
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, $b=1$	2
1.1.2	Trend model, $b=2$	2
1.2	IV estimation method and asymptotic property	3
1.2.1	Mean group IV estimator	4
2	Construction optimal instruments	5
3	Monte Carlo simulation design	7
3.1	dynamic heterogeneous panels data model without multi-factor error structure	7
A	Asymptotic property of LS estimator	8
A.0.1	Mean group LS estimator	10
	Reference	12

1 Model, asymptotic property of IV estimator

1.1 The models and Assumptions

1.1.1 Fixed effects model, b=1

Consider the dynamic heterogeneous panels data model with fixed effects:

$$\begin{aligned} y_{i,t}^b &= \phi_i y_{i,t-1}^b + \mathbf{x}_{i,t}' \boldsymbol{\beta}_i + \alpha_i + u_{i,t}^b \\ &= \mathbf{w}_{i,t}^{b'} \boldsymbol{\theta}_i + \alpha_i + u_{i,t}^b, \quad \text{for } i = 1, \dots, N; t = 1, \dots, T, \end{aligned} \quad (1)$$

where $\mathbf{x}_{i,t}$ and $\boldsymbol{\beta}_i$ are $k \times 1$ vectors, $\boldsymbol{\theta}_i = (\phi_i, \boldsymbol{\beta}_i')'$ and $\mathbf{w}_{i,t}^b = (y_{i,t-1}^b, \mathbf{x}_{i,t}')'$ are $(1+k) \times 1$ vectors, and $b = 1$ corresponds to the fixed effects model. Stacking the T observations for each i , we have

$$\mathbf{y}_i^b = \mathbf{W}_i^b \boldsymbol{\theta}_i + \alpha_i \boldsymbol{\iota}_T + \mathbf{u}_i^b, \quad (2)$$

where $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,T})'$ is a $T \times 1$ vector, $\mathbf{W}_i^b = (\mathbf{w}_{i,1}^b, \dots, \mathbf{w}_{i,T}^b)'$ is a $T \times (1+k)$ matrix, $\boldsymbol{\iota}_T = (1, \dots, 1)'$ and $\mathbf{u}_i^b = (u_{i,1}^b, \dots, u_{i,T}^b)'$ are $T \times 1$ vectors.

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 $(T-1) \times T$ forward demeaning matrix as

$$\mathbf{F}^b = \text{diag}(c_1^b, c_2^b, \dots, c_{T-1}^b) \begin{bmatrix} 1 & \frac{-1}{T-1} & \cdots & \cdots & \frac{-1}{T-1} \\ \vdots & 1 & \frac{-1}{T-2} & \cdots & \frac{-1}{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}, \quad (3)$$

where $c_t^b = \sqrt{(T-t)(T-t+1)}$, and $b = 1$ corresponds to the forward demeaning matrix under the fixed effects model. Premultiplying the model (2) by \mathbf{F}^b , the model can be expressed as

$$\tilde{\mathbf{y}}_i^b = \tilde{\mathbf{W}}_i^b \boldsymbol{\theta}_i + \tilde{\mathbf{u}}_i^b, \quad (4)$$

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

1.1.2 Trend model, b=2

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

$$\begin{aligned} y_{i,t}^b &= \phi_i y_{i,t-1}^b + \mathbf{x}_{i,t}' \boldsymbol{\beta}_i + \alpha_i + \eta_i t + u_{i,t}^b \\ &= \mathbf{w}_{i,t}^{b'} \boldsymbol{\theta}_i + \alpha_i + \eta_i t + u_{i,t}^b, \quad \text{for } i = 1, \dots, N; t = 1, \dots, T, \end{aligned} \quad (5)$$

where $\mathbf{x}_{i,t}$ and β_i are $k \times 1$ vectors, $\theta_i = (\phi_i, \beta_i')'$ and $\mathbf{w}_{i,t}^b = (y_{i,t-1}^b, \mathbf{x}_{i,t}')'$ are $(1+k) \times 1$ vectors, and $b = 2$ corresponds to the trend model. Stacking the T observations for each i , we have

