Spatial econometrics (3)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM

tinbergen institute

1. Introduction

- **Topics:**
  1. Discrete choice
     - Random utility framework, estimating binary and multinomial regression models
  2. **Spatial econometrics**
     - **Spatial data, autocorrelation, spatial regressions**
  3. **Identification**
     - **Research design, IV, OLS, RDD, quasi-experiments, standard errors**
  4. **Hedonic pricing**
     - **Theory and estimation**
  5. Quantitative spatial economics
     - General equilibrium models in spatial economics

1. Introduction

Wednesday

09:30-10:30	Lecture 1	Discrete Choice I (The random utility framework)
10:45-11:45	Lecture 2	Discrete Choice II (Estimating discrete choice models)
12:00-13:00	Lecture 3	Spatial Econometrics I (Spatial data)
14:00-15:30	Tutorial 1	Assignment 1

Thursday

09:30-10:30	Lecture 4	Spatial Econometrics II (Spatial autocorrelation)
10:45-11:45	Lecture 5	Spatial Econometrics III (Spatial regressions)
12:00-12:30	Lecture 6	Identification I (Research design)
13:30-14:00	Tutorial 2	Discussion of Assignment 1
14:00-15:00	Tutorial 3	Assignment 2

Friday

09:30-10:00	Lecture 7	Identification II (RCTs, OLS, IV, quasi-experiments)
10:00-10:30	Lecture 8	Hedonic pricing I (Theory)
10:45-11:45	Lecture 9	Hedonic pricing II (Estimation)
12:00-12:30	Tutorial 4	Discussion of Assignment 2

2. Spatial regressions (continued)

- **Spatial lag model**
  - $\boldsymbol{y} = \rho \boldsymbol{W y} + \boldsymbol{X \beta} + \boldsymbol{\mu}$ (3)
  - $\rho \neq 0, \boldsymbol{\gamma} = 0, \lambda = 0$
  - **Spatial dependence in dependent variables**

- **Spatial cross-regressive model**
  - $\boldsymbol{y} = \mathbf{X \beta} + \boldsymbol{\gamma W X} + \boldsymbol{\mu}$ (5)

- **Spatial error model**
  - $\boldsymbol{y} = \mathbf{X \beta} + \boldsymbol{\epsilon}, \text{ with } \boldsymbol{\epsilon} = \lambda \mathbf{W \epsilon} + \boldsymbol{\mu}$ (6)

2. Spatial regressions (continued)

- Three issues are on the table
  1. When should you use these models?
  2. Which of the models should you choose?
  3. Can we combine these different spatial models?

1. When should you use these models?

- Test for spatial effects
  - H_0: No spatial dependence

- Estimate standard OLS, $y = X\beta + \epsilon$
  - Calculate Moran's I using $\hat{\epsilon}$
  - $I = \frac{R}{S_0} \times \frac{\hat{\epsilon}' W \hat{\epsilon}}{\hat{\epsilon}' \hat{\epsilon}}$

- Moran's I does have a rather uninformative alternative hypothesis
  - H_A: Spatial dependence…

2. Spatial regressions (continued)

1. When should you use these models?

- Lagrange multiplier tests provide more information
  - LM_ρ - test for presence of spatial lag
  - LM_λ - test for presence of spatial error

1. When should you use these models?

- **Test for spatial lag**
  1. **Run OLS**
  2. **Run LM_ρ-test**

$\mathbf{H_0}$: $\rho = 0$

$\mathbf{H_A}$: $\rho \neq 0$

$$LM_\rho = \frac{\mathbf{1}}{\mathbf{nJ}}\left(\frac{\epsilon' W y}{s^2}\right)^2 \sim \chi_1^2 \qquad (9)$$

with $\boldsymbol{J} = [(\boldsymbol{WX\beta})'\boldsymbol{M}(\boldsymbol{WX\beta}) + \boldsymbol{T}s^2]/ns^2$ and

