Panel Data Models with Interactive Fixed Effects

Jushan Bai*

April, 2005
This version: October, 2005

Abstract

This paper considers large N and large T panel data models with unobservable multiple interactive effects. These models are useful for both micro and macro econometric modelings. 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 responses over cross sections. Since the interactive effects are allowed to be correlated with the regressors, they are treated as fixed effects parameters to be estimated along with the common slope coefficients. The model is estimated by the least squares method, which provides the interactive-effects counterpart of the within estimator.

We first consider model identification, and then derive the rate of convergence and the limiting distribution of the interactive-effects estimator of the common slope coefficients. The estimator is shown to be $\sqrt{NT}$ consistent. This rate is valid even in the presence of correlations and heteroskedasticities in both dimensions, a striking contrast with fixed T framework in which serial correlation and heteroskedasticity imply unidentification. The asymptotic distribution is not necessarily centered at zero. Biased corrected estimators are derived. 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.

We also derive identification conditions for models with grand mean, time-invariant regressors, and common regressors. It is shown that there exists a set of necessary and sufficient identification conditions for those models. Given identification, the rate of convergence and limiting results continue to hold.

Key words and phrases: incidental parameters, additive effects, interactive effects, factor error structure, principal components, serial and cross-sectional correlation, serial and cross-sectional heteroskedasticity, biased-corrected estimator, Hausman tests, Time-invariant regressors, common regressors, and grand mean.

*I thank econometrics seminar participants at the University of Pennsylvania, Rice, and MIT/Harvard for useful comments. In particular, I thank Gary Chamberlain, Yoosoon Chang, Jerry Hausman, James Heckman, Guido Imbens, Gregory Kordas, Arthur Lewbel, Whitney Newey, Joon Park, Frank Schorfheide, Robin Sickles, James Stock, and Motohiro Yogo for helpful discussions.
Email: jushan.bai@nyu.edu; Department of Economics, New York University, New York, NY 10003. This work is supported in part by the NSF grants SES-0137084 and SES-0424540.

1 Introduction

We consider the following panel data model with N cross-sectional units and T time periods

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

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 estimation of β, the common slope coefficients.

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

$$Y_{it} = X_{it}'\beta + \alpha_i + \xi_t + \varepsilon_{it}, \tag{1}$$

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 interact 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

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

That an interactive effects model is more general than fixed effects model is well known for the case of single factor ($r = 1$), e.g., Holtz-Eakin, Newey, and Rosen (1988). This follows from, when $F_t = 1$ for all t, $\lambda_i F_t = \lambda_i$, and when $\lambda_i = 1$ for all i, $\lambda_i F_t = F_t$. However, the general additive effects $\alpha_i + \xi_t$ being a special case of multiple interactive effects appears to be less noticed. But once pointed out, it becomes trivial and obvious. The point is that the class of the interactive effects models is much larger than that of additive effects models. For $r > 2$, there exist non-trivial interactive effects.

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 for unobserved heterogeneity, see Chamberlain (1984) and Arellano and Honore (2001). Indeed, we allow the observable X_{it} to follow

$$X_{it} = \tau_i + \theta_t + \sum_{k=1}^{r} a_k \lambda_{ik} + \sum_{k=1}^{r} b_k F_{kt} + \sum_{k=1}^{r} c_k \lambda_{ik} F_{kt} + \pi_i' G_t + \eta_{it} \tag{2}$$

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 influencing u_{it}. 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 has a zero mean, or whether F_t is independent over time. In fact, F_t can be a dynamic process without zero mean. The same is true for λ_i. In this paper, 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 genuine 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 as $F_t \not\equiv \bar{F}$, where $\bar{Y}_{i.}$, $\bar{X}_{i.}$, and $\varepsilon_{i.}$ are averages over time. Nevertheless, this simple interactive effect can be eliminated by the so called quasi-differencing method.[1] It is noted that quasi-differencing gives rise to undesirable features. For example, it introduces lagged dependent variable and time-varying parameters, and requires F_t be non-zero for each t. It does not appear to work with multiple interactive effects.

Recently, Pesaran (2004) 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 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.

A two-step estimator based on principal components was proposed by Coakey, Futers, and Smith (2002). In the first step, β is estimated by the pooled least squares ignoring the factor structure, and then uses the residuals to estimate F_t by the principal components method. The second step treat the estimated F_t as observable and then estimate β. Pesaran shows this estimator in general is inconsistent. It is, in fact, not surprising to find inconsistency of the two-step estimator because both β and F are inconsistently estimated in the first step when the interactive effects are correlated with regressors. 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 noted 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).

An interesting setup that deviates from traditional factor models is proposed by Kneip, Sickles, and Song (2005). They assume F_t is a smooth function of t and estimate it by

[1] See Chamberlain (1984) and Holtz-Eakin, Newey, and Rosen (1988).

smoothing spline. Given the spline basis, the estimation problem becomes that of ridge regression. Such a setup is useful when the time effects 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, in spite of serial or cross-sectional correlations and heteroskedasticities in ε_{it}, a striking contrast for fixed T framework, in which serial correlation implies nonidentification of the model.

When deriving this new result, we also allow very general data generating processes. Earlier fixed T studies assume that X_{it} are iid over i, ruling out X_{it} that 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.

In standard panel data regression $Y_{it} = X_{it}'\beta + \varepsilon_{it}$, with strictly exogenous regressor X_{it} and with either $N \to \infty$ or $T \to \infty$, the least squares estimator for β is consistent even though ε_{it} is correlated or heteroskedastic (serial or cross-sectional). It is common perception that correlation or heteroskedasticity in ε_{it} does not affect consistency. A fundamental difference occurs in the factor model $Y_{it} = X_{it}'\beta + \lambda_i' F_t + \varepsilon_{it}$, where λ_i and F_t are unobserved and are to be estimated. With fixed N and with correlation over i of unknown form, the model is identifiable, see section 3.1 for explanation. The same is true under fixed T and under serial correlation of unknown form. Therefore, it is a significant result that, with both N and T going to infinity, $\sqrt{NT}$ consistency is attainable under arbitrary (weak) serial or cross-sectional correlation.

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 biases 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. In particular, in the absence of serial correlation and heteroskedasticity in ε_{it}, $\hat{\beta} - \beta$ has a bias of order $O(1/N)$. With serial correlation and heteroskedasticity in ε_{it}, an additional bias of order $O(1/T)$ exists. We show that these biases can be consistently estimated and 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

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

also consider the problem of testing additive effects versus interactive effects. We show that the principle of Hausman test is applicable in this context. We also argue that the number of factors can be consistently estimated. Discriminating between additive and interactive effects can also be performed by determining the number of factors.

In section 2, we briefly 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 estimation. These conditions are quite general, allowing correlations and heteroskedasticities in both dimensions. Section 5 derives the asymptotic representation of the interactive-effects estimator along with its asymptotic distribution. 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 regressors that are common to each cross-section. Monte carlo simulations are given in Section 11 and concluding remarks are given in the last section. All proofs are provided in the appendix.

2 Why multiple interactive effects

A theoretical appeal for interactive-effects models is their inclusion of additive-effects models as special cases. While encompassing traditional models, interactive-effects models are not overly general to still retain a manageable structure. More importantly perhaps, interactive models are of practical relevance. For microeconomic data, when studying earnings for example, the usual fixed effects capture the unobservable innate ability or intelligence. Research suggests that other individual habits or characteristics such as motivation, dedication, perseverance, hard-working, and even self-esteem are important determinants for earnings, see Cawley et al. (2003), and Carneiro, Hansen, and Heckman (2003). Arguably, rewards to these characteristics are not time invariant. Among many possible reasons, we suggest two. First, suppose there are different job types with different types placing different valuations on those individual characteristics. When workers switch job types over time, we expect to see a time varying valuations of individual characteristics. Second, it may take time for employers to recognize these unobservable characteristics. We may consider F_t as the level of employer's knowledge on those traits after a worker has been employed for t periods. In this example, X_{it} typically includes work experience, education, race, gender, etc.

In macroeconomics, F_t underlies the common shocks that drive the co-movement of the variables, λ_i represents the heterogeneous responses to these common shocks; the observable vectors X_{it} are firm or country specific variables such as capital and labor inputs. In stock return data, F_t is a vector of unobservable factor returns and λ_i is a vector of factor loadings, X_{it} are firm specific variables such as book to market ratios and PE ratios; ε_{it} are the idiosyncratic returns. The arbitrage pricing theory of Ross (1976) is rested upon a factor structure.

Additionally, interactive effects model provides a simple way of modelling cross-section correlations or common shocks. In a recent paper, Andrews (2004) demonstrates the adverse consequence on statistical inference of neglecting common shocks.

3 Identification and 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. We discuss the implications and consequences of these two aspects on identification. Some of the issues are not well understood by the existing literature. We clarify the pertinent ones.

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

$$Y = X\beta + F\Lambda' + \varepsilon \tag{3}$$

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

$$F'F/T = I_r \tag{4}$$

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

$$\Lambda'\Lambda = diagonal \tag{5}$$

These two sets of restrictions uniquely determine Λ and F, given the product $F\Lambda'$.[3] The least squares estimators for F and Λ derived below satisfy these restrictions.

Uniqueness is only a necessary condition for identification and itself does not imply identification.[4] It is instructive, at this juncture, to compare with the identification restrictions

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

[4] Due to the fundamental lack of identification of factor models, these restrictions are not meant to be the true data generating process. They are in fact meant for producing a unique set of estimates. Once a unique estimate is available, the factors or factor loadings are then rotated to have structural interpretations, see Lawley and Maxwell (1971).

employed in classical factor analysis. For this purpose, write the model as an N-dimensional time series process

$$\underline{Y}_t = \underline{X}_t\beta + \Lambda F_t + \underline{\varepsilon}_t, \qquad t = 1, 2, ..., T$$

where $\underline{Y}_t = (Y_{1t}, ..., Y_{Nt})'$, $\underline{X}_t = (X_{1t}, ..., X_{Nt})'$ and $\underline{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$. Normalization (4) is replaced by $var(F_t) = I_r$. Let $\Sigma_Y = var(\underline{Y}_t)$ and $\Phi = var(\underline{\varepsilon}_t)$, both are $N \times N$ matrices. We have

$$\Sigma_Y = \Lambda\Lambda' + \Phi \tag{6}$$

Restriction (5) is still applicable. In addition, classical factor analysis also assumes Φ is diagonal. An unrestricted Φ would render the classical factor models unidentifiable because Φ alone would have as many unknown parameters as Σ_Y (which is treated as known for identification).[5] These three sets of restrictions imply identification.

Under large N, there is no need to assume Φ to be diagonal. Indeed, none of the elements of Φ need to be zero, an essence of the approximate factor model of Chamberlain and Rothschild (1983).[6] We do require, however, weak cross-sectional correlation characterized by

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\sigma_{ij}| \leq M$$

for all N and for some finite M not depending on N, where $\sigma_{ij} = E\varepsilon_{it}\varepsilon_{jt}$. Such a restriction is ineffective under fixed N since it is already true. But under $N \to \infty$, it implies nontrivial restrictions. Chamberlain and Rothschild (1983) show that Λ is identifiable under weak cross sectional correlations.

The argument of Chamberlain and Rothschild assumes an known Σ_Y. In our case, the number of observations of T can be much smaller than the number of variables N, so that Σ_Y cannot, in general, be consistently estimated. For example, the rank of Σ_Y can be of full rank (i.e., N), but the rank of a covariance estimator of Σ_Y does not exceed $\min[T, N]$. Thus the possibility of not knowing Σ_Y, even under large samples, is a major distinction from the assumption of classical factor analysis and that of Chamberlain and Rothschild. Still, both Λ and F can be consistently estimated as shown in Bai (2003). This forms the basis for consistent estimation of β when regressors are present. Furthermore, similar to Bai (2003), we allow serial correlation and heteroskedasticity. Therefore, the model considered in this paper is more general than the approximate factor model of Chamberlain and Rothschild. Finally, we point out that weak cross-section correlation is part of model assumptions, and it cannot be imposed in estimation due to correlations' unknown form, unlike classical factor analysis in which diagonality of Ω_ε is imposed to solve for other parameters.

The estimated F and Λ under the preceding restrictions do not necessarily have any meaningful economic interpretations, unless they are subject to further rotation, a standard practice in factor analysis. However, it is possible to derive structurally interpretable identification conditions. First, impose the normalization that $var(F_t)$ is diagonal or $F'F/T$ is

[5]Similarly, under fixed T, unrestricted serial correlation makes the model unidentifiable.

[6]They require that the largest eigenvalue of Φ be bounded.

diagonal. The factor loading matrix is assumed to take the form

$$\Lambda = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \lambda_{21} & 1 & \cdots & 0 \\ & & \vdots & \\ \lambda_{r1} & \lambda_{r2} & \cdots & 1 \\ & & \vdots & \\ \lambda_{N1} & \lambda_{N2} & \cdots & \lambda_{Nr} \end{bmatrix},$$

That is, the first r-rows of Λ is a lower triangular matrix with 1's on the diagonal. In the absence of β, the identification condition implies

$$Y_{1t} = F_{1t} + \varepsilon_{1t}$$

so the first variable is equal to the first factor plus an idiosyncratic error. Thus we can give economic meanings to the first factor, for example, the interest rate factor, if Y_{it} is an interest-rate variable. Note that "1" can be replaced by a vector of ones such that $\iota = (1, 1, ..., 1)'$, thus a group of variables are related to the first factor (e.g., a group of bond yields variables with different maturities). Similarly,

$$Y_{2t} = \lambda_{21}F_{1t} + F_{2t} + \varepsilon_{2t} = \lambda_{21}Y_{1t} + F_{2t} + \varepsilon_{2t}^*$$

so we can give meaning to F_{2t}, and so on. Ahn, Lee, and Schmidt (2001) use a similar identification condition by reversing the role of F and Λ, that is, $F_1 = 1$ and leaving Λ unrestricted for a single factor model. The above identification scheme requires a careful arrangement of variables, especially when structural interpretation of F is the main objective. That is, which variable is assigned to Y_{1t} and which is assigned to Y_{2t}, and so on, are not arbitrary. When the objective is to estimate β, not the structural interpretation of F, cross-sectional ordering of the data should play no role. Therefore, the identification restrictions used in this paper are (4) and (5).

To identify β, sufficient variation in X_{it} is needed. When F is observable, the usual condition is that $\frac{1}{NT}\sum_{i=1}^N X_i' M_F X_i$ is a full rank matrix (with rank of p). 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 is

$$SSR(\beta, F, \Lambda) = \sum_{i=1}^N (Y_i - X_i\beta - F\lambda_i)'(Y_i - X_i\beta - F\lambda_i) \tag{7}$$

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

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

Given β, the variables $W_i = Y_i - X_i\beta$ 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 can be written

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

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

$$tr(W'M_F W) = tr(W'W) - tr(F'WW'F)/T \tag{8}$$

Therefore, minimizing with respect to F is equivalent to maximizing $tr[F'(WW')F]$. It follows that 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

$$\hat{\beta} = \Big(\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\Big)^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}} Y_i, \quad \text{and} \tag{9}$$

$$\Big[\frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'\Big]\hat{F} = \hat{F}V_{NT} \tag{10}$$

where V_{NT} is a diagonal matrix consists of the r largest eigenvalues of the above matrix[7] in the brackets, arranged in decreasing order. The solution $(\hat{\beta}, \hat{F})$ can be simply obtained by iteration. Finally, from the concentrated solution $\Lambda = W'F/T$, $\hat{\Lambda}$ is expressed as 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 (7). The pair $(\hat{\beta}, \hat{F})$ jointly minimizes the concentrated objective function (8), which is equal to, when substituting $Y_i - X_i\beta$ for W_i,

$$tr(W'M_F W) = \sum_{i=1}^{N} W_i' M_F W_i = \sum_{i=1}^{N}(Y_i - X_i\beta)' M_F (Y_i - X_i\beta) \tag{11}$$

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 an iterated estimator such as here is proved by Sargan (1964). In section 11, we elaborate some iteration schemes and suggest an iteration procedure that has much better convergence property than the one implied by formulae (9) and (10).

[7]We divide this matrix by NT to make V_{NT} have a proper limit. The scaling does not affect $\hat{F}$.

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 is positive definite. The presence of the second term is because of unobservable F and Λ. It takes on this particular form is due to the special form of the nonlinearity of the interactive effects.

Define the $T \times p$ vector

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

Z_i is 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 $N, T \to \infty$, $D(F) \to D > 0$. If ε_{it} are iid $(0, \sigma^2)$, then the limiting distribution of $\hat{\beta}$ is 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: $E\|X_{it}\|^4 \leq M$. Let $\mathcal{F} = \{F : F'F/T = I\}$.

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

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-invariance regressors and common regressors is delicate and is postponed to Section 10, there it is shown that a necessary and sufficient condition for identification of β (maintaining other identifying restrictions) is $D(F^0) > 0$, where F^0 is the true factor. For now, it is not difficult to show if X_{it} is characterized by (2), where η_{it} have sufficient variations such as iid with positive variance, then Assumption A is satisfied.

Assumption B:

1. $E\|F_t\|^4 \leq M$ 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. $E\|\lambda_i\|^4 \leq M$ 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 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 iid zero mean process. Similarly, λ_i can be cross-sectionally correlated.

