Spatial Dynamic Panel Data Models with Interactive Effects

Ongoing research agenda with several co-authors: Guowei Cui (HUST), Jia Chen (Macau), Takashi Yamagata (York), Sebastian Kripfganz (Exeter)

Vasilis Sarafidis

Brunel University of London

vasilis.sarafidis@brunel.ac.uk

https://sites.google.com/view/vsarafidis

www.youtube.com/@FrontierEconometrics

2025 Stata Economics Virtual Symposium

Model

$$y_{i,t} = \alpha y_{i,t-1} + \psi \sum_{j=1}^{N} w_{i,j} y_{j,t} + \sum_{\ell=1}^{K} \beta_\ell x_{\ell,i,t} + \underbrace{\boldsymbol{\lambda}_i' \mathbf{f}_t + \varepsilon_{i,t}}_{u_{i,t}}, \tag{1}$$

$$i = 1, \ldots, N; \quad t = 1, \ldots, T,$$

"State dependence": Balestra-Nerlove 1966, Anderson-Hsiao 1982, Arellano-Bond 1991, Blundell-Bond 1998;

Spatial lag: "endogenous interactions" due to networks, peer effects, spillovers etc, e.g. Case 1991, Manski 1993, 1995, Fisher 1935, Tobler 1970, Elhorst 2014;

"Causal effects": K idiosyncratic, time-varying *own* characteristics;

Nonlinear error components: "Latent Common Factors"/"Interactive Effects"; e.g., technological disruptions, energy price shocks, geopolitical conflicts, regulatory changes, financial crises, macro policies etc: e.g. Pesaran 2006, Bai 2009, Sarafidis-Wansbeek 2021.

Comparison with Two-Way (or additive) Effects

The Two-Way Effects model assumes error components enter additively: $u_{i,t} = \eta_i + \tau_t + \varepsilon_{i,t}$. But this is a special case of the interactive effects structure.

$r = 2$, $\boldsymbol{\lambda}_i = (\eta_i, 1)'$ & $\mathbf{f}_t = (1, \tau_t)' \Rightarrow \boldsymbol{\lambda}_i'\mathbf{f}_t = \eta_i + \tau_t$.

Neither first-differencing nor the within transformation will remove the common factor component.

The use of multiplicative time dummies at a group level (e.g. using the absorb function in Stata) is not warranted.

To illustrate, let $r = 1$ and $u_{i,t} = \lambda_i f_t + \varepsilon_{i,t}$.

We can rewrite this as $u_{i,t} = u_{i,t} \pm \lambda_{g_i} f_t = \lambda_{g_i} f_t + \varepsilon_{i,t} + \left(\lambda_i - \lambda_{g_i}\right) f_t$

$\therefore$ Covariates remain correlated w.r.t. the nonlinear error component.

Technical Issues

$$y_{i,t} = \alpha_i y_{i,t-1} + \psi_i \sum_{j=1}^{N} w_{i,j} y_{j,t} + \mathbf{x}'_{i,t} \boldsymbol{\beta}_i + \boldsymbol{\lambda}'_i \mathbf{f}_t + \varepsilon_{i,t}.$$

- 3 potential sources of endogeneity
  - Lagged dependent variable is endogenous by construction;
  - Spatial lag variable is endogenous by construction;
  - Covariates may also be correlated with the composite error;
- "How many factors?"
- "Incidental parameters problem"
- "Asymptotic analysis with $N, T \to \infty$ jointly"
- "Heterogeneous Slope Parameters"

Econometric literature: large N, T asymptotics

$$y_{i,t} = \alpha_i y_{i,t-1} + \psi_i \sum_{j=1}^{N} w_{i,j} y_{j,t} + \mathbf{x}'_{i,t} \boldsymbol{\beta}_i + \boldsymbol{\lambda}'_i \mathbf{f}_t + \varepsilon_{i,t}$$

Dynamic panels with additive fixed effects: $(\psi = 0,\ \boldsymbol{\lambda}'_i \mathbf{f}_t = \eta_i + \tau_t)$ e.g. Hahn-Kuersteiner 2002, Alvarez-Arellano 2003, Hayakawa 2015;

Spatial panels with additive fixed effects: $(\alpha = 0,\ \boldsymbol{\lambda}'_i \mathbf{f}_t = \eta_i + \tau_t)$ e.g. Yu-De Jong-Lee 2008, Korniotis 2010, Lee-Yu 2014;

Dynamic panels with "interactive fixed effects": $(\psi = 0)$ e.g. Chudik-Pesaran 2015, Moon-Weidner 2017, Juodis-Sarafidis 2021;

Present agenda sits on the intersection of the above strands of literature. Most approaches available are based on QMLE (Shi-Lee 2017, Bai-Li 2021);

No literature on spatial, dynamic panels with **heterogeneous coefficients**.

Contribution (Homogeneous case)

- Cui-Sarafidis-Yamagata 2023 put forward a two-stage spatial IV (2SIV) estimator; consistent and asy. normal under N, $T \to \infty$, s.t. $N/T \to c$, $0 < c < \infty$;
- 2SIV is **linear and easy to compute**;

  In contrast, for QMLE, estimation of the Jacobian matrix of likelihood function is computational intensive. MC: for $N = T = 200$, QMLE is 60 times slower.
- 2SIV is **asymptotically unbiased**;

  QMLE suffers from asy. bias due to incidental parameters $\Rightarrow$ severe size distortions. Bias-correction is based on true $\#$ of factors;
- 2SIV can accommodate **endogenous regressors** (provided there are exogenous instruments available.)

