Errors in variables and spatial effects in hedonic house price models of ambient air quality

Luc Anselin and Nancy Lozano-Gracia

Presented by

Julia Beckhusen and Kosuke Tamura

February 29, 2008
AGEC 691T: *SHaPE* Seminar

Objective

- **Use hedonic model to estimate effects of pollution on house prices:**
  - House prices would be benefited by clean air
  - Focus on Ozone and total suspended particulate matter (TSP)
- **Examine different specifications of the spatial hedonic model**
  - OLS
  - Instrumental variables without space: Endogenous pollution variables
  - Spatial dependent lag model without endogeneity
  - Spatial dependent lag model with endogeneity
- **Develop robust standard error estimates**
  - Heteroskedasticity and Autocorrelation robust (HAC)

Three main data and sources

1. Individual house price and characteristics: Experian Company
   - 115,729 samples and geocoded location
2. Neighborhood characteristics: 2000 US Census of Population and Housing
3. Measures of ozone (OZ) and particulate matter concentration (TSP):
   - South Coast Air Quality Management District

Classifications of variables

- House attributes: Elevation, Livarea, Landarea, Baths, Pool, Age, AC, Heat
- Locational attributes (accessibility): Avdisp, Beach, Highway1, Highway2,
- Neighborhood characteristics: Poverty, White, Over65, College, Income, Vcrime, API,
- County dummies: LA, Riverside, San Bernardino, and Orange county
- Interpolated air pollution values: OZ and TSP

Use of ArcGIS: Buffering and Kriging

- Highways1: House within a 0.25 km from a highway
- Highways2: House within 0.25 to 1 km from a highway
- OZ and TSP

Variable
Elevation
Livarea
Landarea
Baths
Fireplace
Pool
Age
AC
Heat
Beach
Avdistp
Highway1
Highway2
Traveltime
Poverty
White
Over65
College
Income
Vcrime
API
Riverside
San Bern.
Orange
OZ
TSP