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

1. $E(\varepsilon_{it}) = 0$, $E|\varepsilon_{it}|^8 \leq M$;
2. $E(\varepsilon_{it}\varepsilon_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ for all (t,s) and $|\sigma_{ij,ts}| \leq \tau_{ts}$ for all (i,j) such that

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

   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), $E|N^{-1/2}\sum_{i=1}^{N}\Big[\varepsilon_{is}\varepsilon_{it} - E(\varepsilon_{is}\varepsilon_{it})\Big]|^4 \leq M$.
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 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 that is beyond scope of this paper.

5 Asymptotic representation and limiting theory

We use (β^0, F^0) to denote the true parameters for easy of exposition, and we still use λ_i without the superscript 0 as it is not directly estimated thus not necessary.

Define $S_{NT}(\beta, F)$ as the concentrated objective function in (11) 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

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

As explained in the precious 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 eigenvalues of this matrix, arranged in decreasing order.

Proposition 5.1 *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

$$(F^{0\prime}\hat{F}/T)(\hat{F}' F^0/T) - (F^{0\prime}F^0/T) \xrightarrow{p} 0$$

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 $T \times 1$ 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. Consistency can be stated in terms of some average norm, or can be stated for componentwise consistency. This is done in the next proposition. Nevertheless, part (ii) contains certain consistency property. In fact, (ii) is equivalent to $\|P_{\hat{F}} - P_{F^0}\| = o_p(1)$, i.e, the space spanned by $\hat{F}$ and F^0 are asymptotically the same.

Further development needs the invertibility of the matrix V_{NT}, which we establish below. In addition, we show that the limit of V_{NT} is the diagonal matrix consisting of the eigenvalues of the matrix $\Sigma_\Lambda \Sigma_F$, defined in assumption B. Note that for any positive definite matrices, A and B, the eigenvalues of AB are the same as those of BA and $A^{1/2}BA^{1/2}$, etc, therefore, all eigenvalues are positive.

Proposition 5.2 *Under assumptions A-D*

(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* $\Sigma_\Lambda \Sigma_F$.

(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*

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

Part (ii) shows the average (norm) consistency of $\hat{F}$ for F^0 , and it extends the result of Bai and Ng (2002) to include the estimated β. Since V_{NT} and $F^{0\prime}\hat{F}/T$ are both invertible for all large N and T, the matrix H is invertible. Thus $\hat{F}$ is a full rank rotation of F^0. This is one of key results that leads to $\sqrt{NT}$ consistency for β^0. In contrast, the augmented regressors in Pesaran (2004) do not guarantee a full rank rotation of F^0. Therefore, the Pesaran estimator is in general $\sqrt{N}$ consistent. We now characterize the behavior of $\hat{\beta}$.

Proposition 5.3 *Assume assumptions A-D hold. In addition,* ε_{it} *have no time series correlation and heteroskedasticity, i.e.,* $E(\varepsilon_{it}\varepsilon_{js}) = 0$ *for* $t \neq s$ *and* $E\varepsilon_{it}\varepsilon_{jt} = \sigma_{ij}$. *If* $T/N^2 \to 0$, *then*

$$\sqrt{NT}(\hat{\beta} - \beta^0) = D(\hat{F})^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Big[X_i'M_{\hat{F}} - \frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k'M_{\hat{F}}\Big]\varepsilon_i + o_p(1)$$

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

The representation above still involves estimated F. If we assume N is much larger than T such that $T/N \to 0$, the estimated F can be replaced by the true F^0 in the limit. We have

Proposition 5.4 *Under the conditions of the previous proposition, if $T/N \to 0$, then $\hat{F}$ can be replaced by F^0 such that*

$$\sqrt{NT}(\hat{\beta} - \beta^0) = D(F^0)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Big[X_i' M_{F^0} - \frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k' M_{F^0} \Big] \varepsilon_i + o_p(1)$$

A more compact representation is

$$\sqrt{NT}(\hat{\beta} - \beta^0) = \Big(\frac{1}{NT} \sum_{i=1}^{N} Z_i' Z_i \Big)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i' \varepsilon_i + o_p(1) \tag{12}$$

where

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

The above result assumes the absence of time series correlation and heteroskedasticity for ε_{it} but it permits cross-section correlation and heteroskedasticity. 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.

Proposition 5.4 requires N to be much larger than T, a reasonable assumption for microeconomic data sets. The role of this requirement is to make negligible an asymptotic bias term that is order of $\sqrt{NT}/N$. Thus the purpose of $T/N \to 0$ is to center the asymptotic distribution at zero. When the main concern is not the asymptotic distribution but the rate of convergence, we can allow serial correlation and heteroskedasticity and still obtain $\sqrt{NT}$ consistency under the assumption of equal order of magnitude of N and T,

Proposition 5.5 *Assume assumptions A-D hold. In the presence of correlations and heteroskedasticities in both dimensions (serial and cross-sectional), if N and T are comparable such that $T/N \to \rho > 0$, then*

$$\sqrt{NT}(\hat{\beta} - \beta^0) = O_p(1).$$

Although the estimator is $\sqrt{NT}$ consistent, the underlying limiting distribution will not be centered at zero; asymptotic biases exist. In the next section, we derive the forms of biases and and show they can be consistently estimated and corrected, as is done in Hahn and Kuersteiner (2002) and Hahn and Newey (2004).

The focus so far has been the $o_p(1)$ representation. These representations are more informative than limiting distributions, as the latter disregards "closeness" in norms. Nevertheless, the limiting distributions are useful for inference. To this end, we need additional assumptions.

In view of the representation (12), in order to have asymptotic normality, we need the central limit theorem for $(NT)^{-1/2} \sum_{i=1}^{N} Z_i' \varepsilon_i$. Its variance is given by

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

This variance is indeed $O(1)$ because $\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |\sigma_{ij}| \leq M$ by assumption.

Assumption E: For some nonrandom positive definite matrix D_Z

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

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

In the absence of cross-section correlation such that $E(\varepsilon_i\varepsilon_j') = 0$ for $i \neq j$, we assume

$$\text{plim}\frac{1}{NT}\sum_{i=1}^{N}\sigma_i^2\sum_{t=1}^{T}Z_{it}Z_{it}' = D_Z \tag{13}$$

Theorem 5.6 *Assume assumptions A-E hold. In addition,* $E(\varepsilon_{it}\varepsilon_{js}) = 0$ *for* $t \neq s$*, and* $E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij}$ *for all* i, j *and* t*. As* $T, N \to \infty$ *with* $T/N \to 0$*, then*

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

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

As a corollary of the theorem, noting $D_Z = \sigma^2 D_0$ under iid assumption of ε_{it}, it follows that

Corollary 5.7 *Under the assumptions of Theorem 5.6, 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).

Theorem 5.6 requires $T/N \to 0$. If N and T are comparable such that $T/N \to \rho > 0$, then the limiting distribution is not centered at zero. We have

Theorem 5.8 *Under the assumptions of Theorem 5.6, together with* $T/N \to \rho > 0$*,*

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

where

$$B_0 = \text{plim}\, D(F^0)^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{(X_i - V_i)'F^0}{T}\Big(\frac{F^{0\prime}F^0}{T}\Big)^{-1}\Big(\frac{\Lambda'\Lambda}{N}\Big)^{-1}\lambda_k\sigma_{ik}$$

and $V_i = \frac{1}{N}\sum_{j=1}^{N}a_{ij}X_j$ *and* $a_{ij} = \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_j$.

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. This fact alone entitles its interpretation as a within group estimator. It is more instructive, however, to compare the mathematical expressions of the two estimators. Write the additive effects model (1) in matrix form:

$$Y = \beta_1 X^1 + \beta_2 X^2 + \cdots + \beta_p X^p + \iota_T\, \alpha' + \xi\, \iota_N' + \varepsilon \tag{14}$$

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 (14) 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 (3) 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 follows from some elementary 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 Proposition 5.4 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. Left multiply Z_i' on each side of the following

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

we obtain, noting $Z_i'F=0$,

$$Z_i'Y_i = Z_i'X_i\beta + Z_i'\varepsilon.$$

Summing over i and solving for β we obtain the instrument variable estimator

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

Moreover, it is easy to show $\sum_{i=1}^N Z_i'X_i=\sum_{i=1}^N Z_i'Z_i$. Thus the instrumental variable estimator has the same form as the asymptotic representation of the interactive effects estimator. It follows that the latter estimator is an asymptotically IV estimator with Z_i as instruments.

7 Bias corrected estimator

Unless $T/N\to 0$, asymptotic bias exists as stated in Theorem 5.8. The asymptotic $o_p(1)$ representation leading to Theorem 5.8 is the following

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

where

$$\xi_{NT}=-D(F^0)^{-1}\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^N\frac{(X_i-V_i)'F^0}{T}\Big(\frac{F^{0\prime}F^0}{T}\Big)^{-1}(\frac{\Lambda'\Lambda}{N})^{-1}\lambda_k(\frac{1}{T}\sum_{t=1}^T\varepsilon_{it}\varepsilon_{kt}) \tag{15}$$

with $V_i=\frac{1}{N}\sum_{j=1}^N a_{ij}X_j$. It is easy to show that $\xi_{NT}=O_p(1)$ and thus $\sqrt{T/N}\xi_{NT}$ does not affect the limiting distribution when $T/N\to 0$. But it becomes non-negligible if $T/N\to\rho\neq 0$. The expected value of ξ_{NT} is equal to, assuming no cross-section correlation in ε_{ik} such that $\sigma_{ik}=0$ for $i\neq k$ and $\sigma_{ii}=\sigma_i^2$,

$$B=-D(F^0)^{-1}\frac{1}{N}\sum_{i=1}^N\frac{(X_i-V_i)'F^0}{T}\Big(\frac{F^{0\prime}F^0}{T}\Big)^{-1}(\frac{\Lambda'\Lambda}{N})^{-1}\lambda_i\sigma_i^2 \tag{16}$$

This term represents the asymptotic bias. The bias can be estimated by replacing F^0 with $\hat{F}$, λ_i by $\hat{\lambda}_i$, and σ_i^2 by $\frac{1}{T}\sum_{t=1}^T\hat{\varepsilon}_{it}^2$. This gives, in view of $\hat{F}'\hat{F}/T=I_r$,

$$\hat{B}=-\hat{D}_0^{-1}\frac{1}{N}\sum_{i=1}^N\frac{(X_i-\hat{V}_i)'\hat{F}}{T}(\frac{\hat{\Lambda}'\hat{\Lambda}}{N})^{-1}\hat{\lambda}_i\hat{\sigma}_i^2 \tag{17}$$

We show in the appendix that $[\sqrt{T/N}](\hat{B}-B)=o_p(1)$. The biased corrected estimator is

$$\hat{\hat{\beta}}=\hat{\beta}-\frac{1}{N}\hat{B}$$

Theorem 7.1 *Assume assumptions A-E hold. In addition, $E(\varepsilon_{it}^2) = \sigma_i^2$, and $E(\varepsilon_{it}\varepsilon_{js}) = 0$ for $i \neq j$ or $t \neq s$. If $T/N^2 \to 0$,*

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

In comparison with Theorem 5.6, the condition $T/N \to 0$ is replaced by $T/N^2 \to 0$. In comparison with Theorem 5.8, the bias is removed and the distribution is centered at zero.

We also show in the appendix that $\sqrt{T/N}(\xi_{NT} - B) = o_p(1)$, so that biased corrected estimator does not increase variance.

The preceding analysis assumes no time series correlation and heteroskedasticity for ε_{it}. When they are present, additional bias arises. In the proof of Proposition 5.5, we show that $\sqrt{NT}(\hat{\beta} - \beta)$ has a bias term being order of $\sqrt{NT}/T$ when serial correlation or heteroskedasticity exists. We consider correcting for time series heteroskedasticity, maintaining the assumption of no serial correlation. Thus it is assumed that $E(\varepsilon_{it}^2) = \sigma_{i,t}^2$ and $E(\varepsilon_{it}\varepsilon_{js}) = 0$ for $i \neq j$ or $t \neq s$, ruling out correlations in either dimension but allowing heteroskedasticities in both dimensions.

The corresponding bias term is equal to $[\sqrt{NT}/T]C$, where [8]

$$C = -D(F^0)^{-1} \frac{1}{NT} \sum_{i=1}^{N} X_i' M_{F^0}\, \Omega\, F^0 (F^{0\prime} F^0/T)^{-1} (\Lambda'\Lambda/N)^{-1} \lambda_i \tag{18}$$

and $\Omega = \text{diag}(\frac{1}{N}\sum_{k=1}^N \sigma_{k,1}^2, ..., \frac{1}{N}\sum_{k=1}^N \sigma_{k,T}^2)$. Note that if $\sigma_{k,t}^2$ does not vary with t (no heteroskedasticity in the time dimension), then Ω is a scalar multiple of an identity matrix I_T. From $M_{F^0}F^0 = 0$, we have $C = 0$. Term C can be estimated by

$$\hat{C} = -\hat{D}_0^{-1} \frac{1}{NT} \sum_{i=1}^{N} X_i' M_{\hat{F}}\, \hat{\Omega}\, \hat{F} (\hat{\Lambda}'\hat{\Lambda}/N)^{-1} \hat{\lambda}_i \tag{19}$$

where $\hat{\Omega} = \text{diag}(\frac{1}{N}\sum_{k=1}^N \hat{\varepsilon}_{k,1}^2, ..., \frac{1}{N}\sum_{k=1}^N \hat{\varepsilon}_{k,T}^2)$.

Define

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

Theorem 7.2 *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*

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

where $D_2 = \text{plim} \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T Z_{it} Z_{it}' \sigma_{i,t}^2$.

In this theorem, the limiting variance involves D_2 instead of D_Z. This is due to no-constant variance in the time dimension, not due to biased correction. The correction does not contribute to the variance of the limiting distribution. Also note that, an additional condition

[8] Assume $\text{plim} C = C_0$ for some C_0, then in the presence of time series heteroskedasticity, we have in combination with Theorem 5.8,

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

The asymptotic bias is $\rho^{1/2}B_0 + \rho^{-1/2}C_0$.

$N/T^2 \to 0$ is added. Clearly, the conditions $T/N^2 \to 0$ and $N/T^2 \to 0$ are much 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.

Consistent estimation of 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_Z. We only limit our attention to the case of no cross-section correlation for ε_{it}, but heteroskedasticity is allowed. In this case, a consistent estimator for D_Z

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

where $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$ and $\hat{Z}_{it}$ is defined earlier. In the further presence of time series heteroskedasticity, we need an estimate for D_2,

$$\hat{D}_2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}'\hat{\varepsilon}_{it}^2$$

Proposition 7.3 *Assume assumptions A-E hold. Then as $N, T \to \infty$,*

(i) $\hat{D}_0 \xrightarrow{p} D_0$, *where* $D_0 = \operatorname{plim} D(F^0)$

(ii) $\hat{D}_Z \xrightarrow{p} D_Z$, *where D_Z is defined in (13)*

(iii) $\hat{D}_2 \xrightarrow{p} D_2$, *where D_2 is defined in Theorem 7.2.*

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, we expect the resulting estimator is less efficient. This is indeed the case and is useful in discerning additive versus interactive effects, a topic to be discussed in the next section. 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

$$Y_{it} = X_{it}'\beta + \mu + \alpha_i + \xi_t + \lambda_i'F_t + \varepsilon_{it} \tag{20}$$

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

$$\sum_{i=1}^{N}\alpha_i = 0, \quad \sum_{t=1}^{T}\xi_t = 0 \tag{21}$$

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

$$F'F/T = I_r, \quad \Lambda'\Lambda = diagonal. \tag{22}$$

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

$$\sum_{i=1}^{N} \lambda_i = 0, \quad \sum_{t=1}^{T} F_t = 0. \tag{23}$$

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 $F_t^\dagger = F_t - 2\bar{F}$, we have

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

where $\alpha_i^\dagger = \alpha_i + 2\bar{F}'\lambda_i - 2\bar{\lambda}'\bar{F}$, and $\xi_t^\dagger = \xi_t + 2\bar{\lambda}'F_t - 2\bar{\lambda}'\bar{F}$. Then 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 (20) if (23) is not imposed.

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

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

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

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 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 (21)-(23). Concentrating out $(\mu,\ \{\alpha_i\},\ \{\xi_t\})$ is equivalent to using $\dot{Y}_{it}$ and $\dot{X}_{it}$ to estimate the remaining parameters. That is, the concentrated objective function becomes

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

The dotted variable for $\lambda_i'F_t$ is itself, i.e., $\dot{\overline{\lambda_i'F_t}} = \lambda_i'F_t$ due to restriction (23). This objective function is the same as (7), 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 side is zero, and so $\iota_T'\hat{F} = 0$. The same argument leads to $\sum_{i=1}^{N}\hat{\lambda}_i = 0$.