Contribution cont. (Heterogeneous case)

- Chen-Cui-Sarafidis-Yamagata 2025 put forward a Mean Group spatial IV estimator; consistent and asy. normal under N, $T \rightarrow \infty$, s.t. $N/T^2 \rightarrow 0$;
- MGIV is given by

$$\widehat{\boldsymbol{\theta}}_{MGIV} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\theta}}_i,$$

  where $\widehat{\boldsymbol{\theta}}_i = \left(\widehat{\alpha}_i, \widehat{\psi}_i, \widehat{\boldsymbol{\beta}}_i'\right)'$.

- There exist no alternative methods in this case.
- Both 2SIV and MGIV are now available in Stata with the *spxtivdfreg* command
- *net install xtivdfreg, from(https://www.kripfganz.de/stata) replace*

Generalisations

$$y_{i,t} = \alpha y_{i,t-1} + \psi \sum_{j=1}^{N} w_{i,j} y_{j,t} + \mathbf{x}_{i,t}' \boldsymbol{\beta} + \boldsymbol{\lambda}_i' \mathbf{f}_t + \varepsilon_{i,t}$$

$$+ \psi_1 \sum_{j=1}^{N} w_{i,j} y_{j,t-1} + \gamma' \sum_{j=1}^{N} w_{i,j} \mathbf{x}_{j,t} + \sum_{s=2}^{p} \alpha_s y_{i,t-s}.$$

"Spatial-time lag" (sometimes "contagion")

"Exogenous network effects", or "contextual effects" (the "Spatial Durbin model")

"AR process of order > 1"

Could also think of "spatially-correlated unobservables", through ε_{it}. But not jointly identified (Manski, 1993).

Model in stacked form

Stacking the T observations for each i yields

$$\mathbf{y}_i = \alpha \mathbf{y}_{i,-1} + \psi \sum_{j=1}^{N} w_{i,j} \mathbf{y}_j + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{F} \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i; \tag{2}$$

or

$$\mathbf{y}_i = \mathbf{C}_i \boldsymbol{\theta} + \mathbf{u}_i, \tag{3}$$

where $\mathbf{C}_i = (\mathbf{y}_{i,-1}, \sum_{j=1}^{N} w_{i,j} \mathbf{y}_j, \mathbf{X}_i)$, $\boldsymbol{\theta} = (\alpha, \psi, \boldsymbol{\beta}')'$ and $\mathbf{u}_i = \mathbf{F} \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i$.

We assume

$$\mathbf{X}_i = \mathbf{F}_x \boldsymbol{\Gamma}_i + \mathbf{V}_i. \tag{4}$$

$\mathbf{F}$ & $\mathbf{F}_x$ can overlap, be identical or completely different. Similarly, $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Gamma}_i$ can be mutually correlated.

Outline of two-stage IV procedure (Stage 1)

Remove $\mathbf{F}_x$ from $\mathbf{X}$ (i.e. we "defactor" $\mathbf{X}$) using PCA (e.g. Bai 2003);

Let $\mathbf{M}_{\widehat{\mathbf{F}}_{x,-\tau}} = \mathbf{I}_T - \widehat{\mathbf{F}}_{x,-\tau}\left(\widehat{\mathbf{F}}'_{x,-\tau}\widehat{\mathbf{F}}_{x,-\tau}\right)^{-1}\widehat{\mathbf{F}}'_{x,-\tau}$, where

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_{i,-\tau}\mathbf{X}'_{i,-\tau}\right)\widehat{\mathbf{F}}_{x,-\tau} = \widehat{\mathbf{F}}_{x,-\tau}\widehat{\boldsymbol{\Omega}}_{\tau}. \tag{5}$$

We formulate $\widehat{\mathbf{Z}}_i$ based on defactored covariates:

$$\underbrace{\widehat{\mathbf{Z}}_i}_{T\times 3K} = \left(\mathbf{M}_{\widehat{\mathbf{F}}_x}\mathbf{X}_i, \quad \mathbf{M}_{\widehat{\mathbf{F}}_{x,-1}}\mathbf{X}_{i,-1}, \quad \sum_{j=1}^{N} w_{ij}\mathbf{M}_{\widehat{\mathbf{F}}_x}\mathbf{X}_j\right), \tag{6}$$

and estimate $\boldsymbol{\theta}$ using

$$\widehat{\boldsymbol{\theta}}_{1SIV} = \left(\widehat{\mathbf{A}}'\widehat{\mathbf{B}}^{-1}\widehat{\mathbf{A}}\right)^{-1}\widehat{\mathbf{A}}'\widehat{\mathbf{B}}^{-1}\widehat{\mathbf{c}}, \tag{7}$$

$$\widehat{\mathbf{A}} = \frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}'_i\mathbf{C}_i; \quad \widehat{\mathbf{B}} = \frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}'_i\widehat{\mathbf{Z}}_i; \quad \widehat{\mathbf{c}} = \frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}'_i\mathbf{y}_i\,.$$

Outline of two-stage IV procedure (Stage 2)

Project out $\mathbf{F}$ from $\mathbf{u}_i$ ($\mathbf{y}_i$) using the 1SIV residuals, $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{C}_i \widehat{\boldsymbol{\theta}}_{1SIV}$;

