

NOTES AND COMMENTS

FIXED-EFFECTS DYNAMIC PANEL MODELS,  
A FACTOR ANALYTICAL METHOD

BY JUSHAN BAI<sup>1</sup>

We consider the estimation of dynamic panel data models in the presence of incidental parameters in both dimensions: individual fixed-effects and time fixed-effects, as well as incidental parameters in the variances. We adopt the factor analytical approach by estimating the sample variance of individual effects rather than the effects themselves. In the presence of cross-sectional heteroskedasticity, the factor method estimates the average of the cross-sectional variances instead of the individual variances. The method thereby eliminates the incidental-parameter problem in the means and in the variances over the cross-sectional dimension. We further show that estimating the time effects and heteroskedasticities in the time dimension does not lead to the incidental-parameter bias even when  $T$  and  $N$  are comparable. Moreover, efficient and robust estimation is obtained by jointly estimating heteroskedasticities.

KEYWORDS: Incidental parameters in means, incidental parameters in variances, fixed- $T$  and large- $T$  dynamic panels, heteroskedasticity, robust estimation, efficiency.

1. INTRODUCTION

FIXED-EFFECTS DYNAMIC PANEL MODELS are usually estimated by either the within-group method or the generalized method of moments (GMM). The within-group estimator is biased and inconsistent under fixed  $T$  (Nickell (1981), Kiviet (1995)) and the Arellano–Bond GMM estimator has a bias of order  $1/N$  (Alvarez and Arellano (2003)). In this paper, we argue that the fixed-effects dynamic panel data models can be estimated by the factor analytical method, which entails estimating the sample variance of the individual fixed effects instead of the individual effects themselves, thereby eliminating the incidental-parameter problem. The factor estimator is consistent under fixed  $T$  and it does not have an asymptotic bias of order  $1/N$  or order  $1/T$  even under large  $T$ . Broadly speaking, the factor estimator is consistent irrespective of the way in which  $N$  and  $T$  go to infinity.

In the presence of cross-sectional heteroskedasticities, the factor method provides an estimate for the average variance (over the cross sections) instead of individual variances, thus eliminating the incidental-parameter problem in the cross-sectional variance. We also consider incidental parameters in

<sup>1</sup>The author thanks a co-editor and three anonymous referees for constructive comments, and Marcelo Moreira for useful discussions. He also thanks seminar participants at Rutgers University, Michigan State University, and Columbia Econometrics Colloquium for helpful comments. Partial results of this paper were also presented at the 10th World Congress of the Econometric Society, Shanghai, China and the Asian Meetings of Econometric Society, Seoul, Korea, both as paired-invited talks, and at the 16th African Econometric Society meetings, Nairobi, Kenya, as a keynote talk. Financial support from the NSF (Grant SES-0962410) is acknowledged.

the time dimension. We show that estimating the time effects does not lead to biases for the autoregressive coefficient estimator. We further study time series heteroskedasticity motivated with two considerations. First, a changing variance appears to be an important empirical fact (e.g., [Moffitt and Gottschalk \(2002\)](#)). Second, unlike standard regression analysis where heteroskedasticity is often a problem more of efficiency than of consistency, imposing homoskedasticity leads to inconsistency for dynamic panel models if heteroskedasticity exists under fixed  $T$ . This has led to the robust consideration of [Alvarez and Arellano \(2004\)](#) by allowing heteroskedasticity. However, a concern arises as to whether estimating a large number of variance parameters (under large  $T$ ) will lead to an incidental-parameters bias, similar to the Arellano–Bond estimator analyzed by [Alvarez and Arellano \(2003\)](#). We show that bias does not arise. Also, efficient and robust estimation is obtained by allowing heteroskedasticity. We also provide a novel argument of consistency in the presence of an increasing number of parameters.

## 2. THE MODEL, NOTATION, AND ASSUMPTION

We start with a simple dynamic model without regressors. This simplifies notation and exposition. The case with regressors is examined in Section 5. Consider

$$y_{it} = \eta_i + \delta_t + \rho y_{it-1} + u_{it}, \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T,$$

where  $\eta_i$  are individual effects,  $\delta_t$  are time effects, and  $u_{it}$  are unobservable errors. This model has been widely studied, for example, by [Anderson and Hsiao \(1981\)](#), [Arellano and Bond \(1991\)](#), [Blundell and Smith \(1991\)](#), [Ahn and Schmidt \(1995\)](#), and [Alvarez and Arellano \(2003\)](#). We assume  $u_{it}$  are independent and identically distributed (i.i.d.) over  $i$ , zero mean, and variance  $\sigma_i^2$ . The model allows for time series heteroskedasticity for the purpose of robust estimation, motivated by the work of [Alvarez and Arellano \(2004\)](#). To see that the model can be expressed as a special case of a factor model, we first assume  $y_{i0} = 0$  as in [Moreira \(2009\)](#). The case of  $y_{i0} \neq 0$  is considered in the Supplemental Material ([Bai \(2013\)](#)). Note that assumptions on initial conditions for dynamic panels are important for fixed- $T$  analysis; see, for example, [Anderson and Hsiao \(1981\)](#), [Bhargava and Sargan \(1983\)](#), [Blundell and Smith \(1991\)](#), and [Hahn \(1999\)](#). Rewrite the model as  $By_i = \delta + 1_T \eta_i + u_i$ , or

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

where  $\delta = (\delta_1, \dots, \delta_T)'$  and  $1_T = (1, 1, \dots, 1)'$ , both are  $T \times 1$ , and

$$B = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\rho & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\rho & 1 \end{bmatrix}, \quad \Gamma = B^{-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \rho & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \rho^{T-1} & \cdots & \rho & 1 \end{bmatrix}.$$

This is a factor model with a single factor, and with factor loading  $\Gamma 1_T$  ( $T \times 1$ ) and factor score  $\eta_i$ . A general factor structure is (Anderson and Rubin (1956), Lawley and Maxwell (1971))  $y_i = \mu + \Lambda f_i + \varepsilon_i$  ( $i = 1, 2, \dots, N$ ). For a dynamic panel data model with fixed effects, we have  $\mu = \Gamma \delta$  (a vector of free parameters since  $\delta$  is),  $\Lambda = \Gamma 1_T$ ,  $f_i = \eta_i$ , and  $\varepsilon_i = \Gamma u_i$ . This is an identifiable factor structure (for  $T \geq 3$ ) since the first element of  $\Gamma 1_T$  is equal to 1.

In factor analysis, the vector  $\mu$  is estimated by the sample mean  $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ . Define

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

a  $T \times T$  matrix, where  $n = N - 1$ . From  $y_i - \bar{y} = \Gamma 1_T(\eta_i - \bar{\eta}) + \Gamma(u_i - \bar{u})$ , with  $\bar{\eta} = \frac{1}{N} \sum_{i=1}^N \eta_i$  and  $\bar{u} = \frac{1}{N} \sum_{i=1}^N u_i$ , it is easy to verify that

$$E(S_N) = \Gamma(1_T 1_T' \pi_N + \Phi) \Gamma',$$

where  $\Phi = \text{diag}(\sigma_1^2, \dots, \sigma_T^2)$  and  $\pi_N = \frac{1}{n} \sum_{i=1}^N (\eta_i - \bar{\eta})^2$ ; the latter of which is the sample variance of the individual fixed effects, a scalar quantity. Let  $\theta_N = (\rho, \pi_N, \sigma_1^2, \dots, \sigma_T^2)$  denote the vector of parameters. We also write  $\theta$  for  $\theta_N$  for succinctness. Let

$$(2) \quad \Sigma(\theta) = \Gamma(1_T 1_T' \pi_N + \Phi) \Gamma'.$$

To estimate the parameters, consider the discrepancy function between  $S_N$  and  $\Sigma(\theta)$ ,

$$(3) \quad Q_N(\theta) = \log |\Sigma| + \text{tr}[S_N \Sigma^{-1}],$$

where  $\Sigma = \Sigma(\theta)$ . Multiplying by  $-n/2$ , this discrepancy has the same form as the likelihood function for a central Wishart distribution. It also has the same form as the random-effects likelihood function when  $\eta_i$  are i.i.d. normal. This means that the results also hold for random-effects dynamic panels. Under the fixed-effects assumption, the above function is not a likelihood function, but the factor literature uses it as a distance measure between  $S_N$  and  $\Sigma(\theta)$ ; see Amemiya, Fuller, and Pantula (1987). Other discrepancy functions may be used. Recently, motivated by the maximum invariance principle under the orthogonal transformation, Moreira (2009) used the noncentral Wishart distribution as the discrepancy. The analysis of the noncentral Wishart likelihood involves approximating the underlying Bessel function by a tractable form, which is nontrivial. Moreira considered the case of  $\Phi = \sigma^2 I_T$  and no time effects. The objective function (3) has a simple and familiar form, directly adopted from the factor analysis, and can be justified by a decision theoretical argument of Chamberlain and Moreira (2009) with a suitable choice of a loss function and prior distributions. For further analysis, we make the following assumption:

ASSUMPTION A:  $u_i$  are i.i.d. over  $i$ ;  $E(u_{it}) = 0$ ,  $\text{var}(u_{it}) = \sigma_i^2 > 0$ , and  $u_{it}$  are independent over  $t$  and have bounded fourth moments;  $|\rho| < 1$ ;  $\eta_i$  are fixed effects with  $\pi_N = \frac{1}{n} \sum_{i=1}^N (\eta_i - \bar{\eta})^2 \rightarrow \pi > 0$ .

Throughout, we shall assume that the true parameter  $\theta_N^0$  is an interior point of a compact set  $\Theta$  contained in  $(-1, 1) \times (0, \infty)^{T+1}$  so that  $|\rho| < 1$ , and  $\pi_N$  and  $\sigma_i^2$  ( $t = 1, 2, \dots, T$ ) are strictly positive. Let  $\hat{\theta}$  be the estimator of  $\theta_N^0$  by minimizing the loss function over the parameter space  $\Theta$ , that is,  $\hat{\theta} = \text{argmin}_{\theta} Q_N(\theta)$ . Invoking the result of factor analysis (e.g., Amemiya, Fuller, and Pantula (1987), Anderson and Amemiya (1988)), we immediately obtain consistency and asymptotic normality of  $\sqrt{N}(\hat{\theta} - \theta_N^0)$  for large  $N$  and fixed  $T$ . We thus focus on the large- $T$  setting, in which time effects and heteroskedasticities become incidental parameters.

### 3. INCIDENTAL PARAMETERS: TIME SERIES HETEROSKEDASTICITY UNDER LARGE $T$

In classical factor analysis, the consistency and asymptotic normality are for fixed  $T$  only. The analysis is based on the premise that  $\sqrt{N}(S_N - \Sigma(\theta_N^0))$  is asymptotically normal with a positive definite limiting covariance, where  $\theta_N^0$  is of fixed dimension. This, together with the delta method (Taylor expansion), is the main inference tool for classical factor analysis. As  $T$  increases, however, the dimension of the matrix is also increasing, so this argument will not be applicable. Furthermore, when  $T$  is larger than  $N$ ,  $S_N$  is in fact not of full rank even though  $\Sigma(\theta_N^0)$  is of full rank. The limiting covariance cannot be positive definite in this case. Thus, the large- $T$  analysis requires new arguments, and is considerably more difficult and more delicate since it involves two-dimensional limits. The large- $T$  analysis is important for several reasons. First, many panel data sets have  $T$  not very small. As information cumulates over time, more large- $T$  data sets become available. Second, some estimators such as the crude instrument-variables estimator discussed in Alvarez and Arellano (2003) are consistent under fixed  $T$ , but become inconsistent for large  $T$ . Third, large  $T$  analysis provides a guidance on the relative performance for even small- $T$  settings.

We consider the same model as in Section 2 but without the time effects  $\delta_t$ , confining the incidental-parameter problem within the variance only. Time effects will be considered at the end of this section. From  $y_i = \Gamma 1_T \eta_i + \Gamma u_i$ , and  $S_N = \frac{1}{N} \sum_{i=1}^N y_i y_i'$ , we have  $E(S_N) = \Gamma(1_T 1_T' \pi_N + \Phi) \Gamma'$ , where  $\pi_N = \frac{1}{N} \sum_{i=1}^N \eta_i^2$  and  $\Phi = \text{diag}(\sigma_1^2, \dots, \sigma_T^2)$ . The parameter vector is  $\theta_{NT} = (\rho, \pi_N, \sigma_1^2, \dots, \sigma_T^2)$ , which is indexed by both  $N$  and  $T$ . The same discrepancy (objective) function as in Section 2 is used. So the estimator also takes the same form. We also use  $(\hat{\rho}, \hat{\pi}_N, \hat{\Phi})$  to denote the estimator. The dimension of  $\theta_{NT}$  increases with  $T$ . We can no longer appeal to the fixed- $T$  factor analysis. Consistency of  $\hat{\theta}$  requires a new framework.

