Econometrica, Vol. 77, No. 4 (July, 2009), 1229–1279

PANEL DATA MODELS WITH INTERACTIVE FIXED EFFECTS

BY JUSHAN BAI[1]

This paper considers large N and large T panel data models with unobservable multiple interactive effects, which are correlated with the regressors. In earnings studies, for example, workers' motivation, persistence, and diligence combined to influence the earnings in addition to the usual argument of innate ability. In macroeconomics, interactive effects represent unobservable common shocks and their heterogeneous impacts on cross sections. We consider identification, consistency, and the limiting distribution of the interactive-effects estimator. Under both large N and large T, the estimator is shown to be $\sqrt{NT}$ consistent, which is valid in the presence of correlations and heteroskedasticities of unknown form in both dimensions. We also derive the constrained estimator and its limiting distribution, imposing additivity coupled with interactive effects. The problem of testing additive versus interactive effects is also studied. In addition, we consider identification and estimation of models in the presence of a grand mean, time-invariant regressors, and common regressors. Given identification, the rate of convergence and limiting results continue to hold.

KEYWORDS: Additive effects, interactive effects, factor error structure, bias-corrected estimator, Hausman tests, time-invariant regressors, common regressors.

1. INTRODUCTION

WE CONSIDER THE FOLLOWING PANEL DATA MODEL with N cross-sectional units and T time periods

$$Y_{it} = X_{it}'\beta + u_{it} \tag{1}$$

and

$$u_{it} = \lambda_i' F_t + \varepsilon_{it} \qquad (i = 1, 2, \ldots, N, t = 1, 2, \ldots, T),$$

where X_{it} is a $p \times 1$ vector of observable regressors, β is a $p \times 1$ vector of unknown coefficients, u_{it} has a factor structure, λ_i $(r \times 1)$ is a vector of factor loadings, and F_t $(r \times 1)$ is a vector of common factors so that $\lambda_i' F_t = \lambda_{i1} F_{1t} + \cdots + \lambda_{ir} F_{rt}$; ε_{it} are idiosyncratic errors; λ_i, F_t, and ε_{it} are all unobserved. Our interest is centered on the inference for the slope coefficient β, although inference for λ_i and F_t will also be discussed.

[1]I am grateful to a co-editor and four anonymous referees for their constructive comments, which led to a much improved presentation. I am also grateful for comments and suggestions from seminar participants at the University of Pennsylvania, Rice, MIT/Harvard, Columbia Econometrics Colloquium, New York Econometrics Camp (Saratoga Springs), Syracuse, Malinvaud Seminar (Paris), European Central Bank/Center for Financial Studies Joint Workshop, Cambridge University, London School of Economics, Econometric Society European Summer Meetings (Vienna), Quantitative Finance and Econometrics at Stern, and the Federal Reserve Bank of Atlanta. This work is supported in part by NSF Grants SES-0551275 and SES-0424540.

 DOI: 10.3982/ECTA6135

The preceding set of equations constitutes the interactive-effects model in light of the interaction between λ_i and F_t. The usual fixed-effects model takes the form

$$Y_{it} = X_{it}'\beta + \alpha_i + \xi_t + \varepsilon_{it}, \tag{2}$$

where the individual effects α_i and the time effects ξ_t enter the model additively instead of interactively; accordingly, it will be called the additive-effects model for comparison and reference. It is noted that multiple interactive effects include additive effects as special cases. For $r = 2$, consider the special factor and factor loading such that, for all i and all t,

$$F_t = \begin{bmatrix} 1 \\ \xi_t \end{bmatrix} \quad \text{and} \quad \lambda_i = \begin{bmatrix} \alpha_i \\ 1 \end{bmatrix}.$$

Then

$$\lambda_i' F_t = \alpha_i + \xi_t.$$

The case of $r = 1$ has been studied by Holtz-Eakin, Newey, and Rosen (1988) and Ahn, Lee, and Schmidt (2001), among others.

Owing to potential correlations between the unobservable effects and the regressors, we treat λ_i and F_t as fixed-effects parameters to be estimated. This is a basic approach to controlling the unobserved heterogeneity; see Chamberlain (1984) and Arellano and Honore (2001). We allow the observable X_{it} to be written as

$$X_{it} = \tau_i + \theta_t + \sum_{k=1}^{r} a_k \lambda_{ik} + \sum_{k=1}^{r} b_k F_{kt} + \sum_{k=1}^{r} c_k \lambda_{ik} F_{kt} + \pi_i' G_t + \eta_{it}, \tag{3}$$

where a_k, b_k, and c_k are scalar constants (or vectors when X_{it} is a vector), and G_t is another set of common factors that do not enter the Y_{it} equation. So X_{it} can be correlated with λ_i alone or with F_t alone, or can be simultaneously correlated with λ_i and F_t. In fact, X_{it} can be a nonlinear function of λ_i and F_t. We make no assumption on whether F_t has a zero mean or whether F_t is independent over time: it can be a dynamic process without zero mean. The same is true for λ_i. We directly estimate λ_i and F_t, together with β subject to some identifying restrictions. We consider the least squares method, which is detailed in Section 3.

While additive effects can be removed by the within-group transformation (least squares dummy variables), the scheme fails to purge interactive effects. For example, consider $r = 1$, $Y_{it} = X_{it}'\beta + \lambda_i F_t + \varepsilon_{it}$. Then $Y_{it} - \bar{Y}_{i.} = (X_{it} - \bar{X}_{i.})'\beta + \lambda_i(F_t - \bar{F}) + \varepsilon_{it} - \bar{\varepsilon}_{i.}$, where $\bar{Y}_{i.}$, $\bar{X}_{i.}$, and $\bar{\varepsilon}_{i.}$ are averages over time. Because $F_t \not\equiv \bar{F}$, the within-group transformation with cross-section dummy variable is unable to remove the interactive effects. Similarly, the interactive effects cannot be removed with time dummy variable. Thus the within-group estimator is inconsistent since the unobservables are correlated with the regressors. However, the interactive effects can be eliminated by the quasi-differencing method, as in Holtz-Eakin, Newey, and Rosen (1988). Further details are provided in Section 3.3.

Recently, Pesaran (2006) proposed a new estimator that allows for multiple factor error structure under large N and large T. His method augments the model with additional regressors, which are the cross-sectional averages of the dependent and independent variables, in an attempt to control for F_t. His estimator requires a certain rank condition, which is not guaranteed to be met, that depends on data generating processes. Peseran showed $\sqrt{N}$ consistency irrespective of the rank condition, and a possible faster rate of convergence when the rank condition does hold. Coakley, Fuertes, and Smith (2002) proposed a two-step estimator, but this estimator was found to be inconsistent by Pesaran. The two-step estimator, while related, is not the least squares estimator. The latter is an iterated solution.

Ahn, Lee, and Schmidt (2001) considered the situation of fixed T and noted that the least squares method does not give a consistent estimator if serial correlation or heteroskedasticity is present in ε_{it}. Then they explored the consistent generalized method of moments (GMM) estimators and showed that a GMM method that incorporates moments of zero correlation and homoskedasticity is more efficient than least squares under fixed T. The fixed T framework was also studied earlier by Kiefer (1980) and Lee (1991).

Goldberger (1972) and Jöreskog and Goldberger (1975) are among the earlier advocates for factor models in econometrics, but they did not consider correlations between the factor errors and the regressors. Similar studies include MaCurdy (1982), who considered random effects type of generalized least squares (GLS) estimation for fixed T, and Phillips and Sul (2003), who considered SUR-GLS (seemingly unrelated regressions) estimation for fixed N. Panel unit root tests with factor errors were studied by Moon and Perron (2004). Kneip, Sickles, and Song (2005) assumed F_t is a smooth function of t and estimated F_t by smoothing spline. Given the spline basis, the estimation problem becomes that of ridge regression. The regressors X_{it} are assumed to be independent of the effects.

In this paper, we provide a large N and large T perspective on panel data models with interactive effects, permitting the regressor X_{it} to be correlated with either λ_i or F_t, or both. Compared with the fixed T analysis, the large T perspective has its own challenges. For example, an incidental parameter problem is now present in both dimensions. Consequently, a different argument is called for. On the other hand, the large T setup also presents new opportunities. We show that if T is large, comparable with N, then the least squares estimator for β is $\sqrt{NT}$ consistent, despite serial or cross-sectional correlations and heteroskedasticities of unknown form in ε_{it}. This presents a contrast to the fixed T framework, in which serial correlation implies inconsistency. Earlier fixed T studies assume independent and identically distributed (i.i.d.)

X_{it} over i, disallowing X_{it} to contain common factors, but permitting X_{it} to be correlated with λ_i. Earlier studies also assume ε_{it} are i.i.d. over i. We allow ε_{it} to be weakly correlated across i and over t, thus, u_{it} has the approximate factor structure of Chamberlain and Rothschild (1983). Additionally, heteroskedasticity is allowed in both dimensions.

Controlling fixed effects by directly estimating them, while often an effective approach, is not without difficulty—known as the incidental parameter problem, which manifests itself in bias and inconsistency at least under fixed T, as documented by Neyman and Scott (1948), Chamberlain (1980), and Nickell (1981). Even for large T, asymptotic bias can persist in dynamic or nonlinear panel data models with fixed effects.[2] We show that asymptotic bias arises under interactive effects, leading to nonzero centered limiting distributions.

We also show that bias-corrected estimators can be constructed in a way similar to Hahn and Kuersteiner (2002) and Hahn and Newey (2004), who argued that bias-corrected estimators may have desirable properties relative to instrumental variable estimators.

Because additive effects are special cases of interactive effects, the interactive-effects estimator is consistent when the effects are, in fact, additive, but the estimator is less efficient than the one with additivity imposed. In this paper, we derive the constrained estimator together with its limiting distribution when additive and interactive effects are jointly present. We also consider the problem of testing additive effects versus interactive effects.

In Section 2, we explain why incorporating interactive effects can be a useful modelling paradigm. Section 3 outlines the estimation method and Section 4 discusses the underlying assumptions that lead to consistent estimator. Section 5 derives the asymptotic representation and the asymptotic distribution of the estimator. Section 6 provides an interpretation of the estimator as a generalized within-group estimator. Section 7 derives the bias-corrected estimators. Section 8 considers estimators with additivity restrictions and their limiting distributions. Section 9 studies Hausman tests for additive effects versus interactive effects. Section 10 is devoted to time-invariant regressors and regressors that are common to each cross-sectional unit. Monte Carlo simulations are given in Section 11. All proofs are provided either in the Appendix or in the Supplemental Material (Bai (2009)).

2. SOME EXAMPLES

Macroeconometrics

Here Y_{it} is the output (or growth rate) for country i in period t, X_{it} is the input such as labor and capital, F_t represents common shocks (e.g., technological shocks and financial crises), λ_i represents the heterogeneous impact of

[2]See Nickell (1981), Anderson and Hsiao (1982), Kiviet (1995), Hsiao (2003, pp. 71–74), and Alvarez and Arellano (2003) for dynamic panel data models; see Hahn and Newey (2004) for nonlinear panel models.

common shocks on country i, and, finally, ε_{it} is the country-specific error term of output (or growth rate). In general, common shocks not only affect the output directly (through the total factor productivity or Solow resdidual), but also affect the amount of input in the production process (through investment decisions). When common shocks have homogeneous effects on the output, that is, $\lambda_i = \lambda$ for all i, the model collapses to the usual time effect by letting $\delta_t = \lambda' F_t$, where δ_t is a scalar. It is the heterogeneity that gives rise to a factor structure.

Recently, Giannone and Lenza (2005) provided an explanation for the Feldstein–Horioka (1980) puzzle, one of the six puzzles in international macroeconomics (Obstfeld and Rogoff (2000)). The puzzle refers to the excessively high correlation between domestic savings and domestic investments in open economies. In their model, Y_{it} is the investment and X_{it} is the savings for country i, F_t is the common shock that affects both investment and savings decisions. Giannone and Lenza found that the high correlation is a consequence of the strong assumption that shocks have homogeneous effects across countries (additive effects); it disappears when shocks are allowed to have heterogeneous impacts (interactive effects).

Microeconometrics

In earnings studies, Y_{it} represents the wage rate for individual i with age (or age cohort) t and X_{it} is a vector of observable characteristics, such as education, experience, gender, and race. Here λ_i represents a vector of unobservable characteristics or unmeasured skills, such as innate ability, perseverance, motivation, and industriousness, and F_t is a vector of prices for the unmeasured skills. The model assumes that the price vector for the unmeasured skills is time-varying. If $F_t = f$ for all t, the standard fixed-effects model is obtained by letting $\alpha_i = \lambda_i' f$. In this example, t is not necessarily the calendar time, but age or age cohort. Applications in this area were given by Cawley, Connelly, Heckman, and Vytlacil (1997) and Carneiro, Hansen, and Heckman (2003). As explained in a previous version, the model of Abowd, Kramarz, and Margolis (1999) can be extended to disentangle the worker and the firm effects, while incorporating interactive effects. Ahn, Lee, and Schmidt (2001) provided a theoretical motivation for a single factor model based on the work of Altug and Miller (1990) and Townsend (1994).

In the setup of Holtz-Eakin, Newey, and Rosen (1988), the slope coefficient β is also time-varying. Their model can be considered as a projection of Y_{it} on $\{X_{it}, \lambda_i\}$; see Chamberlain (1984). Pesaran (2006) allowed β to be heterogeneous over i such that $\beta_i = \beta + v_i$ with v_i being i.i.d.. In this regard, the constant slope coefficient is restrictive. To partially alleviate the restriction, it would be useful to allow additional individual and time effects as

$$Y_{it} = X_{it}' \beta + \alpha_i + \delta_t + \lambda_i' F_t + \varepsilon_{it}. \tag{4}$$

Model (4) will be considered in Section 8.

Finance

Here Y_{it} is the excess return of asset i in period t; X_{it} is a vector of observable factors such as dividend yields, dividend payout ratio, and consumption gap as in Lettau and Ludvigson (2001) or book and size factors as in Fama and French (1993); F_t is a vector of unobservable factor returns; λ_i is the factor loading; ε_{it} is the idiosyncratic return. The arbitrage pricing theory of Ross (1976) is built upon a factor model for asset returns. Campbell, Lo, and MacKinlay (1997) provided many applications of factor models in finance.

Cross-Section Correlation

Interactive-effects models provide a tractable way to model cross-section correlations. In the error term $u_{it} = \lambda_i' F_t + \varepsilon_{it}$, each cross section shares the same F_t, causing cross-correlation. If $\lambda_i = 1$ for all i, and ε_{it} are i.i.d. over i and t, an equal correlation model is obtained. In a recent paper, Andrews (2005) showed that cross-section correlation induced by common shocks can be problematic for inference. Andrews' analysis is confined within the framework of a single cross-section unit. In the panel data context, as shown here, consistency and proper inference can be obtained.

3. ESTIMATION

3.1. *Issues of Identification*

Even in the absence of regressors X_{it}, the lack of identification for factor models is well known; see Anderson and Rubin (1956) and Lawley and Maxell (1971). The current setting differs from classical factor identification in two aspects. First, both factor loadings and factors are treated as parameters, as opposed to factor loadings only. Second, the number of individuals N is assumed to grow without bound instead of being fixed, and it can be much larger than the number of observations T.

Write the model as

$$Y_i = X_i\beta + F\lambda_i + \varepsilon_i,$$

where

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix}, \quad X_i = \begin{bmatrix} X'_{i1} \\ X'_{i2} \\ \vdots \\ X'_{iT} \end{bmatrix}, \quad F = \begin{bmatrix} F'_1 \\ F'_2 \\ \vdots \\ F'_T \end{bmatrix}, \quad \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}.$$

Similarly, define $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)'$, an $N \times r$ matrix. In matrix notation,

$$Y = X\beta + F\Lambda' + \varepsilon, \tag{5}$$

where $Y = (Y_1, \dots, Y_N)$ is $T \times N$ and X is a three-dimensional matrix with p sheets ($T \times N \times p$), the ℓth sheet of which is associated with the ℓth element of β ($\ell = 1, 2, \dots, p$). The product $X\beta$ is $T \times N$ and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)$ is $T \times N$.

In view of $F\Lambda' = FAA^{-1}\Lambda'$ for an arbitrary $r \times r$ invertible A, identification is not possible without restrictions. Because an arbitrary $r \times r$ invertible matrix has r^2 free elements, the number of restrictions needed is r^2. The normalization[3]

$$F'F/T = I_r \tag{6}$$

yields $r(r+1)/2$ restrictions. This is a commonly used normalization; see, for example, Connor and Korajczyk (1986), Stock and Watson (2002), and Bai and Ng (2002). Additional $r(r-1)/2$ restrictions can be obtained by requiring

$$\Lambda'\Lambda = \text{diagonal}. \tag{7}$$

These two sets of restrictions uniquely determine Λ and F, given the product $F\Lambda'$.[4] The least squares estimators for F and Λ derived below satisfy these restrictions.

With either fixed N or fixed T, factor analysis would require additional restrictions. For example, the covariance matrix of ε_i is diagonal or the covariance matrix depends on a small number of parameters via parameterization. Under large N and large T, the cross-sectional covariance matrix of ε_{it} or the time series covariance matrix can be of an unknown form. In particular, none of the elements is required to be zero. However, the correlation—either cross sectional or serial—must be weak, which we assume to hold. This is known as the approximate factor model of Chamberlain and Rothschild (1983).

Sufficient variation in X_{it} is also needed. The usual identification condition for β is that the matrix $\frac{1}{NT}\sum_{i=1}^{N} X_i' M_F X_i$ is of full rank, where $M_F = I_T - F(F'F)^{-1}F'$. Because F is not observable and is estimated, a stronger condition is required. See Section 4 for details.