Next, run IV regression again using the same instruments:

$$\widehat{\boldsymbol{\theta}}_{2SIV} = (\widetilde{\mathbf{A}}' \widetilde{\mathbf{B}}^{-1} \widetilde{\mathbf{A}})^{-1} \widetilde{\mathbf{A}}' \widetilde{\mathbf{B}}^{-1} \widetilde{\mathbf{c}} \tag{8}$$

$$\widetilde{\mathbf{A}} = \frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{C}_i \,, \widetilde{\mathbf{B}} = \frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \widehat{\mathbf{Z}}_i \,, \widetilde{\mathbf{c}} = \frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{y}_i \,.$$

$\widehat{\boldsymbol{\theta}}_{2SIV}$ is free from asy. bias. QMLE is subject to two sources of asy. bias: (i) "Nickell bias" due to the presence of $\mathbf{y}_{i,-1}$; (ii) bias due to the defactoring of $\mathbf{X}_i$; estimation error in $\widehat{\mathbf{F}}$ depends on $\mathbf{U}_i$, and this feeds in $\mathbf{X}_i$ due to defactoring. That is, defactored regressors are asy. correlated with the error term.

Assumption (idiosyncratic error in $\mathbf{y}$)

$\varepsilon_{i,t}$ is independently distributed across i and over t, with mean zero, $\mathbb{E}(\varepsilon_{i,t}^2) = \sigma_\varepsilon^2 > 0$ and $\mathbb{E}|\varepsilon_{i,t}|^{8+\delta} \leq C < \infty$ for some $\delta > 0$.

Assumption (idiosyncratic error in $\mathbf{x}$)

1. *$\mathbf{v}_{i,t}$ is group-wise independent from $\varepsilon_{i,t}$, $\mathbb{E}(\mathbf{v}_{i,t}) = 0$ and $\mathbb{E}\|\mathbf{v}_{i,t}\|^{8+\delta} \leq C < \infty$;*
2. *Let $\boldsymbol{\Sigma}_{i,j,s,t} \equiv \mathbb{E}\left(\mathbf{v}_{i,s}\mathbf{v}_{j,t}'\right)$. There exist $\bar{\sigma}_{i,j}$ and $\tilde{\sigma}_{s,t}$, $\|\boldsymbol{\Sigma}_{i,j,s,t}\| \leq \bar{\sigma}_{i,j}$ for all (s,t), and $\|\boldsymbol{\Sigma}_{i,j,s,t}\| \leq \tilde{\sigma}_{s,t}$ for all (i,j), such that*

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\bar{\sigma}_{i,j} \leq C < \infty\,, \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{\sigma}_{s,t} \leq C < \infty\,,$$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\boldsymbol{\Sigma}_{i,j,s,t}\| \leq C < \infty\,.$$

Assumption (factors)

$\mathbb{E}\|\mathbf{f}_{x,t}\|^4 \leq C < \infty$, $T^{-1}\mathbf{F}_x'\mathbf{F}_x \xrightarrow{p} \mathbf{\Sigma}_{F_x} > 0$ *as* $T \to \infty$ *for some non-random positive definite matrix* $\mathbf{\Sigma}_{F_x}$. $\mathbb{E}\|\mathbf{f}_t\|^4 \leq C < \infty$, $T^{-1}\mathbf{F}'\mathbf{F} \xrightarrow{p} \mathbf{\Sigma}_F > 0$ *as* $T \to \infty$ *for some non-random positive definite matrix* $\mathbf{\Sigma}_F$. $\mathbf{f}_{x,t}$ *and* $\mathbf{f}_t$ *are group-wise independent from* $\mathbf{v}_{i,t}$ *and* $\varepsilon_{i,t}$.

Assumption (factor loadings)

$\mathbf{\Gamma}_i \sim \mathrm{i.i.d}(\mathbf{0}, \mathbf{\Sigma}_\Gamma)$, $\boldsymbol{\lambda}_i \sim \mathrm{i.i.d}(\mathbf{0}, \mathbf{\Sigma}_\lambda)$, *where* $\mathbf{\Sigma}_\Gamma$ *and* $\mathbf{\Sigma}_\lambda$ *are positive definite.* $\mathbb{E}\left\|\mathbf{\Gamma}_i\right\|^4 \leq C < \infty$, $\mathbb{E}\left\|\boldsymbol{\lambda}_i\right\|^4 \leq C < \infty$. *In addition,* $\mathbf{\Gamma}_i$ *and* $\boldsymbol{\lambda}_i$ *are independent groups from* $\varepsilon_{i,t}$, $\mathbf{v}_{i,t}$, $\mathbf{f}_{x,t}$ *and* $\mathbf{f}_t$.

