Large Panel Data Models with Cross-Sectional Dependence: A Surevey

Hashem Pesaran
University of Southern California, CAFE, USA, and Trinity College, Cambridge, UK

A Course on Panel Data Models, University of Cambridge, 29 May 2013

Introduction

- Early panel data literature assumed cross sectionally independent errors and slope homogeneity; and heterogeneity across units was modelled by using unit-specific intercepts only, treated as fixed or random.
- Cross-sectional error dependence was only considered in spatial models, but not in standard panels. However, with an increasing availability of data (across countries, regions, or industries), panel literature moved from predominantly micro panels, where the cross section dimension (N) is large and the time series dimension (T) is small, to models with both N and T large, and it has been recognized that, even after conditioning on unit-specific regressors, individual units, in general, need not be cross-sectionally independent.

- Ignoring error cross-sectional dependence can have serious consequences, and the presence of some form of cross-sectional correlation of errors in panel data applications in economics is likely to be the rule rather than the exception. Cross correlations of errors could be due to omitted common effects, spatial effects, or could arise as a result of interactions within socioeconomic networks.
- Conventional panel estimators such as fixed or random effects can result in misleading inference and even inconsistent estimators, depending on the extent of cross-sectional dependence and on whether the source generating the cross-sectional dependence (such as an unobserved common shock) is correlated with regressors.
- The problem of testing for the extent of cross-sectional correlation of panel residuals and modelling the cross-sectional dependence of errors are therefore important issues.

Outline

- Types of cross-sectional dependence
- Modeling cross-sectional dependence by a multi-factor error structure
- Estimation and inference on large panels with strictly exogenous regressors and a factor error structure
- Estimation and inference on large dynamic panel data models with a factor error structure
- Tests of error cross-sectional dependence

Types of cross-sectional (CS) dependence

- Let $\{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ be a double index process defined on a suitable probability space and assume:

$$E(\mathbf{x}_t) = \mathbf{0},\ Var(\mathbf{x}_t) = \mathbf{\Sigma}_x,$$

  where $\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{Nt})'$ and the elements of $\mathbf{\Sigma}_x$, denoted as $\sigma_{x,ij}$ for $i, j = 1, 2, ..., N$, are uniformly bounded in N, namely $|\sigma_{x,ij}| < K$.
- These assumptions could be relaxed, by considering conditional expectations and variances, nonzero time-varying means, and time-varying variances.
- Various summary measures of the matrix $\mathbf{\Sigma}_x$ have been considered in the literature.
- The largest eigenvalue of $\mathbf{\Sigma}_x$, denoted as $\lambda_1(\mathbf{\Sigma}_x)$, has received a great deal of attention in the literature, but $\lambda_1(\mathbf{\Sigma}_x)$ is difficult to estimate when the cross section dimension, N, is large compared to the time dimension, T.

- Chudik, Pesaran and Tosetti (2011) summarize the extent of CS correlations based on the behavior of CS averages. Let $\bar{x}_{wt} = \sum_{i=1}^{N} w_i x_{it}$, where the weights $\mathbf{w} = (w_1, w_2, ..., w_N)'$ satisfy the following 'granularity' conditions:

$$\|\mathbf{w}\| = \sqrt{\mathbf{w}'\mathbf{w}} = O\left(N^{-1/2}\right), \text{ and}$$

$$\frac{w_i}{\|\mathbf{w}\|} = O\left(N^{-1/2}\right) \text{ uniformly in } i \in \mathbb{N}.$$

- $\{x_{it}\}$ is cross-sectionally **weakly dependent** (CWD) if for any sequence of granular weights $\mathbf{w}$, we have

$$\lim_{N\to\infty} Var\left(\bar{x}_{wt}\right) = 0.$$

  Otherwise, $\{x_{it}\}$ is **cross-sectionally strongly dependent** (CSD).

- Bailey, Kapetanios and Pesaran (2012) characterize the pattern of cross-sectional dependence further. Let $w_i = N^{-1}$, for all i, and consider

$$Var\left(\bar{x}_t\right) = Var\left(\frac{1}{N}\sum_{i=1}^{N} x_{it}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sigma_{x,ii} + \underbrace{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sigma_{x,ij}}_{\kappa_{x,N}},$$

  where $\left|\sigma_{x,ij}\right| < K$ and therefore $0 \leq Var\left(N\bar{x}_t\right) < KN^2$.
- The extent of cross-sectional dependence relates directly to the term, κ_N, and Bailey et al. parametrize this term by an **exponent of cross-sectional dependence** $\alpha \in [0,1]$ that satisfies

$$\lim_{N\to\infty} N^{2-2\alpha}\kappa_{x,N} = K \text{ for some constant } 0 < K < \infty.$$

Spatial examples

- Leading examples of a cross sectionally dependent processes are factor models or spatial models.
- Spatial models of the error vector $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$ can be written as

$$\mathbf{u}_t = \mathbf{R}\varepsilon_t,\ \varepsilon_t \sim (0, \mathbf{I}_N)$$

- For instance, $\mathbf{R} = (\mathbf{I}_N - \rho\mathbf{W})^{-1}\,\tilde{}^{1/2}$, in the first order spatial autoregressive model $\mathbf{u}_t = \rho\mathbf{W}\mathbf{u}_t + \tilde{}^{1/2}\varepsilon_t$, where $\tilde{}$ is a diagonal matrix. It is easy to see that $\mathbf{u}_t$ is CWD when row and column matrix norms of $\mathbf{R}$ are both bounded.

Modelling cross-sectional dependence by a factor error structure

- Consider the m factor model for $\{z_{it}\}$

$$z_{it} = \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + ... + \gamma_{im} f_{mt} + e_{it}, \quad i = 1, 2, ..., N,$$

or, in matrix notations

$$\mathbf{z}_t = \mathbf{\Gamma} \mathbf{f}_t + \mathbf{e}_t, \tag{1}$$

where $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})'$, $\mathbf{e}_t = (e_{1t}, e_{2t}, ..., e_{Nt})'$, and $\mathbf{\Gamma} = (\gamma_{ij})$, for $i = 1, 2, ..., N$, $j = 1, 2, ..., m$, is an $N \times m$ matrix of fixed coefficients, known as factor loadings.

- The common factors, $\mathbf{f}_t$, simultaneously affect all cross section units, albeit with different degrees as measured by $\gamma_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{im})'$.
- Examples of observed common factors that tend to affect all households and firms consumption and investment decisions include interest rates and oil prices. Aggregate demand and supply shocks represent examples of common unobserved factors.
- In multifactor models interdependence arises from common correlated *reaction* of units to some external events. Further, according to this representation, correlation between any pair of units does not depend on how far these observations are apart, and violates the distance decay effect that underlies the spatial interaction model.

Assumptions of exact factor model

- The following assumptions are typically made regarding the common factors, $f_{\ell t}$, and the idiosyncratic errors, e_{it}.
  - ASSUMPTION CF.1: The $m \times 1$ vector $\mathbf{f}_t$ is a zero mean covariance stationary process, with absolute summable autocovariances, distributed independently of $e_{it'}$ for all i, t, t', such that $E(f_{\ell t}^2 \,|\Omega_{t-1}) = 1$ and $E(f_{\ell t} f_{pt} \,|\Omega_{t-1}) = 0$, for $\ell \neq p = 1, 2, ..., m$.
  - ASSUMPTION CF.2: $Var\left(e_{it} \,|\Omega_{t-1}\right) = \sigma_i^2 < K < \infty$, e_{it} and e_{jt} are independently distributed for all $i \neq j$ and for all t. Specifically, $\max_i \left(\sigma_i^2\right) = \sigma_{\max}^2 < K < \infty$.
- Assumption CF.1 is an identification condition, since it is not possible to separately identify $\mathbf{f}_t$ and $\mathbf{\Gamma}$. Under the above assumptions, the covariance of $\mathbf{z}_t$ conditional on Ω_{t-1} is given by

$$E\left(\mathbf{z}_t \mathbf{z}_t' \,|\Omega_{t-1}\right) = \mathbf{\Gamma}\mathbf{\Gamma}' + \mathbf{V},$$

  where $\mathbf{V}$ is a diagonal matrix with elements σ_i^2 on the main diagonal.

