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 Heteroskedasticity and autocorrelation consistent variance estimator **6**

3 Construction optimal instruments **7**

4 Monte Carlo simulation design **8**
4.1 dynamic heterogeneous panels data model without multi-factor error structure 9

A Asymptotic property of LS estimator **10**
A.1 Asymptotic property of Mean group LS estimator 12

B Asymptotic property of IV estimator **13**

Reference **13**

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:

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

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

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

where $\boldsymbol{y}_i = (y_{i,1}, \dots, y_{i,T})^{'}$ is a $T \times 1$ vector, $\boldsymbol{W}_i^b = \left(\boldsymbol{w}_{i,1}^b, \dots, \boldsymbol{w}_{i,T}^b\right)^{'}$ is a $T \times (1+k)$ matrix, $\iota_T = (1, \dots, 1)^{'}$ and $\boldsymbol{u}_i^b = \left(u_{i,1}^b, \dots, u_{i,T}^b\right)^{'}$ 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

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

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 $\boldsymbol{F}^b$, the model can be expressed as

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

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

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

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

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

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

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

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

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

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

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

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

Assumption 2 *$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{\lambda}_i$, $\boldsymbol{\lambda}_i \overset{i.i.d.}{\sim} (\mathbf{0}, \boldsymbol{\Sigma}_\lambda)$, where $\boldsymbol{\Sigma}_\lambda$ 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 $\boldsymbol{X}_i$ as instruments, as

$$\tilde{\boldsymbol{Z}}_i^b = \left(\boldsymbol{F}^b \boldsymbol{X}_i, \boldsymbol{F}^b \boldsymbol{X}_{i,-1}\right), \tag{9}$$

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

Assumption 3 *$\boldsymbol{A}_i^b = \text{plim}_{T\to\infty} \tilde{\boldsymbol{A}}_{i,T}^b$ has full column rank, $\boldsymbol{B}_i^b = \text{plim}_{T\to\infty} \tilde{\boldsymbol{B}}_{i,T}^b$ and $\boldsymbol{\Sigma}_i^b = \text{plim}_{T\to\infty} T^{-1} \tilde{\boldsymbol{Z}}_i^{b'} \tilde{\boldsymbol{u}}_i^b \tilde{\boldsymbol{u}}_i^{b'} \tilde{\boldsymbol{Z}}_i^b$ has positive definite, uniformly, where $\tilde{\boldsymbol{A}}_{i,T}^b = \frac{1}{T} \tilde{\boldsymbol{Z}}_i^{b'} \tilde{\boldsymbol{W}}_i^b$ and $\tilde{\boldsymbol{B}}_{i,T}^b = \frac{1}{T} \tilde{\boldsymbol{Z}}_i^{b'} \tilde{\boldsymbol{Z}}_i^b$ 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

$$\hat{\boldsymbol{\theta}}_i^b = \left(\tilde{\boldsymbol{A}}_{i,T}^{b\prime} \tilde{\boldsymbol{B}}_{i,T}^{b-1} \tilde{\boldsymbol{A}}_{i,T}^{b} \right)^{-1} \tilde{\boldsymbol{A}}_{i,T}^{b\prime} \tilde{\boldsymbol{B}}_{i,T}^{b-1} \tilde{\boldsymbol{g}}_{i,T}^{b}, \tag{10}$$

where

$$\tilde{\boldsymbol{g}}_{i,T}^{b} = \frac{1}{T} \tilde{\boldsymbol{Z}}_i^{b\prime} \tilde{\boldsymbol{y}}_i^b. \tag{11}$$

From above equation, we have

$$\hat{\boldsymbol{\theta}}_i^b = \boldsymbol{\theta}_i + \left(\tilde{\boldsymbol{A}}_{i,T}^{b\prime} \tilde{\boldsymbol{B}}_{i,T}^{b-1} \tilde{\boldsymbol{A}}_{i,T}^{b} \right)^{-1} \tilde{\boldsymbol{A}}_{i,T}^{b\prime} \tilde{\boldsymbol{B}}_{i,T}^{b-1} \left(T^{-1} \tilde{\boldsymbol{Z}}_i^{b\prime} \tilde{\boldsymbol{u}}_i^b \right) \tag{12}$$

From assumption, we know $\boldsymbol{x}_{i,t}$ is strictly exogenous regressors. Then, we know $E\left(\boldsymbol{z}_{i,t}\boldsymbol{u}_{it}\right) = 0$. Therefore, we can show that

$$\hat{\boldsymbol{\theta}}_i^b \xrightarrow{p} \boldsymbol{\theta}_i. \tag{13}$$