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

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

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

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

We assume

$$\inf_F \dot{D}(F) > 0. \tag{24}$$

Define $\dot{Z}_i(\hat{F})$ as $\dot{Z}_i(F)$ with F replaced by $\hat{F}$, and let $\dot{Z}_i = \dot{Z}_i(F^0)$. Noticing 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,

Proposition 8.1 *Assume assumptions of Proposition 5.3 hold together with (21)-(24).*
(i) If $T/N^2 \to 0$, then

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

(ii) If $T/N \to 0$ then $\hat{F}$ can be replaced by F^0

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

In the appendix, we show the following mathematical identity

$$\sum_{i=1}^{N} \dot{Z}_i(\hat{F})'\dot{\varepsilon}_i \equiv \sum_{i=1}^{N} \dot{Z}_i(\hat{F})'\varepsilon_i \tag{25}$$

that is, $\dot{\varepsilon}$ can be replaced by ε_i. Under the restrictions (21) and (23), the following is also a mathematical identity,

$$\sum_{i=1}^{N} \dot{Z}_i'\dot{\varepsilon}_i \equiv \sum_{i=1}^{N} \dot{Z}_i'\varepsilon_i \tag{26}$$

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 $\dot{D}_Z > 0$ and $\dot{D}_Z = \operatorname{plim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N} \sigma_{ij}\dot{Z}_i'\dot{Z}_j$

Theorem 8.2 *Under assumptions A-F, we have if $T/N \to 0$,*

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

If $T/N \to \rho > 0$, the asymptotic distribution is not centered at zero. Bias corrected estimators can also be considered. Because the analysis is the same as before with X_i replaced by $\dot{X}_i$, the details are omitted.

9 Testing interactive versus non-interactive effects

Two approaches will be considered to evaluate which specification, fixed effects or interactive effect, gives better description of the data. The first approach is based on Hausman test statistic (Hausman, 1978) and the second is based on the number of factors. Throughout this section, for simplicity, we assume ε_{it} are iid over i and t, and $E(\varepsilon_{it}^2) = \sigma^2$. Also, we assume $T/N \to 0$ so that the limiting distribution of the interactive effects estimator is centered at zero. We discuss Hausman's test first.

9.1 Time-invariant vs time-varying individual effects

Consider the null hypothesis of fixed effects model:

$$Y_{it} = X_{it}'\beta + \lambda_i + \varepsilon_{it} \tag{27}$$

where λ_i is an unobservable scalar. The alternative hypothesis is that the fixed effects is time-varying

$$Y_{it} = X_{it}'\beta + \lambda_i F_t + \varepsilon_{it}. \tag{28}$$

where F_t is also an unobservable scalar. This is a single factor interactive effects model. If $F_t = 1$ for all t, fixed effects model is obtained.

The interactive effects estimator for β is consistent under both models (27) and (28), but is less efficient than the least squares dummy variable estimator for model (27), as the latter imposes the restriction $F_t = 1$ for all t. Nevertheless, the fixed effects estimator is inconsistent under model (28). The principle of the Hausman test is applicable here.

The least squares dummy variable estimator is

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

where $M_T = I_T - \iota_T\iota_T'/T$. For the interactive model, the estimator is

$$\sqrt{NT}(\hat{\beta}_{IE} - \beta) = D(F^0)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Big[X_i'M_{F^0} - \frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k'M_{F^0}\Big]\varepsilon_i + o_p(1)$$

Let

$$\eta = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i'M_{F^0}\varepsilon_i, \quad \xi = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Big[\frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k'M_{F^0}\Big]\varepsilon_i. \tag{29}$$

By Proposition 5.4

$$\sqrt{NT}(\hat{\beta}_{IE} - \beta) = D(F^0)^{-1}(\eta - \xi) + o_p(1). \tag{30}$$

Under the null hypothesis, $F^0 = \iota_T$, and thus $M_T = M_{F^0}$ and

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

where $A = (\frac{1}{NT}\sum_{i=1}^{N} X_i'M_T X_i)$.

The variances of the two estimators (the conditional variance to be precise) are

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

respectively. To show the variance of the difference in estimators is equal to the difference in variances, i.e.,

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

it suffices to show

$$E(\eta\xi') = E(\xi\xi') \tag{31}$$

This is proved in the appendix. Note that $E\xi\xi'$ is positive definite, this is, $A - D(F^0) = E\xi\xi'$ is positive definite. This implies that $\text{var}(\sqrt{NT}(\hat{\beta}_{IE} - \hat{\beta}_{FE}) = \sigma^2[D(F^0)^{-1} - A^{-1}]$ is matrix of full rank (positive definite). Thus

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

Replacing $D(F^0)$ and σ^2 by their consistent estimators, the above is still true. Proposition 7.3 shows that $D(F^0)$ is consistently estimated by $\hat{D}_0$, and 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 $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

9.2 Homogeneous vs heterogeneous time effects

For reasons of comparison, the usual time effects is called homogeneous-time effects since it is the same across individuals:

$$Y_{it} = X_{it}\beta + F_t + \varepsilon_{it},$$

where F_t is unobservable scalar. The heterogeneous time-effects model is the following

$$Y_{it} = X_{it}\beta + \lambda_i F_t + \varepsilon_{it}$$

which is a simple interactive effects model with $r = 1$. The least-squares dummy-variable method for the homogeneous effects gives

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

where $B = (\frac{1}{NT}\sum_{i=1}^N (X_i - \bar{X})'(X_i - \bar{X}))$ and $\psi = \frac{1}{\sqrt{NT}}\sum_{i=1}^N (X_i - \bar{X})'\varepsilon_i$, and $\bar{X} = \frac{1}{N}\sum_{i=1}^N X_i$, a $T \times 1$ vector. The interactive effects estimator has the same representation as in (30). Under the null hypothesis of homogeneous time effect, we have $\lambda_i = 1$ for all i and hence $a_{ik} = 1$. It follows that

$$\mathrm{var}(\eta - \xi) = \sigma^2 D(F^0) = \sigma^2 \frac{1}{NT}\sum_{i=1}^N X_i' M_{F^0} X_i - \sigma^2 \frac{1}{T}\bar{X}' M_{F^0} \bar{X}$$

In the appendix, it is shown that

$$E\eta\psi' = \mathrm{var}(\eta - \xi) = \sigma^2 D(F^0), \quad E(\xi\psi') = 0 \tag{32}$$

This implies that

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

Thus Hausman's test takes the form

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

The above still holds with $D(F^0)$ and σ^2 replaced by $\hat{D}_0$ and $\hat{\sigma}^2$.

9.3 Additive vs interactive effects

The null hypothesis is the additive effects model

$$Y_{it} = X_{it}\beta + \alpha_i + \xi_t + \mu + \varepsilon_{it} \tag{33}$$

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

$$Y_{it} = X_{it}\beta + \lambda_i' F_t + \varepsilon_{it} \tag{34}$$

The null model is nested in the general model with

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

The least squares estimator of β in (33) 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 $\psi = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \dot{X}_i'\varepsilon_i$. Note that

$$var[\sqrt{NT}(\hat{\beta}_{FE} - \beta)] = \sigma^2 C^{-1}$$

The interactive effects estimator again can be written as

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

where η and ξ have the same expression as in (29), although F^0 is now a matrix instead of a vector. In the appendix, we show, under the null hypothesis,

$$E[(\eta - \xi)\psi'] = \sigma^2 D(F^0) \tag{35}$$

This again implies

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

Thus Hausman's test takes the form

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

9.4 The number of factors

In this section we argue why the number of factors can be consistently estimated and how to use this fact to discern additive and interactive effects. For pure factor models, Bai and Ng (2002) show that the number of factors can be consistently estimated based on information criterion approach. Their analysis can be amended to our current setting. Details will not be presented to avoid repetition, but intuition will be given.

We assume that $r \leq \bar{k}$, where $\bar{k}$ is given. Suppose r is unknown, but we entertain k factors in the estimation. It can be shown that as long as $k \geq r$, we have $\hat{\beta}_{IE}^{(k)} - \beta = O_p(1/\sqrt{NT})$, where the superscript k indicates k factors are estimated. Let $\hat{u}_{it}(k) = Y_{it} - X_{it}'\hat{\beta}_{IE}^{(k)}$, and $\hat{\varepsilon}_{it}(k) = \hat{u}_{it}(k) - \hat{\lambda}_i(k)'\hat{F}_t(k)$. Then

$$\hat{u}_{it}(k) = \lambda_i' F_t + \varepsilon_{it} + O_p(1/\sqrt{NT})$$

thus $\hat{u}_{it}$ has a pure factor model; the $O_p(1/\sqrt{NT})$ error will not affect the analysis of Bai and Ng (2002). This means that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2(k) - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2 = O_p(1/\min[N,T])$$

Since $\overline{k} \geq r$, the above is true when k is replaced by $\overline{k}$. Thus,

$$\hat{\sigma}^2(k) - \hat{\sigma}^2(\overline{k}) = O_p(1/\min[N,T])$$

where $\hat{\sigma}^2(k) = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \hat{\varepsilon}_{it}^2(k)$.

If $k < r$, unless $\lambda_i' F_t$ are uncorrelated with the regressors and $E(\lambda_i) = 0$ and $E(F_t) = 0$, β cannot be consistently estimated. In any case, F^0 cannot be consistently estimated since F^0 is $T \times r$ and $\hat{F}(k)$ is only $T \times k$. The consequence of inconsistency is

$$\hat{\sigma}^2(k) - \hat{\sigma}^2(\overline{k}) > c > 0$$

for some $c > 0$, not depending on N and T. This implies that any penalty function that converges to zero but is of greater magnitude than $O_p(1/\min[N,T])$ will lead to consistent estimation of the number of factors. In particular,

$$CP(k) = \hat{\sigma}^2(k) + \hat{\sigma}^2(\overline{k})\Big[k(N+T) - k^2\Big]\frac{\log(NT)}{NT}$$

or

$$IC(k) = \log\hat{\sigma}^2(k) + \Big[k(N+T) - k^2\Big]\frac{\log(NT)}{NT}$$

will work. That is, let $\hat{k} = \text{argmin}_{k \leq \overline{k}} PC(k)$, or $\hat{k} = \text{argmin}_{k \leq \overline{k}} IC(k)$, then $P(\hat{k} = r) \to 1$ as $N, T \to 0$. Although the usual BIC criterion only assumes either $T \to \infty$ or $N \to \infty$ but not both, the $IC(k)$ has the same from as the BIC criterion as there are a total of NT observations. With k factors, the number of parameters is $k(N+T) - k^2 + p$, where k^2 reflects the restriction $F'F/T = I$ and $\Lambda'\Lambda = diagonal$, but p does not vary with k so can be excluded in the penalty function. The CP criterion is similar to that of Mallows' C_p.

Ignore k^2 for a moment (since it is dominated by $k(N+T)$ for large N and T), the penalty function in $IC(k)$ is $k \cdot g(N,T)$, where $g(N,T) = (N+T)\frac{\log(NT)}{NT}$. Clearly, the penalty function goes to zero as $N, T \to 0$, unless $N = \exp(T)$ or $T = \exp(N)$ (these are the rare situations where BIC breaks down. Bai and Ng (2002) suggest several alternative criteria). In addition, $g(N,T)$ is of larger magnitude than $1/\min[N,T]$ since $g(N,T) * \min[N,T] \to \infty$. These two properties of a penalty function imply consistency, as shown by Bai and Ng (2002).

Given that the number of factors can be consistently estimated, we can determine whether an additive model or interactive model is more appropriate. Suppose the null hypothesis postulates time-invariant fixed effects as $Y_{it} = X_{it}'\beta + \lambda_i + \varepsilon_{it}$. Then

$$Y_{it} - \bar{Y}_{i.} = (X_{it} - \bar{X}_{i.})'\beta + \varepsilon_{it} - \bar{\varepsilon}_{i.}$$

Under the time-varying fixed effects model $Y_{it} = X_{it}'\beta + \lambda_i F_t + \varepsilon_{it}$, we have

$$Y_{it} - \bar{Y}_{i.} = (X_{it} - \bar{X}_{i.})'\beta + \lambda_i(F_t - \bar{F}) + \varepsilon_{it} - \bar{\varepsilon}_{i.}$$

Under the null hypothesis, no factor exists, and under the alternative, there exists one factor.

The same argument works for the fixed time effects model, in which we use $Y_{it} - \bar{Y}_{.t}$ as the left-hand side variable and $X_{it} - \bar{X}_{.t}$ as the right hand side variable.

Nest consider the additive vs the interactive model:

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

or

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

where $\dot{Y}_{it}$ and $\dot{X}_{it}$ are defined previously. Therefore, the transformed data exhibit no factors. Under the interactive model (34), the transformed data obey

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

The factor structure is unscathed by the transformation and the number of factors is still two.

10 Time-invariant and common regressors

In panel data analysis, time-invariant regressors and common regressors are more often than not the variables of primary interest. In earnings studies, time-invariant regressors include education, gender, race, eta; common variables are those representing trends or policies. In consumption studies, common regressors include price variables which are the same for each individual. For models with additive fixed effects, those variables are removed along with the fixed effects 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, an approach that may not be desirable. While interactive-effect models permit correlation between factor loadings and regressors, desirable identification conditions should be those that are valid also under the ideal situation in which factor loadings are not correlated with regressors. This section examines such identification conditions when the model is estimated by the least squares method.

A general model can be written as

$$Y_{it} = X_{it}'\varphi + x_i'\gamma + w_t'\delta + \lambda_i' F_t + \varepsilon_{it} \tag{36}$$

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 over all possible F. This assumption fails when time invariant regressors and common regressors exist. This follows from $D(\iota_T)$ and $D(W)$ are not full rank matrices. Fortunately, 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. Given all other assumptions and identifying restrictions, we show that this condition is in effect a necessary and sufficient condition for identification.

We now explain the meaning of $D(F^0) > 0$ and argue that it can be segregated into some intuitive and reasonable conditions. To simply notation and for easy 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 cannot contain ι_T; Λ or its rotation cannot contain ι_N. Otherwise, we are back into the environment of Hausman and Taylor, instrumental variables must be used to identify β. In 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

$$\frac{1}{N}(\underline{x}, \iota_N)' M_\Lambda (\underline{x}, \iota_N) > 0 \tag{37}$$

$$\frac{1}{T}(\iota_T, W)' M_{F^0} (\iota_T, W) > 0 \tag{38}$$

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 $D(F)$,

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

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 (37) and (38) are 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. This implies that the four identification conditions, which are necessary, are also sufficient for identification since $D(F^0) > 0$ implies identification (to be shown later).

Lemma 10.1 *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

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

Remark: $\bar{B}$ needs not be positive definite. For example, for $\ell = 1$, $\bar{B}$ can be the 2×2 matrix with each entry being 1. Then $\bar{A} \diamondsuit \bar{B} = \bar{A} > 0$. The lemma holds if $\bar{A} \geq 0$ with $A > 0$ and $\tau > 0$, but $\bar{B} > 0$ (reversing the role of $\bar{A}$ and $\bar{B}$). Moreover, from $\bar{A} \diamondsuit \bar{B} > 0$, one can deduce the condition of the lemma (or the conditions reversing the role of $\bar{A}$ and $\bar{B}$). In this sense, the condition is necessary and sufficient. The operator $\diamondsuit$ is analogous to the Hadamard product, which requires equal size for $\bar{A}$ and $\bar{B}$ and is defined as componentwise multiplication. We are not aware of any matrix result in this nature. The lemma can be proved for $\ell = 1$ and for arbitrary q, then with induction over ℓ (the proof is available from the author).

Apply the lemma with $A = \frac{1}{N}\underline{x}'M_\Lambda\underline{x} > 0$, $\tau = \iota_N'M_\Lambda\iota_N > 0$, $\nu = \frac{1}{T}\iota_T'M_{F^0}\iota_T > 0$, and $B = W'M_{F^0}W/T > 0$. For $\bar{A} = \frac{1}{N}(\underline{x}, \iota_N)'M_\Lambda(\underline{x}, \iota_N)$ and $\bar{B} = \frac{1}{T}(\iota_T, W)'M_{F^0}(\iota_T, W)$, we have $D(F^0) = \bar{A} \diamondsuit \bar{B} > 0$ by the lemma. Thus the four conditions imply $D(F^0) > 0$.