Approximate factor models

- The assumption that the idiosyncratic errors, e_{it}, are cross-sectionally independent is not necessary and can be relaxed. The factor model that allows the idiosyncratic shocks, e_{it}, to be cross-sectionally weakly correlated is known as the approximate factor model. See Chamberlein (1983).
- In general, the correlation patterns of the idiosyncratic errors can be characterized by

$$\mathbf{e}_t = \mathbf{R}\varepsilon_t,$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})' \sim (\mathbf{0}, \mathbf{I}_N)$. In the case of this formulation $\mathbf{V} = \mathbf{R}\mathbf{R}'$, which is no longer diagonal, and further identification restrictions are needed.
- To this end it is typically assumed that the matrix $\mathbf{R}$ has bounded row and column sum matrix norms (so that the cross-sectional dependence of $\mathbf{e}_t$ is sufficiently weak) and the factor loadings are such that $\lim_{N\to\infty}(N^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Gamma})$ is a full rank matrix.

Strong and weak common factors

- To ensure that the factor component of (1) represents strong cross-sectional dependence (so that it can be distinguished from the idiosyncratic errors) it is sufficient that the absolute column sum matrix norm of $\|\boldsymbol{\Gamma}\|_1 = \max_{i\in\{1,2,\ldots,N\}} \sum_{j=1}^{N} |\gamma_{ij}|$ rises with N at the rate N, which is necessary for $\lim_{N\to\infty}(N^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Gamma})$ to be a full rank matrix, as required earlier.
- The factor $f_{\ell t}$ is said to be strong if

$$\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} |\gamma_{i\ell}| = K > 0.$$

- The factor $f_{\ell t}$ is said to be weak if

$$\lim_{N\to\infty} \sum_{i=1}^{N} |\gamma_{i\ell}| = K < \infty.$$

Intermediate cases

- It is also possible to consider intermediate cases of semi-weak or semi-strong factors. In general, let α_ℓ be a positive constant in the range $0 \leq \alpha_\ell \leq 1$ and consider the condition

$$\lim_{N \to \infty} N^{-\alpha_\ell} \sum_{i=1}^{N} |\gamma_{i\ell}| = K < \infty. \tag{2}$$

- Strong and weak factors correspond to the two values of $\alpha_\ell = 1$ and $\alpha_\ell = 0$, respectively. For any other values of $\alpha_\ell \in (0, 1)$ the factor $f_{\ell t}$ can be said to be semi-strong or semi-weak. It will prove useful to associate the semi-weak factors with values of $0 < \alpha_\ell < 1/2$, and the semi-strong factors with values of $1/2 \leq \alpha_\ell < 1$. In a multi-factor set up the overall exponent of cross-sectional dependence can be defined by $\alpha = \max(\alpha_1, \alpha_2, ..., a_m)$.

- The relationship between the notions of CSD and CWD and the definitions of weak and strong factors are explored in the following theorem.

Theorem 1

Consider the factor model (1) and suppose that Assumptions CF.1-CF.2 hold, and there exists a positive constant $\alpha = \max(\alpha_1, \alpha_2, ..., a_m)$ *in the range* $0 \leq \alpha \leq 1$, *such that condition (2) is met for any* $\ell = 1, 2, .., m$. *Then the following statements hold:*

(i) *The process* $\{z_{it}\}$ *is cross-sectionally weakly dependent at a given point in time* $t \in \mathcal{T}$ *if* $\alpha < 1$, *which includes cases of weak, semi-weak or semi-strong factors,* $f_{\ell t}$, *for* $\ell = 1, 2, ..., m$.

(ii) *The process* $\{z_{it}\}$ *is cross-sectionally strongly dependent at a given point in time* $t \in \mathcal{T}$ *if and only if there exists at least one strong factor.*

- Consistent estimation of factor models with weak or semi-strong factors may be problematic, as evident from the following example.

Example 2

Consider the following single factor model with known factor loadings

$$z_{it} = \gamma_i f_t + \varepsilon_{it}, \quad \varepsilon_{it} \sim IID\left(0, \sigma^2\right).$$

The least squares estimator of f_t, which is the best linear unbiased estimator, is given by

$$\hat{f}_t = \frac{\sum_{i=1}^{N} \gamma_i z_{it}}{\sum_{i=1}^{N} \gamma_i^2}, \quad Var\left(\hat{f}_t\right) = \frac{\sigma^2}{\sum_{i=1}^{N} \gamma_i^2}.$$

If for example $\sum_{i=1}^{N} \gamma_i^2$ is bounded, as in the case of weak factors, then $Var\left(\hat{f}_t\right)$ does not vanish as $N \to \infty$, for each t. See also Onatski (2012).

- Weak, strong and semi-strong common factors may be used to represent very general forms of cross-sectional dependence.

Estimation and inference on large panels with strictly exogenous regressors and a factor error structure

- Consider the following heterogeneous panel data model

$$y_{it} = \boldsymbol{\alpha}_i' \mathbf{d}_t + \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}, \tag{3}$$

where $\mathbf{d}_t$ is a $n \times 1$ vector of observed common effects, $\mathbf{x}_{it}$ is a $k \times 1$ vector of observed individual-specific regressors on the ith cross-section unit at time t, and disturbances, u_{it}, have the following common factor structure

$$u_{it} = \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + ... + \gamma_{im} f_{mt} + e_{it} = \boldsymbol{\gamma}_i' \mathbf{f}_t + e_{it}, \tag{4}$$

in which $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})'$ is an m-dimensional vector of unobservable common factors, and $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{im})'$ is the associated $m \times 1$ vector of factor loadings. The number of factors, m, is assumed to be fixed relative to N, and in particular $m << N$.

- The idiosyncratic errors, e_{it}, could be CWD.
- The factor loadings, γ_i, could be either considered draws from a random distribution, or fixed unknown coefficients.
- We distinguish between the homogenous coefficient case where $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i, and the heterogenous case where $\boldsymbol{\beta}_i$ are random draws from a given distribution. In the latter case, we assume that the object of interest is the mean coefficients $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$.
- When the regressors, $\mathbf{x}_{it}$, are strictly exogenous and the deviations $\boldsymbol{v}_i = \boldsymbol{\beta}_i - \boldsymbol{\beta}$ are distributed independently of the errors and the regressors, the mean coefficients, $\boldsymbol{\beta}$, can be consistently estimated using pooled as well as mean group estimation procedures. But only mean group estimation will be consistent if the regressors are weakly exogenous and/or if the deviations are correlated with the regressors/errors.