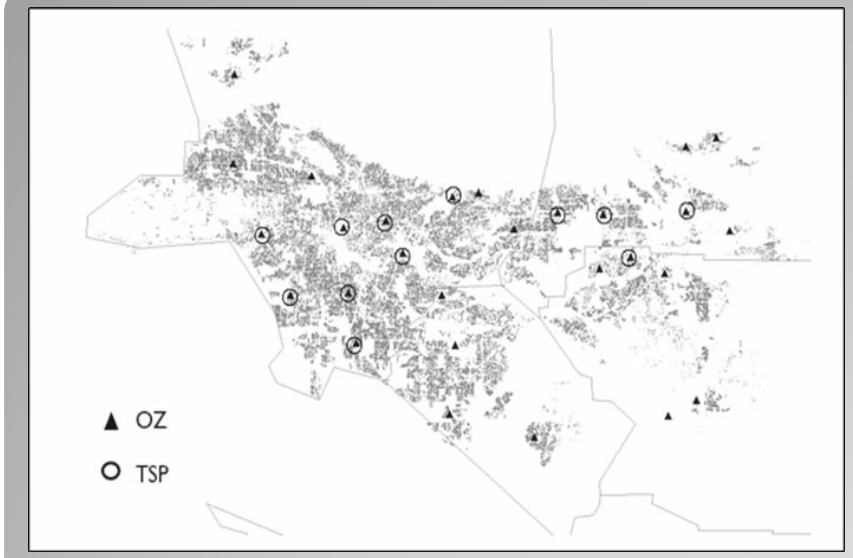


Fig. 1 Spatial distribution of houses and location of monitoring stations

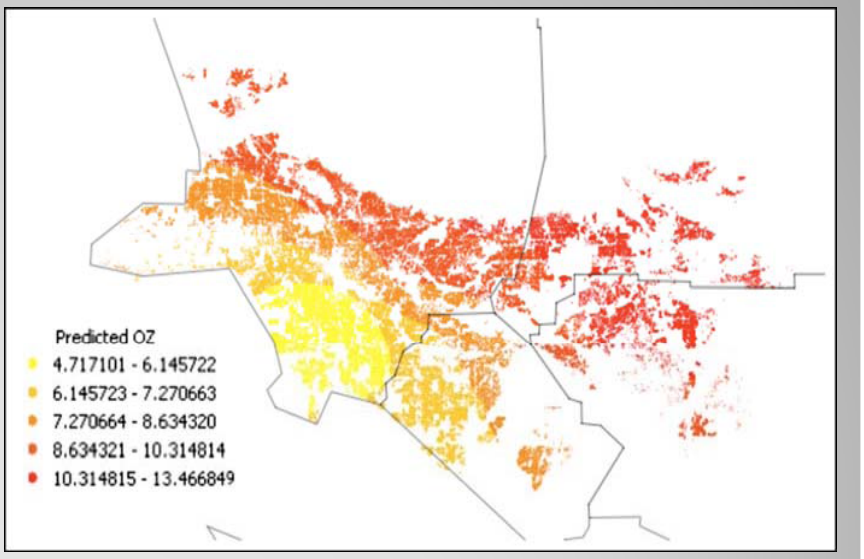


Fig. 2 Kriging Interpolation: OZ

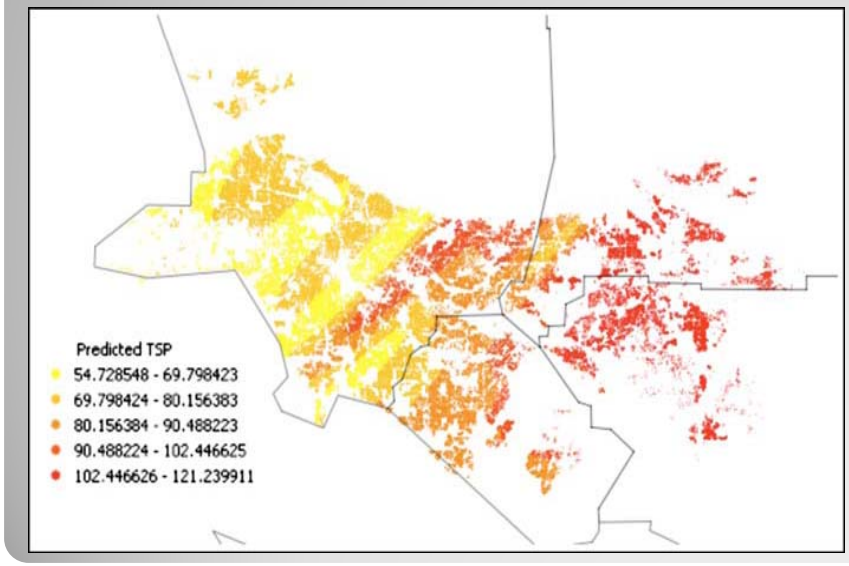


Fig. 3 Kriging interpolation: TSP

Source: Anselin and Lozano-Gracia (2007)

Two distinctive approaches: Spatial hedonic model

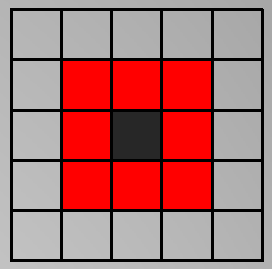
1st distinctive approach

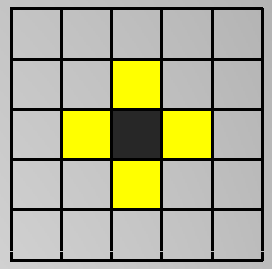
- Use of a spatial two-stage least squares estimator (S2SLS)
  - Spatial lag model without endogeneity (LAG)
  - Allowing a spatial lag and endogenous variables (LAG-end)

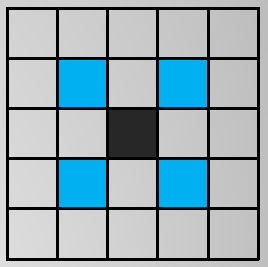
2nd distinctive approach

- Considering spatial error autocorrelation
  - Specification tests: presence of autocorrelation
- Application of heteroskedasticity and autocorrelation (HAC):
  - Kelejian and Prucha (2006a)

1st distinctive methodological approach







- Spatial lag model: $y = \rho W y + X\beta + u,$

where

y = a (n x 1) vector of observations

X = a (n x k) matrix of observations

W = a (n x n) spatial weights matrix

u = a (n x 1) vector of i. i. d. error terms

ρ = the spatial autoregressive coefficients

β = a (k x 1) vector of regression parameters

- Reduced form of Spatial lag model:

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} u,$$

where, standard regularity condition: Inverse $(I - \rho W)^{-1}$

$$(I - \rho W)^{-1} = I + \rho W + (\rho W)^2 + \cdots.$$

- House prices as a function of house and neighbors characteristics

Spatial lag model with endogenous variables:

$$y = \rho W y + \boxed{Y \nu} + X \beta + u ,$$

where

Y = a (n x p) matrix of endogenous variables

ν = coefficient vector

y = a (n x 1) vector of observations

X = a (n x k) matrix of observations

W = a (n x n) spatial weights matrix

u = a (n x 1) vector of i. i. d. error terms

ρ = the spatial autoregressive coefficients

β = a (k x 1) vector of regression parameters

Note: Endogenous variables are air quality variables: OZ and TSP

Endogenous variables: y_2 and y_3

- Air quality variables
- True pollution is not obtained at location *i* of house
- Replacement of pollution by spatially interpolated value: Kriging prediction
- Interpolated value: True pollution containing error at location *i*

$$y_{2i} = y_{2i}^{*} + \psi_i$$

where

y_{2i}^{*} = True air quality entering into the agent's utility function

y_{2i} = Observed value (interpolated value)