$\mathbf{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

1. When should you use these models?

- **Test for spatial error**
  1. **Run OLS**
  2. **Run LM_λ-test**

$\mathbf{H_0}: \lambda = 0$

$\mathbf{H_A}: \lambda \neq 0$

$$LM_\lambda = \frac{\mathbf{1}}{\mathbf{T}}\left(\frac{\boldsymbol{\epsilon}' \boldsymbol{W} \boldsymbol{\epsilon}}{\boldsymbol{s}^2}\right)^2 \sim \chi_1^2 \qquad (8)$$

$$\textbf{with } \mathbf{T} = \mathbf{tr}\left((\mathbf{W}' + \mathbf{W})\boldsymbol{W}\right) \textbf{ and } \boldsymbol{s} = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}\ /n$$

2. Spatial regressions (continued)

1. When should you use these models?

- However,
  - Spatial errors and lags may be correlated
  - May also be both present

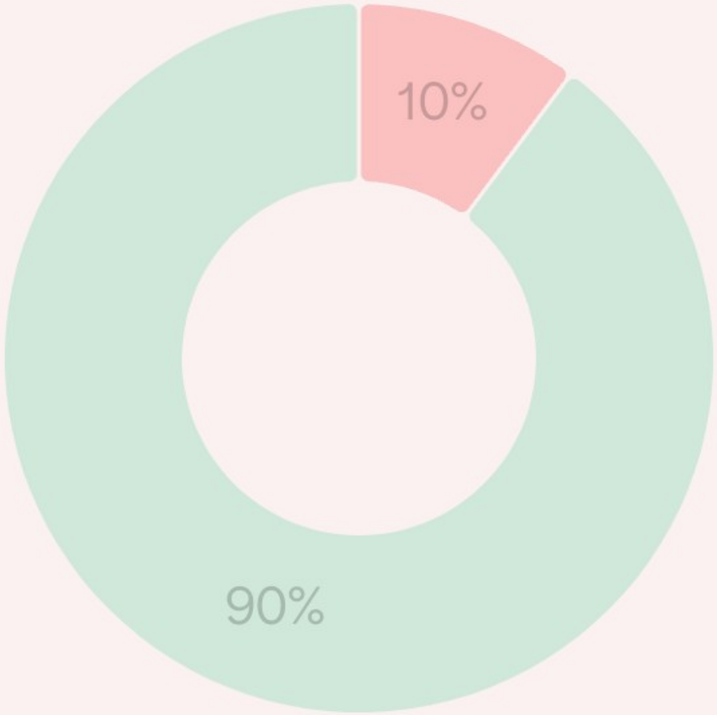
- Use ***robust*** LM tests
  - LM_ρ^* adds correction factor for potential spatial error
  - LM_λ^* adds correction factor for potential spatial lag
  - Complex formulae!

2. Which of the models should you choose?

- Estimate *robust* LM tests using software

- Common practice
  - Choose and estimate the model for which the statistic is the most significant

Why may we not discuss a test for the importance of spatial cross-regressive model?



0%	Because y is irrelevant when you are interested in spatial effects	×
10%	Because the test statistic is simply too complex to show here	×
90%	Because you can simply look at whether y is/are statistically significant	×

2. Which of the models should you choose?

- Estimate *robust* LM tests using software

- Common practice
  - Choose and estimate the model for which the statistic is the most significant

- When LM_λ^* and LM_ρ^* are statistically insignificant we may use OLS
- Robust LM-tests are typically provided in STATA output

3. Can we combine these spatial models?

- In practice, both a spatial lag and spatial error may be present

- How to estimate?
  - Use Kelejian and Prucha's GS2SLS method
  - Three-stage procedure!

3. Can we combine these spatial models?

- **Complicated stuff!**

- **Let software do the difficult calculations!**
  - SPAUTOREG **in STATA**
  - SPIVREG **in STATA**

3. Mostly pointless spatial econometrics?

- **Gibbons and Overman (2012)**
  - ***"Mostly pointless spatial econometrics?"***

- **We are interested to identify causal impacts β:**
  $y = X\beta + \mu$

- **Typical features of spatial data**
  - **Unobserved variables correlated with X**
  - **Omitted variable bias!**
  - **Large datasets**