- The assumption of slope homogeneity is also crucially important for the derivation of the asymptotic distribution of the pooled or the mean group estimators of $\boldsymbol{\beta}$. Under slope homogeneity the asymptotic distribution of the estimator of $\boldsymbol{\beta}$ typically converges at the rate of $\sqrt{NT}$, whilst under slope heterogeneity the rate is $\sqrt{N}$.
- We review the following estimators:
  - The **Principal Components (PC) approach** proposed by Coakley, Fuertes and Smith (2002) and Bai (2008)
  - The **Common Correlated Effects (CCE) approach** proposed by Pesaran (2006) and extended by Kapetanios, Pesaran and Yagamata (2011), Pesaran and Tosetti (2011) and Chudik, Pesaran and Tosetti (2011).

Principal components estimators

- PC approach implicitly assumes that all the unobserved common factors are strong by requiring that $N^{-1}\mathbf{\Gamma}'\mathbf{\Gamma}$ tends to a positive definite matrix
- Coakley, Fuertes and Smith (2002) consider the panel data model with strictly exogenous regressors and homogeneous slopes (i.e., $\boldsymbol{\beta}_i = \boldsymbol{\beta}$), and propose a two-stage estimation procedure:
    1. PCs are extracted from the OLS residuals as proxies for the unobserved variables.
    2. The following augmented regression is estimated

    $$y_{it} = \boldsymbol{\alpha}_i'\mathbf{d}_t + \boldsymbol{\beta}'\mathbf{x}_{it} + \gamma_i'\hat{\mathbf{f}}_t + \varepsilon_{it}, \text{ for } i = 1, 2, ..., N; \ t = 1, 2, ..., T, \quad (5)$$

    where $\hat{\mathbf{f}}_t$ is an $m \times 1$ vector of principal components of the residuals computed in the first stage.

- The resultant estimator is consistent for N and T large, but only when $\mathbf{f}_t$ and the regressors, $\mathbf{x}_{it}$, are uncorrelated.
- Bai (2008) has proposed an iterative method which consists of alternating the PC method applied to OLS residuals and the least squares estimation of (5), until convergence. In particular, to simplify the exposition suppose $\boldsymbol{\alpha}_i = \mathbf{0}$. Then the least squares estimator of $\boldsymbol{\beta}$ and $\mathbf{F}$ is the solution of:

$$\hat{\boldsymbol{\beta}}_{PC} = \left(\sum_{i=1}^{N} \mathbf{X}_i \mathbf{M}_{\hat{F}} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i \mathbf{M}_{\hat{F}} \mathbf{y}_i,$$

$$\frac{1}{NT} \sum_{i=1}^{N} \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{PC} \right) \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{PC} \right)' \hat{\mathbf{F}} = \hat{\mathbf{F}} \hat{\mathbf{V}},$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})'$, $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$, $\mathbf{M}_{\hat{F}} = \mathbf{I}_T - \hat{\mathbf{F}} \left(\hat{\mathbf{F}} \hat{\mathbf{F}}' \right)^{-1} \hat{\mathbf{F}}'$, $\hat{\mathbf{F}} = \left(\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, ..., \hat{\mathbf{f}}_T \right)'$, and $\hat{\mathbf{V}}$ is a diagonal matrix with the m largest eigenvalues of the matrix $\sum_{i=1}^{N} \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{PC} \right) \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{PC} \right)'$ arranged in a decreasing order.

- The solution $\hat{\beta}_{PC}$, $\hat{\mathbf{F}}$ and $\hat{\gamma}_i = \left(\hat{\mathbf{F}}'\hat{\mathbf{F}}\right)^{-1} \hat{\mathbf{F}}' \left(\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_{PC}\right)$ minimizes the sum of squared residuals function,

$$SSR_{NT} = \sum_{i=1}^{N} \left(\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{F}\gamma_i\right)' \left(\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{F}\gamma_i\right),$$

  where $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$.
- This function is a Gaussian quasi maximum likelihood function of the model and in this respect, Bai's iterative principal components estimator can also be seen as a quasi maximum likelihood estimator, since it minimizes the quasi likelihood function.

- Bai (2008) shows that such an estimator is consistent even if common factors are correlated with the explanatory variables. Specifically, the least square estimator of $\boldsymbol{\beta}$ obtained from the above procedure, $\hat{\boldsymbol{\beta}}_{PC}$, is consistent if both N and T go to infinity, without any restrictions on the ratio T/N. When in addition $T/N \to K > 0$, $\hat{\boldsymbol{\beta}}_{PC}$ converges at the rate $\sqrt{NT}$, but the limiting distribution of $\sqrt{NT}\left(\hat{\boldsymbol{\beta}}_{PC} - \hat{\boldsymbol{\beta}}\right)$ does not necessarily have a zero mean. Nevertheless, Bai shows that the asymptotic bias can be consistently estimated and proposes a bias corrected estimator.
- A shortcoming of the iterative PC estimator is that it requires the determination of the unknown number of factors (PCs) to be included in the second stage, since estimation of m can introduce a certain degree of sampling uncertainty into the analysis.

Common Correlated Effects estimators

- Pesaran (2006) suggests the CCE approach, which consists of approximating the linear combinations of the unobserved factors by cross section averages of the dependent and explanatory variables, and then running standard panel regressions augmented with these cross section averages.
- Both pooled and mean group versions are proposed, depending on the assumption regarding the slope homogeneity.
- Under slope heterogeneity the CCE approach assumes that $\boldsymbol{\beta}_i's$ follow the random coefficient model

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i, \quad \boldsymbol{v}_i \sim IID(\mathbf{0}, \boldsymbol{\Omega}_v) \quad \text{for } i = 1, 2, ..., N,$$

  where the deviations, $\boldsymbol{v}_i$, are distributed independently of $e_{jt}, \mathbf{x}_{jt}$, and $\mathbf{d}_t$, for all i, j and t.

- The following model for the individual-specific regressors in (3) is adopted

$$\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{d}_t + {}_i'\mathbf{f}_t + \mathbf{v}_{it}, \tag{6}$$

where $\mathbf{A}_i$ and ${}_i$ are $n \times k$ and $m \times k$ factor loading matrices with fixed components, $\mathbf{v}_{it}$ is the idiosyncratic component of $\mathbf{x}_{it}$ distributed independently of the common effects $\mathbf{f}_{t'}$ and errors $e_{jt'}$ for all i, j, t and t'. However, $\mathbf{v}_{it}$ is allowed to be serially correlated, and cross-sectionally weakly correlated.

- Equations (3), (4) and (6) can be combined into the following system of equations

$$\mathbf{z}_{it} = \left(y_{it}, \mathbf{x}_{it}'\right)' = \mathbf{B}_i'\mathbf{d}_t + \mathbf{C}_i'\mathbf{f}_t + \boldsymbol{\xi}_{it},$$

where $\boldsymbol{\xi}_{it} = \left(e_{it} + \boldsymbol{\beta}_i'\mathbf{v}_{it}, \mathbf{v}_{it}'\right)'$,

$$\mathbf{B}_i = \begin{pmatrix} \boldsymbol{\alpha}_i & \mathbf{A}_i \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}, \ \mathbf{C}_i = \begin{pmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Gamma}_i \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}.$$

- Consider the weighted average of $\mathbf{z}_{it}$, $\bar{\mathbf{z}}_{wt} = \sum_{i=1}^{N} w_i \mathbf{z}_{it}$, using the weights w_i satisfying the granularity conditions:

$$\bar{\mathbf{z}}_{wt} = \bar{\mathbf{B}}_w \mathbf{d}_t + \bar{\mathbf{C}}_w \mathbf{f}_t + \bar{\boldsymbol{\xi}}_{wt},$$

where $\bar{\mathbf{B}}_w = \sum_{i=1}^{N} w_i \mathbf{B}_i$, $\bar{\mathbf{C}}_w = \sum_{i=1}^{N} w_i \mathbf{C}_i$, $\bar{\boldsymbol{\xi}}_{wt} = \sum_{i=1}^{N} w_i \boldsymbol{\xi}_{it}$.

