Estimation methods in panel data models with observed and unobserved components: a Monte Carlo study

Castagnetti, Carolina and Rossi, Eduardo

University of Pavia

December 2008

Quaderni di Dipartimento

Estimation Methods in Panel Data Models with Observed and Unobserved Components: a Monte Carlo Study

Carolina Castagnetti
(Università di Pavia)

Eduardo Rossi
(Università di Pavia)

211 (12-08)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Dicembre 2008

Estimation Methods in Panel Data Models with Observed and Unobserved Components: a Monte Carlo Study

Carolina Castagnetti *
Università di Pavia

Eduardo Rossi†
Università di Pavia

December 20, 2008

Abstract

Recently some new techniques have been proposed for the estimation of the slope coefficients in presence of unobserved components. Though, the presence of common observed and unobserved factors is neither considered or the estimation of their impacts is not taken into account. In this work a range of estimators is surveyed and their finite-sample properties are examined by means of Monte Carlo experiments. We consider both the properties of estimators for the individual specific components and for the observed common effects.

Keywords: factor error structure; principal component; common regressors; cross-section dependence; large panels, Monte Carlo simulations.
JEL - Classification: C23,C32,C33

*Dipartimento di Economia Politica e Metodi Quantitativi, Via San Felice 5, 27100 Pavia, Italy, `castca@eco.unipv.it, tel.++390382986217, corresponding author`

†Dipartimento di Economia Politica e Metodi Quantitativi, `erossi@eco.unipv.it`

1 Introduction

Recently, the large panel literature has focused on the presence of cross section dependence that may stem from omitted common variables. These unobserved common factors affect each cross-section unit heterogeneously and, when correlated with the regressors, lead to inconsistent regression coefficient estimates. Allowing the errors to be correlated makes the framework suited for a wider range of economic applications. Moreover, the large-dimensional nature of the panel data permits consistent estimation of the factors.

The traditional factor analysis is not an implementable strategy for factor models in large panels. Strict factor models do not work directly with typical macro or finance time series because the characteristics of the data usually conflict with the assumptions. Indeed, the classical factor analysis was developed for cross-sectional data where the assumptions are often reasonable.[1] The assumptions underlying the classical factor analysis are in general restrictive for economic problems. The *i.i.d.* assumption for the error terms, for instance, and diagonality of the idiosyncratic covariance matrix, which rules out cross-section correlation, are too strong for economic time series data. Moreover, classical factor analysis can consistently estimate the factor loadings but not the common factors. However, in economics, it is often the common factors (representing the factor returns, common shocks, diffusion indices, etc.) instead of the factor loadings that are of direct interest.

Approximate factor models (static and dynamic) abandon the assumption that the covariance matrix of the idiosyncratic disturbances is diagonal. In the *approximate K-factor model*, the assumption is that the variance covariance matrix of the idiosyncratic disturbances is no more diagonal: it's possible to have correlation among the disturbance terms. These models use many time series and have relatively few underlying factors. A typical application is asset returns data with both a large period of observations and a large number of assets. Starting from the seminal work of Stock and Watson (2002), a new setting for *approximate factor models* has been developed. See, for example Bai (2003), Bai and Ng (2002), Bai (2005), Bai and Ng (2006a) and Bai and Ng (2006b). The main model assumptions may be summarized as follows. First, both factor loadings and the factors are treated as parameters, as opposed to the factor loadings only as in the classical factor analysis. Second, in general the number of observations is large in both the cross-section and the time series dimensions. Last, the idiosyncratic errors can be weakly serially and cross-sectionally correlated.

In this paper we address the issue of how to estimate panel data models with a multifactor error structure. In fact, whenever an unobserved common factor structure exists the estimates of individual slope coefficients are inconsistent. Recently, different papers propose methods to consistently estimate the effects of the observed individual components. First, we present different methods for the consistent estimation of the coefficients of the individual specific regressors in presence of a multifactor error structure.

Second, we consider also the estimation of the observed common components. This aspect has been quite neglected by the literature which focus only on the estimation of the individual components. However, in panel data analysis, common regressors are more often than not the variables of primary interest. In financial empirical applications, for

[1] See, for instance Wansbeek and Meijer (2000), chapter 7 and Mardia, Kent, and Bibby (1979).

instance, is common practice to consider generalization of standard APT models that allows individual asset returns to be affected both by observed and unobserved common factors. Examples are given in Kapetanios and Pesaran (2005) which consider, in addition to unobserved common factors, the rate of change of oil prices in US Dollars for modelling stock return. Ludvigson and Ng (2008) include the linear combination of five forward spreads obtained by Cochrane and Piazzesi (2005) to explain the excess returns of U.S. government bonds. Bai (2005) suggests to add to the factor structure either the common risk factors identified by Fama and French (1993) or the dividend yields, dividend payout ratio, and consumption gap as in Lettau and Ludvigson (2001) to model asset returns.

Third, we compare via a Monte Carlo simulation exercise the small sample properties of the various estimators considered.

2 Panel data with unobserved and observed common factors

In this section we present a model with unobserved and observed common factors. Consider a linear model, where we include observed common factors along with unobserved ones, we have:

$$y_{it} = \boldsymbol{\alpha}'\boldsymbol{d}_t + \boldsymbol{\beta}'\boldsymbol{x_{it}} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \epsilon_{it} \qquad i = 1, \ldots, I \qquad t = 1, \ldots, T \tag{1}$$

$\boldsymbol{d}_t$ is the $(n \times 1)$ vector of observed common factors, $\boldsymbol{x_{it}}$ is the $(k \times 1)$ vector of individual-specific components, and $\boldsymbol{f}_t$ is the $(r \times 1)$ vector of unobserved factors. Coakley, Fuertes, and Smith (2002) and Kao, Trapani, and Urga (2008) assume that the response of y_{it} to $\boldsymbol{f_t}$ being homogeneous across individuals through $\boldsymbol{\lambda}$. Pesaran (2006) assumes that the individual specific factors are correlated with common (observed and unobserved) factors through:

$$\boldsymbol{x_{it}} = \boldsymbol{\Pi_i}'\boldsymbol{d_t} + \boldsymbol{\Lambda_i}'\boldsymbol{f_t} + \boldsymbol{v}_{it} \qquad i = 1, \ldots, I \qquad t = 1, \ldots, T \tag{2}$$

where $\boldsymbol{d_t}$ are independent of $\boldsymbol{v}_{it}$, $\boldsymbol{\Pi_i}$ and $\boldsymbol{\Lambda_i}$ are $n \times k$ and $r \times k$, factor loading matrices with fixed components and $\boldsymbol{v}_{it}$ are the specific components of $\boldsymbol{x_{it}}$ distributed independently of the common effects and across i.

Bai (2005) considers the case where $\boldsymbol{x_{it}}$ is correlated with $\boldsymbol{\lambda_i}$ alone, or with $\boldsymbol{f_t}$ alone, or simultaneously with both. In this case expression (2) becomes:

$$\boldsymbol{x_{it}} = \boldsymbol{\Pi_i}'\boldsymbol{d}_t + \boldsymbol{A}'\boldsymbol{\lambda_i} + \boldsymbol{B}'\boldsymbol{f_t} + \boldsymbol{c}\boldsymbol{\lambda_i}'\boldsymbol{f_t} + \boldsymbol{v_{it}} \qquad i = 1, \ldots, I \qquad t = 1, \ldots, T \tag{3}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$ are $(r \times K)$ constant matrices and $\boldsymbol{c}$ is a $(K \times 1)$ constant vector.

It is worth noticing that both specifications (2) and (3), allow for $\boldsymbol{x_{it}}$ being dependent upon $\boldsymbol{f_t}$ through heterogeneous loadings $\boldsymbol{\Lambda_i}$. Coakley, Fuertes, and Smith (2002) consider the case where the unobserved factors, $\boldsymbol{f_t}$, may be correlated with the individual specific components $\boldsymbol{x_{it}}$ through:

$$\boldsymbol{x_{it}} = \boldsymbol{\Lambda_i}'\boldsymbol{f_t} + \boldsymbol{v}_{it} \tag{4}$$

where $\boldsymbol{\Lambda_i}$ is a $r \times k$ factor loading matrix and v_{it} are the specific components of $\boldsymbol{x_{it}}$ distributed independently across i. The estimation procedures of the above models proposed in the literature focus on the estimation of $\boldsymbol{\beta}$, the common slope coefficients, under large time series T and cross section dimension N. Section 3 briefly illustrates the methods.

