
Model Average of Dynamic Heterogeneous Panels with multi-factor error structure

Contents

1	Introduction	2
2	Model, asymptotic property of IV estimator	2
2.1	The models and Assumptions	2
2.1.1	Fixed effects model, $b=1$	2
2.1.2	Trend model, $b=2$	3
2.2	IV estimation method and asymptotic property	3
3	Construction optimal instruments	5
3.1	Mean group IV estimator	6
4	Dynamic heterogeneous panel data model with multifactor error structure	7
4.1	Optimal number of instrument variables	8
5	Monte Carlo simulation design	9
5.1	dynamic heterogeneous panels data model without multi-factor error structure	9
5.2	Dynamic heterogeneous panels data model with multi-factor error structure	10
A	Consistency of model average IV estimator	13
B	Asymptotic property of LS estimator	15
B.1	Asymptotic property of Mean group LS estimator	17
C	Asymptotic property of IV estimator	18
	Reference	18

1 Introduction

How many instrument variables should we use in estimation ? How to balance the trade-off between bias and efficiency ? Is our model slope heterogeneity or slope homogeneous ? In practical, we often have these open questions and we need to give the reasonable explanation from theory or data. Therefore, it might cost time to control all of these to get the reasonable explain in applied research. In this study, we would want to develop the statistical methods to reduce the model uncertainty. Our contribution in this study are reducing the uncertainty of number of instrument variables and reducing the model selection problems. Also, this method is useful to largest set of instrument variables in high dimensional.

2 Model, asymptotic property of IV estimator

2.1 The models and Assumptions

2.1.1 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 $t = 1, \dots, T-1$.

2.1.2 Trend model, b=2

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

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

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 $b = 2$ corresponds to the trend model. Stacking the T observations for each i , we have

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

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 expressed as

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

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 $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{\lambda}_i$, $\boldsymbol{\lambda}_i \stackrel{i.i.d.}{\sim} (\mathbf{0}, \boldsymbol{\Sigma}_\lambda)$, where $\boldsymbol{\Sigma}_\lambda$ is a fixed positive definite matrix.

2.2 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

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

where $\tilde{\mathbf{Z}}_i^b$ is a $(T-1) \times 2k$ 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, $\mathbf{B}_i^b = \text{plim}_{T \rightarrow \infty} \tilde{\mathbf{B}}_{i,T}^b$ and $\Sigma_i^b = \text{plim}_{T \rightarrow \infty} T^{-1} \tilde{\mathbf{Z}}_i^{b'} \tilde{\mathbf{u}}_i^b \tilde{\mathbf{u}}_i^{b'} \tilde{\mathbf{Z}}_i^b$ has positive definite, uniformly, where $\tilde{\mathbf{A}}_{i,T}^b = \frac{1}{T} \tilde{\mathbf{Z}}_i^{b'} \tilde{\mathbf{W}}_i^b$ and $\tilde{\mathbf{B}}_{i,T}^b = \frac{1}{T} \tilde{\mathbf{Z}}_i^{b'} \tilde{\mathbf{Z}}_i^b$ 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

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

where

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

From above equation, we have

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

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

$$\hat{\boldsymbol{\theta}}_i^b \xrightarrow{p} \boldsymbol{\theta}_i. \quad (13)$$

From (12), we know

$$\sqrt{T} \left(\hat{\boldsymbol{\theta}}_i^b - \boldsymbol{\theta}_i \right) = \left(\tilde{\mathbf{A}}_{i,T}^{b'} \tilde{\mathbf{B}}_{i,T}^{b-1} \tilde{\mathbf{A}}_{i,T}^b \right)^{-1} \tilde{\mathbf{A}}_{i,T}^{b'} \tilde{\mathbf{B}}_{i,T}^{b-1} \left(T^{-1/2} \tilde{\mathbf{Z}}_i^{b'} \tilde{\mathbf{u}}_i^b \right) \quad (14)$$

Then, the property of $T^{-1/2} \tilde{\mathbf{Z}}_i^{b'} \tilde{\mathbf{u}}_i^b$ 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

$$T^{-1/2} \tilde{\mathbf{Z}}_i^{b'} \tilde{\mathbf{u}}_i^b \xrightarrow{d} N(\mathbf{0}, \Sigma_i^b). \quad (15)$$

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 $N/T \rightarrow c$ with $0 < c < \infty$. for each i ,

$$\sqrt{T} \left(\hat{\boldsymbol{\theta}}_i^b - \boldsymbol{\theta}_i \right) \xrightarrow{d} N \left(\mathbf{0}, \left(\mathbf{A}_i^{b'} \tilde{\mathbf{B}}_i^{b-1} \mathbf{A}_i^b \right)^{-1} \mathbf{A}_i^{b'} \tilde{\mathbf{B}}_i^{b-1} \Sigma_i^b \tilde{\mathbf{B}}_i^{b-1} \mathbf{A}_i^b \left(\mathbf{A}_i^{b'} \tilde{\mathbf{B}}_i^{b-1} \mathbf{A}_i^b \right) \right). \quad (16)$$

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

$$\tilde{\mathbf{Z}}_i^{b(j)} = (\mathbf{F}^b \mathbf{X}_i, \mathbf{F}^b \mathbf{X}_{i,-1}, \dots, \mathbf{F}^b \mathbf{X}_{i,-j}), \quad (17)$$

where $\tilde{\mathbf{Z}}_i^{b(j)}$ is $(T-1) \times (j+1)k$ 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

$$\hat{\boldsymbol{\theta}}_i^b = \left(\tilde{\mathbf{W}}_i^{b'} \mathbf{P}_i^{b(j)} \tilde{\mathbf{W}}_i^b \right)^{-1} \tilde{\mathbf{W}}_i^{b'} \mathbf{P}_i^{b(j)} \tilde{\mathbf{y}}_i^b, \quad (18)$$

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

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

$$\boldsymbol{\omega}^{(i)} = \left(\omega_1^{(i)}, \dots, \omega_J^{(i)} \right)', \quad (19)$$