- Assume that $Rank(\bar{\mathbf{C}}_w) = m \leq k+1$ (this condition can be relaxed). We have

$$\mathbf{f}_t = (\bar{\mathbf{C}}_w \bar{\mathbf{C}}_w')^{-1} \bar{\mathbf{C}}_w \left(\bar{\mathbf{z}}_{wt} - \bar{\mathbf{B}}_w' \mathbf{d}_t - \bar{\boldsymbol{\xi}}_{wt} \right).$$

- Under the assumption that e_{it}'s and $\mathbf{v}_{it}$'s are CWD processes, it is possible to show that

$\bar{\boldsymbol{\xi}}_{wt} \overset{q.m.}{\rightarrow} \mathbf{0}$, which implies $\mathbf{f}_t - (\bar{\mathbf{C}}_w \bar{\mathbf{C}}_w')^{-1} \bar{\mathbf{C}}_w \left(\bar{\mathbf{z}}_{wt} - \bar{\mathbf{B}}' \mathbf{d}_t \right) \overset{q.m.}{\rightarrow} 0$,

as $N \rightarrow \infty$, where $\mathbf{C} = \lim_{N \rightarrow \infty} (\bar{\mathbf{C}}_w) = \widetilde{\boldsymbol{\Gamma}} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta} & \mathbf{I}_k \end{pmatrix}$,

$\widetilde{\boldsymbol{\Gamma}} = [E(\gamma_i), E(\boldsymbol{\Gamma}_i)]$ and $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$.

- Therefore, the unobservable common factors, $\mathbf{f}_t$, can be approximated by a linear combination of observed effects, $\mathbf{d}_t$, the cross section averages of the dependent variable, $\bar{y}_{wt}$, and those of the individual-specific regressors, $\bar{\mathbf{x}}_{wt}$.
- When the parameters of interest are the cross section means of the slope coefficients, $\boldsymbol{\beta}$, we can consider two alternative estimators, the CCE Mean Group (CCEMG) estimator and the CCE Pooled (CCEP) estimator.
- Let $\overline{\mathbf{M}}_w$ be defined by

$$\overline{\mathbf{M}}_w = \mathbf{I}_T - \overline{\mathbf{H}}_w(\overline{\mathbf{H}}'_w\overline{\mathbf{H}}_w)^{-}\overline{\mathbf{H}}'_w,$$

where $\overline{\mathbf{H}}_w = (\mathbf{D}, \overline{\mathbf{Z}}_w)$, and $\mathbf{D}$ and $\overline{\mathbf{Z}}_w$ are, respectively, the matrices of the observations on $\mathbf{d}_t$ and $\bar{\mathbf{z}}_{wt} = (\bar{y}_{wt}, \bar{\mathbf{x}}'_{wt})'$.

The CCEMG estimator

- The CCEMG is a simple average of the estimators of the individual slope coefficients

$$\hat{\boldsymbol{\beta}}_{CCEMG} = N^{-1} \sum_{i=1}^{N} \hat{\boldsymbol{\beta}}_{CCE,i},$$

where

$$\hat{\boldsymbol{\beta}}_{CCE,i} = (\mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{X}_i)^{-1} \mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{y}_i.$$

- Pesaran (2006) shows that, under some general conditions, $\hat{\boldsymbol{\beta}}_{CCEMG}$ is asymptotically unbiased for $\boldsymbol{\beta}$, and, as $(N, T) \to \infty$,

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{CCEMG} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{CCEMG}),$$

where $\boldsymbol{\Sigma}_{CCEMG} = \boldsymbol{\Omega}_v$. A consistent estimator of $\boldsymbol{\Sigma}_{CCEMG}$, can be obtained by adopting the non-parametric estimator:

$$\hat{\boldsymbol{\Sigma}}_{CCEMG} = \frac{1}{(N-1)} \sum_{i=1}^{N} (\hat{\boldsymbol{\beta}}_{CCE,i} - \hat{\boldsymbol{\beta}}_{CCEMG})(\hat{\boldsymbol{\beta}}_{CCE,i} - \hat{\boldsymbol{\beta}}_{CCEMG})'.$$

The CCEP estimator

- The CCEP estimator is given by

$$\hat{\beta}_{CCEP} = \left(\sum_{i=1}^{N} w_i \mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} w_i \mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{y}_i .$$

- Under some general conditions, Pesaran (2006) proves that $\hat{\beta}_{CCEP}$ is asymptotically unbiased for β, and, as $(N, T) \to \infty$,

$$\left(\sum_{i=1}^{N} w_i^2 \right)^{-1/2} \left(\hat{\beta}_{CCEP} - \beta \right) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{CCEP}),$$

where $\boldsymbol{\Sigma}_{CCEP} = \boldsymbol{\Psi}^{*-1} \mathbf{R}^* \boldsymbol{\Psi}^{*-1}$,

$$\boldsymbol{\Psi}^* = \lim_{N\to\infty} \left(\sum_{i=1}^{N} w_i \boldsymbol{\Sigma}_i \right), \quad \mathbf{R}^* = \lim_{N\to\infty} \left[N^{-1} \sum_{i=1}^{N} \tilde{w}_i^2 (\boldsymbol{\Sigma}_i \boldsymbol{\Omega}_v \boldsymbol{\Sigma}_i) \right],$$

$$\boldsymbol{\Sigma}_i = p \lim_{T\to\infty} \left(T^{-1} \mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{X}_i \right), \text{ and } \tilde{w}_i = \frac{w_i}{\sqrt{N^{-1} \sum_{i=1}^{N} w_i^2}}.$$

- A consistent estimator of $Var\left(\hat{\beta}_{CCEP}\right)$, denoted by $\widehat{Var}\left(\hat{\beta}_{CCEP}\right)$, is given by

$$\widehat{Var}\left(\hat{\beta}_{CCEP}\right) = \left(\sum_{i=1}^{N} w_i^2\right)^{-1} \hat{\mathbf{\Sigma}}_{CCEP} = \left(\sum_{i=1}^{N} w_i^2\right)^{-1} \hat{\mathbf{\Psi}}^{*-1}\hat{\mathbf{R}}^{*}\hat{\mathbf{\Psi}}^{*-1},$$

where

$$\hat{\mathbf{\Psi}}^{*} = \sum_{i=1}^{N} w_i \left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\right),$$

$$\hat{\mathbf{R}}^{*} = \frac{1}{N-1}\sum_{i=1}^{N} \tilde{w}_i^2 \mathbf{\Delta}_i\mathbf{\Delta}_i', \text{ where } \mathbf{\Delta}_i = \left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\right)(\hat{\beta}_{CCE,i} - \hat{\beta}_{CC$$

- The rate of convergence of $\hat{\beta}_{CCEMG}$ and $\hat{\beta}_{CCEP}$ is $\sqrt{N}$ when $\mathbf{\Omega}_v \neq \mathbf{0}$. Note that even if β_i were observed for all i, then the estimate of $\beta = E\left(\beta_i\right)$ cannot converge at a faster rate than $\sqrt{N}$. If the individual slope coefficients β_i are homogeneous (namely if $\mathbf{\Omega}_v = \mathbf{0}$), $\hat{\beta}_{CCEMG}$ and $\hat{\beta}_{CCEP}$ are still consistent and converge at the rate $\sqrt{NT}$ rather than $\sqrt{N}$.