2.1 Assumption on the factor loadings, $\boldsymbol{\lambda_i}$

Coakley, Fuertes, and Smith (2002) adopt a panel model with common unobserved components which are time-varying but constant across i, i.e. the factor loadings $\boldsymbol{\lambda}$ are constant:

$$y_{it} = \boldsymbol{\beta}'\boldsymbol{x_{it}} + \boldsymbol{\lambda}'\boldsymbol{f_t} + \epsilon_{it} \tag{5}$$

Kao, Trapani, and Urga (2008) share the same framework while Bai (2003), Bai (2005) and Pesaran (2006) have a heterogeneous factor-loading specification. Specification (5) rules out the presence among the regressors of either individual or time-invariant regressors. For instance in case of individual-invariant regressors ($\boldsymbol{d_t}$)

$$y_{it} = \boldsymbol{\beta}'\boldsymbol{x_{it}} + \boldsymbol{\alpha}'\boldsymbol{d_t} + \boldsymbol{\lambda}'\boldsymbol{f}_t + \epsilon_{it}$$

where $\boldsymbol{d_t}$ is a $(r \times 1)$ vector of observed common factors. The coefficients ($\boldsymbol{\alpha}$) of the individual-invariant regressors ($\boldsymbol{d_t}$) are obviously not identified if $\boldsymbol{\lambda}$ is constant among individuals.[2]

Pesaran (2006) assumes that the unobserved factor loadings, $\boldsymbol{\lambda_i}$ and $\boldsymbol{\Lambda_i}$ in equations (1) and (2), are independently and identically distributed across i, and of the individual specific errors, ϵ_{jt} and $\boldsymbol{v}_{jt}$, the common factors, $\boldsymbol{d_t}$ and $\boldsymbol{f_t}$, for all i, j and t. In particular, the factor loadings, $\boldsymbol{\lambda_i}$, follow the random coefficient model:

$$\boldsymbol{\lambda_i} = \boldsymbol{\lambda} + \boldsymbol{\eta_i}, \;\; \boldsymbol{\eta_i} \sim IID(\mathbf{0}, \boldsymbol{\Omega}_\eta), \;\; \text{for } \; i = 1, 2, \ldots, I. \tag{6}$$

Though, Bai (2005) treats $\boldsymbol{f_t}$ and $\boldsymbol{\lambda_i}$ as fixed effects parameters to be estimated along with the common slope coefficients $\boldsymbol{\beta}$.[3]

2.2 Assumption on the $\boldsymbol{f}_t$

All the models cited above assume that the number of unobserved factors is fixed but unknown. Once a consistent estimator of the slope parameters, $\widehat{\boldsymbol{\beta}}$, is provided, the consistent

[2]Ahn, Lee, and Schmidt (2001) allow for time-invariant regressors, although they do not consider the joint presence of common regressors: $y_{it} = \boldsymbol{\beta}'\boldsymbol{x_{it}} + \boldsymbol{\alpha}'\boldsymbol{z_i} + \boldsymbol{\lambda}'\boldsymbol{f_t} + \epsilon_{it}$. Again, the coefficients ($\boldsymbol{\alpha}$) of the individual-invariant regressors ($\boldsymbol{z_i}$) are not identified if $\boldsymbol{\lambda}$ is constant among individuals. Ahn, Lee, and Schmidt (2001) assume non-zero correlation between factor loadings and the regressors to identify the parameters.

[3]Bai assumes that:

$$E\|\boldsymbol{f_t}\|^4 \leq \boldsymbol{M} \;\text{ and }\; \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\boldsymbol{f_t}' \xrightarrow{p} \boldsymbol{\Sigma_f} > 0 \;\text{ as }\; T \to \infty$$

and

$$E\|\boldsymbol{\lambda_i}\|^4 \leq \boldsymbol{M} \;\text{ and }\; \frac{1}{I}\boldsymbol{\Lambda\Lambda}' \xrightarrow{p} \boldsymbol{\Sigma_\Lambda} > 0 \;\text{ as }\; I \to \infty$$

for some finite M not depending on I and T. The assumption that both $\boldsymbol{\Sigma_f}$ and $\boldsymbol{\Sigma_\Lambda}$ are definite positive, i.e. have rank $= r$, rules out redundant components in $\boldsymbol{\lambda_i}$. It is worth noticing that in this set up r is equal to the smallest value of the number of factors that the factor representation $\boldsymbol{\lambda_i f_t}$ holds. In fact, as stressed by Ahn, Lee, and Schmidt (2006), if the factor representation $\boldsymbol{\lambda_i f_t}$ holds for a given r, it also holds for any greater number of factors than r.

residuals show a pure factor model structure:

$$y_{it} - \widehat{\boldsymbol{\beta}}' \boldsymbol{x}_{it} = \boldsymbol{\lambda_i}' \boldsymbol{f}_t + \epsilon_{it} + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \boldsymbol{x}_{it}$$

with an added error given by $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \boldsymbol{x}_{it}$ which does not affect the factor model analysis.[4] Therefore, the number of the unobserved common factors can be consistently estimated based on the information criteria approach developed by Bai and Ng (2002).

It's worth noticing that Bai (2003) shows that the distribution of the estimated factors does not depend on whether the number of factors is known or estimated as long as the number of factors is consistently estimated. Hence, once a consistent estimation of the slope coefficients is given, the unobserved factors as well as the factor loadings can be consistently estimated by means of principal components[5] even though up to a non-singular transformation, i.e. a rotation indeterminacy.

3 Alternative panel estimators

3.1 Estimators of the individual specific components, $\boldsymbol{\beta}$.

3.1.1 Common Correlated Effects Estimator (Pesaran, 2006).

In model (1)-(2) Pesaran (2006) put forward, using cross section averages of y_{it} and $\boldsymbol{x_{it}}$ as proxies for the latent factors, $\boldsymbol{f_t}$, a consistent estimator for $\boldsymbol{\beta}$. The basic idea behind the proposed estimation procedure, the Common Correlated Effects (CCE) estimator, is to filter the individual specific regressors by means of cross section aggregates such that asymptotically (as $I \to \infty$) the differential effects of unobserved common factors are eliminated.

For the individual slope coefficients the CCE estimator is given by augmenting the OLS regression of y_{it} on $\boldsymbol{x_{it}}$ and $\boldsymbol{d_t}$ with the cross-section averages $\overline{\boldsymbol{z}}_{\boldsymbol{t}} = \frac{1}{I}\Sigma_{i=1}^{I} \boldsymbol{z_{it}}$ where

$$\boldsymbol{z_{it}} = \begin{bmatrix} y_{it} \\ \boldsymbol{x_{it}} \end{bmatrix}.$$

Although $\overline{\boldsymbol{y}}_t$ and ϵ_{it} are not independent (i.e. endogeneity bias), their correlation goes to zero as $I \to \infty$. Based upon the CCE estimator, Pesaran (2006) proposes two estimators for the means of the individual specific slope coefficients: the Common Correlated Effects Mean Group (CCEMG) estimator, a generalization of the estimator proposed by Pesaran and Smith (1995), and a generalization of the fixed effects estimator, the Common Correlated Effects Pooled (CCEP) estimator.

Considering the model in (1)-(2), the CCEP estimator allows for the possibility of cross-section dependence:

$$\widehat{\boldsymbol{\beta}}_P = \left(\sum_{i=1}^{I} \boldsymbol{X_i}' \boldsymbol{M} \boldsymbol{X_i} \right)^{-1} \sum_{i=1}^{I} \boldsymbol{X_i}' \boldsymbol{M} \boldsymbol{y_i} \tag{7}$$

[4] See Bai (2005), Pesaran (2006) on page 30.