It remains to argue that $D(F^0) > 0$ (or equivalently, the four conditions above) implies identification and consistent estimation. Denote the true value by (β^0, F^0). Recall the objective function can be written as $S_{NT}(\beta, F) = \tilde{S}_{NT}(\beta, F) + o_p(1)$, where

$$\tilde{S}_{NT}(\beta, F) = (\beta - \beta^0)'D(F)(\beta - \beta^0) + \theta'B\theta$$

where $B = [(\Lambda'\Lambda/N)^{-1} \otimes I_T] > 0$, and θ is a function of (β, F) such that

$$\theta = vec(M_F F^0) + B^{-1}\frac{1}{NT}\sum_{i=1}^{N}(\lambda_i \otimes M_F X_i)(\beta - \beta^0), \tag{39}$$

see the proof of Proposition 5.1 in the appendix. Since $D(F)$ is semi-positive definite for any F, and B is positive definite,

$$\tilde{S}_{NT}(\beta, F) \geq 0$$

for all (β, F). On the other hand, $\tilde{S}_{NT}(\beta^0, F^0) = 0$. We show (β^0, F^0) is the unique point at which $\tilde{S}_{NT}(\beta, F)$ achieves its minimum, where uniqueness with respect to F^0 is up to a rotation (identification restrictions on F and Λ in fact fixes the rotation). Let

$$(\beta^*, F^*) = \text{argmin}\tilde{S}_{NT}(\beta, F)$$

we show $(\beta^*, F^*) = (\beta^0, F^0)$. Since $\tilde{S}_{NT}(\beta^*, F^*) = 0$, we must have

$$(\beta^* - \beta^0)'D(F^*)(\beta^* - \beta^0) = 0 \quad \text{and} \quad \theta^* = \theta(\beta^*, F^*) = 0$$

If $D(F^*)$ is of full rank, then $\beta^* - \beta^0 = 0$. In this case, from $0 = \theta^* = vec(M_{F^*}F^0)$, we have $F^* = F^0$. Only when $D(F^*)$ is not full rank is it possible for $\beta^* \neq \beta^0$. The matrix $D(F^*)$ will not be full rank if F^* or its rotation contains the column ι_T, or contains a column of W. We show this is not possible under $D(F^0) > 0$. If F^* contains the column ι_T, then

$$D(F^*) = \begin{bmatrix} 0 & 0 \\ 0 & (\frac{1}{N}\iota_N' M_\Lambda \iota_N)(W'M_{F^*}W)/T \end{bmatrix}$$

and it follows that

$$0 = (\beta^* - \beta^0)D(F^*)(\beta^* - \beta^0) = a(\delta^* - \delta^0)'(W'M_{F^*}W/T)(\delta^* - \delta^0)$$

where $a = \frac{1}{N}\iota_N' M_\Lambda \iota_N > 0$. The above implies that

$$M_{F^*}W(\delta^* - \delta^0) = 0$$

since $x'x = 0$ implies $x = 0$. Therefore,

$$M_{F^*}X_i(\beta^* - \beta^0) = (M_{F^*}\iota_T x_i', M_{F^*}W)(\beta^* - \beta^0) = (0, M_{F^*}W)(\beta^* - \beta^0) = M_{F^*}W(\delta^* - \delta^0) = 0$$

Thus, by (39), $0 = \theta^* = vec(M_{F^*}F^0)$. It follows that $F^* = F^0$, thus F^* cannot contain ι_T since F^0 does not contain ι_T, a contradiction. Next, suppose that F^* contains at least one column of W. Partition $W = (W_1, W_2)$ and suppose, without loss of generality, F^* contains W_2. Then $M_{F^*}W = (M_{F^*}W_1, 0)$, and

$$D(F^*) = \begin{bmatrix} (\frac{1}{N}\underline{x}'M_\Lambda \underline{x})(\iota_T' M_{F^*}\iota_T/T) & (\frac{1}{N}\underline{x}'M_\Lambda \iota_N)(\iota_T' M_{F^*}W_1/T) & 0 \\ (W_1' M_{F^*}\iota_T/T)(\frac{1}{N}\iota_N' M_\Lambda \underline{x}) & (\frac{1}{N}\iota_N' M_\Lambda \iota_N)(W_1' M_{F^*}W_1/T) & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Under $\frac{1}{T}\iota_T' M_{F^*}\iota_T > 0$, the non-zero diagonal block of $D(F^*)$ is positive definite by Lemma 10.1. Partition $\delta = (\delta_1', \delta_2')'$ so $\beta = (\gamma', \delta_1', \delta_2')'$. Partition β^* and β^0 correspondingly. From

$$(\beta^* - \beta^0)' D(F^*)(\beta^* - \beta^0) = 0$$

we have $\gamma^* - \gamma^0 = 0$ and $\delta_1^* - \delta_1^0 = 0$. Thus $\beta^* - \beta^0 = (0', 0', \delta_2^{*\prime} - \delta_2^{0\prime})$. Together with $M_{F^*}W_2 = 0$, we have

$$M_{F^*}X_i(\beta^* - \beta^0) = (M_F^* \iota_T x_i', M_{F^*}W_1, 0)(\beta^* - \beta^0) = 0$$

In view of (39), $0 = \theta^* = vec(M_{F^*}F^0)$. It follows that $F^* = F^0$, again a contradiction. In summary, under the assumption that $D(F^0) > 0$, the optimal solution of $\tilde{S}_{NT}(\beta, F)$ is achieved uniquely at (β^0, F^0). This implies that $\hat{\beta}$ is consistent estimation for β^0, see the proof of Proposition 5.1 in the appendix.

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 Finite sample properties via simulations

Data are generated according to:

$$Y_{it} = X_{it,1}\beta_1 + X_{it,2}\beta_2 + a\lambda_i' F_t + \varepsilon_{it}$$

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

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

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

with $\iota' = (1, 1)$. The variables λ_{ij}, F_{tj}, $\eta_{it,j}$ are all iid $N(0, 1)$. The important parameters are

$$(\beta_1, \beta_2) = (1, 3).$$

We set $c_1 = c_2 = \mu_1 = \mu_2 = 1$ and $a = 1$. We first consider the case of

$$\varepsilon_{it} \quad iid \quad N(0, 4)$$

then extend to correlated errors.

To estimate $(\hat{\beta}_{IE}, \hat{F})$, consider the iteration scheme in (9) and (10). A staring value for β or F is needed. The least squares objective function is not globally convex, there is no guarantee that an arbitrary starting value will lead to the global optimal solution. Two natural choices exist. The first is the simple least squares estimator of β, ignoring the interactive effects. The second is the principal components estimator for F, ignoring the regressors. If λ_i and F_t have unusually large non-zero means (arbitrarily stretching the model), the first choice can fail, but the second choice leads to the optimal solution. This is because as the interactive effects become dominant, it makes sense to estimate the factor structure first. In this case, using the within estimator β as a starting value will also work. To minimize the chance of local minimum, both choices are used. Upon convergence, we choose the estimator that gives a smaller value of the objective function. Iterations based on (9) and (10) have difficulty of achieving convergence for models with time-invariant and common regressors.

A more robust iteration scheme (having much better convergence property) is the following: given F and Λ, 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, 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 $(\sum_{i=1}^N X_i' X_i)^{-1}$, no need of updating during iteration, unlike the scheme of $\hat{\beta}(F) = (\sum_{i=1}^N X_i' M_F X_i)^{-1} \sum_{i=1}^N X_i' M_F Y_i$. Furthermore, if $N > T$, we do principal components analysis using WW' $(T \times T)$; and if $N < T$, we use $W'W$ $(N \times N)$ to speed up computation. They give the same product $\hat{F}\hat{\lambda}_i$ no matter which matrix is used. For the model associated with Table 4, the iteration method in previous paragraph has many realizations not converging to global optimum, but for the iteration scheme here all lead to global solution.

For comparison, we also compute two additional estimators: (i) the usual within estimator $\hat{\beta}_{LSDU}$, (ii) infeasible estimator $\hat{\beta}(F)$, assuming F is observable.

From Table 1, we can draw several conclusions. First, the within estimator is biased and inconsistent. Biases become more severe when the interactive effects are magnified by setting a larger a. For example, if $a = 10$, the biases are also almost ten times larger (not reported). The infeasible estimator and the interactive effects estimator are virtually unaffected by the value of a. Second, both the feasible and interactive effects estimators are unbiased and consistent. The interactive effects estimator is less efficient than the infeasible estimator, as can be seen from the larger standard errors, which is consistent with the theory. Third, even with small N and T, the interactive effects estimator performs quite well, and both N and T increases, the standard deviation becomes smaller.

Table 2 gives results for cross-sectionally correlated ε_{it}. For cross-sectional data in reality, a large value of $|i - j|$ does not necessarily mean the correlation between ε_{it} and ε_{jt} is small. Nevertheless, for the purpose of introducing cross-section correlation, ε_{it} is generated as AR(1) for each fixed t such that

$$\varepsilon_{it} = \rho\varepsilon_{i-1,t} + e_{it}$$

where $\rho = 0.7$. Once cross-section correlation is introduced, the data can be permuted cross-sectionally if wanted, but the results do not depend on any particular permutation. We generate stationary data by discarding the first 100 observations. This implies that $var(\varepsilon_{it}) = \sigma_e^2/(1 - \rho^2) \approx 4$ for $\sigma_e^2 = 2$ and $\rho = 0.7$. Thus the variance of ε_{it} is approximately the same as the variance for Table 1. Theorem 5.6 claims that for $N >> T$, cross-section correlation does not affect consistency. On the other hand, for small N (no matter how large is T), the estimates are inconsistent. In fact, the model is unidentified as explained following equation (6), as long as the cross-section correlation is regarded as having an unknown form. The simulation results are consistent with those predictions.

Table 3 reports results when the true model has additive effects such that

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

Three estimators are computed: (1) the within estimator, which is efficient given additivity, (2) the infeasible estimator, which assumes $F_t = (1, \xi_t)'$ is observed; (3) the interactive effects estimator, which treats the additive effects as if they were interactive effects. Data are generated as in Table 1, except that effects are additive. All three estimators are consistent. For small N or small T, the interactive effects estimator shows some bias. These findings are consistent with the theory.

Table 4 presents results for models with grand mean, time-invariant regressors, common regressors, and regressors having variations in both dimensions. The model is

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

$$(\beta_1, \beta_2, \mu, \gamma, \delta) = (1, 3, 5, 2, 4)$$

where all variables are generated as in Table 1, and additionally, $x_i \sim \iota'\lambda_i + e_i$, $w_t = \iota' F_t + \eta_i$, with e_i and η_i are iid N(0,1) independent of all other regressors. The within estimator can only estimate β_1 and β_2 and is inconsistent (not reported). The cases of very small N or small T (say $T = 3$ and $T = 5$) have convergence problem, we thus consider cases with N and T no smaller than 10. The infeasible estimators and interactive effects estimators are all consistent, but the latter is less efficient than the former, as expected.

12 Concluding remarks

In this paper, we have derived the rate of convergence and the limiting distribution of the estimated common slope coefficients in panel data models with interactive effects. 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 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 show 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.

Appendix: 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, $T^{-1/2}\|F^0\| = O_p(1)$, and $T^{-1}\|\hat{F}\|^2 = r$, $T^{-1/2}\|\hat{F}\| = \sqrt{r}$, and $T^{-1}\|X_i'F^0\| = O_p(1)$, etc. Throughout, we define $\delta_{NT} = \min[\sqrt{N}, \sqrt{T}]$ so that $\delta_{NT}^2 = \min[N, T]$.

Lemma 12.1 *Under assumptions A-D,*

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

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

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

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

Proof. From $\frac{1}{NT}\sum_{i=1}^N X_i'\varepsilon_i = o_p(1)$, it is sufficient to show $\sup_F \frac{1}{NT}\sum_{i=1}^N X_i' P_F \varepsilon_i = o_p(1)$. Using $P_F = FF'/T$,

$$\frac{1}{NT}\|\sum_{i=1}^N X_i P_F \varepsilon_i\| = \|\frac{1}{N}\sum_{i=1}^N (\frac{X_i'F}{T})\frac{1}{T}\sum_{t=1}^T F_t \varepsilon_{it}\| \le \frac{1}{N}\sum_{i=1}^N \|\frac{X_i'F}{T}\| \cdot \|\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\|$$

Note $T^{-1}\|X_i'F\| \le T^{-1}\|X_i\| \cdot \|F\| = \sqrt{r}T^{-1/2}\|X_i\| \le \sqrt{r}(\frac{1}{T}\sum_{t=1}^T \|X_{it}\|^2)^{1/2}$ because $T^{-1/2}\|F\| = \sqrt{r}$. Thus the above is bounded by, using the Cauchy-Schwarz inequality,

$$\sqrt{r}\Big(\frac{1}{N}\sum_{i=1}^N \frac{1}{T}\sum_{t=1}^T \|X_{it}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N \|\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\|^2\Big)^{1/2}$$

The first expression is $O_p(1)$. It suffices to show the second term is $o_p(1)$ uniformly in F. Now

$$\frac{1}{N}\sum_{i=1}^N \|\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\|^2 = tr\Big(\frac{1}{N}\sum_{i=1}^N \frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s'\varepsilon_{it}\varepsilon_{is}\Big)$$

$$= tr\Big(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s'\frac{1}{N}\sum_{i=1}^N[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})]\Big) + tr\Big(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s'\frac{1}{N}\sum_{i=1}^N \sigma_{ii,ts}\Big)$$

where $\sigma_{ii,ts} = E(\varepsilon_{it}\varepsilon_{is})$. The first expression is bounded by Cauchy-Schwarz inequality,

$$\Big(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T \|F_t\|^2\|F_s\|^2\Big)^{1/2} N^{-1/2}\Big(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T \Big[\frac{1}{\sqrt{N}}\sum_{i=1}^N[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})]\Big]^2\Big)^{1/2}$$

But $T^{-1}\sum_{t=1}^T \|F_t\|^2 = \|F'F/T\| = r$. Thus above is equal to $rN^{-1/2}O_p(1)$. Next, $|\frac{1}{N}\sum_{i=1}^N \sigma_{ii,ts}| \le \tau_{ts}$ by Assumption C2. Again by the Cauchy-Schwarz inequality,

$$\|\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s'\frac{1}{N}\sum_{i=1}^N \sigma_{ii,ts}\| \le \Big(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T \|F_t\|^2\|F_s\|^2\Big)^{1/2}\Big(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T \tau_{ts}^2\Big)^{1/2}$$

$$= rT^{-1/2}\Big(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{ts}^2\Big)^{1/2} = rO(T^{-1/2})$$

the last equality follows from $\tau_{ts}^2 \le M\tau_{ts}$ and Assumption C2. The proof for the remaining statements are the same, thus omitted. Stock and Watson (2002) have similar results but they require $\|F_t\|$ be bounded uniformly in t, ruling out F_t with unbounded support. Our proof does not need bounded F_t, nor our optimization with respect to F_t needs to be taken over bounded set.

Proof of Proposition 5.1. Without loss of generality, assume $\beta^0 = 0$ (purely for notational simplicity), and from $Y_i = X_i\beta^0 + F^0\lambda_i + \varepsilon_i = F^0\lambda_i + \varepsilon_i$, expanding $S_{NT}(\beta, F)$,

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

where

$$\tilde{S}_{NT}(\beta,F) = \beta'\Big(\frac{1}{NT}\sum_{i=1}^{N}X_i'M_FX_i\Big)\beta + tr[(\frac{F^{0\prime}M_FF^0}{T})(\frac{\Lambda'\Lambda}{N})] + 2\beta'\frac{1}{NT}\sum_{i=1}^{N}X_i'M_FF^0\lambda_i \quad (40)$$

By Lemma 12.1

$$S_{NT}(\beta,F) = \tilde{S}_{NT}(\beta,F) + o_p(1) \quad (41)$$

uniformly over bounded β and over F such that $F'F/T = I$. Bounded β is in fact not necessary because the objective funtion 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 specify a unique H. We next show that for any $(\beta, F) \neq (\beta^0, F^0H)$, $\tilde{S}_{NT}(\beta,F) > 0$, thus $S_{NT}(\beta,F)$ attains its unique minimum value 0 at $(\beta^0, F^0H) = (0, F^0H)$.

Define

$$A = \frac{1}{NT}\sum_{i=1}^{N}X_i'M_FX_i, \quad B = (\frac{\Lambda'\Lambda}{N}\otimes I_T), \quad C = \frac{1}{NT}\sum_{i=1}^{N}(\lambda_i'\otimes M_FX_i)$$

and let $\eta = vec(M_FF^0)$ then

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

Completing square, we have

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

$$= \beta'D(F)\beta + \theta'B\theta$$

where $\theta = (\eta + B^{-1}C\beta)$. By Assumption A, $D(F)$ is positive definite, and B is also positive definite, so $\tilde{S}_{NT}(\beta,F) \ge 0$. In addition, if either $\beta \neq \beta^0 = 0$ or $F \neq F^0H$, then $\tilde{S}_{NT}(\beta,F) > 0$. Thus, $\tilde{S}_{NT}(\theta,F)$ achieves its unique minimum at (β^0, F^0H). Further, for $\|\beta\| \ge c > 0$, $\tilde{S}_{NT}(\beta,F) \ge \rho_{min}c^2 > 0$, where ρ_{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^0H. This is because F^0 is $T \times 1$, and as $T \to \infty$, the number of elements going to infinity, the usual consistency is not a well defined. Other notion 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 (41)

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

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 (40), it follows that the above implies

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

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

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

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

In all remaining proofs, β and β^0 are used interchangeably, and so are F and F^0.

Proof of Proposition 5.2. From

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

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

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

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

$$\hat{F}\,V_{NT} - F^0\,(\Lambda'\Lambda/N)(F^{0\prime}\hat{F}/T) = I1 + \cdots + I8 \qquad (42)$$

Multiplying $(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$ on each side of (42)

$$\hat{F}\left[V_{NT}(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}\right] - F^0 = (I1 + \cdots + I8)(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1} \tag{43}$$

Note 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

$$T^{-1/2}\|\hat{F}\left[V_{NT}(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}\right] - F^0\| \le T^{-1/2}(\|I1\| + \cdots \|I8\|)\|\cdot\|(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}\|$$

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

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

because $\|\hat{\beta} - \beta\| = o_p(1)$. Using the same argument, it is easy to prove next four terms ($I2$ to $I5$) 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 is proved in Bai and Ng (Theorem 1). The proof there only uses the property that $\hat{F}'\hat{F}/T = I$, and the assumptions on ε_i. The proof there needs no modification. In summary, we have

$$T^{-1/2}\|\hat{F}V_{NT}(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1} - F^0\| = O_p(\|\hat{\beta} - \beta\|) + O_p(1/\min[\sqrt{N}, \sqrt{T}]). \tag{44}$$

Proof of part (i). Left multiplying (42) 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 $\hat{F}'F^0/T$ is invertible, thus V_{NT} is invertible. To obtain the limit of V_{NT}, left multiplying (42) by $F^{0\prime}$ and then dividing by T,

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

because $T^{-1}F^{0\prime}(I1 + ... + I8) = o_p(1)$. The above equality shows that the columns of $F^{0\prime}\hat{F}/T$ are the (non-normalized) 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 $\Sigma_F\Sigma_\Lambda$.

