Panel Data Models with Interactive Fixed Effects

Jushan Bai*

October 31, 2006

Abstract

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, the interactive effects represent unobservable common shocks and their heterogeneous impacts over the cross sections. We consider identification, consistency, and the limiting distribution of the interactive effects estimator. 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.

Key words and phrases: additive effects, interactive effects, factor error structure, bias-corrected estimator, Hausman tests, Time-invariant regressors, common regressors.

*I am 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), and Quantitative Finance and Econometrics at Stern. This work is supported in part by the NSF grants SES-0137084 and SES-0424540.
Email: jushan.bai@nyu.edu; Department of Economics, New York University, New York, NY 10003.

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

$$u_{it} = \lambda_i' F_t + \varepsilon_{it}$$

$$i = 1, 2, ..., N; t = 1, 2, ..., T$$

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. The interest is centered on the inference for the slope coefficient β, although inference for λ_i and F_t will also be discussed.

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, and accordingly, it will be called 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). Indeed, we allow the observable X_{it} to follow

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

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 not entering the Y_{it} equation). Thus X_{it} can be correlated with λ_i alone, or with F_t alone, or 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 having a zero mean, or whether F_t being 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 to be detailed in Section 3 below.

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 $Y_{it} - \bar{Y}_{i.} = (X_{it} - \bar{X}_{i.})'\beta + \lambda_i(F_t - \bar{F}) + \varepsilon_{it} - \bar{\varepsilon}_{i.}$, leaving escaped the interactive effects because $F_t \not\equiv \bar{F}$, where $\bar{Y}_{i.}$, $\bar{X}_{i.}$, and $\bar{\varepsilon}_{i.}$ are averages over time. Thus the within estimator is inconsistent since the unobservables are correlated with the regressors. However, the interactive effect can be eliminated by the quasi-differencing method, as in Holtz-Eakin, Newey, and Rosen (1988).

Recently, Pesaran (2006) propose 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 certain rank condition, which is not guaranteed to be met, depending on data generating processes. Peseran shows $\sqrt{N}$ consistency irrespective of the rank condition, and possible faster rate of convergence when the rank condition does hold. Coakey, Futers, and Smith (2002) propose a two-step estimator. But this estimator is 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) consider the situation of fixed T and note that the least squares method does not give consistent estimator if serial correlation or heteroskedasticity is present in ε_{it}. Then they explore the GMM estimators and show that GMM method that incorporates moments of zero correlation and homoskedasticity is more efficient than the 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 do not consider correlations between the factor errors and the regressors. Similar studies include MaCurdy (1982), who considers random effects type of GLS estimation for fixed T and Phillips and Sul (2003), who consider SUR-GLS estimation for fixed N. Panel unit root tests with factor errors are studied by Moon and Perron (2004). The method of Kneip, Sickles, and Song (2005) assumes F_t is a smooth function of t and estimates F_t by smoothing spline. Given the spline basis, the estimation problem becomes that of ridge regression. Such a setup is useful when the time effect is slowly varying. 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, large T perspective has its own challenges. For example, 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}. Earlier fixed T studies assume iid X_{it} over i, not allowing X_{it} to contain common factors, but permitting X_{it} to be correlated with λ_i. Earlier studies also assume ε_{it} are iid over i and t. 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 also 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.[1] 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 argue 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. We show that the principle of Hausman test is applicable in this context.

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 within and IV 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 testing additive effects versus interactive effects. Section 10 is devoted to time-invariant regressors and those that are common to each cross sectional unit. All proofs are provided either in the appendix or in a separate accompanying document.

[1] See Nickel (1981), Anderson and Hsiao (1982), Kiviet (1995), Hsiao (2003, 71-74), and Alvarez and Arellano (2003) for dynamic panel data models, and Hahn and Newey (2004) for nonlinear panel models.

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), and λ_i represents the heterogeneous impact of 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, say), but also affect the amount of input in the production process (through investment decisions). When common shocks have homogeneous effects on the output, i.e., $\lambda_i = \lambda$ for all i, the model collapses to the usual time effect by letting $\delta_t = \lambda' F_t$, and δ_t is a scalar. It is the heterogeneity that gives rise to a factor structure.

Recently, using a similar model, Giannone and Lenza (2005) provide an explanation for the Feldstein-Horioka puzzle, one of the six puzzles in international macroeconomics (Obstfeld and Rogoff 2000). In open economies, domestic savings and domestic investments are not tied together, with capital flowing to countries with higher returns. But Feldstein and Horioka (1980) find excessively high correlation between domestic savings and domestic investments among OECD countries. In the model of Giannone and Lenza (2005), 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 find that additive effect models as in previous studies will lead to the puzzle, but an interactive effect model largely nullifies the puzzle. They note that additive effect models imply a strong assumption that shocks have homogeneous effects across countries.

Microeconometrics. In earning studies, Y_{it} represents the wage rate for individual i with age (or age cohort) t, 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 hardworking; 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 effect model is obtained by letting $\alpha_i = \lambda_i' f$. It is noted that t in this example is not necessarily the calendar time, but age or age cohort.[2]

In the setup of Holtz-Eakin, Newey, and Rosen (1988), the slope coefficient β is also time varying. The model can be considered as a projection of Y_{it} on $\{X_{it}, \lambda_i\}$; see Chamberlain (1984). Pesaran (2004) allows β to be heterogenous over i such that $\beta_i = \beta + v_i$ with v_i being iid. In this regard, the constant slope coefficient is restrictive. To partially alleviate the restriction, as suggested by a referee, it would be useful to allow additional individual and

[2]The author thanks Frances Kramarz for suggesting this model.

