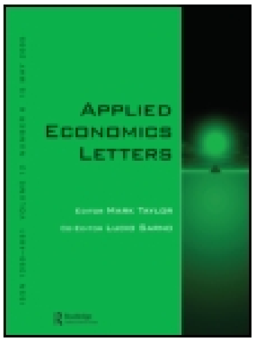

Applied Economics Letters

ISSN: 1350-4851 (Print) 1466-4291 (Online) Journal homepage: https://www.tandfonline.com/loi/rael20

A mean squared error estimator of market size in hedonic price analysis

Camilo Sarmiento

To cite this article: Camilo Sarmiento (2005) A mean squared error estimator of market size in hedonic price analysis, Applied Economics Letters, 12:6, 355-359, DOI: 10.1080/13504850500044187

To link to this article: https://doi.org/10.1080/13504850500044187

Published online: 17 Aug 2006.

Applied Economics Letters, 2005, **12**, 355–359

Routledge
Taylor & Francis Group

A mean squared error estimator of market size in hedonic price analysis

Camilo Sarmiento

Pacific Institute for Research and Evaluation, 11710 Beltsville Drive, Suite 300, Calverton, MD 20705, USA
E-mail: csarmiento@pire.org

This paper presents a spatial kernel estimator that allows coefficients and market size to be estimated at the observation of interest. In economic terms, the bandwidth length of the spatial kernel estimator captures the size of the market incorporated in estimation at the location of interest, and this optimal size of the market included in regression minimizes mean square error (MSE).

I. Introduction

Sources of structural change in cross-sectional analysis are analogous to those in time series analysis: technology, endowments, and institutions (Sarmiento, 2004). In the valuation of product characteristics, for example, the structure of parameter stability in cross-sectional analysis implicitly assumes a single market for all observations (Wang, 2003) while, in contrast, models with spatially varying coefficients allow for spatial market fragmentation. This paper presents a spatial kernel estimator that permits coefficients and market size to be estimated at the observation of interest. The size of the market selected in the spatial kernel is evaluated with a mean square error (MSE) criterion. In effect, the bandwidth length of the spatial kernel estimator captures the size of the market incorporated in estimation at the location of interest.

Application of the spatial kernel estimator clearly elicits heterogeneity in local markets for coal, and systematic spatial patterns are revealed in the variation of the implicit valuation of coal quality across location.

II. Spatial Structural Change in Price Analysis

The law of one price states that the price of a homogeneous product should be the same across space in the absence of transfer costs (e.g. transportation cost and institutional factors). In reality, transfer costs, which restrict trade, competitiveness and mobility of resources across space, contribute to significant regional price differences (Haining, 1990; Case, 1991; Florkowski and Sarmiento, 2005). For example, the dissimilarities in the markets for coal of different qualities in the USA hinges partly upon the distance of the plant to transportation centres and mines and, in particular, the plants' location relative to mines providing high quality coal.

In electric generation, coal with more heat content, which is commonly measured in British thermal units (BTU), is more productive because it increases combustion needed in the production of electricity; and the relation between sulfur content and coal price captures the plant's valuation of switching to cleaner coal in the context of environmental regulation.[1] The relation between quality and price, however, depends on local peculiarities in the supply and demand for different types of coal (Rosen, 1974).

[1] The most common mechanism used by power plants to comply with emission caps is fuel switching to coal with lower sulfur content.

Applied Economics Letters ISSN 1350–4851 print/ISSN 1466–4291 online © 2005 Taylor & Francis Group Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/13504850500044187

A semiparametric formulation of the local market valuation of coal quality that specifies the price of coal paid by electric power plants Y_s in terms of its characteristics (the coal heat content h_{1s} and the coal sulfur content su_{2s}) and the location of the plant is, therefore,

$$\begin{aligned} Y =_s &= \Gamma_s' \begin{bmatrix} 1 \\ \ln h_{1s} \\ \ln su_{2s} \end{bmatrix} + u_s \\ &= \Gamma_s' \mathbf{x}_s + u_s \end{aligned} \tag{1}$$

where the econometric residual $u_s \sim (0, \sigma_s^2)$ depends on market conditions at location s; and the spatially varying coefficients $\Gamma_s{}'$ capture different local conditions in the supply and demand markets for coal.

