Do Atlanta residents value MARTA?
Selecting an autoregressive model to recover willingness-to-pay

Gregory Macfarlane, Laurie Garrow, and Juan Moreno-Cruz

gregmacfarlane@gatech.edu, laurie.garrow@ce.gatech.edu, juan.moreno-cruz@econ.gatech.edu

Hedonic Models

Understanding what homeowners are willing to pay to live close to public transit infrastructure is important for policy and forecasting. This is usually estimated with a hedonic linear regression model,

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

with y_i the price of a home and $\boldsymbol{x}_i$ its attributes, and $\boldsymbol{\beta}$ the marginal willingness-to-pay (MWTP) for the attributes. **But this model may be inadequate** under the following conditions:

Spatial Dependence The price of a home is relative to the prices of homes nearby. This creates a missing variable bias on $\hat{\boldsymbol{\beta}}$. This is remedied with a spatial autoregressive lag (SAR) model,

$$\boldsymbol{y} = \rho W\boldsymbol{y} + X\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2}$$

Spatial Correlation Homes near each other have similar characteristics, and thus correlated errors. This invalidates OLS estimates of σ. This can be remedied with the spatial error model (SEM),

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}, \boldsymbol{u} = \lambda W\boldsymbol{u} + \boldsymbol{\epsilon} \tag{3}$$

Spatial Endogeneity Unobservable neighborhood attributes are important to price. This can cause both a missing variable bias and correlated errors. This is remedied with the spatial Durbin model (SDM)

$$\boldsymbol{y} = \rho W\boldsymbol{y} + X\boldsymbol{\beta} + WX\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{4}$$

In these models, the matrix W defines the spatial relationship between observations in the dataset. If i and j are neighbors, $[W]_{ij} \neq 0$. **Identifying the appropriate model for a given situation is an unresolved empirical problem.**

Empirical Application

4,812 Targeted marketing records of households within 5 miles of a Metropolitan Atlanta Rapid Transit Authority (MARTA) rail station. Variables:

- (log) **Market value of home — dependent variable**
- (log) **Euclidean distance to MARTA station**
- (log) Property acreage
- (log) Home square footage
- (log) Age of home
- (log) Household income
- (log) Euclidean distance to freeway entrance.
- Property type (condo/single)
- Ethnicity

We estimate the models by maximum likelihood with the `spdep` package for R [3].

Model Selection

Misspecification Consequences

	Estimated Model			
True DGP	OLS	SAR	SEM	SDM
OLS: $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$	-	inefficient	inefficient	inefficient
SAR: $\boldsymbol{y} = \rho W\boldsymbol{y} + X\boldsymbol{\beta} + \boldsymbol{\epsilon}$	$\hat{\boldsymbol{\beta}}$ biased	-	$\hat{\boldsymbol{\beta}}$ biased	inefficient
SEM: $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} = \lambda W\boldsymbol{\epsilon} + \boldsymbol{u}$	$\hat{\sigma}^2$ invalid	$\hat{\sigma}^2$ invalid	-	inefficient
SDM: $\boldsymbol{y} = \rho W\boldsymbol{y} + X\boldsymbol{\beta} + WX\boldsymbol{\gamma} + \boldsymbol{u}$	$\hat{\boldsymbol{\beta}}$ biased	$\hat{\sigma}^2$ invalid	$\hat{\boldsymbol{\beta}}$ biased	-

Classical Framework

Spatial models can be expensive to estimate, can OLS residuals show need?

$$H_0 : \rho, \lambda, \boldsymbol{\gamma} = 0; \qquad \text{No spatial effects.}$$
$$H_a : \rho \text{ or } \lambda \text{ or } \boldsymbol{\gamma} \neq 0; \qquad \text{Spatial effects.}$$