For technical reasons, we assume there exist  $a > 0$  and  $b > 0$  such that  $a < \sigma_i^2 < b < \infty$  for all  $\sigma_i^2$  and the maximization with respect to  $\sigma_i^2$  is taken over this set for  $t = 1, 2, \dots, T$ . We also assume the existence of a limit for

$$(4) \quad \omega_T = \frac{1}{T} 1'_T \Phi^{-1} 1_T = \frac{1}{T} \sum_{t=1}^T \frac{1}{\sigma_t^2} \longrightarrow \varpi > 0, \quad \text{say.}$$

The following results will be useful and the limits do not depend on matrix  $\Phi$ .

LEMMA 1: *Assuming the limit in (4) is positive, then, as  $T \rightarrow \infty$ ,*

$$(a) \quad \frac{1'_T \Phi^{-1} L 1_T}{1'_T \Phi^{-1} 1_T} \rightarrow \frac{1}{(1 - \rho)},$$

$$(b) \quad \frac{1'_T L' \Phi^{-1} L 1_T}{1'_T \Phi^{-1} 1_T} \rightarrow \frac{1}{(1 - \rho)^2},$$

where  $L$  is defined in (21) in the Appendix. Furthermore, the results hold when  $L$  is replaced by  $\Gamma$ .

Let  $\ell_{NT}(\theta) = -\frac{N}{2} Q_N(\theta)$ , which will be referred to as the likelihood function. It can be written as

$$\ell_{NT}(\theta) = -\frac{N}{2} \log |\Phi| - \frac{N}{2} \text{tr}(BS_N B' \Phi^{-1}) - \frac{N}{2} \log(1 + T \pi_N \omega_T) + \frac{N}{2} \frac{\pi_N}{1 + T \omega_T \pi_N} (1'_T \Phi^{-1} BS_N B' \Phi^{-1} 1_T).$$

We concentrate out  $\pi_N$ . Solving for  $\pi_N$  from its first order condition gives

$$\tilde{\pi}_N = \frac{1'_T \Phi^{-1} BS_N B' \Phi^{-1} 1_T}{(1'_T \Phi^{-1} 1_T)^2} - \frac{1}{1'_T \Phi^{-1} 1_T},$$

which is a function of  $\rho$  and  $\Phi$ . The concentrated likelihood function is

$$\ell^c(\rho, \Phi) = -\frac{N}{2} \log |\Phi| - \frac{N}{2} \text{tr}(BS_N B' \Phi^{-1}) - \frac{N}{2} \log \left( \frac{1'_T \Phi^{-1} BS_N B' \Phi^{-1} 1_T}{1'_T \Phi^{-1} 1_T} \right) + \frac{N}{2} \left( \frac{1'_T \Phi^{-1} BS_N B' \Phi^{-1} 1_T}{1'_T \Phi^{-1} 1_T} - 1 \right).$$

Let  $\Theta_1 = \Theta_\rho \times [a, b]^T$ , where  $\Theta_\rho$  is a compact subset of  $(-1, 1)$ , and  $0 < a < b < \infty$  are arbitrary; we assume the true parameter  $(\rho^0, \sigma_1^{02}, \dots, \sigma_T^{02})$  is an interior point of  $\Theta_1$ . Then we have the following.

LEMMA 2: Under Assumption A, as  $T, N \rightarrow \infty$ ,

$$(5) \quad \frac{1}{NT} \ell^c(\rho, \Phi) = -\frac{1}{2T} \sum_{t=1}^T \left[ \log(\sigma_t^2) + \frac{\sigma_t^{02}}{\sigma_t^2} \right] \\ - \frac{1}{2} (\rho - \rho^0)^2 \left[ \frac{1}{T} \text{tr}(L^0 \Phi^0 L^{0'} \Phi^{-1}) + \Delta \right] + o_p(1),$$

where  $L^0$  is defined in (21) evaluated at  $\rho^0$ ,  $\Delta \geq 0$ , and  $o_p(1)$  is uniform on  $\Theta_1$ .

The term  $T^{-1} \text{tr}(L^0 \Phi^0 L^{0'} \Phi^{-1})$  is strictly positive (uniformly bounded away from zero) on  $\Theta_1$ . Ignoring the  $o_p(1)$  term, the right hand side of the concentrated likelihood function is uniquely maximized at  $\rho = \rho^0$  and  $\sigma_t^2 = \sigma_t^{02}$  for all  $t$  (i.e.,  $\Phi = \Phi^0$ ). From this lemma, we can deduce the consistency of  $\hat{\rho}$ , but not the consistency of  $\hat{\sigma}_t^2$  because  $T \rightarrow \infty$ . Nevertheless, the lemma implies the average consistency in the sense of

$$(6) \quad \frac{1}{T} \sum_{t=1}^T (\hat{\sigma}_t^2 - \sigma_t^{02})^2 = o_p(1).$$

Average consistency is equivalent to individual consistency under fixed  $T$ . The equivalence breaks down if  $T \rightarrow \infty$ , indicating the complexity of the large- $T$  analysis. Result (6) and the consistency of  $\hat{\rho}$  allow us to show that  $\hat{\pi} = \pi_N^0 + o_p(1)$ . These consistency results, in turn, allow us to deduce the individual consistency of  $\hat{\sigma}_t^2$  for each  $t$  by exploring the relationships among  $\hat{\rho}$ ,  $\hat{\pi}$ , and  $\hat{\Phi}$ . This is the basic idea of our consistency analysis under large  $T$ .

Using the preceding lemma, the centered concentrated-likelihood can be written as

$$\frac{1}{NT} \ell^c(\rho, \Phi) - \frac{1}{NT} \ell^c(\rho^0, \Phi^0) \\ = -\frac{1}{2T} \sum_{t=1}^T \left[ \log(\sigma_t^2) + \frac{\sigma_t^{02}}{\sigma_t^2} - \log(\sigma_t^{02}) - 1 \right] \\ - \frac{1}{2} (\rho - \rho^0)^2 \left[ \frac{1}{T} \text{tr}(L^0 \Phi^0 L^{0'} \Phi^{-1}) + \Delta \right] + o_p(1).$$

Each of the first two expressions on the right hand side is nonpositive. Note that the function  $f(x) = \log(x) + \frac{\sigma_t^{02}}{x} - \log(\sigma_t^{02}) - 1$  is nonnegative for  $x > 0$ . Evaluate the above equation at  $(\hat{\rho}, \hat{\Phi})$ , and notice that  $\ell^c(\hat{\rho}, \hat{\Phi}) - \ell^c(\rho^0, \Phi^0)$  must be nonnegative. This can only be true if each of the right hand side ex-

pressions, evaluated at  $(\widehat{\rho}, \widehat{\Phi})$ , is  $o_p(1)$ , that is,

$$(\widehat{\rho} - \rho^0)^2 \left[ \frac{1}{T} \text{tr}(L^0 \Phi^0 L^0 \widehat{\Phi}^{-1}) + \Delta \right] = o_p(1),$$

$$\frac{1}{T} \sum_{t=1}^T \left[ \log(\widehat{\sigma}_t^2) + \frac{\sigma_t^{02}}{\widehat{\sigma}_t^2} - \log(\sigma_t^{02}) - 1 \right] = o_p(1).$$

The first equality implies  $\widehat{\rho} - \rho^0 = o_p(1)$  and the second equality implies the average consistency in (6). To see this, the function  $f(x)$  introduced earlier satisfies  $f(x) \geq c(x - \sigma_t^{02})^2$  on a compact set such that  $x$  and  $\sigma_t^{02} \in [a, b]$ , where  $c > 0$  only depends on  $a$  and  $b$ . Evaluating the function at  $x = \widehat{\sigma}_t^2$  and using the inequality, we obtain (6). Given these results, we prove in the Appendix that  $\widehat{\pi}$  is consistent for  $\pi_N^0$  and  $\widehat{\sigma}_t^2$  is consistent for  $\sigma_t^{02}$  for each  $t$ . We state this result as a lemma.

LEMMA 3: Under Assumption A, as  $N, T \rightarrow \infty$ , the factor-based MLE with fixed effects and heteroskedasticity is such that  $\widehat{\rho} = \rho^0 + o_p(1)$ ,  $\widehat{\pi} = \pi_N^0 + o_p(1)$ , and  $\widehat{\sigma}_t^2 = \sigma_t^{02} + o_p(1)$  for each  $t$ .

Given consistency, we drop the superscript “0” from the true parameters. Therefore,  $\rho, \sigma_t^2, L, \Phi$ , etc., represent the true parameters or matrices evaluated at the true parameters. We next study the asymptotic representations and the limiting distributions.

After a considerable amount of analysis, the asymptotic representation of  $\widehat{\rho}$  is found to be

$$(7) \quad \sqrt{NT}(\widehat{\rho} - \rho) = \left[ \frac{1}{T} \text{tr}(L\Phi L'\Phi^{-1}) \right]^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N u_i' \Phi^{-1} L u_i + o_p(1),$$

where  $o_p(1)$  holds if  $N, T \rightarrow \infty$  with  $N/T^3 \rightarrow 0$ . Again,  $L$  is given in (21). The variance of the numerator  $\frac{1}{\sqrt{NT}} \sum_{i=1}^N u_i' \Phi^{-1} L u_i$  is equal to  $\frac{1}{T} \text{tr}(L\Phi L'\Phi^{-1})$ , the same as the denominator. Notice that

$$(8) \quad \frac{1}{T} \text{tr}(L\Phi L'\Phi^{-1})$$

$$= \frac{1}{T} \sum_{t=2}^T \frac{1}{\sigma_t^2} (\sigma_{t-1}^2 + \rho^2 \sigma_{t-2}^2 + \dots + \rho^{2(t-2)} \sigma_1^2) \longrightarrow \gamma, \quad \text{say,}$$

where we assume the preceding limit exists. Then the representation of  $\widehat{\rho}$  implies

$$\sqrt{NT}(\widehat{\rho} - \rho) \xrightarrow{d} N(0, 1/\gamma).$$

If  $\sigma_t^2 = \sigma^2$  does not vary with  $t$ , then  $\gamma = 1/(1 - \rho^2)$ . It follows that, under homoskedasticity but without enforcing it, we obtain  $\sqrt{NT}(\hat{\rho} - \rho) \xrightarrow{d} N(0, 1 - \rho^2)$ . Therefore, by permitting heteroskedasticity for robust estimation, there is no loss of asymptotic efficiency under large  $T$ . Also, the incidental-parameter problem in the variance does not cause asymptotic bias. The limiting distribution is centered at zero, even scaled by the fast rate of  $\sqrt{NT}$ .

The estimator  $\hat{\pi}$  has the form

$$(9) \quad \hat{\pi} = \frac{1'_T \hat{\Phi}^{-1} \hat{B} S_N \hat{B}' \hat{\Phi}^{-1} 1_T}{(1'_T \hat{\Phi}^{-1} 1_T)^2} - \frac{1}{1'_T \hat{\Phi}^{-1} 1_T}.$$

The asymptotic representation of  $\hat{\pi}$  is found to be, assuming  $T/N^2 \rightarrow 0$ ,

$$(10) \quad \begin{aligned} \sqrt{NT} \left( \hat{\pi} - \pi_N - \frac{1}{N} b \right) &= -2\pi_N \left( \frac{1'_T \Phi^{-1} L 1_T}{1'_T \Phi^{-1} 1_T} \right) \sqrt{NT} (\hat{\rho} - \rho) \\ &\quad + 2 \left( \frac{1}{T} 1'_T \Phi^{-1} 1_T \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N u'_i \Phi^{-1} 1_T \eta_i \\ &\quad + o_p(1), \end{aligned}$$

where  $b$  is the bias term given by

$$b = 2 \left( \frac{1}{T \omega_T} \sum_{t=1}^T \frac{\nu_t}{\sigma_t^4} \right) \left( \frac{1}{N} \sum_{i=1}^N \eta_i \right)$$

with  $\nu_t = E(u_{it}^3)$ . The joint limiting distribution of  $\hat{\rho}$  and  $\hat{\pi}$  can be found from their representations.

**THEOREM 1:** *Under Assumption A, as  $N, T \rightarrow \infty$  with  $T/N^2 \rightarrow 0$  and  $N/T^3 \rightarrow 0$ , we have*

$$\begin{aligned} &\sqrt{NT} \begin{bmatrix} \hat{\rho} - \rho \\ \hat{\pi} - \pi_N - \frac{1}{N} b \end{bmatrix} \\ &\xrightarrow{d} N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/\gamma & -2\pi \frac{1}{\gamma(1-\rho)} \\ -2\pi \frac{1}{\gamma(1-\rho)} & \frac{4\pi^2}{\gamma(1-\rho)^2} + \frac{4\pi}{\varpi} \end{bmatrix} \right), \end{aligned}$$

where  $\gamma$  is defined in (8),  $\varpi$  is given in (4), and  $\pi$  is the limit of  $\pi_N$ .

