

# Instrumental Variables for Dynamic Spatial Models with Interactive Effects

---

Ayden Higgins

November 2025

# Dynamic Spatial Model

- Determinants of economic growth.

# Dynamic Spatial Model

- Determinants of economic growth.
- Observations on a cross-section of countries over multiple time periods.
  - $y_t$  growth rate of real GDP per capita.
  - $X_t$  other factors, e.g. life expectancy, population growth, consumption, investment and government expenditure shares.
  - $W$  binary indicators of whether countries share a border.

# Dynamic Spatial Model

- Determinants of economic growth.
- Observations on a cross-section of countries over multiple time periods.
  - $y_t$  growth rate of real GDP per capita.
  - $X_t$  other factors, e.g. life expectancy, population growth, consumption, investment and government expenditure shares.
  - $W$  binary indicators of whether countries share a border.
- Model

$$y_t = \rho W y_t + \alpha y_{t-1} + \phi W y_{t-1} + X_t \beta + \eta_t.$$

# Dynamic Spatial Model

- Determinants of economic growth.
- Observations on a cross-section of countries over multiple time periods.
  - $y_t$  growth rate of real GDP per capita.
  - $X_t$  other factors, e.g. life expectancy, population growth, consumption, investment and government expenditure shares.
  - $W$  binary indicators of whether countries share a border.
- Model

$$y_t = \rho W y_t + \alpha y_{t-1} + \phi W y_{t-1} + X_t \beta + \eta_t.$$

- Parameter of interest  $\theta := (\rho, \alpha, \phi, \beta^\top)^\top$ .

# Interactive Effects

- Controlling for unobserved heterogeneity critically important.

# Interactive Effects

- Controlling for unobserved heterogeneity critically important.
- Exposure to common shocks or a common environment can generate correlation in outcomes.

# Interactive Effects

- Controlling for unobserved heterogeneity critically important.
- Exposure to common shocks or a common environment can generate correlation in outcomes.
- The error term  $\eta_t$  may be decomposed as

$$\eta_t = \Lambda f_t + \varepsilon_t.$$



# Interactive Effects

- Controlling for unobserved heterogeneity critically important.
- Exposure to common shocks or a common environment can generate correlation in outcomes.
- The error term  $\eta_t$  may be decomposed as

$$\eta_t = \Lambda f_t + \varepsilon_t.$$

- Note  $\Lambda$ ,  $f_t$  and  $\varepsilon_t$  are all unobserved.

# Interactive Effects

- Controlling for unobserved heterogeneity critically important.
- Exposure to common shocks or a common environment can generate correlation in outcomes.
- The error term  $\eta_t$  may be decomposed as

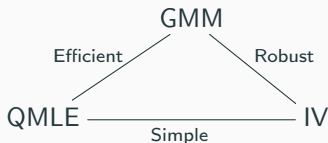
$$\eta_t = \Lambda \mathbf{f}_t + \varepsilon_t.$$

- Note  $\Lambda$ ,  $\mathbf{f}_t$  and  $\varepsilon_t$  are all unobserved.
- Generalisation of classic models such as individual, time or group effects:

$$\Lambda = \begin{pmatrix} \lambda_1 & 1 \\ \vdots & \vdots \\ \lambda_n & 1 \end{pmatrix}, \quad \mathbf{f}_t = \begin{pmatrix} 1 \\ f_t \end{pmatrix}.$$

- $\Lambda \mathbf{F}^\top$  is low rank.

- Spatial models with interactive effects:
  - (Q)MLE - [Shi and Lee \(2017\)](#); [Bai and Li \(2021\)](#). (Large  $n$ , Large  $T$ )
  - Adjusted Score - [Li and Yang \(2023\)](#). (Large  $n$ , Fixed  $T$ )
  - GMM - [Kuersteiner and Prucha \(2020\)](#). (Large  $n$ , Fixed  $T$ )
  - IV - [Cui, Sarafidis and Yamagata \(2022\)](#). (Large  $n$ , Large  $T$ )



- Introduce an IV estimator which provides an easier way to handle both the factor structure and also possible heteroskedasticity of the errors.

# Overview

- Introduce an IV estimator which provides an easier way to handle both the factor structure and also possible heteroskedasticity of the errors.
- Study the properties of the estimator as  $n \rightarrow \infty$ , but also allowing for the possibility that both the number of time periods  $T$  and the number of instruments  $m$  also increase.

# Overview

- Introduce an IV estimator which provides an easier way to handle both the factor structure and also possible heteroskedasticity of the errors.
- Study the properties of the estimator as  $n \rightarrow \infty$ , but also allowing for the possibility that both the number of time periods  $T$  and the number of instruments  $m$  also increase.
- Find that the estimator can exhibit considerable bias, depending on the weights matrix. Bias correction significantly ameliorates this issue.

# Overview

- Introduce an IV estimator which provides an easier way to handle both the factor structure and also possible heteroskedasticity of the errors.
- Study the properties of the estimator as  $n \rightarrow \infty$ , but also allowing for the possibility that both the number of time periods  $T$  and the number of instruments  $m$  also increase.
- Find that the estimator can exhibit considerable bias, depending on the weights matrix. Bias correction significantly ameliorates this issue.
- Provide some additional results which are useful for applying the estimator in practice.

- Introduce an IV estimator which provides an easier way to handle both the factor structure and also possible heteroskedasticity of the errors.
- Study the properties of the estimator as  $n \rightarrow \infty$ , but also allowing for the possibility that both the number of time periods  $T$  and the number of instruments  $m$  also increase.
- Find that the estimator can exhibit considerable bias, depending on the weights matrix. Bias correction significantly ameliorates this issue.
- Provide some additional results which are useful for applying the estimator in practice.
- Apply the method the study the relationship between economic growth, civil liberties and political rights.