Test: Lagrange multipliers [1, 2]

$$LM_\rho = \frac{(\hat{\boldsymbol{\epsilon}}'W\boldsymbol{y}/\hat{\sigma}^2)^2}{nJ}$$
$$LM_\lambda = \frac{(\hat{\boldsymbol{\epsilon}}'W\hat{\boldsymbol{\epsilon}}/\hat{\sigma}^2)^2}{T}$$

with

$$J = \frac{1}{n\hat{\sigma}^2}[(WX\hat{\boldsymbol{\beta}})'(I - X(X'X)^{-1}X')(WX\hat{\boldsymbol{\beta}}) + T\hat{\sigma}^2]$$

and T the trace of the matrix $W'W + W^2$

General Framework

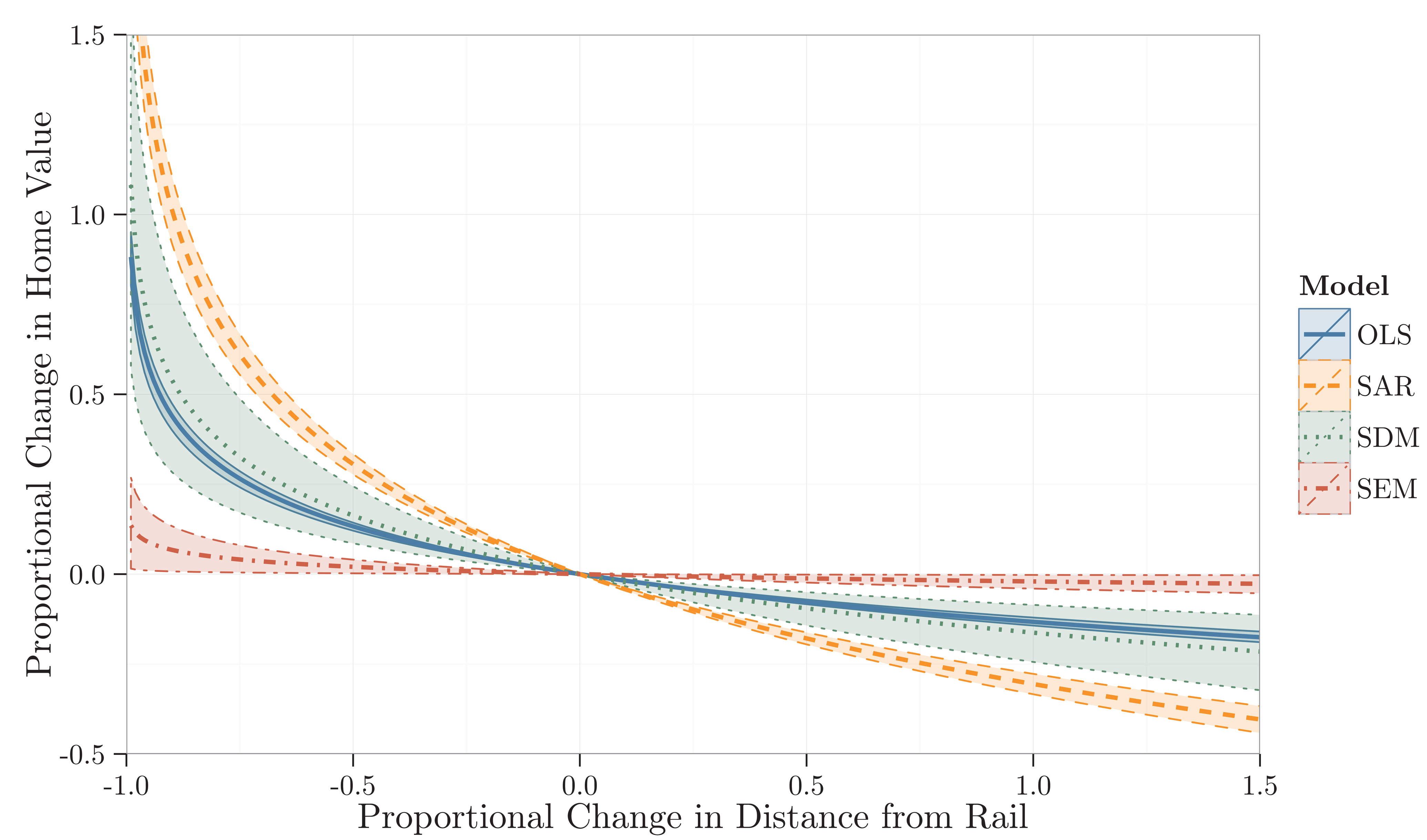
The SDM is a general model that subsumes the other three.

$$H_0 : \rho, \lambda, \boldsymbol{\gamma} \neq 0; \qquad \text{Spatial effects.}$$
$$H_a : \rho \text{ or } \lambda \text{ or } \boldsymbol{\gamma} = 0; \qquad \text{No spatial effects.}$$