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.

The model in its present form is not appropriate for estimating firm effects since Y_{it} does not reflect the firm in which individual i works. For this purpose, we add a subscript to index the firms. Let Y_{ijt} be the wage for worker i in firm j with experience level t. It is important to note that these indices do not vary independently. In general, $j = J(i,t)$, the firm in which worker i with experience level t works (alternatively, $i = I(j,t)$). Abowd, Kramarz, and Margolis (1999) use such a data set to disentangle the individual effect and the firm effect. Their model may be extended to allow for interactive effects

$$y_{ijt} = \beta' x_{ijt} + \alpha_i + \phi_j + \tau_i' \psi_j + \gamma_j' \xi_t + \pi_i' f_t + \varepsilon_{ijt} \tag{5}$$

where α_i is the individual effect and ϕ_j is the firm effect as in Abowd et al (1999); $\tau_i'\psi_j$ is the worker-firm interaction, capturing the matching effect; $\gamma_j'\xi_t$ represents the firm-experience interaction, and $\pi_i' f_t$ has the usual unmeasured skill-price interaction[3]

A further theoretical motivation for the interactive effects is given by Ahn, Lee, and Schmidt (2001). Let $U_{it}(c_{it}, h_{it})$ be agent i's within-period utility function, a separable function of consumption (c_{it}) and labor hours (h_{it}). Under the setup of Altug and Miller (1990), the first order condition for consumption satisfies

$$\frac{\partial U_{it}}{\partial c_{it}} = \mu_i f_t$$

where μ_i is the Lagrange multiplier associated with agent i's life-time budget constraint and f_t is the price of contingent claims at time t. Assuming that the consumption part of the utility function is of constant absolute risk aversion (CARA), $U_{it} = -(1/\sigma_i)\exp[-\sigma_i(c_{it} - d_{it})]$ with $d_{it} = X_{it}'\beta + \varepsilon_{it}$ (the labor part of the utility is suppressed). The risk aversion parameter σ is heterogenous over i, as in Townsend (1994). The first order condition for consumption, upon taking logarithms, implies

$$c_{it} = X_{it}'\beta + \sigma_i^{-1}\log(\mu_i) + \sigma_i^{-1}\log(f_t) + \varepsilon_{it}$$

Let $Y_{it} = c_{it}$, $\alpha_i = \sigma_i^{-1}\log(\mu_i)$, $\lambda_i = \sigma_i^{-1}$, and $F_t = \log(f_t)$, we obtain (4) with $\delta_t = 0$.

Finance. Here Y_{it} is the excess return of asset i in period t, and X_{it} is a vector of observable factors such as dividend yields, dividend payout ratio, and consumption gap as in

[3]By relabelling the indices, model (5) can be written as the interactive effect model with two indices (i,t) but i will have a more general meaning (it is a combination of workers and firms). Abowd, Kramarz, and Margolis (1999) use two indices: worker i and time t on the left hand side, but firms are indexed by $j = J(i,t)$ on the right hand side. They estimate the fixed effects for workers and firms. In (5), the three interactive effects can also be written as $\lambda_i' F_t$. But in this case, there are many restrictions on λ_i and on F_t. Efficient estimation should take into account these restrictions.

Lettau and Ludvigson (2001) or book and size factors as in Fama and French (1993); F_t is a vector of unobservable factor returns, and λ_i is the factor loading and ε_{it} is the idiosyncratic return. Arbitrage Pricing Theory of Ross (1976) is built upon a factor model for asset returns. Campbell, Lo, and MacKinlay (1997) provide many applications of factor models in finance.

Cross-section correlation. Interactive effects model provides a tractable way of modelling 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 iid over i and t, an equal correlation model is obtained. In a recent paper, Andrews (2005) shows 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 model 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 the factors are treated as parameters, as opposed to the factor loadings only. Second, the number of variables N is assumed to grow without bound instead of 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 $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_N)'$, an $N \times r$ matrix. In matrix notation

$$Y = X\beta + F\Lambda' + \varepsilon \tag{6}$$

where $Y = (Y_1, ..., Y_N)$ is $T \times N$; X is a three-dimensional matrix with p sheets $(T \times N \times p)$, the ℓ-th sheet is associated with the ℓ-th element of β ($\ell = 1, 2, ..., p$). The product $X\beta$ is $T \times N$, and $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)$ 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[4]

$$F'F/T = I_r \tag{7}$$

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

$$\Lambda'\Lambda = diagonal. \tag{8}$$

These two sets of restrictions uniquely determine Λ and F, given the product $F\Lambda'$.[5] 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 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 are required to be zero. However, the correlation, either cross sectional or serial, must be weak. This is known as the approximate factor model of Chamberlain and Rothschilds (1981). Under restrictions of (7) and (8), together with weak correlation in both dimensions, we show the model parameters can be consistently estimated.

To identify β, sufficient variation in X_{it} is needed. When F is observable, the usual identification condition is that the matrix $\frac{1}{NT}\sum_{i=1}^N X_i' M_F X_i$ is of full rank. Because F is not observable and is estimated, a stronger condition is required. Further details are given in Section 4.

