Asymptotic Distribution of Factor Augmented Estimators for Panel Regression*

Ryan Greenaway-McGrevy
Bureau of Economic Analysis
Washington, D.C.

Chirok Han
Department of Economics
Korea University

Donggyu Sul
School of Economic, Political and Policy Sciences
University of Texas at Dallas

December 2010

Abstract

In this paper we derive asymptotic theory for linear panel regression augmented with estimated common factors. We give conditions under which the estimated factors can be used in place of the latent factors in the regression equation. For the principal component estimates of the factor space it is shown that these conditions are satisified when $T/N \to 0$ and $N/T^3 \to 0$ under regularity. Monte-Carlo studies verify the asymptotic theory.

Keywords: Factor Augmented Panel Regression, Factor Augmented Estimator, Principal Component Augmented Estimator, Cross Section Dependence, Interactive Fixed Effects,

JEL Classification: C33

*The views expressed herein are those of the authors and not necessarily those of the Bureau of Economic Analysis or the Department of Commerce. We thank I. Choi, Sainan Jin, C.J. Kim, H. R. Moon, Liangjun Su, K. Tanaka, Y. J. Whang, the editors and anonymous referees for helpful comments. A previous version of this paper was circulated under the title "Estimating and Testing Idiosyncratic Equations using Cross Section Dependent Panel Data". Research for the paper was supported under a Marsden grant.

1 Introduction

Recent panel-data research has incorporated strong cross-section dependence into the conventional panel regression model by introducing a factor structure into the regression error. A prototypical panel regression with a factor error structure is given by

(1) $\quad y_{it} = \beta' X_{it} + u_{it}, \;\; u_{it} = \lambda_i^{u\prime} F_t^u + \varepsilon_{it}$

where F_t^u is a r-vector of latent common factors in the regression error u_{it}, λ_i^u is a vector of factor loadings, and X_{it} is a vector of explanatory variables (see, e.g., Pesaran, 2006; Bai, 2009a). No restriction is imposed on the relationship between the regressor X_{it} and the common component $\lambda_i^{u\prime} F_t^u$, so the conventional least squares (LS) or least-squares dummy variable (LSDV) methods may yield inconsistent estimators due to endogeneity.

Several methods have recently been developed to consistently estimate the β parameter. Ahn, Lee and Schmidt (ALS, 2006) extend the single factor model of Ahn, Lee and Schmidt (2001) to allow multiple factors, and provide estimation methods based on moment restrictions on the error term (e.g., white noise or parametric ARMA structure) for small T (time-series observations). Pesaran (2006) proposes filtering out common factors by including the cross-sectional averages of $(y_{it}, X_{it}')'$ in a regression. Under regularity this 'common correlated effects' (CCE) estimator is consistent provided a rank condition is satisfied (e.g., the number of observed variables in the equation is at least as large as r). Recently, Bai (2009a) proposes estimating β jointly with the factor space $\{F_1^u, \ldots, F_T^u\}$ and factor loadings $\{\lambda_1^u, \ldots, \lambda_N^u\}$ by minimizing

$$SSR(\beta, F_1^{uk}, \ldots, F_T^{uk}, \lambda_1^{uk}, \ldots, \lambda_N^{uk}) = \textstyle\sum_{i=1}^N \sum_{t=1}^T (y_{it} - \lambda_i^{uk\prime} F_t^{uk} - \beta' X_{it})^2,$$

subject to the normalizations $T^{-1} \sum_{t=1}^T F_t^{uk} F_t^{uk\prime} = I_k$ and $N^{-1} \sum_{i=1}^N \lambda_i^{uk} \lambda_i^{uk}$ is diagonal, $F_t^{uk} \in \mathbb{R}^k$, and $\lambda_i^{uk} \in \mathbb{R}^k$ for given k. Bai (2009a) shows that this LS estimator is consistent as $N, T \to \infty$ as long as $k \geq r$, without requiring the rank condition of Pesaran (2006), and permitting general weak dependence and heteroskedasticity in the error term ε_{it}. Note that Bai (2009a) controls for common factors of the regression 'residuals' (i.e., $y_{it} - X_{it}'\beta$), whereas Pesaran (2006), by using the cross-section averages of $(y_{it}, X_{it}')'$, controls for the common factors to the 'observable' variables (i.e. y_{it} and X_{it}).

Alternatively one may augment the panel regression with some other factor estimate from the observable variables, such as the principal components (PC) estimate. In fact, Kapetanios and Pesaran (2007) consider a version of this factor augmented estimator and study the finite sample properties by means of small Monte Carlo experiments. Giannone and Lenza (2008) use a similar factor augmented panel regression to estimate the international saving-investment relationship, in which global factors are extracted from the observables.

One purpose of the present paper is to establish formal asymptotics for these factor augmented panel regressions. As yet, rigorous asymptotics have not been derived, although some authors have suggested required conditions under which the PC estimate can replace the common factors in the panel regression without affecting the limiting distribution of the LS estimator. Specifically, based on his Theorem 1, Bai (2003, p. 146) states that the conditions

$\sqrt{T}/N \to 0$ and $N, T \to \infty$ are sufficient under regularity for replacing unobservable common factors with the PC estimate in time series models (Stock and Watson, 2002; Bernanke, Boivin, and Eliasz, 2005; Bai and Ng, 2006). Kapetanios and Pesaran (2007) and Giannone and Lenza (2008) conjecture that this condition also applies to panel regression models such as (1). However, in the pooled panel regression, the fact that the factor loadings are individually specific confounds this generalization, and $\sqrt{T}/N \to 0$ is in fact not sufficient for using PC estimated factors in place of the true factors. Instead we find that $T/N \to 0$ and $N/T^3 \to 0$ are sufficient for the replacement of the unobservable common factors with the PC estimates under regularity. This finding is supported by intuitive explanation and is verified by simulation. In establishing these results we provide general conditions under which any factor estimate can replace the true common factors in the regression, so that our results can be straightforwardly applied to various other factor estimates such as GMM estimators (ALS, 2006) or efficient PC estimates (Choi, 2008). This is our first contribution.