The limiting distribution for  $\hat{\rho}$  is centered at zero, even when a large number of incidental parameters are estimated and when  $T$  and  $N$  are comparable. In

contrast, the within-group estimator has a bias of order  $1/T$  and has a larger asymptotic variance (less efficient). While GMM permits heteroskedasticity, explicit limiting distribution for the large- $T$  GMM with heteroskedasticity does not appear to be stated in the literature (despite abundant results for fixed- $T$  GMM). In general, the GMM estimator has a bias of order  $1/N$ , as shown by Alvarez and Arellano (2003). The condition  $N/T^3 \rightarrow 0$  in Theorem 1 is required for the limiting distribution to have this simple form. Under fixed  $T$ ,  $\sqrt{N}(\hat{\rho} - \rho)$  is still asymptotically normal, as argued in Section 2. The condition  $T/N^2 \rightarrow 0$  is used only for the distribution of  $\hat{\pi}$ , not for  $\hat{\rho}$ , and is only required when  $E(u_{it}^3) \neq 0$ .

REMARK: If one of the following three conditions holds: (i)  $T/N \rightarrow 0$ , (ii)  $E(u_{it}^3) = 0$ , (iii) random effects with  $E(\eta_i) = 0$ , then the term  $b$  is negligible, so that we can omit  $b/N$  from the distribution of  $\hat{\pi}$ . Furthermore, when one of the three conditions holds, it is relatively easy to show that  $\sqrt{NT}(\hat{\rho} - \rho)$  does not have asymptotic bias. It requires additional argument to show that there is still no asymptotic bias when none of the three conditions holds. It turns out that two bias terms arising from the estimation of  $\Phi$  are of equal magnitude with opposite signs, so they offset each other. Details are provided in the Appendix.

The estimated variance matrix  $\Phi = \text{diag}(\sigma_1^2, \dots, \sigma_T^2)$  is linked with  $\hat{\rho}$  and  $\hat{\pi}$  through

$$\hat{\Phi} = \text{diag}(\hat{B}S_N\hat{B}' - 1_T1_T'\hat{\pi} - 2[\hat{B}S_N\hat{B}' - \Omega(\hat{\theta})]\hat{\Phi}^{-1}1_T1_T'\hat{\pi}/(1 + T\hat{\omega}_T\hat{\pi})),$$

where  $\Omega(\hat{\theta}) = 1_T1_T'\hat{\pi} + \hat{\Phi}$  and  $T\hat{\omega}_T = 1_T'\hat{\Phi}^{-1}1_T$ . For a matrix  $A$ ,  $\text{diag}(A)$  is defined as a diagonal matrix by keeping the diagonal elements of  $A$ . The above expression for  $\hat{\Phi}$  is different from the standard factor analysis, which would imply that  $\hat{\Phi}$  is equal to  $\text{diag}(\hat{B}S_N\hat{B}' - 1_T1_T'\hat{\pi})$  (Lawley and Maxwell (1971, p. 30, equation (4.19))). The dynamic panel model implies many restrictions on the factor loadings, and the standard formula does not apply. However, the expression here is shown to be asymptotically equivalent to that of classical factor analysis because the last term in our expression is negligible. The asymptotic representation for  $\hat{\sigma}_i^2$  is found to be

$$(11) \quad \hat{\sigma}_i^2 - \sigma_i^2 = \frac{1}{N} \sum_{i=1}^N (u_{it}^2 - \sigma_i^2) + O_p(T^{-1}) + o_p(N^{-1/2}).$$

Note that the term  $O_p(T^{-1})$  on the right hand side does not mean that  $\hat{\sigma}_i^2$  is inconsistent under fixed  $T$ . The estimator  $\hat{\sigma}_i^2$  is still consistent with fixed  $T$ , as guaranteed by the results of Section 2. However, for the representation to have

this simple form, we require large  $T$ . It follows from the representation that if  $\sqrt{N}/T \rightarrow 0$ , as  $N, T \rightarrow \infty$ , then for each  $t$ ,

$$(12) \quad \sqrt{N}(\hat{\sigma}_t^2 - \sigma_t^2) \xrightarrow{d} N(0, 2\sigma_t^4),$$

under normality for  $u_{it}$ . It is also very easy to obtain the limiting distribution under nonnormality using the preceding asymptotic representation for  $\hat{\sigma}_t^2$ .

### *Joint Presence of Time Effects and Heteroskedasticity*

In the Supplemental Material (Bai (2013)), we also derive the result when time effects are present. All estimators have the same limiting distributions. Additionally, the higher order bias  $\frac{1}{N}b$  for  $\hat{\pi}$  does not exist. This may appear counterintuitive. However, the bias in Theorem 1 depends on the average  $\bar{\eta}$ . When time effects are estimated, the data depend on the individual effects only through the deviations  $\eta_i - \bar{\eta}$ , whose average is zero. Of course, with time effects, we estimate the sample variance  $\frac{1}{n} \sum_{i=1}^N (\eta_i - \bar{\eta})^2$  instead of the sample moment  $\frac{1}{N} \sum_{i=1}^N \eta_i^2$ . Hsiao and Tahmiscioglu (2008) also found that time effects do not yield asymptotic bias (under homoskedasticity) for a feasible generalized least squares (GLS) estimator. We state the results in the following theorem but will omit the proof to avoid repetition.

**THEOREM 2:** *When the time effects are estimated, Theorem 1 still holds with  $b = 0$ . In addition, (12) also holds.*

The issue of cross-sectional heteroskedasticity  $E(u_{it}^2) = \sigma_{it}^2$  is discussed in the Supplemental Material (Bai (2013)).

## 4. EFFICIENCY

### *Efficiency Under Large $T$*

This section considers the semiparametric efficiency bound under normality of  $u_{it}$ . Even with normality, the model still has two nonparametric components. The first component is the individual fixed effects and the second is the heteroskedasticity. Under large  $N$  and large  $T$ , each component has an infinite number of parameters in the limit. In this sense, the model is semiparametric.

Using arguments similar to those of Hahn and Kuersteiner (2002), we show in the Supplemental Material (Bai (2013)) that the semiparametric efficiency bound for regular estimators of  $\rho$  is given by  $1/\gamma$ , where  $\gamma$  is defined in (8). Regular estimators rule out superefficient ones such as those of Hodges and Stein; see, for example, Bickel, Klaassen, Ritov, and Wellner (1993) and van der Vaart and Wellner (1996). We state this result as a proposition.

PROPOSITION 1: *Under Assumption A and normality of  $u_{it}$ , the asymptotic semiparametric efficiency bound for regular estimators of  $\rho$  is  $1/\gamma$ . Furthermore, the factor estimator achieves the semiparametric efficiency bound.*

This means that there exist no regular estimators that can have a smaller asymptotic variance than  $1/\gamma$ . It follows that the factor estimator is efficient because it achieves the efficiency bound.

*Efficiency Under Fixed T*

It is more difficult to derive the semiparametric efficiency bound under fixed  $T$ . However, we can prove a certain optimality result: the factor estimator is asymptotically equivalent to the optimal GMM estimator based on the moment conditions

$$(13) \quad E[s - g(\theta)] = 0,$$

where  $s = \text{vech}(S_N)$  and  $g(\theta) = \text{vech}(\Sigma(\theta))$ . Let  $G_N = \partial g / \partial \theta'$ , evaluated at  $\theta_N^0$ , and let  $\Omega_N$  denote the variance of  $\sqrt{n}[s - g(\theta_N^0)]$ . Let  $\hat{\theta}_{\text{GMM}}$  denote the optimal GMM estimator based on (13) and let  $G'\Omega^{-1}G$  denote the probability limit of  $G'_N\Omega_N^{-1}G_N$ . It is well known that

$$\sqrt{N}(\hat{\theta}_{\text{GMM}} - \theta_N^0) \xrightarrow{d} N(0, (G'\Omega^{-1}G)^{-1}).$$

The factor estimator  $\hat{\theta}$  has the same limiting distribution.

PROPOSITION 2: *Under Assumption A and normality of  $u_{it}$ , then under fixed T,*

$$\sqrt{N}(\hat{\theta} - \theta_N^0) \xrightarrow{d} N(0, (G'\Omega^{-1}G)^{-1}).$$

Let  $v_\rho$  denote the (1, 1)th entry of  $(G'\Omega G')^{-1}$ . We conjecture that  $v_\rho$  is the semiparametric efficiency bound of regular estimators of  $\rho$  that are functions of  $S_N$  under normality of  $u_{it}$ .

Here is a proof of Proposition 2. Notice

$$(14) \quad s - g(\theta_N^0) = \text{vech}(H + H' + \Gamma[S_{uu} - \Phi]\Gamma'),$$

where  $H = \Gamma \frac{1}{n} \sum_{i=1}^N 1_T(u_i - \bar{u})'(\eta_i - \bar{\eta})\Gamma'$ ,  $S_{uu} = \frac{1}{n} \sum_{i=1}^N [(u_i - \bar{u})(u_i - \bar{u})']$ , and  $n = N - 1$ . The variance of  $\sqrt{n}[s - g(\theta_N^0)]$  is

$$\Omega_N = 2D^+(\Gamma \otimes \Gamma)(2(1_T 1_T' \otimes \Phi)\pi_N + \Phi \otimes \Phi)(\Gamma' \otimes \Gamma')D^{+'},$$

where  $D^+$  is  $T(T + 1)/2 \times T^2$ , the generalized inverse of the duplication matrix  $D$  (e.g., Magnus and Neudecker (1999)) associated with a  $T$  dimensional

square matrix. In fact, under normality, the variance of  $\sqrt{n}[\text{vech}(\Gamma[S_{uu} - \Phi]\Gamma')]$  is  $\mathcal{D}^+(\Gamma\Phi\Gamma' \otimes \Gamma\Phi\Gamma')\mathcal{D}^+$ . From  $\text{vec}(H + H') = 2\mathcal{D}^+ \text{vec}(H) = 2\mathcal{D}^+(\Gamma \otimes \Gamma)[1_T \otimes \frac{1}{n} \sum_{i=1}^N (u_i - \bar{u})(\eta_i - \bar{\eta})]$ , its variance is  $4\mathcal{D}^+(\Gamma \otimes \Gamma)(1_T 1_T' \otimes \Phi \pi_N)(\Gamma' \otimes \Gamma')/n$ . Summing up the two terms gives  $\Omega_N$ . Next, introduce

$$W_N = 2\mathcal{D}^+[\Sigma(\theta_N^0) \otimes \Sigma(\theta_N^0)]\mathcal{D}^{+'}$$

The factor estimator maximizes the Wishart likelihood and it is asymptotically equivalent to the following GMM problem:

$$\underset{\theta}{\text{argmin}} n[s - g(\theta)]' W_N^{-1} [s - g(\theta)]$$

(see, e.g., Chamberlain (1984)). It follows that the factor estimator  $\hat{\theta}$  satisfies  $\sqrt{n}(\hat{\theta} - \theta_N^0) \xrightarrow{d} N(0, V)$  with  $V = (G'W^{-1}G)^{-1}G'W^{-1}\Omega W^{-1}G(G'W^{-1}G)^{-1}$ , where  $W$  is the limit of  $W_N$ . Proposition 2 is a consequence of the identity

$$(15) \quad (G'W^{-1}G)^{-1}G'W^{-1}\Omega W^{-1}G(G'W^{-1}G)^{-1} = (G'\Omega^{-1}G)^{-1}.$$

By Rao and Mitra (1971, Chap. 8), the above identity holds if  $W = \Omega + GRG'$  for some  $R$ , where  $R$  is a symmetric matrix subject to the condition that  $W$  is positive definite. We next show that  $W$  is indeed of this form. To see this, from the expressions of  $W_N$  and  $\Sigma(\theta_N^0)$ , we have

$$W_N = \Omega_N + 2\mathcal{D}^+(\Gamma \otimes \Gamma)(1_T 1_T' \otimes 1_T 1_T')\pi_N^2(\Gamma' \otimes \Gamma')\mathcal{D}^{+'}$$

The second term on the right hand side is a quadratic form of the derivative of  $\partial g(\theta)/\partial \pi_N$ , which is equal to  $\mathcal{D}^+(\Gamma \otimes \Gamma)(1_T \otimes 1_T)$ . This implies that  $W_N = \Omega_N + G_N R_N G_N'$  with  $R_N = \text{diag}(0, 2\pi_N^2, 0, \dots, 0)$ . Taking limits gives  $W = \Omega + GRG'$ . This proves Proposition 2.

