \Phi_N)\Gamma'$ and $\theta_N = (\rho, \pi_N, \bar{\sigma}_{N1}^2, \ldots, \bar{\sigma}_{NT}^2)$. The only difference is that the variance parameters also depend on N. Consistency and asymptotical normality remain the same. Estimating the average variance has not been noted in the factor literature. Also in this case, the factor approach cannot be viewed as a likelihood approach even under normal distributions. This is because the normal likelihood function would involve a term $\sum_{i=1}^N (y_i - \bar{y})'\Sigma_i^{-1}(y_i - \bar{y})$, which is not equal to $\operatorname{tr}(S_N \Sigma^{-1})$ when Σ_i depends on i (non common). The factor approach overcomes the incidental-parameter problem caused by the cross-sectional heteroskedasticity. This is especially useful for large N and small T.

S.5. LARGE-T RESULT UNDER HOMOSKEDASTICITY

The result of this section provides a useful benchmark for comparison with the existing literature. We shall first omit the time effects to avoid the complication from the incidental-parameter problem under large T. Consider

$$y_{it} = \eta_i + \rho y_{it-1} + u_{it},$$

with $E(u_{it}) = 0$ and $E(u_{it}^2) = \sigma^2$ for all i and t. We further assume $y_{i0} = 0$ to simplify the notations. This will not affect consistency and the limiting distribution under large T. Moreover, the unrestricted initial conditions discussed in Section S.3 are still applicable. With homoskedasticity, matrix Φ now becomes $\Phi = \sigma^2 I_T$, a special case of the factor model. In GMM estimation, Ahn and Schmidt (1995) imposed the homoskedasticity restriction through additional moment conditions. Although not explicitly treated, we still permit cross-sectional heteroskedasticity. Under $u_{it} \sim N(0, \sigma_i^2)$, with σ_i^2 uniformly bounded, we use $\bar{\sigma}_N^2 = \frac{1}{N}\sum_{i=1}^N \sigma_i^2$ in place of σ^2. In the absence of time effects, we define $S_N = \frac{1}{N}\sum_{i=1}^N y_i y_i'$ (no need to subtract the mean vector). Then $E(S_N) = \Gamma(1_T 1_T' \pi_N + \sigma^2 I_T)\Gamma'$ with $\pi_N = \frac{1}{N}\sum_{i=1}^N \eta_i^2$. We reparameterize it as

$$\Sigma(\theta) = \sigma^2 \Gamma\big(1_T 1_T' \tau_N + I_T\big)\Gamma',$$

where $\tau_N = \pi_N/\sigma^2 = \frac{1}{N}\sum_{i=1}^N \eta_i^2/\sigma^2$. Let $\theta_N = (\rho, \sigma^2, \tau_N)$. Since $|\Gamma| = 1$, we have

$$|\Sigma| = (1 + T\tau_N)\sigma^{2T},$$
$$\log|\Sigma| = T\log\sigma^2 + \log(1 + T\tau_N).$$

From

$$\begin{aligned}\Sigma^{-1} &= \Gamma'^{-1}\big(1_T 1_T'\tau_N + I_T\big)^{-1}\Gamma^{-1}/\sigma^2 \\ &= \frac{1}{\sigma^2}B'\left[I_T - 1_T 1_T' \frac{\tau_N}{1+T\tau_N}\right]B \\ &= \frac{1}{\sigma^2}B'B - B'1_T 1_T' B\frac{\tau_N}{(1+T\tau_N)\sigma^2},\end{aligned}$$

the likelihood function is equal to

$$\begin{aligned}\ell_{NT}(\theta) = &-\frac{NT}{2}\log\sigma^2 - \frac{N}{2\sigma^2}\operatorname{tr}\big(BS_N B'\big) \\ &- \frac{N}{2}\log(1+T\tau_N) + \frac{1}{2}\frac{N}{\sigma^2}\frac{\tau_N}{1+T\tau_N}\big(1_T' BS_N B' 1_T\big).\end{aligned}$$

We shall refer to the above objective function as the likelihood function even though it is not a likelihood function under the fixed-effects setup. Let $\widehat{\theta} = (\widehat{\rho}, \widehat{\sigma}^2, \widehat{\tau})$ be the estimator of θ_N by maximizing the objective function. We first establish consistency of the estimator. Working with the concentrated likelihood function turns out to be convenient. Only the last two terms of the objective function depend on τ_N. By setting the first order condition with respect to τ_N to zero, we obtain

$$\widetilde{\tau}_N = \frac{1}{\sigma^2}\frac{1}{T^2}\big(1_T' BS_N B' 1_T\big) - \frac{1}{T}.$$

Substitute this expression into $\ell_{NT}(\theta)$ to obtain the concentrated objective function as

$$\begin{aligned}\ell_c\big(\rho, \sigma^2\big) = &-\frac{N(T-1)}{2}\log\sigma^2 - \frac{N}{2\sigma^2}\operatorname{tr}\big(BS_N B'\big) \\ &- \frac{N}{2}\log\bigg(\frac{1}{T}1_T' BS_N B' 1_T\bigg) + \frac{N}{2}\bigg(\frac{1_T' BS_N B' 1_T}{\sigma^2 T} - 1\bigg).\end{aligned}$$

Let (ρ^0, σ^{02}) denote the true parameter, and let Θ_1 be a compact subset of $(-1, 1) \times (0, \infty)$ containing (ρ^0, σ^{02}) as an interior point. We show in Section S.8 that the preceding objective function divided by NT converges uniformly on Θ_1.

LEMMA S.1: *Under Assumption* A *and homoskedasticity*, *uniformly on the compact set* Θ_1, *as* $N, T \to \infty$,

$$\text{(S.3)} \qquad \frac{1}{NT}\ell_c\big(\rho, \sigma^2\big) \xrightarrow{p} -\frac{1}{2}\log\sigma^2 - \frac{1}{2}\frac{\sigma^{02}}{\sigma^2} - \frac{1}{2}\frac{\sigma^{02}}{\sigma^2}\big(\rho^0 - \rho\big)^2\frac{1}{1-\rho^{02}},$$

irrespective of how N, T *go to infinity*.

The objective function is uniquely maximized at (ρ^0, σ^{02}). This leads to the consistency of $(\widehat{\rho}, \widehat{\sigma}^2)$. Consistency of $(\widehat{\rho}, \widehat{\sigma}^2)$ implies consistency of $\widehat{\tau}$ for τ_N^0, that is, $\widehat{\tau} = \tau_N^0 + o_p(1)$, as is shown in Section S.8. This argument provides a direct and simple proof of consistency under large N and large T. We state the result as a theorem.

THEOREM S.1: *Under Assumption* A *and homoskedasticity of u_{it} over t, for $\theta_N^0 = (\rho^0, \sigma^{02}, \tau_N^0)$, we have, as $N, T \to \infty$,*

$$\widehat{\theta} = \theta_N^0 + o_p(1),$$

irrespective of how N and T go to infinity.

Except for the proof of consistency (Lemma S.1), it is unnecessary to make a distinction between (ρ, σ^2, τ_N) and $(\rho^0, \sigma^{02}, \tau_N^0)$. So we simply use (ρ, σ^2, τ_N) to denote the true parameter θ_N^0.

We show in the Appendix that, with $\theta_1 = (\rho, \sigma^2)'$,

$$-\frac{1}{NT} \frac{\partial^2 \ell_c}{\partial \theta_1 \, \partial \theta_1'} \bigg|_{\theta = \theta_N^0} \xrightarrow{p} \begin{bmatrix} \dfrac{1}{1-\rho^2} & 0 \\ 0 & \dfrac{1}{2\sigma^4} \end{bmatrix}.$$

It is tempting at this stage to appeal to the result of Amemiya (1985, p. 125) on concentrated likelihood function, which states that the joint distribution of $(\widehat{\rho}, \widehat{\sigma}^2)$ (with an appropriate normalization) is asymptotically normal with variance given by the inverse of the Hessian matrix. Such an approach cannot reveal the requirement on the relative rate at which N and T should tend to infinity. As demonstrated in the technical section of this supplement, we find it necessary to impose the condition $N/T^3 \to 0$. We state this result as a theorem.

THEOREM S.2: *Under the assumption of Theorem* S.1 *and normality of u_i, as $N, T \to \infty$ with $N/T^3 \to 0$, the estimator $\widehat{\theta}$ under the fixed-effects setup satisfies*

$$\sqrt{NT} \begin{bmatrix} \widehat{\rho} - \rho \\ \widehat{\sigma}^2 - \sigma^2 \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1-\rho^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \right).$$

There is no asymptotic bias, despite the fixed-effects setup; the estimators are centered at zero even under the scaling of $\sqrt{NT}$. The condition $N/T^3 \to 0$ is weaker than $0 \le \lim(N/T) \to c < \infty$. The latter condition is assumed by Theorem 5 of Alvarez and Arellano (2003) for the random-effects maximum likelihood estimator. The condition here is similar to the bias-corrected within-group estimator of Hahn and Kuersteiner (2002). However, unlike the bias-corrected within-group estimator, this estimator remains consistent under fixed T, as analyzed in Section S.2.

REMARK: The normality of u_i is only used in obtaining the limiting variance. Given the asymptotic representation for $(\widehat{\rho}, \widehat{\sigma}^2)$ in Section S.8 (see Lemma S.4 and (S.21), which do not assume normality), it is trivial to obtain the limiting distribution under nonnormality. We highlight that Theorem S.2 also holds for fixed N.

Given the rate of convergence, it is easy to establish that $\widehat{\tau}$ is also consistent for τ_N with the same rate of convergence. In fact, $\widehat{\tau}$ is a linear combination of $\widehat{\sigma}^2$ and $\widehat{\rho}$ plus an extra term (see Section S.8):

$$\sqrt{NT}(\widehat{\tau}-\tau_N) = -\frac{\tau}{\sigma^2}\sqrt{NT}\big(\widehat{\sigma}^2-\sigma^2\big) - \frac{2\tau}{1-\rho}\sqrt{NT}(\widehat{\rho}-\rho) + \frac{2}{\sigma^2\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}\eta_i + o_p(1), \tag{S.4}$$

where τ is the limit of τ_N. The third term on the right hand side converges in distribution to $N(0, 4\tau)$, which is also asymptotically independent of $\widehat{\sigma}^2$ and $\widehat{\rho}$ under normality of u_{it}. Therefore, $\widehat{\tau}$ is asymptotically normal with variance being the sum of the variances of the three terms on the right hand side of (S.4), which is equal to $2\tau^2 + 4\tau^2\frac{1+\rho}{1-\rho} + 4\tau$. Furthermore, from $\pi_N = \tau_N\sigma^2$ with $\pi_N = \frac{1}{N}\sum_{i=1}^N \eta_i^2$, if we define $\widehat{\pi} = \widehat{\tau}\widehat{\sigma}^2$, which is the MLE of π_N, then $\widehat{\pi}-\pi_N$ can be written as

$$\sqrt{NT}(\widehat{\pi}-\pi_N) = -\frac{2\pi_N}{1-\rho}\sqrt{NT}(\widehat{\rho}-\rho) + \frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}\eta_i + o_p(1).$$

The limiting variance is equal to $4\pi^2\frac{1+\rho}{1-\rho} + 4\pi\sigma^2$. We state the results as a corollary.