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

$$\mathbf{P}_i^b = \sum_{j=1}^J \omega_j^{(i)} \mathbf{P}_i^{b(j)}, \quad (20)$$

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, $S_{\boldsymbol{\eta}^{(i)}}^b(\boldsymbol{\omega}^{(i)})$, 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 $\mathbf{D}_i^b = (\tilde{\mathbf{w}}_{i,1}^b, \dots, \tilde{\mathbf{w}}_{i,T-1}^b)'$ is a $(T-1) \times (1+k)$ matrix. Let $\hat{\mathbf{H}}_i$ is some estimator of $\mathbf{H}_i = \frac{\mathbf{D}_i^{b'} \mathbf{D}_i^b}{T-1}$. 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 $(T-1) \times (1+k)$ 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

$$\hat{\sigma}_{u,i}^{b2} = \frac{\hat{\mathbf{u}}_i^{b'} \hat{\mathbf{u}}_i^b}{T-1}, \quad \hat{\sigma}_{\eta,i}^{b2} = \frac{\tilde{\mathbf{v}}_{\eta,i}^{b'} \tilde{\mathbf{v}}_{\eta,i}^b}{T-1}, \quad \hat{\sigma}_{\eta u,i}^b = \frac{\tilde{\mathbf{v}}_{\eta,i}^{b'} \hat{\mathbf{u}}_i^b}{T-1}. \quad (21)$$

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

$$\hat{S}_{\eta^{(i)}}^b(\boldsymbol{\omega}^{(i)}) = \hat{\sigma}_{\eta u, i}^{b2} \frac{(K' \boldsymbol{\omega}^{(i)})^2}{T-1} + \hat{\sigma}_{u, i}^{b2} \frac{\boldsymbol{\omega}^{(i)'} \hat{\mathbf{U}}_i \boldsymbol{\omega}^{(i)} - \hat{\sigma}_{\eta, i}^{b2} (J - 2K' \boldsymbol{\omega}^{(i)} + \boldsymbol{\omega}^{(i)'} \boldsymbol{\Gamma}_i \boldsymbol{\omega}^{(i)})}{T-1}. \quad (22)$$

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

$$\hat{\boldsymbol{\theta}}_i^{*b} = \left(\tilde{\mathbf{W}}_i^{b'} \mathbf{P}_i^b(\boldsymbol{\omega}^{(i)*}) \tilde{\mathbf{W}}_i^b \right)^{-1} \tilde{\mathbf{W}}_i^{b'} \mathbf{P}_i^b(\boldsymbol{\omega}^{(i)*}) \tilde{\mathbf{y}}_i^b. \quad (23)$$

Chen et al. (2016) also provide an alternative weighting scheme based on weighting individual estimators. This method can be also applied in heterogeneous dynamic panel data model. Also this method could be seen as an alternative method for mean group estimator. It would be interested to compare the performance of these two weighting scheme.

3.1 Mean group IV estimator

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

$$\hat{\boldsymbol{\theta}}_{IVMG}^b = \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\theta}}_i^{*b}. \quad (24)$$

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

$$\hat{\boldsymbol{\theta}}_{IVMG}^{*b} - \boldsymbol{\theta} = \frac{1}{N} \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta} \right) = \frac{1}{N} \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta}_i \right) + \frac{1}{N} \sum_{i=1}^N \boldsymbol{\lambda}_i, \quad (25)$$

where the first of right hand side

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta}_i \right) &= \frac{1}{N} \sum_{i=1}^N \left(\tilde{\mathbf{A}}_{i,T}^{b'} \tilde{\mathbf{B}}_{i,T}^{b-1} \tilde{\mathbf{A}}_{i,T}^b \right)^{-1} \tilde{\mathbf{A}}_{i,T}^{b'} \tilde{\mathbf{B}}_{i,T}^{b-1} \left(T^{-1} \tilde{\mathbf{Z}}_i^{b'} \tilde{\mathbf{u}}_i^b \right) \\ &= o_p(1). \end{aligned} \quad (26)$$

Then, we can see

$$\sqrt{N} \left(\hat{\boldsymbol{\theta}}_{IVMG}^b - \boldsymbol{\theta} \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{\lambda}_i + o_p(1). \quad (27)$$

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

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{\lambda}_i \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_\lambda). \quad (28)$$

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

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

$$\hat{\Sigma}_{\lambda}^b = \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i^{*b} - \hat{\boldsymbol{\theta}}_{IVMG}^b \right) \left(\hat{\boldsymbol{\theta}}_i^{*b} - \hat{\boldsymbol{\theta}}_{IVMG}^b \right)' . \quad (29)$$

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 (29) as

$$\begin{aligned} & \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta} + \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{IVMG}^b \right) \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta} + \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{IVMG}^b \right)' = \\ & \sum_{i=1}^N \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i' + \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta}_i \right) \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta}_i \right)' + \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta}_i \right) \boldsymbol{\lambda}_i + \sum_{i=1}^N \boldsymbol{\lambda}_i \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta}_i \right) - \\ & N \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{IVMG}^b \right)' \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{IVMG}^b \right) . \end{aligned} \quad (30)$$

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

$$\begin{aligned} \hat{\Sigma}_{\lambda}^b - \Sigma_{\lambda}^b &= \frac{1}{N-1} \sum_{i=1}^N \left(\boldsymbol{\lambda}_i \boldsymbol{\lambda}_i' - \Sigma_{\lambda}^b \right) + \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta}_i \right) \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta}_i \right)' \\ &+ \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta}_i \right) \boldsymbol{\lambda}_i + \frac{1}{N-1} \sum_{i=1}^N \boldsymbol{\lambda}_i \left(\hat{\boldsymbol{\theta}}_i^{*b} - \boldsymbol{\theta}_i \right) - \\ &\frac{N}{N-1} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{IVMG}^b \right)' \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{IVMG}^b \right) = o_p(1). \end{aligned} \quad (31)$$

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

$$\begin{aligned} \sqrt{N} \left(\hat{\boldsymbol{\theta}}_{IVMG}^b - \boldsymbol{\theta} \right) &\xrightarrow{d} N \left(\mathbf{0}, \Sigma_{\lambda}^b \right); \\ \hat{\Sigma}_{\lambda}^b &\xrightarrow{p} \Sigma_{\lambda}^b. \end{aligned} \quad (32)$$