3.2. *Estimation*

The least squares objective function is defined as

$$\mathrm{SSR}(\beta, F, \Lambda) = \sum_{i=1}^{N} (Y_i - X_i\beta - F\lambda_i)'(Y_i - X_i\beta - F\lambda_i) \tag{8}$$

[3]The normalization still leaves rotation indeterminacy. For example, let G be an $r \times r$ orthogonal matrix, and let $F^* = FG$ and $\Lambda^* = \Lambda G$. Then $F\Lambda' = F^*\Lambda^{*\prime}$ and $F^{*\prime}F^*/T = F'F/T = I$. To remove this indeterminacy, we fix G to make $\Lambda^{*\prime}\Lambda^* = G'\Lambda'\Lambda G$ a diagonal matrix. This is the reason for restriction (7).

[4]Uniqueness is up to a columnwise sign change. For example, $-F$ and $-\Lambda$ also satisfy the restrictions.

subject to the constraint $F'F/T = I_r$ and $\Lambda'\Lambda$ being diagonal. Define the projection matrix

$$M_F = I_T - F(F'F)^{-1}F' = I_T - FF'/T.$$

The least squares estimator for β for each given F is simply

$$\hat{\beta}(F) = \left(\sum_{i=1}^{N} X_i' M_F X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_F Y_i.$$

Given β, the variable $W_i = Y_i - X_i\beta$ has a pure factor structure such that

$$W_i = F\lambda_i + \varepsilon_i. \tag{9}$$

Define $W = (W_1, W_2, \ldots, W_N)$, a $T \times N$ matrix. The least squares objective function is

$$\operatorname{tr}[(W - F\Lambda')(W - F\Lambda')'].$$

From the analysis of pure factor models estimated by the method of least squares (i.e., principal components; see Connor and Korajzcyk (1986) and Stock and Watson (2002)), by concentrating out $\Lambda = W'F(F'F)^{-1} = W'F/T$, the objective function becomes

$$\operatorname{tr}(W'M_F W) = \operatorname{tr}(W'W) - \operatorname{tr}(F'WW'F)/T. \tag{10}$$

Therefore, minimizing with respect to F is equivalent to maximizing $\operatorname{tr}[F'(W \times W')F]$. The estimator for F (see Anderson (1984)), is equal to the first r eigenvectors (multiplied by $\sqrt{T}$ due to the restriction $F'F/T = I$) associated with the first r largest eigenvalues of the matrix

$$WW' = \sum_{i=1}^{N} W_i W_i' = \sum_{i=1}^{N} (Y_i - X_i\beta)(Y_i - X_i\beta)'.$$

Therefore, given F, we can estimate β, and given β, we can estimate F. The final least squares estimator $(\hat{\beta}, \hat{F})$ is the solution of the set of nonlinear equations

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}} Y_i \tag{11}$$

and

$$\left[\frac{1}{NT} \sum_{i=1}^{N} (Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'\right]\hat{F} = \hat{F} V_{NT}, \tag{12}$$

where V_{NT} is a diagonal matrix that consists of the r largest eigenvalues of the above matrix[5] in the brackets, arranged in decreasing order. The solution $(\hat{\beta}, \hat{F})$ can be simply obtained by iteration. Finally, from $\Lambda = W'F/T$, $\hat{\Lambda}$ is expressed as a function of $(\hat{\beta}, \hat{F})$ such that

$$\hat{\Lambda}' = (\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_N) = T^{-1}[\hat{F}'(Y_1 - X_1\hat{\beta}), \ldots, \hat{F}'(Y_N - X_N\hat{\beta})].$$

We may also write

$$\hat{\Lambda}' = T^{-1}\hat{F}'(Y - X\hat{\beta}),$$

where Y is $T \times N$ and X is $T \times N \times p$, a three-dimensional matrix.

The triplet $(\hat{\beta}, \hat{F}, \hat{\Lambda})$ jointly minimizes the objective function (8). The pair $(\hat{\beta}, \hat{F})$ jointly minimizes the concentrated objective function (10), which, when substituting $Y_i - X_i\beta$ for W_i, is equal to

$$\text{(13)} \qquad \operatorname{tr}(W'M_FW) = \sum_{i=1}^{N} W_i'M_FW_i = \sum_{i=1}^{N}(Y_i - X_i\beta)'M_F(Y_i - X_i\beta).$$

This is also the objective function considered by Ahn, Lee, and Schmidt (2001), although a different normalization is used. They as well as Kiefer (1980) discussed an iteration procedure for estimation. Interestingly, convergence to a local optimum for such an iterated estimator was proved by Sargan (1964). Here we suggest a more robust iteration scheme (having a much better convergence property from Monte Carlo evidence) than the one implied by (11) and (12). Given F and Λ, we compute

$$\hat{\beta}(F, \Lambda) = \left(\sum_{i=1}^{N} X_i'X_i\right)^{-1} \sum_{i=1}^{N} X_i'(Y_i - F\lambda_i),$$

and given β, we compute F and Λ from the pure factor model $W_i = F\lambda_i + e_i$ with $W_i = Y_i - X_i\beta$. This iteration scheme only requires a single matrix inverse $(\sum_{i=1}^{N} X_i'X_i)^{-1}$, with no need to update during iteration. Our simulation results are based on this scheme.

REMARK 1: The common factor F is obtained by the principal components method from the matrix WW'/N. Under large N but a fixed T, $WW'/N \to \Sigma_W = F\Sigma_\Lambda F' + \Omega$, where Σ_Λ is the limit of $\Lambda'\Lambda/N$—an $r \times r$ matrix—and Ω is a $T \times T$ matrix of the covariance matrix of ε_i; see (9). Unless Ω is a scalar multiple of I_T (identity matrix), that is, no serial correlation and heteroskedasticity, the first r eigenvectors of Σ_W are not a rotation of F, implying that the

[5] We divide this matrix by NT so that V_{NT} will have a proper limit. The scaling does not affect $\hat{F}$.

first r eigenvectors of WW'/N are not consistent for F (a rotation of F to be more precise). This leads to inconsistent estimation of the product $F\Lambda'$, and, thus, inconsistent estimation of β. However, under large T, Ω does not have to be a scalar multiple of an identity matrix and the principal components estimator for F is consistent. This is the essence of the approximate factor model of Chamberlain and Rothschild (1983).

REMARK 2: Instead of estimating F from (9) by the method of principal components, one can directly use factor analysis. Factor analysis such as the maximum likelihood method allows Ω to be heteroskedastic and to have nonzero off-diagonal elements. The off-diagonal elements (due to serial correlation) must be parametrized to avoid too many free parameters. Serial correlation can also be removed by adding lagged dependent variables as regressors, leaving a diagonal Ω. In contrast, the principal components method is designed for large N and large T. In this setting, there is no need to assume a parametric form for serial correlations. The principal components method is a quick and effective approach to extracting common factors. Note that lagged dependent variables lead to bias, just as serial correlation does; see Section 7. Small T models can also be estimated by the quasi-differencing approach (Section 3.3). This latter approach also parametrizes serial correlations by including lagged dependent variables as regressors so that ε_{it} has no serial correlation. The reason to parametrize serial correlation under quasi-differencing is that, with unrestricted serial correlation, lagged dependent variables will not be valid instruments.

REMARK 3: The estimation procedure can be modified to handle unbalanced data. The procedure is elaborated in the Supplemental material; the details are omitted here.

3.3. *Alternative Estimation Methods*

While the analysis is focused on the method of least squares, we discuss several alternative estimation strategies.

METHOD 1: The quasi-differencing method in Holtz-Eakin, Newey, and Rosen (1988) can be adapted for multiple factors. Consider the case of two factors:

$$y_{it} = x_{it}\beta + \lambda_{i1}f_{t1} + \lambda_{i2}f_{t2} + \varepsilon_{it}.$$

Multiplying the equation $y_{i,t-1}$ by $\phi_t = f_{t1}/f_{t-1,1}$ and then subtracting it from the equation y_{it}, we obtain

$$y_{it} = \phi_t y_{i,t-1} + x_{it}'\beta - x_{i,t-1}'\beta\phi_t + \lambda_{i2}\delta_t + \varepsilon_{it}^*,$$

where $\delta_t = f_{t2} - f_{t-1,2}\phi_t$ and $\varepsilon_{it}^* = \varepsilon_{it} - \phi_t\varepsilon_{i,t-1}$. The resulting model has a single factor. If we apply the quasi-differencing method one more times to the resulting equation, then the factor error will be eliminated. This approach was used by Ahn, Lee, and Schmidt (2006). The GMM method as in Holtz-Eakin, Newey, and Rosen (1988) and Ahn, Lee, and Schmidt (2001) can be used to estimate the model parameters consistently under some identification conditions. For the case of $r = 1$, GMM was also discussed by Arellano (2003) and Baltagi (2005). While not always necessary, there may be a need to recover the original model parameters; see Holtz-Eakin, Newey, and Rosen (1988) for details.

This estimator is consistent under fixed T despite serial correlation and heteroskedasticity in ε_{it}. In contrast, the least squares estimator will be inconsistent under this setting. On the other hand, as T increases, due to the many-parameter and many-instrument problem, the GMM method tends to yield bias, a known issue from the existing literature (e.g., Newey and Smith (2004)). The least squares estimator is consistent under large N and large T with unknown form of correlation and heteroskedasticity, and the bias is decreasing in N and T. Furthermore, the least squares method directly estimates all parameters, including the factor processes F_t and the factor loadings λ_i, so there is no need to recover the original parameters. In many applications, the estimated factor processes are used as inputs for further analysis, for example, the diffusion index forecasting of Stock and Watson (2002) and factor-augmented vector autoregression of Bernanke, Boivin, and Eliasz (2005). The estimated loadings are useful objects in finance; see Connor and Korajczyk (1986). Recovering those original parameters becomes more involved under multiple factors with the quasi-differencing approach. The least squares method is simple and effective in handling multiple factors. Furthermore, computation of the least squares method under large N and large T is quite fast.

In summary, with small T and with potential serial correlation and time series heteroskedasticity, the quasi-differencing method is recommended in view of its consistency properties. With large T, the least squares method is a viable alternative. Remark 2 suggests another alternative under small T.

METHOD 2: We can extend the argument of Mundlak (1978) and Chamberlain (1984) to models with interactive effects. When λ_i is correlated with the regressors, it can be projected onto the regressors such that $\lambda_i = A\bar{X}_{i\cdot} + \eta_i$, where $\bar{X}_{i\cdot}$ is the time average of X_{it} and A is $r \times p$, so that model (1) can be rewritten as

$$Y_{it} = X_{it}'\beta + \bar{X}_{i\cdot}'\delta_t + \eta_i'F_t + \varepsilon_{it},$$

where $\delta_t = A'F_t$. The above model still has a factor error structure. However, when F_t is assumed to be uncorrelated with the regressors, the aggregated error $\eta_i'F_t + \varepsilon_{it}$ is now uncorrelated with the regressors, so we can use a random-effects GLS to estimate $(\beta, \delta_1, \ldots, \delta_T)$. Similarly, when F_t is correlated with the regressors, but λ_i is not, one can project F_t onto the cross-sectional averages such that $F_t = B\bar{X}_{.t} + \xi_t$ to obtain

$$Y_{it} = X_{it}'\beta + \bar{X}_{.t}'\rho_i + \lambda_i'\xi_t + \varepsilon_{it}$$

with $\rho_i = B\lambda_i$. Again, a random-effects GLS can be used. When both λ_i and F_t are correlated with regressors, we apply both projections and augment the model with cross-products of $\bar{X}_{i.}$ and $\bar{X}_{.t}$, in addition to $\bar{X}_{i.}$ and $\bar{X}_{.t}$ so that, with $\rho_i = B'\eta_i$ and $\delta_t = A'\xi_t$,

$$Y_{it} = X_{it}'\beta + \bar{X}_{i.}'\delta_t + \bar{X}_{.t}'\rho_i + \bar{X}_{i.}'C\bar{X}_{.t} + \eta_i'\xi_t + \varepsilon_{it}, \tag{14}$$

where C is a matrix. The above can be estimated by the random-effects GLS.

METHOD 3: The method of Pesaran (2006) augments the model with regressors $(\bar{Y}_{.t}, \bar{X}_{.t})$ under the assumption of F_t being correlated with regressors, where $\bar{Y}_{.t}$ and $\bar{X}_{.t}$ attempt to estimate F_t, similar to the projection argument of Mundlak. But in the Mundlak argument, the projection residual ξ_t is assumed to have a fixed variance. In contrast, the variance of ξ_t is assumed to converge to zero as $N \to \infty$ in Pesaran (2006), who assumed X_{it} is of the form $X_{it} = B_iF_t + e_{it}$ so that $\bar{X}_{.t} = BF_t + \xi_t$ with $B = N^{-1}\sum_{i=1}^N B_i$ and $\xi_t = N^{-1}\sum_{i=1}^N e_{it}$. The variance of ξ_t is of $O(N^{-1})$. Thus the factor error $\lambda_i'\xi_t$ is negligible under large N. He established $\sqrt{N}$ consistency and possible $\sqrt{NT}$ consistency for some special cases. It appears that when λ_i is correlated with the regressors, additional regressors $\bar{Y}_{i.}$ and $\bar{X}_{i.}$ should also be added to achieve consistency.

4. ASSUMPTIONS

In this section, we state assumptions needed for consistent estimation and explain the meaning of each assumption prior to or after its introduction. Throughout, for a vector or matrix A, its norm is defined as $\|A\| = (\mathrm{tr}(A'A))^{1/2}$.

The $p \times p$ matrix

$$D(F) = \frac{1}{NT}\sum_{i=1}^N X_i'M_FX_i - \frac{1}{T}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N X_i'M_FX_ka_{ik}\right],$$

where $a_{ik} = \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$, plays an important role in this paper. Note that $a_{ik} = a_{ki}$ since it is a scalar. The identifying condition for β is that $D(F)$ is positive definite. If F were observable, the identification condition for β would be that the first term of $D(F)$ on the right-hand side is positive definite. The presence of the second term is because of unobservable F and Λ. The reason for this particular form is the nonlinearity of the interactive effects.

Define a $T \times p$ vector

$$Z_i = M_F X_i - \frac{1}{N}\sum_{k=1}^{N} M_F X_k a_{ik},$$

so that Z_i is equal to the deviation of $M_F X_i$ from its mean, but here the mean is a weighted average. Write $Z_i = (Z_{i1}, Z_{i2}, \dots, Z_{iT})'$. Then

$$D(F) = \frac{1}{NT}\sum_{i=1}^{N} Z_i' Z_i = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} Z_{it} Z_{it}'\right).$$

The first equality follows from $a_{ik} = a_{ki}$ and $N^{-1}\sum_{i=1}^{N} a_{ik} a_{ij} = a_{kj}$, and the second equality is by definition. Thus $D(F)$ is at least semipositive definite. Since each $Z_{it} Z_{it}'$ is a rank 1 semidefinite matrix, summation of NT such semidefinite matrices should lead to a positive definite matrix, given enough variations in Z_{it} over i and t. Our first condition assumes $D(F)$ is positive definite in the limit. Suppose that as $N, T \to \infty$, $D(F) \xrightarrow{p} D > 0$. If ε_{it} are i.i.d. $(0, \sigma^2)$, then the limiting distribution of $\hat{\beta}$ can be shown to be

$$\sqrt{NT}(\hat{\beta} - \beta) \to N(0, \sigma^2 D^{-1}).$$

This shows the need for $D(F)$ to be positive definite.

Since F is to be estimated, the identification condition for β is assumed as follows:

ASSUMPTION A—$E\|X_{it}\|^4 \le M$: *Let* $\mathcal{F} = \{F : F'F/T = I\}$. *We assume*

$$\inf_{F \in \mathcal{F}} D(F) > 0.$$

The matrix F in this assumption is $T \times r$, either deterministic or random. This assumption rules out time-invariant regressors and common regressors. Suppose $X_i = x_i \iota_T$, where x_i is a scalar and $\iota_T = (1, 1, \dots, 1)'$. For $\iota_T \in \mathcal{F}$ and $D(\iota_T) = 0$, it follows that $\inf_F D(F) = 0$. A common regressor does not vary with i. Suppose all regressors are common such that $X_i = W$. For $F = W(W'W)^{-1/2} \in \mathcal{F}$, $D(F) = 0$. The analysis of time-invariant regressors and common regressors is postponed to Section 10, where we show that it is sufficient to have $D(F) > 0$, when evaluated at the true factor process F. For now, it is not difficult to show that if X_{it} is characterized by (3), where η_{it} have sufficient variations such as i.i.d. with positive variance, then Assumption A is satisfied.

ASSUMPTION B:

(i) $E\|F_t\|^4 \le M$ *and* $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F > 0$ *for some* $r \times r$ *matrix* Σ_F, *as* $T \to \infty$.

(ii) $E\|\lambda_i\|^4 \le M$ *and* $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\Lambda > 0$ *for some* $r \times r$ *matrix* Σ_Λ, *as* $N \to \infty$.

This assumption implies the existence of r factors. Note that whether F_t or λ_t has zero mean is of no issue since they are treated as parameters to be estimated; for example, it can be a linear trend ($F_t = t/T$). But if it is known that F_t is a linear trend, imposing this fact gives more efficient estimation. Moreover, F_t itself can be a dynamic process such that $F_t = \sum_{i=1}^{\infty} C_i e_{t-i}$, where e_t are i.i.d. zero mean process. Similarly, λ_i can be cross-sectionally correlated.

ASSUMPTION C—Serial and Cross-Sectional Weak Dependence and Heteroskedasticity:

(i) $E(\varepsilon_{it}) = 0$ *and* $E|\varepsilon_{it}|^8 \le M$.