Another purpose of this paper is to propose and establish asymptotics for a straightforward one-step dynamic estimator. This method is of practical interest because it exhibits greater efficiency in the presence of serial correlation in the regression errors.

The remainder of the paper consists of four sections. In the next section we explain the model and the estimators. Section 3 derives the asymptotic properties of those estimators, and provides Monte Carlo studies to verify the established theorems. Section 4 concludes. To save the space, we do not provide technical proofs of the theorems presented herein. The proofs are available from the authors upon request. Throughout, '$\to_p$' denotes convergence in probability, and '$\Rightarrow$' convergence in distribution, as $N \to \infty$ and $T \to \infty$ jointly; $\|A\| = \mathrm{tr}(A'A)$; 'LLN' is an acronym for 'law of large numbers', and 'CLT' for 'central limit theory'.

2 Model and Estimators

In vector notation, (1) is given by

(2) $\quad y_i = X_i\beta + F^u\lambda_i^u + \varepsilon_i,$

where $y_i = (y_{i1}, \ldots, y_{iT})'$, $F^u = (F_1^u, \ldots, F_T^u)'$, $X_i = (X_{i1}, \ldots, X_{iT})'$, and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$. Following convention (e.g. Pesaran, 2006; Bai, 2009a), we permit X_i to be arbitrarily correlated with the unobservable $F^u\lambda_i^u$. In many factor-error regressions this correlation is modelled through a latent factor structure in X_i. That is, when

(3) $\quad X_i = F^X\lambda_i^X + V_i,$

we permit $F^X\lambda_i^X$ to be correlated with $F^u\lambda_i^u$.

Let F^y denote the common factors to y_i, and let λ_i^y denote the associated factor loading vector. We define F to be the $T \times m$ matrix consisting of a subset of the columns of F^X and F^y such that $T^{-1}F'F$ is nonsingular, and $F^X = FA^X$ and $F^y = FA^y$ for some selection matrices A^X and A^y. That is, F contains all unique columns in F^X and F^y, and when X_i and y_i share the same factor, it is included only once as a column in F. Then $F^u\lambda_i^u = F^y\lambda_i^y - F^X\lambda_i^X\beta = F(A^y\lambda_i^y - A^X\lambda_i^X\beta)$, which implies that $F^u = FA^u$ for some selection matrix A^u. Thus, by augmenting the panel regression equation $y_i = X_i\beta + u_i$ with F, we can always control for F^u. But since F is unobservable, this estimation method is infeasible. Instead we consider augmenting the regression with the *estimated* common factors.

It is worth noting that this treatment partials out more variation in X_{it} than necessary for the identification of β, thus it may lead to loss of efficiency.[1] In contrast, Bai (2009a) controls for the residual common factors only (i.e., common factors in $y_{it} - X_{it}'\beta$), and it is conjectured that his estimator is asymptotically efficient if ε_{it} are iid over i and t (Bai, 2009a, Corollary 1). In practical applications, factor number estimation is important, and Bai (2009b) provides a criterion which gives a consistent factor number estimator. However, the small sample performance of the LS estimator with the factor number estimated this way might be compromised, possibly leading to a multi-modal sampling distribution of the β estimate. This problem is especially likely for data generating processes in which fewer common factors exist in $y_{it} - X_{it}'b$ than in $y_{it} - X_{it}'\beta$ for some $b \neq \beta$.

We have defined F as the maximal common factor set of the observable variables $(y_i, X_i')'$ listed without duplication. Let $\lambda_{X,i}$ and $\lambda_{u,i}$ be such that $F^X\lambda_i^X = F\lambda_{X,i}$ and $F^u\lambda_i^u = F\lambda_{u,i}$. (When $F^X = FA^X$ and $F^y = FA^y$ as above, we have $\lambda_{X,i} = A^X\lambda_i^X$ and $\lambda_{u,i} = A^y\lambda_i^y - A^X\lambda_i^X\beta$.) Let $Z_{ab} \equiv N^{-1/2}\sum_{i=1}^N a_i b_i'$ for any column vectors a_i and b_i.

We impose the following restrictions on (2)–(3), which are expressed as high level assumptions for simplicity. Sufficient fundamental conditions are stated in the remarks to follow.

Assumption A *(i)* $T^{-1}F'F$ *is convergent and asymptotically nonsingular;*

(ii) $N^{-1}\sum_{i=1}^N \mathrm{vec}(\lambda_{X,i},\lambda_{u,i})\mathrm{vec}(\lambda_{X,i},\lambda_{u,i})' = O_p(1)$, *and* $N^{-1}\sum_{i=1}^N(\lambda_{X,i}\lambda_{X,i}' + \lambda_{u,i}\lambda_{u,i}')$ *is convergent and asymptotically nonsingular;*

(iii) Let $Q_{\varepsilon\varepsilon} \equiv N^{-1}\sum_{i=1}^N \varepsilon_i\varepsilon_i'$, *and* $Q_{vv} = N^{-1}\sum_{i=1}^N \mathrm{vec}(V_i)\mathrm{vec}(V_i)'$. *We have* $E(\|Q_{\varepsilon\varepsilon}\|^2) = O(T^2)$ *and* $E(\|Q_{vv}\|^2) = O(T^2)$;

(iv) The maximal eigenvalues of $T^{-1}Q_{\varepsilon\varepsilon}$ *and* $T^{-1}Q_{vv}$ *are* $O_p(\tilde\delta_{NT}^{-2})$, *where* $\tilde\delta_{NT}^2 = \min(N,T)$;