Dahm and Fuller (1986) gave a similar efficiency result for unrestricted factor models. Proposition 2 does not need normality. Assumption A is sufficient, but normality simplifies the proof. Furthermore, for  $\hat{\rho}$  alone (not the entire vector  $\hat{\theta}$ ) to be as efficient as the optimal GMM so that  $\sqrt{N}(\hat{\rho} - \rho) \xrightarrow{d} N(0, v_\rho)$ , where  $v_\rho$  is the (1, 1)th entry of  $(G'\Omega^{-1}G)^{-1}$ , the existence of  $2 + \epsilon$  moment ( $\epsilon > 0$ ) for  $u_{it}$  is sufficient. A proof of this claim is given in the Supplemental Material.

## 5. MODELS WITH ADDITIONAL REGRESSORS

### *Predetermined Regressors*

First note that the conclusions obtained so far for panel AR(1) can be extended to panel AR( $p$ ) with fixed effects and heteroskedasticity

$$y_{it} = \eta_i + \delta_t + \rho_1 y_{i,t-1} + \dots + \rho_p y_{i,t-p} + u_{it}.$$

The proof needs to modify the  $B$  matrix as well as the initial conditions, as in Section A.3 of Alvarez and Arellano (2004). The rest of the proof is almost identical to the case of AR(1). Similarly, the conclusions also hold when  $y_{it}$  is a vector process so we have a panel VAR. In this case,  $\rho_1, \dots, \rho_p$  and  $\sigma_i^2$  will be matrices. The semiparametric efficiency bound can also be derived for panel VAR models using the same argument as in the proof of Proposition 1. In fact, the efficiency bound in Hahn and Kuersteiner (2002) is also for panel VAR models. So, when  $y_{it}$  is either an AR( $p$ ) or VAR( $p$ ) with heteroskedasticities, our analysis does not require fundamental changes, but instead more complex notations. This is the main reason for our presentation in terms of an AR(1).

With the above fact, we return to the panel AR(1) model in the presence of additional predetermined regressors  $x_{it-1}$  (using a lag here simplifies the VAR notation below):

$$(16) \quad y_{it} = \eta_i + \delta_t + \rho y_{it-1} + \beta' x_{it-1} + u_{it}.$$

Part of the challenge for dynamic panel with predetermined regressors lies in how to control the correlation between the regressors and the fixed effects. For general predetermined regressors, modeling this correlation is a difficult problem. The usual Mundlak and Chamberlain type of projections (period by period, since, for the  $t$ th equation,  $\alpha_i$  can only be projected onto  $x_{i1}, \dots, x_{it-1}$ ) will yield  $O(T^2)$  number of nuisance parameters. However, for a class of predetermined regressors, where  $x_{it}$  is VAR(1) or VAR( $p$ ), the problem of incidental parameters is easy to handle. Consider

$$(17) \quad x_{it} = \tau_i + b_t + \rho_x x_{it-1} + \beta_x y_{it-1} + e_{it},$$

where the  $\tau_i$  are individual effects affecting  $x_{it}$  and  $\tau_i$  can depend on  $\eta_i$ ,  $b_t$  are the time effects, and  $\rho_x$  and  $\beta_x$  are parameters. Let  $z_{it} = (y_{it}, x'_{it})'$ . Combining the two equations gives

$$z_{it} = Az_{i,t-1} + \alpha_i + d_t + \varepsilon_{it},$$

where

$$A = \begin{bmatrix} \rho & \beta' \\ \beta_x & \rho_x \end{bmatrix}, \quad \alpha_i = \begin{bmatrix} \eta_i \\ \tau_i \end{bmatrix}, \quad d_t = \begin{bmatrix} \delta_t \\ b_t \end{bmatrix}, \quad \varepsilon_{it} = \begin{bmatrix} u_{it} \\ e_{it} \end{bmatrix}.$$

So the model becomes a panel VAR(1) model. This panel VAR(1) allows individual effects, time effects, and heteroskedasticity. The factor approach does not estimate the individual effects  $\alpha_i$ , but only  $\pi_N = \frac{1}{n} \sum_{i=1}^N (\alpha_i - \bar{\alpha})(\alpha_i - \bar{\alpha})'$ . Let  $\Phi_t = E(\varepsilon_{it} \varepsilon'_{it})$ . The model parameters are  $(A, \pi_N, \Phi_1, \dots, \Phi_T)$ . Under the assumption that the maximum eigenvalue of  $A$  is less than 1 in absolute value, and a simple modification of Assumption A, the conclusion that there is no asymptotic bias for estimating matrix  $A$  and that the factor estimator achieves the semiparametric efficiency bound under large  $T$  also holds for this model.

*Strictly Exogenous Regressors*

In equation (17), if we set  $\beta_x = 0$  and assume  $\{u_{it}\}$  is independent of  $\{e_{it}\}$ , then  $x_{it}$  becomes strictly exogenous, so that  $x_{it} = \tau_i + b_t + \rho_x x_{i,t-1} + e_{it}$ . With this strictly exogenous regressor, we can still use the panel VAR model to solve for the incidental parameters because this model is a restricted version of the panel VAR(1) described earlier. A further special case is non-dynamic regressors such that  $\rho_x = 0$ . The likelihood method is suitable for imposing these restrictions, and the model is easily estimated using the panel VAR method. We conclude that whether  $T$  is fixed or large, there are no biases stemming from estimating the time effects and heteroskedasticities.

However, dynamic regressors limit the way in which the regressors are correlated with the individual effects  $\eta_i$ . We consider next more general strictly exogenous regressors with

$$E(u_{it} | x_{i1}, x_{i2}, \dots, x_{iT}, \eta_i) = 0,$$

but  $x_{it}$  might be arbitrarily correlated with  $\eta_i$ . Consider the linear projection

$$(18) \quad \eta_i = c_0 + c'_1 x_{i1} + c'_2 x_{i2} + \dots + c'_T x_{iT} + \tau_i.$$

This is known as the Mundlak–Chamberlain projection; see [Mundlak \(1978\)](#), [Chamberlain \(1984\)](#), and [Chamberlain and Moreira \(2009\)](#). We may regard  $\tau_i$  as regression residuals so that  $\sum_{i=1}^N x_{it} \tau_i = 0$  for each  $t$ . If we take this view instead of population projection, then the coefficients  $c_t$  will depend on  $N$  and  $T$ . But this dependence on  $N$  and  $T$  presents no difficulty, in view that  $\pi_N$  and  $\sigma_T^2$  also depend on  $N$  and  $T$ . The focus is a consistent estimation of  $\rho$  and  $\beta$  and the heteroskedasticities  $\sigma_i^2$ . After absorbing  $c_0$  into  $\delta_t$ , we rewrite the model  $y_{it} = \delta_t + \rho y_{it-1} + \beta' x_{it} + \eta_i + u_{it}$  as

$$y_{it} = \delta_t + \rho y_{it-1} + \beta' x_{it} + c'_1 x_{i1} + c'_2 x_{i2} + \dots + c'_T x_{iT} + \tau_i + u_{it}.$$

In matrix notation,

$$By_i = \delta + x_i \beta + 1_T w'_i c + 1_T \tau_i + u_i,$$

where  $x_i = (x_{i1}, x_{i2}, \dots, x_{iT})'$ ,  $w_i = \text{vec}(x'_i)$ , and  $c = (c'_1, c'_2, \dots, c'_T)'$ . Again, for simplicity, we assume  $y_{i0} = 0$ . Rewrite the model as

$$y_i = \Gamma \delta + \Gamma x_i \beta + \Gamma 1_T w'_i c + \Gamma 1_T \tau_i + \Gamma u_i.$$

Define

$$(19) \quad S_N = \frac{1}{n} \sum_{i=1}^N (\dot{y}_i - \Gamma \dot{x}_i \beta - \Gamma 1_T \dot{w}'_i c) (\dot{y}_i - \Gamma \dot{x}_i \beta - \Gamma 1_T \dot{w}'_i c)',$$

where  $\dot{y}_i = y_i - \bar{y}$  and  $\dot{x}_i = x_i - \bar{x}$ , etc. Then, conditional on  $x = (x_1, \dots, x_N)$  and  $\eta = (\eta_1, \dots, \eta_N)$ , the expected value of  $S_N$  is

$$\Sigma(\theta) = \Gamma(1_T 1_T' \pi_N + \Phi)\Gamma',$$

where  $\pi_N = \frac{1}{n} \sum_{i=1}^N (\tau_i - \bar{\tau})^2$ , and  $\theta = (\rho, \beta, c, \pi_N, \sigma_1^2, \dots, \sigma_T^2)$ . We estimate the model by maximizing

$$(20) \quad \ell(\theta) = -\frac{n}{2} \log |\Sigma(\theta)| - \frac{n}{2} \text{tr}[S_N \Sigma(\theta)^{-1}].$$

Under fixed  $T$ , there are no incidental parameters because we only estimate  $\pi_N$  instead of individual  $\tau_i$ . Moreover, the objective function is standard, so root- $N$  consistency and asymptotic normality of  $\hat{\theta}$  follow from the usual argument. No further theoretical proof is needed.

Under large  $T$ , we have additional incidental parameters in the vector  $c$ . We conjecture that under certain conditions such as the rank of regressors and a relative rate between  $T$  and  $N$ , the asymptotic bias arising from estimating these incidental parameters will be negligible. Proof of this conjecture appears to be nontrivial and we leave this as a future research topic.

However, under large  $T$ , we can replace the projection in (18) by  $\eta_i = c_0 + \bar{x}_i' c + \tau_i$ , where  $\bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}$ . This is the standard Mundlak projection by imposing  $c_1 = c_2 = \dots = c_T$ . The dimension of  $c$  is equal to the number of regressors, so there is no incidental-parameters problem. The Mundlak projection works even if the true  $c_t$  are time varying, provided that  $u_{it}$  is homoskedastic ( $\Phi = \sigma_u^2 I_T$ ) and  $N/T^3 \rightarrow 0$  (see the Supplemental Material). With heteroskedasticities in  $u_{it}$ , we need to modify the Mundlak projection by using a weighted average

$$\bar{x}_i(\Phi) = (1_T' \Phi^{-1} 1_T)^{-1} x_i' \Phi^{-1} 1_T = \left( \sum_{t=1}^T \sigma_t^{-2} \right)^{-1} \sum_{t=1}^T \sigma_t^{-2} x_{it}.$$

Absorbing  $(\sum_{t=1}^T \sigma_t^{-2})^{-1}$  into the slope coefficient, we consider the projection

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

Again, the incidental-parameters problem does not occur because  $\Phi$  is already part of the parameters being estimated. After absorbing  $c_0$  into the time effects  $\delta$ , the only additional parameter is  $c$ , whose dimension is fixed. As always, we do not estimate the individual  $\tau_i$ , but  $\pi_N = \frac{1}{n} \sum_{i=1}^N (\tau_i - \bar{\tau})^2$ . Define  $S_N$  as in (19) but with  $\dot{w}_i$  replaced by  $(\dot{x}_i' \Phi^{-1} 1_T)$ . The parameters are estimated by maximizing the objective function (20). In the Supplemental Material, we show that this extended Mundlak projection requires  $N/T \rightarrow 0$  for the bias to not influence the limiting distribution. This implies a bias of order  $O(1/T)$ . We

discuss in the Supplemental Material a further generalization of the Mundlak projection under which the condition  $N/T^3 \rightarrow 0$  becomes sufficient to remove the bias.

## 6. CONCLUSION

The dynamic panel models considered in this paper have had a huge impact on the empirical research in economics and continue to be the workhorses that researchers rely on in contemporaneous empirical studies (e.g., Guiso, Pistaferri, and Schivadi (2005), Blundell, Pistaferri, and Preston (2008), Guvenen (2009)). There has also been enormous advancement in the theoretical analysis over the past three decades. Much of the progress has been summarized in three widely read monographs: Arellano (2003), Baltagi (2005), and Hsiao (2003). A central aspect of the development concerns the problem of incidental parameters (Neyman and Scott (1948), Lancaster (2000, 2002)). Despite the progress, the incidental-parameter problem remains an obstacle to efficient estimation. In this paper, we use the factor analytical perspective to shed light on the analysis of dynamic panel models with fixed effects. We consider additional incidental parameters: time effects and heteroskedasticity, under both fixed and large  $T$ .

We establish some desirable and excellent properties for the factor approach. Whether for large or small  $T$ , the method produces a consistent and efficient estimator. No asymptotic bias exists for the dynamic parameter even when  $T$  and  $N$  are comparable. Neither mean-stationarity nor covariance stationarity is required to establish these properties. Thus the factor approach provides a useful paradigm to deal with the incidental-parameter problem occurring in both dimensions.

## APPENDIX A: PROOFS

Introduce two matrices to be used throughout:

$$(21) \quad J_T = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \quad L = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \rho & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{T-2} & \cdots & \rho & 1 & 0 \end{bmatrix},$$

where  $J_T$  is the matrix derivative  $-dB/d\rho$  and  $L = J_T \Gamma$ .