From (12), we know

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i^b - \boldsymbol{\theta}_i \right) = \left(\tilde{\boldsymbol{A}}_{i,T}^{b\prime} \tilde{\boldsymbol{B}}_{i,T}^{b-1} \tilde{\boldsymbol{A}}_{i,T}^{b} \right)^{-1} \tilde{\boldsymbol{A}}_{i,T}^{b\prime} \tilde{\boldsymbol{B}}_{i,T}^{b-1} \left(T^{-1/2} \tilde{\boldsymbol{Z}}_i^{b\prime} \tilde{\boldsymbol{u}}_i^b \right) \tag{14}$$

Then, the property of $T^{-1/2} \tilde{\boldsymbol{Z}}_i^{b\prime} \tilde{\boldsymbol{u}}_i^b$ 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*

$$\begin{gathered} T^{-1/2} \tilde{\boldsymbol{Z}}_i^{b\prime} \tilde{\boldsymbol{u}}_i^b = O_p(1) \\ \textit{and} \\ T^{-1/2} \tilde{\boldsymbol{Z}}_i^{b\prime} \tilde{\boldsymbol{u}}_i^b \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_i^b\right). \end{gathered} \tag{15}$$

Thus, IV estimator, $\hat{\boldsymbol{\theta}}_i^b$ 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 $N/T \to c$ with $0 < c < \infty$. for each i,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i^b - \boldsymbol{\theta}_i \right) \xrightarrow{d} N\left(\boldsymbol{0}, \left(\boldsymbol{A}_i^{b\prime} \tilde{\boldsymbol{B}}_i^{b-1} \boldsymbol{A}_i^b \right)^{-1} \boldsymbol{A}_i^{b\prime} \tilde{\boldsymbol{B}}_i^{b-1} \boldsymbol{\Sigma}_i^b \tilde{\boldsymbol{B}}_i^{b-1} \boldsymbol{A}_i^b \left(\boldsymbol{A}_i^{b\prime} \tilde{\boldsymbol{B}}_i^{b-1} \boldsymbol{A}_i^b \right) \right). \tag{16}$$

1.2.1 Mean group IV estimator

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

$$\hat{\boldsymbol{\theta}}_{IVMG}^b = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{\theta}}_i^b. \tag{17}$$

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

$$\hat{\boldsymbol{\theta}}^b_{IVMG} - \boldsymbol{\theta} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}_i\right) + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_i, \tag{18}$$

where the first of right hand side

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}_i\right) &= \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{\boldsymbol{A}}^{b\prime}_{i,T}\tilde{\boldsymbol{B}}^{b-1}_{i,T}\tilde{\boldsymbol{A}}^{b}_{i,T}\right)^{-1}\tilde{\boldsymbol{A}}^{b\prime}_{i,T}\tilde{\boldsymbol{B}}^{b-1}_{i,T}\left(T^{-1}\tilde{\boldsymbol{Z}}^{b\prime}_i\tilde{\boldsymbol{u}}^b_i\right)\\ &= o_p\left(1\right).\end{aligned} \tag{19}$$

Then, we can see

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}^b_{IVMG} - \boldsymbol{\theta}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\lambda}_i + o_p\left(1\right). \tag{20}$$

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

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\lambda}_i \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_\lambda\right). \tag{21}$$

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

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

$$\hat{\boldsymbol{\Sigma}}^b_\lambda = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}^b_i - \hat{\boldsymbol{\theta}}^b_{IVMG}\right)\left(\hat{\boldsymbol{\theta}}^b_i - \hat{\boldsymbol{\theta}}^b_{IVMG}\right)'. \tag{22}$$

Follow Norkute et al. (2019), we can show that $\hat{\boldsymbol{\Sigma}}^b_\lambda$ is consistent and it does not have small T bias. Firstly, we decompose (22) as

$$\begin{aligned}&\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta} + \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^b_{IVMG}\right)\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta} + \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^b_{IVMG}\right)' =\\ &\sum_{i=1}^{N}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i' + \sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}_i\right)' + \sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}_i\right)\boldsymbol{\lambda}_i + \sum_{i=1}^{N}\boldsymbol{\lambda}_i\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}_i\right) -\\ &N\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^b_{IVMG}\right)'\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^b_{IVMG}\right).\end{aligned} \tag{23}$$