4 Dynamic heterogeneous panel data model with multifactor error structure

Reconsider model (1) here,

$$\begin{aligned} y_{i,t}^b &= \phi_i y_{i,t-1}^b + \mathbf{x}_{i,t}' \boldsymbol{\beta}_i + u_{i,t}^b \\ &= \mathbf{w}_{i,t}' \boldsymbol{\theta}_i + u_{i,t}^b, \end{aligned} \quad \text{for } i = 1, \dots, N; t = 1, \dots, T, \quad (33)$$

with

$$u_{i,t}^b = \boldsymbol{\gamma}_{y,i}' \mathbf{f}_{y,t} + \varepsilon_{i,t}, \quad (34)$$

where $\gamma_{y,i}$ is a $m_y \times 1$ vector of unobservable factors and $\mathbf{f}_{y,t}$ is a $m_x \times 1$ vector of factor loading and $\varepsilon_{i,t}$ is an idiosyncratic error. Regressor $\mathbf{x}_{i,t}$ follow the process as

$$\mathbf{x}_{i,t} = \mathbf{\Gamma}_{x,i}' \mathbf{f}_{x,t} + \mathbf{v}_{i,t}. \quad (35)$$

Again, we stack the T observations for each i as

$$\mathbf{y}_i = \phi_i \mathbf{y}_i + \mathbf{W}_i \boldsymbol{\theta}_i + \mathbf{u}_i \quad \text{with } \mathbf{u}_i = \mathbf{F}_y \gamma_{y,i} + \boldsymbol{\varepsilon}_i, \quad (36)$$

where $\mathbf{F}_y = (\mathbf{f}_{y,1}, \mathbf{f}_{y,2}, \dots, \mathbf{f}_{y,T})'$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \varepsilon_{i,2}, \dots, \varepsilon_{i,T})'$. And the regressor \mathbf{X}_i as

$$\mathbf{X}_i = \mathbf{F}_x \mathbf{\Gamma}_{x,i} + \mathbf{V}_i, \quad (37)$$

where $\mathbf{F}_x = (\mathbf{f}_{x,1}, \mathbf{f}_{x,2}, \dots, \mathbf{f}_{x,T})'$ and $\mathbf{V}_i = (\mathbf{v}_{i,1}, \mathbf{v}_{i,2}, \dots, \mathbf{v}_{i,T})$. Follow Bai (2003) and Bai (2009), we can use principal components approach to estimate \mathbf{F}_x and $\mathbf{F}_{x,-1}$. Then, we can build the projection matrices as

$$\mathbf{M}_{\hat{\mathbf{F}}_x} = \mathbf{I}_T - \hat{\mathbf{F}}_x \left(\hat{\mathbf{F}}_x' \hat{\mathbf{F}}_x \right)^{-1} \hat{\mathbf{F}}_x'; \quad \mathbf{M}_{\hat{\mathbf{F}}_{x,-1}} = \mathbf{I}_T - \hat{\mathbf{F}}_{x,-1} \left(\hat{\mathbf{F}}_{x,-1}' \hat{\mathbf{F}}_{x,-1} \right)^{-1} \hat{\mathbf{F}}_{x,-1}'. \quad (38)$$

Then, we can project out factor structure in \mathbf{X}_i . Premultiplying \mathbf{X}_i by $\mathbf{M}_{\hat{\mathbf{F}}_x}$, we have $\mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{X}_i = \mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{V}_i$. And we assume \mathbf{V}_i is independent of $\boldsymbol{\varepsilon}$, \mathbf{F}_x , \mathbf{F}_y and $\gamma_{y,i}$. Then, we can see that $E(\mathbf{X}_i' \mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{u}_i) = E(\mathbf{V}_i' \mathbf{M}_{\mathbf{F}_x} \mathbf{u}_i)$. Therefore, using the similar approach in section 1, we can collect the set of instrumental variables:

$$\hat{\mathbf{Z}}_i^j = \left(\mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{X}_i, \mathbf{M}_{\hat{\mathbf{F}}_{x,-1}} \mathbf{X}_{i,-1}, \dots, \mathbf{M}_{\hat{\mathbf{F}}_{x,-j}} \mathbf{X}_{i,-j} \right) \quad (39)$$

Define $\mathbf{P}_i^{(j)}$ as

$$\mathbf{P}_i^{(j)} = \hat{\mathbf{Z}}_i^{(j)} \left(\hat{\mathbf{Z}}_i^{(j)'} \mathbf{M}_{\hat{\mathbf{F}}_x} \hat{\mathbf{Z}}_i^{(j)} \right)^{-1} \hat{\mathbf{Z}}_i^{(j)'} \quad (40)$$

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 show 2SLS estimator of $\boldsymbol{\theta}_i$:

$$\hat{\boldsymbol{\theta}}_i = \left(\mathbf{W}_i' \mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{P}_i^{(j)} \mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{W}_i \right)^{-1} \mathbf{W}_i' \mathbf{P}_i^{(j)} \mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{y}_i, \quad (41)$$

4.1 Optimal number of instrument variables

Again, to select the number of instrument variables, we can follow the approach by Kuersteiner and Okui (2010)

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

$$\boldsymbol{\omega}^{(i)} = \left(\omega_1^{(i)}, \dots, \omega_J^{(i)} \right)', \quad (42)$$

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

$$P_i = \sum_{j=1}^J \omega_j^{(i)} P_i^{(j)}, \quad (43)$$

where J is maximum number of lagged variables that we can observed. Our goal is to select $\omega^{(i)}$ to minimize the approximate mean square error, $S_{\eta^{(i)}}(\omega^{(i)})$, where $\eta^{(i)}$ 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 $D_i = (\tilde{w}_{i,1}, \dots, \tilde{w}_{i,T-1})'$ is a $T \times (1+k)$ matrix. Let \hat{H}_i is some estimator of $H_i = \frac{D_i' D_i}{T}$. Let $\hat{\theta}_i$ is some preliminary estimator, and define the residuals $\hat{u}_i = y_i - W_i \hat{\theta}_i$ that does not depend on the weighting vector. Let $T \times (1+k)$ matrix, V_i , be some preliminary residual from first stage regression, and $v_{\eta,i} = V_i \hat{H}_i^{-1} \eta^{(i)}$.

