ISSN: 2281-1346

Department of Economics and Management

DEM Working Paper Series

A Two-Stage Estimator for Heterogeneous Panel Models with Common Factors

Carolina Castagnetti
(Università di Pavia)

Eduardo Rossi
(Università di Pavia)

Lorenzo Trapani
(City University, London)

 # 66 (01-14)

Via San Felice, 5
I-27100 Pavia
http://epmq.unipv.eu/site/home.html

Janauary 2014

A Two-Stage Estimator for Heterogeneous Panel Models with Common Factors*

Carolina Castagnetti[1], Eduardo Rossi[†1], and Lorenzo Trapani[2]

[1] *Dipartimento di Scienze Economiche e Aziendali, Università di Pavia, Italy*

[2] *Faculty of Finance,Cass Business School, City University, London (UK)*

January 31, 2014

Abstract

This paper considers estimation in a stationary heterogeneous panel model where common unknown factors are present. A two-stage estimator is proposed. This estimator is based on the CCE estimator (Pesaran, 2006) in the first stage and on a similar approach to the Interactive Effect estimator (Bai, 2009) in the second stage. The asymptotic properties of this estimator are provided alongside of the comparative finite-sample properties of a range of estimators by means of Monte Carlo experiments.

Keywords: Large panels; Factor error structure; Principal components; Common regressors; Cross-section dependence.

JEL - Classification: C33, C38

*We wish to thank the participants to the 2013 PRIN MISURA workshop (University of Venice, September 2013); to the 19th International Panel Data Conference (Cass Business School, London, July 2013). Special thanks to Luca Fanelli for providing us with valuable comments. Carolina Castagnetti and Eduardo Rossi acknowledge financial support from PRIN MISURA. The usual disclaimer applies.

†Correspondence to: Dipartimento di Scienze Economiche e Aziendali, Via San Felice 5, 27100 Pavia. E-mail: erossi@eco.unipv.it

1 Introduction

Consider the model

$$y_{it} = \beta_i' x_{it} + \gamma_i' f_t + \epsilon_{it} \qquad i = 1, \ldots, n \quad t = 1, \ldots, T, \tag{1}$$

where $i = 1, \ldots, n$ and $t = 1, \ldots, T$; x_{it} is a $k \times 1$ vector of individual-specific components, and f_t is the $r \times 1$ vector of unobserved factors. We further allow the x_{it}s to be cross sectionally correlated by being linearly dependent on a set of common unobserved factor

$$x_{it} = A_i f_t + u_{it}; \tag{2}$$

A_i is $k \times r$ is a factor loading matrix with nonrandom components and u_{it} are the specific components of x_{it} distributed independently of the common effects and across i. The main advantage of model (1)-(2) is its generality: the model considers the presence of cross sectional correlation in the y_{it}s; it does not rule out the possibility that the idiosyncratic regressors x_{it} are also strongly cross correlated; and it also allows for nontrivial correlation between the observable regressors x_{it} and the unobserved ones f_t. By virtue of its generality, model (1)-(2) nests, as pointed out in Pesaran (2006), several popular specifications.

On account of the huge application potential of (1)-(2), the literature on panel data has focused on developing the inferential theory for the slope coefficients β_i in presence of cross sectional dependence arising from a common factor structure. Several contributions have pointed out that slope estimation may be inconsistent when such common factors are correlated with the other regressors - see e.g. Pesaran (2006). Pesaran (2006) suggests a new approach, based on running a regression augmented with the cross sectional averages of dependent and independent variables, known as the Common Correlated Coefficients (CCE) estimator. The set up in Pesaran (2006) is very general and his results can also be extended to the case of non stationary common factors (Kapetanios, Pesaran, and Yamagata, 2011) and also hold in presence of spatial correlation in the error term (Pesaran and Tosetti, 2011). In another seminal contribution, Bai (2009) proposes an alternative estimator, called Interactive Fixed Effects (IFE), which combines standard OLS with the Principal Components estimation of the unobservable common factors. Whilst the contribution by Bai (2009) assumes homogeneous slopes, a recent contribution by Song (2013) extends the IFE estimator to the case of heterogeneous panels with cross-section dependence, extending the estimation theory to the case of dynamic panels. A common feature to the contributions by Pesaran (2006), Bai (2009) and Song (2013) is that they deal mainly with the estimation of the slope parameters, and inference on the unobserved common factors and their loadings is not fully developed.

There are, however, examples where the parameters of interest are not only the slope coefficients but also the common factors f_t and their loadings λ_i. For example, several contributions have studied the possibility of augmenting a standard regression model with the estimates of common factors from a large set of data (see, for example, Kapetanios and Pesaran, 2007, Ludvigson and Ng, 2007 and Ludvigson and Ng, 2009). Estimated factors may also be used e.g. as inputs for diffusion index forecasting (see Stock and Watson, 2005) and as regressors in factor augmented VARs (see Bernanke, Boivin, and Eliasz, 2005). Castagnetti, Rossi, and Trapani (2012) develop an inferential theory for the unobserved common factors and loadings.

In this paper, we build on (1)-(2), and extend the existing inferential theory on the average slope - say $\beta = E(\beta_i)$, in a random coefficient framework. Specifically, we propose a two-stage estimator of the average slope coefficient, based on the CCE estimator in the first stage. Such estimator can be thought of as the IFE estimator with only one iteration, and thus it complements the results in Bai (2009) and Song (2013) by proposing a simpler version of the estimators defined therein. Heuristically, the two-stage approach should combine the efficiency that can be naturally expected from an iterative procedure such as the IFE, with a greater computational simplicity and no risk of the iterative procedure to fail to achieve convergence. The small sample properties of the proposed estimator, analysed through synthetic data, show that the proposed estimator has good finite sample properties. The Monte Carlo experiments show that the two-stage estimator is, in finite samples, less biased than IFE and more efficient. The two-stage procedure, which represents a simple alternative to the IFE, offers the best trade-off between efficiency and computational reliability.

The remainder of the paper is organized as follows. Section 2 discusses the model and the main assumptions. Section 2illustrates the two-stage estimators, and reports the asymptotic results. Section 3 reports the details of the Monte Carlo experiments as well as the main results from the simulations. Section 4 concludes. Preliminary Lemmas are in Appendix A, and the proof of the main result is in Appendix B.

NOTATION. We use "$\longrightarrow$" to denote the ordinary limit; "$\xrightarrow{d}$" and "$\xrightarrow{p}$" to denote convergence in distribution and in probability respectively, and "a.s." for "almost surely". The Frobenius norm of a matrix A is denoted as $\|A\| = \sqrt{tr\,(A'A)}$, where $tr\,(A)$ denotes the trace of A. Finite constants that do not depend on the sample size are denoted as M, M', M'', etc... Other notation is defined throughout the paper and in Appendix.

2 The two stage estimator

In this section, we define the two-stage estimator of the average slope, and we derive its asymptotics properties.

Recall model (1)-(2):

$$\begin{aligned} y_{it} &= \beta_i' x_{it} + \gamma_i' f_t + \epsilon_{it}, \\ x_{it} &= A_i f_t + u_{it}, \end{aligned}$$

and its matrix form:

$$\begin{aligned} y_i &= X_i \beta_i + F \gamma_i + \epsilon_i, \\ X_i &= F A_i + u_i, \end{aligned}$$

with $y_i = [y_{i1}, \ldots, y_{iT}]'$, $X_i = [x_{i1}| \ldots |x_{iT}]'$, $F = [f_1| \ldots |f_T]'$ and $\epsilon_i = [\epsilon_{i1}, \ldots, \epsilon_{iT}]'$. In order to estimate the individual slopes β_i, the CCE estimator proposed in Pesaran (2006) is given by augmenting the OLS regression of y_{it} on x_{it} with the cross-section averages $\bar{z}_t = \frac{1}{n}\Sigma_{i=1}^{n} z_{it}$, with $z_{it} = [y_{it}, x_{it}']'$. Based on this intuition, the CCE estimator is defined as

$$\tilde{\beta}_i = (X_i' \bar{M} X_i)^{-1} X_i' \bar{M} y_i \tag{3}$$

with $\bar{M} = I_T - \bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'$, and $\bar{Z}$ is the $T \times (k+1)$ matrix of observations on $\bar{z}_t$. 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. The CCEMG estimator is a simple average of the individual CCE estimators, $\tilde{\beta}_i$ of β_i:

$$\tilde{\beta}_{CCEMG} = \frac{1}{n} \sum_{i=1}^{n} \tilde{\beta}_i \tag{4}$$

Assuming $\beta_i = \beta$, the CCEP estimator, which allows for the possibility of cross-section dependence, in its simplest form is given by:

$$\tilde{\beta}_{CCEP} = (\sum_{i=1}^{n} X_i' \bar{M} X_i)^{-1} \sum_{i=1}^{n} X_i' \bar{M} y_i \tag{5}$$

The two-stage estimation procedure of the average slope $\beta = E(\beta_i)$ is based on the following two steps:

Step 1 Obtain an estimate of the common factors f_t in (1)-(2):

1.(a) estimate the β_is using the CCE estimator defined in (3), and compute the residuals $\tilde{v}_i = y_i - X_i \tilde{\beta}_i$;

1.(b) Apply the Principal Component estimator (henceforth, PC) to $\tilde{v}_i$, obtaining $\hat{F}$ with the restriction $\hat{F}'\hat{F} = TI_r$.

Step 2 Re-estimate the slopes β_i and compute their average:

2.(a) Apply OLS to

$$y_{it} = \beta_i' x_{it} + \gamma_i' \hat{f}_t + \epsilon_{it}, \tag{6}$$

obtaining

$$\hat{\beta}_i = \left(X_i' \hat{M}_F X_i \right)^{-1} \left(X_i' \hat{M}_F y_i \right), \tag{7}$$

with $\hat{M}_F = I_T - \hat{F}\hat{F}'/T$;

2.(b) define the Augmented Mean Group (AMG hereafter) estimator of β

$$\hat{\beta}^{AMG} = \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_i. \tag{8}$$

In Step 1.(b), it is worth noting that the number of unobserved factors, r, can be assumed to be fixed but it is unknown, and it needs to be estimated as well. Indeed, given a consistent estimator of the β_is such as e.g. the CCE estimator, the residuals $\tilde{v}_i$ have a pure factor model structure, viz. $\tilde{v}_i = {\gamma_i}' f_t + \epsilon_{it} + (\beta_i - \tilde{\beta}_i)' x_{it}$. Bai (2009) and Pesaran (2006, p.30) show that the error component $(\beta_i - \tilde{\beta}_i)' x_{it}$ does not affect the determination of the number of common factors, which can be therefore estimated using e.g. the information criteria developed by Bai and Ng (2002). Similarly, it is well-known that common factors and loadings are not separately identified, and therefore can be estimated only up to a rotation. As far as our setup is concerned, knowing a rotation of the common factors and loadings is as good as knowing the true factors and loadings.