(ii) $E(\varepsilon_{it}\varepsilon_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \le \bar{\sigma}_{ij}$ *for all* (t, s) *and* $|\sigma_{ij,ts}| \le \tau_{ts}$ *for all* (i, j) *such that*

$$\frac{1}{N}\sum_{i,j=1}^{N} \bar{\sigma}_{ij} \le M, \quad \frac{1}{T}\sum_{t,s=1}^{T} \tau_{ts} \le M, \quad \frac{1}{NT}\sum_{i,j,t,s=1} |\sigma_{ij,ts}| \le M.$$

The largest eigenvalue of $\Omega_i = E(\varepsilon_i \varepsilon_i')$ $(T \times T)$ *is bounded uniformly in i and* T.

(iii) *For every* (t, s), $E|N^{-1/2}\sum_{i=1}^{N}[\varepsilon_{is}\varepsilon_{it} - E(\varepsilon_{is}\varepsilon_{it})]|^4 \le M$.

(iv) *Moreover*

$$T^{-2}N^{-1}\sum_{t,s,u,v}\sum_{i,j} |\operatorname{cov}(\varepsilon_{it}\varepsilon_{is}, \varepsilon_{ju}\varepsilon_{jv})| \le M,$$

$$T^{-1}N^{-2}\sum_{t,s}\sum_{i,j,k,\ell} |\operatorname{cov}(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})| \le M.$$

Assumption C is about weak serial and cross-sectional correlation. Heteroskedasticity is allowed, but ε_{it} is assumed to have a uniformly bounded eighth moment. The first three conditions are relatively easy to understand and are assumed in Bai (2003). We explain the meaning of C(iv). Let $\eta_i = (T^{-1/2}\sum_{t=1}^{T} \varepsilon_{it})^2 - E(T^{-1/2}\sum_{t=1}^{T} \varepsilon_{it})^2$. Then $E(\eta_i) = 0$ and $E(\eta_i^2)$ is bounded. The expected value $(N^{-1/2}\sum_{i=1}^{N} \eta_i)^2$ is equal to $T^{-2}N^{-1}\sum_{t,s,u,v}\sum_{i,j} \operatorname{cov}(\varepsilon_{it}\varepsilon_{is}, \varepsilon_{ju}\varepsilon_{jv})$, that is, the left-hand side of the first inequality without the absolute sign. So the first part of C(iv) is slightly stronger than the assumption that the second moment of $N^{-1/2}\sum_{i=1}^{N} \eta_i$ is bounded. The meaning of the second part is similar. It can be easily shown that if ε_{it} are independent over i and t with $E\varepsilon_{it}^4 \le M$ for all i and t, then C(iv) is true. If ε_{it} are i.i.d. with zero mean and $E\varepsilon_{it}^8 \le M$, then Assumption C holds.

ASSUMPTION D: *ε_{it} is independent of X_{js}, λ_j, and F_s for all i, t, j, and s.*

This assumption rules out dynamic panel data models and is given for the purpose of simplifying the proofs. The procedure works well even with lagged dependent variables; see Table V in the Supplemental Material. We do allow X_{it}, F_t, and ε_{it} to be dynamic processes. If lagged dependent variables are included in X_{it}, then ε_{it} cannot be serially correlated. Also note that X_{it}, λ_i, and ε_{it} are allowed to be cross-sectionally correlated.

5. LIMITING THEORY

We use (β^0, F^0) to denote the true parameters, and we still use λ_i without the superscript 0 as it is not directly estimated and thus not necessary. Here F^0 denotes the true data generating process for F that satisfies Assumption B. This F^0 in general has economic interpretations (e.g., supply shocks and demand shocks). The estimator $\hat{F}$ below estimates a rotation of F^0.[6] Define $S_{NT}(\beta, F)$ as the concentrated objective function in (13) divided by NT together with centering, that is,

$$S_{NT}(\beta, F) = \frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\beta)'M_F(Y_i - X_i\beta) - \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i' M_{F^0}\varepsilon_i.$$

The second term does not depend on β and F, and is for the purpose of centering, where $M_F = I - P_F = I - FF'/T$ with $F'F/T = I$. We estimate β^0 and F^0 by

$$(\hat{\beta}, \hat{F}) = \arg\min_{\beta, F} S_{NT}(\beta, F).$$

As explained in the previous section, $(\hat{\beta}, \hat{F})$ satisfies

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}} Y_i,$$

$$\left[\frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'\right]\hat{F} = \hat{F}V_{NT},$$

where $\hat{F}$ is the the matrix that consists of the first r eigenvectors (multiplied by $\sqrt{T}$) of the matrix $\frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'$ and where V_{NT} is a diagonal

[6]If (6) and (7) hold for the data generating processes (i.e., $F^{0\prime}F^0/T = I$ and $\Lambda^{0\prime}\Lambda^0$ is diagonal) rather than being viewed as estimation restrictions, then $\hat{F}$ estimates F^0 itself instead of a rotation of F^0.

matrix that consists of the first r largest eigenvalues of this matrix. Denote $P_A = A(A'A)^{-1}A'$ for a matrix A.

PROPOSITION 1—Consistency: *Under Assumptions A–D, as $N, T \to \infty$, the following statements hold*:

(i) *The estimator $\hat{\beta}$ is consistent such that $\hat{\beta} - \beta^0 \xrightarrow{p} 0$.*
(ii) *The matrix $F^{0\prime}\hat{F}/T$ is invertible and $\|P_{\hat{F}} - P_{F^0}\| \xrightarrow{p} 0$.*

The usual argument of consistency for extreme estimators would involve showing $S_{NT}(\beta, F) \xrightarrow{p} S(\beta, F)$ uniformly on some bounded set of β and F, and then showing that $S(\beta, F)$ has a unique minimum at β^0 and F^0; see Newey and McFadden (1994). This argument needs to be modified to take into account the growing dimension of F. As F is a $T \times r$ vector, the limit S would involve an infinite number of parameters as N, T going to infinity so the limit as a function of F is not well defined. Furthermore, the concept of bounded F is not well defined either. In this paper, we only require $F'F/T = I$. The modification is similar to Bai (1994), where the parameter space (the break point) increases with the sample size. We show there exists a function $\tilde{S}_{NT}(\beta, F)$, depending on (N, T) and generally still a random function, such that $\tilde{S}_{NT}(\beta, F)$ has a unique minimum at β^0 and F^0. In addition, we show the difference is uniformly small,

$$S_{NT}(\beta, F) - \tilde{S}_{NT}(\beta, F) = o_p(1),$$

where $o_p(1)$ is uniform. This implies the consistency of $\hat{\beta}$ for β^0. However, we cannot claim the consistency of $\hat{F}$ for F^0 (or a rotation of F^0) owing to its growing dimension. Part (ii) claims that the spaces spanned by $\hat{F}$ and F^0 are asymptotically the same. Alternative consistency concepts, including componentwise consistency or average norm consistency, are provided in the Appendix, as these consistency concepts are also needed.

Given consistency, we can further establish the rate of convergence.

THEOREM 1—Rate of Convergence: *Assume Assumptions A–D hold. For comparable N and T such that $T/N \to \rho > 0$, then $\sqrt{NT}(\hat{\beta} - \beta^0) = O_p(1)$.*

The theorem allows cross-section and serial correlations, as well as heteroskedasticities in both dimensions. This is important for applications in macroeconomics, say cross-country studies, or in finance, where the factors may not fully capture the cross-section correlations, and therefore the approximate factor model of Chamberlain and Rothschild (1983) is relevant. For microeconomic data, cross-section heteroskedasticity is likely to be present.

Although the estimator is $\sqrt{NT}$ consistent, the underlying limiting distribution will not be centered at zero; asymptotic biases exist. The next two theorems provide the limiting behavior of the estimator. The first theorem deals with some special cases in which asymptotic bias is absent. This is obtained by requiring stronger assumptions: the absence of either cross-correlation or serial correlation and heteroskedasticity. The second theorem deals with the most general case that allows for correlation and heteroskedasticity in both dimensions.

Introduce

$$Z_i = M_{F^0} X_i - \frac{1}{N} \sum_{k=1}^{N} a_{ik} M_{F^0} X_k.$$

Then in the absence of serial correlation and heteroskedasticity in one of the dimensions, and given an appropriate relative rate for T and N, it is shown in the Appendix that the estimator has the representation

$$\sqrt{NT}(\hat{\beta} - \beta^0) = \left(\frac{1}{NT} \sum_{i=1}^{N} Z_i' Z_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i' \varepsilon_i + o_p(1). \tag{15}$$

If correlation and heteroskedasticity are present in both dimensions, there will be an $O_p(1)$ bias term in the above representation; see (21) in Section 7. In all cases, we need the central limit theorem for $(NT)^{-1/2} \sum_{i=1}^{N} Z_i' \varepsilon_i = (NT)^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} Z_{it} \varepsilon_{it}$. Assuming correlation and heteroskedasticity in both dimensions, its variance is given by

$$\operatorname{var}\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i' \varepsilon_i \right) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \sigma_{ij,ts} E(Z_{it} Z_{js}'),$$

where $\sigma_{ij,ts} = E(\varepsilon_{it} \varepsilon_{js})$. This variance is $O(1)$ because $\frac{1}{NT} \sum_{i,j,t,s} |\sigma_{ij,ts}| \leq M$ by assumption.

ASSUMPTION E: *For some nonrandom positive definite matrix D_Z,*

$$\operatorname{plim} \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \sigma_{ij,ts} Z_{it} Z_{js}' = D_Z, \tag{16}$$

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i' \varepsilon_i \xrightarrow{d} N(0, D_Z).$$

In the absence of serial correlation and heteroskedasticity, we let $\sigma_{ij} = \sigma_{ij,tt} = E(\varepsilon_{it} \varepsilon_{jt})$ since it does not depend on t, and we denote D_Z by D_1. Likewise, with no cross-section correlation and heteroskedasticity, we let $\omega_{ts} =$

$\sigma_{ii,ts} = E(\varepsilon_{it}\varepsilon_{is})$ since it does not depend on i, and we denote D_Z by D_2. That is, D_1 and D_2 are the probability limits of

$$\text{plim}\, \frac{1}{NT} \sum_{i=1}^{N}\sum_{j=1}^{N} \sigma_{ij} \sum_{t=1}^{T} Z_{it} Z_{jt}' = D_1, \tag{17}$$

$$\text{plim}\, \frac{1}{NT} \sum_{t=1}^{T}\sum_{s=1}^{T} \omega_{ts} \sum_{i=1}^{N} Z_{it} Z_{is}'. = D_2$$

The corresponding central limit theorem will be denoted by $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i'\varepsilon_i \xrightarrow{d} N(0, D_1)$ and $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i'\varepsilon_i \xrightarrow{d} N(0, D_2)$, respectively.

THEOREM 2: *Assume Assumptions A–E hold. As $T, N \to \infty$, the following statements hold*:

(i) *In the absence of serial correlation and heteroskedasticity and with $T/N \to 0$,*

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, D_0^{-1} D_1 D_0'^{-1}).$$

(ii) *In the absence of cross-section correlation and heteroskedasticity and with $N/T \to 0$,*

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, D_0^{-1} D_2 D_0'^{-1}),$$

where $D_0 = \text{plim}\, D(F^0) = \text{plim}\, \frac{1}{NT} \sum_{i=1}^{N} Z_i' Z_i$.

Noting that $D_1 = D_2 = \sigma^2 D_0$ under i.i.d. assumption of ε_{it}, the following statement holds:

COROLLARY 1: *Under the assumptions of Theorem 1, if ε_{it} are i.i.d. over t and i, with zero mean and variance σ^2, then $\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, \sigma^2 D_0^{-1})$.*

It is conjectured that $\hat{\beta}$ is asymptotically efficient if ε_{it} are i.i.d. $N(0, \sigma^2)$, based on the argument of Hahn and Kuersteiner (2002).

Part (i) of Theorem 1 still permits cross-section correlation and heteroskedasticity, and part (ii) still permits serial correlation and heteroskedasticity. The theorem also requires an appropriate rate for N and T. If T/N converges to a constant, there will be a bias term due to correlation and heteroskedasticity. The next theorem is concerned with this bias. We shall deal with the more general case in which correlation and heteroskedasticity exist in both dimensions.

THEOREM 3: *Assume Assumptions A–E hold and* $T/N \to \rho > 0$. *Then*

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N\big(\rho^{1/2}B_0 + \rho^{-1/2}C_0, D_0^{-1}D_Z D_0'^{-1}\big),$$

where B_0 *is the probability limit of* B *with*

$$(18)\qquad B = -D(F^0)^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{(X_i - V_i)'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1} \times \left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_k\left(\frac{1}{T}\sum_{t=1}^{T}\sigma_{ik,tt}\right)$$

and C_0 *is the probability limit of* C *with*

$$(19)\qquad C = -D(F^0)^{-1}\frac{1}{NT}\sum_{i=1}^{N}X_i' M_{F^0}\Omega F^0\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i,$$

and $V_i = \frac{1}{N}\sum_{j=1}^{N} a_{ij}X_j$, $a_{ij} = \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_j$, *and* $\Omega = \frac{1}{N}\sum_{k=1}^{N}\Omega_k$ *with* $\Omega_k = E(\varepsilon_k\varepsilon_k')$.

There will be no biases in the absence of correlations and heteroskedasticities. In particular, bias $B_0 = 0$ when cross-sectional correlation and heteroskedasticity are absent, and similarly $C_0 = 0$ when serial correlation and heteroskedasticity are absent. To see this, consider C in (19). The absence of serial correlation and heteroskedasticity implies $\Omega_k = \sigma_k^2 I_T$; thus, $M_{F^0}\Omega F^0 = (\sum_k \sigma_k^2)M_{F^0}F^0 = 0$. It follows that $C = 0$ and hence $C_0 = 0$. The parametric form of serial correlations is usually removed by adding lagged dependent variables. However, lagged dependent variables lead to bias with fixed-effect estimators; see Hahn and Kuersteiner (2002) in a different context. The bias will take a different form and is not studied here. The argument for $B = 0$ is not so obvious and is provided in the proof of Theorem 2(ii). When ε_{it} are i.i.d. over t and over i, then both B_0 and C_0 become zero, and the result specializes to Corollary 1. These results assume no lagged dependent variables, whose presence will lead to additional bias, which is not studied in this paper.

REMARK 4: Suppose that k factors are allowed in the estimation, with k fixed but $k \geq r$. Then $\hat{\beta}$ remains $\sqrt{NT}$ consistent, albeit less efficient than $k = r$. Consistency relies on controlling the space spanned by Λ and that of F, which is achieved when $k \geq r$.

REMARK 5: Due to $\sqrt{NT}$ consistency for $\hat{\beta}$, estimation of β does not affect the rates of convergence and the limiting distributions of $\hat{F}_t$ and $\hat{\lambda}_i$. That is, they are the same as that of the pure factor model of Bai (2003). This follows from $Y_{it} - X_{it}'\hat{\beta} = \lambda_i' F_t + e_{it} + X_{it}'(\hat{\beta} - \beta)$, which is a pure factor model with an added error $X_{it}'(\hat{\beta} - \beta) = (NT)^{-1/2} O_p(1)$. An error of this order of magnitude does not affect the analysis.

6. INTERPRETATIONS OF THE ESTIMATOR

The Meaning of $D(F)$ and the Within-Group Interpretation

Like the least squares dummy-variable (LSDV) estimator, the interactive-effects estimator $\hat{\beta}$ is a result of least squares with the effects being estimated. In this sense, it is a within estimator. It is more instructive, however, to compare the mathematical expressions of the two estimators. Write the additive-effects model (2) in matrix form:

$$Y = \beta_1 X^1 + \beta_2 X^2 + \cdots + \beta_p X^p + \iota_T \alpha' + \xi \iota_N' + \varepsilon, \tag{20}$$

where Y and X^k $(k = 1, 2, \ldots, p)$ are matrices of $T \times N$, with X^k being the regressor matrix associated with parameter β_k (a scalar); ι_T is a $T \times 1$ vector with all elements being 1 and similarly for ι_N; $\alpha' = (\alpha_1, \ldots, \alpha_N)$ and $\xi = (\xi_1, \ldots, \xi_T)'$. Define

$$M_T = I_T - \iota_T \iota_T'/T, \quad M_N = I_N - \iota_N \iota_N'/N.$$

Multiplying equation (20) by M_T from the left and by M_N from the right yields

$$M_T Y M_N = \beta_1 (M_T X^1 M_N) + \cdots + \beta_p (M_T X^p M_N) + M_T \varepsilon M_N.$$

The least squares dummy-variable estimator is simply the least squares applied to the above transformed variables. The interactive-effects estimator has a similar interpretation. Rewrite the interactive-effects model (5) as

$$Y = \beta_1 X^1 + \cdots + \beta_p X^p + F\Lambda' + \varepsilon.$$

Then left multiply M_F and right multiply M_Λ to obtain

$$M_F Y M_\Lambda = \beta_1 (M_F X^1 M_\Lambda) + \cdots + \beta_p (M_F X^p M_\Lambda) + M_F \varepsilon M_\Lambda.$$

Let $\hat{\beta}_{\text{Asy}}$ be the least squares estimator obtained from the above transformed variables, treating F and Λ as known. That is,

$$\hat{\beta}_{\text{Asy}} = \begin{bmatrix} \operatorname{tr}[M_\Lambda X^{1\prime} M_F X^1] & \cdots & \operatorname{tr}[M_\Lambda X^{1\prime} M_F X^p] \\ \vdots & \vdots & \vdots \\ \operatorname{tr}[M_\Lambda X^{p\prime} M_F X^1] & \cdots & \operatorname{tr}[M_\Lambda X^{p\prime} M_F X^p] \end{bmatrix}^{-1} \times \begin{bmatrix} \operatorname{tr}[M_\Lambda X^{1\prime} M_F Y] \\ \vdots \\ \operatorname{tr}[M_\Lambda X^{p\prime} M_F Y] \end{bmatrix}.$$

The square matrix on the right without inverse is equal to $D(F)$ up to a scaling constant, that is,

$$D(F) = \frac{1}{TN}\sum_{i=1}^{N} Z_i' Z_i$$
$$= \frac{1}{TN}\begin{bmatrix} \mathrm{tr}[M_\Lambda X^{1\prime} M_F X^1] & \cdots & \mathrm{tr}[M_\Lambda X^{1\prime} M_F X^p] \\ \vdots & \vdots & \vdots \\ \mathrm{tr}[M_\Lambda X^{p\prime} M_F X^1] & \cdots & \mathrm{tr}[M_\Lambda X^{p\prime} M_F X^p] \end{bmatrix}.$$

This can be verified by some calculations. The estimator $\hat{\beta}_{\mathrm{Asy}}$ can be rewritten as

$$\hat{\beta}_{\mathrm{Asy}} = \left(\sum_{i=1}^{N} Z_i' Z_i\right)^{-1} \sum_{i=1}^{N} Z_i' Y_i.$$

It follows from (15) that $\sqrt{NT}(\hat{\beta}-\beta) = \sqrt{NT}(\hat{\beta}_{\mathrm{Asy}}-\beta) + o_p(1)$. To purge the fixed effects, the LSDV estimator uses M_T and M_N to transform the variables, whereas the interactive-effects estimator uses M_F and M_Λ to transform the variables.