$$\mathbf{y}_i^b = \mathbf{W}_i^b \theta_i + \alpha_i \mathbf{1}_T + \eta_i \boldsymbol{\tau}_T + \mathbf{u}_i^b, \quad (6)$$

where $\boldsymbol{\tau}_T = (1, 2, \dots, T)'$. Then, we define the forward demeaning matrix to the trend model as

$$\mathbf{F}^b = \text{diag}(c_1^b, c_2^b, \dots, c_{T-2}^b) \begin{bmatrix} 1 & \frac{2(-2(T-2))}{(T-1)(T-2)} & \frac{2(-2(T-2)+3)}{(T-1)(T-2)} & \dots & \frac{2(-2(T-2)+3(T-2))}{(T-1)(T-2)} \\ 0 & 1 & \frac{2(-2(T-3))}{(T-2)(T-3)} & \dots & \frac{2(-2(T-3)+3(T-3))}{(T-3)(T-4)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \frac{2(-2+3)}{2 \cdot 1} \end{bmatrix}, \quad (7)$$

where $c_t^b = ((T-t)(T-t-1)/(T-t-1)(T-t+2))^{1/2}$, and $b = 2$ corresponds to the forward demeaning matrix under the trend model.

Multiply \mathbf{F}^b to model (6), the model can be expressed as

$$\tilde{\mathbf{y}}_i^b = \tilde{\mathbf{W}}_i^b \theta_i + \tilde{\mathbf{u}}_i^b, \quad (8)$$

where $\tilde{\mathbf{y}}_i^b = \mathbf{F}^b \mathbf{y}_i^b = (\tilde{y}_{i,1}^b, \dots, \tilde{y}_{i,T-2}^b)'$, $\tilde{\mathbf{W}}_i^b = \mathbf{F}^b \mathbf{W}_i^b$ and $\tilde{\mathbf{u}}_i^b = \mathbf{F}^b \mathbf{u}_i^b$, and $b = 2$.

Assumption 1 $\mathbf{x}_{i,t}$ and $u_{i,t}^b$ are independently distributed for all t, i and b .

Assumption 2 $\theta_i = \theta + \lambda_i$, $\lambda_i \stackrel{i.i.d.}{\sim} (\mathbf{0}, \Sigma_\lambda)$, where Σ_λ is a fixed positive definite matrix.

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 \mathbf{X}_i as instruments, as

$$\mathbf{Z}_i = (\mathbf{X}_i, \mathbf{X}_{i,-1}), \quad (9)$$

where \mathbf{Z}_i is a $T \times 2k$ vector with $\mathbf{X}_i = (\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,T})'$ and $\mathbf{X}_{i,-1} = (\mathbf{x}_{i,0}, \dots, \mathbf{x}_{i,T-1})'$.

Assumption 3 $\mathbf{A}_i^b = \text{plim}_{T \rightarrow \infty} \tilde{\mathbf{A}}_{i,T}^b$ has full column rank, $\mathbf{B}_i = \text{plim}_{T \rightarrow \infty} \mathbf{B}_{i,T}$ and $\Sigma_i^b = \text{plim}_{T \rightarrow \infty} T^{-1} \mathbf{Z}_i' \tilde{\mathbf{u}}_i^b \tilde{\mathbf{u}}_i^{b'} \mathbf{Z}_i$ has positive definite, uniformly, where $\tilde{\mathbf{A}}_{i,T}^b = \frac{1}{T} \mathbf{Z}_i' \tilde{\mathbf{W}}_i^b$ and $\mathbf{B}_{i,T} = \frac{1}{T} \mathbf{Z}_i' \mathbf{Z}_i$ with $b = 1$ corresponds to the fixed effects model while $b = 2$ corresponds to the trend models.

