Urban vs. rural land use in large and small cities

Gianni Guastella Stefano Pareglio Paolo Sckokai
March 31, 2016

Università Cattolica

Table of contents

Introduction

Motivation

Land take opponents argue that the sealing of soil to the detriment of agriculture and the environment is excessive and often unnecessary.

Findings about large cities linked urban spatial expansion to the socio-economic dynamics, implicitly rejecting this hypothesis.

We extend this framework of analysis to small cities, where more space is often available, and negative environmental externalities are less internalized.

Background

The question: to what extent is the spreading of cities driven by markets?

Individual benefits vs. collective loss

Negative sentiment about urban sprawl

- soil sealing as a consequence
- noise and pollution generated from transport
- threatens the ecological equilibrium
- and the potential for rural development through farmland loss and price increase

Urban economists traditionally argued in favour of a rational justification (increasing income and population, declining transportation costs)

The Mills-Muth model of urban spatial structure

- households maximixe utility (housing, consumption)
- transportation costs (distance from CBD)
- the equilibrium urban fringe maximizes the households' utility

The optimal city size depends on *population*, *income*, *transport costs*, *farmland value*. These variables explain ∼80% of variance of city size (based on U.S. data)

Background

Accordingly

- land take is not unnecessary, evidence suggests
- land use restrictions could narrow people's utility by limiting housing supply

However

- the housing market does not fully internalize the externalities
  - landscape is a public good
  - congestion costs
- there is evidence of expanding cities with declining population
- cities become larger and their density lower, especially in the peripheries
- people value the high fragmentation of residential land use

In this paper

- we contend that the empirical approach proved robust since applied to large metropolitan areas
- we argue that urban sprawl (increasing land consumption and declining urbanization density) is most important in medium size and small cities
- we estimate the standard model allowing scale-heterogeneity and spatial relations

In small cities

- more available space, less institutional attention on efficient allocation
- the countryside is more easily accessible and the landscape less valued
- low congestion, increase in commuting not perceived as a problem

Empirical Approach

The equilibrium size of a city (u) depends

- positively on the total population (p) (HD and the price shift up, people move to the fringe, the fringe expands)
- positively on the average income (i) (people demand larger houses)
- negatively on the average farmland value (r) (less convenient to live far from CBD)
- negatively on the cost of transportation (t) (discourages commuting and lowers the disposable income for housing)

$$\mathbf{u} = \beta_0 + \beta_1 \mathbf{p} + \beta_2 \mathbf{i} + \beta_3 \mathbf{t} + \beta_4 \mathbf{r} + \mathbf{e}$$

Specification issues

The hypothesis of uncorrelated residuals is violated by the spatial structure of the data

- price contagion and cross-border transmission
- in modern cities the trade-off between commuting and house-prices extends beyond the city boarder (cross-city commuting)

The hypothesis of structural stability of coefficients might provide a poor representation

Spatial Econometrics

Spatial Econometric Models

Let

$$\mathsf{X} = [\mathsf{p}, \mathsf{i}, \mathsf{t}, \mathsf{r}]$$

be the matrix of the model covariates and

$$\beta^T = [\beta_1, \beta_2, \beta_3, \beta_4]$$

rewrite the model in equation 1 in compact form:

$$\mathsf{u} = \beta_0 + \mathsf{X}\beta + \mathsf{e}$$

To include information on neighboring units we make use of the operator W

Contiguity (spatial weight) matrix W

for $i = j =$ the number of spatial units in the sample

$$W_{i,j} = \begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,2} & \cdots & w_{m,n} \end{pmatrix}$$

$w_{i,j} = 1$ if units i and j are neighbors

$w_{i,j} = 0$ otherwise

$w_{i,i} = 0$ is a standard practice to exclude self-contiguity

Contiguity (spatial weight) matrix W

Distance weights (closest neighbors weight more)

$$w_{ij} = d_{i,j}^{-1}$$

Higher power distance

$$w_{ij} = d_{i,j}^{-k}$$

Gravity weight

$$w_{ij} = M_i \cdot M_j \cdot d_{i,j}^{-1}$$

Contiguity (spatial weight) matrix W

How to compute neighbors (using geographical coordinates)

common boundary two units share an administrative boundary

k-nearest all the k-nearest units are neighbors

cut-off all the units with a critical distance d are neighbors

W is usually row-standardized: Wx is the average value of x in the neighbors

NOTE

a simple shapefile of the spatial units is required

Having defined W and the linear model $\mathbf{u} = \beta_0 + \mathbf{X}\beta + \mathbf{e}$

the most generic spatial specification is the Manski model

$$\begin{aligned} \mathbf{u} &= \rho \mathbf{W}\mathbf{u} + \beta_0 + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\theta + \varepsilon \\ \varepsilon &= \lambda \mathbf{W}\varepsilon + \mathbf{e} \end{aligned}$$