7. BIAS-CORRECTED ESTIMATOR

The interactive-effect estimator is shown to have the representation (see Proposition A.3 in the Appendix)

$$\sqrt{NT}(\hat{\beta}-\beta^0) = D(F^0)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} Z_i'\varepsilon_i + \left(\frac{T}{N}\right)^{1/2} B + \left(\frac{N}{T}\right)^{1/2} C + o_p(1), \tag{21}$$

where B and C are given by (18) and (19), respectively, and they give rise to the biases. Their presence arises from correlations and heteroskedasticities in ε_{it}. We show that B and C can be consistently estimated so that a bias-corrected estimator can be constructed, as in the framework of Hahn and Kuersteiner (2002) and Hahn and Newey (2004). Attention is paid to heteroskedasticities in both dimensions, assuming no correlation in either dimension to simplify the presentation. We do point out how to estimate the biases consistently and outline the idea of the proof when correlation exists in either dimension.

Under the assumption of $E(\varepsilon_{it}^2)=\sigma_{i,t}^2$ and $E(\varepsilon_{it}\varepsilon_{js})=0$ for $i\neq j$ or $t\neq s$, term B becomes

$$B=-D(F^0)^{-1}\frac{1}{N}\sum_{i=1}^{N}\frac{(X_i-V_i)'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i\bar{\sigma}_i^2, \tag{22}$$

where $\bar{\sigma}_i^2=\frac{1}{T}\sum_{t=1}^{T}\sigma_{i,t}^2$. The bias can be estimated by replacing F^0 by $\hat{F}$, λ_i by $\hat{\lambda}_i$, and $\bar{\sigma}_i^2$ by $\hat{\bar{\sigma}}_i^2=\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$. This gives, in view of $\hat{F}'\hat{F}/T=I_r$,

$$\hat{B}=-\hat{D}_0^{-1}\frac{1}{N}\sum_{i=1}^{N}\frac{(X_i-\hat{V}_i)'\hat{F}}{T}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i\hat{\bar{\sigma}}_i^2. \tag{23}$$

The expression C is still given by (19), but Ω now becomes a diagonal matrix under no correlation, that is, $\Omega=\operatorname{diag}(\frac{1}{N}\sum_{k=1}^{N}\sigma_{k,1}^2,\dots,\frac{1}{N}\sum_{k=1}^{N}\sigma_{k,T}^2)$. Let $\hat{\Omega}=\operatorname{diag}(\frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_{k,1}^2,\dots,\frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_{k,T}^2)$ be an estimator for Ω. We estimate C by

$$\hat{C}=-\hat{D}_0^{-1}\frac{1}{NT}\sum_{i=1}^{N}X_i'M_{\hat{F}}\hat{\Omega}\hat{F}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i. \tag{24}$$

In the Appendix we prove $(T/N)^{1/2}(\hat{B}-B)=o_p(1)$ and $(N/T)^{1/2}(\hat{C}-C)=o_p(1)$. Define

$$\hat{\beta}^\dagger=\hat{\beta}-\frac{1}{N}\hat{B}-\frac{1}{T}\hat{C}.$$

THEOREM 4: *Assume Assumptions A–E hold. In addition,* $E(\varepsilon_{it}^2)=\sigma_{i,t}^2$ *and* $E(\varepsilon_{it}\varepsilon_{js})=0$ *for* $i\neq j$ *or* $t\neq s$. *If* $T/N^2\to 0$ *and* $N/T^2\to 0$, *then*

$$\sqrt{NT}(\hat{\beta}^\dagger-\beta^0)\xrightarrow{d}N(0,D_0^{-1}D_3D_0^{-1}),$$

where $D_3=\operatorname{plim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it}Z_{it}'\sigma_{i,t}^2$.

The limiting variance D_3 is a special case of D_Z due to the no correlation assumption. Bias correction does not contribute to the limiting variance. Also note that conditions $N/T^2\to 0$ and $T/N^2\to 0$ are added. Clearly, these conditions are less restrictive than T/N converging to a positive constant. There exist other bias correction procedures (e.g., panel jackknife) that could be used; see Arellano and Hahn (2005) and Hahn and Newey (2004). An alternative to bias correction in the case of $T/N\to\rho>0$ is to use the Bekker (1994) standard errors to improve inference accuracy. This strategy was studied by Hansen, Hausman, and Newey (2005) in the context of many instruments.

REMARK 6: Consider estimating C in the presence of serial correlation. We need consistent estimators for $T^{-1}X_i'\Omega_k F^0$ and $T^{-1}F^{0\prime}\Omega_k F^0$, where $\Omega_k = E\varepsilon_k\varepsilon_k'$ $(T\times T)$, and then we take (weighted) averages over i and over k. Thus consider estimating them for each given (i,k). These terms are standard expressions in the usual heteroskedasticity and autocorrelation (HAC) robust limiting covariance. To see this, let $W_i = (X_i, F^0)$ which is $T\times(p+r)$. Then the long-run variance of $T^{-1/2}W_i'\varepsilon_k = T^{-1/2}\sum_{t=1}^T W_{it}\varepsilon_{kt}$ is the limit of $\frac{1}{T}W_i'\Omega_k W_i$, which contains $\frac{1}{T}X_i'\Omega_k F^0$ and $\frac{1}{T}F^{0\prime}\Omega_k F^0$ as subblocks. A consistent estimator for $T^{-1}W_i'\Omega W_i$ can be constructed by the truncated kernel method of Newey and West (1987) based on the sequence $\hat{W}_{it}\hat{\varepsilon}_{kt}$ $(t=1,\dots,T)$. Similar argument has been made in Bai (2003).

REMARK 7: While estimating B in the presence of cross-section correlation is not difficult, the underlying theory for consistency requires a different argument. In the time series dimension, the data are naturally ordered and distant observations have less correlations. The kernel method puts small weights for autocovariances with large lags, leading to consistent estimation. In the cross-section dimension, such an ordering of data is not available, unless an economic distance can be constructed so that the data can be ordered. In general, large $|i-j|$ does not mean smaller correlation between ε_{it} and ε_{jt}. Bai and Ng (2006) studied the estimation of an object similar to B. They showed that if the whole cross-sample is used in the estimation, the estimator is inconsistent. A partial sample estimator, with N being replaced by n such that $n/N\to 0$ and $n/T\to 0$, is consistent. Thus, B can be estimated by

$$\hat{B} = -\hat{D}_0^{-1}\frac{1}{n}\sum_{i=1}^n\sum_{k=1}^n \frac{(X_i-\hat{V}_i)'\hat{F}}{T}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_k\left(\frac{1}{T}\sum_{t=1}^T\hat{\varepsilon}_{it}\hat{\varepsilon}_{kt}\right), \tag{25}$$

where $n/N\to 0$ and $n/T\to 0$. The argument of Bai and Ng (2006) can be adapted to show that $\hat{B}$ is consistent for B.

Estimating the Covariance Matrices

To estimate D_0, we define

$$\hat{D}_0 = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{Z}_{it}\hat{Z}_{it}',$$

where $\hat{Z}_{it}$ is equal to Z_{it} with F^0, λ_i, and Λ replaced with $\hat{F}$, $\hat{\lambda}_i$, and $\hat{\Lambda}$, respectively. Next consider estimating D_j, $j=1,2,3$. For all cases, we limit our attention to the presence of heteroskedasticity, but no correlation. Thus D_j

$(j = 1, 2, 3)$ are covariance matrices when heteroskedasticity exists in the cross-section dimension only, in the time dimension only, and in both dimensions, respectively. Thus we define

$$\hat{D}_1 = \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i^2\left(\frac{1}{T}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}'\right),$$

$$\hat{D}_2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\omega}_t^2\left(\frac{1}{N}\sum_{i=1}^{N}\hat{Z}_{it}\hat{Z}_{it}'\right),$$

$$\hat{D}_3 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}'\hat{\varepsilon}_{it}^2,$$

where $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$, $\hat{\omega}_t^2 = \frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_{it}^2$, and $\hat{Z}_{it}$ was defined previously.

PROPOSITION 2: *Assume Assumptions A–E hold. Then as $N, T \to \infty$, $\hat{D}_0 \xrightarrow{p} D_0$. In addition, in the absence of serial and cross-section correlations, $\hat{D}_j \xrightarrow{p} D_j$, where D_1 and D_2 are defined in Theorem 2 with no correlation, and D_3 is defined in Theorem 4.*

REMARK 8: When cross-section correlation exists, we estimate D_1 in (17) by

$$\hat{D}_1 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{jt}'\hat{\varepsilon}_{it}\hat{\varepsilon}_{jt},$$

where n satisfies $n/N \to 0$ and $n/T \to 0$; see Remark 7. It can be shown that $\hat{D}_1$ is consistent for D_1. When serial correlation exists, we estimate D_2 of (17) by estimating the long-run variance of the sequence $\{\hat{Z}_{it}\hat{\varepsilon}_{it}\}$ using the truncated kernel of Newey and West (1987); see Remark 6. It can be shown that $\hat{D}_2$ is consistent for D_2. For estimating D_Z—the covariance matrix when correlation exists in both dimensions—we need to use the partial sample method together with the Newey–West procedure. More specifically, let $\hat{\xi}_t = n^{-1/2}\sum_{i=1}^{n}\hat{Z}_{it}\hat{\varepsilon}_{it}$, where n is chosen as before. The estimated long-run variance (e.g., truncated kernel) for the sequence $\hat{\xi}_t$ is an estimator for D_Z. While we conjecture the estimator is consistent, a formal proof remains to be explored.

8. MODELS WITH BOTH ADDITIVE AND INTERACTIVE EFFECTS

Although interactive-effects models include the additive models as special cases, additivity has not been imposed so far, even when it is true. When additivity holds but is ignored, the resulting estimator is less efficient. In this section, we consider the joint presence of additive and interactive effects, and show how to estimate the model by imposing additivity and derive the limiting distribution of the resulting estimator. Consider

$$Y_{it} = X_{it}'\beta + \mu + \alpha_i + \xi_t + \lambda_i' F_t + \varepsilon_{it}, \tag{26}$$

where μ is the grand mean, α_i is the usual fixed effect, ξ_t is the time effect, and $\lambda_i' F_t$ is the interactive effect. Restrictions are required to identify the model. Even in the absence of the interactive effect, the restrictions

$$\sum_{i=1}^{N} \alpha_i = 0, \quad \sum_{t=1}^{T} \xi_t = 0 \tag{27}$$

are needed; see Greene (2000, p. 565). The following restrictions are maintained:

$$F'F/T = I_r, \quad \Lambda'\Lambda = \text{diagonal}. \tag{28}$$

Further restrictions are needed to separate the additive and interactive effects. They are

$$\sum_{i=1}^{N} \lambda_i = 0, \quad \sum_{t=1}^{T} F_t = 0. \tag{29}$$

To see this, suppose that $\bar{\lambda} = \frac{1}{N}\sum_{i=1}^{N} \lambda_i \neq 0$ or $\bar{F} = \frac{1}{T}\sum_{t=1}^{T} F_t \neq 0$, or both are not zero. Let $\lambda_i^{\dagger} = \lambda_i - 2\bar{\lambda}$ and $F_t^{\dagger} = F_t - 2\bar{F}$. Then

$$Y_{it} = X_{it}'\beta + \mu + \alpha_i^{\dagger} + \xi_t^{\dagger} + \lambda_i^{\dagger\prime} F_t^{\dagger} + \varepsilon_{it},$$

where $\alpha_i^{\dagger} = \alpha_i + 2\bar{F}'\lambda_i - 2\bar{\lambda}'\bar{F}$ and $\xi_t^{\dagger} = \xi_t + 2\bar{\lambda}'F_t - 2\bar{\lambda}'\bar{F}$. It is easy to verify that $F^{\dagger\prime}F^{\dagger}/T = F'F/T = I_r$ and $\Lambda^{\dagger\prime}\Lambda^{\dagger} = \Lambda'\Lambda$ is diagonal, and at the same time, $\sum_{i=1}^{N} \alpha_i^{\dagger} = 0$ and $\sum_{t=1}^{T} \xi_t^{\dagger} = 0$. Thus the new model is observationally equivalent to (26) if (29) is not imposed.

To estimate the general model under the given restrictions, we introduce some standard notation. For any variable ϕ_{it}, define

$$\bar{\phi}_{\cdot t} = \frac{1}{N}\sum_{i=1}^{N} \phi_{it}, \quad \bar{\phi}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} \phi_{it}, \quad \bar{\phi}_{\cdot\cdot} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \phi_{it},$$
$$\dot{\phi}_{it} = \phi_{it} - \bar{\phi}_{i\cdot} - \bar{\phi}_{\cdot t} + \bar{\phi}_{\cdot\cdot}$$

and its vector form $\dot{\phi}_i = \phi_i - \iota_T \bar{\phi}_{i\cdot} - \bar{\phi} + \iota_T \bar{\phi}_{\cdot\cdot}$, where $\bar{\phi} = (\bar{\phi}_{\cdot 1}, \ldots, \bar{\phi}_{\cdot T})'$.

The least squares estimators are

$$\hat{\mu} = \bar{Y}_{..} - \bar{X}'_{..}\hat{\beta},$$
$$\hat{\alpha}_i = \bar{Y}_{i.} - \bar{X}'_{i.}\hat{\beta} - \hat{\mu},$$
$$\hat{\xi}_t = \bar{Y}_{.t} - \bar{X}'_{.t}\hat{\beta} - \hat{\mu},$$
$$\hat{\beta} = \left[\sum_{i=1}^{N} \dot{X}'_i M_{\hat{F}} \dot{X}_i\right]^{-1} \sum_{i=1}^{N} \dot{X}'_i M_{\hat{F}} \dot{Y}_i,$$

and $\hat{F}$ is the $T \times r$ matrix consisting of the first r eigenvectors (multiplied by $\sqrt{T}$) associated with the first r largest eigenvalues of the matrix $\frac{1}{NT}\sum_{i=1}^{N}(\dot{Y}_i - \dot{X}_i\hat{\beta})(\dot{Y}_i - \dot{X}_i\hat{\beta})'$. Finally, $\hat{\Lambda}$ is expressed as a function of $(\hat{\beta}, \hat{F})$ such that

$$\hat{\Lambda}' = (\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_N) = T^{-1}[\hat{F}'(\dot{Y}_1 - \dot{X}_1\hat{\beta}), \ldots, \hat{F}'(\dot{Y}_N - \dot{X}_N\hat{\beta})].$$

Iterations are required to obtain $\hat{\beta}$ and $\hat{F}$. The remaining parameters $\hat{u}$, $\hat{\alpha}_i$, $\hat{\xi}_t$, and $\hat{\Lambda}$ require no iteration, and they can be computed once $\hat{\beta}$ and $\hat{F}$ are obtained. The solutions for $\hat{\mu}$, $\hat{\alpha}_i$, and $\hat{\xi}_t$ have the same form as the usual fixed-effects model; see Greene (2000, p. 565).

We shall argue that $(\hat{\mu}, \{\hat{\alpha}_i\}, \{\hat{\xi}_t\}, \hat{\beta}, \hat{F}, \hat{\Lambda})$ are indeed the least squares estimators from minimization of the objective function

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - X'_{it}\beta - \mu - \alpha_i - \xi_t - \lambda'_i F_t)^2$$

subject to the restrictions (27)–(29). Concentrating out $(\mu, \{\alpha_i\}, \{\xi_t\})$ is equivalent to using $(\dot{Y}_{it}, \dot{X}_{it})$ to estimate the remaining parameters. So the concentrated objective function is

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(\dot{Y}_{it} - \dot{X}'_{it}\beta - \lambda'_i F_t)^2.$$

The dotted variable for $\lambda'_i F_t$ is itself, that is, $\dot{c}_{it} = c_{it}$, where $c_{it} = \lambda'_i F_t$ due to restriction (29). This objective function is the same as (8), except Y_{it} and X_{it} are replaced by their dotted versions. From the analysis in Section 3, the least squares estimators for β, F, and Λ are as prescribed above. Given these estimates, the least squares estimators for $(\mu, \{\alpha_i\}, \{\xi_t\})$ are also immediately obtained as prescribed.