Turning to Step 2, note that in Step 2.(a), on account of (8), we have

$$\hat{\beta}_i - \beta_i = \left(\frac{X_i' \hat{M}_F X_i}{T} \right)^{-1} \left[\frac{X_i' \hat{M}_F F}{T} \gamma_i + \frac{X_i' \hat{M}_F \epsilon_i}{T} \right];$$

this can be compared with the IFE estimator in Song (2013). As a consequence, in Step 2.(b) the AMG estimator can be expressed as

$$\hat{\beta}^{AMG} - \beta = \frac{1}{n} \sum_{i=1}^{n} \nu_i + \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_i' \hat{M}_F X_i}{T} \right)^{-1} \left[\frac{X_i' \hat{M}_F F}{T} \gamma_i + \frac{X_i' \hat{M}_F \epsilon_i}{T} \right]. \tag{9}$$

Finally, in addition to $\hat{\beta}^{AMG}$ we define the Augmented Pooled (AP) estimator

$$\hat{\beta}^{AP} = \left(\frac{1}{n} \sum_{i=1}^{n} X_i' \hat{M}_F X_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} X_i' \hat{M}_F y_i \right). \tag{10}$$

In order to study the asymptotic properties of $\hat{\beta}^{AMG}$ and of $\hat{\beta}^{AP}$, consider the following assumptions.

Assumption 1. *[error terms: serial and cross sectional dependence] (i)* $E(\epsilon_{it}) = 0$ and $E|\epsilon_{it}|^{12} < \infty$; *(ii)* (a) $\sum_{t=1}^{T} |E(\epsilon_{it}\epsilon_{is})| \leq M$ for all i and s, (b) $\sum_{i=1}^{n} \sum_{j=1}^{n} |E(\epsilon_{it}\epsilon_{js})| \leq Mn$ for all t and s, (c) $\sum_{t=1}^{T} \sum_{s=1}^{T} |E(\epsilon_{it}\epsilon_{is})| \leq MT$ for all i, (d) $\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} |E(\epsilon_{it}\epsilon_{js})| \leq M(nT)$; *(iii)* (a) $E\left|(nT)^{-1/2} \sum_{i=1}^{n} \sum_{t=1}^{T} \epsilon_{it}\right|^2 \leq M$, (b) $\sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{v=1}^{T} \sum_{u=1}^{T} |E(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv})| \leq MT^2$, (c) $\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{u=1}^{T} |E(\epsilon_{it}\epsilon_{is}\epsilon_{ju}\epsilon_{js})| \leq M(nT)$ for all u, (d) $\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} |E(\epsilon_{it}\epsilon_{kt}\epsilon_{js}\epsilon_{ks})| \leq M(nT)$ for all k; *(iv)* (a) $E\left|\sum_{t=1}^{T} \epsilon_{it}\right|^r \leq ME\left|\sum_{t=1}^{T} \epsilon_{it}^2\right|^{r/2}$ for all i, $r < 12$, (b) $E\left|\sum_{i=1}^{n} \epsilon_{it}\right|^r \leq ME\left|\sum_{i=1}^{n} \epsilon_{it}^2\right|^{r/2}$ for all t, $r < 12$.

Assumption 2. *[regressors and common factors] (i)* $E\|\epsilon_{it}^x\|^{12} < \infty$ and $E\|f_t\|^{12} < \infty$; *(ii)* $T^{-1} \sum_{t=1}^{T} f_t f_t' \xrightarrow{p} \Sigma_f$ as $T \to \infty$ with Σ_f non-singular; *(iii)* $\{\epsilon_{it}^x, f_t\}$ and $\{\epsilon_{js}\}$ are mutually independent for all i, j, t, s; *(iv)* $E\left|\sum_{t=1}^{T} x_{it}\epsilon_{it}\right|^r \leq ME\left|\sum_{t=1}^{T} (x_{it}\epsilon_{it})^2\right|^{r/2}$ for all i, $r \leq 6$.

Assumption 3. *[slopes and loadings] (i)* $\beta_i = \beta + v_i$ with $\|\beta\| \leq M$ and v_i is *i.i.d.* across i with mean zero and independent of $\left\{\epsilon_{jt}, \epsilon_{jt}^x, f_t\right\}$ for all i, j, t; *(ii)* $E\|v_i\|^{2+\delta} < \infty$ for some $\delta > 0$; *(iii)* the γ_is are non stochastic and such that $\max_i \|\gamma_i\| < \infty$ and $n^{-1} \sum_{i=1}^{n} \gamma_i\gamma_i' \to \Sigma_\gamma$ as $n \to \infty$

with Σ_γ non-singular.

Assumption 4. *[Step 1 estimation] (i)* $l_{\min}\left(\frac{X_i'\bar{M}_w X_i}{T}\right) > 0$; $l_{\min}\left(\frac{X_i' M_F X_i}{T}\right) > 0$ and $l_{\min}\left(\frac{F' M_{X_i} F}{T}\right) > 0$ a.s. for all i, where $l_{\min}(\cdot)$ denotes the smallest eigenvalue; *(ii)* $C \equiv n^{-1}\sum_{i=1}^n C_i$ has rank $r \leq m+1$.

Assumptions 1-4 are the same set of assumptions as in Castagnetti, Rossi, and Trapani (2012); basically, they are needed in order to prove the consistency of the estimated common factors and loadings. Assumption 4 is specific to the CCE estimator, employed in Step 1. In Assumption 1, serial and cross sectional dependence are allowed for the error term ϵ_{it}. The rest of the assumption is similar to those in Bai (2009) and Pesaran (2006), and hold immediately if ϵ_{it} is assumed to be independent. Note that existence of the 12-th moment of ϵ_{it} is required - this is stronger than what the literature normally considers, and in our context it is needed in order to derive consistency of $\hat{\gamma}_i$ and $\hat{f}_t$. Finally, part *(iv)* contains Burkholder-type inequalities: these could be shown directly under more specific assumptions on the degree of serial and cross sectional dependence. For example, part (a) holds immediately if one assumes that ϵ_{it} is a Martingale Difference Sequence (MDS) across t (the same holds for part (b), under the MDS assumption across i) - see e.g. Lin and Bai (2010, p.108). As far as Assumption 2 is concerned, we allow for serial and cross sectional dependence in both the ϵ_{it}^xs and in the common factors f_t. The requirement in part *(ii)* is standard in the literature (see e.g. Assumption B in Bai, 2009), and it entails that common factors are "strong" in the sense of Chudik, Pesaran, and Tosetti (2011) (see in particular Assumption 3). Finally, according to part *(iii)*, the x_{it}s are strictly exogenous. Assumption 3 is standard. Assumption 4 is specific to the CCE estimator of the β_is, employed in Step 1. Particularly, the rank condition in part *(ii)* is the same as equation (21) in Pesaran (2006), and it guarantees the consistency of the $\tilde{\beta}_i$s.

We are now ready to present the asymptotics (rates of convergence and limiting distribution) for the two-stage estimators $\hat{\beta}^{AMG}$ and $\hat{\beta}^{AP}$. Let $E(v_i v_i') = \Omega_v$, and consider the Augmented Mean Group estimator defined in (8).

Theorem 1 *Let Assumptions 1-4 hold with $\Omega_v \neq 0$. Then, as $(n, T) \to \infty$*

$$\hat{\beta}^{AMG} - \beta = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\delta_{nT}^2}\right); \tag{11}$$

similar results hold for $\hat{\beta}^{AP}-\beta$*. Under* $\frac{\sqrt{n}}{T}\to 0$*, it holds that*

$$\sqrt{n}\left(\hat{\beta}^{AMG}-\beta\right)\xrightarrow{d}N\left(0,\Omega_v\right), \tag{12}$$

$$\sqrt{n}\left(\hat{\beta}^{AP}-\beta\right)\xrightarrow{d}N\left(0,\tilde{\Omega}_v\right), \tag{13}$$

where

$$\tilde{\Omega}_v=\lim_{n,T\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'M_FX_i\right)^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}X_i'M_FX_i\Omega_vX_i'M_FX_i\right]\left(\frac{1}{n}\sum_{i=1}^{n}X_i'M_FX_i\right)^{-1}. \tag{14}$$

If Assumptions 1-4 hold and $\Omega_v=0$*, then as* $(n,T)\to\infty$ *we have* $\hat{\beta}^{AMG}-\beta=O_p\left(\delta_{nT}^{-2}\right)$*. The same rates holds for* $\hat{\beta}^{AP}$*.*

Theorem 1 reports rates of convergence and asymptotic distributions of both the AMG estimator $\hat{\beta}^{AMG}$ and the AP one $\hat{\beta}^{AP}$; further, rates are studied under slope homogeneity. From a theoretical point of view, the theorem states that, in essence, there is no improvement with respect to the first stage estimator (see Pesaran, 2006). Indeed, the limiting distribution is driven by the term $n^{-1/2}\sum_{i=1}^{n}(\beta_i-\beta)$, which the iterative procedure does not improve. Indeed, as far as rates are concerned, equation (11) also contains the term $O_p\left(\frac{1}{T}\right)$, which arises from having the generated regressors $\hat{f}_t$ in the second stage equation (6). Similarly, under slope homogeneity, both $\hat{\beta}^{AMG}$ and $\hat{\beta}^{AP}$ are consistent at a rate $O_p\left(\delta_{nT}^{-2}\right)$, slower than the Mean Group and the Pooled estimators defined from $\tilde{\beta}_i$, which are consistent at a rate $O_p\left(\frac{1}{\sqrt{nT}}\right)$.

The covariance matrix Ω_v can be estimated using $\hat{\Omega}_v=n^{-1}\sum_{i=1}^{n}\left(\hat{\beta}_i-\hat{\beta}^{AMG}\right)\left(\hat{\beta}_i-\hat{\beta}^{AMG}\right)'$; the same passages as in Pesaran (2006) would entail $\hat{\Omega}_v\xrightarrow{p}\Omega_v$ as $(n,T)\to\infty$. By (14), $\tilde{\Omega}_v$ can also be readily estimated using $\hat{\Omega}_v$ as a plug-in estimator.

3 Monte Carlo Experiments

In this section, we compare the finite sample properties of several estimators (mainly based on the CCE of Pesaran, 2006, and the IFE of Bai, 2009, and Song, 2013) with the two-stage estimator proposed in Section 2, through synthetic data.

Simulations are carried out under different DGPs; specifically, we consider (1) a homogeneous panel with observed and unobserved common factors; (2) a heterogeneous panel with observed and unobserved common factors; and (3) a heterogeneous panel with unobserved common factors only. We refer to the subsections hereafter for a more through discussion of each specification. In each subsection, we consider combinations of n and T based on $n, T \in \{10, 30, 50, 100, 200\}$. Each experiment involves 5,000 replications.

It is worth noting that the estimators considered in this section require the determination of the number of the common factors r. This would be customarily done by using some information criteria such as the ones in Bai and Ng (2002). However, such method could, in small samples, produce an inconsistent estimate of r, thereby leading to incorrect inference. We refer to the simulation results in Coakley, Fuertes, and Smith (2002) and Kapetanios and Pesaran (2007). Thus, we firstly assume that the number of factors is known throughout the simulations, to examine the small sample properties of the estimators without them being affected by the issue of estimating r. We then introduce the number of factors estimation problem in order to take into account its impact on the estimation errors.

We consider the following DGP:

$$y_{it} = \alpha_1 d_{1t} + \alpha_2 d_{2t} + \beta_1 x_{1it} + \beta_2 x_{2it} + \gamma_i f_t + \epsilon_{it}, \tag{15}$$

$$x_{1it} = a_1 + \gamma_i f_t + \gamma_i + f_t + v_{1it}, \tag{16}$$

$$x_{2it} = a_2 + \gamma_i f_t + \gamma_i + f_t + v_{2it}. \tag{17}$$

In (15)-(17), we consider: 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. The error term ϵ_{it} is generated as *i.i.d.* $N(0, 2)$; further we set $a_1 = a_2 = 1$. The observed factors are generated as

$$\begin{aligned} d_{1t} &= 1, \\ d_{2t} &= f_t + v_t^d, \end{aligned}$$

where v_t^d is generated as *i.i.d.* $N(0, 1)$ and independent of all other regressors. The homogeneous coefficients are set to $\beta = (\beta_1, \beta_2) = (1, 3)$ and $\alpha = (\alpha_1, \alpha_2) = (5, 4)$. The variables γ_i, f_t, v_{jit} are all generated as *i.i.d.* $N(0, 1)$. This DGP is identical to Bai (2009) DGP for the case of common regressors.