Then, the IV estimator can be expressed as

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

where

$$\tilde{\mathbf{g}}_{i,T}^b = \frac{1}{T} \mathbf{Z}_i' \tilde{\mathbf{y}}_i^b. \quad (11)$$

From above equation, we have

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

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

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

$$\begin{aligned} T^{-1/2} \mathbf{Z}_i' \tilde{\mathbf{u}}_i^b &= O_p(1) \\ \text{and} \\ T^{-1/2} \mathbf{Z}_i' \tilde{\mathbf{u}}_i^b &\xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_i^b). \end{aligned} \quad (13)$$

Thus, IV estimator, $\boldsymbol{\theta}_i^b$ 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 $(N, T) \xrightarrow{j} \infty$ such that $N/T \rightarrow c$ with $0 < c < \infty$. for each i ,

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

1.2.1 Mean group IV estimator

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

$$\hat{\boldsymbol{\theta}}_{IVMG}^b = \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\theta}}_i^b. \quad (15)$$

From Assumption (2), we can show that the asymptotic property of $\hat{\boldsymbol{\theta}}_{IVMG}^b$, as

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

where the first of right hand side

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i^b - \boldsymbol{\theta}_i \right) &= \frac{1}{N} \sum_{i=1}^N \left(\tilde{\mathbf{A}}_{i,T}^{b'} \mathbf{B}_{i,T}^{-1} \tilde{\mathbf{A}}_{i,T}^b \right)^{-1} \tilde{\mathbf{A}}_{i,T}^{b'} \mathbf{B}_{i,T}^{-1} \left(T^{-1/2} \mathbf{Z}_i' \tilde{\mathbf{u}}_i^b \right) \\ &= o_p(1). \end{aligned} \quad (17)$$

Then, we can see

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

As $N \rightarrow \infty$, we can see

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i \xrightarrow{d} N(\mathbf{0}, \Sigma_\lambda). \quad (19)$$

Therefore, we know that $\hat{\theta}_{IVMG}^b$ is \sqrt{N} consistent.

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

$$\hat{\Sigma}_\lambda^b = \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\theta}_i^b - \hat{\theta}_{IVMG}^b \right) \left(\hat{\theta}_i^b - \hat{\theta}_{IVMG}^b \right)' . \quad (20)$$

Follow [Norkute et al. \(2019\)](#), we can show that $\hat{\Sigma}_\lambda^b$ is consistent and it does not have small T bias. Firstly, we decompose (20) as

$$\begin{aligned} & \sum_{i=1}^N \left(\hat{\theta}_i^b - \theta + \theta - \hat{\theta}_{IVMG}^b \right) \left(\hat{\theta}_i^b - \theta + \theta - \hat{\theta}_{IVMG}^b \right)' = \\ & \sum_{i=1}^N \lambda_i \lambda_i' + \sum_{i=1}^N \left(\hat{\theta}_i^b - \theta_i \right) \left(\hat{\theta}_i^b - \theta_i \right)' + \sum_{i=1}^N \left(\hat{\theta}_i^b - \theta_i \right) \lambda_i + \sum_{i=1}^N \lambda_i \left(\hat{\theta}_i^b - \theta_i \right)' - \\ & N \left(\theta - \hat{\theta}_{IVMG}^b \right)' \left(\theta - \hat{\theta}_{IVMG}^b \right). \end{aligned} \quad (21)$$

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

$$\begin{aligned} \hat{\Sigma}_\lambda^b - \Sigma_\lambda^b &= \frac{1}{N-1} \sum_{i=1}^N \left(\lambda_i \lambda_i' - \Sigma_\lambda^b \right) + \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\theta}_i^b - \theta_i \right) \left(\hat{\theta}_i^b - \theta_i \right)' \\ &+ \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\theta}_i^b - \theta_i \right) \lambda_i + \frac{1}{N-1} \sum_{i=1}^N \lambda_i \left(\hat{\theta}_i^b - \theta_i \right)' - \\ &\frac{N}{N-1} \left(\theta - \hat{\theta}_{IVMG}^b \right)' \left(\theta - \hat{\theta}_{IVMG}^b \right) = o_p(1). \end{aligned} \quad (22)$$