We next argue that all restrictions are satisfied. For example, $\frac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_i = \bar{Y}_{..} - \bar{X}_{..}\hat{\beta} - \hat{\mu} = \hat{\mu} - \hat{\mu} = 0$. Similarly, $\sum_{t=1}^{T}\hat{\xi}_t = 0$. It requires an extra argument to show $\sum_{t=1}^{T}\hat{F}_t = 0$. By definition,

$$\hat{F}V_{NT} = \left[\frac{1}{NT}\sum_{i=1}^{N}(\dot{Y}_i - \dot{X}_i\hat{\beta})(\dot{Y}_i - \dot{X}_i\hat{\beta})'\right]\hat{F}.$$

Multiplying $\iota_T = (1, \dots, 1)'$ on each side yields

$$\iota_T'\hat{F}V_{NT} = \left[\frac{1}{NT}\sum_{i=1}^{N}\iota_T'(\dot{Y}_i - \dot{X}_i\hat{\beta})(\dot{Y}_i - \dot{X}_i\hat{\beta})'\right]\hat{F},$$

but $\iota_T'\dot{Y}_i = \sum_{t=1}^{T}\dot{Y}_{it} = 0$ and, similarly, $\iota_T'\dot{X}_i = 0$. Thus the right-hand side is zero, implying $\iota_T'\hat{F} = 0$. The same argument leads to $\sum_{i=1}^{N}\hat{\lambda}_i = 0$.

To derive the asymptotic distribution for $\hat{\beta}$, we define

$$\dot{Z}_i(F) = M_F\dot{X}_i - \frac{1}{N}\sum_{k=1}^{N}a_{ik}M_F\dot{X}_k \quad \text{and}$$

$$\dot{D}(F) = \frac{1}{NT}\sum_{i=1}^{N}\dot{Z}_i(F)'\dot{Z}_i(F),$$

where $a_{ik} = \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$. We assume

$$\inf_F \dot{D}(F) > 0. \tag{30}$$

Let $\dot{Z}_i = \dot{Z}_i(F^0)$. Notice that

$$\dot{Y}_{it} = \dot{X}_{it}'\beta + \lambda_i'F_t + \dot{\varepsilon}_{it}$$

The entire analysis of Section 4 can be restated here. In particular, under the conditions of Theorem 2, we have the asymptotic representation

$$\sqrt{NT}(\hat{\beta} - \beta^0) = \left[\frac{1}{NT}\sum_{i=1}^{N}\dot{Z}_i'\dot{Z}_i\right]^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{Z}_i'\dot{\varepsilon}_i + o_p(1).$$

In the Supplemental Material, we show the identity (see Lemma A.13) $\sum_{i=1}^{N}\dot{Z}_i'\dot{\varepsilon}_i \equiv \sum_{i=1}^{N}\dot{Z}_i'\varepsilon_i$. That is, $\dot{\varepsilon}_i$ can be replaced by ε_i. It follows that if normality is assumed for $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{Z}_i'\varepsilon_i$, asymptotic normality also holds for $\sqrt{NT}(\hat{\beta} - \beta)$.

ASSUMPTION F: (i) $\operatorname{plim} \frac{1}{NT}\sum_{i=1}^{N}\dot{Z}_i'\dot{Z}_i = \dot{D}_0 > 0$; (ii) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{Z}_i'\varepsilon_i \xrightarrow{d} N(0, \dot{D}_Z)$, *where* $\dot{D}_Z = \operatorname{plim} \frac{1}{NT}\sum_{i,j,t,s}\sigma_{ij,ts}\dot{Z}_{it}\dot{Z}_{js}'$.

THEOREM 5: *Assume Assumptions A–F hold. Then as* $T, N \to \infty$, *the following statements hold*:

(i) *Under the assumptions of Theorem 2*(i),

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, \dot{D}_0^{-1}\dot{D}_1\dot{D}_0^{-1}).$$

(ii) *Under the assumptions of Theorem 2*(ii),

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, \dot{D}_0^{-1}\dot{D}_2\dot{D}_0^{-1}),$$

where $\dot{D}_1$ *and* $\dot{D}_2$ *are special cases of* $\dot{D}_Z$.

An analogous result to Theorem 3 also holds, and bias-corrected estimators can also be considered. Since the analysis holds with X_i replaced by $\dot{X}_i$, details are omitted.

9. TESTING ADDITIVE VERSUS INTERACTIVE EFFECTS

There exist two methods to evaluate which specification—fixed effects or interactive effects—gives a better description of the data. The first method is that of the Hausman test statistic (Hausman (1978)) and the second is based on the number of factors. We detail the Hausman test method, delegating the number-of-factors method to the Supplemental Material. Throughout this section, for simplicity, we assume ε_{it} are i.i.d. over i and t, and that $E(\varepsilon_{it}^2) = \sigma^2$.

The null hypothesis is an additive-effects model

$$Y_{it} = X_{it}\beta + \alpha_i + \xi_t + \mu + \varepsilon_{it} \tag{31}$$

with restrictions $\sum_{i=1}^{N}\alpha_i = 0$ and $\sum_{t=1}^{T}\xi_t = 0$ due to the grand mean parameter μ. The alternative hypothesis—more precisely, the encompassing general model—is

$$Y_{it} = X_{it}\beta + \lambda_i'F_t + \varepsilon_{it}. \tag{32}$$

The null model is nested in the general model with $\lambda_i' = (\alpha_i, 1)$ and $F_t = (1, \xi_t + \mu)'$.

The interactive-effects estimator for β is consistent under both models (31) and (32), but is less efficient than the least squares dummy-variable estimator for model (31), as the latter imposes restrictions on factors and factor loadings. But the fixed-effects estimator is inconsistent under model (32). The principle of the Hausman test is applicable here.

The within-group estimator of β in (31) is

$$\sqrt{NT}(\hat{\beta}_{\mathrm{FE}} - \beta) = \left(\frac{1}{NT}\sum_{i=1}^{N}\dot{X}_i'\dot{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{X}_i\varepsilon_i,$$

where $\dot{X}_i = X_i - \iota_T\bar{X}_{i\cdot} - \bar{X} + \iota_T\bar{X}_{\cdot\cdot}$. Rewrite the fixed-effects estimator more compactly as

$$\sqrt{NT}(\hat{\beta}_{\mathrm{FE}} - \beta) = C^{-1}\psi,$$

where $C = (\frac{1}{NT}\sum_{i=1}^{N}\dot{X}_i'\dot{X}_i)$ and $\psi = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\dot{X}_i'\varepsilon_i$. The interactive-effects estimator can be written as (see Proposition A.3)

$$\sqrt{NT}(\hat{\beta}_{\mathrm{IE}} - \beta) = D(F^0)^{-1}(\eta - \xi) + o_p(1),$$

where

$$\eta = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}X_i'M_{F^0}\varepsilon_i, \quad \xi = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'M_{F^0}\right]\varepsilon_i. \tag{33}$$

The variances of the two estimators are

$$\mathrm{var}(\sqrt{NT}(\hat{\beta}_{\mathrm{FE}} - \beta)) = \sigma^2C^{-1}, \quad \mathrm{var}(\sqrt{NT}(\hat{\beta}_{\mathrm{IE}} - \beta)) = \sigma^2D(F^0)^{-1}.$$

In the accompanying document, we show, under the null hypothesis of additivity,

$$E[(\eta - \xi)\psi'] = \sigma^2D(F^0). \tag{34}$$

This implies $\mathrm{var}(\hat{\beta}_{\mathrm{IE}} - \hat{\beta}_{\mathrm{FE}}) = \mathrm{var}(\hat{\beta}_{\mathrm{IE}}) - \mathrm{var}(\hat{\beta}_{\mathrm{FE}})$. Thus the Hausman test takes the form

$$J = NT\sigma^2(\hat{\beta}_{\mathrm{IE}} - \hat{\beta}_{\mathrm{FE}})'[D(F^0)^{-1} - C^{-1}]^{-1}(\hat{\beta}_{\mathrm{IE}} - \hat{\beta}_{\mathrm{FE}}) \xrightarrow{d} \chi^2_p.$$

Replacing $D(F^0)$ and σ^2 by their consistent estimators, the above is still true. Proposition 2 shows that $D(F^0)$ is consistently estimated by $\hat{D}_0$. Let $\hat{\sigma}^2 = \frac{1}{L}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$, where $L = NT - (N+T)r - p$. Then $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

REMARK 9: The Hausman test is also applicable when there are no time effects but only individual effects (i.e., $\xi_t = 0$). Then it is testing whether the individual effects are time-varying. Similarly, the Hausman test is applicable when $\alpha_i = 0$ in (31) but $\xi_t \neq 0$. Then it is testing whether the common shocks have heterogeneous effects on individuals. Details are given in the Supplemental Materials.

10. TIME-INVARIANT AND COMMON REGRESSORS

In earnings studies, time-invariant regressors include education, gender, race, and so forth; common variables are those that represent trends or policies. In consumption studies, common regressors include price variables, which are the same for each individual. Those variables are removed by the within-group transformation. As a result, identification and estimation must rely on other means such as the instrumental variable approach of Hausman and Taylor (1981). This section considers similar problems under interactive effects. Under some reasonable and intuitive conditions, the parameters of the time-invariant and common regressors are shown to be identifiable and can be consistently estimated. In effect, those regressors act as their own instruments; additional instruments, either within or outside the system, are not necessary. Ahn, Lee, and Schmidt (2001) allowed for time-invariant regressors, although they did not consider the joint presence of common regressors. Their identification condition relies on nonzero correlation between factor loadings and regressors.

A general model can be written as

$$Y_{it} = X_{it}'\varphi + x_i'\gamma + w_t'\delta + \lambda_i'F_t + \varepsilon_{it}, \tag{35}$$

where (X_{it}', x_i', w_t') is a vector of observable regressors, x_i is time invariant, and w_t is cross-sectionally invariant (common). The dimensions of regressors are such that X_{it} is $p \times 1$, x_i is $q \times 1$, w_t is $\ell \times 1$, and F_t is $r \times 1$. Introduce

$$X_i = \begin{bmatrix} X_{i1}' & x_i' & w_1' \\ X_{i2}' & x_i' & w_2' \\ & \vdots & \\ X_{iT}' & x_i' & w_T' \end{bmatrix}, \quad \beta = \begin{bmatrix} \varphi \\ \gamma \\ \delta \end{bmatrix},$$

$$\underline{x} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_N' \end{bmatrix}, \quad W = \begin{bmatrix} w_1' \\ w_2' \\ \vdots \\ w_T' \end{bmatrix}.$$

Then the model can be rewritten as

$$Y_i = X_i\beta + F\lambda_i + \varepsilon_i.$$

Let (β^0, F^0, Λ) denote the true parameters (superscript 0 is not used for Λ). To identify β^0, it was assumed in Section 4 that the matrix

$$D(F) = \frac{1}{NT}\sum_{i=1}^{N} X_i'M_F X_i - \frac{1}{T}\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i'M_F X_k \lambda_i'\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_k\right]$$

is positive definite for all possible F. This assumption fails when time-invariant regressors and common regressors exist. This is because $D(\iota_T)$ and $D(W)$ are not full rank matrices. However, the positive definiteness of $D(F)$ is not a necessary condition. In fact, all that is needed is the identification condition

$$D(F^0) > 0.$$

That is, the matrix $D(F)$ is positive definite when evaluated at the true F^0, a much weaker condition than Assumption A. In the Supplemental Material, we show that the above condition can be decomposed into some intuitive assumptions. First, this means that the interactive effects are genuine (not additive effects); otherwise, we are back to the environment of Hausman and Taylor, and instrumental variables must be used to identify β. Second, there should be no multicollinearity between W and F^0, and no multicollinearity between $\underline{x}$ and Λ. Finally, W and $\underline{x}$ cannot both contain the constant regressor (only one grand mean parameter).

It remains to argue that $D(F^0) > 0$ (or equivalently, the four conditions above) implies consistent estimation. We state this result as a proposition.

PROPOSITION 3: *Assume Assumptions B–D hold. If $D(F^0) > 0$, then $\hat{\beta} \xrightarrow{p} \beta^0$.*

The proof of this proposition is nontrivial and is provided in the Supplemental Material. The proposition implies that $D(F^0) > 0$ is a sufficient condition for consistent estimation.

Given consistency, the rest of the argument for rate of convergence does not hinge on any particular structure of the regressors. Therefore, the rate of convergence of $\hat{\beta}$ and the limiting distribution are still valid in the presence of the grand mean, time-invariant regressors, and common regressors. More specifically, all results up to Section 7 (inclusive) are valid. The result of Section 8 is valid for regressors with variations in both dimensions. Similarly, hypothesis testing in Section 9 can only rely on the subset of coefficients whose regressors have variations in both dimensions.

Discussion

When additive effects are also present, (35) becomes

$$Y_{it} = X_{it}'\varphi + \mu + \alpha_i + \xi_t + x_i'\gamma + w_t'\delta + \lambda_i'F_t + \varepsilon_{it},$$

where μ is the grand mean (explicitly written out), and α_i and ξ_t are, respectively, the individual and the time effects. The parameters γ and δ are no longer directly estimable. Under the restrictions of (27) and (29), the within-group transformation implies $\dot{Y}_{it} = \dot{X}_{it}\phi + \lambda_i'F_t + \dot{\varepsilon}_{it}$. The parameters ϕ and

$\lambda_i'F_t$ are estimable by the interactive-effects estimator, so they can be treated as known in terms of identification. Letting $Y_{it}^* = Y_{it} - X_{it}'\phi - \lambda_i'F_t$, we have

$$Y_{it}^* = \mu + \alpha_i + \xi_t + x_i'\gamma + w_t'\delta + \varepsilon_{it}, \tag{36}$$

which is a standard model. As in Hausman and Taylor (1981), if we assume a subset of X_{it}'s whose time averages are uncorrelated with α_i but correlated with x_i, time averages can be used as instruments for x_i. We can assume a similar instrument for w_t to estimate both γ and δ. This is a direct extension of the Hausman and Taylor framework for interactive-effects models.

A more interesting setup is to allow time-dependent coefficients for the time-invariant regressors and, similarly, allow individual-dependent coefficients for the common regressors; namely

$$Y_{it} = X_{it}'\phi + \mu + \alpha_i + \xi_t + x_i'\gamma_t + w_t'\delta_i + \lambda_i'F_t + \varepsilon_{it}. \tag{37}$$

The observable variables are Y_{it}, X_{it}, x_i, and w_t. Again the levels of γ_t and δ_i are not directly estimable due to $\alpha_i + \xi_t$. For example, $\alpha_i + x_i'\gamma_t = (\alpha_i + x_i'c) + x_i'(\gamma_t - c) = \alpha_i^* + x_i'\gamma_t^*$. However, the deviations of γ_t from its time average $\gamma_t - E(\gamma_t)$ and the deviations of δ_i from its individual average $\delta_i - E(\delta_i)$ are directly estimable. In practice, these deviations or contrasts may be of more importance than the levels, since they reveal the patterns across individuals or changes over time, just like coefficients on dummy variables. Restrictions are needed to estimate $(\phi, \mu, \gamma_t, \delta_i)$. For example, we need to impose $E(\gamma_t) = 0$ or $\sum_t \gamma_t = 0$, and we also need similar restrictions for δ_i, α_i, ξ_t, λ_i, and F_t, together with some normalization and multicollinearity restrictions. Unreported simulations show that these deviations can be well estimated. To estimate the levels $E(\gamma_t)$ and $E(\delta_i)$, the Hausman–Taylor approach appears to be applicable as well. In this case, Y_{it}^* in (36) is replaced by $Y_{it}^* = Y_{it} - X_{it}'\phi - x_i'\gamma_t^* - w_t'\delta_i^* - \lambda_i'F_t$, where $\gamma_t^* = \gamma_t - E(\gamma_t)$ and $\delta_i^* = \delta_i - E(\delta_i)$ are the deviations. The large sample theory of this model warrants a separate study.

11. FINITE SAMPLE PROPERTIES VIA SIMULATIONS

We assess the performance of the estimator by Monte Carlo simulations. A general model with common regressors and time-invariant regressors is considered:

$$Y_{it} = X_{it,1}\beta_1 + X_{it,2}\beta_2 + \mu + x_i\gamma + w_t\delta + \lambda_i'F_t + \varepsilon_{it}$$
$$((\beta_1, \beta_2, \mu, \gamma, \delta) = (1, 3, 5, 2, 4)),$$

where $\lambda_i = (\lambda_{i1}, \lambda_{i2})'$ and $F_t = (F_{t1}, F_{t2})'$. The regressors are generated according to

$$X_{it,1} = \mu_1 + c_1\lambda_i'F_t + \iota'\lambda_i + \iota'F_t + \eta_{it,1},$$
$$X_{it,2} = \mu_2 + c_2\lambda_i'F_t + \iota'\lambda_i + \iota'F_t + \eta_{it,2}$$

TABLE I

MODELS WITH GRAND MEAN, TIME-INVARIANT REGRESSORS AND COMMON REGRESSORS (TWO FACTORS, $r = 2$)

N	T	Mean $\beta_1 = 1$	SD	Mean $\beta_2 = 3$	SD	Mean $\mu = 5$	SD	Mean $\gamma = 2$	SD	Mean $\delta = 4$	SD
					Infeasible Estimator						
100	10	1.003	0.061	2.999	0.061	4.994	0.103	1.998	0.060	4.003	0.087
100	20	1.001	0.039	2.998	0.041	5.002	0.065	2.000	0.040	4.000	0.054
100	50	1.000	0.025	3.002	0.024	5.000	0.039	1.999	0.024	4.000	0.030
100	100	1.000	0.017	3.000	0.017	5.000	0.029	1.999	0.017	3.999	0.020
10	100	0.998	0.056	3.002	0.055	4.998	0.098	2.002	0.066	4.001	0.063
20	100	1.000	0.039	2.998	0.039	5.000	0.064	2.002	0.040	3.999	0.046
50	100	1.000	0.024	3.001	0.025	4.999	0.040	2.001	0.025	4.000	0.029
					Interactive-Effects Estimator						
100	10	1.104	0.135	3.103	0.138	4.611	0.925	1.952	0.242	3.939	0.250
100	20	1.038	0.083	3.036	0.084	4.856	0.524	1.996	0.104	3.989	0.114
100	50	1.010	0.036	3.012	0.037	4.981	0.156	1.995	0.098	3.999	0.058
100	100	1.006	0.032	3.006	0.033	4.992	0.115	1.996	0.066	3.997	0.061
10	100	1.105	0.133	3.108	0.135	4.556	0.962	1.939	0.240	3.949	0.259
20	100	1.038	0.083	3.037	0.084	4.859	0.479	1.991	0.109	3.996	0.082
50	100	1.009	0.035	3.010	0.037	4.974	0.081	2.000	0.041	4.000	0.033

with $\iota' = (1, 1)$. The regressors are correlated with λ_i, F_t, and the product $\lambda_i' F_t$. The variables λ_{ij}, F_{tj}, and $\eta_{it,j}$ are all i.i.d. $N(0, 1)$ and the regression error ε_{it} is i.i.d. $N(0, 4)$. We set $\mu_1 = \mu_2 = c_1 = c_2 = 1$. Further, $x_i \sim \iota'\lambda_i + e_i$ and $w_t = \iota' F_t + \eta_i$, with e_i and η_i being i.i.d. $N(0, 1)$, so that x_i is correlated with λ_i and w_t is correlated with F_t.

