# The Range of Interaction in Spatial Autoregressive Econometric Models

Ruxin Chen

Thesis Submitted for Master of Computational Social Science

Supervised by Pr.Luc Anselin

2019

# Research Question

- Motivation
  - Tobler's first law of geography: everything is related to everything else, but near things are more related than distant things.
  - Modern spatial econometric models take into account of the correlation for units that are closely located. The range of interaction of models indicates how far two units should be accounted for their impact on each other.
  - Commonly used spatial econometric models rely heavily on the assumption of the range of interaction. However, the spatial dependence structure is an unknown priori.
- Reseach Question
  - Conduct Monte Carlo simulations to find the effect on inference if the model is mis-specified
  - Summarize the discussion in the most recent literature related to specification and estimation of spatial dependence structure

# Spatial Econometric Model 1: SAR

- The Spatial Autoregressive Model (SAR)
  - Model specification

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X}\beta + \epsilon \tag{1}$$

  - Reduced form

$$\mathbf{y} = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X}\beta + (\mathbf{I} - \rho \mathbf{W})^{-1} \epsilon \tag{2}$$

  - Leontief expansion

$$(\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X}\beta = \mathbf{X}\beta + \rho \mathbf{W}\mathbf{X}\beta + \rho^2 \mathbf{W}^2 \mathbf{X}\beta^2 + \rho^3 \mathbf{W}^3 \mathbf{X}\beta^3 + \cdots \tag{3}$$

- Global Interaction: the value of $y$ at location $i$ is determined by the value of $x$ at location $i$ and all other locations through the their dependence with location $i$ such dependence structure is specified by the spatial weight matrix $\mathbf{W}$.

# Alternative Autoregressive Models

- The Spatial Error Model (SEM)
  - Model specification

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u} \tag{4}$$
$$\mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \epsilon \tag{5}$$

- The Spatial Durbin Model (SDM)
  - Model specification

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\gamma + \epsilon \tag{6}$$

  - Problem of Identification

# The Spatial Econometric Model 4: SLX

- The Spatial Lag Model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{W}\mathbf{X}\gamma + \epsilon \tag{7}$$

  - Unlike models mentioned previously, SLX imply local interaction: spatial effect is only limited to direct neighbors
  - Advantage over SAR
    1. No endogeneity problem. Hence, it can be estimated directly using OLS
    2. Furthermore, it allows the spatial weight matrix to be parameterized – easy to combine with more complicated methods
    3. Parameters are easy to interpret

# Problems for Spatial Autoregressive Models

- Unknown spatial dependence structrue
  - Spatial weight matrix should reflect the prior information corresponding to a specific research question, the applied work mainly follows the same routine of specifying this matrix, such as using contiguity, k nearest neighbors (KNN) or some other heuristic procedures, without considering the idiosyncracy for each case
  - Symmetric spatial weight matrix ?
  - Different spatial models imply different assumption on the range of interaction, but it is hard to distinguish them emprically
- Economic Interpretation
  - The omitted variables often are highly correlated over space. McMillen(2012) recognized that SAR is just a form of spatial smoothing and used as a panacea for model misspefication issue

# Monte Carlo Simulation

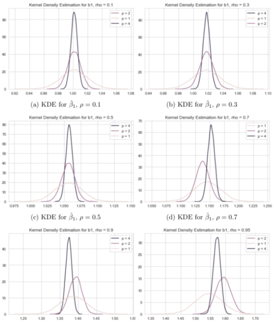
- What if choose an incorrect spatial model?
  - Specify spatial autoregressive parameters $\rho = (0.1, 0.3, 0.5, 0.7, 0.9, 0.95)$ and $\frac{\sigma_x}{\sigma_u} = (1, 2, 4)$
  - Conduct 10,000 Monte Carlo simulations for each case
    1. SLX-SAR, SAR-SLX
    2. SDM-SAR, SAR-SDM
    3. SDM-SLX, SLX-SDM
    4. SEM-SDM, SEM-SAR, SEM-SLX
  - Randomly generate $\mathbf{X}$ and $\mathbf{u}$, generate $\mathbf{y}$ using the reduced form specification for each model
  - Geometry: US counties (3085 counties, mainland excluding Alaska). Generate spatial weight matrix using Queen, Rook contiguity and Block by State
  - Test nulls

$$H_0 : \theta_i = \theta_{i0} \tag{8}$$

Compare the rejection probability with the true DGP

- Empirical distribution for $\hat{\beta}_0$, $\hat{\beta}_1$ and the probability of rejection for nulls