(v) $E(\|Z_{\varepsilon\lambda}\|^2) = O(T)$ *and* $E(\|Z_{V\lambda}^a\|^2) = O(T)$ *for all* a, *where* $Z_{V\lambda}^a \equiv N^{-1/2}\sum_{i=1}^N V_{a,i}\lambda_i'$ *with* $V_{a,i}$ *denoting the* a*th column of* V_i *and* $\lambda_i \equiv (\lambda_{X,i}', \lambda_{u,i}')'$;

(vi) $E(\|F'Z_{\varepsilon\lambda}\|^2) = O(T)$ *and* $E(\|F'Z_{V\lambda}^a\|^2) = O(T)$ *for all* a;

(vii) $E(\|Z_{V\varepsilon}^a\|^2) = O(T^2)$ *for all* a, *where* $Z_{V\varepsilon}^a \equiv N^{-1/2}\sum_{i=1}^N V_{a,i}\varepsilon_i'$;

(viii) $E(\|Z_{V\varepsilon}^a F\|^2) = O(T^2)$ *and* $E(\|F'Z_{V\varepsilon}^a\|^2) = O(T^2)$ *for all* a.

Assumptions A(i) and A(ii) are conventional in the approximate factor model literature. The second part of condition (ii) ensures that each common factor has a nontrivial contribution to the variance of at least one of the elements of $(X_{it}', u_{it})'$, so that the regularity conditions of Bai (2003) are satisfied (see Kapetanios and Pesaran, 2007). Assumption A(iii) means that

$$\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{jt}\varepsilon_{js}] = O(1),$$

[1] This was pointed out by an anonymous referee to one of our previous drafts of this paper.

which holds if $E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{jt}\varepsilon_{js}]$ is uniformly bounded. The same arguments apply to the V_i part. Assumption A(iv) for ε_i is taken from a result in Yin, Bai and Krishnaiah (1988, Theorem 3.1) and Bai and Ng (2002b). For example, it holds for ε_{it} if ε_{it} is an element of $A_T e B_N$, where e is the $T \times N$ matrix of iid random variables with finite fourth moments, and where the eigenvalues of $A_N' A_N$ and $B_T' B_T$ are uniformly bounded, because then the maximal eigenvalue of $T^{-1}Q_{\varepsilon\varepsilon}$ is of order $(T^{-1/2} + N^{-1/2})^2 \leq 4/\min(N,T)$. A similar treatment can be made to permit weak dependence and heteroskedasticity among the elements of V_{it}. Assumption A(v) is motivated as follows: Note that $\|Z_{\varepsilon\lambda}\|^2 = N^{-1}\sum_i\sum_j \lambda_i'\lambda_j\varepsilon_i'\varepsilon_j$. If $|E(\varepsilon_i'\varepsilon_j/T|\lambda_1,\ldots,\lambda_N)| \leq \omega_{ij}$ and $|E(\lambda_i'\lambda_j)| \leq \bar{C}_\lambda$ for some universal constant $\bar{C}_\lambda$, then we have

$$E(\|Z_{\varepsilon\lambda}\|^2) \leq \frac{T}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_{ij}\bar{C}_\lambda,$$

which is $O(T)$ if $N^{-1}\sum_{i=1}^N\sum_{j=1}^N \omega_{ij} < \infty$. The same remarks apply to the second part. This assumption relates to Lemmas 1(ii) and (iv) of Bai and Ng (2002). Assumption A(vi) means

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E\Big[F_t\varepsilon_{it}\lambda_i'\lambda_j\varepsilon_{js}F_s'\Big] = O(1),$$

which holds if $E(F_s'F_t\lambda_i'\lambda_j)$ is uniformly bounded and

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E(\varepsilon_{it}\varepsilon_{js}) = O(1),$$

under the assumption that ε_{it} is independent of the common factors and the factor loadings. Similar arguments apply to the second part of the assumption. Assumption A(vii) is satisfied if the second moments of the elements of $Z_{V\varepsilon}^a$, i.e., $N^{-1}\sum_{i=1}^N\sum_{j=1}^N E(V_{it}^a V_{jt}^a \varepsilon_{is}\varepsilon_{js})$, are uniformly bounded. This would be satisfied in general unless cross section dependence is too strong so, e.g., $Z_{V\varepsilon}^a$ does not follow a CLT element-wise. The first part of Assumption A(viii) can be written as $E(F' Z_{V\varepsilon}^{a\prime} Z_{V\varepsilon}^a F) = O(T^2)$. Intuitively, each element of $T^{-1/2}Z_{V\varepsilon}^a F$ is likely bounded and asymptotically random (if $EZ_{V\varepsilon}^a F = 0$), thus the sum of its squared elements would be bounded in the mean. To illustrate this more rigorously, let X_{it} be scalar. Assumption A(viii) means

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T} E[F_s V_{is}\varepsilon_{it}\varepsilon_{jt}V_{jr}F_r'] = O(1).$$

If $|E(\varepsilon_{it}\varepsilon_{jt}|V_1,\ldots,V_N,F)| \leq \omega_{ij}^\varepsilon$ for all t, then the left hand side (which is nonnegative) is bounded by

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\omega_{ij}^\varepsilon|E(V_{is}V_{jr}F_r'F_s)|.$$

If furthermore $|E(V_{is}V_{jr}|F)| \leq \omega_{sr}^V$ for all i and j, and if $|E(F_r'F_s)|$ is uniformly bounded, then the above quantity is bounded by a universal constant times

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\omega_{ij}^{\varepsilon}\omega_{rs}^{V}=\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_{ij}^{\varepsilon}\right]\cdot\left[\frac{1}{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\omega_{sr}^{V}\right].$$

Thus the first part of Assumption A(viii) holds if the right hand of the above displayed expression is bounded, which is a weak condition. The second part can be examined in a similar manner.

Our interest is a pooled regression of y_i on X_i augmented with an estimate $\hat{F}$ of the common factors F. This *factor augmented estimator* (FAE) is defined as

(4) $\quad \hat{b}_{\text{FAE}} \equiv \left(\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}} y_i,$

where $M_A \equiv I - A(A'A)^{-1}A'$ for any full column rank matrix A.