**PROOF OF LEMMA 1:** The limits are the same when  $L$  is replaced by  $\Gamma$ . Note that  $1'_T \Phi^{-1} \Gamma 1_T = \sum_{t=1}^T \frac{1}{\sigma_t^2} (1 + \rho + \cdots + \rho^{t-1})$ . But  $(1 + \rho + \cdots + \rho^{t-1}) \rightarrow 1/(1 - \rho)$  as  $t \rightarrow \infty$  and  $\sum_{t=1}^T \frac{1}{\sigma_t^2} = T \omega_T \rightarrow \infty$ . By the Toeplitz lemma (Hall and Heyde (1980, p. 31)),  $1'_T \Phi^{-1} \Gamma 1_T / (T \omega_T) \rightarrow 1/(1 - \rho)$ , proving part (a).

For part (b), denote its limit by  $C$ ; we show  $C = 1/(1 - \rho)^2$ . Notice  $\Gamma = I_T + \rho L$ . Thus

$$1'_T \Gamma' \Phi^{-1} \Gamma 1_T = 1'_T \Phi^{-1} 1_T + 2\rho(1'_T \Phi^{-1} L 1_T) + \rho^2(1'_T L' \Phi^{-1} L 1_T).$$

Divide by  $1'_T \Phi^{-1} 1_T$  on each side and take the limit; we have  $C = 1 + 2\rho \frac{1}{1-\rho} + \rho^2 C$ , implying  $C = 1/(1 - \rho)^2$ . Note that the limits are the same when  $L$  is replaced by  $\Gamma$  and vice versa. *Q.E.D.*

To prove consistency, we need to distinguish the true parameters  $(\rho^0, \pi_N^0, \Phi^0)$  from the variables  $(\rho, \pi_N, \Phi)$  in the likelihood function. So  $\Gamma^0$  denotes  $\Gamma$  when  $\rho = \rho^0$ .

LEMMA A.1: *As  $N, T \rightarrow \infty$ , uniformly for  $(\rho, \sigma_1^2, \dots, \sigma_T^2) \in \Theta_1$ ,*

- (i)  $\frac{1}{T} \text{tr}[B\Gamma^0 \frac{1}{N} \sum_{i=1}^N (u_i u'_i - \Phi^0) \Gamma^{0'} B' \Phi^{-1}] = o_p(1)$ ,
- (ii)  $\frac{1}{T} \text{tr}[B\Gamma^0 1_T \frac{1}{N} \sum_{i=1}^N u'_i \Gamma^{0'} B' \Phi^{-1} \eta_i] = o_p(1)$ ,
- (iii)  $\frac{1}{T^2} 1'_T (\Phi^{-1} B\Gamma^0 \frac{1}{N} \sum_{i=1}^N (u_i u'_i - \Phi^0) \Gamma^{0'} B' \Phi^{-1}) 1_T = o_p(1)$ ,
- (iv)  $\frac{1}{T^2} 1'_T (\Phi^{-1} B\Gamma^0 1_T \frac{1}{N} \sum_{i=1}^N u'_i \Gamma^{0'} B' \Phi^{-1} \eta_i) 1_T = o_p(1)$ .

PROOF: Consider (i). Let  $W = \frac{1}{N} \sum_{i=1}^N (u_i u'_i - \Phi^0)$  and use  $B\Gamma^0 = I_T + (\rho^0 - \rho)L^0$ ; we have

$$\begin{aligned} & \text{tr} \left[ B\Gamma^0 \frac{1}{N} \sum_{i=1}^N (u_i u'_i - \Phi^0) \Gamma^{0'} B' \Phi^{-1} \right] \\ &= \text{tr}(W\Phi^{-1}) + 2(\rho^0 - \rho) \text{tr}(L^0 W \Phi^{-1}) + (\rho^0 - \rho)^2 \text{tr}(L^0 W L^0 \Phi^{-1}). \end{aligned}$$

Note that  $\Phi^{-1}$  is a diagonal matrix:

$$\begin{aligned} \frac{1}{T} |\text{tr}(W\Phi^{-1})| &= \frac{1}{TN} \left| \sum_{i=1}^N \sum_{t=1}^T (u_{it}^2 - \sigma_t^{02}) / \sigma_t^2 \right| \\ &\leq N^{-1/2} \left( \frac{1}{T} \sum_{t=1}^T \frac{1}{\sigma_t^4} \right)^{1/2} \\ &\quad \times \left( \frac{1}{T} \sum_{t=1}^T \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^N (u_{it}^2 - \sigma_t^{02}) \right]^2 \right)^{1/2} \\ &= N^{-1/2} O_p(1) \\ &= o_p(1) \end{aligned}$$

uniformly for all  $\sigma_i^2 \in [a, b]$  since  $1/\sigma_i^4 \leq 1/a^2 < \infty$ . Next,

$$\frac{1}{T} \text{tr}(L^0 W \Phi^{-1}) = \sum_{k=0}^{T-2} (\rho^0)^k \frac{1}{NT} \sum_{t=k+2}^T \sum_{i=1}^N \frac{u_{it-k-1} u_{it}}{\sigma_i^2}.$$

Thus

$$\begin{aligned} \left| \frac{1}{T} \text{tr}(L^0 W \Phi^{-1}) \right| &\leq N^{-1/2} \sum_{k=0}^{T-2} |\rho^0|^k \left( \frac{1}{T} \sum_{t=k+2}^T \frac{1}{\sigma_t^4} \right)^{1/2} \\ &\quad \times \left( \frac{1}{T} \sum_{t=k+2}^T \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{it-k-1} u_{it} \right]^2 \right)^{1/2}. \end{aligned}$$

Thus

$$\sup_{\Phi} \left| \frac{1}{T} \text{tr}(L^0 W \Phi^{-1}) \right| \leq N^{-1/2} \frac{1}{a} A_{NT},$$

where  $A_{NT} = \sum_{k=0}^{T-2} |\rho^0|^k \left( \frac{1}{T} \sum_{t=k+2}^T \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{it-k-1} u_{it} \right]^2 \right)^{1/2}$ . Note that  $A_{NT} = O_p(1)$  because its expected value is bounded by Assumption A. Thus  $\sup_{\Phi} \left| \frac{1}{T} \text{tr}(L^0 W \Phi^{-1}) \right| = o_p(1)$ .

Next, let  $W_{i,hk}$  denote the  $(h, k)$ th entry of the matrix  $u_i u_i' - \Phi^0$ . Then

$$\text{tr}(L^0 W L^0 \Phi^{-1}) = \sum_{t=1}^{T-1} \frac{1}{\sigma_{t+1}^2} \sum_{h=1}^t (\rho^0)^{t-h} \sum_{k=1}^t (\rho^0)^{t-k} \frac{1}{N} \sum_{i=1}^N W_{i,hk}.$$

It follows that

$$\begin{aligned} \frac{1}{T} \left| \text{tr}(L^0 W L^0 \Phi^{-1}) \right| &\leq N^{-1/2} \left( \frac{1}{T} \sum_{t=1}^{T-1} \frac{1}{\sigma_{t+1}^4} \right)^{1/2} \\ &\quad \times \left( \frac{1}{T} \sum_{t=1}^{T-1} \left[ \sum_{h=1}^t (\rho^0)^{t-h} \sum_{k=1}^t (\rho^0)^{t-k} \frac{1}{\sqrt{N}} \sum_{i=1}^N W_{i,hk} \right]^2 \right)^{1/2}. \end{aligned}$$

Again,  $\frac{1}{T} \sum_{t=1}^{T-1} \frac{1}{\sigma_{t+1}^4} \leq 1/a^2$  uniformly on  $\Theta_1$ . The above is  $O_p(N^{-1/2})$ . Summarizing results, we prove (i).

The proofs for (ii)–(iv) are similar, and the details are omitted. *Q.E.D.*

PROOF OF LEMMA 2: Notice that

$$\begin{aligned}
 BS_N B' \Phi^{-1} &= B \Sigma(\theta_N^0) B' \Phi^{-1} + B \Gamma^0 \frac{1}{N} \sum_{i=1}^N (u_i u_i' - \Phi^0) \Gamma^{0'} B' \Phi^{-1} \\
 &\quad + B \Gamma^0 1_T \frac{1}{N} \sum_{i=1}^N u_i \Gamma^{0'} B' \Phi^{-1} \eta_i \\
 &\quad + B \Gamma^0 \frac{1}{N} \sum_{i=1}^N u_i 1_T' \Gamma^{0'} B' \Phi^{-1} \eta_i,
 \end{aligned}$$

where  $\Sigma(\theta_N^0) = \Gamma^0 (1_T 1_T' \pi_N^0 + \Phi^0) \Gamma^{0'}$ . Dividing by  $T$  and taking the trace, the last three terms are  $o_p(1)$  by Lemma A.1(i) and (ii). Thus, uniformly on  $\Theta_1$ ,

$$(22) \quad \frac{1}{T} \text{tr}(BS_N B' \Phi^{-1}) = \frac{1}{T} \text{tr}[B \Sigma(\theta_N^0) B' \Phi^{-1}] + o_p(1).$$

Similarly, by Lemma A.1(iii) and (iv), we have, uniformly on  $\Theta_1$ ,

$$\frac{1}{T^2} (1_T' \Phi^{-1} BS_N B' \Phi^{-1} 1_T) = \frac{1}{T^2} [1_T' \Phi^{-1} B \Sigma(\theta_N^0) B' \Phi^{-1} 1_T] + o_p(1).$$

The following also holds because  $1_T' \Phi^{-1} 1_T$  is of order  $T$  uniformly on  $\Theta_1$ :

$$(23) \quad \frac{1}{T} \frac{(1_T' \Phi^{-1} BS_N B' \Phi^{-1} 1_T)}{1_T' \Phi^{-1} 1_T} = \frac{1}{T} \frac{[1_T' \Phi^{-1} B \Sigma(\theta_N^0) B' \Phi^{-1} 1_T]}{1_T' \Phi^{-1} 1_T} + o_p(1).$$

From  $B \Gamma^0 = I_T + (\rho^0 - \rho) L^0$ , we obtain

$$\begin{aligned}
 &\text{tr}[B \Sigma(\theta_N^0) B' \Phi^{-1}] \\
 &= 1_T' \Phi^{-1} 1_T \pi_N^0 + \text{tr}(\Phi^0 \Phi^{-1}) + 2(\rho^0 - \rho) (1_T' \Phi^{-1} L^0 1_T) \pi_N^0 \\
 &\quad + (\rho^0 - \rho)^2 [(1_T' L^0 \Phi^{-1} L^0 1_T) \pi_N^0 + \text{tr}(L^0 \Phi^0 L^0 \Phi^{-1})].
 \end{aligned}$$

We have used  $\text{tr}(L^0 \Phi^0 \Phi^{-1}) = 0$  because  $L^0$  is lower triangular and  $\Phi^0$  and  $\Phi$  are diagonal. Similarly,

$$\begin{aligned}
 (24) \quad &1_T' \Phi^{-1} B \Sigma(\theta_N^0) B' \Phi^{-1} 1_T \\
 &= (1_T' \Phi^{-1} 1_T)^2 \pi_N^0 + 1_T' \Phi^{-1} \Phi^0 \Phi^{-1} 1_T \\
 &\quad + 2(\rho^0 - \rho) [1_T' \Phi^{-1} L^0 1_T] (1_T' \Phi^{-1} 1_T) \pi_N^0 + (1_T' \Phi^{-1} L^0 \Phi^0 \Phi^{-1} 1_T) \\
 &\quad + (\rho^0 - \rho)^2 [(1_T' \Phi^{-1} L^0 1_T)^2 \pi_N^0 + 1_T' \Phi^{-1} L^0 \Phi^0 L^0 \Phi^{-1} 1_T].
 \end{aligned}$$