Following Bai (2003), we compute for each iteration the correlation coefficient between $\{\hat{f}_{AP,t}\}_{t=1}^T$ and $\{f_t\}_{t=1}^T$, where $\hat{f}_{AP,t}$ is the estimated factor obtained from the residuals $\tilde{v}_{it} = y_{it} - x'_{it}\tilde{\beta}_{CCEP}$. We compare them with those obtained using the IFE estimate for the factor, $\hat{f}_{IFE,t}$. Table 1 below reports the average correlation coefficients for both estimation methods

[Insert Table 1 somewhere here]

The results in Table 1 suggest that both factor estimates are highly correlated with the unobserved factor. The last two columns of the table report the average number of iterations and the number of failures for each Monte Carlo simulation, respectively. Failure to achieve convergence for iterative procedures seems to occur particularly for small samples, which is particularly evident for $n = 30, T = 10$.

The bias and the RMSE of the estimates are in Tables 2 and 3 respectively. Further, bias and RMSE are also computed for the infeasible pooled estimator, calculated as:

$$\widehat{\beta}_{inf,pooled} = \left[\sum_{i=1}^{n} (X_i{}' M_F X_i) \right]^{-1} \sum_{i=1}^{n} (X_i{}' M_F y_i).$$

In principle, the infeasible estimator constitutes a lower bound to the bias and efficiency.

[Insert Tables 2 and 3 somewhere here]

We observe that for $n \leq 30$, the AP estimator slightly outperforms both the CCEP and the IFE estimator in terms of bias and RMSE. The situation is reversed when n is larger than 30. In this case, CCEP and IFE are less biased than AP. In general, the AP estimator and the IFE are quite close in terms of bias and RMSE. For all estimators considered, the RMSE decreases when n increases, for each value of T. As expected, the RMSE gets smaller as long as both n and T increase.

3.1 Heterogeneous panel with observed and unobserved common factors

The setup in (15)-(17) is extended to the case of heterogenous slopes, viz.

$$\alpha_{1i} = 5 + \eta_{i3}, \quad \alpha_{2i} = 4 + \eta_{i4}, \tag{18}$$

and

$$\beta_{1i} = 1 + \eta_{i1}, \quad \beta_{2i} = 3 + \eta_{i2}, \tag{19}$$

where $\eta_{ij} \sim i.i.d.N(0, 0.04)$ for $j = 1, \ldots, 4$. For each experiment we compute the CCEMG by Pesaran (2006), the Augmented Mean Group (AMG) and the Song (2013) estimator as well as the infeasible MG estimator, assuming f_t is observable. Namely the infeasible MG estimator is computed as:

$$\widehat{\beta}_{inf,MG} = \frac{1}{n}\sum_{i=1}^{n}(X_i{}' M_F X_i)^{-1}(X_i{}' M_F y_i).$$

The Song (2013) estimator is computed by allowing up to 200 iterations for each simulation; as a convergence criterion, we employ $\|\text{vec}(\hat{\beta}) - \text{vec}(\beta)\|$, where $\text{vec}(\beta) = (\beta_1', \ldots, \beta_n')'$ and $\beta_i = (\beta_{1i}, \beta_{2i})'$, and we set the tolerance coefficient equal to 0.0001.

Bias and RMSE of the various estimators are in Tables 4 and 5 respectively.

[Insert Tables 4 and 5 somewhere here]

The tables show that, when slope heterogeneity is considered, results are radically different from the homogeneous case in the previous subsection. In terms of bias, the AMG dominates both the CCEMG and the Mean Group estimator based on the IFE individual estimates studied in Song (2013). Similar conclusions can be drawn when considering the RMSEs in Table 5, although the differences between CCEMG and AMG are very small.

Finally, we consider the average correlation between estimated factors and true ones, reported in Table 6.

[Insert Table 6 somewhere here]

Interestingly, Table 6 shows that the two-step procedure yields estimated factors that are more correlated with the true ones than other procedures. Again, iterative procedures are computationally costly, even more than in the case of a homogeneous panel.

3.2 Heterogeneous panel with unobserved common factors only

We finally consider the following DGP, where only unobservable common factors are present:

$$y_{it} = \beta_{0i} + \beta_{1i}x_{1it} + \beta_{2i}x_{2it} + \gamma_i f_t + \epsilon_{it}, \tag{20}$$

and

$$x_{it} = a_{1i} + a_{2i}f_t + u_{it}, \tag{21}$$

with $\beta_{0i} = 5 + 0.2z_i$ and z_i generated as *i.i.d.* $N(0,1)$; further, we also generate a_{1i} and a_{2i} as *i.i.d.* $N(0,1)$. We contrast the same estimators as in the previous subsection.

Bias and RMSEs are reported in Tables 7 and 8 respectively.

[Insert Tables 7 and 8 somewhere here]

Table 7 shows that all the estimators considered do not exhibit significant differences as far as the bias is concerned. Conversely, based on Table 8, the two-step procedure has a lower RMSE than the other procedures.

Finally, correlations between true and estimated factors, reported in Table 9 shows that in the case of Song's iterative procedure, the average correlation is always smaller than that obtained with the AMG procedure. The two average correlations converge only when n becomes larger than 200. Finally, it should be noted how the Song's iterative procedure fail systematically to achieve convergence.

[Insert Table 9 somewhere here]

3.3 Unknown number of Factors

An important question is how robust are the various estimators to the knowledge of the true number of factors. We investigate the estimation of the number of factors problem via simulations. We extend the setup in Section 3.1 to the case of more factors, namely two factors. The factors $F_t = (F_{1t}, F_{2t})$ are generated as $i.i.d. N(0,1)$. We estimate the number of factors by using the IC_{p2} criteria of Bai and Ng (2002).[1] Table 10 reports the estimated number of factors averaged over $5,000$ replications. The first estimator, $\hat{r}_{AMG}$, is obtained from the residuals after the first stage CCEMG estimator. The estimator $\hat{r}_{Song}$ is obtained from the residuals of the Song's estimator. For the Song estimator we update the estimation of the factors at each iteration, until convergence of $\hat{\beta}$ according to the convergence criteria of Section 3.1. The estimate $\hat{r}_{AMG}$ stabilizes when $n, T > 30$. On the contrary, the estimates of r obtained with the Song's iterative procedure are more unstable and biased.

[1] We observe that changing the Bai and Ng (2002) criteria does not substantially alter the simulation results.

[Insert Table 10 somewhere here]

The knowledge of the true number of factors does not play a very important role in improving the performance of the estimators, as it is evident from the bias and RMSE reported in Tables 11 and 12, respectively. Further, there is no clear indication that one of the estimators is more affected than others by the uncertainty in the number of the unknown factors.

[Insert Table 11 and Table 12 somewhere here]

4 Conclusions

This paper considers inference in a stationary panel model where slopes are allowed to be heterogeneous and common unknown factors are present. A two-stage estimator is proposed, based on the CCE estimator (Pesaran, 2006) in the first stage and on a similar approach to the Interactive Effect estimator (Bai, 2009) in the second stage. This affords the estimation of the common factors coefficients, and, it also yields an alternative estimator for the individual slopes and for the average slope. The properties of the slope estimators are analyzed. The finite sample properties are investigated by means of Monte Carlo simulations, under different data-generating processes. The results show that the two-step estimator proposed has remarkable properties when compared to the Bai (2009) iterative estimator and its extension to the case of heterogeneous panels provided by Song (2013). The IFE estimator is computationally more demanding and less accurate when we consider the fact that it fails to achieve convergence in a relevant number of cases. In conclusion, there is no clear advantage in finite sample in adopting the iterative procedure with respect to the much more reliable two-stage procedure proposed here.

Appendix A: Preliminary lemmas

This section contains some preliminary Lemmas to prove Theorem 1. Lemmas 1-4 have been derived in Castagnetti, Rossi and Trapani (2012), and we refer to that paper for the proofs; Lemmas 5 and 6 are new, and the proofs are reported here. Their proofs are based on very similar arguments as in Castagnetti, Rossi, and Trapani (2012), and we only report the main passages when possible for the sake of a concise discussion.

Henceforth, we use the notation $\delta_{nT} = \min\left\{\sqrt{n}, \sqrt{T}\right\}$ and $\phi_{nT} = \min\{n, \sqrt{T}\}$. Note that, by equation (45) in Pesaran (2006, p.980)

$$\begin{aligned}
\tilde{\beta}_i - \beta_i &= \left(X_i' M_F X_i\right)^{-1}\left(X_i' M_F \epsilon_i\right) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right) &\quad (22)\\
&= O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right). &\quad (23)
\end{aligned}$$

Both expressions (22) and (23) will be used frequently; similarly, we henceforth define $\Upsilon_i = \left(X_i' M_F X_i\right)^{-1}\left(X_i' M_F \epsilon_i\right)$, so that we can write

$$\tilde{\beta}_i - \beta_i = \Upsilon_i + \bar{\Upsilon}_i, \quad (24)$$

for every i; by construction, (23) yields $\bar{\Upsilon}_i = O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right)$.

All proofs rely upon the decomposition - see Proposition A.1 in Bai (2009)

$$\begin{aligned}
\hat{F} - F &= \frac{1}{nT}\sum_{j=1}^{n} X_j\left(\tilde{\beta}_j - \beta_j\right)\left(\tilde{\beta}_j - \beta_j\right)' X_j'\hat{F}\\
&\quad - \frac{1}{nT}\sum_{j=1}^{n} X_j\left(\tilde{\beta}_j - \beta_j\right)\gamma_j' F'\hat{F} - \frac{1}{nT}\sum_{j=1}^{n} X_j\left(\tilde{\beta}_j - \beta_j\right)\epsilon_j'\hat{F}\\
&\quad - \frac{1}{nT}\sum_{j=1}^{n} F\gamma_j\left(\tilde{\beta}_j - \beta_j\right)' X_j'\hat{F} - \frac{1}{nT}\sum_{j=1}^{n} \epsilon_j\left(\tilde{\beta}_j - \beta_j\right)' X_j'\hat{F}\\
&\quad + \frac{1}{nT}\sum_{j=1}^{n} F\gamma_j\epsilon_j'\hat{F} + \frac{1}{nT}\sum_{j=1}^{n} \epsilon_j\gamma_j' F'\hat{F} + \frac{1}{nT}\sum_{j=1}^{n} \epsilon_j\epsilon_j'\hat{F} \qquad (25)\\
&= F1 + F2 + F3 + F4 + F5 + F6 + F7 + F8.
\end{aligned}$$

For the sake of notational simplicity, we set rotation matrix H in $\hat{F} - FH$ equal to the identity matrix. In view of (25), the only difference with Bai (2009) is the presence of the unit specific estimates, $\tilde{\beta}_j$.

Lemma 1 *Let Assumptions 1-4 hold. Then, as* $(n, T) \to \infty$

$$X_i' \left(\hat{F} - F \right) = O_p \left(T \delta_{nT}^{-2} \right), \tag{26}$$
$$\hat{F}' \left(\hat{F} - F \right) = O_p \left(T \delta_{nT}^{-2} \right). \tag{27}$$

for all i*.*

Lemma 2 *Let Assumptions 1-4 hold. Then, as* $(n, T) \to \infty$ *it holds that, for every* i

(i) $T^{-1} \epsilon_i' \left(\hat{F} - F \right) = O_p \left(\delta_{nT}^{-2} \right)$;

(ii) $n^{-1/2} T^{-1} \sum_{i=1}^n \epsilon_i' \left(\hat{F} - F \right) = O_p \left(n^{-1/2} \right) + O_p \left(T^{-1} \right)$.

