Modelling the Cross-section Dependence, the Spatial Heterogeneity and the Network Diffusion in the Multi-dimensional Dataset

Yongcheol Shin, University of York
The Workshop at University of Bari

18-21 September 2018

Cross-section Dependence (CSD)

- CSD's are pervasive in panels, though it was usual to assume the absence of CSD.
- Consequences of ignoring CSD can be serious: pooling may introduce severe biases (Phillips and Sul, 2003).
- In spatial econometrics a natural way to characterise dependence in terms of distance, but for most economic problems no obvious distance measure.
- Trade between countries reflects not just geographical distance, but transport costs, policy and historical factors as well as the multilateral barriers (Anderson and van Wincoop, 2003).
- For large T straightforward to test for CSD (Pesaran, 2015).

- We consider the generic panel data model (Serlenga and Shin, 2007):

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\lambda}'\mathbf{z}_i + \boldsymbol{\pi}_i'\mathbf{s}_t + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, ..., T, \quad (1)$$

- $\mathbf{x}_{it} = (x_{1,it}, ..., x_{k,it})'$ is a $k \times 1$ vector of variables that vary over individuals and time periods $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)'$;
- $\mathbf{z}_i = (z_{1,i}, ..., z_{g,i})'$ is a $g \times 1$ vector of individual-specific variables with $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_g)'$;
- $\mathbf{s}_t = (s_{1,t}, ..., s_{s,t})'$ is an $s \times 1$ vector of observed factors with $\boldsymbol{\pi}_i = (\pi_{1i}, ..., \pi_{si})'$.

- To address heterogeneous individual and time effects, consider one- and two-way error components specifications:

$$\varepsilon_{it} = \alpha_i + u_{it} \tag{2}$$

$$\varepsilon_{it} = \alpha_i + \theta_t + u_{it} \tag{3}$$

- α_i is an individual effect correlated with $\mathbf{x}_{it}$ and $\mathbf{z}_i$;
- θ_t is the common time effects;
- u_{it} is a zero mean idiosyncratic disturbance.

Representation of CSD via Unobserved Factors

- The most popular approach is to add heterogeneous factors:

$$\varepsilon_{it} = \alpha_i + \boldsymbol{\gamma}_i' \boldsymbol{f}_t + u_{it}, \tag{4}$$

  where $\boldsymbol{f}_t$ is the $c \times 1$ vector of unobserved factors with heterogeneous parameter vector $\boldsymbol{\gamma}_i = (\gamma_{1i}, ..., \gamma_{ci})'$, and u_{it} is a zero mean idiosyncratic disturbance.
- Factors expected to provide good proxy for any remaining complex time-varying patterns associated with multilateral resistance and globalisation trends.
- CSD explicitly allowed through heterogeneous loadings $\boldsymbol{\gamma}_i$.
- If $\boldsymbol{f}_t$ is correlated with $\boldsymbol{x}_{it}$, then not allowing for CSD by omitting $\boldsymbol{f}_t$ causes the conventional FE estimates of $\boldsymbol{\beta}_i$ to be biased (Pesaran, 2006; Bai, 2009).

Representations of CSD via Spatial Effects

- Spatial models allow the $N \times 1$ vector of errors, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ to follow:

$$\boldsymbol{\varepsilon}_t = \boldsymbol{W}\boldsymbol{\varepsilon_t} + \boldsymbol{u}_t$$

  where $\boldsymbol{u}_t = (u_{1t}, ..., u_{Nt})'$ is cross-sectionally independent and $\boldsymbol{W}$ is a known (possibly time-varying) spatial weights matrix.
- The structure of CSD is related to the location and the distance on the basis of a pre-specified weight matrix, $\boldsymbol{W}$.
- Hence, cross section correlation is represented by means of a spatial process, which relates each unit to its neighbours.

- Consider a spatial gravity (SARAR) model (MSS, 2015):

$$y_{it} = \rho y_{it}^{*} + \boldsymbol{\beta}'\boldsymbol{x}_{it} + \boldsymbol{\lambda}'\mathbf{z}_i + \alpha_i + v_{it},\ i = 1, ..., N,\ t = 1, ..., T, \quad (5)$$

$$v_{it} = \lambda v_{it}^{*} + u_{it} \quad (6)$$

  where $y_{it}^{*} = \sum_{j\neq i}^{N} w_{ij} y_{jt}$ is the spatial lagged variable and $v_{it}^{*} = \sum_{j\neq i}^{N} w_{ij} v_{jt}$ is the spatial autoregressive error term, w_{ij}'s are the spatial weight with the row-sum normalisation, $\sum_i w_{ij} = 1$, and u_{it} is a zero mean idiosyncratic disturbance.
- ρ is the spatial lag coefficient and λ the spatial error coefficient.
- These capture spatial spillover effects and measure the influence of the weighted average of neighboring observations.

Weak and Strong CSD

- With weak CSD, dependence is local and decline with N; each unit is spatially correlated only with near neighbors.
- With strong CSD, the dependence influences all units (e.g. common or dominant factors).
- For weak dependence, all the eigenvalues of the covariance matrix of the errors are bounded as $N \to \infty$.
- For strong dependence, the largest eigenvalue $\to \infty$ with N.
- Bailey et al. (2016) characterise strength of dependence as $\alpha = \ln(n)/\ln(N)$, where n is the number of units with nonzero factor loadings.
- For strong factor, $\alpha = 1$.
- $\alpha < 1/2$ indicates the weak factor.
- $\frac{1}{2} \leq \alpha \leq \frac{3}{4}$ represent a moderate degree of CSD.
- The implications are different depending on whether $\boldsymbol{f}_t$'s are nuisance parameters or they are the parameters of interest.

Cross-section Dependence (CD) Test by Pesaran (2015)

- CSD captured by non-zero covariance between ε_{it} and ε_{jt}, which relates to rate at which $\frac{1}{N}\sum_{i=1}^{N}\sigma_{ijt}$ declines with N.
- We compute the pair-wise residual correlations by

$$\hat{\rho}_{ij} = \frac{\hat{\boldsymbol{\varepsilon}}_i' \hat{\boldsymbol{\varepsilon}}_j}{\sqrt{\left(\hat{\boldsymbol{\varepsilon}}_i' \hat{\boldsymbol{\varepsilon}}_i\right)\left(\hat{\boldsymbol{\varepsilon}}_j' \hat{\boldsymbol{\varepsilon}}_j\right)}}, \quad i, j = 1, \dots, N \text{ and } i \neq j,$$

  where $\hat{\boldsymbol{\varepsilon}}_i = \left(\hat{\varepsilon}_{i1}, \dots, \hat{\varepsilon}_{iT}\right)'$.
- We construct the CD statistic by

$$CD = \sqrt{\frac{2}{N\left(N-1\right)}}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sqrt{T}\hat{\rho}_{ij} \qquad (7)$$

- CD test has the limiting $N(0,1)$ distribution under the null of residual cross-section independence, $H_0 : \hat{\rho}_{ij} = 0$.

The Factor-based Models

- A large number of estimators suggested to deal with CSD.
- The market leader appears to be CCE type estimators.
- There are a number of issues of interpretation.
- It is common in a lot of time-series applications of PCs to transform the data to make it stationary by differencing.

The correlated common effect estimator

- If one treats factors as nuisance parameters and removes CSD, a simple procedure is the CCE estimator (Pesaran, 2006).
- Consider the panel data model:

$$y_{it} = \boldsymbol{\delta}_i' \boldsymbol{d}_t + \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \varepsilon_{it} \text{ with } \varepsilon_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + u_{it} \tag{8}$$

where y_{it} is a dependent variable, $\boldsymbol{d}_t$ is a $k_d \times 1$ vector of variables that do not differ over units (intercept and trend), $\boldsymbol{x}_{it}$ is a $k_x \times 1$ vector of regressors which differ over units, $\boldsymbol{f}_t$ is an $r \times 1$ vector of unobserved factors, which may influence each unit differently and are correlated with $\boldsymbol{x}_{it}$, and u_{it} is an unobserved disturbance with $E(u_{it}) = 0$ and $E(u_{it}^2) = \sigma_i^2$, which is independently distributed across i and t.

- Pesaran suggests to include the cross-section means of y_{it} and $\boldsymbol{x}_{it}$ as additional regressors, to remove the effect of the factors.
- The consistency holds for any linear combination of the dependent variable and the regressors subject to the assumptions that weights w_i satisfy:

$$(i) : w_i = O\left(\frac{1}{N}\right); \; (ii) : \sum_{i=1}^{N} |w_i| < K; \; (iii) : \sum_{i=1}^{N} w_i \gamma_i \neq 0$$

- These clearly hold for the mean:

$$w_i = \frac{1}{N}; \; \sum_{i=1}^{N} |w_i| = 1; \; \sum_{i=1}^{N} w_i \gamma_i = N^{-1} \sum_{i=1}^{N} \gamma_i \neq 0.$$

- This involves adding CS means of the dependent and independent variables:

$$y_{it} = \boldsymbol{\delta}_i' \boldsymbol{d}_t + \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \boldsymbol{\pi}_{yi} \bar{y}_t + \boldsymbol{\pi}_{xi}' \bar{\boldsymbol{x}}_t + u_{it} \tag{9}$$

- Assume a single factor and average (8) across units:

$$\bar{y}_t = \bar{\boldsymbol{\delta}}' \boldsymbol{z}_t + \bar{\boldsymbol{\beta}}_i' \bar{\boldsymbol{x}}_t + \bar{\gamma} f_t + \bar{u}_t + N^{-1} \sum \left(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}} \right)' \boldsymbol{x}_{it} \tag{10}$$

and thus

$$f_t = \frac{1}{\bar{\gamma}} \left\{ \bar{y}_t - \bar{\boldsymbol{\delta}}' \boldsymbol{z}_t - \bar{\boldsymbol{\beta}}_i' \bar{\boldsymbol{x}}_t - \bar{u}_t - N^{-1} \sum \left(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}} \right)' \boldsymbol{x}_{it} \right\} \tag{11}$$

so $\bar{y}_t$ and $\bar{\boldsymbol{x}}_t$ provide a proxy for the unobserved factor.

- For large N there is no endogeneity problem as the covariance between $\bar{y}_t$ and u_{it} goes to zero.
- CCE generalises to many factors and lagged dependent variables, but requires that $\bar{\gamma}$ is non-zero.

- Pesaran (2006) shows that $\boldsymbol{\beta}_i$ can be consistently estimated by

$$\hat{\boldsymbol{\beta}}_i = \left(\boldsymbol{X}_i' \boldsymbol{M}_D \boldsymbol{X}_i\right)^{-1} \boldsymbol{X}_i' \boldsymbol{M}_D \boldsymbol{y}_i$$

where $\boldsymbol{y}_i$ is a $T \times 1$ vector of the dependent variable for the ith unit, $\boldsymbol{X}_i$ is a $T \times k_x$ vector of regressors, and $\boldsymbol{M}_D = \boldsymbol{I}_T - \boldsymbol{D}\left(\boldsymbol{D}'\boldsymbol{D}\right)^{-1} \boldsymbol{D}$ with $\boldsymbol{D}$ consisting of observed common factor and CS averages, $\bar{y}_t$ and $\bar{\boldsymbol{x}}_t$.

- For the mean of $\boldsymbol{\beta}_i$ we can use the mean group estimator:

$$\hat{\boldsymbol{\beta}}_{MG} = \frac{1}{n} \sum_{i=1}^{n} \hat{\boldsymbol{\beta}}_i.$$

- Also the pooled version:

$$\hat{\beta}_P = \left(\sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{M}_D \boldsymbol{X}_i \right)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{M}_D \boldsymbol{y}_i$$

- The CCE is valid with a single or multiple unobserved factors and does not require the number of factors to be smaller than the number of observed CS averages.
- CCE is easy to compute as OLS and no iteration is needed. Desirable small sample properties of CCE are demonstrated.
- Estimating factors by CS means seems to work better than estimating them by the PC estimator.
- CCE determines the weights *a priori* rather than estimating them by PCs. Not estimating the weights seems to improve the performance of the procedure (Kapetanios et al. 2011).
- **Remark**: Westerlund and Urbain (2013) show that CCE becomes inconsistent when the factor loadings in y equation are correlated with the factor loadings in x equation.

Panel Data Models with Interactive Fixed Effects

- Bai (2009) considers the large N large T panel data model with interactive fixed effects:

$$y_{it} = \boldsymbol{X}'_{it}\boldsymbol{\beta} + u_{it} \text{ with } \; u_{it} = \boldsymbol{\lambda}'_i \boldsymbol{F}_t + \varepsilon_{it} \tag{12}$$

  where $\boldsymbol{X}_{it}$ is a $k \times 1$ vector of regressors and $\boldsymbol{f}_t$ is an $r \times 1$ vector of unobserved factors.
- This model assumes homogeneous parameters.
- It is a generalisation of the additive model, nesting FE model:

$$y_{it} = \boldsymbol{X}'_{it}\boldsymbol{\beta} + \alpha_i + \xi_t + \varepsilon_{it} = \boldsymbol{X}'_{it}\boldsymbol{\beta} + \boldsymbol{\lambda}'_i \boldsymbol{F}_t + \varepsilon_{it}$$

  where $\boldsymbol{\lambda}_i = (1, \alpha_i)'$ and $\boldsymbol{F}_t = (t, 1)'$.
- It allows a richer form of unobserved heterogeneity: e.g. F_t can represent a vector of macroeconomic common shocks and $\boldsymbol{\lambda}_i$ individual i's heterogeneous responses.

- We estimate the model by minimizing:

$$SSR(\beta, F, \Lambda) = \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{F}\boldsymbol{\lambda}_i)' (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{F}\boldsymbol{\lambda}_i)$$

$$st \ \frac{\boldsymbol{F}'\boldsymbol{F}}{T} = \boldsymbol{I}_r, \ \boldsymbol{\Lambda}'\boldsymbol{\Lambda} \text{ is diagonal}$$

- No closed-form solution, but consistent estimators obtained by iterations.
- Obtain some initial values $\boldsymbol{\beta}^{(0)}$, such as LS estimators from regressing $\boldsymbol{Y}_i$ on $\boldsymbol{X}_i$.
- Perform PC analysis for $\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}^{(0)}$ to obtain $\boldsymbol{F}^{(1)}$ and $\Lambda^{(1)}$.
- Next, regress $\boldsymbol{Y}_i - \boldsymbol{F}^{(1)}\boldsymbol{\lambda}_i^{(1)}$ on $\boldsymbol{X}_i$ to obtain $\boldsymbol{\beta}^{(1)}$.
- Iterate such steps until convergence.

- The limiting distribution for $\hat{\beta}$ depends on assumptions on ε_{it} as well as on the ratio T/N.
- If $T/N \to 0$, the limiting distribution of $\hat{\beta}$ will be centered around zero, given that $E\left(\varepsilon_{it}\varepsilon_{js}\right) = 0$ for $t \neq s$, and $E\left(\varepsilon_{it}\varepsilon_{js}\right) = \sigma_{ij}$ for all i, j, t.
- If $T/N \to K > 0$, the limiting distribution will not be centered around zero, a challenge for inference.
- Bai (2009) provided a bias-corrected estimator, $\tilde{\beta}$, whose limiting distribution is centered around zero.
- Assume that $T/N^2 \to 0$ and $N/T^2 \to 0$, $E\left(\varepsilon_{it}^2\right) = \sigma_{it}^2$, and $E\left(\varepsilon_{it}\varepsilon_{js}\right) = 0$ for $i \neq j$ and $t \neq s$, then

$$\sqrt{NT}\left(\tilde{\beta} - \beta\right) \to_d N\left(0, \Sigma_\beta\right)$$

- The issue of choosing r remains.

Extensions

- Moon and Weidner (2014) consider the model with lagged dependent variable as regressors.
- Moon and Weidner (2015) show that the limiting distribution of the LS estimator is not affected by the number of factors, as long as it is no smaller than the true number.
- Lu and Su (2015) propose the adaptive group LASSO (least absolute shrinkage and selection operator), which can simultaneously select the regressors and the number of factors.
- Westerlund and Urbain (2015) provide a comparison of the cross-sectional average and principal component estimators.

- Chudik and Pesaran (2015) propose CCE estimation of heterogeneous dynamic panel data models with weakly exogenous regressors.
- De Vos and Everaert (2016) extend the CCEP to homogeneous dynamic panels, and develop a bias-corrected estimator that is consistent as N tends to infinity, both for T fixed or large.
- Karabiyik Reese and Westerlund (2017) point to a problem with the CCE approach when the number of factors is strictly less than the number of observables. The use of too many observables causes the second moment matrix of the estimated factors to become asymptotically singular.
- More...

Nonlinear Panel Data Model of CSD

- Kapetanios, Mitchell and Shin (2014) propose a nonlinear panel data model which can endogenously generate both 'weak' and 'strong' CSD.
- A given agent's behaviour is influenced by an aggregation of the views or actions of those around them.
- The model allows for considerable flexibility in terms of the genesis of herding or clustering type behaviour.
- At an econometric level, the model nests various extant dynamic panel data models. These include panel AR models, spatial models, and panel-factor models.

- We propose dynamic nonlinear panel data models:

$$x_{i,t} = \rho \sum_{j=1}^{N} w_{ij} \left(x_{-i,t-1}, x_{i,t-1}; \gamma \right) x_{j,t-1} + \epsilon_{i,t}, \quad i = 1, ...N, \quad t = 1, ... \tag{13}$$

  where $x_{-i,t} = \left(x_{1,t}, x_{2,t}, ..., x_{i-1,t}, x_{i+1,t}, ..., x_{Nt} \right)'$ and $\sum_{j=1}^{N} w_{ij} \left(x_{-i,t-1}, x_{i,t-1}; \gamma \right) = 1$.
- $x_{i,t}$ depends in a nonlinear fashion depending on how w_{ij} is parameterised, on weighted averages of past values of $x_t = (x_{1,t}, ..., x_{Nt})'$, where the weights depend on x_{t-1}.
- It mimics social interactions between economic units.
- This model can accommodate generic forms of CSD.
- We place particular emphasis on specifications where weights depend on x_{t-1} only through distances, $|x_{j,t-1} - x_{i,t-1}|$.
- This is easy to analyse, based on a threshold mechanism.

- Our models provide an intuitive means by which many forms of CSD can arise in a large panel comprised of variables of a 'similar' nature that relate to different agents/units.
- These variables might be the disaggregates underlying often studied macroeconomic or financial aggregates, such as economy-wide inflation or the S&P500 index.
- The model allows different economic units to cluster and for these clusters (and their number) to evolve over time.
- The degree of CSD can vary, from a case where it is similar to factor models to the case of very weak structure.
- Our model constitutes the first attempt to introduce endogenous CSD.

- Let $x_{i,t}$ denote the agent's income or the agent's view of the future value of some macroeconomic variable, at time t.
- Then, we specify:

$$x_{i,t} = \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(\left|x_{i,t-1} - x_{j,t-1}\right| \leq r\right) x_{j,t-1} + \epsilon_{i,t}, \qquad (14)$$

where

$$m_{i,t} = \sum_{j=1}^{N} \mathcal{I}\left(\left|x_{i,t-1} - x_{j,t-1}\right| \leq r\right),$$

$\{\epsilon_{i,t}\}_{t=1}^{T}$ is an error process, $\mathcal{I}(.)$ is the indicator function and $-1 < \rho < 1$.

- $x_{i,t}$ is influenced by CS average of a selection of $x_{j,t-1}$ and the relevant $x_{j,t-1}$ are those closest to $x_{i,t-1}$.

- A deterministic form of the above model has been analysed in the mathematical and system engineering literature.
- They analyse a continuous form of the restricted version:

$$x_{i,t} = \frac{1}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(\left|x_{i,t-1} - x_{j,t-1}\right| \leq 1\right) x_{j,t-1}, \tag{15}$$

where $m_{i,t} = \sum_{j=1}^{N} \mathcal{I}\left(\left|x_{i,t-1} - x_{j,t-1}\right| \leq 1\right)$.

- By setting $r = 0$, we obtain a simple panel autoregressive model:

$$x_{i,t} = \rho x_{i,t-1} + \epsilon_{i,t}. \tag{16}$$

- Letting $r \to \infty$, we obtain the following model

$$x_{i,t} = \frac{\rho}{N} \sum_{j=1}^{N} x_{j,t-1} + \epsilon_{i,t}, \tag{17}$$

  where past cross-sectional averages of opinions inform, in similar fashions, current opinions.
- The use of such cross-sectional averages has been advocated by Pesaran (2006) as a means of modelling CSD in the form of unobserved factors.
- In our case, the use of cross-sectional averages is a limiting case of a 'structural' nonlinear model.

- Factor models have the property that both the maximum eigenvalue and the row/column sum norm of the covariance matrix of x_t tend to infinity as $N \to \infty$.
- For spatial models, these quantities are bounded.
- We show that the column sum norm of the variance covariance matrix of x_t is $O(N)$. The model is more similar to factor models.
- Interestingly, there are versions of (14) that resemble spatial models.
- Our model has a clear parametric structure, which is a feature shared by dynamic spatial model and more general than spatial models, as the weighting schemes are estimated endogenously.
- The nonlinear model can lie between the two extremes characterised by weak spatial- and strong factor models.

- We allow different weights to the selected neighbors as follows:

$$x_{i,t} = \nu_i + \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) w_{ij} x_{j,t-1} + \epsilon_{i,t} \tag{18}$$

  where we may consider the following weights

$$w_{ij} = \frac{d_{ij}^{-2}}{\sum_{j=1}^{N} d_{ij}^{-2}}, \quad d_{ij} = |x_{i,t-1} - x_{j,t-1}| \tag{19}$$

- The estimation can be done in two steps:
- First, the consistent estimate of r is obtained;
- Then, construct the weights by (271) and estimate the model, (270).

- Mastromarco, Serlenga and Shin (2015) propose a framework for accommodating both time- and cross-section dependence in modelling technical efficiency in stochastic frontier.
- The approach enables us to deal with both weak and strong forms of CSD by introducing exogenously driven common factors and an endogenous threshold selection mechanism.
- Using the dataset of 26 OECD countries over 1970-2010, we provide the satisfactory estimation results for the production technology parameters and the efficiency ranking of individual countries.
- We find positive spillover effect on efficiency, supporting the hypothesis that knowledge spillover is more likely to be induced by technological proximity.
- Our approach enables us to identify efficiency clubs endogenously.

- We begin with the Cobb-Douglas production function:

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}, \; i = 1, ..., N, \; t = 1, ..., T, \tag{20}$$

where y_{it} is a logarithm of output of country i at time t, $\mathbf{x}_{it}$ a $k \times 1$ vector of (logged) production inputs, $\boldsymbol{\beta}$ a $k \times 1$ vector of structural parameters, and ε_{it} is the errors including the idiosyncratic disturbance (v_{it}) and time varying (logged) technical inefficiency (u_{it}):

$$\varepsilon_{it} = v_{it} - u_{it}. \tag{21}$$

- Mastromarco *et al.* (2013) propose the stochastic frontier model with unobserved factors for modelling u_{it}:

$$u_{it} = \alpha_i + \boldsymbol{\lambda}_i'\mathbf{f}_t, \; i = 1, ..., N, \; t = 1, ..., T, \tag{22}$$

where α_i is (unobserved) individual effects, and $\mathbf{f}_t$ is an $r \times 1$ vector of unobserved factors that provide a proxy for complex trending patterns.

- Individual country's total factor productivity (TFP) is likely to be significantly affected by economic performance of neighboring or frontier countries.
- The productivity shocks are assumed to be spatially correlated:

$$\boldsymbol{\varepsilon}_t = \rho \mathbf{W} \boldsymbol{\varepsilon}_t + \mathbf{e_t},\ t = 1, ..., T, \tag{23}$$

  where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$, $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{N}$ is the $N \times N$ spatial weight matrix, ρ is a spatial AR parameter, and $\mathbf{e_t} = (e_{1t}, ..., e_{Nt})'$ idiosyncratic disturbances.
- The spatial-based approach is likely to produce biased estimates in the presence of strong CSD.
- Factor-based models impose an assumption that the strong CSD is driven by an exogenously given unobserved factors.
- KMS propose an approach that allows the CSD to be determined endogenously.

- Suppose that the product of a country i at time t, Y_{it}, is determined by labor input and private capital, L_{it} and K_{it}. It is also affected by the Hicks-neutral multi-factor productivity:

$$Y_{it} = TFP_{it}F(L_{it}, K_{it}), \tag{24}$$

- TFP_{it} can be decomposed into the level of technology A_{it}, a measurement error w_{it}, and the efficiency measure τ_{it}:

$$TFP_{it} = A_{it}\tau_{it}w_{it}. \tag{25}$$

- By writing (24) in log form:

$$y_{it} = \alpha + \beta_1 k_{it} + \beta_2 l_{it} - u_{it} + v_{it}, \tag{26}$$

with the two-way error components structure given by

$$\varepsilon_{it} = v_{it} - u_{it}, \tag{27}$$

where $v_{it} = \ln w_{it}$ and $u_{it} = -\ln(\tau_{it})$ is the (time-varying) technical inefficiency.

- Innovators consider the behaviour of other agents as:

$$u_{it} = \alpha_i + \rho \tilde{u}_{it}(r) + \boldsymbol{\lambda}_i' \mathbf{f}_t. \tag{28}$$

$$\tilde{u}_{it}(r) = \frac{1}{m_{it}} \sum_{j=1}^{N} I\left(\left|u_{t-1}^{*} - u_{jt-1}\right| \leq r\right) u_{jt-1}, \tag{29}$$

  and r is the threshold parameter that is determined endogenously and u_{t-1}^{*} is the efficiency of the best performing country and $m_{it} = \sum_{j=1}^{N} I\left(\left|u_{t-1}^{*} - u_{jt-1}\right| \leq r\right)$.
- $\tilde{u}_{it}(r)$ is a spatial interaction term capturing CS local average of the best technology.
- Such externalities can be captured by a negative ρ.
- We measure individual inefficiency (Schmidt and Sickles, 1984)

$$e_{it} = \max_i (u_{it}) - (u_{it}) = \max_i \left(\alpha_i + \rho \tilde{u}_{it}(r) + \boldsymbol{\lambda}_i' \mathbf{f}_t\right) - \left(\alpha_i + \rho \tilde{u}_{it}(r) + \right. \tag{30}$$

- Rewrite the models:

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, ..., T, \tag{31}$$

$$\varepsilon_{it} = v_{it} - u_{it}, \tag{32}$$

$$u_{it} = \alpha_i + \rho \tilde{u}_{it}(r) + \boldsymbol{\lambda}_i'\mathbf{f}_t, \tag{33}$$

$$\tilde{u}_{it}(r) = \frac{1}{m_{it}} \sum_{j=1}^{N} I\left(\left|u_{t-1}^{*} - u_{jt-1}\right| \leq r\right) u_{jt-1}, \tag{34}$$

  where α_i is (unobserved) individual-specific effect, $\mathbf{f}_t$ is an $r \times 1$ vector of unobserved factors and $\boldsymbol{\lambda}_i$ is an $r \times 1$ vector of the heterogeneous loading, $\tilde{u}_{it}(r)$ represents a cluster effect which is equal to the average efficiency of countries which are close to the frontier where $u_{t-1}^{*} = \min_j \left(u_{jt-1}\right)$.
- The distinguishing feature is the use of unit-specific aggregates, which summaries past values of efficiency, and connects the units that are close to the technology frontier.

- We estimate $\hat{\boldsymbol{\beta}}$ in (31) by PCCE or IPC, and derive $\hat{\varepsilon}_{it} = y_{it} - \mathbf{x}_{it}\hat{\boldsymbol{\beta}}$ with $\hat{v}_{it} = v_{it} - \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\mathbf{x}_{it} = v_{it} + o_p(1)$.
- We get a first proxy of inefficiency as $\hat{e}_{it} = \max_i(\hat{\varepsilon}_{it}) - (\hat{\varepsilon}_{it})$.
- We consider the threshold estimation, where a grid for r is constructed. We estimate $\hat{r}$ and $\hat{\rho}$ jointly by minimising:

$$\mathbf{V}(r,\rho) = \min_{r,\rho} \sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{e}_{it} - \rho\frac{1}{m_{it}}\sum_{j=1}^{N} I\left(\left|\hat{e}^{*}_{t-1} - \hat{e}_{jt-1}\right| \le r\right)\hat{e}_{jt-1}\right) \tag{35}$$

- The time-varying individual technical inefficiencies can be consistently estimated by

$$\hat{e}_{it} = \max_i(\hat{u}_{it}) - (\hat{u}_{it}) = \max_i\left(\hat{\alpha}_i + \hat{\rho}\tilde{u}_{it}(\hat{r}) + \hat{\boldsymbol{\lambda}}_i'\mathbf{f}_t\right) - \left(\hat{\alpha}_i + \hat{\rho}\tilde{u}_{it}(\hat{r}) - \right. \tag{36}$$

- Convert $\hat{e}_{it}$ to time-varying individual technical efficiency:

$$\hat{\tau}_{it} = \exp(-\hat{e}_{it}). \tag{37}$$

For empirical implementations, we follow Bailey *et al.* (2016) who propose a multi-step procedure to deal with both strong and weak forms of CSD as follows:

1. Test for the existence of CSD by applying the Pesaran (2015) CD test;
2. If the null of CSD is rejected, we apply the factor-based model to control for strong CSD.
3. We apply the CD test again to the (de-factored) residuals.
4. If the null of no CSD is rejected, we also apply spatial or network modelling to the residuals (see (33)).

The Spatial Autoregressive (SAR) Process

- Consider the first-order spatial autoregressive (SAR) process:

$$y_i = \lambda \boldsymbol{w}_{in} \boldsymbol{Y}_n + \varepsilon_i, \quad i = 1, ..., n, \tag{38}$$

where $\boldsymbol{Y}_n = (y_1, ..., y_n)'$ is an $n \times 1$ vector of dependent variable, $\boldsymbol{w}_{in}$ is a $1 \times n$ vector of weights, and $\varepsilon_i \sim iid(0, \sigma^2)$.

- We write the model in the matrix form:

$$\boldsymbol{Y}_n = \lambda \boldsymbol{W}_n \boldsymbol{Y}_n + \boldsymbol{E}_n. \tag{39}$$

$\boldsymbol{W}_n \boldsymbol{Y}_n$ is called ‘the spatial lag’.

- Assuming that $\boldsymbol{S}_n(\lambda) = \boldsymbol{I}_n - \lambda \boldsymbol{W}_n$ is nonsingular, we have:

$$\boldsymbol{Y}_n = \boldsymbol{S}_n(\lambda)^{-1} \boldsymbol{E}_n.$$

- Consider the regression with SAR disturbance:

$$\boldsymbol{Y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{U}_n, \quad \boldsymbol{U}_n = \rho\boldsymbol{W}_n\boldsymbol{U}_n + \boldsymbol{E}_n \tag{40}$$

  where $\boldsymbol{E}_n$ has zero mean and variance $\sigma^2\boldsymbol{I}_n$.
- $\boldsymbol{U}_n$ are spatially-correlated across units.
- The variance of $\boldsymbol{U}_n$ is $\sigma^2\boldsymbol{S}_n(\rho)^{-1}\boldsymbol{S}_n(\rho)^{-1\prime}$.
- The off-diagonal elements of $\boldsymbol{S}_n(\rho)^{-1}\boldsymbol{S}_n(\rho)^{-1\prime}$ may be nonzero, and u_i's are cross-sectionally correlated.

Spatial autoregressive model with covariates

- We generalise SAR process by incorporating exogenous variables $\boldsymbol{x}_i$. In matrix form,

$$\boldsymbol{Y}_n = \lambda \boldsymbol{W}_n \boldsymbol{Y}_n + \boldsymbol{X}_n \boldsymbol{\beta} + \boldsymbol{E}_n \tag{41}$$

  where $\boldsymbol{E}_n \sim iid(0, \sigma^2 \boldsymbol{I}_n)$.

- Its reduced form is:

$$\boldsymbol{Y}_n = \boldsymbol{S}_n(\lambda)^{-1} \boldsymbol{X}_n \boldsymbol{\beta} + \boldsymbol{S}_n(\lambda)^{-1} \boldsymbol{E}_n.$$

Some Intuitions on Spatial Weights Matrix

- The jth element of $\boldsymbol{w}_{in}$, $w_{n,ij}$, represents the link between the neighbor j and the spatial unit i.
- The diagonal of $\boldsymbol{W}_n$ is set to zero ($w_{n,ii} = 0$) because $\lambda \boldsymbol{w}_{in}$ represents the effect of other spatial units on the spatial unit i.
- It is common to make $\boldsymbol{W}_n$ row-normalised (the sum of each row is unity): e.g. the ith row $\boldsymbol{w}_{in}$ constructed as $\boldsymbol{w}_{in} = (d_{i1}, d_{i2}, ..., d_{in}) / \sum_{j=1}^{n} d_{ij}$, where $d_{ij} \geq 0$ is a function of the distance.
- When neighbors are adjacent ones, the correlation is local in the sense that correlations will be stronger for neighbors but weak for units far away.

- Suppose that $\|\rho \boldsymbol{W}_n\| \leq 1$ for matrix norm $\|.\|$. Then

$$\boldsymbol{S}_n(\rho)^{-1} = \boldsymbol{I}_n + \sum_{i=1}^{\infty} \rho^i \boldsymbol{W}_n^i$$

- Notice that

$$\left\| \sum_{i=m}^{\infty} \rho^i \boldsymbol{W}_n^i \right\| \leq |\rho \boldsymbol{W}_n|^i \left\| \boldsymbol{S}_n(\rho)^{-1} \right\|$$

- If $\boldsymbol{W}_n$ is row-normalized, then

$$\left\| \sum_{i=m}^{\infty} \rho^i \boldsymbol{W}_n^i \right\|_\infty \leq \sum_{i=m}^{\infty} \left|\rho^i\right| = \frac{|\rho|^m}{1-|\rho|}$$

will become small as m gets larger.

- $\boldsymbol{U}_n$ can be represented as

$$\boldsymbol{U}_n = \boldsymbol{E}_n + \rho \boldsymbol{W}_n \boldsymbol{E}_n + \rho^2 \mathbf{W}_n^2 \boldsymbol{E}_n + ...,$$

  where $\rho \boldsymbol{W}_n$ may represent the influence of neighbors, $\rho^2 \mathbf{W}_n^2$ represents the second layer neighborhood influence, *etc*.
- $\boldsymbol{W}_n \boldsymbol{S}_n (\rho)^{-1}$ is a vector of measures of centrality, which summaries the position of each spatial unit in a network.

Other generalizations

- We may combine SAR with SAR disturbances:

$$\boldsymbol{Y}_n = \lambda \boldsymbol{W}_n \boldsymbol{Y}_n + \boldsymbol{X}_n \boldsymbol{\beta} + \boldsymbol{U}_n, \quad \boldsymbol{U}_n = \rho \boldsymbol{M}_n \boldsymbol{U}_n + \boldsymbol{E}_n \tag{42}$$

  where $\boldsymbol{W}_n$ and $\boldsymbol{M}_n$ are spatial weights matrices, which may not be identical.

- Further extension may allow high-order spatial lags:

$$\boldsymbol{Y}_n = \sum_{j=1}^{p} \lambda_j \boldsymbol{W}_{jn} \boldsymbol{Y}_n + \boldsymbol{X}_n \boldsymbol{\beta} + \boldsymbol{E}_n \tag{43}$$

  where $\boldsymbol{W}_{jn}$'s are p distinct spatial weights matrices.

Estimation Methods

- We consider the QML, the 2SLS, and the GMM.
- QML has usually good finite sample properties.
- MLE is not computationally attractive for higher spatial lags model, in which case IV and GMM may be feasible (Lee, 2007).

MLE

- For the SAR process, we have the log-likelihood function:

$$\ln L_n\left(\lambda,\sigma^2\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln\left|\boldsymbol{S}_n\left(\lambda\right)\right| - \frac{1}{2\sigma^2}\mathbf{Y}_n'\boldsymbol{S}_n\left(\lambda\right)'\boldsymbol{S}_n\left(\lambda\right)\mathbf{Y}_n \quad (44)$$

- For the model with SAR disturbances:

$$\ln L_n\left(\rho,\beta,\sigma^2\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln\left|\boldsymbol{S}_n\left(\rho\right)\right| \quad (45)$$

$$-\frac{1}{2\sigma^2}\left(\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta}\right)'\boldsymbol{S}_n\left(\rho\right)'\boldsymbol{S}_n\left(\rho\right)\left(\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta}\right)$$

- The LLF for the SAR model with covariates is

$$\ln L_n\left(\lambda, \beta, \sigma^2\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln\left|\boldsymbol{S}_n\left(\lambda\right)\right| \quad (46)$$

$$-\frac{1}{2\sigma^2}\left(\boldsymbol{Y}_n\boldsymbol{S}_n\left(\lambda\right) - \boldsymbol{X}_n\boldsymbol{\beta}\right)'\left(\boldsymbol{Y}_n\boldsymbol{S}_n\left(\lambda\right) - \boldsymbol{X}_n\boldsymbol{\beta}\right)$$

- The LF involves computation of the determinant of $\boldsymbol{S}_n\left(\lambda\right) = \boldsymbol{I}_n - \lambda\boldsymbol{W}_n$, which is a function of the unknown parameter λ, and may have a large dimension n.
- Anselin suggest an iterative algorithm between λ and $\boldsymbol{\beta}$.
- A computationally tractable alternative method is due to Ord (1975).

2SLS estimation

- The spatial lag $\boldsymbol{W}_n\boldsymbol{Y}_n$ can be correlated with the disturbance, $\boldsymbol{E}_n$. OLS may not be a consistent estimator.
- Rewrite (41) as

$$\boldsymbol{Y}_n = \boldsymbol{Z}_n\boldsymbol{\theta} + \boldsymbol{E}_n \tag{47}$$

where $\boldsymbol{Z}_n = (\boldsymbol{W}_n\boldsymbol{Y}_n, \boldsymbol{X}_n)$ and $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}')'$.

- Kelejian and Prucha (1998) suggest the 2SLS estimator of $\boldsymbol{\theta}$ using IVs, $\boldsymbol{Q}_n$:

$$\hat{\boldsymbol{\theta}}_{2SLS} = \left[\boldsymbol{Z}_n'\boldsymbol{Q}_n\left(\boldsymbol{Q}_n'\boldsymbol{Q}_n\right)^{-1}\boldsymbol{Q}_n'\boldsymbol{Z}_n\right]\left[\boldsymbol{Z}_n'\boldsymbol{Q}_n\left(\boldsymbol{Q}_n'\boldsymbol{Q}_n\right)^{-1}\boldsymbol{Q}_n'\boldsymbol{Y}_n\right]$$

- The asymptotic distribution of $\hat{\boldsymbol{\theta}}_{2SLS}$ follows:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{2SLS}-\boldsymbol{\theta}\right)\rightarrow_{d}$$
$$N\left(0,\sigma^{2}\left(\boldsymbol{G}_{n}\boldsymbol{X}_{n}\boldsymbol{\beta},\boldsymbol{X}_{n}\right)'\boldsymbol{Q}_{n}\left(\boldsymbol{Q}_{n}'\boldsymbol{Q}_{n}\right)^{-1}\boldsymbol{Q}_{n}'\left(\boldsymbol{G}_{n}\boldsymbol{X}_{n}\boldsymbol{\beta},\boldsymbol{X}_{n}\right)\right)$$

  where $\boldsymbol{G}_n = \boldsymbol{W}_n \boldsymbol{S}_n^{-1}$.
- Kelejian and Prucha suggest the use of linearly independent variables in $(\boldsymbol{X}_n, \boldsymbol{W}_n\boldsymbol{X}_n)$ for the construction of $\boldsymbol{Q}_n$.
- The optimum IV matrix is $(\boldsymbol{G}_n\boldsymbol{X}_n\boldsymbol{\beta}, \boldsymbol{X}_n)$.
- This 2SLS cannot be used for the estimation of the (pure) SAR process with $\boldsymbol{\beta} = 0$.

GMM

- Kelejian and Prucha (1999) suggest GMM estimation:

$$\min_{\theta} \boldsymbol{g}_n' (\boldsymbol{\theta}) \boldsymbol{g}_n (\boldsymbol{\theta}) .$$

  based on three moment equations:

$$\begin{aligned} E\left(\boldsymbol{E}_n' \boldsymbol{E}_n\right) &= n\sigma^2; E\left(\boldsymbol{E}_n' \boldsymbol{W}_n' \boldsymbol{W}_n \boldsymbol{E}_n\right) = \sigma^2 tr\left(\boldsymbol{W}_n' \boldsymbol{W}_n\right); \\ E\left(\boldsymbol{E}_n' \boldsymbol{W}_n \boldsymbol{E}_n\right) &= 0 \end{aligned}$$

- In this case we have:

$$g_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{Y}_n' \boldsymbol{S}_n(\lambda)' \boldsymbol{S}_n(\lambda) \boldsymbol{Y}_n - n\sigma^2 \\ \boldsymbol{Y}_n' \boldsymbol{S}_n(\lambda)' \boldsymbol{W}_n' \boldsymbol{W}_n \boldsymbol{S}_n(\lambda) \boldsymbol{Y}_n - \sigma^2 tr\left(\boldsymbol{W}_n' \boldsymbol{W}_n\right), \\ \boldsymbol{Y}_n' \boldsymbol{S}_n(\lambda)' \boldsymbol{W}_n \boldsymbol{S}_n(\lambda) \boldsymbol{Y}_n \end{pmatrix}$$

- The orthogonality conditions, $\boldsymbol{Q}_n' \boldsymbol{\varepsilon}_n(\boldsymbol{\theta}) = 0$ provide the $k_x \times 1$ vector of moment conditions, where

$$\boldsymbol{\varepsilon}_n(\boldsymbol{\theta}) = \boldsymbol{S}_n(\lambda) \boldsymbol{Y}_n - \boldsymbol{X}_n \beta$$

- For SAR with covariates, we can obtain other moment equations.
- Consider a finite number of $n \times n$ constant matrices, $\boldsymbol{P}_{1n}, ..., \boldsymbol{P}_{mn}$, each of which has a zero diagonal. Then, $(\boldsymbol{P}_{jn}\boldsymbol{\varepsilon}_n(\boldsymbol{\theta}))' \boldsymbol{\varepsilon}_n(\boldsymbol{\theta})$ can be used as the moment functions.
- We have the following moment conditions vector:

$$\boldsymbol{g}_n(\boldsymbol{\theta}) = (\boldsymbol{P}_{1n}\boldsymbol{\varepsilon}_n(\boldsymbol{\theta}), ..., \boldsymbol{P}_{mn}\boldsymbol{\varepsilon}_n(\boldsymbol{\theta}), \boldsymbol{Q}_n)' \boldsymbol{\varepsilon}_n(\boldsymbol{\theta})$$

- Intuitively, as

$$\boldsymbol{W}_n\boldsymbol{Y}_n = \boldsymbol{G}_n\boldsymbol{X}_n\boldsymbol{\beta}_0 + \boldsymbol{G}_n\boldsymbol{\varepsilon}_n, \ \boldsymbol{G}_n = \boldsymbol{W}_n\boldsymbol{S}_n^{-1}, \boldsymbol{S}_n = \boldsymbol{S}_n(\lambda_0),$$

and $\boldsymbol{G}_n\boldsymbol{\varepsilon}_n$ is correlated with the disturbance $\boldsymbol{\varepsilon}_n$ in

$$\boldsymbol{Y}_n = \lambda\boldsymbol{W}_n\boldsymbol{Y}_n + \boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{\varepsilon}_n$$

- Any $\boldsymbol{P}_{jn}\boldsymbol{\varepsilon}_n$, uncorrelated with $\boldsymbol{\varepsilon}_n$, can be used as IV for $\boldsymbol{W}_n\boldsymbol{Y}_n$ as long as $\boldsymbol{P}_{jn}\boldsymbol{\varepsilon}_n$ and $\boldsymbol{G}_n\boldsymbol{\varepsilon}_n$ are correlated.

The SDPD Models

- The dynamic panel data with spatial effect is called the SDPD model (Yu et al., 2008):

$$\begin{aligned} \boldsymbol{Y}_{nt} &= \lambda_0 \boldsymbol{W}_n \boldsymbol{Y}_{nt} + \gamma_0 \boldsymbol{Y}_{n,t-1} + \rho_0 \boldsymbol{W}_n \boldsymbol{Y}_{n,t-1} \\ &\quad + \boldsymbol{X}_{nt} \boldsymbol{\beta}_0 + \boldsymbol{c}_{n_0} + \boldsymbol{\alpha}_{t0} \boldsymbol{l}_n + \boldsymbol{V}_{nt}, \end{aligned} \tag{48}$$

  where $\boldsymbol{c}_{n0}$ is $n \times 1$ column vector of fixed effects and $\boldsymbol{\alpha}_{t0}$'s are time effects.
- γ_0 captures pure dynamic effect and ρ_0 the spatial-time or diffusion effect.

- Define

$$\boldsymbol{S}_n(\lambda) = \boldsymbol{I}_n - \lambda \boldsymbol{W}_n \text{ and } \boldsymbol{S}_n \equiv \boldsymbol{S}_n(\lambda_0) = \boldsymbol{I}_n - \lambda_0 \boldsymbol{W}_n.$$

- (48) can be rewritten as

$$\boldsymbol{Y}_{nt} = \boldsymbol{A}_n \boldsymbol{Y}_{n,t-1} + \boldsymbol{S}_n^{-1}(\boldsymbol{X}_{nt}\boldsymbol{\beta}_0 + \boldsymbol{c}_{n0} + \boldsymbol{\alpha}_{t0}\boldsymbol{l}_n + \boldsymbol{V}_{nt}) \quad (49)$$

- where $\boldsymbol{A}_n = \boldsymbol{S}_n^{-1}(\gamma_0 \boldsymbol{I}_n + \rho_0 \boldsymbol{W}_n)$.
- Let $\boldsymbol{\varpi}_n = diag\{\varpi_{n1}, ..., \varpi_{nn}\}$ be the $n \times n$ diagonal eigenvalues matrix of $\boldsymbol{W}_n$ such that $\boldsymbol{W}_n = \boldsymbol{\Gamma}_n \boldsymbol{\varpi}_n \boldsymbol{\Gamma}_n$ where $\boldsymbol{\Gamma}_n$ is the eigenvector matrix.