These equations imply, after canceling the common terms,

$$\begin{aligned} & \frac{1}{T} \operatorname{tr}[B\Sigma(\theta_N^0)B'\Phi^{-1}] - \frac{1}{T} \frac{(1_T'\Phi^{-1}BS_NB'\Phi^{-1}1_T)}{1_T'\Phi^{-1}1_T} \\ &= \frac{1}{T} \operatorname{tr}(\Phi^0\Phi^{-1}) + (\rho^0 - \rho)^2 \left[ \frac{1}{T} \operatorname{tr}(L^0\Phi^0L^0\Phi^{-1}) + \Delta \right] + O\left(\frac{1}{T}\right), \end{aligned}$$

where

$$\Delta = \frac{1}{T} \left[ (1_T'L^0\Phi^{-1}L^01_T) - \frac{(1_T'\Phi^{-1}L^01_T)^2}{1_T'\Phi^{-1}1_T} \right] \pi_N^0,$$

and  $\Delta \geq 0$  by the Cauchy–Schwarz inequality. The  $O(T^{-1})$  term represents

$$2(\rho^0 - \rho)(1_T'\Phi^{-1}L^0\Phi^0\Phi^{-1}1_T) + (\rho^0 - \rho)^2 1_T'\Phi^{-1}L^0\Phi^0L^0\Phi^{-1}1_T$$

divided by  $T(1_T'\Phi^{-1}1_T)$ . The above is uniformly  $O(T)$  on  $\Theta_1$ . But  $T(1_T'\Phi^{-1} \times 1_T) \geq T^2/b = O(T^2)$ . Thus the ratio is  $O(T^{-1})$ . Summarizing results, and in view of (22) and (23), we have

$$\begin{aligned} & \frac{1}{T} \operatorname{tr}(BS_NB'\Phi^{-1}) - \frac{1}{T} \frac{(1_T'\Phi^{-1}BS_NB'\Phi^{-1}1_T)}{1_T'\Phi^{-1}1_T} \\ &= \frac{1}{T} \operatorname{tr}(\Phi^0\Phi^{-1}) + (\rho^0 - \rho)^2 \left[ \frac{1}{T} \operatorname{tr}(L^0\Phi^0L^0\Phi^{-1}) + \Delta \right] + o_p(1) \end{aligned}$$

uniformly on  $\Theta_1$ . Further, the preceding analysis shows that the left hand side of (23) is  $O_p(1)$  uniformly on  $\Theta_1$ . Multiplying by  $T$ , it is  $O_p(T)$  uniformly on  $\Theta_1$ . Thus

$$\frac{1}{T} \log \left( \frac{1_T'\Phi^{-1}BS_NB'\Phi^{-1}1_T}{1_T'\Phi^{-1}1_T} \right) = O_p \left( \frac{\log(T)}{T} \right)$$

uniformly on  $\Theta_1$ . Finally, from  $\log |\Phi| = \sum_{t=1}^T \log(\sigma_t^2)$ ,  $\operatorname{tr}(\Phi^0\Phi^{-1}) = \sum_{t=1}^T \frac{\sigma_t^{02}}{\sigma_t^2}$ , and the definition of concentrated likelihood function, we obtain the asymptotic representation in Lemma 2. *Q.E.D.*

**PROOF OF LEMMA 3:** Given Lemma 2, we already argued the consistency of  $\hat{\rho}$  and the average consistency of (6). It remains to show  $\hat{\pi} = \pi_N^0 + o_p(1)$  and  $\hat{\sigma}_t^2 = \sigma_t^{02} + o_p(1)$  for each  $t$ . Consider  $\hat{\pi}$  in (9). Equation (23) holds with  $1/T$  replaced by  $1/1_T'\Phi^{-1}1_T$  because  $1/(Tb) \leq (1_T'\Phi^{-1}1_T)^{-1} \leq 1/(Ta)$  uniformly on  $\Theta_1$ . Thus

$$(25) \quad \frac{(1_T'\Phi^{-1}BS_NB'\Phi^{-1}1_T)}{(1_T'\Phi^{-1}1_T)^2} = \frac{[1_T'\Phi^{-1}B\Sigma(\theta_N^0)B'\Phi^{-1}1_T]}{(1_T'\Phi^{-1}1_T)^2} + o_p(1).$$

Because the above holds uniformly on  $\Theta_1$ , it holds at  $(\widehat{\rho}, \widehat{\Phi})$ . Evaluate the preceding equation at  $(\widehat{\rho}, \widehat{\Phi})$ , and subtract  $1/(1'_T \widehat{\Phi}^{-1} 1_T) = O_p(1/T) = o_p(1)$  on each side of the equation; we get

$$\widehat{\pi} = \frac{[1'_T \widehat{\Phi}^{-1} \widehat{B} \Sigma(\theta_N^0) \widehat{B}' \widehat{\Phi}^{-1} 1_T]}{(1'_T \widehat{\Phi}^{-1} 1_T)^2} + o_p(1).$$

Evaluate (24) at  $(\widehat{\rho}, \widehat{\Phi})$  and divide it by  $(1'_T \widehat{\Phi}^{-1} 1_T)^2$ ; we get

$$\begin{aligned} \widehat{\pi} &= \pi_N^0 + 2(\rho^0 - \widehat{\rho}) \left[ \frac{1'_T \widehat{\Phi}^{-1} L^0 1_T}{1'_T \widehat{\Phi}^{-1} 1_T} \right] \pi_N^0 \\ &\quad + (\rho^0 - \widehat{\rho})^2 \left[ \frac{(1'_T \widehat{\Phi}^{-1} L^0 1_T)^2}{(1'_T \widehat{\Phi}^{-1} 1_T)^2} \right] \pi_N^0 + o_p(1); \end{aligned}$$

here  $o_p(1)$  absorbs some  $O_p(1/T)$  terms. The terms inside the square brackets are  $O_p(1)$ . From the consistency of  $\widehat{\rho}$ , the preceding equation implies  $\widehat{\pi} = \pi_N^0 + o_p(1)$ .

We next show that  $\widehat{\sigma}_i^2$  is consistent for  $\sigma_i^{02}$  for every  $i$ . The first order condition for  $\Phi$  implies that

$$(26) \quad \widehat{\Phi} = \text{diag}(\widehat{B} S_N \widehat{B}' - 2 \widehat{B} S_N \widehat{B}' \widehat{\Phi}^{-1} 1_T 1'_T c_T + 1_T 1'_T \widehat{\pi}),$$

where  $c_T = \widehat{\pi}/(1 + T \widehat{\omega}_T \widehat{\pi})$  with  $\widehat{\omega}_T = 1'_T \widehat{\Phi}^{-1} 1_T$ . It is easy to show that  $2 \Omega(\widehat{\theta}) \widehat{\Phi}^{-1} 1_T 1'_T c_T = 2 1_T 1'_T \widehat{\pi}$ , where  $\Omega(\widehat{\theta}) = 1_T 1'_T \widehat{\pi} + \widehat{\Phi}$ . So we can rewrite  $\widehat{\Phi}$  as

$$\widehat{\Phi} = \text{diag}(\widehat{B} S_N \widehat{B}' - 1_T 1'_T \widehat{\pi} - 2[\widehat{B} S_N \widehat{B}' - \Omega(\widehat{\theta})] \widehat{\Phi}^{-1} 1_T 1'_T c_T).$$

Given the consistency of  $\widehat{\rho}$  and  $\widehat{\pi}$ , and with  $c_T$  being  $O_p(T^{-1})$ , we can show that every diagonal element of  $[\widehat{B} S_N \widehat{B}' - \Omega(\widehat{\theta})] \widehat{\Phi}^{-1} 1_T 1'_T c_T$  is  $o_p(1)$ . So to establish the consistency of  $\widehat{\sigma}_i^2$ , it suffices to show that the diagonal elements of  $\widehat{B} S_N \widehat{B}' - 1_T 1'_T \widehat{\pi} - \Phi^0$  are  $o_p(1)$ . Using a similar argument leading to (22), every diagonal element of  $\widehat{B} S_N \widehat{B}' - \widehat{B} \Sigma(\theta_N^0) \widehat{B}'$  is  $o_p(1)$  (here no trace, thus no need to divide by  $T$ ), where  $\Sigma(\theta_N^0) = \Gamma^0 \Omega^0 \Gamma^{0\prime}$  with  $\Omega^0 = (1_T 1'_T \pi_N^0 + \Phi^0)$ . Using  $\widehat{B} \Gamma^0 = I_T + (\rho^0 - \widehat{\rho}) L^0$ , we have

$$\widehat{B} \Sigma(\theta_N^0) \widehat{B}' = \Omega^0 + (\rho^0 - \widehat{\rho}) \Omega^0 L^{0\prime} + (\rho^0 - \widehat{\rho}) L^0 \Omega^0 + (\rho^0 - \widehat{\rho})^2 L^0 \Omega^0 L^{0\prime}.$$

Owing to the consistency of  $\widehat{\rho}$ , every diagonal element of the last three matrices is  $o_p(1)$ . Thus, the diagonal elements of  $\widehat{B} S_N \widehat{B}' - \Omega^0 = \widehat{B} S_N \widehat{B}' - 1_T 1'_T \pi_N^0 - \Phi^0$  are  $o_p(1)$ . By the consistency of  $\widehat{\pi}$ , the diagonal elements of  $\widehat{B} S_N \widehat{B}' - 1_T 1'_T \widehat{\pi} - \Phi^0$  are also  $o_p(1)$ . This proves the consistency of  $\widehat{\sigma}_i^2$ . Q.E.D.

*Asymptotic Representations of  $\widehat{\Phi}$ ,  $\widehat{\pi}$ , and  $\widehat{\rho}$*

PROOF OF (11): Consider (26). Again using  $\widehat{B}\Gamma^0 = I_T + (\rho^0 - \widehat{\rho})L^0$ , it can be shown that

$$\begin{aligned} \widehat{B}S_N\widehat{B}' &= 1_T1_T'\pi_N^0 + \frac{1}{N}\sum_{i=1}^N u_i1_T'\eta_i + \frac{1}{N}\sum_{i=1}^N 1_Tu_i'\eta_i + \frac{1}{N}\sum_{i=1}^N u_iu_i' \\ &\quad + (\widehat{\rho} - \rho^0)C_{NT}, \end{aligned}$$

where  $C_{NT}$  is  $T \times T$ , with each diagonal element being  $O_p(1)$ . It follows that

$$\begin{aligned} &\widehat{B}S_N\widehat{B}'\widehat{\Phi}^{-1}1_T1_T'c_T \\ &= 1_T1_T'(1_T'\widehat{\Phi}^{-1}1_T)c_T\pi_N^0 + \frac{1}{N}\sum_{i=1}^N u_i1_T'\eta_i(1_T'\widehat{\Phi}^{-1}1_T)c_T \\ &\quad + 1_T1_T'\left(\frac{1}{N}\sum_{i=1}^N (u_i'\widehat{\Phi}^{-1}1_T)\eta_i\right)c_T + \frac{1}{N}\sum_{i=1}^N u_i1_T'(u_i'\widehat{\Phi}^{-1}1_T)c_T \\ &\quad + (\widehat{\rho} - \rho^0)D_{NT}; \end{aligned}$$

here  $D_{NT}$  is  $T \times T$ , with diagonal elements being  $O_p(1)$ . Using  $(1_T'\widehat{\Phi}^{-1}1_T)c_T = 1 + O_p(1/T)$ ,

$$\begin{aligned} &\widehat{B}S_N\widehat{B}' - 2\widehat{B}S_N\widehat{B}'\widehat{\Phi}^{-1}1_T1_T'c_T \\ &= \frac{1}{N}\sum_{i=1}^N u_iu_i' - 1_T1_T'\pi_N^0 + (\widehat{\rho} - \rho^0)(C_{NT} - D_{NT}) + O_p(1/T) \\ &\quad + \frac{1}{N}\sum_{i=1}^N (1_Tu_i' - u_i1_T')\eta_i - 21_T1_T'\left(\frac{1}{N}\sum_{i=1}^N (u_i'\widehat{\Phi}^{-1}1_T)\eta_i\right)c_T \\ &\quad - 2\frac{1}{N}\sum_{i=1}^N u_i1_T'(u_i'\widehat{\Phi}^{-1}1_T)c_T. \end{aligned}$$

The first term on the third line has zero diagonal elements. Using the average consistency (6) and  $c_T = O_p(1/T)$ , each of the last two expressions has diagonal elements being of  $o_p(N^{-1/2})$ . Thus, adding  $1_T1_T'\widehat{\pi}$  on each side, and by (26) together with  $\text{diag}(1_T1_T') = I_T$ , we get

$$\begin{aligned} (27) \quad \widehat{\Phi} &= \text{diag}\left(\frac{1}{N}\sum_{i=1}^N u_iu_i'\right) + I_T(\widehat{\pi} - \pi_N^0) \\ &\quad + (\widehat{\rho} - \rho^0)O_p(1) + O_p(1/T) + o_p(N^{-1/2}). \end{aligned}$$

This representation of  $\widehat{\Phi}$ , together with the consistency of  $\widehat{\pi}$  and  $\widehat{\rho}$ , allows us to show (below)  $\widehat{\pi} - \pi_N^0 = O_p(1/\sqrt{NT}) + O_p(1/N)$  and  $\widehat{\rho} - \rho^0 = O_p(1/\sqrt{NT})$ . Inserting these rates into (27) and subtracting  $\Phi^0$  on each side give (11).  
*Q.E.D.*