We also consider improving efficiency by estimating an equation that includes lagged defactored variables as regressors. More specifically, one can estimate β by fitting

(5) $\quad \ddot{y}_{it} = \ddot{X}_{it}'\beta + \ddot{X}_{it-1}'\phi + \rho\ddot{y}_{it-1} + \text{error}_{it},$

where $\ddot{y}_{it}$ and $\ddot{X}_{it}'$ are the t-th rows of $\ddot{y}_i \equiv M_{\hat{F}} y_i$ and $\ddot{X}_i \equiv M_{\hat{F}} X_i$ respectively. The β estimate from this regression, denoted as $\hat{b}_{\text{SFAE}}$, is called the *FAE under serial correlation* (SFAE in short).

When the common factors F are estimated using the PC method, we call the resulting feasible estimator (4) the *principal component augmented estimator* (PCAE) and denote it by $\hat{b}_{\text{PCA}}$. Similarly, the SFAE estimator from (5) using the PC factor estimates is called the *PCAE under serial correlation* (SPCAE in short), and it is denoted by $\hat{b}_{\text{SPCA}}$.

3 Asymptotics

Replacing F^u with the full factor set F, we can write (2) as

(6) $\quad y_i = X_i\beta + F\lambda_{u,i} + \varepsilon_i,$

where $\lambda_{u,i}$ satisfies $F\lambda_{u,i} = F^u\lambda_i^u$ as explained previously. In this section, we provide asymptotics for the factor augmented estimators $\hat{b}_{\text{PCA}}$ and $\hat{b}_{\text{SPCA}}$ defined in the previous section.

3.1 Ordinary Factor Augmented Estimator

We first consider an *infeasible* factor-augmented estimator, denoted as $\hat{b}_{\text{I,FAE}}$, which is obtained from a pooled regression of $M_F y_i$ on $M_F X_i$. It satisfies:

$$\hat{b}_{\text{I,FAE}} = \beta + \left(\sum_{i=1}^{N} X_i' M_F X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_F \varepsilon_i.$$

The estimator is infeasible since the factors are assumed to be known. When ε_{it} is stationary over t, a $\sqrt{NT}$ convergence rate would be obtained under regularity, and the limit distribution of $\sqrt{NT}(\hat{b}_{\text{I,FAE}} - \beta)$ is also naturally obtained from the behavior of $(NT)^{-1}\sum_{i=1}^{N} X_i' M_F X_i$ and $(NT)^{-1/2}\sum_{i=1}^{N} X_i' M_F \varepsilon_i$. We maintain the following assumption.

Assumption B *As* $N, T \to \infty$, $(NT)^{-1}\sum_{i=1}^{N} X_i' M_F X_i \to_p \Sigma_X$ *which is nonsingular, and* $(NT)^{-1/2}\sum_{i=1}^{N} X_i' M_F \varepsilon_i \Rightarrow N(0, A_{\text{FAE}})$ *for some* A_{FAE}.

When F is replaced with an estimate $\hat{F}$, the corresponding FAE satisfies

(7) $$\hat{b}_{\text{FAE}} = \beta + \left(\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}} (\varepsilon_i + F\lambda_{u,i}).$$

The properties of $\hat{b}_{\text{FAE}}$ depend heavily on the last term. In particular, consistency requires that $plim (NT)^{-1}\sum_{i=1}^{N} X_i' M_{\hat{F}}(\varepsilon_i + F\lambda_{u,i}) = 0$, while in order for $\hat{b}_{\text{FAE}}$ to have an unbiased limiting distribution, we require $(NT)^{-1/2}\sum_{i=1}^{N} X_i' M_{\hat{F}}(\varepsilon_i + F\lambda_{u,i})$ to be centered at zero.

We will consider 'consistent' factor estimates $\hat{F}$ in the sense that

(8) $$T^{-1}(\hat{F} - FH)'(\hat{F} - FH) = O_p(\delta_{NT}^{-2}) \text{ for some } \delta_{NT} \to \infty,$$

where H is asymptotically nonsingular. (Note that $F\lambda_{u,i} = FHH^{-1}\lambda_{u,i}$, so any nonsingular transformation of the columns of F can also be regarded as common factors. See Bai and Ng, 2002a.) In (8), δ_{NT} is a function of N and T. For example, Bai and Ng (2002a) show that for the PC estimator, $\delta_{NT} = \min[N, T]^{1/2}$.

Theorem 1 below gives asymptotic theory for the standard FAE. Given that the PC factor estimation method is popular in practice, we also provide theory for the PCAE $\hat{b}_{\text{PCA}}$. For the PCAE, we take the results given in Bai and Ng (2002a) as a high level assumption in Theorem 1. We have the following results as $N, T \to \infty$.

Theorem 1 *Under Assumption A: (i) if (8) is satisfied for some* $\delta_{NT} \to \infty$ *and asymptotically nonsingular* H*, then* $plim\, \hat{b}_{\text{FAE}} = \beta$*; (ii) If the conditions in (i) hold, and if (a)* $T^{1/2}\delta_{NT}^{-2} \to 0$*, (b)* $F' M_{\hat{F}} F = o_p(1)$*, and (c)* $F' M_{\hat{F}} F = o_p(\sqrt{T/N})$*, then* $\sqrt{NT}(\hat{b}_{\text{FAE}} - \hat{b}_{\text{I,FAE}}) \to_p 0$*; (iii) If* $T/N \to 0$, $N/T^3 \to 0$ *and if the PC estimator* $\hat{F}$ *satisfies (8) for some* $\delta_{NT}^2 = \min(N, T)$*, then* $\sqrt{NT}(\hat{b}_{\text{PCA}} - \hat{b}_{\text{I,FAE}}) \to_p 0$.

Remarks.