3. Mostly pointless spatial econometrics?

- **Tempting to 'fix' omitted variable bias by including a spatial lag**

- **Let's consider again:**

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

- **Reduced-form:**

$$\boldsymbol{y} = \rho \boldsymbol{W}(\rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}) + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

$$\boldsymbol{y} = \rho \boldsymbol{W}(\rho \boldsymbol{W}(\rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}) + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}) + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

…

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\rho + \boldsymbol{W}^2\boldsymbol{X}\rho^2 + \boldsymbol{W}^3\boldsymbol{X}\rho^3 + [\dots] + \tilde{\boldsymbol{\mu}}$$

… The last equation suggests that in the end y is just a non-linear function of the X-variables

3. Mostly pointless spatial econometrics?

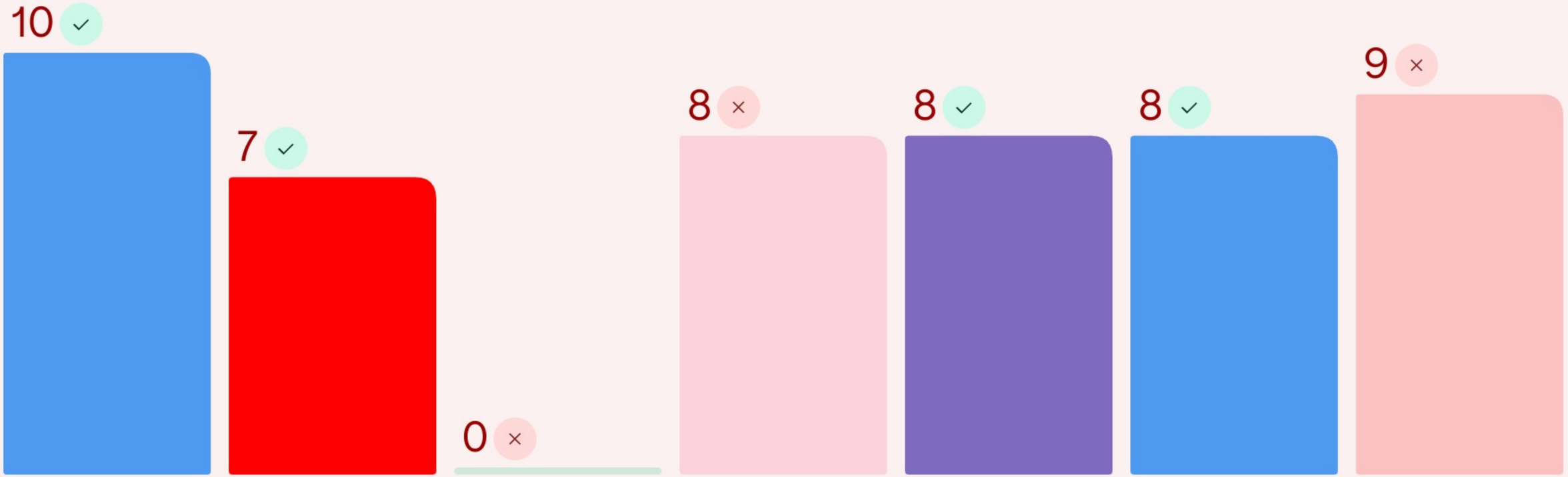
- **Reduced-form of spatial lag model ≈ spatial cross-regressive model**
  - **It is hard to prove that the spatial lag model is the 'right' model**
  - **So, it is hard to distinguish empirically between the two types of models**
  - **Only when there is a structural (network) model, a spatial lag may be appropriate**

3. Mostly pointless spatial econometrics?

- **The spatial lag model *does not* solve the problem of omitted variable bias!**
  - Think of real exogenous sources of variation in X to identify β
  - Use instruments or quasi-experiments
  - More discussion on identification strategies in last week!

- **Estimate spatial error model?**
  - Spatial datasets are typically large
  - Efficiency issues are *usually* not so important

When would you use spatial econometric techniques *(multiple answers can be correct)*?