Spatial effect may enter (one set to zero to allow identification)

- as a spatial autoregressive term ($\rho \mathbf{W}\mathbf{u}$) to reflect the impact of the the process in neighbors
- exogenously to reflect the consequence for each unit of the change in exogenous variables ($\mathbf{W}\mathbf{X}\theta$)
- with a spatial autocorrelation structure ($\lambda \mathbf{W}\varepsilon$) due to unobserved environmental effects common to neighboring units.

Common specifications

Spatial Auto-Regressive model (SAR)

$$\mathbf{u} = \rho \mathbf{W}\mathbf{u} + \beta_0 + \mathbf{X}\beta$$

Spatial Error Model (SEM)

$$\begin{aligned} \mathbf{u} &= \beta_0 + \mathbf{X}\beta + \varepsilon \\ \varepsilon &= \lambda \mathbf{W}\varepsilon + \mathbf{e} \end{aligned}$$

Spatial Durbin Model (SDM)

$$\mathbf{u} = \rho \mathbf{W}\mathbf{u} + \beta_0 + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\theta + \varepsilon$$

Model selection

SDM

$$\mathbf{u} = \rho \mathbf{W}\mathbf{u} + \beta_0 + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\theta + \varepsilon$$

if $\theta = 0$: SDM = SAR

$$\mathbf{u} = \rho \mathbf{W}\mathbf{u} + \beta_0 + \mathbf{X}\beta + \varepsilon$$

if $\theta = -\lambda\beta$: SDM = SER

$$\begin{aligned} \mathbf{u} &= \beta_0 + \mathbf{X}\beta + \varepsilon \\ \varepsilon &= \lambda \mathbf{W}\varepsilon + \mathbf{e} \end{aligned}$$

$$\mathbf{u} = \beta_0 + \mathbf{X}\beta + (1 - \lambda \mathbf{W})^{-1}\mathbf{e}$$

$$(1 - \lambda \mathbf{W})\mathbf{u} = (1 - \lambda \mathbf{W})\beta_0 + (1 - \lambda \mathbf{W})\mathbf{X}\beta + \mathbf{e}$$

$$\mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \beta_0 - \lambda \mathbf{W}\beta_0 + \mathbf{X}\beta - \lambda\beta \mathbf{W}\mathbf{X}$$

Model Selection

General-to-Specific procedure (Elhorst, 2010, Spatial Economic Analysis)

1. Estimate the unrestricted Spatial Durbin Model
2. Estimate the restricted SAR and SEM
3. Compare the models using Likelihood ratio statistics
4. Pick the SDM if the both restrictions are statistically rejected
5. Otherwise pick the restricted model relative to the unrejected hypothesis
6. using the model estimates compute the direct and indirect effects

$$\frac{\partial u}{\partial x}; \frac{\partial u}{\partial Wx}$$

Estimation

Maximum Likelihood very common, implemented in a number of statistical and econometric packages

Generalized Methods of Moments also common, uses internal instruments

Bayesian methods new frontier

ML is fast and reliable, GMM is useful when some covariates are endogenous

Software

GeoDa basic spatial econometrics, ML estimation, Exploratory Spatial Data Analysis

GeoDaSpace SAR and SEM, endogenous variables, heteroskedasticity , structural instability

STATA SPPACK (spatial weights, linear regression) + XSMLE (panel data)

MATLAB SE toolbox (almost everything + panel data models)

R-Geo SPDEP (almost everything) SPHET (heteroschedasticity in spatial models, GMM estimation) SPLM (panel data) MCSPATIAL (non-parametric spatial regression) + link to other libraries for spatial data (extensions)

Extensions

- Spatially varying parameters models (North-South regimes, group-specific parameters)
- Geographically Weighted Regression (parameters function of geographical coordinates)
- Coordinates in the regression - Generalized Additive Models (spatial trends, unobserved spatial heterogeneity in non-linear models)
- Non-linear spatial models (logit, probit, spatial quantile regression)