COROLLARY S.1: *Under the assumptions of Theorem S.2, with $\pi_N \to \pi$ and $\tau_N \to \tau = \pi/\sigma^2$, then*

$$\sqrt{NT}(\widehat{\tau}-\tau_N) \xrightarrow{d} N\left(0, 2\tau^2 + 4\tau^2\frac{1+\rho}{1-\rho} + 4\tau\right),$$

and

$$\sqrt{NT}(\widehat{\pi}-\pi_N) \xrightarrow{d} N\left(0, 4\pi^2\frac{1+\rho}{1-\rho} + 4\pi\sigma^2\right).$$

So the sample moment of the individual effects π_N and its ratio over the idiosyncratic variance $\tau_N = \pi_N/\sigma^2$ are all consistently estimable and asymptotically normal. These quantities are of practical interest.

The representation for $\widehat{\pi}$ also implies the asymptotic covariances between $\widehat{\pi}$ and $(\widehat{\rho}, \widehat{\sigma})$. For example, its asymptotic covariance with $\widehat{\rho}$ is equal to $-2\pi(1+\rho)$ (i.e., $-2\pi/(1-\rho)$ times the variance of $\sqrt{NT}(\widehat{\rho}-\rho)$). Therefore, the joint limiting distribution has the following form.

COROLLARY S.2: *Under the assumptions of Theorem* S.2,

$$\sqrt{NT}\begin{bmatrix} \widehat{\rho}-\rho \\ \widehat{\sigma}^2-\sigma^2 \\ \widehat{\pi}-\pi_N \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1-\rho^2 & 0 & -2\pi(1+\rho) \\ 0 & 2\sigma^4 & 0 \\ -2\pi(1+\rho) & 0 & 4\pi^2\dfrac{1+\rho}{1-\rho}+4\pi\sigma^2 \end{bmatrix}\right).$$

Similarly, we can easily derive the joint limit of $(\widehat{\rho}, \widehat{\sigma}^2, \widehat{\tau})$ from the representations for $\widehat{\rho}$, $\widehat{\sigma}^2$, and $\widehat{\tau}$. It is also straightforward to derive the limiting distributions under nonnormality given (S.21) and (S.4) in Section S.8. The joint distribution will depend on the skewness and kurtosis coefficients. Corollary S.2 holds under fixed N, and in this case, we can replace π occurring in the limit by π_N since no limit is taken with respect to N.

Throughout the analysis that led to the preceding results, we do not make any assumption about zero mean for η_i because we do not assume they are random variables, but rather fixed constants. This does make a difference in our analysis. For example, under i.i.d. zero mean, $N^{-1/2}\sum_{i=1}^N \eta_i$ is stochastically bounded, but with η_i being fixed constants, this quantity is $O(N^{1/2})$. A concise and yet self-contained proof for the theorems and the corollaries is provided in Section S.8 of this supplement.

S.6. INCIDENTAL PARAMETERS: TIME EFFECTS UNDER LARGE T

We consider the same model as in the previous section with the addition of time effects:

$$y_i = \Gamma\delta + \Gamma 1_T \eta_i + \Gamma u_i.$$

We estimate the time effects by subtracting the cross-section mean. Therefore, $S_N = \frac{1}{n}\sum_{i=1}^N (y_i - \bar{y})(y_i - \bar{y})'$ with $n = N-1$. Then $E(S_N) = \Sigma(\theta) = \sigma^2\Gamma(1_T 1_T'\tau_N + I_T)\Gamma'$, where $\tau_N = \pi_N/\sigma^2$ and $\pi_N = \frac{1}{n}\sum_{i=1}^N(\eta_i - \bar{\eta})^2$. The estimator is defined exactly the same as in the previous section except with the newly defined S_N. Despite the incidental-parameter problem over the time dimension, we show that, for the parameter $\widehat{\rho}$, the same limiting distribution holds. For the variance parameter $\widehat{\sigma}^2$, there is a bias term of $O(1/N)$ arising from estimating the time effects. We state this result as a theorem.

THEOREM S.3: *Under Assumption* A *and normality of* u_i, *as* $N, T \to \infty$ *with* $N/T^3 \to 0$, *then*

$$\sqrt{NT}\begin{bmatrix} \widehat{\rho}-\rho \\ \widehat{\sigma}^2 - \sigma^2\left(1-\dfrac{1}{N}\right) \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1-\rho^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix}\right).$$

Again, the normality assumption is used only for the limiting variance. Given the literature on the incidental-parameter problem (e.g., Neyman and Scott (1948)) and the GMM results of Alvarez and Arellano (2003), it is natural to conjecture that there might be some bias for $\widehat{\rho}$. Theorem S.3 proves otherwise. The variance estimator exhibits a bias of order $1/N$, which conforms with that of Neyman and Scott (1948). So under fixed N, $\widehat{\sigma}^2$ is inconsistent due to the time effects. Using representation (S.4), it is easy to show that $\sqrt{NT}(\widehat{\tau} - \tau_N(1+\frac{1}{N}))$ has the limiting distribution given in Corollary S.1. But there is no asymptotic bias for $\widehat{\pi}$. This implies that Corollary S.2 holds when σ^2 in the left hand side is replaced by $\sigma^2(1-1/N)$. We have the following.

COROLLARY S.3: *Under the assumptions of Theorem* S.3,

$$\sqrt{NT}\begin{bmatrix} \widehat{\rho}-\rho \\ \widehat{\sigma}^2 - \sigma^2\left(1-\dfrac{1}{N}\right) \\ \widehat{\pi}-\pi_N \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1-\rho^2 & 0 & -2\pi(1+\rho) \\ 0 & 2\sigma^4 & 0 \\ -2\pi(1+\rho) & 0 & 4\pi^2\dfrac{1+\rho}{1-\rho}+4\pi\sigma^2 \end{bmatrix}\right).$$

By imposing either $\frac{1}{T}\sum_{t=1}^T \delta_t = 0$ or $\bar{\eta} = \frac{1}{N}\sum_{i=1}^N \eta_i = 0$, each component of the time effects δ is estimated with $\sqrt{N}$ consistency. Recall that $\widehat{\delta} = \widehat{\Gamma}^{-1}\bar{y}$. By assuming $\bar{\eta} = 0$, then

$$\sqrt{N}(\widehat{\delta}_t - \delta_t) \xrightarrow{d} N(0, \sigma^2) \tag{S.5}$$

for each t. Under $\sum_{t=1}^T \delta_t = 0$, we define $\widehat{\delta}_t^\dagger = \widehat{\delta}_t - 1_T'\widehat{\delta}/T$ as the estimate for δ_t. The same limit holds for $\widehat{\delta}_t^\dagger$.

S.7. FIXED-T EFFICIENCY OF $\widehat{\rho}$ UNDER WEAKER ASSUMPTIONS AND NONNORMALITY

We shall assume that the $T\times 1$ vector u_i are i.i.d. over i. We show that $\widehat{\rho}$ is efficient under very mild conditions on u_{it}. Efficiency is in the sense that $\widehat{\rho}$ has the same limiting distribution as the optimal GMM discussed in the main text. Under fixed effects, no moment beyond $2+\epsilon$ order ($\epsilon > 0$) for u_{it} is required, and under random effects, no moment beyond the second order is required. Our argument is based on Anderson and Amemiya (1988). Although our factor model is different from the one considered by Anderson and Amemiya in that the factor loadings and the factor residual variance in dynamic panel model share the common parameters ρ, and that the factor residual variance is non-diagonal ($\Gamma\Phi\Gamma'$), their argument goes through.

Let $\widetilde{\sigma}_t^2 = \frac{1}{n}\sum_{i=1}^N (u_{it} - \bar{u}_t)^2$, the tth diagonal element of S_{uu} defined earlier. Let $\widetilde{\theta}_N^0 = (\rho, \pi_N, \widetilde{\sigma}_1^2, \dots, \widetilde{\sigma}_T^2)$. Anderson and Amemiya (1988) considered the distribution of $\widehat{\theta}$ centered at $\widetilde{\theta}_N^0$, instead of θ_N^0. Define $\widetilde{\Phi} = \operatorname{diag}(\widetilde{\sigma}_1^2, \dots, \widetilde{\sigma}_T^2)$ and $\Sigma(\widetilde{\theta}_N^0) = \Gamma(1_T 1_T' \pi_N + \widetilde{\Phi})\Gamma'$. Notice that

$$s - g(\widetilde{\theta}_N^0) = \operatorname{vech}[H + H' + \Gamma(S_{uu} - \widetilde{\Phi})\Gamma'].$$

The diagonal elements of $S_{uu} - \widetilde{\Phi}$ are zero, so $s - g(\widetilde{\theta}_N^0)$ does not involve u_{it}^2. The estimator based on the Wishart likelihood has the asymptotic representation, under fixed T,

$$\text{(S.6)} \qquad \sqrt{n}(\widehat{\theta} - \widetilde{\theta}_N^0) = D(\theta^0)'\sqrt{n}[s - g(\widetilde{\theta}_N^0)] + o_p(1),$$

where θ^0 is the limit of θ_N^0 (also the limit of $\widetilde{\theta}_N^0$), and $D(\theta^0)$ has full column rank. The above is easy to show and is based on the consistency of $\widehat{\theta}$ and the Delta-method; see Anderson and Amemiya (1988). So if $\sqrt{n}[s - g(\widetilde{\theta}_N^0)]$ is asymptotically normal, then $\sqrt{n}(\widehat{\theta} - \widetilde{\theta}_N^0)$ is asymptotically normal. Note that $s - g(\widetilde{\theta}_N^0)$ only involves elements of H and the elements of S_{uu} that are strictly below the diagonal. So asymptotic normality for $s - g(\widetilde{\theta}_N^0)$ can be achieved under mild conditions. We next state some primitive conditions, under which the limiting distribution of $\sqrt{n}[s - g(\widetilde{\theta}_N^0)]$ is asymptotically normal and, furthermore, the limiting variance is the same as if normality of u_i were assumed. Since under normality, we have $\sqrt{N}(\widehat{\rho} - \rho) \xrightarrow{d} N(0, v_\rho)$, where v_ρ is the $(1,1)$th element of $(G'\Omega G)^{-1}$, it follows that, under the weak conditions, we still have the same limiting result for $\widehat{\rho}$ because the first component of $\sqrt{n}(\widehat{\theta} - \widetilde{\theta}_N^0)$ is equal to $\sqrt{n}(\widehat{\rho} - \rho)$.