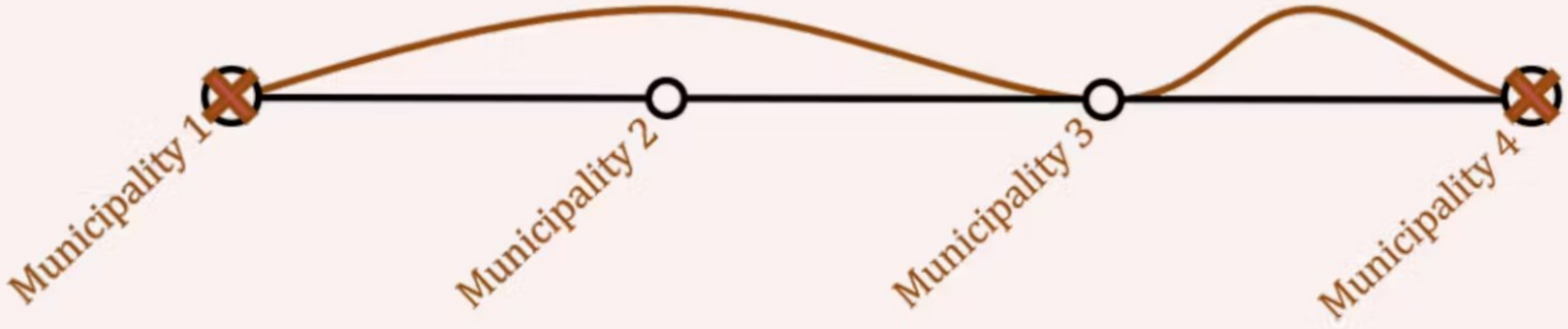
3. Mostly pointless spatial econometrics?

- **Why then use spatial econometrics?**
  1. **Exploratory tool to investigate spatial autocorrelation**
  2. **Test for spatial dependence and heterogeneity, also in quasi-experiments and when using instruments**
  3. **Investigate whether results are robust to spatial autocorrelation (using different W)**
  4. **Spatial cross-regressive models are often useful and straightforward to interpret**

3. Mostly pointless spatial econometrics?

Koster, Tabuchi & Thisse (2022, *JoEG*)

- Modern economies invest a sizable amount of money into high-speed rail
- We study the impact of high-speed rail stations on 'intermediate' places
- Local policy makers lobby for the opening of a station, but is this a good idea?