Assumption (spatial weighting matrix)

1. *All diagonal elements of* $\mathbf{W}_N$ *are zeros;*
2. *The row and column sums of the matrices* $\mathbf{W}_N$ *and* $(\mathbf{I}_N - \psi \mathbf{W}_N)^{-1}$ *are bounded uniformly in absolute value.*
3. $$\sum_{\ell=0}^{\infty} \|[\alpha(\mathbf{I}_N - \psi \mathbf{W}_N)^{-1}]^{\ell}\|_{\infty} \leq C\,; \quad \sum_{\ell=0}^{\infty} \|[\alpha(\mathbf{I}_N - \psi \mathbf{W}_N)^{-1}]^{\ell}\|_{1} \leq C$$

Assumption (identification)

1. $\overline{\mathbf{A}} = plim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{C}_i$ *is fixed with full column rank, and* $\overline{\mathbf{B}} = plim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i$ *is fixed and positive definite.*
2. $\mathbb{E}\left\| T^{-1} \mathbf{Z}_i' \mathbf{Z}_i \right\|^{2+2\delta} \leq C < \infty$ *and* $\mathbb{E}\left\| T^{-1} \mathbf{Z}_i' \mathbf{C}_i \right\|^{2+2\delta} \leq C < \infty$ *for all* i, T.

1SIV - Asymptotic Properties

Proposition

Under Assumptions A-F, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{u}_i + \sqrt{\frac{T}{N}}\mathbf{b}_1 + \sqrt{\frac{N}{T}}\mathbf{b}_2 + o_p(1).$$

Theorem

Under Assumptions A-F, as $N, T \to \infty$ such that $N/T \to c$, where $0 < c < \infty$, we have

$$\sqrt{NT}\left(\widehat{\boldsymbol{\theta}}_{1SIV} - \boldsymbol{\theta}\right) = O_p(1)\ . \tag{9}$$

2SIV - Asymptotic Properties

Theorem

Under Assumptions A-F, as $N, T \to \infty$ such that $N/T \to c$, where $0 < c < \infty$, $\widehat{\boldsymbol{\theta}}_{2SIV}$ is consistent and

$$\sqrt{NT}\left(\widehat{\boldsymbol{\theta}}_{2SIV} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Psi}\right)$$

where $\boldsymbol{\Psi} = \sigma_\varepsilon^2 \left(\mathbf{A}_0' \mathbf{B}_0^{-1} \mathbf{A}_0\right)^{-1}$, $\mathbf{A}_0 = \operatorname{plim}_{N,T\to\infty} \mathbf{A}$, $\mathbf{B}_0 = \operatorname{plim}_{N,T\to\infty} \mathbf{B}$, *with*

$$\mathbf{A} = \frac{1}{NT}\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{C}_i \, , \mathbf{B} = \frac{1}{NT}\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i \, .$$

Moreover, $\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi} \xrightarrow{p} \mathbf{0}$ *as* $N, T \to \infty$, *where* $\widetilde{\boldsymbol{\Psi}} = \tilde{\sigma}_\varepsilon^2 \left(\widetilde{\mathbf{A}}' \widetilde{\mathbf{B}}^{-1} \widetilde{\mathbf{A}}\right)^{-1}$.

Outline of MGIV procedure

We further assume that $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \mathbf{e}_i$, where $\boldsymbol{\theta} = (\alpha, \psi, \boldsymbol{\beta}')'$ and $\mathbf{e}_i$ is a random i.i.d. error with mean zero and variance $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$.

First, recover estimates of $\boldsymbol{\theta}_i$ as follows:

$$\widehat{\boldsymbol{\theta}}_i = \left(\widehat{\mathbf{A}}_i'\widehat{\mathbf{B}}_i^{-1}\widehat{\mathbf{A}}_i\right)^{-1}\widehat{\mathbf{A}}_i'\widehat{\mathbf{B}}_i^{-1}\widehat{\mathbf{c}}_i,$$

where

$$\widehat{\mathbf{A}}_i = T^{-1}\widehat{\mathbf{Z}}_i'\mathbf{C}_i, \quad \widehat{\mathbf{B}}_i = T^{-1}\widehat{\mathbf{Z}}_i'\widehat{\mathbf{Z}}_i, \quad \widehat{\mathbf{c}}_i = T^{-1}\widehat{\mathbf{Z}}_i'\mathbf{y}_i,$$

and $\widehat{\mathbf{Z}}_i = (\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{i,-1},\ \sum_{j=1}^{N} w_{ij}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_j,\ \mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i)$ as the instruments in the estimation of $\boldsymbol{\theta}_i$.

Subsequently, we construct the sample average of $\widehat{\boldsymbol{\theta}}_i$ as

$$\widehat{\boldsymbol{\theta}}_{MGIV} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\theta}}_i.$$

MGIV - Asymptotic Properties

Theorem

Under aforementioned assumptions, as $N, T \to \infty$ *such that* $N/T^2 \to 0$, *then* $\widehat{\boldsymbol{\theta}}_{MGIV}$ *is consistent for the population mean* $\boldsymbol{\theta}$. *If, it further holds that* $N/T^{6/5} \to 0$, *then* $\widehat{\boldsymbol{\theta}}_{MGIV}$ *has the following asymptotic distribution*

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{MGIV} - \boldsymbol{\theta}) \xrightarrow{d} N\left(\mathbf{0},\ \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right), \quad \text{as } N, T \to \infty.$$

Remarks

- DGP for covariates: $x_{i,t} = \boldsymbol{\gamma}_i' \mathbf{f}_t + v_{i,t}$;

  - Frequently employed in economics or econometrics (e.g. Pesaran et al. 2013, Westerlund-Urbain 2015, Hansen-Liao 2018 etc.)

