Misspecification Issues of Spatial Econometric Models: Application to Regional Unemployment in Russia and Germany

Elena Semerikova, Olga Demidova

National Research University Higher School of Economics

24-26 September, Nizhny Novgorod, 2015

Table of Contents

Introduction

Motivation

Research plan

Data& Variables

Models estimation on real data

Analysis on simulated data

Conclusions

Introduction

Motivation

Research plan

Data& Variables

Models estimation on real data

Analysis on simulated data

Conclusions

Spatial Matrix

Spatial Matrix - formal specification of spatial effects in econometric models (Anselin 1988).
Spatial Autoregressive Model:

$$Y = \rho WY + X\beta + \epsilon$$

Spatial Matrices

Boundary matrix

$$w_{ik} = \begin{cases} 1, \text{if } i \text{ and } k \text{ share a border} \\ 0, \text{if } i \text{ and } k \text{ do not have a common border} \end{cases}.$$

Matrix of inverse distances

$$W = \begin{pmatrix} 0 & \frac{1}{d_{12}} & \cdots & \frac{1}{d_{1N}} \\ \frac{1}{d_{21}} & 0 & \cdots & \frac{1}{d_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{d_{N1}} & \frac{1}{d_{N2}} & \cdots & 0 \end{pmatrix}.$$

Matrices are row normalized

The influence of the spatial matrix specification to the estimates

- Arbia (2008) criticizes spatial models for their sensitivity to the choice of spatial weighting matrices
- LeSage and Pace (2010): sensitivity to the choice of spatial matrix is the myth of spatial modelling

Bell and Bockstael (2000):

- the choice of the weighting matrix affects the estimates more than estimation method

Stakhovych and Bijmolt (2009):

- The estimates of spatial coefficient ρ and βs are biased
- The higher the ρ, the higher the bias for ρ and βs.

Coefficient estimator

$Y = \rho WY + X\beta + \epsilon$ - Spatial Autoregressive Model

Maximum Likelihood Estimator: $\hat{\beta} = (X'X)^{-1}X'(I - \rho W)Y$

Bias when choosing wrong W

$Y = \rho WY + X\beta + \epsilon_1$ - data generating model
$Y = \tilde{\rho}\widetilde{W}Y + X\tilde{\beta} + \epsilon_2$ - estimated model

$$E(\hat{\beta}) = \beta + \underbrace{(X'X)^{-1}X'(I - \tilde{\rho}\widetilde{W})[(I - \rho W)^{-1} - (I - \tilde{\rho}\tilde{W})^{-1}]X\beta}_{\text{Bias}}$$

Bias when we do not account for W

$Y = \rho WY + X\beta + \epsilon$ - data generating model
$Y = X\beta + \epsilon$ - estimated model

$$E(\hat{\beta}) = \beta + \underbrace{\rho(X'X)^{-1}X'W(I - \rho W)^{-1}X\beta}_{\text{Bias}}$$

Introduction

Motivation

Research plan

Data& Variables

Models estimation on real data

Analysis on simulated data

Conclusions

Motivation

How sensitive are the parameter estimates to the type of weighting matrix chosen for the estimation?

Which matrix to choose?

Introduction

Motivation

Research plan

Data& Variables

Models estimation on real data

Analysis on simulated data

Conclusions

Research plan

1. Estimate spatial models with different types of weight matrices on real data.
2. Decide whether the estimates are different.
3. Estimate models with different types of weight matrices on simulated data.
4. Asses the severity of the bias.

Introduction

Motivation

Research plan

Data& Variables

Models estimation on real data

Analysis on simulated data

Conclusions

Regional Unemployment
Panel data

- 370 counties, Germany (NUTS III)
- 75 regions, Russia
- From 2005 till 2010

Variables for Germany

Demographic Characteristics:

- share of young people (from 15 to 25) (young) and share of people of ages from 55 to 65 (old) in the population of working ages (from 15 to 65)
- Share of employed people with only school education (edus)
- Share of employed people with university education (eduu)
- Gross Regional Product per capita (gdppc)

Sectoral structure:

- Share of employed people in different sectors (agricultural, industrial, manufacturing, construction, trade, financial, sevices)

Variables for Russia

- Gross Regional Product per capita (gdp base pur)
- Share of old people in the population of working ages (old work)
- Share of young people in the population of working ages (young work)
- Share of people with higher education in the labor force

Sectoral structure:

- Share of sectoral production in gross value added (raw materials, manufacturing, government and nongovernment services)

Introduction

Motivation

Research plan

Data& Variables

Models estimation on real data

Analysis on simulated data

Conclusions

Spatial Autoregressive Model (SAR)

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + X_{it}\beta + \mu_i + \sum_{k=7}^{10} \gamma_k d_{200k} + \epsilon_{it},$$

where ρ - spatial coefficient,
X_{it} - matrix of regressors,
μ_i - individual effects,
d_{200k} - dummies for each year, errors $\epsilon_{it} \sim iid(0, \sigma^2)$.