III. A Spatial Kernel

To estimate the relative implicit price of quality Γ_s, both coefficients and market size at the power plant of interest need to be estimated. A spatial kernel that estimates market size simultaneously with the parameters of the model is next formulated by adapting a kernel estimator of varying coefficients in Sarmiento (2005) to spatial modelling. In particular, a spatial kernel estimator of the varying coefficient model in Equation 1 evaluated at observation q minimizes the weighted sum of squares:

$$\underset{\Gamma_q}{\text{Min}} = \sum_{s=1}^{M} \left\{ k\left(\frac{d(s,q)}{h}\right) \left(y_s - \Gamma_q' \mathbf{x}_s\right)^2 \right\} \tag{2}$$

where the distance function between firms s and q is $d(s, q)$, and the bandwidth length h determines the size of the spatial weights allotted to each observation, while the density function k of the kernel captures the shape of the weights. For example, a Gaussian weight function for the kernel in Equation 2 evaluated at the reference point q is

$$k\left(\frac{d(s,q)}{h}\right) = (2\pi)^{-1/2} \exp\left[-\left(\frac{1}{2}\right)\left(\frac{d_{sq}}{h}\right)^2\right] \tag{3}$$

where $d_{sq} = d(s, q)$.

Silverman (1986) shows that nonparametric regression is robust to the structure of the weights in Equation 3 (i.e. a Gaussian kernel) but it is sensitive to the bandwidth length h. In spatial modelling, the extent of the market incorporated in estimation of coefficients at location s depends on the bandwidth length and, thus, selection of the market size in regression is the most important factor in the spatial kernel. By construction, the bias is inversely related to h (market size included), and the assumption of parameter stability, $h \to \infty$, assumes a single market for all observations.

Following Sarmiento (2005), if $k(d(s, q)/h) = \beta_{sqh}^2$, then the nonparametric estimator of varying coefficients is

$$\mathbf{Q}_q' \mathbf{E}_q = 0 \tag{4a}$$

$$\frac{\mathbf{E}_q' \mathbf{E}_q}{\sum_s \beta_{sqh}^2} = \sigma_q^2 \tag{4b}$$

where

$$\mathbf{E}_q = \left\{ \beta_{1qh} y_1 - \beta_{1qh}(\Gamma_q' \mathbf{x}_1), \ldots, \beta_{Mqh} y_M - \beta_{Mqh}(\Gamma_q' \mathbf{x}_M) \right\}$$

$$\mathbf{Q}_q = \left\{ \beta_{1qh} \mathbf{x}_1, \ldots, \beta_{Mqh} \mathbf{x}_M \right\}$$

and, for inference testing, the associated variance for an estimated parameter at reference point q is the qth element of the vector

$$\text{Var}\left(\Gamma_q\right) = \frac{\mathbf{E}_q' \mathbf{E}_q}{\sum_s \beta_{sqh}^2} \left\{ \text{Diag}\left(\mathbf{Q}_q' \mathbf{Q}_q\right)^{-1} \right\}$$

where the standard errors depend on the bandwidth length.

Yet, the extent of the market captured by the bandwidth length in estimation of parameters at an arbitrary location s cannot be simultaneously estimated with the other coefficients in Equation 4 since in that case $h \to 0$. Because of this difficulty, the criterion of *cross-validation* is often used to estimate the bandwidth (Engle *et al.*, 1986; Schmalensee and Stoker, 1999). Cross-validation, a mean squared error criterion, is frequently implemented by minimizing the estimated prediction error:

$$J(h) = M^{-1} \sum_{q=1}^{M} \left(y_q - \hat{\hat{y}}_q \right)^2$$

where $\hat{\hat{y}}_q = \hat{\hat{\Gamma}}_q \mathbf{x}_q$ is computed as the 'leave-one-out' estimator, which deletes the qth observation in Equation 1 when estimating $\hat{\hat{\Gamma}}_q$. Thus, different from $\hat{\Gamma}_q$, the vector $\hat{\hat{\Gamma}}_q$ is estimated as in Equation 4, but without the qth observation. Hence, the bandwidth length is selected by

$$\underset{h}{\text{Min}} \quad J(h) \tag{5}$$

where, given Equation 5, coefficient estimates at the desired reference point from the spatial kernel are obtained from Equation 4.

The solution of the optimization in Equation 5 thus generates a bandwidth h^* – the optimal coverage of the market when estimating the economic relation for an arbitrary location s – which can be used in Equation 4 to estimate coefficients at the observation of interest. The derived spatial kernel estimator, therefore, corresponds to a Nadaraya–Watson local estimator of the gradient at observation q when the source of heterogeneity is location. The presence of spatially varying coefficients and heteroscedastic residuals could then be incorporated at any desired location by varying the reference point in the spatial kernel estimator.