- As least one element of $\boldsymbol{\beta}$ must not equal to zero;

- Endogeneity of $\mathbf{X}_i \Rightarrow$ use external instruments, e.g. $\widetilde{\mathbf{Z}}_i = \widetilde{\mathbf{F}} \widetilde{\boldsymbol{\Gamma}}_i + \widetilde{\mathbf{V}}_i$;

Simulation Experiments - Design

$$y_{i,t} = \eta_i + \alpha y_{i,t-1} + \psi \sum_{j=1}^{N} w_{i,j} y_{j,t} + \sum_{\ell=1}^{2} \beta_\ell x_{\ell,i,t} + \sum_{s=1}^{r_y} \lambda_{s,i} f_{s,t} + \varepsilon_{i,t}; \quad r_y = 3. \tag{10}$$

$$x_{\ell,i,t} = \mu_{\ell,i} + \sum_{s=1}^{r_x} \gamma_{\ell,s,i} f_{(x),s,t} + v_{\ell,i,t}; \quad r_x = 2; \quad i = 1, \ldots, N; t = -49, \ldots, T. \tag{11}$$

$i = 1, \ldots, N; t = -49, \ldots, T.$

$\varepsilon_{i,t}$, is non-normal and heteroskedastic across both i and t, such that $\varepsilon_{i,t} = \varsigma_\varepsilon \sigma_{i,t}(\epsilon_{i,t} - 1)/\sqrt{2}$, $\epsilon_{i,t} \sim i.i.d.\chi^2_1$, with $\sigma^2_{i,t} = g_i \phi_t$, $g_i \sim i.i.d.\chi^2_2/2$, and $\phi_t = t/T$ for $t = 0, 1, ..., T$ and unity otherwise.

All remaining time-varying error components are AR(1) processes with parameter equal to .5. All individual-specific error components are correlated, $\alpha_{\eta,\mu} = 0.5$.

We set ς^2_ε s.t. $\pi_u \in \{1/4, 3/4\}$. We set $var(v_{\ell it})$ s.t. $SNR = 4$. We set $\alpha = 0.4$, $\psi = 0.25$, and $\beta_1 = 3$ and $\beta_2 = 1$, following Bai (2009).

Variance estimator for 2SIV

In order to allow for cross-section and time series heteroskedasticity, the variance estimator for 2SIV is

$$\widetilde{\boldsymbol{\Psi}} = \left(\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}}\right)^{-1}\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widehat{\boldsymbol{\Omega}}\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}}\left(\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}}\right)^{-1}, \tag{12}$$

with

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\hat{\mathbf{Z}}_i, \tag{13}$$

and $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{C}_i\hat{\theta}_{1SIV}$.

$$\hat{\mathbf{Z}}_i = \left(\mathbf{M}_{\widehat{\mathbf{F}}}\underline{\mathbf{X}}_i, \quad \mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\underline{\mathbf{X}}_{i,-1}, \quad \mathbf{M}_{\widehat{\mathbf{F}}}\sum_{j=1}^{N}w_{ij}\underline{\mathbf{X}}_j, \quad \mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\sum_{j=1}^{N}w_{ij}\underline{\mathbf{X}}_{j,-1}\right). \tag{14}$$

Results (homogeneous case)

Table 1A. RMSE, $\pi_u = 3/4$

$\alpha(= 0.4)$	2SIV			QMLE		
τ	1	2	4	1	2	4
$N = 100\tau,\ T = 25\tau$	.017	.007	.004	.014	.006	.003
$N = 25\tau,\ T = 100\tau$	.014	.007	.003	.008	.004	.002
$N = 50\tau,\ T = 50\tau$	.015	.003	.003	.009	.005	.002

$\psi(= 0.25)$	2SIV			QMLE		
τ	1	2	4	1	2	4
$N = 100\tau,\ T = 25\tau$	.019	.008	.004	.033	.015	.006
$N = 25\tau,\ T = 100\tau$	.017	.008	.004	.023	.010	.005
$N = 50\tau,\ T = 50\tau$	.017	.008	.004	.025	.011	.005

$\beta_2(= 1)$	2SIV			QMLE		
τ	1	2	4	1	2	4
$N = 100\tau,\ T = 25\tau$	.066	.025	.012	.093	.029	.012
$N = 25\tau,\ T = 100\tau$	.049	.025	.012	.118	.040	.014
$N = 50\tau,\ T = 50\tau$	.050	.024	.012	.086	.028	.012

Results ctd

Table 1B. ARB, $\pi_u = 3/4$, $ARB \equiv \left(|\widehat{\theta}_\ell - \theta_\ell|/\theta_\ell\right) \times 100$

$\alpha(= 0.4)$		2SIV			QMLE	
τ	1	2	4	1	2	4
$N = 100\tau,\ T = 25\tau$	.082	.104	.054	2.52	1.08	.551
$N = 25\tau,\ T = 100\tau$	.116	.033	.008	.169	.087	.080
$N = 50\tau,\ T = 50\tau$	.074	.026	.017	.870	.493	.225

$\psi(= 0.25)$		2SIV			QMLE	
τ	1	2	4	1	2	4
$N = 100\tau,\ T = 25\tau$	.119	.112	.080	.216	.159	.113
$N = 25\tau,\ T = 100\tau$	.054	.009	.016	.156	.226	.007
$N = 50\tau,\ T = 50\tau$	.218	.039	.012	.026	.092	.001

$\beta_2(= 1)$		2SIV			QMLE	
τ	1	2	4	1	2	4
$N = 100\tau,\ T = 25\tau$	.947	.087	.039	6.41	1.35	.379
$N = 25\tau,\ T = 100\tau$	.043	.043	.003	7.10	1.81	.361
$N = 50\tau,\ T = 50\tau$	.222	.023	.023	5.18	1.15	.228

