Applying Geographically Weighted Regression to Conjoint Analysis: Empirical Findings from Urban Park Amenities

Katsuya Tanaka
Research Center for Sustainability and Environment
Shiga University
1-1-1 Bamba, Hikone, Shiga 522-8522 Japan
Phone/Fax: +81 (749) 27-1154
Email: katsuyat@gmail.com

Kentaro Yoshida
Graduate School of Systems and Information Engineering
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573 Japan
Phone/Fax: +81 (29) 853-5220
Email: yoshidak@sk.tsukuba.ac.jp

Yasushi Kawase
Hitachi High-Technologies Corporation
1-24-14 Nishi-shimbashi, Minato, Tokyo 105-8717 Japan
Phone: +81 (3) 3504-7111
Email: clickon_clickoff@ybb.ne.jp

Applying Geographically Weighted Regression to Conjoint Analysis:

Empirical Findings from Urban Park Amenities

by

Katsuya Tanaka, Kentaro Yoshida, and Yasushi Kawase

Abstract

The objective of this study is to develop spatially-explicit choice model and investigate its validity and applicability in CA studies. This objective is achieved by applying locally-regressed geographically weighted regression (GWR) and GIS to survey data on hypothetical dogrun facilities (off-leash dog area) in urban recreational parks in Tokyo, Japan. Our results show that spatially-explicit conditional logit model developed in this study outperforms traditional model in terms of data fit and prediction accuracy. Our results also show that marginal willingness-to-pay for various attributes of dogrun facilities has significant spatial variation. Analytical procedure developed in this study can reveal spatially-varying individual preferences on attributes of urban park amenities, and facilitates area-specific decision makings in urban park planning.

Key words:

Choice experiments; conjoint analysis; dogrun; geographically weighted regression; spatial econometrics

_____

Tanaka is Associate Professor in Research Center for Sustainability and Environment, Shiga University, Japan. Yoshida is Associate Professor in Graduate School of Systems and Information Engineering, University of Tsukuba, Japan. Kawase is an employee of Hitachi High-Technologies Corporation, Japan.

1. Introduction

Most existing valuation studies using conjoint analysis (CA) have paid limited attention on spatial stability of choice models. If economic, social, and physical characteristics are different over space, individuals in different locations may exhibit different preferences on specific attribute of environment or natural resources. Although recent models (e.g. mixed logit and latent class models) take unobserved heterogeneity into account, they do not address spatial heterogeneity in explicit manner. Furthermore, the estimated parameters using the mixed logit models are meaningful if and only if the standard deviations of disturbances are constant for all observations. However, a number of studies reports that this condition is rarely satisfied (Louviere, Eagle, and Cohen 2005).

The objective of this study is to develop spatially-explicit choice model and investigate its validity and applicability in CA studies. This objective is achieved by applying locally-regressed geographically weighted regression (GWR) and GIS to survey data on hypothetical dogrun facilities (off-leash dog area) in urban recreational parks in Tokyo, Japan. Our results show that spatially-explicit conditional logit model developed in this study outperforms traditional model in terms of data fit and prediction accuracy. Our results also show that marginal willingness-to-pay for various attributes of dogrun facilities has significant spatial variation. Thus, this approach can reveal spatially-varying individual preferences on attributes of urban park amenities, and facilitates area-specific decision makings in urban park planning.

2. Empirical Procedure

The model

Geographically weighted conditional logit (GWCL), the model used in this study, is an extension of geographically weighted regression (GWR) developed by Fotheringham, Brunsdon, and Charlton (2002). First consider first a global (conventional) conditional logit (CL) model. Probability that individual *i* chooses choice *j* among *J* alternatives ($j = 1, 2, ..., J$) is defined by:

$$\Pr\left(y_i = m \mid \mathrm{x}_i\right) = \frac{\exp\left(\mathrm{x}_{im}\beta\right)}{\sum_{j=m}^{J} \exp\left(\mathrm{x}_{im}\beta\right)} \tag{1}$$

where y_i is a choice made by individual *i*, x_{im} is a vector of attributes, and β is a vector of coefficients. Under the global model settings in equation (1), it is assumed that all individuals place the same attitude (i.e. same coefficient) on each of attributes. However, this assumption

may be too strong to derive relationship between dependent and independent variables. It is more reasonable to assume that such relationship is changing over space.