3.2 Estimation

The least squares objective function is defined as

$$SSR(\beta, F, \Lambda) = \sum_{i=1}^N (Y_i - X_i\beta - F\lambda_i)'(Y_i - X_i\beta - F\lambda_i) \tag{9}$$

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

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

[4]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 (8).

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

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

$$W_i = F\lambda_i + \varepsilon_i$$

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

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

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), concentrating out $\Lambda = W'F(F'F)^{-1} = W'F/T$, the objective function becomes

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

Therefore, minimizing with respect to F is equivalent to maximizing $tr[F'(WW')F]$. 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 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 following set of nonlinear equations

$$\hat{\beta} = \Big(\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\Big)^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}} Y_i, \quad \text{and} \tag{11}$$

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

where V_{NT} is a diagonal matrix consists of the r largest eigenvalues of the above matrix[6] 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

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

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

$$tr(W'M_F W) = \sum_{i=1}^{N} W_i' M_F W_i = \sum_{i=1}^{N} (Y_i - X_i\beta)' M_F (Y_i - X_i\beta) \tag{13}$$

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

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) discuss an iteration procedure for estimation. Interestingly, convergence to a local optimum for such an iterated estimator is proved by Sargan (1964). Here we suggest a more robust iteration scheme (having much better convergence property) than the one implied by (11) and (12). Given F and Λ, we compute

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

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 of updating during iteration. The tables in the accompanying materials are based on this scheme.

Unbalanced panel. The estimation procedure can be modified to handle unbalanced data. Stock and Watson (1998) presented a method for estimating unbalanced factor models based on EM algorithm. Their procedure is extended here to models with regressors. Two sets of iterations are needed: outer iterations and inner iterations. Outer iterations are those between β and the factor model, similar to balanced data. Inner iterations are those within the factor model inherited from the EM method. For cross-section i, suppose we have observations for $t = 1, 2, ..., T_i$ (missing observations could occur at the beginning of the sample or at both ends). Suppose λ_i and F_t are observable for the moment, then the least squares estimator for β is

$$\hat{\beta} = \Big(\sum_{i=1}^{N}\sum_{t=1}^{T_i} X_{it}X_{it}'\Big)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T_i} X_{it}(Y_{it} - \lambda_i' F_t) \tag{14}$$

Assuming β is known, let $W_{it} = Y_{it} - X_{it}'\beta$, then $W_{it} = \lambda_i' F_t + \varepsilon_{it}$ is a pure factor model with unbalanced panel. Let $T = \max\{T_1, T_2, ..., T_N\}$ and define $I_{it} = 1$ for observable (i,t), and $I_{it} = 0$, otherwise. The EM algorithm in Stock and Watson (1998), at each stage of iteration, imputes the missing values using estimates from the prior stage. More specifically, let $\hat{\lambda}_i^{(h-1)}$, and $\hat{F}_t^{(h-1)}$ $(i = 1, ..., N; t = 1, 2, ..., T)$ be the estimates at stage $h-1$. Let $W_{it}^{(h)} = W_{it}$ for $I_{it} = 1$, and $W_{it}^{(h)} = \lambda_i^{(h-1)\prime} \cdot F_t^{(h-1)}$ for $I_{it} = 0$ (with starting value $W_{it}^{(0)} = 0$). Finally, let $W^{(h)} = (W_{it}^{(h)})$ be the $T \times N$ matrix. The h stage estimate for $\hat{F}^{(h)}$ is the first r eigenvectors of the matrix $W^{(h)}W^{(h)\prime}$ subject to the constraint $\hat{F}^{(h)\prime}\hat{F}^{(h)}/T = I$ and $\Lambda^{(h)} = T^{-1}W^{(h)\prime}\hat{F}^{(h)}$. This process continues until convergence. Let λ_i^* and F_t^* be the final stage estimates; these values are then plugged into (14) to obtain a new estimate of β (outer iteration). With the new β, we recompute $W_{it} = Y_{it} - X_{it}\beta$ for $I_{it} = 1$, readying for another around of inner iterations. Within the inner iterations, the starting value for $W_{it}^{(0)}$ when $I_{it} = 0$ is now $W_{it}^{(0)} = \lambda_i^{*\prime} F_t^*$ (instead of zero for faster convergence), where λ_i^* and F_t^* are the stopping values in the previous round of inner iterations.

3.3 Alternative estimation methods

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

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

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

Multiply the equation of $y_{i,t-1}$ by $\phi_t = f_{t1}/f_{t-1,1}$, and then subtract 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 one more time of the quasi-differencing method to the resulting equation, then the factor error will be eliminated. The GMM method as in Holtz-Eakin et al (1988) and Ahn, Lee and Schmidt (2001) can be used to consistently estimate the model parameters under some identification conditions. For the case of $r = 1$, GMM is also discussed by Arellano (2003) and Baltagi (2005).

2. Using the argument of Mundlak (1978) and Chamberlain (1984), 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}, 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-effect GLS to estimate $(\beta, \delta_1, ..., \delta_T)$. 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 $F_t = B\bar{X}_{\cdot t} + \xi_t$ to obtain

$$Y_{it} = X'_{it}\beta + \bar{X}'_{\cdot t}\rho_i + \lambda'_i \xi_t + \varepsilon_{it}$$

with $\rho_i = B\lambda_i$. Again, a random-effect GLS may 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 $\bar{X}_{i\cdot}$ and $\bar{X}_{\cdot t}$, in addition to $\bar{X}_{i\cdot}$ and $\bar{X}_{\cdot t}$.