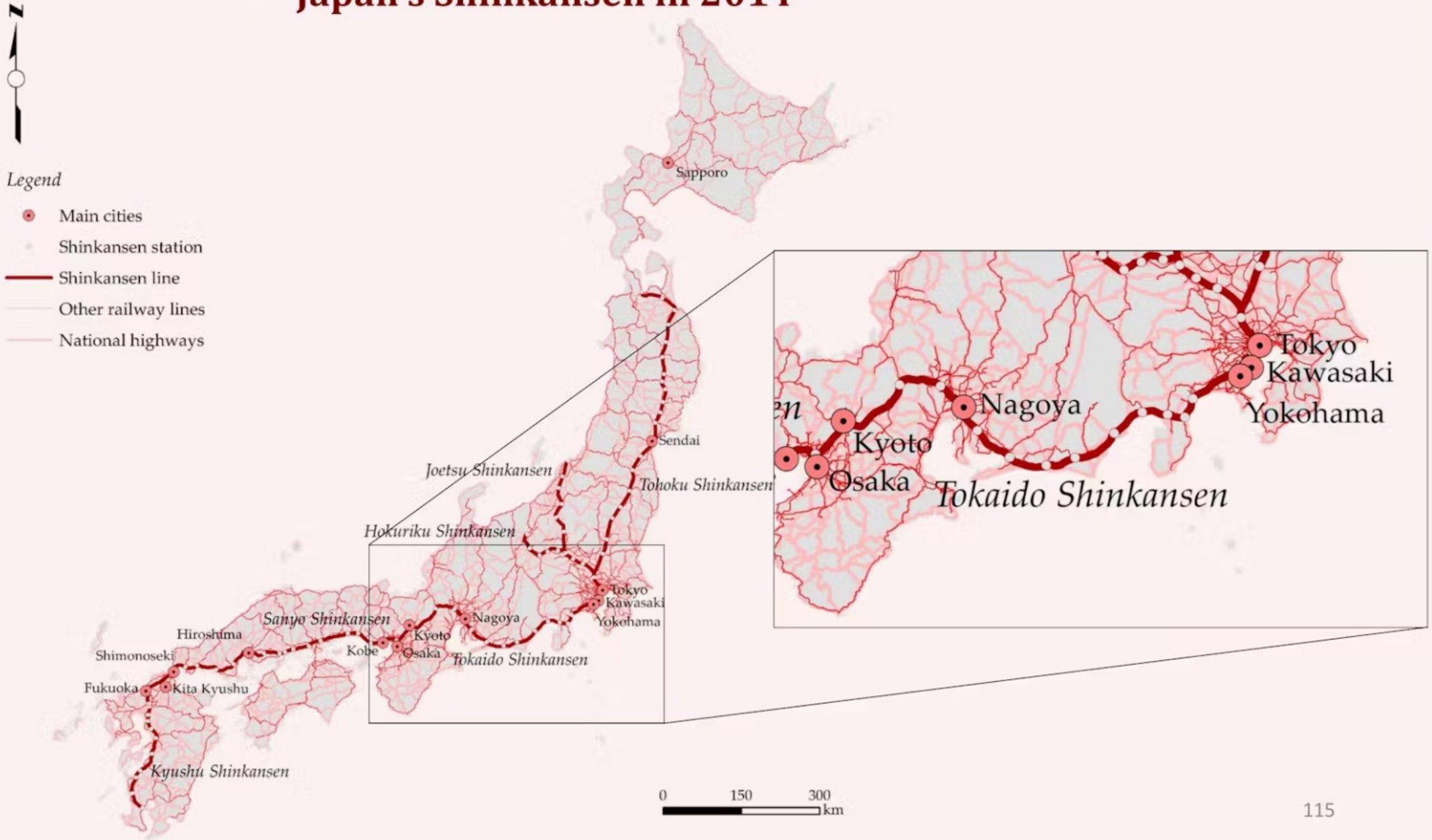
3. Mostly pointless spatial econometrics?

Three potential effects on employment in intermediate places:

- \+ **A better connection reduces the need for firms to locate near large markets with high demand for goods and services**
- – **A better connection to local markets reduces the need to locate near local markets**
- – **When firms start to concentrate in local markets, competition becomes tougher**

3. Mostly pointless spatial econometrics?

- **Japan's Shinkansen in 2014**



3. Mostly pointless spatial econometrics?

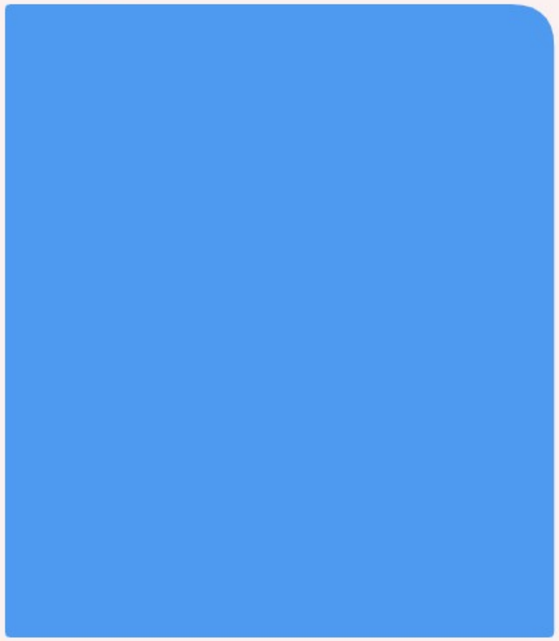
- We estimate the following regression

$$\Delta \log e = \alpha + \beta s + \boldsymbol{X}\gamma + \epsilon$$

where s captures a dummy whether a municipality has a station

With $\Delta \log e = \alpha + \beta s + \mathbf{X}\gamma + \epsilon$, what does β capture?