[5] See Bai (2003).

$$\boldsymbol{M} = \boldsymbol{I_T} - \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'$$

where $\boldsymbol{X_i}$ is a $T \times k$ matrix of observed specific regressors for unit i, $\boldsymbol{y_i}$ is a $T \times 1$ vector of observed specific regressors for unit i. $\boldsymbol{M}$ is the orthogonal projection matrix with respect to $\boldsymbol{H} = (\overline{\boldsymbol{y}}, \overline{\boldsymbol{X}})$, $\boldsymbol{D}$, $\overline{\boldsymbol{y}}$ and $\overline{\boldsymbol{X}}$ being, respectively, the $(T \times I)$, $(T \times 1)$ and $(T \times k)$ matrices of observations on $\boldsymbol{d_t}$, $\overline{\boldsymbol{y}}_t$ and $\overline{\boldsymbol{x}}'_t$ where $\overline{\boldsymbol{y}}_t = \frac{1}{I}\sum_{i=1}^{I} y_{it}$ and $\overline{\boldsymbol{x}}'_t = \frac{1}{I}\sum_{i=1}^{I} \boldsymbol{x_{it}}$. Pesaran (2006), pag. 67, suggests to use $\widetilde{\boldsymbol{e}}_i = \boldsymbol{M}(\boldsymbol{y_i} - \boldsymbol{X_i}\widehat{\boldsymbol{\beta}}_P)$, the consistent estimates of the errors $e_{it} = y_{it} - \boldsymbol{\alpha}'\boldsymbol{d}_t - \boldsymbol{beta}'\boldsymbol{x}_{it}$, to obtain consistent estimates of the factors, $\widehat{\boldsymbol{f}}_t$. Last, the factor loadings can be easily estimated in the regression equation:

$$y_{it} = \boldsymbol{\alpha}'\boldsymbol{d_t} + \boldsymbol{\beta}'\boldsymbol{x_{it}} + \boldsymbol{\lambda_i}'\widehat{\boldsymbol{f}}_t + \zeta_{it} \tag{8}$$

However, the estimates of the unobserved common factors $\widehat{\boldsymbol{f}}_t$, obtained as linear combinations of the vectors $\hat{\boldsymbol{e}}_t$, are by construction orthogonal to $\overline{\boldsymbol{z}}_t$ In particular $\overline{\boldsymbol{y}}'\hat{\boldsymbol{f}} = 0$ where $\overline{\boldsymbol{y}}$ is the $(T \times 1)$ vector whose $t - th$ element is given by $\overline{y}_t = \frac{1}{I}\sum_{i=1}^{I} y_{it}$ and $\hat{f}$ is the $(T \times 1)$ vector of the estimated unobserved common factors. The previous relation implies:

$$\boldsymbol{y}'(\boldsymbol{\iota}_I \otimes \hat{\boldsymbol{f}}) = 0$$

where $\boldsymbol{y}$ is the $(IT \times 1)$ vector of observations over y_{it} and $\boldsymbol{\iota}_I$ is the unit vector of length I. Thus, if we estimate a familiar panel model as the fixed or random effects models, the estimated factors $\widehat{\boldsymbol{f_t}}$ in equation 8 is orthogonal to the dependent variable y_{it} bringing no gain in explaining it.[6]

3.1.2 Quasi-Maximum Likelihood Estimator (Bai, 2005).

Bai (2005) considers the Concentrated Least-Squares (CLS) estimation of the linear model (1). The CLS $\widehat{\boldsymbol{\beta}}_{CLS}$ estimator minimizes the following concentrated least-squares function:

$$CLS_{IT}(\boldsymbol{\beta}) = \min_{\boldsymbol{\Lambda},\boldsymbol{F}} \sum_{i=1}^{I} (\boldsymbol{y_i} - \boldsymbol{X_i}\boldsymbol{\beta} - \boldsymbol{F}\boldsymbol{\lambda_i})'(\boldsymbol{y_i} - \boldsymbol{X_i}\boldsymbol{\beta} - \boldsymbol{F}\boldsymbol{\lambda_i}) \tag{9}$$

where the function have been already minimized over $\boldsymbol{\Lambda}$ and $\boldsymbol{F}$, treated as parameters. $\boldsymbol{\Lambda}$ and $\boldsymbol{F}$ are subject to the following identification constraints: $\boldsymbol{F}'\boldsymbol{F}/T = \boldsymbol{I}_r$ and $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ being diagonal. Integrating out $\boldsymbol{\Lambda}$ one obtains:

$$CLS_{IT}(\boldsymbol{\beta}) = \min_{\boldsymbol{F}} \sum_{i=1}^{I} (\boldsymbol{y_i} - \boldsymbol{X_i}\boldsymbol{\beta})'\boldsymbol{M_F}(\boldsymbol{y_i} - \boldsymbol{X_i}\boldsymbol{\beta}) \tag{10}$$

where $\boldsymbol{M_F} = \boldsymbol{I}_T - \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'$. Given $\boldsymbol{F}$ the solution $\boldsymbol{\beta}$ of (10) is

$$\widehat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{I} (\boldsymbol{X_i}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{X_i})\right]^{-1} \sum_{i=1}^{I} (\boldsymbol{X_i}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{y_i}) \tag{11}$$

[6]Alternatively, $\widehat{\boldsymbol{f}}_t$ could be used as a regressor in a SURE-GLS model. However when the cross-section dimension I is bigger than the time series T dimension SURE-GLS is not feasible.

and given $\boldsymbol{\beta}$ the solution $\boldsymbol{F}$ of (10) is

$$\left[\sum_{i=1}^{I}(\boldsymbol{y_i} - \boldsymbol{X_i}\widehat{\boldsymbol{\beta}})(\boldsymbol{y_i} - \boldsymbol{X_i}\widehat{\boldsymbol{\beta}})'\right]\widehat{\boldsymbol{F}} = \widehat{\boldsymbol{F}}\boldsymbol{V}_{IT} \tag{12}$$

where $\widehat{\boldsymbol{F}}$ is equal to the first r eigenvectors associated with the first r largest eigenvalues of the above matrix in the brackets and $\boldsymbol{V}_{IT}$ is the corresponding diagonal matrix of eigenvalues.[7] Last, from the concentrated solution of (9), $\boldsymbol{\Lambda}(\boldsymbol{F}'\boldsymbol{F}) = \boldsymbol{Z}'\boldsymbol{F}$ where $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_I)$ and $\boldsymbol{Z}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$. Thus $\widehat{\boldsymbol{\Lambda}} = \frac{\boldsymbol{Z}'\boldsymbol{F}}{T}$ is expressed as function of $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{F}})$.

Under the assumptions that the ϵ_{it} are *iid* normal and if $\boldsymbol{x_{it}}$ are treated as fixed, the CLS estimator is the Maximum Likelihood estimator.

Because the number of $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$ grows with sample size I and T, both the $\boldsymbol{\lambda_i}$ and $\boldsymbol{f}_t$ are *incidental parameters* in the sense of Neyman and Scott (1948). As a consequence the usual results for the asymptotic properties of the MLE (or quasi-MLE) do not apply and the asymptotic properties of the CLS estimator need to be derived directly. (See Ahn, Lee, and Schmidt (2001), Bai (2005) and Moon and Weidner (2008)). Moreover, consistency for both $\boldsymbol{\lambda_i}$ and $\boldsymbol{f_t}$ can only be stated in terms of some average norm or for componentwise consistency (Bai and Ng (2002), Bai (2003), Bai (2005)).