3. The method of Pesaran (2006) augments the model with regressors $(\bar{Y}_{\cdot t}, \bar{X}_{\cdot t})$ under the assumption of F_t being correlated with regressors, where $\bar{Y}_{\cdot t}$ and $\bar{X}_{\cdot t}$ 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 assumes X_{it} is of the form $X_{it} = B_i F_t + e_{it}$

so that $\bar{X}_{\cdot t} = BF_t + \xi_t$ 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 order $O(N^{-1})$. Thus the factor error $\lambda_i'\xi_t$ is negligible under large N. He establishes $\sqrt{N}$ consistency and possible $\sqrt{NT}$ consistency for some special cases. It appears that when λ_i is correlated with the regressors, additional regressors, $\bar{Y}_{i,\cdot}$ and $\bar{X}_{i\cdot}$, 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 $\|A\| = (tr(A'A))^{1/2}$.

The following $p \times p$ matrix plays an important role in the paper,

$$D(F) = \frac{1}{NT}\sum_{i=1}^N X_i'M_F X_i - \frac{1}{T}\Big[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N X_i'M_F X_k a_{ik}\Big]$$

where $a_{ik} = \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$. 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 due to 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 Z_i is equal to the deviation of $M_F X_i$ from its mean, but here the mean is weighted average. Write $Z_i = (Z_{i1}, Z_{i2}, ..., Z_{iT})'$. Then

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

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 semi positive definite. Since each $Z_{it}Z_{it}'$ is a rank one semi-definite matrix, summation of NT such semi-definite 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. In fact, suppose that as $N, T \to \infty$, $D(F) \xrightarrow{p} D > 0$. If ε_{it} are iid $(0, \sigma^2)$, 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

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 $\iota_T = (1, 1, ..., 1)'$. 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 $F = W(W'W)^{-1/2} \in \mathcal{F}$, $D(F) = 0$. Assumption A is sufficient but not necessary. The analysis of time-invariant regressors and common regressors is postponed to Section 10, where it is shown that a necessary and sufficient condition for identification of β (maintaining other identifying restrictions) is $D(F) > 0$, when evaluated at the true factor process F. For now, it is not difficult to show if X_{it} is characterized by (3), where η_{it} have sufficient variations such as iid with positive variance, then Assumption A is satisfied.

Assumption B:

1. $E\|F_t\|^4 \le M$ 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 $T \to \infty$.
2. $E\|\lambda_i\|^4 \le M$ and $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\Lambda > 0$, for some $r \times r$ matrix Σ_Λ, as $N \to \infty$.

This assumption implies existence of r factors. Note that whether F_t or λ_t having 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 iid zero mean process. Similarly, λ_i can be cross-sectionally correlated.

Assumption C: serial and cross-sectional weak dependence and heteroskedasticity

1. $E(\varepsilon_{it}) = 0$, $E|\varepsilon_{it}|^8 \le M$;

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

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

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

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

4.

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

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

Assumption C is about weak serial and cross-sectional correlation. Heteroskedasticity is allowed but ε_{it} is assumed to have uniformly bounded eighth moment. The first three conditions are relatively easy to understand and are assumed in Bai (2003). We explain the meaning of C4. 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 $T^{-2}N^{-1}\sum_{t,s,u,v}\sum_{i,j}\text{cov}(\varepsilon_{it}\varepsilon_{is},\varepsilon_{ju}\varepsilon_{jv})$, i.e., the left hand side of the first inequality without the absolute sign. So part 1 of C4 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 part 2 is similar. It can be easily shown that if ε_{it} are independent over i and t with $E\varepsilon_{it}^4\leq M$ for all i and t, then C4 is true. If ε_{it} are iid with zero mean and $E\varepsilon_{it}^8\leq M$, then all assumptions in C hold.

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

Therefore, X_{it} is strictly exogenous. This rules out dynamic panel data models, a topic not considered in this paper.

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 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 is estimating a rotation of F^0.[7] Define $S_{NT}(\beta,F)$ as the concentrated objective function in (13) divided by NT together with centering, i.e.,

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

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

$$(\hat{\beta},\hat{F}) = \text{argmin}_{\beta,F}S_{NT}(\beta,F)$$

[7] In factor analysis, the estimated factors are then rotated to find economic interpretations. If one defines F^0 as the unique F that satisfies the identification restrictions (7) and (8), then $\hat{F}$ is directly estimating F^0. The objective function $S_{NT}(\beta,F)$ is uniquely defined irrespective of the definition of F^0. The fact that we can define F^0 differently is owing to the fundamental indeterminacy of factor models.

