Dynamic Heterogeneous Panels

Contents

1 Model, asymptotic property of IV estimator **2**
- 1.1 The models and Assumptions . . . 2
  - 1.1.1 Fixed effects model . . . 2
  - 1.1.2 Trend model . . . 2
- 1.2 IV estimation method and asymptotic property . . . 3
  - 1.2.1 Mean group IV estimator . . . 4

2 Construction optimal instruments **5**

3 Estimation method on dynamic heterogeneous panel data model with multifactor error structure **6**
- 3.1 Norkutes' (2019) IVMG estimator . . . 7

4 Monte Carlo simulation design **7**
- 4.1 dynamic heterogeneous panels data model without error factor structure 7
- 4.2 Dynamic heterogeneous panels data model with multi-factor error structure . . . 8

5 Monte Carlo simulation results **10**
- 5.1 Dynamic Heterogeneous Panels without multifactor error structure . 10
- 5.2 Dynamic Heterogeneous Panels with multifactor error structure . . . 10

6 Short summary **11**
- 6.1 Dynamic Heterogeneous Panels without multifactor error structure . 11
- 6.2 Dynamic Heterogeneous Panels with multifactor error structure . . . 11

A Asymptotic property of LS estimator **12**
  - A.0.1 Mean group LS estimator . . . 14

Reference **16**

1 Model, asymptotic property of IV estimator

1.1 The models and Assumptions

1.1.1 Fixed effects model

Consider the dynamic heterogeneous panels data model with fixed effects:

$$\begin{aligned} y_{i,t} &= \alpha_i + \phi_i y_{i,t-1} + \boldsymbol{x}_{i,t}^{'}\boldsymbol{\beta}_i + u_{i,t},\ \ for\ i = 1, \ldots N;\ t = 1, \ldots, T\,, \\ &= \boldsymbol{w}_{i,t}^{'}\boldsymbol{\theta}_i + \alpha_i + u_{i,t}, \end{aligned} \tag{1}$$

where $\boldsymbol{x}_{i,t}$ and $\boldsymbol{\beta}_i$ are $k \times 1$ vectors, $\boldsymbol{\theta}_i = \left(\phi_i, \boldsymbol{\beta}_i^{'}\right)^{'}$ and $\boldsymbol{w}_{i,t} = \left(y_{i,t-1}, \boldsymbol{x}_{i,t}^{'}\right)^{'}$ is a $(1+k) \times 1$ vector. Stacking the T observations for each i, we have

$$\boldsymbol{y}_i = \boldsymbol{W}_i \boldsymbol{\theta}_i + \alpha_i \boldsymbol{\iota}_T + \boldsymbol{u}_i, \tag{2}$$

where $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,T})^{'}$, $\boldsymbol{W}_i = (\boldsymbol{w}_{i,1}, \ldots, \boldsymbol{w}_{i,T})^{'}$, $\iota_T = (1, \ldots, 1)^{'}$ and $\boldsymbol{u}_i = (u_{i,1}, \ldots, u_{i,T})^{'}$.

Due to the incidental parameters problem arise, we use forward filter to the model by Moon and Phillips (2000), Hayakawa (2009) and Hayakawa et al. (2019). We define the $(T-1) \times 1$ forward demeaning matrix as

$$\boldsymbol{F} = diag(c_1, c_2, \ldots c_{T-1}) \begin{bmatrix} 1 & \frac{-1}{T-1} & \cdots & \cdots & \frac{-1}{T-1} \\ \vdots & 1 & \frac{-1}{T-2} & \cdots & \frac{-1}{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}, \tag{3}$$

where $c_t = \sqrt{(T-t)(T-t+1)}$. Multiply $\boldsymbol{F}$ to model (2), the model can expressed as

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{W}}_i \boldsymbol{\theta} + \tilde{\boldsymbol{u}}_i, \tag{4}$$

where $\tilde{\boldsymbol{y}}_i = \boldsymbol{F}\boldsymbol{y}_i = \left(\tilde{\boldsymbol{y}}_{i,1}, \ldots, \tilde{\boldsymbol{y}}_{i,T-1}\right)^{'}$, $\tilde{\boldsymbol{W}}_i = \boldsymbol{F}\boldsymbol{W}_i = (\tilde{\boldsymbol{w}}_{i,1}, \ldots, \tilde{\boldsymbol{w}}_{i,T-1})^{'}$ and $\tilde{\boldsymbol{u}}_i = \boldsymbol{F}\boldsymbol{u}_i = (\tilde{\boldsymbol{u}}_{i,1}, \ldots, \tilde{\boldsymbol{u}}_{i,T-1})^{'}$ with $\tilde{\boldsymbol{y}}_{i,t} = c_t \left[y_{i,t} - (y_{i,t+1} + \cdots + y_{i,T}) / (T-t)\right]$, for $t = 1, \ldots, T-1$.

1.1.2 Trend model

Consider the dynamic heterogeneous panels data model with fixed effects and heterogeneous time trends:

$$\begin{aligned} y_{i,t} &= \alpha_i + \eta_i t + \phi_i y_{i,t-1} + \boldsymbol{x}_{i,t}^{'}\boldsymbol{\beta}_i + u_{i,t},\ \ for\ i = 1, \ldots N;\ t = 1, \ldots, T\,, \\ &= \boldsymbol{w}_{i,t}^{'}\boldsymbol{\theta} + \alpha_i + \eta_i t + u_{i,t}, \end{aligned} \tag{5}$$

where $\boldsymbol{x}_{i,t}$ and $\boldsymbol{\beta}_i$ are $k \times 1$ vectors and $\boldsymbol{w}_{i,t} = \left(\phi_i, \boldsymbol{\beta}_i^{'}\right)^{'}$ is a $(1+k) \times 1$ vector. Stacking the T observations for each i, we have

$$\boldsymbol{y}_i = \boldsymbol{W}_i \boldsymbol{\theta}_i + \alpha_i \boldsymbol{\iota}_T + \eta_i \boldsymbol{\tau}_T + \boldsymbol{u}_i, \tag{6}$$

where $\boldsymbol{\tau}_T = (1, 2, \ldots, T)^{'}$. Again, we define the forward demeaning matrix as

$$\boldsymbol{F}^{\tau} = diag(c_1^{\tau}, c_2^{\tau}, \ldots c_{T-2}^{\tau}) \begin{bmatrix} 1 & \frac{2(-2(T-2))}{(T-1)(T-2)} & \frac{2(-2(T-2)+3)}{(T-1)(T-2)} & \cdots & \frac{2(-2(T-2)+3(T-2))}{(T-1)(T-2)} \\ 0 & 1 & \frac{2(-2(T-3))}{(T-2)(T-3)} & \cdots & \frac{2(-2(T-3)+3(T-3))}{(T-3)(T-4)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \frac{2(-2+3)}{2\cdot 1} \end{bmatrix}, \tag{7}$$

where $c_t^{\tau} = \sqrt{(}(T-t)(T-t-1)/(T-t-1)(T-t+2))$.