Simulation results are reported in Table I (based on 1000 repetitions). The infeasible estimator in this table assumes observable F_t. Both the infeasible and interactive-effects estimators are consistent, but the latter is less efficient than the former, as expected. The coefficients for the common regressors and time-invariant regressors are estimated well. The within-group estimator can only estimate β_1 and β_2 and is not reported.

We next investigate what happens when interactive-effects estimator is used when the underlying effects are additive. That is, $\lambda_i = (\alpha_i, 1)'$ and $F_t = (1, \xi_t)'$ so that $\lambda_i' F_t = \alpha_i + \delta_t$. With regressors $X_{it,1}$ and $X_{it,2}$ generated with the earlier formula, the model is

$$Y_{it} = X_{it,1}\beta_1 + X_{it,2}\beta_2 + \alpha_i + \xi_t + \varepsilon_{it}.$$

We consider three estimators: (i) the within-group estimator, (ii) the infeasible estimator, and (iii) the interactive-effects estimator. All three are consistent. The results are reported in Table II. The interactive-effects estimator remains valid under additive effects, but is less efficient than the within-group estimator, as expected.

TABLE II

MODELS OF ADDITIVE EFFECTS

		Within-Group Estimator				Infeasible Estimator				Interactive-Effects Estimator			
N	T	Mean $\beta_1 = 1$	SD	Mean $\beta_2 = 3$	SD	Mean β_1	SD	Mean β_2	SD	Mean β_1	SD	Mean β_2	SD
100	3	1.002	0.146	2.997	0.144	1.001	0.208	2.998	0.206	1.155	0.253	3.164	0.259
100	5	1.001	0.099	3.002	0.100	1.001	0.114	3.003	0.118	1.189	0.194	3.190	0.186
100	10	1.000	0.068	2.996	0.066	1.000	0.072	2.995	0.072	1.110	0.167	3.106	0.167
100	20	0.999	0.048	2.999	0.046	0.998	0.048	2.998	0.047	1.017	0.083	3.016	0.080
100	50	1.001	0.029	2.999	0.029	1.001	0.029	2.999	0.029	1.003	0.029	3.000	0.029
100	100	0.999	0.021	3.000	0.021	0.999	0.021	3.000	0.021	1.000	0.021	3.001	0.021
3	100	1.001	0.142	2.995	0.143	1.002	0.113	2.996	0.116	1.163	0.240	3.165	0.251
5	100	1.000	0.102	3.005	0.100	1.000	0.093	3.006	0.092	1.179	0.190	3.180	0.189
10	100	1.000	0.069	2.999	0.069	1.001	0.066	2.999	0.065	1.106	0.167	3.106	0.164
20	100	1.001	0.047	3.000	0.047	1.001	0.045	3.000	0.046	1.018	0.080	3.017	0.080
50	100	0.998	0.030	3.002	0.029	0.998	0.030	3.002	0.028	1.000	0.030	3.004	0.029

Additional simulations are reported in the Supplemental Material, where we consider cross-sectionally correlated e_{it}. Under cross-section correlation in e_{it} and with a fixed N, the interactive-effects estimator is inconsistent. The estimator becomes consistent as N going to infinity. These theoretical results are confirmed by the simulations. A primary use of the factor model in practice is to account for cross-sectional correlations. With a sufficient number of factors included, much of the correlation in the error terms will either be removed or be reduced, making the correlation a less critical issue. We also report results that include lagged dependent variables as regressors. The idea is to parametrize and to control for serial correlation. The parameters are well estimated in the simulation. The interactive-effects estimator is effective under large N and large T. The computation is fast and the bias is decreasing with N and T, as shown in the theory and confirmed in the simulation.

12. CONCLUDING REMARKS

In this paper, we have examined issues related to identification and inference for panel data models with interactive effects. In earnings studies, the interactive effects are a result of changing prices for a vector of unmeasured skills. The model can also be motivated from an optimal choice of consumption and labor supply for heterogeneous agents under a competitive economy with complete markets. In macroeconomics, interactive effects represent common shocks and heterogeneous impacts on the cross-units. In finance, the common factors represent marketwide risks and the loadings reflect assets' exposure to the risks. A factor model is also a useful approach to controlling cross-section correlations. This paper focuses on some of the underlying econometric issues. We showed that the convergence rate for the interactive-effects estimator is $\sqrt{NT}$, and this rate holds in spite of correlations and heteroskedasticity in both dimensions. We also derived bias-corrected estimator and estimators under additivity restrictions and their limiting distributions. We further studied the problem of testing additive effects against interactive effects. The interactive-effects estimator is easy to compute, and both the factor process F_t and the factor loadings λ_i can also be consistently estimated up to a rotation. Under interactive effects, we showed that the grand mean, the coefficients of time-invariant regressors, and the coefficients of common regressors are identifiable and can be consistently estimated.

A useful extension is the large N and large T dynamic panel data model with multiple interactive effects. The argument for consistency and rate of convergence remains the same, but the asymptotic bias will take a different form. Another broad extension is nonstationary panel data analysis, particularly panel data cointegration, a subject that recently attracted considerable attention. In this setup, X_{it} is a vector of an integrated variable and F_t can be either integrated or stationary. When F_t is integrated, then Y_{it}, X_{it}, and F_t are cointegrated. Neglecting F_t is equivalent to spurious regression and the estimation of β will not be consistent. However, the interactive-effect approach can be applied by jointly estimating the unobserved common stochastic trends F_t and the model coefficients, leading to consistent estimation. Finally, the models introduced in Section 10 (see Discussion) warrant further investigation.

APPENDIX A: PROOFS

We use the following facts throughout: $T^{-1}\|X_i\|^2 = T^{-1}\sum_{t=1}^{T}\|X_{it}\|^2 = O_p(1)$ or $T^{-1/2}\|X_i\| = O_p(1)$. Averaging over i, $(TN)^{-1}\sum_{i=1}^{N}\|X_i\|^2 = O_p(1)$. Similarly, $T^{-1/2}\|F^0\| = O_p(1)$, $T^{-1}\|\hat{F}\|^2 = r$, $T^{-1/2}\|\hat{F}\| = \sqrt{r}$, $T^{-1}\|X_i'F^0\| = O_p(1)$, and so forth. Throughout, we define $\delta_{NT} = \min[\sqrt{N}, \sqrt{T}]$ so that $\delta_{NT}^2 = \min[N, T]$. The proofs of the lemmas are given in the Supplemental material.

LEMMA A.1: *Under Assumptions A–D*,

$$\sup_F \left\| \frac{1}{NT}\sum_{i=1}^{N} X_i' M_F \varepsilon_i \right\| = o_p(1),$$

$$\sup_F \left\| \frac{1}{NT}\sum_{i=1}^{N} \lambda_i' F^{0\prime} M_F \varepsilon_i \right\| = o_p(1),$$

$$\sup_F \left\| \frac{1}{NT}\sum_{i=1}^{N} \varepsilon_i' P_F \varepsilon_i \right\| = o_p(1),$$

where the sup *is taken with respect to* F *such that* $F'F/T = I$.

PROOF OF PROPOSITION 1: Without loss of generality, assume $\beta^0 = 0$ (purely for notational simplicity). From $Y_i = X_i\beta^0 + F^0\lambda_i + \varepsilon_i = F^0\lambda_i + \varepsilon_i$, expanding $S_{NT}(\beta, F)$, we obtain

$$S_{NT}(\beta, F) = \tilde{S}_{NT}(\beta, F) + 2\beta' \frac{1}{NT}\sum_{i=1}^{N} X_i' M_F \varepsilon_i + 2\frac{1}{NT}\sum_{i=1}^{N} \lambda_i' F^{0\prime} M_F \varepsilon_i + \frac{1}{NT}\sum_{i=1}^{N} \varepsilon_i'(P_F - P_{F^0})\varepsilon_i,$$

where

$$\tilde{S}_{NT}(\beta, F) = \beta'\left(\frac{1}{NT}\sum_{i=1}^{N} X_i' M_F X_i\right)\beta + \operatorname{tr}\left[\left(\frac{F^{0\prime} M_F F^0}{T}\right)\left(\frac{\Lambda'\Lambda}{N}\right)\right] + 2\beta'\frac{1}{NT}\sum_{i=1}^{N} X_i' M_F F^0 \lambda_i. \tag{38}$$

By Lemma A.1,

$$S_{NT}(\beta, F) = \tilde{S}_{NT}(\beta, F) + o_p(1) \tag{39}$$

uniformly over bounded β and over F such that $F'F/T = I$. Bounded β is in fact not necessary because the objective function is quadratic in β (that is, it is easy to argue that the objective function cannot achieve its minimum for very large β).

Clearly, $\tilde{S}_{NT}(\beta^0, F^0H) = 0$ for any $r \times r$ invertible H, because $M_{F^0H} = M_{F^0}$ and $M_{F^0}F^0 = 0$. The identification restrictions implicitly fix an H. We next show that for any $(\beta, F) \neq (\beta^0, F^0H)$, $\tilde{S}_{NT}(\beta, F) > 0$; thus, $\tilde{S}_{NT}(\beta, F)$ attains its unique minimum value 0 at $(\beta^0, F^0H) = (0, F^0H)$. Define

$$A = \frac{1}{NT}\sum_{i=1}^{N} X_i' M_F X_i, \quad B = \left(\frac{\Lambda'\Lambda}{N} \otimes I_T\right),$$

$$C = \frac{1}{NT}\sum_{i=1}^{N} (\lambda_i' \otimes M_F X_i),$$

and let $\eta = \operatorname{vec}(M_F F^0)$. Then

$$\tilde{S}_{NT}(\beta, F) = \beta' A\beta + \eta' B\eta + 2\beta' C'\eta.$$

Completing the square, we have

$$\begin{aligned}\tilde{S}_{NT}(\beta, F) &= \beta'(A - C'B^{-1}C)\beta + (\eta' + \beta'CB^{-1})B(\eta + B^{-1}C\beta) \\ &= \beta' D(F)\beta + \theta' B\theta,\end{aligned}$$

where $\theta = (\eta + B^{-1}C\beta)$. By Assumption A, $D(F)$ is positive definite and B is also positive definite, so $\tilde{S}_{NT}(\beta, F) \geq 0$. In addition, if either $\beta \neq \beta^0 = 0$ or $F \neq F^0 H$, then $\tilde{S}_{NT}(\beta, F) > 0$. Thus, $\tilde{S}_{NT}(\theta, F)$ achieves its unique minimum at $(\beta^0, F^0 H)$. Further, for $\|\beta\| \geq c > 0$, $\tilde{S}_{NT}(\beta, F) \geq \rho_{\min} c^2 > 0$, where $\rho_{\min}$ is the minimum eigenvalue of the positive definite matrix $\inf_F D(F)$. This implies that $\hat{\beta}$ is consistent for $\beta^0 = 0$. However, we cannot deduce that $\hat{F}$ is consistent for $F^0 H$. This is because F^0 is $T \times r$ and as $T \to \infty$, the number of elements goes to infinity, so the usual consistency is not well defined. Other notions of consistency will be examined.

To prove part (ii), note that the centered objective function satisfies $S_{NT}(\beta^0, F^0) = 0$ and, by definition, $S_{NT}(\hat{\beta}, \hat{F}) \leq 0$. Therefore, in view of (39),

$$0 \geq S_{NT}(\hat{\beta}, \hat{F}) = \tilde{S}_{NT}(\hat{\beta}, \hat{F}) + o_p(1).$$

Combined with $\tilde{S}_{NT}(\hat{\beta}, \hat{F}) \geq 0$, it must be true that

$$\tilde{S}_{NT}(\hat{\beta}, \hat{F}) = o_p(1).$$

From $\hat{\beta} \xrightarrow{p} \beta^0 = 0$ and (38), it follows that the above implies

$$\operatorname{tr}\left[\frac{F^{0\prime} M_{\hat{F}} F^0}{T} \frac{\Lambda'\Lambda}{N}\right] = o_p(1).$$

Because $\Lambda'\Lambda/N > 0$ and $(F^{0\prime} M_{\hat{F}} F^0)/T \geq 0$, the above implies the latter matrix is $o_p(1)$, that is,

$$\frac{F^{0\prime} M_{\hat{F}} F^0}{T} = \frac{F^{0\prime} F^0}{T} - \frac{F^{0\prime} \hat{F}}{T} \frac{\hat{F}' F^0}{T} = o_p(1). \tag{40}$$

By Assumption B, $F^{0\prime} F^0/T$ is invertible, so it follows that $F^{0\prime} \hat{F}/T$ is invertible. Next,

$$\|P_{\hat{F}} - P_{F^0}\|^2 = \operatorname{tr}[(P_{\hat{F}} - P_{F^0})^2] = 2\operatorname{tr}(I_r - \hat{F}' P_{F^0} \hat{F}/T).$$

But (40) implies $\hat{F}' P_{F^0} \hat{F}/T \xrightarrow{p} I_r$, which is equivalent to $\|P_{\hat{F}} - P_{F^0}\| \xrightarrow{p} 0$.

Q.E.D.

Note that for any positive definite matrices, A and B, the eigenvalues of AB are the same as those of BA, $A^{1/2}BA^{1/2}$, and so forth; therefore, all eigenvalues are positive. In all remaining proofs, β and β^0 are used interchangeably, and so are F and F^0.

PROPOSITION A.1: *Under Assumptions A–D, we can make the following statements*:

(i) *V_{NT} is invertible and $V_{NT} \xrightarrow{p} V$, where V $(r \times r)$ is a diagonal matrix consisting of the eigenvalues of $\Sigma_\Lambda \Sigma_F$; V_{NT} is defined in* (12).

(ii) *Let $H = (\Lambda'\Lambda/N)(F^{0\prime}\hat{F}/T)V_{NT}^{-1}$. Then H is an $r \times r$ invertible matrix and*

$$\frac{1}{T}\|\hat{F} - F^0 H\|^2 = \frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H'F_t^0\|^2$$
$$= O_p(\|\hat{\beta} - \beta\|^2) + O_p\left(\frac{1}{\min[N, T]}\right).$$

PROOF: From

$$\left[\frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'\right]\hat{F} = \hat{F}V_{NT}$$

and $Y_i - X_i\hat{\beta} = X_i(\beta - \hat{\beta}) + F^0\lambda_i + \varepsilon_i$, by expanding terms, we obtain

$$\begin{aligned}
\hat{F}V_{NT} &= \frac{1}{NT}\sum_{i=1}^{N} X_i(\beta - \hat{\beta})(\beta - \hat{\beta})'X_i'\hat{F} + \frac{1}{NT}\sum_{i=1}^{N} X_i(\beta - \hat{\beta})\lambda_i'F^{0\prime}\hat{F} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N} X_i(\beta - \hat{\beta})\varepsilon_i'\hat{F} + \frac{1}{NT}\sum_{i=1}^{N} F^0\lambda_i(\beta - \hat{\beta})'X_i'\hat{F} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N} \varepsilon_i(\beta - \hat{\beta})'X_i'\hat{F} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N} F^0\lambda_i\varepsilon_i'\hat{F} + \frac{1}{NT}\sum_{i=1}^{N} \varepsilon_i\lambda_i'F^{0\prime}\hat{F} + \frac{1}{NT}\sum_{i=1}^{N} \varepsilon_i\varepsilon_i'\hat{F} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N} F^0\lambda_i\lambda_i'F^{0\prime}\hat{F} \\
&= I1 + \cdots + I9.
\end{aligned}$$

The last term on the right is equal to $F^0(\Lambda'\Lambda/N)(F^{0\prime}\hat{F}/T)$. Letting $I1, \ldots, I8$ denote the eight terms on the right, the above can be rewritten as

$$\hat{F}V_{NT} - F^0(\Lambda'\Lambda/N)(F^{0\prime}\hat{F}/T) = I1 + \cdots + I8. \tag{41}$$

Multiplying $(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$ on each side of (41), we obtain

$$\begin{aligned}\hat{F}[V_{NT}(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}] - F^0 \\ = (I1 + \cdots + I8)(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}.\end{aligned} \tag{42}$$

Note that the matrix $V_{NT}(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$ is equal to H^{-1}, but the invertibility of V_{NT} is not proved yet. We have

$$\begin{aligned}&T^{-1/2}\big\|\hat{F}[V_{NT}(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}] - F^0\big\| \\ &\quad \le T^{-1/2}(\|I1\| + \cdots + \|I8\|) \cdot \|(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}\|.\end{aligned}$$