- Original model in matrices:

$$\mathbf{S}(\rho)\mathbf{Y} = \alpha\mathbf{Y}_{-1} + \phi\mathbf{W}\mathbf{Y}_{-1} + \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{F}^\top + \boldsymbol{\varepsilon},$$

with  $\mathbf{S}(\rho) := \mathbf{I}_n - \rho\mathbf{W}$ .

- Original model in matrices:

$$\mathbf{S}(\rho)\mathbf{Y} = \alpha\mathbf{Y}_{-1} + \phi\mathbf{W}\mathbf{Y}_{-1} + \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{F}^\top + \boldsymbol{\varepsilon},$$

with  $\mathbf{S}(\rho) := \mathbf{I}_n - \rho\mathbf{W}$ .

- Under some conditions we can expand

$$\mathbf{Y} = \sum_{h=0}^{\infty} (\rho\mathbf{W})^h (\alpha\mathbf{Y}_{-1} + \phi\mathbf{W}\mathbf{Y}_{-1} + \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{F}^\top + \boldsymbol{\varepsilon}).$$

- Original model in matrices:

$$\mathbf{S}(\rho)\mathbf{Y} = \alpha\mathbf{Y}_{-1} + \phi\mathbf{W}\mathbf{Y}_{-1} + \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{F}^\top + \boldsymbol{\varepsilon},$$

with  $\mathbf{S}(\rho) := \mathbf{I}_n - \rho\mathbf{W}$ .

- Under some conditions we can expand

$$\mathbf{Y} = \sum_{h=0}^{\infty} (\rho\mathbf{W})^h (\alpha\mathbf{Y}_{-1} + \phi\mathbf{W}\mathbf{Y}_{-1} + \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{F}^\top + \boldsymbol{\varepsilon}).$$

- Generate instruments for  $\mathbf{Y}$ , i.e.  $\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}, \dots$

- With  $S$  weights matrices

$$\mathbf{y}_t = \sum_{s=1}^S \rho_s \mathbf{W}_s \mathbf{y}_t + \alpha \mathbf{y}_{t-1} + \sum_{s=1}^S \phi_s \mathbf{W}_s \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\eta}_t.$$

- Produces a great many instruments

$$\begin{aligned} S^{-1}(\boldsymbol{\rho}) &= (\mathbf{I}_n - \boldsymbol{\rho} \cdot \mathbf{W})^{-1} \\ &= \mathbf{I}_n + \sum_{s=1}^S \rho_{s_1} \mathbf{W}_{s_1} + \sum_{s_1=1}^S \sum_{s_2=1}^S \rho_{s_1} \rho_{s_2} \mathbf{W}_{s_1} \mathbf{W}_{s_2} \\ &\quad + \sum_{s_1=1}^S \sum_{s_2=1}^S \sum_{s_3=1}^S \rho_{s_1} \rho_{s_2} \rho_{s_3} \mathbf{W}_{s_1} \mathbf{W}_{s_2} \mathbf{W}_{s_3} + \cdots . \end{aligned}$$

- Re-write

$$\begin{aligned} Y &= \rho WY + \alpha Y_{-1} + \phi WY_{-1} + X \cdot \beta + \Lambda F^\top + \epsilon, \\ &= Z \cdot \theta + \Lambda F^\top + \epsilon. \end{aligned}$$

- Re-write

$$\begin{aligned} Y &= \rho WY + \alpha Y_{-1} + \phi WY_{-1} + X \cdot \beta + \Lambda F^\top + \epsilon, \\ &= Z \cdot \theta + \Lambda F^\top + \epsilon. \end{aligned}$$

- Let  $\mathcal{V}$  be an  $n \times m$  matrix containing a set of instruments.

- Re-write

$$\begin{aligned} Y &= \rho WY + \alpha Y_{-1} + \phi WY_{-1} + X \cdot \beta + \Lambda F^\top + \varepsilon, \\ &= Z \cdot \theta + \Lambda F^\top + \varepsilon. \end{aligned}$$

- Let  $\mathcal{V}$  be an  $n \times m$  matrix containing a set of instruments.
- Construct an  $n \times m$  matrix  $Q_{\mathcal{V}}$  as  $\mathcal{V}(\mathcal{V}^\top \mathcal{V})^{-\frac{1}{2}}$ .

- Re-write

$$\begin{aligned} Y &= \rho WY + \alpha Y_{-1} + \phi WY_{-1} + X \cdot \beta + \Lambda F^\top + \varepsilon, \\ &= Z \cdot \theta + \Lambda F^\top + \varepsilon. \end{aligned}$$

- Let  $\mathcal{V}$  be an  $n \times m$  matrix containing a set of instruments.
- Construct an  $n \times m$  matrix  $Q_{\mathcal{V}}$  as  $\mathcal{V}(\mathcal{V}^\top \mathcal{V})^{-\frac{1}{2}}$ .
- Premultiply by  $Q_{\mathcal{V}}^\top$  to give

$$Q_{\mathcal{V}}^\top Y = \tilde{Y} = \tilde{Z} \cdot \theta + \tilde{\Lambda} F^\top + \tilde{\varepsilon}.$$

- Dimension reduction:  $n \times T \rightarrow m \times T$ .



- Concentrated objective function:

$$\mathcal{Q}(\boldsymbol{\theta}) := \frac{1}{nT} \sum_{r=R+1}^T \mu_r \left( \left( \tilde{\mathbf{Y}} - \tilde{\mathbf{Z}} \cdot \boldsymbol{\theta} \right)^\top \left( \tilde{\mathbf{Y}} - \tilde{\mathbf{Z}} \cdot \boldsymbol{\theta} \right) \right),$$

where  $\mu_r(\cdot)$  denotes the  $r$ -th largest eigenvalue of a matrix.