Multiply $\boldsymbol{F}$ to model (6), the model can expressed as

$$\tilde{\boldsymbol{y}}_i^{\tau} = \tilde{\boldsymbol{W}}_i^{\tau} \boldsymbol{\theta} + \tilde{\boldsymbol{u}}_i^{\tau}, \tag{8}$$

where $\tilde{\boldsymbol{y}}_i^{\tau} = \boldsymbol{F}^{\tau}\boldsymbol{y}_i = \left(\tilde{\boldsymbol{y}}_{i,1}^{\tau}, \ldots, \tilde{\boldsymbol{y}}_{i,T-2}^{\tau}\right)^{'}$, $\tilde{\boldsymbol{W}}_i^{\tau} = \boldsymbol{F}^{\tau}\boldsymbol{W}_i$ and $\tilde{\boldsymbol{u}}_i^{\tau} = \boldsymbol{F}^{\tau}\boldsymbol{u}_i$.

Assumption 1 *$x_{i,t}$ and $u_{i,t}$ are independently distributed for all t and s .*

1.2 IV estimation method and asymptotic property

Norkute et al. (2019) propose an IV estimator in dynamic heterogeneous panel data model. They use current and lagged values of $\boldsymbol{x}_i$ as instruments, as

$$\boldsymbol{Z}_{i,t} = (\boldsymbol{x}_i, \boldsymbol{x}_{i,-1})^{'}, \tag{9}$$

where $\boldsymbol{Z}_i$ is $T \times 2k$ vector.

Assumption 2 *$\boldsymbol{A}_i = \text{plim}_{T\to\infty} \tilde{\boldsymbol{A}}_{i,T}$ has full column rank, $\boldsymbol{B}_i = \text{plim}_{T\to\infty} \tilde{\boldsymbol{B}}_{i,T}$ and $\boldsymbol{\Sigma}_i = \text{plim}_{T\to\infty} T^{-1}\boldsymbol{Z}_i^{'}\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{Z}$ has positive definite, uniformly, where $\tilde{\boldsymbol{A}}_{i,T} = \frac{1}{T}\boldsymbol{Z}_i^{'}\tilde{\boldsymbol{W}}_i^{b}$ and $\tilde{\boldsymbol{B}}_{i,T} = \frac{1}{T}\tilde{\boldsymbol{Z}}_i^{'}\boldsymbol{Z}_i$,*
where $b = 1$ corresponds to the fixed effects model while $b = 2$ corresponds to the trend models.

Then, the IV estimator can be expressed as

$$\hat{\boldsymbol{\theta}}_{IV,i}^{b} = \left(\tilde{\boldsymbol{A}}_{i,T}^{'}\tilde{\boldsymbol{B}}_{i,T}^{-1}\tilde{\boldsymbol{A}}_{i,T}\right)\tilde{\boldsymbol{A}}_{i,T}^{'}\tilde{\boldsymbol{B}}_{i,T}^{-1}\tilde{\boldsymbol{g}}_{i,T}, \tag{10}$$

where

$$\tilde{\boldsymbol{g}}_{i,T} = \frac{1}{T}\boldsymbol{Z}_i^{'}\tilde{\boldsymbol{y}}_i^{b}, \tag{11}$$

and $\tilde{\boldsymbol{W}}_i^{b} = \left(\tilde{w}_{i,1}^{b}, \ldots, \tilde{w}_{i,T}^{b}\right)^{'}$ is $T \times 2k$ matrix

From above equation, we have

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b} - \boldsymbol{\theta}_i\right) = \left(\tilde{\boldsymbol{A}}_{i,T}^{'}\tilde{\boldsymbol{B}}_{i,T}^{-1}\tilde{\boldsymbol{A}}_{i,T}\right)\tilde{\boldsymbol{A}}_{i,T}^{'}\tilde{\boldsymbol{B}}_{i,T}^{-1}\left(T^{-1/2}\boldsymbol{Z}_i^{'}\tilde{\boldsymbol{u}}_i\right) \tag{12}$$

Then, the property of $T^{-1/2}\boldsymbol{Z}_i^{'}\tilde{\boldsymbol{u}}_i$ is given by following proposition.

proposition 1 *Under above assumptions, as* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$*, for each* i*, we have*

$$\begin{aligned} & N^{-1}\boldsymbol{Z}_i^{'}\tilde{\boldsymbol{u}}_i \xrightarrow{p} \mathbf{0}, \\ & \quad \textit{and} \\ & T^{-1/2}\boldsymbol{Z}_i^{'}\tilde{\boldsymbol{u}}_i \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right). \end{aligned} \tag{13}$$

Thus, IV estimator, $\boldsymbol{\theta}_{IV,i}^{b}$ is $\sqrt{T}$ consistent to $\boldsymbol{\theta}_i$ and this estimator does not have Nickell's bias. Then, we have following theorem

Theorem 1

As $(N,T) \to \infty$ such that $N/T \to c$ with $0 < c < \infty$. for each i,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b} - \boldsymbol{\theta}_i\right) \xrightarrow{d} N\left(\mathbf{0}, \left(\boldsymbol{A}_i^{'}\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\right)^{-1}\boldsymbol{A}_i^{'}\boldsymbol{B}_i^{-1}\boldsymbol{\Sigma}_i\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\left(\boldsymbol{A}_i^{'}\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\right)\right). \tag{14}$$

1.2.1 Mean group IV estimator

Now, we define the mean group estimator of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_{IVMG}^{b} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{IVi}^{b}. \tag{15}$$

And we can show that the asymptotic property of $\hat{\boldsymbol{\theta}}_{IVMG}$, as

$$\hat{\boldsymbol{\theta}}_{IVMG}^{b} - \boldsymbol{\theta} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b} - \boldsymbol{\theta}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b} - \boldsymbol{\theta}_i\right) + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_i, \tag{16}$$

where

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b} - \boldsymbol{\theta}_i\right) &= \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{\boldsymbol{A}}_{i,T}^{'}\tilde{\boldsymbol{B}}_{i,T}^{-1}\tilde{\boldsymbol{A}}_{i,T}\right)^{-1}\tilde{\boldsymbol{A}}_{i,T}^{'}\tilde{\boldsymbol{B}}_{i,T}^{-1}\left(T^{-1}\tilde{\boldsymbol{Z}}_i^{'}\tilde{\boldsymbol{u}}_i\right) \\ &= O_p\left(\delta_{NT}^{-2}\right), \end{aligned} \tag{17}$$

where $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. We note that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\lambda}_i = O_p\left(N^{-1/2}\right)$, if $\delta_{NT}^{-(2+\varsigma)} \to 0$ for any $\varsigma > 0$, we have

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{IVMG}^{b} - \boldsymbol{\theta}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\lambda}_i + o_p(1). \tag{18}$$