Results ctd

Table 1C. Size (nominal size = 5%), $\pi_u = 3/4$

$\alpha(=0.4)$	2SIV			QMLE		
τ	1	2	4	1	2	4
$N = 100\tau,\ T = 25\tau$	.065	.059	.052	.361	.293	.286
$N = 25\tau,\ T = 100\tau$	.084	.066	.048	.138	.085	.079
$N = 50\tau,\ T = 50\tau$	.052	.056	.053	.146	.121	.109

$\psi(=0.25)$	2SIV			QMLE		
τ	1	2	4	1	2	4
$N = 100\tau,\ T = 25\tau$	.062	.051	.052	.392	.258	.134
$N = 25\tau,\ T = 100\tau$	.094	.062	.054	.156	.078	.070
$N = 50\tau,\ T = 50\tau$	.076	.060	.054	.213	.139	.068

$\beta_2(=1)$	2SIV			QMLE		
τ	1	2	4	1	2	4
$N = 100\tau,\ T = 25\tau$	.109	.057	.055	.415	.142	.079
$N = 25\tau,\ T = 100\tau$	.086	.074	.059	.531	.306	.134
$N = 50\tau,\ T = 50\tau$	.063	.052	.050	.369	.156	.090

Spillovers and Drivers of U.S. Population Growth

Kripfganz-Sarafidis 2025 specify:

$$\Delta y_{i,t} = \delta_i y_{i,t-1} + \psi_i \sum_{j=1}^{N} w_{i,j} \Delta y_{j,t} + \sum_{\ell=1}^{4} \beta_{\ell,i} x_{\ell,i,t} + \boldsymbol{\lambda}_i \mathbf{f}_t + \eta_i + \tau_t + \varepsilon_{i,t}, \tag{15}$$

$$i = 1, \ldots, N(= 49) \quad \text{and } t = 1, \ldots, T(= 52), \text{ i.e. 1966 - 2017.}$$

$\Delta y_{i,t} \equiv \ln(pop_{i,t}) - \ln(pop_{i,t-1})$: population growth;
$y_{i,t-1} \equiv \ln(pop_{i,t-1})$: lagged level of logged population;
$\Delta y_{j,t} \equiv \ln(pop_{j,t})$: population growth of state j;
$x_{1,i,t} \equiv TFP_{i,t}$: "Total Factor Productivity";
$x_{2,i,t} \equiv \ln(amenities)_{i,t}$: logged "amenities";
$x_{3,i,t} \equiv \ln(income_{i,t})$: logged labour income;
$x_{4,i,t} \equiv \ln(MigCost)_{i,t}$: logged migration frictions.

$\mathbf{f}_t :=$ nationwide demographic/institutional shocks (changes in fertility/aging patterns, shifts in federal immigration policy, national housing and credit cycles).

$\delta_i :=$ conditional convergence: pop. converges (diverges) if $\delta_i < 0$ ($\delta_i > 0$) $\forall\ i$.

Network Specification

Our baseline model estimates $\mathbf{W}$ from the data using the method of Juodis-Kapetanios-Sarafidis 2025.

We also consider contiguity and distance-based networks:

$$w_{i,j} = \begin{cases} 1, & \text{if } j \text{ shares common border with } i \\ 0, & \text{otherwise,} \end{cases} \tag{16}$$

and

$$w_{i,j} = \begin{cases} 1/d_{i,j}, & \text{if } d_{i,j} \leq c \\ 0, & \text{otherwise,} \end{cases} \tag{17}$$

where c denotes the cth percentile of the empirical distribution of $d_{i,j}$.

Finally, we set $w_{i,j,t} = 1/(\text{trade-inflows})_{i,j,t}$, where $(\text{trade-inflows})_{i,j,t}$ represents the value of trade inflows from state j to state i at time t.

Instruments & Marginal Effects

We use the following instruments:

$$\widehat{\mathbf{Z}}_i = \Big(\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i,\;\; \mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{i,-1},\;\; \mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-2}}\mathbf{X}_{i,-2},$$
$$\mathbf{M}_{\widehat{\mathbf{F}}}\sum_{j=1}^{N} w_{i,j}\mathbf{X}_j,\;\; \mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\sum_{j=1}^{N} w_{i,j}\mathbf{X}_{j,-1},\;\; \mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-2}}\sum_{j=1}^{N} w_{i,j}\mathbf{X}_{j,-2}\Big). \tag{18}$$

Distinguish among **direct, indirect and total effects**:

$$\mathbf{y}_{(t)} = [\mathbf{I}_N - \psi\mathbf{W}]^{-1}\left(\delta\mathbf{y}_{(t-1)} + \sum_{\ell=1}^{k}\beta_\ell\mathbf{x}_{\ell(t)} + \mathbf{u}_{(t)}\right), \tag{19}$$

$$\left[\frac{\partial E(\mathbf{y})}{\partial x_{\ell 1}} \dots \frac{\partial E(\mathbf{y})}{\partial x_{\ell N}}\right] = [\mathbf{I}_N - \psi\mathbf{W}]^{-1}\beta_\ell. \tag{20}$$

spxtivdfreg command in Stata

Kripfganz, S. and V. Sarafidis (2025). Estimating Spatial Dynamic Panel Data Models with Common Factors, *Journal of Statistical Software*, 113(6), 1–27.