- Concentrated objective function:

$$\mathcal{Q}(\boldsymbol{\theta}) := \frac{1}{nT} \sum_{r=R+1}^T \mu_r \left( \left( \tilde{\mathbf{Y}} - \tilde{\mathbf{Z}} \cdot \boldsymbol{\theta} \right)^\top \left( \tilde{\mathbf{Y}} - \tilde{\mathbf{Z}} \cdot \boldsymbol{\theta} \right) \right),$$

where  $\mu_r(\cdot)$  denotes the  $r$ -th largest eigenvalue of a matrix.

- IV-IFE estimator:

$$\hat{\boldsymbol{\theta}} := \arg \min_{\boldsymbol{\theta} \in \Theta} \mathcal{Q}(\boldsymbol{\theta}).$$

## Assumption MD (Model)

- (i) The parameter vector  $\theta_0$  is in the interior of  $\Theta$ , where  $\Theta$  is a compact subset of  $\mathbb{R}^P$ .
- (ii) The weights matrix  $\mathbf{W}$  has a zero diagonal, is nonstochastic and UB.
- (iii) For all  $\rho \in \Theta_\rho$ ,  $\alpha \in \Theta_\alpha$  and  $\phi \in \Theta_\phi$ ,  $|\det(\mathbf{S}(\rho))| \geq c > 0$ ,  $|\det(\bar{\mathbf{B}}(\rho, \alpha, \phi))| \geq c > 0$ , and  $\|\mathbf{A}(\rho, \alpha, \phi)\|_2 < 1 - c$  holds for all  $(n, T)$ , and  $\mathbf{S}^{-1}(\rho)$ ,  $\bar{\mathbf{B}}^{-1}(\rho, \alpha, \phi)$  and  $\sum_{h=1}^{\infty} |\mathbf{A}^h(\rho, \alpha, \phi)|$  are UB.
- (iv)  $x_{kit}$ ,  $\lambda_{0,ir}$  and  $f_{0,tr}$  have uniformly bounded fourth moments.
- (v) The errors  $\varepsilon_{it}$  are independent of the factors, the loadings, and the covariates, and are also independent over  $i$  and  $t$  with  $\mathbb{E}[\varepsilon_{it}] = 0$ ,  $\mathbb{E}[\varepsilon_{it}^2] =: \sigma_{it}^2 > 0$  and uniformly bounded fourth moments.

## Assumption CS (Consistency)

- (i)  $R \geq R_0 := \text{rank}(\tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top)$ .
- (ii)  $\min_{\boldsymbol{\delta} \in \mathbb{R}^P: \|\boldsymbol{\delta}\|_2=1} \sum_{r=R+R_0+1}^T \mu_r\left(\frac{1}{nT}(\boldsymbol{\delta} \cdot \tilde{\mathbf{Z}})^\top (\boldsymbol{\delta} \cdot \tilde{\mathbf{Z}})\right) \geq c > 0$ ,  
w.p.a.1 as  $n \rightarrow \infty$ .

## Proposition 1 (Consistency)

*Under Assumptions MD and CS, as  $n \rightarrow \infty$ ,*

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = \mathcal{O}_p \left( \sqrt{\frac{m}{n}} \right).$$

## Assumption AD (Asymptotic Distribution)

- (i)  $R = R_0$ .
- (ii)  $\frac{1}{n} \tilde{\Lambda}_0^\top \tilde{\Lambda}_0 \xrightarrow{p} \Sigma_{\tilde{\Lambda}_0}$  as  $n \rightarrow \infty$ , with  $\mu_{R_0}(\Sigma_{\tilde{\Lambda}_0}) > 0$  and  $\mu_1(\Sigma_{\tilde{\Lambda}_0}) < \infty$ .
- (iii)  $\mu_{R_0}(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0) > 0$  and  $\mu_1(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0) < \infty$ .
- (iv) The elements of the matrices  $\bar{M} \tilde{\mathcal{H}}$  and  $\bar{M} \tilde{\mathcal{R}}$  have uniformly bounded fourth moments for all  $(n, T)$ .

The matrices  $\mathcal{H}$  and  $\mathcal{R}$  are  $nm \times P$  and  $\bar{M} := (M_{\mathbf{F}_0} \otimes Q_{\mathbf{V}} M_{\tilde{\Lambda}_0} Q_{\mathbf{V}}^\top)$ .

# Asymptotic Distribution I

## Theorem 1 (Asymptotic Distribution)

*Under Assumptions MD, CS and AD, with  $\gamma_{nm}^2 := m^2 T/n \rightarrow c \geq 0$  as  $n, m \rightarrow \infty$ ,*

$$\sqrt{nT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\boldsymbol{\Delta}^{-1}\boldsymbol{\psi}, \boldsymbol{\Delta}^{-1}\boldsymbol{\Omega}\boldsymbol{\Delta}^{-1}),$$

*where,*

$$\boldsymbol{\psi}_n := \frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\boldsymbol{\Sigma}\bar{\mathbf{M}}\bar{\mathbf{W}}\bar{\mathbf{B}}^{-1}) \\ \mathbf{0}_{(K+2)} \end{pmatrix},$$

$$\boldsymbol{\psi} := \text{plim}_{n \rightarrow \infty} \boldsymbol{\psi}_n, \quad \boldsymbol{\Delta}_n := (nT)^{-1}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})^\top \bar{\mathbf{M}}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}}),$$

$$\boldsymbol{\Delta} = \text{plim}_{n \rightarrow \infty} \boldsymbol{\Delta}_n, \quad \boldsymbol{\Omega}_n := (nT)^{-1}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})^\top \bar{\mathbf{M}}\boldsymbol{\Sigma}\bar{\mathbf{M}}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}}),$$

$$\boldsymbol{\Omega} := \text{plim}_{n \rightarrow \infty} \boldsymbol{\Omega}_n, \text{ and } \boldsymbol{\Sigma} \text{ is an } nT \times nT \text{ matrix with diagonal elements } \sigma_{11}^2, \dots, \sigma_{nT}^2 \text{ and remaining elements equal to zero.}$$