Proof of part (ii). Since V_{NT} is invertible, the left side (44) can be written as $T^{-1/2}\|\hat{F}H^{-1} - F^0\|$, thus (44) 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 the cross product term from expanding the squares has the same bound.

Lemma 12.2 *Under assumptions A-C, there exist an $M < \infty$, such that*
(i)

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

(ii) for all $i = 1, 2..., N$ and $h = 1, 2, ..., r$,

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

Proof of (i). Denote the term inside $\|\cdot\|^2$ as A. Then the left hand side is equal to $Etr(AA')$. Using $E\|F_t\|^4 \le M$ and Assumption C.4, (i) follows readily. The proof of (ii) is similar.

Lemma 12.3 *Under assumptions of Proposition 5.3,*

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

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

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

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

Proof of (i). This part extends Lemma B.2 of Bai (2003). Using (43), it is easy to see that the first five terms are each $O_p(\hat{\beta}-\beta)$. In fact, the first, third and fifth are $o_p(\hat{\beta}-\beta)$, the second and fourth are $O_p(\hat{\beta}-\beta)$. The next three terms are considered in Bai (2003) and each is shown to be $O_p(\delta_{NT}^{-2})$ in the absence of β. With the estimation of β, they are each to be $O_p(\hat{\beta}-\beta)O_p(\delta_{NT}^{-1})+O_p(\delta_{NT}^{-2})$ due to Proposition 5.2(ii) instead of Lemma A.1 of Bai (2003). But $O_p(\hat{\beta}-\beta)O_p(\delta_{NT}^{-1})$ is dominated by $O_p(\hat{\beta}-\beta)$, the order of the first five terms. Thus summing over the eight terms, we obtain part (i).

For part (ii),

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

by part (i) and Proposition 5.2(ii). The proof of part (iii) is identical to part (i).

For (iv)

$$\frac{1}{NT}\sum_{i=1}^{N}X_i'M_{\hat{F}}(\hat{F}-F^0H) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}X_i'(\hat{F}-F^0H)+\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'\hat{F}}{T}\hat{F}'(\hat{F}-F^0H)$$

The first term on the right is an average of (iii) over i, and thus is still that order of magnitude. The second term is bounded by $\frac{1}{N}\sum_{i=1}^{N}\|X_i/\sqrt{T}\|^2\sqrt{r}\|T^{-1}\hat{F}'(\hat{F}-F^0H)\| = O_p(1)\|T^{-1}\hat{F}'(\hat{F}-F^0H)\|$. Thus (iv) follows from part (ii).

Lemma 12.4 *Under assumptions of Proposition 3,*
(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) $\frac{1}{NT}\sum_{k=1}^{N}\lambda_k'(\hat{F}H^{-1}-F^0)'\varepsilon_k = (NT)^{-1/2}O_p(\hat{\beta}-\beta)+O_p(N^{-1})+N^{-1/2}O_p(\delta_{NT}^{-2})$
(iv)

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

Proof of (i). Part (i) extends lemma B.1 of Bai (2003). The proof is omitted as it is easier than the proof part (ii) (a proof can be found in a working version). Now consider the proof of (ii). From (43) and denoting $G = (F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$ for the moment

$$T^{-1}N^{-1/2}\sum_{k=1}^{N}\varepsilon_k'(\hat{F}H^{-1} - F^0) = T^{-1}N^{-1/2}\sum_{k=1}^{N}\varepsilon_k'(I1 + \cdots + I8)G = a1 + \cdots + a8$$

We show that first four terms are each $T^{-1/2}O_p(\hat{\beta} - \beta)$.

$$\|a1\| \le T^{-1/2}\|G\|\left(\frac{1}{N}\sum_{i=1}^{N}\|(\frac{1}{\sqrt{TN}}\sum_{k=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}X_{it})\|(\|X_i\|^2/T)\right)\|\hat{\beta}-\beta\|^2 = T^{-1/2}\|\hat{\beta}-\beta)\|^2 O_p(1)$$

$$a2 = \frac{1}{NT}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'X_i(\beta-\hat{\beta})\lambda_i(\Lambda'\Lambda/N)^{-1} = \frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}X_{it}\varepsilon_{kt}O_p(\hat{\beta}-\beta)\lambda_i(\Lambda'\Lambda/N)^{-1}$$

$$= T^{-1/2}(\hat{\beta} - \beta)$$

$$\|a3\| \le T^{-1/2}\|G\|\left(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}X_{it}\|(\|\varepsilon_i\|^2/T)\|\right)\hat{\beta} - \beta\| = T^{-1/2}O_p(\|\hat{\beta} - \beta\|)$$

$$a4 = T^{-1/2}(\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}F_t')(\beta - \hat{\beta})'\left(\frac{1}{N}\sum_{i=1}^{N}(X_i'\hat{F}/T)\right)G = T^{-1/2}O_p(\hat{\beta} - \beta)$$

For $a5$, let $W_i = X_i'\hat{F}/T$, note that $\|W_i\|^2 \le \|X_i\|^2/T$,

$$a5 = \frac{1}{NT}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i(\beta - \hat{\beta})'W_iG$$

$$= \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_{kt})(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{it}(\hat{\beta} - \beta)W_i)G = N^{-1/2}O_p(\hat{\beta} - \beta)$$

For $a6$,

$$a6 = \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_k'F^0\sum_{i=1}^{N}\lambda_i\varepsilon_i'\hat{F}G = \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_k'F^0\sum_{i=1}^{N}\lambda_i\varepsilon_i'F^0HG$$

$$+\frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_k'F^0\sum_{i=1}^{N}\lambda_i\varepsilon_i'(\hat{F} - F^0H)G = a6.1 + a6.2$$

$$a6.1 = \frac{1}{\sqrt{N}T}\Big(\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}F_t^{0\prime}\varepsilon_{kt}\Big)\Big(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_iF_t^{0\prime}\varepsilon_{it}\Big)HG = O_p(T^{-1}N^{-1/2})$$

$$a6.2 = T^{-1/2}\Big(\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}F_t^{0\prime}\varepsilon_{kt}\Big)\frac{1}{TN}\sum_{i=1}^{N}\lambda_i\varepsilon_i'(\hat{F} - F^0H)G$$

$$\|a6.2\| \le T^{-1/2}O_p(1)\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|\|\varepsilon_i/\sqrt{T}\|\frac{\|\hat{F} - F^0H\|}{\sqrt{T}}\|G\|$$

$$= T^{-1/2}[O_p(\hat{\beta} - \beta) + O_p(\delta_{NT}^{-1})] = T^{-1/2}(\hat{\beta} - \beta) + O_p(\delta_{NT}^{-2})$$

Next consider $a7$.

$$a7 = \frac{1}{NT}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i\lambda_i'(\Lambda'\Lambda/N)^{-1}$$

$$= N^{-1/2}\frac{1}{T}\sum_{t=1}^{T}\Big[(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_{kt})(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{it}\lambda_i')\Big](\Lambda'\Lambda/N)^{-1} = O_p(N^{-1/2})$$

$$a8 = \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i(\varepsilon_i'\hat{F})G = \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i(\varepsilon_i'F^0)HG + \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i(\varepsilon_i'(\hat{F}-F^0H))G$$

$$= b8 + c8$$

$$b8 = \frac{1}{NT}\sum_{i=1}^{N}\Big[\Big(\frac{1}{\sqrt{TN}}\sum_{k=1}^{N}\sum_{t=1}^{T}(\varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it})\Big)\Big(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{is}F_s^0H\Big)\Big]G$$

$$+\frac{1}{\sqrt{NT}}\frac{1}{NT}\sum_{k=1}^{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_{ki,t}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{is}F_s^0HG = O_p(T^{-1}) + O_p((NT)^{-1/2})$$

Ignore G,

$$c8 = T^{-1/2}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{TN}}\sum_{k=1}^{N}\sum_{t=1}^{T}[\varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it})]\frac{\varepsilon_i'(\hat{F}-F^0H)}{T} + \frac{1}{N^{3/2}T}\sum_{k=1}^{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_{ki,t}\frac{\varepsilon_i'(\hat{F}-F^0H)}{T}$$

$$= c8.1 + c8.2$$

$$\|c8.1\| \le T^{-1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big[\frac{1}{\sqrt{TN}}\sum_{k=1}^{N}\sum_{t=1}^{T}[\varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it})\Big]^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\varepsilon_i\|^2/T)\Big)^{1/2}\|\hat{F}-F^0H\|/\sqrt{T}$$

$$= T^{-1/2}O_p(\|\hat{\beta}-\beta\|) + T^{-1/2}O_p(\delta_{NT}^{-1}) = T^{-1/2}O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-2})$$

$$\|c8.2\| \le \frac{1}{\sqrt{N}}\frac{\|\hat{F}-F^0H\|}{\sqrt{T}}\frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{N}\|\gamma_{ki}\|\varepsilon_i\|/\sqrt{T}$$

$$= [O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-1})]N^{-1/2} = N^{-1/2}O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-2})$$

Note that $EN^{-1}\sum_{k=1}^{N}\sum_{i=1}^{N}\|\gamma_{ki}\|\varepsilon_i\|/\sqrt{T} \le \max_i E(\|\varepsilon_i\|/\sqrt{T})N^{-1}\sum_{k=1}^{N}\sum_{i=1}^{N}|\gamma_{ki}| = O(1)$.

Part (iii) is derived from (ii) with the division by $\sqrt{N}$. The presence of λ_k does not alter the results. A direct proof would be similar to that of (ii). The details are omitted.

Part (iv) is the same as (iii) with λ_k replaced by $(X_k'F^0/T)(F^{0\prime}F^0/T) = O_p(1)$. The first term on the right is an elaboration of the corresponding $O_p(N^{-1})$ term appearing in (iii). This elaborated expression will be used later.

Proof of Proposition 5.3. From $Y_i = X_i\beta^0 + F^0\lambda_i + \varepsilon_i$

$$\hat{\beta} - \beta^0 = (\sum_{i=1}^{N}X_i'M_{\hat{F}}X_i)^{-1}\sum_{i=1}^{N}X_i'M_{\hat{F}}F^0\lambda_i + (\sum_{i=1}^{N}X_i'M_{\hat{F}}X_i)^{-1}\sum_{i=1}^{N}X_i'M_{\hat{F}}\varepsilon_i$$

or

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

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. Choose $A = H^{-1}$, from (43),

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

It follows that

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

where $J1$ up to $J8$ are implicitly defined vis-a-vis $I1 - I8$. For example,

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

Term $J1$ is bounded in norm by $O_p(1)\|\hat{\beta} - \beta\|^2$ and thus $J1 = o_p(1)(\hat{\beta} - \beta)$. Consider $J2$.

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

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

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

Write $\varepsilon_k'\hat{F}/T = \varepsilon_k' F^0 H/T + \varepsilon_k'(\hat{F} - F^0 H)/T = O_p(T^{-1/2}) + O_p(\hat{\beta} - \beta) + O_p(1/\min[\sqrt{N}, \sqrt{T}])$, by Lemma 12.4, it easy to see $J3 = o_p(1)(\hat{\beta} - \beta)$.

Next

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

Write $M_{\hat{F}}F^0 = M_{\hat{F}}(F^0 - \hat{F}H^{-1})$ and using $T^{-1/2}\|F^0 - \hat{F}H^{-1}\|$ being small, $J4$ is equal to $o_p(1)(\hat{\beta} - \beta)$. It is easy to show $J5 = o_p(1)(\hat{\beta} - \beta)$ and thus 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 $J6$.

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

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

$$J6 \quad = \quad -\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}}(F^0 - \hat{F}H^{-1})\Big(\frac{1}{N}\sum_{k=1}^{N}\lambda_k(\epsilon_k'\hat{F}/T)\Big)G\lambda_i$$

Now

$$\frac{1}{NT}\sum_{k=1}^{N}\lambda_k\varepsilon_k'\hat{F} = \frac{1}{NT}\sum_{k=1}^{N}\lambda_k\varepsilon_k' F^0 H + \frac{1}{NT}\sum_{k=1}^{N}\lambda_k\varepsilon_k'(\hat{F} - F^0 H)$$

$$= O_p(\frac{1}{\sqrt{NT}}) + (NT)^{-1/2}O_p(\hat{\beta} - \beta) + O_p(N^{-1}) + N^{-1/2}O_p(\delta_{NT}^{-2})$$

$$= O_p(\frac{1}{\sqrt{NT}}) + O_p(N^{-1}) + N^{-1/2}O_p(\delta_{NT}^{-2})$$

by Lemma 12.4(iii). The last equality is because $(NT)^{-1/2}$ dominates $(NT)^{-1/2}(\hat{\beta} - \beta)$. Furthermore, by Lemma 12.3 $\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}}(\hat{F} - F^0 H)\lambda_{i\ell} = O_p(\hat{\beta} - \beta) + O_p(\delta_{NT}^{-2})$ for $\ell = 1, 2, ..., r$, and noting G does not depend on i and $\|G\| = O_p(1)$, we have

$$J6 = \Big[O_p(\hat{\beta} - \beta) + O_p(\delta_{NT}^{-2})\Big]\Big[O_p(\frac{1}{\sqrt{NT}}) + O_p(N^{-1}) + N^{-1/2}O_p(\delta_{NT}^{-2})\Big]$$

$$= o_p(\hat{\beta} - \beta) + o_p(\frac{1}{\sqrt{NT}}) + O_p(\delta_{NT}^{-2})N^{-1} + N^{-1/2}O_p(\delta_{NT}^{-4}).$$

The term $J7$ is simply

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

Next consider $J8$, which has the expression

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

Let $E(\varepsilon_k\varepsilon_k') = \Omega_k$, a $T \times T$ matrix. Denote $G = (F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$, $\|G\| = O_p(1)$. Rewrite

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

When no serial correlation and heteroskedasticity, $\Omega_k = \sigma_k^2 I_T$. From $M_{\hat{F}}\hat{F} = 0$, the first term is equal to zero. The second term being small does not rely on $\Omega_k = \sigma_k^2 I_T$. Lemma 12.5 below deals with general Ω_k that will also be applicable for Proposition 5.5. By Lemma 12.5, we have

$$J8 = O_p(\frac{1}{T\sqrt{N}}) + (NT)^{-1/2}[O_p(\hat{\beta} - \beta) + O_p(\delta_{NT}^{-1})] + \frac{1}{\sqrt{N}}O_p(\|\hat{\beta} - \beta\|^2) + \frac{1}{\sqrt{N}}O_p(\delta_{NT}^{-2}).$$

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

$$\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} F^0 \lambda_i$$

$$= J2 + J7 + o_p(\hat{\beta}-\beta) + o_p((NT)^{-1/2}) + O_p(\frac{1}{T\sqrt{N}}) + N^{-1/2}O_p(\delta_{NT}^{-2})$$

Thus,

$$\Big(\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i + o_p(1)\Big)(\hat{\beta}-\beta) - J2 = \frac{1}{NT}\sum_{i=1}^{N} X_i M_{\hat{F}} \varepsilon_i + J7$$

$$+ o_p((NT)^{-1/2}) + O_p(\frac{1}{T\sqrt{N}}) + N^{-1/2}O_p(\delta_{NT}^{-2})$$

Combining terms and multiplying $\sqrt{NT}$

$$\Big[D(\hat{F}) + o_p(1)\Big]\sqrt{NT}(\hat{\beta}-\beta) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Big[X_i' M_{\hat{F}} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_{\hat{F}}\Big]\varepsilon_i + o_p(1)$$

$$+ O_p(T^{-1/2}) + T^{1/2}O_p(\delta_{NT}^{-2})$$

Thus, if $T/N^2 \to 0$, the last term is also $o_p(1)$. Noting $[D(\hat{F}) + o_p(1)]^{-1} = D(\hat{F})^{-1} + o_p(1)$, we proved the proposition.