Ψ_i = An error term related to interpolated error

Practical perspective

- Kriging predictor: prediction error will be spatially structured

Suggestion

- Mimic spatially structured equation disturbance *u*:

Second distinctive approach: HAC approach in Kelejian and Prucha(2006a)

- Allows for spatial error autocorrelation of unspecified form as well as heteroskedasticity
  - More powerful than traditional white standard error which only accounts for heteroskedasticity
- HAC (heteroskedastic and autocorrelation robust): uses a non-parametric estimator for spatial covariance with weighted averages of cross-products of residuals:

$$\psi = Q'\Sigma Q$$

  - Where, Σ is the non-diagonal spatial variance-covariance matrix for the error terms and *Q* is a matrix of instruments

  - The individual elements of the matrix are estimated by

$$\widehat{\psi}_{r,s} = (1/n)\sum_i \sum_j q_{ir} q_{js} \hat{u}_i \hat{u}_j K(d_{ij}/d)$$

  - where K is the kernel function, three different kernel functions used in estimation. û is the residual vector

- This new var/cov matrix is used to compute the HAC variance (14) which the authors argue will give us a more realistic measure of standard errors of the estimates.

Estimation Results (table 3)

- OLS
  - Overall, parameter estimates are significant and of the correct sign
  - Evidence of high residual spatial autocorrelation
  - LM test results indicate the lag specification as the proper alternative

- Instrumental Variables (IV) – coordinates of house location as instrument
  - Non-spatial 2SLS
  - Pollutants treated as endogenous
  - Results very similar to OLS

- Lagged Dependent Variable (LAG)
  - Spatial 2SLS with a spatially lagged dependent variable
  - Results very similar to OLS and IV
  - Spatial autocorrelation still present

- LAG-end
  - Spatial 2SLS with a spatially lagged dependent variable and pollutants treated as endogenous
  - Results still very similar, estimated coefficient on Poverty becomes significant and negative.
  - Spatial autocorrelation still exists

Table 4 Pollutant coefficients by estimator—queen weights

Variable Name	OLS	IV	LAG	LAG-end
OZ	−0.0253	−0.0137	−0.0195	−0.0099
TSP	−0.0047	−0.0102	−0.0032	−0.0073
ρ	–	–	0.3314	0.3266
RLM-LAG	2357.271	–	–	–
RLM-ERR	1339.671	–	-	–
DWH	2,540	–	–	–
A-K		18323.48	60.24	137.46

Comparing pollution coefficients across models

- Coefficients are significant and negative
- Controlling for endogeneity has a strong affect on estimates
- A 1 ppb reduction in OZ would increase house prices by between 2.5% (OLS) and 0.99% (LAG-end)
  - 8hr OZ concentrations > 105 ppd is “unhealthy” to the general public
- A 1 μ/m^3 reduction in TSP would increase house prices by between 0.47% (OLS) and 1.02% (IV)
  - Daily allowable average by EPA is 50 μ/m3

Table 5 Standard errors: OZ

		Coeff.	Standard errors				
		OZ	Classical	White	HAC-Ep	HAC-Tr	HAC-Bi
OLS		−0.0253	0.0008	0.0008	0.0018	0.0016	0.0016
IV		−0.0137	0.0010	0.0011	0.0026	0.0023	0.0024
LAG	Queen	−0.0195	0.0007	0.0008	0.0012	0.0011	0.0011
LAG-end		−0.0099	0.00099	0.0010	0.0017	0.0016	0.0016
LAG	Knn6	−0.01822	0.00078	0.00087	0.00125	0.00115	0.00115
LAG-end		−0.00895	0.00099	0.00108	0.00175	0.00157	0.00158
LAG	Knn12	−0.01802	0.00078	0.00086	0.00123	0.00113	0.00114
LAG-end		−0.00853	0.00098	0.00107	0.00170	0.00154	0.00155

LAG-end with HAC variance estimates

- Uses three spatial weights matrices and three kernel functions
  - Choice of spatial weight and kernal function has little effect on estimates
- HAC estimates have twice as large standard error estimates than either classical or White results.
- This wider bound allows for the presence of remaining spatial autocorrelation that wasn’t included in the model

Conclusions

- Models that account for both spatial correlation and endogeneity do a much better job than OLS but still can not explain all existing spatial correlation
  - Specification of spatial weights and kernel functions have little effect across models
- A robust standard error estimator (HAC) should be used to estimate more realistic bounds on the estimates.
  - Extremely relevant to policy makers who may use these estimates to calculate MWTP for air pollution reduction or other derived welfare measures
- Future research: perhaps changing the scale of analysis to reduce heterogeneity and spatial correlation. How would this help?
- How much time was spent developing the 9 cases when in the end, the results barely changed?
- What does this tell us about the importance of proper weight matrix and kernel function specification?