The off-diagonal elements of $\Gamma[\frac{1}{n}\sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})' - \widetilde{\Phi}]\Gamma'$ are linear combinations of $\frac{1}{n}\sum_{i=1}^N (u_{it} - \bar{u}_t)(u_{ih} - \bar{u}_h) = \frac{1}{n}\sum_{i=1}^N u_{it}u_{ih} - \bar{u}_t\bar{u}_h = \frac{1}{n}\sum_{i=1}^N u_{it}u_{ih} + O_p(\frac{1}{N})$, where $t \neq h$. The combination coefficients depend on the matrix Γ. Upon multiplying by $\sqrt{n}$, the limit of $\frac{1}{n}\sum_{i=1}^N u_{it}u_{ih}$ is asymptotically normal $N(0, \sigma_t^2\sigma_h^2)$, for $t \neq h$, if u_{it} are independent over t. This limiting distribution only requires the existence of a second moment, and the limit is the same as if normality of u_i were assumed. A typical element of H is a linear combination (over t) of $\frac{1}{n}\sum_{i=1}^{N}(u_{it}-\bar{u}_t)(\eta_i-\bar{\eta})=\frac{1}{n}\sum_{i=1}^{N}u_{it}\eta_i-\bar{u}_t\bar{\eta}$. If we assume η_i are i.i.d. and independent of u_i, then, for each given t, $u_{it}\eta_i$ are i.i.d. over i, so the CLT holds for $n^{-1/2}\sum_{i=1}^{N}u_{it}\eta_i$ provided that the variance of u_{it} and of η_i is finite. If we treat η_i as fixed constants, then $u_{it}\eta_i$ are independent, but not identically distributed. The Lyapunov sufficient condition is $E|u_{it}|^{2+\epsilon}<\infty$ and $\sum_{i=1}^{N}|\eta_i|^{2+\epsilon}/(\sum_{i=1}^{N}\eta_i^2)^{2+\epsilon}\to 0$. A sufficient condition for the latter is $\frac{1}{N}\sum_{i=1}^{N}|\eta_i|^{2+\epsilon}$ and $\frac{1}{N}\sum_{i=1}^{N}\eta_i^2$ have positive limits. In either case (random or fixed η_i), we have $n^{-1/2}\sum_{i=1}^{N}(u_{it}-\bar{u}_t)(\eta_i-\bar{\eta})$ converging weakly to $N(0,\sigma_t^2\pi)$, the same limit as if normality of u_{it} were assumed. Summarizing the preceding argument, we state the conditions under which $\sqrt{n}[s-g(\widetilde{\theta}_N^0)]$ behaves as if u_{it} were normal.

ASSUMPTION S1: *The $T\times 1$ vector u_i are i.i.d. with zero mean and diagonal covariance matrix Φ. Furthermore, u_{it} are independent over t.*

ASSUMPTION S2: *One of the following conditions holds*:

(i) *the individual effects η_i are i.i.d. with finite second moment, and are independent of u_i.*

(ii) *η_i are fixed constants; for some $\epsilon>0$, both $\frac{1}{N}\sum_{i=1}^{N}|\eta_i|^{2+\epsilon}$ and $\frac{1}{N}\sum_{i=1}^{N}\eta_i^2$ have positive limits; and for each t, $E|u_{it}|^{2+\epsilon}<\infty$.*

THEOREM S.4: *Assume that Assumptions S1 and S2 hold. Under fixed T, $\sqrt{N}(\widehat{\rho}-\rho)\xrightarrow{d}N(0,v_\rho)$, where v_ρ is the $(1,1)$th element of $(G'\Omega G)^{-1}$, where G and Ω are defined in the main text.*

If Assumptions S1 and S2(i) hold, the theorem holds without the requirement of a moment beyond the second order. This is a strong and perhaps somewhat surprising result, as most of the random-effects literature on dynamic model requires finite fourth moment, and relies on large T as well as stationarity. Here efficiency is obtained for fixed T without time series stationarity (neither mean stationarity nor covariance stationarity). On the other hand, the result should come as no surprise. In a pure time series regression model, $y_t=\rho y_{t-1}+\varepsilon_t$, if ε_t are i.i.d., then finite variance of ε_t is sufficient for $\widehat{\rho}$ to have asymptotic result identical to normal ε_t provided that the time series is strictly stationary and T grows to infinity. If y_t is not strictly stationary, then $2+\epsilon$ moment is sufficient; see Davidson (2000, p. 129). Here, with fixed T, the time series is nonstationary because of heteroskedasticity, and also because of the first observation not being drawn from a stationary distribution. But the cross-sectional central limit theorem for u_i drives the underlying results.

Because only $2+\epsilon$ moment (at most) is required instead of the usual fourth moment, Assumptions S1 and S2 are satisfied by a large class of distributions. These assumptions are sufficient for $\sqrt{N}(\widehat{\rho}-\rho)$ to be as efficient as under normality. However, if one is interested in the distribution of $\sqrt{N}(\widehat{\sigma}_t^2 - \sigma_t^2)$, fourth moment will be required.

S.8. TECHNICAL DETAILS

PROOF OF LEMMA S.1: From

$$y_i y_i' = \Gamma 1_T 1_T' \Gamma' \eta_i^2 + \Gamma u_i u_i' \Gamma' + \Gamma 1_T u_i' \Gamma' \eta_i + \Gamma u_i 1_T' \Gamma' \eta_i,$$

we have

$$\frac{1}{N}\sum_{i=1}^{N} y_i y_i' = \sigma^2 \Gamma 1_T 1_T' \Gamma' \tau_N + \sigma^2 \Gamma \Gamma' + \Gamma \frac{1}{N}\sum_{i=1}^{N}\big(u_i u_i' - \sigma^2 I_T\big)\Gamma' + \Gamma 1_T \frac{1}{N}\sum_{i=1}^{N} u_i' \Gamma' \eta_i + \Gamma \frac{1}{N}\sum_{i=1}^{N} u_i 1_T' \Gamma' \eta_i.$$

For the proof of consistency, we need to distinguish the true parameters $(\rho^0, \sigma^{02}, \tau_N^0)$ from the variables (ρ, σ^2, τ_N) in the likelihood function. So let Γ^0 denote the Γ matrix when $\rho = \rho^0$. Then

$$S_N = \Sigma\big(\theta_N^0\big) + \Gamma^0 \frac{1}{N}\sum_{i=1}^{N}\big(u_i u_i' - \sigma^{02} I_T\big)\Gamma^{0\prime} + \Gamma^0 1_T \frac{1}{N}\sum_{i=1}^{N} u_i' \Gamma^{0\prime} \eta_i + \Gamma^0 \frac{1}{N}\sum_{i=1}^{N} u_i 1_T' \Gamma^{0\prime} \eta_i,$$

where $\Sigma(\theta_N^0) = \sigma^{02}\Gamma^0(1_T 1_T' \tau_N^0 + I_T)\Gamma^{0\prime}$. Thus

$$BS_N B' = B\Sigma\big(\theta_N^0\big)B' + B\Gamma^0 \frac{1}{N}\sum_{i=1}^{N}\big(u_i u_i' - \sigma^{02} I_T\big)\Gamma^{0\prime}B' + B\Gamma^0 1_T \frac{1}{N}\sum_{i=1}^{N} u_i' \Gamma^{0\prime} B' \eta_i + B\Gamma^0 \frac{1}{N}\sum_{i=1}^{N} u_i 1_T' \Gamma^{0\prime} B' \eta_i.$$

Note matrix B is a function of ρ. Let d_1 and d_2 denote the constants defined by

$$\mathrm{tr}\big(B\Sigma\big(\theta_N^0\big)B'\big) = \sigma^{02} d_1, \quad 1_T' B\Sigma\big(\theta_N^0\big)B' 1_T = \sigma^{02} d_2.$$

Notice that $B\Gamma^0 = I_T + (\rho^0 - \rho)L^0$, where L^0 denotes the matrix L when $\rho = \rho^0$; we obtain

$$B\Sigma\big(\theta_N^0\big)B' = \sigma^{02}\big[I_T + \big(\rho^0 - \rho\big)L^0\big]\big(1_T 1_T' \tau_N^0 + I_T\big)\big[I_T + \big(\rho^0 - \rho\big)L^0\big]'.$$

Then

$$
\begin{aligned}
d_1 &= \operatorname{tr}\bigl(B\Sigma\bigl(\theta_N^0\bigr)B'\bigr)/\sigma^{02} \\
&= \bigl(1+\tau_N^0\bigr)T + 2\bigl(\rho^0-\rho\bigr)\tau_N^0 1_T' L^0 1_T \\
&\quad + \bigl(\rho^0-\rho\bigr)^2\bigl[1_T' L^0 L^{0\prime} 1_T \tau_N^0 + \operatorname{tr}\bigl(L^0 L^{0\prime}\bigr)\bigr], \\
d_2 &= \bigl(1_T' B\Sigma\bigl(\theta_N^0\bigr)B' 1_T\bigr)/\sigma^{02} \\
&= T\bigl(T\tau_N^0+1\bigr) + 2\bigl(\rho^0-\rho\bigr)\bigl(T\tau_N^0+1\bigr)1_T' L^0 1_T \\
&\quad + \bigl(\rho^0-\rho\bigr)^2\bigl[\bigl(1_T' L^0 1_T\bigr)^2\tau_N^0 + 1_T' L^0 L^{0\prime} 1_T\bigr].
\end{aligned}
$$

Using the following limits, proved in Moreira (2009):

$$
\begin{aligned}
&\frac{1_T' L^0 1_T}{T} \to \frac{1}{1-\rho^0}, \\
&\frac{1_T' L^{0\prime} L^0 1_T}{T} \to \frac{1}{(1-\rho^0)^2}, \\
&\frac{\operatorname{tr}(L^{0\prime} L^0)}{T} \to \frac{1}{1-\rho^{02}},
\end{aligned}
$$

we obtain

$$
\begin{aligned}
&\frac{d_1}{T} \to 1+\tau_N^0 + 2\tau_N^0\left(\frac{\rho^0-\rho}{1-\rho^0}\right) + \bigl(\rho^0-\rho\bigr)^2\left[\frac{\tau_N^0}{(1-\rho^0)^2} + \frac{1}{1-\rho^{02}}\right], \\
&\frac{d_2}{T^2} \to \tau_N^0 + 2\tau_N^0\left(\frac{\rho^0-\rho}{1-\rho^0}\right) + \bigl(\rho^0-\rho\bigr)^2\frac{\tau_N^0}{(1-\rho^0)^2},
\end{aligned}
$$

and

$$
\frac{d_1}{T} - \frac{d_2}{T^2} \to 1 + \bigl(\rho^0-\rho\bigr)^2\frac{1}{1-\rho^{02}}.
$$

Further, it is easy to show the following.