Kiefer (1980) shows that the CLS estimator can be computed by the iterative scheme in (11) and (12). For a given value of $\boldsymbol{\beta}$, the estimator of $\boldsymbol{F}$ is the first r eigenvectors associated with the first r largest eigenvalues of $\sum_{i=1}^{I}(\boldsymbol{y_i} - \boldsymbol{X_i}\widehat{\boldsymbol{\beta}})(\boldsymbol{y_i} - \boldsymbol{X_i}\widehat{\boldsymbol{\beta}})'$. Conversely, for a given value of $\boldsymbol{F}$, the estimator of $\boldsymbol{\beta}$ is obtained by regressing $\boldsymbol{M_F y_i}$ on $\boldsymbol{M_F X_i}$

A starting value for $\boldsymbol{F}$ or $\boldsymbol{\beta}$ is needed. The two natural candidates are the principal components estimator for $\boldsymbol{F}$ (ignoring the regressors $\boldsymbol{X}_i$) and the simple least squares for $\boldsymbol{\beta}$ (ignoring the unobserved common effects), respectively. Bai (2005) proposes also the following iteration scheme which shows better convergence features especially for the case of time-invariant and common regressors included in $\boldsymbol{X}$. Given $\boldsymbol{F}$ and $\boldsymbol{\Lambda}$, compute

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{I}\boldsymbol{X_i}'\boldsymbol{X_i}\right)^{-1}\sum_{i=1}^{I}\boldsymbol{X_i}(\boldsymbol{Y_i} - \boldsymbol{F}\boldsymbol{\lambda}_i)$$

and given $\boldsymbol{\beta}$, compute $\boldsymbol{F}$ and $\boldsymbol{\Lambda}$ from the pure factor model $\boldsymbol{W_i} = \boldsymbol{F\lambda_i} + \epsilon_i$ with $\boldsymbol{W_i} = \boldsymbol{Y_i} - \boldsymbol{X_i\beta}$.

3.1.3 Two-step estimator (Coakley, Fuertes, and Smith, 2002).

Coakley, Fuertes, and Smith (2002) propose a two-step estimator based on principal components, when $\boldsymbol{\alpha} = \boldsymbol{0}$ in the model (1). They augment the regression of each dependent variable y_{it} on $\boldsymbol{x_{it}}$ with one or more principal components of the estimated OLS residuals

[7] Bai (2005) divides $[\sum_{i=1}^{I}(\boldsymbol{y_i} - \boldsymbol{X_i}\widehat{\boldsymbol{\beta}})(\boldsymbol{y_i} - \boldsymbol{X_i}\widehat{\boldsymbol{\beta}})']$ by IT to make have $\boldsymbol{V}_{IT}$ a proper limit. However the scaling does not affect $\widehat{\boldsymbol{F}}$. The concentrated objective function (10) is the same considered by Ahn, Lee, and Schmidt (2001) although they divide it by I instead of IT. Ahn, Lee, and Schmidt (2001) consider the case of a single unobserved factor for fixed T under the assumption that $\boldsymbol{X_i}$ are iid distributed across individuals.

$\widehat{e}_{it}$, for $i = 1, \ldots, I$ and $t = 1, \ldots, T$ obtained from a first stage OLS regression of y_{it} on $\boldsymbol{x_{it}}$ for each i. The second stage consists of estimating

$$y_{it} = \boldsymbol{\beta}' \boldsymbol{x_{it}} + \boldsymbol{\lambda}' \widehat{\boldsymbol{f}}_{\boldsymbol{t}} + \epsilon_{it} \tag{13}$$

where $\widehat{\boldsymbol{f}}_{\boldsymbol{t}}$ are the r largest Principal Components of the first-stage standardized residuals[8] where r is estimated by the Bai and Ng (2002) selection criteria.

Pesaran (2004) shows that this procedure leads to inconsistent estimation when the cross section mean of the included regressors, $\overline{\boldsymbol{x}}_t = \frac{1}{I}\sum_{i=1}^{I} \boldsymbol{x_{it}}$, and the unobserved factors are correlated. It is, in fact, not surprising to find inconsistency of the two-step estimator because both $\boldsymbol{\beta}$ and $\boldsymbol{f_t}$ are inconsistently estimated in the first step.

3.1.4 Two-step estimator: PCA augmented estimators (Kapetanios and Pesaran, 2005)

Kapetanios and Pesaran (2005) propose an alternative two-stage estimation method: in the first step principal components of all the economic variables in the panel data model (y_{it} and $\boldsymbol{x_{it}}$) are obtained as in Stock and Watson (2002), and in the second step the model is estimated augmenting the observed regressors with the estimated PCs:[9]

$$y_{it} = \boldsymbol{\alpha}' \boldsymbol{d_t} + \boldsymbol{\beta}' \boldsymbol{x_{it}} + \boldsymbol{\lambda_i}' \widehat{\boldsymbol{f}}_t + \epsilon_{it} \tag{14}$$

In this setup the estimated factors $\widehat{\boldsymbol{f}}_t$ are linear combinations of both the unobserved and the observed common factors $\boldsymbol{f_t}$ and $\boldsymbol{d_t}$. Therefore, $r + n$ rather than just r factors, must be extracted from $\boldsymbol{z_{it}} = (y_{it}, \boldsymbol{x_{it}})$, where the number of unobserved factors (r) is estimated by the Bai and Ng (2002) selection criteria. This can introduce some sampling uncertainty into the analysis as stressed by Kapetanios and Pesaran (2005) which show substantial small sample bias on the estimations when the number of factors to be included in the regression has to be estimated.

3.2 Estimators of the common observed components, $\boldsymbol{\alpha}$

To the best of our knowledge, only Bai (2005) explicitly considers the issue of identification and estimation of the common observed components when the errors have a factor structure. Ahn, Lee, and Schmidt (2001) allow for time-invariant regressors, although they do not consider the joint presence of common regressors, in the case of a single unobserved factor. In the following subsections we present two estimation methods which deal not only with the estimation of the individual specific components but also with the common observed ones.

[8]The estimated factor matrix $\widehat{\boldsymbol{F}} = (\widehat{\boldsymbol{f}}_1, \widehat{\boldsymbol{f}}_2 \ldots, \widehat{\boldsymbol{f}}_T)'$, is the $(T \times r)$ eigenvectors matrix corresponding to the r largest eigenvalues of the $(T \times T)$ matrix $\widehat{\boldsymbol{E}}\widehat{\boldsymbol{E}}'$ where $\widehat{\boldsymbol{E}} = (\widehat{\epsilon_1}, \widehat{\epsilon_2}, \ldots, \widehat{\epsilon_T})'$.

[9]Bai (2003) shows that as long as $\frac{\sqrt{t}}{I} \to 0$, the error in the estimated factor is negligible. Thus he justifies to augment the equation with the estimated factors.

3.2.1 Two-step estimator: CCEP+PCA (Castagnetti and Rossi, 2008)

When the analysis is not concerned only with the estimation of $\boldsymbol{\beta}$, the slope coefficients, but also with $\boldsymbol{\alpha}$, the coefficients of the observed common components, is important to rely on a *consistent* estimator of $\boldsymbol{\beta}$, which is obtained using suitable proxies for the unobservable factors. Based on these estimates is possible to compute *consistent* estimates of the errors $\boldsymbol{e}_{it}$, which can be used as observed data to obtain estimates of the unobserved factors, $\boldsymbol{f}_t$. In a previous work, we propose an estimation procedure which heavily relies on both Bai (2005) and Pesaran (2006) estimators. First, we estimate the individual slope coefficients by means of the Pesaran (2006) CCEP estimator.

1. we consistently estimate the slope parameter $\widehat{\boldsymbol{\beta}}$ by means of the CCEP estimator of equation (7), based on an estimate of $\boldsymbol{f_t}$ by means of cross-section averages, $\overline{\boldsymbol{z}}_t$, and $\boldsymbol{d}_t$.

2. for $i = 1, \ldots, I$ we estimate the residuals as:

$$\widehat{\boldsymbol{e}}_i = \overline{\boldsymbol{M}}_d(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_p) \tag{15}$$

   where $\overline{\boldsymbol{M}}_d$ is given by

$$\overline{\boldsymbol{M}}_d = \boldsymbol{I}_T - \boldsymbol{D}(\boldsymbol{D'D})^{-1}\boldsymbol{D'}$$

3. The unobserved common factors are estimated, up to a non-singular transformation (i.e. *rotation indeterminacy*), by the method of least squares. The estimator of $\boldsymbol{F}$ is equal to the first J eigenvectors associated with the first J largest eigenvalues of the matrix $\widehat{\boldsymbol{E}}\widehat{\boldsymbol{E}}'$ where $\widehat{\boldsymbol{E}}$ is the $(T \times I)$ matrix: $\widehat{\boldsymbol{E}} = (\widehat{\boldsymbol{e}}_1, \widehat{\boldsymbol{e}}_2, \ldots, \widehat{\boldsymbol{e}}_I)$.