Define

$$\hat{\sigma}_{u,i}^2 = \frac{\hat{u}_i' \hat{u}_i}{T}, \quad \hat{\sigma}_{\eta,i}^2 = \frac{v_{\eta,i}' v_{\eta,i}}{T}, \quad \hat{\sigma}_{\eta u,i} = \frac{v_{\eta,i}' \hat{u}_i}{T}. \quad (44)$$

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

$$\hat{S}_{\eta^{(i)}}(\omega^{(i)}) = \hat{\sigma}_{\eta u,i}^2 \frac{(K' \omega^{(i)})^2}{T} + \hat{\sigma}_{u,i}^2 \frac{\omega^{(i)'} \hat{U}_i \omega^{(i)} - \hat{\sigma}_{\eta,i}^2 (J - 2K' \omega^{(i)} + \omega^{(i)'} \Gamma_i \omega^{(i)})}{T}. \quad (45)$$

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

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

$$\hat{\theta}_i^* = \left(\tilde{W}_i' P_i(\omega^{(i)*}) \tilde{W}_i \right)^{-1} \tilde{W}_i' P_i(\omega^{(i)*}) \tilde{y}_i. \quad (46)$$

5 Monte Carlo simulation design

It has been found that the IV estimator exhibit systematic bias in the model with heterogeneous slopes 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.

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

The data generating process:

$$\begin{aligned} y_{i,t} &= \phi_i y_{i,t-1} + \beta_{1i} x_{1i,t} + \alpha_i + u_{i,t}, \text{ for } i = 1, \dots, N; t = -49 - (J+1), \dots, T, \\ x_{1i,t} &= \rho x_{1i,t-1} + \tau_\alpha \alpha_i + v_{i,t}, \end{aligned} \quad (47)$$

where $u_{i,t} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, and $v_{i,t} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $\alpha_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $\rho \in \{0, 0.5\}$, $\tau_\alpha = 0.5$, $J = 3, 4, 5, 6$.

The slope coefficients are generated as

$$\phi_i = \phi + \eta_{\phi i}; \beta_{1,i} = \beta_1 + \eta_{\beta_1 i}. \quad (48)$$

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

$$\eta_{\beta_1 i} = (1 - \rho_\beta^2)^{1/2} \eta_{\phi i}. \quad (49)$$

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 $N \in \{25, 50, 100, 200\}$.

5.2 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

$$y_{i,t} = \alpha_i + \phi_i y_{i,t-1} + \sum_{\ell=1}^k \beta_{\ell i} x_{\ell i,t} + u_{i,t}, \text{ for } i = 1, \dots, N; t = -49, \dots, T. \quad (50)$$

$$(51)$$

We allow error factor structure in the model as

$$u_{i,t} = \sum_{s=1}^{m_y} \gamma_{si}^0 f_{s,t}^0 + \varepsilon_{i,t}, \quad (52)$$

where

$$f_{s,t}^0 = \rho_{s,t}^0 f_{s,t-1}^0 + (1 - \rho_{s,t}^0)^{1/2} \zeta_{s,t}, \quad (53)$$

with $\zeta_{s,t} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ for $s = 1, \dots, m_y$. 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

$$\varepsilon_{i,t} = \varsigma_\varepsilon \sigma_{it} (\epsilon_{it} - 1) / \sqrt{2}, \quad (54)$$

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

$$\varsigma_\varepsilon = \frac{\pi_\mu}{1 - \pi_\mu} m_y. \quad (55)$$

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

The process of regressors is

$$x_{\ell i,t} = \mu_{\ell i} + \sum_{\ell=1}^k \phi_{\ell i} x_{\ell i,t-1} + \sum_{s=1}^{m_x} \gamma_{\ell si}^0 f_{s,t}^0 + v_{\ell i,t}, \text{ for } i = 1, \dots, N; t = -49, \dots, T; \ell = 1, 2. \quad (56)$$

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

$$v_{\ell i,t} = \rho_{v,\ell} v_{\ell i,t-1} + (1 - \rho_{v,\ell}^2)^{\frac{1}{2}} \varpi_{\ell i,t}, \text{ for } \ell = 1, 2, \quad (57)$$

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

$$\alpha_i^* \stackrel{i.i.d.}{\sim} N(0, (1 - \rho_i)^2), \mu_{\ell i}^* = \rho_{\mu,\ell} \alpha_i^* + (1 - \rho_{\mu,\ell}^2)^{1/2} \omega_{\ell i}, \quad (58)$$

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

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

$$\begin{aligned} \gamma_{1si}^{0*} &= \rho_{\gamma,1s} \gamma_{3i}^{0*} + (1 - \rho_{\gamma,1s}^2)^{1/2} \xi_{1si}; \xi_{1si} \stackrel{i.i.d.}{\sim} N(0, 1); \\ \gamma_{2si}^{0*} &= \rho_{\gamma,2s} \gamma_{3i}^{0*} + (1 - \rho_{\gamma,2s}^2)^{1/2} \xi_{2si}; \xi_{2si} \stackrel{i.i.d.}{\sim} N(0, 1); \end{aligned} \quad (59)$$

for $s = 1, \dots, m_x = 2$. 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

$$\mathbf{\Gamma} = \mathbf{\Gamma}^0 + \mathbf{\Gamma}_i^{0*} \quad (60)$$

where

$$\mathbf{\Gamma}_i^0 = \begin{bmatrix} \gamma_{1i}^0 & \gamma_{11i}^0 & \gamma_{21i}^0 \\ \gamma_{2i}^0 & \gamma_{12i}^0 & \gamma_{22i}^0 \\ \gamma_{3i}^0 & 0 & 0 \end{bmatrix} \quad (61)$$

and

$$\mathbf{\Gamma}_i^{0*} = \begin{bmatrix} \gamma_{1i}^{0*} & \gamma_{11i}^{0*} & \gamma_{21i}^{0*} \\ \gamma_{2i}^{0*} & \gamma_{12i}^{0*} & \gamma_{22i}^{0*} \\ \gamma_{3i}^{0*} & 0 & 0 \end{bmatrix}. \quad (62)$$

We set

$$\mathbf{\Gamma}^0 = \begin{bmatrix} 1/4 & 1/4 & -1 \\ 1/2 & -1 & 1/4 \\ 1/2 & 0 & 0 \end{bmatrix}. \quad (63)$$

And

$$\alpha_i = \alpha + \alpha_i^*, \mu_{\ell i} = \mu_{\ell} + \mu_{\ell i}^*, \quad (64)$$

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

The slope coefficients are generated as

$$\phi_i = \phi + \eta_{\phi i}, \beta_{1,i} = \beta_1 + \eta_{\beta 1i} \text{ and } \beta_{2,i} = \beta_2 + \eta_{\beta 2i}. \quad (65)$$