- The eigenvalues matrix of $\boldsymbol{A}_n$ is

$$\boldsymbol{D}_n = (\boldsymbol{I}_n - \lambda_0 \boldsymbol{W}_n)^{-1} (\gamma_0 \boldsymbol{I}_n + \rho_0 \boldsymbol{W}_n)$$

  such that $\boldsymbol{A}_n = \boldsymbol{\Gamma}_n \boldsymbol{D}_n \boldsymbol{\Gamma}_n$.
- As $\boldsymbol{W}_n$ is row-normalized, all the eigenvalues are less than or equal to 1 in absolute value.
- Let the first m_n eigenvalues of $\boldsymbol{W}_n$ be the unity.
- $\boldsymbol{D}_n$ decomposed into two parts, one corresponding to unit eigenvalues of $\boldsymbol{W}_n$ and the other corresponding to eigenvalues smaller than 1.

- Define $\boldsymbol{J}_n = diag\{l'_{m_n}, 0, ..., 0\}$ with l_{m_n} being an $m_n \times 1$ vector of ones and $\boldsymbol{D}_n = diag\{0, ..., 0, d_{n,m_n+1}, ..., d_{nn}\}$, where $|d_{ni}| < 1$ for $i = m_n + 1, ..., n$.
- We have

$$\boldsymbol{A}_n = \left(\frac{\gamma_0 + \rho_0}{1 - \lambda_0}\right)^h \boldsymbol{\Gamma}_n \boldsymbol{J}_n \boldsymbol{\Gamma}_n + \boldsymbol{B}_n^h \text{ with } \boldsymbol{B}_n = \boldsymbol{\Gamma}_n \boldsymbol{D}_n \boldsymbol{\Gamma}_n$$

- Four cases:
- Stable case: $\gamma_0 + \rho_0 + \lambda_0 < 1$ (with some other restrictions).
- Spatial cointegration case: $\gamma_0 + \rho_0 + \lambda_0 = 1$ but $\gamma_0 < 1$.
- Unit roots case: $\gamma_0 + \rho_0 + \lambda_0 = 1$ and $\gamma_0 = 1$.
- Explosive case: $\gamma_0 + \rho_0 + \lambda_0 > 1$.

The direct and indirect impacts

- A space-time multiplier in (49) specifies how the joint determination of the dependent variables is a function of both spatial and time lags of explanatory variables.
- LeSage and Pace (2009) introduce the concept of direct impact, total impact, and indirect impact.
- In a SAR model:

$$\boldsymbol{Y}_n = \boldsymbol{\alpha}_0 \boldsymbol{l}_n + \lambda_0 \boldsymbol{W}_n \boldsymbol{Y}_n + \sum_{k=1}^{k_x} \beta_{k0} \boldsymbol{X}_{nk} + \boldsymbol{\varepsilon}_n$$

- The impact of $\boldsymbol{X}_{nk}$ on Y_n is:

$$\frac{\partial \boldsymbol{Y}_n}{\partial \boldsymbol{X}'_{nk}} = (\boldsymbol{I}_n - \lambda_0 \boldsymbol{W}_n)^{-1}_{k0} \beta_{k0} \quad \text{for the } k\text{th regressor.}$$

- The average direct impact, average total impact and average indirect impact are defined as

$$f_{k,direct}\left(\theta_0\right) \equiv \frac{1}{n} tr\left(\left(I_n - \lambda_0 \mathbf{W}_n\right)^{-1} \beta_{k0}\right),$$

$$f_{k,total}\left(\theta_0\right) \equiv \frac{1}{n} l_n' \left(\left(I_n - \lambda_0 \mathbf{W}_n\right)^{-1} \beta_{k0}\right) l_n,$$

$$f_{k,indirect}\left(\theta_0\right) \equiv f_{k,total}\left(\theta_0\right) - f_{k,direct}\left(\theta_0\right),$$

  with l_n being an n-dimensional column of ones.
- Debarsy, Ertur and LeSage (2012) extend to spatial dynamic panel models.
- LeSage and Chin (2016) extend to heterogenous spatial models.

The Spatial Durbin Model

- Elhorst (2012) propose the general spatial Durbin model:

$$\begin{aligned} \boldsymbol{Y}_t &= \tau \boldsymbol{Y}_{t-1} + \delta \boldsymbol{W}\boldsymbol{Y}_t + \eta \boldsymbol{W}\boldsymbol{Y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta}_1 \\ &\quad + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\beta}_2 + \boldsymbol{X}_{t-1}\boldsymbol{\beta}_3 + \boldsymbol{W}\boldsymbol{X}_{t-1}\boldsymbol{\beta}_4 + \boldsymbol{Z}_t\boldsymbol{\theta} + \boldsymbol{v}_t \end{aligned} \tag{50}$$

$$\boldsymbol{v}_t = \gamma \boldsymbol{v}_{t-1} + \rho \boldsymbol{W} v_t + \boldsymbol{\mu} + \boldsymbol{\lambda}_t \boldsymbol{i}_N + \boldsymbol{\varepsilon}_t \text{ with } \boldsymbol{\mu} = \kappa \boldsymbol{W}\boldsymbol{\mu} + \boldsymbol{\xi}$$

  where $\boldsymbol{Y}_t$ is an $N \times 1$ vector of the dependent variable, $\boldsymbol{X}_t$ is an $N \times K$ matrix of exogenous regressors, and $\boldsymbol{Z}_t$ is an $N \times L$ matrix of endogenous regressors.

- τ, δ and η are scalar parameters on $\boldsymbol{Y}_{t-1}$, $\boldsymbol{W}\boldsymbol{Y}_t$ and $\boldsymbol{W}\boldsymbol{Y}_{t-1}$.

- $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, and $\boldsymbol{\beta}_4$ are the vectors of the parameters on exogenous regressors and $\boldsymbol{\theta}$ is the $L \times 1$ vector of the parameters on endogenous regressors.
- $\boldsymbol{v}_t$ is the $N \times 1$ vector of error term, which may be serially and spatially correlated.
- γ and ρ are the serial and spatial autocorrelation coefficient.
- $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)'$ is the $N \times 1$ vector of the spatial-specific effects, and $\boldsymbol{\lambda}_t$ $(t = 1, \ldots T)$ is time effects.
- $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ and $\boldsymbol{\xi}$ are vectors of *iid* disturbance terms, with zero mean and finite variance σ^2 and σ_ξ^2.
- Elhorst (2012) acknowledges the estimation complexity due to numerous identification issues.

Stationarity conditions

- The characteristic roots of the matrix $(\boldsymbol{I} - \delta \boldsymbol{W})^{-1} (\tau \boldsymbol{I} + \eta \boldsymbol{W})$ should lie within the unit circle:

$$\begin{aligned} \tau &< 1 - (\delta + \eta)\,\omega_{\max}, \quad if \quad \delta + \eta \geq 0 \\ \tau &< 1 - (\delta + \eta)\,\omega_{\min}, \quad if \quad \delta + \eta < 0 \\ -1 + (\delta - \eta)\,\omega_{\max} &< \tau, \quad if \quad \delta - \eta \geq 0 \\ -1 + (\delta - \eta)\,\omega_{\min} &< \tau, \quad if \quad \delta - \eta < 0 \end{aligned} \tag{51}$$

  where $\omega_{\min}$ denotes the smallest (most negative) and $\omega_{\max}$ the largest real characteristic root of $\boldsymbol{W}$.
- Stationarity conditions on the spatial and temporal parameters in (51) are considerably more difficult.

Estimation methods

- The spatial model estimated mainly by the iterative ML estimation by Anselin (1988).
- Three estimation methods follow:
  1. the bias-corrected QML estimator
  2. the IV or GMM estimator
  3. the Bayesian Markov Chain Monte Carlo (MCMC) approach.
- Yu et al. (2008) construct a bias-corrected estimator for a dynamic model with $(\boldsymbol{Y}_{t-1}, \boldsymbol{W}\boldsymbol{Y}_t, \boldsymbol{W}\boldsymbol{Y}_{t-1})$ and fixed effects.
- Lee and Yu (2010d) include time effects, and provide the bias-corrected LSDV estimator when (N, T) tend to infinity, but T cannot be too small relative to N.

- Elhorst (2010d) proposes the FD-GMM to accommodate endogenous interaction effects; severely biased.
- Lee and Yu (2010c) show that a 2SLS estimator is inconsistent due to too many moments; the dominant bias caused by endogeneity of $\boldsymbol{W}\boldsymbol{Y}_t$. They propose an optimal GMM estimator.
- Kukenova and Monteiro (2009) and Jacobs et al. (2009) consider a dynamic panel data model with $(\boldsymbol{Y}_{t-1}, \boldsymbol{W}\boldsymbol{Y}_t)$ and extend the system GMM to account for endogenous interaction effects $(\boldsymbol{W}\boldsymbol{Y}_t)$.
- GMM can be used to instrument endogenous regressors.

The Dynamic Spatial Durbin model

- Burridge (1981) recommends the first-order spatial autoregressive distributed lag model, known as the spatial Durbin model: $\boldsymbol{Y}$ is regressed on $\boldsymbol{WY}$, $\boldsymbol{X}$ and $\boldsymbol{WX}$.
- The cost of ignoring spatial dependence in the dependent/independent variables is relatively high (biased) whilst ignoring spatial dependence in the disturbances will only cause a loss of efficiency (LeSage and Pace, 2009).
- The spatial Durbin model produces unbiased estimates, even if the DGP contains a spatial error.

- Elhorst et al. (2010b) propose a dynamic spatial Durbin model:

$$\boldsymbol{Y}_t = \tau \boldsymbol{Y}_{t-1} + \delta \boldsymbol{W} \boldsymbol{Y}_t + \eta \boldsymbol{W} \boldsymbol{Y}_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta}_1 + \boldsymbol{W} \boldsymbol{X}_t \boldsymbol{\beta}_2 + \boldsymbol{v}_t \tag{52}$$

- Rewriting (52) as

$$\begin{aligned} \boldsymbol{Y} &= (\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1} (\tau \boldsymbol{I}_N + \eta \boldsymbol{W}) \boldsymbol{Y}_{t-1} \\ &\quad + (\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1} (\boldsymbol{X}_t \boldsymbol{\beta}_1 + \boldsymbol{W} \boldsymbol{X}_t \boldsymbol{\beta}_2) + (\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1} \boldsymbol{v}_t, \end{aligned} \tag{53}$$

  we can derive the partial derivatives of $\boldsymbol{Y}$ with respect to the kth explanatory variable of $\boldsymbol{X}$ by

$$\left[\begin{array}{ccc} \frac{\partial Y}{\partial x_{1k}} & \cdots & \frac{\partial Y}{\partial x_{Nk}} \end{array} \right]_t = (\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1} (\beta_{1k} \boldsymbol{I}_N + \beta_{2k} \boldsymbol{W}) \tag{54}$$

- The long-term effects can be:

$$[(1-\tau) \boldsymbol{I}_N - (\delta + \eta) \boldsymbol{W}]^{-1} (\beta_{1k} \boldsymbol{I}_N + \beta_{2k} \boldsymbol{W}) \tag{55}$$

- By continuous substitution of $\boldsymbol{Y}_{t-1}$ up to $\boldsymbol{Y}_1$ in (53),

$$\begin{aligned}\boldsymbol{Y} &= (\boldsymbol{I} - \delta \boldsymbol{W})^{-T} (\tau \boldsymbol{I} + \eta \boldsymbol{W})^{T} \boldsymbol{Y}_{t-T} \\ &+ \sum_{p=1}^{T} (\boldsymbol{I} - \delta \boldsymbol{W})^{-p} (\tau \boldsymbol{I} + \eta \boldsymbol{W})^{p-1} \\ &\times \left(\boldsymbol{X}_{t-(p-1)} \boldsymbol{\beta}_1 + \boldsymbol{W} \boldsymbol{X}_{t-(p-1)} \boldsymbol{\beta}_2 + \boldsymbol{v}_{t-(p-1)} \right) .\end{aligned} \quad (56)$$

- Two global spatial multiplier matrices, $(\boldsymbol{I} - \delta \boldsymbol{W})^{-p}$ and $(\tau \boldsymbol{I} + \eta \boldsymbol{W})^{p-1}$, are at work in conjunction with one process that produces local spatial spillover effects, $\boldsymbol{W} \boldsymbol{X}_{t-(p-1)} \boldsymbol{\beta}_2$.
- Elhorst (2001) suggests to regress $\boldsymbol{Y}_t$ on $\boldsymbol{Y}_{t-1}$, $\boldsymbol{W} \boldsymbol{Y}_t$, $\boldsymbol{W} \boldsymbol{Y}_{t-1}$, $\boldsymbol{X}_t$, $\boldsymbol{W} \boldsymbol{X}_t$, $\boldsymbol{X}_{t-1}$ and $\boldsymbol{W} \boldsymbol{X}_{t-1}$. This extension worsens the identification problem.

The following four restrictions are imposed:

1. $\boldsymbol{\beta}_2 = 0$: the local indirect effects (spatial spillover) set to 0. The indirect effects over the direct effects become the same for all X both in the short- and the long-term.
2. $\delta = 0$ such that $(\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1} = \boldsymbol{I}_N$. The global short-term indirect effect is zero.

- $\eta = -\tau\delta$ (Parent and LeSage, 2011). The impact of the explanatory variables can be decomposed into a spatial effect and a time effect; the impact over space falls by $\delta \boldsymbol{W}$ for every higher-order neighbor and over time by the factor τ for every period. The disadvantage is that the indirect effects remain constant over time.
- $\eta = 0$. Although this limits the flexibility of the ratio between indirect and direct effects, it seems to be the least restrictive.

HSAR model

- Aquaro, Bailey and Pesaran (2015) extend the SAR panel data model to the case where the spatial coefficients differ across the spatial units.
- QML estimators are consistent and asymptotically normally distributed when both T and N are large.
- QML estimators have satisfactory small sample properties with moderate time dimensions and irrespective of the number of cross section units, under certain sparsity conditions on the spatial weight matrix.

Heterogeneous spatial autoregressive (HSAR) model

- Consider the HSAR model:

$$y_{it} = \psi_i \sum_{j=1}^{N} w_{ij} y_{jt} + \varepsilon_{it}, \; i = 1, ..., N; t = 1, ..., T \tag{57}$$

  where $\boldsymbol{w}_i' \boldsymbol{y}_t = \sum_{j=1}^{N} w_{ij} y_{jt}$ with $\boldsymbol{y}_t = (y_{1t}, ..., y_{Nt})'$.
- $\boldsymbol{w}_i = (w_{i1}, ..., w_{iN})'$ with $w_{ii} = 0$ denotes an $N \times 1$ non-stochastic vector.
- Stacking the observations on individual units for each t:

$$(\boldsymbol{I}_N - \boldsymbol{\Psi} \boldsymbol{W}) \boldsymbol{y}_t = \boldsymbol{\varepsilon}_t, \; t = 1, ..., T$$

  where $\Psi = diag(\boldsymbol{\psi})$ with $\boldsymbol{\psi} = (\psi_1, ..., \psi_N)'$.

- Define: $\boldsymbol{S}(\psi) = (\boldsymbol{I}_N - \boldsymbol{\Psi W})$ and $\boldsymbol{S}_0 = (\boldsymbol{I}_N - \boldsymbol{\Psi}_0 \boldsymbol{W})$.
- Under the condition that $\boldsymbol{S}(\psi) = (\boldsymbol{I}_N - \boldsymbol{\Psi W})$ is non-singular, (57) can be expressed as

$$\boldsymbol{y}_t = \boldsymbol{S}^{-1}(\psi)\,\boldsymbol{\varepsilon}_t;\ t = 1, ..., T \tag{58}$$

- The (quasi) log-likelihood function can be written as

$$\begin{aligned} \ell(\boldsymbol{\theta}) \quad = \quad & -\frac{NT}{2}\ln 2\pi - \frac{NT}{2}\ln\sigma^2 + T\ln|\mathbf{S}(\psi)| \\ & -\frac{T}{2\sigma^2}\left[\mathbf{S}(\psi)'\,\mathbf{S}(\psi)\,\hat{\Sigma}_T\right] \end{aligned} \tag{59}$$

where $\boldsymbol{\theta} = (\psi', \sigma^2)'$.

- **Proposition 2** Consider the HSAR model (57) and suppose that (a) Assumptions 1 to 5 hold, (b) the invertibility condition (58) is met, (c) the $N \times N$ information matrix

$$\mathbf{H}_{11,2} = \left(\mathbf{G}_0 \odot \mathbf{G}_0'\right) + diag\left(\mathbf{G}_0\mathbf{G}_0'\right) - \frac{2}{N} diag\left(\mathbf{G}_0\right) \boldsymbol{\tau}_N' \boldsymbol{\tau}_N diag\left(\mathbf{G}_0'\right)$$

is full rank, where $\mathbf{G}_0 = \mathbf{W}(\mathbf{I}_N - \Psi\mathbf{W})^{-1}$, $\boldsymbol{\Psi}_0 = diag(\boldsymbol{\psi}_0)$, $\boldsymbol{\psi}_0 = (\psi_{10}, ..., \psi_{N0})'$, and (d) $\varepsilon_{it} \sim IIDN(0, \sigma_0^2)$. The MLE of $\boldsymbol{\psi}_0$ has the following asymptotic distribution:

$$\sqrt{T}\left(\hat{\boldsymbol{\psi}}_T - \boldsymbol{\psi}_0\right) \rightarrow_d N\left(0, AsyVar\left(\hat{\boldsymbol{\psi}}_T\right)\right)$$

$$AsyVar\left(\hat{\boldsymbol{\psi}}_T\right) = \begin{bmatrix} \left(\mathbf{G}_0 \odot \mathbf{G}_0'\right) + diag\left(\mathbf{G}_0\mathbf{G}_0'\right) \\ -\frac{2}{N} diag\left(\mathbf{G}_0\right) \boldsymbol{\tau}_N' \boldsymbol{\tau}_N diag\left(\mathbf{G}_0'\right) \end{bmatrix}^{-1}$$

- HSAR extended to include exogenous regressors and heteroskedastic errors:

$$y_{it} = \psi_i \sum_{j=1}^{N} w_{ij} y_{jt} + \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \varepsilon_{it} \tag{60}$$

- $\boldsymbol{x}_{it} = (x_{i1,t}, ..., x_{ik,t})'$ is a $k \times 1$ vector of exogenous regressors with $\boldsymbol{\beta}_i = (\beta_{i1}, ..., \beta_{ik})'$.
- We allow ε_{it} to be cross-sectionally heteroskedastic, $Var(\varepsilon_{it}) = \sigma_i^2$ for $i = 1, ..., N$.
- Stacking by individual units for each t, (60) becomes

$$\boldsymbol{y}_t = \boldsymbol{\Psi} \boldsymbol{W} \boldsymbol{y}_t + \boldsymbol{B} \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t \tag{61}$$

where $\boldsymbol{\Psi} = diag(\boldsymbol{\psi})$, $\boldsymbol{\psi} = (\psi_1, ..., \psi_N)'$, $\boldsymbol{B} = diag\left(\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_N'\right)'$, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$.
- (61) can be written as

$$\boldsymbol{y}_t = \left(\boldsymbol{I}_N - \boldsymbol{\Psi} \boldsymbol{W}\right)^{-1} \left(\boldsymbol{B} \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t\right)$$

- The log-likelihood function can be written as

$$\ell(\boldsymbol{\theta}) = -\frac{NT}{2}\ln 2\pi - \frac{T}{2}\sum_{i=1}^{N}\ln\sigma_i^2 + T\ln\left|\boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}\right|$$

$$-\frac{1}{2}\sum_{t=1}^{T}\left[\left(\boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}\right)\boldsymbol{y}_t - \boldsymbol{B}\boldsymbol{x}_t\right]'\boldsymbol{\Sigma}_\varepsilon\left[\left(\boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}\right)\boldsymbol{y}_t - \boldsymbol{B}\boldsymbol{x}_t\right]$$

- It is convenient to write the LF as

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= -\frac{NT}{2}\ln 2\pi - \frac{T}{2}\sum_{i=1}^{N}\ln\sigma_i^2 + T\ln\left|\boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}\right| \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\frac{\left(\boldsymbol{y}_i - \psi_i\boldsymbol{y}_i^* - \boldsymbol{X}_i\boldsymbol{\beta}_i\right)'\left(\boldsymbol{y}_i - \psi_i\boldsymbol{y}_i^* - \boldsymbol{X}_i\boldsymbol{\beta}_i\right)}{\sigma_i^2}\end{aligned}$$

where $\boldsymbol{\theta} = \left(\psi, \boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_N', \sigma_1^2, ..., \sigma_N^2\right)'$, $\boldsymbol{\beta}_i = (\beta_{i1}, ..., \beta_{iN})'$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT})'$ is the $T \times k$ matrix of regressors with $\boldsymbol{x}_{it} = (x_{i1,t}, ..., x_{ik,t})'$, $\boldsymbol{y}_i = (y_{i1}, ..., y_{iT})'$, $\boldsymbol{y}_i^* = (y_{i1}^*, ..., y_{iT}^*)'$.

- **Proposition 3** Consider the HSAR model (60) and suppose that (a) Assumptions 1, 4, 5, and 6, 7, and 8 hold, (b) the invertibility condition (58) is met, (c) $\lambda_{min}\left(\tilde{\mathbf{H}}_{11,2}\right) > \epsilon > 0$ for all N, where $\tilde{\mathbf{H}}_{11,2}$ is the $N \times N$ matrix

$$\begin{aligned}\tilde{\mathbf{H}}_{11,2} &= (\mathbf{G}_0 \odot \mathbf{G}_0') + diag\left(-g_{0,ii} + \sum_{s=1,s\neq i}^{N} \frac{\sigma_s^2}{\sigma_i^2} g_{0,is}^2\right) \\ &+ diag\left[\frac{1}{\sigma_i^2}\sum_{r=1}^{N}\sum_{s=1}^{N} g_{0,is} g_{0,ir} \beta_r' \left(\Sigma_{rs} - \Sigma_{ri}\Sigma_{ii}^{-1}\Sigma_{it}\right)\beta_s\right]\end{aligned}$$

$\mathbf{G}_0 = \mathbf{W}(\mathbf{I}_N - \Psi\mathbf{W})^{-1}$, $\mathbf{\Psi}_0 = diag(\boldsymbol{\psi}_0)$;
$\boldsymbol{\psi}_0 = (\psi_{10}, ..., \psi_{N0})'$, and (d) $\varepsilon_{it} \sim IIDN(0, \sigma_i^2)$. Then,

$$\sqrt{T}\left(\hat{\boldsymbol{\psi}}_T - \boldsymbol{\psi}_0\right) \rightarrow_d N\left(0, AsyVar\left(\hat{\boldsymbol{\psi}}_T\right)\right) \ as \ T \rightarrow \infty,$$

where $AsyVar\left(\hat{\boldsymbol{\psi}}_T\right) = \tilde{\mathbf{H}}_{11,2}^{-1}$.

BHP (2016) application to US housing prices

- Consider the heterogeneous equation:

$$x_{it} = \psi_i x_{it}^* + u_{it},\ i = 1, ..., N, t = 1, ..., T$$

  where $x_{it}^* = \mathbf{w}_i' \mathbf{x}_{\circ t}$.
- In the spatial econometrics literature it is assumed that all units have at least one neighbour, which ensures that $\mathbf{w}_i' \tau = 1$ for all i.
- It is possible for some units not to have any connection. Then, $x_{it}^* = 0$ and ψ_i is unidentified, and thus we set $\psi_i = 0$.

- In matrix notation we have

$$\mathbf{x}_{ot} = \mathbf{\Psi}\mathbf{W}\mathbf{x}_{ot} + \mathbf{u}_{ot}; \ \ for \ t = 1, ..., T;$$

where $\mathbf{\Psi} = diag\,(\psi)$, $\psi = (\psi_1, ..., \psi_N)'$, and $\sigma^2_{ui} = var\,(u_{it})$.

- An extension that incorporates richer temporal and spatial dynamics and accommodates negative and positive connections is:

$$\mathbf{x}_{ot} = \sum_{j=1}^{h_\lambda} \mathbf{\Lambda}_j \mathbf{x}_{ot-j} + \sum_{j=1}^{h_\psi^+} \mathbf{\Psi}_j^+ \mathbf{W}^+ \mathbf{x}_{ot-j} + \sum_{j=1}^{h_\psi^-} \mathbf{\Psi}_j^- \mathbf{W}^- \mathbf{x}_{ot-j} + \mathbf{u}_{ot}$$

where $h_\lambda = \max\,(h_{\lambda 1}, ..., h_{\lambda N})'$; $h_\psi^+ = \left(h_{\psi 1}^+, ..., h_{\psi N}^+\right)'$; $h_\psi^- = \left(h_{\psi 1}^-, ..., h_{\psi N}^-\right)'$; $\mathbf{\Lambda}_j$, $\mathbf{\Psi}_j^+$, $\mathbf{\Psi}_j^-$ are $N \times N$ diagonal matrices with λ_{ij}, ψ_{ij}^+ and ψ_{ij}^-.

- $\mathbf{W}^+$ and $\mathbf{W}^-$ are $N \times N$ network matrices for positive and negative connections such that $\mathbf{W} = \mathbf{W}^+ + \mathbf{W}^-$.

- We set $h_\lambda = h_\psi^+ = h_\psi^- = 1$ for simplicity.
- The concentrated log-likelihood function can be used:

$$\ell\left(\boldsymbol{\psi}_0^+, \boldsymbol{\psi}_0^-\right) \propto T \ln \left| I_N - \boldsymbol{\Psi}_0^+ \mathbf{W}^+ - \boldsymbol{\Psi}_0^- \mathbf{W}^- \right| - \frac{T}{2} \sum_{i=1}^{N} \left(\frac{1}{T} \tilde{\mathbf{x}}_i' \mathbf{M}_i \tilde{\mathbf{x}}_i \right)$$

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i - \boldsymbol{\psi}_{i0}^+ \mathbf{x}_i^+ - \boldsymbol{\psi}_{i0}^- \mathbf{x}_i^-$$

$$\mathbf{M}_i = \mathbf{I}_T - \mathbf{Z}_i \left(\mathbf{Z}_i' \mathbf{Z}_i\right)^{-1} \mathbf{Z}_i, \mathbf{Z}_i = \left(\mathbf{x}_{i,-1}, \mathbf{x}_{i,-1}^+, \mathbf{x}_{i,-1}^-\right)$$

$$\boldsymbol{\psi}_0^+ = \left(\psi_{10}^+, ..., \psi_{N0}^+\right)', \boldsymbol{\psi}_0^- = \left(\psi_{10}^-, ..., \psi_{N0}^-\right)',$$

- λ_1, ψ_1^+ and ψ_1^-, can be estimated by least squares applied to the individual equations conditional on ψ_{i0}^+ and ψ_{i0}^-.

- For inference the analysis must be carried out with respect to the unconcentrated LLF in terms of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_N'\right)'$, where $\boldsymbol{\theta}_i = \left(\psi_{i0}^+, \psi_{i0}^-, \psi_{i1}^+, \psi_{i1}^-, \lambda_{i1}, \sigma_{ui}^2\right)'$.
- The covariance matrix of $\hat{\boldsymbol{\theta}}_{ML}$ is computed as

$$\hat{\Sigma}_{\hat{\boldsymbol{\theta}}_{ML}} = \left[-\frac{1}{T}\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}_{ML}\right)}{\partial\hat{\boldsymbol{\theta}}_{ML}\partial\hat{\boldsymbol{\theta}}_{ML}'}\right]^{-1}$$

Motivations for STARDL

- How best to model spatial heterogeneity and diffusion/network dependence, possibly with observed or unobserved factors.
- In the literature dynamic spatial Durbin models are most general, but with homogeneous parameters; e.g. Elhorst (2012).
- Given the availability of large spatial datasets with a large time dimension it is important to explore spatial heterogeneity and diffusion dynamics in details.
- To date, the spatial parameters assumed to be homogeneous; ABP only considering heterogeneous parameters and proposing the QML but without any diffusion dynamics.

- We propose to generalise the spatial panel data model through the spatio-temporal autoregressive distributed lag (STARDL) model.
- We aim to:
- (a) develop the STARDL model and derive the QML and CF estimator;
- (b) provide the asymptotic theory;
- (c) develop the spatio-temporal dynamic and diffusion multipliers and apply network analyses.

The STARDL Model

- Consider the STARDL model with heterogeneous parameters:

$$\begin{aligned} y_{it} &= \phi_i y_{it-1} + \boldsymbol{\pi}'_{i0}\boldsymbol{x}_{it} + \boldsymbol{\pi}'_{i1}\boldsymbol{x}_{i,t-1} \\ &\quad + \phi^*_{i0} y^*_{it} + \phi^*_{i1} y^*_{it-1} + \boldsymbol{\pi}^{*\prime}_{i0}\boldsymbol{x}^*_{it} + \boldsymbol{\pi}^{*\prime}_{i1}\boldsymbol{x}^*_{i,t-1} + u_{it} \end{aligned} \tag{62}$$

- y_{it} is the dependent variable of the ith spatial unit at time t;
- $\boldsymbol{x}_{it} = (x^1_{it}, ..., x^K_{it})'$ is a $K \times 1$ vector of exogenous regressors with $\boldsymbol{\pi}_{i0} = (\pi^1_{i0}, ..., \pi^K_{i0})'$.

- The spatial variables, y_{it}^* and $\boldsymbol{x}_{it}^*$, are defined by

$$y_{it}^* = \sum_{j=1}^{N} w_{ij} y_{jt} = \boldsymbol{w}_i \boldsymbol{y}_t \text{ with } \underset{N\times 1}{\boldsymbol{y}_t} = \left(y_{1t}, ..., y_{Nt}\right)',$$

$$\begin{aligned} \underset{K\times 1}{\boldsymbol{x}_{it}^*} &= (x_{it}^{1*}, ..., x_{it}^{k*})' = \left(\sum_{j=1}^{N} w_{ij} x_{jt}^1, ..., \sum_{j=1}^{N} w_{ij} x_{jt}^K \right)' \\ &= (\boldsymbol{w}_i \otimes \boldsymbol{\iota}_K)\, \boldsymbol{x}_t \text{ with } \underset{NK\times 1}{\boldsymbol{x}_t} = \left(\boldsymbol{x}_{1t}', ..., \boldsymbol{x}_{Nt}'\right)' \end{aligned}$$

  where $\boldsymbol{w}_i = (w_{i1}, ..., w_{iN})$ is a $1 \times N$ row vector of spatial weights and $\boldsymbol{\iota}_K$ is a $K \times 1$ vector of unity.
- $\boldsymbol{y}_t^* = (y_{1t}^*, ..., y_{Nt}^*)'$ and $\boldsymbol{x}_t^* = (\boldsymbol{x}_{1t}^{*\prime}, ..., \boldsymbol{x}_{Nt}^{*\prime})'$ can be expressed as

$$\underset{N\times 1}{\boldsymbol{y}_t^*} = \boldsymbol{W} \boldsymbol{y}_t \text{ and } \underset{NK\times 1}{\boldsymbol{x}_t^*} = (\boldsymbol{W} \otimes \boldsymbol{I}_K)\, \boldsymbol{x}_t \qquad (63)$$

  where $\boldsymbol{W}$ is the $N \times N$ matrix of the spatial weights.

The STARDL (p,q) model

- It is straightforward to develop the STARDL(p, q) model:

$$y_{it} = \sum_{h=1}^{p} \phi_{ih} y_{i,t-h} + \sum_{h=0}^{q} \boldsymbol{\pi}'_{ih} \boldsymbol{x}_{i,t-h} + \sum_{h=0}^{p} \phi^*_{ih} y^*_{i,t-h} + \sum_{h=0}^{q} \boldsymbol{\pi}'^*_{ih} \boldsymbol{x}^*_{i,t-h} + u_{it} \quad (64)$$

- If the lag orders p and q are selected sufficiently large, u_{it}'s are free from serial correlations.

- Stacking the individual regressions, (64), we have:

$$\boldsymbol{y}_t = \sum_{h=1}^{p} \boldsymbol{\Phi}_h \boldsymbol{y}_{t-h} + \sum_{h=0}^{q} \boldsymbol{\Pi}_h \boldsymbol{x}_{t-h} + \sum_{h=0}^{p} \boldsymbol{\Phi}_h^* \boldsymbol{W} \boldsymbol{y}_{t-h} \quad (65)$$
$$+ \sum_{h=0}^{q} \boldsymbol{\Pi}_h^* \left(\boldsymbol{W} \otimes \boldsymbol{I}_k\right) \boldsymbol{x}_{t-h} + \boldsymbol{u}_t$$

where for $h = 0, 1, ..., q$,

$$\underset{N\times N}{\boldsymbol{\Phi}_h} = \begin{bmatrix} \phi_{1h} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_{Nh} \end{bmatrix} \quad \underset{N\times N}{\boldsymbol{\Phi}_h^*} = \begin{bmatrix} \phi_{1h}^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_{Nh}^* \end{bmatrix}$$

$$\underset{N\times NK}{\boldsymbol{\Pi}_h} = \begin{bmatrix} \boldsymbol{\pi}'_{1h} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\pi}'_{Nh} \end{bmatrix}, \quad \underset{N\times NK}{\boldsymbol{\Pi}_h^*} = \begin{bmatrix} \boldsymbol{\pi}'^*_{1h} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\pi}'^*_{Nh} \end{bmatrix}$$

Stability conditions and assumptions

- We rewrite (64) compactly as

$$y_{it} = \phi^*_{i0} y^*_{it} + \boldsymbol{\theta}'_i \boldsymbol{\chi}_{it} + u_{it} \tag{66}$$

where $\boldsymbol{\chi}_{it} = \left(y_{i,t-1}, ..., y_{i,t-p}, y^*_{i,t-1}, ..., y^*_{i,t-p}, \boldsymbol{x}'_{it}, ..., \boldsymbol{x}'_{i,t-q}, \boldsymbol{x}^{*\prime}_{it}, ..., \boldsymbol{x}^{*\prime}_{i,t-q}, 1 \right)'$
and $\boldsymbol{\theta}_i = \left(\boldsymbol{\phi}'_i, \boldsymbol{\phi}^{*\prime}_i, \boldsymbol{\pi}'_i, \boldsymbol{\pi}'^*_i, \alpha_i \right)'$ with $\boldsymbol{\phi}_i = \left(\phi_{i1}, ..., \phi_{ip} \right)'$,
$\boldsymbol{\phi}^*_i = \left(\phi^*_{i1}, ..., \phi^*_{ip} \right)'$, $\boldsymbol{\pi}_i = \left(\boldsymbol{\pi}'_{i0}, ..., \boldsymbol{\pi}'_{iq} \right)'$,
$\boldsymbol{\pi}^*_i = \left(\boldsymbol{\pi}^{*\prime}_{i0}, ..., \boldsymbol{\pi}^{*\prime}_{iq} \right)'$.
- Stacking (66), we have

$$\boldsymbol{y}_t = \boldsymbol{\Phi}^*_0 \boldsymbol{W} \boldsymbol{y}_t + \boldsymbol{\Theta} \boldsymbol{\chi}_t + \boldsymbol{u}_t \tag{67}$$

where $\boldsymbol{\Phi}^*_0 = diag\left(\phi^*_{10}, ..., \phi^*_{N0} \right)$, $\boldsymbol{\Theta} = diag\left(\boldsymbol{\theta}'_1, ..., \boldsymbol{\theta}'_N \right)$, and
$\boldsymbol{\chi}_t = \left(\boldsymbol{\chi}'_{1t}, \cdots \boldsymbol{\chi}'_{Nt} \right)'$.

Assumption 1: $\{u_{it}\}$ are independent across i and t with zero mean, heterogeneous variance $\sigma_i^2 > 0$. $E\left|u_{it}\right|^{4+\epsilon} < \infty$.
Assumption 2: The true parameter vector $\left(\boldsymbol{\phi}_0^{*\prime}, \boldsymbol{\theta}', \boldsymbol{\sigma}'\right)'$ are in a compact set.
Assumption 3: The spatial weights matrix $\boldsymbol{W}$ is non-stochastic with zero diagonals and uniformly bounded for all N with absolute row and column sums.
Assumption 4: (a) as $N \to \infty$, $(\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1}$ exists for all N; or (b) for bounded N, the eigenvalues of $\boldsymbol{\Phi}_0^* \boldsymbol{W}$ lie inside the unit circle such that $\boldsymbol{S}\left(\boldsymbol{\Phi}_0^*\right) = \boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W}$ is invertible for all $\boldsymbol{\Phi}_0^* \in \Theta_{\Phi_0^*}$, where $\Theta_{\Phi_0^*}$ is the compact parameter space.

As a result of Assumption 4, we rewrite (65) as

$$\tilde{\boldsymbol{\Phi}}(L)\,\boldsymbol{y}_t = \tilde{\boldsymbol{\Pi}}_\ell(L)\,\boldsymbol{x}_t + \tilde{\boldsymbol{u}}_t, \tag{68}$$

where $\tilde{\boldsymbol{\Phi}}(z) = \boldsymbol{I} - \sum_{\ell=1}^{p} \tilde{\boldsymbol{\Phi}}_\ell z^\ell$, and $\tilde{\boldsymbol{\Pi}}_\ell(z) = \sum_{\ell=0}^{q} \tilde{\boldsymbol{\Pi}}_\ell z^\ell$ with
$\tilde{\boldsymbol{\Phi}}_\ell = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1}(\boldsymbol{\Phi}_\ell + \boldsymbol{\Phi}_\ell^* \boldsymbol{W})$,
$\tilde{\boldsymbol{\Pi}}_\ell = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1}[\boldsymbol{\Pi}_\ell + \boldsymbol{\Pi}_\ell^*(\boldsymbol{W} \otimes \boldsymbol{I}_K)]$, and
$\tilde{\boldsymbol{u}}_t = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1}\boldsymbol{u}_t$.

Assumption 5 (Time stability): The roots of the characteristic equation $\left|\tilde{\boldsymbol{\Phi}}(z)\right| = \left|\boldsymbol{I}_N - \sum_{\ell=1}^{p} \tilde{\boldsymbol{\Phi}}_\ell z^\ell\right|$ lie outside the unit circle.
We rewrite (68) as an infinite order MA:

$$\boldsymbol{y}_t = \tilde{\boldsymbol{\Phi}}(L)^{-1}\left(\sum_{\ell=0}^{q} \tilde{\boldsymbol{\Pi}}_\ell \boldsymbol{x}_{t-\ell} + \tilde{\boldsymbol{u}}_t\right) \equiv \sum_{\ell=0}^{\infty} \tilde{\boldsymbol{B}}_\ell \boldsymbol{x}_{t-\ell} + \sum_{\ell=0}^{\infty} \boldsymbol{B}_\ell \tilde{\boldsymbol{u}}_{t-\ell}, \tag{69}$$

where $\tilde{\boldsymbol{B}}(L) = \tilde{\boldsymbol{\Phi}}(L)^{-1}\tilde{\boldsymbol{\Pi}}_\ell(L)$.
Then, $\sum_{\ell=0}^{\infty} \|\boldsymbol{B}_\ell\|_1$ and $\sum_{\ell=0}^{\infty} \|\boldsymbol{B}_\ell\|_\infty$ are bounded by C.

Assumption 6: $E(\|\boldsymbol{x}_{it}\|^4) \leq C$ for all i and t, and $\boldsymbol{x}_{it}$ are independent of idiosyncratic errors u_{js} for all (i, j, t, s).
Assumption 7: Identification condition:
$p\lim_{N,T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left(\boldsymbol{\chi}_{it}' \boldsymbol{\chi}_{it}\right)$ is strictly positive definite with the largest eigenvalue bounded by C. At the true parameter values of $\left(\phi_0^*, \boldsymbol{\theta}, \boldsymbol{\Sigma}_u\right)$, the following matrix

$$\lim_{N,T\to\infty} \left\{ \frac{1}{N(T-q)} \left[\begin{array}{c} \sum_{t=q}^{T} E\left(\boldsymbol{V}_t' \boldsymbol{V}_t\right) \\ -\sum_{t=q}^{T} E\left(\boldsymbol{V}_t' \boldsymbol{\chi}_t\right) \left(\sum_{t=q}^{T} E\left(\boldsymbol{\chi}_t' \boldsymbol{\chi}_t\right)\right)^{-1} \sum_{t=q}^{T} E\left(\boldsymbol{\chi}\right. \end{array} \right. \right.$$

is positive definite, where $\boldsymbol{V}_t = \boldsymbol{G}\boldsymbol{\chi}_t\boldsymbol{\theta}$ with $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{S}^{-1}\left(\boldsymbol{\Phi}_0^*\right)$.

The QML and CF Estimation

- To deal with the endogeneity of y_{it}^{*}, we apply the QML and the control function approach.
- The QML estimator constructed as the optimiser of

$$\mathcal{L}\left(\boldsymbol{\phi}_0^{*}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_u\right) = -\frac{nT}{2}\ln\left(2\pi\right) - \frac{T}{2}\ln\left|\boldsymbol{\Sigma}_u\right| + T\ln\left|S(\boldsymbol{\Phi}_0^{*})\right| - \frac{1}{2}\boldsymbol{u}_t'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{u}_t \quad (70)$$

  where $\boldsymbol{\phi}_0^{*} = (\phi_{01}^{*}, \ldots, \phi_{0N}^{*})'$, $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_N'\right)'$ and $\boldsymbol{\Sigma}_u = diag\left(\sigma_1^2, ..., \sigma_N^2\right)$.
- By extending ABP, it is straightforward to derive Theorem 1.

- **Theorem 1** (**The QML estimator**) Consider the STARDL model (66) and suppose that (i) Assumptions 1-7 hold, and (ii) $\lambda_{min}\left(\tilde{\mathbf{H}}_{11,2}\right) > \epsilon > 0$ for all N, where $\tilde{\mathbf{H}}_{11,2}$ is the $N \times N$ matrix given by

$$\tilde{\boldsymbol{H}}_{11,2} = (\boldsymbol{G} \odot \boldsymbol{G}') + diag\left(-g_{ii} + \sum_{s=1,s\neq i}^{N} \frac{\sigma_s^2}{\sigma_i^2} g_{is}^2,\ i = 1, ..., N\right) + diag\left[\frac{1}{\sigma_i^2} \sum_{r=1}^{N} \sum_{s=1}^{N} g_{is} g_{ir} \boldsymbol{\theta}_r' \left(\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri}^{-1} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{it}\right) \boldsymbol{\theta}_s,\ i =$$

$\boldsymbol{G} = \boldsymbol{W}\left(\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W}\right)^{-1} = \{g_{ij}\}$, and $\boldsymbol{\Sigma}_{ij} = E\left(\boldsymbol{\chi}_{it} \boldsymbol{\chi}_{it}'\right)$. As $T \to \infty$,

$$\sqrt{T}\left(\hat{\phi}_0^* - \phi_0^*\right) \to_d N\left(0, AVar\left(\hat{\phi}_0^*\right)\right)$$

where $AVar\left(\hat{\phi}_0^*\right) = \tilde{\boldsymbol{H}}_{11,2}^{-1}$.

- Then, $\boldsymbol{\theta}_i$ and σ_i^2 can be estimated by LSE applied to individual equations conditional on $\hat{\phi}_{i0}^*$ for $i = 1, ..., N$.
- Let $\boldsymbol{\xi} = \left(\phi_0^{*\prime}, \boldsymbol{\theta}', \boldsymbol{\sigma}^{2\prime}\right)'$.
- As $T \to \infty$,

$$\begin{aligned} \sqrt{T}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\right) &\to {}_d N\left(0, AVar\left(\hat{\boldsymbol{\xi}}\right)\right) \\ AVar\left(\hat{\boldsymbol{\xi}}\right) &= \boldsymbol{H}_T^{-1}\left(\hat{\boldsymbol{\xi}}\right)\boldsymbol{J}_T\left(\hat{\boldsymbol{\xi}}\right)\boldsymbol{H}_T^{-1}\left(\hat{\boldsymbol{\xi}}\right) \end{aligned}$$

where

$$\boldsymbol{J}_T\left(\boldsymbol{\xi}\right) = \frac{-1}{T}\left(\frac{\partial \ell\left(\boldsymbol{\xi}\right)}{\partial \boldsymbol{\xi}}\right)\left(\frac{\partial \ell\left(\boldsymbol{\xi}\right)}{\partial \boldsymbol{\xi}}\right)'; \boldsymbol{H}_T\left(\boldsymbol{\xi}\right) = \frac{-1}{T}\frac{\partial^2 \ell\left(\boldsymbol{\xi}\right)}{\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}'}.$$

- Maximimising (70) is equivalent to maximising:

$$\mathcal{L}_C\left(\boldsymbol{\phi}_0^*\right) \propto T \ln \left|S(\boldsymbol{\Phi}_0^*)\right| - \frac{1}{2}\sum_{i=1}^{N}\left(\boldsymbol{y}_i - \phi_{i0}^*\boldsymbol{y}_i^*\right)' \boldsymbol{M}_{\chi_i}\left(\boldsymbol{y}_i - \phi_{i0}^*\boldsymbol{y}_i^*\right) \tag{71}$$

where $\boldsymbol{y}_i = (y_{i1}, ..., y_{iT})'$, $\boldsymbol{y}_i^* = (y_{i1}^*, ..., y_{iT}^*)'$, $\boldsymbol{M}_{\chi_i} = \boldsymbol{I}_T - \boldsymbol{\chi}_i\left(\boldsymbol{\chi}_i'\boldsymbol{\chi}_i\right)^{-1}\boldsymbol{\chi}_i'$, $\boldsymbol{\chi}_i = \left(\boldsymbol{y}_{i,-1}, ..., \boldsymbol{y}_{i,-p}, \boldsymbol{y}_{i,-1}^*, ..., \boldsymbol{y}_{i,-p}^*, \boldsymbol{X}_i, ..., \boldsymbol{X}_{i,-q}, \boldsymbol{X}_i^*, ..., \boldsymbol{X}_{i,-q}^*, \mathbf{1}_T\right)$.

