Instrumental Variables for Dynamic Spatial Models with Interactive Effects

Ayden Higgins

October 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.
  - $\boldsymbol{y}_t$ growth rate of real GDP per capita.
  - $\boldsymbol{X}_t$ other factors, e.g. life expectancy, population growth, consumption, investment and government expenditure shares.
  - $\boldsymbol{W}$ 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.
  - $\boldsymbol{y}_t$ growth rate of real GDP per capita.
  - $\boldsymbol{X}_t$ other factors, e.g. life expectancy, population growth, consumption, investment and government expenditure shares.
  - $\boldsymbol{W}$ binary indicators of whether countries share a border.
- Model

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

Dynamic Spatial Model

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

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

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

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 $\boldsymbol{\eta}_t$ may be decomposed as

$$\boldsymbol{\eta}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t.$$

Interactive Effects

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

$$\boldsymbol{\eta}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t.$$

- Note $\boldsymbol{\Lambda}$, $\boldsymbol{f}_t$ and $\boldsymbol{\varepsilon}_t$ 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 $\boldsymbol{\eta}_t$ may be decomposed as

$$\boldsymbol{\eta}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t.$$

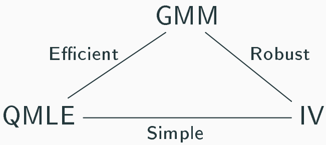
- Note $\boldsymbol{\Lambda}$, $\boldsymbol{f}_t$ and $\boldsymbol{\varepsilon}_t$ are all unobserved.
- Generalisation of classic models such as individual, time or group effects:

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

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

Related Work

- 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)



Overview

- 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 $n \to \infty$, 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 $n \to \infty$, 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 $n \to \infty$, 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.

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 $n \to \infty$, 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.

Outline I

- Original model in matrices:

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

with $\boldsymbol{S}(\rho) \coloneqq \boldsymbol{I}_n - \rho\boldsymbol{W}$.

Outline I

- Original model in matrices:

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

  with $\boldsymbol{S}(\rho) \coloneqq \boldsymbol{I}_n - \rho\boldsymbol{W}$.
- Under some conditions we can expand

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

Outline I

- Original model in matrices:

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

  with $\boldsymbol{S}(\rho) \coloneqq \boldsymbol{I}_n - \rho\boldsymbol{W}$.
- Under some conditions we can expand

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

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

Outline I

- With S weights matrices

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

- Produces a great many instruments

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

Outline II

- Re-write

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

Outline II

- Re-write

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

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

Outline II

- Re-write

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

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

- Re-write

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

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

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

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

Outline III

- Concentrated objective function:

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

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

Outline III

- Concentrated objective function:

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

  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 $\boldsymbol{\theta}_0$ is in the interior of Θ, where Θ is a compact subset of $\mathbb{R}^P$.

(ii) The weights matrix $\boldsymbol{W}$ has a zero diagonal, is nonstochastic and UB.

(iii) For all $\rho \in \Theta_\rho$, $\alpha \in \Theta_\alpha$ and $\phi \in \Theta_\phi$, $|\det(\boldsymbol{S}(\rho))| \geq c > 0$, $|\det(\bar{\boldsymbol{B}}(\rho, \alpha, \phi))| \geq c > 0$, and $\|\boldsymbol{A}(\rho, \alpha, \phi)\|_2 < 1 - c$ holds for all (n, T), and $\boldsymbol{S}^{-1}(\rho)$, $\bar{\boldsymbol{B}}^{-1}(\rho, \alpha, \phi)$ and $\sum_{h=1}^{\infty} |\boldsymbol{A}^h(\rho, \alpha, \phi)|$ are UB.

(iv) $x_{kit}, \lambda_{0,ir}$ and $f_{0,tr}$ have uniformly bounded fourth moments.

(v) The errors ε_{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

Assumption CS (Consistency)

(i) $R \geq R_0 := \text{rank}(\tilde{\boldsymbol{\Lambda}}_0 \boldsymbol{F}_0^\top)$.

(ii) $\min_{\boldsymbol{\delta} \in \mathbb{R}^P : \|\boldsymbol{\delta}\|_2 = 1} \sum_{r=R+R_0+1}^{T} \mu_r(\frac{1}{nT}(\boldsymbol{\delta} \cdot \tilde{\boldsymbol{Z}})^\top (\boldsymbol{\delta} \cdot \tilde{\boldsymbol{Z}})) \geq c > 0$, w.p.a.1 as $n \to \infty$.