1. Theorem 1(i) states that the consistency of an standard FAE requires only the consistency of the associated factor estimator in the sense of (8). For example, the PC estimator is consistent when $N, T \to \infty$ in the presence of serial correlation and weak cross-sectional dependence in the idiosyncratic errors V_{it} and ε_{it} (Bai and Ng, 2002a). If T is fixed and the errors are uncorrelated over t, then various $\sqrt{N}$-consistent factor estimates are available (e.g., ALS, 2006, classical PC estimates, etc.).

2. Theorem 1(ii) means that under three additional conditions, the standard FAE and its infeasible counterpart are asymptotically equivalent up to order $(NT)^{-1/2}$, thus having the same asymptotic distribution. When $T/N \to 0$, condition (c) implies condition (b). Note that conditions (a)–(c) are not automatically satisfied by consistent F estimates and should be checked for each factor estimate. If T is fixed and $\hat{F}$ is $\sqrt{N}$-consistent (e.g., ALS, 2006), then all the conditions are satisfied.

3. The condition that $\delta_{NT}^2 = \min(N,T)$ is a result derived by Bai and Ng (2002a, Theorem 1) from a set of fundamental assumptions. See Bai and Ng (2002a) for full discussion.

4. Theorem 1(iii) gives conditions under which the PC estimate satisfies the requirements of Theorem 1(ii), such that $\sqrt{NT}(\hat{b}_{\text{PCA}} - \hat{b}_{\text{I,FAE}}) \to_p 0$. These conditions merit some discussion. Notably these conditions are different from the $\sqrt{T}/N \to 0$ condition given in Kapetanios and Pesaran (2007) and Giannone and Lenza (2008). The condition that $N/T^3 \to 0$ and $T/N \to 0$ is justified as follows. When the idiosyncratic errors V_{it} and ε_{it} are serially correlated, the factor estimate may be biased for small T. If N increases fast while T increases too slowly, then the remaining small bias can be amplified greatly (when multiplied by $\sqrt{NT}$), so the limiting distribution can be biased. The condition that $N/T^3 \to 0$ precludes this possibility. On the other hand, if T grows too fast compared to N, then the discrepancy between FH and $\hat{F}$ may accumulate, possibly resulting in a biased asymptotic distribution. This possibility is precluded by the condition that $T/N \to 0$.

5. Theorem 1(iii) suggests that if N is large compared to T and if T is not too small, then the asymptotic distribution of the PCAE is equivalent to that of its infeasible counterpart. But if $T > N$, the asymptotic distribution of $\sqrt{NT}(\hat{b}_{\text{PCA}} - \beta)$ is biased. But in that case, one can simply estimate the factor loadings of $(y_{it}, X_{it}')'$ first and then augment the panel regression with the estimated factor loadings. By switching the roles of N and T, and of the common factors and the factor loadings, we can see that this 'factor loading augmented' estimator is asymptotically equivalent to the corresponding infeasible estimator if $N/T \to 0$ and $T/N^3 \to 0$. However if $N/T \to c > 0$, then the asymptotic distribution of the FAE is biased. A bias correction as proposed in Bai, 2009a, for this case would be an interesting future research topic.

We next consider the computation of standard errors for the standard FAE under the assumption that the random variables are independent across i, such that

(9) $$A_{\text{FAE}} = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^N E(X_i' M_F \varepsilon_i \varepsilon_i' M_F X_i).$$

When the FAE is equivalent to its infeasible counterpart, the variance of the asymptotic distribution of $\sqrt{NT}(\hat{b}_{\text{FAE}} - \beta)$ is $V_{\text{FAE}} \equiv \Sigma_X^{-1} A_{\text{FAE}} \Sigma_X^{-1}$, because of Theorem 1(ii) and Assumption B. The Σ_X term is naturally estimated by $\hat{\Sigma}_X \equiv (NT)^{-1} \sum_{i=1}^N X_i' M_{\hat{F}} X_i$, and A_{FAE} in (9) is estimated by

(10) $$\hat{A}_{\text{FAE}} \equiv \frac{1}{NT} \sum_{i=1}^N X_i' M_{\hat{F}} \hat{\varepsilon}_i \hat{\varepsilon}_i' M_{\hat{F}} X_i, \quad \hat{\varepsilon}_i \equiv y_i - X_i \hat{b}_{\text{FAE}},$$

under the assumption of cross sectional independence for ε_{it}.

Pesaran (2006) proposes another method of estimating the asymptotic variance. Let $\hat{b}_i$ be the individual feasible β estimate, i.e., $\hat{b}_i = (X_i' M_{\hat{F}} X_i)^{-1} X_i' M_{\hat{F}} y_i$. Then we have $\hat{b}_i = \beta + (X_i' M_{\hat{F}} X_i)^{-1} X_i' M_{\hat{F}} (\varepsilon_i + F\lambda_{u,i})$, thus

$$X_i' M_{\hat{F}} \varepsilon_i + X_i' M_{\hat{F}} F \lambda_{u,i} = X_i' M_{\hat{F}} X_i (\hat{b}_i - \beta),$$

where the second term on the left hand side is negligible in the sense that its second sample moment $(NT)^{-1} \sum_{i=1}^N X_i' M_{\hat{F}} F \lambda_{u,i} \lambda_{u,i}' F' M_{\hat{F}} X_i$ asymptotically disappears. Thus A_{FAE} can also be estimated by

$$\text{(11)} \quad \tilde{A}_{\text{FAE}} = \frac{1}{NT} \sum_{i=1}^N X_i' M_{\hat{F}} X_i (\hat{b}_i - \bar{b})(\hat{b}_i - \bar{b})' X_i' M_{\hat{F}} X_i, \quad \bar{b} = \frac{1}{N} \sum_{i=1}^N \hat{b}_i.$$

According to supplementary simulations (not reported) this variance estimate performs quite well even in small samples.