- Advantage of the nonparametric estimators $\hat{\boldsymbol{\Sigma}}_{CCEMG}$ and $\hat{\boldsymbol{\Sigma}}_{CCEP}$ is that they do not require knowledge of the weak cross-sectional dependence of e_{it} (provided it is sufficiently weak) nor the knowledge of serial correlation of e_{it}.
- An important question is whether the non-parametric variance estimators $\widehat{Var}\left(\hat{\beta}_{CCEMG}\right) = N^{-1}\hat{\boldsymbol{\Sigma}}_{CCEMG}$ and $\widehat{Var}\left(\hat{\beta}_{CCEP}\right)$ can be used in both cases of homogenous and heterogenous slopes.
- As established in Pesaran and Tosetti (2011), the asymptotic distribution of $\hat{\beta}_{CCEMG}$ and $\hat{\beta}_{CCEP}$ depends on nuisance parameters when slopes are homogenous ($\boldsymbol{\Omega}_v = \mathbf{0}$), including the extent of cross-sectional correlations of e_{it} and their serial correlation structure.
- However, it can be shown that the robust non-parametric estimators $\widehat{Var}\left(\hat{\beta}_{CCEMG}\right)$ and $\widehat{Var}\left(\hat{\beta}_{CCEP}\right)$ are consistent when the regressor-specific components, $\mathbf{v}_{it}$, are independently distributed across i.

- The CCE continues to be applicable even if the rank condition is not satisfied. This could happen if, for example, the factor in question is weak, in the sense defined above. Another possible reason for failure of the rank condition is if the number of unobservable factors, m, is larger than $k+1$, where k is the number of the unit-specific regressors included in the model.
- In such cases, common factors cannot be estimated from cross section averages. However, the cross section means of the slope coefficients, $\boldsymbol{\beta}_i$, can still be consistently estimated, under the additional assumption that the unobserved factor loadings, γ_i, are independently and identically distributed across i, and of e_{jt}, $\mathbf{v}_{jt}$, and $\mathbf{g}_t = (\mathbf{d}_t', \mathbf{f}_t')'$ for all i, j and t. No assumptions are required on the loadings attached to the regressors, $\mathbf{x}_{it}$.
- Advantage of the CCE approach is that it does not require an a priori knowledge of the number of unobserved common factors.

- Further advantage of the CCE approach is that it yields consistent estimates under a variety of situations:
    - Kapetanios, Pesaran and Yagamata (2011) consider the case where the unobservable common factors follow unit root processes and could be cointegrated.
    - Pesaran and Tosetti (2011) prove consistency and asymptotic normality for CCE estimators when $\{e_{it}\}$ are generated by a spatial process.
    - Chudik, Pesaran and Tosetti (2011) prove consistency and asymptotic normality of the CCE estimators when errors are subject to a finite number of unobserved strong factors and an infinite number of weak and/or semi-strong unobserved common factors, provided that certain conditions on the loadings of the infinite factor structure are satisfied.

- In a Monte Carlo (MC) study, Coakley, Fuertes and Smith (2006) compare ten alternative estimators for the mean slope coefficient in a linear heterogeneous panel regression with strictly exogenous regressors and unobserved common (correlated) factors. Their results show that, overall, the mean group version of the CCE estimator stands out as the most efficient and robust.
- These conclusions are in line with those in Kapetanios, Pesaran and Yagamata (2011) and Chudik, Pesaran and Tosetti (2011), who investigate the small sample properties of CCE estimators and the estimators based on principal components. The MC results show that PC augmented methods do not perform as well as the CCE approach, and can lead to substantial size distortions, due, in part, to the small sample errors in the number of factors selection procedure.

Estimation and inference on large dynamic panel data models with a factor error structure

- Consider the following heterogeneous dynamic panel data model

$$
\begin{aligned}
y_{it} &= \lambda_i y_{i,t-1} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}, \qquad (7) \\
u_{it} &= \gamma_i' \mathbf{f}_t + e_{it}, \qquad (8)
\end{aligned}
$$

  for $i = 1, 2, ..., N$; $t = 1, 2, ..., T$. It is assumed that $|\lambda_i| < 1$, and the dynamic processes have started a long time in the past.
- Fixed effects and observed common factors (denoted by $\mathbf{d}_t$ previously) can also be included in the model. They are excluded to simplify the notations.
- The problem of estimation of panels subject to cross-sectional error dependence becomes much more complicated once the assumption of strict exogeneity of the unit-specific regressors is relaxed.

- As before, we distinguish between the case of homogenous coefficients, where $\lambda_i = \lambda$ and $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i, and the heterogenous case, where λ_i and $\boldsymbol{\beta}_i$ are randomly distributed across units and the object of interest are the mean coefficients $\lambda = E(\lambda_i)$ and $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$.
- This distinction is more important for dynamic panels, since not only the rate of convergence is affected by the presence of coefficient heterogeneity, but, as shown by Pesaran and Smith (1995), pooled least squares estimators are no longer consistent in the case of dynamic panel data models with heterogenous coefficients.
- It is convenient to define the vector of regressors $\boldsymbol{\zeta}_{it} = (y_{i,t-1}, \mathbf{x}'_{it})'$ and the corresponding parameter vector $\boldsymbol{\pi}_i = \left(\lambda_i, \boldsymbol{\beta}'_i\right)'$ so that (7) can be written as

$$y_{it} = \boldsymbol{\pi}'_i \boldsymbol{\zeta}_{it} + u_{it}.$$

- We review the following estimators:
  - **Quasi Maximum Likelihood Estimator (QMLE)** proposed by Moon and Weidner (2010a,b).
  - Extension of the **Principal Components (PC) approach** to dynamic heterogenous panels proposed Song (2013)
  - Extension of the **Common Correlated Effects (CCE) approach** to dynamic heterogeneous panels by Chudik and Pesaran (2013b).

QMLE approach

- Moon and Weidner (2010a,b) assume $\boldsymbol{\pi}_i = \boldsymbol{\pi}$ for all i and develop a Gaussian QMLE of the homogenous coefficient vector $\boldsymbol{\pi}$:

$$\hat{\boldsymbol{\pi}}_{QMLE} = \arg\min_{\boldsymbol{\pi}\in\mathbb{B}} L_{NT}\left(\boldsymbol{\pi}\right),$$

where $\mathbb{B}$ is a compact parameter set assumed to contain the true parameter values, and the objective function is the profile likelihood function.

$$L_{NT}\left(\boldsymbol{\pi}\right) = \min_{\{\gamma_i\},\{\mathbf{f}_t\}} \frac{1}{NT}\sum_{i=1}^{N}\left(\mathbf{y}_i - \boldsymbol{\Xi}_i\boldsymbol{\pi} - \mathbf{F}\gamma_i\right)'\left(\mathbf{y}_i - \boldsymbol{\Xi}_i\boldsymbol{\pi} - \mathbf{F}\gamma_i\right),$$

where

$$\boldsymbol{\Xi}_i = \begin{pmatrix} y_{i1} & \mathbf{x}'_{i,2} \\ y_{i,2} & \mathbf{x}'_{i,3} \\ \vdots & \vdots \\ y_{i,T-1} & \mathbf{x}'_{iT} \end{pmatrix}$$

- Both $\hat{\pi}_{QMLE}$ and $\hat{\beta}_{PC}$ minimize the same objective function and therefore, when the same set of regressors is considered, these two estimators are numerically the same, but there are important differences in their bias-corrected versions and in other aspects of the analysis of Bai and the analysis of Moon and Weidner (MW).
- MW allow for more general assumptions on regressors, including the possibility of weak exogeneity, and adopt a quadratic approximation of the profile likelihood function, which allows the authors to work out the asymptotic distribution and to conduct inference on the coefficients.
- MW show that $\hat{\pi}_{QMLE}$ is a consistent estimator of π, as $N, T \to \infty$ without any restrictions on the ratio T/N.
- To derive the asymptotic distribution of $\hat{\pi}_{QMLE}$, MW require $T/N \to \varkappa$, $0 < \varkappa < \infty$, as $N, T \to \infty$, and assume that the idiosyncratic errors, e_{it}, are cross-sectionally independent.