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

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

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

where $\hat{F}$ is the the matrix consisting 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 V_{NT} is a diagonal matrix consisting 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, we have, as $N, T \to \infty$,*

(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 show $S(\beta, F)$ 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 space 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 (1981) 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 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 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:

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

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 $(NT)^{-1/2}\sum_{i=1}^{N} Z_i'\varepsilon_i = (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} Z_{it}\varepsilon_{it}$. Its variance is given by, assuming correlation and heteroskedasticity in both dimensions

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

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

Assumption E: For some nonrandom positive definite matrix D_Z,

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

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

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 the following:

$$\text{plim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}\sum_{t=1}^{T}Z_{it}Z'_{jt} = D_1, \quad \text{plim}\frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\omega_{ts}\sum_{i=1}^{N}Z_{it}Z'_{is} = D_2 \tag{17}$$

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

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, D_0^{-1}D_1D_0'^{-1})$$

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

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, D_0^{-1}D_2D_0'^{-1})$$

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

Noting that $D_1 = D_2 = \sigma^2 D_0$ under iid assumption of ε_{it}, it follows that

Corollary 1 *Under the assumptions of Theorem 1, if ε_{it} are iid over t and i, 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 iid $N(0, \sigma^2)$, 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. It also requires an appropriate rate for N and T. If T/N converges to a constant, then bias will be present. 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*

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N\Big(\rho^{1/2}B_0 + \rho^{-1/2}C_0,\ D_0^{-1}D_ZD_0'^{-1}\Big)$$

where B_0 is the probability limit of B with

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

and C_0 is the probability limit of C with

$$C = -\,D(F^0)^{-1}\frac{1}{NT}\sum_{i=1}^{N}X'_iM_{F^0}\,\Omega\,F^0(F^{0\prime}F^0/T)^{-1}(\Lambda'\Lambda/N)^{-1}\lambda_i \tag{19}$$

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 $\Omega_k = E(\varepsilon_k\varepsilon'_k)$.

The bias $B_0 = 0$ when cross-section correlation and heterokedasticity are absent, and similarly $C_0 = 0$ when serial correlation and heteroskedasticity are absent. To see this, consider C of (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 argument for $B = 0$ is not so obvious, and is provided in the proof of Theorem 2(ii). When ε_{it} are iid over t and over i, both B_0 and C_0 are zero, the result specializes to Corollary 1.

Remark 1. Suppose k factors are allowed in the estimation, where $k \geq r$ but fixed. Then $\hat{\beta}$ remains to be $\sqrt{NT}$ 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 2. 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 a 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 $(k = 1, 2, ..., p)$ are matrices of $T \times N$ with X^k being the regressor matrix associated with parameter β_k (a scalar); ι_T is $T \times 1$ vector with all elements being 1, similarly for ι_N; $\alpha' = (\alpha_1, ..., \alpha_N)$ and $\xi = (\xi_1, ..., \xi_T)'$. 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 left and by M_N from 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 (6) as

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

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

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

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

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

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

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

It follows from (15) that

$$\sqrt{NT}(\hat{\beta} - \beta) = \sqrt{NT}(\hat{\beta}_{Asy} - \beta) + o_p(1).$$

Therefore, to purge the fixed effects, 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.

Instrumental variable interpretation. Treat Z_i as if it were an instrumental variable, and let $\hat{\beta}_{IV} = (\sum_{i=1}^{N} Z_i' X_i)^{-1} \sum_{i=1}^{N} Z_i' Y_i$. Then $\hat{\beta}_{IV}$ is exactly equal to the asymptotic representation of the interactive effect estimator. Thus the latter is an asymptotically IV estimator. Details are given in the accompanying document.

7 Bias corrected estimator

The interactive effect estimator is shown to have the following representation (see Proposition **??** in the appendix)

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

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 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 consistently estimate the biases and outline the idea of 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

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

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 with $\hat{F}$, 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 $\hat{F}'\hat{F}/T = I_r$,

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

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

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

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* $i \neq j$ *and* $t \neq s$. *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 $D_3 = \mathrm{plim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it}Z_{it}'\sigma_{i,t}^2$.

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. An alternative to biases correction in the case of $T/N \to \rho > 0$ is to use the Bekker (1994) standard errors to improve inference accuracy. This strategy is studied by Hansen, Hausman, and Newey (2005) in the context of many instruments.

Remark 3. Consider estimating C in the presence of serial correlation. We need consistent estimators for $T^{-1}X_i'\Omega_kF^0$ and $T^{-1}F^{0\prime}\Omega_kF^0$, where $\Omega_k = E\varepsilon_k\varepsilon_k'$ $(T \times T)$ and then 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 sub-blocks. 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 $\hat{W}_{it} \hat{\varepsilon}_{kt}$ $(t = 1, ..., T)$. Similar argument has been made in Bai (2003).

Remark 4. 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 is naturally ordered and far-apart 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 (2005) study the estimation of a similar object as B. They show 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

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

where $n/N \to 0$ and $n/T \to 0$. The argument of Bai and Ng (2005) can be adapted to show $\hat{B}$ 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

$$\hat{D}_1 = \frac{1}{N} \sum_{i=1}^N \hat{\sigma}_i^2 (\frac{1}{T} \sum_{t=1}^T \hat{Z}_{it} \hat{Z}_{it}'), \quad \hat{D}_2 = \frac{1}{T} \sum_{t=1}^T \hat{\omega}_t^2 (\frac{1}{N} \sum_{i=1}^N \hat{Z}_{it} \hat{Z}_{it}'), \quad \hat{D}_3 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \hat{Z}_{it} \hat{Z}_{it}' \hat{\varepsilon}_{it}^2$$

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 $\hat{Z}_{it}$ is defined earlier.