   Under the assumption that $E[\boldsymbol{f_t d_t}'] = 0 \ \forall t$ we show in a previous work (Castagnetti and Rossi, 2008) that $\widehat{\boldsymbol{F}}$ is a consistent estimator (in average norm) for $\boldsymbol{F}$.

4. Finally, $\widehat{\boldsymbol{f_t}}$ are used as regressors in the model. Bai (2003) shows that as long as $\sqrt{T}/I \to 0$ the error in the estimated factor is negligible, and for large I, $\boldsymbol{f_t}$ can be treated as known.

   The two-step estimator of $\boldsymbol{\delta} = (\boldsymbol{\alpha'}, \boldsymbol{\beta'})'$ is given by:

$$\widehat{\boldsymbol{\delta}}_{2step} = \left[\sum_{i=1}^{I} \boldsymbol{Q_i}'\boldsymbol{M}_{(\widehat{\boldsymbol{F}})}\boldsymbol{Q_i}\right]^{-1} \sum_{i=1}^{I}(\boldsymbol{Q_i}'\boldsymbol{M}_{(\widehat{\boldsymbol{F}})}\boldsymbol{y_i}) \tag{16}$$

   where $\boldsymbol{Q_i} = (\boldsymbol{X_i}, \boldsymbol{D})$ and $\boldsymbol{M}_{(\widehat{\boldsymbol{F}})}$ is the orthogonal projection matrix with respect to $\widehat{\boldsymbol{F}}$.[10].

[10] Or any linear combination of them, i.e. $\widehat{\boldsymbol{F}}\boldsymbol{H}$, where $\boldsymbol{H}$ is an invertible matrix such that $\widehat{\boldsymbol{F}}$ is an estimator of $\boldsymbol{FH}$ and $\boldsymbol{H}^{-1}\widehat{\boldsymbol{\Lambda}}'$ is an estimator of $\boldsymbol{\Lambda'}$

3.2.2 Quasi-Maximum Likelihood Estimator (Bai, 2005)

Bai (2005) explicitly considers the case of observed common factors included in the regressors.[11] The conditions that guarantee both the identification as well as the consistent estimation of the parameters can be summarized as follows:

- neither $\boldsymbol{F}$ nor its rotation can contain the unit vector; the same for $\boldsymbol{\Lambda}$
- absence of multicollinearity between $\boldsymbol{F}$ and $\boldsymbol{D}$.

The first condition guarantees what Bai (2005) defines the presence of a *genuine* factor structure in the error terms. The second condition is a standard identification condition for the common components coefficients, $\boldsymbol{\alpha}$.

The estimation method is the same described in section 3.1.2, equations (11)-(12), where $\boldsymbol{X_{it}}' = (\boldsymbol{d_t}', \boldsymbol{x_{it}}')$ and $\boldsymbol{\beta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$.

3.3 Estimation of the number of factors

Unlike the Pesaran (2006) estimator, the implementation of all the other methods presented above require the determination of the number of factors to be included in the regression. This is usually done by means of the criteria advanced in Bai and Ng (2002). They formulate the problem of selecting the number of factors in approximate factor models as a model selection problem therefore by minimizing information criteria. This method is designed for data where the number of observations is large in both the cross-section (I) and the time series (T) dimensions. However this method could produce inconsistent estimators if either I or T is small. Simulation results reported in Bai and Ng (2002) indicate that the number of factors is not accurately estimated if I or T is less than 20. Ahn and Perez (2008) present a generalized method of moment (GMM) estimator of the number of factors which requires just one of the data dimensions (I or T) to be large.[12]

Both Coakley, Fuertes, and Smith (2002) and Kapetanios and Pesaran (2005) evaluate the impact of selecting the factors on the accuracy of the second-step estimation. Coakley, Fuertes, and Smith (2002) extract the factors from estimated disturbances rather than observed variables as considered by Bai and Ng (2002). They observe that the selecting information criteria of Bai and Ng (2002) are quite accurate. On the contrary, Kapetanios and Pesaran (2005) show substantial small sample bias on the estimations due to the need to selecting the number of factors to be included in the regression.

4 Monte Carlo Experiments

The purpose of this section is to compare the small sample properties of the estimators discussed in Section 3 when unobserved common factors are present. Each experiment involves 1,000 replications of $(I, T+T_0)$ observations where the first $T_0 = 50$ observations

[11] Bai (2005) considers also the presence of individual-invariant regressors.

[12] The method proposed by Ahn and Perez (2008) is designed for exact factor model but it can be easily used to estimate some approximate factor models.

are discarded for each time series to avoid dependence on the initial conditions (set equal to zero). We consider combinations of $T =$ and $I =$. Therefore we consider both the case of I much larger than T and of T much larger than I.

At each iteration we generate the following DGP:

$$y_{it} = \alpha_1 + \alpha_2 d_{2t} + \beta_1 x_{1it} + \beta_2 x_{2it} + \lambda_i f_t + \epsilon_{it} \tag{17}$$

$$x_{1it} = a_{11} + a_{21} d_{2t} + \lambda_1 f_t + v_{1it} \tag{18}$$

$$x_{2it} = a_{12} + a_{22} d_{2t} + \lambda_2 f_t + v_{2it} \tag{19}$$

for $i = 1, \ldots, I$, and $t = 1, \ldots, T$. This DGP considers only two individual specific components, x_{1it} and x_{2it}, two observed common factors, d_{1t} and d_{2t}, and one unobserved common factor f_t. $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) = (0.8, 0.5)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2) = (1, 3)$.

The parameters

$$A = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$$

and

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

are generated as $vec(A) \sim IIDN(\mathbf{0}, 0.5 \times \mathbf{I}_4)$, and $IIDN(\mathbf{0}, 0.5 \times \mathbf{I}_2)$ respectively, and are kept constant across replications. $\lambda_i = IIDN(1, 0.2)$. The common factors and the individual specific errors are generated as independent stationary AR(1) processes with zero means and unit variances:

$$\begin{aligned} d_{1t} &= 1 \\ d_{2t} &= \rho_d d_{2,t-1} + v_{dt}, \quad t = -49, \ldots 1, \ldots, T. \\ v_{dt} &\sim IIDN(0, 1 - \rho_d^2), \quad \rho_d = 0.5, \quad d_{2,-50} = 0 \end{aligned}$$

$$\begin{aligned} f_t &= \rho_f f_{t-1} + v_{ft}, \quad t = -49, \ldots 1, \ldots, T. \\ v_{ft} &\sim IIDN(0, 1 - \rho_f^2), \quad \rho_f = 0.5, \quad f_{-50} = 0 \end{aligned}$$

$$\begin{aligned} v_{jit} &= \rho_{v_{ij}} v_{ji,t-1} + \nu_{ijt}, \quad t = -49, \ldots 1, \ldots, T. \\ \nu_{ijt} &\sim IIDN(0, 1 - \rho_{v_{ij}}^2), \quad v_{ji,-50} = 0 \;\; j = 1, 2 \end{aligned}$$

and

$$\rho_{v_{ij}} \sim IIDU\,(0.05, 0.95)\,, \quad j = 1, 2.$$

The errors of y_{it} are generated as stationary AR(1) processes:

$$\begin{aligned} \epsilon_{it} &= \rho_{i\epsilon} \epsilon_{i,t-1} + \sigma_i (1 - \rho_{i\epsilon}^2)^{1/2} \zeta_{it} \quad \text{for} \quad i = 1, \ldots, I \\ \rho_{i\epsilon} &\sim IIDU\,(0.05, 0.95) \\ \sigma_i^2 &\sim IIDU\,(0.5, 1.5) \\ \zeta_{it} &\sim IIDN\,(0, 1) \end{aligned}$$

For each experiment we computed the CCEP and the Bai (2005) estimators as well as the infeasible estimator, assuming $\boldsymbol{f_t}$ is observable, and the *naive* estimator that excludes the factor. The infeasible estimator provides an upper bound to the unbiasedness and efficiency of the CCEP and the Bai (2005) estimators. The *naive* estimator illustrates the extent of bias and size distortions that can occur if the error cross section dependence is ignored.