Then we can show consistent of $\hat{\boldsymbol{\Sigma}}^b_\lambda$ as

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}^b_\lambda - \boldsymbol{\Sigma}^b_\lambda &= \frac{1}{N-1}\sum_{i=1}^{N}\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i' - \boldsymbol{\Sigma}^b_\lambda\right) + \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}_i\right)'\\ &+ \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}_i\right)\boldsymbol{\lambda}_i + \frac{1}{N-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\left(\hat{\boldsymbol{\theta}}^b_i - \boldsymbol{\theta}_i\right) -\\ &\frac{N}{N-1}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^b_{IVMG}\right)'\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^b_{IVMG}\right) = o_p(1).\end{aligned} \tag{24}$$

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

$$\begin{aligned}\sqrt{N}\left(\hat{\boldsymbol{\theta}}^b_{IVMG} - \boldsymbol{\theta}\right) &\xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}^b_\lambda\right);\\ \hat{\boldsymbol{\Sigma}}^b_\lambda &\xrightarrow{p} \boldsymbol{\Sigma}^b_\lambda.\end{aligned} \tag{25}$$

2 Heteroskedasticity and autocorrelation consistent variance estimator

In practice we difficult to know the slop is homogeneous or heterogeneous. Therefore, we want to construct the heteroskedasticity and autocorrelation consistent (HAC) variance estimator to reduce the model selection problem. This approach is already provided by **?** and **?**. From the model (4) or 8, we know

$$\begin{aligned}\tilde{\boldsymbol{y}}_i^b &= \tilde{\boldsymbol{W}}_i^b \boldsymbol{\theta}_i + \tilde{\boldsymbol{u}}_i^b, \\ &= \tilde{\boldsymbol{W}}_i^b \boldsymbol{\theta} + \tilde{\boldsymbol{\varepsilon}}_i^b, \quad \tilde{\boldsymbol{\varepsilon}}_i^b = \tilde{\boldsymbol{W}}_i^b \boldsymbol{\lambda}_i + \boldsymbol{u}_i.\end{aligned} \tag{26}$$

Then, we use the lag of regressors as instruments to get the IV estimator for $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}}^b = \left(\sum_{i=1}^N \tilde{\boldsymbol{A}}_i^{b'} \tilde{\boldsymbol{B}}_i^{b-1} \tilde{\boldsymbol{A}}_i^b\right)^{-1} \sum_{i=1}^N \tilde{\boldsymbol{A}}_i^{b'} \tilde{\boldsymbol{B}}_i^{b-1} \tilde{\boldsymbol{g}}_i^b. \tag{27}$$

We define

$$\bar{\boldsymbol{D}}_{NT} = \frac{1}{N}\sum_{i=1}^N \bar{\boldsymbol{D}}_{it}, \quad \bar{\boldsymbol{D}}_{iT} = T^{-1} \tilde{\boldsymbol{A}}_i^{b'} \tilde{\boldsymbol{B}}_i^{b-1} \tilde{\boldsymbol{A}}_i^b. \tag{28}$$

From above setting, we have

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^b = \left(\sum_{i=1}^N \tilde{\boldsymbol{A}}_i^{b'} \tilde{\boldsymbol{B}}_i^{b-1} \tilde{\boldsymbol{A}}_i^b\right)^{-1} \sum_{i=1}^N \tilde{\boldsymbol{A}}_i^{b'} \tilde{\boldsymbol{B}}_i^{b-1} \tilde{\boldsymbol{Z}}_i^{b'} \tilde{\boldsymbol{\varepsilon}}_i^b \tag{29}$$

$$= \bar{\boldsymbol{D}}_{NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^N \tilde{\boldsymbol{A}}_i^{b'} \tilde{\boldsymbol{B}}_i^{b-1} \tilde{\boldsymbol{Z}}_i^{b'} \tilde{\boldsymbol{u}}_i^b + \frac{1}{N} \tilde{\boldsymbol{A}}_i^{b'} \tilde{\boldsymbol{B}}_i^{b-1} \tilde{\boldsymbol{A}}_i^b \tilde{\boldsymbol{\lambda}}_i\right). \tag{30}$$

Then, from **?** and **?**, we have

Assumption 4 $\boldsymbol{C}_{iT} = T^{-2}E\left(\boldsymbol{D}_i \boldsymbol{\Sigma}_{\lambda,i} \boldsymbol{D}_i\right)$, $\boldsymbol{C} = \lim_{N,T\to\infty} \boldsymbol{C}_{NT}$ *and* $\boldsymbol{C}_{NT} = N^{-1}\sum_{i=1}^N \boldsymbol{C}_{iT}$.

Assumption 5 $\boldsymbol{E}_{iT} = T^{-1}E\left(\boldsymbol{Z}_i' \boldsymbol{u}_i \boldsymbol{u}_i' \boldsymbol{Z}_i\right)$, $\boldsymbol{E} = \lim_{N,T\to\infty} \boldsymbol{E}_{NT}$ *and* $\boldsymbol{E}_{NT} = N^{-1}\sum_{i=1}^N \boldsymbol{E}_{iT}$.