Lemma 12.5 *Under the assumptions of Proposition 5.3 and for* $G = (F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$

$$\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i' M_{\hat{F}}(\varepsilon_k\varepsilon_k' - \Omega_k)\hat{F}G\lambda_i$$

$$= O_p(\frac{1}{T\sqrt{N}}) + (NT)^{-1/2}[O_p(\hat{\beta}-\beta) + O_p(\delta_{NT}^{-1})] + \frac{1}{\sqrt{N}}O_p(\|\hat{\beta}-\beta\|^2) + \frac{1}{\sqrt{N}}O_p(\delta_{NT}^{-2}).$$

Proof: Rewrite the left hand side as

$$\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i'(\varepsilon_k\varepsilon_k' - \Omega_k)\hat{F}G\lambda_i - \frac{1}{N}\sum_{i=1}^{N}(\frac{X_i'\hat{F}}{T})\frac{1}{NT^2}\sum_{k=1}^{N}\hat{F}'(\varepsilon_k\varepsilon_k' - \Omega_k)\hat{F}G\lambda_i = I + II$$

Adding and subtracting terms

$$I = \frac{1}{N^2T^2}\sum_{k=1}^{N}\sum_{i=1}^{N} X_i'(\varepsilon_k\varepsilon_k' - \Omega_k)F^0HG\lambda_i + \frac{1}{N^2T^2}\sum_{k=1}^{N}\sum_{i=1}^{N} X_i'(\varepsilon_k\varepsilon_k' - \Omega_k)(\hat{F} - F^0H)G\lambda_i$$

The first term on the right is equal to

$$(\frac{1}{N^2T^2})\sum_{i=1}^{N}\sum_{k=1}^{N}\Big\{\sum_{t=1}^{T}\sum_{s=1}^{T} X_{it}[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]F_s^{0\prime}HG\lambda_i\Big\}$$

$$= \frac{1}{T\sqrt{N}}\frac{1}{N}\sum_{i=1}^{N}\Big[N^{-1/2}\sum_{k=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} X_{it}[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]F_s^{0\prime}\Big]HG\lambda_i = O_p(\frac{1}{T\sqrt{N}})$$

by Lemma 12.2(ii). Denote

$$a_s = \Big(\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T} X_{it}[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]\Big) = O_p(1)$$

Then the second term of I is

$$\frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{s=1}^{T} a_s(\hat{F}_s - F_s^0 H)$$

$$\|\frac{1}{T}\sum_{s=1}^{T} a_s(\hat{F}_s - F_s^0 H)\| \le (\frac{1}{T}\sum_{s=1}^{T}\|a_s\|^2)^{1/2}(\frac{1}{T}\sum_{s=1}^{T}\|\hat{F}_s - F_s^0 H\|^2)^{1/2} = O_p(\hat{\beta}-\beta) + O_p(\delta_{NT}^{-1})$$

Thus the second term of I is $(NT)^{-1/2}[O_p(\hat{\beta}-\beta) + O_p(\delta_{NT}^{-1})]$. Consider II.

$$\|II\| \le \frac{1}{N}\sum_{i=1}^{N}\|\frac{X_i\hat{F}}{T}\|\|G\lambda_i\|\cdot\|\frac{1}{NT^2}\sum_{k=1}^{N}\hat{F}'(\varepsilon_k\varepsilon_k' - \Omega_k)\hat{F}\| = O_p(1)\|\frac{1}{NT^2}\sum_{k=1}^{N}\hat{F}'(\varepsilon_k\varepsilon_k' - \Omega_k)\hat{F}\|$$

But

$$\begin{aligned}
&\frac{1}{NT^2}\sum_{k=1}^{N}\hat{F}'(\varepsilon_k\varepsilon_k' - \Omega_k)\hat{F} \\
&= H\frac{1}{NT^2}\sum_{k=1}^{N}F^{0\prime}(\varepsilon_k\varepsilon_k' - \Omega_k)F^0 H \\
&+ H\frac{1}{NT^2}\sum_{k=1}^{N}F^{0\prime}(\varepsilon_k\varepsilon_k' - \Omega_k)(\hat{F} - F^0 H) \\
&+ \frac{1}{NT^2}\sum_{k=1}^{N}(\hat{F} - F^0 H)'(\varepsilon_k\varepsilon_k' - \Omega_k)F^0 H \\
&+ \frac{1}{NT^2}\sum_{k=1}^{N}(\hat{F} - F^0 H)'(\varepsilon_k\varepsilon_k' - \Omega_k)(\hat{F} - F^0 H) \\
&= b1 + b2 + b3 + b4
\end{aligned}$$

Now

$$b1 = H(\frac{1}{T^2 N})\sum_{k=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} F_s F_t'[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]H = O_p(\frac{1}{T\sqrt{N}})$$

by Lemma 12.2(i). Next

$$b2 = H\frac{1}{\sqrt{NT}}\frac{1}{T}\sum_{s=1}^{T}\Big[\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{k=1}^{N}F_t^0[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]\Big](\hat{F}_s - H'F_s^0)$$

Thus if we let $A_s = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{k=1}^{N}F_t^0[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]$,

$$\|b2\| \le \|H\|\frac{1}{\sqrt{NT}}\Big(\frac{1}{T}\sum_{s=1}^{T}\|A_s\|^2\Big)^{1/2}\Big(\frac{1}{T}\sum_{s=1}^{T}\|\hat{F}_s - H'F_s^0\|^2\Big)^{1/2} = \frac{1}{\sqrt{NT}}[O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-1})]$$

The term $b3$ has the same upper bound because it is the transpose of $b2$. The last term

$$b4 = \frac{1}{\sqrt{N}}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{F}_t - H'F_t^0)(\hat{F}_s - H'F_s^0)'[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]$$

Thus by the Cauchy-Schwarz inequality,

$$\|b4\| \leq \frac{1}{\sqrt{N}}(\frac{1}{T}\sum_{t=1}^{T}\|F_t - H'F_t^0\|^2)\Big(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\Big[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]\Big]^2\Big)^{1/2}$$

$$= \frac{1}{\sqrt{N}}O_p(\|\hat{\beta} - \beta\|^2) + \frac{1}{\sqrt{N}}O_p(\delta_{NT}^{-2}).$$

Now collecting terms yields the lemma.

Lemma 12.6 *Under the assumptions of Proposition 5.4,*

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

Proof: The first two results of Lemma 12.3 can be rewritten as

$$F^{0\prime}\hat{F}/T - (F^{0\prime}F^0/T)H = O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2}), \quad \text{and}$$

$$I - (\hat{F}'F^0/T)H = O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2})$$

Left multiply the first equation by H' and use the transpose of the second equation to obtain

$$I - H'(F^{0\prime}F^0/T)H = O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2})$$

Right multiplying by H' and left multiplying by H'^{-1}, we obtain

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

This is equivalent to the lemma.

Lemma 12.7 *Under the assumptions of Proposition 5.4,*

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

Proof

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

Proposition 5.1(ii) already implies $I_r - \hat{F}'P_{F^0}\hat{F}/T = o_p(1)$. By rewriting $T^{-1}\hat{F}'F^0 = T^{-1}\hat{F}(F^0 - \hat{F}H^{-1}) + H$, we can easily show using earlier lemmas $I_r - \hat{F}'P_{F^0}\hat{F}/T = O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2})$. The details are omitted.

Proof of Proposition 5.4. We first show $D(\hat{F}) - D(F^0) = o_p(1)$.

$$D(\hat{F}) - D(F^0) = \frac{1}{NT}\sum_{i=1}^{N}X_i'(M_{\hat{F}} - M_{F^0})X_i - \frac{1}{T}\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}X_i'(M_{\hat{F}} - M_{F^0})X_k a_{ik}\Big]$$

The norm of the first term on the right is bounded by

$$\|\frac{1}{NT}\sum_{i=1}^{N} X_i'(P_{\hat{F}} - P_{F^0})X_i\| \le \frac{1}{N}\sum_{i=1}^{N}(\|X_i\|^2/T)\|P_{\hat{F}} - P_{F^0}\| = o_p(1)$$

by Lemma 12.7. The proof for the second term is $o_p(1)$ is the same.

Much more involved is the proof of the following.

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Big[X_i'M_{\hat{F}} - \frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k'M_{\hat{F}}\Big]\varepsilon_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Big[X_i'M_{F^0} - \frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k'M_{F^0}\Big]\varepsilon_i + o_p(1)$$

The above is implied by the following two results, as $T/N \to 0$,

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i(M_{F^0} - M_{\hat{F}})\varepsilon_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i'(P_{\hat{F}} - P_{F^0})\varepsilon_i = o_p(1) \tag{47}$$

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k'(M_{F^0} - M_{\hat{F}})\varepsilon_i = o_p(1) \tag{48}$$

Consider (47). By adding and subtracting terms

$$\begin{aligned}
& \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{X_i'\hat{F}}{T}\hat{F}'\varepsilon_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i'P_{F^0}\varepsilon_i \\
= \; & \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{X_i'(\hat{F} - F^0H)}{T}H'F^{0\prime}\varepsilon_i + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{X_i'(\hat{F} - F^0H)}{T}(\hat{F} - F^0H)'\varepsilon_i \\
& + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{X_i'F^0H}{T}(\hat{F} - F^0H)'\varepsilon_i + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\Big[HH' - (F^{0\prime}F^0/T)^{-1}\Big]F^{0\prime}\varepsilon_i \\
= \; & a + b + c + d
\end{aligned}$$

Consider (a). Note $(\hat{F}_s - H'F_s^0)'H'F_t^0$ is scalar thus commutable with X_{it}.

$$a = \frac{1}{T}\sum_{s=1}^{T}(\hat{F}_s - H'F_s^0)'H'\Big(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t^0 X_{is}\varepsilon_{it}\Big)$$

Thus

$$\begin{aligned}
\|a\| &\le \Big[\frac{1}{T}\sum_{s=1}^{T}\|F_s - H'F_s^0\|^2\Big]^{1/2}\|H\|\Big[\frac{1}{T}\sum_{s=1}^{T}\|\Big(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t^0 X_{is}\varepsilon_{it}\Big)\|^2\Big]^{1/2} \\
&= [O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-1})]O_p(1) = o_p(1)
\end{aligned}$$

Similarly,

$$b = T^{1/2}\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}(\hat{F}_s - H'F_s^0)'(\hat{F}_t - H'F_t^0)\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_{is}\varepsilon_{it}\Big)$$

By the Schwarz inequality

$$\|b\| \le \sqrt{T}(\frac{1}{T}\sum_{t=1}^{T}\|F_t - H'F_t\|^2)\Big(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_{it}\varepsilon_{it}\|^2\Big)^{1/2}$$

$$= \sqrt{T}[O_p(\|\hat\beta - \beta\|^2) + O_p(\delta_{NT}^{-2})]O_p(1)$$

which is $o_p(1)$ if $T/N^2 \to 0$.

Consider c.

$$c = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \frac{X_i' F^0}{T} HH'(\hat F H^{-1} - F^0)'\varepsilon_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \Big(\frac{F^{0\prime} F^0}{T}\Big)^{-1} (\hat F H^{-1} - F^0)'\varepsilon_i$$

$$+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \frac{X_i' F^0}{T} \Big[HH' - (\frac{F^{0\prime} F^0}{T})^{-1}\Big] (\hat F H^{-1} - F^0)'\varepsilon_i = c1 + c2$$

Denote $Q = HH' - (\frac{F^{0\prime}F^0}{T})^{-1}$ for the moment. We show $c2 = o_p(1)$.

$$c2 = \sqrt{NT}\Big(\frac{1}{NT}\sum_{i=1}^{N} [\varepsilon_i'(\hat F H^{-1} - F^0) \otimes (X_i' F^0/T)]\Big) vec(Q)$$

$$= \sqrt{NT}\Big[(NT)^{-1/2}(\|\hat\beta - \beta)\| + O_p(N^{-1}) + N^{-1/2} O_p(\delta_{NT}^{-2})\Big] vec(Q)$$

by the argument of Lemma 12.4(iii) and (iv). By Lemma 12.6, $vec(B) = O_p(\|\hat\beta - \beta\|) + O_p(\delta_{NT}^{-2})$. Thus $c2 = O_p(\hat\beta - \beta) + \sqrt{T/N} O_p(\delta_{NT}^{-2}) + \sqrt{T} O_p(\delta_{NT}^{-4}) \xrightarrow{p} 0$ if $T/N^3 \to 0$.

By Lemma 12.4 (iv), switching the role of i and k,

$$c1 = (\sqrt{NT}/N)\psi_{NT} + O_p(\hat\beta - \beta) + \sqrt{T} O_p(\delta_{NT}^{-2})$$

where

$$\psi_{NT} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N} \frac{X_i' F^0}{T}\Big(\frac{F^{0\prime}F^0}{T}\Big)^{-1}(\frac{\Lambda'\Lambda}{N})^{-1}\lambda_k(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}) \tag{49}$$

But $\psi_{NT} = O_p(1)$ because

$$\psi_{NT} = \frac{1}{T}\sum_{t=1}^{T}\Big(N^{-1/2}\sum_{i=1}^{N} A_i \varepsilon_{it}\Big)\Big(N^{-1/2}\sum_{k=1}^{N} B_k \varepsilon_{kt}\Big) = O_p(1)$$

with $A_i = (X_i' F_0/T)(F^{0\prime}F^0/T)^{-1}$ and $B_k = (\Lambda'\Lambda/N)^{-1}\lambda_k$. Thus $c1 \to 0$ if $T/N \to 0$.

For (d), again let $Q = HH' - (F^{0\prime}F^0/T)^{-1}$. Then

$$d = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}[\varepsilon_i' F^0 \otimes (X_i' F^0/T)] vec(Q) = \Big(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t^0 \varepsilon_{it} \otimes (X_i' F^0/T)\Big) vec(Q) = O_p(1) vec(Q)$$

which is $o_p(1)$ because $vec(Q) = O_p(\|\hat\beta - \beta\|) + O_p(\delta_{NT}^{-2})$ by Lemma 12.6. In summary, ignore dominated terms

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i'(M_{F^0} - M_{\hat F})\varepsilon_i = (\sqrt{NT}/N)\psi_{NT} + O_p(\hat\beta - \beta) + \sqrt{T} O_p(\delta_{NT}^{-2}) \tag{50}$$

with $\psi_{NT} = O_p(1)$ being defined in (49). The above is $o_p(1)$ if $T/N \to 0$, proving (47).

It remains to prove (48). This is obtained by replacing X_i in the earlier proof with $V_i = \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k$. Then (48) becomes

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} V_i'(M_{F^0} - M_{\hat F})\varepsilon_i = (\sqrt{NT}/N)\psi_{NT}^* + O_p(\hat\beta - \beta) + \sqrt{T} O_p(\delta_{NT}^{-2}) \tag{51}$$

where $\psi^*_{NT} = O_p(1)$ is defined as

$$\psi^*_{NT} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{V_i'F^0}{T}\Big(\frac{F^{0\prime}F^0}{T}\Big)^{-1}(\frac{\Lambda'\Lambda}{N})^{-1}\lambda_k(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}) \tag{52}$$

Thus (51) is $o_p(1)$ if $T/N \to 0$.

Proof of Proposition 5.5. In the presence of serial correlation or heteroskedasticity, the first term of (46) is no longer zero. Except for this special term, the proof of preceding two propositions is still valid under serial correlation or heteroskedasticity. Ignore that term for a moment, the proof of Proposition 5.4 shows, combining (50) and (51),

$$\sqrt{NT}(\hat\beta-\beta) = D(F^0)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}Z_i'\varepsilon_i + \sqrt{T/N}\xi_{NT} + O_p(\hat\beta-\beta) + \sqrt{T}O_p(\delta_{NT}^{-2})$$

where

$$\begin{aligned}\xi_{NT} &= -D(F^0)^{-1}(\psi_{NT}-\psi^*_{NT})\\ &= -D(F^0)^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{(X_i-V_i)'F^0}{T}\Big(\frac{F^{0\prime}F^0}{T}\Big)^{-1}(\frac{\Lambda'\Lambda}{N})^{-1}\lambda_k(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt})\end{aligned}$$

As argued earlier, $\xi_{NT} = O_p(1)$. Thus if $T/N = O(1)$, then $\sqrt{NT}(\hat\beta-\beta) = O_p(1)$.

In the presence of serial correlation or time series heteroskedasticity, $M_{\hat F}\Omega_k\hat F \neq 0$, so the first term on the right of (46) is nonzero. Denote that term by A_{NT}. This means that $\sqrt{NT}(\hat\beta-\beta)$ has an extra term $D(F^0)^{-1}\sqrt{NT}A_{NT}$. That is,

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

where

$$A_{NT} = -\frac{1}{NT^2}\sum_{i=1}^{N}X_i'M_{\hat F}(\frac{1}{N}\sum_{k=1}^{N}\Omega_k)\hat F G\lambda_i = -\frac{1}{NT^2}\sum_{i=1}^{N}X_i'\Omega\hat F G\lambda_i + \frac{1}{NT^2}\sum_{i=1}^{N}\frac{X_i'\hat F}{T}(\hat F'\Omega\hat F)G\lambda_i \tag{53}$$

and $\Omega = \frac{1}{N}\sum_{k=1}^{N}\Omega_k$. We now show that $A_{NT} = O_p(1/T)$ so that $\sqrt{NT}A_{NT} = O_p((N/T)^{1/2})$, which is $O_p(1)$ when N/T is bounded. Note that $\|\hat F'\Omega\hat F\| \le \lambda_{\max}(\Omega)\|\hat F'\hat F\| = \lambda_{\max}(\Omega)rT = O(T)$, where $\lambda_{\max}(\Omega)$ is the largest eigenvalue of Ω and is bounded by assumption. Thus, Similarly, $\|X_i'\Omega\hat F\| \le \frac{1}{2}\lambda_{\max}(\Omega)[\|X_i\|^2 + \|\hat F\|^2]$.

$$\|\frac{1}{NT^2}\sum_{i=1}^{N}X_i'\Omega\hat F G\lambda_i\| \le \frac{1}{2T}\lambda_{\max}(\Omega)\Big(\frac{1}{N}\sum_{i=1}^{N}(\|X_i\|^2/T + r)\|G\|\|\lambda_i\|\Big) = O_p(1/T)$$

$$\|\frac{1}{NT^2}\sum_{i=1}^{N}\frac{X_i'\hat F}{T}(\hat F'\Omega\hat F)G\lambda_i\| \le \frac{1}{T}r^{3/2}\lambda_{\max}(\Omega)\Big(\|\frac{1}{N}\sum_{i=1}^{N}\frac{\|X_i\|}{\sqrt{T}}\|G\|\|\lambda_i\|\Big) = O_p(1/T)$$

This completes the proof of Proposition 5.5.