The analysis so far is based on the supposition that the factor numbers are known or correctly estimated. When the factor numbers are unknown, they can be consistently estimated using the selection criteria suggested by Bai and Ng (2002a), for example.

3.2 FAE under Serial Correlation

In this subsection we provide asymptotic results for the SFAE, which is proposed in order to enhance efficiency in the presence of serial correlation in ε_{it}. The analysis is similar to the previous case: We consider an infeasible estimator first, and then show that a feasible estimator is asymptotically equivalent to the infeasible counterpart under certain conditions. Let $\dot{y}_i \equiv M_F y_i$ and $\dot{X}_i \equiv M_F X_i$, and let $\dot{y}_{it}$ and $\dot{X}_{it}'$ denote the t-th rows of $\dot{y}_i$ and $\dot{X}_i$ respectively. The infeasible estimator $\hat{b}_{\text{I,SFAE}}$ of β under serial correlation is the estimated coefficient of $\dot{X}_{it}$ when $\dot{y}_{it}$ is regressed on $\dot{X}_{it}$, $\dot{X}_{it-1}$ and $\dot{y}_{it-1}$ by pooled least squares. (To specify a higher AR order, one can simply use more lagged variables on the right hand side. For example, if an AR(2) specification is to be fitted, $\dot{X}_{it}$, $\dot{X}_{it-1}$, $\dot{X}_{it-2}$, $\dot{y}_{it-1}$ and $\dot{y}_{it-2}$ appear on the right hand side.) Importantly, β is consistently estimated by this autoregressive estimation, despite $\dot{y}_{it-1}$ being correlated with the regression error. (This is shown by Phillips and Sul, 2007, p. 169, for the case when case for $F_t = (1,t)'$.)

Under regularity similar to Assumption B for a panel CLT, the infeasible estimator $\hat{b}_{\text{I,SFAE}}$ is asymptotically normal. The assumption below ensures this holds. Let $W_{it} \equiv [\dot{X}_{it}', \dot{X}_{it-1}', \dot{y}_{it-1}]'$, which is the infeasible de-factored regressor vector. Let $\dot{\varepsilon}_{it}$ be the $t-$th row of $\dot{\varepsilon}_i \equiv M_F \varepsilon_i$. We assume the following.

Assumption C *As $N \to \infty$ and $T \to \infty$, $(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T W_{it} W_{it}' \to_p \Sigma_W$, which is nonsingular, and $(NT)^{-1/2} \sum_{i=1}^N \sum_{t=1}^T (\xi_{it} - E\xi_{it})$, where $\xi_{it} \equiv W_{it}(\dot{\varepsilon}_{it}, \dot{\varepsilon}_{it-1})$, is asymptotically normal.*

The SFAE $\hat{b}_{\text{SFAE}}$ is obtained by replacing F with an estimate $\hat{F}$ satisfying (8). That is, letting $\ddot{y}_i \equiv M_{\hat{F}} y_i$ and $\ddot{X}_i \equiv M_{\hat{F}} X_i$ and $M_{\hat{F}} = I - \hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'$, the feasible estimator $\hat{b}_{\text{SFAE}}$ is obtained by regressing $\ddot{y}_{it}$ on $\ddot{X}_{it}$, $\ddot{X}_{it-1}$ and $\ddot{y}_{it-1}$, where $\ddot{y}_{it}$ and $\ddot{X}_{it}'$ are the t-th rows of $\ddot{y}_i$ and $\ddot{X}_i$ respectively. The PCAE estimator for this autoregressive case is the feasible estimator using the PC estimator. As stated above, $\hat{b}_{\text{SPCA}}$ denotes the feasible estimator with PC estimated factors. We have the following result.

Theorem 2 *The results in Theorem 1 also hold when $\hat{b}_{\text{FAE}}$, $\hat{b}_{\text{I,FAE}}$ and $\hat{b}_{\text{PCA}}$ are replaced with $\hat{b}_{\text{SFAE}}$, $\hat{b}_{\text{I,SFAE}}$ and $\hat{b}_{\text{SPCA}}$, respectively.*

The required conditions of Theorem 2 are identical to those of Theorem 1.

To evaluate the asymptotic variance, we note that the autoregressive estimator is algebraically identical to the slope estimate from the regression on $\ddot{z}_{it} = \ddot{X}_{it}'\beta + \eta_{it}$ where $\ddot{z}_{it} \equiv \ddot{y}_{it} - \ddot{X}_{it-1}'\hat{b}_1 - \hat{\rho}\ddot{y}_{it-1}$ with $\hat{\rho}$ and $\hat{b}_1$ being the estimated coefficients of $\ddot{y}_{it-1}$ and $\ddot{X}_{it-1}$ respectively from the autoregressive feasible regression, viz.,

$$(12) \quad \sqrt{NT}(\hat{b}_{\text{SFAE}} - \beta) = \left[\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \ddot{X}_{it}\ddot{X}_{it}'\right]^{-1} \frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T \ddot{X}_{it}(\ddot{z}_{it} - \ddot{X}_{it}'\beta).$$

Because of the asymptotic equivalence of the feasible and infeasible estimators under Theorem 2, and under the assumption that ε_{it} are cross section independent, the variance of the numerator is approximated by

$$(13) \quad \frac{1}{NT}\sum_{i=1}^N \left[\sum_{t=1}^T \ddot{X}_{it}(\ddot{z}_{it} - \ddot{X}_{it}'\beta)\right]\left[\sum_{t=1}^T \ddot{X}_{it}(\ddot{z}_{it} - \ddot{X}_{it}'\beta)\right]'.$$

Now (13) can be estimated by replacing β with $\hat{b}_{\text{SFAE}}$, and the asymptotic variance of (12) is estimated using the usual sandwich form.