Consistency

Proposition 1 (Consistency)

Under Assumptions MD and CS, as $n \to \infty$,

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

Assumption AD (Asymptotic Distribution)

(i) $R = R_0$.

(ii) $\frac{1}{n}\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \xrightarrow{p} \mathbf{\Sigma}_{\tilde{\mathbf{\Lambda}}_0}$ as $n \to \infty$, with $\mu_{R_0}(\mathbf{\Sigma}_{\tilde{\mathbf{\Lambda}}_0}) > 0$ and $\mu_1(\mathbf{\Sigma}_{\tilde{\mathbf{\Lambda}}_0}) < \infty$.

(iii) $\mu_{R_0}(\frac{1}{T}\boldsymbol{F}_0^\top \boldsymbol{F}_0) > 0$ and $\mu_1(\frac{1}{T}\boldsymbol{F}_0^\top \boldsymbol{F}_0) < \infty$.

(iv) The elements of the matrices $\bar{\boldsymbol{M}}\tilde{\boldsymbol{\mathcal{H}}}$ and $\bar{\boldsymbol{M}}\tilde{\boldsymbol{\mathcal{R}}}$ have uniformly bounded fourth moments for all (n, T).

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

Theorem 1 (Asymptotic Distribution)

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

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

where,

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

$\boldsymbol{\psi} := \operatorname{plim}_{n\to\infty} \boldsymbol{\psi}_n$, $\boldsymbol{\Delta}_n := (nT)^{-1}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})^\top \bar{\boldsymbol{M}}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})$, $\boldsymbol{\Delta} = \operatorname{plim}_{n\to\infty} \boldsymbol{\Delta}_n$, $\boldsymbol{\Omega}_n := (nT)^{-1}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})^\top \bar{\boldsymbol{M}}\boldsymbol{\Sigma}\bar{\boldsymbol{M}}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})$, $\boldsymbol{\Omega} := \operatorname{plim}_{n\to\infty} \boldsymbol{\Omega}_n$, *and* $\boldsymbol{\Sigma}$ *is an* $nT \times nT$ *matrix with diagonal elements* $\sigma_{11}^2, \ldots, \sigma_{nT}^2$ *and remaining elements equal to zero.*

Asymptotic Distribution II

- Consider more closely

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

Asymptotic Distribution II

- Consider more closely

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

- In general

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

Asymptotic Distribution II

- Consider more closely

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

- 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

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

- 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:

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

with

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

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

Asymptotic Distribution III

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

$$\begin{aligned}\psi_{n,1} &= \frac{1}{\sqrt{nT}} \mathrm{tr}(\boldsymbol{\Sigma}\bar{\boldsymbol{W}}(\boldsymbol{I}_{nT} - \rho_0 \bar{\boldsymbol{W}})^{-1}\bar{\boldsymbol{P}}) + \mathcal{O}(1) \\ &=: \xi + \mathcal{O}(1),\end{aligned}$$

and establish bounds

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

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

Asymptotic Distribution III

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

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

Asymptotic Distribution III

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

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

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

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

Asymptotic Distribution III

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

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

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

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

- A spectrum of cases lie in between.

Asymptotic Distribution III

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

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

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

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

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

Asymptotic Distribution IV

- Construct a bias correct estimator
$$\tilde{\boldsymbol{\theta}} \coloneqq \hat{\boldsymbol{\theta}} - \frac{1}{\sqrt{nT}} \hat{\boldsymbol{\Delta}}^{-1} \hat{\boldsymbol{\psi}}.$$
- Under assumptions MD, CS and AD, as $n, m \to \infty$ with $m^2 T/n \to c \geq 0$
$$\sqrt{nT}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{P\times 1}, \boldsymbol{\Delta}^{-1}\boldsymbol{\Omega}\boldsymbol{\Delta}^{-1}).$$

Illustration I

Design:

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

with

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

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

before being row-normalised.
- Easy to verify that rank$(\boldsymbol{W}) = 2G$ 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 - $G = 5$