Lemma 3 *Let Assumptions 1-4 hold. Then it holds that, for every* i

(i) $T^{-1} X_i' \left(\hat{F} - FH \right) = O_p \left(\delta_{nT}^{-2} \right)$;

(ii) $T^{-1} F' \left(\hat{F} - FH \right) = O_p \left(\delta_{nT}^{-2} \right)$;

(iii) $T^{-1} \left(\hat{F} - FH \right)' \left(\hat{F} - FH \right) = O_p \left(\delta_{nT}^{-2} \right)$.

Lemma 4 *Under Assumptions 1-4, it holds that, for every* i*,* $E \left\| \tilde{\beta}_i - \beta_i \right\|^r = O \left(\phi_{nT}^{-r} \right)$*, for any* $r \leq 3$.

Lemma 5 *Let Assumptions 1-4 hold. Then it holds that, for every* i

(i) $T \left(X_i' \hat{M}_F X_i \right)^{-1} = T \left(X_i' M_F X_i \right)^{-1} + O_p \left(\delta_{nT}^{-2} \right)$;

(ii) $\frac{1}{n} \sum_{i=1}^n \left(X_i' \hat{M}_F X_i \right)^{-1} X_i' \hat{M}_F \epsilon_i = O_p \left(\delta_{nT}^{-2} \right)$.

Proof. Part *(i)* is similar to Lemma A.7 in Bai (2009). Consider $T^{-1} X_i' \hat{M}_F X_i = T^{-1} X_i' M_F X_i + T^{-1} X_i' \left(\hat{M}_F - M_F \right) X_i$. Using the definition of $\hat{M}_F$ and the identification restrictions $\hat{F}'\hat{F} = T I_k$ and $F'F = T I_k$, we have

$$\frac{X_i' \left(\hat{M}_F - M_F \right) X_i}{T} = \frac{X_i' F}{T} \frac{\left(F - \hat{F} \right)' X_i}{T} + \frac{X_i' \left(F - \hat{F} \right)}{T} \frac{F' X_i}{T}$$
$$- \frac{X_i' \left(\hat{F} - F \right)}{T} \frac{\left(\hat{F} - F \right)' X_i}{T} = I + I' - II.$$

Using Lemma 1*(i)*, $I = O_p\left(\delta_{nT}^{-2}\right)$; similarly, $II = O_p\left(\delta_{nT}^{-4}\right)$. Since $T^{-1}X_i'M_FX_i$ is invertible by Assumption 4*(i)*, part *(i)* follows.

We now turn to part *(ii)*; using the short-hand notation $\Psi_i = T\left(X_i'M_FX_i\right)^{-1}$ we have

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\left(X_i'\hat{M}_FX_i\right)^{-1}X_i'\hat{M}_F\epsilon_i \\
= & \frac{1}{n}\sum_{i=1}^{n}\Psi_i\frac{X_i'\hat{M}_F\epsilon_i}{T}+\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{X_i'\hat{M}_FX_i}{T}\right)^{-1}-\left(\frac{X_i'M_FX_i}{T}\right)^{-1}\right]\frac{X_i'\hat{M}_F\epsilon_i}{T}+o_p(1) \\
= & \frac{1}{n}\sum_{i=1}^{n}\Psi_i\frac{X_i'M_F\epsilon_i}{T}+\frac{1}{n}\sum_{i=1}^{n}\Psi_i\frac{X_i'\left(\hat{M}_F-M_F\right)\epsilon_i}{T}+o_p(1)=I+II+o_p(1),
\end{aligned}$$

where the second equality follows from part *(i)* of the Lemma. Consider I; using Assumption 4

$$E\left\|\frac{1}{n}\sum_{i=1}^{n}\Psi_i\frac{X_i'M_F\epsilon_i}{T}\right\|^2=\frac{1}{n^2}\sum_{i=1}^{n}E\left\|\Psi_i\frac{X_i'M_F\epsilon_i}{T}\right\|^2\leq\frac{1}{n^2T}\sum_{i=1}^{n}E\left\|\frac{X_i'\epsilon_i}{\sqrt{T}}\right\|^2.$$

Note that

$$E\left\|\frac{X_j'\epsilon_j}{\sqrt{T}}\right\|^2=E\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}x_{jt}x_{js}'\epsilon_{jt}\epsilon_{js}\right\|\leq\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\|x_{jt}x_{js}'\right\|E\left|\epsilon_{jt}\epsilon_{js}\right|\leq M\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left|\epsilon_{jt}\epsilon_{js}\right|\leq M',$$

using, respectively, Assumptions 2*(iii)*, 2*(i)* and 1*(ii)*(c). Thus, $I = O_p\left(n^{-1/2}T^{-1/2}\right)$. As far as II is concerned, note

$$\begin{aligned}
-II = & -\frac{1}{n}\sum_{i=1}^{n}\Psi_i\frac{X_i'\left(\hat{M}_F-M_F\right)\epsilon_i}{T} \\
= & \frac{1}{n}\sum_{i=1}^{n}\Psi_i\left(\frac{X_i'F}{T}\right)\frac{\left(\hat{F}-F\right)'\epsilon_i}{T}+\frac{1}{n\sqrt{T}}\sum_{i=1}^{n}\Psi_i\frac{X_i'\left(\hat{F}-F\right)}{T}\frac{F'\epsilon_i}{\sqrt{T}} \\
& -\frac{1}{n}\sum_{i=1}^{n}\Psi_i\frac{X_i'\left(\hat{F}-F\right)}{T}\frac{\left(\hat{F}-F\right)'\epsilon_i}{T}=II_a+II_b-II_c.
\end{aligned}$$

Consider II_a; this is bounded by $E\left[\left\|\frac{X_i'F}{T}\right\|\left\|\frac{(\hat{F}-F)'\epsilon_i}{T}\right\|\right]$. Using Lemma 2*(i)* and the same passages as above, $II_a = O_p\left(\delta_{nT}^{-2}\right)$. Similarly, $\|II_b\|$ is bounded by $T^{-1/2}E\left[\left\|\frac{X_i'(\hat{F}-F)}{T}\right\|\left\|\frac{F'\epsilon_i}{\sqrt{T}}\right\|\right] = O_p\left(T^{-1/2}\delta_{nT}^{-2}\right)$. Similarly, $\|II_c\| = O_p\left(\delta_{nT}^{-4}\right)$. Putting all together, $II = O_p\left(\delta_{nT}^{-2}\right)$. Thus, $I + II = O_p\left(n^{-1/2}T^{-1/2}\right) + O_p\left(\delta_{nT}^{-2}\right) = O_p\left(\delta_{nT}^{-2}\right)$. QED

Lemma 6 *Under Assumptions 1-4, it holds that* $\frac{1}{n}\sum_{i=1}^{n}\left(X_i'\hat{M}_FX_i\right)^{-1}X_i'\hat{M}_FF\gamma_i = O_p\left(\delta_{nT}^{-2}\right)$.

Proof. Using Lemma 5*(i)*, $\frac{1}{n}\sum_{i=1}^{n}\left(X_i'\hat{M}_F X_i\right)^{-1} X_i'\hat{M}_F F\gamma_i = \frac{1}{n}\sum_{i=1}^{n}\left(X_i' M_F X_i\right)^{-1} X_i'\hat{M}_F F\gamma_i$ $+o_p(1)$. By the definition of $\hat{M}_F$, $\frac{1}{n}\sum_{i=1}^{n}\left(X_i'\hat{M}_F X_i\right)^{-1} X_i'\hat{M}_F F\gamma_i = \frac{1}{n}\sum_{i=1}^{n}\left(X_i'\hat{M}_F X_i\right)^{-1} X_i'\hat{M}_F\left(F-\hat{F}\right)\gamma_i$. Using (25), we obtain

$$
\begin{aligned}
& -\frac{1}{n}\sum_{i=1}^{n}\Psi_i\frac{X_i'\hat{M}_F\left(F-\hat{F}\right)}{T}\gamma_i \\
= & \ \frac{1}{n}\sum_{i=1}^{n}\Psi_i\frac{1}{n}\sum_{j=1}^{n}\frac{X_i'\hat{M}_F X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}\gamma_i \\
& -\frac{1}{n^2T}\sum_{i=1}^{n}\Psi_i X_i'\hat{M}_F\sum_{j=1}^{n}X_j\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'\frac{F'\hat{F}}{T}\gamma_i \\
& -\frac{1}{n^2T}\sum_{i=1}^{n}\Psi_i X_i'\hat{M}_F\sum_{j=1}^{n}X_j\left(\tilde{\beta}_j-\beta_j\right)\frac{\epsilon_j'\hat{F}}{T}\gamma_i-\frac{1}{n^2T}\sum_{i=1}^{n}\Psi_i X_i'\hat{M}_F\sum_{j=1}^{n}F\gamma_j\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}\gamma_i \\
& -\frac{1}{n^2T}\sum_{i=1}^{n}\Psi_i X_i'\hat{M}_F\sum_{j=1}^{n}\epsilon_j\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}\gamma_i+\frac{1}{n^2T}\sum_{i=1}^{n}\Psi_i X_i'\hat{M}_F\sum_{j=1}^{n}F\gamma_j\frac{\epsilon_j'\hat{F}}{T}\gamma_i \\
& +\frac{1}{n^2T}\sum_{i=1}^{n}\Psi_i X_i'\hat{M}_F\sum_{j=1}^{n}\epsilon_j\gamma_j'\frac{F'\hat{F}}{T}\gamma_i+\frac{1}{n^2T}\sum_{i=1}^{n}\Psi_i X_i'\hat{M}_F\sum_{j=1}^{n}\epsilon_j\frac{\epsilon_j'\hat{F}}{T}\gamma_i \\
= & \ I-II-III-IV-V+VI+VII+VIII.
\end{aligned}
$$

Consider I; we have

$$
I \leq E\left[\left\|\frac{X_i'M_F X_j}{T}\right\|\left\|\frac{X_j'\hat{F}}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|^2\right]+E\left[\left\|\frac{X_i'\left(\hat{M}_F-M_F\right)X_j}{T}\right\|\left\|\frac{X_j'\hat{F}}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|^2\right]=I_a+I_b.
$$