- MW show that $\sqrt{NT}\left(\hat{\pi}_{QMLE}-\pi\right)$ convergences to a normal distribution that is not centered around zero. The nonzero mean is due to two types of asymptotic bias:
  - the first is due to the heteroskedasticity of the error terms, as in Bai (2009), and
  - the second source of bias is due to the presence of weakly exogenous regressors.
- Authors provide consistent estimators of each component of the asymptotic bias.
- Regarding the tests on the estimated parameters, MW propose modified versions of the Wald, the likelihood ratio, and the Lagrange multiplier tests. Modifications are required due to the asymptotic parameter bias.

- Using MC experiments MW show that their bias corrected QMLE preforms well in small samples.
- Same as in Bai (2009), the number of factors is assumed to be known and therefore the estimation of m can introduce a certain degree of sampling uncertainty into the analysis.
- To overcome this problem, MW show, under somewhat more restrictive set of assumptions, that it is sufficient to assume an upper bound $m_{\max}$ on the number of factors and conduct the estimation with $m_{\max}$ principal components so long as $m \leq m_{\max}$.

PC approach

- Song (2013) extends Bai's (2009) approach to dynamic panels with heterogenous coefficients. The focus of Song's analysis is on the estimation of unit-specific coefficients $\boldsymbol{\pi}_i = \left(\lambda_i, \boldsymbol{\beta}_i'\right)'$.
- Song proposes an iterated least squares estimator of $\boldsymbol{\pi}_i$, and as in Bai (2009) shows that the solution can be obtained by alternating the PC method applied to the least squares residuals and the least squares estimation of $y_{it} = \lambda_i y_{i,t-1} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}$ until convergence.
- The least squares estimator of $\boldsymbol{\pi}_i$ and $\mathbf{F}$ is the solution to the following set of non-linear equations

$$\hat{\boldsymbol{\pi}}_{i,PC} = \left(\Xi_i' \mathbf{M}_{\hat{F}} \Xi_i\right)^{-1} \Xi_i' \mathbf{M}_{\hat{F}} \mathbf{y}_i, \text{ for } i = 1, 2, \ldots, N,$$

$$\frac{1}{NT} \sum_{i=1}^{N} \left(\mathbf{y}_i - \Xi_i \hat{\boldsymbol{\pi}}_{i,PC}\right) \left(\mathbf{y}_i - \Xi_i \hat{\boldsymbol{\pi}}_{i,PC}\right)' \hat{\mathbf{F}} = \hat{\mathbf{F}} \hat{\mathbf{V}}.$$

- Song establishes consistency of $\hat{\pi}_{i,PC}$ when $N, T \to \infty$ without any restrictions on T/N. If in addition $T/N^2 \to 0$, Song shows that $\hat{\pi}_{i,PC}$ is $\sqrt{T}$ consistent, but derives the asymptotic distribution only under some additional requirements including the cross-sectional independence of e_{it}.
- Song does not provide theoretical results on the estimation of the mean coefficients $\pi = E(\pi_i)$, but considers the mean group estimator,

$$\hat{\pi}^s_{PCMG} = \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}_{i,PC},$$

  in a Monte Carlo study.
- Results on the asymptotic distribution of $\hat{\pi}^s_{PCMG}$ are not yet established in the literature, but results of Monte Carlo study presented in Chudik and Pesaran (2013b) suggest that $\sqrt{N}\left(\hat{\pi}^s_{PCMG} - \pi\right)$ is asymptotically normally distributed with mean zero and a covariance matrix that can be estimated nonparametrically in the same way as in the case of the CCEMG estimator.

CCE approach

- The CCE approach as it was originally proposed in Pesaran (2006) does not cover the case where the panel includes a lagged dependent variable or weakly exogenous regressors.
- Chudik and Pesaran (2013b, CP) extends the CCE approach to dynamic panels with heterogenous coefficients and weakly exogenous regressors.
- The inclusion of lagged dependent variable amongst the regressors has three main consequences for the estimation of the mean coefficients:
  1. The **time series bias**, which affects the individual specific estimates and is of order $O\left(T^{-1}\right)$.
  2. The **full rank condition becomes necessary** for the consistent estimation of the mean coefficients (unless the factors in $\mathbf{f}_t$ are serially uncorrelated).
  3. The interaction of dynamics and coefficient heterogeneity leads to **infinite lag order relationships** between unobserved common factors and cross section averages of the observables when N is large.

- CP show that there exists the following large N distributed lag relationship between the unobserved common factors and cross section averages of the dependent variable and the regressors, $\bar{\mathbf{z}}_{wt} = (\bar{y}_{wt}, \bar{\mathbf{x}}'_{wt})'$,

$$\boldsymbol{\Lambda}(L)\,\tilde{\boldsymbol{\Gamma}}'\mathbf{f}_t = \mathbf{z}_{wt} + O_p\left(N^{-1/2}\right),$$

  where as before $\tilde{\boldsymbol{\Gamma}} = E(\boldsymbol{\gamma}_i, \Gamma_i)$.
- The existence of a large N relationship between the unobserved common factors and cross section averages of variables is not surprising considering that only the components with the largest exponents of cross-sectional dependence can survive cross-sectional aggregation with granular weights.
- The decay rate of the matrix coefficients in $\boldsymbol{\Lambda}(L)$ depends on the heterogeneity of λ_i and $\boldsymbol{\beta}_i$ and other related distributional assumptions.

- Assuming $\tilde{\boldsymbol{\Gamma}}$ has full row rank, i.e. $rank\left(\tilde{\boldsymbol{\Gamma}}\right) = m$, and the distributions of coefficients are such that $\boldsymbol{\Lambda}^{-1}(L)$ exists and has exponentially decaying coefficients yields following unit-specific cross-sectionally augmented auxiliary regressions,

$$y_{it} = \lambda_i y_{i,t-1} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + \sum_{\ell=0}^{p_T} \delta_{i\ell}' \bar{\mathbf{z}}_{w,t-\ell} + e_{yit}, \tag{9}$$

where $\bar{\mathbf{z}}_{wt}$ and its lagged values are used to approximate $\mathbf{f}_t$.
- The error term e_{yit} consists of three parts: an idiosyncratic term, e_{it}, an error component due to the truncation of possibly infinite distributed lag function, and an $O_p\left(N^{-1/2}\right)$ error component due to the approximation of unobserved common factors based on large N relationships.
- CP consider the least squares estimates of $\boldsymbol{\pi}_i = \left(\lambda_i, \boldsymbol{\beta}_i'\right)'$ based on the cross sectionally augmented regression (9), denoted as $\hat{\boldsymbol{\pi}}_i = \left(\hat{\lambda}_i, \hat{\boldsymbol{\beta}}_i'\right)'$, and the mean group estimate of $\boldsymbol{\pi} = E\left(\boldsymbol{\pi}_i\right)$ based on $\hat{\boldsymbol{\pi}}_i$, denoted as $\widehat{\boldsymbol{\pi}}_{MG} = \frac{1}{N}\sum_{i=1}^{N} \hat{\boldsymbol{\pi}}_i$.