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

$$\eta_{\beta \ell i} = [(2c)^2/12] \rho_{\beta} \xi_{\beta \ell i} + (1 - \rho_{\beta}^2)^{1/2} \eta_{\phi i}, \quad (66)$$

where

$$\xi_{\beta\ell i} = \frac{v_{\ell i}^2 - \bar{v}_\ell^2}{\left[N^{-1} \sum_{i=1}^N (v_{\ell i}^2 - \bar{v}_\ell^2)^2 \right]^{1/2}}, \quad (67)$$

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

$$\varsigma_v^2 = \varsigma_\varepsilon^2 \left[SNR - \frac{\rho_v^2}{1 - \rho_v^2} \right] \left(\frac{\beta_1^2 + \beta_2^2}{1 - \rho_v^2} \right)^{-1}, \quad (68)$$

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

Appendix

A Consistency of model average IV estimator

First we consider IV estimator without model average. From previous section, IV estimator can be written as

$$\sqrt{T}(\hat{\theta}_i - \theta_i) = \left(\mathbf{W}_i' \mathbf{M}_{\hat{F}_x} \hat{\mathbf{Z}}_i^{(j)} \left(\hat{\mathbf{Z}}_i^{(j)'} \mathbf{M}_{\hat{F}_x} \hat{\mathbf{Z}}_i^{(j)} \right)^{-1} \hat{\mathbf{Z}}_i^{(j)'} \mathbf{M}_{\hat{F}_x} \mathbf{W}_i \right)^{-1} \times \mathbf{W}_i' \hat{\mathbf{Z}}_i^{(j)} \left(\hat{\mathbf{Z}}_i^{(j)'} \mathbf{M}_{\hat{F}_x} \hat{\mathbf{Z}}_i^{(j)} \right)^{-1} \left(T^{-1/2} \hat{\mathbf{Z}}_i^{(j)'} \mathbf{M}_{\hat{F}_x} \mathbf{u}_i \right), \quad (69)$$

When $N, T \rightarrow \infty$, we can show $T^{-1/2} \hat{\mathbf{Z}}_i^{(j)'} \mathbf{M}_{\hat{F}_x} \mathbf{u}_i$ is $O_p(1)$ as following. Consider the first element of $\hat{\mathbf{Z}}_i^{(j)} = \left(\mathbf{M}_{\hat{F}_x} \mathbf{X}_i, \mathbf{M}_{\hat{F}_{x,-1}} \mathbf{X}_{i,-1}, \dots, \mathbf{M}_{\hat{F}_{x,-j}} \mathbf{X}_{i,-j} \right)$. Then we have

$$T^{-1/2} \hat{\mathbf{X}}_i' \mathbf{M}_{\hat{F}_x} \mathbf{u}_i = T^{-1/2} \hat{\mathbf{X}}_i' \mathbf{M}_{\hat{F}_x^0} \mathbf{u}_i + T^{-1/2} \hat{\mathbf{X}}_i' \left(\mathbf{M}_{\hat{F}_x} - \mathbf{M}_{\hat{F}_x^0} \right) \mathbf{u}_i. \quad (70)$$

$$\begin{aligned} T^{-1/2} \hat{\mathbf{X}}_i' \left(\mathbf{M}_{\hat{F}_x} - \mathbf{M}_{\hat{F}_x^0} \right) \mathbf{u}_i &= -T^{-1/2} \hat{\mathbf{X}}_i' \left(\frac{\hat{\mathbf{F}}_x' \hat{\mathbf{F}}_x}{T} - \mathbf{P}_{F_x^0} \right) \mathbf{u}_i = \\ &= -T^{-1/2} \frac{\mathbf{X}_i' \left(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x \right)}{T} \mathbf{G}_x' \mathbf{F}_x^{0'} \mathbf{u}_i - T^{-1/2} \frac{\mathbf{X}_i' \left(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x \right)}{T} \left(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 - \mathbf{F}_x^0 \mathbf{G}_x \right)' \mathbf{u}_i \\ &= -T^{-1/2} \frac{\mathbf{X}_i' \mathbf{F}_x^0}{T} \mathbf{G}_x \left(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x \right)' \mathbf{u}_i - T^{-1/2} \frac{\mathbf{X}_i' \mathbf{F}_x^0}{T} \left(\mathbf{G}_x \mathbf{G}_x' - \left(\frac{\mathbf{F}_x^{0'} \mathbf{F}_x^0}{T} \right)^{-1} \right) \mathbf{F}_x^{0'} \mathbf{u}_i \\ &= -(b_1 + b_2 + b_3 + b_4). \end{aligned} \quad (71)$$

By Cauchy-Schwarz inequality we know

$$\begin{aligned} |b_1| &\leq T^{1/2} \left\| \frac{\mathbf{X}_i' \left(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x \right)}{T} \right\| \left\| \mathbf{G}_x \right\| \left\| \frac{\mathbf{F}_x^{0'} \mathbf{u}_i}{T} \right\| \leq T^{1/2} \left\| \boldsymbol{\Gamma}_{x,i} \right\| \left\| \frac{\mathbf{F}_x' \left(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x \right)}{T} \right\| \left\| \mathbf{G}_x \right\| \left\| \frac{\mathbf{F}_x^0}{\sqrt{T}} \right\| \left\| \frac{\mathbf{u}_i}{\sqrt{T}} \right\| \\ &+ T^{1/2} \left\| \frac{\mathbf{V}_i' \left(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x \right)}{T} \right\| \left\| \mathbf{G}_x \right\| \left\| \frac{\mathbf{F}_x^0}{\sqrt{T}} \right\| \left\| \frac{\mathbf{u}_i}{\sqrt{T}} \right\| = \sqrt{T} O_p(c_{NT}^{-2}), \end{aligned} \quad (72)$$

where $\left\| \frac{\mathbf{F}_x^{0'} \left(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x \right)}{T} \right\| = O_p(c_{NT}^{-2})$, $\left\| \frac{\mathbf{V}_i' \left(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x \right)}{T} \right\| = O_p(c_{NT}^{-2})$, $\left\| \mathbf{G}_x \right\| = O_p(c_{NT}^{-2})$, $\left\| \boldsymbol{\Gamma}_{x,i} \right\| = O_p(c_{NT}^{-2})$, $\left\| \frac{\mathbf{F}_x^0}{\sqrt{T}} \right\| = O_p(c_{NT}^{-2})$, $\left\| \frac{\mathbf{u}_i}{\sqrt{T}} \right\| \leq \left\| \boldsymbol{\gamma}_i^0 \right\| \frac{\left\| \mathbf{F}_y^0 \right\|}{\sqrt{T}} + \left\| \boldsymbol{\lambda}_i^0 \right\| \frac{\left\| \mathbf{F}_y^0 \right\|}{\sqrt{T}} + \frac{\left\| \varepsilon_i \right\|}{\sqrt{T}} = O_p(c_{NT}^{-2})$.

