A Panel Data Approach for Spatial and Network Selection Models

Sophia Ding[1], Peter H. Egger[2]
[1]**ETH Zurich;** ETH Zurich, CEPR, CESifo

1 Introduction

Motivation:

- Common features of economic data: **spatial/network pattern** (cross-sectional interdependence), non-randomly missing observations (**sample selection/treatment selection**), **panel-data**
- No model exists to deal with all three simultaneously!
- Neglecting selection and/or spatial/network correlation results in **biased** coefficient estimates!

	Cross Section	Panel
Non-Spatial/ Non-Network	Heckman (1976, 1979)	Wooldridge (1995)
Spatial/ Network	McMillen (1995), Flores-Lagunes, Schnier (2012), Doğan,Taşpinar (2017)	**This paper!**

Example/Future Application: Export-Wage Premium

- Empirical and theoretical evidence that **exporters pay higher wage/worker** than non-exporting firms (**treatment effect** of exporter status)
- Exporting decision as well as wage/worker depends on latent export profitability → **treatment ≠ random**
- Wages may have a **spatial pattern** due to local labor markets, commuting, etc. → Shocks to wages are **correlated across firms**!
- Profitability of exporting may have **network pattern** due to input/output linkages or industry affiliation → Shocks to profitability are **correlated across firms**!

This paper:
Develop **two-step approach** towards (sample/treatment) selection on unobservables akin to Heckman (1976, 1979) and Wooldridge (1995) but for **panel-data** with **spatial or network interdependencies**.
Focus here: **Spatial/Network Treatment Selection Model**

2 Econometric Model

Selection equation

- (1) Fixed Effects/corr. Random Effects (Mundlak 1978, Wooldridge 1995)
- (2)Panel Spatial autoregr. process (Kapoor, Kelejian, and Prucha, 2007)

$$y_{ti}^{A*} = x_{ti}^{A\prime}\beta^A + e_{ti}^A, \quad y_{ti}^A = 1[y_{ti}^{A*} > 0]$$

$$e_{ti}^A = \rho^A \sum_{j=1}^{N} w_{tij} e_{tj}^A + \bar{x}_i^{A\prime}\delta^A + \underbrace{\mu_i^A + \varepsilon_{ti}^A}_{=\xi_{ti}^A}$$

$$e_{ti}^A = \sum_{j=1}^{N} r_{tij}^A \bar{x}_j^{A\prime}\delta^A + \underbrace{\sum_{j=1}^{N} r_{tij}^A \xi_{tj}^A}_{=u_{ti}^A}, \quad \text{using} \quad R_t^A = (I_N - \rho^A W_t)^{-1} = (r_{tij}^A)$$

Outcome equation

- (1) Fixed Effects + (2) panel SAR + (3) joint normality of errors

$$y_{ti}^{B*} = \alpha y_{ti}^A + x_{ti}^{B\prime}\beta^B + \sum_{j=1}^{N} r_{tij}^B \bar{x}_j^{B\prime}\delta^B + u_{ti}^B, \quad y_{ti}^B = \begin{cases} y^{B*} & \text{if } y_{ti}^A = 1 \\ y^{B*} & \text{if } y_{ti}^A = 0 \end{cases}$$

$$\begin{pmatrix} u_{ti}^A \\ u_{ti}^B \end{pmatrix} | x^A, x^B \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\xi^A}^2 \sum_{j=1}^N (r_{tij}^A)^2 & \sigma_{\xi^{AB}} \sum_{j=1}^N r_{tij}^B r_{tij}^A \\ \sigma_{\xi^{BA}} \sum_{j=1}^N r_{tij}^B r_{tij}^A & \sigma_{\xi^B}^2 \sum_{j=1}^N (r_{tij}^B)^2 \end{pmatrix} \right)$$

Correcting for Selection Bias

- Conditional Expectation

$$E[y_{ti}^B | y_{ti}^A, x^{A0}, x^B] = \alpha y_{ti}^A + x_{ti}^{B\prime}\beta^B + \sum_{j=1}^{N} r_{tij}^B \bar{x}_j^{B\prime}\delta^B + \boxed{E[u_{ti}^B | y_{ti}^A, x^{A0}, x^B]}$$

- Adjusted Generalized Inverse Mills Ratio

$$E[u_{ti}^B | y_{ti}^A, x^{A0}, x^B] = \frac{\sigma_{\xi^{BA}}}{\sqrt{\sigma_{\xi^A}^2}} \frac{\sum_{j=1}^N r_{tij}^B r_{tij}^A}{\sum_j^N (r_{tij}^A)^2} \left[y_{ti}^A \frac{\phi(z_{ti})}{\Phi(z_{ti})} + (1 - y_{ti}^A) \frac{\phi(z_{ti})}{1 - \Phi(z_{ti})} \right]$$

$$= \tau \psi_{ti} \lambda_{ti}^g$$

3 Estimation Strategy (Outline)

Step 1: Estimate selection equation using **Pooled Bayesian Spatial/ Network Error Probit** model to obtain $\hat{\theta}_A = \{\hat{\tilde{\beta}}^A, \hat{\tilde{\delta}}^B, \hat{\rho}^A\}$, where $\tilde{\beta}^A = \frac{\beta^A}{\sigma_{\xi^A}}$, $\tilde{\delta}^A = \frac{\delta^A}{\sigma_{\xi^A}}$.

Step 2: Use estimated parameters to construct spatially/network adjusted (generalized) Inverse Mills' Ratio.