IV. Estimation

Implementation of the spatial kernel estimator to the fuel upgrading model in Equation 1 uses available data on coal quality, the price of coal (per ton), and the location of coal-powered plants for 268 observations. The *Electric Power Annual* (published by the Federal Energy Commission) reports firm-level data for the year 2000 on coal price, and the coal heat content measured in BTU and sulfur content used by each firm, while data on location for each firm used in the sample is extracted from *EPA*'s emissions and generation resource integrated database (E-GRID).[2] Yet, to implement the spatial kernel, geographical coordinates need to be transformed onto relative distances using polar coordinates (Marsen and Tromba, 1988).

To estimate Equation 1 using the spatial kernel and available data, a grid search is applied over the bandwidth parameter h of the kernel defined in Equation 2. Table 1 shows the estimated prediction error in Equation 5 computed as the 'leave-one-out' estimator for different values of window length. This criterion of cross-validation captures goodness of fit through observations not included in the regression capturing the effect of the variance as well as the biases from the model specification. Results show that the estimated prediction error in Equation 5, which corresponds to a mean square error criterion, is minimized at a bandwidth length equal to 0.08. The estimate of the bandwidth in Table 1 thus clearly elicits heterogeneity in local markets for coal. After selecting the bandwidth length of the spatial kernel estimator, the valuation of coal quality in Equation 1, i.e. the implicit price of fuel upgrading, can be estimated at the location of interest.

Given the selected bandwidth in Table 1, coefficients estimates can be obtained at the observation of interest. Selected estimates in Tables 2 and 3 indicate that differences in the relative implicit price of upgrading BTU across power plants are more pronounced from south to north than from east to west, while the firm longitude explains better than firm latitude differences in the implicit price of upgrading sulfur content. The spatial kernel estimates, moreover, identify that firms that face the lowest relative implicit price of upgrading fuel heat content BTU are located East of Lousiana (in the intersection of latitude $<32^\circ$ and longitude $>96^\circ$), while firms that face the lower relative implicit price for an additional unit of less sulfur content (clean fuel) are located at longitudes $>108^\circ$ (i.e. West of Colorado Springs). Estimates also indicate that the relative implicit price of fuel upgrading in terms of coal productivity is highest among US power plants located North of Champaign, IL, while the relative implicit price of fuel upgrading in terms of lower sulfur content is largest for US power plants located in Virginia.

More specifically, Tables 2 and 3 indicate that the relative implicit price of upgrading coal productivity is highest among US power plants located in the vicinity of Matinette, WI (45° 6′N with 87° 38′W and at 41° 80′N with 90° 23′W in Dubuque, IA), and lowest at 31° 56′N with 96° 05′W (South of Dallas, TX), while the relative implicit price of fuel upgrading in terms of lower sulfur content is largest for US power plants located in the vicinity of 38° 56′N with 76° 68′W (West of Norfolk, VA) and lowest at 39° 51′N

Table 1. Estimated prediction error under different bandwidth lengths

Bandwidth	EPE
1	0.0184
0.9	0.0184
0.8	0.0183
0.7	0.0183
0.6	0.0182
0.5	0.0181
0.4	0.018
0.3	0.0178
0.2	0.0174
0.1	0.0167
0.09	0.0167
0.08	**0.0166**
0.07	0.0167
0.06	0.0168
0.05	0.0171
0.04	0.0177
0.03	0.0188
0.02	0.02

Notes: The estimated prediction error (EPE) is calculated using the 'leave-one-out' estimator.

[2] The 268 observations stem from the common denominator of coal-fired plants reported in the *Electric Power Annual* and the *EPA*'s emissions and generation resource integrated database (E-GRID).

Table 2. Selected spatial kernel coefficient estimates categorized by firms' longitude

Firm location	32°13′33″N	32°03′56″N	31°82′06″N	32°05′56″N	32°16′00″N
	81°13′33″W	93°56′44″W	96°05′47″W	109°88′61″W	110°90′44″W
Constant	6.00[c]	5.51	5.56	5.50	5.52
	(46.93)	(46.29)	(45.70)	(34.83)	(34.10)
Sulfur content	−0.06	−0.01	−0.00	0.02	0.03
	(−2.89)	(−0.47)	(−0.07)	(0.91)	(0.96)
BTU	1.79	1.46	1.43	1.47	1.48
	(20.71)	(19.37)	(19.07)	(14.73)	(14.45)
Firm location	38°56′39″N	38°56′78″N	38°52′67″N	38°55′83″N	32°16′00″N
	76°68′06″W	85°41′39″W	87°25′22″W	90°83′61″W	104°70′56″W
Constant	6.13	6.14	6.16	6.13	5.67
	(43.01)	(40.11)	(39.79)	(38.06)	(30.17)
Sulfur content	−0.07	−0.07	−0.06	−0.05	0.05
	(−3.58)	(−3.40)	(−3.25)	(−2.15)	(1.75)
BTU	1.91	1.93	1.94	1.92	1.63
	(19.86)	(19.03)	(19.06)	(18.22)	(13.47)
Firm location	41°72′39″N	41°79′14″N	41°72′19″N	41°80′75″N	41°75′72″N
	81°25′22″W	83°44′86″W	86°90′92″W	90°23′33″W	110°59′86″W
Constant	6.22	6.23	6.25	6.26	5.28
	(41.50)	(40.97)	(40.04)	(38.96)	(31.21)
Sulfur content	−0.07	−0.06	−0.05	−0.04	0.06
	(−3.35)	(−3.14)	(−2.62)	(−1.95)	(1.92)
BTU	1.97	1.98	1.99	2.00	1.75
	(19.78)	(19.66)	(19.42)	(19.10)	(14.74)