## Asymptotic Distribution II

- Consider more closely

$$\psi_n := \frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\Sigma \bar{W} \bar{B}^{-1} \bar{M}) \\ \mathbf{0}_{(K+2) \times 1} \end{pmatrix}.$$



# Asymptotic Distribution II

- Consider more closely

$$\psi_n := \frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\Sigma \bar{W} \bar{B}^{-1} \bar{M}) \\ \mathbf{0}_{(K+2) \times 1} \end{pmatrix}.$$

- In general

$$\psi_{n,1} = \mathcal{O} \left( \sqrt{\frac{m^2 T}{n}} \right).$$

# Asymptotic Distribution II

- Consider more closely

$$\psi_n := \frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\Sigma \bar{W} \bar{B}^{-1} \bar{M}) \\ \mathbf{0}_{(K+2) \times 1} \end{pmatrix}.$$

- In general

$$\psi_{n,1} = \mathcal{O} \left( \sqrt{\frac{m^2 T}{n}} \right).$$

- The potential magnitude of this bias lies behind the restrictions on the growth of  $n$ ,  $m$  and  $T$ .

# Asymptotic Distribution II

- Consider more closely

$$\psi_n := \frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\Sigma \bar{W} \bar{B}^{-1} \bar{M}) \\ \mathbf{0}_{(K+2) \times 1} \end{pmatrix}.$$

- In general

$$\psi_{n,1} = \mathcal{O} \left( \sqrt{\frac{m^2 T}{n}} \right).$$

- The potential magnitude of this bias lies behind the restrictions on the growth of  $n$ ,  $m$  and  $T$ .
- The exact magnitude of this bias will depend in part on the structure of cross-sectional dependence, represented by the weights matrix.

# Asymptotic Distribution III

- Suppose that:

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 & 0 & \dots & 0 \\ 0 & \mathbf{W}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{W}_G \end{pmatrix},$$

with

$$\mathbf{W}_g = \begin{pmatrix} 0 & n_g^{-1} & \dots & n_g^{-1} \\ n_g^{-1} & 0 & \dots & n_g^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ n_g^{-1} & n_g^{-1} & \dots & 0 \end{pmatrix},$$

for  $g = 1, \dots, G$ .

- One can show (with  $n \rightarrow \infty$ ,  $m/n$  and  $T/n \rightarrow 0$ )

$$\begin{aligned}\psi_{n,1} &= \frac{1}{\sqrt{nT}} \text{tr}(\Sigma \bar{W} (I_{nT} - \rho_0 \bar{W})^{-1} \bar{P}) + o(1) \\ &=: \xi + o(1),\end{aligned}$$

and establish bounds

$$-\sigma_{\min}^2 \sqrt{\frac{T}{n}} \times \frac{m}{n_{\min} + (\rho - 1)} \leq \xi \leq \sigma_{\max}^2 \sqrt{\frac{T}{n}} \left( \frac{m \wedge G}{1 - \rho} - \frac{0 \vee (m - G)}{n_{\max} + (\rho - 1)} \right),$$

where  $n_{\max}$  and  $n_{\min}$  denote maximum and minimum group sizes, respectively.

- If  $n, m, G \rightarrow \infty$  and  $n_{\max}, n_{\min}$  remain fixed

$$\psi_{n,1} = \mathcal{O}\left(\sqrt{\frac{m^2 T}{n}}\right).$$

# Asymptotic Distribution III

- If  $n, m, G \rightarrow \infty$  and  $n_{\max}, n_{\min}$  remain fixed

$$\psi_{n,1} = \mathcal{O} \left( \sqrt{\frac{m^2 T}{n}} \right).$$

- If  $n, m \rightarrow \infty$ ,  $G$  remains fixed and  $m/n_{\min} \rightarrow 0$

$$\psi_{n,1} = \mathcal{O} \left( \sqrt{\frac{T}{n}} \right).$$

# Asymptotic Distribution III

- If  $n, m, G \rightarrow \infty$  and  $n_{\max}, n_{\min}$  remain fixed

$$\psi_{n,1} = \mathcal{O} \left( \sqrt{\frac{m^2 T}{n}} \right).$$

- If  $n, m \rightarrow \infty$ ,  $G$  remains fixed and  $m/n_{\min} \rightarrow 0$

$$\psi_{n,1} = \mathcal{O} \left( \sqrt{\frac{T}{n}} \right).$$

- A spectrum of cases lie in between.



# Asymptotic Distribution III

- If  $n, m, G \rightarrow \infty$  and  $n_{\max}, n_{\min}$  remain fixed

$$\psi_{n,1} = \mathcal{O} \left( \sqrt{\frac{m^2 T}{n}} \right).$$

- If  $n, m \rightarrow \infty$ ,  $G$  remains fixed and  $m/n_{\min} \rightarrow 0$

$$\psi_{n,1} = \mathcal{O} \left( \sqrt{\frac{T}{n}} \right).$$

- A spectrum of cases lie in between.
- Direct relationship between the properties of the estimator and the structure of the weights matrix.

- Construct a bias correct estimator

$$\tilde{\boldsymbol{\theta}} := \hat{\boldsymbol{\theta}} - \frac{1}{\sqrt{nT}} \hat{\boldsymbol{\Delta}}^{-1} \hat{\boldsymbol{\psi}}.$$

- Under assumptions MD, CS and AD, as  $n, m \rightarrow \infty$  with  $m^2 T/n \rightarrow c \geq 0$

$$\sqrt{nT}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{P \times 1}, \boldsymbol{\Delta}^{-1} \boldsymbol{\Omega} \boldsymbol{\Delta}^{-1}).$$