Theorem
As $(N,T)\to\infty$,

$$\bar{\boldsymbol{D}}_{NT}^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \tilde{\boldsymbol{A}}_i^{b'} \tilde{\boldsymbol{B}}_i^{b-1} \tilde{\boldsymbol{Z}}_i^{b'} \tilde{\boldsymbol{u}}_i^b \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{D}^{-1}\boldsymbol{E}\boldsymbol{D}^{-1}\right), \tag{31}$$

Theorem
As $(N,T)\to\infty$,

$$\bar{\boldsymbol{D}}_{NT}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{\boldsymbol{A}}_i^{b'} \tilde{\boldsymbol{B}}_i^{b-1} \tilde{\boldsymbol{A}}_i^b \tilde{\boldsymbol{\lambda}}_i \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{D}^{-1}\right) \tag{32}$$

We can estimate the variance of $\boldsymbol{\theta}$ by Chudik and Pesaran (2015) and **?** as

$$\hat{\boldsymbol{\Sigma}}_{\hat{\theta}} = \frac{1}{N}\bar{\boldsymbol{D}}_{NT}^{-1}\boldsymbol{G}_{NT}\bar{\boldsymbol{D}}_{NT}^{-1}, \tag{33}$$

where

$$\boldsymbol{G}_{NT} = \frac{1}{N}\sum_{i=1}^{N}\sum_{i=1}^{N}\bar{\boldsymbol{D}}_{iT}\left(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}}\right)\left(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}}\right)'\bar{\boldsymbol{D}}_{iT}. \tag{34}$$

From above, we have

$$\begin{aligned}
&\frac{1}{NT^2}\sum_{i=1}^{N}E\left(\tilde{\boldsymbol{A}}_i^{b'}\tilde{\boldsymbol{B}}_i^{b-1}\tilde{\boldsymbol{Z}}_i^{b'}\tilde{\boldsymbol{\varepsilon}}_i^{b}\tilde{\boldsymbol{\varepsilon}}_i^{b'}\tilde{\boldsymbol{Z}}_i^{b}\tilde{\boldsymbol{B}}_i^{b-1}\tilde{\boldsymbol{A}}_i^{b}\right) = \frac{1}{NT^2}\sum_{i=1}^{N}E\left(\tilde{\boldsymbol{A}}_i^{b'}\tilde{\boldsymbol{B}}_i^{b-1}\tilde{\boldsymbol{A}}_i^{b}\boldsymbol{\Sigma}_{\lambda,i}\tilde{\boldsymbol{A}}_i^{b'}\tilde{\boldsymbol{B}}_i^{b-1}\tilde{\boldsymbol{A}}_i^{b}\right) + \\
&\frac{1}{NT^2}\sum_{i=1}^{N}E\left(\tilde{\boldsymbol{A}}_i^{b'}\tilde{\boldsymbol{B}}_i^{b-1}\boldsymbol{Z}_i'\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{Z}_i\tilde{\boldsymbol{B}}_i^{b-1}\tilde{\boldsymbol{A}}_i^{b}\right) = \\
&\frac{1}{NT^2}\sum_{i=1}^{N}E\left(\tilde{\boldsymbol{A}}_i^{b'}\tilde{\boldsymbol{B}}_i^{b-1}\tilde{\boldsymbol{A}}_i^{b}\boldsymbol{\Sigma}_{\lambda,i}\tilde{\boldsymbol{A}}_i^{b'}\tilde{\boldsymbol{B}}_i^{b-1}\tilde{\boldsymbol{A}}_i^{b}\right) + O(T^{-1}).
\end{aligned} \tag{35}$$

Then, we have HAC variance estimator as

$$\hat{\boldsymbol{\Sigma}}_{\hat{\lambda}} = \left(\sum_{i=1}^{N}\tilde{\boldsymbol{A}}_i^{b'}\tilde{\boldsymbol{B}}_i^{b-1}\tilde{\boldsymbol{A}}_i^{b}\right)^{-1}\left(\sum_{i=1}^{N}\tilde{\boldsymbol{A}}_i^{b'}\tilde{\boldsymbol{B}}_i^{b-1}\boldsymbol{E}\tilde{\boldsymbol{B}}_i^{b-1}\tilde{\boldsymbol{A}}_i^{b}\right)\left(\sum_{i=1}^{N}\tilde{\boldsymbol{A}}_i^{b'}\tilde{\boldsymbol{B}}_i^{b-1}\tilde{\boldsymbol{A}}_i^{b}\right)^{-1} \tag{36}$$