	$\hat{\rho}$			$\tilde{\rho}$		
$n \setminus T$	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 - $G = 25$

	$\hat{\rho}$			$\tilde{\rho}$		
$n \setminus T$	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

Specification Testing I

- IV-IFE estimator is based the moment condition

$$\mathbb{E}\left[(\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{Q}_{\boldsymbol{\mathcal{V}}} \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0} \boldsymbol{Q}_{\boldsymbol{\mathcal{V}}}^\top)\mathsf{vec}(\boldsymbol{\eta})\right] = \boldsymbol{0}_{nT}.$$

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

$$\mathcal{J} := \mathsf{vec}(\hat{\boldsymbol{\eta}})^\top (\boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{Q}_{\boldsymbol{\mathcal{V}}} \boldsymbol{M}_{\hat{\tilde{\boldsymbol{\Lambda}}}} \boldsymbol{Q}_{\boldsymbol{\mathcal{V}}}^\top)\mathsf{vec}(\hat{\boldsymbol{\eta}}),$$

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

- Define

$$\begin{aligned}
\boldsymbol{M}_{\mathcal{J}} &:= \bar{\boldsymbol{M}} - \boldsymbol{P}_{\mathcal{J}}, \\
\boldsymbol{P}_{\mathcal{J}} &:= \bar{\boldsymbol{M}}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})((\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})^\top \bar{\boldsymbol{M}}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}}))^{-1}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})^\top \bar{\boldsymbol{M}}, \\
\ell &:= (T - R_0)(m - R_0) - P, \\
\sigma^2_{\mathcal{J}} &:= \operatorname{tr}((\boldsymbol{\mathcal{M}}^{(4)} - 3\boldsymbol{\Sigma}^2)(\boldsymbol{M}_{\mathcal{J}} \odot \boldsymbol{M}_{\mathcal{J}})) \\
&\quad + \; 2\boldsymbol{\iota}_{nT}^\top(\boldsymbol{\Sigma}(\boldsymbol{M}_{\mathcal{J}} \odot \boldsymbol{M}_{\mathcal{J}})\boldsymbol{\Sigma})\boldsymbol{\iota}_{nT},
\end{aligned}$$

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

Assumption JS (J-Test)

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

Specification Testing III

Theorem 2 (J-Test)

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

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

where

$$\varphi_{\mathcal{J}} := \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{M}_{\mathcal{J}}) - \boldsymbol{\psi}_n^\top \boldsymbol{\Delta}_n^{-1}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})^\top \bar{\boldsymbol{M}}(\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{R}})\boldsymbol{\Delta}_n^{-1}\boldsymbol{\psi}_n.$$

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

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 $\sigma^2_{\mathcal{J}} = 2\ell$.
- Then if $R < R_0$ it can be shown that under the assumptions of Proposition 2 (except AD(i))

$$\xi_R \coloneqq \frac{\mathcal{J}_R - \ell_R}{\sqrt{2\ell_R}} \to \infty.$$

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 $\sigma^2_{\mathcal{J}} = 2\ell$.
- Then if $R < R_0$ it can be shown that under the assumptions of Proposition 2 (except AD(i))

$$\xi_R \coloneqq \frac{\mathcal{J}_R - \ell_R}{\sqrt{2\ell_R}} \to \infty.$$

- Then,

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

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

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

$$\boldsymbol{\varepsilon}_t = \rho_\varepsilon \boldsymbol{W} \boldsymbol{\varepsilon}_t + \alpha_\varepsilon \boldsymbol{\varepsilon}_{t-1} + \phi_\varepsilon \boldsymbol{W} \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{u}_t.$$

Spatially/Serially Correlated Errors II

Assumption ER (Error)