LEMMA S.2: *As $T\to\infty$, regardless of N, uniformly for ρ in a compact subset of $(-1,1)$, we have*

(i) $\frac{1}{T}\operatorname{tr}[B\Gamma^0\frac{1}{N}\sum_{i=1}^N(u_iu_i' - \sigma^{02}I_T)\Gamma^{0\prime}B'] = o_p(1)$,
(ii) $\frac{1}{T}\operatorname{tr}[B\Gamma^0 1_T\frac{1}{N}\sum_{i=1}^N u_i'\Gamma^{0\prime}B'\eta_i] = o_p(1)$,
(iii) $\frac{1}{T^2}(1_T' B\Gamma^0\frac{1}{N}\sum_{i=1}^N(u_iu_i' - \sigma^{02}I_T)\Gamma^{0\prime}B')1_T = o_p(1)$,
(iv) $\frac{1}{T^2}1_T'(B\Gamma^0 1_T\frac{1}{N}\sum_{i=1}^N u_i'\Gamma^{0\prime}B'\eta_i)1_T = o_p(1)$.

The proof is easy, and thus omitted.

By the definition of the concentrated objective function,

$$\frac{1}{nT}\ell_c(\rho,\sigma^2) = -\frac{1}{2}\log\sigma^2 - \frac{1}{2}\frac{\sigma^{02}}{\sigma^2}\left(\frac{d_1}{T} - \frac{d_2}{T^2}\right) + o_p(1)$$
$$= -\frac{1}{2}\log\sigma^2 - \frac{1}{2}\frac{\sigma^{02}}{\sigma^2} - \frac{1}{2}\frac{\sigma^{02}}{\sigma^2}(\rho^0-\rho)^2\frac{1}{1-\rho^{02}} + o_p(1).$$

The first equality (especially the $o_p(1)$ term) follows from Lemma S.2 and $\frac{1}{T}\log(\frac{1}{T}1_T' BS_N B' 1_T) = O_p(\log(T)/T) = o_p(1)$ uniformly on the compact set Θ_1. The second equality follows from the limit for $(d_1/T - d_2/T^2)$ derived earlier. This proves Lemma S.1. *Q.E.D.*

Except for Lemmas S.1 and S.2, there is no need to carry the superscript 0 for the true parameters ρ^0, σ^{02}, and τ_N^0. In what follows, the MLE is denoted by $(\widehat{\rho}, \widehat{\sigma}^2, \widehat{\tau})$, and the true parameter vector θ_N^0 denotes (ρ, σ^2, τ_N), and the matrices B, Γ, and L are all evaluated at the true parameter ρ, except when indicated otherwise.

PROOF OF THEOREM S.1: Lemma S.1 implies the consistency of $(\widehat{\rho}, \widehat{\sigma}^2)$. This follows from standard argument as in Amemiya (1985) or Newey and McFadden (1994). It remains to show that $\widehat{\tau}$ is consistent. Subtracting and adding terms,

$$\widehat{\tau} - \tau_N = \frac{1}{\widehat{\sigma}^2}\frac{1_T'\widehat{B}S_N\widehat{B}'1_T}{T^2} - \tau_N - \frac{1}{T}$$
$$= \left(\frac{\sigma^2 - \widehat{\sigma}^2}{\widehat{\sigma}^2\sigma^2}\right)\frac{1_T'\widehat{B}S_N\widehat{B}'1_T}{T^2} + \frac{2}{\sigma^2}\left(\frac{1_T'(\widehat{B}-B)S_N B'1_T}{T^2}\right)$$
$$+ \frac{1}{\sigma^2}\left(\frac{1_T'(\widehat{B}-B)S_N(\widehat{B}-B)'1_T}{T^2}\right)$$
$$+ \left(\frac{1_T'BS_N B'1_T}{\sigma^2 T^2} - \tau_N - \frac{1}{T}\right).$$

The first term on the right being $o_p(1)$ follows from the consistency of $\widehat{\sigma}^2$. Owing to the consistency of $\widehat{\rho}$, the next two terms are $o_p(1)$. For example, consider the second term. Using $\widehat{B} - B = -(\widehat{\rho} - \rho)J_T$ and $1_T'J_T S_N B' 1_T/T^2 = O_p(1)$ (see Lemma S.3(vi) below), the second term is $o_p(1)$ since $\widehat{\rho} - \rho = o_p(1)$. For the last term, notice that

$$S_N = \sigma^2\Gamma(1_T 1_T'\tau_N + I_T)\Gamma' + \Gamma 1_T \frac{1}{N}\sum_{i=1}^N u_i'\Gamma'\eta_i \tag{S.7}$$
$$+ \Gamma\frac{1}{N}\sum_{i=1}^N u_i 1_T'\Gamma'\eta_i + \Gamma\frac{1}{N}\sum_{i=1}^N (u_i u_i' - \sigma^2 I_T)\Gamma'.$$

Since B is evaluated at the true parameter ρ, we have $B\Gamma = I_T$ and

$$\text{(S.8)} \qquad 1_T' B S_N B' 1_T = \sigma^2\big(T^2\tau_N + T\big) + 2T\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}\eta_i + \frac{1}{N}\sum_{i=1}^{N}\left[\left(\sum_{t=1}^{T} u_{it}\right)^2 - T\sigma^2\right]$$

and

$$\text{(S.9)} \qquad \frac{1}{\sigma^2 T^2} 1_T' B S_N B' 1_T - \tau_N - \frac{1}{T} = \frac{2}{\sigma^2 T}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}\eta_i + \frac{1}{\sigma^2 TN}\sum_{i=1}^{N}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}\right)^2 - \sigma^2\right],$$

which is in fact $O_p(1/\sqrt{NT})$. Combining results, we obtain $\widehat{\tau} - \tau_N = o_p(1)$. *Q.E.D.*

First Order Conditions and the Hessian Matrix

The first order conditions for the concentrated likelihood function are

$$\frac{\partial \ell_c}{\partial \rho} = \frac{N}{\sigma^2}\operatorname{tr}\big(J_T S_N B'\big) - \frac{N}{\sigma^2}\left[\frac{T\widetilde{\tau}_N}{1 + T\widetilde{\tau}_N}\right]\frac{1}{T}\big(1_T' J_T S_N B' 1_T\big),$$

$$\frac{\partial \ell_c}{\partial \sigma^2} = -\frac{N(T-1)}{2}\frac{1}{\sigma^2} + \frac{N}{2\sigma^4}\operatorname{tr}\big(B S_N B'\big) - \frac{N}{2\sigma^4}\frac{1}{T}\big(1_T' B S_N B' 1_T\big),$$

where $\widetilde{\tau}_N$ is a function of (ρ, σ^2) as a result of concentration:

$$\text{(S.10)} \qquad \widetilde{\tau}_N = \frac{1}{\sigma^2}\frac{1}{T^2}\big(1_T' B S_N B' 1_T\big) - \frac{1}{T}.$$

To derive the Hessian matrix, using

$$\frac{\partial \widetilde{\tau}_N}{\partial \rho} = -2\frac{1}{\sigma^2}\frac{1}{T^2}\big(1_T' J_T S_N B' 1_T\big),$$

we obtain

$$\frac{\partial^2 \ell_c}{\partial \rho^2} = -\frac{N}{\sigma^2}\operatorname{tr}\big(J_T S_N J_T'\big) + \frac{N}{\sigma^2}\left[\frac{T\widetilde{\tau}_N}{1 + T\widetilde{\tau}_N}\right]\frac{1}{T}\big(1_T' J_T S_N J_T' 1_T\big) + \frac{N}{\sigma^4}\frac{2T^2}{(1 + T\widetilde{\tau}_N)^2}\left(\frac{1_T' J_T S_N B' 1_T}{T^2}\right)^2,$$

$$\frac{\partial^2 \ell_c}{\partial(\sigma^2)^2} = \frac{N(T-1)}{2}\frac{1}{\sigma^4} - \frac{N}{\sigma^6}\operatorname{tr}\bigl(BS_N B'\bigr) + \frac{N}{\sigma^6}\frac{1}{T}\bigl(1_T' BS_N B' 1_T\bigr),$$

$$\frac{\partial^2 \ell_c}{\partial\sigma^2\,\partial\rho} = -\frac{N}{\sigma^4}\operatorname{tr}\bigl(J_T S_N B'\bigr) + \frac{N}{\sigma^4}\frac{1}{T}\bigl(1_T' J_T S_N B' 1_T\bigr).$$

LEMMA S.3: *Evaluated at the true parameters θ_N^0, as $T \to \infty$, regardless of N (fixed or going to infinity),*

(i) $\frac{1}{T}\operatorname{tr}(J_T S_N J_T') \xrightarrow{p} \sigma^2\tau\frac{1}{(1-\rho)^2} + \frac{\sigma^2}{1-\rho^2}$,

(ii) $\frac{1}{T^2}(1_T' J_T S_N J_T' 1_T) \xrightarrow{p} \sigma^2\tau\frac{1}{(1-\rho)^2}$,

(iii) $\frac{1}{T}\operatorname{tr}(BS_N B') \xrightarrow{p} \sigma^2(1+\tau)$,

(iv) $\frac{1}{T^2}(1_T' BS_N B' 1_T) \xrightarrow{p} \sigma^2\tau$,

(v) $\frac{1}{T}\operatorname{tr}(J_T S_N B') \xrightarrow{p} \sigma^2\tau\frac{1}{1-\rho}$,

(vi) $\frac{1}{T^2}(1_T' J_T S_N B' 1_T) \xrightarrow{p} \sigma^2\tau\frac{1}{1-\rho}$,

where τ is the limit of τ_N. If N is fixed, we use τ_N in place of τ.

PROOF: The proof of this lemma uses the following facts:

$$S_N = \sigma^2 \Gamma 1_T 1_T' \Gamma' \tau_N + \Gamma 1_T \frac{1}{N}\sum_{i=1}^N u_i' \Gamma' \eta_i + \Gamma\frac{1}{N}\sum_{i=1}^N u_i 1_T' \Gamma' \eta_i + \Gamma\frac{1}{N}\sum_{i=1}^N u_i u_i' \Gamma',$$

and at the true parameters, $J_T\Gamma = L$ and $B = \Gamma^{-1}$.