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 5: 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 4. 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 3. 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

While interactive effects models include the additive models as special cases, additivity is not 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 following restrictions are needed

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

see Greene (2000, page 565). The following restrictions are maintained:

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

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 notations. For any variable ϕ_{it}, define

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

$$\dot{\phi}_{it} = \phi_{it} - \bar{\phi}_{i.} - \bar{\phi}_{.t} + \bar{\phi}_{..}$$

and its vector form

$$\dot{\phi}_i = \phi_i - \iota_T\bar{\phi}_{i.} - \bar{\phi} + \iota_T\bar{\phi}_{..}$$

where $\bar{\phi} = (\bar{\phi}_{.1}, ..., \bar{\phi}_{.T})'$.

The least squares estimators are

$$\begin{aligned}
\hat{\mu} &= \bar{Y}_{..} - \bar{X}_{..}'\hat{\beta} \\
\hat{\alpha}_i &= \bar{Y}_{i.} - \bar{X}_{i.}'\hat{\beta} - \hat{\mu} \\
\hat{\xi}_t &= \bar{Y}_{.t} - \bar{X}_{.t}'\hat{\beta} - \hat{\mu} \\
\hat{\beta} &= \Big[\sum_{i=1}^N \dot{X}_i' M_{\hat{F}} \dot{X}_i\Big]^{-1} \sum_{i=1}^N \dot{X}_i' M_{\hat{F}} \dot{Y}_i
\end{aligned}$$

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

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

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, page 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 (μ, $\{\alpha_i\}$, $\{\xi_t\}$) is equivalent to using $(\dot{Y}_{it}, \dot{X}_{it})$ to estimate the remaining parameters. That is, 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, i.e., $\dot{\lambda_i'F_t} = \lambda_i'F_t$ due to restriction (29). This objective function is the same as (9), except Y_{it} and X_{it} are replaced by their dotted versions. From the analysis of 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,

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

Multiply $\iota_T = (1, ..., 1)'$ on each side,

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

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

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

In the accompanying document, we show the identity (see Lemma **??**), $\sum_{i=1}^N \dot{Z}_i'\dot{\varepsilon}_i \equiv \sum_{i=1}^N \dot{Z}_i'\varepsilon_i$. 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) $\text{plim}\frac{1}{NT}\sum_{i=1}^N \dot{Z}_i'\dot{Z}_i = \dot{D}_0 > 0$

(ii) $\frac{1}{\sqrt{NT}}\sum_{i=1}^N \dot{Z}_i'\varepsilon_i \xrightarrow{d} N(0, \dot{D}_Z)$, where $\dot{D}_Z = \text{plim}\frac{1}{NT}\sum_{i,j,t,s}\sigma_{ij,ts}\dot{Z}_{it}\dot{Z}_{js}'$

Theorem 5 *Assume Assumptions A-F hold. Then as $T, N \to \infty$*
(i) under the assumptions of part (i) of Theorem 2

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

(ii) Under the assumptions of part (ii) of Theorem 2

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

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 is the same as before 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 better description of the data. The first method is that of Hausman test statistic (Hausman, 1978) and the second is based on the number of factors. We detail the Hansman test method, delegating the number-of-factors method to the accompanying materials. Throughout this section, for simplicity, we assume ε_{it} are iid over i and t, and $E(\varepsilon_{it}^2) = \sigma^2$.

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 $\lambda_i' = (\alpha_i, 1), \quad F_t = (1, \xi_t + \mu)'$.

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 estimator of β in (31) is

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

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

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

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 **??**

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

where

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

The variances of the two estimators are

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

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

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

This implies

$$\mathrm{var}(\hat{\beta}_{IE} - \hat{\beta}_{FE}) = \mathrm{var}(\hat{\beta}_{IE}) - \mathrm{var}(\hat{\beta}_{FE})$$

Thus the Hausman test takes the form

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

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 $L = NT - (N+T) - p + 1$. Then $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

Remark 6. 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, Hansamn 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 accompanying materials.

10 Time-invariant and common regressors

In earnings studies, time-invariant regressors include education, gender, race, etc; common variables are those representing 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 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) allow for time-invariant regressors, although they do not consider the joint presence of common regressors. Their identification condition relies on non-zero correlation between factor loadings and the 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 as follows: X_{it} is $p \times 1$, x_i is $q \times 1$, w_t is $\ell \times 1$, F_t is $r \times 1$. Introduce

$$X_i = \begin{bmatrix} X_{i1}' & x_i' & w_1' \\ X_{i2}' & x_i' & w_2 \\ & \vdots & \\ X_{iT}' & x_i & w_T \end{bmatrix}, \quad \beta = \begin{bmatrix} \varphi \\ \gamma \\ \delta \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_N' \end{bmatrix}, \quad W = \begin{bmatrix} w_1' \\ w_2 \\ \vdots \\ w_T \end{bmatrix}$$

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

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

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 needed is the following 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.