Alternatively Pesaran's (2006) method can again be employed. Specifically, for each i, let $\tilde{b}_i = (\sum_t \ddot{X}_{it}\ddot{X}_{it}')^{-1}\sum_t \ddot{X}_{it}\ddot{z}_{it}$, which is the individual autoregressive estimator is obtained by regressing $\ddot{z}_{it}$ on $\ddot{X}_{it}$. Then we have

$$\sum_{t=1}^T \ddot{X}_{it}\ddot{X}_{it}'(\tilde{b}_i - \beta) = \sum_{t=1}^T \ddot{X}_{it}(\ddot{z}_{it} - \ddot{X}_{it}'\beta),$$

so (13) equals $(NT)^{-1}\sum_{i=1}^N \left[\sum_{t=1}^T \ddot{X}_{it}\ddot{X}_{it}'(\tilde{b}_i - \beta)\right]\left[\sum_{t=1}^T \ddot{X}_{it}\ddot{X}_{it}'(\tilde{b}_i - \beta)\right]'$. This can be estimated by replacing β with the average of $\tilde{b}_i$ over i. According to simulations this alternative estimator performs well even when the sample size is small, a result similar to that obtained for the standard FAE.

3.3 Monte-Carlo Study

Monte Carlo experiments are used to evaluate both the finite sample and asymptotic properties of the PCAEs. We generate data from $y_{it} = X_{it}\beta + \gamma\sum_{j=1}^r \lambda_{ji}F_{jt} + \varepsilon_{it}$ with scalar $X_{it} = \sum_{j=1}^r \lambda_{ji}F_{jt} + \sum_{j=1}^p \kappa_{ji}F_{jt}^x + V_{it}$. Here ε_{it}, V_{it}, $\{F_{jt}\}_{j=1}^r$ and $\{F_{jt}^x\}_{j=1}^p$ follow independent AR(1) processes based on iid $N(0,1)$ innovations with coefficients equal to 0.5; and $\{\lambda_{ji}\}_{j=1}^r$ and

$\{\kappa_{ji}\}_{j=1}^{p}$ are independently drawn from $iid\ N(1,1)$ for each i. We set $\beta = 1$ in all simulations and vary r, p and γ. Each simulation is replicated 2,000 times.

Table 1 reports the simulated finite sample properties of several estimators of β: The infeasible standard and dynamic FAE (with true factors $\{F_{jt}\}_{j=1}^{r}$); the standard and dynamic PCAE (using $\{\hat{F}_{jt}\}_{j=1}^{r}$ estimated from the pooled observable variables); and the estimator proposed in Bai (2009a). For the feasible estimators the factor number is known.

We consider two different pairs of factor numbers. The first setting is $r = 2$ and $p = 0$, so that the PCAEs and Bai (2009a) estimator should exhibit similar bias and variance in the large sample since both methods partial out the same factors. Panel A in Table 1 confirms this conjecture. Panel A also shows that the finite sample performance of the two estimators is dependent on the 'signal' of the common factors. For the Bai (2009a) estimator the factors are estimated using the regression residual, and thus the 'signal' is increasing in the magnitude of γ. Thus the bias and variance of the estimator decrease as $|\gamma|$ increases. The signal of the common factors in observable y_{it} increases as the magnitude of $\gamma + \beta$ increases for this particular DGP. Thus as $\gamma + \beta$ gets larger, the bias and variance of the PCAEs decrease. Overall, the small sample performance of the PCAEs relative to the Bai (2009a) estimator is dependent on the β and γ parameters. For example, for $\gamma = -0.5$, the PCAEs outperform the Bai (2009a) estimator in terms of bias and variance, but for $\gamma = 1$, the Bai (2009a) estimator exhibits smaller bias. For $N = T = 100$ all estimators have variance close to that of the infeasible estimator.

In the second DGP we set $r = 1$ and $p = 5$, so the regressor has additional common factors and the PCAEs partial out more variation than necessary for consistency. As discussed in section 2, the Bai (2009a) estimator is more efficient in this DGP. Panel B in Table 1 confirms that the Bai (2009a) estimator has smaller variance as well as bias than the PCAEs.

Additional results available from the authors upon request consider the case where the factor number is estimated using the Bai and Ng (2002a) and Bai (2009b) criteria. In general, with large enough N and T the factor number is estimated consistently so that there is little, if any, effect on estimator performance. However for small N or T, the Bai (2009a) estimates often exhibit large bias when γ is small due to the underestimation of the number of factors in the error. In contrast the number of factors in the observable y_{it} and X_{it} is overestimated for the considered DGPs. Thus all the endogenous factors are partialled out in the PCAE procedures and the point estimates consequently exhibit much less bias.

Table 2 reports the mean and variance of $\sqrt{NT}(\hat{b}_{\text{PCA}} - \hat{b}_{\text{I,FAE}})$ in order to verify Theorem 1(iii). Again the factor number is known. (Our conclusions do not change when the factor number is estimated.) Because the infeasible estimator is unbiased, the mean of $\sqrt{NT}(\hat{b}_{\text{PCA}} - \hat{b}_{\text{I,FAE}})$ is the normalized bias of the PCAE. Note the (absolute) bias has a "U" shape as N increases for given T (this is particularly prominent for small T). This would imply that bias results when growth in T is too slow (compared to N), which partly illustrates the necessity of the $N/T^3 \to 0$ condition. Also, as $T \to \infty$ for fixed N, the absolute bias either increases for small N, or has a "U" shape for large N, which suggests the requirement that $T/N \to 0$ in order for the bias to diminish. In addition the bias does not dissipate as N and T grow with $N = T$, which is also in accordance with the $T/N \to 0$ condition. Additional simulation results available online also verify Theorem 2(iii) for this DGP.

4 Concluding Remarks

In this paper we establish asymptotics for linear panel regression estimators augmented with estimated common factors. A specific rate condition ($T/N \to 0$ and $N/T^3 \to 0$) is derived for the asymptotic equivalence of PC-augmented panel estimators. These conditions are different from those for time series models augmented with factors (i.e., that $\sqrt{T}/N \to 0$; see Bai and Ng, 2006). Monte-Carlo studies support these asymptotic results. We also derive asymptotics for a one-step dynamic estimator which can achieve efficiency gains in the presence of serial correlation in the idiosyncratic regression error.