- $\boldsymbol{\phi}_0^* = (\phi_{i0}^*, ..., \phi_{N0}^*)'$ can be estimated by

$$\hat{\boldsymbol{\phi}}_0^* = \arg \max_{\boldsymbol{\phi}_0^* \in \Theta_{\boldsymbol{\phi}_0^*}} L_C\left(\boldsymbol{\phi}_0^*\right)$$

- The repeated evaluation of the $N \times N$ matrix, $\boldsymbol{I}_N - \boldsymbol{\Phi}_0^*\boldsymbol{W}$, can still make the maximisation of (71) numerically burdensome for large N.

The CF Estimation

- Let $\boldsymbol{z}_{it}$ be the $L \times 1$ vector of exogenous variables:

$$\boldsymbol{z}_{it} = \left(\boldsymbol{z}_{it}^{1\prime}, \boldsymbol{z}_{it}^{2\prime}\right);$$

  $\boldsymbol{z}_{it}^{1} = \boldsymbol{\chi}_{it}$ the $L_1 \times 1$ vector of exogenous variables included and $\boldsymbol{z}_{it}^{2}$ the $L_2 \times 1$ vector of exogenous variables excluded.

- We run the reduced form regression of y_{it}^{*} on $\boldsymbol{z}_{it}$:

$$y_{it}^{*} = \boldsymbol{\varphi}_i^{\prime} \boldsymbol{z}_{it} + v_{it} \quad \text{with } E\left(\boldsymbol{z}_{it}^{\prime} v_{it}\right) = \mathbf{0} \tag{72}$$

- Apply the linear projection of u_{it} on v_{it}:

$$u_{it} = \rho_i v_{it} + e_{it} \tag{73}$$

  where $\rho_i = E\left(v_{it} u_{it}\right) / E\left(v_{it}^2\right)$.

- Replacing u_{it} by (73), we obtain the transformation:

$$y_{it} = \sum_{h=1}^{p} \phi_{ih} y_{i,t-h} + \sum_{h=0}^{q} \boldsymbol{\pi}'_{ih} \boldsymbol{x}_{i,t-h} + \sum_{h=0}^{p} \phi^{*}_{ih} y^{*}_{i,t-h} + \sum_{h=0}^{q} \boldsymbol{\pi}'^{*}_{ih} \boldsymbol{x}^{*}_{i,t-h} + \rho_i v_{it} + e_{it} \quad (74)$$

where v_{it} is the control variable, rendering e_{it} uncorrelated with y^{*}_{it} as well as v_{it} and other regressors.

The two-step procedure

- (i) Obtain the residuals, $\hat{v}_{it} = y_{it}^* - \hat{\boldsymbol{\varphi}}_i' \boldsymbol{z}_{it}$ from (72).
- (ii) Run the following augmented regression:

$$\begin{aligned} y_{it} &= \sum_{h=1}^{p} \phi_{ih} y_{i,t-h} + \sum_{h=0}^{q} \boldsymbol{\pi}_{ih}' \boldsymbol{x}_{i,t-h} + \sum_{h=0}^{p} \phi_{ih}^* y_{i,t-h}^* \quad (75) \\ &+ \sum_{h=0}^{q} \boldsymbol{\pi}_{ih}'^* \boldsymbol{x}_{i,t-h}^* + \rho_i \hat{v}_{it} + e_{it}^* \end{aligned}$$

where $e_{it}^* = e_{it}^* + \rho_i (\hat{\boldsymbol{\varphi}}_i - \boldsymbol{\varphi}_i)' \boldsymbol{z}_{it}$ depends on the sampling error in $\hat{\boldsymbol{\varphi}}_i$ unless $\rho_i = 0$.

We rewrite (74) compactly as

$$y_{it} = \boldsymbol{\beta}_i' \boldsymbol{q}_{it} + e_{it}, \; t = 1, ..., T, \tag{76}$$

where $\boldsymbol{q}_{it} = \left(y_{it}^*, \boldsymbol{z}_{it}^{1\prime}, 1, v_{it}\right)'$ denotes the tth row of the matrix of the regressors, $\boldsymbol{q}_i = (\boldsymbol{q}_{i1}', ..., \boldsymbol{q}_{iT}')'$, and $\boldsymbol{\beta}_i = \left(\phi_{i0}^*, \boldsymbol{\theta}_i', \rho_i\right)'$.

Assumption 8: There exist a set of instruments such that $E\left(\boldsymbol{z}_{it} u_{it}\right) = \mathbf{0}$, $T^{-1} \sum_{t=1}^{T} \boldsymbol{z}_{it} \boldsymbol{z}_{it}' \rightarrow_p \boldsymbol{Q}_{zz}$, and $T^{-1} \sum_{t=1}^{T} \boldsymbol{z}_{it} \left(\boldsymbol{y}_{it}^*, \boldsymbol{z}_{it}^{1\prime}\right) \rightarrow \boldsymbol{Q}_{z*}$.

Assumption 9 e_{it} in (74) is *iid* across all i, t with zero mean and heterogeneous variance, σ_{ei}^2. v_{it} is *iid* control variable with zero mean and variance, σ_{vi}^2. $\boldsymbol{q}_{it} e_{it}$ and $\boldsymbol{z}_{it} v_{it}$, are stationary and ergodic mixingales of size -1.

- **Theorem 2** Under Assumptions 1-9, as $T \to \infty$, the OLS estimator from (75) is consistent and asymptotically normally distributed as

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\right)) \to_d N\left(0, AVar\left(\hat{\boldsymbol{\beta}}_i\right)\right),$$

where

$$AVar\left(\hat{\boldsymbol{\beta}}_i\right) \to_p \hat{\sigma}_i^2 \left(\tilde{\boldsymbol{X}}_i{}' \tilde{\boldsymbol{X}}_i\right)^{-1},$$

where $\tilde{\boldsymbol{X}}'_{it} = \left(y^*_{i,t} - \hat{v}_{it}, \boldsymbol{z}^{1\prime}_{it}\right)$ denotes the t'th row of the matrix $\tilde{\boldsymbol{X}}_i = \left(\tilde{\boldsymbol{X}}'_{i1}, ..., \tilde{\boldsymbol{X}}'_{iT}\right)'$, and $\hat{\sigma}_i^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_{it}^2$, where $\hat{u}_{it} = \hat{e}_{it} + \hat{v}_{it}\hat{\rho}$.

The selection of internal IVs

- By modelling the spatial and dynamic effects jointly, we can obtain the valid IVs internally as follows: Under Assumption 4,

$$\boldsymbol{y}_t^* = \boldsymbol{G}\left[\begin{array}{c} \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \boldsymbol{y}_{t-\ell} + \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell^* \boldsymbol{W} \boldsymbol{y}_{t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\Pi}_\ell \boldsymbol{x}_{t-\ell} \\ + \sum_{\ell=0}^{q} \boldsymbol{\Pi}_\ell^* \left(\boldsymbol{W} \otimes \boldsymbol{I}_K\right) \boldsymbol{x}_{t-\ell} + \boldsymbol{\alpha} + \boldsymbol{u}_t \end{array}\right] \tag{77}$$

where $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{S}(\boldsymbol{\Phi}_0^*)^{-1}$. This suggests:

$$\left[\sum_{\ell=1}^{p} \mathbf{W}^2 \boldsymbol{y}_{t-\ell}, \sum_{\ell=1}^{p} \mathbf{W}^3 \boldsymbol{y}_{t-\ell}, \ldots, \sum_{\ell=0}^{q} \mathbf{W}^2 \boldsymbol{x}_{t-\ell}, \sum_{\ell=0}^{q} \mathbf{W}^3 \boldsymbol{x}_{t-\ell}, \ldots\right] \tag{78}$$

can be used as the IV for $\boldsymbol{y}_t^*$.

- Next, we can derive the additional IVs from the higher time lags as

$$\left[\boldsymbol{W}^2\boldsymbol{y}_{t-p-1}, \boldsymbol{W}^2\boldsymbol{y}_{t-p-2}, ..., \boldsymbol{W}\boldsymbol{x}_{t-p-1}, \boldsymbol{W}\boldsymbol{x}_{t-p-2}, ...\right] \quad (79)$$

- We employ the following set of IVs for y_{it}^* in the individual STARDL regression, (64):

$$\boldsymbol{z}_{it} = \left(\sum_{\ell=1}^{p} y_{i,t-\ell}^{**}, \sum_{\ell=1}^{p} y_{i,t-\ell}^{***}, ..., \sum_{\ell=0}^{q} x_{i,t-\ell}^{**}, \sum_{\ell=0}^{q} x_{i,t-\ell}^{***}, ... \right)$$

where $y_{i,t-\ell}^{**} = \sum_{j=1}^{N} w_{ij}^{(2)} y_{j,t-\ell}$, $y_{i,t-\ell}^{***} = \sum_{j=1}^{N} w_{ij}^{(3)} y_{j,t-\ell}$, $x_{i,t-\ell}^{**} = \sum_{j=1}^{N} w_{ij}^{(2)} x_{j,t-\ell}$ and $x_{i,t-\ell}^{***} = \sum_{j=1}^{N} w_{ij}^{(3)} x_{j,t-\ell}$ with $w_{ij}^{(2)}$ and $w_{ij}^{(3)}$ being the (i,j)th element of $\boldsymbol{W}^2$ and $\boldsymbol{W}^3$.

General Remarks on the STARDL approach

- The STARDL is more general and its system representation encompasses the (homogeneous) dynamic spatial Durbin model:

$$\begin{aligned} \boldsymbol{y}_t &= \phi \boldsymbol{y}_{t-1} + \pi_0 \boldsymbol{x}_t + \pi_1 \boldsymbol{x}_{t-1} + \phi_0^* \boldsymbol{W} \boldsymbol{y}_t + \phi_1^* \boldsymbol{W} \boldsymbol{y}_{t-1} \\ &\quad + \pi_0^* \left(\boldsymbol{W} \otimes \boldsymbol{\iota}_K \right) \boldsymbol{x}_t + \pi_1^* \left(\boldsymbol{W} \otimes \boldsymbol{\iota}_K \right) \boldsymbol{x}_{t-1} + \boldsymbol{u}_t \end{aligned}$$

- In practice difficult to provide meaningful interpretation on the homogeneous spatial parameter.
- Our proposed approach can deliver much more flexible and sensible interpretations.

- Requires large T, but no bias correction even for large N.
- We can accommodate the time-varying weight matrix.
- When we apply different weight matrices to y and x, we can also use $\boldsymbol{W}\boldsymbol{x}_t$ as an Internal IV.
- Estimation seems less sensitive to the sparsity condition usually imposed on the weight matrix, see MC results below.

The spatio-temporal dynamic multipliers

- Due to the large number of estimation results, it is important to provide the succinct summary outputs.
- First, straightforward to derive dynamic multipliers associated with unit changes in y_t^*, x_{it} and x_{it}^* on y_{it}.
- Rewrite the STARDL(p, q) model, (64) as

$$\phi_i(L)\, y_{it} = \phi_i^*(L)\, y_{it}^* + \boldsymbol{\pi}_i(L)\, \boldsymbol{x}_{it} + \boldsymbol{\pi}_i^*(L)\, \boldsymbol{x}_{it}^* + u_{it} \quad (80)$$

$$\begin{aligned}
\phi_i(L) &= 1 - \sum_{h=1}^{p} \phi_{ih} L^h;\ \phi_i^*(L) = 1 - \sum_{h=0}^{p} \phi_{ih}^* L^h; \\
\boldsymbol{\pi}_i(L) &= \sum_{h=0}^{q} \boldsymbol{\pi}_{ih}' L^h;\ \boldsymbol{\pi}_i^*(L) = \sum_{h=0}^{q} \boldsymbol{\pi}_{ih}^{*\prime} L^h.
\end{aligned}$$

- Premultiplying (80) by the inverse of $\phi_i(L)$, we obtain:

$$y_{it} = \tilde{\phi}_i^*(L)\, y_{it}^* + \tilde{\boldsymbol{\pi}}_i(L)\, \boldsymbol{x}_{it} + \tilde{\boldsymbol{\pi}}_i^*(L)\, \boldsymbol{x}_{it}^* + \tilde{u}_{it} \tag{81}$$

where $\tilde{u}_{it} = [\phi_i(L)]^{-1} u_{it}$,

$$\tilde{\phi}_i^*(L) \left(= \sum_{h=0}^{\infty} \tilde{\phi}_{ih}^* L^h \right) = [\phi_i(L)]^{-1} \tilde{\phi}_i^*(L)$$

$$\tilde{\boldsymbol{\pi}}_i(L) \left(= \sum_{h=0}^{\infty} \tilde{\boldsymbol{\pi}}'_{ih} L^h \right) = [\phi_i(L)]^{-1} \boldsymbol{\pi}_i(L)$$

$$\tilde{\boldsymbol{\pi}}_i^*(L) \left(= \sum_{h=0}^{\infty} \tilde{\boldsymbol{\pi}}_{ih}^{*\prime} L^h \right) = [\phi_i(L)]^{-1} \boldsymbol{\pi}_i^*(L)$$

- $\tilde{\phi}^*_{ij}$, $\tilde{\boldsymbol{\pi}}'_{ij}$ and $\tilde{\boldsymbol{\pi}}^{*\prime}_{ij}$ can be evaluated using the following recursive relationships for $j = 0, 1, ...$:

$$\tilde{\phi}^*_{ij} = \phi_{i1}\tilde{\phi}^*_{i,j-1} + \phi_{i2}\tilde{\phi}^*_{i,j-2} + \cdots + \phi_{i,j-1}\tilde{\phi}^*_{i1} + \phi_{ij}\tilde{\phi}^*_{i0} + \phi^*_{ij}$$

where $\phi_{ij} = 0$ for $j < 1$ and $\tilde{\phi}^*_{i0} = \phi^*_{i0}$, $\tilde{\phi}^*_{ij} = 0$ for $j < 0$;

$$\tilde{\boldsymbol{\pi}}'_{ij} = \phi_{i1}\tilde{\boldsymbol{\pi}}'_{i,j-1} + \phi_{i2}\tilde{\boldsymbol{\pi}}'_{i,j-2} + \cdots + \phi_{i,j-1}\tilde{\boldsymbol{\pi}}'_{i,1} + \phi_{ij}\tilde{\boldsymbol{\pi}}'_{i0} + \boldsymbol{\pi}'_{ij}$$

where $\tilde{\boldsymbol{\pi}}'_{i0} = \boldsymbol{\pi}'_{i0}$, $\tilde{\boldsymbol{\pi}}'_{ij} = 0$ for $j < 0$, and

$$\tilde{\boldsymbol{\pi}}^{*\prime}_{ij} = \phi_{i1}\tilde{\boldsymbol{\pi}}^{*\prime}_{i,j-1} + \phi_{i2}\tilde{\boldsymbol{\pi}}^{*\prime}_{i,j-2} + \cdots + \phi_{i,j-1}\tilde{\boldsymbol{\pi}}^{*\prime}_{i,1} + \phi_{ij}\tilde{\boldsymbol{\pi}}^{*\prime}_{i0} + \boldsymbol{\pi}'_{ij}$$

where $\tilde{\boldsymbol{\pi}}^{*\prime}_{i0} = \boldsymbol{\pi}^{*\prime}_{i0}$, $\tilde{\boldsymbol{\pi}}^{*\prime}_{ij} = 0$ for $j < 0$.

- The cumulative dynamic multiplier effects of y_{it}^{*}, $\boldsymbol{x}_{it}$ and $\boldsymbol{x}_{it}$ on $y_{i,t+h}$ evaluated as

$$m_{y_i}\left(y_i^{*}, H\right) = \sum_{h=0}^{H} \frac{\partial y_{i,t+h}}{\partial y_{it}^{*}} = \sum_{h=0}^{H} \tilde{\phi}_{ih}^{*}, \; H = 0, 1, \ldots$$

$$\boldsymbol{m}_{y_i}\left(\boldsymbol{x}_i, H\right) = \sum_{h=0}^{H} \frac{\partial y_{i,t+h}}{\partial \boldsymbol{x}_{it}} = \sum_{h=0}^{H} \tilde{\boldsymbol{\pi}}_{ih}', \quad H = 0, 1, \ldots$$

$$\boldsymbol{m}_{y_i}\left(\boldsymbol{x}_i^{*}, H\right) = \sum_{h=0}^{H} \frac{\partial y_{i,t+h}}{\partial \boldsymbol{x}_{it}^{*}} = \sum_{h=0}^{H} \tilde{\boldsymbol{\pi}}_{ih}^{*\prime}, \quad H = 0, 1, \ldots$$

- As $H \to \infty$,

$$m_{y_i}(y_i^*, H) \to \beta_{yi};\ \boldsymbol{m}_{y_i}(\boldsymbol{x}_i, H) \to \boldsymbol{\beta}'_{xi};\ \boldsymbol{m}_{y_i}(\boldsymbol{x}_i^*, H) \to \boldsymbol{\beta}^{*\prime}_{xi}$$

where β_{yi}, $\boldsymbol{\beta}_{xi}$ and $\boldsymbol{\beta}^*_{xi}$ are the long-run multipliers.

- Suppose that y_{it} is the domestic policy variable. An important feature is to capture three different dynamic adjustments from initial to the new equilibrium following an economic perturbation wrt domestic conditions (x_{it}), overseas conditions (x^*_{it}) and overseas policy decisions (y^*_{it}).
- We may apply the MGE of the dynamic multipliers to investigate the overall average pattern provided with the bootstrap-based confidence intervals.

The system diffusion multipliers

- We develop the spatial-temporal diffusion multipliers in terms of the spatial system (65), which can be rewritten as

$$\tilde{\mathbf{\Phi}}\left(L\right)\boldsymbol{y}_t = \tilde{\mathbf{\Pi}}\left(L\right)\boldsymbol{x}_t + \tilde{\boldsymbol{u}}_t, \tag{82}$$

where $\tilde{\boldsymbol{u}}_t = \left(\boldsymbol{I}_N - \mathbf{\Phi}_0^* \boldsymbol{W}\right)^{-1} \boldsymbol{u}_t$,

$$\tilde{\mathbf{\Phi}}\left(L\right) = \boldsymbol{I}_N - \sum_{j=1}^{p} \tilde{\mathbf{\Phi}}_j L^j,\ \tilde{\mathbf{\Phi}}_j = \left(\boldsymbol{I}_N - \mathbf{\Phi}_0^* \boldsymbol{W}\right)^{-1}\left(\mathbf{\Phi}_j + \mathbf{\Phi}_j^* \boldsymbol{W}\right)$$

$$\tilde{\mathbf{\Pi}}\left(L\right) = \sum_{j=0}^{q} \tilde{\mathbf{\Pi}}_j L^j,\ \tilde{\mathbf{\Pi}}_j = \left(\boldsymbol{I}_N - \mathbf{\Phi}_0^* \boldsymbol{W}\right)^{-1}\left[\mathbf{\Pi}_h + \mathbf{\Pi}_h^* \left(\boldsymbol{W} \otimes \boldsymbol{\iota}_K\right)\right]$$

- Premultiplying (82) by $\left[\tilde{\mathbf{\Phi}}(L)\right]^{-1}$, we obtain:

$$\boldsymbol{y}_t = \boldsymbol{B}(L)\boldsymbol{x}_t + \left[\tilde{\mathbf{\Phi}}(L)\right]^{-1}\tilde{\boldsymbol{u}}_t, \tag{83}$$

where $\boldsymbol{B}(L)\left(=\sum_{j=0}^{\infty}\boldsymbol{B}_j L^j\right) = \left[\tilde{\mathbf{\Phi}}(L)\right]^{-1}\tilde{\mathbf{\Pi}}(L)$.

- $\boldsymbol{B}_j$ can be evaluated recursively as

$$\boldsymbol{B}_j = \tilde{\mathbf{\Phi}}_1\boldsymbol{B}_{j-1}+\tilde{\mathbf{\Phi}}_2\boldsymbol{B}_{j-2}+\cdots+\tilde{\mathbf{\Phi}}_{j-1}\boldsymbol{B}_1+\tilde{\mathbf{\Phi}}_j\boldsymbol{B}_0+\tilde{\mathbf{\Pi}}_j,\ j=1,2,... \tag{84}$$

where $\boldsymbol{B}_0 = \tilde{\mathbf{\Pi}}_0$ and $\boldsymbol{B}_j = 0$ for $j < 0$.

- Cumulative diffusion multipliers can be evaluated as:

$$\boldsymbol{m}_x(H) = \sum_{h=0}^{H} \frac{\partial \boldsymbol{y}_{t+h}}{\partial \boldsymbol{x}'_t} = \sum_{h=0}^{H} \boldsymbol{B}_h, \;\; H = 0, 1, 2, \ldots$$

- For homogeneous spatial panels, LeSage and Pace (2009) propose average diagonal elements as own-partial derivatives. This includes some feedback effects that arise as a result of impacts passing through neighboring regions. They propose an average of off diagonal elements as a summary indirect effect.
- See also Debarsy et al. (2012) for dynamic space-time panel data, and LeSage and Chin (2016) for the case with heterogeneous spatial coefficients.
- It is not possible to separate out time and spillover dependence from diffusion effects.

- We follow Greenwood-Nimmo, Nguyen and Shin (2015, GNS) and apply the generalised connectedness measures.
- At horizon, h, one cross-tabulates the impacts of x_{jt}^{k} on the $N \times 1$ vector of $\boldsymbol{y}_t$ in the $N \times N$ connectedness matrix:

$$\mathbb{C} = \begin{bmatrix} \phi_{1\leftarrow 1} & \phi_{1\leftarrow 2} & \cdots & \phi_{1\leftarrow N} \\ \phi_{2\leftarrow 1} & \phi_{2\leftarrow 2} & \cdots & \phi_{2_1\leftarrow N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{N\leftarrow 1} & \phi_{N\leftarrow 2} & \cdots & \phi_{N\leftarrow N} \end{bmatrix} \tag{85}$$

- The main diagonal elements represent own-region impacts that arise from both time and spatial dependence.
- Off-diagonal elements reflect spillovers measuring contemporaneous cross-partial derivatives and diffusion measuring cross-partial derivatives that involve different periods.

- We start with the (cumulative) own-region impacts that arise from both time and spatial dependence:

$$H_{j\leftarrow j} = \phi_{j\leftarrow j} \tag{86}$$

- We write the cross-from or *spill-in* contribution as

$$F_{j\leftarrow \bullet} = \sum_{i=1, i\neq j}^{N} \phi_{j\leftarrow i} \tag{87}$$

- Similarly, define the total contributions *to* all other countries (or *spill-out* contributions) as

$$T_{\bullet\leftarrow j} = \sum_{i=1, i\neq j}^{N} \phi_{i\leftarrow j} \tag{88}$$

- The net directional connectedness is defined as

$$N_{\bullet\leftarrow j} = T_{\bullet\leftarrow j} - F_{j\leftarrow \bullet}. \tag{89}$$

- It is straightforward to develop the aggregate (non-directional) connectedness measures:

$$H = \sum_{j=1}^{N} H_{j\leftarrow j};\ S = \sum_{j=1}^{N} F_{j\leftarrow \bullet} = \sum_{j=1}^{N} T_{\bullet\leftarrow j}, \sum_{j=1}^{N} N_{\bullet\leftarrow j} = 0 \tag{90}$$

$$H + S = TOT_{\bullet\leftarrow\bullet} = \sum_{j=1}^{N} TOT_{j\leftarrow\bullet} \tag{91}$$

- H and S are the aggregate direct own-region impacts and the aggregate cross-spillover contribution.
- The sum of the aggregate heatwave and spillover measures, $TOT_{\bullet\leftarrow\bullet}$ accounts for all of the impacts in the entire system.

- We define a pair of indices to address 2 questions of interest: (i) 'how dependent is the j-th country on external conditions?' and (ii) 'to what extent does the j-th country influence/is the j-th country influenced by the system as a whole?'.
- We first propose the dependence index :

$$O_j = \frac{F_{j\leftarrow\bullet}}{W_{j\leftarrow\bullet} + F_{j\leftarrow\bullet}}, \; j = 1, ..., N, \; 0 \leq O_j \leq 1$$

As $O_j \rightarrow 1$, conditions in j dominated by external shocks.

- We develop the influence index:

$$I_j = \frac{N_{j\leftarrow\bullet}}{T_{j\leftarrow\bullet} + F_{j\leftarrow\bullet}},\ j = 1, ..., N,\ -1 \leq I_j \leq 1$$

- The j-th group is a net shock recipient if $-1 \leq I_j \leq 0$, a net shock transmitter if $0 \leq I_j \leq 1$.
- The coordinate pair (O_j, I_j) provides a representation of country i's role in the global system.
- A classic small open economy located close to $(1, -1)$ while an overwhelmingly dominant economy would exist in the locale of $(0, 1)$.

Monte Carlo Design

- The Monte Carlo exercise is based on the heterogeneous parameter STARDL(1,1) model:

$$\begin{aligned} y_{it} &= \phi_i y_{i,t-1} + \pi_{i0} x_{it} + \pi_{i1} x_{i,t-1} + \phi_{i0}^* y_{it}^* + \phi_{i1}^* y_{i,t-1}^* \\ &\quad + \pi_{i0}^* x_{it}^* + \pi_{i1}^* x_{i,t-1}^* + u_{it} \end{aligned}$$

  where $y_{it}^* = \sum_{j=1}^N w_{ij} y_{jt}$ and $x_{it}^* = \sum_{j=1}^N w_{ij} x_{jt}$ with $w_{ii} = 0$.
- $\mathbf{W}$ is a row-normalised b-nearest neighbours weighting matrix, with $b = (2, 4, 10)$, with null elements apart from the $b/2$ either side of the principle diagonal.
- ϕ_i, ϕ_{i0}^*, ϕ_{i1}^* are independent draws from $U(0, 0.4)$ while π_{i0}, π_{i1}, π_{i0}^*, π_{i1}^* are independent draws from $U(0, 1)$.

- We consider two experiments.
- In experiment 1 both regressors and exogenous variables are draws from $N(0, 1)$.
- Experiment 2 uses a set of serially correlated exogenous variables:

$$x_{i,t} = \rho_i x_{i,t-1} + v_{i,t},\ v_{i,t} \sim N\left(0, 1 - \rho_i^2\right), \tag{92}$$

where $\rho_i \sim U[0.4, 0.6]$, and heteroskedastic errors $u_{it} \sim N(0, \sigma_i^2)$, where $\sigma_i^2 = 0.5 + 0.25 \times \eta_i$ and $\eta_i \sim \chi_2^2$.

- We report the following statistics:
- Average bias $= N^{-1} \sum_{i=1}^{N} R^{-1} \sum_{j=1}^{R} (\hat{\alpha}_{ij} - \alpha_{i0})$;
- Average RMSE $= N^{-1} \sum_{i=1}^{N} \sqrt{R^{-1} \sum_{j=1}^{R} (\hat{\alpha}_{ij} - \alpha_{i0})^2}$.
- Average Size $= N^{-1} \sum_{i=1}^{N} R^{-1} \sum_{j=1}^{R} \mathbb{I}\left(\left|\frac{\hat{\alpha}_{ij}-\alpha_{i0}}{\sigma_\alpha}\right| > t_{0.975}\right)$, where $\mathbb{I}(.)$ is the indicator function and σ_α is the standard error.
- We consider $N = (25, 50, 75, 100)$ over $T = (50, 100, 200)$ with $R = 1,000$ replications.

- CF estimates based on the instrument set of $y^{**}_{i,t-1}$ and x^{**}_{it}, where $y^{**}_{i,t-1} = \boldsymbol{w}_i \boldsymbol{W} \boldsymbol{y}_{t-1}$ and $x^{**}_{it} = \boldsymbol{w}_i \boldsymbol{W} \boldsymbol{x}_t$ are the second spatial lags.
- The price for including extra instruments is potential multi-collinearity and we find that two instruments was often the best choice.
- QML initial values were provided by (inconsistent) OLS estimates.
- The exogenous variables were concentrated out and leaving an iteration over the N vector $\boldsymbol{\phi}^*$, with estimates of other parameters recovered by least squares regression conditional on $\hat{\boldsymbol{\phi}}^*$.

- Both CF and QML estimates perform reasonably well in terms of bias.
- Bias falls as T increases and is not greatly affected by N, supporting our theoretical prediction.
- For small T the QMLE has a slightly lower bias, but as T becomes large the results of the two estimators are comparable with the CF estimator having a strong computational advantage.
- The estimates of all parameters have biases of similar magnitude with ϕ^*, π and π^*, exhibiting the lower biases than their equivalents on lagged terms and ϕ_1. This is not too surprising as the time dynamics open more channels through which a time lagged variable may potentially impact on a y_{it}.
- Both methods robust to serial dependence in exogenous variables - slightly improves some control function estimates because instrument set closer to optimum.

RMSE

- The QMLE is the more efficient estimator.
- This is not too surprising: it should be.

Size

- CF estimates tend to be under-sized, particularly for ϕ^* and ϕ_1^*.
- The size for the QMLE is much closer to 5 per cent but the others are slightly over-sized.
- Performance deteriorates with the number of connections, b, with the CF becoming more under-sized while QML becomes over-sized.
- For biases and RMSE there is no noticeable deterioration in the performance of either estimator as b rises.
- The QMLE recovers more successfully as T rises and is always within a percentage point for $T = 200$.

Civilian Casualties in Iraqi War during 2004-2010

- The 2003 Iraq War is rarely considered as a success in terms of its prolonged duration and the considerable human cost.
- The U.S. Department of Defense compiled civilian deaths but also insurgents and Iraqi security forces killed by armed violence with exact location and time.
- This Pentagon's archive (Iraq War Logs) would provide a rare opportunity to infer the intensity of armed violence and its spatio-temporal diffusion during the war period.

- Thomas Schelling's noncooperative game framework (The Strategy of Conflict, 1960) and Ivan Arreguin-Toft's strategic interaction thesis (2001) suggest that warring groups choose between killing civilians and battling with armed opponents.
- Weaker actors (commonly being rebels) tend to target civilians to increase the probability of winning in asymmetric war, and to build up local support.
- The spatio-temporal analysis of civilian victimisation would shed lights on the effects of US forces' methods in acknowledging civilian loss upon the ensuing development of war.

Major findings of STARDL estimation

- It is based on governorate level STARDL(1,1) estimation with $y =$ civilian casualties and $x =$ enemy casualties and with $N = 18$ and $T = 70$.
- A row standardised inverse distance $\boldsymbol{W}$ matrix has been employed. The results appear to be qualitatively similar when using different W matrices.
- The system is stable with the largest eigenvalue of time stability matrix being 0.866.
- LR spill-over effects (i.e. y^* on y) are positive for all 18 provinces; a rise in civilian casualties in neighbours leads to a cumulative increase in civilian deaths in any province.
- The LR effects of x on y are more diverse, but positive for 13 provinces and negative for 5, partially due to the strategic decision to target other provinces with a lower level of security infrastructure.

- Those of x^* on y are positive for 10 provinces only.
- The most striking is Basrah, which appears the most open to the impact of insurgent deaths outside its locality, implying that armed violence against civilians in the neighbouring provinces lead to an increase in civilian casualties in Basrah.
- Our model highlights an interesting comparison between Basrah and Baghdad.
- Baghdad shows strong temporal diffusion (i.e., y_{t-1} on y_t).
- Baghdad, the major military post for the US forces, was heavily armoured due to severe insurgency against military personnel, civilians and foreign contractors.

Cumulative dynamic and diffusion multipliers

- CDMs for Basrah and Baghdad show strong opposite patterns.
- CDMs of y^* in Basrah are strongly positive and reach the long-run elasticity of 2.5 within 2-3 months, reflecting an openness to the effects of casualties elsewhere.
- Those of Baghdad are initially negative, converging to the small positive value gradually, reflecting the tighter security.
- CDMs with respect to x show a similar pattern: insurgent deaths in other provinces making Baghdad relatively safer.
- MGE CDMs tend to the intermediate figures.

- Again direct, spill_in and spill_out CDMs of Basrah and Baghdad show strong opposite patterns.
- Direct CDMs of y to x in Basrah are negative and reach the long-run value quickly while those of Baghdad are positive, converging to the long-run value gradually.
- Spill_in CDMs of y in Basrah are substantially positive while those of Baghdad are slightly negative.
- Spill_out CDMs in Basrah are slightly negative while those of Baghdad are large and positive.
- Net effects show a mirror image, displaying that net effects of Baghdad are large and positive while those of Basrah are large and negative.

- The UK and US had been the major political stakeholders which invaded Iraq in spite of strong disagreement with the UN.
- The military forces of the two countries appear to have faced substantially different challenges, with the UK forces being tested by spatial diffusion of armed violence, and the US forces challenged by self-generated temporal persistency in violence.
- Our analysis indicates that much of the climate for the insurgency was made in Baghdad but that its greatest effects were felt, with relatively little delay, in Basrah.

Joint Modelling of Spatial Dependence and Unobserved Factors

- Recently, a few studies attempted to develop a combined approach that can accommodate both weak and strong CSD, e.g. Bailey et al. (2016) and Mastromarco et al. (2015).
- Shi and Lee (2017), Bai and Li (2015) and Kuersteiner and Prucha (2015) have also developed the framework for jointly modelling spatial effects and interactive effects.

STARDL models with observed common factors

- Consider the STARDL(p, q) model with the $G \times 1$ vector of observed global factors, $\boldsymbol{g}_t = \left(g_t^1, ..., g_t^G\right)'$ (e.g. oil prices):

$$
\begin{aligned}
y_{it} \quad = \quad & \sum_{h=1}^{p} \phi_{ih} y_{i,t-h} + \sum_{h=0}^{q} \boldsymbol{\pi}'_{ih} \boldsymbol{x}_{i,t-h} + \sum_{h=0}^{p} \phi^*_{ih} y^*_{i,t-h} \quad (93) \\
& + \sum_{h=0}^{q} \boldsymbol{\pi}'^*_{ih} \boldsymbol{x}^*_{i,t-h} + \sum_{h=0}^{q} \boldsymbol{\psi}'_{ih} \boldsymbol{g}_{t-h} + u_{it}
\end{aligned}
$$

- Straightforward to derive the QML and CF estimators and develop the individual dynamic multipliers, $\boldsymbol{m}_g^H$ and the system diffusion multipliers, $\boldsymbol{d}_g^H$ esp. with respect to $\boldsymbol{g}_t$.
- The difference between $\boldsymbol{d}_{\boldsymbol{g}}^H$ and $\boldsymbol{m}_{\boldsymbol{g}}^H$ may indicate the additional spatial impacts of $\boldsymbol{g}_t$ at each forecast horizon.

STARDL models with unobserved factors

- Consider the STARDL(p, q) model with an $r \times 1$ vector of unobserved common factors, $\boldsymbol{f}_t = (f_{1t}, ..., f_{rt})'$:

$$y_{it} = \sum_{\ell=1}^{p} \phi_{i\ell} y_{i,t-\ell} + \sum_{\ell=0}^{p} \phi_{i\ell}^{*} y_{i,t-\ell}^{*} + \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}' \boldsymbol{x}_{i,t-\ell} \quad (94)$$
$$+ \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}'^{*} \boldsymbol{x}_{i,t-\ell}^{*} + \alpha_i + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + u_{it},$$

  where $\boldsymbol{\lambda}_i = (\lambda_{1,i}, ..., \lambda_{r,i})'$is the heterogenous loadings.
- We develop the two estimation procedures.
- In the first approach we allow $\boldsymbol{x}_{it}$'s to be correlated arbitrarily with the common factors and/or the factor loadings.
- In the second approach, we assume that $\boldsymbol{x}_{it}$'s follow the VAR processes but also share the unobserved common factors.

- We write (94) compactly as

$$y_{it} = \phi_{i0}^{*} y_{it}^{*} + \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it} + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + u_{it} \tag{95}$$

where $\boldsymbol{\theta}_i = \left(\boldsymbol{\phi}_i', \boldsymbol{\phi}_i^{*\prime}, \boldsymbol{\pi}_i', \boldsymbol{\pi}_i'^{*}\right)'$ and
$\boldsymbol{\chi}_{it} = \left(\boldsymbol{y}_{i,-\ell}', \boldsymbol{y}_{i,-\ell}^{*\prime}, \boldsymbol{x}_{i,-\ell}', \boldsymbol{x}_{i,-\ell}^{*\prime}, 1\right)'$ with $\boldsymbol{\phi}_i = \left(\phi_{i1}, ..., \phi_{ip}\right)'$,
$\boldsymbol{\phi}_i^{*} = \left(\phi_{i1}^{*}, ..., \phi_{ip}^{*}\right)'$, $\boldsymbol{\pi}_i = \left(\boldsymbol{\pi}_{i0}', ..., \boldsymbol{\pi}_{iq}'\right)'$,
$\boldsymbol{\pi}_i^{*} = \left(\boldsymbol{\pi}_{i0}^{*\prime}, ..., \boldsymbol{\pi}_{iq}^{*\prime}\right)'$, and $\boldsymbol{y}_{i,-p} = \left(y_{i,t-1}, ..., y_{i,t-p}\right)'$,
$\boldsymbol{y}_{i,-p}^{*} = \left(y_{i,t-1}^{*}, ..., y_{i,t-p}^{*}\right)'$, $\boldsymbol{x}_{i,-q} = \left(\boldsymbol{x}_{it}', ..., \boldsymbol{x}_{i,t-q}'\right)'$,
$\boldsymbol{x}_{i,-q}^{*} = \left(\boldsymbol{x}_{it}^{*\prime}, ..., \boldsymbol{x}_{i,t-q}^{*\prime}\right)'$.

- Stacking (95), we have

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_0^{*} \boldsymbol{W} \boldsymbol{y}_t + \boldsymbol{\Theta} \boldsymbol{\chi}_t + \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{u}_t \tag{96}$$

where $\boldsymbol{\Phi}_0^{*} = diag\left(\phi_{10}^{*}, ..., \phi_{N0}^{*}\right)$, $\boldsymbol{\Theta} = diag\left(\boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_N'\right)$,
$\boldsymbol{\chi}_t = \left(\boldsymbol{\chi}_{1t}', ...\boldsymbol{\chi}_{Nt}'\right)'$ and $\boldsymbol{\Lambda} = \left[\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_N\right]'$.

The IPC-QML estimator

- We allow $\boldsymbol{x}_{it}$ to be arbitrarily correlated with $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$, and estimate both $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$ as parameters.
- **Assumption F1**: $\boldsymbol{f}_t$ are random and independent of u_{is} for all t and s.
- **Assumption F2**: The factor loadings $\boldsymbol{\lambda}_i$ are random such that $E\left(\|\Gamma_i\|^4\right) \leq C$ for all i and $N^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_{u\ u}\boldsymbol{\Lambda} \rightarrow_p \Omega_r$, where $\boldsymbol{\Sigma}_u = diag\left(\sigma_{u1}^2, ..., \sigma_{uN}^2\right)$, and $\boldsymbol{\Omega}_r$ is positive definite. $\boldsymbol{\lambda}_i$s are independent of the idiosyncratic errors u_{jt} for all i and j.

- The (quasi) log-likelihood function of (96) can be derived as

$$\mathcal{L}\left(\boldsymbol{\Phi}_0^*, \boldsymbol{\Theta}, \boldsymbol{F}\right) = -\frac{NT}{2}\ln 2\pi - \frac{T}{2}\sum_{i=1}^{N}\ln\sigma_i^2 + T\ln\left|\boldsymbol{I}_N - \boldsymbol{\Phi}_0^*\boldsymbol{W}\right| \tag{97}$$

$$-\frac{1}{2}\sum_{t=1}^{T}\left[\left(\boldsymbol{I}_N - \boldsymbol{\Phi}_0^*\boldsymbol{W}\right)\boldsymbol{y}_t - \boldsymbol{\Theta}\boldsymbol{\chi}_t - \boldsymbol{\Lambda}\boldsymbol{f}_t\right]'\Sigma_{uu}^{-1}\left[\left(\boldsymbol{I}_N - \boldsymbol{\Phi}_0^*\boldsymbol{W}\right)\boldsymbol{y}_t - \boldsymbol{\Theta}\right.$$

- Given $\boldsymbol{\Phi}_0^*$, $\boldsymbol{\Theta}$, $\boldsymbol{\Lambda}$, $\boldsymbol{f}_t$ maximize $\mathcal{L}\left(\boldsymbol{\Phi}_0^*, \boldsymbol{\Theta}, \boldsymbol{F}\right)$ at

$$\boldsymbol{f}_t = \left(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_{uu}^{-1}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_{uu}^{-1}\left[\left(\boldsymbol{I}_N - \boldsymbol{\Phi}_0^*\boldsymbol{W}\right)\boldsymbol{y}_t - \boldsymbol{\Theta}\boldsymbol{\chi}_t\right]. \tag{98}$$

- Substituting (98) in (97), we obtain:

$$\mathcal{L}\left(\mathbf{\Phi}_0^*, \mathbf{\Theta}\right) = -\frac{NT}{2}\ln 2\pi - \frac{T}{2}\sum_{i=1}^{N}\ln\sigma_i^2 + T\ln\left|\boldsymbol{I}_N - \mathbf{\Phi}_0^*\boldsymbol{W}\right| \tag{99}$$

$$-\frac{1}{2}\sum_{t=1}^{T}\left[\left(\boldsymbol{I}_N - \mathbf{\Phi}_0^*\boldsymbol{W}\right)\boldsymbol{y}_t - \mathbf{\Theta}\boldsymbol{\chi}_t\right]'\boldsymbol{M}_{uu}\left[\left(\boldsymbol{I}_N - \mathbf{\Phi}_0^*\boldsymbol{W}\right)\boldsymbol{y}_t - \mathbf{\Theta}\boldsymbol{\chi}_t\right]$$

where

$$\boldsymbol{M}_{uu} = \mathbf{\Sigma}_{uu}^{-1} - \mathbf{\Sigma}_{uu}^{-1}\mathbf{\Lambda}\left(\mathbf{\Lambda}'\mathbf{\Sigma}_{uu}^{-1}\mathbf{\Lambda}\right)^{-1}\mathbf{\Lambda}'\mathbf{\Sigma}_{uu}^{-1} = \mathbf{\Sigma}_{uu}^{-1} - \frac{1}{N}\mathbf{\Sigma}_{uu}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}'\mathbf{\Sigma}_{uu}^{-1}$$

with $\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Sigma}_{uu}^{-1}\mathbf{\Lambda} = \boldsymbol{I}_r$.

- The QML estimator of $\left(\mathbf{\Phi}_0^*, \mathbf{\Theta}\right)$ is defined by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta}\mathcal{L}\left(\mathbf{\Phi}_0^*, \mathbf{\Theta}\right)$$

The IPC-QML algorithm

- **Step 1**: Given $\hat{\Phi}_0^{*(s)}$, $\hat{\Theta}^{(s)}$ and $\hat{\Sigma}_{uu}^{(s)}$, we estimate $\hat{\Lambda}^{(s+1)}$ as the first r eigenvectors associated with the first r largest eigenvalues of the $N \times N$ matrix, $\hat{G} =$

$$\left[\frac{1}{NT}\sum_{t=1}^{T}\left(y_t - \hat{\Phi}_0^{*(s)} W y_t - \hat{\Theta}^{(s)} \chi_t\right)\left(y_t - \hat{\Phi}_0^{*(s)} W y_t - \hat{\Theta}^{(s)} \chi_t\right)'\right.$$

and $\hat{f}_t^{(s+1)}$ by

$$\hat{f}_t^{(s+1)} = \frac{1}{N}\hat{\Lambda}^{(s+1)\prime}\left(\hat{\Sigma}_{uu}^{(s)}\right)^{-1}\left(y_t - \hat{\Phi}_0^{*(s)} W y_t - \hat{\Theta}^{(s)} \chi_t\right)$$

We construct

$$\hat{C}_{it}^{(s+1)} = \hat{\lambda}_i^{(s+1)\prime}\hat{f}_t^{(s+1)} \text{ and } \hat{C}_i^{(s+1)} = \left(\hat{C}_{i1}^{(s+1)}, \ldots, \hat{C}_{iT}^{(s+1)}\right)'$$

- **Step 2**: Given $\hat{C}_{it}^{(s+1)}$ and $\hat{\boldsymbol{C}}_i^{(s+1)}$, update $\hat{\boldsymbol{\Sigma}}_{uu}^{(s+1)}$ and $\hat{\boldsymbol{\theta}}_i^{(s+1)}$:

$$\left(\hat{\sigma}_i^{2)}\right)^{(s+1)} = \frac{1}{T}\sum_{t=1}^{T}\left(\left(y_{it} - \hat{C}_{it}^{(s+1)}\right) - \hat{\phi}_{i0}^{*(s)} y_{it}^{*} - \hat{\boldsymbol{\theta}}_i^{(s)\prime}\boldsymbol{\chi}_{it}\right)^2$$

$$\hat{\boldsymbol{\theta}}_i^{(s+1)} = \left(\boldsymbol{\chi}_i'\boldsymbol{\chi}_i\right)^{-1}\left(\boldsymbol{\chi}_i'\boldsymbol{\chi}_i\right)\left\{\left(\boldsymbol{y}_i - \hat{\boldsymbol{C}}_i^{(s+1)}\right) - \hat{\phi}_{i0}^{*(s)}\boldsymbol{y}_i^{*}\right\}, i = 1, ..., N.$$

- Finally, update $\hat{\boldsymbol{\Phi}}_0^{*(s+1)}$ by maximizing $\mathcal{L}\left(\boldsymbol{\Phi}_0^*, \boldsymbol{\Theta}\right)$ in (99) directly with respect to $\boldsymbol{\Phi}_0^*$ at $\boldsymbol{\Lambda} = \hat{\boldsymbol{\Lambda}}^{(s+1)}, \boldsymbol{\Sigma}_{uu} = \hat{\boldsymbol{\Sigma}}_{uu}^{(s+1)}$, and $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}^{(s+1)}$.
- We repeat Steps 1 and 2 until convergence.
- There may be a bias term of order $O_p\left(N^{-1}\right)$ as IPC-QML estimator is not consistent under fixed N, e.g. Lu (2017).