We now explain the meaning of $D(F^0) > 0$ and argue that it can be segregated into some intuitive and reasonable conditions. To simplify notation and for ease of discussion, we assume the only regressors are time invariant or common (no X_{it}), i.e.,

$$X_i = (\iota_T x_i', W), \quad \beta' = (\gamma', \delta')$$

The condition $D(F^0) > 0$ implies the following four restrictions:

1. (Genuine interactive effects) F^0 or its rotation does not contain ι_T; Λ or its rotation does not contain ι_N. Otherwise, we are back into the environment of Hausman and Taylor, instrumental variables must be used to identify β. In algebraic notation

$$\frac{1}{T}\iota_T' M_{F^0}\iota_T > 0 \quad \text{and} \quad \frac{1}{N}\iota_N' M_\Lambda \iota_N > 0$$

2. (No multicollinearity between W and F^0) The following matrix is positive definite,

$$\frac{1}{T}W' M_{F^0} W > 0.$$

   Without this assumption, even if F^0 is observable, we cannot identity β and Λ due to multicollinearity.

3. (No multicollinearity between $\underline{x}$ and Λ)

$$\frac{1}{N}\underline{x}' M_\Lambda \underline{x} > 0$$

   This is required for identification of β and F^0.

4. (Identification of grand mean, if exists). At least one of the following holds

$$\frac{1}{N}(\underline{x}, \iota_N)' M_\Lambda (\underline{x}, \iota_N) > 0 \tag{36}$$

$$\frac{1}{T}(\iota_T, W)' M_{F^0} (\iota_T, W) > 0 \tag{37}$$

   That is, either $\underline{x}$ does not contain ι_N or W does not contain ι_T. If both contain the constant regressor, there will be two grand mean parameters, thus not identifiable.

To see that $D(F^0) > 0$ implies the above four conditions, we simply compute $D(F)$,

$$D(F) = \begin{bmatrix} (\frac{1}{N}\underline{x}' M_\Lambda \underline{x})(\iota_T' M_F \iota_T/T) & (\frac{1}{N}\underline{x}' M_\Lambda \iota_N)(\iota_T' M_F W/T) \\ (W' M_F \iota_T/T)(\frac{1}{N}\iota_N' M_\Lambda \underline{x}) & (\frac{1}{N}\iota_N' M_\Lambda \iota_N)(W' M_F W/T) \end{bmatrix}$$

For a positive definite matrix, the diagonal block matrices must be positive definite. This leads to the first three conditions immediately. To see that $D(F^0) > 0$ also implies 4, we use contradiction argument. Suppose neither of the matrices in (36) and (37) is positive definite and since they are semi-positive definite, their determinants must be zero. Then it is not difficult to show that the determinant of $D(F^0)$ is also zero. This contradicts with $D(F^0) > 0$.

More interestingly, the four conditions above are also sufficient for $D(F^0) > 0$, a consequence of the Lemma below:

Lemma 1 *Let A be a $q \times q$ symmetric matrix. Assume the following $(q+1) \times (q+1)$ matrix is positive definite,*

$$\bar{A} = \begin{bmatrix} A & \alpha \\ \alpha' & \tau \end{bmatrix} > 0$$

so $A > 0$ and $\tau > 0$ (a scalar). Suppose $\bar{B}$ below is semi-positive definite

$$\bar{B} = \begin{bmatrix} \nu & b' \\ b & B \end{bmatrix} \geq 0, \quad \text{with } \nu > 0, B > 0$$

where B is $\ell \times \ell$ and ν is scalar. Then the following $(q+\ell) \times (q+\ell)$ matrix is positive definite

$$\bar{A} \Diamond \bar{B} = \begin{bmatrix} A\nu & \alpha\, b' \\ b\, \alpha' & \tau B \end{bmatrix} > 0$$

The role of $\bar{A}$ and $\bar{B}$ can be reversed. The lemma only requires one of them to be positive definite, not both. Now suppose (36) holds. Let $\bar{A} = \frac{1}{N}(\underline{x}, \iota_N)' M_\Lambda (\underline{x}, \iota_N)$ and $\bar{B} = \frac{1}{T}(\iota_T, W)' M_{F^0}(\iota_T, W)$, and $\bar{A} > 0$. In addition, $A = \frac{1}{N}\underline{x}' M_\Lambda \underline{x} > 0$, $\tau = \iota_N' M_\Lambda \iota_N > 0$, $\nu = \frac{1}{T}\iota_T' M_{F^0} \iota_T > 0$, and $B = W' M_{F^0} W/T > 0$, all following from the first three conditions. Thus the assumptions of the Lemma 1 hold. It follows that $\bar{A} \Diamond \bar{B} > 0$. But $\bar{A} \Diamond \bar{B} = D(F^0)$. Thus the four conditions imply $D(F^0) > 0$. We now summarize the result.

Lemma 2 *The matrix $D(F^0) > 0$ if and only if the above four conditions hold.*

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 accompanying document. The proposition implies that $D(F^0) > 0$ is a sufficient condition for consistent estimation.

Given consistency, the rest 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 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.

11 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 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 market wide risks, and the loadings reflect assets' exposure to the risks. 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. Under genuine interact effects, we showed that the grand mean, the coefficients of time-invariant regressors and those of common regressors are identifiable and can be consistently estimated.

Many important and interesting issues remain to be examined. A useful extension is large N-large T dynamic panel data model with multiple interactive effects. Another broad extension is nonstationary panel data analysis, particulary panel data cointegration, a subject that recently attracts considerable attention. In this setup, X_{it} is a vector of 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, interactive effect approach can be applied by jointly estimating the unobserved common stochastic trends F_t and the model coefficients, leading to consistent estimation.