Spatial Durbin Model (SDM)

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + X_{it}\beta + \sum_{j=1}^{N} w_{ij} X_{jt}\theta + \mu_i + \sum_{k=7}^{10} \gamma_k d_{200k} + \epsilon_{it},$$

where ρ - spatial coefficient,
X_{it} - matrix of regressors,
μ_i - individual effects,
d_{200k} - dummies for each year, errors $\epsilon_{it} \sim iid(0, \sigma^2)$.

Spatial Error Model (SEM)

$$y_{it} = X_{it}\beta + \mu_i + \sum_{k=7}^{10} \gamma_k d_{200k} + v_{it}, \qquad v_{it} = \lambda \sum_{j=1}^{N} w_{ij} v_{jt} + \epsilon_{it}$$

where ρ - spatial coefficient,
X_{it} - matrix of regressors,
μ_i - individual effects,
d_{200k} - dummies for each year, errors $\epsilon_{it} \sim iid(0, \sigma^2)$.

Spatial Autoregressive with Spatially Autocorrelated Errors Model (SAC)

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + X_{it}\beta + \mu_i + \sum_{k=7}^{10} \gamma_k d_{200k} + v_{it}, \quad v_{it} = \lambda \sum_{j=1}^{N} w_{ij} v_{jt} + \epsilon_{it}$$

where ρ - spatial coefficient,
X_{it} - matrix of regressors,
μ_i - individual effects,
d_{200k} - dummies for each year, errors $\epsilon_{it} \sim iid(0, \sigma^2)$.

Results of the real data estimation

- The choice of the weighting matrix influences the coefficient estimates and its' standard errors
- Russia - almost no difference
- Germany - obvious differences