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

$$\begin{aligned} \sqrt{N} \left(\hat{\theta}_{IVMG}^b - \theta \right) &\xrightarrow{d} N(\mathbf{0}, \Sigma_\lambda^b); \\ \hat{\Sigma}_\lambda^b &\xrightarrow{p} \Sigma_\lambda^b. \end{aligned} \quad (23)$$

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

$$\mathbf{Z}_i^{(j)} = (\mathbf{X}_i, \mathbf{X}_{i,-1}, \dots, \mathbf{X}_{i,-j}), \quad (24)$$

where $\mathbf{Z}_i^{(j)}$ is $T \times (j+1)k$ matrix with $1 \leq j \leq J$ with J is the maximum lags of variables that we can observe.

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

$$\hat{\boldsymbol{\theta}}_i^b = \left(\tilde{\mathbf{W}}_i^{b'} \mathbf{P}_i^{(j)} \tilde{\mathbf{W}}_i^b \right)^{-1} \tilde{\mathbf{W}}_i^{b'} \mathbf{P}_i^{(j)} \tilde{\mathbf{y}}_i^b, \quad (25)$$

where $\mathbf{P}_i^{(j)} = \mathbf{Z}_i^{(j)} \left(\mathbf{Z}_i^{(j)'} \mathbf{Z}_i^{(j)} \right)^{-1} \mathbf{Z}_i^{(j)'}$.

If we can observe long lagged length from data, we can use more instruments. 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

$$\boldsymbol{\omega}^{(i)} = \left(\omega_1^{(i)}, \dots, \omega_J^{(i)} \right)', \quad (26)$$

where $\boldsymbol{\omega}^{(i)}$ is a $J \times 1$ vector and $\sum_{j=1}^J \omega_j^{(i)} = 1$. Then, we can weight $\mathbf{P}_i^{(j)}$ as

$$\mathbf{P}_i = \sum_{j=1}^J \omega_j^{(i)} \mathbf{P}_i^{(j)}, \quad (27)$$

where J is maximum number of lagged variables that we can observed. Our goal is to select $\boldsymbol{\omega}^{(i)}$ to minimize the approximate mean square error, $S_{\eta^{(i)}}^b(\boldsymbol{\omega}^{(i)})$, where $\boldsymbol{\eta}^{(i)}$ is a $(1+k) \times 1$ fixed parameters vector. In beginning, we consider positive weights, such that $\omega_j^{(i)} \in [0, 1]$. And we define $\mathbf{D}_i = (\mathbf{w}_{i,1}^b, \dots, \mathbf{w}_{i,T}^b)'$ is a $T \times (1+k)$ matrix. Let $\hat{\mathbf{H}}_i$ is some estimator of $\mathbf{H}_i = \frac{\mathbf{D}_i' \mathbf{D}_i}{T}$. Let $\hat{\boldsymbol{\theta}}_i^b$ is some preliminary estimator, and define the residuals $\hat{\mathbf{u}}_i^b = \tilde{\mathbf{y}}_i^b - \tilde{\mathbf{W}}_i^b \hat{\boldsymbol{\theta}}_i^b$ that does not depend on the weighting vector. Let $T \times (1+k)$ matrix, $\tilde{\mathbf{V}}_i^b$, be some preliminary residual from first stage regression, and $\tilde{\mathbf{v}}_{\eta,i}^b = \tilde{\mathbf{V}}_i^b \hat{\mathbf{H}}_i^{-1} \boldsymbol{\eta}^{(i)}$.