- CP show that $\widehat{\boldsymbol{\pi}}_i$ and $\widehat{\boldsymbol{\pi}}_{MG}$ are consistent estimators of $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}$, respectively assuming that the rank condition is satisfied and $(N, T, p_T) \to \infty$ such that $p_T^3/T \to \varkappa$, $0 < \varkappa < \infty$, but without any restrictions on the ratio N/T.
- The rank condition is necessary for the consistency of $\widehat{\boldsymbol{\pi}}_i$ because the unobserved factors are allowed to be correlated with the regressors. If the unobserved common factors were serially uncorrelated (but still correlated with $\mathbf{x}_{it}$), then $\widehat{\boldsymbol{\pi}}_{MG}$ is consistent also in the rank deficient case, despite the inconsistency of $\widehat{\boldsymbol{\pi}}_i$, so long as factor loadings are independently, identically distributed across i.
- The convergence rate of $\widehat{\boldsymbol{\pi}}_{MG}$ is $\sqrt{N}$ due to the heterogeneity of the slope coefficients. CP show that $\widehat{\boldsymbol{\pi}}_{MG}$ converges to a normal distribution as $(N, T, p_T) \to \infty$ such that $p_T^3/T \to \varkappa_1$ and $T/N \to \varkappa_2$, $0 < \varkappa_1, \varkappa_2 < \infty$.
- The ratio N/T needs to be restricted for conducting inference, due to the presence of small time series bias.

- In the full rank case, the asymptotic variance of $\widehat{\boldsymbol{\pi}}_{MG}$ is given by the variance of $\boldsymbol{\pi}_i$ alone. When the rank condition does not hold, but factors are serially uncorrelated, then the asymptotic variance depends also on other parameters, including the variance of factor loadings.
- In both cases the asymptotic variance can be consistently estimated non-parametrically as before.
- Monte Carlo experiments in Chudik and Pesaran (2013b) show that extension of the CCE approach to dynamic panels with a multi-factor error structure performs reasonably well (in terms of bias, RMSE, size and power).
- This is particularly the case when the parameter of interest is the average slope of the regressors ($\boldsymbol{\beta}$), where the small sample results are quite satisfactory even if N and T are relatively small (around 40).

- The situation is different if the parameter of interest is the mean coefficient of the lagged dependent variable (λ), where the CCEMG estimator suffers form the well known time series bias and tests based on it tend to be over-sized, unless T is sufficiently large.
- To mitigate the consequences of this bias, Chudik and Pesaran (2013b) consider application of half-panel jackknife procedure (Dhaene and Jochmansy, 2012), and the recursive mean adjustment procedure (So and Shin, 1999), both of which are easy to implement.
- The proposed jackknife bias-corrected CCEMG estimator is found to be more effective in mitigating the time series bias, but it can not fully deal with the size distortion when T is relatively small.
- Improving the small T sample properties of the CCEMG estimator of λ in the heterogeneous panel data models still remains a challenge to be taken on in the future.

Further extensions of the CCE approach

- The application of the CCE approach to static panels with weakly exogenous regressors (namely without lagged dependent variables) has not yet been investigated in the literature.
- Monte Carlo study by Chudik and Pesaran (2013a) suggests for this case that:
  - The CCE mean group estimator performs very well (in terms of bias and RMSE) for $T > 50$ (for all values of N considered). Also tests based on this estimator are correctly sized and have good power properties. These results are obtained in experiments where the rank condition does not hold.
  - The CCE pooled estimator, in contrast, is no longer consistent in the case of weakly exogenous regressors with heterogenous coefficients, due to the bias caused by the correlation between the slope coefficients and the regressors.

Tests of error cross-sectional dependence

- Consider the following panel data model

$$y_{it} = a_i + \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}, \tag{10}$$

  where a_i and $\boldsymbol{\beta}_i$ for $i = 1, 2, ..., N$ are assumed to be fixed unknown coefficients, and $\mathbf{x}_{it}$ is a k-dimensional vector of regressors.
- We provide on overview of alternative approaches to testing the cross-sectional independence or weak dependence of the errors u_{it}.
- We consider both cases where the regressors are strictly and weakly exogenous, as well as when they include lagged values of y_{it}.

- The literature on testing for error cross-sectional dependence in large panels follow two separate strands, depending on whether the cross section units are ordered or not. In what follows we review the various attempts made in the literature to develop tests of cross-sectional dependence when the cross-section units are unordered.
- In the case of cross section observations that do not admit an ordering, tests of cross-sectional dependence are typically based on estimates of pair-wise error correlations (ρ_{ij}) and are applicable when T is sufficiently large so that relatively reliable estimates of ρ_{ij} can be obtained.
- An early test of this type is the Lagrange multiplier (LM) test of Breusch and Pagan (1980) which tests the null hypothesis that *all* pair-wise correlations are zero. This test is based on the average of the *squared* estimates of pair-wise correlations, and under standard regularity conditions it is shown to be asymptotically (as $T \to \infty$) distributed as χ^2 with $N(N-1)/2$ degrees of freedom.

- The LM test tends to be highly over-sized in the case of panels with relatively large N.
- In the remainder, we review the various attempts made in the literature to develop tests of cross-sectional dependence when N is large.
- When N is relatively large and rising with T, it is unlikely to matter if out of the total $N(N-1)/2$ pair-wise correlations only a few are non-zero. Accordingly, Pesaran (2013) argues that the null of cross-sectionally uncorrelated errors, defined by

$$H_0 : E\left(u_{it}u_{jt}\right) = 0, \text{for all } t \text{ and } i \neq j, \tag{11}$$

is restrictive for large panels and the null of a sufficiently weak cross-sectional dependence could be more appropriate since mere incidence of isolated dependencies are of little consequence for estimation or inference about the parameters of interest, such as the individual slope coefficients, $\boldsymbol{\beta}_i$, or their average value, $E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$.

- Let $\hat{u}_{it}$ be the OLS estimator of u_{it} defined by

$$\hat{u}_{it} = y_{it} - \hat{a}_i - \hat{\boldsymbol{\beta}}_i' \mathbf{x}_{it},$$

with $\hat{a}_i$, and $\hat{\boldsymbol{\beta}}_i$ being the OLS estimates of a_i and $\boldsymbol{\beta}_i$, based on the T sample observations, $y_t, \mathbf{x}_{it}$, for $t = 1, 2, ..., T$.

- Consider the sample estimate of the pair-wise correlation of the residuals, $\hat{u}_{it}$ and $\hat{u}_{jt}$, for $i \neq j$

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T} \hat{u}_{it} \hat{u}_{jt}}{\left(\sum_{t=1}^{T} \hat{u}_{it}^2\right)^{1/2} \left(\sum_{t=1}^{T} \hat{u}_{jt}^2\right)^{1/2}}.$$

- It is known that, under the null (11) and when N is finite,

$$\sqrt{T}\hat{\rho}_{ij} \overset{a}{\sim} N(0,1),$$

for a given i and j, as $T \to \infty$, and $T\hat{\rho}_{ij}^2$ is asymptotically distributed as a χ_1^2.