PROOF OF (10): From (9), direct calculation shows

$$\begin{aligned}
 (28) \quad \widehat{\pi}_N - \pi_N^0 &= 2(\rho^0 - \widehat{\rho}) \left[ \frac{1'_T \widehat{\Phi}^{-1} L^0 1_T}{1'_T \widehat{\Phi}^{-1} 1_T} \right] \pi_N^0 + 2 \frac{\frac{1}{N} \sum_{i=1}^N u'_i \widehat{\Phi}^{-1} 1_T \eta_i}{1'_T \widehat{\Phi}^{-1} 1_T} \\
 &\quad + \frac{1'_T \widehat{\Phi}^{-1} \left[ \frac{1}{N} \sum_{i=1}^N (u_i u'_i - \Phi^0) \right] \widehat{\Phi}^{-1} 1_T}{(1'_T \widehat{\Phi}^{-1} 1_T)^2} \\
 &\quad + (\widehat{\rho} - \rho^0) o_p(1) + o_p\left(\frac{1}{\sqrt{NT}}\right);
 \end{aligned}$$

here  $(\widehat{\rho} - \rho^0) o_p(1)$  combines many terms of the form  $(\widehat{\rho} - \rho^0)[O_p(T^{-1}) + O_p(N^{-1/2})] + (\widehat{\rho} - \rho^0)^2 O_p(1)$ , which can be ignored because it is dominated by the first term on the right hand side. The first two terms on the right hand side determine the limiting distribution of  $\widehat{\pi}$ . The second term has a bias of order  $(1/N)$  if  $E(u_{it}^3) \neq 0$  due to the estimation of  $\Phi^0$ . To see this,

$$\sum_{i=1}^N u'_i \widehat{\Phi}^{-1} 1_T \eta_i = \sum_{i=1}^N u'_i \Phi^{0-1} 1_T \eta_i + \sum_{i=1}^N u'_i (\widehat{\Phi}^{-1} - \Phi^{0-1}) 1_T \eta_i.$$

Dividing by  $NT$ , we have  $\frac{1}{TN} \sum_{i=1}^N u'_i (\widehat{\Phi}^{-1} - \Phi^{0-1}) 1_T \eta_i = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{u_{it}}{\widehat{\sigma}_t^2 \sigma_t^{02}} \times (\widehat{\sigma}_t^2 - \sigma_t^{02})$ . Using the expression of  $\widehat{\Phi}$  in (27) (subtract  $\Phi^0$  on each side), we can show

$$\begin{aligned}
 \frac{1}{TN} \sum_{i=1}^N u'_i (\widehat{\Phi}^{-1} - \Phi^{0-1}) 1_T \eta_i &= \frac{1}{TN} \sum_{i=1}^N \sum_{t=1}^T \frac{u_{it} \eta_i}{\sigma_t^{04}} \left[ \frac{1}{N} \sum_{k=1}^N (u_{kt}^2 - \sigma_t^{02}) \right] \\
 &\quad + o_p(1/\sqrt{NT}),
 \end{aligned}$$

where  $o_p(1/\sqrt{NT})$  contains  $(\widehat{\pi} - \pi_N^0) O_p(1/\sqrt{NT}) + (\widehat{\rho} - \rho^0) O_p(1/\sqrt{NT})$  as one of the components. The first term on the right is dominated by its expected value, which is  $O(1/N)$  and is given by

$$\frac{1}{N} \left( \frac{1}{T} \sum_{t=1}^T \frac{\nu_t}{(\sigma_t^0)^4} \right) \left( \frac{1}{N} \sum_{i=1}^N \eta_i \right),$$

where  $\nu_i = E(u_{it}^3)$  (assuming not depending on  $i$ ).

The third term on the right hand side of  $\hat{\pi} - \pi_N^0$  can be shown to be  $o_p(1/\sqrt{NT}) + O_p(1/N^{3/2})$ . If  $E(u_{it}^3) = 0$ , the term  $O_p(1/N^{3/2})$  will be absent. It is easy to show that

$$\frac{1'_T \hat{\Phi}^{-1} L^0 1_T}{1'_T \hat{\Phi}^{-1} 1_T} = \frac{1'_T \Phi^{0-1} L^0 1_T}{1'_T \Phi^{0-1} 1_T} + o_p(1), \quad \frac{1'_T \Phi^{0-1} 1_T}{1'_T \hat{\Phi}^{-1} 1_T} = 1 + o_p(1).$$

To summarize results,

$$(29) \quad \hat{\pi} - \pi_N^0 = 2(\rho^0 - \hat{\rho}) \left[ \frac{1'_T \Phi^{0-1} L^0 1_T}{1'_T \Phi^{0-1} 1_T} \right] \pi_N^0 + 2 \frac{\frac{1}{N} \sum_{i=1}^N u'_i \Phi^{0-1} 1_T \eta_i}{1'_T \Phi^{0-1} 1_T} + \frac{b}{N} + O_p(1/N^{3/2}) + o_p(1/\sqrt{NT}),$$

where the bias term  $b$  is defined in the main text. Thus if  $T/N^2 \rightarrow 0$ ,

$$(30) \quad \sqrt{NT} \left( \hat{\pi} - \pi_N^0 - \frac{1}{N} b \right) = 2\sqrt{NT}(\rho^0 - \hat{\rho}) \left[ \frac{1'_T \Phi^{0-1} L^0 1_T}{1'_T \Phi^{0-1} 1_T} \right] \pi_N^0 + 2 \frac{\frac{1}{\sqrt{NT}} \sum_{i=1}^N u'_i \Phi^{0-1} 1_T \eta_i}{T^{-1} 1'_T \Phi^{0-1} 1_T} + o_p(1),$$

which is the asymptotic representation of  $\hat{\pi}$  in (10).

*Q.E.D.*

PROOF OF (7): The first order condition implies that  $\hat{\rho} = A_{NT}^{-1} \cdot B_{NT}$ , where

$$A_{NT} = \text{tr}[J_T S_N J'_T \hat{\Phi}^{-1}] - c_T (1'_T \hat{\Phi}^{-1} J_T S_N J'_T \hat{\Phi}^{-1} 1_T),$$

$$B_{NT} = \text{tr}[J_T S_N \hat{\Phi}^{-1}] - c_T (1'_T \hat{\Phi}^{-1} J_T S_N \hat{\Phi}^{-1} 1_T),$$

and  $c_T = \hat{\pi}/(1 + T\hat{\omega}_T \hat{\pi})$ . Using  $I_T = B^{0'} + \rho^0 J'_T$ , we can write  $S_N = S_N B^{0'} + \rho^0 S_N J'_T$ , so that  $B_{NT} = \rho^0 A_{NT} + B_{NT}^*$ , with

$$B_{NT}^* = \text{tr}(J_T S_N B^{0'} \hat{\Phi}^{-1}) - c_T (1'_T \hat{\Phi}^{-1} J_T S_N B^{0'} \hat{\Phi}^{-1} 1_T).$$

Thus  $\hat{\rho} - \rho^0 = A_{NT}^{-1} \cdot B_{NT}^*$ , or equivalently,

$$\sqrt{NT}(\hat{\rho} - \rho^0) = \left( \frac{1}{T} A_{NT} \right)^{-1} \cdot \left( \frac{N}{T} \right)^{1/2} B_{NT}^*.$$

We shall show

$$(31) \quad \frac{1}{T} A_{NT} = \frac{1}{T} \text{tr}(L^0 \Phi^0 L^{0'} \Phi^{0-1}) + o_p(1),$$

$$(32) \quad \left(\frac{N}{T}\right)^{1/2} B_{NT}^* = \frac{1}{\sqrt{NT}} \sum_{i=1}^N u_i' \Phi^{0-1} L^0 u_i + o_p(1).$$

Because  $A_{NT}$  occurs in the denominator of  $\widehat{\rho}$ , the analysis of (31) is relatively easy since it converges to a positive constant (all that is needed is a consistency argument). In contrast, the analysis of the numerator (32), which determines the limiting distribution, is much more delicate. Potential biases must be taken care of. Note that the two equations together imply (7).

Consider (31). It can be shown that replacing  $\widehat{\Phi}$  by  $\Phi^0$  does not affect the limit. Using  $J_T I^0 = L^0$ ,

$$\begin{aligned} \frac{1}{T} \text{tr}(J_T S_N J_T' \widehat{\Phi}^{-1}) &= \frac{1}{T} (1_T' L^0 \Phi^{0-1} L^0 1_T) \pi_N^0 \\ &\quad + \frac{1}{T} \text{tr}(L^0 \Phi^0 L^{0'} \Phi^{0-1}) + o_p(1), \\ \frac{1}{T} c_T (1_T' \widehat{\Phi}^{-1} J_T S_N B^0 \widehat{\Phi}^{-1} 1_T) &= \frac{1}{T} c_T (1_T' \Phi^{0-1} L^0 1_T)^2 \pi_N^0 \\ &\quad + \frac{1}{T} c_T (1_T' \Phi^{0-1} L^0 \Phi^0 L^{0'} \Phi^{0-1} 1_T) \\ &\quad + o_p(1). \end{aligned}$$

The difference between the leading terms in the preceding two equations is  $o_p(1)$  because they have the same limit. This follows from Lemma 1 [applied with  $T$  replacing  $(1_T' \Phi^{0-1} 1_T)$  in the denominator; also note we replace  $c_T$  by  $1/(1_T' \Phi^{0-1} 1_T)$  without affecting the limit]. The second term in the second equation is  $O_p(1/T)$ . We thus obtain (31).

Next consider (32). First rewrite  $c_T$  as

$$c_T = \frac{1}{1_T' \widehat{\Phi}^{-1} 1_T} - \frac{1}{1_T' \widehat{\Phi}^{-1} \widehat{B} S_N \widehat{B}' \widehat{\Phi}^{-1} 1_T} = \frac{1}{T \widehat{\omega}_T} - \frac{1}{d_T},$$

where  $d_T$  is implicitly defined and is  $O_p(T^2)$ . We rewrite  $B_{NT}^*$  as

$$(33) \quad \begin{aligned} B_{NT}^* &= \text{tr}(J_T S_N B^0 \widehat{\Phi}^{-1}) - \frac{1}{T \widehat{\omega}_T} 1_T' \widehat{\Phi}^{-1} J_T S_N B^0 \widehat{\Phi}^{-1} 1_T \\ &\quad + \frac{1}{d_T} 1_T' \widehat{\Phi}^{-1} J_T S_N B^0 \widehat{\Phi}^{-1} 1_T. \end{aligned}$$

Using  $J_T \Gamma^0 = L^0$  and  $\Gamma^{0r} B^{0r} = I_T$ , the first two terms in the preceding equation are

$$\begin{aligned} & \text{tr}(J_T S_N B^{0r} \widehat{\Phi}^{-1}) - \frac{1}{T \widehat{\omega}_T} 1'_T \widehat{\Phi}^{-1} J_T S_N B^{0r} \widehat{\Phi}^{-1} 1_T \\ &= \frac{1}{N} \sum_{i=1}^N u'_i \widehat{\Phi}^{-1} L^0 u_i + \frac{1}{N} \sum_{i=1}^N (u'_i \widehat{\Phi}^{-1} L^0 1_T) \eta_i \\ & \quad - \frac{1}{T \widehat{\omega}_T} 1'_T \widehat{\Phi}^{-1} L^0 \frac{1}{N} \sum_{i=1}^N u_i u'_i \widehat{\Phi}^{-1} 1_T \\ & \quad - \frac{1}{T \widehat{\omega}_T} (1'_T \widehat{\Phi}^{-1} L^0 1_T) \frac{1}{N} \sum_{i=1}^N u'_i \widehat{\Phi}^{-1} 1_T \eta_i. \end{aligned}$$

We make the following three claims. The difference between the second and the fourth term, multiplied by  $(N/T)^{1/2}$ , is negligible, that is,

$$(34) \quad (N/T)^{1/2} \left[ \frac{1}{N} \sum_{i=1}^N (u'_i \widehat{\Phi}^{-1} L^0 1_T) \eta_i - \frac{1}{T \widehat{\omega}_T} (1'_T \widehat{\Phi}^{-1} L^0 1_T) \frac{1}{N} \sum_{i=1}^N u'_i \widehat{\Phi}^{-1} 1_T \eta_i \right] = o_p(1).$$

The third term, after centering (replacing  $u_i u'_i$  by  $u_i u'_i - \Phi^0$ ), is negligible, that is,

$$(35) \quad -(N/T)^{1/2} \left[ \frac{1}{T \widehat{\omega}_T} 1'_T \widehat{\Phi}^{-1} L^0 \frac{1}{N} \sum_{i=1}^N (u_i u'_i - \Phi^0) \widehat{\Phi}^{-1} 1_T \right] = o_p(1).$$

The centering amounts to adding the term  $\frac{1}{T \widehat{\omega}_T} 1'_T \widehat{\Phi}^{-1} L^0 \Phi^0 \widehat{\Phi}^{-1} 1_T$ . Now subtract this term from the last term of (33) and assume  $N/T^3 \rightarrow 0$ ; then

$$(36) \quad (N/T)^{1/2} \left[ \frac{1}{d_T} 1'_T \widehat{\Phi}^{-1} J_T S_N B^{0r} \widehat{\Phi}^{-1} 1_T - \frac{1}{T \widehat{\omega}_T} 1'_T \widehat{\Phi}^{-1} L^0 \Phi^0 \widehat{\Phi}^{-1} 1_T \right] = o_p(1).$$

The preceding three equations are equivalent to

$$(N/T)^{1/2} B_{NT}^* = \frac{1}{\sqrt{NT}} \sum_{i=1}^N u'_i \widehat{\Phi}^{-1} L^0 u_i + o_p(1).$$

To prove (32), it suffices to show that  $(NT)^{-1/2} \sum_{i=1}^N u'_i(\widehat{\Phi}^{-1} - \Phi^{0-1})L^0u_i = o_p(1)$ . The product  $u'_i(\widehat{\Phi}^{-1} - \Phi^{0-1})$  involves terms of the form  $u_{it}u_{kt}^2$ . However,  $L^0u_i$  involves terms of past values of  $u_{it}$ . Because the expected value  $E(u_{is}u_{it}u_{kt}^2) = 0$  for  $s < t$ , asymptotic bias due to estimating  $\Phi^0$  does not arise in this expression. The leading term of  $(NT)^{-1/2} \sum_{i=1}^N u'_i(\widehat{\Phi}^{-1} - \Phi^{0-1})L^0u_i$  is  $O_p(N^{-1/2}) = o_p(1)$ . This argument can be made precisely as in the analysis of the bias in  $\widehat{\pi} - \pi_N^0$ . We omit the details to avoid repetition. Thus, given (34)–(36), we obtain (32).