Define

$$\hat{\sigma}_{u,i}^{b2} = \frac{\hat{\mathbf{u}}_i^{b'} \hat{\mathbf{u}}_i^b}{T}, \quad \hat{\sigma}_{\eta,i}^{b2} = \frac{\tilde{\mathbf{v}}_{\eta,i}^{b'} \tilde{\mathbf{v}}_{\eta,i}^b}{T}, \quad \hat{\sigma}_{\eta u,i}^b = \frac{\tilde{\mathbf{v}}_{\eta,i}^{b'} \hat{\mathbf{u}}_i^b}{T}. \quad (28)$$

Let $\hat{\mathbf{v}}_{\eta,i}^{(j)} = \left(\mathbf{P}_i^{(j)} - \mathbf{P}_i \right) \tilde{\mathbf{W}}_i^b \hat{\mathbf{H}}_i^{-1} \boldsymbol{\eta}^{(i)}$ is a $T \times 1$ vector, and $\hat{\mathbf{U}}_i = \left(\hat{\mathbf{v}}_{\eta,i}^{(1)}, \dots, \hat{\mathbf{v}}_{\eta,i}^{(J)} \right)' \left(\hat{\mathbf{v}}_{\eta,i}^{(1)}, \dots, \hat{\mathbf{v}}_{\eta,i}^{(J)} \right)$ is $J \times J$ matrix. Define $\boldsymbol{\Gamma}^{(i)}$ be the $J \times J$ matrix whose (j', j) element is $\min(j', j)$ and let $K = (1, \dots, J)'$. The criterion $\hat{S}_{\eta^{(i)}}^b(\boldsymbol{\omega}^{(i)})$ is

$$\hat{S}_{\eta^{(i)}}^b(\boldsymbol{\omega}^{(i)}) = \hat{\sigma}_{\eta u,i}^{b2} \frac{(K' \boldsymbol{\omega}^{(i)})^2}{T} + \hat{\sigma}_{u,i}^{b2} \frac{\boldsymbol{\omega}^{(i)'} \hat{\mathbf{U}}_i \boldsymbol{\omega}^{(i)} - \hat{\sigma}_{\eta,i}^{b2} (J - 2K' \boldsymbol{\omega}^{(i)} + \boldsymbol{\omega}^{(i)'} \boldsymbol{\Gamma}_i \boldsymbol{\omega}^{(i)})}{T}. \quad (29)$$

Then, we can find the optimal weight, $\omega^{(i)*}$, by minimize $\hat{S}_{\eta^{(i)}}^b(\omega^{(i)})$.

Therefore, we have mean average 2SLS estimator, $\hat{\theta}_i^b$, as

$$\hat{\theta}_i^{*b} = \left(\tilde{W}_i^{b'} P_i(\omega^{(i)*}) \tilde{W}_i^b \right)^{-1} \tilde{W}_i^{b'} P_i(\omega^{(i)*}) \tilde{y}_i^b. \quad (30)$$

3 Monte Carlo simulation design

3.1 dynamic heterogeneous panels data model without multi-factor error structure

The data generating process:

$$\begin{aligned} y_{i,t} &= \phi_i y_{i,t-1} + \beta_{1i} x_{1i,t} + \alpha_i + u_{i,t}, \text{ for } i = 1, \dots, N; t = -49 - (J + 1), \dots, T, \\ x_{1i,t} &= \rho x_{1i,t-1} + \tau_\alpha \alpha_i + \theta_u u_{i,t-1} + v_{i,t}, \end{aligned} \quad (31)$$

where $u_{i,t} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, and $v_{i,t} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $\alpha_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $\rho = 0.5$, $\theta_u = -0.2$, $\tau_\alpha = 0.5$, $J = 4$.

The slope coefficients are generated as

$$\phi_i = \phi + \eta_{\phi i}; \quad \beta_{1,i} = \beta_1 + \eta_{\beta 1 i}. \quad (32)$$

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

$$\eta_{\beta 1 i} = (1 - \rho_\beta^2)^{1/2} \eta_{\phi i}. \quad (33)$$

Here, we set $c = 0.2$, $\rho_\beta = 0.4$.

The first 50 observations are discarded. For the (T, N) , we consider $T \in \{25, 50, 100, 200\}$ and $N \in \{25, 50, 100, 200\}$.