Introduction

Motivation

Research plan

Data& Variables

Models estimation on real data

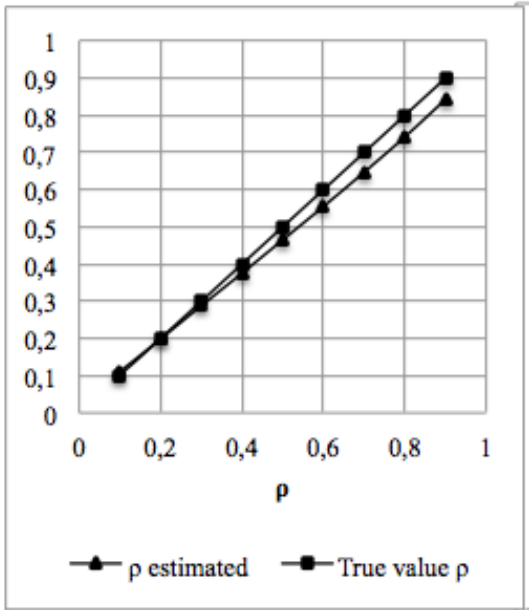
Analysis on simulated data

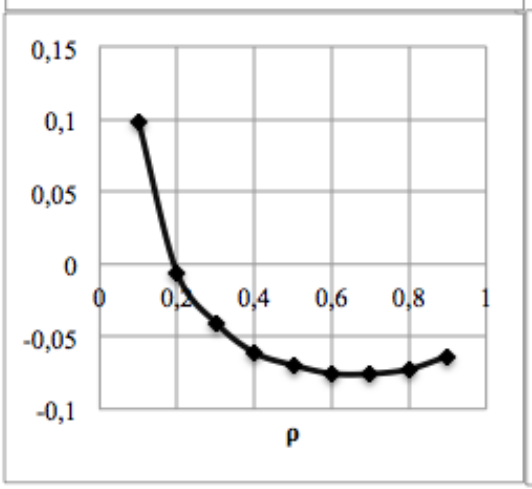
Conclusions

Simulation Design

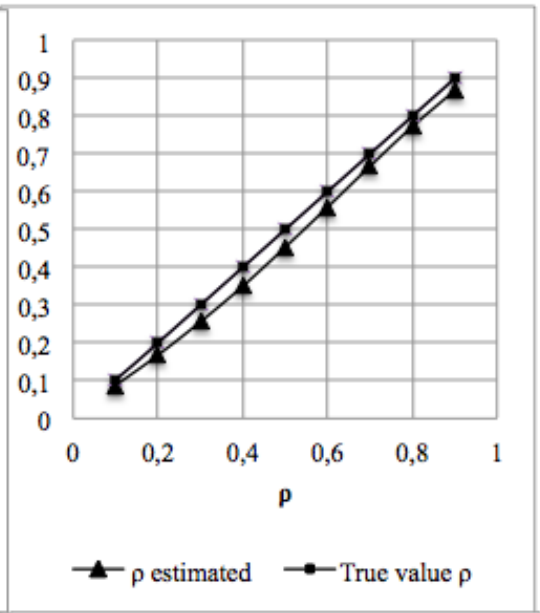
- Real values of Xs
- Real matrices W (boundary and inverse distances) for 75 Russian regions
- $\rho \in [0.1; 0.9]$
- 10000 replications
- Model generating Y: SAR model
- $Y_{it} = (I - \rho W)^{-1}(X\beta)_{it} + (I - \rho W)^{-1}\mu_i + (I - \rho W)^{-1}\tau_t + (I - \rho W)^{-1}\epsilon_{it}$

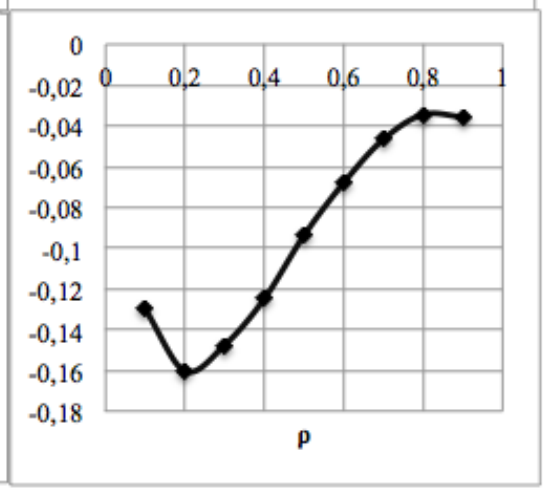
Simulation - Boundary matrix
Estimation - Inverse distance





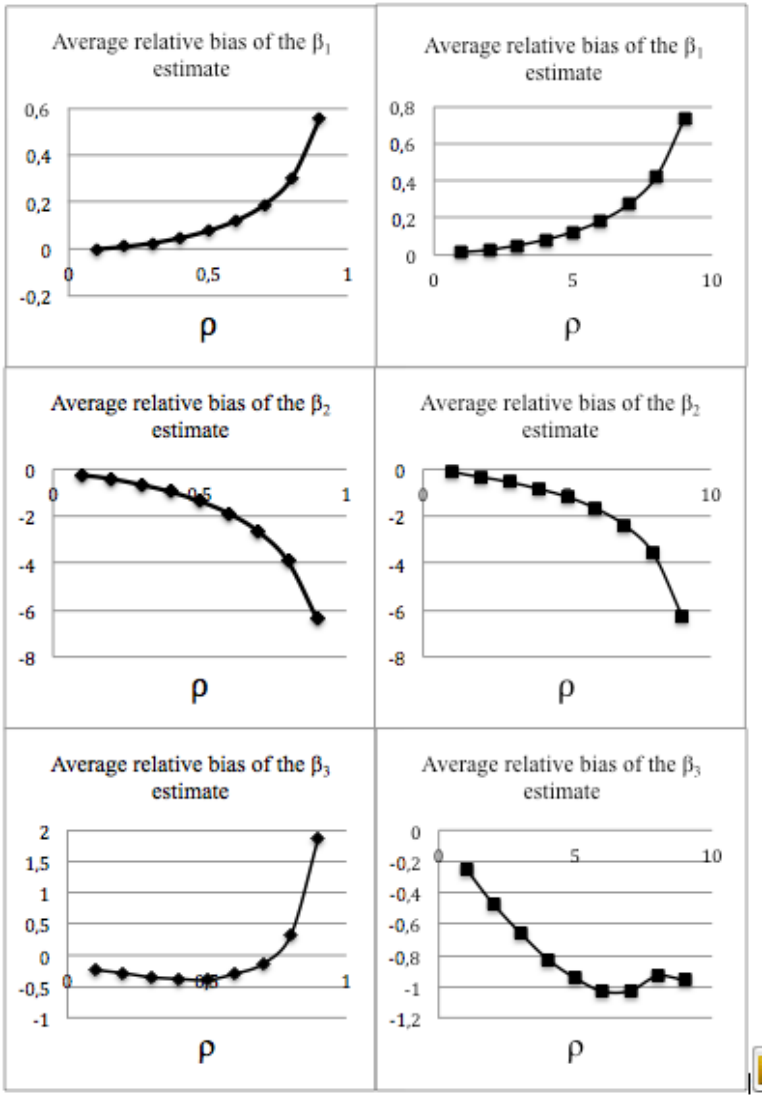
Simulation - Inverse distance matrix
Estimation - Boundary matrix