3 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

$$\tilde{\boldsymbol{Z}}_i^{b(j)} = \left(\boldsymbol{F}^b\boldsymbol{X}_i, \boldsymbol{F}^b\boldsymbol{X}_{i,-1}, \ldots, \boldsymbol{F}^b\boldsymbol{X}_{i,-j}\right), \tag{37}$$

where $\tilde{\boldsymbol{Z}}_i^{b(j)}$ is $T1 \times (j+1)k$ 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

$$\hat{\boldsymbol{\theta}}_i^{b} = \left(\tilde{\boldsymbol{W}}_i^{b'}\boldsymbol{P}_i^{b(j)}\tilde{\boldsymbol{W}}_i^{b}\right)^{-1}\tilde{\boldsymbol{W}}_i^{b'}\boldsymbol{P}_i^{b(j)}\tilde{\boldsymbol{y}}_i^{b}, \tag{38}$$

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

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

$$\boldsymbol{\omega}^{(i)} = \left(\omega_1^{(i)}, \ldots, \omega_J^{(i)}\right)', \tag{39}$$

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

$$\boldsymbol{P}_i^b = \sum_{j=1}^{J} \omega_j^{(i)} \boldsymbol{P}_i^{b(j)}, \tag{40}$$

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, $S_{\eta^{(i)}}^b\left(\boldsymbol{\omega^{(i)}}\right)$, 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 $\boldsymbol{D}_i^b = \left(\tilde{\boldsymbol{w}}_{i,1}^b, \ldots, \tilde{\boldsymbol{w}}_{i,T}^b\right)'$ is a $T \times (1+k)$ matrix. Let $\hat{\boldsymbol{H}}_i$ is some estimator of $\boldsymbol{H}_i = \frac{\boldsymbol{D}_i^{b'} \boldsymbol{D}_i^b}{T}$. Let $\hat{\boldsymbol{\theta}}_i^b$ is some preliminary estimator, and define the residuals $\hat{\tilde{\boldsymbol{u}}}_i^b = \tilde{\boldsymbol{y}}_i^b - \tilde{\boldsymbol{W}}_i^b \hat{\boldsymbol{\theta}}_i^b$ that does not depend on the weighting vector. Let $T \times (1+k)$ matrix, $\tilde{\boldsymbol{V}}_i^b$, be some preliminary residual from first stage regression, and $\tilde{\boldsymbol{v}}_{\eta,i}^b = \tilde{\boldsymbol{V}}_i^b \hat{\boldsymbol{H}}_i^{-1} \boldsymbol{\eta}^{(i)}$.

Define

$$\hat{\sigma}_{u,i}^{b2} = \frac{\hat{\tilde{\boldsymbol{u}}}_i^{b'} \hat{\tilde{\boldsymbol{u}}}_i^b}{T}, \; \hat{\sigma}_{\eta,i}^{b2} = \frac{\tilde{\boldsymbol{v}}_{\eta,i}^{b'} \tilde{\boldsymbol{v}}_{\eta,i}^b}{T}, \; \hat{\boldsymbol{\sigma}}_{\eta u,i}^b = \frac{\tilde{\boldsymbol{v}}_{\eta,i}^{b'} \hat{\tilde{\boldsymbol{u}}}_i^b}{T}. \tag{41}$$

Let $\hat{\boldsymbol{v}}_{\eta,i}^{(j)} = \left(\boldsymbol{P}_i^{(J)} - \boldsymbol{P}_i^{(j)}\right) \tilde{\boldsymbol{W}}_i^b \hat{\boldsymbol{H}}^{-1} \boldsymbol{\eta}^{(i)}$ is a $T \times 1$ vector, and $\hat{\boldsymbol{U}}_i = \left(\hat{\boldsymbol{v}}_{\eta,i}^{(1)}, \ldots, \hat{\boldsymbol{v}}_{\eta,i}^{(J)}\right)' \left(\hat{\boldsymbol{v}}_{\eta,i}^{(1)}, \ldots, \hat{\boldsymbol{v}}_{\eta,i}^{(J)}\right)$ 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 $K = (1, \ldots, J)'$. The criterion $\hat{S}_{\eta^{(i)}}^b\left(\boldsymbol{\omega^{(i)}}\right)$ is

$$\hat{S}_{\eta^{(i)}}^b\left(\boldsymbol{\omega^{(i)}}\right) = \hat{\sigma}_{\eta u,i}^{b2} \frac{(K'\boldsymbol{\omega}^{(i)})^2}{T} + \hat{\sigma}_{u,i}^{b2} \frac{\boldsymbol{\omega}^{(i)'} \hat{\boldsymbol{U}}_i \boldsymbol{\omega}^{(i)} - \hat{\sigma}_{\eta,i}^{b2}(J - 2K'\boldsymbol{\omega}^{(i)} + \boldsymbol{\omega}^{(i)'} \Gamma_i \boldsymbol{\omega}^{(i)})}{T}. \tag{42}$$