Test: likelihood ratio to estimate posterior probabilities $SAR + SEM = SDM : \pi_a + \pi_b = 1$ [4, 6]

$$\begin{aligned}
\boldsymbol{y}_c &= \pi_a \boldsymbol{y}_a + \pi_b \boldsymbol{y}_b \\
\boldsymbol{y}_c &= \pi_a((I - \rho W)^{-1}(X\beta + \boldsymbol{\epsilon})) + \pi_b(X\beta + (I - \rho W)^{-1}\boldsymbol{\epsilon}) \\
(I - \rho W)\boldsymbol{y}_c &= X(\pi_a \boldsymbol{\beta}) + (I - \rho W)(X\pi_b\boldsymbol{\beta}) + (\pi_a + \pi_b)\boldsymbol{\epsilon} \\
\boldsymbol{y}_c &= \rho W\boldsymbol{y_c} + (\pi_a + \pi_b)X\boldsymbol{\beta} + WX(-\rho\pi_b\boldsymbol{\beta}) + \boldsymbol{\epsilon}
\end{aligned}$$

Estimated MWTP with 95% Confidence Intervals



Other Studies

Statistic	Osland [8]	Löchl [7]	Ibeas [5]	This Study
$\ln(\mathcal{L}_{SDM})$	96.9	4192.6	−2.1	−272.7
$\ln(\mathcal{L}_{SEM})$	80.9	4118.0	−34.8	−338.1
$\ln(\mathcal{L}_{SAR})$	65.8		−75.0	−838.8
$\ln(\mathcal{L}_{OLS})$	52.0	3183.3	−111.9	−2006.7
LR statistic	31.9*	149.2*	65.5*	130.9*
Classical	SEM†	SEM	Unknown‡	SEM
General	SDM	SDM	SDM	SDM

* Reject null hypothesis with $p < 0.01$.
† Selected SDM after likelihood ratio test.
‡ Did not report *LM* statistics, but selected SEM.

Conclusions

The econometric treatment of **spatial effects matters for policy interpretations**; the SDM estimates of MWTP are substantially less certain than the OLS estimates in this case.

Existing **selection frameworks lead to different models**; the classical framework selected the SEM which in this case appears inconsistent.

The general framework **minimizes the consequences of Type I error** error; hypothesis testing is built to minimize the risk of Type I error.

The **SDM is general and conservative**:

- It is a linear combination of the SAR and SEM models.
- It is the outcome of a spatial fixed effects process.
- It is the outcome of a spatially endogenous variables process.

The SDM may be inefficient; if researchers wish for a more efficient model they can test restrictions to the SDM.

References

[1] Luc Anselin. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, 20(1):1–17, September 1988. ISSN 00167363. doi: 10.1111/j.1538-4632.1988.tb00159.x. URL http://doi.wiley.com/10.1111/j.1538-4632.1988.tb00159.x.

[2] Luc Anselin, Anil K. Bera, Raymond J G M Florax, and Mann J. Yoon. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26:77–104, 1996.

[3] Roger Bivand. spdep: Spatial Dependence, Weighting Schemes, Statistics, and Models, 2013. URL http://cran.r-project.org/package=spdep.

[4] Raymond J G M Florax, Hendrik Folmer, and Sergio J Rey. Specification searches in spatial econometrics: The relevance of Hendry's methodology. *Regional Science and Urban Economics*, 33:557–579, 2003.

[5] Ángel Ibeas, Ruben Cordera, Luigi Dell'Olio, Pierluigi Coppola, and Alberto Dominguez. Modelling transport and real-estate values interactions in urban systems. *Journal of Transport Geography*, 24:370–382, September 2012. ISSN 09666923. doi: 10.1016/j.jtrangeo.2012.04.012. URL http://linkinghub.elsevier.com/retrieve/pii/S096669231200124X.

[6] James P. LeSage and Robert Kelley Pace. *Introduction to Spatial Econometrics*. Chapman and Hall/CRC, 2009.

[7] Michael Löchl and Kay W. Axhausen. Modeling hedonic residential rents for land use and transport simulation while considering spatial effects. *Journal of Transport and Land Use*, 3(2):39–63, 2010. doi: 10.1598/jtlu.v3i2.117.

[8] Liv Osland. An application of spatial econometrics in relation to hedonic house price modeling. *The Journal of Real Estate Research*, 32(3):289–320, 2010.