References

Ahn, S.C., Lee, Y.-H., Schmidt, P. 2001. GMM estimation of linear panel data models with time-varying individual effects. Journal of Econometrics 101, 219–255.

Ahn, S.C., Lee, Y.-H., Schmidt, P. 2006. Panel data models with multiple time-varying individual effects. Working Paper Series 0702, University of Crete.

Bai, J. 2003. Inferential theory for factor models of large dimensions. Econometrica 71,135–171.

Bai, J. 2009a. Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

Bai, J. 2009b. Supplement to "Panel data models with interactive fixed effects": Technical details and proofs (*Econometrica*, Vol. 77, No. 4, July 2009, 1229-1279). Econometrica Supplementary Material.

Bai, J., Ng, S. 2002a. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bai, J., Ng, S. 2002b. Determining the number of factors in approximate factor models, Errata. http://www.columbia.edu/~sn2294/papers/correctionEcta2.pdf.

Bai, J., Ng, S. 2006. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. Econometrica 74, 1133–1150.

Bernanke, B., Boivin, J., Eliasz, P. 2005. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. The Quarterly Journal of Economics 120, 387-422.

Choi, I. 2008. Efficient estimation of factor models. Mimeo, Sogang University.

Giannone, D., Lenza, M. 2008. The Feldstein-Horioka fact. European Central Bank working paper No. 873.

Kapetanios, G., Pesaran, M. H. 2007. Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. In Phillips, G., Tzavalis, E. (Eds.), The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis. Cambridge University Press, Cambridge, pp 239-281 **(**Chapter 10**).**

Pesaran, M.H. 2006. Estimation and inference in large heterogenous panels with a multi factor error structure. Econometrica 74, 967–1012.

Phillips, P.C.B., Sul, D. 2007. Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. Journal of Econometrics 137, 162-188.

Stock, J.H., Watson, M.W. 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97 1167–1179.

Yin, Y.Q., Bai, Z.D., Krishnaiah, P.R. 1988. On the limit of the largest eigenvalue of large dimensional sample covariance matrix. Probability Theory and Related Fields 78, 509–521.

Table 1: Simulated properties of factor-augmented panel estimators

			Bias					Variance $\times$ 1000				
γ	N	T	A	B	C	D	E	A	B	C	D	E
Panel A: $r=2$, $p=0$												
-0.5	25	25	0.003	0.001	-0.076	-0.076	-0.132	2.867	2.093	3.521	2.678	21.74
-0.5	25	100	0.001	0.000	-0.056	-0.056	-0.011	0.693	0.428	0.781	0.493	1.026
-0.5	100	25	0.000	0.000	-0.023	-0.023	-0.054	0.680	0.485	0.733	0.529	5.525
-0.5	100	100	0.000	0.000	-0.013	-0.013	-0.003	0.157	0.103	0.161	0.106	0.169
1	25	25	0.003	0.001	-0.038	-0.039	0.012	2.867	2.093	3.073	2.265	3.553
1	25	100	0.001	0.000	-0.040	-0.040	0.002	0.693	0.428	0.746	0.463	0.764
1	100	25	0.000	0.000	-0.010	-0.009	0.003	0.680	0.485	0.695	0.494	0.719
1	100	100	0.000	0.000	-0.010	-0.010	0.001	0.157	0.103	0.159	0.105	0.160
Panel B: $r=1$, $p=5$												
-0.5	25	25	0.000	0.000	-0.187	-0.185	-0.002	0.253	0.185	4.055	3.672	0.405
-0.5	25	100	0.000	0.000	-0.157	-0.155	0.000	0.061	0.038	0.982	0.629	0.087
-0.5	100	25	0.000	0.000	-0.043	-0.043	-0.001	0.065	0.047	0.848	0.760	0.100
-0.5	100	100	0.000	0.000	-0.034	-0.034	0.000	0.015	0.010	0.189	0.121	0.019
1	25	25	0.000	0.000	-0.169	-0.167	0.001	0.253	0.185	3.886	3.473	0.347
1	25	100	0.000	0.000	-0.154	-0.152	0.000	0.061	0.038	0.946	0.613	0.085
1	100	25	0.000	0.000	-0.035	-0.035	0.000	0.065	0.047	0.820	0.721	0.085
1	100	100	0.000	0.000	-0.033	-0.033	0.000	0.015	0.010	0.189	0.121	0.019

$A = \hat{b}_{\text{I,FAE}},\ B = \hat{b}_{\text{I,SFAE}},\ C = \hat{b}_{\text{PCA}},\ D = \hat{b}_{\text{SPCA}},\ E =$ Bai (2009a) estimator

Table 2: Mean (left) and variance (right) of $\sqrt{NT}(\hat{b}_{\text{PCA}} - \hat{b}_{\text{I,FAE}})$; $\gamma = -0.5$, $r=2$, $p=0$

$N \backslash T$	15	25	50	100	200	15	25	50	100	200
25	-1.904	-1.867	-2.198	-2.849	-3.879	1.391	0.817	0.561	0.455	0.438
50	-1.505	-1.353	-1.513	-1.925	-2.617	0.998	0.456	0.261	0.215	0.215
100	-1.272	-1.060	-1.091	-1.352	-1.795	0.782	0.241	0.133	0.101	0.097
200	-1.268	-0.927	-0.819	-0.979	-1.281	0.781	0.194	0.066	0.054	0.046
1000	-1.893	-1.039	-0.599	-0.510	-0.595	2.115	0.204	0.033	0.014	0.010
2000	-2.431	-1.305	-0.638	-0.440	-0.448	2.904	0.382	0.034	0.009	0.006
4000	-3.373	-1.706	-0.749	-0.417	-0.352	5.280	0.536	0.048	0.008	0.003