Hence, using a similar logic to the proofs in Castagnetti, Rossi, and Trapani (2012)

$$
\begin{aligned}
I_a & \leq \left[E\left(\left\|\tilde{\beta}_j-\beta_j\right\|^{2p}\right)\right]^{1/p}\left[E\left(\left\|\frac{X_i'M_F X_j}{T}\right\|\left\|\frac{X_j'\hat{F}}{T}\right\|\right)^q\right]^{1/q} \\
& \leq \left[E\left\|\tilde{\beta}_j-\beta_j\right\|^3\right]^{2/3}\left[E\left\|\frac{X_i'X_j}{T}\right\|^6\right]^{1/6}\left[E\left\|\frac{X_j'\hat{F}}{T}\right\|^6\right]^{1/6},
\end{aligned}
$$

using Holder's inequality in the first line (with $p=\frac{3}{2}$ and $q=3$), and the Cauchy-Schwartz inequality in the second line. Lemma 4 yields that the first term is $O\left(\phi_{nT}^{-2}\right)$; Assumption 2 ensures that the second and the third term are both $O(1)$. As far as I_b is concerned, by Lemma A.3*(i)*-*(ii)* in Bai (2009), $\left\|\hat{M}_F-M_F\right\|=o_p(1)$, and therefore I_b is dominated. Thus, $I=O_p\left(\phi_{nT}^{-2}\right)$. Turning to II, using (24)

$$II=\frac{1}{n^2T^2}\sum_{i=1}^{n}\Psi_i X_i' M_F\left[\sum_{j=1}^{n}X_j\Upsilon_j\gamma_j'F'\hat{F}\gamma_i+\sum_{j=1}^{n}X_j\bar{\Upsilon}_j\gamma_j'F'\hat{F}\gamma_i\right]+o_p(1)=II_a+II_b+o_p(1),$$

where the $o_p(1)$ term comes from $\left\|\hat{M}_F-M_F\right\|=o_p(1)$. Consider II_a; this is bounded by

$$\begin{aligned}&\frac{1}{\sqrt{nT}}\left\|\frac{F'\hat{F}}{T}\right\|\left[E\left\|\frac{X_i}{\sqrt{T}}\right\|^2\right]^{1/2}\left[E\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}X_j\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_j'\bar{M}_w\epsilon_j}{T}\right)\right\|^2\right]^{1/2}\\ =\;& O_p\left(\frac{1}{\sqrt{nT}}\right),\end{aligned}$$

based on the same passages as above. Turning to II_b, this is bounded by $\left\|\frac{F'\hat{F}}{T}\right\|\left[E\left\|\frac{X_i'X_j}{T}\right\|^2\right]^{1/2}$ $\left[E\left\|\bar{\Upsilon}_j\right\|^2\right]^{1/2}$. Based on Lemma 3, Assumption 2, and (24), we have that this is $O_p\left(\frac{1}{n}\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)$. Hence, $II=O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$. As far as III is concerned, it is bounded by

$$\begin{aligned}&\frac{1}{n^2\sqrt{T}}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\Psi_i\|\left\|\frac{X_i'M_FX_j}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|\left\|\frac{\epsilon_j'F}{\sqrt{T}}\right\|\\ &+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\Psi_i\|\left\|\frac{X_i'M_FX_j}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|\left\|\frac{\epsilon_j'\left(\hat{F}-F\right)}{T}\right\|\\ =\;& III_a+III_b.\end{aligned}$$

Term III_a is bounded by $T^{-1/2}\ E\left[\left\|\frac{X_i'X_j}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|\left\|\frac{\epsilon_j'F}{\sqrt{T}}\right\|\right]$; we have

$$III\leq\left[E\left\|\tilde{\beta}_j-\beta_j\right\|^3\right]^{2/3}\left[E\left\|\frac{X_i'X_j}{T}\right\|^6\right]^{1/6}\left[E\left\|\frac{\epsilon_j'F}{\sqrt{T}}\right\|^6\right]^{1/6};$$

by a similar logic as above, $E\left\|\frac{\epsilon_j'F}{\sqrt{T}}\right\|^6$ is bounded, and hence similar passages as above yield III_a $=O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$. Similarly, by Lemma 2*(i)*, $III_b=O_p\left(\left\|\tilde{\beta}_j-\beta_j\right\|\right)O_p\left(\delta_{nT}^{-2}\right)$. This entails $III=O_p\left(\phi_{nT}^{-1}\delta_{nT}^{-2}\right)$. As far as IV and V are concerned, they have the same order of magnitude as II and

III respectively. Considering VI, it is bounded by

$$\begin{aligned} & \frac{1}{n\sqrt{nT}}\sum_{i=1}^{n}\|\Psi_i\|\left\|\frac{X_i'\left(\hat{M}_F-M_F\right)F}{T}\right\|\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\epsilon_j'F}{\sqrt{T}}\right\| \\ & +\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\left\|\frac{X_i'\left(\hat{M}_F-M_F\right)F}{T}\right\|\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\epsilon_j'\left(\hat{F}-F\right)}{T}\right\| \\ = & \; VI_a+VI_b. \end{aligned}$$

Note that $E\left\|\frac{X_i'F}{T}\right\|$ can be shown to be bounded by using the Cauchy-Schwartz inequality and by using Assumption 2; further,

$$\begin{aligned} E\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\epsilon_j'F}{\sqrt{T}}\right\|^2 &= E\left\|\frac{1}{nT}\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\epsilon_{it}\epsilon_{js}f_tf_s\right\| \\ &\leq M\frac{1}{nT}\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\|\epsilon_{it}\epsilon_{js}f_tf_s\right\| \\ &\leq M'\frac{1}{nT}\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left|\epsilon_{it}\epsilon_{js}\right|E\left\|f_tf_s\right\| \\ &\leq M''\frac{1}{nT}\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left|\epsilon_{it}\epsilon_{js}\right|\leq M''', \end{aligned}$$

using Assumption 2 and 1*(ii)*(d). Therefore, VI_a is bounded by $\frac{1}{\sqrt{nT}}\left\|\hat{M}_F-M_F\right\|$; since $\left\|\hat{M}_F-M_F\right\|$ $=o_p(1)$, $VI_a=o_p\left(n^{-1/2}T^{-1/2}\right)$. Turning to VI_b, note that $\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\epsilon_j'(\hat{F}-F)}{T}\right\|=O_p\left(\delta_{nT}^{-2}\right)$ by Lemma 2*(ii)*; thus, $VI_b=o_p\left(n^{-1/2}\delta_{nT}^{-2}\right)$. Therefore, $VI=o_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$. Turning to VII, it is bounded by $n^{-1/2}\left\|\frac{F'\hat{F}}{T}\right\|E\left\|n^{-1/2}\sum_{j=1}^{n}\Psi_i\frac{X_i'\hat{M}_F\epsilon_j}{T}\right\|$; by Lemma 5*(ii)*, this entails that $VII=O_p\left(\delta_{nT}^{-2}\right)$. Finally, turning to $VIII$, this is bounded by

$$\frac{1}{n^2\sqrt{T}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|\Psi_i\frac{X_i'\hat{M}_F\epsilon_j}{T}\right\|\left\|\frac{\epsilon_j'F}{\sqrt{T}}\right\|+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|\Psi_i\frac{X_i'\hat{M}_F\epsilon_j}{T}\right\|\left\|\frac{\epsilon_j'\left(\hat{F}-F\right)}{T}\right\|=VIII_a+VIII_b.$$

As above, $VIII_a$ is bounded by $T^{-1/2}\left[E\left\|\Psi_i\frac{X_i'\hat{M}_F\epsilon_j}{T}\right\|^2\right]^{1/2}\left[E\left\|\frac{\epsilon_j'F}{\sqrt{T}}\right\|^2\right]^{1/2}$. Given that $\Psi_i\frac{X_i'M_F\epsilon_j}{T}=O_p\left(T^{-1/2}\right)$, and that, using Lemma 3, $\left\|\hat{M}_F-M_F\right\|=O_p\left(\delta_{nT}^{-2}\right)$, we have $VIII_a=O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$. Also, by Lemma 2*(i)*, $VIII_b=O_p\left(\phi_{nT}^{-1}\right)O_p\left(\delta_{nT}^{-2}\right)$. Therefore, $VIII=O_p\left(\phi_{nT}^{-2}\right)$. Putting all together, the Lemma follows. QED

Appendix B

Proof of Theorem 1. Consider (11) and (12). It holds that, by definition

$$\begin{aligned} \hat{\beta}^{AMG} - \beta &= \frac{1}{n}\sum_{i=1}^{n}(\beta_i - \beta) + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_i - \beta_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}(\beta_i - \beta) + \frac{1}{n}\sum_{i=1}^{n}\left(X_i'\hat{M}_F X_i\right)^{-1}\left(X_i'\hat{M}_F\epsilon_i + X_i'\hat{M}_F F\gamma_i\right), \end{aligned} \tag{28}$$

and

$$\hat{\beta}^{AP} - \beta = \left[\frac{1}{n}\sum_{i=1}^{n}X_i'\hat{M}_F X_i\right]^{-1} \times \left[\frac{1}{n}\sum_{i=1}^{n}X_i'\hat{M}_F X_i(\beta_i - \beta) + \frac{1}{n}\sum_{i=1}^{n}X_i'\hat{M}_F\epsilon_i + \frac{1}{n}\sum_{i=1}^{n}X_i'\hat{M}_F F\gamma_i\right]. \tag{29}$$

Equation (11) follows immediately from Lemmas 3 and 5, and by noting that, using Assumption 3,

$$E\left\|\frac{1}{n}\sum_{i=1}^{n}(\beta_i - \beta)\right\|^2 = \frac{1}{n^2}\sum_{i=1}^{n}E\left\|(\beta_i - \beta)\right\|^2 = O\left(\frac{1}{n}\right),$$

and

$$E\left\|\frac{1}{n}\sum_{i=1}^{n}X_i'\hat{M}_F X_i(\beta_i - \beta)\right\|^2 = \frac{1}{n^2}\sum_{i=1}^{n}E\left\|X_i'\hat{M}_F X_i\right\|^2 E\left\|(\beta_i - \beta)\right\|^2 = O\left(\frac{1}{n}\right).$$

Consider now (12). The term that dominates in (28) is $\frac{1}{n}\sum_{i=1}^{n}(\beta_i - \beta)$; under Assumption 3, a Central Limit Theorem (CLT) holds, so that (12) follows immediately. As far as (13) is concerned, in (29) the term that dominates is $\left[\frac{1}{n}\sum_{i=1}^{n}X_i'\hat{M}_F X_i\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}X_i'\hat{M}_F X_i(\beta_i - \beta)\right]$. Consider the denominator; from the above we have

$$\frac{1}{n}\sum_{i=1}^{n}X_i'\hat{M}_F X_i = \frac{1}{n}\sum_{i=1}^{n}X_i'M_F X_i + o_p(1);$$

note that the $X_i'M_F X_i$s are a conditionally independent sequence with finite second moment - indeed, by Assumption 2, the sequence has moments up to the 6th. Hence a conditional Law of Large Numbers can be applied (see e.g. Rao (2009)), and $\frac{1}{n}\sum_{i=1}^{n}X_i'M_F X_i \stackrel{d}{\to} \frac{1}{n}\sum_{i=1}^{n}E(X_i'M_F X_i)$, where E represents the expected value operator conditional upon the σ-field $\{f_t\}_{t=1}^{T}$. As far as the numerator is concerned, note that $X_i'\hat{M}_F X_i(\beta_i - \beta) = X_i'M_F X_i(\beta_i - \beta) + o_p(1)$; further, the sequence $X_i'M_F X_i(\beta_i - \beta)$ is, conditional upon $\{f_t\}_{t=1}^{T}$, independent across i. A Lyapunov condition can be shown to hold, with $E\left\|X_i'M_F X_i(\beta_i - \beta)\right\|^{2+\delta} \leq E\left\|X_i'X_i\right\|^{2+\delta} E\left\|\beta_i - \beta\right\|^{2+\delta}$, on account of Assumptions 2 and 3. Thus, the conditional CLT (Rao, 2009) yields

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i' M_F X_i \left(\beta_i - \beta\right) \xrightarrow{d} V^{1/2} \times Z,$$

with $Z \sim N\left(0, I_k\right)$ independent of $\left\{f_t\right\}_{t=1}^{T}$ and

$$\begin{aligned} V &= \lim_{n,T\to\infty} E\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{X_i' M_F X_i}{T}\left(\beta_i - \beta\right)\left(\beta_j - \beta\right)' \frac{X_j' M_F X_j}{T}\right] \\ &= \lim_{n,T\to\infty} E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{X_i' M_F X_i}{T}\left(\beta_i - \beta\right)\left(\beta_i - \beta\right)' \frac{X_i' M_F X_i}{T}\right] \\ &= \lim_{n,T\to\infty} E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{X_i' M_F X_i}{T}\Omega_v \frac{X_i' M_F X_i}{T}\right]. \end{aligned}$$

Putting all together, (13) follows.