$$\begin{aligned}
|b_2| &\leq \sqrt{T} \left\| \frac{\mathbf{X}'_i (\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x)}{T} \right\| \left\| \frac{(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 - \mathbf{F}_x^0 \mathbf{G}_x)' \mathbf{u}_i}{T} \right\| \\
&\leq \sqrt{T} \|\Gamma_{x,i}\| \left\| \frac{\mathbf{F}'_x (\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x)}{T} \right\| \left\| \frac{(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 - \mathbf{F}_x^0 \mathbf{G}_x)' \mathbf{u}_i}{T} \right\| + \\
&\quad \sqrt{T} \|\Gamma_{x,i}\| \left\| \frac{\mathbf{V}'_i (\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x)}{T} \right\| \left\| \frac{(\hat{\mathbf{F}}_x - \mathbf{F}_x^0 - \mathbf{F}_x^0 \mathbf{G}_x)' \mathbf{u}_i}{T} \right\| = \sqrt{T} O_p(c_{NT}^{-2}),
\end{aligned} \tag{73}$$

where $\left\| \frac{\mathbf{u}'_i (\hat{\mathbf{F}}_x - \mathbf{F}_x^0 - \mathbf{F}_x^0 \mathbf{G}_x)}{T} \right\| \leq \|\gamma_i^0\| \left\| \frac{\mathbf{V}'_i (\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x)}{T} \right\| + \|\lambda_i^0\| \left\| \frac{\mathbf{F}_y^{0'} (\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x)}{T} \right\| + \left\| \frac{\epsilon'_i (\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x)}{T} \right\|$

$$\begin{aligned}
|b_3| &\leq \sqrt{T} \left\| \frac{\mathbf{X}'_i \mathbf{F}_x^0}{T} \right\| \|\mathbf{G}_x\| \left\| (\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x)' \mathbf{u}_i \right\| \\
&\leq \sqrt{T} \|\Gamma_{x,i}\| \left\| \frac{\mathbf{F}'_x \mathbf{F}_x}{T} \right\| \|\mathbf{G}_x\| \left\| (\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x)' \mathbf{u}_i \right\| + \\
&\quad \sqrt{T} \left\| \frac{\mathbf{V}_i}{\sqrt{T}} \right\| \left\| \frac{\mathbf{F}_x^0}{\sqrt{T}} \right\| \|\mathbf{G}_x\| \left\| (\hat{\mathbf{F}}_x - \mathbf{F}_x^0 \mathbf{G}_x)' \mathbf{u}_i \right\| = \sqrt{T} O_p(c_{NT}^{-2}).
\end{aligned} \tag{74}$$

$$\begin{aligned}
|b_4| &\leq \sqrt{T} \left\| \frac{\mathbf{X}'_i \mathbf{F}_x^0}{T} \right\| \|\mathbf{G}_x \mathbf{G}'_x - \left(\frac{\mathbf{F}_x^0 \mathbf{F}_x^0}{T} \right)^{-1} \left\| \frac{\mathbf{F}_x^{0'} \mathbf{u}_i}{T} \right\| \\
&\quad \sqrt{T} \|\mathbf{G}_x\| \left\| \frac{\mathbf{F}'_x \mathbf{F}_x}{T} \right\| \|\mathbf{G}_x \mathbf{G}'_x - \left(\frac{\mathbf{F}_x^{0'} \mathbf{F}_x^0}{T} \right)^{-1} \left\| \frac{\mathbf{F}_x^{0'} \mathbf{u}_i}{T} \right\| = \sqrt{T} O_p(c_{NT}^{-2}).
\end{aligned} \tag{75}$$

Then, we can see

$$\|T^{-1/2} \hat{\mathbf{X}}'_i (\mathbf{M}_{\hat{\mathbf{F}}_x} - \mathbf{M}_{\hat{\mathbf{F}}_x^0}) \mathbf{u}_i\| = \sqrt{T} O_p(c_{NT}^{-2}) \tag{76}$$

and

$$T^{-1/2} \mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{u}_i = T^{-1/2} \mathbf{X}'_i \mathbf{M}_{\mathbf{F}_x^0} \mathbf{u}_i + \sqrt{T} O_p(c_{NT}^{-2}) \tag{77}$$

Thus, we know when $N, T \rightarrow \infty$, $T^{-1/2} \mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{u}_i$ is $O_p(1)$. Next, we consider the second element of $\hat{\mathbf{Z}}_i^{(j)} = (\mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{X}_i, \mathbf{M}_{\hat{\mathbf{F}}_{x,-1}} \mathbf{X}_{i,-1}, \dots, \mathbf{M}_{\hat{\mathbf{F}}_{x,-j}} \mathbf{X}_{i,-j})$. Then we have

$$\begin{aligned}
T^{-1/2} \mathbf{X}'_{i,-1} \mathbf{M}_{\hat{\mathbf{F}}_{x,-1}} \mathbf{M}_{\mathbf{F}_x^0} \mathbf{u}_i &= T^{-1/2} \mathbf{X}'_{i,-1} \mathbf{M}_{\mathbf{F}_{x,-1}^0} \mathbf{M}_{\mathbf{F}_x^0} \mathbf{u}_i + \\
T^{-1/2} \mathbf{X}'_{i,-1} (\mathbf{M}_{\hat{\mathbf{F}}_{x,-1}} - \mathbf{M}_{\mathbf{F}_{x,-1}^0}) \mathbf{M}_{\mathbf{F}_x^0} \mathbf{u}_i &+ T^{-1/2} \mathbf{X}'_{i,-1} \mathbf{M}_{\hat{\mathbf{F}}_{x,-1}} (\mathbf{M}_{\hat{\mathbf{F}}_x} \mathbf{M}_{\mathbf{F}_x^0}) \mathbf{u}_i
\end{aligned} \tag{78}$$