Namely the infeasible estimator is given by:

$$\widehat{\boldsymbol{\beta}}_{inf} = \left[\sum_{i=1}^{I}(\boldsymbol{X_i}'\boldsymbol{M}_{(\boldsymbol{D},\boldsymbol{F})}\boldsymbol{X_i})\right]^{-1}\sum_{i=1}^{I}(\boldsymbol{X_i}'\boldsymbol{M}_{(\boldsymbol{D},\boldsymbol{F})}\boldsymbol{y_i})$$

while the *naive* estimator is given by:

$$\widehat{\boldsymbol{\beta}}_{naive} = \left[\sum_{i=1}^{I}(\boldsymbol{X_i}'\boldsymbol{M}_{(\boldsymbol{D})}\boldsymbol{X_i})\right]^{-1}\sum_{i=1}^{I}(\boldsymbol{X_i}'\boldsymbol{M}_{(\boldsymbol{D})}\boldsymbol{y_i})$$

The Bai (2005) estimator is computed by allowing up to 500 iterations for each simulation and by setting the tolerance coefficient equal to 0.0001.

4.1 Simulation results

- **finite sample properties of $\boldsymbol{\beta}$**

  First we compare the estimators of β, i.e. the individual specific components coefficients. Tables 1 and 2 report the bias and the root mean squared errors (RMSE) of the two estimators, respectively. Overall, the Pesaran Common Correlated Effects Pooled (CCEP) estimator bias is the closest to the one realized by the infeasible estimator. Bai estimator and 2step estimator show mixed results. For what concerns the efficiency, as measured by the root mean square error, the Pesaran estimator turns out to the best, while the Bai and the 2step estimators have very close performances.

  [Tables 1 and 2 here]

- **finite sample properties of $\boldsymbol{\alpha}$**

  We now compare the Bai (2005) and the Castagnetti and Rossi (2008) estimator $\widehat{\alpha}$ for the common observed components. The two-step estimation method of section 3.2.1 relies on a first-step *consistent* estimation of the errors $\boldsymbol{e_{it}}$, which can be used as observed data to obtain consistent (in average norm) estimates of the unobserved factors, $\boldsymbol{f_t}$.

  Like in Bai (2003) to evaluate the estimate of a transformation of f_t, $\widehat{f}_{t,2step}$, we compute the correlation coefficient between $\{f_{t,\hat{2}step}\}_{t=1}^{T}$ and $\{f_t\}_{t=1}^{T}$, for each Monte Carlo simulation. We compare these correlation coefficients with those obtain using the Bai (2005) estimator in equation (12). Table 3 below reports the average correlation coefficients for both estimation methods. The results suggest that both factor estimates are highly correlated with the unobserved factor. This seems to confirm the results in Bai (2003), obtained in a different context, that is as $\sqrt{T}/I \rightarrow 0$, the estimation error in the factor estimates is negligible. It is worth noticing that in many cases the iteration method of Bai (2005) does not converge. The last two columns of table 3 report the average number of iterations and the number of failure for each Monte Carlo simulation, respectively.

[Table 3 here]

Tables 4 and 5 report the bias and the RMSE of the the Bai (2005) and the two-step estimator of section 3.2.1 estimators of $\boldsymbol{\alpha}$, respectively. As before we also present estimation results for the *naive* as well as for the infeasible estimator. For what concerns the bias, when $I > T$ we observe a slightly superior performance of the two-step estimator with respect to the CLS estimator. For the root mean square error we observe a mixed situation. However, we should take into account that the CLS estimator is more computationally intensive than the two-step estimator and that the number of cases in which the iterative procedure fails in achieving the convergence is quite high.

[Tables 4 and 5 here]

- **finite sample properties of $\boldsymbol{\beta}$ when the factor loadings λ_i are correlated with the regressors**

  Bai (2005) suggests that the method proposed by Pesaran (2006) does not provide consistent estimates of $\boldsymbol{\beta}$ when $\boldsymbol{\lambda_i}$ is correlated with the regressors. Using the projection argument of Mundlak (1978), Bai (2005) suggests that additional regressors, the time-series averages $\overline{\boldsymbol{z}}_{\boldsymbol{i}} = \frac{1}{T}\Sigma_{t=1}^{T}\boldsymbol{z_{it}}$ where $\boldsymbol{z_{it}} = (y_{it}, \boldsymbol{x}'_{\boldsymbol{it}})'$ should also be added to achieve consistency. Appendix 6 shows that the general argument used by Pesaran (2006) to justify the CCEP estimator holds not only when $\boldsymbol{F_t}$ is correlated with the regressors but also when $\boldsymbol{\lambda_i}$ is correlated with the regressors.

  Here we investigate the small sample properties of the CCEP estimator when $\boldsymbol{\lambda_i}$ is correlated with $\boldsymbol{x_{it}}$. Following Bai (2005) we adopt the following DGP:

$$y_{it} = \alpha_1 d_{1t} + \alpha_2 d_{2t} + \beta_1 x_{1it} + \beta_2 x_{2it} + \lambda_i f_t + \epsilon_{it} \tag{20}$$

$$x_{1it} = a_1 + \lambda_i f_t + \lambda_i + f_t + v_{1it} \tag{21}$$

$$x_{2it} = a_2 + \lambda_i f_t + \lambda_i + f_t + v_{2it} \tag{22}$$

  for $i = 1, \ldots, I$, and $t = 1, \ldots, T$. This DGP considers only two individual specific components, x_{1it} and x_{2it}, two observed common factors, d_{1t} and d_{2t}, and one unobserved common factor f_t. ϵ_{it} is $IIDN(0, 2)$. $\boldsymbol{a} = (a_1, a_2) = (1, 1)$ $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) = (5, 4)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2) = (1, 3)$.

$$\begin{aligned} d_{1t} &= 1 \\ d_{2t} &= f_t + v_{dt} \end{aligned}$$

  v_{dt} is $IIDN(0, 1)$ independent of all other regressors. The variables λ_i, f_t, v_{jit} are all $IIDN(0, 1)$.

This DGP is identical to Bai's (2005) DGP for the case of common regressors. Tables 6 and 7 report the Monte Carlo results in terms of bias and RMSE for $\boldsymbol{\beta}$.

[Tables 6 and 7 here]

The Monte Carlo results in table 6 and 7 show that Pesaran (2006) estimator has better bias and root mean square error performances than the Bai estimator when the $\boldsymbol{x_{it}}$ are simultaneously correlated with λ_i and f_t.

5 Conclusions

In this paper we review the estimation techniques adopted in panel data models with individual and common factors. In particular we consider the presence of unobserved and observed common factors. We present also a new approach to the estimation of individual-specific components along with the estimation of the common factors coefficients which is based on a two-step estimation procedure. The finite sample properties are investigated by means of Monte Carlo simulations, under different data-generating processes. The results show that the CCEP estimator by Pesaran (2006) has remarkable properties under different DGPs. Furthermore, the two-step estimator by Castagnetti and Rossi (2008) shows good finite sample properties when compared to the iterative CLS estimator by Bai (2005), which is computationally more demanding and less accurate when we consider the fact that it fails to achieve convergence in a relevant number of cases.