Finally, note that, when $\beta_i = \beta$ (and, therefore, $\Omega_v = 0$) for all i, $\hat{\beta}^{AMG} - \beta$ reduces to $n^{-1}\sum_{i=1}^{n}\left(X_i'\hat{M}_F X_i\right)^{-1} X_i'\hat{M}_F \epsilon_i + n^{-1}\sum_{i=1}^{n}\left(X_i'\hat{M}_F X_i\right)^{-1} X_i'\hat{M}_F F\gamma_i$, which as mentioned above is $O_p\left(\delta_{nT}^{-2}\right)$. The same passages yield the same result for $\hat{\beta}^{AP}$.

n	T	Corr$(F_t, \hat{F}_{IFE,t})$	Corr$(F_t, \hat{F}_{AP,t})$	iter	fail
10	10	0.5939	0.6463	43.102	58
10	30	0.6878	0.7194	41.775	32
10	50	0.7354	0.7564	40.756	12
10	100	0.7744	0.7820	44.141	4
10	200	0.7974	0.7922	45.345	3
30	10	0.7900	0.7815	41.345	29
30	30	0.9017	0.8970	42.350	0
30	50	0.9196	0.9131	43.200	0
30	100	0.9287	0.9231	43.013	0
30	200	0.9323	0.9275	43.452	0
50	10	0.8679	0.8421	42.504	24
50	30	0.9458	0.9321	44.192	0
50	50	0.9536	0.9443	43.239	0
50	100	0.9574	0.9521	43.231	0
50	200	0.9596	0.9561	42.895	0
100	10	0.9419	0.8957	43.584	2
100	30	0.9745	0.9576	43.817	0
100	50	0.9773	0.9671	42.990	0
100	100	0.9788	0.9737	42.678	0
100	200	0.9796	0.9768	42.192	0
200	10	0.9732	0.9183	45.069	1
200	30	0.9876	0.9702	43.386	0
200	50	0.9888	0.9784	43.043	0
200	100	0.9894	0.9843	42.290	0
200	200	0.9898	0.9872	42.080	0

Table 1: Average correlation coefficients between $\{F_t\}_{t=1}^T$ and $\{\hat{F}_{IFE,t}\}_{t=1}^T$ and $\{\hat{F}_{AP,t}\}_{t=1}^T$ in the case of homogeneous slope. *Iter* and *fail* indicate the average number of iterations and the number of failures (lack of convergence) of the iterative process for the estimation method of Bai (2009), respectively. The DGP is in (15)-(17).

		IFE		CCEP		Inf Pooled		AP	
n	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.0940	0.0917	0.0548	0.0473	-0.0003	-0.0010	0.0665	0.0658
10	30	0.0606	0.0585	0.0515	0.0481	0.0010	-0.0011	0.0409	0.0388
10	50	0.0440	0.0461	0.0500	0.0509	-0.0011	-0.0003	0.0296	0.0310
10	100	0.0317	0.0302	0.0526	0.0504	0.0007	-0.0009	0.0242	0.0229
10	200	0.0166	0.0173	0.0522	0.0526	0.0001	0.0003	0.0189	0.0196
30	10	0.0655	0.0606	0.0201	0.0142	0.0009	-0.0021	0.0517	0.0470
30	30	0.0268	0.0258	0.0175	0.0163	0.0003	-0.0011	0.0175	0.0163
30	50	0.0155	0.0141	0.0183	0.0167	0.0010	-0.0003	0.0118	0.0103
30	100	0.0083	0.0072	0.0179	0.0167	0.0005	-0.0004	0.0082	0.0071
30	200	0.0037	0.0040	0.0173	0.0175	0.0001	0.0004	0.0047	0.0050
50	10	0.0512	0.0536	0.0082	0.0150	-0.0004	0.0026	0.0412	0.0458
50	30	0.0150	0.0152	0.0097	0.0103	-0.0004	-0.0002	0.0130	0.0133
50	50	0.0083	0.0079	0.0103	0.0094	0.0002	-0.0003	0.0081	0.0078
50	100	0.0036	0.0041	0.0098	0.0103	-0.0003	0.0002	0.0045	0.0050
50	200	0.0021	0.0025	0.0103	0.0107	-0.0001	0.0004	0.0027	0.0031
100	10	0.0287	0.0299	0.0055	0.0042	-0.0007	0.0006	0.0331	0.0348
100	30	0.0066	0.0078	0.0040	0.0054	-0.0005	0.0008	0.0094	0.0106
100	50	0.0039	0.0036	0.0051	0.0045	0.0000	-0.0002	0.0067	0.0064
100	100	0.0023	0.0022	0.0053	0.0052	-0.0002	-0.0002	0.0035	0.0034
100	200	0.0012	0.0011	0.0051	0.0050	0.0000	-0.0001	0.0017	0.0017
200	10	0.0160	0.0163	0.0019	0.0023	-0.0002	0.0002	0.0298	0.0303
200	30	0.0040	0.0035	0.0027	0.0023	0.0001	-0.0003	0.0090	0.0087
200	50	0.0018	0.0025	0.0023	0.0029	-0.0002	0.0005	0.0054	0.0062
200	100	0.0010	0.0013	0.0024	0.0028	0.0000	0.0003	0.0020	0.0024
200	200	0.0007	0.0007	0.0026	0.0026	0.0001	0.0001	0.0012	0.0012

Table 2: Bias of β estimators in the case of homogeneous slope. The DGP is in (15)-(15). *IFE* is the Interactive Fixed Effects estimator by Bai (2009), *CCEP* is the Pooled CCE estimator by Pesaran (2006), *Inf Pooled* is the infeasible pooled estimator obtained with pooling and unobserved common factors. *AP* is the Augmented Pooled estimator.

		IFE		CCEP		Inf Pooled		AP	
n	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.2563	0.2469	0.3054	0.2911	0.1767	0.1751	0.2182	0.2090
10	30	0.1594	0.1594	0.1499	0.1467	0.0968	0.1022	0.1225	0.1253
10	50	0.1245	0.1281	0.1123	0.1139	0.0754	0.0723	0.0919	0.0916
10	100	0.0940	0.0965	0.0945	0.0939	0.0508	0.0537	0.0650	0.0668
10	200	0.0668	0.0669	0.0766	0.0785	0.0366	0.0378	0.0464	0.0485
30	10	0.1584	0.1558	0.1691	0.1675	0.1005	0.0981	0.1378	0.1391
30	30	0.0894	0.0877	0.0752	0.0767	0.0550	0.0570	0.0762	0.0787
30	50	0.0657	0.0637	0.0581	0.0592	0.0434	0.0433	0.0601	0.0610
30	100	0.0446	0.0419	0.0436	0.0429	0.0311	0.0306	0.0402	0.0424
30	200	0.0282	0.0288	0.0332	0.0332	0.0217	0.0214	0.0302	0.0307
50	10	0.1260	0.1310	0.1303	0.1277	0.0736	0.0749	0.1136	0.1194
50	30	0.0625	0.0651	0.0546	0.0582	0.0403	0.0430	0.0661	0.0700
50	50	0.0477	0.0478	0.0438	0.0444	0.0323	0.0348	0.0524	0.0530
50	100	0.0313	0.0314	0.0316	0.0308	0.0234	0.0229	0.0383	0.0376
50	200	0.0216	0.0218	0.0230	0.0234	0.0166	0.0167	0.0270	0.0271
100	10	0.0963	0.0943	0.0941	0.0898	0.0561	0.0533	0.1044	0.1067
100	30	0.0436	0.0438	0.0407	0.0402	0.0318	0.0313	0.0644	0.0628
100	50	0.0300	0.0310	0.0297	0.0305	0.0232	0.0228	0.0489	0.0501
100	100	0.0214	0.0214	0.0211	0.0210	0.0167	0.0171	0.0351	0.0357
100	200	0.0147	0.0144	0.0150	0.0145	0.0119	0.0111	0.0252	0.0250
200	10	0.0668	0.0688	0.0606	0.0610	0.0375	0.0378	0.0990	0.0984
200	30	0.0289	0.0291	0.0285	0.0289	0.0218	0.0218	0.0609	0.0613
200	50	0.0214	0.0209	0.0210	0.0208	0.0164	0.0162	0.0484	0.0490
200	100	0.0147	0.0150	0.0145	0.0153	0.0116	0.0117	0.0343	0.0345
200	200	0.0102	0.0103	0.0105	0.0104	0.0082	0.0083	0.0242	0.0249

Table 3: RMSE of β estimators in the case of homogeneous slope. The DGP is in (15)-(15). *IFE* is the Interactive Fixed Effects estimator by Bai (2009), *CCEP* is the Pooled CCE estimator by Pesaran (2006), *Inf Pooled* is the infeasible pooled estimator obtained with pooling and unobserved common factors. *AP* is the Augmented Pooled estimator.

		Inf MG		CCEMG		AMG		Song	
n	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.0327	0.1679	0.0763	0.2155	0.0778	0.2120	0.1374	0.2583
10	30	-0.1060	0.0210	-0.0626	0.0648	-0.0740	0.0516	-0.0282	0.0979
10	50	-0.0183	-0.0250	0.0275	0.0197	0.0092	0.0004	0.0513	0.0430
10	100	0.1007	-0.0208	0.1446	0.0235	0.1226	0.0016	0.1551	0.0349
10	200	0.0543	-0.0062	0.0983	0.0377	0.0752	0.0143	0.0944	0.0336
30	10	-0.0141	0.0104	0.0016	0.0221	0.0122	0.0373	0.0863	0.1109
30	30	-0.0202	-0.0463	-0.0032	-0.0305	-0.0087	-0.0344	0.0389	0.0118
30	50	-0.0362	0.0194	-0.0198	0.0357	-0.0268	0.0287	0.0078	0.0632
30	100	0.0127	0.0171	0.0290	0.0333	0.0190	0.0234	0.0388	0.0433
30	200	0.0033	0.0013	0.0196	0.0177	0.0087	0.0067	0.0173	0.0152
50	10	-0.0079	0.0231	0.0048	0.0317	0.0138	0.0437	0.0870	0.1137
50	30	0.0218	-0.0217	0.0316	-0.0119	0.0296	-0.0139	0.0681	0.0248
50	50	0.0071	-0.0281	0.0170	-0.0181	0.0124	-0.0228	0.0377	0.0023
50	100	0.0129	0.0231	0.0227	0.0328	0.0165	0.0268	0.0291	0.0394
50	200	0.0384	0.0223	0.0483	0.0322	0.0412	0.0252	0.0463	0.0304
100	10	-0.0009	-0.0320	0.0047	-0.0251	0.0141	-0.0174	0.0829	0.0523
100	30	-0.0436	0.0376	-0.0385	0.0426	-0.0395	0.0417	-0.0142	0.0670
100	50	-0.0261	0.0216	-0.0213	0.0264	-0.0232	0.0244	-0.0088	0.0389
100	100	-0.0077	-0.0145	-0.0027	-0.0096	-0.0059	-0.0127	0.0006	-0.0062
100	200	0.0175	-0.0098	0.0225	-0.0048	0.0187	-0.0085	0.0215	-0.0058
200	10	0.0173	-0.0028	0.0195	0.0019	0.0249	0.0056	0.0883	0.0685
200	30	-0.0223	-0.0030	-0.0195	-0.0006	-0.0197	-0.0006	-0.0057	0.0136
200	50	-0.0027	0.0016	-0.0001	0.0041	-0.0013	0.0030	0.0062	0.0104
200	100	0.0076	0.0222	0.0101	0.0247	0.0084	0.0230	0.0117	0.0264
200	200	-0.0011	0.0051	0.0013	0.0076	-0.0006	0.0057	0.0008	0.0071

Table 4: Bias of β_i estimators in the case of heterogeneous slope, see (18) and (19). *Inf MG* is the mean group estimators obtained with unobserved common factors. *CCEMG* is the mean group CCE estimator by Pesaran (2006). *AMG* is the mean group estimator in (8). *Song* is the Interactive Fixed Effects estimator by Bai (2009) extended to the heterogenous case by Song (2013).