And the variance estimator of $\hat{\boldsymbol{\theta}}_{IVMG}$ is given by

$$\hat{\boldsymbol{\Sigma}}_{IV,\lambda} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b} - \hat{\boldsymbol{\theta}}_{IVMG}^{b}\right)\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b} - \hat{\boldsymbol{\theta}}_{IVMG}^{b}\right)^{'}. \tag{19}$$

Follow Norkute et al. (2019), we can show that $\hat{\boldsymbol{\Sigma}}_{IV,\lambda}$ is consistent and it does not have small T bias. Firstly, we decompose (19) as

$$\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b}-\boldsymbol{\theta}+\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{IVMG}^{b}\right)\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b}-\boldsymbol{\theta}+\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{IVMG}^{b}\right)^{\prime}=$$
$$\sum_{i=1}^{N}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i^{\prime}+\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b}-\boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b}-\boldsymbol{\theta}_i\right)^{\prime}+\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b}-\boldsymbol{\theta}_i\right)\boldsymbol{\lambda}_i+\sum_{i=1}^{N}\boldsymbol{\lambda}_i\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b}-\boldsymbol{\theta}_i\right)-$$
$$N\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{IVMG}^{b}\right)^{\prime}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{IVMG}^{b}\right). \tag{20}$$

Then we can show consistent of $\hat{\boldsymbol{\Sigma}}_{IV,\lambda}^{b}$ as

$$\hat{\boldsymbol{\Sigma}}_{IV,\lambda}^{b}-\boldsymbol{\Sigma}_{IV,\lambda}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i^{\prime}-\boldsymbol{\Sigma}_{IV,\lambda}\right)+\frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b}-\boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b}-\boldsymbol{\theta}_i\right)^{\prime}$$
$$+\frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b}-\boldsymbol{\theta}_i\right)\boldsymbol{\lambda}_i+\frac{1}{N-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\left(\hat{\boldsymbol{\theta}}_{IV,i}^{b}-\boldsymbol{\theta}_i\right)-$$
$$\frac{N}{N-1}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{IVMG}^{b}\right)^{\prime}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{IVMG}^{b}\right)=o_p(1). \tag{21}$$

Then, we can see that the asymptotic property of $\hat{\boldsymbol{\theta}}_{IVMG}$ as,

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{IVMG}^{b}-\boldsymbol{\theta}\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Sigma}_{IV,\lambda}\right). \tag{22}$$

2 Construction optimal instruments

How to select optimal instruments is important issue in empirical studies. Donald and Newey (2001), Kuersteiner and Okui (2010), Okui (2011), Kang (2019) and Lee and Shin (2019) propose different methods to construct optimal instruments. Here, we try to apply model average approach by Kuersteiner and Okui (2010) in here.

If we can observed long past periods variables, we have many instruments. In this case we define our instruments, as

$$\boldsymbol{Z}_i^{(j)}=\left(\boldsymbol{x}_{i,\cdot},\boldsymbol{x}_{i,-1},\ldots,\boldsymbol{x}_{i,-j}\right)^{\prime}, \tag{23}$$

where $\boldsymbol{Z}_i^{(j)}$ is $T\times(j+1)k$ matrix with $j\geq 1$.

Then, we can define IV estimator, $\boldsymbol{\theta}_{IV,i}^{b}$, as

$$\hat{\boldsymbol{\theta}}_{IV,i}^{b}=\left(\tilde{\boldsymbol{W}}_i^{\prime}\boldsymbol{P}_i^{(j)}\tilde{\boldsymbol{W}}_i\right)^{-1}\tilde{\boldsymbol{W}}_i^{\prime}\boldsymbol{P}_i^{(j)}\tilde{\boldsymbol{y}}_i, \tag{24}$$

where $\boldsymbol{P}_i^{(j)}=\boldsymbol{Z}_i^{(j)}\left(\boldsymbol{Z}_i^{(j)\prime}\boldsymbol{Z}_i^{(j)}\right)^{-1}\boldsymbol{Z}_i^{(j)\prime}$ and $b=1$ corresponds to the fixed effects model while $b=2$ corresponds to the trend models.

If we can observe long lagged length from data, we have many instruments that can be used. In empirical study, researchers do not use all past variables as instruments because there are trade off between efficiency and bias. But we have not clearly know that the effect of using long lagged length IV estimator in heterogeneous dynamic panel data model. And how to select the instruments to balance the bias and efficiency.

Kuersteiner and Okui (2010) provided model averaging two stage least squares estimator to balance the bias and efficiency. We try to follow this method to construct the optimal instruments. We define a weighting vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_J)'$. Then, we can weight $\boldsymbol{P}_i^{(j)}$ as

$$\boldsymbol{P}_i\left(\boldsymbol{\omega}\right) = \sum_{j=1}^{J} \omega_j \boldsymbol{P}_i^{(j)}, \tag{25}$$

where J is maximum number of lagged variables that we can observed. Our goal is to select $\boldsymbol{\omega}$ to minimize the approximate mean square error, $S_\lambda\left(\boldsymbol{\omega}\right)$, where λ are some fixed parameters. In beginning, we consider positive weights, such that $\omega_j \in [0,1]$. And we define $\boldsymbol{f}_i = \left(\boldsymbol{Z}_i^{(1)}, \ldots, \boldsymbol{Z}_i^{(J)}\right)'$. Let $\hat{\boldsymbol{H}}$ is some estimator of $\boldsymbol{H} = \frac{\boldsymbol{f}_i'\boldsymbol{f}_i}{n}$, and $\tilde{v}_i$ be some preliminary residual from first stage regression. Also, let $\tilde{u}_i$ be some preliminary residual that does not depend on the weighting vector. Let $\tilde{v}_{\lambda,i} = \tilde{v}_i \hat{H}_i^{-1}\lambda$. Define

$$\hat{\sigma}_{u,i}^2 = \frac{\tilde{u}_i'\tilde{u}_i}{N}, \ \hat{\sigma}_{\lambda,i}^2 = \frac{\tilde{v}_{\lambda,i}'\tilde{v}_{\lambda,i}}{N}, \ \hat{\sigma}_{\lambda u,i}^2 = \frac{\tilde{v}_{\lambda,i}'\tilde{u}_i}{N}. \tag{26}$$

Let $\hat{v}_{\lambda,i}^j = \left(\boldsymbol{P}_i^{(J)} - \boldsymbol{P}_i^{(j)}\right)\boldsymbol{W}_i\hat{\boldsymbol{H}}^{-1}\lambda$ and $\hat{U} = \left(\hat{v}_{\lambda,i}^1, \ldots, \hat{v}_{\lambda,i}^J\right)'\left(\hat{v}_{\lambda,i}^1, \ldots, \hat{v}_{\lambda,i}^J\right)$. Let $\boldsymbol{\Gamma}$ be the $J \times J$ matrix whose (i,j) element is $min(i,j)$ and let $K = (1, \ldots, J)'$. The criterion $\hat{S}_\lambda(\boldsymbol{\omega})$ is

$$\hat{S}_\lambda(\boldsymbol{\omega}) = \hat{a}_\lambda \frac{(K'\boldsymbol{\omega})^2}{N} + \hat{\sigma}_u^2 \frac{\boldsymbol{\omega}'\hat{U}\boldsymbol{\omega} - \sigma_\lambda^2(J - 2K'\boldsymbol{\omega} + \boldsymbol{\omega}'\Gamma\boldsymbol{\omega})}{N}. \tag{27}$$

Then, we can find the optimal weight by minimize $\hat{S}_\lambda(\boldsymbol{\omega})$.