Model and Data

Spatial specification of the city size model

The predictions of the monocentric models describe an equilibrium condition which reflects also additional sources of spatial heterogeneity over which households have defined preferences

- presence of local public goods such as open spaces
- environmental amenities
- a combination of both
- agricultural amenities

that and are likely clustered in space but unobservable

There is indication for the the Spatial Error Model

Concerning the spatial spillover effects, a spatial effect in the covariates appears more reasonable

If the income increases in Milan, Urbanization will increase in Cernusco sul Naviglio

If the urbanization increases in Milan, not necessarily the urbanization will increase in CsN also

Testing

Tests suggest including a spatial structure in the error term and not in the dependent variable

The spatial model for city size

$$\begin{aligned} u &= X\beta + WX\gamma + Z\delta + \varepsilon \\ \varepsilon &= \lambda W\varepsilon + e \end{aligned}$$

X: population, income, transport cost, farmland value in the spatial unit

$\beta = \frac{\partial u}{\partial x}$ direct effects

Z: controls

WX: X in the neighbors

$\gamma = \frac{\partial u}{\partial Wx}$ spillover effects

$W\varepsilon$: unobserved spatial heterogeneity

Structural Instability

Problem: market forces determining the equilibrium city size may operate differently in large and small cities

Solution: sample splitting

- ex-ante: you have knowledge about the nature of splitting
- : split the sample in groups
- estimate coefficients separately for each group
- how large is a large city?

Endogenous sample splitting

1. define a threshold variable, population (p) in this case
2. take its m unique values
3. for each value generate a dummy variable $\mathbf{d}_m(p_i > p^*_m)$ equal to one if the population of the city (p_i) is larger than that critical value (p^*)
4. estimate

$$
\begin{aligned}
\mathbf{u} &= \mathbf{X}\beta + \mathbf{WX}\gamma + \mathbf{Z}\delta + \varepsilon \\
\varepsilon &= \lambda \mathbf{W}\varepsilon + \mathbf{e} \\
\beta &= \beta^0 + \beta^1 \mathbf{d}_m \\
\gamma &= \gamma^0 + \gamma^1 \mathbf{d}_m
\end{aligned}
$$

5. store the concentrated sum of squares $S_m(p^*_m)$
6. the estimator of the threshold p^1 is the value of p^*_m that minimizes $S_m(p^*_m)$

We use the spatial version of the Chow test for structural instability. It is a LR test comparing the restricted ($\beta^1 = \gamma^1 = 0$) model with the unrestricted one

If structural stability is rejected repeat steps 3-6 using the model

$$\begin{aligned} \mathbf{u} &= \mathbf{X}\beta + \mathbf{W}\mathbf{X}\gamma + \mathbf{Z}\delta + \varepsilon \\ \varepsilon &= \lambda \mathbf{W}\varepsilon + \mathbf{e} \\ \beta &= \beta^0 + \beta^1 \mathbf{d}^1 + \beta^2 \mathbf{d}_m \\ \gamma &= \gamma^0 + \gamma^1 \mathbf{d}^1 + \gamma^2 \mathbf{d}_m \end{aligned}$$

to find the estimator of the second threshold, p^2, the value of p of that minimizes $S_m\left(p_m^* | p^1\right)$

and test ($\beta^2 = \gamma^2 = 0$)

Data

- Lombardy region (IT), 24,000 km2, 600 km of motorways and 11,000 km of other roads
- 14.5% (12.6% in 2009) of land is urbanized, almost twice the national average
- in less than 15 years approximately 45,000 ha of land have been urbanized
- 90,000 square meters, almost 9 football fields, of agricultural and natural land are being lost every day to leave space to commercial and residential areas (Legambiente)