Boundary matrix for generating data

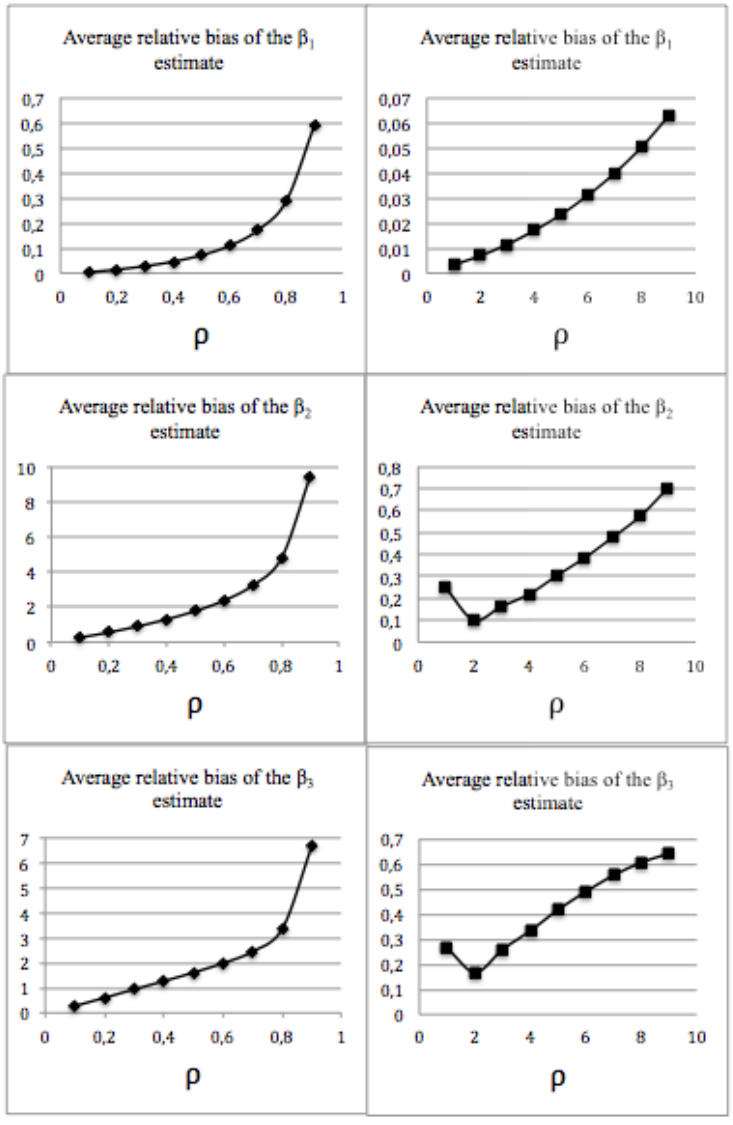
Inverse distance for testing data



No W matrix for testing data

Inverse distance matrix for generating data

Boundary matrix for testing data



No W matrix for testing data

Mean Squared Error of Spatial Parameter

Table : MSE of $\hat{\rho}$

		W for testing data	
		W boundary	W of inverse distances
W for generating data	W boundary	0.0077	0.0253
	W of inverse distances	0.0215	0.0220
	Total	0.0146	0. 0237

Mean Squared Error of Regression Parameters

Table : MSE of $\hat{\beta}_1$

		W for testing data	
		W boundary	W of inverse distances
W for generating data	W boundary	0.0075	0.0077
	W of inverse distances	0.0082	0.0001
	Total	0.0079	0. 0040

Table : MSE of $\hat{\beta}_2$

		W for testing data	
		W boundary	W of inverse distances
W for generating data	W boundary	0.2770	0.9938
	W of inverse distances	1.8621	0.0298
	Total	1.0695	0. 5118

Table : MSE of $\hat{\beta}_3$

		W for testing data	
		W boundary	W of inverse distances
W for generating data	W boundary	35.23	28.9908
	W of inverse distances	151.2275	17.5
	Total	93.2321	23.2453

Introduction

Motivation

Research plan

Data& Variables

Models estimation on real data

Analysis on simulated data

Conclusions

Conclusions

- Non-zero bias in coefficient estimates when choosing wrong matrix and when not accounting for spatial structure
- The choice of the matrix (as well as accounting for spatial structure) is more important when we deal with smaller regions
- The higher the ρ, the bigger the bias.
- Inverse distance matrix performs better for the analysis of regional unemployment in Russia