# Illustration I

Design:

$$\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \alpha \mathbf{Y}_{-1} + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \mathbf{\Lambda} \mathbf{F}^\top + \boldsymbol{\varepsilon},$$

with

- $R_0 = 2, R = 2, f_{tr}, \lambda_{ir} \sim \mathcal{N}(0, 1),$
- $\mathbf{X}_1 = \mathbf{\Lambda} \mathbf{F}^\top + \mathbf{e},$  where  $e_{it} \sim \mathcal{N}(0, 1),$
- $X_{2,it} \sim \mathcal{N}(0, 1),$
- $\text{vec}(\boldsymbol{\varepsilon}) := \boldsymbol{\Sigma}^{\frac{1}{2}} \text{vec}(\mathbf{u}), u_{it} \sim \mathcal{N}(0, 1), \boldsymbol{\Sigma}$  diagonal with elements  $\sigma_{it}^2 \in (0, 2),$
- $\mathcal{V} := (\mathbf{X}_1, \mathbf{X}_2, \mathbf{W} \mathbf{X}_1, \mathbf{W} \mathbf{X}_2).$
- Notice that  $m = \mathcal{O}(T).$

## Illustration II

- Partition the cross-section into  $G$  disjoint groups.
- Within each group all units are connected only to a single central unit who reciprocates the link.
- Produces a block diagonal weights matrix representing multiple stars.
- E.g. with  $G = 1$

$$\mathbf{W} = \begin{pmatrix} 0 & \boldsymbol{\iota}_{n-1}^\top \\ \boldsymbol{\iota}_{n-1} & \mathbf{0}_{n-1 \times n-1} \end{pmatrix},$$

before being row-normalised.

- Easy to verify that  $\text{rank}(\mathbf{W}) = 2G$  whereby

$$\psi_{n,1} = \mathcal{O} \left( (m \wedge G) \times \sqrt{\frac{T}{n}} \right) = \mathcal{O} \left( (T \wedge G) \times \sqrt{\frac{T}{n}} \right).$$

## Illustration III

**Table 1a:** Coverage 95% Confidence Intervals -  $G = 5$

$n \setminus T$	$\hat{\rho}$			$\tilde{\rho}$		
	6	9	12	6	9	12
100	0.902	0.843	0.853	0.932	0.951	0.964
300	0.928	0.921	0.916	0.953	0.944	0.960
500	0.939	0.939	0.929	0.954	0.949	0.947

**Table 1b:** Coverage 95% Confidence Intervals -  $G = 25$

$n \setminus T$	$\hat{\rho}$			$\tilde{\rho}$		
	6	9	12	6	9	12
100	0.815	0.798	0.353	0.956	0.951	0.962
300	0.918	0.892	0.698	0.956	0.961	0.961
500	0.935	0.913	0.861	0.948	0.965	0.960

- IV-IFE estimator is based the moment condition

$$\mathbb{E} \left[ (M_{F_0} \otimes Q_{\mathbf{V}} M_{\tilde{\Lambda}_0} Q_{\mathbf{V}}^\top) \text{vec}(\eta) \right] = \mathbf{0}_{nT}.$$

- A total of  $(T - R_0)(m - R_0)$  linearly independent restrictions.
- Motivates the following statistic:

$$\mathcal{J} := \text{vec}(\hat{\eta})^\top (M_{\hat{F}} \otimes Q_{\mathbf{V}} M_{\hat{\Lambda}} Q_{\mathbf{V}}^\top) \text{vec}(\hat{\eta}),$$

with  $\hat{\eta} := \mathbf{y} - \mathbf{Z} \cdot \hat{\boldsymbol{\theta}}$ .

- Define

$$\mathbf{M}_{\mathcal{J}} := \bar{\mathbf{M}} - \mathbf{P}_{\mathcal{J}},$$

$$\mathbf{P}_{\mathcal{J}} := \bar{\mathbf{M}}(\mathcal{H} + \mathcal{R})((\mathcal{H} + \mathcal{R})^{\top} \bar{\mathbf{M}}(\mathcal{H} + \mathcal{R}))^{-1}(\mathcal{H} + \mathcal{R})^{\top} \bar{\mathbf{M}},$$

$$\ell := (T - R_0)(m - R_0) - P,$$

$$\begin{aligned} \sigma_{\mathcal{J}}^2 &:= \text{tr}((\mathcal{M}^{(4)} - 3\Sigma^2)(\mathbf{M}_{\mathcal{J}} \odot \mathbf{M}_{\mathcal{J}})) \\ &\quad + 2\boldsymbol{\iota}_{nT}^{\top}(\Sigma(\mathbf{M}_{\mathcal{J}} \odot \mathbf{M}_{\mathcal{J}})\Sigma)\boldsymbol{\iota}_{nT}, \end{aligned}$$

where  $\mathcal{M}^{(4)}$  is an  $nT \times nT$  matrix with diagonal elements  $\mathbb{E}[\varepsilon_{11}^4], \dots, \mathbb{E}[\varepsilon_{nT}^4]$  and all remaining elements equal to zero.

## Assumption JS (*J*-Test)

- (i) The errors  $\varepsilon_{it}$  have uniformly bound eighth moments.
- (ii)  $\ell^{-1}\sigma_{\mathcal{J}}^2 \geq c > 0$  w.p.a.1.

## Theorem 2 (*J*-Test)

*Under Assumptions MD, CS, AD and JS, with  $\gamma_{nm}^2 \rightarrow c \geq 0$  as  $n, \ell \rightarrow \infty$ ,*

$$\frac{\mathcal{J} - \varphi_{\mathcal{J}}}{\sigma_{\mathcal{J}}} \xrightarrow{d} \mathcal{N}(0, 1),$$

*where*

$$\varphi_{\mathcal{J}} := \text{tr}(\Sigma \mathbf{M}_{\mathcal{J}}) - \psi_n^{\top} \Delta_n^{-1} (\mathcal{H} + \mathcal{R})^{\top} \bar{M} (\mathcal{H} + \mathcal{R}) \Delta_n^{-1} \psi_n.$$

- E.g. if  $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ ,  $\text{tr}(\Sigma \mathbf{M}_{\mathcal{J}}) = \ell$  and  $\sigma_{\mathcal{J}}^2 = 2\ell$ .