		Inf MG		CCEMG		AMG		Song	
n	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.3190	0.3203	0.4517	0.4626	0.3462	0.3453	0.4861	0.4880
10	30	0.1283	0.1253	0.1463	0.1473	0.1439	0.1450	0.1908	0.1853
10	50	0.0943	0.0962	0.1131	0.1158	0.1087	0.1111	0.1471	0.1505
10	100	0.0649	0.0658	0.0859	0.0857	0.0769	0.0767	0.1129	0.1127
10	200	0.0450	0.0453	0.0707	0.0704	0.0565	0.0562	0.0863	0.0849
30	10	0.1846	0.1840	0.2583	0.2573	0.1923	0.1952	0.2363	0.2384
30	30	0.0751	0.0754	0.0812	0.0802	0.0808	0.0799	0.1057	0.1052
30	50	0.0556	0.0556	0.0602	0.0608	0.0600	0.0600	0.0800	0.0808
30	100	0.0376	0.0374	0.0426	0.0423	0.0405	0.0402	0.0538	0.0542
30	200	0.0264	0.0264	0.0326	0.0318	0.0290	0.0288	0.0357	0.0358
50	10	0.1401	0.1438	0.1992	0.1990	0.1476	0.1494	0.1857	0.1881
50	30	0.0573	0.0565	0.0608	0.0603	0.0601	0.0590	0.0805	0.0798
50	50	0.0426	0.0427	0.0451	0.0447	0.0446	0.0448	0.0582	0.0580
50	100	0.0291	0.0293	0.0314	0.0317	0.0304	0.0308	0.0377	0.0375
50	200	0.0199	0.0206	0.0229	0.0233	0.0212	0.0218	0.0238	0.0242
100	10	0.1019	0.1003	0.1453	0.1402	0.1067	0.1059	0.1355	0.1369
100	30	0.0411	0.0405	0.0434	0.0427	0.0426	0.0419	0.0549	0.0543
100	50	0.0305	0.0304	0.0317	0.0316	0.0316	0.0314	0.0379	0.0379
100	100	0.0206	0.0206	0.0215	0.0215	0.0211	0.0213	0.0236	0.0237
100	200	0.0144	0.0144	0.0154	0.0154	0.0148	0.0149	0.0155	0.0156
200	10	0.0721	0.0708	0.0992	0.1005	0.0742	0.0731	0.1043	0.1037
200	30	0.0289	0.0292	0.0307	0.0304	0.0296	0.0296	0.0357	0.0353
200	50	0.0211	0.0210	0.0218	0.0217	0.0215	0.0213	0.0239	0.0238
200	100	0.0145	0.0148	0.0149	0.0151	0.0148	0.0150	0.0156	0.0158
200	200	0.0102	0.0101	0.0106	0.0105	0.0104	0.0102	0.0106	0.0105

Table 5: RMSE of β_i estimators in the case of heterogeneous slope, see (18) and (19). *Inf MG* is the mean group estimators obtained with unobserved common factors. *CCEMG* is the mean group CCE estimator by Pesaran (2006). *AMG* is the mean group estimator in (8). *Song* is the Interactive Fixed Effects estimator by Bai (2009) extended to the heterogenous case by Song (2013).

n	T	$\text{Corr}(F_t, \hat{F}_{Song,t})$	$\text{Corr}(F_t, \hat{F}_{AMG,t})$	iter	fail
10	10	0.490	0.687	145.659	2300
10	30	0.592	0.752	92.385	884
10	50	0.646	0.781	91.724	860
10	100	0.699	0.803	95.889	876
10	200	0.739	0.811	104.351	1028
30	10	0.562	0.816	155.686	2487
30	30	0.791	0.908	126.797	1438
30	50	0.852	0.921	133.215	1463
30	100	0.893	0.931	142.441	1590
30	200	0.915	0.935	147.912	1506
50	10	0.601	0.856	168.865	3020
50	30	0.867	0.939	152.717	2157
50	50	0.911	0.949	159.180	2173
50	100	0.938	0.957	163.199	1987
50	200	0.950	0.960	161.316	1655
100	10	0.679	0.896	184.589	3817
100	30	0.931	0.962	183.717	3495
100	50	0.954	0.970	183.428	3198
100	100	0.968	0.976	180.592	2640
100	200	0.975	0.979	177.441	2154
200	10	0.771	0.919	195.511	4556
200	30	0.958	0.971	195.709	4393
200	50	0.973	0.980	193.661	3946
200	100	0.982	0.985	189.554	3244
200	200	0.986	0.988	188.616	2991

Table 6: Average correlation coefficients between $\{F_t\}_{t=1}^T$ and $\{\hat{F}_{Song,t}\}_{t=1}^T$ and $\{\hat{F}_{AMG,t}\}_{t=1}^T$ in the case of heterogeneous slope. *Iter* and *Fail* indicate the average number of iterations and the number of failures (lack of convergence) of the iterative process for the estimation method of Song (2013), respectively.

		Inf MG		CCEMG		AMG		Song	
n	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.0285	0.0871	0.0336	0.0864	0.0313	0.0901	0.0309	0.0892
10	30	-0.0104	0.0009	-0.0082	0.0019	-0.0083	0.0003	-0.0088	0.0019
10	50	-0.0299	0.0066	-0.0307	0.0066	-0.0311	0.0067	-0.0314	0.0068
10	100	-0.0153	-0.0186	-0.0172	-0.0185	-0.0161	-0.0192	-0.0175	-0.0178
10	200	-0.0274	-0.0132	-0.0254	-0.0134	-0.0267	-0.0134	-0.0245	-0.0135
30	10	0.0001	-0.0200	-0.0016	-0.0235	-0.0015	-0.0204	-0.0004	-0.0206
30	30	0.0186	-0.0199	0.0191	-0.0194	0.0190	-0.0197	0.0197	-0.0199
30	50	-0.0350	0.0174	-0.0353	0.0175	-0.0352	0.0177	-0.0354	0.0180
30	100	-0.0130	-0.0109	-0.0131	-0.0102	-0.0128	-0.0111	-0.0132	-0.0099
30	200	-0.0472	-0.0107	-0.0465	-0.0105	-0.0474	-0.0107	-0.0463	-0.0107
50	10	0.0296	-0.0404	0.0315	-0.0383	0.0318	-0.0407	0.0304	-0.0398
50	30	-0.0335	0.0002	-0.0328	-0.0008	-0.0330	-0.0001	-0.0326	-0.0007
50	50	-0.0070	0.0288	-0.0065	0.0288	-0.0068	0.0289	-0.0061	0.0291
50	100	-0.0475	0.0016	-0.0467	0.0021	-0.0476	0.0018	-0.0468	0.0023
50	200	-0.0057	-0.0136	-0.0056	-0.0138	-0.0059	-0.0136	-0.0051	-0.0140
100	10	-0.0155	-0.0312	-0.0164	-0.0295	-0.0160	-0.0309	-0.0145	-0.0295
100	30	-0.0003	-0.0219	-0.0007	-0.0221	-0.0004	-0.0217	-0.0002	-0.0217
100	50	0.0006	-0.0039	0.0004	-0.0043	0.0004	-0.0042	0.0003	-0.0046
100	100	-0.0146	-0.0024	-0.0151	-0.0022	-0.0146	-0.0024	-0.0149	-0.0019
100	200	-0.0234	-0.0093	-0.0236	-0.0097	-0.0234	-0.0093	-0.0237	-0.0101
200	10	0.0171	-0.0082	0.0179	-0.0092	0.0175	-0.0084	0.0182	-0.0090
200	30	0.0061	0.0198	0.0057	0.0200	0.0060	0.0198	0.0062	0.0199
200	50	-0.0073	-0.0018	-0.0074	-0.0018	-0.0072	-0.0017	-0.0074	-0.0022
200	100	0.0238	-0.0008	0.0241	-0.0005	0.0239	-0.0007	0.0241	-0.0005
200	200	0.0249	0.0082	0.0253	0.0082	0.0250	0.0083	0.0256	0.0083

Table 7: Bias of β_i estimators in the case of heterogeneous slope with unobserved common factors only. The DGP is in (20) and (21). *Inf MG* is the mean group estimators obtained with unobserved common factors. *CCEMG* is the mean group CCE estimator by Pesaran (2006). *AMG* is the mean group estimator (8). *Song* is the Interactive Fixed Effects estimator by Bai (2009) extended to the heterogenous case by Song (2013).

		Inf MG		CCEMG		AMG		Song	
n	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.2797	0.2861	0.3646	0.3596	0.3014	0.2985	0.2562	0.2550
10	30	0.1283	0.1265	0.1484	0.1456	0.1407	0.1378	0.1541	0.1493
10	50	0.0934	0.0930	0.1124	0.1153	0.1048	0.1056	0.1284	0.1306
10	100	0.0669	0.0648	0.0900	0.0897	0.0752	0.0745	0.1135	0.1135
10	200	0.0447	0.0460	0.0796	0.0794	0.0544	0.0546	0.1071	0.1063
30	10	0.1630	0.1624	0.1998	0.2005	0.1644	0.1653	0.1462	0.1437
30	30	0.0728	0.0732	0.0809	0.0818	0.0756	0.0763	0.0855	0.0872
30	50	0.0543	0.0543	0.0639	0.0632	0.0570	0.0564	0.0743	0.0740
30	100	0.0378	0.0375	0.0528	0.0511	0.0400	0.0395	0.0677	0.0659
30	200	0.0263	0.0262	0.0441	0.0436	0.0281	0.0278	0.0608	0.0603
50	10	0.1262	0.1258	0.1574	0.1560	0.1285	0.1277	0.1137	0.1132
50	30	0.0555	0.0565	0.0623	0.0613	0.0570	0.0579	0.0667	0.0661
50	50	0.0424	0.0421	0.0494	0.0497	0.0439	0.0435	0.0580	0.0582
50	100	0.0291	0.0295	0.0400	0.0392	0.0301	0.0301	0.0513	0.0506
50	200	0.0201	0.0198	0.0341	0.0341	0.0210	0.0206	0.0478	0.0476
100	10	0.0892	0.0906	0.1095	0.1111	0.0896	0.0916	0.0812	0.0804
100	30	0.0401	0.0401	0.0433	0.0436	0.0407	0.0405	0.0467	0.0466
100	50	0.0295	0.0294	0.0347	0.0348	0.0299	0.0300	0.0410	0.0409
100	100	0.0206	0.0206	0.0274	0.0283	0.0209	0.0209	0.0353	0.0365
100	200	0.0144	0.0143	0.0244	0.0239	0.0146	0.0146	0.0338	0.0335
200	10	0.0639	0.0636	0.0769	0.0778	0.0640	0.0631	0.0570	0.0570
200	30	0.0283	0.0284	0.0315	0.0305	0.0285	0.0286	0.0337	0.0328
200	50	0.0208	0.0206	0.0244	0.0242	0.0210	0.0208	0.0288	0.0285
200	100	0.0147	0.0147	0.0196	0.0199	0.0148	0.0148	0.0257	0.0258
200	200	0.0100	0.0101	0.0163	0.0170	0.0101	0.0102	0.0232	0.0242