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

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

$$\hat{\boldsymbol{\theta}}_i^{*b} = \left(\tilde{\boldsymbol{W}}_i^{b'} \boldsymbol{P}_i^b\left(\boldsymbol{\omega^{(i)*}}\right) \tilde{\boldsymbol{W}}_i^b\right)^{-1} \tilde{\boldsymbol{W}}_i^{b'} \boldsymbol{P}_i^b\left(\boldsymbol{\omega^{(i)*}}\right) \tilde{\boldsymbol{y}}_i^b. \tag{43}$$

4 Monte Carlo simulation design

It has been found that the IV estimator exhibit systematic bias in the model with heterogeneous slops by Norkute et al. (2019). And $IVMG^c$ is outperform $CCEMG$ in terms of bias, RMSE, size and power. Therefore, we would like to see the behaviour of $IV^a, IV^c, IVMG^a, IVMG^c, IV^{opt}$, and $IVMG^{opt}$ estimators in this simple simulation, where a corresponds to two instruments are used and c corresponds to three instruments are used while *opt* corresponds MA2SLS estimator.

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

The data generating process:

$$
\begin{aligned}
y_{i,t} &= \phi_i y_{i,t-1} + \beta_{1i} x_{1i,t} + \alpha_i + u_{i,t},\ \ for\ i = 1, \ldots N;\ t = -49 - (J+1), \ldots, T\,, \\
x_{1i,t} &= \rho x_{1i,t-1} + \tau_\alpha \alpha_i + \theta_u u_{i,t-1} + v_{i,t},
\end{aligned}
\tag{44}
$$

where $u_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,\ 1)$, and $v_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,\ 1), \alpha_i \overset{i.i.d.}{\sim} \mathcal{N}(0,\ 1)$, $\rho = 0.5$, $\theta_u = -0.2$, $\tau_\alpha = 0.5$, $J = 2, 3, 4, 5$.

The slope coefficients are generated as

$$
\phi_i = \phi + \eta_{\phi i};\ \ \beta_{1,i} = \beta_1 + \eta_{\beta_1 i}. \tag{45}
$$

Here we consider $\phi \in \{0.2, 0.5, 0.8\}$, $\beta_1 = 3$. For the design of heterogenous slopes, $\eta_{\phi i} \overset{i.i.d.}{\sim} U(-c, c)$, and

$$
\eta_{\beta_1 i} = \left(1 - \rho_\beta^2\right)^{1/2} \eta_{\phi i}. \tag{46}
$$

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

$$\hat{\boldsymbol{\theta}}_{LS,i} = \begin{pmatrix} \hat{\phi}_i \\ \hat{\beta}_i \end{pmatrix} = \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1}^2}{T} & \frac{\sum_{t=1}^T \tilde{y}_{i,t-1}\tilde{x}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t}\tilde{y}_{i,t-1}}{T} & \frac{\sum_{t=1}^T \tilde{x}_{i,t}^2}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1}\tilde{y}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t}\tilde{y}_{i,t}}{T} \end{pmatrix}, \tag{47}$$

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

$$\begin{pmatrix} \hat{\phi}_i - \phi_i \\ \hat{\beta}_i - \beta_i \end{pmatrix} = \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1}^2}{T} & \frac{\sum_{t=1}^T \tilde{y}_{i,t-1}\tilde{x}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t}\tilde{y}_{i,t-1}}{T} & \frac{\sum_{t=1}^T \tilde{x}_{i,t}^2}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1}\tilde{u}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t}\tilde{u}_{i,t}}{T} \end{pmatrix}, \tag{48}$$

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

$$A_{\phi i}^{(1)} = \operatorname*{plim}_{T\to\infty} \left(\frac{\sum_{t=1}^T \tilde{y}_{i,t-1}\tilde{u}_{i,t}}{T} \right). \tag{49}$$