For time-invariant networks, we specify:

```
spxtivdfreg D.lnPop L.lnPop TFP Amenity
lnLaborInc_pc MigTradeCost, absorb(statefp
year) splag spmatrix("W_contiguity.xlsx",
import) iv(TFP Amenity lnLaborInc_pc
MigTradeCost, splags lag(2)) mg factmax(4)
```

For time-varying networks we use:

```
spxtivdfreg D.lnPop L.lnPop TFP Amenity
lnLaborInc_pc MigTradeCost, absorb(statefp
year) splag spmatrix("W_*_trade.xlsx", import)
iv(TFP Amenity lnLaborInc_pc MigTradeCost,
splags lag(2)) mg factmax(4)
```

spxtivdfreg command in Stata

If we want to include: (i) spatial-time lags; (ii) spatial lags of covariates (SDM); and (iii) lags of the dependent variable, we specify:

```
spxtivdfreg D.lnPop L.lnPop TFP Amenity
lnLaborInc_pc MigTradeCost, absorb(statefp
year) splag sptlags(#) spindepvars(varlist)
tlags(#) spmatrix("W_contiguity.xlsx", import)
iv(TFP Amenity lnLaborInc_pc MigTradeCost,
splags lag(2)) mg factmax(4)
```

The command is very flexible, as it allows: defactoring individual covariates (useful for common covariates without cross-sectional variation), saving individual-specific estimated coefficients, computing direct/indirect/total effects, computing spillins/spillouts, and many other options.

Table 1: Homogeneous (2SIV) Slope Coefficients

	2SIV					
	$\widehat{\mathbf{W}}$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_5$
$y_{i,t-1}$ (lagged population)	-0.016***	-0.022***	-0.014***	-0.019***	-0.025***	-0.025***
	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
$x_{1,i,t}$ (TFP)	0.000	0.002	0.002**	0.003**	0.002**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$x_{2,i,t}$ (amenities)	0.005***	0.006***	0.006***	0.006***	0.007***	0.006***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
$x_{3,i,t}$ (labour income)	0.038***	0.030***	0.033***	0.031***	0.032***	0.033***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.003)	(0.004)
$x_{4,i,t}$ (migration cost)	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\sum_{j=1}^{N} w_{i,j} \Delta y_{j,t}$ (spatial lag)	0.335***	0.083***	0.169***	0.135***	0.230*	0.538***
	(0.038)	(0.013)	(0.037)	(0.034)	(0.125)	(0.141)
N	49	49	49	49	49	49
χ^2_J	36.504	35.064	33.895	33.261	34.584	34.678
p_J	0.006	0.009	0.013	0.016	0.011	0.010
r_x	2	2	2	2	2	2
r_y	1	1	1	1	1	1
ρ	0.614	0.702	0.745	0.744	0.739	0.700

$\widehat{\mathbf{W}}$ denotes the data-driven network obtained based on Juodis-Kapetanios-Sarafidis 2025. $\mathbf{W}_1$ is contiguity; $\mathbf{W}_2$–$\mathbf{W}_3$ are inverse-distance (10th/5th percentile cutoffs); $\mathbf{W}_4$–$\mathbf{W}_5$ are trade-based (time-averaged/time-varying).

Heterogeneity in Conditional Convergence

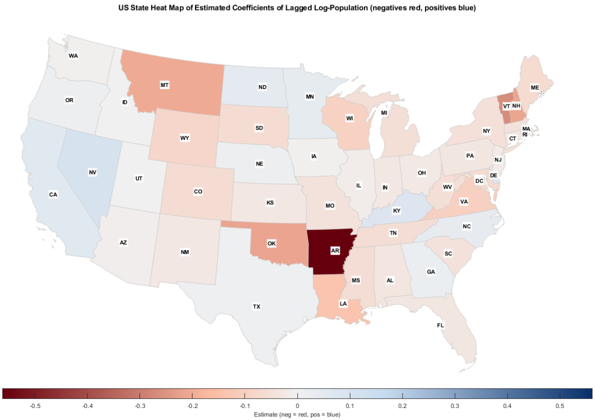

Figure: U.S. Heat Map of Estimated Coefficients of Lagged Log-Population (negatives red, positives blue).

Average Marginal Effects - Direct/Indirect

	MGIV					
	$\widehat{\mathbf{W}}$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_5$
Direct effects						
Lagged population	-0.051***	-0.046***	-0.048***	-0.050***	-0.050***	-0.052
	(0.015)	(0.013)	(0.013)	(0.014)	(0.014)	—
TFP	0.005**	0.001	0.003	0.003	0.005	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	—
Amenities	0.007***	0.006***	0.006***	0.006***	0.006***	0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	—
Labour income	0.051***	0.051***	0.052***	0.053***	0.052***	0.051
	(0.012)	(0.011)	(0.012)	(0.013)	(0.011)	—
Migration cost	-0.004***	-0.002***	-0.003***	-0.002***	-0.003***	-0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	—
Indirect effects						
Lagged population	-0.028**	-0.049***	-0.039***	-0.030***	-0.127	-0.075
	(0.011)	(0.018)	(0.013)	(0.009)	(0.135)	—
TFP	0.003**	0.001	0.002	0.002	0.014	0.005
	(0.001)	(0.003)	(0.003)	(0.002)	(0.016)	—
Amenities	0.004***	0.006***	0.005***	0.004***	0.016	0.008
	(0.001)	(0.002)	(0.002)	(0.001)	(0.017)	—
Labour income	0.028***	0.054**	0.043***	0.032***	0.131	0.074
	(0.010)	(0.022)	(0.014)	(0.009)	(0.137)	—
Migration cost	-0.002***	-0.002***	-0.002***	-0.001***	-0.008	-0.004
	(0.000)	(0.001)	(0.001)	(0.000)	(0.008)	—