Variable	Description of the variable	Mean
u	Urbanized (residential, industrial and commercial) area - hundreds of hectares (DUSAF 2012)	2.79
p	Total Population - thousands of inhabitants (ISTAT 2011)	6.31
i	Average income - thousands of euros (MEF 2012)	19.51
t	Transport Costs - inverse of the number of vehicles (cars) per inhabitant (ACI 2012)	1.65
r	Farmland Value - thousands of euro per hectare (INEA 2012 and DUSAF 2012)	31.02
road	Area occupied by the road network - hundreds of hectares (DUSAF 2012)	7.22
train	Area Occupied by the rail network - hundreds of hectares (DUSAF 2012)	1.80
$aero_D$	Dummy - 1 if a portion of soil is occupied by airports (DUSAF 2012)	0.03
$port_D$	Dummy - 1 if a portion of soil is occupied by ports (DUSAF 2012)	0.05
constr	Area occupied by construction sites- hundreds of hectares (DUSAF 2012)	3.75

Findings

	Linear	SDEM	SEM	SAR
$\beta_{constant}$	-0.904**	3.481*	-0.876*	-0.499
	(0.400)	(2.029)	(0.482)	(0.395)
β_p	0.113***	0.112***	0.112***	0.111***
	(0.006)	(0.007)	(0.007)	(0.006)
β_i	0.137***	0.13***	0.139***	0.087***
	(0.018)	(0.026)	(0.023)	(0.019)
β_t	-0.448***	-0.386***	-0.334**	-0.369**
	(0.15)	(0.148)	(0.146)	(0.146)
β_r	0.024***	-0.004	0.018***	0.012***
	(0.004)	(0.01)	(0.006)	(0.004)
γ_p		-0.012		
		(0.012)		
γ_i		0.003		
		(0.059)		
γ_t		-2.803***		
		(0.992)		
γ_r		0.039***		
		(0.013)		
λ		0.523***	0.54***	
		(0.049)	(0.048)	
ρ				0.304***
				(0.038)

Spatial Chow tests

Threshold	regimes	Spatial Chow stat.	DoF	p-value
47,000	2	2,204.96	14	0.000
12,000	3	451.24	23	0.000
4,000	4	93.72	32	0.000
34,000	5	66.93	41	0.006
21,000	6	52.3	50	0.394

Spatial city size model with parameter heterogeneity

Threshold	$\geq 47,000$	$< 47,000$ $\geq 34,000$	$< 34,000$ $\geq 12,000$	$< 12,000$ $\geq 4,000$	$< 4,000$
$\beta_{constant}$	191.47***	89.137***	34.994***	11.639***	1.948**
	(14.472)	(9.208)	(2.664)	(1.523)	(0.978)
Direct effects					
β_p	0.101***	0.058	0.272***	0.424***	0.606***
	(0.003)	(0.065)	(0.014)	(0.021)	(0.029)
β_i	-0.216	0.47	0.127***	-0.013	0.01
	(0.182)	(0.363)	(0.042)	(0.021)	(0.013)
β_t	-22.608***	-16.549***	-4.492***	-2.055***	-0.147**
	(5.183)	(3.434)	(0.789)	(0.425)	(0.062)
β_r	-0.177**	-0.862**	-0.022	-0.001	-0.003
	(0.083)	(0.389)	(0.015)	(0.006)	(0.005)
Indirect effects					
γ_p	-0.082**	-0.09***	-0.046***	-0.048***	-0.024
	(0.035)	(0.027)	(0.007)	(0.009)	(0.023)
γ_i	-1.627***	-0.298	-0.477***	-0.152***	-0.065**
	(0.401)	(0.335)	(0.059)	(0.038)	(0.032)
γ_t	-70.028***	-31.378***	-12.033***	-2.901***	-0.257
	(7.851)	(5.627)	(1.56)	(0.848)	(0.458)
γ_r	0.637***	0.824***	0.086***	0.027***	0.006
	(0.074)	(0.263)	(0.018)	(0.009)	(0.007)

Conclusion

- the relationship between the size of urban area and the market-related variables varies significantly across the regimes of cities of different sizes
- the response of cities to population growth monotonically increases as the size of the city decreases: an increase of population by 1000 inhabitants translates into an increase of urbanized area by 10 hectares in large cities; the figure is estimated three to six times larger in small and medium-size cities.
- the effect of transport costs on urbanization appears sizable only in medium and large cities
- a higher farmland value, which reflects higher agricultural productivity, can effectively contrast urban spreading in large cities, while it is altogether ineffective in small cities

Conclusion

- the size of large cities is determined by movements of citizens from neighboring cities
- larger size is associated to low levels of income and population and to the availability of cars for cross-cities commuting in the neighboring municipalities

One remark

empirical models should fit the reality more than the theory