For the right hand side of above equation, we have

$$\begin{aligned}
T^{-1/2} \mathbf{X}'_{i,-1} \mathbf{M}_{\hat{\mathbf{F}}_{x,-1}} (\mathbf{M}_{\hat{\mathbf{F}}_x} - \mathbf{M}_{\mathbf{F}_x^0}) \mathbf{u}_i &= \\
T^{-1/2} \mathbf{X}'_{i,-1} \mathbf{M}_{\hat{\mathbf{F}}_{x,-1}} (\mathbf{M}_{\hat{\mathbf{F}}_x} - \mathbf{M}_{\mathbf{F}_x^0}) \mathbf{u}_i &+ T^{-1/2} \mathbf{X}'_{i,-1} \mathbf{P}_{\hat{\mathbf{F}}_{x,-1}} (\mathbf{M}_{\hat{\mathbf{F}}_x} - \mathbf{M}_{\mathbf{F}_x^0}) \mathbf{u}_i \\
&= \sqrt{T} O_p(c_{NT}^{-2})
\end{aligned} \tag{79}$$

and

$$\begin{aligned}
& T^{-1/2} \mathbf{X}'_{i,-1} \left(\mathbf{M}_{\hat{F}_{x,-1}} - \mathbf{M}_{F_{x,-1}^0} \right) \mathbf{M}_{F_x^0} \mathbf{u}_i = \\
& T^{-1/2} \mathbf{X}'_{i,-1} \left(\mathbf{M}_{\hat{F}_x} - \mathbf{M}_{F_x^0} \right) \mathbf{u}_i - T^{-1/2} \mathbf{X}'_{i,-1} \left(\mathbf{M}_{\hat{F}_{x,-1}} - \mathbf{M}_{F_{x,-1}^0} \right) \mathbf{P}_{\hat{F}_x^0} \mathbf{u}_i \quad (80) \\
& = \sqrt{T} O_p(c_{NT}^{-2})
\end{aligned}$$

B Asymptotic property of LS estimator

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

$$\hat{\boldsymbol{\theta}}_{LS,i} = \begin{pmatrix} \hat{\phi}_i \\ \hat{\beta}_i \end{pmatrix} = \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1}^2}{T} & \frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{x}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t} \tilde{y}_{i,t-1}}{T} & \frac{\sum_{t=1}^T \tilde{x}_{i,t}^2}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{y}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t} \tilde{y}_{i,t}}{T} \end{pmatrix}, \quad (81)$$

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

$$\begin{pmatrix} \hat{\phi}_i - \phi_i \\ \hat{\beta}_i - \beta_i \end{pmatrix} = \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1}^2}{T} & \frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{x}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t} \tilde{y}_{i,t-1}}{T} & \frac{\sum_{t=1}^T \tilde{x}_{i,t}^2}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{u}_{i,t}}{T} \\ \frac{\sum_{t=1}^T \tilde{x}_{i,t} \tilde{u}_{i,t}}{T} \end{pmatrix}, \quad (82)$$

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

$$A_{\phi i}^{(1)} = \text{plim}_{T \rightarrow \infty} \left(\frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{u}_{i,t}}{T} \right). \quad (83)$$

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

$$\begin{aligned}
A_{\phi i}^{(1)} &= E(y_{i,t-1} - \bar{y}_{i,-1})(u_{i,t} - \bar{u}_i) \\
&= E(y_{i,t-1} u_{i,t}) - E(y_{i,t-1} \bar{u}_i) - E(\bar{y}_{i,-1} u_{i,t}) + E(\bar{y}_{i,-1} \bar{u}_i), \quad (84)
\end{aligned}$$

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

$$y_{i,t} = \frac{\alpha_i}{(1 - \phi_i)} + \sum_{s=0}^{\infty} \beta_i \phi_i^s x_{i,t-s} + \sum_{s=0}^{\infty} \phi_i^s u_{i,t-s}, \quad (85)$$

Then, we have

$$\begin{aligned}
A_{\phi i}^{(1)} &= -E \left(\left(\sum_{s=0}^{\infty} \phi_i^s u_{i,t-s-1} \right) \left(\frac{1}{T} \sum_{t=1}^T u_{i,t} \right) \right) - E \left(\frac{u_{i,t}}{T} \sum_{t=1}^T \sum_{s=0}^{\infty} \phi_i^s u_{i,t-s-1} \right) + \\
&\quad \left(\frac{1}{T} \sum_{t=1}^T \sum_{s=0}^{\infty} \phi_i^s u_{i,t-s-1} \right) \left(\frac{1}{T} \sum_{t=1}^T u_{i,t} \right). \quad (86)
\end{aligned}$$

Hence, from above equation, we have

$$\begin{aligned}
A_{\phi_i}^{(1)} &= -\frac{1}{T}E\left\{(u_{i,t-1} + u_{i,t-2}\phi_i^1 + u_{i,t-3}\phi_i^2 + \dots)(u_{i,1} + \dots + u_{i,t-1} + u_{i,t} + \dots + u_{i,T})\right\} - \\
&\quad \frac{1}{T}E\left\{u_{i,t}\sum_{s=1}^T(u_{i,s-1}\phi_i^0 + u_{i,s-2}\phi_i^1 + \dots + u_{i,s-t-1}\phi_i^t + \dots + u_{i,s-T-1}\phi_i^T + \dots)\right\} + \\
&\quad + \frac{1}{T}E\left\{\left(\sum_{t=1}^T u_{i,t-1}\phi_i^0 + \sum_{t=1}^T u_{i,t-2}\phi_i^1 + \dots + \sum_{t=1}^T u_{i,-1}\phi_i^t + \dots + \sum_{t=1}^T u_{it-T-1}\phi_i^T + \dots\right)\right. \\
&\quad \left.\left(\frac{1}{T}\sum_{t=1}^T u_{i,t}\right)\right\} \\
&= -\frac{\sigma_u^2(1-\phi_i^{t-1})}{T(1-\phi_i)} - \frac{\sigma_u^2(1-\phi_i^{T-t})}{T(1-\phi_i)} + \frac{\sigma_u^2}{T}\left(\frac{1}{1-\phi_i} - \frac{1}{T}\frac{(1-\phi_i^T)}{(1-\phi_i)^2}\right) \\
&= -\frac{\sigma_u^2}{T(1-\phi_i)}\left(1 - \phi_i^{t-1} - \phi_i^{T-t} + \frac{1}{T}\frac{(1-\phi_i^T)}{(1-\phi_i)}\right).
\end{aligned} \tag{87}$$