The QML-EM estimator

- We assume that $\boldsymbol{x}_{it}$ follows the VAR(p) process and shares the same unobserved factors, $\boldsymbol{f}_t$:

$$\boldsymbol{x}_{it} = \sum_{\ell=1}^{p} \Psi_i \boldsymbol{x}_{i,t-\ell} + \boldsymbol{b}_i + \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \boldsymbol{v}_{it} \tag{100}$$

- (94) and (100) can be written as

$$\begin{bmatrix} y_{it} - \sum_{\ell=1}^{p} \phi_{i\ell} y_{i,t-\ell} - \sum_{\ell=0}^{p} \phi_{i\ell}^* y_{i,t-\ell}^* - \sum_{\ell=0}^{p} \pi_{i\ell}' x_{i,t-\ell} - \sum_{\ell=0}^{p} \cdots \\ x_{it} - \sum_{\ell=1}^{p} \Psi_i x_{i,t-\ell} \end{bmatrix}$$

$$= \quad \boldsymbol{\mu}_i + \boldsymbol{\Phi}_i' \boldsymbol{f}_t + \boldsymbol{\epsilon}_{it}$$

where

$$\boldsymbol{\mu}_i = \begin{bmatrix} \alpha_i \\ \boldsymbol{v}_i \end{bmatrix}; \boldsymbol{\Phi}_i' = \begin{bmatrix} \boldsymbol{\lambda}_i' \\ \boldsymbol{\gamma}_i' \end{bmatrix}; \boldsymbol{\epsilon}_{it} = \begin{bmatrix} u_{it} \\ \boldsymbol{v}_{it} \end{bmatrix}$$

- Let $\boldsymbol{z}_t = (\boldsymbol{z}_{1t}', \boldsymbol{z}_{2t}', \ldots, \boldsymbol{z}_{Nt}')'$ with $\boldsymbol{z}_{it} = (y_{it}, x_{it}')'$.
- We can write (101) compactly as

$$\boldsymbol{D}(L) \underset{N(k+1)\times 1}{\boldsymbol{z}_t} = \boldsymbol{\mu} + \underset{N(k+1)\times r}{\boldsymbol{\Phi}} \underset{r\times 1}{\boldsymbol{f}_t} + \boldsymbol{\epsilon}_t \qquad (102)$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_N')'$, $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_N)'$, $\boldsymbol{\epsilon}_t = (\boldsymbol{\epsilon}_{1t}', \ldots, \boldsymbol{\epsilon}_{Nt}')'$, and $\boldsymbol{D}(L) = \boldsymbol{D}_0 - \sum_{\ell=1}^{p} \boldsymbol{D}_\ell L^\ell$ with (i, j) sub-blocks of $\boldsymbol{D}_0$ and $\boldsymbol{D}_\ell$ given by

$$\underset{(k+1)\times(k+1)}{\boldsymbol{D}_{0,ij}} = \left\{ \begin{array}{c} \begin{bmatrix} 1 & -\boldsymbol{\pi}_{i0}' \\ 0 & \boldsymbol{I}_k \end{bmatrix}, \ if \ i = j \\ \begin{bmatrix} \phi_{i0}^* w_{ij} & \boldsymbol{\pi}_{i0}^{*\prime} w_{ij} \\ 0 & 0 \end{bmatrix}, \ if \ i \neq j \end{array} \right\}$$

$$\underset{(k+1)\times(k+1)}{\boldsymbol{D}_{\ell,ij}} = \left\{ \begin{array}{c} \begin{bmatrix} \phi_{i\ell} & \boldsymbol{\pi}_{i\ell}' \\ 0 & \Psi_{i\ell} \end{bmatrix}, \ if \ i = j \\ \begin{bmatrix} \phi_{i\ell}^* w_{ij} & \boldsymbol{\pi}_{i\ell}^{*\prime} w_{ij} \\ 0 & 0 \end{bmatrix}, \ if \ i \neq j \end{array} \right\}, \ell = 1, \ldots, p$$

- **Assumption Q1**: $\boldsymbol{\epsilon}_{it} = (u_{it}, \boldsymbol{v}'_{it})'$ are such that (i) u_{it} is *iid* distributed over t and uncorrelated over i with $E(u_{it}) = 0$ and $E(u_{it}^4) \leq \infty$. (ii) $\boldsymbol{v}_{it}$ is *iid* distributed over t and uncorrelated over i with $E(\boldsymbol{v}_{it}) = 0$ and $E(\|\boldsymbol{v}_{it}\|^4) \leq \infty$. (iii) u_{it} is independent of $\boldsymbol{v}_{js}$ for all (i, j, t, s). Let $\boldsymbol{\Sigma}_{ii} = diag\left(\sigma_i^2, \boldsymbol{\Sigma}_{iiv}\right)$ denote the variance matrix of $\boldsymbol{\epsilon}_{it}$, where σ_i^2 is the variance of e_{it} and Σ_{iiv} the variance matrix of $\boldsymbol{v}_{it}$.
- **Assumption Q2**: There exists a $C > 0$ such that (i) $\|\Phi_i\| \leq C$; (ii) $C^{-1} \leq \tau_{\min}\left(\Sigma_{jj}\right) \leq \tau_{\max}\left(\Sigma_{jj}\right) \leq C$, where $\tau_{\min}\left(\Sigma_{jj}\right)$ and $\tau_{\max}\left(\Sigma_{jj}\right)$ denote the smallest and largest eigenvalues of Σ_{jj}; (iii) There exists an $r \times r$ positive matrix $\boldsymbol{Q}$ such that $\boldsymbol{Q} = \lim_{N\to\infty} N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\Phi}$, and $\boldsymbol{\Sigma}_{\epsilon\epsilon} = diag\left(\Sigma_{11}, ..., \Sigma_{NN}\right)$.

- **Assumption Q3**: The variances Σ_{ii} for all i and $\boldsymbol{M}_{ff}$ are estimated in a compact set, *i.e.* all the eigenvalues of $\hat{\Sigma}_{ii}$ and $\hat{\boldsymbol{M}}_{ff}$ are in an interval $[C^{-1}, C]$.
- **Assumption Q4: Identification conditions**. We impose the normalization restrictions: (i) $\bar{\boldsymbol{f}} = T^{-1}\sum_{t=1}^{T}\boldsymbol{f}_t = 0$; (ii) $\boldsymbol{M}_{ff} = T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{f}_t - \bar{\boldsymbol{f}}\right)\left(\boldsymbol{f}_t - \bar{\boldsymbol{f}}\right)' = \boldsymbol{I}_r$; and (iii) $N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{\Phi}$ is diagonal with the diagonal elements being distinct and arranged in descending order.

- The objective function for the model (102) is:

$$\mathcal{L}(\boldsymbol{\xi}) = -\frac{1}{2N}\ln|\boldsymbol{\Sigma}_{zz}| + \frac{1}{N}\ln|\boldsymbol{I}_N - \boldsymbol{\Phi}_0^*\boldsymbol{W}| \qquad (103)$$

$$-\frac{1}{2NT}\sum_{t=1}^{T}\left(\boldsymbol{D}_0\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)'\boldsymbol{\Sigma}_{zz}^{-1}\left(\boldsymbol{D}_0\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)$$

where $\boldsymbol{\xi} = (\boldsymbol{\phi}_0^*, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}_{\epsilon\epsilon})$ with $\boldsymbol{\phi}_0^* = (\phi_{10}^*, ..., \phi_{N0}^*)'$, $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_N'\right)'$ and $\Sigma_{zz} = \Phi\Phi' + \Sigma_{\varepsilon\varepsilon}$.

The QML-EM algorithms

- **Step 1**: Update $\mathbf{\Phi}^{(s)}, \mathbf{\Sigma}_{\epsilon\epsilon}^{(s)}, \boldsymbol{\theta}^{(s)}$ according to EM algorithm:

$$\Phi^{(s+1)} = \left[\frac{1}{T}\sum_{t=1}^{T} E\left(\left(\boldsymbol{D}_0^{(s)}\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell^{(s)}\boldsymbol{z}_{t-\ell}\right)\boldsymbol{f}_t'|\theta_s\right)\right] \times \left[\frac{1}{T}\sum_{t=1}^{T} E\left(\boldsymbol{f}_t\boldsymbol{f}_t'|\theta_s\right)\right]^{-1}$$

$$\Sigma_{\epsilon\epsilon}^{(s+1)} = Dg\left[\begin{array}{c} \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{D}_0^{(s)}\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell^{(s)}\boldsymbol{z}_{t-\ell}\right)\left(\boldsymbol{D}_0^{(s} \right. \\ -\mathbf{\Phi}^{(s+1)}\mathbf{\Phi}^{(s)\prime}\left(\mathbf{\Sigma}_{zz}^{(s)}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{D}_0^{(s)}\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell^{(s)} \right. \end{array}\right. \quad (104)$$

- and

$$\begin{aligned}\boldsymbol{\theta}_i^{(s+1)} &= \left[\sum_{t=1}^{T}\frac{1}{\left(\sigma_i^{(s+1)}\right)^2}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right]^{-1} \\ &\times\left[\sum_{t=1}^{T}\frac{1}{\left(\sigma_i^{(s+1)}\right)^2}\boldsymbol{\chi}_{it}\left(y_{it}-\phi_{i0}^{*(s)}y_{it}^{*}-\boldsymbol{\lambda}_i^{(s+1)\prime}\boldsymbol{f}_t^{(s)}\right)\right] \\ &= \left[\sum_{t=1}^{T}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right]^{-1}\left[\sum_{t=1}^{T}\boldsymbol{\chi}_{it}\left(y_{it}-\phi_{i0}^{*(s)}y_{it}^{*}-\boldsymbol{\lambda}_i^{(s+1)\prime}\boldsymbol{f}_t^{(s)}\right)\right]\end{aligned}$$

where Dg is the operator which sets the entries of its argument to zeros if the counterparts of $E\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right)$ are zeros;

- $\left(\sigma_i^{(s+1)}\right)^2$ is the $[(i-1)(k+1)+1]$th diagonal element of $\boldsymbol{\Sigma}_{\epsilon\epsilon}^{(s+1)}$. $\boldsymbol{\lambda}_i^{(s+1)\prime}$ is the transpose of the $[(i-1)(k+1)+1]$th row of $\boldsymbol{\Phi}^{(s+1)}$.

- In addition,

$$\frac{1}{T}\sum_{t=1}^{T} E\left(\left(\boldsymbol{D}_0\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)\boldsymbol{f}'_t|\theta_s\right)$$

$$= \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{D}_0\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)\left(\boldsymbol{D}_0\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)'\left(\boldsymbol{\Sigma}_{zz}^{(s)}\right)$$

$$\frac{1}{T}\sum_{t=1}^{T} E\left(\boldsymbol{f}_t\boldsymbol{f}'_t|\theta_s\right) = \boldsymbol{I}_r - \boldsymbol{\Phi}^{(s)\prime}\left(\boldsymbol{\Sigma}_{zz}^{(s)}\right)^{-1}\boldsymbol{\Phi}^{(s)}$$

$$+\boldsymbol{\Phi}^{(s)\prime}\left(\boldsymbol{\Sigma}_{zz}^{(s)}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{D}_0\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)\left(\boldsymbol{D}_0\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-}\right.$$

$$\boldsymbol{f}_t^{(s)} = \boldsymbol{\Phi}^{(s)\prime}\left(\boldsymbol{\Sigma}_{zz}^{(s)}\right)^{-1}\left(\boldsymbol{D}_0\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right).$$

- **Step 2**: We update ϕ_0^* by maximising (103) with respect to ϕ_0^* at $\boldsymbol{\theta}_i = \hat{\boldsymbol{\theta}}_i^{(s+1)}$, $\boldsymbol{\Phi} = \hat{\boldsymbol{\Phi}}^{(s+1)}$ and $\boldsymbol{\Sigma}_{\epsilon\epsilon} = \hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon}^{(s+1)}$ with an initial value of ϕ_0^* at $\hat{\phi}_0^{*(s)}$.
- Alternatively, we may combine the STARDL-CF estimator with the EM algorithm.
- Given $\hat{C}_{it}$, we update all other parameters including ϕ_{i0}^* by running the following augmented regression:

$$\left(y_{it} - \hat{C}_{it}\right) = \sum_{\ell=1}^{p} \phi_{i\ell} y_{i,t-\ell} + \sum_{\ell=0}^{p} \phi_{i\ell}^* y_{i,t-\ell}^* + \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}' \boldsymbol{x}_{i,t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}'^* \boldsymbol{x}_{i,t-\ell}^* + \alpha_i + \rho \hat{v}_{it} + e_{it}^*,$$

where $\hat{v}_{it} = y_{it}^* - \hat{\boldsymbol{\varphi}}_i' \boldsymbol{z}_{it}$ and
$e_{it}^* = e_{it} + \rho \left(\hat{\boldsymbol{\varphi}}_i - \boldsymbol{\varphi}_i\right)' \boldsymbol{z}_{it} + \left(C_{it} - \hat{C}_{it}\right)$.

Special Case: STAR models with factors

- Consider the STAR model with the heterogeneous parameters:

$$y_{it} = \phi_i y_{it-1} + \phi_{i0}^* y_{it}^* + \phi_{i1}^* y_{it-1}^* + u_{it} \tag{105}$$

- Stacking the individual STAR(1) regressions, we have:

$$\boldsymbol{y}_t = \boldsymbol{\Phi} \boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_0^* \boldsymbol{W} \boldsymbol{y}_t + \boldsymbol{\Phi}_1^* \boldsymbol{W} \boldsymbol{y}_{t-1} + \boldsymbol{u}_t \tag{106}$$

$$\underset{N\times N}{\boldsymbol{\Phi}} = \begin{bmatrix} \phi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_N \end{bmatrix}, \underset{N\times N}{\boldsymbol{\Phi}_h^*} = \begin{bmatrix} \phi_{1h}^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_{Nh}^* \end{bmatrix} h = 0, 1$$

- It is straightforward to develop the general STAR(p) model with the heterogeneous parameters:

$$y_{it} = \sum_{h=1}^{p} \phi_{ih} y_{i,t-h} + \sum_{h=0}^{p} \phi_{ih}^{*} y_{i,t-h}^{*} + u_{it} \tag{107}$$

- Stacking the individual STAR(p) regressions:

$$\boldsymbol{y}_t = \sum_{h=1}^{p} \boldsymbol{\Phi}_h \boldsymbol{y}_{t-h} + \sum_{h=0}^{p} \boldsymbol{\Phi}_h^{*} \boldsymbol{W} \boldsymbol{y}_{t-h} + \boldsymbol{u}_t \tag{108}$$

$$\underset{N\times N}{\boldsymbol{\Phi}_h} = \begin{bmatrix} \phi_{1h} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_{Nh} \end{bmatrix}, \underset{N\times N}{\boldsymbol{\Phi}_h^{*}} = \begin{bmatrix} \phi_{1h}^{*} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_{Nh}^{*} \end{bmatrix}$$

- **Remark**: **Spatial stability**: The eigenvalues of $\mathbf{\Phi}_0^* \boldsymbol{W}$ lie inside the unit circle.
- **Remark**: **Time stability**: We rewrite the STAR(p) regression as

$$\boldsymbol{y}_t = \sum_{h=1}^{p} \tilde{\mathbf{\Phi}}_h \boldsymbol{y}_{t-h} + \tilde{\boldsymbol{u}}_t, \tag{109}$$

where $\tilde{\mathbf{\Phi}}_h = \left(\boldsymbol{I}_N - \mathbf{\Phi}_0^* \boldsymbol{W}\right)^{-1} \left(\mathbf{\Phi}_h + \mathbf{\Phi}_h^* \boldsymbol{W}\right)$, and $\tilde{\boldsymbol{u}}_t = \left(\boldsymbol{I}_N - \mathbf{\Phi}_0^* \boldsymbol{W}\right)^{-1} \boldsymbol{u}_t$.
The roots of the $N \times N$ matrix polynomial $\tilde{\mathbf{\Phi}}(z) = \boldsymbol{I}_N - \sum_{h=1}^{p} \tilde{\mathbf{\Phi}}_h z$ lie outside the unit circle.

- To deal with endogeneity of y_{it}^* we apply the CF approach by considering the following CF DGP:

$$y_{it}^* = \boldsymbol{\varphi}_i' \boldsymbol{z}_{it} + v_{it} \quad \text{with } E\left(\boldsymbol{z}_{it}' v_{it}\right) = \mathbf{0} \tag{110}$$

where $\boldsymbol{z}_{it}$ are the $L \times 1$ vector of exogenous variables.

- The two-step procedure: (i) obtain the reduced form residuals, $\hat{v}_{it} = y_{it}^* - \hat{\boldsymbol{\varphi}}_i' \boldsymbol{z}_{it}$ and (ii) run the following regression:

$$y_{it} = \sum_{h=1}^{p} \phi_{ih} y_{i,t-h} + \sum_{h=0}^{p} \phi_{ih}^* y_{i,t-h}^* + \rho \hat{v}_{it} + e_{it}^* \tag{111}$$

where $e_{it}^* = e_{it}^* + \rho \left(\hat{\boldsymbol{\varphi}}_i - \boldsymbol{\varphi}_i\right)' \boldsymbol{z}_{it}$ depends on the sampling error in $\hat{\boldsymbol{\varphi}}_i$ unless $\rho = 0$ (exogeneity test).

- The OLS estimator will be consistent.

- Rewrite the STAR(p) model as

$$\phi_i(L)\, y_{it} = \phi_i^*(L)\, y_{it}^* + u_{it} \tag{112}$$

$$\phi_i(L) = 1 - \sum_{h=1}^{p} \phi_{ih} L^h; \;\; \phi_i^*(L) = 1 - \sum_{h=0}^{p} \phi_{ih}^* L^h.$$

- Premultiplying by the inverse of $\phi_i(L)$:

$$y_{it} = \tilde{\phi}_i^*(L)\, y_{it}^* + \tilde{u}_{it} \tag{113}$$

where $\tilde{\phi}_i^*(L) \left(= \sum_{j=0}^{\infty} \tilde{\phi}_{ij}^* L^j \right) = [\phi_i(L)]^{-1} \phi_i^*(L)$, and $\tilde{u}_{it} = [\phi_i(L)]^{-1} u_{it}$.

- Dynamic multipliers can be evaluated as

$$\tilde{\phi}_{ij}^* = \phi_{i1}\tilde{\phi}_{i,j-1}^* + \phi_{i2}\tilde{\phi}_{i,j-2}^* + \cdots + \phi_{i,j-1}\tilde{\phi}_{i1}^* + \phi_{ij}\tilde{\phi}_{i0}^* + \phi_{ij}^*, \; j = 1, 2, \ldots \tag{114}$$

where $\phi_{ij} = 0$ for $j < 1$ and $\tilde{\phi}_{i0}^* = \phi_{i0}^*$, $\tilde{\phi}_{ij}^* = 0$ for $j < 0$.

- The cumulative dynamic multiplier effects of y_{it}^* on $y_{i,t+h}$ can be evaluated as

$$m_{y_i}\left(y_i^*, H\right) = \sum_{h=0}^{H} \tilde{\phi}_{ih}^*, \quad H = 0, 1, \ldots$$

- As $H \to \infty$, $m_{y_i}\left(y_i^*, H\right) \to \beta_{yi}^*$ (the long-run coefficient).
- Define the dynamic multiplier effects as

$$\frac{\partial \boldsymbol{y}_{t+h}}{\partial \boldsymbol{y}_t^{*\prime}} = \begin{bmatrix} \frac{\partial y_{1,t+h}}{\partial y_{1t}^*} & 0 & \cdots & 0 \\ 0 & \frac{\partial y_{2,t+h}}{\partial y_{2t}^*} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial y_{N,t+h}}{\partial y_{Nt}^*} \end{bmatrix}_{N \times N}$$

- What's $\frac{\partial y_{i,t+h}}{\partial y_{jt}}$? Say,

$$\frac{\partial y_{i,t+h}}{\partial y_{jt}} = \frac{\partial y_{i,t+h}}{\partial y_{it}^{*}} \times w_{ij} \text{ for } i \neq j$$

Then, what about $\frac{\partial y_{i,t+h}}{\partial y_{it}}$? Simply set to zero?

Diffusion IRF and FEVD

- We rewrite the STAR(p) model as

$$\tilde{\boldsymbol{\Phi}}(L)\,\boldsymbol{y}_t = \tilde{\boldsymbol{u}}_t, \quad (115)$$

where $\tilde{\boldsymbol{u}}_t = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1} \boldsymbol{u}_t$,

$$\tilde{\boldsymbol{\Phi}}(L) = \boldsymbol{I}_N - \sum_{j=1}^{p} \tilde{\boldsymbol{\Phi}}_j L^j \text{ with } \tilde{\boldsymbol{\Phi}}_j = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1} (\boldsymbol{\Phi}_j + \boldsymbol{\Phi}_j^* \boldsymbol{W})$$

- Premultiplying by the inverse of $\tilde{\boldsymbol{\Phi}}(L)$:

$$\boldsymbol{y}_t = \left[\tilde{\boldsymbol{\Phi}}(L)\right]^{-1} \tilde{\boldsymbol{u}}_t = \left[\tilde{\boldsymbol{\Phi}}(L)\right]^{-1} (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1} \boldsymbol{u}_t \quad (116)$$

from which we can construct (diffusion) IRF and FEVD.

- $\{\mathbf{Q}\}$ with respect to $\tilde{\boldsymbol{u}}_t$ or $\boldsymbol{u}_t$, probably $\boldsymbol{u}_t$ (can we assume $\boldsymbol{u}_t$ as structural?) More to follow:

With Factors

- In the spatial modelling it is implicitly assumed that u_{it} is iid across spatial units or spatially correlated:

$$u_{it} = \lambda \mathbf{W} u_t + \varepsilon_{it}.$$

- Now we introduce the common factor structure such that

$$u_t = \mathbf{\Lambda} \boldsymbol{f}_t + \boldsymbol{v}_t.$$

- Then, the STAR model can be extended as

$$y_{it} = \phi_i y_{it-1} + \phi_{i0}^* y_{it}^* + \phi_{i1}^* y_{it-1}^* + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + v_{it} \qquad (117)$$

where v_{it} contains the idiosyncratic components which are mutually uncorrelated across (i, j).

- Generally, we have $STAR(p)$ with (observed) factors:

$$y_{it} = \sum_{h=1}^{p} \phi_{ih} y_{i,t-h} + \sum_{h=0}^{p} \phi^*_{ih} y^*_{i,t-h} + \boldsymbol{\lambda}'_i \boldsymbol{f}_t + v_{it} \tag{118}$$

- To deal with the endogeneity of y^*_{it} we apply the CF approach.
- Then, (118) can be expressed as

$$\tilde{\boldsymbol{\Phi}}(L)\, \boldsymbol{y}_t = \left(\boldsymbol{I}_N - \boldsymbol{\Phi}^*_0 \boldsymbol{W}\right)^{-1} \left(\boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{v}_t\right) \tag{119}$$

  from which we can construct IRF and FEVD.
- **Remark**: This is a parsimonious specification, implying that we can include the large N spatial units, so another way of circumventing the curse of input dimensionality.

GVAR-SPVAR Model

- Consider a global economy consisting of N economies and denote the country-specific variables by an $m_i \times 1$ vector $\boldsymbol{y}_{it}$, and the country-specific foreign variables by an $m_i^* \times 1$ vector $\boldsymbol{y}_{it}^* = \sum_{j=1}^{N} w_{ij} \boldsymbol{y}_{jt}$ where $w_{ij} \geq 0$ is the set of granular weights with $\sum_{j=1}^{N} w_{ij} = 1$, and $w_{ii} = 0$ for all i.
- The country-specific VARX* $(2,2)$ model can be written as

$$\begin{aligned} \boldsymbol{y}_{it} &= \boldsymbol{h}_{i0} + \boldsymbol{h}_{i1} t + \boldsymbol{\Phi}_{i1} \boldsymbol{y}_{i,t-1} + \boldsymbol{\Phi}_{i2} \boldsymbol{y}_{i,t-2} \\ &\quad + \boldsymbol{\Psi}_{i0} \boldsymbol{y}_{it}^* + \boldsymbol{\Psi}_{i1} \boldsymbol{y}_{i,t-1}^* + \boldsymbol{\Psi}_{i2} \boldsymbol{y}_{i,t-2}^* + \boldsymbol{u}_{it} \end{aligned} \tag{120}$$

  where the dimension of $\boldsymbol{h}_{ij}$ and $\boldsymbol{\delta}_{ij}$ is $m_i \times 1$ while those of $\boldsymbol{\Phi}_{ij}$ and $\boldsymbol{\Psi}_{ij}$ are $m_i \times m_i$ and $m_i \times m_i^*$.
- $\boldsymbol{u}_{it} \sim iid\,(0, \boldsymbol{\Sigma}_{ii})$ where $\boldsymbol{\Sigma}_{ii}$ is an $m_i \times m_i$ PD matrix.

The spatial representation of GVAR

- Define the $m \times 1$ vector of the global variables:

$$\underset{m\times 1}{\boldsymbol{y}_t} = \left(\boldsymbol{y}'_{1t}, ..., \boldsymbol{y}'_{Nt}\right)' \text{ with } \underset{k\times 1}{\boldsymbol{y}_{it}} = \left(y_{1,it}, ..., y_{k,it}\right)'$$

- Define the $N \times N$ weight matrix:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & & w_{1N} \\ w_{21} & w_{22} & & w_{2N} \\ & & \ddots & \\ w_{N1} & w_{N2} & & w_{NN} \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_N \end{bmatrix}$$

- Then, we have:

$$\underset{k\times 1}{\boldsymbol{y}^*_{it}} = \left(y^*_{1,it}, ..., y^*_{k,it}\right)' = \left(\mathbf{w}_i \otimes \mathbf{I}_k\right) \boldsymbol{y}_t, \ \underset{m\times 1}{\boldsymbol{y}^*_t} = \left(\mathbf{W} \otimes \mathbf{I}_k\right) \boldsymbol{y}_t$$

- Thus, (120) can be written as

$$\begin{aligned} \boldsymbol{y}_{it} &= \boldsymbol{\Phi}_{i1}\boldsymbol{y}_{i,t-1} + \boldsymbol{\Phi}_{i2}\boldsymbol{y}_{i,t-2} + \boldsymbol{\Psi}_{i0}\left(\mathbf{w}_i \otimes \mathbf{I}_{m_i}\right)\boldsymbol{y}_{it} \\ &\quad + \boldsymbol{\Psi}_{i1}\left(\mathbf{w}_i \otimes \mathbf{I}_{m_i}\right)\boldsymbol{y}_{it-1} + \boldsymbol{\Psi}_{i2}\left(\mathbf{w}_i \otimes \mathbf{I}_{m_i}\right)\boldsymbol{y}_{it-2} + \boldsymbol{u}_{it} \end{aligned} \tag{121}$$

- Stacking these results, we have:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\Phi}_1\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_2\boldsymbol{y}_{t-2} + \boldsymbol{\Psi}_0\boldsymbol{y}_t^* + \boldsymbol{\Psi}_1\boldsymbol{y}_{t-1}^* + \boldsymbol{\Psi}_2\boldsymbol{y}_{t-2}^* + \boldsymbol{u}_t \\ &= \boldsymbol{\Phi}_1\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_2\boldsymbol{y}_{t-2} + \boldsymbol{\Psi}_0\left(\mathbf{W} \otimes \mathbf{I}_k\right)\boldsymbol{y}_t + \boldsymbol{\Psi}_1\left(\mathbf{W} \otimes \mathbf{I}_k\right)\boldsymbol{y}_{t-1} \\ &\quad + \boldsymbol{\Psi}_2\left(\mathbf{W} \otimes \mathbf{I}_k\right)\boldsymbol{y}_{t-2} + \boldsymbol{u}_t \end{aligned} \tag{122}$$

- Alternatively, (122) can be written as

$$\begin{aligned} & \left(\boldsymbol{I}_m - \boldsymbol{\Psi}_0 \left(\mathbf{W} \otimes \mathbf{I}_k\right)\right) \boldsymbol{y}_t \\ = \quad & \left(\boldsymbol{\Phi}_1 + \boldsymbol{\Psi}_1 \left(\mathbf{W} \otimes \mathbf{I}_k\right)\right) \boldsymbol{y}_{t-1} + \left(\boldsymbol{\Phi}_2 + \boldsymbol{\Psi}_2 \left(\mathbf{W} \otimes \mathbf{I}_k\right)\right) \boldsymbol{y}_{t-2} + \boldsymbol{u}_t \end{aligned} \tag{123}$$

or

$$\left\{ \left(\boldsymbol{I}_m - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2 L^2\right) - \left(\boldsymbol{\Psi}_0 + \boldsymbol{\Psi}_1 L + \boldsymbol{\Psi}_2 L^2\right) \left(\mathbf{W} \otimes \mathbf{I}_k\right) \right\} \boldsymbol{y}_t = \boldsymbol{u}_t \tag{124}$$

- **Remark**: This shows that SPVAR is the special case of GVAR. We are interested in IRFs in terms of $\frac{\partial \boldsymbol{y}_{t+h}}{\partial \boldsymbol{u}_t}$, which are the combination of spatial and dynamic ones. So it would be an important issue how to decompose overall IRFs into the spatial and dynamic components.
- **Remark**: We may be interested in evaluating the dynamic multipliers in terms of $\frac{\partial \boldsymbol{y}_{1,t+h}}{\partial \boldsymbol{y}_{2t}}$ or *vice versa* where $\boldsymbol{y}_t = (\boldsymbol{y}_{1t}' \boldsymbol{y}_{2t}')'$, but this is not quite straightforward.
- When we add the global factors, $\boldsymbol{f}_t$, such as the oil or commodity prices, it is straightforward to derive the dynamic multipliers in terms of $\frac{\partial \boldsymbol{y}_{t+h}}{\partial \boldsymbol{f}_t}$.

Multi-dimensional Panel Data Modelling with CSD

- Given the growing availability of the big dataset, the recent literature attempted to extend the error components models to the multidimensional setting.
- Balazsi, Matyas and Wansbeek (2015) develop the 3D within estimator; Balazsi, Baltagi, Matyas and Pus (2016) the 3D random effects approach.
- This multi-dimensional approach becomes an essential tool for the analysis of complex interconnectedness of the big dataset;
  - the bilateral flows such as trade, FDI, capital or migration flows (e.g. Feenstra, 2004; Bertoli and Fernandez-Huertas Moraga, 2013; Gunnella et al., 2015);
  - matched dataset that link the employer-employee and pupils-teachers (Abowd et al., 1999; Kramarz et al., 2008).
- No study to address an issue of controlling CSD in 3D or higher-dimensional data, despite strong CSD evidence in 2D

- 2 main approaches in modelling CSD in 2D panels;
  - the factor-based approach (Pesaran, 2006; Bai, 2009)
  - the spatial econometrics techniques (Baltagi, 2005; Behrens et al., 2012).
  - the factor-based models exhibit strong CSD while the spatial models weak CSD only (Chudik et al., 2011).
- See also Bailey et al. (2016), Le Gallo and Pirotte (2017), Baltagi, Egger and Erhardt (2017).

- Following this research trend, we generalise the multi-dimensional error components by incorporating unobserved heterogeneous global factors.
- The country-time fixed (CTFE) and random effects (CTRE) estimators fail to remove heterogenous global factors; inconsistent in the presence of nonzero correlation between the regressors and unobserved factors.
- We develop the 2-step estimation procedure.
  1. We augment the 3D model with cross-section averages of dependent variable and regressors, as proxies for unobserved global factors.
  2. We apply the 3D-within transformation to the augmented specification and obtain consistent estimators (the 3D-PCCE estimator).
- Our approach is the first attempt to accommodate strong CSD in multi-dimensional panels.

- We discuss the extent of CSD under 3 different error components with CTFE, with 2-way heterogeneous factor, and with both.
- We develop a diagnostic test for the null of (pairwise) residual cross-section independence or weak dependence; a modified CD test in 2D panels by Pesaran (2015)
- We provide extensions into unbalanced panels and 4D models.
- Monte Carlo studies confirm that the 3D-PCCE estimators perform well.
- On the contrary CTFE displays severe biases and size distortions.

- We apply the 3D PCCE estimation to the dataset over 1960-2008 for 91 country-pairs amongst 14 EU countries.
- Based on the CD test results, and the predicted signs and statistical significance of the coefficients, we find that the 3D PCCE estimation results are most reliable.
- The trade effect of currency union is rather modest.
- This suggests that the trade increase within the Euro area may reflect a continuation of a long-run historical trend linked to the broader set of EU's economic integration policies.

Notations

- $\mathbf{I}_N$ is an $N \times N$ identity matrix, $\mathbf{J}_N$ the $N \times N$ identity matrix of ones, and $\boldsymbol{\iota}_N$ the $N \times 1$ vector of ones.
- $\mathbf{M}_A$ projects the $N \times N$ matrix $\mathbf{A}$ into its null-space, i.e., $\mathbf{M}_A = \mathbf{I}_N - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$.
- $y_{.jt} = N_1^{-1} \sum_{i=1}^{N_1} y_{ijt}$, $y_{i.t} = N_2^{-1} \sum_{j=1}^{N_2} y_{ijt}$ and $y_{ij.} = T^{-1} \sum_{t=1}^{T} y_{ijt}$ denote the average of y over the index i, j and t, respectively, with the definition extending to other quantities such as $y_{..t}$, $y_{.j.}$, $y_{i..}$ and $y_{...}$.

- Consider the 3D country-time fixed effects panel data model:

$$y_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \boldsymbol{\gamma}'\mathbf{s}_{it} + \boldsymbol{\delta}'\mathbf{d}_{jt} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\mathbf{z}_{ij} + u_{ijt}, \tag{125}$$

for $i = 1, ..., N_1, j = 1, ..., N_2, t = 1, ..., T$, with errors:

$$u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \varepsilon_{ijt} \tag{126}$$

- y_{ijt} is the dependent variable across 3 indices (e.g. the import of country j from country i at period t);
- $\mathbf{x}_{ijt}$, $\mathbf{s}_{it}$, $\mathbf{d}_{jt}$, $\mathbf{q}_t$, $\mathbf{z}_{ij}$ are the $k_x \times 1$, $k_s \times 1$, $k_d \times 1$, $k_q \times 1$, $k_z \times 1$ vectors of covariates covering all measurements across 3 indices;
- The multi-error components contain pair-fixed effects (μ_{ij}) as well as origin and destination CTFEs, v_{it} and ζ_{jt}.

- To remove all unobserved FEs, BMW derive the 3D within transformation:

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.} - y_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t} + \bar{y}_{.j.} + \bar{y}_{i..} - \bar{y}_{...} \quad (127)$$

- Estimate consistently $\boldsymbol{\beta}$ from the transformed regression:

$$\tilde{y}_{ijt} = \boldsymbol{\beta}' \tilde{\mathbf{x}}_{ijt} + \tilde{\varepsilon}_{ijt}, \quad (128)$$

where $\tilde{\mathbf{x}}_{ijt} = \mathbf{x}_{ijt} - \bar{\mathbf{x}}_{ij.} - \bar{\mathbf{x}}_{.jt} - \bar{\mathbf{x}}_{i.t} + \bar{\mathbf{x}}_{..t} + \bar{\mathbf{x}}_{.j.} + \bar{\mathbf{x}}_{i..} - \bar{\mathbf{x}}_{...}$.

- We write (128) compactly as

$$\tilde{\mathbf{Y}}_{ij} = \tilde{\mathbf{X}}_{ij}\boldsymbol{\beta} + \tilde{\mathbf{E}}_{ij} \quad (129)$$

$$\underset{T\times 1}{\tilde{\mathbf{Y}}_{ij}} = \begin{bmatrix} \tilde{y}_{ij1} \\ \vdots \\ \tilde{y}_{ijT} \end{bmatrix}, \quad \underset{T\times k_x}{\tilde{\mathbf{X}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ij1} \\ \vdots \\ \tilde{\mathbf{x}}'_{ijT} \end{bmatrix}, \quad \underset{T\times 1}{\tilde{\mathbf{E}}_{ij}} = \begin{bmatrix} \tilde{\varepsilon}_{ij1} \\ \vdots \\ \tilde{\varepsilon}_{ijT} \end{bmatrix}.$$

- The 3D-within estimator of β is obtained by

$$\hat{\boldsymbol{\beta}}_W = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}_{ij}' \tilde{\mathbf{X}}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}_{ij}' \tilde{\mathbf{Y}}_{ij} \right). \tag{130}$$

- As $(N_1, N_2, T) \to \infty$,

$$\sqrt{N_1 N_2 T} \left(\hat{\boldsymbol{\beta}}_W - \boldsymbol{\beta} \right) \tag{131}$$
$$\overset{a}{\sim} N \left(\mathbf{0}, \sigma_\varepsilon^2 \lim_{(N_1, N_2, T) \to \infty} \left(\frac{1}{N_1 N_2 T} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}_{ij}' \tilde{\mathbf{X}}_{ij} \right)^{-1} \right).$$

- The 3D within transformation wipes out all other covariates, $\mathbf{x}_{it}$, $\mathbf{x}_{jt}$, $\mathbf{x}_{t}$, and $\mathbf{x}_{ij}$.
- It would be worthwhile to develop an extension of the Hausman-Taylor (1981) estimation, popular in the two-way panels in the presence of CSD (e.g. Serlenga and Shin, 2007).
- Balazsi, Bun, Chan and Harris (2017) develop an extended HT estimator for multi-dimensional panels.

- We consider a couple of 3D error components that can accommodate CSD.
- υ_{it} and ζ_{jt} are supposed to measure the (local) origin and destination CTFEs, so natural to add the global factor λ_t:

$$u_{ijt} = \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \lambda_t + \varepsilon_{ijt} \tag{132}$$

- The 3D-within transformation, (127) also removes λ_t, because λ_t is proportional to $\sum_{i=1}^{N_1} \upsilon_{it}$ or $\sum_{j=1}^{N_2} \zeta_{jt}$.

- First, consider the following error components specification:

$$u_{ijt} = \mu_{ij} + \pi_{ij}\lambda_t + \varepsilon_{ijt}. \tag{133}$$

- Similar to the 2-way heterogeneous factor model by Serlenga and Shin (2007).
- We apply the cross-section averages of (125) over i and j:

$$\bar{y}_{..t} = \boldsymbol{\beta}'\bar{\mathbf{x}}_{..t} + \boldsymbol{\gamma}'\bar{\mathbf{s}}_{.t} + \boldsymbol{\delta}'\bar{\mathbf{d}}_{.t} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\bar{\mathbf{z}}_{..} + \bar{\mu}_{..} + \bar{\pi}_{..}\lambda_t + \bar{\varepsilon}_{..t} \tag{134}$$

where $\bar{y}_{..t} = (N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} y_{ijt}$, $\bar{\mathbf{s}}_{.t} = N_1^{-1}\sum_{i=1}^{N_1}\mathbf{s}_{it}$,
$\bar{\mathbf{d}}_{.t} = N_2^{-1}\sum_{j=1}^{N_2}\mathbf{d}_{jt}$, $\bar{\mathbf{z}}_{..} = (N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathbf{z}_{ij}$,
$\bar{\mu}_{..} = (N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mu_{ij}$,
$\bar{\pi}_{..} = (N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\pi_{ij}$.

- Hence, we have:

$$\lambda_t = \frac{1}{\bar{\pi}_{..}}\left\{\bar{y}_{..t} - \left(\boldsymbol{\beta}'\bar{\mathbf{x}}_{..t} + \boldsymbol{\gamma}'\bar{\mathbf{s}}_{.t} + \boldsymbol{\delta}'\bar{\mathbf{d}}_{.t} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\bar{\mathbf{z}}_{..} + \bar{\mu}_{..} + \bar{\varepsilon}_{..t}\right)\right\}$$

- We augment the model (125) with the cross-section averages:

$$y_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \boldsymbol{\gamma}'\mathbf{s}_{it} + \boldsymbol{\delta}'\mathbf{d}_{jt} + \boldsymbol{\psi}'_{ij}\mathbf{f}_t + \tau_{ij} + \mu^*_{ij} + \varepsilon^*_{ijt}, \quad (135)$$

where

$$\boldsymbol{\psi}'_{ij} = \left(\frac{\pi_{ij}}{\bar{\pi}_{..}}, \frac{-\pi_{ij}\boldsymbol{\beta}'}{\bar{\pi}_{..}}, \frac{-\pi_{ij}\boldsymbol{\gamma}'}{\bar{\pi}_{..}}, \frac{-\pi_{ij}\boldsymbol{\delta}'}{\bar{\pi}_{..}}, \left(1 - \frac{\pi_{ij}}{\bar{\pi}_{..}}\right)\boldsymbol{\kappa}' \right)$$

$$\mathbf{f}_t = \left(\bar{y}_{..t}, \bar{\mathbf{x}}'_{..t}, \bar{\mathbf{s}}'_{.t}, \bar{\mathbf{d}}'_{.t}, \mathbf{q}'_t\right)' \quad (136)$$

$$\tau_{ij} = \boldsymbol{\varphi}'\mathbf{z}_{ij} - \frac{-\pi_{ij}}{\bar{\pi}_{..}}\boldsymbol{\varphi}'\mathbf{z}_{..}, \ \mu^*_{ij} = \mu_{ij} - \frac{\pi_{ij}\mu_{..}}{\bar{\pi}_{..}}, \ \varepsilon^*_{ijt} = \varepsilon_{ijt} - \frac{\pi_{ij}}{\bar{\pi}_{..}}\bar{\varepsilon}_{..t}.$$

- We write (267) compactly as

$$\mathbf{Y}_{ij} = \mathbf{W}_{ij}\boldsymbol{\theta} + \mathbf{H}\psi^*_{ij} + \mathbf{E}^*_{ij},\ i = 1, ..., N_1, j = 1, ..., N_2 \quad (137)$$

$$\underset{T\times 1}{\mathbf{Y}_{ij}} = \begin{bmatrix} y_{ij1} \\ \vdots \\ y_{ijT} \end{bmatrix},\ \underset{T\times k_x}{\mathbf{X}_{ij}} = \begin{bmatrix} \mathbf{x}'_{ij1} \\ \vdots \\ \mathbf{x}'_{ijT} \end{bmatrix},\ \underset{T\times k_s}{\mathbf{S}_i} = \begin{bmatrix} \mathbf{s}'_{i1} \\ \vdots \\ \mathbf{s}'_{iT} \end{bmatrix},$$

$$\underset{T\times k_d}{\mathbf{D}_j} = \begin{bmatrix} \mathbf{d}'_{j1} \\ \vdots \\ \mathbf{d}'_{jT} \end{bmatrix},\ \underset{T\times k_f}{\mathbf{F}} = \begin{bmatrix} \mathbf{f}'_1 \\ \vdots \\ \mathbf{f}'_T \end{bmatrix},\ \underset{T\times 1}{\mathbf{E}^*_{ij}} = \begin{bmatrix} \varepsilon^*_{ij1} \\ \vdots \\ \varepsilon^*_{ijT} \end{bmatrix},$$

$$\mathbf{W}_{ij} = (\mathbf{X}_{ij}, \mathbf{S}_i, \mathbf{D}_j),\ \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta}' & \boldsymbol{\gamma}' & \boldsymbol{\delta}' \end{pmatrix}',$$

$\boldsymbol{\psi}^*_{ij} = \left(\boldsymbol{\psi}'_{ij}, \left(\tau_{ij} + \mu^*_{ij}\right)\right)'$ and $\mathbf{H} = [\mathbf{F}, \boldsymbol{\iota}_T]$.

- We derive the 3D-PCCE estimator of $\boldsymbol{\theta}$ by

$$\hat{\boldsymbol{\theta}}_{PCCE} = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbf{W}_{ij}' \mathbf{M}_H \mathbf{W}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbf{W}_{ij}' \mathbf{M}_H \mathbf{Y}_{ij} \right) \tag{138}$$

where $\mathbf{M}_H = \mathbf{I}_T - \mathbf{H} \left(\mathbf{H}'\mathbf{H} \right)^{-1} \mathbf{H}'$.