References

[1] Abowd, J.M., F. Kramarz, and D.N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67 251-333.

[2] 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*, 102, 219-255.

[3] Altug, S. and R.A. Miller (1990): "Household Choices in equilibrium," *Econometrica*, 58, 543-570.

[4] Anderson, T. W. 1984, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

[5] Anderson, T.W., and C. Hsiao (1982): "Formulation and estimation of dynamic Models with Error Components," *Journal of Econometrics*, 76, 598-606.

[6] Anderson, T.W. and H. Rubin (1956): "Statistical Inference in Factor Analysis," in J. Neyman, ed., *Proceedings of Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Vol 5, 111-150.

[7] Andrews, D. W. K. (2004), Cross-section Regression with Common Shocks, forthcoming in *Econometrica*.

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

[9] 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.

[10] Alvarez, J. and M. Arellano (2003): The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators. *Econometrica* 71, 1121-1159.

[11] Bai, J. (1994): "Least squares estimation of shift in linear processes," *Journal of Time Series Analysis*, 15, 453-472.

[12] Bai, J. (2003): "Inferential Theory for Factor Models of Large Dinensions," *Econometrica*, 71, 135-173.

[13] Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191-221.

[14] Bai, J., and S. Ng (2005): "Confidence intervals for diffusion index forecasting," http://www-personal.umich.edu/ ~ngse/pub/forecast_long.pdf.

[15] Baltagi, B.H. (2005): *Econometric Analysis of Panel Data*, John Wiley: Chichester.

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

[17] Campbell, J.Y, A.W. Lo, and A.C. MacKinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.

[18] 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," NBER Working Paper 9546, forthcoming in *International Economic Review*.

[19] Cawley, J., K. Conneely, J. Heckman and E. Vytlacil (1997): "Cognitive Ability, Wages, and Meritocracy," in *Intelligence Genes, and Success: Scientists Respond to the Bell Curve*, edited by B. Devlin, S. E. Feinberg, D. Resnick and K. Roeder, Copernicus: Springer-Verlag, 179-192.

[20] Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225-238.

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

[22] Chamberlain, G., and M. Rothschild (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica* 51, 1305-1324.

[23] 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.

[24] 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.

[25] Fama, E. and French, K. 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*.

[26] Feldstein, M., and C. Horioka. (1980): "Domestic savings and international capital flows," *Economic Journal* 90, 314-329.

[27] Giannone, D. and M. Lenza (2005): "The Feldstein Horioka Fact," Unpublished manuscript, European Central Bank.

[28] Goldberger, A.S. (1972): "Structural Equations Methods in the Social Sciences," *Econometrica*, 40, 979-1001.

[29] Greene, W. (2000): *Econometric Analysis*, 4th edition, Prentice Hall, New Jersey.

[30] 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.

[31] Hahn, J. and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models." *Econometrica* 72:4, 1295-1319

[32] Hansen, C., J. Hausman, and W. Newey (2005): Estimation with many instruments, unpublished manuscript, Department of Economics, MIT.

[33] Hausman, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1271.

[34] Hausman, J.A and W.E. Taylor (1981): "Panel data and unobservable individual effects," *Econometrica*, 49,1377-1398.

[35] Holtz-Eakin, D., W. Newey, and H. Rosen (1988): "Estimating Vector Autoregressions with Panel Data", *Econometrica*, 56, 1371-1395.

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

[37] Jöreskog, K.G., and Goldberger, A.S. (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.

[38] 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.

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

[40] 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.

[41] Lawley, D.N. and A.E. Maxwell (1971): *Factor Analysis as a Statistical Method*, London: Butterworth.

[42] 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.

[43] Lettau, M. and S. Ludvigson (2001): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia are Time Varying," *Journal of Politcal Economy* 109, 1238-1287.

[44] 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.

[45] Moon, R. and B. Perron (2004): "Testing for a unit root in panels with dynamic factors," *Journal of Econometrics*, 122, 81-126.

[46] Mundlak, Y. (1978): "On the pooling of time series and cross section data," *Econometrica,* 46, 69-85.

[47] Neyman, J., and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1-32.

[48] Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in Engle, R.F. and D. McFadden (eds.) *Handbook of Econometrics,* North Holland.

[49] Newey, W.K. and K.D. West (1987): "A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica, 55,* 703-708.

[50] Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417-1426.

[51] Obstfeld, M. and K. Rogoff (2000): "The six major puzzles in international macroeconomics: is there a common cause?" NBER Macroeconomic Annual, 339-390.

[52] Pesaran, M. H. (2004): "Estimation and Inference in Large Heterogeneous panels with a Multifactor Error Structure," unpublished manuscript, University of Cambridge,

[53] Phillips, P.C.B., and Sul, D. (2003): "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence," *The Econometrics Journal*, 6,217-259.

[54] Ross, S. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Finance* 13, 341–360.

[55] Sargan, J.D. (1964): "Wages and prices in the United Kingdom: a study in ecnometrics methodology, in *Econometric Analysis of National Economic Planning,* Hart PG, Mill G, and Whitaker JK (eds). Butterworths: London, 25-54.

[56] Stock, J. H. and M. W. Watson (1998): "Diffusion indexes," NBER working paper, no. 6702.

[57] 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.

[58] Townsend, R. (1994): "Risk and Insurance in Village India," *Econometrica*, 62, 539-592.