Following Donald and Newey (2001) and Kuersteiner and Okui (2010), an approximation of MSE is conditional on the exogenous variable by $\sigma_{u,i}^2 Q^{-1} + S\left(W\right)$, where

3 Estimation method on dynamic heterogeneous panel data model with multifactor error structure

For convenient, we assume the number of regressor is 1 and we express the model as

$$y_{i,t} = \phi_i y_{i,t-1} + \sum_{\ell=1}^{k} \beta_{\ell i} x_{\ell i,t} + u_{i,t}, \ for\ i = 1, \ldots N;\ t = 1, \ldots, T,\ \ell = 1, \ldots, k. \tag{28}$$

Consider the model (20), we drawn $x_{\ell i,t}$ as

$$x_{\ell i,t} = \boldsymbol{\gamma}_{xi}^{0\prime} \boldsymbol{f}_{xt}^{0} + \varepsilon_{xi,t} \tag{29}$$

and the idiosyncratic errors of the process for $y_{i,t}$ as

$$u_{i,t} = \boldsymbol{\gamma}_{yi}^{0\prime} \boldsymbol{f}_{yt}^{0} + \varepsilon_{yi,t}, \tag{30}$$

where $\boldsymbol{\gamma}_{yi}^{0}$ and $\boldsymbol{\gamma}_{xi}^{0}$ are $m_y \times 1$ and $m_x \times 1$ true factor loading respectively, $\boldsymbol{f}_{yt}^{0}$ and $\boldsymbol{f}_{xt}^{0}$ are $m_y \times 1$ and $m_x \times 1$ true vector of unobservable factors respectively.

3.1 Norkutes' (2019) IVMG estimator

We asymptotically eliminate the common factor in $\boldsymbol{x}_i$ by projecting matrix, $\boldsymbol{M}_{F_x^0}$.

$$\boldsymbol{M}_{F_x^0} = \boldsymbol{I}_T - \boldsymbol{F}_x^0 \left(\boldsymbol{F}_x^{0\prime} \boldsymbol{F}_x^0 \right)^{-1} \boldsymbol{F}_x^{0\prime}; \boldsymbol{M}_{F_{x,-1}^0} = \boldsymbol{I}_T - \boldsymbol{F}_{x,-1}^0 \left(\boldsymbol{F}_{x,-1}^{0\prime} \boldsymbol{F}_{x,-1}^0 \right)^{-1} \boldsymbol{F}_{x,-1}^{0\prime} \tag{31}$$

And using the defactored covariates as instruments, as

$$\boldsymbol{Z}_{IVi} = \left(\boldsymbol{M}_{F_X^0} \boldsymbol{x}_i, \boldsymbol{M}_{F_{x,-1}^0} \boldsymbol{X}_{i,-1} \right) \tag{32}$$

The first step IV estimator can be expressed as

$$\begin{aligned} \hat{\boldsymbol{\varphi}}_{IVi} = &\left(\left(\frac{\boldsymbol{Z}_i' \boldsymbol{M}_{F_X^0} \boldsymbol{W}_i}{T} \right)' \left(\frac{\boldsymbol{Z}_i' \boldsymbol{M}_{F_X^0} \boldsymbol{Z}_i}{T} \right)^{-1} \left(\frac{\boldsymbol{Z}_i' \boldsymbol{M}_{F_X^0} \boldsymbol{W}_i}{T} \right) \right)^{-1} \\ &\left(\left(\frac{\boldsymbol{Z}_i' \boldsymbol{M}_{F_X^0} \boldsymbol{W}_i}{T} \right)' \left(\frac{\boldsymbol{Z}_i' \boldsymbol{M}_{F_X^0} \boldsymbol{Z}_i}{T} \right)^{-1} \left(\frac{\boldsymbol{Z}_i' \boldsymbol{M}_{F_X^0} \boldsymbol{y}_i}{T} \right) \right). \end{aligned} \tag{33}$$

4 Monte Carlo simulation design

4.1 dynamic heterogeneous panels data model without error factor structure

The data generating process:

$$\begin{aligned} y_{i,t} &= \phi_i y_{i,t-1} + \sum_{\ell=1}^{k} \beta_{\ell i} x_{\ell i,t} + u_{i,t}, \;\; for \; i = 1, \dots N; \; t = -49, \dots, T\,, \\ x_{\ell i,t} &= \sum_{\ell=1}^{k} \phi_{\ell i} x_{\ell i,t-1} + v_{\ell i,t}, \end{aligned} \tag{34}$$

where $u_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,\, 1)$, and $v_{\ell i,t} = \rho_{v,\ell} v_{\ell i,t-1} + \left(1 - \rho_{v,\ell}^2\right)^{\frac{1}{2}} \varpi_{\ell i,t}, \varpi_{\ell i,t} \overset{i.i.d.}{\sim} U(0.5, 1.5)\,, \rho_{v,\ell} = 0.5$.

The slope coefficients are generated as

$$\phi_i = \phi + \eta_{\phi i}, \;\; \beta_{1,i} = \beta_1 + \eta_{\beta_1 i} \, and \, \beta_{2,i} = \beta_2 + \eta_{\beta_2 i}. \tag{35}$$

Here we consider $\phi \in \{0.5\}$, $\beta_1 = 3$ and $\beta_2 = 1$. For the design of heterogenous slopes, $\eta_{\phi i} \overset{i.i.d.}{\sim} U(-c, c)$, and

$$\eta_{\beta_\ell i} = \left(1 - \rho_\beta^2\right)^{1/2} \eta_{\phi i}. \tag{36}$$

Here, we set $c = 0.2$, $\rho_\beta = 0.4$ for $\ell = 1, 2$.