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

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

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{W}}_i \boldsymbol{\theta}_i + \tilde{\mathbf{u}}_i, \tag{88}$$

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

$$(\phi_i, \beta_i)' = \boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{\lambda}_i, \tag{89}$$

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

$$\hat{\boldsymbol{\theta}}_{LS,i} = \left(\frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{W}}_i}{T}\right)^{-1} \frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{y}}_i}{T}. \tag{90}$$

From above discussion and assumptions, we have following theorem

Theorem 1

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{LS,i} - \boldsymbol{\theta}_i) \xrightarrow{d} N(\mathbf{0}, \mathbf{Q}_i^{-1} \boldsymbol{\Sigma}_{LS,i} \mathbf{Q}_i^{-1}), \tag{91}$$

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 $\mathbf{Q}_i = \text{plim}_{T \rightarrow \infty} T^{-1} \tilde{\mathbf{W}}_i' \tilde{\mathbf{W}}_i$

B.1 Asymptotic property of Mean group LS estimator

Now, we define the mean group estimator of θ :

$$\hat{\theta}_{LSMG} = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_{LSi}. \quad (92)$$

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

$$\begin{aligned} \hat{\theta}_{LSMG} &= N^{-1} \sum_{i=1}^N \left(\frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{W}}_i}{T} \right)^{-1} \frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{y}}_i}{T} \\ &= \bar{\theta} + N^{-1} \sum_{i=1}^N \left(\frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{W}}_i}{T} \right)^{-1} \frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{u}}_i}{T}, \end{aligned} \quad (93)$$

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

$$\text{plim}_{T \rightarrow \infty} \hat{\theta}_{LSMG} = \bar{\theta} + N^{-1} \sum_{i=1}^N \text{plim}_{T \rightarrow \infty} \left(\frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{W}}_i}{T} \right)^{-1} \text{plim}_{T \rightarrow \infty} \left(\frac{\tilde{\mathbf{W}}_i' \tilde{\mathbf{u}}_i}{T} \right) \quad (94)$$

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

$$\text{plim}_{T \rightarrow \infty} \hat{\theta}_{LSMG} = \bar{\theta}. \quad (95)$$

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

$$\text{plim}_{T \rightarrow \infty, N \rightarrow \infty} \hat{\theta}_{LSMG} = \theta. \quad (96)$$

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

$$\hat{\Sigma}_{LS,\lambda} = \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\theta}_{LS,i} - \hat{\theta}_{LSMG} \right) \left(\hat{\theta}_{LS,i} - \hat{\theta}_{LSMG} \right)'. \quad (97)$$

Firstly, we decompose (97) as

$$\begin{aligned} &\sum_{i=1}^N \left(\hat{\theta}_{LS,i} - \theta + \theta - \hat{\theta}_{LSMG} \right) \left(\hat{\theta}_{LS,i} - \theta + \theta - \hat{\theta}_{LSMG} \right)' = \\ &\sum_{i=1}^N \lambda_i \lambda_i' + \sum_{i=1}^N \left(\hat{\theta}_{LS,i} - \theta_i \right) \left(\hat{\theta}_{LS,i} - \theta_i \right)' + \sum_{i=1}^N \left(\hat{\theta}_{LS,i} - \theta_i \right) \lambda_i + \sum_{i=1}^N \lambda_i \left(\hat{\theta}_{LS,i} - \theta_i \right) - \\ &N \left(\theta - \hat{\theta}_{LSMG} \right)' \left(\theta - \hat{\theta}_{LSMG} \right). \end{aligned} \quad (98)$$

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

$$\begin{aligned}\hat{\Sigma}_{LS,\lambda} - \Sigma_{LS,\lambda} &= \frac{1}{N-1} \sum_{i=1}^N \left(\lambda_i \lambda_i' - \Sigma_{LS,\lambda} \right) + \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\theta}_{LS,i} - \theta_i \right) \left(\hat{\theta}_{LS,i} - \theta_i \right)' \\ &+ \frac{1}{N-1} \sum_{i=1}^N \left(\hat{\theta}_{LS,i} - \theta_i \right) \lambda_i + \frac{1}{N-1} \sum_{i=1}^N \lambda_i \left(\hat{\theta}_{LS,i} - \theta_i \right) - \\ &\frac{N}{N-1} \left(\theta - \hat{\theta}_{LSMG} \right)' \left(\theta - \hat{\theta}_{LSMG} \right) = o_p(1).\end{aligned}\tag{99}$$

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

$$\sqrt{N} \left(\hat{\theta}_{LSMG} - \theta \right) \xrightarrow{d} N(0, \Sigma_{LS,\lambda}).\tag{100}$$

C Asymptotic property of IV estimator

From the section 1, we know the IV estimator is

$$\begin{aligned}\hat{\theta}_i^b &= \left(\tilde{A}_{i,T}^{b'} \tilde{B}_{i,T}^{b-1} \tilde{A}_{i,T}^b \right)^{-1} \tilde{A}_{i,T}^{b'} \tilde{B}_{i,T}^{b-1} \tilde{g}_{i,T}^b, \\ \hat{\theta}_i^b &= \theta_i^b + \left(\tilde{A}_{i,T}^{b'} \tilde{B}_{i,T}^{b-1} \tilde{A}_{i,T}^b \right)^{-1} \tilde{A}_{i,T}^{b'} \tilde{B}_{i,T}^{b-1} \left(T^{-1/2} \tilde{Z}_i^{b'} \tilde{u}_i^b \right)\end{aligned}\tag{101}$$

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

$$\hat{\theta}_i^b \xrightarrow{p} \theta_i^b.\tag{102}$$

References

- Bai, J. (2003). *Econometrica. Inferential Theory for Factor Models of Large Dimensions* 71, 135–171.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.
- Chen, X., D. T. Jacho-Chávez, and O. Linton (2016). Averaging of an increasing number of moment condition estimators. *Econometric Theory* 32, 30–70.
- 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. ISEER Discussion Paper No. 1019.
- Okui, R. (2011). Instrumental variable estimation in the presence of many moment conditions. *Journal of Econometrics* 165, 70–86.