## Specification Testing IV

- This result can also be used to test for correct number of factors.

## Specification Testing IV

- This result can also be used to test for correct number of factors.
- For simplicity assume that  $\varepsilon_{it} \sim \mathcal{N}(0, 1)$  and  $\rho_0 = 0$  such that  $\varphi_{\mathcal{J}} = \ell$  and  $\sigma_{\mathcal{J}}^2 = 2\ell$ .

## Specification Testing IV

- This result can also be used to test for correct number of factors.
- For simplicity assume that  $\varepsilon_{it} \sim \mathcal{N}(0, 1)$  and  $\rho_0 = 0$  such that  $\varphi_{\mathcal{J}} = \ell$  and  $\sigma_{\mathcal{J}}^2 = 2\ell$ .
- Then if  $R < R_0$  it can be shown that under the assumptions of Proposition 2 (except AD(i))

$$\xi_R := \frac{\mathcal{J}_R - \ell_R}{\sqrt{2\ell_R}} \rightarrow \infty.$$

# Specification Testing IV

- This result can also be used to test for correct number of factors.
- For simplicity assume that  $\varepsilon_{it} \sim \mathcal{N}(0, 1)$  and  $\rho_0 = 0$  such that  $\varphi_{\mathcal{J}} = \ell$  and  $\sigma_{\mathcal{J}}^2 = 2\ell$ .
- Then if  $R < R_0$  it can be shown that under the assumptions of Proposition 2 (except AD(i))

$$\xi_R := \frac{\mathcal{J}_R - \ell_R}{\sqrt{2\ell_R}} \rightarrow \infty.$$

- Then,

$$\hat{R} := \min_{R=1, \dots, R_{\max}} \{R : \xi_r \geq c_{1-\delta_n} \text{ } r = 1, \dots, R-1, \xi_r < c_{1-\delta_n}\},$$

is a consistent estimator of  $R_0$ , with  $c_{1-\delta_n}$  being the  $100(1 - \delta_n)$  percentile of  $\Phi(x)$  and  $\delta_n \rightarrow 0$ ,  $\ln(\delta) = \mathcal{O}(m/nT)$  as  $n, m \rightarrow \infty$  and  $m^2T/n \rightarrow c \geq 0$ .

# Spatially/Serially Correlated Errors I

- Can extend the model of cross-sectional and serial dependence to the error term.
- Suppose the errors are generated according to

$$\epsilon_t = \rho_\epsilon \mathbf{W} \epsilon_t + \alpha_\epsilon \epsilon_{t-1} + \phi_\epsilon \mathbf{W} \epsilon_{t-1} + u_t.$$

## Assumption ER (Error)

1. The vector  $\boldsymbol{\theta}_{\varepsilon,0} := (\rho_\varepsilon, \alpha_\varepsilon, \phi_\varepsilon)^\top$  lies in the interior of  $\Theta_\varepsilon$ , where  $\Theta_\varepsilon$  is a compact subset of  $\mathbb{R}^3$  in which  $\inf_{\theta_\varepsilon \in \Theta_\varepsilon} \det(\mathbf{S}_\varepsilon(\rho_\varepsilon)) \neq 0$  and  $\inf_{\theta_\varepsilon \in \Theta_\varepsilon} \det(\mathbf{B}_\varepsilon(\boldsymbol{\theta}_\varepsilon)) \neq 0$ ,  $\mathbf{S}_\varepsilon^{-1}(\rho_\varepsilon)$ , and  $\bar{\mathbf{B}}_\varepsilon^{-1}(\boldsymbol{\theta}_\varepsilon)$  are UB, and  $\|\mathbf{A}_\varepsilon^h(\boldsymbol{\theta}_\varepsilon)\|_2 < 1 - c$  for some  $c > 0$ .
2. The errors  $u_{it}$  are independent of the factors, the loadings, and the covariates, and are also independent over  $i$  and  $t$  with  $\mathbb{E}[u_{it}] = 0$ ,  $\mathbb{E}[u_{it}^2] =: \sigma_{u,it}^2 > 0$  and uniformly bounded fourth moments.

# Spatially/Serially Correlated Errors III

## Assumption IC (Initial Condition)

Assume the initial conditions are generated as

$$\mathbf{y}_0 = \Sigma_{y_0}^{\frac{1}{2}} \boldsymbol{\nu}_1$$

$$\boldsymbol{\varepsilon}_0 = \Sigma_{\varepsilon_0}^{\frac{1}{2}} \boldsymbol{\nu}_2$$

where  $\{\nu_{1,j}, \nu_{2,j}\}$  are independent of the independent of the factors, the loadings, the covariates, and the errors  $\mathbf{u}$ , are also independent over  $j$  with  $\mathbb{E}[\nu_{i,1}] = \mathbb{E}[\nu_{i,2}] = 0$ ,  $\mathbb{E}[\nu_{i,1}^2] = \mathbb{E}[\nu_{i,2}^2] = 1$  and uniformly bounded fourth moments. Moreover,

$$\mathbb{E} \left[ \begin{pmatrix} \mathbf{y}_0 \\ \boldsymbol{\varepsilon}_0 \end{pmatrix} \begin{pmatrix} \mathbf{y}_0 \\ \boldsymbol{\varepsilon}_0 \end{pmatrix}^\top \right] := \begin{pmatrix} \Sigma_{\varepsilon_0} & \Sigma_{\varepsilon_0 y_0} \\ \Sigma_{\varepsilon_0 y_0} & \Sigma_{y_0} \end{pmatrix} =: \Sigma_0,$$

where  $\Sigma_0$  is UB with  $\mu_{\min}(\Sigma_0) \geq c > 0$ .