Proof of Theorem 5.6. This follows immediately from Proposition 5.4 and Assumption F.

Proof of Theorem 5.8. The proof of Proposition 5.4 shows that

$$\sqrt{NT}(\hat{\beta}-\beta^0) = D(F^0)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} Z_i'\varepsilon_i + \sqrt{T/N}\xi_{NT} + o_p(1) \tag{54}$$

where ξ_{NT} is defined in (15). The expected value of ξ_{NT} is given

$$B = -D(F^0)^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{(X_i-V_i)'F^0}{T}\Big(\frac{F^{0\prime}F^0}{T}\Big)^{-1}\Big(\frac{\Lambda'\Lambda}{N}\Big)^{-1}\lambda_k\sigma_{ik}^2$$

We show $\sqrt{T/N}(\xi_{NT}-B) = o_p(1)$. Let $A_{ik} = \lambda_k'\otimes[(X_i-V_i)'F^0/T]$ and $G = (F^{0\prime}F^0/T)^{-1}(\Lambda'\Lambda/N)^{-1}$, then $E\|A_{ik}\|^2 \le M$ and $\|G\| = O_p(1)$. We have

$$D(F^0)[\xi_{NT}-B] = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\Big[A_{ik}\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{it}\varepsilon_{kt}-\sigma_{ik}^2)\Big]vec(G)$$

Assumption C4 implies the above is $O_p(T^{-1/2})$. Thus $\sqrt{T/N}(\xi_{NT}-B) = O_p(N^{-1/2})$ does not affect the limiting distribution. Since $\sqrt{T/N}\to\sqrt{\rho}$ and $\text{plim}B = B_0$ by assumption, the theorem follows.

Biased correction

Lemma 12.8

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

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

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

Proof: By definition, $\hat{\Lambda}' = \frac{1}{T}\hat{F}'(Y-X\hat{\beta})$, where $Y=(Y_1,...,Y_N)$ is $T\times N$ and X is $T\times N\times p$ (three dimensional matrix), so that $X\hat{\beta}$ is $T\times N$ (readers may consider β is a scalar so that X is simply $T\times N$). Thus from $Y - X\hat{\beta} = F^0\Lambda' + \varepsilon - X(\hat{\beta}-\beta^0)$,

$$\hat{\Lambda}' = T^{-1}\hat{F}'F^0\Lambda' + T^{-1}\hat{F}'\varepsilon - T^{-1}\hat{F}'X(\hat{\beta}-\beta^0)$$

From $F^0 = F^0 - \hat{F}H^{-1} + \hat{F}H^{-1}$ and use $\hat{F}'\hat{F}/T = I$, we have

$$\hat{\Lambda}' - H^{-1}\Lambda' = T^{-1}\hat{F}'(F^0-\hat{F}H^{-1})\Lambda' + T^{-1}\hat{F}'\varepsilon - T^{-1}\hat{F}'X(\hat{\beta}-\beta) \tag{55}$$

Thus

$$N^{-1/2}\|\hat{\Lambda}'-H^{-1}\Lambda'\| \le \sqrt{r}\frac{\|\hat{F}^0-\hat{F}H^{-1}\|}{\sqrt{T}}\frac{\|\Lambda\|}{\sqrt{N}} + T^{-1/2}\|\frac{1}{\sqrt{NT}}\hat{F}'\varepsilon\| + \sqrt{r}\frac{\|X\|}{\sqrt{NT}}\|\hat{\beta}-\beta\|$$

The first term is $O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-1})$ by Proposition 5.2(ii), and the second term is $O_p(T^{-1/2})$, third term is $O_p(\|\hat{\beta}-\beta\|)$ in view $\|X\|/\sqrt{NT} = O_p(1)$. Thus $N^{-1/2}\|\hat{\Lambda}'-H^{-1}\Lambda'\| = O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-1})$. This is equivalent to (i).

For (ii), left multiplying Λ on each side and then dividing by N,

$$N^{-1}(\hat{\Lambda}' - H^{-1}\Lambda')\Lambda = T^{-1}\hat{F}'(F^0 - \hat{F}H^{-1})(\Lambda'\Lambda/N) + (TN)^{-1}\hat{F}'\varepsilon\Lambda - (TN)^{-1}\hat{F}'X(\hat{\beta} - \beta)\Lambda$$

The first term on the right is $O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2})$ by Lemma 12.3. The second term is

$$(TN)^{-1}(\hat{F} - F^0H)'\varepsilon\Lambda + (TN)^{-1}HF^{0\prime}\varepsilon\Lambda = a + b$$

But (a) is the left hand side of Lemma 12.4(iii), thus having the desired result. Term (b) is simply $(TN)^{-1/2}O_p(1)$, also as desired. Finally,

$$\|(TN)^{-1}\hat{F}'X(\hat{\beta} - \beta)\Lambda\| \leq \sqrt{r}\|X/\sqrt{TN}\| \cdot \|\Lambda/\sqrt{N}\| \cdot \|\hat{\beta} - \beta\| = O_p(\|\hat{\beta} - \beta\|)$$

proving (ii). For (iii), adding and subtracting terms, (iii) follows from (i) and (ii). For (iv) follows from $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ and (iii).

Lemma 12.9 *Under assumptions of Theorem 7.1, for B and $\hat{B}$ defined in (16) and (17),*

$$\sqrt{T/N}(\hat{B} - B) = o_p(1)$$

Proof: The denominator of B is $D(F^0)$. Equation (60) shows that $\sqrt{T/N}[\hat{D}_0 - D(F^0)] = o_p(1)$. Thus it is sufficient to consider the numerator only. We shall prove

$$(\sqrt{T/N})\left[\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'\hat{F}}{T}(\frac{\hat{\Lambda}'\hat{\Lambda}}{N})^{-1}\hat{\lambda}_i\hat{\sigma}_i^2 - \frac{1}{N}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\Big(\frac{F^{0\prime}F^0}{T}\Big)^{-1}(\frac{\Lambda'\Lambda}{N})^{-1}\lambda_i\sigma_i^2\right] = o_p(1) \tag{56}$$

and

$$(\sqrt{T/N})\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{V}_i'\hat{F}}{T}(\frac{\hat{\Lambda}'\hat{\Lambda}}{N})^{-1}\hat{\lambda}_i\hat{\sigma}_i^2 - \frac{1}{N}\sum_{i=1}^{N}\frac{V_i'F^0}{T}\Big(\frac{F^{0\prime}F^0}{T}\Big)^{-1}(\frac{\Lambda'\Lambda}{N})^{-1}\lambda_i\sigma_i^2\right] = o_p(1) \tag{57}$$

Consider (56). There are four items being estimated, namely, F, $\Lambda'\Lambda/N$, λ_i, and σ_i^2. Using the identity $\hat{a}\hat{b}\hat{c}\hat{d} - abcd = (\hat{a} - a)\hat{b}\hat{c}\hat{d} + a(\hat{b} - b)\hat{c}\hat{d} + ab(\hat{c} - c)\hat{d} + abc(\hat{d} - d)$. The first corresponding term is

$$\|\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'(\hat{F} - F^0H)}{T}(\frac{\hat{\Lambda}'\hat{\Lambda}}{N})^{-1}\hat{\lambda}_i\hat{\sigma}_i^2\| \leq \frac{\|\hat{F} - F^0H\|}{\sqrt{T}}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\|X_i\|}{\sqrt{T}}(\frac{\hat{\Lambda}'\hat{\Lambda}}{N})^{-1}\hat{\lambda}_i\hat{\sigma}_i^2\|\right) = O_p(\delta_{NT}^{-1})$$

The second corresponding term is $O_p(\delta_{NT}^{-1})$, which follows from Lemma 12.8(iv). The term HH' arises in the interim, which just matches $(F^{0\prime}F^0/T)^{-1}$ by Lemma 12.6 and $HH' - (F^{0\prime}F^0/T)^{-1} = O_p(\delta_{NT}^{-1})$.

For the third corresponding term, from (55)

$$\begin{aligned} \hat{\lambda}_i - H^{-1}\lambda_i &= T^{-1}\hat{F}'(F^0 - \hat{F}H^{-1})\lambda_i + T^{-1}\hat{F}'\varepsilon_i - T^{-1}\hat{F}'X_i'(\hat{\beta} - \beta) \\ &= T^{-1}\hat{F}'(F^0 - \hat{F}H^{-1})\lambda_i + T^{-1}(\hat{F} - \hat{F}^0H)'\varepsilon_i + T^{-1}HF^{0\prime}\varepsilon_i - T^{-1}\hat{F}'X_i(\hat{\beta} - \beta) \end{aligned} \tag{58}$$

This means that the corresponding third term is also split into four expressions. Each expression can be easily shown to be dominated by $O_p(\delta_{NT}^{-1})$.

Next

$$\hat{\varepsilon}_{it} = \varepsilon_{it} + X_{it}'(\hat{\beta} - \beta) + (\hat{F}_t - H'F_t^0)H^{-1}\lambda_i + \hat{F}_t'(\hat{\lambda}_i - H^{-1}\lambda_i) \tag{59}$$

It is easy to show that $\frac{1}{T}\sum_{t=1}^T \hat{\varepsilon}_{it}^2 - \frac{1}{T}\sum_{t=1}^T \varepsilon_{it}^2 = O_p(\delta_{NT}^{-1})$. Furthermore, $\frac{1}{T}\sum_{t=1}^T \varepsilon_{it}^2 - \sigma_i^2 = \frac{1}{T}\sum_{t=1}^T [\varepsilon_{it}^2 - E(\varepsilon_{it}^2)] = O_p(T^{-1/2})$. In summary, (56) is equal to $\sqrt{T/N}O_p(\delta_{NT}^{-1}) = o_p(1)$ if $T/N^2 \to 0$.

Consider (57). The only difference between (57) and (56) is X_i replaced by $\hat{V}_i$. Thus it is sufficient to prove

$$(\sqrt{T/N})\frac{1}{N}\sum_{i=1}^N \frac{(\hat{V}_i - V_i)'F^0}{T}A_i = o_p(1)$$

where $A_i = (F^{0\prime}F^0/T)^{-1}(\Lambda'\Lambda/N)^{-1}\lambda_i\sigma_i^2 = O_p(1)$.

$$\|\frac{1}{N}\sum_{i=1}^N \frac{(\hat{V}_i - V_i)'F^0}{T}A_i\| \leq \Big(\frac{1}{N}\sum_{i=1}^N T^{-1/2}\|\hat{V}_i - V_i\|\|A_i\|\Big)\|T^{-1/2}F^0\|$$

Now $\hat{V}_i - V_i = \frac{1}{N}\sum_{k=1}^N (\hat{a}_{ik} - a_{ik})X_k$, where

$$\hat{a}_{ik} - a_{ik} = (\hat{\lambda}_i - H^{-1}\lambda_i)'(\hat{\Lambda}'\hat{\Lambda}/N)^{-1}\hat{\lambda}_k + \lambda_i'H'^{-1}\Big[(\hat{\Lambda}'\hat{\Lambda}/N)^{-1} - H'(\Lambda'\Lambda/N)^{-1}H\Big]\hat{\lambda}_k$$
$$+\lambda_i'(\Lambda'\Lambda/N)^{-1}H(\hat{\lambda}_k - H^{-1}\lambda_k)$$

Thus

$$\frac{1}{N}\sum_{i=1}^N T^{-1/2}\|\hat{V}_i - V_i\|\|A_i\| \leq \Big(\frac{1}{N}\sum_{i=1}^N \|\hat{\lambda}_i - H^{-1}\lambda_i\|\|A_i\|\Big)\|(\hat{\Lambda}'\hat{\Lambda}/N)^{-1}\|\Big(\frac{1}{N}\sum_{k=1}^N \|\hat{\lambda}_k\|\|X_k/\sqrt{T}\|\Big)$$
$$+\|\Big[(\hat{\Lambda}'\hat{\Lambda}/N)^{-1} - H'(\Lambda'\Lambda/N)^{-1}H\Big]\|\Big(\frac{1}{N}\sum_{i=1}^N \|H^{-1}\lambda_i\|\|A_i\|\Big)\Big(\frac{1}{N}\sum_{k=1}^N \|\hat{\lambda}_k\|\|X_k/\sqrt{T}\|\Big)$$
$$+\frac{1}{N}\sum_{i=1}^N \|\lambda_i(\Lambda'\Lambda/N)^{-1}H\|\|A_i\|\frac{1}{N}\sum_{k=1}^N \|(\hat{\lambda}_k - H^{-1}\lambda_k)\|\|X_k/\sqrt{T}\|$$

Each term on the right is bounded $O_p(\delta_{NT}^{-1})$ by Lemmas 12.8 and 12.11. Thus (57) is equal to $\sqrt{T/N}O_p(\delta_{NT}^{-1})$, which is $o_p(1)$ if $T/N^2 \to 0$.

Proof of Theorem 7.1. This follows from (54), Lemma 12.9, and $\sqrt{T/N}(\xi_{NT} - B) = o_p(1)$, as shown in the proof of Theorem 5.8.

Lemma 12.10 *Under assumptions of Theorem 7.2, for C and $\hat{C}$ defined in (18) and (19),*

$$\sqrt{N/T}(\hat{C} - C) = o_p(1)$$

Proof: We only analyze terms involving the difference $\hat{\Omega} - \Omega$ because expressions involve other estimates are analyzed in the proof of Lemma 12.9. Consider

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

$$+\frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'\hat{F}}{T}\hat{F}'(\hat{\Omega}-\Omega)\hat{F}(\hat{\Lambda}'\hat{\Lambda}/N)^{-1}\hat{\lambda}_i = a+b$$

$$a = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}X_{it}\hat{F}_t'\Big(\frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_{kt}^2-\sigma_{k,t}^2\Big)(\hat{\Lambda}'\hat{\Lambda}/N)^{-1}\hat{\lambda}_i$$

$$\|a\| \le \Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\|\frac{1}{N}\sum_{i=1}^{N}X_{it}\hat{F}_t'((\hat{\Lambda}'\hat{\Lambda}/N)^{-1}\hat{\lambda}_i\|\Big)^2\Big]^{1/2}\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_{kt}^2-\sigma_{k,t}^2\Big)^2\Big]^{1/2}$$

But $\frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_{kt}^2-\sigma_{k,t}^2 = \frac{1}{N}\sum_{k=1}^{N}[\hat{\varepsilon}_{kt}^2-\varepsilon_{kt}^2]+\frac{1}{N}\sum_{k=1}^{N}[\varepsilon_{kt}^2-\sigma_{k,t}^2] = \frac{1}{N}\sum_{k=1}^{N}[\hat{\varepsilon}_{kt}^2-\varepsilon_{kt}^2]+O_p(N^{-1/2})$. Moreover, $\frac{1}{N}\sum_{k=1}^{N}[\hat{\varepsilon}_{kt}^2-\varepsilon_{kt}^2] = O_p(\delta_{NT}^{-1})$ and so is the average over t. Thus $a = O_p(\delta_{NT}^{-1})$. Next

$$\|b\| \le T^{-1}\|\hat{F}'(\hat{\Omega}-\Omega)\hat{F}\|\frac{1}{N}\sum_{i=1}^{N}\|X_i'\hat{F}/T\|\|(\hat{\Lambda}'\hat{\Lambda}/N)^{-1}\hat{\lambda}_i\| = T^{-1}\|\hat{F}'(\hat{\Omega}-\Omega)\hat{F}\|O_p(1).$$

But $\frac{1}{T}\|\hat{F}'(\hat{\Omega}-\Omega)\hat{F}\| = \frac{1}{T}\|\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'(\frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_{kt}^2-\sigma_{k,t}^2)\| \le \sqrt{r}\{\frac{1}{T}\sum_{t=1}^{T}[\frac{1}{N}\sum_{k=1}^{N}(\hat{\varepsilon}_{kt}^2-\sigma_{k,t}^2)]^2\}^{1/2} = O_p(\delta_{NT}^{-1})$, i.e., $b = O_p(\delta_{NT}^{-1})$. Thus $\sqrt{N/T}(\hat{C}-C) = \sqrt{N/T}O_p(\delta_{NT}^{-1}) \to 0$ if $N/T^2 \to 0$.

Proof of Theorem 7.2. With time series heteroskedasticity,

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

Thus the theorem follows from Theorem 7.1 and Lemma 12.10.

Lemma 12.11

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

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

Proof. Part (i) follows from (58) and Lemmas 12.3. For part (ii), multiply (58) by $\|T^{-1/2}X_i\|$ on each side and then take the sum. The bound is the same as in (i).