- Consider the following statistic

$$CD_{LM} = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(T\hat{\rho}_{ij}^2 - 1 \right). \quad (12)$$

- Based on the Euclidean norm of the matrix of sample correlation coefficients, (12) is a version of the Lagrange Multiplier test statistic due to Breusch and Pagan (1980).
- Frees (1995) first explored the finite sample properties of the LM statistic, calculating its moments for fixed values of T and N, under the normality assumption. He advanced a non-parametric version of the LM statistic based on the Spearman rank correlation coefficient.
- Dufour and Khalaf (2002) have suggested to apply Monte Carlo exact tests to correct the size distortions of CD_{LM} in finite samples. However, these tests, being based on the bootstrap method applied to the CD_{LM}, are computationally intensive, especially when N is large.

- An alternative adjustment to the LM test is proposed by Pesaran and Ullah and Yamagata (2008), where the LM test is centered to have a zero mean for a fixed T. These authors also propose a correction to the variance of the LM test.
- The basic idea is generally applicable, but analytical bias corrections can be obtained only under the assumption that the regressors, $\mathbf{x}_{it}$, are strictly exogenous and the errors, u_{it} are normally distributed. The adjusted LM statistic is now given by

$$LM_{Adj} = \sqrt{\frac{2}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{(T-k)\,\hat{\rho}_{ij}^2 - \mu_{Tij}}{v_{Tij}},$$

where μ_{Tij} and v_{Tij} depends on T, k, and $\{\mathbf{x}_{it}\}$ and their expressions are provided in Pesaran and Ullah and Yamagata (2008).

- LM_{Adj} is asymptotically $N(0,1)$ under H_0, when $T \to \infty$ followed by $N \to \infty$.

- The application of the LM_{Adj} test to dynamic panels or panels with weakly exogenous regressors is further complicated by the fact that the bias corrections depend on the true values of the unknown parameters and will be difficult to implement. The implicit null of LM tests when T and $N \to \infty$, jointly rather than sequentially could also differ from the null of uncorrelatedness of all pair-wise correlations.
- To overcome some of these difficulties Pesaran (2004) has proposed a test that has exactly mean zero for fixed values of T and N. This test is based on the average of pair-wise correlation coefficients

$$CD_P = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \right).$$

- As $N, T \to \infty$ in any order, CD_P tends approximately to a standardized normal. One important advantage of the CD_P test is that it is applicable also to autoregressive heterogeneous panels, even for a fixed T, so long as u_{it} are symmetrically distributed around zero. The CD test can also be applied to unbalanced panels.
- Pesaran (2013) extends the analysis of CD_P test and shows that the implicit null of the test is that of weak cross-sectional dependence and it depends on the relative expansion rates of N and T.
- In particular, using the exponent of cross-sectional dependence, α, developed in Bailey, Kapetanios and Pesaran (2011) and discussed above, Pesaran shows that when $T = O\left(N^{\epsilon}\right)$ for some $0 < \epsilon \leq 1$ the implicit null of the CD_P test is given by $0 \leq \alpha < (2 - \epsilon)/4$. This yields the range $0 \leq \alpha < 1/4$ when $N, T \to \infty$ at the same rate such that $T/N \to \varkappa$ for some finite positive constant $\varkappa$, and the range $0 \leq \alpha < 1/2$ when T is small relative to N.

- For larger values of α, as shown by Bailey, Kapetanios and Pesaran (2011), α can be estimated consistently using the variance of the cross-sectional averages.
- Monte Carlo experiments reported in Pesaran (2013) show that the *CD* test has good small sample properties for values of α in the range $0 \leq \alpha \leq 1/4$, even in cases where regressors are weakly exogenous.

- Other statistics have also been proposed in the literature to test for zero contemporaneous correlation in the errors u_{it}.
- Using results from the literature on *spacing* discussed in Pyke (1965), Ng (2006) considers a statistic based on the q^{th} differences of the cumulative normal distribution associated to the $N(N-1)/2$ pair-wise correlation coefficients ordered from the smallest to the largest, in absolute value.
- Building on the work of John (1971), and under the assumption of normal disturbances, strictly exogenous regressors, and homogenous slopes, Baltagi, Feng and Kao (2011) propose a test of the null hypothesis of sphericity, defined by $H_0^{BFK} : \mathbf{u}_t \sim IIDN\left(\mathbf{0}, \sigma_u^2 \mathbf{I}_N\right)$. Joint assumption of homoskedastic errors and homogenous slopes is quite restrictive in applied work and therefore the use of the J_{BFK} statistics as a test of cross-sectional dependence should be approached with care.

Conclusions

- We have characterized the cross-sectional dependence as weak or strong, and defined the exponent of cross-sectional dependence, α.
- We have also considered estimation and inference on large panels with a factor error structure. We have distinguished between panels with strictly exogenous regressors and dynamic panels.
- Last but not least, we have provided an overview of the literature on tests of error cross-sectional dependence when N is large and units are unordered.

Reading list

This lecture is based on the following survey paper:

- Chudik, A. and M. H. Pesaran (2013a). Large Panel Data Models with Cross-Sectional Dependence: A Survey. Mimeo, May 2013.

For further reading on the topics covered in this lecture, see the following papers.
- On the types of cross-sectional dependence:

- Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *The Econometrics Journal* 14, C45–C90.
- Bailey, N., G. Kapetanios, and M. H. Pesaran (2012). Exponents of cross-sectional dependence: Estimation and inference. CESifo Working Paper No. 3722, revised October 2012.

- **On the estimation and inference on large panels with strictly exogenous regressors and a factor error structure:**
(**a**) PC methods

- Coakley, J., A. M. Fuertes, and R. Smith (2002). A principal components approach to cross-section dependence in panels. Birkbeck College Discussion Paper 01/2002.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.

(**b**) CCE approach

- Pesaran, M. H. (2006). Estimation and inference in large heterogenous panels with multifactor error structure. *Econometrica* 74, 967–1012.

Extensions of CCE approach:

- Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *The Econometrics Journal* 14, C45–C90
- Kapetanios, G., M. H. Pesaran, and T. Yagamata (2011). Panels with nonstationary multifactor error structures. *Journal of Econometrics* 160, 326–348.
- Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics* 161 (2), 182–202.

(**c**) Monte Carlo study

- Coakley, J., A. M. Fuertes, and R. Smith (2006). Unobserved heterogeneity in panel time series. *Computational Statistics and Data Analysis* 50, 2361–2380.

- On the estimation and inference on large dynamic panel data models with a factor error structure:

(**a**) QMLE approach

- Moon, H. R. and M. Weidner (2010a). Dynamic linear panel regression models with interactive fixed effects. Mimeo, July 2010.
- Moon, H. R. and M. Weidner (2010b). Linear regression for panel with unknown number of factors as interactive fixed effects. Mimeo, July 2010.

(**b**) PC approach

- Song, M. (2013). Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Mimeo, 30 January 2013.

(**c**) CCE approach

- Chudik, A. and M. H. Pesaran (2013b). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. CESifo Working Paper No. 4232.

- **On the tests of error cross-sectional dependence:**

- Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. CESifo Working Paper No. 1229.
- Pesaran, M. H. (2013). Testing weak cross-sectional dependence in large panels. forthcoming in *Econometric Reviews*.
- Pesaran, M. H., A. Ullah, and T. Yamagata (2008). A bias-adjusted LM test of error cross section independence. *The Econometrics Journal* 11, 105–127.