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

$$\begin{aligned} A_{\phi i}^{(1)} &= E\left(y_{i,t-1} - \bar{y}_{i,-1}\right)\left(u_{i,t} - \bar{u}_i\right) \\ &= E\left(y_{i,t-1}u_{i,t}\right) - E\left(y_{i,t-1}\bar{u}_i\right) - E\left(\bar{y}_{i,-1}u_{i,t}\right) + E\left(\bar{y}_{i,-1}\bar{u}_i\right), \end{aligned} \tag{50}$$

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

$$y_{i,t} = \frac{\alpha_i}{(1-\phi_i)} + \sum_{s=0}^{\infty} \beta_i \phi_i^s x_{i,t-s} + \sum_{s=0}^{\infty} \phi_i^s u_{i,t-s}, \tag{51}$$

Then, we have

$$\begin{aligned} A_{\phi i}^{(1)} = &-E\left(\left(\sum_{s=0}^{\infty} \phi_i^s u_{i,t-s-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} u_{i,t}\right)\right) - E\left(\frac{u_{i,t}}{T}\sum_{t=1}^{T}\sum_{s=0}^{\infty} \phi_i^s u_{i,t-s-1}\right) + \\ &\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=0}^{\infty} \phi_i^s u_{i,t-s-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} u_{i,t}\right). \end{aligned} \tag{52}$$

Hence, from above equation, we have

$$
\begin{aligned}
A_{\phi i}^{(1)} =& -\frac{1}{T}E\left\{\left(u_{i,t-1}+u_{i,t-2}\phi_i^1+u_{i,t-3}\phi_i^2+\ldots\right)\left(u_{i,1}+\ldots+u_{i,t-1}+u_{i,t}+\ldots+u_{i,T}\right)\right\}- \\
& \frac{1}{T}E\left\{u_{i,t}\sum_{s=1}^{T}\left(u_{i,s-1}\phi_i^0+u_{i,s-2}\phi_i^1+\ldots+u_{i,s-t-1}\phi_i^t+\ldots+u_{i,s-T-1}\phi_i^T+\ldots\right)\right\}+ \\
& +\frac{1}{T}E\{\left(\sum_{t=1}^{T}u_{i,t-1}\phi_i^0+\sum_{t=1}^{T}u_{i,t-2}\phi_i^1+\ldots+\sum_{t=1}^{T}u_{i,-1}\phi_i^t+\ldots+\sum_{t=1}^{T}u_{it-T-1}\phi_i^T+\ldots\right) \\
& \left(\frac{1}{T}\sum_{t=1}^{T}u_{i,t}\right)\} \\
=& -\frac{\sigma_u^2}{T}\frac{\left(1-\phi_i^{t-1}\right)}{1-\phi_i}-\frac{\sigma_u^2}{T}\frac{\left(1-\phi_i^{T-t}\right)}{\left(1-\phi_i\right)}+\frac{\sigma_u^2}{T}\left(\frac{1}{1-\phi_i}-\frac{1}{T}\frac{\left(1-\phi_i^T\right)}{(1-\phi_i)^2}\right) \\
=& -\frac{\sigma_u^2}{T(1-\phi_i)}\left(1-\phi_i^{t-1}-\phi_i^{T-t}+\frac{1}{T}\frac{\left(1-\phi_i^T\right)}{\left(1-\phi_i\right)}\right).
\end{aligned}
\tag{53}
$$

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,

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{W}}_i\boldsymbol{\theta}_i + \tilde{\boldsymbol{u}}_i, \tag{54}$$

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

$$\left(\phi_i,\beta_i\right)' = \boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{\lambda}_i, \tag{55}$$

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

$$\hat{\boldsymbol{\theta}}_{LS,i} = \left(\frac{\tilde{\boldsymbol{W}}_i'\tilde{\boldsymbol{W}}_i}{T}\right)^{-1}\frac{\tilde{\boldsymbol{W}}_i'\tilde{\boldsymbol{y}}_i}{T}. \tag{56}$$

From above discussion and assumptions, we have following theorem

Theorem 1

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{LS,i}-\boldsymbol{\theta}_i\right) \xrightarrow{d} N\left(\mathbf{0},\boldsymbol{Q}_i^{-1}\boldsymbol{\Sigma}_{LS,i}\boldsymbol{Q}_i^{-1}\right), \tag{57}$$

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.1 Asymptotic property of Mean group LS estimator

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

$$\hat{\boldsymbol{\theta}}_{LSMG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{LSi}. \tag{58}$$

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

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{LSMG} &= N^{-1}\sum_{i=1}^{N}\left(\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{W}}_i}{T}\right)^{-1}\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{y}}_i}{T} \\ &= \bar{\boldsymbol{\theta}} + N^{-1}\sum_{i=1}^{N}\left(\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{W}}_i}{T}\right)^{-1}\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{u}}_i}{T},\end{aligned} \tag{59}$$