Proof of Proposition 7.3.
Proof of (i). Because $D(F^0) \xrightarrow{p} D_0$, it is sufficient to prove $\hat{D}_0 - D(F^0) \xrightarrow{p} 0$, where

$$D(F^0) = \frac{1}{NT}\sum_{i=1}^{N}X_i'M_{F^0}X_i' - \frac{1}{TN^2}\sum_{i=1}^{N}\sum_{k=1}^{N}X_i'M_{F^0}X_k a_{ik}$$

and $\hat{D}_0$ is the same as $D(F^0)$) with F^0 and a_{ik} replaced by $\hat{F}$ and $\hat{a}_{ik}$. The proof of Proposition 5.3 shows that $\|\frac{1}{NT}\sum_{i=1}^{N}X_i'(M_{\hat{F}}-M_{F^0})X_i\| = O_p(1)\|P_{\hat{F}}-P_{F^0}\| \le O_p(\|\hat{\beta}-\beta\|^{1/2})+O_p(\delta_{NT}^{-1}) = O_p(\delta_{NT}^{-1})$ by Lemma 12.7. It remains to show

$$\delta = \frac{1}{TN^2}\sum_{i=1}^{N}\sum_{k=1}^{N}X_i'M_{\hat{F}}X_k[\hat{a}_{ik}-a_{ik}] = o_p(1).$$

Noticing $a_{ik} = \lambda_k'(\Lambda'\Lambda/N)^{-1}\lambda_i$, adding and subtracting terms,

$$\hat{a}_{ik}-a_{ik} = (\hat{\lambda}_k - H^{-1}\lambda_k)'(\hat{\Lambda}'\hat{\Lambda}/N)^{-1}\hat{\lambda}_i + \lambda_k'H'^{-1}\Big[(\hat{\Lambda}'\hat{\Lambda}/N)^{-1} - H'(\Lambda'\Lambda/N)^{-1}H\Big]\hat{\lambda}_i$$

$$+\lambda_k'(\Lambda'\Lambda/N)^{-1}H(\hat{\lambda}_i - H^{-1}\lambda_i)$$
$$= b_{ik} + c_{ik} + d_{ik}$$

Decompose δ into $\delta_1+\delta_2+\delta_3$, where δ_1, δ_2 and δ_3 are defined the same way as δ but with $\hat{a}_{ik}-a_{ik}$ replaced by b_{ik}, c_{ik} and d_{ik}, respectively. From $T^{-1}\|X_i'M_{\hat{F}}X_k\| \le \|T^{-1/2}X_i\|\|T^{-1/2}X_k\|$,

$$\|\delta_1\| \le \left(\frac{1}{N}\sum_{i=1}^N \|T^{-1/2}X_i\|\|(\hat{\Lambda}'\hat{\Lambda}/N)^{-1}\hat{\lambda}_i\|\right)\left(\frac{1}{N}\sum_{k=1}^N \|T^{-1/2}X_k\|\|\hat{\lambda}_k - H^{-1}\lambda_k\|\right)$$

by Lemma 12.11, $\|\delta_1\| = O_p(\delta_{NT}^{-1}) + O_p(\|\hat{\beta}-\beta\|) = o_p(1)$. Next,

$$\|\delta_2\| \le \left(\frac{1}{N}\sum_{i=1}^N \|T^{-1/2}X_i\|\|\hat{\lambda}_i\|\right)\left(\frac{1}{N}\sum_{k=1}^N \|T^{-1/2}X_k\|\|\lambda_k\|\|H^{-1}\|\right)\|(\hat{\Lambda}'\hat{\Lambda}/N)^{-1} - H'(\Lambda'\Lambda/N)^{-1}H\|$$

which is $O_p(\delta_{NT}^{-2}) + O_p(\|\hat{\beta}-\beta\|)$ by Lemma 12.8(iii). Finally, $\delta_3 = o_p(1)$ using the same argument for δ_1. In summary $\hat{D}_0 - D(F^0) = o_p(1)$. In fact, we obtain stronger results $\hat{D}_0 - D(F^0) = O_p(\delta_{NT}^{-1})$. Thus

$$\sqrt{T/N}[\hat{D}_0 - D(F^0)] = \sqrt{T/N}O_p(\delta_{NT}^{-1}) = o_p(1) \tag{60}$$

provided that $T/N^2 \to 0$. Similarly

$$\sqrt{N/T}[\hat{D}_0 - D(F^0)] = \sqrt{N/T}O_p(\delta_{NT}^{-1}) = o_p(1)$$

provided that $N/T^2 \to 0$. These two results are used in the biased corrected estimators.

Proof of (ii). Let $D_Z^* = \frac{1}{NT}\sum_{i=1}^N \sigma_i^2 \sum_{t=1}^T Z_{it}Z_{it}'$. From $D_Z^* \xrightarrow{p} D_Z$, we only need to show $\hat{D}_Z - D_Z^* \xrightarrow{p} 0$.

$$\hat{D}_Z - D_Z^* = \frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2 - \sigma_i^2)\frac{1}{T}\sum_{t=1}^T \hat{Z}_{it}\hat{Z}_{it}' + \frac{1}{N}\sum_{i=1}^N \sigma_i^2 \frac{1}{T}\sum_{t=1}^T(\hat{Z}_{it}\hat{Z}_{it}' - Z_{it}Z_{it}') = a + b$$

$$\|a\| \le \frac{1}{N}\sum_{i=1}^N |\hat{\sigma}_i^2 - \sigma_i^2)|\frac{1}{T}\sum_{t=1}^T \|\hat{Z}_{it}\|^2$$

Using (59), we can show that

$$\frac{1}{T}\sum_{t=1}^T \hat{\varepsilon}_{it}^2 - \frac{1}{T}\sum_{t=1}^T \varepsilon_{it}^2 = O_p(\delta_{NT}^{-1})v_i$$

where $O_p(\delta_{NT}^{-1})$ does not depend on i, and v_i is such that $\frac{1}{N}\sum_{i=1}^N |v_i|^2 = O_p(1)$. Now $\hat{\sigma}_i^2 - \sigma_i^2 = \frac{1}{T}\sum_{t=1}^T \hat{\varepsilon}_{it}^2 - \frac{1}{T}\sum_{t=1}^T \varepsilon_{it}^2 + \frac{1}{T}\sum_{t=1}^T[\varepsilon_{it}^2 - \sigma_i^2] = O_p(\delta_{NT}^{-1})v_i + T^{-1/2}w_i$, where $w_i = T^{-1/2}\sum_{t=1}^T[\varepsilon_{it}^2 - \sigma_i^2] = O_p(1)$. Thus

$$\|a\| \le O_p(\delta_{NT}^{-1})\frac{1}{N}\sum_{i=1}^N |v_i|(\frac{1}{T}\sum_{t=1}^T \|\hat{Z}_{it}\|^2) + T^{-1/2}\frac{1}{N}\sum_{i=1}^N |w_i|(\frac{1}{T}\sum_{t=1}^T \|\hat{Z}_{it}\|^2) = O_p(\delta_{NT}^{-1}).$$

The proof of b is $o_p(1)$ is the same as that of part (i); the factor σ_i^2 does not affect the proof.

Proof of (iii). Let $D_2^* = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T Z_{it}Z_{it}'\sigma_{i,t}^2$, from $D_2^* \xrightarrow{p} D_2$, it is sufficient to show $\hat{D}_2 - D_2^* = o_p(1)$.

$$\hat{D}_2 - D_2^* = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{Z}_{it}\hat{Z}_{it}'(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2) + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{Z}_{it}\hat{Z}_{it}' - Z_{it}Z_{it})\varepsilon_{it}^2 + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T Z_{it}Z_{it}'(\varepsilon_{it}^2 - \sigma_{i,t}^2)$$

The first term is bounded by

$$\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\|\hat{Z}_{it}\|^4\Big)^{1/2}\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2)^2\Big)^{1/2}$$

it is easy to show $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2)^2 = o_p(1)$. The second term on the right is essentially analyzed in part (i); the extra factor ε_{it}^2 does not affect the analysis. The third term being $o_p(1)$ is due to the law of large number numbers, as in White's heteroskedasticity estimator.

Proof of (25) and (26). First note that

$$\dot{\varepsilon}_i = \varepsilon_i - \iota_T\bar{\varepsilon}_{i.} - \bar{\varepsilon} + \iota_T\bar{\varepsilon}_{..}$$

where $\bar{\varepsilon} = (\bar{\varepsilon}_{.1}, \bar{\varepsilon}_{.2}, ..., \bar{\varepsilon}_{.T})'$ not depending on i. From the constraint $\sum_{t=1}^T \hat{F}_t = \hat{F}'\iota_T = 0$, we have $M_{\hat{F}}\iota_T = \iota_T$. Also, $\dot{X}_i'\iota_T = 0$ for all i. It follows $\dot{Z}_i'\iota_T = 0$ in view of

$$\dot{Z}_i = \dot{X}_i'M_{\hat{F}} - \frac{1}{N}\sum_{k=1}^N a_{ik}\dot{X}_k'M_{\hat{F}}.$$

From $\sum_{i=1}^N \dot{X}_i = 0$ and $\sum_{i=1}^N a_{ik} = 0$, we have $\sum_{i=1}^N \dot{Z}_i = 0$. It follows that $\sum_{i=1}^N \dot{Z}_i'\bar{\varepsilon} = 0$. Thus $\sum_{i=1}^N \dot{Z}_i'\dot{\varepsilon}_i = \sum_{i=1}^N \dot{Z}_i'\varepsilon_i$. We note that $\sum_{i=1}^N a_{ik} = 0$ because $\sum_{i=1}^N \lambda_i = 0$. This proves (25). For (26), noting that $F^{0\prime}\iota_T = 0$ due to the restriction (23), the proof is identical to (25).

Proof of (31). Under iid assumptions for ε_{it}, using $E\varepsilon_i\varepsilon_j' = 0$ for $i \neq j$ and $E\varepsilon_i\varepsilon_i' = \sigma^2 I_T$,

$$E(\eta\xi') = \sigma^2\frac{1}{N^2T}\sum_{i=1}^N\sum_{k=1}^N a_{ik}X_k'M_FX_i$$

$$\begin{aligned}
E(\xi\xi') &= \frac{1}{N^3T}\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{\ell=1}^N a_{ik}X_k'M_FE(\varepsilon_i\varepsilon_j)M_FX_\ell a_{j\ell} \\
&= \sigma^2\frac{1}{N^2T}\sum_{k=1}^N\sum_{\ell=1}^N\Big(\frac{1}{N}\sum_{i=1}^N a_{ik}a_{i\ell}\Big)X_k'M_FX_\ell \\
&= \sigma^2\frac{1}{N^2T}\sum_{k=1}^N\sum_{\ell=1}^N a_{k\ell}X_k'M_FX_\ell = E\eta\xi'
\end{aligned}$$

since $\frac{1}{N}\sum_{i=1}^N a_{ik}a_{i\ell} = a_{k\ell}$.

Proof of (32).

$$E(\eta\phi') = \sigma^2\frac{1}{NT}\sum_{i=1}^N X_i'M_F[X_i - \bar{X}] = \sigma^2\frac{1}{NT}\sum_{i=1}^N X_i'M_FX_i - \sigma^2\frac{1}{T}\bar{X}'M_F\bar{X}$$

$$E(\xi\psi') = \sigma^2 \frac{1}{N^2 T} \sum_{i=1}^{N} \sum_{k=1}^{N} a_{ik} X_k' M_F [X_i - \bar{X}] = \sigma^2 \frac{1}{T} \bar{X}' M_F (\bar{X} - \bar{X}) = 0$$

note that $a_{ik} = 1$ for all i and k under the null hypothesis.

Proof of (35). Under the iid assumption, $E(\varepsilon_i \varepsilon_j') = 0$ and $E\varepsilon_i \varepsilon_i' = \sigma^2 I_T$. Thus

$$E(\eta\psi') = \sigma^2 \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F [X_i - \iota_T \bar{X}_{i.} - \bar{X} + \iota_T \bar{X}_{..}] = \sigma^2 \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F X_i - \sigma^2 \frac{1}{T} \bar{X}' M_F \bar{X}$$

from $M_F \iota_T = 0$ because F contains ι_T as one of its columns. Next,

$$\begin{aligned} E\xi\psi' &= \sigma^2 \frac{1}{NT} \sum_{i=1}^{N} \Big[\frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k' M_F \Big] [X_i - \iota_T \bar{X}_{i.} - \bar{X} + \iota_T \bar{X}_{..}] \\ &= \sigma^2 \frac{1}{TN^2} \sum_{i=1}^{N} \sum_{k=1}^{N} a_{ik} X_k' M_F X_i - \sigma^2 \frac{1}{NT} \sum_{k=1}^{N} \Big[\frac{1}{N} \sum_{i=1}^{N} a_{ik} \Big] X_k' M_F \bar{X} \\ &= \sigma^2 \frac{1}{TN^2} \sum_{i=1}^{N} \sum_{k=1}^{N} a_{ik} X_k' M_F X_i - \frac{1}{T} \sigma^2 \bar{X}' M_F \bar{X} \end{aligned}$$

because $\sum_{i=1}^{N} a_{ik} = 1$ under the null hypothesis. This follows from $\lambda_i = (1, \alpha_i)'$ with $\sum_i \alpha_i = 0$. Thus

$$E[(\eta - \xi)\psi'] = \sigma^2 \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F X_i - \frac{1}{T} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{k=1}^{N} a_{ik} X_k' M_F X_i = \sigma^2 D(F^0).$$

References

[1] Ahn, S.G., Y-H. Lee and Schmidt, P. (2001): GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects, *Journal of Econometrics*, 102, 219-255.

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

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

[4] 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.

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

[6] 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.

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

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

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

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

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

[12] 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.*

[13] 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.

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

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

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

[17] 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.

[18] 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.

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

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

[21] 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.

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

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

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

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

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

[27] 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.

[28] 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.

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

[30] 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.

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

[32] 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.

[33] 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.

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

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

[36] 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.

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

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

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

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

[41] 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.

[42] 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.

Table 1. Various estimators, iid errors

		Within 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	1.363	0.145	3.364	0.145	0.990	0.161	3.008	0.158	1.022	0.236	3.025	0.229
100	5	1.382	0.096	3.382	0.098	1.000	0.089	3.000	0.086	1.021	0.133	3.021	0.129
100	10	1.388	0.064	3.393	0.063	0.998	0.055	3.002	0.054	1.011	0.071	3.014	0.067
100	20	1.396	0.043	3.399	0.042	0.997	0.034	3.002	0.035	1.002	0.040	3.006	0.040
100	50	1.399	0.027	3.400	0.027	1.000	0.021	3.001	0.021	1.002	0.024	3.003	0.024
100	100	1.399	0.020	3.399	0.020	1.000	0.015	2.999	0.015	1.001	0.017	3.000	0.017
3	100	1.360	0.150	3.361	0.136	0.996	0.090	2.999	0.093	1.039	0.240	3.032	0.231
5	100	1.384	0.098	3.380	0.095	1.003	0.071	2.998	0.070	1.025	0.132	3.019	0.128
10	100	1.389	0.062	3.393	0.063	0.998	0.046	3.002	0.048	1.009	0.066	3.011	0.069
20	100	1.394	0.043	3.395	0.042	0.999	0.034	3.001	0.035	1.004	0.041	3.006	0.041
50	100	1.399	0.027	3.398	0.028	1.000	0.021	3.000	0.021	1.002	0.024	3.002	0.024

Table 2. Various estimators, cross-sectionally correlated errors

		Within 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	1.368	0.136	3.366	0.142	1.005	0.176	2.996	0.172	1.062	0.235	3.061	0.242
100	5	1.381	0.094	3.382	0.092	0.995	0.092	2.999	0.093	1.064	0.152	3.069	0.157
100	10	1.390	0.061	3.393	0.061	0.998	0.056	3.002	0.058	1.053	0.105	3.057	0.107
100	20	1.397	0.043	3.395	0.042	1.001	0.039	2.999	0.038	1.033	0.078	3.031	0.076
100	50	1.397	0.026	3.400	0.026	0.999	0.023	3.001	0.022	1.010	0.046	3.013	0.046
100	100	1.399	0.020	3.399	0.019	1.000	0.016	3.000	0.016	1.006	0.030	3.005	0.030
3	100	1.368	0.110	3.370	0.105	1.002	0.089	2.999	0.091	1.176	0.166	3.181	0.171
5	100	1.382	0.075	3.385	0.076	1.000	0.070	3.000	0.070	1.222	0.117	3.218	0.117
10	100	1.394	0.053	3.392	0.056	1.002	0.050	2.998	0.049	1.237	0.089	3.238	0.090
20	100	1.396	0.040	3.395	0.041	1.000	0.038	2.999	0.037	1.227	0.089	3.227	0.088
50	100	1.399	0.027	3.398	0.027	1.001	0.024	3.000	0.023	1.072	0.116	3.071	0.117

Matlab programs are available from the author

Table 3. Models of Additive Effects

		Within 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

Table 4. Models with grand mean, time-invariant regressors, and common regressors

		Infeasible Estimator									
N	T	Mean $\beta_1 = 1$	SD	Mean $\beta_2 = 3$	SD	Mean $\mu = 5$	SD	Mean $\gamma = 2$	SD	Mean $\delta = 4$	SD
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