1. The vector $\boldsymbol{\theta}_{\varepsilon,0} := (\rho_\varepsilon, \alpha_\varepsilon, \phi_\varepsilon)^\top$ lies in the interior of Θ_ε, where Θ_ε is a compact subset of $\mathbb{R}^3$ in which $\inf_{\theta_\varepsilon \in \Theta_\varepsilon} \det(\boldsymbol{S}_\varepsilon(\rho_\varepsilon)) \neq 0$ and $\inf_{\boldsymbol{\theta}_\varepsilon \in \Theta_\varepsilon} \det(\boldsymbol{B}_\varepsilon(\boldsymbol{\theta}_\varepsilon)) \neq 0$, $\boldsymbol{S}_\varepsilon^{-1}(\rho_\varepsilon)$, and $\bar{\boldsymbol{B}}_\varepsilon^{-1}(\boldsymbol{\theta}_\varepsilon)$ are UB, and $\|\boldsymbol{A}_\varepsilon^h(\boldsymbol{\theta}_\varepsilon)\|_2 < 1 - c$ 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

$$\boldsymbol{y}_0 = \boldsymbol{\Sigma}_{y_0}^{\frac{1}{2}} \boldsymbol{\nu}_1$$

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

where $\{\nu_{1,j}, \nu_{2,j}\}$ are independent of the independent of the factors, the loadings, the covariates, and the errors $\boldsymbol{u}$, 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,

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

where $\boldsymbol{\Sigma}_0$ is UB with $\mu_{\min}(\boldsymbol{\Sigma}_0) \geq c > 0$.

Spatially/Serially Correlated Errors IV

Theorem 3 (Asymptotic Distribution - CE)

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

$$\sqrt{nT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\Delta}_*^{-1}(\boldsymbol{\psi}_{*,1} + \boldsymbol{\psi}_{*,2} + \boldsymbol{\psi}_{*,3}), \boldsymbol{\Delta}_*^{-1}\boldsymbol{\Omega}_*\boldsymbol{\Delta}_*^{-1}\right).$$

- $\boldsymbol{\psi}_{*,1}$ generalises $\boldsymbol{\psi}$:

$$\boldsymbol{\psi}_{*,1} := \frac{1}{\sqrt{nT}} \begin{pmatrix} \operatorname{tr}\left(\boldsymbol{\Sigma}_u \bar{\boldsymbol{B}}_\varepsilon^{-\top} \bar{\boldsymbol{M}} \boldsymbol{W} \bar{\boldsymbol{B}}^{-1} \bar{\boldsymbol{B}}_\varepsilon^{-1}\right) \\ \operatorname{tr}\left(\boldsymbol{\Sigma}_u \bar{\boldsymbol{B}}_\varepsilon^{-\top} \bar{\boldsymbol{M}} \bar{\boldsymbol{\Pi}} \bar{\boldsymbol{B}}^{-1} \bar{\boldsymbol{B}}_\varepsilon^{-1}\right) \\ \operatorname{tr}\left(\boldsymbol{\Sigma}_u \bar{\boldsymbol{B}}_\varepsilon^{-\top} \bar{\boldsymbol{M}} \bar{\boldsymbol{W}} \bar{\boldsymbol{\Pi}} \bar{\boldsymbol{B}}^{-1} \bar{\boldsymbol{B}}_\varepsilon^{-1}\right) \\ \mathbf{0}_{K\times 1} \end{pmatrix}.$$

- $\boldsymbol{\psi}_{*,2}$ and $\boldsymbol{\psi}_{*,3}$ arise due to correlation with the initial condition $\boldsymbol{\varepsilon}_0$ and are $\mathcal{O}(T^{-\frac{1}{2}})$.

Spatially/Serially Correlated Errors V

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

$$\varphi(\boldsymbol{\theta}_\varepsilon) := \frac{1}{nT} \begin{pmatrix} \hat{\boldsymbol{\xi}}(\boldsymbol{\theta}_\varepsilon)^\top \boldsymbol{\Psi}_1 \hat{\boldsymbol{\xi}}(\boldsymbol{\vartheta}_\varepsilon) \\ \vdots \\ \hat{\boldsymbol{\xi}}(\boldsymbol{\theta}_\varepsilon)^\top \boldsymbol{\Psi}_L \hat{\boldsymbol{\xi}}(\boldsymbol{\theta}_\varepsilon) \end{pmatrix},$$

  where $\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_L$ are a series of $n \times n$ matrices with zero diagonals.
- Estimate of $\boldsymbol{\theta}_{0,\varepsilon}$ can be obtained as

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

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

Higher Order Bias

- 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 $\boldsymbol{Y}_{-1}$ and $\boldsymbol{W}\boldsymbol{Y}_{-1}$, 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 $\boldsymbol{Y}_{-1}$ and $\boldsymbol{W}\boldsymbol{Y}_{-1}$, 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 $\boldsymbol{Y}_{-1}$ and $\boldsymbol{W}\boldsymbol{Y}_{-1}$, 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 $\boldsymbol{Y}_{-1}$ and $\boldsymbol{W}\boldsymbol{Y}_{-1}$, 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 21st 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 δ_{ij} denote the shortest distance between two countries, and let e denote half the distance of the equator.
- The $n \times n$ weights matrix $\boldsymbol{W}$ 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 τ is a cut-off point set to 0.1.