6 Appendix

When the factor loadings are correlated with the regressors, model (1-2) becomes:

$$y_{it} = \boldsymbol{\beta}'\boldsymbol{x_{it}} + \boldsymbol{\lambda_i}'\boldsymbol{f_t} + \epsilon_{it} \tag{23}$$

$$\boldsymbol{x_{it}} = \boldsymbol{\Lambda_i}'\boldsymbol{f_t} + \boldsymbol{A}\boldsymbol{\lambda_i} + \boldsymbol{v_{it}} \tag{24}$$

where $\boldsymbol{\Lambda_i}$ is a $(r \times K)$ factor loading matrix with fixed components, $\boldsymbol{\Lambda}$, and $\boldsymbol{A}$ is a $(K \times r)$ matrix of parameters. Combining (23-24) we have the system of equations:

$$\boldsymbol{z_{it}} = \begin{pmatrix} y_{it} \\ \boldsymbol{x_{it}} \end{pmatrix} = \boldsymbol{C_i}'\boldsymbol{f_t} + \boldsymbol{B}\boldsymbol{\lambda_i} + \boldsymbol{u_{it}} \tag{25}$$

where

$$\boldsymbol{C_i} = \begin{pmatrix} \boldsymbol{\lambda_i} & \boldsymbol{\Lambda_i} \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{\beta} & \boldsymbol{I_K} \end{pmatrix}$$

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{\beta}' \\ \boldsymbol{I_K} \end{pmatrix} \boldsymbol{A}$$

and

$$\boldsymbol{u_{it}} = \begin{pmatrix} \boldsymbol{\beta}'\boldsymbol{v_{it}} + \epsilon_{it} \\ \boldsymbol{v_{it}} \end{pmatrix} \tag{26}$$

Consider the cross-section averages of the equations in (25):

$$\bar{\boldsymbol{z}}_{\boldsymbol{t}} = \bar{\boldsymbol{C}}' \boldsymbol{f}_{\boldsymbol{t}} + \boldsymbol{B}\bar{\boldsymbol{\lambda}} + \bar{\boldsymbol{u}}_{\boldsymbol{t}} \tag{27}$$

suppose that $Rank(\bar{\boldsymbol{C}}) = r \leq K+1$ for all I, then we have:

$$\boldsymbol{f}_{\boldsymbol{t}} = (\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}')^{-1}\bar{\boldsymbol{C}}(\bar{\boldsymbol{z}}_{\boldsymbol{t}} - \boldsymbol{B}\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{u}}_{\boldsymbol{t}}) \tag{28}$$

Pesaran (2006), (Lemma 1) shows that $\bar{\boldsymbol{u}}_{\boldsymbol{t}} \stackrel{q.m.}{\to} 0$ as $I \to \infty$ for every t. Moreover, as in Pesaran (2006), $\bar{\boldsymbol{\lambda}} \stackrel{p}{\to} \boldsymbol{\lambda}$ as $I \to \infty$ where $\boldsymbol{\lambda} = E(\boldsymbol{\lambda}_{\boldsymbol{i}})$ and $\bar{\boldsymbol{C}} \stackrel{p}{\to} \boldsymbol{C}$ where

$$\boldsymbol{C} = \begin{pmatrix} E(\boldsymbol{\lambda}_{\boldsymbol{i}}) & E(\boldsymbol{\Lambda}_{\boldsymbol{i}}) \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta} & \boldsymbol{I}_{\boldsymbol{K}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\lambda} & \boldsymbol{\Lambda} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta} & \boldsymbol{I}_{\boldsymbol{K}} \end{pmatrix}$$

Therefore, from (28) we obtain:

$$\boldsymbol{f}_{\boldsymbol{t}} - (\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}')^{-1}\bar{\boldsymbol{C}}(\bar{\boldsymbol{z}}_{\boldsymbol{t}} - \boldsymbol{B}\boldsymbol{\lambda}) \stackrel{p}{\to} 0 \tag{29}$$

Therefore the argument of Pesaran (2006), pag. 976 of using $\bar{\boldsymbol{z}}_{\boldsymbol{t}}$ as observable proxies for $\boldsymbol{f}_{\boldsymbol{t}}$ still holds also when the factor loadings $\boldsymbol{\lambda}_{\boldsymbol{i}}$ are correlated with the regressors.

References

AHN, S., Y. LEE, AND P. SCHMIDT (2001): "GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects," *Journal of Econometrics*, 101, 219–255.

——— (2006): "Panel Data Models with Multiple Time-Varying Individual Effects," mimeo.

AHN, S., AND M. PEREZ (2008): "GMM Estimation of the Number of Latent Factors," mimeo.

BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71(1), 135–171.

——— (2005): "Panel Data Models with Interactive Fixed Effects," mimeo.

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

——— (2006a): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions.," *Econometrica*, 74(4), 1133–1150.

——— (2006b): "Evaluating Latent and Observed Factors in Macroeconomics and Finance.," *Journal of Econometrics*, 113(1), 507–537.

CASTAGNETTI, C., AND E. ROSSI (2008): "Euro Corporates Bonds Risk Factors," Mimeo.

COAKLEY, J., A. FUERTES, AND R. SMITH (2002): "A Principal Component Approach to Cross-Section Dependence in Panels," Discussion paper, Birkbeck College.

COCHRANE, J., AND M. PIAZZESI (2005): "Bond Risk Premia," *The American Economic Review*, 95, 138–160.

FAMA, E., AND K. FRENCH (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33(1), 3–56.

KAO, C., L. TRAPANI, AND G. URGA (2008): "The Asymptotics for Panel Models with Common Shocks," mimeo.

KAPETANIOS, G., AND M. PESARAN (2005): *Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns*vol. forthcoming in The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis. Garry Phillips and Elias Tzavalis, Cambridge, cambridge university press edn.

KIEFER, N. (1980): "A Time Series-Cross Section Model with Fixed Effects with an Intertemporal Factor Structure," Unpuplished manuscript, Cornell University.

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

LUDVIGSON, S., AND S. NG (2008): "Macro Factors in Bond Risk Premia," *The Review of Financial Studies,* forthcoming.

MARDIA, K., J. KENT, AND J. BIBBY (1979): *Multivariate Analysis*. Academic Press.

MOON, H. R., AND M. WEIDNER (2008): "Asymptotic Analysis of the quasi-MLE of Panel Regression Models with Interactive Fixed Effects," Department of Economics, UCS.

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

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

PESARAN, M. (2004): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," mimeo, Cambridge University.

——— (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012.

PESARAN, M., AND R. SMITH (1995): "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics*, 68(1), 79–113.

STOCK, J., AND M. WATSON (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147–162.

WANSBEEK, T., AND E. MEIJER (2000): *Measurement Error and Latent Variables in Economics*. North-Holland.

Table 1: **BIAS of β estimators**

		CLS		***CCEP***		***infeasible***		***2step***		***naive***	
I	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.0589	-0.0108	-0.0130	0.0031	0.0024	0.0017	0.0063	-0.0463	0.2749	-0.1621
10	30	0.0011	-0.0157	-0.0016	-0.0082	-0.0011	-0.0007	0.0298	-0.0040	0.4868	-0.5624
10	100	0.0009	-0.0009	0.0012	0.0015	-0.0012	0.0025	-0.0100	0.0031	-0.1788	-0.7481
30	10	-0.0201	0.0492	-0.0015	0.0088	-0.0025	0.0003	0.0299	0.0117	0.0345	0.3747
30	30	-0.0179	-0.0012	0.0029	0.0017	-0.0004	0.0025	-0.0041	0.0046	-0.3070	0.5571
30	100	-0.0014	-0.0213	-0.0012	-0.0008	-0.0012	-0.0011	-0.0026	-0.0109	-0.0902	-0.4141
100	10	-0.0210	0.0072	-0.0016	0.0009	-0.0017	0.0027	0.0003	0.0050	-0.0345	0.4657
100	30	0.0150	-0.0172	-0.0008	-0.0009	0.0006	-0.0015	0.0025	-0.0037	0.2270	-0.3850
100	100	-0.0005	-0.0001	-0.0004	0.0005	0.0002	0.0002	-0.0019	0.0011	-0.4112	-0.0598
200	10	0.0097	-0.0032	0.0004	0.0003	0.0021	-0.0002	0.0021	0.0064	-0.3819	-0.3717
200	30	-0.0021	-0.0103	-0.0002	0.0002	-0.0004	0.0003	0.0017	0.0245	-0.0366	0.5980
200	100	0.0007	0.0000	0.0000	0.0005	0.0002	0.0007	0.0013	0.0039	0.2251	0.4504