Step 3: Add estimated spatially/network adjusted generalized Inverse Mills' Ratio in outcome equation and estimate using **Pooled Non-linear Least Squares** to obtain $\hat{\theta}^B = \{\hat{\alpha}, \hat{\beta}^B, \hat{\delta}^B, \hat{\tau}, \hat{\rho}^B\}$.

4 Variance-Covariance Matrix

- Account for **estimated** first-stage parameters: **Murphy-Topel** (1985, 2002) type of **correction** for two-step estimators.
- Corrected VC-Matrix is a function of the **truncated variance** and **truncated covariance** of the spatial error components: outline estimation procedure along the lines of Heckman (1979).

5 Monte Carlo Evidence (Selected Results)

Case 1: Medium Spatial/Network Correlation

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	α	β_1^B	δ_1^B	τ	ρ^B
		True	0.707	0.707	0.707	0.707	0.5	1	1	3	0.707	0.5
N=250	SNTS	Mean	0.732	0.735	0.736	0.734	0.449	1.006	1.002	3.005	0.694	0.494
		Bias	0.025	0.028	0.029	0.027	-0.051	0.006	0.002	0.005	-0.013	-0.006
		RMSE	0.091	0.084	0.198	0.185	0.131	0.207	0.060	0.138	0.146	0.091
	WPS	Mean	0.653	0.670	0.681	0.708		0.849	1.023	3.076	0.873	
Ignore spatial/ network correlation		Bias	-0.054	-0.037	-0.026	0.001		-0.151	0.023	0.076	0.166	
		RMSE	0.096	0.081	0.140	0.134		0.257	0.065	0.159	0.231	
	NLLS	Mean						1.379	0.940	2.951		0.528
Ignore sample selection		Bias						0.379	-0.059	-0.049		0.028
		RMSE						0.411	0.082	0.142		0.094
N=500	SNTS	Mean	0.716	0.719	0.721	0.711	0.482	1.004	0.999	3.002	0.698	0.495
		Bias	0.009	0.012	0.014	0.004	-0.018	0.004	-0.001	0.002	-0.009	-0.005
		RMSE	0.055	0.060	0.134	0.140	0.074	0.160	0.040	0.105	0.108	0.055
	WPS	Mean	0.660	0.657	0.686	0.671		1.027	0.995	3.160	0.779	
Ignore spatial/ network correlation		Bias	-0.047	-0.050	-0.022	-0.036		0.027	-0.005	0.160	0.072	
		RMSE	0.069	0.074	0.098	0.109		0.163	0.040	0.193	0.140	
	NLLS	Mean						1.390	0.939	2.938		0.490
Ignore sample selection		Bias						0.390	-0.061	-0.062		-0.010
		RMSE						0.408	0.071	0.117		0.060

Case 2: No Spatial/Network Correlation

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	α	β_1^B	δ_1^B	τ	ρ^B
		True	0.707	0.707	0.707	0.707	0	1	1	3	0.707	0
N=250	SNTS	Mean	0.739	0.740	0.743	0.744	-0.114	1.005	1.000	3.006	0.697	-0.009
		Bias	0.032	0.033	0.036	0.037	-0.114	0.005	0.000	0.006	-0.010	-0.009
		RMSE	0.092	0.085	0.193	0.177	0.245	0.112	0.051	0.132	0.106	0.137
	WPS	Mean	0.715	0.716	0.719	0.718		1.000	0.999	3.009	0.706	
Ignore spatial/ network correlation		Bias	0.008	0.009	0.011	0.011		0.000	-0.001	0.009	-0.001	
		RMSE	0.082	0.075	0.143	0.140		0.112	0.050	0.130	0.106	
	NLLS	Mean						1.350	0.942	2.973		0.067
Ignore sample selection		Bias						0.350	-0.058	-0.027		0.067
		RMSE						0.362	0.076	0.130		0.148
N=500	SNTS	Mean	0.721	0.723	0.730	0.718	-0.066	1.003	0.999	3.002	0.701	-0.006
		Bias	0.014	0.016	0.023	0.011	-0.066	0.003	-0.001	0.002	-0.006	-0.006
		RMSE	0.055	0.059	0.133	0.133	0.176	0.086	0.033	0.099	0.076	0.087
	WPS	Mean	0.709	0.711	0.714	0.706		1.001	0.999	3.002	0.705	
Ignore spatial/ network correlation		Bias	0.002	0.004	0.007	-0.001		0.001	-0.001	0.002	-0.002	
		RMSE	0.052	0.056	0.102	0.102		0.086	0.033	0.096	0.076	
	NLLS	Mean						1.358	0.942	2.935		-0.016
Ignore sample selection		Bias						0.358	-0.058	-0.065		-0.016
		RMSE						0.365	0.066	0.114		0.093

References

Doğan, O., Taşpinar, S., 2017. Bayesian Inference in Spatial Sample Selection Models. Oxford Bulleting of Economics and Statistics 80, 90-121.

Flores-Lagunes, A., Schnier, K.E., 2012. Estimation of Sample Selection Models with Spatial Dependence. Journal of Applied Econometrics 27, 173-204.

Heckman, J., 1976. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. The Annals of Economic and Social Measurement 5, 475-492.

Heckman, J., 1979. Sample Selection Bias as a Specification Error. Econometrica 47 (1), 153-61.

Kapoor, M, Kelejian, H.H. Prucha, I.R., 2007. Panel Data Models with Spatially Correlated Error Components. Journal of Econometrics 140, 97-130.

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46, 69-85.

McMillen, D.P., 1995. Selection Models in Spatial Econometric Models. Journal of Regional Science 35 (3), 417-436.

Wooldridge, J., 1995. Selection corrections for panel data models under conditional mean independence assumptions. Journal of Econometrics 68 (1), 115-132.