4.2 Dynamic heterogeneous panels data model with multi-factor error structure

This Monte Carlo simulation design same as Norkute et al. (2019). For convenience, we rewrite the data generating process as bellow

$$y_{i,t} = \alpha_i + \phi_i y_{i,t-1} + \sum_{\ell=1}^{k} \beta_{\ell i} x_{\ell i,t} + u_{i,t},\ for\ i = 1, \ldots N;\ t = -49, \ldots, T\,. \tag{37}$$

$$\tag{38}$$

We allow error factor structure in the model as

$$u_{i,t} = \sum_{s=1}^{m_y} \gamma_{si}^0 f_{s,t}^0 + \varepsilon_{i,t}, \tag{39}$$

where

$$f_{s,t}^0 = \rho_{s,t}^0 f_{s,t-1}^0 + \left(1 - \rho_{fs}^2\right)^{1/2} \zeta_{s,t}, \tag{40}$$

with $\zeta_{s,t} \overset{i.i.d.}{\sim} N(0, 1)$ for $s = 1, \ldots m_y$. We assume $k = 2$ and $m_y = 1 + k = 3$ and set $\rho_{s,t}^0 = 0.5$ for all s. The error term, $\varepsilon_{i,t}$, setting as

$$\varepsilon_{i,t} = \varsigma_\varepsilon \sigma_{it} \left(\epsilon_{it} - 1\right) / \sqrt{2}, \tag{41}$$

where $\epsilon_{it} \overset{i.i.d.}{\sim} \chi_1^2$, $\sigma_{it}^2 = \eta_i \varphi_t$, $\eta_i \overset{i.i.d.}{\sim} \chi_2^2/2$, and $\varphi_t = t/T\, for\ t = 0, \ldots, T$. And we set

$$\varsigma_\varepsilon = \frac{\pi_\mu}{1 - \pi_\mu} m_y. \tag{42}$$

we set $\pi_\mu \in \{3/4\}$.

The process of regressors is

$$x_{\ell it} = \mu_{\ell i} + + \sum_{\ell=1}^{k} \phi_{\ell i} x_{\ell i,t-1} + \sum_{s=1}^{m_x} \gamma_{\ell si}^0 f_{s,t}^0 + v_{\ell it},\ for\ i = 1, \ldots N;\ t = -49, \ldots, T\,;\ \ell = 1, 2. \tag{43}$$

We set number of factor, m_x, is 2. Therefore, $\boldsymbol{f}_{y,t}^0 = \left(f_{1t}^0, f_{2t}^0, f_{3t}^0\right)'$ and $\boldsymbol{f}_{x,t}^0 = \left(f_{1t}^0, f_{2t}^0\right)'$. We set

$$v_{\ell i,t} = \rho_{v,\ell} v_{\ell i,t-1} + \left(1 - \rho_{v,\ell}^2\right)^{\frac{1}{2}} \varpi_{\ell i,t}, for\ \ell = 1, 2, \tag{44}$$

where $\rho_{v,\ell} = 0.5$ for all ℓ. The individual effect is

$$\alpha_i^* \overset{i.i.d.}{\sim} N\left(0, (1-\rho_i)^2\right), \; \mu_{\ell i}^* = \rho_{\mu,\ell}\alpha_i^* + \left(1-\rho_{\mu,\ell}^2\right)^{1/2}\omega_{\ell i}, \tag{45}$$

where $\omega \overset{i.i.d.}{\sim} N\left(0, (1-\rho_i)^2\right)$ and $\rho_{\mu,\ell} = 0.5$.

Now, we define the factor loading in $u_{i,t}$ are generated as $\gamma_{si}^{0*} \overset{i.i.d.}{\sim} N(0,1)$, for $s = 1, \ldots, m_y = 3$, and the factor loading in x_{1it} and x_{2it} are drawn as

$$\begin{aligned} \gamma_{1si}^{0*} &= \rho_{\gamma,1s}\gamma_{3i}^{0*} + \left(1-\rho_{\gamma,1s}^2\right)^{1/2}\xi_{1si};\, \xi_{1si} \overset{i.i.d.}{\sim} N(0,1)\,; \\ \gamma_{2si}^{0*} &= \rho_{\gamma,2s}\gamma_{si}^{0*} + \left(1-\rho_{\gamma,2s}^2\right)^{1/2}\xi_{2si};\, \xi_{2si} \overset{i.i.d.}{\sim} N(0,1)\,; \end{aligned} \tag{46}$$

for $s = 1, \ldots, m_x = 2$. We set $\rho_{\gamma,11} = \rho_{\gamma,12} \in \{0.5\}$ and $\rho_{\gamma,21} = \rho_{\gamma,22} = 0.5$. The factor loading are generated as

$$\mathbf{\Gamma} = \mathbf{\Gamma}^0 + \mathbf{\Gamma}_i^{0*} \tag{47}$$

where

$$\mathbf{\Gamma}_i^0 = \begin{bmatrix} \gamma_{1i}^0 & \gamma_{11i}^0 & \gamma_{21i}^0 \\ \gamma_{2i}^0 & \gamma_{12i}^0 & \gamma_{22i}^0 \\ \gamma_{3i}^0 & 0 & 0 \end{bmatrix} \tag{48}$$

and

$$\mathbf{\Gamma}_i^{0*} = \begin{bmatrix} \gamma_{1i}^{0*} & \gamma_{11i}^{0*} & \gamma_{21i}^{0*} \\ \gamma_{2i}^{0*} & \gamma_{12i}^{0*} & \gamma_{22i}^{0*} \\ \gamma_{3i}^{0*} & 0 & 0 \end{bmatrix}. \tag{49}$$

We set

$$\mathbf{\Gamma}^0 = \begin{bmatrix} 1/4 & 1/4 & -1 \\ 1/2 & -1 & 1/4 \\ 1/2 & 0 & 0 \end{bmatrix}. \tag{50}$$

And

$$\alpha_i = \alpha + \alpha_i^*,\, \mu_{\ell i} = \mu_\ell + \mu_{\ell i}^*, \tag{51}$$

where $\alpha = 1/2$, $\mu_1 = 1$, $\mu_2 = -1/2$.

The slope coefficients are generated as

$$\phi_i = \phi + \eta_{\phi i},\;\; \beta_{1,i} = \beta_1 + \eta_{\beta_1 i}\, and\, \beta_{2,i} = \beta_2 + \eta_{\beta_2 i}. \tag{52}$$

Here we consider $\phi \in \{0.5\}$, $\beta_1 = 3$ and $\beta_2 = 1$. For the design of heterogenous slopes, $\eta_{\phi i} \overset{i.i.d.}{\sim} U(-c, c)$, and

$$\eta_{\beta_\ell i} = \left[(2c)^2/12\right]\rho_\beta \xi_{\beta \ell i} + \left(1-\rho_\beta^2\right)^{1/2}\eta_{\phi i}, \tag{53}$$

where

$$\xi_{\beta\ell i} = \frac{\bar{v}_{\ell i}^2 - \bar{v}_\ell^2}{\left[N^{-1}\sum_{i=1}^N \left(\bar{v}_{\ell i}^2 - \bar{v}_\ell^2\right)^2\right]^{1/2}}, \tag{54}$$

with $\bar{v_{elli}^2} = T^{-1} \sum_{t=1}^{T} v_{\ell it}^2$, $\bar{v_\ell^2} = N^{-1} \sum_{i=1}^{N} \bar{v_{\ell i}^2}$, for $\ell = 1, 2$. Here, we set $c = 0.2$, $\rho_\beta = 0.4$ for $\ell = 1, 2$. And

$$\varsigma_v^2 = \varsigma_\varepsilon^2 \left[SNR - \frac{\rho_v^2}{1 - \rho_v^2} \right] \left(\frac{\beta_1^2 + \beta_2^2}{1 - \rho_v^2} \right)^{-1}, \tag{55}$$

where $SNR = 4$. For the (T, N), we consider $T \in \{25, 50, 100, 200\}$ and $N \in \{25, 50, 100, 200\}$.