Appendix

A Asymptotic property of LS estimator

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

$$\hat{\theta}_{LS,i} = \begin{pmatrix} \hat{\phi}_i \\ \hat{\beta}_i \end{pmatrix} = \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1}^2}{T} & \frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{x}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t} \tilde{y}_{i,t-1}}{T} & \frac{\sum_{t=1}^T \tilde{x}_{i,t}^2}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{y}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t} \tilde{y}_{i,t}}{T} \end{pmatrix}, \quad (34)$$

where $\tilde{y}_{i,t} = y_{i,t} - \bar{y}_i$, $\tilde{y}_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1}$ 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

$$\begin{pmatrix} \hat{\phi}_i - \phi_i \\ \hat{\beta}_i - \beta_i \end{pmatrix} = \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1}^2}{T} & \frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{x}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t} \tilde{y}_{i,t-1}}{T} & \frac{\sum_{t=1}^T \tilde{x}_{i,t}^2}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{u}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t} \tilde{u}_{i,t}}{T} \end{pmatrix}, \quad (35)$$

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

$$A_{\phi i}^{(1)} = \text{plim}_{T \rightarrow \infty} \left(\frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{u}_{i,t}}{T} \right). \quad (36)$$

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

$$\begin{aligned} A_{\phi i}^{(1)} &= E(y_{i,t-1} - \bar{y}_{i,-1})(u_{i,t} - \bar{u}_i) \\ &= E(y_{i,t-1} u_{i,t}) - E(y_{i,t-1} \bar{u}_i) - E(\bar{y}_{i,-1} u_{i,t}) + E(\bar{y}_{i,-1} \bar{u}_i), \end{aligned} \quad (37)$$

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

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

$$y_{i,t} = \frac{\alpha_i}{(1 - \phi_i)} + \sum_{s=0}^{\infty} \beta_i \phi_i^s x_{i,t-s} + \sum_{s=0}^{\infty} \phi_i^s u_{i,t-s}, \quad (38)$$

Then, we have

$$\begin{aligned} A_{\phi i}^{(1)} &= -E \left(\left(\sum_{s=0}^{\infty} \phi_i^s u_{i,t-s-1} \right) \left(\frac{1}{T} \sum_{t=1}^T u_{i,t} \right) \right) - E \left(\frac{u_{i,t}}{T} \sum_{t=1}^T \sum_{s=0}^{\infty} \phi_i^s u_{i,t-s-1} \right) + \\ &\quad \left(\frac{1}{T} \sum_{t=1}^T \sum_{s=0}^{\infty} \phi_i^s u_{i,t-s-1} \right) \left(\frac{1}{T} \sum_{t=1}^T u_{i,t} \right). \end{aligned} \quad (39)$$

Hence, from above equation, we have

$$\begin{aligned}
A_{\phi_i}^{(1)} &= -\frac{1}{T}E \left\{ (u_{i,t-1} + u_{i,t-2}\phi_i^1 + u_{i,t-3}\phi_i^2 + \dots) (u_{i,1} + \dots + u_{i,t-1} + u_{i,t} + \dots + u_{i,T}) \right\} - \\
&\quad \frac{1}{T}E \left\{ u_{i,t} \sum_{s=1}^T (u_{i,s-1}\phi_i^0 + u_{i,s-2}\phi_i^1 + \dots + u_{i,s-t-1}\phi_i^t + \dots + u_{i,s-T-1}\phi_i^T + \dots) \right\} + \\
&\quad + \frac{1}{T}E \left\{ \left(\sum_{s=1}^T u_{i,s-1}\phi_i^0 + \sum_{s=1}^T u_{i,s-2}\phi_i^1 + \dots + \sum_{s=1}^T u_{i,s-t-1}\phi_i^t + \dots + \sum_{s=1}^T u_{i,s-T-1}\phi_i^T + \dots \right) \right. \\
&\quad \left. \left(\frac{1}{T} \sum_{s=1}^T u_{i,s} \right) \right\} \\
&= -\frac{\sigma_u^2}{T} \frac{(1 - \phi_i^{t-1})}{1 - \phi_i} - \frac{\sigma_u^2}{T} \frac{(1 - \phi_i^{T-t})}{(1 - \phi_i)} + \frac{\sigma_u^2}{T} \left(\frac{1}{1 - \phi_i} - \frac{1}{T} \frac{(1 - \phi_i^T)}{(1 - \phi_i)^2} \right) \\
&= -\frac{\sigma_u^2}{T(1 - \phi_i)} \left(1 - \phi_i^{t-1} - \phi_i^{T-t} + \frac{1}{T} \frac{(1 - \phi_i^T)}{(1 - \phi_i)} \right).
\end{aligned} \tag{40}$$