Application III

- Outcome equation: $\boldsymbol{y}_t = \alpha \boldsymbol{y}_{t-1} + \beta \boldsymbol{d}_t + \boldsymbol{\eta}_t$.
- Instruments: $\boldsymbol{\mathcal{V}} = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_T)$.
- Long term effect: $\gamma := (1-\alpha)^{-1}\beta$.

-	FE	IV-IFE	ANRR
β	**0.0161**	**0.0141**	**0.0078 - 0.0097**
t-stat	3.9000	2.5343	
α	**0.9228**	**0.8435**	**0.938 - 0.973**
t-stat	146.3841	24.7520	
γ	**0.2086**	**0.0901**	**0.1264 - 0.3558**
t-stat	3.7379	2.7803	
J-stat	-	**0.2096**	-

Application III

- Outcome equation: $\boldsymbol{y}_t = \alpha \boldsymbol{y}_{t-1} + \beta \boldsymbol{d}_t + \boldsymbol{\eta}_t$.
- Instruments: $\boldsymbol{\mathcal{V}} = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_T, \boldsymbol{W}\boldsymbol{d}_1, \ldots, \boldsymbol{W}\boldsymbol{d}_T)$.
- Long term effect: $\gamma := (1-\alpha)^{-1}\beta$.

-	FE	IV-IFE	ANRR
β	**0.0161**	**0.0124**	**0.0078 - 0.0097**
t-stat	3.9000	76.8928	
α	**0.9228**	**0.9335**	**0.938 - 0.973**
t-stat	146.3841	24.7520	
γ	**0.2086**	**0.1865**	**0.1264 - 0.3558**
t-stat	3.7379	2.4492	
J-stat	-	**0.4632**	-

Application IV

- Outcome equation:

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

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

Application IV

-	FE	IV-IFE	IV-IFE-BC
β	**0.0158**	**0.0111**	**0.0112**
t-stat	3.8585	2.0618	2.0781
α	**0.9224**	**0.9347**	**0.9374**
t-stat	145.5072	78.0726	78.2993
ρ	**0.0125**	**0.0164**	**0.0082**
t-stat	6.9085	8.0257	4.0043
ϕ	**-0.0118**	**-0.0156**	**-0.0084**
t-stat	-7.3287	-8.0175	-4.3520
γ_D	**0.2042**	**0.1716**	**0.1793**
t-stat	3.7029	2.1954	2.2006
γ_I	**0.0645**	**0.1142**	**-0.0184**
t-stat	0.9342	0.8839	-0.5809
J-stat	-	**0.1592**	**0.0938**

Application V

- Preliminary findings suggest results in ANRR are robust.

Application V

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

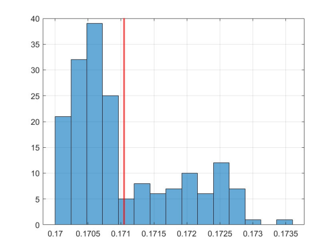
Application V

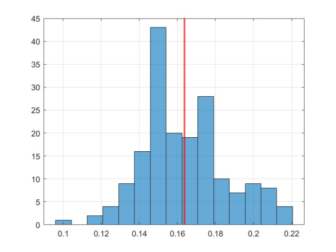
- Preliminary findings suggest results in ANRR are robust.
- Spatial dependence a feature, but oscillates. Why?
    - Misspecification.
    - Unit roots.
    - Genuine feature of the data.
- Explore more complex specification and utilise the additional results.

Application V

- 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





$\beta = 0.0111$	$\beta = 0.0111$
$\alpha = 0.9347$	$\alpha = 0.9347$
$\rho = 0.0164$	$\rho = 0.0164$
$\phi =$ -0.0156	$\phi =$ -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.