5 Monte Carlo simulation results

5.1 Dynamic Heterogeneous Panels without multifactor error structure

We consider ARDL(1,0) model.
$\phi \in \{0.5\}$.
$\beta_1 = 3$.
$\beta_2 = 1$.
$u_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,\ 1)$.
$\varpi_{\ell i,t} \overset{i.i.d.}{\sim} U(0.5, 1.5)$.
$\rho_{v,\ell} = 0.5$.
$c = 0.2$.
$\rho_\beta = 0.4$.
$T \in \{25, 50, 100, 200\}$.
$N \in \{25, 50, 100, 200\}$.

LSMG estimator is provided in sheet 1 of MC.xlsx file.
IVMG estimator is provided in sheet 2 of MC.xlsx file.

5.2 Dynamic Heterogeneous Panels with multifactor error structure

We consider ARDL(1,0) model.
$\phi \in \{0.5\}$.
$\beta_1 = 3$.
$\beta_2 = 1$.
$k = 2$.
$m_y = 1 + k = 3$.
$m_x = k = 2$.
$\zeta_{s,t} \overset{i.i.d.}{\sim} N(0, 1)$
$\pi_\mu \in \{3/4\}$.
$\rho_{s,t}^0 = 0.5$.
$\rho_{v,\ell} = 0.5$.
$\rho_{\mu,\ell} = 0.5$.
$\gamma_{si}^{0*} \overset{i.i.d.}{\sim} N(0, 1)$.

$\xi_{1si} \overset{i.i.d.}{\sim} N(0,1)$.
$\xi_{2si} \overset{i.i.d.}{\sim} N(0,1)$.
$\rho_{\gamma,11} = \rho_{\gamma,12} \in \{0.5\}$.
$\rho_{\gamma,21} = \rho_{\gamma,22} = 0.5$.
$\mathbf{\Gamma}^0 = \begin{bmatrix} 1/4 & 1/4 & -1 \\ 1/2 & -1 & 1/4 \\ 1/2 & 0 & 0 \end{bmatrix}$.
$\alpha = 1/2$.
$\mu_1 = 1$.
$\mu_2 = -1/2$.
$c = 0.2$.
$\rho_\beta = 0.4$.
$SNR = 4$.
$T \in \{25, 50, 100, 200\}$.
$N \in \{25, 50, 100, 200\}$.

IVMG estimator is provided in sheet 3 of MC.xlsx file.

6 Short summary

6.1 Dynamic Heterogeneous Panels without multifactor error structure

1. The performance of IVMG estimator is better than LSMG estimator in bias and RMSE.

6.2 Dynamic Heterogeneous Panels with multifactor error structure

1. When N and T increase, the performance of IVMG estimator is good in bias and RMSE.
Related literature to dynamic Heterogeneous Panels with multifactor error structure: Chudik and Pesaran (2015) and Norkute et al. (2019).
Related literature to choosing number of instruments: Donald and Newey (2001), Swanson (2005), Carrasco (2012), Bai and Ng (2010) and Kang (2019).

Appendix

A Asymptotic property of LS estimator

Based on heterogenous dynamic panel data model (1), we can obtain fixed effect estimator as

$$\hat{\boldsymbol{\theta}}_{LS,i} = \begin{pmatrix} \hat{\phi}_i \\ \hat{\beta}_i \end{pmatrix} = \begin{pmatrix} \frac{\sum_{t=1}^{T} \tilde{y}_{i,t-1}^2}{T} & \frac{\sum_{t=1}^{T} \tilde{y}_{i,t-1}\tilde{x}_{i,t}}{T} \\ \frac{\sum_{t=1}^{T} \tilde{x}_{i,t}\tilde{y}_{i,t-1}}{T} & \frac{\sum_{t=1}^{T} \tilde{x}_{i,t}^2}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sum_{t=1}^{T} \tilde{y}_{i,t-1}\tilde{y}_{i,t}}{T} \\ \frac{\sum_{t=1}^{T} \tilde{x}_{i,t}\tilde{y}_{i,t}}{T} \end{pmatrix}, \tag{56}$$

where $\tilde{y}_{i,t} = y_{i,t} - \bar{y}_i$, $\tilde{y}_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1}\iota_T$ and $\tilde{x}_{i,t} = x_{i,t} - \bar{x}_i$ with $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{i,t}$, $\bar{y}_{i,-1} = \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}$, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{i,t}$. Under equation (1), we have

$$\begin{pmatrix} \hat{\phi}_i - \phi_i \\ \hat{\beta}_i - \beta_i \end{pmatrix} = \begin{pmatrix} \frac{\sum_{t=1}^{T} \tilde{y}_{i,t-1}^2}{T} & \frac{\sum_{t=1}^{T} \tilde{y}_{i,t-1}\tilde{x}_{i,t}}{T} \\ \frac{\sum_{t=1}^{T} \tilde{x}_{i,t}\tilde{y}_{i,t-1}}{T} & \frac{\sum_{t=1}^{T} \tilde{x}_{i,t}^2}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sum_{t=1}^{T} \tilde{y}_{i,t-1}\tilde{u}_{i,t}}{T} \\ \frac{\sum_{t=1}^{T} \tilde{x}_{i,t}\tilde{u}_{i,t}}{T} \end{pmatrix}, \tag{57}$$

where $\tilde{u}_{i,t}$ is $u_{i,t} - \bar{u}_i$ with $\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{i,t}$.

Now, we can investigate asymptotic bias by taking the probability limit as

$$A_{\phi i}^{(1)} = \operatorname*{plim}_{T\to\infty} \left(\frac{\sum_{t=1}^{T} \tilde{y}_{i,t-1}\tilde{u}_{i,t}}{T} \right). \tag{58}$$

Then $A_i^{(1)}$ can be taken expectations as

$$\begin{aligned} A_{\phi i}^{(1)} &= E\left(y_{i,t-1} - \bar{y}_{i,-1}\right)\left(u_{i,t} - \bar{u}_i\right) \\ &= E\left(y_{i,t-1}u_{i,t}\right) - E\left(y_{i,t-1}\bar{u}_i\right) - E\left(\bar{y}_{i,-1}u_{i,t}\right) + E\left(\bar{y}_{i,-1}\bar{u}_i\right), \end{aligned} \tag{59}$$

where $E\left(y_{i,t-1}u_{i,t}\right) = 0$.