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

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

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{W}}_i \boldsymbol{\theta}_i + \tilde{\mathbf{u}}_i, \tag{41}$$

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

Also, we define our interested parameter as

$$(\phi_i, \beta_i)' = \boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{\lambda}_i, \tag{42}$$

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

$$\hat{\boldsymbol{\theta}}_{LS,i} = \left(\frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{W}}_i}{T} \right)^{-1} \frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{y}}_i}{T}. \tag{43}$$

From above discussion and assumptions, we have following theorem

Theorem 1

$$\sqrt{T} \left(\hat{\boldsymbol{\theta}}_{IV,i} - \boldsymbol{\theta}_i \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{Q}_i^{-1} \boldsymbol{\Sigma}_{LS,i} \mathbf{Q}_i^{-1}), \tag{44}$$

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

A.0.1 Mean group LS estimator

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

$$\hat{\boldsymbol{\theta}}_{LSMG} = \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\theta}}_{LSi}. \quad (45)$$

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

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{LSMG} &= N^{-1} \sum_{i=1}^N \left(\frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{W}}_i}{T} \right)^{-1} \frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{y}}_i}{T} \\ &= \bar{\boldsymbol{\theta}} + N^{-1} \sum_{i=1}^N \left(\frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{W}}_i}{T} \right)^{-1} \frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{u}}_i}{T}, \end{aligned} \quad (46)$$

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

$$\text{plim}_{T \rightarrow \infty} \hat{\boldsymbol{\theta}}_{LSMG} = \bar{\boldsymbol{\theta}} + N^{-1} \sum_{i=1}^N \text{plim}_{T \rightarrow \infty} \left(\frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{W}}_i}{T} \right)^{-1} \text{plim}_{T \rightarrow \infty} \left(\frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{u}}_i}{T} \right) \quad (47)$$

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

$$\text{plim}_{T \rightarrow \infty} \hat{\boldsymbol{\theta}}_{LSMG} = \bar{\boldsymbol{\theta}}. \quad (48)$$

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

$$\text{plim}_{T \rightarrow \infty, N \rightarrow \infty} \hat{\boldsymbol{\theta}}_{LSMG} = \boldsymbol{\theta}. \quad (49)$$

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

$$\hat{\boldsymbol{\Sigma}}_{LS,\lambda} = \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG} \right) \left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG} \right)'. \quad (50)$$

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

$$\begin{aligned} \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG} \right) \left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG} \right)' &= \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_{LS,i} - E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) \right) \left(\hat{\boldsymbol{\theta}}_{LS,i} - E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) \right)' \\ &+ \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_{LSMG} - E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) \right) \left(\hat{\boldsymbol{\theta}}_{LSMG} - E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) \right)' \\ &- \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_{LSMG} - E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) \right) \left(\hat{\boldsymbol{\theta}}_{LS,i} - E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) \right)' \\ &- \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_{LS,i} - E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) \right) \left(\hat{\boldsymbol{\theta}}_{LSMG} - E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) \right)'. \end{aligned} \quad (51)$$