Consider (i):

$$J_T S_N J_T' = \sigma^2 L 1_T 1_T' L' \tau_N + L\frac{1}{N}\sum_{i=1}^N u_i u_i' L' + L 1_T \frac{1}{N}\sum_{i=1}^N u_i' L' \eta_i + L\frac{1}{N}\sum_{i=1}^N u_i 1_T' L' \eta_i.$$

Thus

$$\text{(S.11)}\qquad \frac{1}{T}\operatorname{tr}\bigl(J_T S_N J_T'\bigr) = \sigma^2 \frac{1_T' L' L 1_T}{T}\tau_N + \sigma^2\frac{1}{T}\operatorname{tr}\bigl(L'L\bigr)$$

$$+\frac{1}{TN}\sum_{i=1}^{N}\left[u_i'L'Lu_i-\sigma^2\operatorname{tr}(L'L)\right] \tag{S.12}$$

$$+2\frac{1}{TN}\sum_{i=1}^{N}u_i'L'L1_T\eta_i.$$

The first two terms have the stated limit. The last two terms are $o_p(1)$.

Consider (ii). Using the expression for $J_TS_NJ_T'$ in (i),

$$\frac{1}{T^2}1_T'J_TS_NJ_T'1_T=\sigma^2\left(\frac{1_T'L1_T}{T}\right)^2\tau_N \tag{S.13}$$

$$+\frac{1}{T^2N}\sum_{i=1}^{N}\left(1_T'Lu_i\right)^2+2\frac{1_T'L1_T}{T}\frac{1}{TN}\sum_{i=1}^{N}\left(1_T'Lu_i\right)\eta_i.$$

The first term has the stated limit, the second is $O_p(\frac{1}{T})$, and the third is $O_p((NT)^{-1/2})$.

For (iii), using $B\Gamma=I_T$, we have

$$\operatorname{tr}\left(BS_NB'\right)=\sigma^2(1+\tau_N)T+2\frac{1}{N}\sum_{i=1}^{N}u_i'1_T\eta_i+\frac{1}{N}\sum_{i=1}^{N}\left(u_i'u_i-T\sigma^2\right). \tag{S.14}$$

Divided by T, the last two terms are $O_p((NT)^{-1/2})$, and the first term has the stated limit.

Result (iv) is already implied by (S.9).

Consider (v). At the true parameters, $J_T\Gamma=L$ and $B=\Gamma^{-1}$,

$$J_TS_NB'=\sigma^2L1_T1_T'\tau_N+L\frac{1}{N}\sum_{i=1}^{N}u_iu_i'$$

$$+L1_T\frac{1}{N}\sum_{i=1}^{N}u_i'\eta_i+L\frac{1}{N}\sum_{i=1}^{N}u_i1_T'\eta_i,$$

$$\operatorname{tr}\left(J_TS_NB'\right)=\sigma^2\left(1_T'L1_T\right)\tau_N+\frac{1}{N}\sum_{i=1}^{N}u_i'Lu_i \tag{S.15}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\left[u_i'(L+L')1_T\right]\eta_i,$$

$$u_i'Lu_i=\sum_{t=2}^{T}u_{it}\left(\sum_{s=0}^{t}\rho^ju_{it-s}\right)=\sum_{t=2}^{T}u_{it}w_{it-1},$$

where $w_{it} = \rho w_{it-1} + u_{it}$ with $w_{i0} = 0$. Divided by T, the first term has the stated limit and the last two terms are $O_p((NT)^{-1/2})$.

Consider (vi):

$$1_T' J_T S_N B' 1_T = \sigma^2 T \big(1_T' L 1_T\big) \tau_N + \frac{1}{N} \sum_{i=1}^{N} \big(1_T' L u_i\big)\big(u_i' 1_T\big) + \big(1_T' L 1_T\big) \frac{1}{N} \sum_{i=1}^{N} \big(u_i' 1_T\big) \eta_i + T \frac{1}{N} \sum_{i=1}^{N} \big(1_T' L u_i\big) \eta_i. \tag{S.16}$$

Divided by T^2,

$$\frac{1_T' J_T S_N B' 1_T}{T^2} = \sigma^2 \frac{1_T' L 1_T}{T} \tau_N + \frac{1}{T^2 N} \sum_{i=1}^{N} \big(1_T' L u_i\big)\big(u_i' 1_T\big) + \frac{(1_T' L 1_T)}{T} \frac{1}{TN} \sum_{i=1}^{N} \big(u_i' 1_T\big) \eta_i + \frac{1}{TN} \sum_{i=1}^{N} \big(1_T' L u_i\big) \eta_i. \tag{S.17}$$

The first term on the right hand side has the stated limit, the second term is $O_p(T^{-1})$, and the last terms are each $O_p((NT)^{-1/2})$. *Q.E.D.*

LEMMA S.4: *Under the assumptions of Theorem S.2, as $T \to \infty$, regardless of N, with $\theta_1 = (\rho, \sigma^2)'$,*

$$-\frac{1}{NT} \frac{\partial^2 \ell_c}{\partial \theta_1 \, \partial \theta_1'} \bigg|_{\theta = \theta_N^0} \xrightarrow{p} \begin{bmatrix} \dfrac{1}{1-\rho^2} & 0 \\ 0 & \dfrac{1}{2\sigma^4} \end{bmatrix}.$$

PROOF: This follows from the expressions for the second order derivatives, Lemma S.3, and $T^2/(1 + T\widetilde{\tau}_N)^2 \xrightarrow{p} 1/\tau^2$ when evaluated at the true parameters. Note that the limit of $\frac{1}{NT} \frac{\partial^2 \ell_c}{\partial \rho^2}$ is determined by the first two terms; the third term is $O_p(T^{-1})$. *Q.E.D.*

LEMMA S.5: *Under the assumptions of Theorem S.2, evaluated at the true parameters, as $T \to \infty$ with arbitrary N (including fixed N) such that $N/T^3 \to 0$,*

$$\frac{1}{\sqrt{NT}} \begin{bmatrix} \dfrac{\partial \ell_c}{\partial \rho} \\ \dfrac{\partial \ell_c}{\partial \sigma^2} \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \dfrac{1}{1-\rho^2} & 0 \\ 0 & \dfrac{1}{2\sigma^4} \end{bmatrix} \right).$$

PROOF:

$$\frac{\partial \ell_c}{\partial \rho} = \frac{N}{\sigma^2}\operatorname{tr}(J_T S_N B') - \frac{N}{\sigma^2}\frac{1}{T}(1_T' J_T S_N B' 1_T) + \frac{N}{\sigma^2}\left[\frac{1}{1+T\widetilde{\tau}_N}\right]\frac{1}{T}(1_T' J_T S_N B' 1_T).$$

Thus

$$\frac{1}{\sqrt{NT}}\frac{\partial \ell_c}{\partial \rho} = \frac{1}{\sigma^2}\left(\frac{N}{T}\right)^{1/2}\left\{\operatorname{tr}(J_T S_N B') - \frac{1}{T}(1_T' J_T S_N B' 1_T) + \frac{1_T' L 1_T}{T}\sigma^2\right\} + \frac{1}{\sigma^2}\left(\frac{N}{T}\right)^{1/2}\left\{\left[\frac{T}{1+T\widetilde{\tau}_N}\right]\frac{1}{T^2}(1_T' J_T S_N B' 1_T) - \frac{1_T' L 1_T}{T}\sigma^2\right\}.$$

In the above, we add and subtract the term $1_T' L 1_T \sigma^2/T$. We show that, as $T \to \infty$, the first term converges to $N(0, 1/(1-\rho^2))$ regardless of N, and the second term is negligible when N is fixed or $N \to \infty$ with $N/T^3 \to 0$. From (S.16), dividing by T,

$$\frac{1_T' J_T S_N B' 1_T}{T} = \sigma^2(1_T' L 1_T)\tau_N + \frac{1}{TN}\sum_{i=1}^N (1_T' L u_i)(u_i' 1_T) + \frac{(1_T' L 1_T)}{T}\frac{1}{N}\sum_{i=1}^N (u_i' 1_T)\eta_i + \frac{1}{N}\sum_{i=1}^N (1_T' L u_i)\eta_i.$$

Together with (S.15),

$$\begin{aligned}
&\operatorname{tr}(J_T S_N B') - \frac{1_T' J_T S_N B 1_T}{T} \\
&\quad = \frac{1}{N}\sum_{i=1}^N u_i' L u_i + \frac{1}{N}\sum_{i=1}^N u_i' L 1_T \eta_i \\
&\qquad - \frac{1}{TN}\sum_{i=1}^N (1_T' L u_i)(u_i' 1_T) - \frac{(1_T' L 1_T)}{T}\frac{1}{N}\sum_{i=1}^N (u_i' 1_T)\eta_i.
\end{aligned}$$

Except for the third term, all terms have zero expectation. Subtract the expectation of the third term, which is $\sigma^2 1_T' L 1_T/T$:

$$\begin{aligned}&\frac{1}{TN}\sum_{i=1}^{N}(1_T' L u_i)(u_i' 1_T) - \sigma^2 1_T' L 1_T/T \\ &\quad = \frac{1}{N}\sum_{i=1}^{N}[a_i - E(a_i)] = O_p\left(\frac{1}{\sqrt{N}}\right),\end{aligned}$$

where $a_i = (\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1} w_{it-1})(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it})$. We have used the fact that $u_i' 1_T = \sum_{t=1}^{T} u_{it}$ and $1_T' L u_i = \sum_{t=1}^{T-1} w_{it-1}$, where $w_{it} = \rho w_{it-1} + u_{it}$ with $w_{i0} = 0$. Thus,

$$\begin{aligned}&(N/T)^{1/2}\left\{\operatorname{tr}(J_T S_N B') - \frac{1_T' J_T S_N B 1_T}{T} + \frac{1_T' L 1_T}{T}\sigma^2\right\} \\ &\quad = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} u_i' L u_i + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} u_i' L 1_T \eta_i \\ &\qquad - \frac{(1_T' L 1_T)}{T}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(u_i' 1_T)\eta_i + O_p(T^{-1/2}).\end{aligned}$$

Note that $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} u_i' L 1_T \eta_i - \frac{(1_T' L 1_T)}{T}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(u_i' 1_T)\eta_i = O_p(T^{-1/2})$ because its variance is

$$\left[\frac{1_T' L' L 1_T}{T} - \left(\frac{1_T' L 1_T}{T}\right)^2\right]\tau_N \sigma^2 = O\left(\frac{1}{T}\right)\tau_N\sigma^2 \to 0.$$

Next,

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} u_i' L u_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} w_{it-1} u_{it} \xrightarrow{d} \sigma^2 N\big(0, 1/(1-\rho^2)\big).$$

We shall show that

$$\begin{aligned}\text{(S.18)}\quad &\frac{1}{\sigma^2}\left(\frac{N}{T}\right)^{1/2}\left\{\left[\frac{T}{1+T\widetilde{\tau}_N}\right]\frac{1}{T^2}(1_T' J_T S_N B' 1_T) - \frac{1_T' L 1_T}{T}\sigma^2\right\} \\ &\quad = O_p(1/T) + O_p(N^{1/2}/T^{3/2}).\end{aligned}$$