Table 2: **RMSE of β estimators**

		CLS		***CCEP***		***infeasible***		***2step***		***naive***	
I	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.1749	0.1192	0.1459	0.1174	0.1428	0.1172	0.1615	0.1365	0.2885	0.2036
10	30	0.0892	0.1018	0.0816	0.0971	0.0874	0.0986	0.0975	0.1032	0.4942	0.5701
10	100	0.0505	0.0663	0.0502	0.0666	0.0478	0.0633	0.0513	0.0656	0.1848	0.7489
30	10	0.0780	0.0906	0.0904	0.0858	0.0844	0.0866	0.0973	0.0929	0.0945	0.3801
30	30	0.0672	0.0591	0.0589	0.0494	0.0687	0.0570	0.0697	0.0589	0.3136	0.5579
30	100	0.0403	0.0910	0.0339	0.0296	0.0351	0.0304	0.0356	0.0327	0.0962	0.4143
100	10	0.0449	0.0397	0.0472	0.0434	0.0498	0.0435	0.0522	0.0449	0.0602	0.4670
100	30	0.0325	0.0333	0.0285	0.0281	0.0325	0.0316	0.0328	0.0320	0.2289	0.3857
100	100	0.0174	0.0168	0.0169	0.0166	0.0173	0.0167	0.0174	0.0169	0.4112	0.0620
200	10	0.0317	0.0328	0.0305	0.0312	0.0337	0.0340	0.0338	0.0345	0.3829	0.3728
200	30	0.0209	0.0238	0.0200	0.0209	0.0213	0.0221	0.0225	0.0339	0.0428	0.5981
200	100	0.0114	0.0116	0.0108	0.0109	0.0114	0.0116	0.0115	0.0123	0.2254	0.4505

Table 3: **Average correlation coefficients between $\{\hat{f}_t\}_{t=1}^T$ and $\{f_t\}_{t=1}^T$.**

I	T	$\{\hat{f}_t\}_{ccep}$	$\{\hat{f}_t\}_{qmle}$	iter	fail
10	10	0.7999	0.9251	127.161	368
10	30	0.9632	0.9808	107.648	187
10	100	0.9694	0.9743	81.453	52
30	10	0.8583	0.7975	194.651	931
30	30	0.9621	0.9873	193.881	898
30	100	0.9408	0.9654	148.134	399
100	10	0.9318	0.9956	199.284	988
100	30	0.9555	0.9920	197.297	948
100	100	0.9962	0.9960	53.425	0
200	10	0.9926	0.9708	184.396	796
200	30	0.9980	0.9476	118.122	1
200	100	0.9967	0.9968	58.621	0

Iter and fail indicate the average number of iteration and the number of failure of the iterative process for the qmle estimation method of Bai (2005), respectively.

Table 4: **BIAS of α estimators**

		CLS		***infeasible***		***2step***		***naive***	
I	T	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2
10	10	-0.3800	0.0135	-0.0082	-0.0002	0.5099	0.0564	-0.0441	0.1844
10	30	0.3163	-0.0359	0.0008	-0.0019	-0.4956	0.0721	-0.6562	0.0321
10	100	-0.0761	-0.0201	0.0032	-0.0009	0.2061	0.0547	0.3242	0.4322
30	10	0.5592	0.0125	0.0009	0.0008	-0.5102	-0.4649	-0.5183	-0.4575
30	30	0.8892	-0.1330	-0.0017	0.0013	-0.4511	-0.1406	0.1399	-0.0052
30	100	0.2176	0.0692	0.0006	0.0013	-0.1546	-0.2050	0.5849	-0.1189
100	10	-1.3010	-0.3470	0.0002	-0.0042	0.3402	-0.0215	0.0462	-0.2003
100	30	1.0918	0.1580	0.0016	-0.0005	-0.0793	0.0228	0.2678	-0.1667
100	100	0.0353	0.0108	0.0006	-0.0007	-0.0924	-0.0425	0.0038	0.1640
200	10	0.5647	0.0454	0.0019	-0.0010	-0.2575	0.3264	-0.4850	0.1927
200	30	-0.4202	0.0204	-0.0001	0.0003	0.6019	-0.3455	0.2108	0.2318
200	100	-0.1268	0.0164	0.0005	-0.0003	0.2605	0.0212	0.7366	-0.4138

Table 5: **RMSE of α estimators**

		CLS		***infeasible***		***2step***		***naive***	
I	T	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2
10	10	0.9248	0.3412	0.2891	0.1831	0.6562	0.1913	0.2980	0.2387
10	30	0.9111	0.2602	0.1572	0.0827	0.5183	0.1135	0.6736	0.0890
10	100	0.4882	0.1562	0.0989	0.0595	0.2279	0.0820	0.3381	0.4357
30	10	1.8646	0.4592	0.1851	0.0789	0.5424	0.4697	0.5474	0.4623
30	30	1.7113	0.3374	0.0888	0.0648	0.4621	0.1562	0.1650	0.0649
30	100	0.6986	0.1780	0.0887	0.0387	0.1812	0.2090	0.5907	0.1252
100	10	1.8175	0.4661	0.0634	0.0485	0.3467	0.0540	0.0785	0.2058
100	30	1.5220	0.2366	0.0520	0.0477	0.0955	0.0541	0.2722	0.1729
100	100	0.0956	0.0510	0.0346	0.0379	0.0987	0.0572	0.0347	0.1681
200	10	1.3146	0.1603	0.0486	0.0309	0.2615	0.3277	0.4869	0.1949
200	30	0.4394	0.0528	0.0329	0.0233	0.6030	0.3469	0.2133	0.2328
200	100	0.1516	0.0327	0.0273	0.0163	0.2618	0.0269	0.7370	0.4141

Table 6: **BIAS of β estimators**

		CLS		***CCEP***		***infeasible***		***naive***	
I	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.2358	0.2398	-0.0121	-0.0237	0.0001	-0.0028	0.2914	0.2913
10	30	0.2022	0.2034	0.0054	0.0007	-0.0013	-0.0021	0.3011	0.2952
10	100	0.1301	0.1254	0.0080	0.0024	0.0026	-0.0022	0.3054	0.3018
30	10	0.1766	0.1807	0.0095	0.0113	-0.0022	0.0029	0.2856	0.2887
30	30	0.0732	0.0783	0.0009	0.0047	-0.0028	0.0028	0.2980	0.3036
30	100	0.0115	0.0108	0.0040	0.0023	0.0017	0.0011	0.3044	0.3034
100	10	0.0632	0.0630	0.0028	0.0029	0.0009	0.0008	0.2897	0.2895
100	30	0.0067	0.0061	0.0030	0.0018	0.0006	-0.0002	0.3003	0.2998
100	100	0.0018	0.0016	0.0004	0.0000	0.0000	-0.0003	0.3037	0.3022
200	10	0.0260	0.0225	0.0050	0.0026	0.0023	-0.0007	0.2905	0.2879
200	30	0.0028	0.0033	-0.0007	0.0002	-0.0007	-0.0002	0.2999	0.3001
200	100	0.0017	0.0011	0.0000	-0.0006	0.0002	-0.0004	0.3035	0.3032

Table 7: **RMSE of β estimators**

		CLS		***CCEP***		***infeasible***		***naive***	
I	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.3264	0.3371	0.3173	0.3158	0.1802	0.1821	0.3345	0.3363
10	30	0.2719	0.2727	0.1356	0.1377	0.1008	0.0999	0.3169	0.3110
10	100	0.2077	0.2086	0.0686	0.0689	0.0541	0.0528	0.3101	0.3064
30	10	0.2493	0.2485	0.1695	0.1690	0.1004	0.1012	0.3012	0.3043
30	30	0.1594	0.1654	0.0738	0.0719	0.0551	0.0551	0.3028	0.3084
30	100	0.0591	0.0584	0.0394	0.0395	0.0309	0.0316	0.3061	0.3049
100	10	0.1415	0.1396	0.0940	0.0894	0.0561	0.0533	0.2952	0.2952
100	30	0.0437	0.0422	0.0415	0.0392	0.0301	0.0289	0.3019	0.3014
100	100	0.0205	0.0208	0.0204	0.0208	0.0163	0.0163	0.3042	0.3027
200	10	0.0780	0.0783	0.0609	0.0609	0.0375	0.0378	0.2941	0.2914
200	30	0.0288	0.0275	0.0284	0.0277	0.0215	0.0212	0.3009	0.3010
200	100	0.0144	0.0145	0.0145	0.0144	0.0115	0.0118	0.3038	0.3035