- Following Pesaran (2006), it is straightforward to show that as $(N_1, N_2, T) \to \infty$,

$$\sqrt{N_1 N_2 T}\left(\hat{\boldsymbol{\theta}}_{PCCE} - \boldsymbol{\theta}\right) \overset{a}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}_\theta\right), \tag{139}$$

where the (robust) consistent estimator of $\boldsymbol{\Sigma}_\theta$ is given by

$$\hat{\boldsymbol{\Sigma}}_\theta = \frac{1}{N_1 N_2}\mathbf{S}_\theta^{-1}\mathbf{R}_\theta\mathbf{S}_\theta^{-1}, \tag{140}$$

$$\mathbf{R}_\theta = \frac{1}{N_1 N_2 - 1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{\mathbf{W}_{ij}'\mathbf{M}_H\mathbf{W}_{ij}}{T}\right)\left(\hat{\boldsymbol{\theta}}_{ij} - \hat{\boldsymbol{\theta}}_{MG}\right)\left(\hat{\boldsymbol{\theta}}_{ij} - \hat{\boldsymbol{\theta}}_{MG}\right.$$

$$\mathbf{S}_\theta = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{\mathbf{W}_{ij}'\mathbf{M}_H\mathbf{W}_{ij}}{T}\right), \quad \hat{\boldsymbol{\theta}}_{MG} = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\hat{\boldsymbol{\theta}}_{ij},$$

where $\hat{\boldsymbol{\theta}}_{ij}$ is the (ij) pairwise OLS estimator obtained from the individual regression of $\mathbf{Y}_{ij}$ on $(\mathbf{W}_{ij}, \mathbf{H})$ in (137).

- Next, we consider the 3D model with more general errors:

$$u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \pi_{ij}\lambda_t + \varepsilon_{ijt}. \tag{141}$$

- The 3D-within transformation fails to remove $\pi_{ij}\lambda_t$, because

$$\tilde{u}_{ijt} = \tilde{\pi}_{ij}\tilde{\lambda}_t + \tilde{\varepsilon}_{ijt}$$

  where $\tilde{\lambda}_t = \lambda_t - \bar{\lambda}$ with $\bar{\lambda} = T^{-1}\sum_{t=1}^{T}\lambda_t$ and $\tilde{\pi}_{ij} = \pi_{ij} - \pi_{.j} - \pi_{i.} + \pi_{..}$ with $\pi_{.j} = N_1^{-1}\sum_{i=1}^{N_1}\pi_{ij}$ and $\pi_{i.} = N_2^{-1}\sum_{j=1}^{N_2}\pi_{ij}$.[1]
- In the presence of the nonzero correlation between $\mathbf{x}_{ijt}$ and λ_t, the 3D-within estimator of $\boldsymbol{\beta}$ is biased.

[1] Unless $\tilde{\pi}_{ij} = 0, \tilde{u}_{ijt} \neq \tilde{\varepsilon}_{ijt}$. This holds only if factor loadings, π_{ij} are homogeneous.

- We develop the two-step estimation procedure.
- First, taking the cross-section averages of (125) over i and j,

$$\bar{y}_{..t} = \boldsymbol{\beta}'\bar{\mathbf{x}}_{..t} + \boldsymbol{\gamma}'\bar{\mathbf{s}}_{.t} + \boldsymbol{\delta}'\bar{\mathbf{d}}_{.t} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\mathbf{z}_{..} + \mu_{..} + \bar{v}_{.t} + \bar{\zeta}_{.t} + \bar{\pi}_{..}\lambda_t + \bar{\varepsilon}_{..t} \tag{142}$$

where $\bar{v}_{.t} = N_1^{-1}\sum_{i=1}^{N_1} v_{it}$, $\bar{\zeta}_{.t} = N_2^{-1}\sum_{j=1}^{N_2} \zeta_{jt}$.
- We augment the model (125) with the cross-section averages:

$$y_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \boldsymbol{\gamma}'\mathbf{s}_{it} + \boldsymbol{\delta}'\mathbf{d}_{jt} + \boldsymbol{\psi}'_{ij}\mathbf{f}_t + \tau_{ij} + \mu^*_{ij} + v^*_{ijt} + \zeta^*_{ijt} + \varepsilon^*_{ijt}, \tag{143}$$

where $v^*_{ijt} = v_{it} - \frac{\pi_{ij}\bar{v}_{.t}}{\bar{\pi}_{..}}$, $\zeta^*_{ijt} = \zeta_{jt} - \frac{\pi_{ij}\bar{\zeta}_{.t}}{\bar{\pi}_{..}}$.

- We rewrite (143) as

$$y_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \boldsymbol{\gamma}'\mathbf{s}_{it} + \boldsymbol{\delta}'\mathbf{d}_{jt} + \boldsymbol{\psi}'_{ij}\mathbf{f}_t + \tau_{ij} + \mu^*_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon^{**}_{ijt}, \tag{144}$$

where $\varepsilon^{**}_{ijt} = \varepsilon_{ijt} - \frac{\pi_{ij}}{\bar{\pi}_{..}}\bar{\varepsilon}_{..t} - \frac{\pi_{ij}\bar{\upsilon}_{.t}}{\bar{\pi}_{..}} - \frac{\pi_{ij}\bar{\zeta}_{.t}}{\bar{\pi}_{..}}$.
- As $N_1, N_2 \to \infty$, $\varepsilon^{**}_{ijt} \to_p \varepsilon_{ijt}$.
- We apply the 3D-within transformation (127) to (144):

$$\tilde{y}_{ijt} = \boldsymbol{\beta}'\tilde{\mathbf{x}}_{ijt} + \tilde{\boldsymbol{\psi}}'_{ij}\tilde{\mathbf{f}}_t + \tilde{\varepsilon}^{**}_{ijt}, \tag{145}$$

where $\tilde{\boldsymbol{\psi}}_{ij} = \boldsymbol{\psi}_{ij} - \boldsymbol{\psi}_{.j} - \boldsymbol{\psi}_{j.} + \boldsymbol{\psi}_{..}$, $\tilde{\mathbf{f}}_t = \mathbf{f}_t - \bar{\mathbf{f}}$ with $\bar{\mathbf{f}} = T^{-1}\sum_{t=1}^{T}\mathbf{f}_t$.

- Rewriting (145) compactly as

$$\tilde{\mathbf{Y}}_{ij} = \tilde{\mathbf{X}}_{ij}\boldsymbol{\beta} + \tilde{\mathbf{F}}\tilde{\boldsymbol{\psi}}_{ij} + \tilde{\mathbf{E}}^{**}_{ij}, \; i = 1, ..., N_1, j = 1, ..., N_2 \quad (146)$$

$$\underset{T\times 1}{\tilde{\mathbf{Y}}_{ij}} = \begin{bmatrix} \tilde{y}_{ij1} \\ \vdots \\ \tilde{y}_{ijT} \end{bmatrix}, \underset{T\times k_x}{\tilde{\mathbf{X}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ij1} \\ \vdots \\ \tilde{\mathbf{x}}'_{ijT} \end{bmatrix}, \underset{T\times k_f}{\tilde{\mathbf{F}}} = \begin{bmatrix} \tilde{\mathbf{f}}'_1 \\ \vdots \\ \tilde{\mathbf{f}}'_T \end{bmatrix}, \tilde{\mathbf{E}}^{**}_{ij} = \begin{bmatrix} \tilde{\varepsilon}^{**}_{ij1} \\ \vdots \\ \tilde{\varepsilon}^{**}_{ijT}
\end{bmatrix}$$

- The 3D-PCCE estimator of $\boldsymbol{\beta}$ is obtained by

$$\hat{\boldsymbol{\beta}}_{PCCE} = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij}\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij}\mathbf{M}_{\tilde{F}}\tilde{\mathbf{Y}}_{ij} \right) \quad (147)$$

where $\mathbf{M}_{\tilde{F}} = \mathbf{I}_T - \tilde{\mathbf{F}}\left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}}\right)^{-1}\tilde{\mathbf{F}}'$ is the $T \times T$ idempotent matrix.

- As $(N_1, N_2, T) \to \infty$,

$$\sqrt{N_1 N_2 T}\left(\hat{\boldsymbol{\beta}}_{PCCE} - \boldsymbol{\beta}\right) \overset{a}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{\beta}\right), \qquad (148)$$

where the (robust) consistent estimator of $\boldsymbol{\Sigma}_{\beta}$ is given by

$$\hat{\boldsymbol{\Sigma}}_{\beta} = \frac{1}{N^2}\mathbf{S}_{\beta}^{-1}\mathbf{R}_{\beta}\mathbf{S}_{\beta}^{-1}, \qquad (149)$$

$$\mathbf{R}_{\beta} = \frac{1}{N_1 N_2 - 1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}}{T}\right)\left(\hat{\boldsymbol{\beta}}_{ij} - \hat{\boldsymbol{\beta}}_{MG}\right)\left(\hat{\boldsymbol{\beta}}_{ij} - \hat{\boldsymbol{\beta}}_{MG}\right.$$

$$\mathbf{S}_{\beta} = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}}{T}\right), \ \hat{\boldsymbol{\beta}}_{MG} = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\hat{\boldsymbol{\beta}}_{ij},$$

where $\hat{\boldsymbol{\beta}}_{ij}$ is the (ij) pairwise OLS estimator from the individual regression of $\tilde{\mathbf{Y}}_{ij}$ on $\left(\tilde{\mathbf{X}}_{ij}, \tilde{\mathbf{F}}\right)$ in (146).

- Extend to 3D panels with heterogeneous parameters:

$$y_{ijt} = \boldsymbol{\beta}'_{ij}\mathbf{x}_{ijt} + \boldsymbol{\gamma}'_j\mathbf{s}_{it} + \boldsymbol{\delta}'_i\mathbf{d}_{jt} + \boldsymbol{\kappa}'_{ij}\mathbf{q}_t + \boldsymbol{\varphi}'\mathbf{z}_{ij} + u_{ijt} \quad (150)$$

- We develop the mean group estimators:

$$\hat{\boldsymbol{\beta}}_{W,MG} = \frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\tilde{\mathbf{X}}'_{ij}\tilde{\mathbf{X}}_{ij}\right)^{-1}\left(\tilde{\mathbf{X}}'_{ij}\mathbf{Y}_{ij}\right) \quad (151)$$

$$\hat{\boldsymbol{\theta}}_{MGCCE} = \frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\mathbf{W}'_{ij}\mathbf{M}_H\mathbf{W}_{ij}\right)^{-1}\left(\mathbf{W}'_{ij}\mathbf{M}_H\mathbf{Y}_{ij}\right) \quad (152)$$

$$\hat{\boldsymbol{\beta}}_{MGCCE} = \frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\tilde{\mathbf{X}}'_{ij}\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}\right)^{-1}\left(\tilde{\mathbf{X}}'_{ij}\mathbf{M}_{\tilde{F}}\tilde{\mathbf{Y}}_{ij}\right) \quad (153)$$

- The extent of CSD is captured by non-zero covariance between u_{ijt} and $u_{i'j't}$, which relates to rate at which $\frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sigma_{ijt,u}$ declines with $N_1 N_2$.
- First, consider the 3D model (125) with CTFEs (126).
- We make the random effects assumptions (BBMP):

$$\begin{aligned} \mu_{ij} &\sim iid\left(0,\sigma_{\mu}^{2}\right), v_{it} \sim iid\left(0,\sigma_{v}^{2}\right), \\ \zeta_{jt} &\sim iid\left(0,\sigma_{\zeta}^{2}\right), \varepsilon_{ijt} \sim iid\left(0,\sigma_{\varepsilon}^{2}\right) \end{aligned} \tag{154}$$

where μ_{ij}, v_{it}, ζ_{jt} and ε_{ijt} are pairwise uncorrelated.

- We rewrite (126) sequentially as

$$\mathbf{u}_{ij} = \mu_{ij}\boldsymbol{\iota}_T + \mathbf{v}_i + \boldsymbol{\zeta}_j + \boldsymbol{\varepsilon}_{ij},\ i = 1, ..., N_1,\ j = 1, ..., N_2,$$
$$\mathbf{u}_i = \boldsymbol{\mu}_i \otimes \boldsymbol{\iota}_T + \boldsymbol{\iota}_{N_2} \otimes \mathbf{v}_i + \boldsymbol{\zeta} + \boldsymbol{\varepsilon}_i, i = 1, ..., N_1,$$
$$\mathbf{u} = \boldsymbol{\mu} \otimes \boldsymbol{\iota}_T + \mathbf{V} + \boldsymbol{\iota}_{N_1} \otimes \boldsymbol{\zeta} + \boldsymbol{\varepsilon} \qquad (155)$$

$$\underset{T\times 1}{\mathbf{u}_{ij}} = \begin{bmatrix} u_{ij1} \\ \vdots \\ u_{ijT} \end{bmatrix}, \underset{T\times 1}{\mathbf{v}_i} = \begin{bmatrix} v_{i1} \\ \vdots \\ v_{iT} \end{bmatrix}, \underset{T\times 1}{\boldsymbol{\zeta}_j} = \begin{bmatrix} \zeta_{j1} \\ \vdots \\ \zeta_{jT} \end{bmatrix}, \underset{T\times 1}{\boldsymbol{\varepsilon}_{ij}} = \begin{bmatrix} \varepsilon_{ij1} \\ \vdots \\ \varepsilon_{ijT} \end{bmatrix}$$

$$\underset{N_2T\times 1}{\mathbf{u}_i} = \begin{bmatrix} \mathbf{u}_{i1} \\ \vdots \\ \mathbf{u}_{iN_2} \end{bmatrix}, \underset{N_2\times 1}{\boldsymbol{\mu}_i} = \begin{bmatrix} \mu_{i1} \\ \vdots \\ \mu_{iN_2} \end{bmatrix}, \underset{N_2T\times 1}{\boldsymbol{\zeta}} = \begin{bmatrix} \boldsymbol{\zeta}_1 \\ \vdots \\ \boldsymbol{\zeta}_{N_2} \end{bmatrix}, \underset{N_2T\times 1}{\boldsymbol{\varepsilon}_i} =$$

$$\underset{N_1N_2T\times 1}{\mathbf{u}} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{N_1} \end{bmatrix}, \underset{N_1N_2\times 1}{\boldsymbol{\mu}} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_{N_1} \end{bmatrix}, \underset{N_1N_2T\times 1}{\mathbf{V}} = \begin{bmatrix} \boldsymbol{\iota}_{N_2} \otimes \mathbf{v}_1 \\ \vdots \\ \boldsymbol{\iota}_{N_2} \otimes \mathbf{v}_{N_1} \end{bmatrix}$$

(156)

- It is easily seen that

$$\begin{aligned} Cov\left(\mathbf{u}\right) &= \mathbf{I}_{N_1 N_2} \otimes \left(\sigma_\mu^2 \mathbf{J}_T\right) + \mathbf{I}_{N_1} \otimes \mathbf{J}_{N_2} \otimes \left(\sigma_v^2 \mathbf{I}_T\right) \\ &+ \mathbf{J}_{N_1} \otimes \left(\sigma_\zeta^2 \mathbf{I}_{N_2 T}\right) + \sigma_\varepsilon^2 \mathbf{I}_{N_1 N_2 T} \end{aligned} \quad (157)$$

- CTRE imposes very limited structure of CSD, because for $i \neq i'$ and $j \neq j'$, we have:

$$E[u_{ijt}u_{ij't}] = \sigma_v^2,\ E[u_{ijt}u_{ij't}] = \sigma_\zeta^2 \text{ and } E[u_{ijt}u_{i'j't}] = 0. \quad (158)$$

- Next, consider the 3D model with the 2-way heterogeneous factor, (133). It is straightforward to derive:

$$\mathbf{u} = \boldsymbol{\mu} \otimes \boldsymbol{\iota}_T + \boldsymbol{\pi} \otimes \boldsymbol{\lambda}_T + \boldsymbol{\varepsilon} \tag{159}$$

$$\underset{N_1 N_2 \times 1}{\boldsymbol{\pi}} = \begin{bmatrix} \boldsymbol{\pi}_1 \\ \vdots \\ \boldsymbol{\pi}_{N_1} \end{bmatrix}, \quad \underset{N_2 \times 1}{\boldsymbol{\pi}_i} = \begin{bmatrix} \pi_{i1} \\ \vdots \\ \pi_{iN_2} \end{bmatrix}, \quad \underset{T \times 1}{\boldsymbol{\lambda}_T} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_T \end{bmatrix} \tag{160}$$

- The covariance matrix for $\mathbf{u}$ in (159) is:

$$Cov(\mathbf{u}) = \mathbf{I}_{N_1 N_2} \otimes \left(\sigma_\mu^2 \mathbf{J}_T\right) + \left(\boldsymbol{\pi}\boldsymbol{\pi}'\right) \otimes \left(\sigma_\lambda^2 \mathbf{I}_T\right) + \sigma_\varepsilon^2 \mathbf{I}_{N_1 N_2 T} \tag{161}$$

- It captures CSD by

$$\begin{aligned} E[u_{ijt}u_{ij't}] &= \pi_{ij}\pi_{ij'}\sigma_\lambda^2, \; E[u_{ijt}u_{i'jt}] = \pi_{ij}\pi_{i'j}\sigma_\lambda^2, \quad (162) \\ E[u_{ijt}u_{i'j't}] &= \pi_{ij}\pi_{i'j'}\sigma_\lambda^2. \end{aligned}$$

- Consider the 3D model with general error components, (141). It is straightforward to derive:

$$\mathbf{u} = \boldsymbol{\mu} \otimes \boldsymbol{\iota}_T + \mathbf{V} + \boldsymbol{\iota}_{N_1} \otimes \boldsymbol{\zeta} + \boldsymbol{\pi} \otimes \boldsymbol{\lambda}_T + \boldsymbol{\varepsilon}. \tag{163}$$

- The covariance matrix for $\mathbf{u}$ is given by

$$\begin{aligned} Cov(\mathbf{u}) &= \mathbf{I}_{N_1N_2} \otimes \left(\sigma_\mu^2 \mathbf{J}_T\right) + \mathbf{I}_{N_1} \otimes \mathbf{J}_{N_2} \otimes \left(\sigma_v^2 \mathbf{I}_T\right) \\ &\quad + \mathbf{J}_{N_1} \otimes \left(\sigma_\zeta^2 \mathbf{I}_{N_2T}\right) + \left(\boldsymbol{\pi}\boldsymbol{\pi}'\right) \otimes \left(\sigma_\lambda^2 \mathbf{I}_T\right) + \sigma_\varepsilon^2 \mathbf{I}_{N_1N_2T} \end{aligned} \tag{164}$$

- It captures CSD by

$$\begin{aligned} E[u_{ijt}u_{ij't}] &= \pi_{ij}\pi_{ij'}\sigma_\lambda^2 + \sigma_v^2, E[u_{ijt}u_{i'jt}] = \pi_{ij}\pi_{i'j}\sigma_\lambda^2 + \sigma_\xi^2, \\ E[u_{ijt}u_{i'j't}] &= \pi_{ij}\pi_{i'j'}\sigma_\lambda^2. \end{aligned}$$

- **Remark**: CTFE accommodates non-zero covariance locally, but imposes the same covariance for all $i = 1, ..., N_1$ and $j = 1, ..., N_2$. Such restrictions are too strong.
- Our proposed error components (141) accommodates non-zero covariances both locally and globally.
- Consider the heterogeneous local factors specifications:

$$v_{it} = v_i \tau_t \text{ and } \zeta_{jt} = \zeta_j \tau_t^* \tag{165}$$

where τ_t and τ_t^* are the importer and exporter-specific local factors. Then,we replace (141) by

$$u_{ijt} = \mu_{ij} + v_i \tau_t + \zeta_j \tau_t^* + \varepsilon_{ijt}. \tag{166}$$

- Exporter i reacts heterogeneously to the common import market condition τ_t and importer j reacts heterogeneously to the common export market condition τ_t^*.

- KSS propose a hierarchical multi-factor error components specification:

$$u_{ijt} = \mu_{ij} + v_i \tau_{it} + \zeta_j \tau^*_{jt} + \pi_{ij} \lambda_t + \varepsilon_{ijt}. \qquad (167)$$

- We can distinguish between three types of CSD:
  - the strong global factor, λ_t influences the (ij) pairwise interactions (of $N_1 N_2$ dimension);
  - the semi-strong local factors, τ_{it} and τ^*_{jt}, influence exporters or importers separately (each of N_1 or N_2 dimension);
  - the weak CSD idiosyncratic errors, ε_{ijt}.

- We assume that

$$\begin{aligned} \mu_{ij} &\sim iid\left(0, \sigma_\mu^2\right), \tau_{it} \sim iid\left(0, \sigma_\tau^2\right), \tau_{jt}^* \sim iid\left(0, \sigma_{\tau^*}^2\right), \\ \lambda_t &\sim iid\left(0, \sigma_\lambda^2\right), \varepsilon_{ijt} \sim iid\left(0, \sigma_\varepsilon^2\right) \end{aligned} \quad (168)$$

where μ_{ij}, τ_{it}, τ_{jt}^*, λ_t and ε_{ijt} are mutually independent.

- The model (167) captures CSD by

$$E[u_{ijt}u_{ij't}] = v_i^2\sigma_\tau^2 + \pi_{ij}\pi_{ij'}\sigma_\lambda^2, \; E[u_{ijt}u_{i'jt}] = \zeta_j^2\sigma_{\tau^*}^2 + \pi_{ij}\pi_{i'j}\sigma_\lambda^2$$

$$E[u_{ijt}u_{i'j't}] = \pi_{ij}\pi_{i'j'}\sigma_\lambda^2.$$

- The covariance structure is more flexible than (**??**).

- The diagnostic test for the null hypothesis of residual cross–section independence in the 3D panels using the residuals, $\mathbf{e}_{ij} = (e_{ij1}, ..., e_{ijT})'$.
- We have $\mathbf{e}_{ij} = \tilde{\mathbf{Y}}_{ij} - \tilde{\mathbf{X}}_{ij}\hat{\boldsymbol{\beta}}_W$ for the model (129), $\mathbf{e}_{ij} = \mathbf{M}_H\mathbf{Y}_{ij} - \mathbf{M}_H\mathbf{W}_{ij}\hat{\boldsymbol{\theta}}_{PCCE}$ for (137), and $\mathbf{e}_{ij} = \mathbf{M}_{\tilde{F}}\tilde{\mathbf{Y}}_{ij} - \mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}\hat{\boldsymbol{\beta}}_{PCCE}$ for (146).
- The cross-section dependence (CD) test is a modified counterpart of an existing CD test by Pesaran (2015).

- We represent $\mathbf{e}_{ij}$ as the (ij) pair using the single index $n = 1, ..., N_1N_2$, and compute the pair-wise residual correlations between n and n' cross-section units by

$$\hat{\rho}_{nn'} \left(= \rho_{n'n}\right) = \frac{\mathbf{e}_n' \mathbf{e}_{n'}}{\sqrt{\left(\mathbf{e}_n' \mathbf{e}_n\right)\left(\mathbf{e}_{n'}' \mathbf{e}_{n'}\right)}}, \ n, n' = 1, ..., N_1N_2 \text{ and } n \neq n'.$$

- We construct the CD statistic by

$$CD = \sqrt{\frac{2}{N_1N_2\left(N_1N_2 - 1\right)}} \sum_{n=1}^{N_1N_2-1} \sum_{n'=n+1}^{N_1N_2} \sqrt{T}\hat{\rho}_{nn'} \quad (169)$$

- CD test has the limiting $N(0,1)$ distribution under the null $H_0 : \hat{\rho}_{nn'} = 0$ for all $n, n' = 1, ..., N_1N_2$ and $n \neq n'$ (Pesaran, 2015).

- An issue of unbalanced panels has been almost neglected even in the 2D panels with unobserved factors.
- Kapetanios and Pesaran (2005) briefly deal with it in their Monte Carlo studies.
- Bai et al. (2015) investigate the unbalanced 2D panel data model with interactive effects, and propose the functional principal components analysis and the EM algorithm.

- BMW derive the complex within transformation, which is computationally demanding as it involves an inversion of $NT \times NT$ matrices.
- We introduce a vector of selection indicators for each pair (i,j), $\mathbf{s}_{ij} = (s_{ij,1},...,s_{ij,T})'$, where $s_{ij,t} = 1$ if time period t for pair (i,j) can be used in estimation.
- Following Wooldridge (2010), we assume that selection is ignorable conditional on $(\mathbf{x}_{ijt}, \mathbf{s}_{it}, \mathbf{d}_{jt}, \mathbf{q}_t, \mathbf{z}_{ij}, \mu_{ij}, \lambda_t)$:

$$\begin{aligned} & E\left(y_{it}|\mathbf{x}_{ijt}, \mathbf{s}_{it}, \mathbf{d}_{jt}, \mathbf{q}_t, \mathbf{z}_{ij}, \mu_{ij}, \lambda_t, \mathbf{s}_i\right) \\ = \ & E\left(y_{it}|\mathbf{x}_{ijt}, \mathbf{s}_{it}, \mathbf{d}_{jt}, \mathbf{q}_t, \mathbf{z}_{ij}, \mu_{ij}, \lambda_t\right). \end{aligned}$$

- Let $n=\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}s_{ij,t}$ be the total number of observations.
- Define $n_t=\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}s_{ij,t}$ and $n_{ij}=\sum_{t=1}^{T}s_{ij,t}$ as the number of cross-section pairs at period t and the number of time periods for pair (i,j).
- Define $n_i=\sum_{j=1}^{N_2}\sum_{t=1}^{T}s_{ij,t}$, $n_j=\sum_{i=1}^{N_1}\sum_{t=1}^{T}s_{ij,t}$, $n_{it}=\sum_{j=1}^{N_2}s_{ij,t}$ and $n_{jt}=\sum_{i=1}^{N_1}s_{ij,t}$.
- We maintain the assumption: $\left(\min_i n_i,\min_j n_j,\min_t n_t,\min_{(ij)} n_{ij}\right)\to\infty$ or $\left(\min_t n_t,\min_{(ij)} n_{ij}\right)\to\infty$.

- We multiply the 3D model (125) with the error components (133) by the selection indicator to get:

$$y^s_{ijt} = \boldsymbol{\beta}'\mathbf{x}^s_{ijt} + \boldsymbol{\gamma}'\mathbf{s}^s_{it} + \boldsymbol{\delta}'\mathbf{d}^s_{jt} + \boldsymbol{\kappa}'\mathbf{q}^s_t + \boldsymbol{\varphi}'\mathbf{z}^s_{ij} + \mu^s_{ij} + \pi^s_{ij}\lambda_t + \varepsilon^s_{ijt}, \quad (170)$$

where $y_{ijt} = s_{ij,t}y_{ijt}$, $\mathbf{x}^s_{ijt} = s_{ij,t}\mathbf{x}_{ijt}$, $\mathbf{s}^s_{it} = s_{ij,t}\mathbf{s}_{it}$, $\mathbf{d}^s_{jt} = s_{ij,t}\mathbf{d}_{jt}$, $\mathbf{q}^s_t = s_{ij,t}\mathbf{q}_t$, $\mathbf{z}^s_{ij} = s_{ij,t}\mathbf{z}_{ij}$, $\mu^s_{ij} = s_{ij,t}\mu_{ij}$, $\pi^s_{ij} = s_{ij,t}\pi_{ij}$, $\varepsilon^s_{ijt} = s_{ij,t}\varepsilon_{ijt}$.

- Applying the cross-section averages of (170) over i and j,

$$\bar{y}^s_{..t} = \boldsymbol{\beta}'\bar{\mathbf{x}}^s_{..t} + \boldsymbol{\gamma}'\bar{\mathbf{s}}^s_{.t} + \boldsymbol{\delta}'\bar{\mathbf{d}}^s_{.t} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\bar{\mathbf{z}}^s_{..t} + \bar{\mu}^s_{..t} + \bar{\pi}^s_{..t}\lambda_t + \bar{\varepsilon}^s_{..t} \quad (171)$$

where $\bar{y}^s_{..t} = \frac{1}{n_t}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} s_{ij,t}y_{ijt} = \sum_{i=1}^{N_1} w_{it}\bar{y}^s_{i.t}$ is a weighted average with $w_{it} = n_{it}/n_t$ and $\bar{y}^s_{i.t} = n_{it}^{-1}\sum_{j=1}^{N_2} s_{ij,t}y_{ijt}$.

- Similarly for $\bar{\mathbf{x}}^s_{..t}$, $\bar{\mathbf{z}}^s_{..t}$, $\bar{\mu}^s_{..t}$, $\bar{\pi}^s_{..t}$ and $\bar{\varepsilon}^s_{..t}$.
- Further, $\bar{\mathbf{s}}^s_{.t} = \sum_{i=1}^{N_1} w_{it}\mathbf{s}_{it}$, $\bar{\mathbf{d}}^s_{.t} = \sum_{j=1}^{N_2} w_{jt}\mathbf{d}_{jt}$ with $w_{jt} = n_{jt}/n_t$, and $\bar{\mathbf{q}}^s_t = \mathbf{q}_t$.

- As $n_t \to \infty$,

$$\bar{\mathbf{z}}^s_{..t} = \bar{\mathbf{z}} + o_p(1), \bar{\mu}^s_{..t} = \bar{\mu} + o_p(1), \ \bar{\pi}^s_{..t} = \bar{\pi} + o_p(1), \bar{\varepsilon}^s_{..t} = \bar{\varepsilon}_{..t} + o_p(1) \tag{172}$$

where $\bar{\mathbf{z}} = (N_1 N_2)^{-1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbf{z}_{ij} \to_p E(\mathbf{z}_{ij})$,
$\bar{\mu} = (N_1 N_2)^{-1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mu_{ij} \to_p 0$,
$\bar{\pi} = (N_1 N_2)^{-1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \pi_{ij} \to_p E(\pi_{ij}) \neq 0$,
$\bar{\varepsilon}_{..t} = (N_1 N_2)^{-1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \varepsilon_{ijt} \to_p 0$.

- Using (172), we rewrite (171) as

$$\bar{y}^s_{..t} = \boldsymbol{\beta}' \bar{\mathbf{x}}^s_{..t} + \boldsymbol{\gamma}' \bar{\mathbf{s}}^s_{.t} + \boldsymbol{\delta}' \bar{\mathbf{d}}^s_{.t} + \boldsymbol{\kappa}' \mathbf{q}_t + \boldsymbol{\varphi}' \bar{\mathbf{z}} + \bar{\mu} + \bar{\pi}\lambda_t + \bar{\varepsilon}_{..t} + o_p(1)$$

- Hence, λ_t can be approximated by

$$\lambda_t \simeq \frac{1}{\bar{\pi}} \left\{ \bar{y}^s_{..t} - \left(\boldsymbol{\beta}' \bar{\mathbf{x}}^s_{..t} + \boldsymbol{\gamma}' \bar{\mathbf{s}}^s_{.t} + \boldsymbol{\delta}' \bar{\mathbf{d}}^s_{.t} + \boldsymbol{\kappa}' \mathbf{q}_t + \boldsymbol{\varphi}' \bar{\mathbf{z}} + \bar{\mu} + \bar{\varepsilon}_{..t} \right) \right\}.$$

- We augment the model (170) with cross-section averages:

$$y^s_{ijt} = \boldsymbol{\beta}'\mathbf{x}^s_{ijt} + \boldsymbol{\gamma}'\mathbf{s}^s_{it} + \boldsymbol{\delta}'\mathbf{d}^s_{jt} + \boldsymbol{\psi}'_{ij}\mathring{\mathbf{f}}^s_t + \tau^s_{ij} + \mu^s_{ij} + \varepsilon^{*s}_{ijt}, \quad (173)$$

where $\tau^s_{ij} = s_{ij,t}\tau_{ij}$, $\varepsilon^{*s}_{ijt} = s_{ij,t}\varepsilon^*_{ijt}$ and

$$\mathring{\mathbf{f}}^s_t = s_{ij,t}\mathbf{f}^s_t \text{ with } \mathbf{f}^s_t = \left(\bar{y}^s_{..t}, \bar{\mathbf{x}}^{s\prime}_{..t}, \bar{\mathbf{s}}^{s\prime}_{.t}, \bar{\mathbf{d}}^{s\prime}_{.t}, \mathbf{q}'_t\right)' \quad (174)$$

- Collecting the n_{ij} observations with $s_{ij,t} = 1$ from (173),

$$\mathbf{Y}_{ij} = \mathbf{W}_{ij}\boldsymbol{\theta} + \mathbf{H}_{ij}\boldsymbol{\psi}^*_{ij} + \mathbf{E}^*_{ij} \tag{175}$$

where $\mathbf{W}_{ij} = \left(\mathbf{X}_{ij}, \mathbf{S}_{ij}, \mathbf{D}_{ij}\right)$, $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta}' & \boldsymbol{\gamma}' & \boldsymbol{\delta}' \end{pmatrix}'$, $\boldsymbol{\psi}^*_{ij} = \left(\boldsymbol{\psi}'_{ij}, \left(\tau_{ij} + \mu_{ij}\right)\right)'$, $\mathbf{H}_{ij} = \left[\mathbf{F}_{ij}, \boldsymbol{\iota}_{n_{ij}}\right]$ and

$$\underset{n_{ij}\times 1}{\mathbf{Y}_{ij}} = \begin{bmatrix} y_{ij(1)} \\ \vdots \\ y_{ij(n_{ij})} \end{bmatrix}, \underset{T_{ij}\times k_x}{\mathbf{X}_{ij}} = \begin{bmatrix} \mathbf{x}'_{ij(1)} \\ \vdots \\ \mathbf{x}'_{ij(n_{ij})} \end{bmatrix}, \underset{T_{ij}\times k_s}{\mathbf{S}_{ij}} = \begin{bmatrix} \mathbf{s}'_{i(1)} \\ \vdots \\ \mathbf{s}'_{i(n_{ij})} \end{bmatrix}$$

$$\underset{n_{ij}\times k_d}{\mathbf{D}_{ij}} = \begin{bmatrix} \mathbf{d}'_{j(1)} \\ \vdots \\ \mathbf{d}'_{j(n_{ij})} \end{bmatrix}, \underset{n_{ij}\times k_f}{\mathbf{F}_{ij}} = \begin{bmatrix} \mathbf{f}^{s\prime}_{(1)} \\ \vdots \\ \mathbf{f}^{s\prime}_{(n_{ij})} \end{bmatrix}, \underset{T_{ij}\times 1}{\mathbf{E}_{ij}} = \begin{bmatrix} \varepsilon^*_{ij(1)} \\ \vdots \\ \varepsilon^*_{ij(n_{ij})} \end{bmatrix}.$$

- We express the time index inside (.) to highlight different initial and last periods for each pair (ij).

- The 3D-PCCE estimator of $\boldsymbol{\theta}$ is obtained by

$$\hat{\boldsymbol{\theta}}_{PCCE} = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbf{W}'_{ij} \mathbf{M}_{H_{ij}} \mathbf{W}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbf{W}'_{ij} \mathbf{M}_{H_{ij}} \mathbf{Y}_{ij} \right) \tag{176}$$

where $\mathbf{M}_{H_{ij}} = \mathbf{I}_{T_{ij}} - \mathbf{H}_{ij} \left(\mathbf{H}'_{ij} \mathbf{H}_{ij} \right)^{-1} \mathbf{H}'_{ij}$.

- Next, we multiply the 3D model (125) with the general error components (141) by $s_{ij,t} = 1$:

$$y_{ijt}^{s} = \boldsymbol{\beta}'\mathbf{x}_{ijt}^{s} + \boldsymbol{\gamma}'\mathbf{s}_{it}^{s} + \boldsymbol{\delta}'\mathbf{d}_{jt}^{s} + \boldsymbol{\kappa}'\mathbf{q}_{t}^{s} + \boldsymbol{\varphi}'\mathbf{z}_{ij}^{s} + \mu_{ij}^{s} + \upsilon_{it}^{s} + \zeta_{jt}^{s} + \pi_{ij}\lambda_{t}^{s} + \varepsilon_{ijt}^{s} \tag{177}$$

where $y_{ijt}^{s} = s_{ij,t}y_{ijt}$ and similarly for others.

- Taking the cross-section averages of (177) over i and j,

$$\bar{y}_{..t}^{s} = \boldsymbol{\beta}'\bar{\mathbf{x}}_{..t}^{s} + \boldsymbol{\gamma}'\bar{\mathbf{s}}_{.t}^{s} + \boldsymbol{\delta}'\bar{\mathbf{d}}_{.t}^{s} + \boldsymbol{\kappa}'\mathbf{q}_{t} + \boldsymbol{\varphi}'\bar{\mathbf{z}}_{..t}^{s} + \bar{\mu}_{..t}^{s} + \bar{\upsilon}_{.t}^{s} + \bar{\zeta}_{.t}^{s} + \bar{\pi}_{..t}^{s}\lambda_{t} + \bar{\varepsilon}_{..t}^{s} \tag{178}$$

where $\bar{\upsilon}_{.t}^{s} = \sum_{i=1}^{N_1} w_{it}\upsilon_{it}$ with $w_{it} = n_{it}/n_{t}$,
$\bar{\zeta}_{.t}^{s} = \sum_{j=1}^{N_2} w_{jt}\zeta_{jt}$ with $w_{jt} = n_{jt}/n_{t}$.

- As $n_t \to \infty$,

$$\bar{\upsilon}^s_{.t} = \bar{\upsilon} + o_p(1) \text{ and } \bar{\zeta}^s_{.t} = \bar{\zeta} + o_p(1) \tag{179}$$

where $\bar{\upsilon} = N_1^{-1} \sum_{i=1}^{N_1} \upsilon_{it} \to_p 0$ and $\bar{\zeta} = N_2^{-1} \sum_{j=1}^{N_2} \zeta_{jt} \to_p 0$.

- Using (172) and (179), we can approximate $\bar{y}^s_{..t}$ and λ_t by

$$\bar{y}^s_{..t} = \boldsymbol{\beta}'\bar{\mathbf{x}}^s_{..t} + \boldsymbol{\gamma}'\bar{\mathbf{s}}^s_{.t} + \boldsymbol{\delta}'\bar{\mathbf{d}}^s_{.t} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\bar{\mathbf{z}} + \bar{\mu} + \bar{\upsilon} + \bar{\zeta} + \bar{\pi}\lambda_t + \bar{\varepsilon}_{..t} + o_p(1)$$

$$\lambda_t = \frac{1}{\bar{\pi}} \left\{ \bar{y}^s_{..t} - \left(\boldsymbol{\beta}'\bar{\mathbf{x}}^s_{.t} + \boldsymbol{\gamma}'\bar{\mathbf{s}}^s_{.t} + \boldsymbol{\delta}'\bar{\mathbf{d}}^s_{.t} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\bar{\mathbf{z}} + \bar{\mu} + \bar{\upsilon} + \bar{\zeta} + \bar{\varepsilon}_{..t}\right.\right.$$

- We augment the model (177) with the cross-section averages:

$$y^s_{ijt} = \boldsymbol{\beta}'\mathbf{x}^s_{ijt} + \boldsymbol{\gamma}'\mathbf{s}^s_{it} + \boldsymbol{\delta}'\mathbf{d}^s_{jt} + \boldsymbol{\psi}'_{ij}\mathbf{f}^s_t + \tau^s_{ij} + \mu^s_{ij} + \upsilon^s_{it} + \zeta^s_{jt} + \varepsilon^{*s}_{ijt}, \tag{180}$$

where $\varepsilon^{*s}_{ijt} = s_{ij,t}\varepsilon^{*}_{ijt}$ with
$\varepsilon^{*}_{ijt} = \varepsilon_{ijt} - \frac{\pi_{ij}}{\bar{\pi}} \left(\bar{\varepsilon}_{..t} + \bar{\mu} + \bar{\upsilon} + \bar{\zeta} \right) \to_p \varepsilon_{ijt}$.

- Consider the simpler specification:

$$y_{ijt}^{s} = \mu_{ij}^{s} + v_{it}^{s} + \xi_{jt}^{s} + \varepsilon_{ijt}^{s}, \tag{181}$$

and examine the transformed data:

$$\tilde{y}_{ijt}^{s} = y_{ijt}^{s} + s_{ij,t}\left(-\bar{y}_{ij\cdot}^{s} - \bar{y}_{\cdot jt}^{s} - \bar{y}_{i\cdot t}^{s} + \bar{y}_{\cdot\cdot t}^{s} + \bar{y}_{\cdot j\cdot}^{s} + \bar{y}_{i\cdot\cdot}^{s} - \bar{y}_{\cdot\cdot\cdot}^{s}\right) \tag{182}$$

- It is straightforward to show:

$$\begin{aligned} & \left(-\bar{y}_{ij\cdot}^{s} - \bar{y}_{\cdot jt}^{s} - \bar{y}_{i\cdot t}^{s} + \bar{y}_{\cdot\cdot t}^{s} + \bar{y}_{\cdot j\cdot}^{s} + \bar{y}_{i\cdot\cdot}^{s} - \bar{y}_{\cdot\cdot\cdot}^{s}\right) \\ = \; & -\left(\mu_{ij} + v_{it} + \xi_{jt}\right) + D_1 + D_2 + D_3 + D_4 + D_5 \end{aligned}$$

- where

$$D_1 = -\left(\bar{\upsilon}^s_{ij.} - \sum_{j=1}^{N_2} \frac{n_{ij}}{n_i} \bar{\upsilon}^s_{ij.}\right) + \left(\sum_{i=1}^{N_1} \frac{n_{ij}}{n_j} \bar{\upsilon}^s_{ij.} - \sum_{i=1}^{N_1} \frac{n_i}{n} \sum_{j=1}^{N_2} \frac{n_{ij}}{n_i} \bar{\upsilon}^s_{ij.}\right)$$

$$D_2 = -\left(\bar{\xi}^s_{ij.} - \sum_{i=1}^{N_1} \frac{n_{ij}}{n_j} \bar{\xi}^s_{ij.}\right) + \left(\sum_{j=1}^{N_2} \frac{n_{ij}}{n_i} \bar{\xi}^s_{ij.} - \sum_{j=1}^{N_2} \frac{n_j}{n} \sum_{i=1}^{N_1} \frac{n_{ij}}{n_j} \bar{\xi}^s_{ij.}\right)$$

$$D_3 = -\left(\bar{\mu}_{.jt} - \sum_{t=1}^{T} \frac{n_{jt}}{n_j} \bar{\mu}_{.jt}\right) - \left(\bar{\mu}_{i.t} - \sum_{t=1}^{T} \frac{n_{it}}{n_i} \bar{\mu}_{i.t}\right) + \left(\bar{\mu}_{..t} - \sum_{t=1}^{T} \frac{n_t}{n}\right.$$

$$D_4 = -\left(\bar{\upsilon}^s_{.jt} - \sum_{j=1}^{N_2} \frac{n_{jt}}{n_t} \bar{\upsilon}^s_{.jt}\right), \quad D_5 = -\left(\bar{\xi}_{i.t} - \sum_{i=1}^{N_1} \frac{n_{it}}{n_t} \bar{\xi}_{i.t}\right)$$

with $\bar{\upsilon}^s_{ij.} = \frac{1}{n_{ij}} \sum_{t=1}^{T} s_{ij,t} \upsilon_{it}$, $\bar{\xi}^s_{ij.} = \frac{1}{n_{ij}} \sum_{t=1}^{T} s_{ij,t} \xi_{jt}$,
$\bar{\mu}^s_{.jt} = \frac{1}{n_{jt}} \sum_{i=1}^{N_1} s_{ij,t} \mu_{ij}$, $\bar{\mu}^s_{i.t} = \frac{1}{n_{it}} \sum_{j=1}^{N_2} s_{ij,t} \mu_{ij}$,
$\bar{\mu}^s_{..t} = \frac{1}{n_t} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} s_{ij,t} \mu_{ij}$, $\bar{\upsilon}^s_{.jt} = \frac{1}{n_{jt}} \sum_{i=1}^{N_1} s_{ij,t} \upsilon_{it}$,

- In the balanced panels $D_1 = D_2 = D_3 = D_4 = D_5 = 0$.
- As $\left(\min_i n_i, \min_j n_j, \min_t n_t, \min_{(ij)} n_{ij}\right) \to \infty$, $D_i \to_p 0$ for $i = 1, ..., 5$.
- Therefore,

$$\begin{aligned} & \left(-\bar{y}^s_{ij\cdot} - \bar{y}^s_{\cdot jt} - \bar{y}^s_{i\cdot t} + \bar{y}^s_{\cdot\cdot t} + \bar{y}^s_{\cdot j\cdot} + \bar{y}^s_{i\cdot\cdot} - \bar{y}^s_{\cdot\cdot\cdot}\right) \\ = & -\left(\mu_{ij} + v_{it} + \xi_{jt}\right) + o_p(1) \end{aligned} \quad (183)$$

- Using (183) and applying (182) to (181), we obtain:

$$\tilde{y}^s_{ijt} = \tilde{\varepsilon}^s_{ijt}, \quad (184)$$

where
$\tilde{\varepsilon}^s_{ijt} = \varepsilon^s_{ijt} - s_{ij,t}\left(\bar{\varepsilon}^s_{ij\cdot} - \bar{\varepsilon}^s_{\cdot jt} - \bar{\varepsilon}^s_{i\cdot t} + \bar{\varepsilon}^s_{\cdot\cdot t} + \bar{\varepsilon}^s_{\cdot j\cdot} + \bar{\varepsilon}^s_{i\cdot\cdot} - \bar{\varepsilon}^s_{\cdot\cdot\cdot}\right)$.