$\widehat{\mathbf{W}}$ denotes the data-driven network obtained based on Juodis-Kapetanios-Sarafidis 2025. $\mathbf{W}_1$ is contiguity; $\mathbf{W}_2$–$\mathbf{W}_3$ are inverse-distance (10th/5th percentile cutoffs); $\mathbf{W}_4$–$\mathbf{W}_5$ are trade-based (time-averaged/time-varying).

Spillouts originating from Midwest States

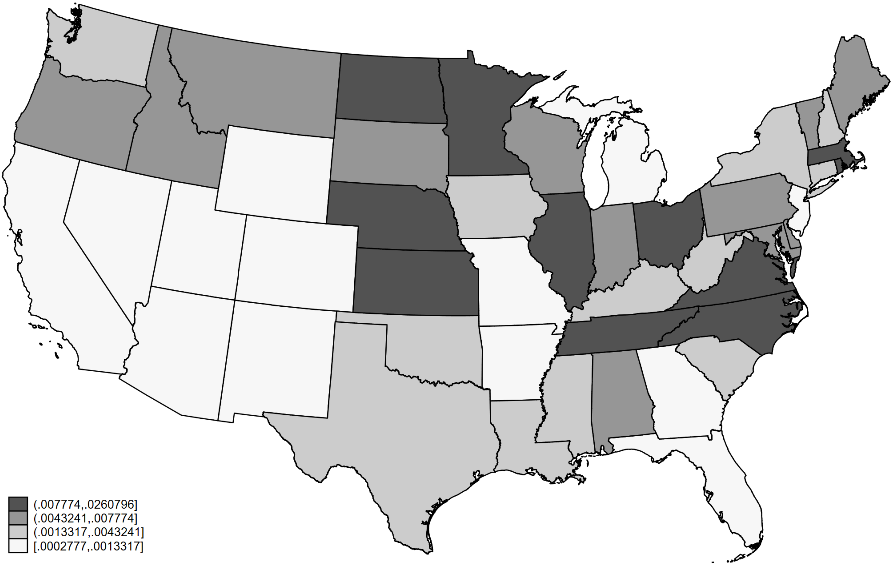

Figure: Spillouts Originating from the Midwest Region.

U.S. Population Growth Network

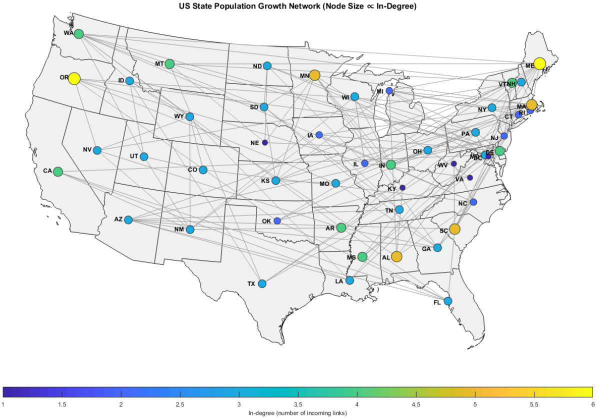

Figure: U.S. States Population Growth Network.

Concluding Remarks

- We put forward two IV estimators for homogeneous and heterogeneous models; Both are asy. unbiased, consistent and asy. normal under N, $T \to \infty$;
- Both IV estimators are **linear and computationally inexpensive**, and can accommodate **endogeneity**;
- *spxtivdfreg* implements these estimators in Stata and is very flexible as it allows for:
  - importing W from a csv/Excel file;
  - W being time-varying;
  - different instrumentation and defactoring strategies;
  - computation of direct/indirect effects and spillins/spillouts.

Pillars of the *spxtivdfreg* command

The *spxtivdfreg* Stata command draws upon the following work:

- Cui, G., Chen, J., Sarafidis, V., and T. Yamagata. 2025. Estimating Spatial Dynamic Panel Data Models with Common Factors in Stata. Unpublished Manuscript, arXiv:2501.18467v1.
- Cui, G., Sarafidis, V., and T. Yamagata. 2023. IV Estimation of Spatial Dynamic Panels with Interactive Effects: Large Sample Theory and an Application on Bank Attitude Toward Risk. *Econometrics Journal*, Vol. 26(2), pp. 124-146.
- Juodis, A., Sarafidis, V., and G. Kapetanios. 2025. Identification and estimation of panel network models using high-dimensional regression. Unpublished Manuscript.
- Kripfganz, S., and V. Sarafidis. 2025. Estimating Spatial Dynamic Panel Data Models with Common Factors in Stata. *Journal of Statistical Software*, Vol. 113(6), pp. 1-27.
- Kripfganz, S., and V. Sarafidis. 2025. Chasing Opportunity: Spillovers and Drivers of U.S. State Population Growth. Invited contribution.

Thank you so much for listening!