Taking expectation on equation (51), we have

$$\begin{aligned}
E \left(\sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG} \right) \left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG} \right)' \right) &= \sum_{i=1}^N Var \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) + \sum_{i=1}^N \hat{\boldsymbol{\theta}}_{LSMG} \hat{\boldsymbol{\theta}}_{LSMG}' + \\
&\sum_{i=1}^N E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right)' - \sum_{i=1}^N \hat{\boldsymbol{\theta}}_{LSMG} \hat{\boldsymbol{\theta}}_{LS,i}' + \sum_{i=1}^N E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) \hat{\boldsymbol{\theta}}_{LS,i}' - \sum_{i=1}^N \hat{\boldsymbol{\theta}}_{LSMG} E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right)' - \\
&\sum_{i=1}^N \hat{\boldsymbol{\theta}}_{LS,i} \hat{\boldsymbol{\theta}}_{LSMG}' + \sum_{i=1}^N \hat{\boldsymbol{\theta}}_{LS,i} E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right)' - \sum_{i=1}^N E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right)' = \\
&\sum_{i=1}^N Var \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) - N E \left(\hat{\boldsymbol{\theta}}_{LSMG} \hat{\boldsymbol{\theta}}_{LSMG}' \right) + \sum_{i=1}^N E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right)'
\end{aligned} \tag{52}$$

and

$$E \left(\hat{\boldsymbol{\theta}}_{LSMG} \hat{\boldsymbol{\theta}}_{LSMG}' \right) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E \left(\hat{\boldsymbol{\theta}}_{LS,i} \hat{\boldsymbol{\theta}}_{LS,i}' \right). \tag{53}$$

And, we also have

$$E \left(\hat{\boldsymbol{\theta}}_{LSMG} \hat{\boldsymbol{\theta}}_{LSMG}' \right) = \frac{1}{N^2} \left(\sum_{i=1}^N Var \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) + \sum_{i=1}^N \sum_{j=1}^N E \left(\hat{\boldsymbol{\theta}}_{LS,i} \hat{\boldsymbol{\theta}}_{LS,i}' \right) \right). \tag{54}$$

Then,

$$\begin{aligned}
E \left(\sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG} \right) \left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG} \right)' \right) &= \\
\left(1 - \frac{1}{N} \right) \sum_{i=1}^N Var \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) + \sum_{i=1}^N E \left(\hat{\boldsymbol{\theta}}_{LS,i} \hat{\boldsymbol{\theta}}_{LS,i}' \right) - \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N E \left(\hat{\boldsymbol{\theta}}_{LS,i} \hat{\boldsymbol{\theta}}_{LS,i}' \right).
\end{aligned} \tag{55}$$

From above equation (55), we can observe the bias term as,

$$\aleph = \sum_{i=1}^N E \left(\hat{\boldsymbol{\theta}}_{LS,i} \hat{\boldsymbol{\theta}}_{LS,i}' \right) - \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N E \left(\hat{\boldsymbol{\theta}}_{LS,i} \hat{\boldsymbol{\theta}}_{LS,i}' \right). \tag{56}$$

Taking expectation on equation (26), we have

$$E \left(\hat{\boldsymbol{\theta}}_{LS,i} \right) = \boldsymbol{\theta} + E \left(b_i \right), \tag{57}$$

From equation (56), we know that

$$\aleph = \sum_{i=1}^N E \left(b_i \right) E \left(b_i' \right) - \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N E \left(b_i \right) E \left(b_j' \right). \tag{58}$$

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

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

1.

$$\sqrt{N} \left(\hat{\boldsymbol{\theta}}_{LSMG} - \boldsymbol{\theta} \right) \xrightarrow{d} N \left(\mathbf{0}, \boldsymbol{\Sigma}_{LS,\lambda} \right). \quad (59)$$

2.

$$\hat{\boldsymbol{\Sigma}}_{LS,\lambda} \xrightarrow{p} \boldsymbol{\Sigma}_{LS,\lambda} \quad (60)$$

where

$$\hat{\boldsymbol{\Sigma}}_{LS,\lambda} = \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG} \right) \left(\hat{\boldsymbol{\theta}}_{LS,i} - \hat{\boldsymbol{\theta}}_{LSMG} \right)'. \quad (61)$$

References

- 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.