Table 8: RMSE of β_i estimators in the case of heterogeneous slope with unobserved common factors only. The DGP is in (20) and (21). *Inf MG* is the mean group estimators obtained with unobserved common factors. *CCEMG* is the mean group CCE estimator by Pesaran (2006). *AMG* is the mean group estimator (8). *Song* is the Interactive Fixed Effects estimator by Bai (2009) extended to the heterogenous case by Song (2013).

n	T	$\text{Corr}(F_t, \hat{F}_{Song,t})$	$\text{Corr}(F_t, \hat{F}_{AMG,t})$	iter	fail
10	10	0.520	0.718	200	5000
10	30	0.627	0.773	200	5000
10	50	0.675	0.792	200	5000
10	100	0.743	0.810	200	5000
10	200	0.783	0.817	200	5000
30	10	0.631	0.874	200	5000
30	30	0.856	0.924	200	5000
30	50	0.898	0.930	200	5000
30	100	0.922	0.935	200	5000
30	200	0.929	0.936	200	5000
50	10	0.698	0.921	200	5000
50	30	0.924	0.955	200	5000
50	50	0.944	0.958	200	5000
50	100	0.954	0.961	200	5000
50	200	0.957	0.961	200	5000
100	10	0.815	0.959	200	5000
100	30	0.965	0.977	200	5000
100	50	0.973	0.979	200	5000
100	100	0.977	0.980	200	5000
100	200	0.979	0.981	200	5000
200	10	0.916	0.981	200	5000
200	30	0.983	0.989	200	5000
200	50	0.986	0.990	200	5000
200	100	0.988	0.990	200	5000
200	200	0.989	0.990	200	5000

Table 9: Average correlation coefficients between $\{F_t\}_{t=1}^T$ and $\{\hat{F}_{Song,t}\}_{t=1}^T$ and $\{\hat{F}_{AMG,t}\}_{t=1}^T$ in the case of heterogeneous slope. *Iter* and *Fail* indicate the average number of iterations and the number of failures (lack of convergence) of the iterative process for the estimation method of Song (2013), respectively.

n	T	$\hat{r}_{AMG}$	$\hat{r}_{Song}$	iter	fail
10	10	3.62	2.35	171.622	3522
10	30	3.55	1.57	108.628	1236
10	50	3.60	1.52	96.416	810
10	100	3.60	1.46	80.605	403
10	200	3.51	1.43	72.032	222
30	10	3.58	1.64	170.493	3067
30	30	2.00	1.08	94.768	369
30	50	2.00	1.39	86.301	159
30	100	2.00	1.75	79.513	41
30	200	2.00	1.90	74.569	0
50	10	3.38	1.29	176.557	3160
50	30	2.00	1.34	108.307	361
50	50	2.00	1.64	95.829	126
50	100	2.00	1.97	84.830	8
50	200	2.00	2.00	79.043	10
100	10	2.74	1.04	186.556	3661
100	30	2.00	1.83	127.252	374
100	50	2.00	2.00	105.139	41
100	100	2.00	2.00	90.936	2
100	200	2.00	2.00	85.241	0
200	10	2.06	1.00	195.032	4338
200	30	2.00	2.00	137.876	337
200	50	2.00	2.00	113.399	32
200	100	2.00	2.00	98.117	2
200	200	2.00	2.00	91.997	2

Table 10: Average number of factors estimated by AMG and Song in the case of two unknown common factors. *Iter* and *Fail* indicate the average number of iterations and the number of failures (lack of convergence) of the iterative process for the estimation method of Song (2013), respectively.

		Inf MG		CCEMG		AMG		Song	
n	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.008	0.032	0.014	0.033	0.013	0.033	0.012	0.032
10	30	0.031	-0.070	0.029	-0.072	0.029	-0.071	0.030	-0.072
10	50	0.123	0.075	0.125	0.074	0.123	0.074	0.124	0.075
10	100	-0.024	-0.013	-0.023	-0.011	-0.024	-0.012	-0.025	-0.012
10	200	-0.063	-0.003	-0.063	-0.002	-0.063	-0.002	-0.063	-0.003
30	10	-0.032	-0.014	-0.031	-0.015	-0.031	-0.012	-0.031	-0.014
30	30	0.047	0.059	0.047	0.058	0.047	0.059	0.048	0.058
30	50	-0.026	-0.044	-0.025	-0.044	-0.026	-0.044	-0.025	-0.044
30	100	-0.018	-0.029	-0.019	-0.030	-0.018	-0.029	-0.017	-0.029
30	200	-0.001	0.046	-0.002	0.045	-0.002	0.046	-0.002	0.046
50	10	0.053	-0.005	0.055	-0.005	0.054	-0.007	0.054	-0.006
50	30	-0.027	0.010	-0.026	0.010	-0.027	0.010	-0.027	0.010
50	50	-0.013	0.022	-0.013	0.023	-0.013	0.022	-0.013	0.022
50	100	-0.008	0.012	-0.008	0.012	-0.008	0.012	-0.008	0.012
50	200	0.007	-0.062	0.007	-0.061	0.007	-0.062	0.007	-0.062
100	10	-0.012	-0.002	-0.011	-0.002	-0.012	-0.003	-0.012	-0.003
100	30	0.001	0.016	0.001	0.015	0.001	0.015	0.002	0.015
100	50	0.023	0.026	0.022	0.026	0.023	0.026	0.023	0.026
100	100	0.022	-0.022	0.022	-0.022	0.022	-0.022	0.022	-0.022
100	200	0.043	0.013	0.043	0.013	0.043	0.013	0.043	0.013
200	10	0.002	0.009	0.003	0.008	0.003	0.009	0.003	0.009
200	30	-0.009	0.006	-0.009	0.006	-0.009	0.006	-0.009	0.006
200	50	-0.015	-0.006	-0.014	-0.006	-0.015	-0.006	-0.015	-0.006
200	100	-0.006	0.018	-0.006	0.018	-0.006	0.018	-0.006	0.018
200	200	0.005	0.005	-0.013	0.017	-0.013	0.017	-0.013	0.017

Table 11: Bias of β_i estimators in the case of heterogeneous slope when the number of common factors f_t is unknown. *Inf MG* is the mean group estimators obtained with unobserved common factors. *CCEMG* is the mean group CCE estimator by Pesaran (2006). *AMG* is the mean group estimator (8). *Song* is the Interactive Fixed Effects estimator by Bai (2009) extended to the heterogenous case by Song (2013).

		Inf MG		CCEMG		AMG		Song	
n	T	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	10	0.160	0.161	0.209	0.208	0.207	0.207	0.283	0.279
10	30	0.064	0.064	0.107	0.110	0.086	0.089	0.124	0.124
10	50	0.048	0.048	0.098	0.097	0.073	0.072	0.098	0.098
10	100	0.033	0.033	0.092	0.091	0.061	0.061	0.081	0.081
10	200	0.023	0.023	0.086	0.088	0.053	0.054	0.072	0.074
30	10	0.092	0.092	0.117	0.122	0.116	0.119	0.124	0.127
30	30	0.037	0.037	0.062	0.061	0.039	0.039	0.064	0.063
30	50	0.028	0.028	0.055	0.055	0.029	0.029	0.049	0.049
30	100	0.019	0.019	0.050	0.051	0.020	0.020	0.032	0.032
30	200	0.013	0.013	0.049	0.049	0.015	0.015	0.020	0.020
50	10	0.071	0.071	0.093	0.092	0.088	0.087	0.085	0.086
50	30	0.029	0.028	0.048	0.048	0.030	0.029	0.044	0.043
50	50	0.022	0.021	0.043	0.041	0.022	0.022	0.032	0.033
50	100	0.014	0.014	0.039	0.039	0.015	0.015	0.016	0.016
50	200	0.010	0.010	0.038	0.038	0.010	0.010	0.010	0.010
100	10	0.050	0.050	0.063	0.065	0.058	0.057	0.055	0.057
100	30	0.021	0.020	0.033	0.033	0.021	0.021	0.024	0.024
100	50	0.015	0.015	0.030	0.030	0.015	0.015	0.015	0.016
100	100	0.010	0.010	0.028	0.028	0.010	0.011	0.010	0.011
100	200	0.007	0.007	0.026	0.026	0.007	0.007	0.007	0.007
200	10	0.035	0.036	0.046	0.046	0.036	0.036	0.038	0.038
200	30	0.015	0.015	0.023	0.024	0.015	0.015	0.015	0.015
200	50	0.011	0.011	0.021	0.021	0.011	0.011	0.011	0.011
200	100	0.007	0.007	0.019	0.019	0.007	0.007	0.007	0.007
200	200	0.005	0.005	0.019	0.019	0.005	0.005	0.005	0.005

Table 12: RMSE of β_i estimators in the case of heterogeneous slope when the number of common factors f_t is unknown. *Inf MG* is the mean group estimators obtained with unobserved common factors. *CCEMG* is the mean group CCE estimator by Pesaran (2006). *AMG* is the mean group estimator (8). *Song* is the Interactive Fixed Effects estimator by Bai (2009) extended to the heterogenous case by Song (2013).

References

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

——— (2009): "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77(4), 1229–1279.

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

Bernanke, B., J. Boivin, and P. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 120(1), 387–422.

Castagnetti, C., E. Rossi, and L. Trapani (2012): "Inference on Factor Structures in Heterogeneous Panels," quaderno di dipartimento, n.211, Università di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi.

Chudik, A., M. Pesaran, and E. Tosetti (2011): "Weak and strong cross-section dependence and estimation of large panels," *Econometrics Journal*, 14(1), 45–90.

Coakley, J., A. Fuertes, and R. Smith (2002): "A Principal Component Approach to Cross-Section Dependence in Panels," Discussion paper, Birkbeck College.

Kapetanios, G., and M. Pesaran (2007): "Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns," in *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, ed. by G. Phillips, and E. Tzavalis. Cambridge University Press, Cambridge.

Kapetanios, G., M. Pesaran, and T. Yamagata (2011): "Panels with nonstationary multifactor error structures," *Journal of Econometrics*, 160(2), 326–348.

Lin, Z., and Z. Bai (2010): *Probability Inequalities.* Science Press and Springer-Verlag.

Ludvigson, S., and S. Ng (2007): "The Empirical Risk-Return Relation: A Factor Analysis Approach," *Journal of Financial Economics*, 83, 171–222.

——— (2009): "Macro Factors in Bond Risk Premia," *The Review of Financial Studies*, 22(12), 5027–5067.

Pesaran, M. (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.

Pesaran, M., and E. Tosetti (2011): "Large panels with common factors and spatial correlation," *Journal of Econometrics*, 161(2), 182–202.

Rao, P. B. (2009): "Conditional independence, conditional mixing and conditional association," *Annals of the Institute of Statistical Mathematics*, 61, 441–460.

Song, M. (2013): "Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications," Mimeo.

Stock, J., and M. Watson (2005): "Implication of Dynamic Factor Models for VAR Analysis," working paper 11467.