Figure A.6: Estimates for $\beta_1$ for SAR-SLX, Queen Contiguity

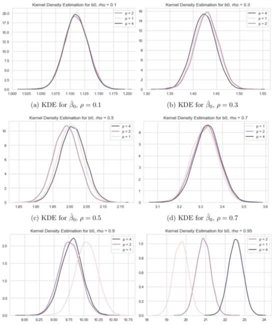


(a) KDE for $\hat{\beta}_1$, $\rho = 0.1$ (b) KDE for $\hat{\beta}_1$, $\rho = 0.3$

(c) KDE for $\hat{\beta}_1$, $\rho = 0.5$ (d) KDE for $\hat{\beta}_1$, $\rho = 0.7$

(e) KDE for $\hat{\beta}_1$, $\rho = 0.9$ (f) KDE for $\hat{\beta}_1$, $\rho = 0.95$

Figure A.5: Estimates for $\beta_0$ for SAR-SLX, Queen Contiguity



(a) KDE for $\hat{\beta}_0$, $\rho = 0.1$ (b) KDE for $\hat{\beta}_0$, $\rho = 0.3$

(c) KDE for $\hat{\beta}_0$, $\rho = 0.5$ (d) KDE for $\hat{\beta}_0$, $\rho = 0.7$

(e) KDE for $\hat{\beta}_0$, $\rho = 0.9$ (f) KDE for $\hat{\beta}_0$, $\rho = 0.95$

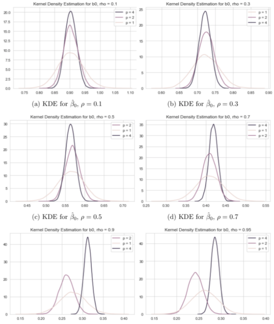
Table A.2: Parameter Estimates: SAR-SLX

| | | True SAR, Estimated , SLX | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\rho = 0.1$ | | | $\rho = 0.3$ | | | $\rho = 0.5$ | | |
| | | $\beta_0$ | $\beta_1$ | $\gamma$ | $\beta_0$ | $\beta_1$ | $\gamma$ | $\beta_0$ | $\beta_1$ | $\gamma$ |
| | $\sigma_x$ =1 | 1.000 | 0.052 | 0.689 | 1.000 | 0.157 | 1.000 | 1.000 | 0.793 | 1.000 |
| Queen | $\sigma_x$ =2 | 1.000 | 0.059 | 0.999 | 1.000 | 0.486 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x$ =4 | 1.000 | 0.066 | 1.000 | 1.000 | 0.970 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x$ =1 | 1.000 | 0.050 | 0.695 | 1.000 | 0.176 | 1.000 | 1.000 | 0.890 | 1.000 |
| Rook | $\sigma_x$ =2 | 1.000 | 0.054 | 0.999 | 1.000 | 0.437 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x$ =4 | 1.000 | 0.066 | 1.000 | 1.000 | 0.975 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x$ =1 | 1.000 | 0.051 | 0.126 | 1.000 | 0.049 | 0.795 | 1.000 | 0.072 | 0.998 |
| Block | $\sigma_x$ =2 | 1.000 | 0.050 | 0.450 | 1.000 | 0.056 | 0.970 | 1.000 | 0.135 | 1.000 |
| | $\sigma_x$ =4 | 1.000 | 0.051 | 0.751 | 1.000 | 0.070 | 1.000 | 1.000 | 0.421 | 1.000 |
| | | $\rho = 0.7$ | | | $\rho = 0.9$ | | | $\rho = 0.95$ | | |
| | | $\beta_0$ | $\beta_1$ | $\gamma$ | $\beta_0$ | $\beta_1$ | $\gamma$ | $\beta_0$ | $\beta_1$ | $\gamma$ |
| | $\sigma_x$ =1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Queen | $\sigma_x$ =2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x$ =4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x$ =1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Rook | $\sigma_x$ =2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x$ =4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x$ =1 | 1.000 | 0.251 | 1.000 | 1.000 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 |
| Block | $\sigma_x$ =2 | 1.000 | 0.766 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x$ =4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

# Simulation Result: SLX-SAR

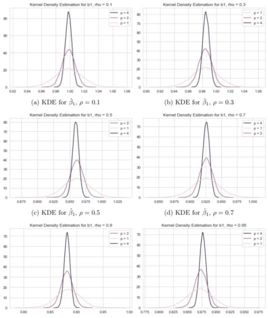
- Empirical distribution for $\hat{\beta}_0$, $\hat{\beta}_1$ and the probability of rejection for nulls