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

$$\operatorname*{plim}_{T\to\infty}\hat{\boldsymbol{\theta}}_{LSMG} = \bar{\boldsymbol{\theta}} + N^{-1}\sum_{i=1}^{N}\operatorname*{plim}_{T\to\infty}\left(\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{W}}_i}{T}\right)^{-1}\operatorname*{plim}_{T\to\infty}\left(\frac{\tilde{\boldsymbol{W}}_i^{'}\tilde{\boldsymbol{u}}_i}{T}\right) \tag{60}$$

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

$$\operatorname*{plim}_{T\to\infty}\hat{\boldsymbol{\theta}}_{LSMG} = \bar{\boldsymbol{\theta}}. \tag{61}$$

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

$$\operatorname*{plim}_{T\to\infty,N\to\infty}\hat{\boldsymbol{\theta}}_{LSMG} = \boldsymbol{\theta}. \tag{62}$$

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

$$\hat{\boldsymbol{\Sigma}}_{LS,\lambda} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i}-\hat{\boldsymbol{\theta}}_{LSMG}\right)\left(\hat{\boldsymbol{\theta}}_{LS,i}-\hat{\boldsymbol{\theta}}_{LSMG}\right)^{'}. \tag{63}$$

Firstly, we decompose (63) as

$$\begin{aligned}&\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i}-\boldsymbol{\theta}+\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{LSMG}\right)\left(\hat{\boldsymbol{\theta}}_{LS,i}-\boldsymbol{\theta}+\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{lSMG}\right)^{'} = \\ &\sum_{i=1}^{N}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i^{'} + \sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i}-\boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}_{LS,i}-\boldsymbol{\theta}_i\right)^{'} + \sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i}-\boldsymbol{\theta}_i\right)\boldsymbol{\lambda}_i + \sum_{i=1}^{N}\boldsymbol{\lambda}_i\left(\hat{\boldsymbol{\theta}}_{LS,i}-\boldsymbol{\theta}_i\right) - \\ &N\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{LSMG}\right)^{'}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{LSMG}\right).\end{aligned} \tag{64}$$

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

$$\hat{\boldsymbol{\Sigma}}_{LS,\lambda} - \boldsymbol{\Sigma}_{LS,\lambda} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i' - \boldsymbol{\Sigma}_{LS,\lambda}\right) + \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i} - \boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}_{LS,i} - \boldsymbol{\theta}_i\right)'$$
$$+ \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{LS,i} - \boldsymbol{\theta}_i\right)\boldsymbol{\lambda}_i + \frac{1}{N-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\left(\hat{\boldsymbol{\theta}}_{LS,i} - \boldsymbol{\theta}_i\right) -$$
$$\frac{N}{N-1}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{LSMG}\right)'\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{LSMG}\right) = o_p(1). \tag{65}$$

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

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{LSMG} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{LS,\lambda}\right). \tag{66}$$

B Asymptotic property of IV estimator

From the section 1, we know the IV estimator is

$$\begin{aligned}\hat{\boldsymbol{\theta}}_i^b &= \left(\tilde{\boldsymbol{A}}_{i,T}^{b'}\tilde{\boldsymbol{B}}_{i,T}^{b-1}\tilde{\boldsymbol{A}}_{i,T}^{b}\right)^{-1}\tilde{\boldsymbol{A}}_{i,T}^{b'}\tilde{\boldsymbol{B}}_{i,T}^{b-1}\tilde{\boldsymbol{g}}_{i,T}^{b}, \\ \hat{\boldsymbol{\theta}}_i^b &= \boldsymbol{\theta}_i^b + \left(\tilde{\boldsymbol{A}}_{i,T}^{b'}\tilde{\boldsymbol{B}}_{i,T}^{b-1}\tilde{\boldsymbol{A}}_{i,T}^{b}\right)^{-1}\tilde{\boldsymbol{A}}_{i,T}^{b'}\tilde{\boldsymbol{B}}_{i,T}^{b-1}\left(T^{-1/2}\tilde{\boldsymbol{Z}}_i^{b'}\tilde{\boldsymbol{u}}_i^b\right)\end{aligned} \tag{67}$$

From assumption, we know $\boldsymbol{x}_{i,t}$ is strictly exogenous regressors. Then, we know $E\left(\boldsymbol{z}_{i,t}\boldsymbol{u}_{it}\right) = 0$. Therefore, we can show that

$$\hat{\boldsymbol{\theta}}_i^b \xrightarrow{p} \boldsymbol{\theta}_i^b. \tag{68}$$

References

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.