It remains to establish (34)–(36). The first result is the most difficult to prove, so we show (34) only. It is easy to show that (34) holds if  $u'_i\Phi^{0-1}$  is in place of  $u'_i\widehat{\Phi}^{-1}$ . So it is sufficient to show

$$(37) \quad \frac{1}{\sqrt{NT}} \sum_{i=1}^N [u'_i(\widehat{\Phi}^{-1} - \Phi^{0-1})L^01_T] \eta_i - \frac{(1'_T \widehat{\Phi}^{-1} L^0 1_T)}{T \widehat{\omega}_T} \frac{1}{\sqrt{NT}} \sum_{i=1}^N [u'_i(\widehat{\Phi}^{-1} - \Phi^{0-1})1_T] \eta_i = o_p(1).$$

We show that each of the two terms on the left has a bias of order  $(T/N)^{1/2}$  arising from the estimation of  $\Phi^0$ . However, the biases are of equal magnitude with opposite signs, so they cancel each other. We first examine the second term of (37). By the Toeplitz lemma (or by a direct argument),

$$(38) \quad \frac{1}{T \widehat{\omega}_T} (1'_T \widehat{\Phi}^{-1} L^0 1_T) = \frac{1}{1'_T \widehat{\Phi}^{-1} 1_T} (1'_T \widehat{\Phi}^{-1} L^0 1_T) = \frac{1}{1 - \rho^0} + O_p\left(\frac{1}{T}\right).$$

Next,

$$(39) \quad \begin{aligned} & \frac{1}{\sqrt{NT}} \sum_{i=1}^N u'_i(\widehat{\Phi}^{-1} - \Phi^{0-1})1_T \eta_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \frac{u_{it} \eta_i}{\widehat{\sigma}_t^2 \sigma_t^{02}} (\widehat{\sigma}_t^2 - \sigma_t^{02}) \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \frac{u_{it} \eta_i}{\sigma_t^{04}} \left[ \frac{1}{N} \sum_{k=1}^N (u_{kt}^2 - \sigma_t^{02}) \right] + o_p(1), \end{aligned}$$

where the second equality uses the representation of  $\widehat{\Phi}$ . The expected value of the above is

$$\left(\frac{T}{N}\right)^{1/2} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{E(u_{it}^3) \eta_i}{\sigma_t^{04}} = \left(\frac{T}{N}\right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^T \frac{\nu_t}{\sigma_t^{04}}\right) \bar{\eta},$$

where  $\nu_t = E(u_{it}^3)$ , not depending on  $i$ . Subtracting its expected value, (39) is negligible. The potential nonnegligible term (bias) is the product of the preceding expression and (38). That is,

$$(40) \quad \left(\frac{T}{N}\right)^{1/2} \frac{1}{1-\rho^0} \left(\frac{1}{T} \sum_{t=1}^T \frac{\nu_t}{\sigma_t^{04}}\right) \bar{\eta} + o_p(1),$$

where  $o_p(1)$  is in fact  $O_p(1/\sqrt{NT})$ . Next, consider the first term of (37),

$$\begin{aligned} & \frac{1}{\sqrt{NT}} \sum_{i=1}^N [u_i'(\widehat{\Phi}^{-1} - \Phi^{0-1})L^0 1_T] \eta_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \frac{u_{it} \eta_i}{\sigma_t^{04}} \left[ \frac{1}{N} \sum_{k=1}^N (u_{kt}^2 - \sigma_t^{02}) \right] \ell_t + o_p(1), \end{aligned}$$

where  $\ell_t$  is the  $t$ th element of  $L^0 1_T$ , so that  $\ell_1 = 0$  and  $\ell_t = 1 + \rho^0 + \dots + (\rho^0)^{t-2}$  for  $t \geq 2$ . We only need to consider its expected value (bias) since the deviation from the expected value is  $o_p(1)$ . The expected value of the above is

$$\left(\frac{T}{N}\right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^T \frac{\nu_t \ell_t}{\sigma_t^{04}}\right) \bar{\eta}.$$

Because  $\ell_t \rightarrow 1/(1-\rho^0)$  as  $t$  grows, by the Toeplitz lemma (or by a direct proof), we have

$$\frac{1}{T} \sum_{t=1}^T \frac{\nu_t \ell_t}{\sigma_t^{04}} = \frac{1}{1-\rho^0} \left(\frac{1}{T} \sum_{t=1}^T \frac{\nu_t}{\sigma_t^{04}}\right) + O\left(\frac{1}{T}\right).$$

Multiplying the preceding expression by  $(T/N)^{1/2} \bar{\eta}$  leads to an identical bias term as in (40). The two bias terms offset each other as they enter (37) with opposite signs. This verifies (37). Q.E.D.

**PROOF OF THEOREM 1:** For the limiting distribution of  $\widehat{\pi}$ , notice that  $\text{var}((NT)^{-1/2} \sum_{i=1}^N u_i' \Phi^{-1} 1_T \eta_i) = \frac{1}{T} (1_T' \Phi^{-1} 1_T) \pi_N \rightarrow \varpi \pi$  by (4) and the assumption  $\pi_N \rightarrow \pi$ . So the limit of the second term on the right hand side of (10) is  $N(0, 4\pi/\varpi)$ , which is asymptotically independent of  $\sqrt{NT}(\widehat{\rho} - \rho)$ . Together with Lemma 1(a), we have  $\sqrt{NT}(\widehat{\pi} - \pi_N - \frac{1}{N}b) \xrightarrow{d} N(0, \frac{4\pi^2}{\gamma(1-\rho)^2} + \frac{4\pi}{\varpi})$ . The joint asymptotic distribution of  $\widehat{\rho}$  and  $\widehat{\pi}$  follows from their representations in (7) and (10) and their marginal limiting distributions. Q.E.D.

The proof of Proposition 1 is given in the Supplemental Material.

## REFERENCES

- AHN, S., AND P. SCHMIDT (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5–27. [286]
- ALVAREZ, J., AND M. ARELLANO (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71, 1121–1159. [285,286,288,293]
- (2004): "Robust Likelihood Estimation of Dynamic Panel Data Models," Unpublished Manuscript, CEMFI. [286,297]
- AMEMIYA, Y., W. A. FULLER, AND S. G. PANTULA (1987): "The Asymptotic Distributions of Some Estimators for a Factor Analysis Model," *Journal of Multivariate Analysis*, 22, 51–64. [287, 288]
- ANDERSON, T. W., AND Y. AMEMIYA (1988): "The Asymptotic Normal Distribution of Estimators in Factor Analysis Under General Conditions," *The Annals of Statistics*, 16, 759–771. [288]
- ANDERSON, T. W., AND C. HSIAO (1981): "Estimation of Dynamic Models With Error Components," *Journal of the American Statistical Association*, 76, 598–606. [286]
- ANDERSON, T. W., AND H. RUBIN (1956): "Statistical Inference in Factor Analysis," in *Proceedings of Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5, ed. by J. Neyman. Oxford, U.K.: University of California Press, 111–150. [287]
- ARELLANO, M. (2003): *Panel Data Econometrics*. Oxford, U.K.: Oxford University Press. [300]
- ARELLANO, M., AND S. R. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297. [286]
- BAI, J. (2013): "Supplement to 'Fixed-Effects Dynamic Panel Models, a Factor Analytical Method'," *Econometrica Supplemental Material*, 81, [http://www.econometricsociety.org/ecta/Supmat/9409\\_Proofs.pdf](http://www.econometricsociety.org/ecta/Supmat/9409_Proofs.pdf). [286,294]
- BALTAGI, B. H. (2005): *Econometric Analysis of Panel Data*. Chichester: Wiley. [300]
- BHARGAVA, A., AND J. D. SARGAN (1983): "Estimating Dynamic Random Effects Models From Panel Data Covering Short Time Periods," *Econometrica*, 51, 1635–1659. [286]
- BICKEL, P. J., C. A. J. KLAASSEN, Y. RITOV, AND J. A. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer. [294]
- BLUNDELL, R., AND R. J. SMITH (1991): "Initial Conditions and Efficient Estimation in Dynamic Panel Data Models," *Annales d'Economie et de Statistique*, 20/21, 109–123. [286]
- BLUNDELL, R., L. PISTAFERRI, AND I. PRESTON (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98, 1887–1921. [300]
- CHAMBERLAIN, G. (1984): "Panel Data," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. Intriligator. Amsterdam: North-Holland. [296,298]
- CHAMBERLAIN, G., AND M. J. MOREIRA (2009): "Decision Theory Applied to a Linear Panel Data Model," *Econometrica*, 77, 107–133. [287,298]
- DAHM, P. F., AND W. A. FULLER (1986): "Generalized Least Squares Estimation of the Functional Multivariate Linear Errors-in-Variables Model," *Journal of Multivariate Analysis*, 19, 132–141. [296]
- GUIZO, L., L. PISTAFERRI, AND F. SCHIVADI (2005): "Insurance Within the Firm," *Journal of Political Economy*, 113, 1053–1087. [300]
- GUVENEN, F. (2009): "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, 12, 58–79. [300]
- HAHN, J. (1999): "How Informative Is the Initial Condition in the Dynamic Panel Model With Fixed Effects?" *Journal of Econometrics*, 93, 309–326. [286]
- HAHN, J., AND G. KUERSTEINER (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model With Fixed Effects When Both  $n$  and  $T$  Are Large," *Econometrica*, 70, 1639–1657. [294,297]
- HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. New York: Academic Press. [300]
- HSIAO, C. (2003): *Analysis of Panel Data*. New York: Cambridge University Press. [300]

- HSIAO, C., AND A. K. TAHMISCIOGLU (2008): "Estimation of Dynamic Panel Data Models With Both Individual and Time-Specific Effects," *Journal of Statistical Planning and Inference*, 138, 2698–2721. [294]
- KIVIET, J. (1995): "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, 68, 53–78. [285]
- LANCASTER, T. (2000): "The Incidental Parameter Problem Since 1948," *Journal of Econometrics*, 95, 391–413. [300]
- (2002): "Orthogonal Parameters and Panel Data," *Review of Economics Studies*, 69, 647–666. [300]
- LAWLEY, D. N., AND A. E. MAXWELL (1971): *Factor Analysis as a Statistical Method*. London: Butterworth. [287,293]
- MAGNUS, J. R., AND H. NEUDECKER (1999): *Matrix Differential Calculus With Applications in Statistics and Econometrics*. New York: Wiley. [295]
- MOFFITT, R., AND P. GOTTSCHALK (2002): "Trends in the Transitory Variance of Earnings in the United States," *The Economic Journal*, 112, C68–C73. [286]
- MOREIRA, M. J. (2009): "A Maximum Likelihood Method for the Incidental Parameter Problem," *The Annals of Statistics*, 37, 3660–3696. [286,287]
- MUNDLAK, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69–85. [298]
- NEYMAN, J., AND E. L. SCOTT (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32. [300]
- NICKELL, S. (1981): "Biases in Dynamic Models With Fixed Effects," *Econometrica*, 49, 1417–1426. [285]
- RAO, C. R., AND S. K. MITRA (1971): *Generalized Inverse of Matrices and Its Applications*. New York: Wiley. [296]
- VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. New York: Springer. [294]

*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.*