- We apply the 3D-within transformation (182) to (180):

$$\tilde{y}^s_{ijt} = \boldsymbol{\beta}'\tilde{\mathbf{x}}^s_{ijt} + \tilde{\boldsymbol{\psi}}'_{ij}\mathring{\mathbf{f}}^s_{ijt} + \tilde{\varepsilon}^{*s}_{ijt}, \tag{185}$$

where $\tilde{\boldsymbol{\psi}}_{ij} = \boldsymbol{\psi}'_{ij} - \left(\frac{1}{n_{jt}}\sum_{i=1}^{N_1}\boldsymbol{\psi}'_{ij}\right) - \left(\frac{1}{n_{it}}\sum_{j=1}^{N_2}\boldsymbol{\psi}'_{ij}\right) + \left(\frac{1}{n_t}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\boldsymbol{\psi}'_{ij}\right)$, $\mathring{\mathbf{f}}^s_{ijt} = s_{ij,t}\tilde{\mathbf{f}}^s_{ij}$ with $\tilde{\mathbf{f}}^s_{ij} = \mathbf{f}^s_t - \bar{\mathbf{f}}^s_{ij}$ and $\bar{\mathbf{f}}^s_{ij} = n^{-1}_{ij}\sum_{t=1}^{T} s_{ij,t}\mathbf{f}^s_t$.

- Collecting only the n_{ij} observations with $s_{ij,t} = 1$ from (185),

$$\tilde{\mathbf{Y}}_{ij} = \tilde{\mathbf{X}}_{ij}\boldsymbol{\beta} + \tilde{\mathbf{F}}_{ij}\tilde{\boldsymbol{\psi}}_{ij} + \tilde{\mathbf{E}}^*_{ij}, \tag{186}$$

where

$$\underset{n_{ij}\times 1}{\tilde{\mathbf{Y}}_{ij}} = \begin{bmatrix} \tilde{y}_{ij(1)} \\ \vdots \\ \tilde{y}_{ij(n_{ij})} \end{bmatrix}, \underset{n_{ij}\times k_x}{\tilde{\mathbf{X}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ij(1)} \\ \vdots \\ \tilde{\mathbf{x}}'_{ij(n_{ij})} \end{bmatrix}, \underset{n_{ij}\times k_f}{\tilde{\mathbf{F}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{f}}^{s\prime}_{ij(1)} \\ \vdots \\ \tilde{\mathbf{f}}^{s\prime}_{ij(n_{ij})} \end{bmatrix}, n$$

- The 3D-PCCE estimators of β are obtained by

$$\tilde{\beta}_{PCCE} = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij} \mathbf{M}_{ij} \tilde{\mathbf{X}}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij} \mathbf{M}_{ij} \tilde{\mathbf{Y}}_{ij} \right) \tag{187}$$

  where $\mathbf{M}_{ij} = \mathbf{I}_T - \tilde{\mathbf{F}}_{ij} \left(\tilde{\mathbf{F}}'_{ij} \tilde{\mathbf{F}}_{ij} \right)^{-1} \tilde{\mathbf{F}}'_{ij}$.

- As $\left(\min_t n_t, \min_{(ij)} n_{(ij)}\right) \to \infty$, both PCCE estimators, (176) and (187), will follow the asymptotic normal distribution.

- BMW propose the following 4D model:

$$y_{ijst} = \mathbf{x}'_{ijst}\boldsymbol{\beta} + u_{ijst}, \tag{188}$$

$$u_{ijst} = \mu_{ijs} + \theta_{ijt} + \zeta_{jst} + v_{ist} + \varepsilon_{ijst} \tag{189}$$

for $i = 1, ..., N_1$, $j = 1, ..., N_2$, $s = 1, ..., N_3$, $t = 1, ..., T$.

- BMW derive the 4D within transformation to eliminate pair-wise interaction effects, μ_{ijs}, v_{ist}, ζ_{jst}, and λ_{ijt}:

$$\begin{aligned} \tilde{y}_{ijst} &= y_{ijst} - \bar{y}_{.jst} - \bar{y}_{i.st} - \bar{y}_{ij.t} - \bar{y}_{ijs.} + \bar{y}_{..st} + \bar{y}_{.j.t} + \bar{y}_{.js.} \\ &\quad + \bar{y}_{i..t} + \bar{y}_{i.s.} + \bar{y}_{ij..} - \bar{y}_{...t} - \bar{y}_{..s.} - \bar{y}_{.j..} - \bar{y}_{i...} - \bar{y}_{....} \end{aligned} \tag{190}$$

and estimate $\boldsymbol{\beta}$ consistently from

$$\tilde{y}_{ijst} = \tilde{\mathbf{x}}'_{ijst}\boldsymbol{\beta} + \tilde{u}_{ijst}. \tag{191}$$

- BBMP propose the feasible GLS estimator under the assumption that error components are pairwise uncorrelated.

- To introduce CSD into (188), consider the following extension:

$$y_{ijst} = \mathbf{x}'_{ijst}\boldsymbol{\beta} + \mu_{ijs} + \theta_{ijt} + \zeta_{jst} + \upsilon_{ist} + \pi_{ijs}\lambda_t + \varepsilon_{ijst} \quad (192)$$

- The 4D-FE and 4D-RE estimators biased due to the correlation between $\mathbf{x}_{ijst}$ and λ_t.
- We develop the two-step consistent estimation procedure.

- Taking the cross-section averages of (192) over i, j and s,

$$\bar{y}_{...t} = \boldsymbol{\beta}'\bar{\mathbf{x}}_{...t} + \bar{\mu}_{...} + \bar{\theta}_{..t} + \bar{\zeta}_{..t} + \bar{\upsilon}_{..t} + \bar{\pi}_{...}\lambda_t + \bar{\varepsilon}_{...t} \quad (193)$$

where $\bar{\mathbf{x}}_{...t} = \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\mathbf{x}_{ijst}$,
$\bar{\varepsilon}_{...t} = \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\varepsilon_{ijst}$,
$\bar{\mu}_{...} = \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\mu_{ijs}$,
$\bar{\pi}_{...} = \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\pi_{ijs}$,
$\bar{\theta}_{..t} = \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\theta_{ijt}$, $\bar{\zeta}_{..t} = \frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\zeta_{jst}$,
and $\bar{\upsilon}_{..t} = \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N}\sum_{s=1}^{N_3}\upsilon_{ist}$.

- From (193) we have:

$$\lambda_t = \frac{1}{\pi_{...}} \left\{ \bar{y}_{...t} - \left(\boldsymbol{\beta}' \bar{\mathbf{x}}_{...t} + \bar{\mu}_{...} + \bar{\theta}_{..t} + \bar{\zeta}_{..t} + \bar{\upsilon}_{..t} + \bar{\pi}_{...} \lambda_t + \bar{\varepsilon}_{...t} \right) \right\}.$$

- We derive the cross-section augmented version of (192) by

$$y_{ijst} = \boldsymbol{\beta}' \mathbf{x}_{ijst} + \boldsymbol{\psi}'_{ijs} \mathbf{f}_t + \mu_{ijs} + \theta_{ijt} + \zeta_{jst} + \upsilon_{ist} + \varepsilon^*_{ijst}, \quad (194)$$

where $\mathbf{f}_t = (\bar{y}_{...t}, \bar{\mathbf{x}}'_{...t})'$,
$\boldsymbol{\psi}'_{ijs} = \left(\psi_{0,ijs}, \boldsymbol{\psi}'_{ijs} \right) = \left(\frac{\pi_{ijs}}{\bar{\pi}_{...}}, -\frac{\pi_{ijs}}{\bar{\pi}_{...}} \boldsymbol{\beta}' \right)$ and
$\varepsilon^*_{ijst} = \varepsilon_{ijst} - \frac{\pi_{ijs}}{\bar{\pi}_{...}} \left(\bar{\varepsilon}_{...t} + \bar{\theta}_{..t} + \bar{\zeta}_{..t} + \bar{\upsilon}_{..t} \right)$.

- As $N_1, N_2, N_3 \to \infty$, $\varepsilon^*_{ijst} \to_p \varepsilon_{ijst}$.

- Apply the 4D-within transformation (190) to (194):

$$\tilde{y}_{ijst} = \boldsymbol{\beta}' \tilde{\mathbf{x}}_{ijst} + \tilde{\boldsymbol{\psi}}'_{ijs} \tilde{\mathbf{f}}_t + \tilde{\varepsilon}^*_{ijt}. \tag{195}$$

where $\tilde{\mathbf{f}}_t = \left(\mathbf{f}_t - \bar{\mathbf{f}}\right)$ and

$$\tilde{\boldsymbol{\psi}}'_{ijs} = \left(\boldsymbol{\psi}_{ijs} - \boldsymbol{\psi}_{.js} - \boldsymbol{\psi}_{i.s} - \boldsymbol{\psi}_{ij.} + \boldsymbol{\psi}_{..s} + \boldsymbol{\psi}_{.j.} + \boldsymbol{\psi}_{i..} - \boldsymbol{\psi}_{...}\right)'$$

- We rewrite (195) as

$$\tilde{\mathbf{Y}}_{ijs} = \tilde{\mathbf{X}}_{ijs} \boldsymbol{\beta} + \tilde{\mathbf{F}} \tilde{\boldsymbol{\psi}}_{ijs} + \tilde{\boldsymbol{\varepsilon}}^*_{ijs}, \tag{196}$$

where

$$\underset{T \times 1}{\tilde{\mathbf{Y}}_{ijs}} = \begin{bmatrix} \tilde{y}_{ijs1} \\ \vdots \\ \tilde{y}_{ijsT} \end{bmatrix}, \quad \underset{T \times k_x}{\tilde{\mathbf{X}}_{ijs}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ijs1} \\ \vdots \\ \tilde{\mathbf{x}}'_{ijsT} \end{bmatrix}, \quad \underset{T \times k_f}{\tilde{\mathbf{F}}} = \begin{bmatrix} \tilde{\mathbf{f}}'_1 \\ \vdots \\ \tilde{\mathbf{f}}'_T \end{bmatrix}.$$

- It is straightforward to derive the PCCE estimator of β by

$$\hat{\beta}_{PCCE} = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{s=1}^{N_3} \tilde{\mathbf{X}}'_{ijs} \mathbf{M}_{\tilde{\mathbf{F}}} \tilde{\mathbf{X}}_{ijs} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{s=1}^{N_3} \tilde{\mathbf{X}}'_{ijs} \mathbf{M}_{\tilde{\mathbf{F}}} \tilde{\mathbf{Y}}_{ijs} \right) \tag{197}$$

where $\mathbf{M}_{\tilde{\mathbf{F}}} = \mathbf{I}_T - \tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}' \tilde{\mathbf{F}} \right)^{-1} \tilde{\mathbf{F}}'$.

- We follow KSS and develop 4D models with the hierarchical multi-factor error structure.
- Define the global factor λ_t which affects all (ijs) pairs, the regional factors τ_{it}, τ^*_{jt}, τ^{**}_{st}, and finally the local factors τ_{ijt}, τ^*_{ist} and τ^{**}_{jst}.
- This suggests the following model:

$$\begin{aligned} y_{ijst} &= \mathbf{x}'_{ijst}\boldsymbol{\beta} + \mu_{ijs} + \upsilon_{js}\tau_{it} + \upsilon^*_{is}\tau_{jt} + \upsilon^{**}_{ij}\tau_{st} + \zeta_s\tau_{ijt} \\ &\quad + \zeta^*_j\tau^*_{ist} + \zeta^{**}_i\tau^{**}_{jst} + \pi_{ij}\lambda_t + \varepsilon_{ijst}. \end{aligned} \tag{198}$$

- Such setups involve several layers of factor specifications (a number that grows with the dimension), rendering their estimation challenging.

- We construct DGP1 by

$$y_{ijt} = \beta' x_{ijt} + \mu_{ij} + \pi_{ij}\lambda_t + \varepsilon_{ijt}, \tag{199}$$

$$x_{ijt} = \mu^x_{ij} + \mu_{ij} + \pi^x_{ij}\lambda_t + v_{ijt}, \tag{200}$$

for $i = 1, ..., N_1$, $j = 1, ..., N_2$, and $t = 1, ..., T$.

- The global factor, λ_t and idiosyncratic errors, ε_{ijt} and v_{ijt} are generated independently as *iid* processes

$$\lambda_t \sim iidN(0,1), \ \varepsilon_{ijt} \sim iidN(0,1), \ v_{ijt} \sim iidN(0,1).$$

- We generate pairwise individual effects independently as

$$\mu_{ij} \sim iidN(0,1), \ \mu^x_{ij} \sim iidN(0,1).$$

- Factor loadings, π_{ij} and π^x_{ij}, are independently generated from $U[1,2]$.

- Next, we construct DGP2 by

$$y_{ijt} = \beta' x_{ijt} + \mu_{ij} + v_{it} + \zeta_{jt} + \pi_{ij}\lambda_t + \varepsilon_{ijt}. \tag{201}$$

$$x_{ijt} = \mu_{ij}^x + \mu_{ij} + \pi_{ij}^x \lambda_t + v_{ijt}, , \tag{202}$$

  for $i = 1, ..., N_1$, $j = 1, ..., N_2$, and $t = 1, ..., T$.

- In addition, we generate v_{it} and ζ_{jt} independently as:

$$v_{it} \sim U(-1, 1) \text{ and } \zeta_{jt} \sim U(-1, 1)$$

- In both DGP1 and DGP2 we set $\beta = 1$.

- We evaluate the following summary statistics:
- Bias: $\hat{\beta}_R - \beta_0\,(=1)$ and $\hat{\beta}_R = R^{-1}\sum_{r=1}^{R}\hat{\beta}_r$.
- RMSE: $\sqrt{R^{-1}\sum_{r=1}^{R}\left(\hat{\beta}_r - \beta_0\right)^2}$.
- Size: the rejection probability of the t-statistic for the null $\beta = \beta_0$ against $\beta \neq \beta_0$ at the 5%.
- We conduct experiment 1,000 times for $N_1, N_2 = \{25, 49, 100\}$ and $T = \{50, 100, 200, 400\}$.

- In Table 1 biases of the 2D PCCE and 3D PCCE estimators of β are mostly negligible even for $(N_1, N_2, T) = (25, 25, 50)$.
- The CTFE estimator displays substantial biases.
- RMSE results are qualitatively similar to the bias pattern.
- CTFE over-rejects the null in all cases and tends to 1 even as N_1 (N_2) or T rises.
- The size of the 2D PCCE is close to the nominal 5% while 3D PCCE slightly over-rejects when N_1 or N_2 is small.
- Overall performance of the 2D PCCE estimator is the best under DGP1.

Table: Simulation results for β under the DGP1

	CTFE						
	Bias						
(N_1N_2, T)	50	100	200	400			
25	0.0829	0.0832	0.0833	0.0822			
49	0.0347	0.0341	0.0338	0.0344			
100	-0.0307	-0.0315	-0.0313	-0.0316			
	RMSE						
(N_1N_2, T)	50	100	200	400			
25	0.0914	0.0871	0.0854	0.0832			
49	0.0420	0.0383	0.0357	0.0353			
100	0.0347	0.0336	0.0324	0.0322			
	Size						
(N_1N_2, T)	50	100	200	400			
25	0.7610	0.9590	0.9980	1.0000			
49	0.4020	0.6290	0.8810	0.9950			
100	0.5530	0.8410	0.9910	1.0000			
	2D PCCE				3D PCCE		
	Bias						
(N_1N_2, T)	50	100	200	400	50	100	200
25	0.0017	0.0011	0.0014	0.0003	0.0017	0.0008	0.0012
49	0.0006	-0.0005	0.0002	0.0004	0.0000	-0.0001	0.0000
100	0.0004	0.0005	0.0001	0.0003	0.0000	0.0002	0.0000

Notes: We report the simulation results for three estimators for the DGP1, (199) and (200). CTFE refers to the 3D within estimator given by (129), 2D PCCE is the PCCE estimator given by (137) and 3D PCCE is the PCCE estimator given by (146).

- Simulation results in Table 2 are qualitatively similar to those in Table 1.
- Biases of PCCE are almost negligible and their RMSEs decrease rapidly with N_1 (N_2) or T.
- Empirical sizes are still close to the nominal 5% level.
- CTFE suffers from substantial biases and size distortions, and its performance does not improve in large samples.
- Good performance of the 2D PCCE is rather surprising as the 3D PCCE estimator is expected to dominate.
- Overall simulation results support the simulation findings reported under the 2D panels by Kapetanios and Pesaran (2005) and Pesaran (2006).

Table: Simulation results for β under the DGP2

	CTFE						
	Bias						
(N_1N_2, T)	50	100	200	400			
25	0.0835	0.0829	0.0830	0.0827			
49	0.0143	0.0144	0.0155	0.0156			
100	-0.0365	-0.0371	-0.0362	-0.0370			
	RMSE						
(N_1N_2, T)	50	100	200	400			
25	0.0921	0.0872	0.0850	0.0839			
49	0.0272	0.0220	0.0194	0.0177			
100	0.0400	0.0388	0.0371	0.0374			
	Size						
(N_1N_2, T)	50	100	200	400			
25	0.7780	0.9450	1.0000	1.0000			
49	0.1420	0.2060	0.3630	0.5650			
100	0.7120	0.9400	0.9940	1.0000			
	2D PCCE				3D PCCE		
	Bias						
(N_1N_2, T)	50	100	200	400	50	100	200
25	-0.0001	0.0008	0.0009	0.0001	0.0012	0.0006	0.0009
49	-0.0002	0.0000	0.0006	0.0001	-0.0001	0.0000	0.0005
100	0.0000	0.0001	0.0001	0.0002	0.0001	0.0003	0.0001

Notes: We report the simulation results for three estimators for the DGP2, (201) and (202). CTFE refers to the 3D within estimator given by (129), 2D PCCE is the PCCE estimator given by (137) and 3D PCCE is the PCCE estimator given by (146).

- Anderson and van Wincoop (2003): "The gravity equation tells us that bilateral trade, after controlling for size, depends on the bilateral trade barriers but relative to the product of their Multilateral Resistance Indices (MTR)."
- Omitting MTR induces severe bias (e.g. Baldwin and Taglioni, 2006).
- Subsequent research focused on estimating the model with directional country-specific fixed effects to control for unobservable MTRs (e.g. Feenstra, 2004).

- A large number of studies established an importance of taking into account multilateral resistance and bilateral heterogeneity in the 2D panels.
- Serlenga and Shin (2007) is the first to develop the panel gravity model by incorporating observed and unobserved factors.
- Behrens et al. (2012) develop the spatial econometric specification, to control for multilateral cross-sectional correlations across trade flows.
- Mastromarco et al. (2015) compare the factor- and the spatial-based gravity models to investigate the Euro impact on intra-EU trade flows over 1960-2008 for 190 country-pairs of 14 EU and 6 non-EU OECD countries.
- The CD test confirms that the factor-based model is more appropriate for controlling for CSD.

- For the 3D models, we should control for source of biases presented by unobserved time-varying MTRs.
- Baltagi et al. (2003) propose the 3D model (128) with CTFE specification (126).
- This approach popular in measuring the impacts of MTRs of the exporters and the importers in the structural gravity studies (e.g. Baltagi et al., 2015).
- CTFE or CTRE estimators fail to accommodate (strong and heterogeneous) CSD.
- The presence of CSD across (ij) pairs suggests that the appropriate econometric techniques be required.

- We apply our approach to the dataset covering the period 1960-2008 (49 years) for 182 country-pairs amongst 14 EU member countries (Austria, Belgium-Luxemburg, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden, United Kingdom).
- Consider the generalised panel gravity specification:

$$\ln EXP_{ijt} = \beta_0 + \beta_1 CEE_{ijt} + \beta_2 EMU_{ijt} + \beta_3 SIM_{ijt} + \beta_4 RLF_{ijt} \quad (203)$$

$$+ \beta_5 \ln GDP_{it} + \beta_6 \ln GDP_{jt} + \beta_7 RER_t$$

$$+ \gamma_1 DIS_{ij} + \gamma_2 BOR_{ij} + \gamma_3 LAN_{ij} + u_{ijt}$$

- The dependent variable EXP_{ijt} is the export flow from country i to country j at time t;
- CEE and EMU are dummies for European Community membership and European Monetary Union;
- SIM and RLF measure similarity in size and difference in relative factor endowments;
- RER represents the logarithm of common real exchange rates;
- GDP_{it} and GDP_{jt} are logged GDPs of exporter and importer;
- The logarithm of geographical distance (DIS) and the dummies for common language (LAN) and for common border (BOR) represent time-invariant bilateral barriers.

- We report the estimation results of (258) for 4 estimators;
  - the two-way within estimator with $u_{ijt} = \mu_{ij} + \lambda_t + \varepsilon_{ijt}$;
  - the CTFE estimator with $u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \varepsilon_{ijt}$;
  - the 2D PCCE estimator with $u_{ijt} = \mu_{ij} + \pi_{ij}\lambda_t + \varepsilon_{ijt}$;
  - the 3D PCCE estimator with $u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \pi_{ij}\lambda_t + \varepsilon_{ijt}$.
- We report the CD test results for the residuals and the estimates of the CSD exponent (α).
- Our focus is on the impacts of t_{ij} that contain both barriers and incentives to trade. We focus on the two dummy variables;
- CEE (one when both countries belong to the European Community);
- EMU (one when both adopt the same currency).
- Both are expected to exert a positive impact on bilateral export flows.

- The empirical evidence is mixed.
- Rose (2001), Frankel and Rose (2002), Glick and Rose (2002) and Frankel (2008), document a huge positive effect;
- A number of studies report negative or insignificant effects (Persson, 2001, Pakko and Wall, 2002, De Nardis and Vicarelli, 2003).
- Recent studies by Serlenga and Shin (2007), Mastromarco et al. (2015) and Gunnella et al. (2015) findng a small but significant effect (7 to 10%) of the euro on intra-EU trade, after controlling for strong CSD.

- Table 3 reports the estimation results.
- The two-way FE estimation results are statistically significant except RER.
- The impacts of home and foreign GDPs on exports are positive, but surprisingly, the former is twice larger than the latter.
- The impact of SIM is negative and significant, inconsistent with *a priori* expectations.
- CEE and EMU significantly boost exports, but their magnitudes seem to be too high.
- The CD test rejects the null of no or weak CSD convincingly.
- $\hat{\alpha}$ is 0.99 with CI containing unity; the residuals strongly correlated and the FE results biased and unreliable.
- This supports our main concern that upward trends in omitted trade determinants may cause them to be upward-biased.

- We turn to the CTFE estimation results.
- CD test results indicate that the CTFE residuals do not suffer from any strong CSD.
- This rather surprising result is not supported by $\hat{\alpha} = 0.91$ (pretty close to 1).
- All the coefficients become insignificant except for CEE.
- The CEE is still substantial (0.29) while the EMU turns negligible (-0.011).
- Overall CTFE results are unreliable.

- The 2D PCCE results are significant with the expected signs except for EMU.
- The impact of foreign GDP on exports is substantially larger than home GDP.
- The RER is positive, confirming that a depreciation of the home currency increases exports.
- The CEE is smaller (0.186), but EMU is insignificant and negligible (0.017).
- The 2D PCCE suffers from strong CSD residuals with $\hat{\alpha} = 0.87$.

- Finally, the 3D PCCE results show that all the coefficients are significant with the expected signs.
- The CD test fails to strongly reject the null, supported by the smaller estimate of $\hat{\alpha} = 0.77$, close to a moderate range of weak CSD.[2]
- CEE still substantial (0.335) while the EMU modest at 0.081, close to the consensus reported in the 2D panel studies (e.g. Baldwin, 2006, Gunnella et al., 2015).
- The 3D PCCE results are mostly reliable, suggesting that the trade-boosting effect of the Euro should be viewed in the long-run historical and multilateral perspectives rather than simply focusing on the formation of a monetary union as an isolated event.

[2] BHP show that the values of $\alpha \in [1/2, 3/4)$ represent a moderate degree of CSD.

Table: 3D panel gravity model estimation results for bilateral export flows

	FE			CTFE		
	Coeff	se	t-ratio	Coeff	se	t-ratio
gdph	2.185	0.041	52.97			
gdpf	1.196	0.041	28.98			
sim	-0.263	0.052	-5.069	-0.055	0.074	-0.754
rlf	0.031	0.006	5.011	0.006	0.005	1.294
rer	0.005	0.007	0.791	0.031	0.072	0.436
cee	0.302	0.014	22.05	0.290	0.017	16.99
emu	0.204	0.019	10.71	-0.011	0.036	-0.315
CD stat	620.1			-2.676		
	$\alpha_{0.05}$	α	$\alpha_{0.95}$	$\alpha_{0.05}$	α	$\alpha_{0.95}$
CSD exponent	0.925	0.992	1.059	0.865	0.914	0.963
	2D PCCE			3D PCCE		
	Coeff	se	t-ratio	Coeff	se	t-ratio
gdph	0.289	0.095	3.033			
gdpf	1.491	0.095	15.69			
sim	0.042	0.105	0.401	1.032	0.111	9.290
rlf	0.007	0.005	1.420	-0.004	0.005	-0.748
rer	0.144	0.019	7.427	0.168	0.114	1.471
cee	0.187	0.014	13.20	0.335	0.022	15.10
emu	0.018	0.015	1.160	0.081	0.045	1.793

Notes: Using the annual dataset over 1960-2008 for 182 country-pairs amongst 14 EU member countries, we estimate the generalised panel gravity specification, (258). FE stands for the standard two-way fixed effects estimator with country-pair and time fixed effects. CTFE refers to the 3D within estimator given by (129). 2D PCCE is the PCCE estimator given by (137) with factors $\mathbf{f}_t = \left\{\overline{gdp}_{.t}, \overline{sim}_{..t}, \overline{rlf}_{..t}, \overline{cee}_{..t}, \overline{rer}_t, t\right\}$. 3D PCCE is the PCCE estimator given by (146) with factors $\mathbf{f}_t = \left\{\overline{sim}_{..t}, \overline{rlf}_{..t}, \overline{rer}_t\right\}$. CD test refers to testing the null hypothesis of residual cross-sectional error independence or weak dependence and is defined in (169). CSD exponent denotes the point estimate of the exponents of CSD α and the 90% level confidence bands.

- The CTFE estimator is proposed to capture (unobserved) multilateral resistance terms and trade costs, but it fails to accommodate strong CSD among MTRs, clearly present in our sample of the EU countries (confirmed by CD tests and CSD exponent estimates).
- We should model the time-varying interdependency of bilateral export flows in a flexible manner than simply introducing deterministic country-time specific dummies.
- MTRs arise from the bilateral country-pair specific reactions to global shocks or the local spillover effects or both.

- We propose novel estimation techniques to accommodate CSD within the 3D panel data models.
- Our framework is a generalisation of the multidimensional country-time fixed and random effects estimators.
- Our approach is the first attempt to introduce strong CSD into the multi-dimensional error components.
- We develop the two-step estimation procedure, the 3D-PCCE estimator.
- The empirical usefulness of the 3D-PCCE estimator is demonstrated via the Monte Carlo studies and the empirical application to the gravity model of the intra-EU trade.

- Extensions and generalisations.
- First, we develop the multi-dimensional heterogenous panel data models with hierarchical multi-factor error structure (KSS).
- Next, we aim to develop the challenging models by combining the spatial- and the factor-based techniques.
  - Bailey et al. (2016) develop the multi-step estimation procedure that can distinguish the relationship between spatial units from that which is due to the effect of common factors.
  - Mastromarco et al. (2015) propose the technique for allowing weak and strong CSD in stochastic frontier panels by combining the exogenously driven factor-based approach and an endogenous threshold regime selection by Kapetanios et al. (2014, KMS).
  - Bai and Li (2015) and Shi and Lee (2014,5) developed the framework for jointly modelling spatial effects and interactive effects.
  - See also Gunnella et al. (2015) and Kuersteiner and Prucha (2015).

Introduction

- Given the growing availability of the multidimensional dataset, recent studies attempted to extend the two-way model to the multidimensional setting.
- Balazsi, Mátyás and Wansbeek (2015, BMW) introduce the 3D within estimators for the three-way fixed effects panel data models and analyses their behavior.
- The 3D models applied to a number of bilateral flows such as trade, FDI, capital or migration as well as a variety of matched dataset (the employer-employees or pupils to teachers).

Introduction

- Consider the 3D country-time fixed effects panel data model:
$$y_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \boldsymbol{\gamma}'\mathbf{s}_{it} + \boldsymbol{\delta}'\mathbf{d}_{jt} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\mathbf{z}_{ij} + u_{ijt}, \quad (204)$$
for $i = 1, ..., N_1, j = 1, ..., N_2, t = 1, ..., T$, with errors:
$$u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \varepsilon_{ijt} \quad (205)$$
where y_{ijt} is the dependent variable across 3 indices (the import of country j from country i at period t); $\mathbf{x}_{ijt}$, $\mathbf{s}_{it}$, $\mathbf{d}_{jt}$, $\mathbf{q}_t$, $\mathbf{z}_{ij}$ are the $k_x \times 1$, $k_s \times 1$, $k_d \times 1$, $k_q \times 1$, $k_z \times 1$ vectors of covariates covering all measurements across 3indices.
- The multi-error components contain pair-fixed effects (μ_{ij}) as well as origin and destination CTFEs, v_{it} and ζ_{jt}.
- The specification (205), proposed by Baltagi et al. (2003), applied to measure the impacts of (unobserved) multilateral resistance of exporters and importers in the structural gravity studies (Feenstra, 2004).

Motivations and extensions

- KMSS consider the following error components:

$$u_{ijt} = \mu_{ij} + \pi_{ij}\lambda_t + \varepsilon_{ijt} \tag{206}$$

$$u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \pi_{ij}\lambda_t + \varepsilon_{ijt} \tag{207}$$

- In this paper we model the error components, u_{ijt} to follow the hierarchical multi-factor structure:

$$u_{ijt} = \boldsymbol{\gamma}'_{ij}\boldsymbol{f}_{t} + \boldsymbol{\gamma}'_{\circ j}\boldsymbol{f}_{i\circ t} + \boldsymbol{\gamma}'_{i\circ}\boldsymbol{f}_{\circ jt} + \varepsilon_{ijt}, \tag{208}$$

- Exporter i reacts heterogeneously to the common import market condition $\boldsymbol{f}_{\circ jt}$ and importer j reacts heterogeneously to the common export market condition $\boldsymbol{f}_{i\circ t}$. Both reacts heterogeneously to the common global market condition $\boldsymbol{f}_{t}$.
- 3 types of CSD: (i) strong global factor, $\boldsymbol{f}_{t}$ influences the (ij) pairwise interactions (of N^2 dimension); (ii) semi-strong local factors, $\boldsymbol{f}_{i\circ t}$ and $\boldsymbol{f}_{\circ jt}$, influence origin or destination (of N dimension); and (iii) weak CSD ε_{ijt}.

- The full estimation can be feasible by combining the Pesaran or the Bai type estimation procedures.
- A few studies attempt to develop an approach that can accommodate both weak and strong CSD in panels.
- Bailey et al. (2015) develop the multi-step estimation procedure.
- Mastromarco et al. (2015) propose the novel technique for allowing weak and strong CSD in modelling technical efficiency of stochastic frontier panels by combining the exogenously driven factor-based approach by SS and an endogenous threshold regime selection by KMS.
- Bai and Li (2015) and Shi and Lee (2017) develop the framework for jointly modelling spatial effects and interactive effects. See also Kuersteiner and Prucha (2015).
- An extension of such joint modelling to the multidimensional data would be challenging but shed further lights on the understanding the complex structure of CSD.

The model setup

- Consider the triple-index heterogeneous panel data model:

$$y_{ijt} = \boldsymbol{\beta}'_{ij}\boldsymbol{x}_{ijt} + \boldsymbol{\delta}'_{ij}\boldsymbol{d}_t + u_{ijt},\ i, j = 1, ..., N,\ t = 1, ..., T, \tag{209}$$

- y_{ijt} is the dependent variable observed across 3 indices, i the origin, j the destination at period t (say, the export from country i to j at t);
- $\boldsymbol{x}_{ijt}$ is the $m_x \times 1$ vector of covariates;
- $\boldsymbol{d}_t$ is the $m_d \times 1$ vector of observed common effects such as constants and trends.
- $\boldsymbol{\beta}_{ij}$ and $\boldsymbol{\delta}_{ij}$ are the $m_x \times 1$ and $m_d \times 1$ vectors of parameters.

- We allow u_{ijt} to follow the hierarchical multi-factor structure:

$$u_{ijt} = \boldsymbol{\gamma}_{ij}'\boldsymbol{f}_{t} + \boldsymbol{\gamma}_{\circ j}'\boldsymbol{f}_{i\circ t} + \boldsymbol{\gamma}_{i\circ}'\boldsymbol{f}_{\circ jt} + \varepsilon_{ijt} \qquad (210)$$

- $\boldsymbol{f}_{t}$, $\boldsymbol{f}_{\circ jt}$ and $\boldsymbol{f}_{i\circ t}$ are $m_f \times 1$, $m_{\circ\bullet} \times 1$ and $m_{\bullet\circ} \times 1$ vectors of unobserved common effects'
- ε_{ijt} are idiosyncratic errors distributed independently of $(\boldsymbol{x}_{ijt}, \boldsymbol{d}_t)$.
- $\boldsymbol{f}_{t}$ are the global factors affecting all of the bilateral pairs;
- $\boldsymbol{f}_{i\circ t}$ and $\boldsymbol{f}_{\circ jt}$ are local origin i and destination j factors.
- They are designed to account for commonality in y_{ijt}; CSD between a given flow and a flow from the exporting region's commonality to the importing region (exporting-based dependence) and another flow from the exporting region to the importing region's commonality (importing-based dependence).
- This can provide an alternative to the existing literature on business cycles, e.g. Kose et al. (2003) and Choi et al. (2016)

- To deal with the general case where $\boldsymbol{f}_t$, $\boldsymbol{f}_{\circ jt}$ and $\boldsymbol{f}_{i\circ t}$, are correlated with $(\boldsymbol{x}_{ijt}, \boldsymbol{d}_t)$, we consider the following DGP:

$$\boldsymbol{x}_{ijt} = \mathcal{D}_{ij}\boldsymbol{d}_t + \boldsymbol{\Gamma}_{ij}\boldsymbol{f}_t + \boldsymbol{\Gamma}_{\circ j}\boldsymbol{f}_{i\circ t} + \boldsymbol{\Gamma}_{i\circ}\boldsymbol{f}_{\circ jt} + \boldsymbol{v}_{ijt}, \qquad (211)$$

where $\mathcal{D}_{ij}$ is the $(m_x \times m_d)$ parameter matrix, $\boldsymbol{\Gamma}_{ij}$, $\boldsymbol{\Gamma}_{\circ j}$ and $\boldsymbol{\Gamma}_{i\circ}$ are $(m_x \times m_f)$, $(m_x \times m_{\bullet\circ})$, $(m_x \times m_{\circ\bullet})$ factor loading matrices, and $\boldsymbol{v}_{ijt}$ are the idiosyncratic errors.

- Combining (209)-(211), we have:

$$\boldsymbol{z}_{ijt} = \begin{pmatrix} y_{ijt} \\ \boldsymbol{x}_{ijt} \end{pmatrix} = \boldsymbol{\Xi}_{ij}\boldsymbol{d}_t + \boldsymbol{\Phi}_{ij}\boldsymbol{f}_t + \boldsymbol{\Phi}_{\circ j}\boldsymbol{f}_{i\circ t} + \boldsymbol{\Phi}_{i\circ}\boldsymbol{f}_{\circ jt} + \boldsymbol{u}_{ijt} \qquad (212)$$

$$\boldsymbol{\Xi}_{ij} = \begin{pmatrix} \boldsymbol{\delta}'_{ij} + \boldsymbol{\beta}'_{ij}\mathcal{D}_{ij} \\ \mathcal{D}_{ij} \end{pmatrix}, \boldsymbol{\Phi}_{ij} = \begin{pmatrix} \boldsymbol{\gamma}'_{ij} + \boldsymbol{\beta}'_{ij}\boldsymbol{\Gamma}_{ij} \\ \boldsymbol{\Gamma}_{ij} \end{pmatrix}, \boldsymbol{\Phi}_{i\circ} = \begin{pmatrix} \boldsymbol{\gamma}'_{i\circ} + \\ \boldsymbol{\Gamma} \end{pmatrix} \qquad (213)$$

$$\boldsymbol{\Phi}_{\circ j} = \begin{pmatrix} \boldsymbol{\gamma}'_{\circ j} + \boldsymbol{\beta}'_{ij}\boldsymbol{\Gamma}_{\circ j} \\ \boldsymbol{\Gamma}_{\circ j} \end{pmatrix}, \boldsymbol{u}_{ijt} = \begin{pmatrix} \varepsilon_{ijt} + \boldsymbol{\beta}'_{ij}\boldsymbol{v}_{ijt} \\ \boldsymbol{v}_{ijt} \end{pmatrix}.$$

- The ranks of $\boldsymbol{\Phi}_{ij}$, $\boldsymbol{\Phi}_{i\circ}$ and $\boldsymbol{\Phi}_{\circ j}$ determined by the ranks of

- Rewrite (209) and (212) in the matrix notation:

$$\boldsymbol{y}_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta}_{ij} + \boldsymbol{D}\boldsymbol{\delta}_{ij} + \boldsymbol{F}\boldsymbol{\gamma}_{ij} + \boldsymbol{F}_{i\circ}\boldsymbol{\gamma}_{\circ j} + \boldsymbol{F}_{\circ j}\boldsymbol{\gamma}_{i\circ} + \boldsymbol{\varepsilon}_{ij}, \quad (214)$$

$$\boldsymbol{z}_{ij} = \boldsymbol{D}\boldsymbol{\Xi}_{ij} + \boldsymbol{F}\boldsymbol{\Phi}_{ij} + \boldsymbol{F}_{i\circ}\boldsymbol{\Phi}_{\circ j} + \boldsymbol{F}_{\circ j}\boldsymbol{\Phi}_{i\circ} + \boldsymbol{u}_{ij}, \quad (215)$$

where

$$\underset{T\times 1}{\boldsymbol{y}_{ij}} = \begin{bmatrix} y_{ij1} \\ \vdots \\ y_{ijT} \end{bmatrix}, \underset{T\times m_x}{\boldsymbol{X}_{ij}} = \begin{bmatrix} \boldsymbol{x}'_{ij1} \\ \vdots \\ \boldsymbol{x}'_{ijT} \end{bmatrix}, \underset{T\times m_d}{\boldsymbol{D}} = \begin{bmatrix} \boldsymbol{d}'_1 \\ \vdots \\ \boldsymbol{d}'_T \end{bmatrix}, \underset{T\times(m_x+1)}{\boldsymbol{z}_{ij}} = \quad (216)$$

$$\underset{T\times m_f}{\boldsymbol{F}} = \begin{bmatrix} \boldsymbol{f}'_1 \\ \vdots \\ \boldsymbol{f}'_T \end{bmatrix}, \underset{T\times m_{\bullet\circ}}{\boldsymbol{F}_{i\circ}} = \begin{bmatrix} \boldsymbol{f}'_{i\circ 1} \\ \vdots \\ \boldsymbol{f}'_{i\circ T} \end{bmatrix}, \underset{T\times m_{\circ\bullet}}{\boldsymbol{F}_{\circ j}} = \begin{bmatrix} \boldsymbol{f}'_{\circ j1} \\ \vdots \\ \boldsymbol{f}'_{\circ jT} \end{bmatrix}, \underset{T\times 1}{\boldsymbol{\varepsilon}_{ij}} = \begin{bmatrix} \varepsilon_{ij1} \\ \vdots \\ \varepsilon_{ijT} \end{bmatrix}$$

Assumption 1. *Common Effects*: The $(m_d + m_f + m_{\bullet\circ} + m_{\circ\bullet}) \times 1$ vector of common factors $\boldsymbol{g}_t = \left(\boldsymbol{d}_t', \boldsymbol{f}'_t, \boldsymbol{f}'_{i\circ t}, \boldsymbol{f}'_{\circ jt}\right)'$, is covariance stationary with absolute summable autocovariances, distributed independently of $\varepsilon_{ijt'}$ and $\boldsymbol{v}_{ijt'}$ for all i, j, t and t'.

Assumption 2. *Individual-specific Errors*: ε_{ijt} and $\boldsymbol{v}_{ijt'}$ are distributed independently for all i, j, t and t', and they are distributed independently of $\boldsymbol{x}_{ijt}$ and $\boldsymbol{d}_t$.

Assumption 3. *Factor Loadings*: Unobserved factor loadings are independently and identically distributed across (i, j), and of ε_{ijt}, $\boldsymbol{v}_{ijt}$, $\boldsymbol{g}_t$ for all i, j, t, with finite means and variances. In particular,

$$\boldsymbol{\gamma}_{ij} = \boldsymbol{\gamma}_{\circ\circ} + \boldsymbol{\eta}_{ij}, \quad \boldsymbol{\gamma}_{i\circ} = \boldsymbol{\gamma}_{\bullet\circ} + \boldsymbol{\eta}_{i\circ}, \quad \boldsymbol{\gamma}_{\circ j} = \boldsymbol{\gamma}_{\circ\bullet} + \boldsymbol{\eta}_{\circ j}, \tag{217}$$

$$\boldsymbol{\Gamma}_{ij} = \boldsymbol{\Gamma}_{\circ\circ} + \boldsymbol{\xi}_{ij}, \quad \boldsymbol{\Gamma}_{i\circ} = \boldsymbol{\Gamma}_{\bullet\circ} + \boldsymbol{\xi}_{i\circ}, \quad \boldsymbol{\Gamma}_{\circ j} = \boldsymbol{\Gamma}_{\circ\bullet} + \boldsymbol{\xi}_{\circ j}, \tag{218}$$

where $\boldsymbol{\eta}_{ij} \sim iid\left(0, \boldsymbol{\Omega}_{\boldsymbol{\eta}_{\circ\circ}}\right)$, $\boldsymbol{\xi}_{ij} \sim iid\left(0, \boldsymbol{\Omega}_{\boldsymbol{\xi}_{\circ\circ}}\right)$, $\boldsymbol{\eta}_{i\circ} \sim iid\left(0, \boldsymbol{\Omega}_{\boldsymbol{\eta}_{\bullet\circ}}\right)$, $\boldsymbol{\xi}_{i\circ} \sim iid\left(0, \boldsymbol{\Omega}_{\boldsymbol{\xi}_{\bullet\circ}}\right)$, $\boldsymbol{\eta}_{\circ j} \sim iid\left(0, \boldsymbol{\Omega}_{\boldsymbol{\eta}_{\circ\bullet}}\right)$ and $\boldsymbol{\xi}_{\circ j} \sim iid\left(0, \boldsymbol{\Omega}_{\boldsymbol{\xi}_{\circ\bullet}}\right)$. Further, $\|\boldsymbol{\gamma}_{\circ\circ}\| < K$, $\|\boldsymbol{\gamma}_{\bullet\circ}\| < K$, $\|\boldsymbol{\gamma}_{\circ\bullet}\| < K$, $\|\boldsymbol{\Gamma}_{\circ\circ}\| < K$, $\|\boldsymbol{\Gamma}_{\bullet\circ}\| < K$, and $\|\boldsymbol{\Gamma}_{\circ\bullet}\| < K$ for positive constant $K < \infty$.

Assumption 4. *Random Slope Coefficients*:

$$\boldsymbol{\beta}_{ij} = \boldsymbol{\beta} + \boldsymbol{\nu}_{i\circ} + \boldsymbol{\nu}_{\circ j} + \boldsymbol{\nu}_{ij}, \boldsymbol{\nu}_{i\circ} \sim iid\left(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\nu}_{\bullet\circ}}\right), \boldsymbol{\nu}_{\circ j} \sim iid\left(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\nu}_{\circ\bullet}}\right), \boldsymbol{\nu}_{ij} \sim i \tag{219}$$

where $\|\boldsymbol{\beta}\| < K$ and $\boldsymbol{\nu}_{ij}$, $\boldsymbol{\nu}_{i\circ}$, $\boldsymbol{\nu}_{\circ j}$ are distributed independently of one another, and of $\boldsymbol{\gamma}_{ij}$, $\boldsymbol{\Gamma}_{ij}$, ε_{ijt}, $\boldsymbol{v}_{ijt}$ and $\boldsymbol{g}_t$ for all i, j and t.

Assumption 5. *Identification of* $\boldsymbol{\beta}_{ij}$ *and* $\boldsymbol{\beta}$: Construct the cross-section averages of $\boldsymbol{z}_{ijt}$ by

$$\bar{\boldsymbol{z}}_t = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{z}_{ijt},\ \ \bar{\boldsymbol{z}}_{i\circ t} = \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{z}_{ijt} \text{ and } \bar{\boldsymbol{z}}_{\circ jt} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{ijt} \tag{220}$$

Let $\bar{\boldsymbol{Z}}_{ij} = \left(\bar{\boldsymbol{Z}}, \bar{\boldsymbol{Z}}_{i\circ}, \bar{\boldsymbol{Z}}_{\circ j}\right)$ and $\bar{\boldsymbol{H}}_{ij} = \left(\boldsymbol{D}, \bar{\boldsymbol{Z}}_{ij}\right)$, where

$$\underset{T\times(m_x+1)}{\bar{\boldsymbol{Z}}} = \begin{bmatrix} \bar{\boldsymbol{z}}_1' \\ \vdots \\ \bar{\boldsymbol{z}}_T' \end{bmatrix},\ \underset{T\times(m_x+1)}{\bar{\boldsymbol{Z}}_{i\circ}} = \begin{bmatrix} \bar{\boldsymbol{z}}_{i\circ 1}' \\ \vdots \\ \bar{\boldsymbol{z}}_{i\circ T}' \end{bmatrix},\ \underset{T\times(m_x+1)}{\bar{\boldsymbol{Z}}_{i\circ}} = \begin{bmatrix} \bar{\boldsymbol{z}}_{\circ j1}' \\ \vdots \\ \bar{\boldsymbol{z}}_{\circ j1}' \end{bmatrix},$$

(i) Identification of $\boldsymbol{\beta}_{ij}$: $\bar{\boldsymbol{\Psi}}_{ij,T} = T^{-1}\left(\boldsymbol{X}_{ij}'\bar{\boldsymbol{M}}_{ij}\boldsymbol{X}_{ij}\right)$ are nonsingular, and $\bar{\boldsymbol{\Psi}}_{ij,T}^{-1}$ have finite second-order moments, where

$$\bar{\boldsymbol{M}}_{ij} = \boldsymbol{I}_T - \bar{\boldsymbol{H}}_{ij}\left(\bar{\boldsymbol{H}}_{ij}'\bar{\boldsymbol{H}}_{ij}\right)^{-1}\bar{\boldsymbol{H}}_{ij}' \tag{221}$$

(ii) Identification of $\boldsymbol{\beta}$: $\bar{\boldsymbol{\Psi}} = N^{-2}\sum^{N}\sum^{N}\bar{\boldsymbol{\Psi}}_{ij,T}$ is nonsingular

Remark 1: It is challenging to develop an appropriate model for accommodating CSD within the multilevel dataset. LeSage and Llano (2015) propose a spatial econometric methodology that introduces spatially-structured origin and destination effects. Choi et al. (2016) develop a multilevel factor model with global and country factors, and propose a sequential principal component estimation procedure. KMSS control CSD in 3D panels by adding unobserved heterogeneous global factors to the CTFE specification, and propose the 3D PCCE estimator. The hierarchical multi-factor error model is more parsimonious and structural.