It is already shown in (S.17) that

$$\frac{1}{T^2}(1_T' J_T S_N B' 1_T) = \frac{(1_T' L 1_T)}{T}\tau_N \sigma^2 + O_p(1/T) + O_p\big((NT)^{-1/2}\big).$$

Let a_N denote the term $O_p(1/T)+O_p((NT)^{-1/2})$ for the moment; then

$$\begin{aligned}
&\frac{T}{1+T\widetilde{\tau}_N}\frac{(1_T'J_TS_NB'1_T)}{T^2}-\frac{(1_T'L1_T)}{T}\sigma^2\\
&=\frac{T}{1+T\widetilde{\tau}_N}\left[\frac{(1_T'L1_T)}{T}\tau_N\sigma^2+a_N\right]-\frac{(1_T'L1_T)}{T}\sigma^2\\
&=\left(\frac{T\tau_N}{1+T\widetilde{\tau}_N}-1\right)\frac{(1_T'L1_T)}{T}\sigma^2+\frac{T}{1+T\widetilde{\tau}_N}a_N\\
&=\frac{T(\tau_N-\widetilde{\tau}_N)}{1+T\widetilde{\tau}_N}\frac{(1_T'L1_T)}{T}+O_p(1/T)+a_NO_p(1)\\
&=(\tau_N-\widetilde{\tau}_N)O_p(1)+O_p\bigl(T^{-1}\bigr)+a_NO_p(1)\\
&=O_p\bigl((NT)^{-1/2}\bigr)+O_p(1/T).
\end{aligned}$$

The third and fourth equalities follow from $T/(1+T\widetilde{\tau}_N)=O_p(1)$, and the last equality follows from $\sqrt{NT}(\tau_N-\widetilde{\tau}_N)=O_p(1)$ when $\widetilde{\tau}_N$ is evaluated at the true parameters (ρ,σ^2). Multiplied by $(N/T)^{1/2}$, the whole expression becomes $O_p(T^{-1})+O_p(N^{1/2}/T^{3/2})$. This proves (S.18). In sum, we have shown that

$$\text{(S.19)}\qquad \begin{aligned}\frac{1}{\sqrt{NT}}\frac{\partial\ell_c}{\partial\rho} &= \frac{1}{\sigma^2\sqrt{NT}}\sum_{i=1}^N u_i'Lu_i+O_p\bigl(T^{-1/2}\bigr)+O_p\bigl(N^{1/2}/T^{3/2}\bigr)\\
&\xrightarrow{d} N\bigl(0,1/\bigl(1-\rho^2\bigr)\bigr).\end{aligned}$$

Next, consider the first order condition with respect to the variance. Using (S.8) and (S.14), we obtain

$$\begin{aligned}
&\operatorname{tr}\bigl(BS_NB'\bigr)-\frac{1}{T}\bigl(1_T'BS_NB'1_T\bigr)-\sigma^2(T-1)\\
&=\frac{1}{N}\sum_{i=1}^N\bigl(u_i'u_i-T\sigma^2\bigr)-\frac{1}{TN}\sum_{i=1}^N\bigl[\bigl(1_T'u_i\bigr)^2-T\sigma^2\bigr].
\end{aligned}$$

Multiply the preceding equation by $N/(2\sigma^4)$ and divide it by $\sqrt{NT}$, and by the definition of $\frac{\partial\ell_c}{\partial\sigma^2}$, we obtain

$$\text{(S.20)}\qquad \frac{1}{\sqrt{NT}}\frac{\partial\ell_c}{\partial\sigma^2}=\frac{1}{2\sigma^4}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\bigl(u_{it}^2-\sigma^2\bigr)+O_p\left(\frac{1}{\sqrt{T}}\right).$$

Under normality for u_{it}, the above converges in distribution to $N(0,\frac{1}{2\sigma^4})$. Moreover, since $w_{it-1}u_{it}$ and $(u_{is}^2-\sigma^2)$ are uncorrelated, $(NT)^{-1/2}\frac{\partial\ell_c}{\partial\rho}$ and

$(NT)^{-1/2}\frac{\partial\ell_c}{\partial\sigma^2}$ are asymptotically independent. This proves the lemma. We in fact prove more than the lemma. Our analysis shows that

$$\text{(S.21)}\qquad \frac{1}{\sqrt{NT}}\begin{bmatrix}\dfrac{\partial\ell_c}{\partial\rho}\\ \dfrac{\partial\ell_c}{\partial\sigma^2}\end{bmatrix}=\begin{bmatrix}\dfrac{1}{\sigma^2\sqrt{NT}}\displaystyle\sum_{i=1}^{N}u_i'Lu_i\\ \dfrac{1}{2\sigma^4\sqrt{NT}}\displaystyle\sum_{i=1}^{N}\sum_{t=1}^{T}\bigl(u_{it}^2-\sigma^2\bigr)\end{bmatrix} + O_p\bigl(T^{-1/2}\bigr)+O_p\bigl(N^{1/2}/T^{3/2}\bigr).$$

This representation allows us to easily obtain the limiting distribution under nonnormality. *Q.E.D.*

PROOF OF THEOREM S.2: This follows from the consistency of $\widehat{\theta}$, Lemmas S.4 and S.5, and Amemiya (1985, Chap. 4). *Q.E.D.*

PROOF OF COROLLARY S.1: We first prove representation (S.4). Since $\widehat{B}-B=-(\widehat{\rho}-\rho)J_T$, using the expression of $\widehat{\tau}-\tau_N$ in the proof of Theorem S.1, we have

$$\begin{aligned}\sqrt{NT}(\widehat{\tau}-\tau_N)&=\left(\frac{\sqrt{NT}(\sigma^2-\widehat{\sigma}^2)}{\widehat{\sigma}^2\sigma^2}\right)\frac{1_T'\widehat{B}S_N\widehat{B}'1_T}{T^2}\\&\quad-\frac{2}{\sigma^2}\sqrt{NT}(\widehat{\rho}-\rho)\left(\frac{1_T'J_TS_NB'1_T}{T^2}\right)\\&\quad+\frac{1}{\sigma^2}\sqrt{NT}(\widehat{\rho}-\rho)^2\left(\frac{1_T'J_TS_NJ_T'1_T}{T^2}\right)\\&\quad+\sqrt{NT}\left(\frac{1_T'BS_NB'1_T}{\sigma^2T^2}-\tau_N-\frac{1}{T}\right).\end{aligned}$$

The third term involves $(\widehat{\rho}-\rho)^2$ and is negligible. Given the consistency of $\widehat{\rho}$,

$$\frac{1_T'\widehat{B}S_N\widehat{B}'1_T}{T^2}=\frac{1_T'BS_NB'1_T}{T^2}+o_p(1)\xrightarrow{p}\tau\sigma^2,$$

by Lemma S.3. Also by part (vi) of the same lemma, the second term is $-2\tau/(1-\rho)\sqrt{NT}(\widehat{\rho}-\rho)+o_p(1)$. Thus

$$\begin{aligned}\sqrt{NT}(\widehat{\tau}-\tau_N)&=-\frac{\tau}{\sigma^2}\sqrt{NT}\bigl(\widehat{\sigma}^2-\sigma^2\bigr)-2\frac{\tau}{1-\rho}\sqrt{NT}(\widehat{\rho}-\rho)\\&\quad+\sqrt{NT}\left(\frac{1_T'BS_NB'1_T}{\sigma^2T^2}-\tau_N-\frac{1}{T}\right)+o_p(1).\end{aligned}$$

By (S.9),

$$\sqrt{NT}\left(\frac{1_T' B S_N B' 1_T}{\sigma^2 T^2} - \tau_N - \frac{1}{T}\right)$$
$$= \frac{2}{\sigma^2}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}\eta_i$$
$$+ T^{-1/2}\frac{1}{\sigma^2}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}\right)^2 - \sigma^2\right].$$

The last term is $O_p(T^{-1/2})$. This proves (S.4). Notice that

$$\begin{aligned} 2\frac{1}{\sigma^2}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}\eta_i &= 2\sqrt{\tau_N}\frac{1}{\sigma^2\sqrt{\tau_N}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}\eta_i \\ &\xrightarrow{d} 2\sqrt{\tau}N(0,1) \\ &\overset{d}{=} N(0, 4\tau). \end{aligned}$$

The above expression is asymptotically uncorrelated with both $\sqrt{NT}(\widehat{\sigma}^2 - \sigma^2)$ and $\sqrt{NT}(\widehat{\rho} - \rho)$. So the limiting distribution of $\sqrt{NT}(\widehat{\tau} - \tau_N)$ is easily obtained. From $\widehat{\pi} = \widehat{\tau}\widehat{\sigma}^2$, it is easy to derive the representation for $\widehat{\pi}$, given the representation for $\widehat{\tau}$. The corollary follows from these representations. *Q.E.D.*

PROOF OF COROLLARY S.2: This again follows from the representation of $\sqrt{NT}(\widehat{\pi} - \pi_N)$. *Q.E.D.*

PROOF OF THEOREM S.3: The proof uses the same argument as in the proof of Theorem S.2, except we replace u_i by $u_i - \bar{u}$ and η_i by $\eta_i - \bar{\eta}$. It is easy to verify that, under large N, Lemmas S.2–S.4 hold when u_i and η_i are replaced by $u_i - \bar{u}$ and $\eta_i - \bar{\eta}$, respectively. It is Lemma S.5 that requires further analysis. Equation (S.19) becomes

$$\frac{1}{\sqrt{NT}}\frac{\partial \ell_c}{\partial \rho} = \frac{1}{\sigma^2\sqrt{NT}}\sum_{i=1}^{N}(u_i - \bar{u})' L (u_i - \bar{u})$$
$$+ O_p\big(T^{-1/2}\big) + O_p\big(N^{1/2}/T^{3/2}\big).$$

Note that the term $O_p(T^{-1/2}) + O_p(N^{1/2}/T^{3/2})$ is not effected. But

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(u_i - \bar{u})' L (u_i - \bar{u}) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} u_i' L u_i - (N/T)^{1/2}\bar{u}' L \bar{u}.$$

We show that $(N/T)^{1/2}\bar{u}'L\bar{u} = O_p(N^{-1/2})$. Its variance is

$$
\begin{aligned}
(N/T)E(\bar{u}'L\bar{u})^2 &= \frac{1}{TN^3}\sum_{i,j,k,l} E\big[(u_i'Lu_j)(u_k'Lu_l)\big] \\
&= \frac{1}{TN^3}\sum_{i=1}^{N} E\big[(u_i'Lu_i)^2\big] + \frac{2}{TN^3}\sum_{i,j;i\neq j} E\big[(u_i'Lu_j)^2\big].
\end{aligned}
$$

Note that the term $E(u_i'Lu_i)E(u_k'Lu_k) = 0$ is omitted. The first term on the right is $O(N^{-2})$, and the second term is $O(N^{-1})$ because $T^{-1}E(u_i'Lu_j)^2 = E(T^{-1/2}\sum_{t=1}^{T} w_{it-1}u_{jt})^2 = O(1)$. To sum up,