# Spatially/Serially Correlated Errors IV

## Theorem 3 (Asymptotic Distribution - CE)

Under Assumptions MD, CS, AD and ER, with  $m^2T/n \rightarrow c \geq 0$  as  $n, m \rightarrow \infty$ ,

$$\sqrt{nT}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(\Delta_*^{-1}(\psi_{*,1} + \psi_{*,2} + \psi_{*,3}), \Delta_*^{-1}\Omega_*\Delta_*^{-1}).$$

- $\psi_{*,1}$  generalises  $\psi$ :

$$\psi_{*,1} := \frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\Sigma_u \bar{B}_\varepsilon^{-\top} \bar{M} W \bar{B}^{-1} \bar{B}_\varepsilon^{-1}) \\ \text{tr}(\Sigma_u \bar{B}_\varepsilon^{-\top} \bar{M} \bar{\Pi} \bar{B}^{-1} \bar{B}_\varepsilon^{-1}) \\ \text{tr}(\Sigma_u \bar{B}_\varepsilon^{-\top} \bar{M} \bar{W} \bar{\Pi} \bar{B}^{-1} \bar{B}_\varepsilon^{-1}) \\ \mathbf{0}_{K \times 1} \end{pmatrix}.$$

- $\psi_{*,2}$  and  $\psi_{*,3}$  arise due to correlation with the initial condition  $\varepsilon_0$  and are  $\mathcal{O}(T^{-\frac{1}{2}})$ .



## Spatially/Serially Correlated Errors V

- Let  $\hat{\xi}(\theta_\varepsilon) := \text{vec}(\hat{\varepsilon}) - \rho_\varepsilon \bar{W} \text{vec}(\hat{\varepsilon}) - \alpha_\varepsilon \text{vec}(\hat{\varepsilon}_{-1}) - \phi_\varepsilon \bar{W} \text{vec}(\hat{\varepsilon}_{-1})$   
where  $\hat{\varepsilon} := (Y - Z \cdot \hat{\theta})M_{\hat{F}}$ , and

$$\varphi(\theta_\varepsilon) := \frac{1}{nT} \begin{pmatrix} \hat{\xi}(\theta_\varepsilon)^\top \Psi_1 \hat{\xi}(\theta_\varepsilon) \\ \vdots \\ \hat{\xi}(\theta_\varepsilon)^\top \Psi_L \hat{\xi}(\theta_\varepsilon) \end{pmatrix},$$

where  $\Psi_1, \dots, \Psi_L$  are a series of  $n \times n$  matrices with zero diagonals.

- Estimate of  $\theta_{0,\varepsilon}$  can be obtained as

$$\hat{\theta}_\varepsilon := \arg \min_{\hat{\theta}_\varepsilon \in \Theta_\varepsilon} \|\varphi(\theta)\|_2^2.$$

- Can be used to construct a test for  $\rho_\varepsilon = \alpha_\varepsilon = \phi_\varepsilon = 0$ .

- The bias described in Theorem 1 is caused by the implicit transformation of the model to purge the factors.

# Higher Order Bias

- The bias described in Theorem 1 is caused by the implicit transformation of the model to purge the factors.
- Similar bias do still arise from  $\mathbf{Y}_{-1}$  and  $\mathbf{W}\mathbf{Y}_{-1}$ , though of a lower order.

# Higher Order Bias

- The bias described in Theorem 1 is caused by the implicit transformation of the model to purge the factors.
- Similar bias do still arise from  $\mathbf{Y}_{-1}$  and  $\mathbf{W}\mathbf{Y}_{-1}$ , though of a lower order.
- These biases do not occur through not knowing the factors and the loadings.

# Higher Order Bias

- The bias described in Theorem 1 is caused by the implicit transformation of the model to purge the factors.
- Similar bias do still arise from  $\mathbf{Y}_{-1}$  and  $\mathbf{W}\mathbf{Y}_{-1}$ , though of a lower order.
- These biases do not occur through not knowing the factors and the loadings.
- The 'usual' LS-IFE biases are also present, though of a lower stochastic order.

# Higher Order Bias

- The bias described in Theorem 1 is caused by the implicit transformation of the model to purge the factors.
- Similar bias do still arise from  $\mathbf{Y}_{-1}$  and  $\mathbf{W}\mathbf{Y}_{-1}$ , though of a lower order.
- These biases do not occur through not knowing the factors and the loadings.
- The 'usual' LS-IFE biases are also present, though of a lower stochastic order.
- It is also possible to correct for these.

# Application I

- Study the relationship economic growth, civil liberties and political rights in the 21<sup>st</sup> century.
- Similar in spirit to Acemoglu et al. (2019). (ANRR for short)
- Data covers a panel of 180 countries observed between 2001 and 2020.
- Outcome  $y_{it}$  log of GDP per capita taken from the World Bank.
- Binary regressor  $d_{it}$  derived from Freedom House index

$$d_{it} = \begin{cases} 0 & \text{if classified as } \textit{not free}, \\ 1 & \text{if classified as } \textit{partially free} \text{ or } \textit{free}. \end{cases}$$

## Application II

- The World Bank provides high resolution latitude and longitude coordinates of international boundaries.
- These are rounded to generate a lower resolution projection which describes the shape of countries using a fewer data points.
- Great-circle distance is calculated between every pair of coordinates.
- For each country pair  $ij$ , let  $\delta_{ij}$  denote the shortest distance between two countries, and let  $e$  denote half the distance of the equator.
- The  $n \times n$  weights matrix  $\mathbf{W}$  is generated by setting element  $w_{ij}$  equal to

$$w_{ij} = \begin{cases} 1 - \delta_{ij}/e & \text{if } \delta_{ij}/e < \tau, \\ 0 & \text{otherwise,} \end{cases}$$

where  $\tau$  is a cut-off point set to 0.1.



## Application III

- Outcome equation:  $y_t = \alpha y_{t-1} + \beta d_t + \eta_t$ .
- Instruments:  $\mathcal{V} = (d_1, \dots, d_T)$ .
- Long term effect:  $\gamma := (1 - \alpha)^{-1}\beta$ .