Remark 2: The weights are not necessarily unique. One could use the equal weight, $1/N$ for reasonably large N.

Remark 3: The number of observed factors and the number of individual-specific regressors are fixed and known.

Represent hierarchical cross-section averages as follows:

$$\bar{\boldsymbol{z}}_t = \bar{\boldsymbol{\Xi}}_{\circ\circ}\boldsymbol{d}_t + \bar{\boldsymbol{\Phi}}_{\circ\circ}\boldsymbol{f}_t + \bar{\boldsymbol{\Phi}}_{\circ\bullet}\boldsymbol{f}_{\bullet\circ t} + \bar{\boldsymbol{\Phi}}_{\bullet\circ}\boldsymbol{f}_{\circ\bullet t} + \bar{\boldsymbol{u}}_t, \tag{222}$$

$$\bar{\boldsymbol{z}}_{i\circ t} = \bar{\boldsymbol{\Xi}}_{i\circ}\boldsymbol{d}_t + \bar{\boldsymbol{\Phi}}_{i\circ}\boldsymbol{f}_t + \bar{\boldsymbol{\Phi}}_{\circ\bullet}\boldsymbol{f}_{i\circ t} + \boldsymbol{\Phi}_{i\circ}\boldsymbol{f}_{\circ\bullet t} + \bar{\boldsymbol{u}}_{i\circ t}, \tag{223}$$

$$\bar{\boldsymbol{z}}_{\circ jt} = \bar{\boldsymbol{\Xi}}_{\circ j}\boldsymbol{d}_t + \bar{\boldsymbol{\Phi}}_{\circ j}\boldsymbol{f}_t + \boldsymbol{\Phi}_{\circ j}\boldsymbol{f}_{\bullet\circ t} + \bar{\boldsymbol{\Phi}}_{\bullet\circ}\boldsymbol{f}_{\circ jt} + \bar{\boldsymbol{u}}_{\circ jt}, \tag{224}$$

$$\bar{\boldsymbol{\Xi}}_{\circ\circ} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Xi}_{ij}, \quad \bar{\boldsymbol{\Xi}}_{i\circ} = \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\Xi}_{ij}, \quad \bar{\boldsymbol{\Xi}}_{\circ j} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Xi}_{ij},$$

$$\bar{\boldsymbol{\Phi}}_{\circ\circ} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Phi}_{ij}, \quad \bar{\boldsymbol{\Phi}}_{i\circ} = \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\Phi}_{ij}, \quad \bar{\boldsymbol{\Phi}}_{\circ j} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_{ij}, \tag{225}$$

$$\bar{\boldsymbol{\Phi}}_{\circ\bullet} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Phi}_{\circ j} = \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\Phi}_{\circ j}, \quad \bar{\boldsymbol{\Phi}}_{\bullet\circ} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Phi}_{i\circ} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi} \tag{226}$$

$$\bar{\boldsymbol{u}}_t = \frac{1}{N^2}\sum^{N}\sum^{N}\boldsymbol{u}_{ijt}, \quad \bar{\boldsymbol{u}}_{i\circ t} = \frac{1}{N}\sum^{N}\boldsymbol{u}_{ijt}, \quad \bar{\boldsymbol{u}}_{\circ jt} = \frac{1}{N}\sum^{N}\boldsymbol{u}_{ijt}$$

Combining (222)-(224), we have:

$$\bar{\boldsymbol{z}}_{ijt} = \bar{\boldsymbol{\Xi}}_{ij} \boldsymbol{d}_t + \bar{\boldsymbol{\Phi}}_{ij} \boldsymbol{f}_{ijt} + \bar{\boldsymbol{u}}_{ijt}, \tag{227}$$

where

$$\underset{3(m_x+1)\times 1}{\bar{\boldsymbol{z}}_{ijt}} = \begin{bmatrix} \bar{\boldsymbol{z}}_t \\ \bar{\boldsymbol{z}}_{i\circ t} \\ \bar{\boldsymbol{z}}_{\circ jt} \end{bmatrix}, \underset{m\times 1}{\boldsymbol{f}_{ijt}} = \begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{i\circ t} \\ \boldsymbol{f}_{\circ jt} \end{bmatrix}, \bar{\boldsymbol{u}}_{ijt} = \begin{bmatrix} \bar{\boldsymbol{u}}_t + \bar{\boldsymbol{\Phi}}_{\circ\bullet} \boldsymbol{f}_{\bullet\circ t} + \bar{\boldsymbol{\Phi}}_{\bullet\circ} \boldsymbol{f} \\ \bar{\boldsymbol{u}}_{i\circ t} + \boldsymbol{\Phi}_{i\circ} \boldsymbol{f}_{\circ\bullet t} \\ \bar{\boldsymbol{u}}_{\circ jt} + \boldsymbol{\Phi}_{\circ j} \boldsymbol{f}_{\bullet\circ t} \end{bmatrix}$$

$$\underset{3(m_x+1)\times m_d}{\bar{\boldsymbol{\Xi}}_{ij}} = \begin{bmatrix} \bar{\boldsymbol{\Xi}}_{\circ\circ} \\ \bar{\boldsymbol{\Xi}}_{i\circ} \\ \bar{\boldsymbol{\Xi}}_{\circ j} \end{bmatrix}, \quad \underset{3(m_x+1)\times m}{\bar{\boldsymbol{\Phi}}_{ij}} = \begin{bmatrix} \bar{\boldsymbol{\Phi}}_{\circ\circ} & \mathbf{0} & \mathbf{0} \\ \bar{\boldsymbol{\Phi}}_{i\circ} & \bar{\boldsymbol{\Phi}}_{\circ\bullet} & \mathbf{0} \\ \bar{\boldsymbol{\Phi}}_{\circ j} & \mathbf{0} & \bar{\boldsymbol{\Phi}}_{\bullet\circ} \end{bmatrix}, \tag{228}$$

with $m = m_f + m_{\circ\bullet} + m_{\bullet\circ}$.

Using (213) and (219), $\bar{\boldsymbol{\Phi}}_{ij}$ can be represented as follows:

$$\underset{(k+1)\times m_f}{\bar{\boldsymbol{\Phi}}_{\circ\circ}} = \tilde{\boldsymbol{B}}\tilde{\boldsymbol{\Gamma}}_{\circ\circ} + \begin{pmatrix} \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\boldsymbol{\nu}_{i\circ}+\boldsymbol{\nu}_{\circ j}+\boldsymbol{\nu}_{ij}\right)'\boldsymbol{\Gamma}_{ij} \\ \mathbf{0} \end{pmatrix} \tag{229}$$

$$\underset{(k+1)\times m_{\circ\bullet}}{\bar{\boldsymbol{\Phi}}_{\circ\bullet}} = \tilde{\boldsymbol{B}}\tilde{\boldsymbol{\Gamma}}_{\circ\bullet} + \begin{pmatrix} \frac{1}{N}\sum_{j=1}^{N}\left(\boldsymbol{\nu}_{i\circ}+\boldsymbol{\nu}_{\circ j}+\boldsymbol{\nu}_{ij}\right)'\boldsymbol{\Gamma}_{\circ j} \\ \mathbf{0} \end{pmatrix} \tag{230}$$

$$\underset{(k+1)\times m_{\bullet\circ}}{\bar{\boldsymbol{\Phi}}_{\bullet\circ}} = \tilde{\boldsymbol{B}}\tilde{\boldsymbol{\Gamma}}_{\bullet\circ} + \begin{pmatrix} \frac{1}{N}\sum_{=1}^{N}\left(\boldsymbol{\nu}_{i\circ}+\boldsymbol{\nu}_{\circ j}+\boldsymbol{\nu}_{ij}\right)'\boldsymbol{\Gamma}_{i\circ} \\ \mathbf{0} \end{pmatrix} \tag{231}$$

$$\underset{(k+1)\times m_f}{\bar{\boldsymbol{\Phi}}_{i\circ}} = \tilde{\boldsymbol{B}}\tilde{\boldsymbol{\Gamma}}_{i\circ} + \begin{pmatrix} \frac{1}{N}\sum_{j=1}^{N}\left(\boldsymbol{\nu}_{i\circ}+\boldsymbol{\nu}_{\circ j}+\boldsymbol{\nu}_{ij}\right)'\boldsymbol{\Gamma}_{ij} \\ \mathbf{0} \end{pmatrix} \tag{232}$$

$$\underset{(k+1)\times m_f}{\bar{\boldsymbol{\Phi}}_{\circ j}} = \tilde{\boldsymbol{B}}\tilde{\boldsymbol{\Gamma}}_{\circ j} + \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\nu}_{i\circ}+\boldsymbol{\nu}_{\circ j}+\boldsymbol{\nu}_{ij}\right)'\boldsymbol{\Gamma}_{ij} \\ \mathbf{0} \end{pmatrix} \tag{233}$$

$$\tilde{B} = \begin{pmatrix} 1 & \boldsymbol{\beta}' \\ 0 & \boldsymbol{I}_k \end{pmatrix}, \tilde{\Gamma}_{\circ\circ} = \begin{pmatrix} \bar{\boldsymbol{\gamma}}'_{\circ\circ} \\ \bar{\boldsymbol{\Gamma}}_{\circ\circ} \end{pmatrix}, \tilde{\Gamma}_{\circ\bullet} = \begin{pmatrix} \bar{\boldsymbol{\gamma}}'_{\circ\bullet} \\ \bar{\boldsymbol{\Gamma}}_{\circ\bullet} \end{pmatrix}, \tilde{\Gamma}_{\bullet\circ} = \begin{pmatrix} \bar{\boldsymbol{\gamma}}'_{\bullet\circ} \\ \bar{\boldsymbol{\Gamma}}_{\bullet\circ} \end{pmatrix}, \tilde{\Gamma}_{i\circ} =$$

- Suppose that the rank condition holds:

$$Rank\left(\bar{\boldsymbol{\Phi}}_{ij}\right) = m \text{ for all } (ij)\,. \tag{234}$$

- Then, we obtain from (227):

$$\boldsymbol{f}_{ijt} = \left(\bar{\boldsymbol{\Phi}}_{ij}'\bar{\boldsymbol{\Phi}}_{ij}\right)^{-1}\bar{\boldsymbol{\Phi}}_{ij}'\left(\bar{\boldsymbol{z}}_{ijt} - \bar{\boldsymbol{\Xi}}_{ij}\boldsymbol{d}_t - \bar{\boldsymbol{u}}_{ijt}\right) \tag{235}$$

- It is easily seen that

$$\bar{\boldsymbol{u}}_{ijt} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \text{ for each } t\text{, as } N \to \infty.$$

- Therefore,

$$\boldsymbol{f}_{ijt} - \left(\bar{\boldsymbol{\Phi}}_{ij}'\bar{\boldsymbol{\Phi}}_{ij}\right)^{-1}\bar{\boldsymbol{\Phi}}_{ij}'\left(\bar{\boldsymbol{z}}_{ijt} - \bar{\boldsymbol{\Xi}}_{ij}\boldsymbol{d}_t\right) = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

- We can use $\bar{\boldsymbol{h}}_{ijt} = \left(\boldsymbol{d}_t', \bar{\boldsymbol{z}}_{ijt}'\right)'$ as observable proxies for $\boldsymbol{f}_{ijt}$, and consistently estimate $\boldsymbol{\beta}_{ij}$ and their mean $\boldsymbol{\beta}$ by augmenting the regression, (209) with $\boldsymbol{d}_t$ and $\bar{\boldsymbol{z}}_{ijt}$.
- These are referred to as the 3DCCE estimators.

Individual Specific Coefficients

- The 3DCCE estimator of $\boldsymbol{\beta}_{ij}$ is given by

$$\hat{\boldsymbol{b}}_{ij} = \left(\boldsymbol{X}'_{ij}\bar{\boldsymbol{M}}_{ij}\boldsymbol{X}_{ij}\right)^{-1}\boldsymbol{X}'_{ij}\bar{\boldsymbol{M}}_{ij}\boldsymbol{y}_{ij} \qquad (236)$$

- We show the dependence of $\hat{\boldsymbol{b}}_{ij}$ on the unobserved factors as:

$$\hat{\boldsymbol{b}}_{ij} - \boldsymbol{\beta}_{ij} = \left(\frac{\boldsymbol{X}_{ij}\bar{\boldsymbol{M}}_{ij}\boldsymbol{X}_{ij}}{T}\right)^{-1}\frac{\boldsymbol{X}'_{ij}\bar{\boldsymbol{M}}_{ij}\boldsymbol{F}_{ij}}{T}\boldsymbol{\gamma}^{*}_{ij} + \left(\frac{\boldsymbol{X}'_{ij}\bar{\boldsymbol{M}}_{ij}\boldsymbol{X}_{ij}}{T}\right)^{-} \qquad (237)$$

$$= \left(\frac{\boldsymbol{X}_{ij}\boldsymbol{M}_{Q,ij}\boldsymbol{X}_{ij}}{T}\right)^{-1}\frac{\boldsymbol{X}'_{ij}\boldsymbol{M}_{Q,ij}\boldsymbol{F}_{ij}}{T}\boldsymbol{\gamma}^{*}_{ij} + \left(\frac{\boldsymbol{X}'_{ij}\boldsymbol{M}_{Q,ij}\boldsymbol{X}}{T}\right.$$

$$+ O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$$

where $\boldsymbol{F}_{ij} = (\boldsymbol{F}, \boldsymbol{F}_{i\circ}, \boldsymbol{F}_{\circ j})$, $\boldsymbol{\gamma}^{*}_{ij} = \left(\boldsymbol{\gamma}'_{ij}, \boldsymbol{\gamma}'_{i\circ}, \boldsymbol{\gamma}'_{\circ j}\right)'$ and

$\boldsymbol{M}_{Q,ij} = \boldsymbol{I}_T - \boldsymbol{Q}_{ij}\left(\boldsymbol{Q}'_{ij}\boldsymbol{Q}_{ij}\right)^{-1}\boldsymbol{Q}'_{ij}$ with

Suppose that the rank condition (234), is satisfied. Then,

Theorem

Consider the triple-index heterogeneous panel data model, (209)-(211). Suppose that Assumptions 1-4 and 5(a) hold. Then, the 3DCCE estimator of the individual slope coefficients given by (236) is consistent. Further, as $N, T \to \infty$ and $T/N \to K < \infty$,

$$\sqrt{T}\left(\hat{\boldsymbol{b}}_{ij} - \boldsymbol{\beta}_{ij}\right) \to^{d} N(0, \boldsymbol{V}_{ij}), \tag{238}$$

where $\boldsymbol{V}_{ij} = \boldsymbol{\Sigma}_{v,ij}^{-1}\boldsymbol{\Sigma}_{ij\varepsilon}\boldsymbol{\Sigma}_{v,ij}^{-1}$, $\boldsymbol{\Sigma}_{v,ij} = Var(\boldsymbol{v}_{ijt})$,
$\boldsymbol{\Sigma}_{ij\varepsilon} = p\lim_{T\to\infty}\left[\frac{\boldsymbol{X}'_{ij}\boldsymbol{M}_{F,ij}\boldsymbol{\Omega}_{ij\varepsilon}\boldsymbol{M}_{F,ij}\boldsymbol{X}_{ij}}{T}\right]$, *and*
$\boldsymbol{\Omega}_{ij\varepsilon} = E\left(\boldsymbol{\varepsilon}'_{ij}\boldsymbol{\varepsilon}_{ij}\right)$.

- **Remark**: If the rank condition (234) does not hold, we need to show that $\frac{1}{T}\boldsymbol{X}'_{ij}\bar{\boldsymbol{M}}_{ij}\left(\boldsymbol{F}\boldsymbol{\gamma}_{ij}+\boldsymbol{F}_{i\circ}\boldsymbol{\gamma}_{\circ j}+\boldsymbol{F}_{\circ j}\boldsymbol{\gamma}_{i\circ}\right)$ converges to zero.
- We can establish that

$$\hat{\boldsymbol{b}}_{ij}-\boldsymbol{\beta}_{ij}=\left(\frac{\boldsymbol{X}'_{ij}\boldsymbol{M}_{Q,ij}\boldsymbol{X}_{ij}}{T}\right)^{-1}\frac{\boldsymbol{X}'_{ij}\boldsymbol{M}_{Q,ij}\boldsymbol{\varepsilon}_{ij}}{T}+o_p\left(1\right).$$

- $\sqrt{T}\left(\hat{\boldsymbol{b}}_{ij}-\boldsymbol{\beta}_{ij}\right)$ will be asymptotically normal if $\sqrt{T}/N\to 0$ as $N,T\to\infty$.

3D Common Correlated Effects Mean Group Estimator

The 3DCCEMG estimator is an average of the individual $\hat{\boldsymbol{b}}_{ij}$:

$$\hat{\boldsymbol{b}}_{MG} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\boldsymbol{b}}_{ij}, \tag{239}$$

Under Assumption 4 and using (237), we decompose $\sqrt{N}\left(\hat{\boldsymbol{b}}_{MG} - \boldsymbol{\beta}\right)$ and analyse each terms to obtain the following Theorem.

Theorem

Consider the 3D model, (209)-(211). Suppose that Assumptions 1-4 and 5(a) hold. Then, the 3D CCEMG, $\hat{\boldsymbol{b}}_{MG}$ is consistent. As $N, T \to \infty$,

$$\sqrt{N}\left(\overline{\boldsymbol{b}}_{MG} - \boldsymbol{\beta}\right) \to^{d} N(0, \boldsymbol{V}_{MG}), \boldsymbol{V}_{MG} = \boldsymbol{\Omega}_{\boldsymbol{\nu}\bullet\circ} + \boldsymbol{\Omega}_{\boldsymbol{\eta}\bullet\circ} + \boldsymbol{\Omega}_{\boldsymbol{\nu}\circ\bullet} + \boldsymbol{\Omega}_{\boldsymbol{\eta}\circ\bullet} \tag{240}$$

$$\Omega_{\boldsymbol{\nu}\bullet\circ} = \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\boldsymbol{A}_{1,i,NT} \boldsymbol{\Omega}_{\boldsymbol{\nu}_{i\circ}} \boldsymbol{A}'_{1,i,NT}\right),$$

$$\Omega_{\boldsymbol{\eta}\bullet\circ} = \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\boldsymbol{A}_{2,i,NT} \boldsymbol{\Omega}_{\boldsymbol{\eta}_{i\circ}} \boldsymbol{A}'_{2,i,NT}\right)$$

$$\Omega_{\boldsymbol{\nu}\circ\bullet} = \lim_{N,T\to\infty} \frac{1}{N} \sum_{j=1}^{N} E\left(\boldsymbol{A}_{1,j,NT} \boldsymbol{\Omega}_{\boldsymbol{\nu}_{\circ j}} \boldsymbol{A}'_{1,j,NT}\right),$$

$\boldsymbol{V}_{MG}$ can be consistently estimated by

$$\hat{V}_{MG} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right)\left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right)' + \frac{1}{N-1}\sum_{j=1}^{N}\left(\hat{\boldsymbol{b}}_j - \hat{\boldsymbol{b}}_{MG}\right)\left(\hat{\boldsymbol{b}}_j - \hat{\boldsymbol{b}}_{MG}\right)', \tag{241}$$

where $\hat{\boldsymbol{b}}_i = \frac{1}{N}\sum_{j=1}^{N}\hat{\boldsymbol{b}}_{ij}$ and $\hat{\boldsymbol{b}}_j = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_{ij}$.

- The dominant terms of $\sqrt{N}\left(\bar{\boldsymbol{b}}_{MG}-\boldsymbol{\beta}\right)$ are those that involve $\boldsymbol{\nu}_{i\circ}$, $\boldsymbol{\nu}_{i\circ}$, $\boldsymbol{\eta}_{i\circ}$ and $\boldsymbol{\eta}_{\circ j}$ only, because the terms associated with $\boldsymbol{\nu}_{ij}$ and $\boldsymbol{\eta}_{ij}$ are asymptotically negligible. This explains the $N^{1/2}$ rate of convergence.
- The nonparametric variance estimator $\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\hat{\boldsymbol{b}}_{ij}-\hat{\boldsymbol{b}}_{MG}\right)\left(\hat{\boldsymbol{b}}_{ij}-\hat{\boldsymbol{b}}_{MG}\right)'$ used by Pesaran (2006), is not consistent since it gives equal weights to the terms containing $\boldsymbol{\nu}_{i\circ}$, $\boldsymbol{\nu}_{i\circ}$, $\boldsymbol{\eta}_{i\circ}$ and $\boldsymbol{\eta}_{\circ j}$, and those containing $\boldsymbol{\nu}_{ij}$ and $\boldsymbol{\eta}_{ij}$.
- The consistent nonparametric estimator, $\hat{\boldsymbol{V}}_{MG}$ ensures that $\boldsymbol{\nu}_{ij}$ and $\boldsymbol{\eta}_{ij}$ are averaged out by the use of $\hat{\boldsymbol{b}}_i$ and $\hat{\boldsymbol{b}}_j$.
- **Remark** Theorem 2 does not require the rank condition as long as the number of factors m is fixed. We do not require any restriction on the relative rate of N and T.

3D Common Correlated Effects Pooled Estimator

- Consider the special case where $\boldsymbol{\beta}_{ij}$ are homogeneous, where efficiency gains from pooling can be achieved.
- We still allow the coefficients on observed and unobserved common effects to differ across (ij).
- We derive the pooled estimator of $\boldsymbol{\beta}$, referred to as the 3D CCEP estimator by

$$\hat{\boldsymbol{b}}_P = \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{X}'_{ij} \bar{\boldsymbol{M}}_{ij} \boldsymbol{X}_{ij} \right)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{X}'_{ij} \bar{\boldsymbol{M}}_{ij} \boldsymbol{y}_{ij}, \quad (242)$$

Theorem

Consider the 3D model, (209)-(211). Suppose that Assumptions 1-4 and 5(b) hold. Then,

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_P - \boldsymbol{\beta}\right) \rightarrow^d N(0, \boldsymbol{\Psi}^{-1}\boldsymbol{R}\boldsymbol{\Psi}^{-1}) \tag{243}$$

$$\boldsymbol{\Psi} = \lim_{N\to\infty} \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Psi}_{ij} \text{ with } \boldsymbol{\Psi}_{ij} = E\left[\left(\frac{\boldsymbol{X}_{ij}\bar{\boldsymbol{M}}_{ij}\boldsymbol{X}_{ij}}{T}\right)^{-1}\right] \tag{244}$$

$$\boldsymbol{R} = \tilde{\boldsymbol{\Omega}}_{\boldsymbol{\nu}\bullet\circ} + \tilde{\boldsymbol{\Omega}}_{\boldsymbol{\eta}\bullet\circ} + \tilde{\boldsymbol{\Omega}}_{\boldsymbol{\nu}\circ\bullet} + \tilde{\boldsymbol{\Omega}}_{\boldsymbol{\eta}\circ\bullet}. \tag{245}$$

$$\tilde{\Omega}_{\boldsymbol{\nu}\bullet\circ} = \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\tilde{\boldsymbol{A}}_{1,i,NT} \boldsymbol{\Omega}'_{\boldsymbol{\nu}_{i\circ}} \tilde{\boldsymbol{A}}'_{1,i,NT} \right),$$

$$\tilde{\Omega}_{\boldsymbol{\eta}\bullet\circ} = \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\tilde{\boldsymbol{A}}_{2,i,NT} \boldsymbol{\Omega}'_{\boldsymbol{\eta}_{i\circ}} \tilde{\boldsymbol{A}}_{2,i,NT} \right)$$

$$\tilde{\Omega}_{\boldsymbol{\nu}\circ\bullet} = \lim_{N,T\to\infty} \frac{1}{N} \sum_{j=1}^{N} E\left(\tilde{\boldsymbol{A}}_{1,j,NT} \boldsymbol{\Omega}_{\boldsymbol{\nu}_{\circ j}} \boldsymbol{A}'_{1,j,NT} \right),$$

$$\tilde{\Omega}_{\boldsymbol{\eta}\circ\bullet} = \lim_{N,T\to\infty} \frac{1}{N} \sum_{j=1}^{N} E\left(\tilde{\boldsymbol{A}}_{2,j,NT} \boldsymbol{\Omega}_{\boldsymbol{\eta}_{\circ j}} \tilde{\boldsymbol{A}}'_{2,j,NT} \right)$$

where $\tilde{\boldsymbol{A}}_{1,i,NT}$, $\tilde{\boldsymbol{A}}_{2,i,NT}$, $\tilde{\boldsymbol{A}}_{1,j,NT}$ and $\tilde{\boldsymbol{A}}_{2,j,NT}$ are defined in Appendix.

The variance $\boldsymbol{\Psi}^{-1}\boldsymbol{R}\boldsymbol{\Psi}^{-1}$ can be consistently estimated by $\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{R}}\hat{\boldsymbol{\Psi}}^{-1}$:

$$\hat{\boldsymbol{\Psi}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\boldsymbol{X}'_{ij}\bar{\boldsymbol{M}}_{ij}\boldsymbol{X}_{ij}}{T}, \tag{246}$$

$$\hat{R}=\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\boldsymbol{X}'_{ij}\bar{\boldsymbol{M}}_{ij}\boldsymbol{X}_{ij}}{T}\right)\left(\hat{\boldsymbol{b}}_{ij}-\hat{\boldsymbol{b}}_{MG}\right)\right]\left[\frac{1}{N}\sum_{j=1}^{N}\left(\hat{\boldsymbol{b}}_{ij}-\hat{\boldsymbol{b}}_{MG}\right)\right.$$
$$+\frac{1}{N}\sum_{j=1}^{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}'_{ij}\bar{\boldsymbol{M}}_{ij}\boldsymbol{X}_{ij}}{T}\right)\left(\hat{\boldsymbol{b}}_{ij}-\hat{\boldsymbol{b}}_{MG}\right)\right]\left[\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{b}}_{ij}-\hat{\boldsymbol{b}}_{MG}\right)\right.$$

Remark The asymptotic variance matrix of $\hat{\boldsymbol{b}}_P$ depends on unobserved factors and loadings, but it is possible to estimate it consistently along lines similar to 3DCCEMG.

The Special Cases

- Better convergence rates can be achieved if the hierarchical structure is simplified.
- We focus on two special cases:

$$\text{Condition S1} : \boldsymbol{\eta}_{i\circ} = \boldsymbol{\eta}_{\circ j} = \boldsymbol{\nu}_{i\circ} = \boldsymbol{\nu}_{\circ j} = \mathbf{0}$$

$$\text{Condition S2} : \boldsymbol{F}_{i\circ} = \boldsymbol{F}_{\circ j} = 0.$$

- S2 is more restrictive and considered by KMSS.
- Under Condition S2, the setup is similar to that of Pesaran (2006) because there is no hierarchical factor structure.
- We can treat the dataset as a $T \times N^2$ panel by amalgamating the two cross-section dimensions into one and applying the 2D CCE estimation procedure.
- The $\sqrt{N}$ rate will be replaced by N, and all the results of Pesaran (2006) and others analysing CCE estimator hold.

- Next, consider the case where S1 holds but not S2. Then,

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_{MG}-\boldsymbol{\beta}\right)=\frac{1}{N^{3/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\nu}_{ij}+\frac{1}{N^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Psi}_{ijT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}'_{ij}\boldsymbol{M}}{T}\right. \tag{247}$$

$$+\frac{1}{N^{3/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\chi_{ij}+\frac{1}{N^{3/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\chi_{ij,\circ\circ}+\frac{1}{N^{3/2}}\sum_{i=}$$

$$+\frac{1}{N^{3/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\chi}_{ij,\circ\bullet}+O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right.$$

- From the proof of Theorem 2, the magnitude of all terms on the RHS of (247) is still N as long as $N/T\rightarrow 0$, since $\frac{1}{\sqrt{NT}}=o\left(\frac{1}{N}\right)$.
- Normality does not follow since the $O_p\left(\frac{1}{N}\right)$ term in RHS of (247) is not negligible.

- In this case, the asymptotic variance estimators in Pesaran (2006) become relevant only if normality holds. In particular,

$$\hat{\boldsymbol{V}}_{MG} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\hat{\boldsymbol{b}}_{ij} - \hat{\boldsymbol{b}}_{MG}\right)\left(\hat{\boldsymbol{b}}_{ij} - \hat{\boldsymbol{b}}_{MG}\right)', \qquad (248)$$

for the mean group estimator, and for the pooled estimator

$$\hat{\boldsymbol{R}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{\boldsymbol{X}_{ij}'\bar{\boldsymbol{M}}_{ij}\boldsymbol{X}_{ij}}{T}\right)\left(\bar{\boldsymbol{b}}_{ij} - \bar{\boldsymbol{b}}_{MG}\right)\left(\bar{\boldsymbol{b}}_{ij} - \bar{\boldsymbol{b}}_{MG}\right)'\left(\frac{\boldsymbol{X}_{ij}'\boldsymbol{M}}{}\right. \qquad (249)$$

- If Condition S1 is considered too restrictive, we may entertain the more general setup:

$$\boldsymbol{\gamma}_{i\circ} = \boldsymbol{\gamma}_{ij\circ} = \boldsymbol{\gamma}_{\bullet\circ} + \boldsymbol{\eta}_{ij\circ}, \quad \boldsymbol{\gamma}_{\circ j} = \boldsymbol{\gamma}_{\circ ij} = \boldsymbol{\gamma}_{\circ\bullet} + \boldsymbol{\eta}_{\circ ij}.$$

- Because of the double cross-section averaging, $\boldsymbol{\eta}_{ij\circ}$ and $\boldsymbol{\eta}_{\circ ij}$ are negligible since terms associated with $\boldsymbol{\chi}_{ij,\bullet\circ}$ and $\boldsymbol{\chi}_{ij,\circ\bullet}$ decay to give the same fast convergence rate as under S1.

Monte Carlo Study

- We generate y_{ijt} and x_{ijt} as follows:

$$y_{ijt} = \beta_{ij}x_{ijt} + \gamma_{1,ij}f_{1,t} + \gamma_{2,ij}f_{2,t} + \gamma_{1,\circ j}f_{1,i\circ t} + \gamma_{2,\circ j}f_{2,i\circ t} + \gamma_{1,i\circ}f_{1,\circ jt} + \tag{250}$$

$$x_{ijt} = \Gamma_{1,ij}f_{1,t} + \Gamma_{2,ij}f_{2,t} + \Gamma_{1,\circ j}f_{1,i\circ t} + \Gamma_{2,\circ j}f_{2,i\circ t} + \Gamma_{1,i\circ}f_{1,\circ jt} + \Gamma_{2,i\circ}f_{2,} \tag{251}$$

- We set $m_d = 0$, $m_x = 1$ and $m_f = m_{\circ\bullet} = m_{\bullet\circ} = 2$.
- $\boldsymbol{f}_t$, $\boldsymbol{f}_{\circ jt}$, $\boldsymbol{f}_{i\circ t}$ are generated independently as stationary AR processes with zero mean and unit variance:

$$f_{h,t} = \rho_{f_h}f_{h,t-1} + \upsilon_{f_h t} \text{ with } \upsilon_{f_h t} \sim iidN\left(0, 1-\rho_{f_h}^2\right) \text{ for } h = 1,2$$

$$f_{h,i\circ t} = \rho_{f_{h,i\circ}}f_{h,i\circ,t-1} + \upsilon_{f_{h,i\circ}t} \text{ with } \upsilon_{f_{h,i\circ}t} \sim iidN\left(0, 1-\rho_{f_{h,i\circ}}\right) \text{ for}$$

$$f_{h,\circ jt} = \rho_{f_{h,\circ j}}f_{h,\circ j,t-1} + \upsilon_{f_{h,\circ j}t} \text{ with } \upsilon_{f_{h,\circ j}t} \sim iidN\left(0, 1-\rho_{f_{h,\circ j}}^2\right) \text{ fo}$$

- ε_{ijt} and v_{ijt}, are generated independently as

$$\varepsilon_{ijt} = \rho_{\varepsilon}\varepsilon_{ij,t-1} + e_{\varepsilon,ijt} \text{ with } e_{\varepsilon,ijt} \sim iidN\left(0, 1-\rho_{\varepsilon}^2\right)$$

- 2 experiments: Experiment A with the full rank and Experiment B with the rank condition (234) violated.
- For x_{ijt} in (251), we draw the factor loadings independently by

$$\Gamma_{1,ij} \sim iidN\left(0.5, 0.5\right) \text{ and } \Gamma_{2,ij} \sim iidN\left(0, 0.5\right) \text{ for } i, j = 1, ..., N$$
$$\Gamma_{1,\circ j} \sim iidN\left(0.5, 0.5\right) \text{ and } \Gamma_{2,\circ j} \sim iidN\left(0, 0.5\right) \text{ for } j = 1, ..., N$$
$$\Gamma_{1,i\circ} \sim iidN\left(0.5, 0.5\right) \text{ and } \Gamma_{2,i\circ} \sim iidN\left(0, 0.5\right) \text{ for } i = 1, ..., N$$

- For y_{ijt} in (250), we consider two experiments. For experiment A,

$$\gamma_{1,ij} \sim iidN\left(1, 0.2\right) \text{ and } \gamma_{2,ij} \sim iidN\left(1, 0.2\right) \text{ for } i, j = 1, ..., N$$
$$\gamma_{1,\circ j} \sim iidN\left(1, 0.2\right) \text{ and } \gamma_{2,\circ j} \sim iidN\left(1, 0.2\right) \text{ for } j = 1, ..., N$$
$$\gamma_{1,i\circ} \sim iidN\left(1, 0.2\right) \text{ and } \gamma_{2,i\circ} \sim iidN\left(1, 0.2\right) \text{ for } i = 1, ..., N.$$

- For experiment B, we generate:

$$\gamma_{1,ij} \sim iidN\left(1, 0.2\right) \text{ and } \gamma_{2,ij} \sim iidN\left(0, 1\right) \text{ for } i, j = 1, ..., N$$
$$\gamma_{1,\circ j} \sim iidN\left(1, 0.2\right) \text{ and } \gamma_{2,\circ j} \sim iidN\left(0, 1\right) \text{ for } j = 1, ..., N$$

- Consider Case 1 with the heterogeneous slopes:

  $\beta_{ij} = \beta + \nu_{i\circ} + \nu_{\circ j} + \nu_{ij}, \nu_{i\circ} \sim iidN(0,1), \nu_{\circ j} \sim iidN(0,1), \nu_{ij} \sim ii$

  and Case 2 with the homogeneous slopes $\beta_{ij} = \beta = 1$.
- We consider the two-way within estimator with $u_{ijt} = \alpha_{ij} + \theta_t + \varepsilon_{ijt}$, and the three versions of 3D estimators: the $3DCCE_G$ with $u_{ijt} = \alpha_{ij} + \boldsymbol{\gamma}'_{ij}\boldsymbol{f}_t + \varepsilon_{ijt}$ where we approximate the heterogenous global factors only by $\bar{\boldsymbol{z}}_t = (\bar{y}_t, \bar{x}_t)'$; the $3DCCE_L$ estimator with $u_{ijt} = \alpha_{ij} + \boldsymbol{\gamma}'_{\circ j}\boldsymbol{f}_{i\circ t} + \boldsymbol{\gamma}'_{i\circ}\boldsymbol{f}_{\circ jt} + \varepsilon_{ijt}$ where we approximate the heterogenous local factors only by $\bar{\boldsymbol{z}}_{i\circ t}$ and $\bar{\boldsymbol{z}}_{\circ jt}$ and the $3DCCE_{GL}$ estimator with $u_{ijt} = \alpha_{ij} + \boldsymbol{\gamma}'_{\circ j}\boldsymbol{f}_{i\circ t} + \boldsymbol{\gamma}'_{i\circ}\boldsymbol{f}_{\circ jt} + \boldsymbol{\gamma}'_{ij}\boldsymbol{f}_t + \varepsilon_{ijt}$ where we approximate the heterogenous global and local factors by $\bar{\boldsymbol{z}}_t$, $\bar{\boldsymbol{z}}_{i\circ t}$ and $\bar{\boldsymbol{z}}_{\circ jt}$.
- We report the bias, the root mean squared error and coverage rates at the 95% confidence with 1,000 replications for (N, T) pairs with $N = \{10, 25, 100\}$ and $T = \{50, 100\}$

- Table 1: simulation results for Experiment A (the full rank) with heterogeneous coefficients (Case 1).
- The biases of $3DCCE_{GL}$ are mostly negligible even for the relatively small samples.
- The performance of both pooled and mean group estimators is almost identical.
- Both FE and $3DCCE_G$ estimators suffer from severe biases.
- The biases of the $3DCCE_L$ are much smaller than those of the $3DCCE_G$, showing that the local factor approximations seem to be more effective than the global counterpart, though they are still non-negligible even for large N and T.
- The 2D CCE estimator advanced by Pesaran (2006) fails to remove correlations between local factors and regressors.
- These provide strong support for our theoretical predictions that the joint approximations of the heterogenous global and local factors can only provide consistent estimation of $E(\beta)$ in the presence of the hierarchical multifactors.

- We find the similar patterns of RMSE.
- The RMSEs of $3DCCE_L$ and $3DCCE_{GL}$ estimators are significantly lower than those of FE and $3DCCE_G$ estimators.
- The difference between $3DCCE_L$ and $3DCCE_{GL}$ is mostly negligible, but the RMSEs of $3DCCE_{GL}$ tends to decline slightly faster with sample sizes.
- Turning to the coverage rates, $3DCCE_L$ and the $3DCCE_{GL}$ estimators perform better than FE and $3DCCE_G$ estimators.
- Coverage rates of $3DCCE_{GL}$ estimator only tend to the nominal 95% as N or T rises.

- Table 2 presents simulation results for Experiment A (the full rank) with homogeneous coefficients (Case 2).
- We find qualitatively similar results for the biases to Table 1, confirming that the $3DCCE_{GL}$ estimator is most reliable.
- RMSEs of FE and $3DCCE_{G}$ estimators are significantly higher than those of $3DCCE_{L}$ and $3DCCE_{GL}$ estimators.
- RMSEs of $3DCCE_{GL}$ is significantly lower than those of $3DCCE_{L}$, but they also fall sharply with sample sizes.
- The relative performance of both pooled and mean group estimators is qualitatively similar.
- Surprisingly, all the estimators produce unsatisfactory coverage rates. FE, $3DCCE_{G}$ and $3DCCE_{L}$ estimators tend to under-estimate coverage substantially even as N rises whilst $3DCCE_{GL}$ over-estimate it.

- Table 3 presents simulation results for Experiment B (the rank deficiency) with heterogeneous coefficients (Case 1).
- We find qualitatively similar results to Table 1; the performance of the estimators are not affected significantly by the rank deficiency; confirming that the $3DCCE_{GL}$ estimator is most reliable.
- Table 4 presents simulation results for Experiment B (the rank deficiency) with homogeneous coefficients (Case 2).
- We find qualitatively similar results to Table 2, and conclude that the $3DCCE_{GL}$ estimator is still most reliable, though it tends to over-estimate coverage rates.
- We conduct the additional simulations under Conditions S1 and S2 described in Section 3.4
- We find the results confirming that the faster convergence rates are achieved in both cases (available on online supplement).

Table: Simulation results for Case 1 - Full Rank (Experiment A)

		FE		$3DCCE_G$		$3DCCE_L$		$3DCCE_{LG}$	
		Panel A: Bias							
		Pooled Estimator							
ρ	N/T	50	100	50	100	50	100	50	100
	10	0.203	0.216	0.232	0.251	0.064	0.079	-0.006	0.008
0	25	0.250	0.217	0.263	0.272	0.062	0.07	-0.004	0.004
	50	0.227	0.220	0.285	0.279	0.077	0.072	0.008	0.002
	100	0.222	0.224	0.282	0.285	0.074	0.076	0.002	0.003
	10	0.233	0.210	0.302	0.253	0.115	0.067	0.04	-0.008
0.5	25	0.220	0.236	0.265	0.289	0.049	0.074	-0.022	0.004
	50	0.251	0.251	0.301	0.29	0.079	0.068	0.005	-0.006
	100	0.244	0.245	0.300	0.300	0.077	0.078	-0.002	-0.001
		Mean Group Estimator							
	10	0.201	0.212	0.219	0.239	0.046	0.063	0.001	0.017
0	25	0.242	0.211	0.249	0.258	0.053	0.061	0.003	0.012
	50	0.219	0.213	0.268	0.264	0.067	0.062	0.011	0.007
	100	0.214	0.216	0.265	0.268	0.062	0.065	0.002	0.005
	10	0.227	0.205	0.292	0.247	0.103	0.056	0.053	0.006
0.5	25	0.214	0.230	0.254	0.276	0.041	0.063	-0.012	0.010
	50	0.241	0.241	0.287	0.276	0.068	0.056	0.009	-0.003
	100	0.236	0.237	0.286	0.287	0.064	0.065	-0.001	0.000
		Panel B: RMSE							
		Pooled Estimator							
ρ	N/T	50	100	50	100	50	100	50	100
	10	0.522	0.524	0.544	0.53	0.488	0.474	0.483	0.468

Notes:FE is the the two-way within estimator, $3DCCE_G$ is the 3D CCE estimator with the global factors approximation only, $3DCCE_L$ is the 3D CCE estimator with the local factors approximation only, and $3DCCE_{GL}$ is the 3D CCE estimator with both global and local factors approximation. $3DCCE$ estimators are defined in (239) and (242). We consider both mean group and pooled estimators. The variance of $3DCCE_G$ is estimated by (248) for the mean group and (249) for the pooled estimator. The variances of $3DCCE_L$ and $3DCCE_{GL}$ are given by (241) for the mean group and (246)-(**??**) for the pooled estimator.