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\frac{\partial \ell_c}{\partial \rho} &= \frac{1}{\sigma^2\sqrt{NT}}\sum_{i=1}^{N} u_i'Lu_i \\
&\quad + O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(N^{1/2}/T^{3/2}) \\
&\xrightarrow{d} N\big(0, 1/(1-\rho^2)\big).
\end{aligned}
$$

Consider the first order condition for the variance. Equation (S.20) becomes

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\frac{\partial \ell_c}{\partial \sigma^2} &= \frac{1}{2\sigma^4}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(u_{it}^2 - \sigma^2) - \frac{1}{2\sigma^4}\left(\frac{N}{T}\right)^{1/2}\bar{u}'\bar{u} \\
&\quad + O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}
$$

But

$$
\begin{aligned}
\left(\frac{N}{T}\right)^{1/2}\bar{u}'\bar{u} &= \left(\frac{N}{T}\right)^{1/2}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} u_{it}\right)^2 \\
&= (T/N)^{1/2}\frac{1}{T}\sum_{t=1}^{T}\left(N^{-1/2}\sum_{i=1}^{N} u_{it}\right)^2 \\
&= (T/N)^{1/2}\sigma^2 + (T/N)^{1/2}\frac{1}{T}\sum_{t=1}^{T}\left[\left(N^{-1/2}\sum_{i=1}^{N} u_{it}\right)^2 - \sigma^2\right] \\
&= (T/N)^{1/2}\sigma^2 + O_p(N^{-1/2}).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\frac{\partial \ell_c}{\partial \sigma^2} + \frac{1}{2\sigma^2}(T/N)^{1/2} &= \frac{1}{2\sigma^4\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(u_{it}^2 - \sigma^2) \\
&\quad + O_p(N^{-1/2}) + O_p(T^{-1/2}).
\end{aligned}
$$

The Hessian matrix is block diagonal; so is its inverse. The bias for $\widehat{\sigma}^2$ equals $2\sigma^4(\frac{1}{2\sigma^2}(T/N)^{1/2}) = \sigma^2(T/N)^{1/2}$. We have

$$\begin{aligned}
&\sqrt{NT}\begin{bmatrix} \widehat{\rho}-\rho \\ \widehat{\sigma}^2-\sigma^2 \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma^2(T/N)^{1/2} \end{bmatrix} \\
&= \begin{bmatrix} 1-\rho^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \begin{bmatrix} \dfrac{1}{\sqrt{NT}}\dfrac{\partial \ell_c}{\partial \rho} \\ \dfrac{1}{\sqrt{NT}}\dfrac{\partial \ell_c}{\partial \sigma^2} + \dfrac{1}{2\sigma^2}(T/N)^{1/2} \end{bmatrix} + o_p(1) \\
&= \begin{bmatrix} 1-\rho^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \begin{bmatrix} \dfrac{1}{\sigma^2\sqrt{NT}}\displaystyle\sum_{i=1}^N u_i' L u_i \\ \dfrac{1}{2\sigma^4\sqrt{NT}}\displaystyle\sum_{i=1}^N\sum_{t=1}^T (u_{it}^2-\sigma^2) \end{bmatrix} + o_p(1).
\end{aligned}$$

This is equivalent to Theorem S.3. This representation also allows us to easily obtain the limiting distribution under nonnormality. *Q.E.D.*

S.9. ROBUSTNESS OF THE MUNDLAK PROJECTION UNDER LARGE T

The Mundlak–Chamberlain projection aims to take into account the arbitrary correlation between the strictly exogenous x_{it} and the effects η_i. The projection takes the form

$$\eta_i = c_0 + c_1' x_{i1} + \cdots + c_T' x_{iT} + \tau_i. \tag{S.22}$$

This projection has too many free parameters. The original Mundlak projection imposes equal coefficients $c_1 = \cdots = c_T = c$ so that $\eta_i = c_0 + c'\bar{x}_i + \tau_i$. For non-dynamic panel models and under homoskedasticity $\mathrm{var}(u_{it}) = \sigma_u^2$, the maximum likelihood estimator with the Mundlak projection coincides with the within-group estimator, which is consistent under both fixed and large T (see Mundlak (1978)). For dynamic panel models, the MLE with the Mundlak projection is no longer the within-group estimator. We show that the maximum likelihood estimator has a negligible bias provided that $N/T^3 \to 0$. This result will be proved as a special case for the general model that allows for heteroskedasticity. Under heteroskedasticity, the usual Mundlak projection is replaced by a weighted average of the strictly exogenous regressors. But the existence of heteroskedasticity requires $N/T \to 0$ for the bias to be negligible (there is a bias of order $1/T$). The proof below shows why the latter condition is needed and why a weaker condition suffices under homoskedasticity. We also suggest a further generalization of the Mundlak projection under which $N/T^3 \to 0$ becomes sufficient again to remove the bias.

With heteroskedasticity, we need to use a modified Mundlak projection

$$\eta_i = c_0 + \left(x_i'\Phi^{-1}1_T\right)'c + \tau_i.$$

For theoretical analysis, we consider

$$\eta_i = c_0 + \bar{x}_i(\Phi)'c + \tau_i, \tag{S.23}$$

where $\bar{x}_i(\Phi) = (1_T'\Phi^{-1}1_T)^{-1}x_i'\Phi^{-1}1_T$. This is a matter of renormalizing c, and it prevents the predictor from becoming unbounded as T increases. Note that the true projection coefficients in (S.22) reflect the relationship between η_i and x_i; they are not related in any way to the heteroskedasticity matrix Φ of u_i. The projection in (S.23) is motivated by the fixed-effects estimates under heteroskedasticity; see Alvarez and Arellano (2004).

Using (S.23), we can rewrite the model as

$$y_i = \Gamma\delta + \Gamma x_i\beta + \Gamma 1_T\bar{x}_i(\Phi)'c + \Gamma 1_T\tau_i + \Gamma u_i.$$

Removing the time effects, we have

$$\dot{y}_i = \Gamma\dot{x}_i\beta + \Gamma 1_T\bar{x}_i(\Phi)'c + \Gamma 1_T\dot{\tau}_i + \Gamma\dot{u}_i.$$

Let $\dot{y}_{i,-1}$ denote the lag of $\dot{y}_i$. Notice that $\dot{y}_{i,-1} = J_T\dot{y}_i$ and $J_T\Gamma = L$, where J_T and L are defined in the main text. We have

$$\dot{y}_{i,-1} = L\dot{x}_i\beta + L1_T\bar{x}_i(\Phi)'c + L1_T\dot{\tau}_i + L\dot{u}_i. \tag{S.24}$$

In the following analysis, we assume that π_N and Φ are known. The validity of our results does not hinge on this assumption, but it simplifies the analysis and provides the key insights. Under this assumption, the matrix $\Omega = 1_T1_T'\pi_N + \Phi$ is known, and the unknown parameters are $\theta = (\rho, \beta', c')'$. Our objective is to show that $\sqrt{NT}(\widehat{\theta} - \theta) = O_p(1)$, where $\widehat{\theta}$ is obtained by maximizing

$$\ell(\theta) = -\frac{n}{2}\log\left|\Sigma(\theta)\right| - \frac{n}{2}\operatorname{tr}\left[S_N\Sigma(\theta)^{-1}\right].$$

Given Ω, the estimator $\widehat{\theta}$ is simply

$$\widehat{\theta} = \left(\sum_{i=1}^N \dot{W}_i'\Omega^{-1}\dot{W}_i\right)^{-1}\sum_{i=1}^N \dot{W}_i'\Omega^{-1}\dot{y}_i,$$

where $\dot{W}_i = [\dot{y}_{i,-1}, \dot{x}_i, 1_T\dot{\bar{x}}_i(\Phi)']$ and

$$\Omega^{-1} = \Phi^{-1} - \Phi^{-1}1_T1_T'\Phi^{-1}a_T,$$

with $a_T = \pi_N/(1 + T\omega_T\pi_N)$ being a scalar and $\omega_T = (1_T'\Phi^{-1}1_T)/T$. We can rewrite the estimator as

$$\sqrt{NT}(\widehat{\theta} - \theta) = \left(\frac{1}{NT}\sum_{i=1}^N \dot{W}_i'\Omega^{-1}\dot{W}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{W}_i'\Omega^{-1}[1_T\dot{\tau}_i + \dot{u}_i].$$

In view of the expression for Ω^{-1}, we want to show that

$$\text{(S.25)}\qquad \frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{W}_i'\big[\Phi^{-1} - \Phi^{-1}1_T1_T'\Phi^{-1}a_T\big][1_T\dot{\tau}_i + \dot{u}_i] = O_p(1)$$

if $N/T \to 0$ for general Φ (heteroskedasticity) and $N/T^3 \to 0$ for $\Phi = \sigma_u^2 I_T$ (homoskedasticity).

Before proceeding, we point out that we take c_0 and c in (S.23) as the least squares coefficients (so c_0 and c_1 depend on N and T). The residuals τ_i satisfy

$$\sum_{i=1}^N \tau_i = 0, \quad \sum_{i=1}^N \bar{x}_i(\Phi)\tau_i = 0.$$

The above further implies that

$$\text{(S.26)}\qquad \sum_{i=1}^N \dot{\bar{x}}_i(\Phi)\dot{\tau}_i = 0, \quad \text{or equivalently,} \quad \sum_{i=1}^N \big(\dot{x}_i'\Phi^{-1}1_T\big)\dot{\tau}_i = 0.$$

We next show that (S.25) holds for each component of W_i. For the last component, $1_T\dot{\bar{x}}_i(\Phi)'$, we need to show that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{\bar{x}}_i(\Phi)1_T'\big[\Phi^{-1} - \Phi^{-1}1_T1_T'\Phi^{-1}a_T\big][1_T\dot{\tau}_i + \dot{u}_i] = O_p(1).$$

The left hand side has four terms, two of which are zero and two are $O_p(1)$. The zero terms are, in view of (S.26),

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{\bar{x}}_i(\Phi)\dot{\tau}_i\big(1_T'\Phi^{-1}1_T\big) = 0$$

and

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{\bar{x}}_i(\Phi)\dot{\tau}_i\big(1_T'\Phi^{-1}1_T\big)^2 a_T = 0.$$

The $O_p(1)$ terms are

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{\ddot{x}}_i(\Phi)1_T'\Phi^{-1}\dot{u}_i = O_p(1)$$

and

$$\left(1_T'\Phi^{-1}1_T\right)a_T\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{\ddot{x}}_i(\Phi)1_T'\Phi^{-1}\dot{u}_i = O_p(1).$$