	<b>FE</b>	<b>IV-IFE</b>	<b>ANRR</b>
$\beta$	<b>0.0161</b>	<b>0.0141</b>	<b>0.0078 - 0.0097</b>
t-stat	3.9000	2.5343	
$\alpha$	<b>0.9228</b>	<b>0.8435</b>	<b>0.938 - 0.973</b>
t-stat	146.3841	24.7520	
$\gamma$	<b>0.2086</b>	<b>0.0901</b>	<b>0.1264 - 0.3558</b>
t-stat	3.7379	2.7803	
$J$ -stat	-	<b>0.2096</b>	-

## Application III

- Outcome equation:  $y_t = \alpha y_{t-1} + \beta d_t + \eta_t$ .
- Instruments:  $\mathcal{V} = (d_1, \dots, d_T, Wd_1, \dots, Wd_T)$ .
- Long term effect:  $\gamma := (1 - \alpha)^{-1}\beta$ .

	<b>FE</b>	<b>IV-IFE</b>	<b>ANRR</b>
$\beta$	<b>0.0161</b>	<b>0.0124</b>	<b>0.0078 - 0.0097</b>
t-stat	3.9000	76.8928	
$\alpha$	<b>0.9228</b>	<b>0.9335</b>	<b>0.938 - 0.973</b>
t-stat	146.3841	24.7520	
$\gamma$	<b>0.2086</b>	<b>0.1865</b>	<b>0.1264 - 0.3558</b>
t-stat	3.7379	2.4492	
$J$ -stat	-	<b>0.4632</b>	-

- Outcome equation:

$$\mathbf{y}_t = \rho \mathbf{W} \mathbf{y}_t + \alpha \mathbf{y}_{t-1} + \phi \mathbf{W} \mathbf{y}_{t-1} + \beta \mathbf{d}_t + \boldsymbol{\eta}_t.$$

- Instruments:  $\mathcal{V} = (\mathbf{d}_1, \dots, \mathbf{d}_T, \mathbf{W} \mathbf{d}_1, \dots, \mathbf{W} \mathbf{d}_T).$
- Long term effects:
  - Direct effect:  $\gamma_D := \text{tr}(((1 - \alpha)\mathbf{I}_n - (\rho + \phi)\mathbf{W})^{-1}\beta)/n.$
  - Indirect effect:  $\gamma_I := \boldsymbol{\iota}_n^\top ((1 - \alpha)\mathbf{I}_n - (\rho + \phi)\mathbf{W})^{-1}\beta \boldsymbol{\iota}_n/n - \gamma_D.$

## Application IV

	FE	IV-IFE	IV-IFE-BC
$\beta$	<b>0.0158</b>	<b>0.0111</b>	<b>0.0112</b>
t-stat	3.8585	2.0618	2.0781
$\alpha$	<b>0.9224</b>	<b>0.9347</b>	<b>0.9374</b>
t-stat	145.5072	78.0726	78.2993
$\rho$	<b>0.0125</b>	<b>0.0164</b>	<b>0.0082</b>
t-stat	6.9085	8.0257	4.0043
$\phi$	<b>-0.0118</b>	<b>-0.0156</b>	<b>-0.0084</b>
t-stat	-7.3287	-8.0175	-4.3520
$\gamma_D$	<b>0.2042</b>	<b>0.1716</b>	<b>0.1793</b>
t-stat	3.7029	2.1954	2.2006
$\gamma_I$	<b>0.0645</b>	<b>0.1142</b>	<b>-0.0184</b>
t-stat	0.9342	0.8839	-0.5809
$J$ -stat	-	<b>0.1592</b>	<b>0.0938</b>

- Preliminary findings suggest results in ANRR are robust.

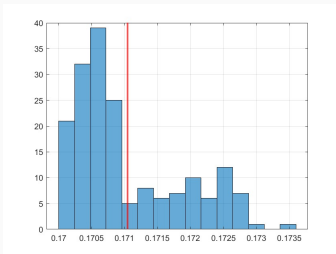
- Preliminary findings suggest results in ANRR are robust.
- Spatial dependence a feature, but oscillates. Why?
  - Misspecification.
  - Unit roots.
  - Genuine feature of the data.

- Preliminary findings suggest results in ANRR are robust.
- Spatial dependence a feature, but oscillates. Why?
  - Misspecification.
  - Unit roots.
  - Genuine feature of the data.
- Explore more complex specification and utilise the additional results.

- Preliminary findings suggest results in ANRR are robust.
- Spatial dependence a feature, but oscillates. Why?
  - Misspecification.
  - Unit roots.
  - Genuine feature of the data.
- Explore more complex specification and utilise the additional results.
- Any comments/suggestions most welcome.



# Application VI

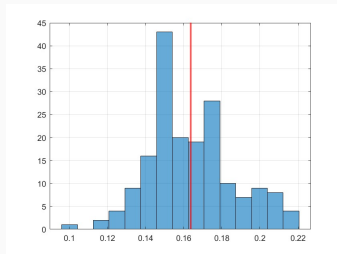


$$\beta = 0.0111$$

$$\alpha = 0.9347$$

$$\rho = 0.0164$$

$$\phi = -0.0156$$



$$\beta = 0.0111$$

$$\alpha = 0.9347$$

$$\rho = 0.0164$$

$$\phi = -0.0100$$

## Closing Remarks

- Introduce a simple IV estimator which is consistent and asymptotically normally distributed as long as the number of cross-sectional units  $n$  grows sufficiently fast relative to the number of instruments  $m$  and the number of time periods  $T$ .
- Circumstances exist where, depending on the weights matrix, the estimator can exhibit considerable bias.
- Constructing a bias corrected estimator significantly ameliorates this issue.
- Application applies the method to study the relationship between economic growth, political rights and civil liberties.
- Multiple extensions.