Table: Simulation Results for Experiment A (Full Rank) with Homogeneous Coefficients

		FE		$3DCCE_G$		$3DCCE_L$		$3DCCE_{LG}$	
		Panel A: Bias							
		Pooled Estimator							
ρ	N/T	50	100	50	100	50	100	50	100
0	10	0.204	0.199	0.240	0.246	0.067	0.073	-0.001	0.006
	25	0.214	0.215	0.269	0.270	0.068	0.068	0.001	0.002
	50	0.218	0.219	0.278	0.278	0.071	0.070	0.001	0.001
	100	0.220	0.220	0.281	0.282	0.074	0.074	0.000	0.000
0.5	10	0.228	0.225	0.267	0.261	0.076	0.072	0.002	-0.003
	25	0.240	0.245	0.292	0.292	0.075	0.075	0.003	0.004
	50	0.245	0.245	0.300	0.299	0.077	0.076	0.002	0.002
	100	0.248	0.248	0.304	0.303	0.081	0.080	0.001	0.000
		Mean Group Estimator							
0	10	0.202	0.198	0.230	0.235	0.057	0.062	0.005	0.011
	25	0.207	0.208	0.255	0.256	0.059	0.060	0.006	0.006
	50	0.211	0.212	0.262	0.262	0.060	0.060	0.002	0.003
	100	0.212	0.212	0.265	0.265	0.061	0.062	0.000	0.001
0.5	10	0.224	0.223	0.256	0.252	0.067	0.064	0.010	0.006
	25	0.233	0.236	0.279	0.278	0.066	0.065	0.008	0.008
	50	0.236	0.237	0.286	0.286	0.066	0.066	0.004	0.004
	100	0.239	0.239	0.289	0.289	0.068	0.067	0.002	0.001
		Panel B: RMSE							
		Pooled Estimator							
ρ	N/T	50	100	50	100	50	100	50	100
	10	0.236	0.230	0.268	0.273	0.129	0.136	0.110	0.112

Table: Simulation results for Case 1 - Rank Deficiency- (Experiment A)

		FE		$3DCCE_G$		$3DCCE_L$		$3DCCE_{LG}$	
		Panel A: Bias							
		Pooled Estimator							
ρ	N/T	50	100	50	100	50	100	50	100
0	10	0.227	0.213	0.220	0.242	0.077	0.099	-0.005	0.018
	25	0.214	0.222	0.223	0.239	0.050	0.066	-0.017	-0.001
	50	0.217	0.208	0.248	0.246	0.061	0.061	0.002	0.001
	100	0.232	0.222	0.260	0.249	0.067	0.057	0.012	0.001
0.5	10	0.227	0.208	0.242	0.249	0.076	0.084	-0.004	0.003
	25	0.246	0.227	0.273	0.274	0.076	0.079	0.009	0.012
	50	0.247	0.241	0.258	0.266	0.049	0.058	-0.012	-0.003
	100	0.246	0.245	0.272	0.271	0.056	0.055	0.000	-0.002
		Mean Group Estimator							
0	10	0.223	0.210	0.211	0.234	0.040	0.062	0.008	0.030
	25	0.204	0.217	0.214	0.228	0.028	0.043	-0.006	0.009
	50	0.209	0.202	0.237	0.235	0.047	0.046	0.009	0.008
	100	0.224	0.214	0.248	0.238	0.057	0.046	0.010	0.005
0.5	10	0.227	0.207	0.230	0.241	0.041	0.054	0.006	0.019
	25	0.241	0.220	0.263	0.263	0.059	0.060	0.023	0.024
	50	0.239	0.234	0.247	0.254	0.038	0.045	-0.003	0.005
	100	0.237	0.236	0.260	0.259	0.047	0.046	0.003	0.002
		Panel B: RMSE							
		Pooled Estimator							
ρ	N/T	50	100	50	100	50	100	50	100
	10	0.497	0.498	0.509	0.520	0.467	0.469	0.461	0.460

Table: Simulation results for Case 2 - Rank Deficiency - (Experiment B)

		FE		$3DCCE_G$		$3DCCE_L$		$3DCCE_{LG}$	
		Panel A: Bias							
		Pooled Estimator							
ρ	N/T	50	100	50	100	50	100	50	100
0	10	0.209	0.205	0.231	0.223	0.076	0.072	0.003	-0.001
	25	0.215	0.214	0.243	0.243	0.063	0.063	0.002	0.002
	50	0.219	0.219	0.247	0.248	0.056	0.058	0.000	0.001
	100	0.220	0.220	0.250	0.251	0.054	0.054	0.000	0.001
0.5	10	0.227	0.230	0.250	0.250	0.078	0.078	0.004	0.004
	25	0.241	0.242	0.270	0.266	0.067	0.064	0.004	0.001
	50	0.246	0.247	0.272	0.271	0.059	0.059	0.001	0.001
	100	0.249	0.248	0.276	0.275	0.056	0.056	0.001	0.000
		Mean Group Estimator							
0	10	0.204	0.201	0.222	0.215	0.053	0.047	0.012	0.006
	25	0.209	0.208	0.232	0.234	0.048	0.049	0.008	0.009
	50	0.211	0.212	0.236	0.237	0.046	0.047	0.005	0.005
	100	0.212	0.213	0.238	0.239	0.045	0.046	0.002	0.003
0.5	10	0.224	0.227	0.242	0.246	0.057	0.060	0.013	0.016
	25	0.233	0.234	0.258	0.256	0.054	0.052	0.011	0.010
	50	0.236	0.238	0.261	0.260	0.051	0.050	0.007	0.006
	100	0.239	0.239	0.264	0.264	0.049	0.049	0.004	0.004
		Panel B: RMSE							
		Pooled Estimator							

Empirical Application

- Anderson and van Wincoop (2003) show that bilateral trade depends on the bilateral trade barriers but relative to the product of their Multilateral Resistance Indices, and derive the system of structural gravity equations:

$$X_{ij} = \frac{Y_i Y_j}{Y} \left(\frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma} \tag{252}$$

$$\Pi_i^{1-\sigma} = \sum_j \left(\frac{t_{ij}}{P_j} \right)^{1-\sigma} \frac{Y_j}{Y} \text{ and } P_j^{1-\sigma} = \sum_i \left(\frac{t_{ij}}{\Pi_i} \right)^{1-\sigma} \frac{Y_i}{Y} \tag{253}$$

  where X_{ij} are exports from i to j, Y_i, Y_j and Y are GPD for i (exporter), j (importer) and the world, $t_{ij}\,(>1)$ is one plus the tariff equivalent of trade costs of imports of j from i, $\sigma\,(>1)$ is the elasticity of substitution with CES preference;
- Π_i is ease of access of exporter i, and P_j is the ease of access of importer j.

- Omitting MTR induces potentially severe bias.
- Consider the log-linearised specification of (252):

$$\ln X_{ij} = \beta_0 + \beta_1 \ln Y_i + \beta_2 \ln Y_j + \beta_3 \ln t_{ij} + \beta_4 \ln P_i + \beta_5 \ln P_j + \varepsilon_{ij} \quad (254)$$

  where P_i and P_j are unobservable MTRs, and t_{ij} contain both barriers and incentives to trade between i and j.
- Subsequent research has focused on estimating (254) with replacing unobservable MTRs by N country-specific dummies, μ_i and μ_j.

- We extend (254) into 3D panels:

$$\ln X_{ijt} = \beta_0 + \beta_1 \ln Y_{it} + \beta_2 \ln Y_{jt} + \beta_3 \ln t_{ijt} + \beta_4 \ln P_{it} + \beta_5 \ln P_{jt} + \varepsilon_{ijt} \quad (255)$$

where we should allow MTRs to vary over time.

- Baltagi et al. (2003) propose:

$$u_{ijt} = \alpha_{ij} + \theta_{it} + \theta^*_{jt} + \varepsilon_{ijt}, \quad (256)$$

which contains bilateral pair-fixed effects α_{ij} as well as origin (exporter) and destination (importer) country-time fixed effects (CTFE) θ_{it} and θ^*_{jt}.

- This approach popular in measuring the impacts of MTRs of exporters and importers in the structural gravity studies.

- The main drawback of the CTFE approach lies in the assumption that bilateral trade flows are independent of what happens to the rest of the trading world.
- Recently, KMSS extend the 3D panel data model (209) with the more general error components:

$$u_{ijt} = \alpha_{ij} + \theta_{it} + \theta^*_{jt} + \pi_{ij}\theta_t + \varepsilon_{ijt}, \tag{257}$$

  that attempts to model residual CSD via unobserved heterogeneous global factor θ_t in addition to CTFEs.
- CTFE estimator is biased because it fails to remove heterogenous global factors correlated with covariates.
- KMSS develop the two-step consistent 3D-PCCE estimation procedure by approximating global factors with double cross-section averages of dependent variable and regressors and applying the 3D-within transformation.
- In this paper, we develop the hierarchical multi-factor error components specification, (210), which is more structural and parsimonious.

The data

- We collect the dataset over the period 1970-2013 (44 years), and consider two control groups:
  1. the 210 country-pairs of the EU15 member countries with 11 Euro countries (Austria, Belgium-Luxemburg, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain) and 4 control countries (Denmark, Norway, Sweden, the UK);
  2. the 320 country-pairs among 19 countries with the EU15 countries and 4 non-EU OECD countries (Australia, Canada, Japan and the US).
- We collect the bilateral export flow from IMF. The Data starts from 1970 as the German data are unavailable in the 60s. There are no missing data so we consider the balanced panel.
- Our sample period consists of several important economic integrations, such as the European Monetary System in 1979 and the Single Market in 1993, all of which can be regarded as promoting intra-EU trades.

Empirical specification:

- We consider the 3D panel gravity specification:

$$\ln EXP_{ijt} = \beta_0 + \beta_1 CEE_{ijt} + \beta_2 EMU_{ijt} + \beta_3 SIM_{ijt} + \beta_4 RLF_{ijt} \tag{258}$$

$$+ \beta_6 \ln GDP_{jt} + \beta_7 RER_t + \gamma_1 DIS_{ij} + \gamma_2 BOR_{ij} + \gamma_3 L$$

- the dependent variable, EXP_{ijt} is the export flow from country i to country j at time t;
- CEE and EMU are dummies for European Community membership and for European Monetary Union;
- SIM is the logarithm of an index that captures the relative size of two countries and bounded between zero (absolute divergence) and 0.5 (equal size);
- RLF is the logarithm of the absolute value of the difference between per capita GDPs of trading countries;
- RER represents the logarithm of common real exchange rates;
- GDP_{it} and GDP_{jt} are logged GDPs of exporter and importer;

- We apply the four estimators considered in the MC simulations, namely the two-way within estimator and the three versions of 3D CCEP estimators.
- We also report the CD results applied to the residuals and the estimates of the CSD exponent (α).
- We focus on investigating the impacts of t_{ij} that contain both barriers and incentives to trade; the two dummy variables CEE (equal to one when both countries belong to the European Community) and EMU (equal to one when both trading partners adopt the same currency).
- Both are expected to exert a positive impact on export flows.
- Empirical evidence is mixed, though recent studies by Mastromarco et al. (2015), and Gunnella et al. (2015) that control for strong CSD in 2D panels, find modest effects (7 to 10%) of the euro on intra-EU trade flows.
- KMSS (2017) apply a 3D PCCE estimator; the EMU impact on exports is about 8%.

The Estimation Results for the EU15 Countries

- Table 5 reports the panel gravity estimation results for the 210 country-pairs among the EU15 member countries over the period 1970-2013 (44 years).
- The FE estimator suffers from strong CSD while the $3DCCEP$ estimators display lower degree CSD.
- CD diagnostic test by Pesaran (2015) fails to reject the null of weak CSD for both $3DCCEP_L$ and $3DCCEP_{GL}$. This is also supported by the smaller estimates of α for $3DCCEP_L$ (0.624) and $3DCCEP_{GL}$ (0.609), close to a moderate range of weak CSD.
- We focus on the $3DCCEP_{GL}$ estimation results with the lowest degree of CSD.

Factor approximations

- Theoretically, we should employ the entire set of cross-section averages to approximate heterogeneous global and local factors.
- In practice, this may raise an issue of multicollinearity. Further, to avoid the curse of dimensionality, we search for an optimal subset of cross sectional averages.
- In the aftermath of the global financial crisis, export flows display a negative average growth as shown below:

Export Growth	70/80	80/90	90/00	00/10	10/13
EU15 + 4 OECD	7.06	6.25	4.35	2.16	-0.34
EU15	8.86	7.37	3.92	2.82	-2.05

- Hence, we also add t^2 as an observed factor, which helps to capture the confounding effect of the crisis.

- All the coefficients are significant and their signs are consistent with our *a priori* expectations.
- The effect of the foreign GDP is substantially higher than the home GDP.
- The effects of SIM and RER are positive while a depreciation of the home currency leads to a significant increase in exports.
- SIM boosts real export flows, which suggests that the intra-industry trade is the main part of the trade in the EU.
- Importantly, the impacts of EMU and CEE are significant, but substantially smaller than the potentially biased FE estimates.
- Both Euro and CEE impacts drop sharply from 0.099 and 0.074 to 0.03 and 0.05.
- Other estimators provide rather unreliable results.[3]

[3] For example, the impacts of home GDP on exports is surprisingly larger than the foreign impact while both Euro and CEE impacts seem to be rather high for the FE. The RER coefficient is significantly negative for the CCEP with the global approximation only whereas the CEE impact is insignificant and the Euro impact is almost negligible for the CCEP with the local approximation

Table: Table 5: Estimation Results for 15 EU Countries

	FE	$3DCCEP_G$	$3DCCEP_L$	$3DCCEP_{GL}$
gdph	1.517	0.230	0.023	0.342
	(0.044)	(0.036)	(0.037)	(0.124)
gdpf	0.953	1.478	0.779	1.498
	(0.044)	(0.037)	(0.057)	(0.031)
sim	-0.045	0.639	-0.012	0.197
	(0.060)	(0.069)	(0.056)	(0.075)
rlf	0.030	-0.002	0.002	0.006
	(0.006)	(0.005)	(0.002)	(0.004)
rer	0.012	-0.046	0.016	0.103
	(0.008)	(0.007)	(0.004)	(0.010)
euro	0.099	0.030	0.012	0.030
	(0.016)	(0.003)	(0.003)	(0.003)
cee	0.074	0.066	0.007	0.050
	(0.014)	(0.007)	(0.007)	(0.013)
CD stat	206.6	4.67	2.33	2.72
α	0.91 (0.90-0.93)	0.78 (0.72-0.84)	0.62 (0.59-0.66)	0.61 (0.57-0.65)

Notes: FE is the two-way fixed effect estimator. $3DCCEP_G$ is the CCEP estimator with only the global factors approximated by $\mathbf{f}_t = \left\{\overline{export}_{..t}, \overline{gdp}_{..t}, \overline{sim}_{..t}, \overline{rlf}_{..t}, \overline{cee}_{..t}, t, t^2\right\}$ $3DCCEP_L$ is the CCEP estimator with only the local factors approximated by $\mathbf{f}_t = \mathbf{f}_{iot} = \left\{\overline{y}_{i.t}, \overline{gdp}_{i.t}\right\}$ and $\mathbf{f}_{ojt} = \left\{\overline{sim}_{.jt}, \overline{rlf}_{.jt}\right\}$. $3DCCEP_{GL}$ is the CCEP estimator with both global and local factors approximated by $\mathbf{f}_t = \left\{\overline{\text{export}}_{..t}, \overline{gdp}_{..t}, \overline{sim}_{..t}, \overline{rlf}_{..t}, \overline{cee}_{..t}, t, t^2\right\}$ and $\mathbf{f}_{iot} = \left\{\overline{sim}_{i.t}, \overline{rlf}_{i.t}, \overline{rer}_{i.t}\right\}$. * and ** stand for significance at 5% and 10% level. CD test refers to testing the null hypothesis of residual cross-section independence or weak dependence (Pesaran, 2015). α is the estimate of CSD exponent with 90% confidence bands inside parenthesis.

- The 3DCCEP estimator wipes out the time invariant regressors.
- Following the 2-step approach as in Serlenga and Shin (2007), we can estimate γ by the between estimator:

$$d_{ijt} = \alpha_{ij} + \gamma_1 DIS_{ij} + \gamma_2 BOR_{ij} + \gamma_3 LAN_{ij} + u_{ijt} \quad (259)$$

where $d_{ijt} = y_{ijt} - \hat{\boldsymbol{\beta}}' \mathbf{x}_{ijt}$ with $\hat{\boldsymbol{\beta}}$ being the 3D CCEP estimator.
- We test the validity of the hypothesis: if the Euro had a positive effect on the EU trade by reducing bilateral barriers and eliminating exchange-related uncertainties and transaction costs, this caused a decrease in trade impacts of bilateral barriers (e.g. Cafiso, 2010).
- A declining trend after 1999 will support the hypothesis that the Euro helps to promote more EU integration.
- To this end we estimate (259) by the cross-section regression at each period, and produce time-varying coefficients of γ.

- Figure 1 shows time varying estimates of γ, using the CCEP estimator with both local and global factors approximation.
- The border effect has been declining until the mid 1980's, and quite stable except the slight dip during the global financial crisis albeit statistically insignificant.
- The language effect decreasing until the end of 1980's, reflecting a progressive lessening of restrictions on labor mobility within EU, that encouraged migration and reduced the relative importance of trade costs and cultural difference.
- Since the introduction of the Euro in 1999, both language and border effects became flat, suggesting that the EU integration may reach near-completion stage. This is consistent with the currency union formation hypothesis by Frankel (2005) that countries, which decide to join a currency union, are self-selected on the basis of distinctive features shared by EU members.
- The effect of distance has been on a declining trend from the

The Estimation Results for the EU15 plus 4 OECD Countries

- Table 3 reports the estimation results for an enlarged sample of the 342 country-pairs among the EU15 member countries plus four more countries (Australia, Canada, Japan and the US).
- Again we focus on the estimation results for the $3DCCEP_{GL}$, which shows the lowest degree of CSD.
- The results are qualitatively similar to those in Table 2.
- All the coefficients are significant with expected signs.
- The effect of the foreign GDP is substantially higher than the home GDP effect.
- The effects of SIM is slightly higher (from 0.2 to 0.22) but RLF becomes negligibly negative.
- The impact of RER is stronger (from 0.10 to 0.18), implying a stronger terms of trade effect.

- The impacts of EMU and CEE are significant, though their magnitudes become smaller than those with the EU15 countries, namely from 3% to 1.5% and from 5% to 3%.
- The smaller effects for the enlarged sample might reflect the trade diversion between the Euro and non-Euro area.
- The effects of the EMU on trade will differ with respect to the selected control group and depend on the composition of treatment and control groups (e.g. Baier and Bergstrand, 2009).
- Other estimators provide rather misleading results. In particular, the FE estimation provides an opposite result that both Euro and CEE impacts increase substantially, from 0.099 and 0.074 to 0.258 and 0.161.

Table: Table 6: Estimation Results for 15 EU plus 4 OECD countries

	FE	$3DCCEP_G$	$3DCCEP_L$	$3DCCEP_{GL}$
gdph	1.066	0.531	0.069	0.169
	(0.019)	(0.016)	(0.010)	(0.055)
gdpf	0.904	1.419	1.262	1.417
	(0.020)	(0.020)	(0.015)	(0.017)
sim	0.332	0.109	0.100	0.220
	(0.029)	(0.021)	(0.013)	(0.023)
rlf	0.027	-0.008	0.010	-0.004
	(0.004)	(0.002)	(0.001)	(0.002)
rer	0.058	0.086	0.074	0.179
	(0.008)	(0.004)	(0.002)	(0.005)
euro	0.258	0.012	0.012	0.014
	(0.009)	(0.003)	(0.001)	(0.002)
cee	0.161	0.021	0.007	0.030
	(0.028)	(0.017)	(0.001)	(0.012)
CD stat	243.33	3.272	2.331	3.201
α	0.90 (0.88-0.920)	0.74 (0.69-0.76)	0.65 (0.61-0.69)	0.62 (0.57-0.66

Notes: FE is the two-way fixed effect estimator. $3DCCEP_G$ is the CCEP estimator with only the global factors approximated by $\mathbf{f}_t = \left\{\overline{export}_{..t}, \overline{gdp}_{..t}, \overline{sim}_{..t}, \overline{rlf}_{..t}, \overline{cee}_{..t}, t\right\}$. $3DCCEP_L$ is the CCEP estimator with only the local factors approximated by $\mathbf{f}_t = \mathbf{f}_{iot} = \left\{\overline{y}_{i.t}, \overline{gdp}_{i.t}\right\}$ and $\mathbf{f}_{ojt} = \left\{\overline{sim}_{.jt}, \overline{rlf}_{.jt}\right\}$. $3DCCEP_{GL}$ is the CCEP estimator with both global and local factors approximated by $\mathbf{f}_t = \left\{\overline{export}_{..t}, \overline{gdp}_{..t}, \overline{sim}_{..t}, \overline{rlf}_{..t}, \overline{cee}_{..t}, t\right\}$ and $\mathbf{f}_{iot} = \left\{\overline{sim}_{i.t}, \overline{rlf}_{i.t}, \overline{rer}_{i.t}\right\}$. * and ** stand for significance at 5% and 10% level. CD test refers to testing the null hypothesis of residual cross-section independence or weak dependence (Pesaran, 2015). α is the estimate of CSD exponent with 90% confidence bands inside parenthesis.

- Figure 2 displays time varying estimates of γ, using the CCEP estimator with both local and global factors approximation.
- Both border and language effect show similar pattern to the case with the EU15 countries.
- Again, we do not observe any evidence in favour of the Euro effect on trade integration, consistent with the currency union formation hypothesis by Frankel (2005).
- The effect of distance has been slightly increasing over the whole period. This is consistent with the meta-study by Disdier and Head (2008), who document that the trade elasticity with respect to distance has not declined, but rather increased recently.

- We propose novel estimation techniques to accommodate CSD within the 3D panel data models.
- Our approach is the first attempt to introduce strong CSD into the multi-dimensional error components.
- The empirical usefulness of the 3D-PCCE estimator is demonstrated via the Monte Carlo studies and the empirical application to the gravity model of the intra-EU trade.
- Extensions and generalisations:
  We aim to develop the challenging models by combining the spatial- and the factor-based techniques.
  Bailey et al. (2016) develop the multi-step estimation procedure that can distinguish the relationship between spatial units from that which is due to the effect of common factors.
  Mastromarco et al. (2015) propose the technique for allowing weak and strong CSD in stochastic frontier panels by combining the exogenously driven factor-based approach and an endogenous threshold regime selection by Kapetanios et al.

Further Issues

- This area is developing very rapidly, with many interesting and often surprising results.
- There is also a general pattern of extending issues in the time-series literature to panels with the multiple indexes.
- Theoretical and applied econometrics are very different activities.
- The former is a deductive activity where you have no data, know the model and derive properties of estimators and tests conditional on that model. There are right and wrong answers.
- The latter is an inductive activity where you do have data, but do not know the model or the questions let alone the answers.
- One must take account of the statistical theory but also the purpose of the activity and the economic context, which define the parameters of interest.
- Different models may be appropriate for different purposes

- There appear to be some general points when using large N large T panels.
- One should be very careful about using pooled estimators to estimate dynamic panels. The dynamic parameters are subject to large potential biases when the parameters differ across groups and the regressors are serially or spatially correlated.
- Pooled regressions can be measuring very different parameters from averages of the corresponding parameters in time-series regressions. This difference can be expressed as a consequence of a dependence between the time-series parameters and the regressors. The interpretation of this difference will depend on the theory related to the substantive application.
- It is important to allow for between group dependence; the CCE estimator is a good start, but you may be able to give the estimates an economic interpretation.

The Joint Modelling of the Spatial Dependence and Unobserved Factors

- Recently, a few studies attempted to develop a combined approach that can accommodate both weak and strong CSD.
- Bailey et al. (2016) develop multi-step estimation procedure that can distinguish the relationship between spatial units that is purely spatial from that which is due to common factors.
- Mastromarco et al. (2015) propose the novel technique in modelling technical efficiency of stochastic frontier panels by combining the exogenously driven factor-based approach and an endogenous threshold regime selection advanced by Kapetanios et al. (2014).
- Shi and Lee (2017), Bai and Li (2015) and Kuersteiner and Prucha (2015) have also developed the framework for jointly modelling spatial effects and interactive effects.

Shi and Lee (2017, JoE)

- Consider the SDPD model:

$$\boldsymbol{Y}_{nt} = \lambda \boldsymbol{W}_n \boldsymbol{Y}_{nt} + \gamma \boldsymbol{Y}_{n,t-1} + \rho \boldsymbol{W}_n \boldsymbol{Y}_{n,t-1} + \boldsymbol{X}_{nt}\boldsymbol{\beta} + \boldsymbol{\Gamma}_n \boldsymbol{f}_t + \boldsymbol{U}_{nt} \tag{260}$$

$$\boldsymbol{U}_{nt} = \alpha \tilde{\boldsymbol{W}}_n \boldsymbol{U}_{nt} + \boldsymbol{\varepsilon}_{nt}$$

  where $\boldsymbol{Y}_{nt}$ is a n-dimensional column vector of dependent variables and $\boldsymbol{X}_{nt}$ is a $n \times (K-2)$ matrix of exogenous regressors, so that the total number of variables in $\boldsymbol{Y}_{n,t-1}$, $\boldsymbol{W}_n \boldsymbol{Y}_{n,t-1}$ and $\boldsymbol{X}_{nt}$ is K.
- The model accommodates two types of CSD; local dependence and global (strong) dependence.

- Individual units are impacted by time-varying unknown factors $\boldsymbol{f}_t$, which captures global (strong) dependence. The effects can be heterogeneous.
- In an earnings regression where Y_{nt} is the wage rate, each row of Γ_n may correspond to a vector of an individual's skills and $\boldsymbol{f}_t$ is its time varying premium.
- The true number of unobserved factors is assumed to be a fixed constant r, much smaller than n and T.
- The matrix of $n \times r$ factor loading Γ_n and the $T \times r$ factors $\boldsymbol{F}_t = (f_1, ..., f_T)'$ are not observed and treated as fixed effects parameters.
- This approach is flexible and allows unknown correlation between factors and regressors.

- The $n \times n$ spatial weights matrices $\boldsymbol{W}_n$ and $\tilde{\boldsymbol{W}}_n$ are used to model spatial dependence, which represents local dependence.
- $\lambda_0 \boldsymbol{W}_n \boldsymbol{Y}_{nt}$ describes contemporaneous spatial interactions.
- $\gamma_0 \boldsymbol{Y}_{n,t-1}$ captures pure dynamic effect.
- $\rho_0 \boldsymbol{W}_n \boldsymbol{Y}_{n,t-1}$ is a spatial time lag of interactions (diffusion).
- The idiosyncratic error $\boldsymbol{U}_{nt}$ with elements of ε_{it} being $iid(0, \sigma_0^2)$ also possesses a spatial structure $\tilde{\boldsymbol{W}}_n$.
- Shi and Lee (2017) propose a QML estimator.
- When n and T are comparable, the estimator is $\sqrt{nT}$ consistent.
- MC experiment shows that QMLE performs well and the proposed bias correction is effective.

Bai and Li (2015)

- Bai and Li (2015) consider jointly modeling spatial interactions, dynamic interactions and common shocks:

$$y_{it} = \alpha_i + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \delta y_{it-1} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\lambda}'_i \boldsymbol{f}_t + e_{it} \quad (261)$$

  where y_{it} is the dependent variable, $\boldsymbol{x}_{it} = (x_{it1}, ..., x_{itk})'$ is a k-dimensional vector of explanatory variables, $\boldsymbol{f}_t$ is an r-dimensional vector of unobservable common shocks; $\boldsymbol{\lambda}_i$ is the corresponding heterogenous response to the common shocks, $\boldsymbol{W}_n$ is a spatial weights matrix whose diagonal elements are 0, and e_{it} are idiosyncratic errors.
- $\boldsymbol{\lambda}'_i \boldsymbol{f}_t$ captures the common-effects, $\sum_{j=1}^{N} w_{ij,N} y_{jt}$ captures the spatial effects, and δy_{it-1} captures the dynamic effects.

- An additional feature is the allowance of cross sectional heteroskedasticity.
- If heteroskedasticity exists but homoskedasticity is imposed, then MLE can be inconsistent. Interestingly, we show that the limiting variance of the MLE is not of a sandwich form if heteroskedasticity is allowed.
- The spatial interaction on the dependent variable gives rise to the endogeneity problem, while the spatial interaction on the errors, in general, does not.
- Existing estimation methods on the common shocks models such as Pesaran (2006) and Bai (2009) cannot be directly applied to model (261) due to the endogeneity from the spatial interactions.

- They consider the pseudo-Gaussian MLE, which simultaneously estimates all parameters including heteroskedasticity.
- The incidental parameters problem (individual-dependent intercepts, interactive effects, heteroskedasticity) exists and the MLE is shown to have a non-negligible bias.
- Following Hahn and Kuersteiner (2002), we conduct bias correction on the MLE to make it center around zero.
- This paper integrates several correlation-modeling techniques and propose dynamic spatial panel data models with common shocks to accommodate possibly complicated correlation structure over cross section and time.
- We propose a bias correction method for the QMLE. The simulations reveal the excellent finite sample properties of the QMLE after bias correction.

Kuersteiner and Prucha (2015)

- Consider panel data $\{y_t, x_t, z_t\}_{t=1}^{T}$, where $y_t = [y_{1t}, ..., y_{nt}]'$, $x_t = [x'_{1t}, ..., x'_{nt}]'$, and $z_t = [z'_{1t}, ..., z'_{nt}]'$ denote the vector of endogenous variables, and matrices of k_x weakly exogenous and k_z strictly exogenous variables.
- They allow for regressors and disturbances to be affected by common shocks.
- Alternatively we allow for CSD from "spatial lags" in endogenous and exogenous variables and in disturbances.
- Spatial lags represent weighted cross sectional averages, where the weights reflect some measure of distance.
- $\varepsilon_t = [\varepsilon_{1t}, ..., \varepsilon_{nt}]'$ is the vector of regression disturbances, $u_t = [u_{1t}, ..., u_{nt}]'$ the vector of unobserved idiosyncratic disturbances, and μ is an $n \times 1$ vector of unobserved factor loadings.

- The $n \times n$ spatial weight matrices denoted as $W_{pt} = (w_{p,ijt})$ and $M_{qt} = (m_{q,ijt})$.
- Let λ and ρ be P, Q dimensional vectors of parameters with typical elements λ_p and ρ_q and define

$$R_t(\lambda) = \sum_{p=1}^{P} \lambda_p W_{pt} \quad \text{for SAR}$$

$$R_t^*(\rho) = I - \sum_{q=1}^{Q} \rho_q M_{qt} \quad \text{for a spatial AR error term}$$

$$R_t^*(\rho) = \left(I + \sum_{q=1}^{Q} \rho_q M_{qt} \right)^{-1} \quad \text{for a spatial MA error term}$$

- The panel data model can be written as

$$y_t = R_t(\lambda) y_t + x_t \beta_x + z_t \beta_z + \varepsilon_t = X_t \delta + \varepsilon_t \qquad (262)$$

$$R_t^* (\rho) \varepsilon_t = \mu f_t + u_t$$

where $X_t = [M_{1t} y_t, ..., M_{Pt} y_t, x_t, z_t]$ and $\delta = [\lambda', \beta']'$.

- As a normalization we take

$$w_{p,iit} = m_{q,iit} = 0 \text{ and } f_T = 1$$

- (262) is a system of n equations describing simultaneous interactions between the individual units.
- The weights are allowed to be endogenous in that they can depend on $\mu_1, ..., \mu_N$ and u_{it}, and can be correlated with ε_t.
- This extension is important to model sequential group formation as in Carrell et al. (2013) or endogenous network formation as in Goldsmith-Pinkham and Imbens (2013).

- The reduced form of the model is given by

$$y_t = (I_n - R_t(\lambda))^{-1} W_t \delta + (I_n - R_t(\lambda))^{-1} \varepsilon_t \tag{263}$$

$$\varepsilon_t = R_t^*(\rho)^{-1} (\mu f_t + u_t)$$

- Applying a Cochrane-Orcutt type transformation by premultiplying the first equation in (262) with $R_t^*(\rho)$ yields

$$R_t^*(\rho)\, y_t = R_t^*(\rho)\, W_t \delta + \mu f_t + u_t \tag{264}$$

Three examples

1. The social interactions model by Graham (2008) illustrates the use of both spatial interaction terms and interactive effects in a social interaction model;
2. The analysis of the group level heterogeneity is based on Carrell et al. (2013), which illustrates the use of higher order, and data-dependent spatial lags to model within-group heterogeneity. By allowing $R_t(\lambda)$ to depend on predetermined outcomes we can accommodate the fact that group membership is not exogenous;
3. The next example is in the area of health, and considers the spread of an infectious disease.

- They consider a class of GMM estimators.
- Significantly expanding the literature by allowing for endogenous spatial weight matrices, time-varying interactive effects, as well as weakly exogenous covariates.
- An important area of application is in social interaction and network models where our specification can accommodate data dependent network formation.
- Identification of spatial interaction parameters is achieved through a combination of linear and quadratic moment conditions.
- We develop an orthogonal forward differencing transformation to estimate factor components while maintaining orthogonality of moment conditions.
- In the social interactions example, orthogonal forward differencing amounts to controlling for unobserved correlated effects by combining multiple outcome measures.

Further suggestions on the KMS model

- The KMS model can be seen to lie between the two extremes characterised by weakly cross-sectionally dependent spatial models and strong factor models.
- **Inflation expectations**: Consider the model with FEs:

$$\pi_{i,t} = \nu_i + \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|\pi_{i,t-1} - \pi_{j,t-1}| \leq r\right) \pi_{j,t-1} + \epsilon_{i,t} \quad (265)$$

  where π is the one-quarter ahead CPI inflation forecast, $\nu_i \sim iid(0, \sigma_\nu^2)$, and obtain the within estimator of ρ along with the consistent estimator of r.
- Notable exceptions are two extreme models, PAR and CSA:

$$\pi_{i,t} = \nu_i + \rho\pi_{i,t-1} + \epsilon_{i,t} \quad (266)$$

$$\pi_{i,t} = \nu_i + \rho\bar{\pi}_{t-1} + \epsilon_{i,t} \quad (267)$$

- We consider the extensions:

$$\pi_{i,t} = \nu_i + \rho_1 \tilde{\pi}_{i,t-1} + \rho_2 \tilde{\pi}^c_{i,t-1} + \epsilon_{i,t} \tag{268}$$

where $\tilde{\pi}_{i,t-1}$ and $\tilde{\pi}^c_{i,t-1}$ are the respective cross-section averages related to similar and dissimilar forecasters given by

$$\tilde{\pi}_{i,t-1} = \frac{1}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(\left|\pi_{i,t-1} - \pi_{j,t-1}\right| \leq r\right) \pi_{j,t-1}$$

$$\tilde{\pi}^c_{i,t-1} = \frac{1}{N - m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(\left|\pi_{i,t-1} - \pi_{j,t-1}\right| > r\right) \pi_{j,t-1}$$

- We can also test the hypothesis of $\rho_2 = 0$ (informational contents arising from dissimilar forecasters).
- If $\rho_2 \neq 0$, what's the predicted sign of ρ_2?

- Anselin et al. (2008) distinguish spatial dynamic models into 4 categories based on the time-space-dynamic specification:

$$x_{i,t} = \rho_0 x_{i,t-1} + \rho_1 \sum_{j \neq i} w_{ij} x_{j,t-1} + \beta \sum_{j \neq i} w_{ij} x_{j,t} + v_i + \epsilon_{i,t} \tag{269}$$

- ρ_0 captures serial dependence, β represents the intensity of a contemporaneous spatial effect and ρ_1 captures *space time autoregressive dependence* (diffusion).
- Most studies focus on the stable case with $\rho_0 + \rho_1 + \beta < 1$.

- if $\rho_0 = \rho_1 = 0$, we are dealing with a SAR model while if $\rho_1 = \beta = 0$ we obtain a 'simple' dynamic model.
- if $\rho_0 = \beta = 0$, we obtain a 'pure-space recursive' model in which dependence results from the neighborhood locations in the previous period;
- if $\beta = 0$, the model is reduced to a 'time space recursive' model in which dependence relates to both the location itself $(x_{i,t-1})$ and its neighbors in the previous period $\sum_{j \neq i} w_{ij} x_{j,t-1}$;
- if $\rho_1 = 0$, we obtain a 'time space simultaneous' model which includes the time lag and the spatial lag.

- The KMS model is similar to the time-space recursive model considered by Korniotis (2010). He applies it to investigate the internal versus external habit formation using the annual consumption data for the U.S. states, and finds that state consumption growth is not significantly affected by its own (lagged) consumption growth but affected by lagged consumption growth of nearby states.
- Notice that the weight w_{ij} measures the importance of $x_{j,t-1}$ on x_{it}. The weights are observed and exogenous.
- Because the spatial-time lag, $\sum_{j=1}^{N} w_{ij} x_{j,t-1}$, is a weighted average of past consumption of other cross-sectional units, it is the measure of the catching-up habit.

- There is a trade-off between KMS model and the time space recursive model by Korniotis (2010).
- In the former the neighbors are selected endogenously but the equal weights are imposed.
- In the Korniotis's model, the neighbors are selected exogenously, but the weights are selected in a flexible manner albeit not time-varying.
- The application of the KMS model to the consumption habit formation will provide an interesting insight.

- We generalise (265) and allow different weights to the selected neighbors as follows:

$$x_{i,t} = \nu_i + \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) w_{ij} x_{j,t-1} + \epsilon_{i,t} \tag{270}$$

where we consider the following weights

$$w_{ij} = \frac{d_{ij}^{-2}}{\sum_{j=1}^{N} d_{ij}^{-2}}, \;\; d_{ij} = |x_{i,t-1} - x_{j,t-1}| \tag{271}$$

- The estimation can be done in two steps: first, the consistent estimate of r is obtained. Then, construct the weights by (271) and the associated cross-section averages, and estimate the model, (270).
- Possibly more complicated due to the grid search over r. The KMS approach would be more general than spatial models.

Endogenous Spatial Weights Matrix and KMS

- The choice of appropriate spatial weights is central for spatial models as it assumes a structure of spatial dependence, which may not correspond closely to reality.
- The choice of weights is arbitrary, and empirical results vary considerably.
- When the spatial weights matrix is constructed with economic/socioeconomic distances, it can be time varying (e.g. Case et al., 1993).
- Lee and Yu (2012b) investigate QMLE of SDPD models with time varying spatial weights. MC results show that a model misspecification of a time invariant spatial weights may cause substantial bias.

- Define the $N \times N$ matrix of the spatial weights:

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{NN} \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{w}_N \end{bmatrix} \text{ with } w_{ii} = 0 \tag{272}$$

- The jth element of $\boldsymbol{w}_i$, w_{ij}, represents the link between the neighbor j and the spatial unit i.
- It is a common practice to have $\boldsymbol{W}$ having a zero diagonal and being row-normalized.
- The ith row $\boldsymbol{w}_i$ may be constructed as $\boldsymbol{w}_i = (d_{i1}, d_{i2}, ..., d_{in}) / \sum_{j=1}^{n} d_{ij}$, where $d_{ij} \geq 0$ represents a function of the spatial distance between ith and jth units.

- **Assumption 6** in Shi and Lee (2017, JoE). (1) The spatial weights satisfy $w_{ij,t} \geq 0$, $w_{ij,t} = 0$, and $w_{ij,t} = 0$ if $\rho_{ij,t} > \rho_c$, i.e., there exists a threshold ρ_c such that the weight is zero if the geographic distance exceeds ρ_c. For $i \neq j$,

$$w_{ij,t} = h_{ij}\left(z_{it}, z_{jt}\right) I\left(\rho_{ij,t} < \rho_c\right) \qquad (273)$$

or the row normalized version that

$$w_{ij,t} = \frac{h_{ij}\left(z_{it}, z_{jt}\right) I\left(\rho_{ij,t} < \rho_c\right)}{\sum_{k=1}^{n} h_{ik}\left(z_{it}, z_{kt}\right) I\left(\rho_{ik,t} < \rho_c\right)}$$

where $h_{ij}()$'s are nonnegative, uniformly bounded functions. (2) The function $h_{ij}(.)$ satisfies the Lipschitz condition,

$$\left|h_{ij}\left(a_1, b_1\right) - h_{ij}\left(a_2, b_2\right)\right| \leq c_0\left(\left|a_1 - a_2\right| + \left|b_1 - b_2\right|\right)$$

for some finite constant c_0.

- For convenience consider the simple SAR model:

$$y_{it} = \lambda \sum_{j=1}^{n} w_{ij,t} y_{jt} + v_{it}. \tag{274}$$

- Different specifications for the spatial weights:
- Fixed weights based on the physical or economic distance. This exogenous assumption may hold when spatial weights are constructed using predetermined geographic distances.
- If economic distance such as the GDP or trade volume is used to construct the weight matrix, it is likely that these elements are correlated with the final outcome.
- BHP use the spatial correlation-based adjacency matrix subject to sparsity. However, it may render the estimation less reliable, e.g. in the UK house price data by Beulah's thesis, the correlation-based weights matrix contains only 0.3% nonzero weights.

- Qu and Lee (2015) and Shi and Lee (2017) consider:

$$h_{ij}\left(z_{it}, z_{jt}\right) = \frac{1}{\left|z_{it} - z_{jt}\right|} \tag{275}$$

  where $\left|z_{it} - z_{jt}\right|$ measures the economic distance.
- Using (273) and (275), the model (274) can be written as

$$y_{it} = \lambda \sum_{j=1}^{n} \frac{1}{\left|z_{it} - z_{jt}\right|} I\left(\rho_{ij,t} < \rho_c\right) y_{jt} + v_{it}, \tag{276}$$

  where $I\left(\rho_{ij,t} < \rho_c\right)$ is predetermined.
- Assuming that geographic distance ρ_c is known and z_{it} is correlated with v_{it}, Shi and Lee develops the CF approach:

$$y_{it} = \lambda \sum_{j=1}^{n} w_{ij,t} y_{jt} + \left(Z_{nt} - X_{2nt}\Gamma\right)'_{it} \delta + \xi_{it}. \tag{277}$$

$$w_{ij,t} = \frac{1}{\left|z_{it} - z_{jt}\right|} I\left(\rho_{ij,t} < \rho_c\right)$$

- Consider the single covariate in z:

$$z_{it} = x_{it}'\beta_z + \gamma_{iz}' f_{zt} + \epsilon_{it} \tag{278}$$

where x_{it} are $k_z \times 1$ regressors, f_{zt} consisting of $R_z \times 1$ factors with loading γ_{izl}' and ϵ_{it} is idiosyncratic error.

- Then, (277) can be written as

$$y_{it} = \lambda \sum_{j=1}^{n} w_{ij,t} y_{jt} + \left(z_{it} - x_{it}'\beta_z - \gamma_{iz}' f_{zt}\right)' \delta + \xi_{it}. \tag{279}$$

- The structure of (276) is similar to KMS, given by

$$y_{it} = \lambda \frac{1}{m_{it}} \sum_{j=1}^{n} I\left(\rho_{ij,t} < \rho_c\right) y_{jt} + v_{it}. \tag{280}$$

where $m_{it} = \sum_{j=1}^{n} I\left(\rho_{ij,t} < \rho_c\right)$.

- (276) assumes the known threshold ρ_c whilst (280) estimates λ and ρ_c jointly, then imposing the equal weight once y_{jt} is selected.

- More generally, we consider:

$$w_{ij,t} = \frac{1}{\rho_{ij,t}} I\left(\rho_{ij,t} < \rho_c\right) \;\; with \;\; \rho_{ij,t} = \left|z_{it} - z_{jt}\right| \tag{281}$$

- We may consider the row-normalisation version as

$$w_{ij,t} = \frac{\frac{1}{\rho_{ij,t}} I\left(\rho_{ij,t} < \rho_c\right)}{\sum_{j=1}^{n} \frac{1}{\rho_{ij,t}} I\left(\rho_{ij,t} < \rho_c\right)} \; with \; \rho_{ij,t} = \left|z_{it} - z_{jt}\right| \tag{282}$$

- We can consider (i) the exogenous case, $E\left(z_{it}' v_{it}\right) = 0$ and (ii) the endogenous case, $E\left(z_{it}' v_{it}\right) \neq 0$.
- We may also consider the time-invariant case using

$$w_{ij} = \frac{1}{\rho_{ij}} I\left(\rho_{ij} < \rho_c\right) \;\; with \;\; \rho_{ij} = \left|z_i - z_j\right| \text{ or } \rho_{ij} = \left|\bar{z}_i - \bar{z}_j\right| \tag{283}$$

where we use time-invariant covariate, z_i or the time-average, $\bar{z}_i = T^{-1} \sum z_{it}$.

- We conjecture that the KMS algorithm to construct the endogenous spatial weights matrix would be useful.
- For example, we consider VAR as the DGP for the $N \times 1$ vector, $\mathbf{z}_t = (z_{1t}, ..., z_{Nt})'$, say

$$\mathbf{z}_t = \sum_{j=1}^{p} \Phi_j \mathbf{z}_{t-j} + \boldsymbol{\epsilon}_t$$

and derive the CF:

$$\mathbf{v}_t = \left(\mathbf{z}_t - \sum_{j=1}^{p} \Phi_j \mathbf{z}_{t-j} \right)' \boldsymbol{\delta} + \boldsymbol{\xi}_t.$$

- Then, the final model will become:

$$y_{it} = \lambda \sum_{j=1}^{n} w_{ij,t} y_{jt} + \boldsymbol{\epsilon}_{it}' \delta + \xi_{it} \tag{284}$$

- **MORE discussions...**
- Relationship to the selection bias corrections using both parametric and semiparametric approaches.
- Also in threshold models and network formation or selection?
- CCE approximation possible for spatial and factors?
- Heterogeneous extension of KMS?
- How to estimate the weights and thresholds jointly in KMS with and without endogeneity?

Regime-switching panel data modelling with spatial effects and factors

- Another important issue is how to model the spatial dependence, the spatial heterogeneity and the spatial nonlinearity, simultaneously.
- See the lecture handouts for a preliminary review of the current literature.

- Eventually, this project aims to
- develop the general econometrics specifications that can accommodate the spatial and factor dependence, the spatial heterogeneity, the endogenous spatial weights matrix as well as the spatial nonlinearity in dynamic heterogenous panels in a rather unified framework by combining all the recent advances made in the related literature.
- extend all the advances to the multi-dimensional dataset, separately and jointly. As the dimension grows, it would be more complicated and challenging to develop the hierarchical and structural model of both spatial effects and factors, jointly.
- These works will be of great applicability to a variety of the big dataset, not only the health economics data...

Further Modelling Issues

- This area is developing very rapidly, with very many interesting and often surprising results emerging.
- There is also a general pattern of extending issues in the time-series literature to panels with the multiple indexes.
- Theoretical and applied econometrics are very different activities.
- The former is a deductive activity where you have no data, know the model and derive properties of estimators and tests conditional on that model. There are right and wrong answers.
- The latter is an inductive activity where you do have data, but do not know the model or the questions let alone the answers.

- In applied econometrics one must take account of not merely the statistical theory but also the purpose of the activity and the economic context, which define the parameters of interest.
- Different models may be appropriate for different purposes, such as forecasting, policy analysis or testing hypotheses and purpose and the economic context (theory, history, institutions) should guide the choice of model.
- There are some general points that applied workers might bear in mind when using large N large T panels.

- First, one should be very careful about using pooled estimators to estimate dynamic panel data models. The dynamic parameters are subject to large potential biases when they differ across groups and the regressors are serially or spatially correlated.
- Second, pooled regressions can be measuring very different parameters from the averages of the corresponding parameters in time-series regressions. This difference can be expressed as a consequence of a dependence between the time-series parameters and the regressors. Its interpretation will depend on the theory related to the substantive application. It is not primarily a matter of statistical technique.
- Third, it is important to allow for between group dependence, the CCE estimator is a good start, but you may also need to be able to give the estimates an economic interpretation.

Further Modelling Issues

- This area is developing very rapidly, with many interesting and often surprising results emerging.
- We assume spatial weighting matrix to be determined exogenously, ruling out a number of exciting research areas in social networks. Our control function approach has the potential to control for this source of endogeneity.
- We have restricted ourselves to linear effects, both in time and across space, and to modelling conditional means rather than other quantiles of the conditional distribution.
- Eventually, this project aims to develop the general econometrics models that can accommodate spatial and factor dependence, spatial heterogeneity, endogenous spatial weight matrix as well as spatial nonlinearity in a unified framework.