Consider each term on the right. For the first term, note that $T^{-1/2}\|\hat{F}\| = \sqrt{r}$ and

$$\begin{aligned}T^{-1/2}\|I1\| &\le \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\|X_i\|^2}{T}\right)\|\hat{\beta} - \beta\|^2\sqrt{r} \\ &= O_p(\|\hat{\beta} - \beta\|^2) = o_p(\|\hat{\beta} - \beta\|)\end{aligned}$$

because $\|\hat{\beta} - \beta\| = o_p(1)$. Using the same argument, it is easy to prove that next four terms ($I2$–$I5$) are each $O_p(\hat{\beta} - \beta)$. The last three terms do not explicitly depend on $\hat{\beta} - \beta$ and they have the same expressions as those in Bai and Ng (2002). Each of these terms is $O_p(1/\min[\sqrt{N}, \sqrt{T}])$, which was proved in Bai and Ng (2002, Theorem 1). The proof there only uses the property that $\hat{F}'\hat{F}/T = I$ and the assumptions on ε_i; thus the proof needs no modification. In summary, we have

$$\begin{aligned}T^{-1/2}\|\hat{F}V_{NT}(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1} - F^0\| \\ = O_p(\|\hat{\beta} - \beta\|) + O_p(1/\min[\sqrt{N}, \sqrt{T}]).\end{aligned} \tag{43}$$

(i) Left multiplying (41) by $\hat{F}'$ and using $\hat{F}'\hat{F} = T$, we have

$$V_{NT} - (\hat{F}'F^0/T)(\Lambda'\Lambda/N)(F^{0\prime}\hat{F}/T) = T^{-1}\hat{F}'(I1 + \cdots + I8) = o_p(1)$$

because $T^{-1/2}\|\hat{F}\| = \sqrt{r}$ and $T^{-1/2}\|(I1 + \cdots + I8)\| = o_p(1)$. Thus

$$V_{NT} = (\hat{F}'F^0/T)(\Lambda'\Lambda/N)(F^{0\prime}\hat{F}/T) + o_p(1).$$

Proposition 1 shows that $\hat{F}'F^0/T$ is invertible; thus V_{NT} is invertible. To obtain the limit of V_{NT}, left multiply (41) by $F^{0\prime}$ and then divide by T to yield

$$(F^{0\prime}F^0/T)(\Lambda'\Lambda/N)(F^{0\prime}\hat{F}/T) + o_p(1) = (F^{0\prime}\hat{F}/T)V_{NT}$$

because $T^{-1}F^{0\prime}(I1+\cdots+I8)=o_p(1)$. The above equality shows that the columns of $F^{0\prime}\hat{F}/T$ are the (nonnormalized) eigenvectors of the matrix $(F^{0\prime}F^0/T)(\Lambda'\Lambda/N)$, and V_{NT} consists of the eigenvalues of the same matrix (in the limit). Thus $V_{NT}\xrightarrow{p}V$, where V is $r\times r$, consisting of the r eigenvalues of the matrix $\Sigma_F\Sigma_\Lambda$.

(ii) Since V_{NT} is invertible, the left-hand side of (43) can be written as $T^{-1/2}\|\hat{F}H^{-1}-F^0\|$; thus (43) is equivalent to

$$T^{-1/2}\|\hat{F}-F^0H\|=O_p(\|\hat{\beta}-\beta\|)+O_p(1/\min[\sqrt{N},\sqrt{T}]).$$

Taking squares on each side gives part (ii). Note that the cross-product term from expanding the square has the same bound.

Q.E.D.

The proofs for the next four lemmas are given in the Supplemental Material.

LEMMA A.2: *Under Assumptions A–C, there exists an $M<\infty$, such that statements* (i) *and* (ii) *hold*:

(i) *We have*

$$E\left\|N^{-1/2}\sum_{k=1}^N\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T F_sF_t'[\varepsilon_{kt}\varepsilon_{ks}-E(\varepsilon_{kt}\varepsilon_{ks})]\right\|^2\leq M.$$

(ii) *For all $i=1,2,\dots,N$ and $h=1,2,\dots,r$, we have*

$$E\left\|N^{-1/2}\sum_{k=1}^N\frac{1}{T}\left\{\sum_{t=1}^T\sum_{s=1}^T X_{it}[\varepsilon_{kt}\varepsilon_{ks}-E(\varepsilon_{kt}\varepsilon_{ks})]F_{hs}\right\}\right\|^2\leq M.$$

LEMMA A.3: *Under Assumptions A–D, we have four equalities*:

(i) $T^{-1}F^{0\prime}(\hat{F}-F^0H)=O_p(\hat{\beta}-\beta)+O_p(\delta_{NT}^{-2})$.

(ii) $T^{-1}\hat{F}'(\hat{F}-F^0H)=O_p(\hat{\beta}-\beta)+O_p(\delta_{NT}^{-2})$.

(iii) $T^{-1}X_k'(\hat{F}-F^0H)=O_p(\hat{\beta}-\beta)+O_p(\delta_{NT}^{-2})$ *for each* $k=1,2,\dots,N$.

(iv) $\frac{1}{NT}\sum_{i=1}^N X_i'M_{\hat{F}}(\hat{F}-F^0H)=O_p(\hat{\beta}-\beta)+O_p(\delta_{NT}^{-2})$.

LEMMA A.4: *Under Assumptions A–D, we also have four equalities*:

(i) $T^{-1}\varepsilon_k'(\hat{F}-F^0H)=T^{-1/2}O_p(\hat{\beta}-\beta)+O_p(\delta_{NT}^{-2})$ *for each* k.

(ii) $\frac{1}{T\sqrt{N}}\sum_{k=1}^N\varepsilon_k'(\hat{F}-F^0H)=T^{-1/2}O_p(\hat{\beta}-\beta)+N^{-1/2}O_p(\hat{\beta}-\beta)+O_p(N^{-1/2})+O_p(\delta_{NT}^{-2})$.

(iii) $\frac{1}{NT}\sum_{k=1}^N\lambda_k'(\hat{F}H^{-1}-F^0)'\varepsilon_k=(NT)^{-1/2}O_p(\hat{\beta}-\beta)+O_p(N^{-1})+N^{-1/2}\times O_p(\delta_{NT}^{-2})$.

(iv) $\frac{1}{NT}\sum_{k=1}^{N}(X_k' F^0/T)(F^{0\prime}F^0/T)(\hat{F}H^{-1}-F^0)'\varepsilon_k = (1/N^2)\sum_{i=1}^{N}\sum_{k=1}^{N}(X_k' \times F^0/T)(F^{0\prime}F^0/T)(\Lambda'\Lambda/N)^{-1}\lambda_i(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}) + (NT)^{-1/2}O_p(\hat{\beta}-\beta) + N^{-1/2}\times O_p(\delta_{NT}^{-2})$.

LEMMA A.5: *Let* $G = (F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$. *Under Assumptions* A–D, *we have*

$$\begin{aligned}
&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i' M_{\hat{F}}(\varepsilon_k\varepsilon_k' - \Omega_k)\hat{F}G\lambda_i \\
&\quad = O_p\left(\frac{1}{T\sqrt{N}}\right) + (NT)^{-1/2}[O_p(\hat{\beta}-\beta) + O_p(\delta_{NT}^{-1})] \\
&\qquad + \frac{1}{\sqrt{N}}O_p(\|\hat{\beta}-\beta\|^2) + \frac{1}{\sqrt{N}}O_p(\delta_{NT}^{-2}).
\end{aligned}$$

PROPOSITION A.2: *Assume Assumptions* A–D *hold*. *If* $T/N^2 \to 0$, *then*

$$\begin{aligned}
\sqrt{NT}(\hat{\beta}-\beta^0) = {}& D(\hat{F})^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i' M_{\hat{F}} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_{\hat{F}}\right]\varepsilon_i \\
& + \sqrt{\frac{N}{T}}\zeta_{NT} + o_p(1),
\end{aligned}$$

where $a_{ik} = \lambda_i'(\Lambda'\Lambda/N)\lambda_k$ *and*

$$\zeta_{NT} = -D(\hat{F})^{-1}\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}}\Omega\hat{F}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i \tag{44}$$

with $\Omega = \frac{1}{N}\sum_{k=1}^{N}\Omega_k$ *and* $\Omega_k = E(\varepsilon_k\varepsilon_k')$.

PROOF: From $Y_i = X_i\beta^0 + F^0\lambda_i + \varepsilon_i$,

$$\begin{aligned}
\hat{\beta}-\beta^0 = {}& \left(\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)^{-1}\sum_{i=1}^{N} X_i' M_{\hat{F}} F^0\lambda_i \\
& + \left(\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)^{-1}\sum_{i=1}^{N} X_i' M_{\hat{F}}\varepsilon_i
\end{aligned}$$

or

$$\left(\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)(\hat{\beta}-\beta) = \frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} F^0\lambda_i + \frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}}\varepsilon_i. \tag{45}$$

In view of $M_{\hat{F}}\hat{F}=0$, we have $M_{\hat{F}}F^0=M_{\hat{F}}(F^0-\hat{F}A)$ for any A. Choosing $A=H^{-1}$, from (42), we get

$$F^0-\hat{F}H^{-1}=-[I1+\cdots+I8](F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}.$$

It follows that

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}X_i'M_{\hat{F}}F^0\lambda_i\\
&\quad=-\frac{1}{NT}\sum_{i=1}^{N}X_i'M_{\hat{F}}[I1+\cdots+I8]\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i\\
&\quad=J1+\cdots+J8,
\end{aligned}$$

where $J1$–$J8$ are implicitly defined vis-à-vis $I1$–$I8$. For example,

$$J1=-\frac{1}{NT}\sum_{i=1}^{N}X_i'M_{\hat{F}}(I1)\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i.$$

Term $J1$ is bounded in norm by $O_p(1)\|\hat{\beta}-\beta\|^2$ and thus $J1=o_p(1)(\hat{\beta}-\beta)$. Consider

$$\begin{aligned}
J2&=-\frac{1}{N^2T}\sum_{i=1}^{N}X_i'M_{\hat{F}}\left[\sum_{k=1}^{N}X_k(\beta-\hat{\beta})\lambda_k'\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\right]\lambda_i\\
&=\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}(X_i'M_{\hat{F}}X_k)\left[\lambda_k'\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i\right](\hat{\beta}-\beta)\\
&=\frac{1}{T}\left[\frac{1}{N}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}X_i'M_{\hat{F}}X_k a_{ik}\right](\hat{\beta}-\beta),
\end{aligned}$$

where $a_{ik}=\lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$ is a scalar and thus commutable with $\hat{\beta}-\beta$. Now consider

$$J3=\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}X_i'M_{\hat{F}}X_k\left(\frac{\varepsilon_k'\hat{F}}{T}\right)\left(\frac{\hat{F}'F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i(\hat{\beta}-\beta).$$

Writing $\varepsilon_k'\hat{F}/T=\varepsilon_k'F^0H/T+\varepsilon_k'(\hat{F}-F^0H)/T=O_p(T^{-1/2})+O_p(\hat{\beta}-\beta)+O_p(1/\min[\sqrt{N},\sqrt{T}])$, by Lemma A.4, it is easy to see that $J3=o_p(1)(\hat{\beta}-\beta)$.

Next

$$J4 = -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i' M_{\hat F} F^0 \lambda_k (\beta - \hat\beta)'\left(\frac{X_k'\hat F}{T}\right) \times \left(\frac{\hat F' F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i.$$

Writing $M_{\hat F}F^0 = M_{\hat F}(F^0 - \hat F H^{-1})$ and using that $T^{-1/2}\|F^0 - \hat F H^{-1}\|$ is small, then $J4$ is equal to $o_p(1)(\hat\beta - \beta)$. It is easy to show $J5 = o_p(1)(\hat\beta - \beta)$ and thus it is omitted.

The last three terms $J6$–$J8$ do not explicitly depend on $\hat\beta - \beta$. Only term $J7$ contributes to the limiting distribution of $\hat\beta - \beta$; the other two terms are $o_p((NT)^{-1/2})$ plus $o_p(\hat\beta - \beta)$. We shall establish these claims. Consider

$$J6 = -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i' M_{\hat F} F^0 \lambda_k \left(\frac{\varepsilon_k'\hat F}{T}\right)\left(\frac{\hat F' F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i.$$

Denote $G = (\hat F' F^0/T)^{-1}(\Lambda'\Lambda/N)^{-1}$ for the moment: it is a matrix of fixed dimension and does not vary with i. Using $M_{\hat F}F^0 = M_{\hat F}(F^0 - \hat F H^{-1})$, we can write

$$J6 = -\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat F}(F^0 - \hat F H^{-1})\left(\frac{1}{N}\sum_{k=1}^{N}\lambda_k\left(\frac{\varepsilon_k'\hat F}{T}\right)\right) G\lambda_i.$$

Now

$$\begin{aligned}
\frac{1}{NT}\sum_{k=1}^{N}\lambda_k\varepsilon_k'\hat F &= \frac{1}{NT}\sum_{k=1}^{N}\lambda_k\varepsilon_k' F^0 H + \frac{1}{NT}\sum_{k=1}^{N}\lambda_k\varepsilon_k'(\hat F - F^0 H)\\
&= O_p\left(\frac{1}{\sqrt{NT}}\right) + (NT)^{-1/2}O_p(\hat\beta - \beta)\\
&\quad + O_p(N^{-1}) + N^{-1/2}O_p(\delta_{NT}^{-2})\\
&= O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(N^{-1}) + N^{-1/2}O_p(\delta_{NT}^{-2})
\end{aligned}$$

by Lemma A.4(iii). The last equality is because $(NT)^{-1/2}$ dominates $(NT)^{-1/2}\times(\hat\beta - \beta)$. Furthermore, by Lemma A.3, $\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat F}(\hat F - F^0 H)\lambda_{i\ell} = O_p(\hat\beta - \beta) + O_p(\delta_{NT}^{-2})$ for $\ell = 1, 2, \ldots, r$, and noting G does not depend on i and

$\|G\| = O_p(1)$, we have

$$\begin{aligned}
J6 &= [O_p(\hat{\beta} - \beta) + O_p(\delta_{NT}^{-2})] \\
&\quad \times \left[O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(N^{-1}) + N^{-1/2} O_p(\delta_{NT}^{-2}) \right] \\
&= o_p(\hat{\beta} - \beta) + o_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(\delta_{NT}^{-2}) N^{-1} \\
&\quad + N^{-1/2} O_p(\delta_{NT}^{-4}).
\end{aligned}$$

The term $J7$ is simply

$$\begin{aligned}
J7 &= -\frac{1}{N^2 T} \sum_{i=1}^{N} X_i' M_{\hat{F}} \left[\sum_{k=1}^{N} \varepsilon_k \lambda_k' \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \right] \lambda_i \\
&= -\frac{1}{N^2 T} \sum_{i=1}^{N} \sum_{k=1}^{N} a_{ik} X_i' M_{\hat{F}} \varepsilon_k.
\end{aligned}$$

Next consider $J8$, which has the expression

$$J8 = -\frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \varepsilon_k' \hat{F} \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1} \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \lambda_i.$$

Let $E(\varepsilon_k \varepsilon_k') = \Omega_k$ $(T \times T)$. Denoting $G = (F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$ and $\|G\| = O_p(1)$, and rewriting gives

$$\begin{aligned}
J8 &= -\frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{k=1}^{N} X_i' M_{\hat{F}} \Omega_k \hat{F} G \lambda_i \\
&\quad - \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{k=1}^{N} X_i' M_{\hat{F}} (\varepsilon_k \varepsilon_k' - \Omega_k) \hat{F} G \lambda_i.
\end{aligned} \tag{46}$$

Denote the first term on the right by A_{NT}. By Lemma A.5, we have

$$\begin{aligned}
J8 &= A_{NT} + O_p\left(\frac{1}{T\sqrt{N}}\right) + \frac{1}{\sqrt{NT}} [O_p(\hat{\beta} - \beta) + O_p(\delta_{NT}^{-1})] \\
&\quad - \frac{1}{\sqrt{N}} O_p(\|\hat{\beta} - \beta\|^2) + \frac{1}{\sqrt{N}} O_p(\delta_{NT}^{-2}).
\end{aligned}$$

Collecting terms from $J1$ to $J8$ with dominated terms ignored gives

$$\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} F^0 \lambda_i = J2 + J7 + A_{NT} + o_p(\hat{\beta}-\beta) + o_p\big((NT)^{-1/2}\big) + O_p\left(\frac{1}{T\sqrt{N}}\right) + N^{-1/2} O_p(\delta_{NT}^{-2}).$$

Thus,

$$\begin{aligned}
&\left(\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i + o_p(1)\right)(\hat{\beta}-\beta) - J2 \\
&\quad = \frac{1}{NT}\sum_{i=1}^{N} X_i M_{\hat{F}} \varepsilon_i + J7 + A_{NT} \\
&\qquad + o_p\big((NT)^{-1/2}\big) + O_p\left(\frac{1}{T\sqrt{N}}\right) + N^{-1/2} O_p(\delta_{NT}^{-2}).
\end{aligned}$$

Combining terms and multiplying by $\sqrt{NT}$ yields

$$\begin{aligned}
&[D(\hat{F}) + o_p(1)]\sqrt{NT}(\hat{\beta}-\beta) \\
&\quad = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i' M_{\hat{F}} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_{\hat{F}}\right]\varepsilon_i + \sqrt{NT} A_{NT} \\
&\qquad + o_p(1) + O_p\big(T^{-1/2}\big) + T^{1/2} O_p(\delta_{NT}^{-2}).
\end{aligned}$$

Thus, if $T/N^2 \to 0$, the last term is also $o_p(1)$. Multiply $D(\hat{F})^{-1}$ on each side of the above and note that $D(\hat{F})^{-1}\sqrt{NT} A_{NT} = \sqrt{N/T}\zeta_{NT}$. Finally, $D(\hat{F})^{-1}[D(\hat{F}) + o_p(1)]^{-1} = I + o_p(1)$, so we have proved the proposition. *Q.E.D.*

LEMMA A.6: *Under Assumptions A–D*, $\zeta_{NT} = O_p(1)$, *where* ζ_{NT} *is given in Proposition A.2.*

LEMMA A.7: *Under Assumptions A–D, we have the following equalities*:

(i) $HH' = (F^{0\prime}F^0/T)^{-1} + O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-2})$.