Figure A.1: Estimates for $\beta_0$ for SLX-SAR, Rook Contiguity




(a) KDE for $\hat{\beta}_0$, $\rho = 0.1$
(b) KDE for $\hat{\beta}_0$, $\rho = 0.3$
(c) KDE for $\hat{\beta}_0$, $\rho = 0.5$
(d) KDE for $\hat{\beta}_0$, $\rho = 0.7$
(e) KDE for $\hat{\beta}_0$, $\rho = 0.9$
(f) KDE for $\hat{\beta}_0$, $\rho = 0.95$

Figure A.2: Estimates for $\beta_1$ for SLX-SAR, Rook Contiguity




(a) KDE for $\hat{\beta}_1$, $\rho = 0.1$
(b) KDE for $\hat{\beta}_1$, $\rho = 0.3$
(c) KDE for $\hat{\beta}_1$, $\rho = 0.5$
(d) KDE for $\hat{\beta}_1$, $\rho = 0.7$
(e) KDE for $\hat{\beta}_1$, $\rho = 0.9$
(f) KDE for $\hat{\beta}_1$, $\rho = 0.95$

Table A.1: Parameter Estimates: SLX_SAR

| | | True SLX, Estimated , SAR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\gamma = 0.1$ | | | $\gamma = 0.3$ | | | $\gamma = 0.5$ | | |
| | | $\beta_0$ | $\beta_1$ | $\rho$ | $\beta_0$ | $\beta_1$ | $\rho$ | $\beta_0$ | $\beta_1$ | $\rho$ |
| | $\sigma_x = 1$ | 0.591 | 0.053 | 0.657 | 1.000 | 0.124 | 1.000 | 1.000 | 0.545 | 1.000 |
| Queen | $\sigma_x = 2$ | 0.960 | 0.055 | 0.998 | 1.000 | 0.312 | 1.000 | 1.000 | 0.964 | 1.000 |
| | $\sigma_x = 4$ | 0.998 | 0.070 | 1.000 | 1.000 | 0.863 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x = 1$ | 0.597 | 0.051 | 0.688 | 1.000 | 0.144 | 1.000 | 1.000 | 0.553 | 1.000 |
| Rook | $\sigma_x = 2$ | 0.966 | 0.051 | 0.997 | 1.000 | 0.403 | 1.000 | 1.000 | 0.982 | 1.000 |
| | $\sigma_x = 4$ | 0.998 | 0.067 | 1.000 | 1.000 | 0.875 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x = 1$ | 0.100 | 0.049 | 0.102 | 0.361 | 0.053 | 0.376 | 0.950 | 0.054 | 0.943 |
| Block | $\sigma_x = 2$ | 0.308 | 0.047 | 0.277 | 0.975 | 0.054 | 0.982 | 1.000 | 0.060 | 1.000 |
| | $\sigma_x = 4$ | 0.633 | 0.049 | 0.708 | 1.000 | 0.058 | 1.000 | 1.000 | 0.096 | 1.000 |
| | | $\gamma = 0.7$ | | | $\gamma = 0.9$ | | | $\gamma = 0.95$ | | |
| | | $\beta_0$ | $\beta_1$ | $\rho$ | $\beta_0$ | $\beta_1$ | $\rho$ | $\beta_0$ | $\beta_1$ | $\rho$ |
| | $\sigma_x = 1$ | 1.000 | 0.950 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Queen | $\sigma_x = 2$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x = 4$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x = 1$ | 1.000 | 0.958 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| Rook | $\sigma_x = 2$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x = 4$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_x = 1$ | 0.986 | 0.062 | 0.990 | 1.000 | 0.066 | 1.000 | 0.999 | 0.076 | 0.998 |
| Block | $\sigma_x = 2$ | 1.000 | 0.092 | 1.000 | 1.000 | 0.119 | 1.000 | 1.000 | 0.126 | 1.000 |
| | $\sigma_x = 4$ | 1.000 | 0.193 | 1.000 | 1.000 | 0.337 | 1.000 | 1.000 | 0.378 | 1.000 |

# Conclusion

- Concluding Remark
  - Models do not deviate severely when the spatial dependence is weak, however, a misspecified model can severely over-reject thenull hypothesis when the dependence is sufficiently strong
  - Parameter estimates for the slope coefficient $\beta_1$ is in general more robust to model misspecification than $\hat{\beta}_0$.
  - The SLX models are unable to fully capture the global effectsin most spatial autoregressive framework, i.e. SAR and SDM. However, this distortion mightbe mitigated if the spatial weight matrix reflects the global feature.