9 ✓



The percentage change (approximately) in employment density due to the opening of the Shinkansen station

0 ×

The absolute change in employment density due to the opening of the Shinkansen station

2 ×

The percentage change (approximately) in employment density due to the opening of the Shinkansen station in nearby municipalities

0 ×

The absolute change in employment density due to the opening of the Shinkansen station in nearby municipalities

3. Mostly pointless spatial econometrics?

- We estimate the following regression

$$\Delta \log e = \alpha + \beta s + \boldsymbol{X}\gamma + \epsilon$$

where s captures a dummy whether a municipality has a Shinkansen station

- (Why) should we apply spatial econometric methods here?

3. Mostly pointless spatial econometrics?

- **We therefore extend the baseline equation**

$$\Delta \log e = \alpha + \beta_0 s + \beta_1 \boldsymbol{W} s + \boldsymbol{X}\gamma + \epsilon$$

where $\epsilon = \lambda \boldsymbol{W}\epsilon + \mu$ **and** $\boldsymbol{W}$ **is a row-standardised inverse-distance weight matrix**

With $\Delta \log e = \alpha + \beta_0 \boldsymbol{s} + \beta_1 \mathbf{W}\boldsymbol{s} + \mathbf{X}\gamma + \epsilon$, what does β_1 capture?

0 ×

The %-change in employment density due to an increase in the spatially weighted employment density in nearby municipalities

0 ✓

The %-change in employment density due to an increase in the spatially weighted number of Shinkansen stations in nearby municipalities

0 ×

The %-change in employment density due to an increase in the spatially weighted independent variables in nearby municipalities

0 ×

The %-change in employment density due to an increase in the spatially weighted residuals in nearby municipalities

0 ×

The %-change in Shinkansen stations as a result of an increase in the spatially weighted employment density

0 ×

The %-change in Shinkansen stations as a result of an increase in the spatially weighted independent variables

3. Mostly pointless spatial econometrics?

▪ Results

Table 5.1: The opening of a Shinkansen station

(Dependent variable: the log of the change in the employment density between 1957 and 2014)

	(1) OLS	(2) OLS	(3) GS2SLS	(4) GS2SLS	(5) GS2SLS
	Baseline OLS	Spatial cross-regressive model	Spatial error model	Spatial lag model	All spatial effects
Shinkansen station in 2014	-0.2796**	-0.2814**	-0.2034*	-0.2167*	-0.2182*
	(0.1218)	(0.1198)	(0.1233)	(0.1246)	(0.1239)
Spatial effects:					
W· Shinkansen station in 2014		-11.1404***			-2.6923
		(2.8048)			(3.1049)
$\mathbf{W}\cdot\epsilon$			2.0174***		0.3840
			(0.3265)		(0.5581)
$\mathbf{W}\cdot\log\Delta e$				1.2501***	1.2290***
				(0.1878)	(0.2483)
Region fixed effects (8)	Yes	Yes	Yes	Yes	Yes
Number of observations	1,412	1,412	1,412	1,412	1,412
R^2	0.206	0.211			
Pseudo-R^2			0.202	0.225	0.226

Notes: **W** is a row-standardized inverse distance-weight matrix. We exclude municipalities that are centres of metropolitan or micropolitan areas. Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

3. Mostly pointless spatial econometrics?

Koster, Tabuchi & Thisse (2022, *JoEG*)

- The impact of a Shinkansen station reduces employment density by ≈20-25%
  - Hence, a Shinkansen station *does not benefit* intermediate places

- Spatial cross-regressive model
  - A standard deviation increase in WS, employment density decreases by 6.8%
  - WS = the spatially weighted number of Shinkansen stations in nearby municipalities

- Spatial error and lag effects are relevant — unrealistically high spatial parameters
  - More importantly, the main effect is hardly influenced by the inclusion of spatial effects

4. Summary

Spatial econometrics:

- Spatial data:
  - No natural origin, reciprocity, multidirectional
  - Define spatial relationships by the spatial weight matrix

- Spatial regressions
  - Spatial lag model
  - Spatial cross-regressive model
  - Spatial error model
  - … Combine using advanced methods

- Spatial econometrics are a useful tool, *but* not a way to identify causal effects

Spatial econometrics (3)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM

tinbergen institute