And we assume $y_{i,t}$ has started from a long time period in the past, so we have

$$y_{i,t} = \frac{\alpha_i}{(1-\phi_i)} + \sum_{s=0}^{\infty} \beta_i\phi_i^s x_{i,t-s} + \sum_{s=0}^{\infty} \phi_i^s u_{i,t-s}, \tag{60}$$

Then, we have

$$\begin{aligned} A_{\phi i}^{(1)} = &-E\left(\left(\sum_{s=0}^{\infty}\phi_i^s u_{i,t-s-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} u_{i,t}\right)\right) - E\left(\frac{u_{i,t}}{T}\sum_{t=1}^{T}\sum_{s=0}^{\infty}\phi_i^s u_{i,t-s-1}\right) + \\ &\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=0}^{\infty}\phi_i^s u_{i,t-s-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} u_{i,t}\right). \end{aligned} \tag{61}$$

Hence, from above equation, we have

$$
\begin{aligned}
A_{\phi i}^{(1)} =& -\frac{1}{T}E\left\{\left(u_{i,t-1}+u_{i,t-2}\phi_i^1+u_{i,t-3}\phi_i^2+\ldots\right)\left(u_{i,1}+\ldots+u_{i,t-1}+u_{i,t}+\ldots+u_{i,T}\right)\right\}- \\
& \frac{1}{T}E\left\{u_{i,t}\sum_{s=1}^{T}\left(u_{i,s-1}\phi_i^0+u_{i,s-2}\phi_i^1+\ldots+u_{i,s-t-1}\phi_i^t+\ldots+u_{i,s-T-1}\phi_i^T+\ldots\right)\right\}+ \\
& +\frac{1}{T}E\{\left(\sum_{s=1}^{T}u_{i,s-1}\phi_i^0+\sum_{s=1}^{T}u_{i,s-2}\phi_i^1+\ldots+\sum_{s=1}^{T}u_{i,s-t-1}\phi_i^t+\ldots+\sum_{s=1}^{T}u_{is-T-1}\phi_i^T+\ldots\right) \\
& \left(\frac{1}{T}\sum_{s=1}^{T}u_{i,s}\right)\} \\
& =-\frac{\sigma_u^2}{T}\frac{\left(1-\phi_i^{t-1}\right)}{1-\phi_i}-\frac{\sigma_u^2}{T}\frac{\left(1-\phi_i^{T-t}\right)}{\left(1-\phi_i\right)}+\frac{\sigma_u^2}{T}\left(\frac{1}{1-\phi_i}-\frac{1}{T}\frac{\left(1-\phi_i^T\right)}{\left(1-\phi_i\right)^2}\right) \\
& =-\frac{\sigma_u^2}{T(1-\phi_i)}\left(1-\phi_i^{t-1}-\phi_i^{T-t}+\frac{1}{T}\frac{\left(1-\phi_i^T\right)}{\left(1-\phi_i\right)}\right).
\end{aligned}
\tag{62}
$$

Therefore, we can see the bias of $\hat{\phi}_i$ is $O\left(T^{-1}\right)$.

To be more compact ,we can rewrite the model as,

$$\tilde{\boldsymbol{y}}_i=\tilde{\boldsymbol{W}}_i\boldsymbol{\theta}_i+\tilde{\boldsymbol{u}}_i, \tag{63}$$

where $\tilde{\boldsymbol{y}}_i=\left(\tilde{\boldsymbol{y}}_{i,1},\ldots,\tilde{\boldsymbol{y}}_{i,T}\right)^{'}$ is $T\times 1$ vector, $\tilde{\boldsymbol{W}}_i=\left(\tilde{\boldsymbol{w}}_{i,1},\ldots,\tilde{\boldsymbol{w}}_{i,T}\right)^{'}$ is $T\times 2$ matrix and $\tilde{\boldsymbol{u}}_i=\left(\tilde{\boldsymbol{u}}_{i,1},\ldots,\tilde{\boldsymbol{u}}_{i,T}\right)$ is $T\times 1$ vector with $\tilde{\boldsymbol{w}}_{i,t}=\left(y_{i,t-1}-\bar{y}_{i,-1},x_{i,t}-\bar{x}_i\right)^{'}$, for $t=1,\ldots,T$ and $i=1,\ldots,N$.

Also, we define our interested parameter as

$$\left(\phi_i,\beta_i\right)^{'}=\boldsymbol{\theta}_i=\boldsymbol{\theta}+\boldsymbol{\lambda}_i, \tag{64}$$

where $\boldsymbol{\lambda}_i\overset{i.i.d.}{\sim}\left(\mathbf{0},\boldsymbol{\Sigma}_\lambda\right)$. The lest square estimator, $\hat{\boldsymbol{\theta}}_{LS,i}$, can be expressed as

$$\hat{\boldsymbol{\theta}}_{LS,i}=\left(\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{W}}_i}{T}\right)^{-1}\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{y}}_i}{T}. \tag{65}$$

From above discussion and assumptions, we have following theorem

Theorem 1

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{IV,i}-\boldsymbol{\theta}_i\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{Q}_i^{-1}\boldsymbol{\Sigma}_{LS,i}\boldsymbol{Q}_i^{-1}\right), \tag{66}$$

where $\boldsymbol{\Sigma}_{LS,i}=\mathrm{plim}_{T\to\infty}T^{-1}\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{u}}_i\tilde{\boldsymbol{u}}_i^{'}\tilde{\boldsymbol{W}}_i$ *and* $\boldsymbol{Q}_i=\mathrm{plim}_{T\to\infty}T^{-1}\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{W}}_i$

A.0.1 Mean group LS estimator

Now, we define the mean group estimator of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_{LSMG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{LSi}. \tag{67}$$

And we can show that the asymptotic property of $\hat{\boldsymbol{\theta}}_{LSMG}$, as

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{LSMG} &= N^{-1}\sum_{i=1}^{N}\left(\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{W}}_i}{T}\right)^{-1}\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{y}}_i}{T} \\ &= \bar{\boldsymbol{\theta}} + N^{-1}\sum_{i=1}^{N}\left(\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{W}}_i}{T}\right)^{-1}\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{u}}_i}{T},\end{aligned} \tag{68}$$

where $\bar{\boldsymbol{\theta}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{\theta}_i$. For fixed N and large T, we have

$$\operatorname*{plim}_{T\to\infty}\hat{\boldsymbol{\theta}}_{LSMG} = \bar{\boldsymbol{\theta}} + N^{-1}\sum_{i=1}^{N}\operatorname*{plim}_{T\to\infty}\left(\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{W}}_i}{T}\right)^{-1}\operatorname*{plim}_{T\to\infty}\left(\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{u}}_i}{T}\right) \tag{69}$$

Then, from section 1.1, we know that $\operatorname{plim}_{T\to\infty}\left(\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{u}}_i}{T}\right) = O_p(1)$. Thus, we can obtain

$$\operatorname*{plim}_{T\to\infty}\hat{\boldsymbol{\theta}}_{LSMG} = \bar{\boldsymbol{\theta}}. \tag{70}$$