Notes: The terms in parentheses refer to the *t*-value.
BTU refers to the heat content of coal measured in British thermal units; and coal sulfur content captures coal volatility.

Table 3. Selected spatial kernel coefficient estimates categorized by firms' latitude

Firm location	26°61′25″N	33°01′58″N	36°28′11″N	39°70′00″N
	80°06′78″W	79°92′97″W	80°06′03″W	79°91′67″W
Constant	5.96[c]	6.04	6.09	6.12
	(52.63)	(46.29)	(43.55)	(40.82)
Sulfur content	−0.05	−0.06	−0.07	−0.07
	(−2.35)	(−3.16)	(−3.55)	(−3.36)
BTU	1.74	1.82	1.88	1.91
	(22.77)	(20.61)	(19.91)	(19.04)
Firm location	30°56′58″N	37°36′36″N	41°64′31″N	46°56′94″N
	87°22′39″W	87°12′14″W	87°[1]12′28″W	87°39′33″W
Constant	5.81	6.09	6.25	6.33
	(49.46)	(40.46)	(39.91)	(46.08)
Sulfur content	−0.04	−0.07	−0.05	−0.06
	(−1.89)	(−3.40)	(−2.57)	(−2.75)
BTU	1.65	1.89	1.99	2.04
	(21.26)	(19.08)	(19.36)	(22.82)
Firm location	30°61′67″N	31°82′06″N	41°33′00″N	46°29′00″N
	96°07′78″W	96°07′78″W	95°94′67″W	96°04′28″W
Constant	5.51	5.46	6.22	6.29
	(48.39)	(45.70)	(35.68)	(46.24)
Sulfur content	−0.01	0.00	−0.01	−0.03
	(−0.30)	(−0.07)	(−0.55)	(−1.46)
BTU	1.46	1.43	1.98	2.01
	(20.48)	(19.07)	(17.57)	(23.42)

Notes: The terms in parentheses refer to the *t*-value.
BTU refers to the heat content of coal measured in British thermal units; and coal sulfur content captures coal volatility.

with 112° 57′W in Richfied, UT (West of Salt Lake City, UT).

Overall, the spatial kernel permits the data to select the extent of the market to be included in regression using a MSE criterion where the least squares estimator is a special case that assumes a single market for all observations. In modelling the valuation of fuel quality made by power plants, the estimated bandwidth length reveals the importance of heterogeneous local markets. Future work may further explore performance of the spatial kernel estimator in other applications, e.g. measuring the value of housing attributes and environmental amenities.

References

Case, A. (1991) Spatial patterns in household demand, *Econometrica*, **59**, 953–66.

Engle, R. F., Granger, W. J., Rice, J. and Weiss, A. (1986) Semiparametric estimates of the relation between weather and electricity sales, *Journal of the American Statistical Association*, **81**, 310–20.

Florkowski, W. and Sarmiento, C. (2005) The examination of pecan price differences with spatial correlation estimation, *Applied Economics*, forthcoming.

Haining, R. (1990) *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, Cambridge.

Marsen, J. E. and Tromba, A. J. (1988) *Vector Calculus*, W.H. Feeman and Company.

Rosen, S. (1974) Hedonic prices and implicit prices, *Journal of Political Economy*, **82**, 34–55.

Sarmiento, C. (2004) Modeling firm heterogeneity with spatial trends, *Applied Economics Letters*, **11**, 271–74.

Sarmiento, C. (2005) A time varying coefficient approach to global flexibility in demand analysis: a semiparametric approximation, *American Journal of Agricultural Economics*, forthcoming.

Schmalensee, R. and Stoker, T. (1999) Household gasoline demand in the U.S., *Econometrica*, **67**, 645–62.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

Wang, Z. (2003) Hedonic price of crude oil, *Applied Economics Letters*, **10**, 857–61.