(ii) $\|P_{\hat{F}} - P_{F^0}\|^2 = O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-2})$.

Proposition A.2 still involves estimated F. To replace $\hat{F}$ by F^0, we need some preliminary results.

LEMMA A.8: *Under Assumptions A–D*,

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i' M_{\hat F} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_{\hat F}\right]\varepsilon_i \\
&\quad = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i' M_{F^0} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_{F^0}\right]\varepsilon_i \\
&\qquad + \left(\sqrt{\frac{T}{N}}\right)\xi_{NT}^{\dagger} + \sqrt{T}O_p(\|\hat\beta-\beta^0\|^2) \\
&\qquad + O_p(\|\hat\beta-\beta^0\|) + \sqrt{T}O_p(\delta_{NT}^{-2}),
\end{aligned}$$

where

$$\begin{aligned}
\xi_{NT}^{\dagger} &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{(X_i - V_i)' F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1} \\
&\quad \times \left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_k\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}\right) = O_p(1).
\end{aligned} \tag{47}$$

Combining Proposition A.2 and Lemma A.8 and noting that $\sqrt{T}O_p(\|\hat\beta - \beta^0\|^2) + O_p(\|\hat\beta-\beta^0\|)$ is dominated by $\sqrt{NT}(\hat\beta-\beta^0)$ and $\sqrt{T}O_p(\delta_{NT}^{-2}) = o_p(1)$ if $T/N^2 \to 0$, we have an additional statement:

COROLLARY A.1: *Under Assumptions A–D and as* $T/N^2 \to 0$,

$$\begin{aligned}
\sqrt{NT}(\hat\beta-\beta^0) &= D(\hat F)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i' M_{F^0} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_{F^0}\right]\varepsilon_i \\
&\quad + \sqrt{\frac{T}{N}}\xi_{NT} + \sqrt{\frac{N}{T}}\zeta_{NT} + o_p(1),
\end{aligned}$$

where $\xi_{NT} = D(\hat F)^{-1}\xi_{NT}^{\dagger}$, $\xi_{NT}^{\dagger}$ *is defined in* (47) *and* ζ_{NT} *is given in* (44).

PROOF OF THEOREM 1: The assumption implies that both T/N and N/T are $O(1)$. Furthermore, $\zeta_{NT} = O_p(1)$ by Lemma A.6, and $\xi_{NT}^{\dagger}$ and hence ξ_{NT} are $O_p(1)$ by Lemma A.8. The theorem follows from the expression for $\sqrt{NT}(\hat\beta - \beta)$ given in Corollary 1. *Q.E.D.*

LEMMA A.9: *Under Assumptions A–D, the following equalities hold*:

(i) $D(\hat F)^{-1} - D(F^0)^{-1} = o_p(1)$.

(ii) $\sqrt{T/N}[D(\hat{F})^{-1} - D(F^0)^{-1}] = o_p(1)$ *if* $T/N^2 \to 0$.
(iii) $\sqrt{N/T}[D(\hat{F})^{-1} - D(F^0)^{-1}] = o_p(1)$ *if* $N/T^2 \to 0$.
(iv) $\sqrt{T/N}(\xi_{NT} - B) = o_p(1)$ *if* $T/N^2 \to 0$, *where* B *is given in* (18).
(v) $\sqrt{N/T}(\zeta_{NT} - C) = o_p(1)$ *if* $N/T^2 \to 0$, *where* C *is given in* (19).

PROPOSITION A.3: *Under Assumptions* A–D, *if* $T/N^2 \to 0$ *and* $N/T^2 \to 0$, *then*

$$\sqrt{NT}(\hat{\beta} - \beta^0) = D(F^0)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left[X_i' M_{F^0} - \frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k' M_{F^0} \right] \varepsilon_i + \sqrt{\frac{T}{N}} B + \sqrt{\frac{N}{T}} C + o_p(1),$$

where B *and* C *are in* (18) *and* (19), *respectively*.

In this representation, the left-hand side involves no estimated quantities.

PROOF OF PROPOSITION A.3: From Lemma A.9(i), $D(\hat{F})^{-1} = D(F^0)^{-1} + o_p(1)$, the matrix $D(\hat{F})^{-1}$ in Corollary 1 can be replaced by $D(F^0)$ since $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [X_i' M_{F^0} - \frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k' M_{F^0}] \varepsilon_i = O_p(1)$. Parts (iv) and (v) of Lemma A.9 together with Corollary 1 again immediately lead to the proposition. *Q.E.D.*

PROOF OF THEOREM 2: (i) We use the representation in Proposition A.3. Without serial correlation or heteroskedasticity, C in Proposition A.3 is zero. This follows from $\Omega_k = \sigma_k^2 I_T$ and $M_{F^0} \Omega F^0 = (\sum_{i=1}^{N} \sigma_k^2) M_{F^0} F^0 = 0$. Furthermore, $\sqrt{T/N} B \xrightarrow{p} 0$ since $T/N \to 0$. Thus, by Proposition A.3,

$$\sqrt{NT}(\hat{\beta} - \beta^0) = D(F^0)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i' \varepsilon_i + o_p(1). \tag{48}$$

The limiting distribution now follows from Assumption E.

(ii) The proof again uses the representation in Proposition A.3. From $N/T \to 0$, we have $\sqrt{N/T} C \to 0$. We next argue that $B = 0$ when the cross-section correlation and heteroskedasticity are absent. Recall that $\sigma_{ik,tt} = E(\varepsilon_{it} \varepsilon_{kt})$. By assumption, $\sigma_{ij,tt} = \sigma_t^2$ for $i = j$ and $\sigma_{ij,tt} = 0$ for $i \neq j$. Thus B in (18) is simplified as

$$B = -D(F^0)^{-1} \frac{1}{N} \sum_{i=1}^{N} \frac{(X_i - V_i)' F^0}{T} \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \left(\frac{\Lambda' \Lambda}{N} \right)^{-1} \lambda_i \bar{\sigma}^2,$$

where $\bar{\sigma}^2 = T^{-1}\sum_{t=1}^T \sigma_t^2$. From $V_i = \frac{1}{N}\sum_{k=1}^N X_k a_{ik}$ and that a_{ik} is a scalar, thus commutable with all matrices, we have $\frac{1}{N}\sum_{i=1}^N \lambda_i a_{ik} = \frac{1}{N}\sum_{i=1}^N \lambda_i\lambda_i'(\Lambda'\Lambda/N)^{-1} \times \lambda_k = \lambda_k$. This means that

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N \frac{V_i'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i \\
&\quad = \frac{1}{N}\sum_{k=1}^N \frac{X_k'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_k,
\end{aligned}$$

which implies $B = 0$. Thus (48) holds and the limiting distribution once again follows from Assumption E. *Q.E.D.*

PROOF OF THEOREM 3: This again follows from the representation of Proposition A.3. Under the assumption that $T/N \to \rho$, $\sqrt{T/N}B \xrightarrow{p} \rho^{1/2}B_0$, where B_0 is the probability limit of B, and $\sqrt{N/T}C \xrightarrow{p} \rho^{-1/2}C_0$, where C_0 is the probability limit of C. Combining with Assumption E, we obtain the theorem. *Q.E.D.*

Bias Correction

LEMMA A.10: *Under Assumptions A–D, the following equalities hold*:

(i) $\frac{1}{N}\|\hat{\Lambda}' - H^{-1}\Lambda'\|^2 = \frac{1}{N}\sum_{i=1}^N \|\hat{\lambda}_i - H^{-1}\lambda_i\|^2 = O_p(\|\hat{\beta}-\beta\|^2) + O_p(\delta_{NT}^{-2})$.

(ii) $N^{-1}(\hat{\Lambda}' - H^{-1}\Lambda')\Lambda = O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-2})$.

(iii) $\hat{\Lambda}'\hat{\Lambda}/N - H^{-1}(\Lambda'\Lambda/N)H'^{-1} = O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-2})$.

(iv) $(\hat{\Lambda}'\hat{\Lambda}/N)^{-1} - H'(\Lambda'\Lambda/N)^{-1}H = O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-2})$.

(v) $\frac{1}{N}\sum_{i=1}^N \|\hat{\lambda}_i - H^{-1}\lambda_i\| = O_p(\delta_{NT}^{-1}) + O_p(\|\hat{\beta}-\beta\|)$.

(vi) $\frac{1}{N}\sum_{k=1}^N \|T^{-1/2}X_i\|\|\hat{\lambda}_i - H^{-1}\lambda_i\| = O_p(\delta_{NT}^{-1}) + O_p(\|\hat{\beta}-\beta\|)$.

LEMMA A.11: *Under the assumptions of Theorem 4*, $\sqrt{T/N}(\hat{B} - B) = o_p(1)$.

LEMMA A.12: *Under the assumptions of Theorem 4*, $\sqrt{N/T}(\hat{C} - C) = o_p(1)$.

The proof of Theorem 4 follows from Proposition A.3, Lemma A.11, and Lemma A.12. For the proof of Proposition 2, see the Supplemental Material.

REFERENCES

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333. [1233]

AHN, S. G., Y. H. LEE, AND P. SCHMIDT (2001): "GMM Estimation of Linear Panel Data Models With Time-Varying Individual Effects," *Journal of Econometrics*, 101, 219–255. [1230,1231,1233, 1237,1239,1258]

——— (2006): "Panel Data Models With Multiple Time-Varying Effects," Mimeo, Arizona State University. [1239]

ALTUG, S., AND R. A. MILLER (1990): "Household Choices in Equilibrium," *Econometrica*, 58, 543–570. [1233]

ALVAREZ, J., AND M. ARELLANO (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71, 1121–1159. [1232]

ANDERSON, T. W. (1984): *An Introduction to Multivariate Statistical Analysis*. New York: Wiley. [1236]

ANDERSON, T. W., AND C. HSIAO (1982): "Formulation and Estimation of Dynamic Models With Error Components," *Journal of Econometrics*, 76, 598–606. [1232]

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. Berkeley: University of California Press, 111–150. [1234]

ANDREWS, D. W. K. (2005): "Cross-Section Regression With Common Shocks," *Econometrica*, 73, 1551–1585. [1234]

ARELLANO, M. (2003): *Panel Data Econometrics*. Oxford: Oxford University Press. [1239]

ARELLANO, M., AND J. HAHN (2005): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," Unpublished Manuscript, CEMFI. [1250]

ARELLANO, M., AND B. HONORE (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics*, Vol. 5, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland. [1230]

BAI, J. (1994): "Least Squares Estimation of Shift in Linear Processes," *Journal of Time Series Analysis*, 15, 453–472. [1244]

——— (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–173. [1242,1247,1251]

——— (2009): "Supplement to 'Panel Data Models With Interactive Fixed Effects'," *Econometrica Supplemental Material*, 77, http://www.econometricsociety.org/ecta/Supmat/6135_proofs.pdf. [1232]

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [1235,1267]

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augment Regressions," *Econometrica*, 74, 1133–1150. [1251]

BALTAGI, B. H. (2005): *Econometric Analysis of Panel Data*. Chichester: Wiley. [1239]

BEKKER, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–681. [1250]

BERNANKE, B., J. BOIVIN, AND P. ELIASZ (2005): "Factor Augmented Vector Autoregression and the Analysis of Monetray Policy," *Quarterly Journal of Economics*, 120, 387–422. [1239]

CAMPBELL, J. Y., A. W. LO, AND A. C. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press. [1234]

CARNEIRO, P., K. T. HANSEN, AND J. J. HECKMAN (2003): "Estimating Distributions of Treatment Effects With an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," Working Paper 9546, NBER; *International Economic Review*, 44, 362–422. [1233]

CAWLEY, J., K. CONNELLY, J. HECKMAN, AND E. VYTLACIL (1997): "Cognitive Ability, Wages, and Meritocracy," in *Intelligence Genes, and Success: Scientists Respond to the Bell Curve*, ed. by B. Devlin, S. E. Feinberg, D. Resnick, and K. Roeder. Berlin: Springer-Verlag, 179–192. [1233]

CHAMBERLAIN, G. (1980): "Analysis of Covariance With Qualitative Data," *Review of Economic Studies*, 47, 225–238. [1232]

——— (1984): "Panel Data," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. Intriligator. Amsterdam: North-Holland. [1230,1233,1239]

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1281–1304. [1232,1235,1238, 1244]

COAKLEY, J., A. FUERTES, AND R. P. SMITH (2002): "A Principal Components Approach to Cross-Section Dependence in Panels," Mimeo, Birkbeck College, University of London. [1231]

CONNOR, G., AND R. KORAJZCYK (1986): "Performance Measurement With the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373–394. [1235,1236,1239]

FAMA, E., AND K. FRENCH (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56. [1234]

FELDSTEIN, M., AND C. HORIOKA (1980): "Domestic Savings and International Capital Flows," *Economic Journal*, 90, 314–329. [1233]

GIANNONE, D., AND M. LENZA (2005): "The Feldstein Horioka Fact," Unpublished Manuscript, European Central Bank. [1233]

GOLDBERGER, A. S. (1972): "Structural Equations Methods in the Social Sciences," *Econometrica*, 40, 979–1001. [1231]

GREENE, W. (2000): *Econometric Analysis* (Fourth Ed.). Englewood Cliffs, NJ: Prentice-Hall. [1253,1254]

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. [1232,1246,1247,1249]

HAHN, J., AND W. NEWEY (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295–1319. [1232,1249,1250]

HANSEN, C., J. HAUSMAN, AND W. NEWEY (2005): "Estimation With Many Instruments," Unpublished Manuscript, Department of Economics, MIT. [1250]

HAUSMAN, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271. [1256]

HAUSMAN, J. A., AND W. E. TAYLOR (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377–1398. [1258,1260]

HOLTZ-EAKIN, D., W. NEWEY, AND H. ROSEN (1988): "Estimating Vector Autoregressions With Panel Data," *Econometrica*, 56, 1371–1395. [1230,1231,1233,1238,1239]

HSIAO, C. (2003): *Analysis of Panel Data*. New York: Cambridge University Press. [1232]

JÖRESKOG, K. G., AND A. S. GOLDBERGER (1975): "Estimation of a Model With Multiple Indicators and Multiple Causes of a Single Latent Variable," *Journal of the American Statistical Association*, 70, 631–639. [1231]

KIEFER, N. M. (1980): "A Time Series-Cross Section Model With Fixed Effects With an Intertemporal Factor Structure," Unpublished Manuscript, Department of Economics, Cornell University. [1231,1237]

KIVIET, J. (1995): "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, 68, 53–78. [1232]

KNEIP, A., R. SICKLES, AND W. SONG (2005): "A New Panel Data Treatment for Heterogeneity in Time Trends," Unpublished Manuscript, Department of Economics, Rice University. [1231]

LAWLEY, D. N., AND A. E. MAXWELL (1971): *Factor Analysis as a Statistical Method*. London: Butterworth. [1234]

LEE, Y. H. (1991): "Panel Data Models With Multiplicative Individual and Time Effects: Application to Compensation and Frontier Production Functions," Unpublished Ph.D. Dissertation, Michigan State University. [1231]

LETTAU, M., AND S. LUDVIGSON (2001): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time Varying," *Journal of Political Economy*, 109, 1238–1287. [1234]

MACURDY, T. (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 18, 83–114. [1231]

MOON, R., AND B. PERRON (2004): "Testing for a Unit Root in Panels With Dynamic Factors," *Journal of Econometrics*, 122, 81–126. [1231]

MUNDLAK, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69–85. [1239]

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, 2111–2245. [1244]

NEWEY, W., AND R. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likeliood Estimators," *Econometrica*, 72, 219–255. [1239]

NEWEY, W. K., AND K. D. WEST (1987): "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708. [1251,1252]

NEYMAN, J., AND E. L. SCOTT (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32. [1232]

NICKELL, S. (1981): "Biases in Dynamic Models With Fixed Effects," *Econometrica*, 49, 1417–1426. [1232]

OBSTFELD, M., AND K. ROGOFF (2000): "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" *NBER Macroeconomic Annual*, 15, 339–390. [1233]

PESARAN, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels With a Multifactor Error Structure," *Econometrica*, 74, 967–1012. [1231,1233,1240]

PHILLIPS, P. C. B., AND D. SUL (2003): "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence," *The Econometrics Journal*, 6, 217–259. [1231]

ROSS, S. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360. [1234]

SARGAN, J. D. (1964): "Wages and Prices in the United Kingdom: A Study in Ecnometrics Methodology," in *Econometric Analysis of National Economic Planning*, ed. by P. G. Hart, G. Mill, and J. K. Whitaker. London: Butterworths, 25–54. [1237]

STOCK, J. H., AND M. W. WATSON (2002): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179. [1235,1236,1239]

TOWNSEND, R. (1994): "Risk and Insurance in Village India," *Econometrica*, 62, 539–592. [1233]

Dept. Economics, New York University, 19 West 4th Street, New York, NY 10012, U.S.A., SEM, Tsinghua University, and CEMA, Central University of Finance and Economics, Beijing, China; jushan.bai@nyu.edu.

Manuscript received October, 2005; final revision received January, 2009.