When $N\to\infty$ and $T\to\infty$ and by the law of large numbers, we can see that

$$\operatorname*{plim}_{T\to\infty, N\to\infty}\hat{\boldsymbol{\theta}}_{LSMG} = \boldsymbol{\theta}. \tag{71}$$

And the variance estimator of $\hat{\boldsymbol{\theta}}_{LSMG}$ is given by

$$\hat{\boldsymbol{\Sigma}}_{LS,\lambda} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG}\right)\left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG}\right)^{'}. \tag{72}$$

We show that $\hat{\boldsymbol{\Sigma}}_{LS,\lambda}$ is consistent when $N\to\infty$ and $T\to\infty$.

$$\begin{aligned}&\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG}\right)\left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG}\right)^{'} = \sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i} - E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)\right)\left(\hat{\boldsymbol{\theta}}_{LS,i} - E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)\right)^{'} \\ &+\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LSMG} - E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)\right)\left(\hat{\boldsymbol{\theta}}_{LSMG} - E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)\right)^{'} \\ &-\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LSMG} - E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)\right)\left(\hat{\boldsymbol{\theta}}_{LS,i} - E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)\right)^{'} \\ &-\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i} - E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)\right)\left(\hat{\boldsymbol{\theta}}_{LSMG} - E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)\right)^{'}.\end{aligned} \tag{73}$$

Taking expectation on equation (73), we have

$$E\left(\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i}-\hat{\boldsymbol{\theta}}_{LSMG}\right)\left(\hat{\boldsymbol{\theta}}_{LS,i}-\hat{\boldsymbol{\theta}}_{LSMG}\right)^{'}\right)=\sum_{i=1}^{N}Var\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)+\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{LSMG}\hat{\boldsymbol{\theta}}_{LSMG}^{'}+$$
$$\sum_{i=1}^{N}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)^{'}-\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{LSMG}\hat{\boldsymbol{\theta}}_{LS,i}^{'}+\sum_{i=1}^{N}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)\hat{\boldsymbol{\theta}}_{LS,i}^{'}-\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{LSMG}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)^{'}-$$
$$\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{LS,i}\hat{\boldsymbol{\theta}}_{LSMG}^{'}+\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{LS,i}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)^{'}-\sum_{i=1}^{N}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)^{'}=$$
$$\sum_{i=1}^{N}Var\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)-NE\left(\hat{\boldsymbol{\theta}}_{LSMG}\hat{\boldsymbol{\theta}}_{LSMG}^{'}\right)+\sum_{i=1}^{N}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)^{'} \tag{74}$$

and

$$E\left(\hat{\boldsymbol{\theta}}_{LSMG}\hat{\boldsymbol{\theta}}_{LSMG}^{'}\right)=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\hat{\boldsymbol{\theta}}_{LS,i}^{'}\right). \tag{75}$$

And, we also have

$$E\left(\hat{\boldsymbol{\theta}}_{LSMG}\hat{\boldsymbol{\theta}}_{LSMG}^{'}\right)=\frac{1}{N^2}\left(\sum_{i=1}^{N}Var\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)+\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\hat{\boldsymbol{\theta}}_{LS,i}^{'}\right)\right). \tag{76}$$

Then,

$$\begin{aligned}&E\left(\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i}-\hat{\boldsymbol{\theta}}_{LSMG}\right)\left(\hat{\boldsymbol{\theta}}_{LS,i}-\hat{\boldsymbol{\theta}}_{LSMG}\right)^{'}\right)=\\&\left(1-\frac{1}{N}\right)\sum_{i=1}^{N}Var\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)+\sum_{i=1}^{N}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\hat{\boldsymbol{\theta}}_{LS,i}^{'}\right)-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\hat{\boldsymbol{\theta}}_{LS,i}^{'}\right).\end{aligned} \tag{77}$$

From above equation (77), we can observe the bias term as,

$$\aleph=\sum_{i=1}^{N}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\hat{\boldsymbol{\theta}}_{LS,i}^{'}\right)-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\hat{\boldsymbol{\theta}}_{LS,i}\hat{\boldsymbol{\theta}}_{LS,i}^{'}\right). \tag{78}$$

Taking expectation on equation (26), we have

$$E\left(\hat{\boldsymbol{\theta}}_{LS,i}\right)=\boldsymbol{\theta}+E\left(b_i\right), \tag{79}$$

From equation (78), we know that

$$\aleph=\sum_{i=1}^{N}E\left(b_i\right)E\left(b_i^{'}\right)-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(b_i\right)E\left(b_j^{'}\right). \tag{80}$$

When $T\rightarrow\infty$, $\aleph=0$. Therefore, we have following theorem

Theorem 2 *When* $(T, N) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$,

1.

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{LSMG} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{LS,\lambda}\right). \tag{81}$$

2.

$$\hat{\boldsymbol{\Sigma}}_{LS,\lambda} \xrightarrow{p} \boldsymbol{\Sigma}_{LS,\lambda} \tag{82}$$

where

$$\hat{\boldsymbol{\Sigma}}_{LS,\lambda} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG}\right)\left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG}\right)'. \tag{83}$$

References

Bai, J. and S. Ng (2010). Instrumental variable estimation in a data rich enviroment. *Econometric Theory 26*, 1577–1606.

Carrasco, M. (2012). A regularization approach to the many instruments problem. *Journal of Econometrics 170*, 383–398.

Chudik, A. and M. H. Pesaran (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics 188*, 393–420.

Donald, S. G. and W. K. Newey (2001). Choosing the number of instruments. *Econometrica 69*, 1161–1191.

Hayakawa, K. (2009). A simple efficient instrumental variable estimator in panel ar(p) models when both n and t are large. *Econometric Theory 25*, 873–890.

Hayakawa, K., M. Qi, and J. Breitung (2019). Double filter instrumental variable estimation of panel data models with weakly exogenous variables. *Econometric Reviews 38*, 1055–1088.

Kang, B. (2019). Higher order approximation of iv estimators with invalid instruments.

Kuersteiner, G. and R. Okui (2010). Constructing optimal instruments by first-stage prediction averaging. *Econometrica 78*, 697–718.

Lee, S. and Y. Shin (2019). Complete subset averaging with many instruments.

Moon, H. R. and P. C. Phillips (2000). Estimation of autoregressive roots near unity using panel data. *Econometric Theory 16*, 927–997.

Norkute, M., V. Sarafidis, T. Yamagata, and G. Cui (2019). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. ISER Discussion Paper No. 1019.

Okui, R. (2011). Instrumental variable estimation in the presence of many moment conditions. *Journal of Econometrics 165*, 70–86.

Swanson, J. C. C. N. R. (2005). Consistent estimation with a large number of weak instruments. *Econometrica 73*, 1673–1692.