Note that $(1_T'\Phi^{-1}1_T)a_T = O(1)$. For the second component of W_i, we need to show that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{x}_i'\left[\Phi^{-1}-\Phi^{-1}1_T1_T'\Phi^{-1}a_T\right][1_T\dot{\tau}_i+\dot{u}_i] = O_p(1).$$

Again, the left hand side has two zero terms and two $O_p(1)$ terms based on the same reasoning as the first component. It is more involved to analyze the first component of W_i, that is,

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{y}_{i,-1}'\left[\Phi^{-1}-\Phi^{-1}1_T1_T'\Phi^{-1}a_T\right][1_T\dot{\tau}_i+\dot{u}_i] = O_p(1).$$

Because $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{y}_{i,-1}'\Phi^{-1}\dot{u}_i = O_p(1)$, we need to show that the remaining three terms are $O_p(1)$. This requires, after rearranging terms,

$$\text{(S.27)}\qquad \left[1-\left(1_T'\Phi^{-1}1_T\right)a_T\right]\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{y}_{i,-1}'\Phi^{-1}1_T\dot{\tau}_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\dot{y}_{i,-1}'\Phi^{-1}1_T\right)\left(1_T'\Phi^{-1}\dot{u}_i\right)a_T = O_p(1).$$

First note that

$$1-\left(1_T'\Phi^{-1}1_T\right)a_T = \frac{1}{1+T\omega_T\pi_N} = O\left(\frac{1}{T}\right).$$

The lag $\dot{y}_{i,-1}$ consists of four terms: $L\dot{x}_i$, $L1_T\dot{\ddot{x}}_i(\Phi)'$, $L1_T\dot{\tau}_i$, and $L\dot{u}_i$. We analyze the property for each of them when substituting into (S.27).

LEMMA S.6: *Let $\dot{\ddot{x}}_i$ denote for $\dot{\ddot{x}}_i(\Phi)$. We have*

(a) $\frac{1}{1+T\omega_T\pi_N}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{x}_i'L'\Phi^{-1}1_T\dot{\tau}_i = O_p(\sqrt{\frac{N}{T}})$,

(b) $\frac{1}{1+T\omega_T\pi_N}\frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{\bar{x}}_i(1_T'L'\Phi^{-1}1_T)\dot{\tau}_i = 0$,

(c) $\frac{1}{1+T\omega_T\pi_N}\frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{\tau}_i 1_T'L'\Phi^{-1}1_T\dot{\tau}_i = \frac{1_T'L'\Phi^{-1}1_T\pi_N}{1+T\omega_T\pi_N}\sqrt{\frac{N}{T}}$,

(d) $\frac{1}{1+T\omega_T\pi_N}\frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{u}_i L'\Phi^{-1}1_T\dot{\tau}_i = O_p(\frac{1}{T})$,

(e) $\frac{1}{\sqrt{NT}}\sum_{i=1}^N(\dot{x}_i'L'\Phi^{-1}1_T)(1_T'\Phi^{-1}\dot{u}_i)a_T = O_p(1)$,

(f) $\frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{\bar{x}}_i(1_T'L'\Phi^{-1}1_T)(1_T'\Phi^{-1}\dot{u}_i)a_T = O_p(1)$,

(g) $\frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{\tau}_i(1_T'L'\Phi^{-1}1_T)(1_T'\Phi^{-1}\dot{u}_i)a_T = O_p(1)$,

(h) $\frac{1}{\sqrt{NT}}\sum_{i=1}^N(\dot{u}_i'L'\Phi^{-1}1_T)(1_T'\Phi^{-1}\dot{u}_i)a_T = \frac{1_T'L'\Phi^{-1}1_T\pi_N}{1+T\omega_T\pi_N}\sqrt{\frac{N}{T}} + O_p(T^{-1/2})$.

PROOF: For part (a), notice that $\sigma_t^{-2} \le 1/a$ because $0 < a \le \sigma_t^2 \le b$ by assumption. Then part (a) is bounded by

$$\frac{1}{1+T\omega_T\pi_N}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \|x_{it}\|\frac{1}{1-\rho}\frac{1}{a} = O_p(\sqrt{N/T}).$$

Part (b) follows from (S.26). Part (c) is a simple identity. Part (d) is trivial. Part (e) follows from $a_T = O(1/T)$, $\dot{x}_i'L'\Phi^{-1}1_T a_T = O_p(1)$, and the latter is uncorrelated with $1_T'\Phi^{-1}\dot{u}_i$. Parts (f) and (g) are similar to (e). For part (h), its expected value is $1_T'L'\Phi^{-1}1_T a_T$, which is equal to the first term on the right hand side. Its difference with its expected value is $O_p(T^{-1/2})$. *Q.E.D.*

Equation (S.27) is equal to the sum of the expressions from (a) to (d) minus the sum of the expressions from (e) to (h). The difference between (c) and (h) is $O_p(T^{-1/2})$ and thus is negligible. In summary,

$$(\text{S.27}) = (a) + (e) - (f) - (g) + O_p(T^{-1/2}).$$

Part (a) determines the bias, and parts (e), (f), and (g) are $O_p(1)$ and they contribute to the limiting distribution. The magnitude of part (a) implies a bias of order $1/T$, that is,

$$\widehat{\theta} - \theta = O\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

REMARK: For non-dynamic panel data models, part (a) is not present in the model. This implies that the modified Mundlak procedure removes the bias induced by the correlation between the regressors and the effects. That is, $\widehat{\theta} - \theta = O_p(\frac{1}{\sqrt{NT}})$.

An Extended Mundlak Projection

A further extension of the Mundlak procedure is to include $x_i'L\Phi^{-1}1_T$ in the projection of (S.23); this will imply that part (a) is negligible. But since

L depends on ρ, this will increase the nonlinearity of the model. Here is an alternative solution. Let $\widetilde{\rho}$ be a preliminary estimator (say the within-group estimator). Let $\widetilde{L}$ be the corresponding matrix. Define

$$\bar{x}_{2i}(\Phi) = \frac{x_i'\widetilde{L}'\Phi^{-1}1_T}{1_T'\widetilde{L}'\Phi^{-1}1_T}.$$

Consider the projection

$$\eta_i = c_0 + \bar{x}_i(\Phi)'c_1 + \bar{x}_{2i}(\Phi)'c_2 + \tau_i. \tag{S.28}$$

In addition to (S.26), the projection residuals satisfy

$$\sum_{i=1}^{N} \dot{x}_i'\widetilde{L}\Phi^{-1}1_T\dot{\tau}_i = 0.$$

The above implies that part (a) is negligible. To see this,

$$\begin{aligned}
\text{(a)} &= \frac{1}{1+T\omega_T\pi_N}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{x}_i'L'\Phi^{-1}1_T\dot{\tau}_i \\
&= \frac{1}{1+T\omega_T\pi_N}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{x}_i'(L-\widetilde{L})'\Phi^{-1}1_T\dot{\tau}_i.
\end{aligned}$$

But $L-\widetilde{L}$ only depends on $\rho-\widetilde{\rho}$. It is easy to show that (a) is bounded by $|\rho-\widetilde{\rho}|O(\sqrt{N/T})$. It follows that if $\rho-\widetilde{\rho} = O_p(1/T) + O_p(1/N)$, then (a) $= O_p(N^{1/2}/T^{3/2}) + O_p(1/\sqrt{NT})$, which is negligible if $N/T^3 \to 0$. In summary, under the projection (S.28) and $N/T^3 \to 0$, we have $\sqrt{NT}(\widehat{\theta}-\theta) = O_p(1)$.

The Case of Homoskedasticity

Under homoskedasticity, projection (S.28) is not required. Part (a) is in fact already $O_p(\sqrt{N}/T^{3/2})$ under projection (S.23) with $\bar{x}_i(\Phi) = \frac{1}{T}\sum_{t=1}^{T} x_{it}$. This implies that, with $N/T^3 \to 0$,

$$\sqrt{NT}(\widehat{\theta}-\theta) = O_p(1).$$

To see this, under homoskedasticity, $\Phi = \sigma_u^2 I_T$,

$$L'\Phi^{-1}1_T = \frac{1}{\sigma_u^2}L'1_T = \frac{1}{\sigma_u^2}\frac{1}{1-\rho}1_T - \frac{1}{\sigma_u^2}\frac{1}{1-\rho}\begin{bmatrix}\rho^{T-1}\\ \rho^{T-2}\\ \vdots \\ 1\end{bmatrix}.$$

It follows that

$$\dot{x}_i' L' \Phi^{-1} 1_T = \frac{1}{\sigma_u^2} \frac{T}{1-\rho} \dot{\bar{x}}_i - \frac{1}{\sigma_u^2} \frac{1}{1-\rho} \sum_{t=1}^{T} \dot{x}_{it} \rho^{T-t}.$$

Thus,

$$\begin{aligned}
&\frac{1}{1+T\omega_T \pi_N} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \dot{x}_i' L' \Phi^{-1} 1_T \dot{\tau}_i \\
&\quad = \frac{T}{\sigma_u^2 + T\pi_N} \frac{1}{1-\rho} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \dot{\bar{x}}_i \dot{\tau}_i \\
&\qquad - \frac{1}{\sigma_u^2 + T\pi_N} \frac{1}{1-\rho} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \dot{x}_{it} \rho^{T-t} \dot{\tau}_i.
\end{aligned}$$

The first term on the right hand side is 0 because $\sum_{i=1}^{N} \dot{\bar{x}}_i \dot{\tau}_i = 0$. For the second term, using $\| \sum_{t=1}^{T} \dot{x}_{it} \rho^{T-t} \| \leq \sum_{t=1}^{T} \|\dot{x}_{it}\| |\rho|^{T-t} = O_p(1)$, we have

$$\frac{1}{N} \left| \sum_{i=1}^{N} \sum_{t=1}^{T} \dot{x}_{it} \rho^{T-t} \right| = O_p(1).$$

This implies that the second term is $O_p(N^{1/2}/T^{3/2})$ and so is part (a). In summary, under homoskedasticity and $N/T^3 \to 0$, the Mundlak projection (S.23) removes the bias induced by the arbitrary correlation between the regressors and the effects.

ADDITIONAL REFERENCES

AMEMIYA, T. (1985): *Advanced Econometrics*. Cambridge, MA: Harvard University Press. [3,7,15,23]

ARNOLD, S. F. (1981): *The Theory of Linear Models and Multivariate Analysis*. New York: Wiley. [3]

DAVIDSON, J. (2000): *Econometrics Theory*. Oxford: Blackwell. [12]

NEWEY, W., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle and D. McFadden. Amsterdam: North-Holland. [3,15]

Dept. of Economics, Columbia University, New York, NY 10012, U.S.A. and CEMA, Central University of Finance and Economics, Beijing, China; jushan.bai@columbia.edu.

Manuscript received July, 2010; final revision received June, 2012.