The GWR approach extends this conventional discrete choice framework by allowing local rather than global coefficients to be estimated. The GWR methodology is a modified version of locally linear regression methods introduced by McMillen (1996) and is developed by Brunsdon, Fotheringham, and Charlton (1996). The GWR methodology is the use of distance-weighted sub-samples of the data to produce locally linear regression estimates for every point in space. Each set of parameter estimates is based on a distance-weighted sub-sample of "neighboring observations," which has a great deal of intuitive appeal in spatial econometrics (LeSage 1999). Although the GWR methodology is relatively new approach in spatial econometrics, it has gained rapidly increasing attention and has been applied to number of studies in applied economics (for some most recent applications, see Bitter 2007; Cho 2007; Partridge 2007).

This study extends the GWR methodology and applies to discrete-choice conditional logit model settings (geographically weighted conditional logit; GWCL hereafter). In GWCL model, probability that individual in location i chooses choice j among J alternatives can be defined by:

$$\Pr\left(y\left(u_i,v_i\right)=m \mid \mathrm{x}_i \right) = \frac{\exp\left(\mathrm{x}_{im}\beta\left(u_i,v_i\right)\right)}{\sum_{j=m}^{J}\exp\left(\mathrm{x}_{im}\beta\left(u_i,v_i\right)\right)} \tag{2}$$

where $y\left(u_i,v_i\right)$ denotes an individual locating in ith point in space and $\beta\left(u_i,v_i\right)$ is the vector of parameter coefficients for the ith location. Typically, $\left(u_i,v_i\right)$ is recorded as latitude and longitude or x and y coordinates in location i. In the GWCL model, a log-likelihood function for location i is defined as follows:

$$\ln LL\left(u_i,v_i\right) = \sum_i \sum_j d_{ij}\ln\frac{\exp\left(\mathrm{x}_{im}W\left(i\right)\beta_i\right)}{\sum_{j=m}^{J}\exp\left(\mathrm{x}_{im}W\left(i\right)\beta_i\right)} \tag{3}$$

where $W\left(i\right)$ is a spatial weight matrix for location $\left(u_i,v_i\right)$. This weight matrix assigns weights based on their spatial proximity to location i in order to account for the fact that an observation near location i has more of an influence in the estimation of the $\beta_k\left(u,v\right)$ than do observations located farther from i.

$$W(i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix} \tag{4}$$

where w_{ij} is the weight given to data point j in the calibration of the model for location i. The diagonal elements of the weight matrix, w_{ij}, are equal to:

$$w_{ij} = \begin{cases} \left[1 - \left(d_{ij}/b \right)^2 \right]^2 & \text{if } d_{ij} < b \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where d_{ij} is the Euclidean distance between point i and j and b is a chosen bandwidth. At the regression point i, the weight of the data point is unity and falls to zero when the distance between i and j equals the bandwidth or higher.

As b tends to be infinity, w_{ij} approaches 1 regardless of d_{ij} in which case the parameter estimates become uniform and locally weighted regression is equivalent to OLS. Conversely, as b becomes smaller, the parameter estimates will increasingly depend on observations in close proximity to location i and hence have increased variance. This study chooses bandwidth which minimizes the Akaike Information Criterion (AIC). AIC takes the following form:

$$\text{AIC} = n\log\left(\hat{\sigma}^2 \right) + n\log\left(2\pi \right) + n\left(\frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \right) \tag{6}$$

where n is the number of observations, $\hat{\sigma}^2$ is the variance of residuals, and $\text{tr}(S)$ is the trace of hat matrix (Fotheringham, Brunsdon, and Charlton 2002). The Golden Selection method is used for the function minimization process. The model presented above is estimated using MATLAB[1].

Experimental design

To investigate the validity of geographically weighted conjoint analysis, this study uses the dog-run facility data collected by Yoshida and Kawase (under review). They conducted a survey regarding hypothetical dogrun (off-leash dog area) facility development in Tokyo metropolitan area. In November 2007, they conducted a survey in Komazawa Olympic park (413,000 square

meters), one of the most major recreational parks in Tokyo and where one of the largest dog-run facility (1,200 square meters) is available.
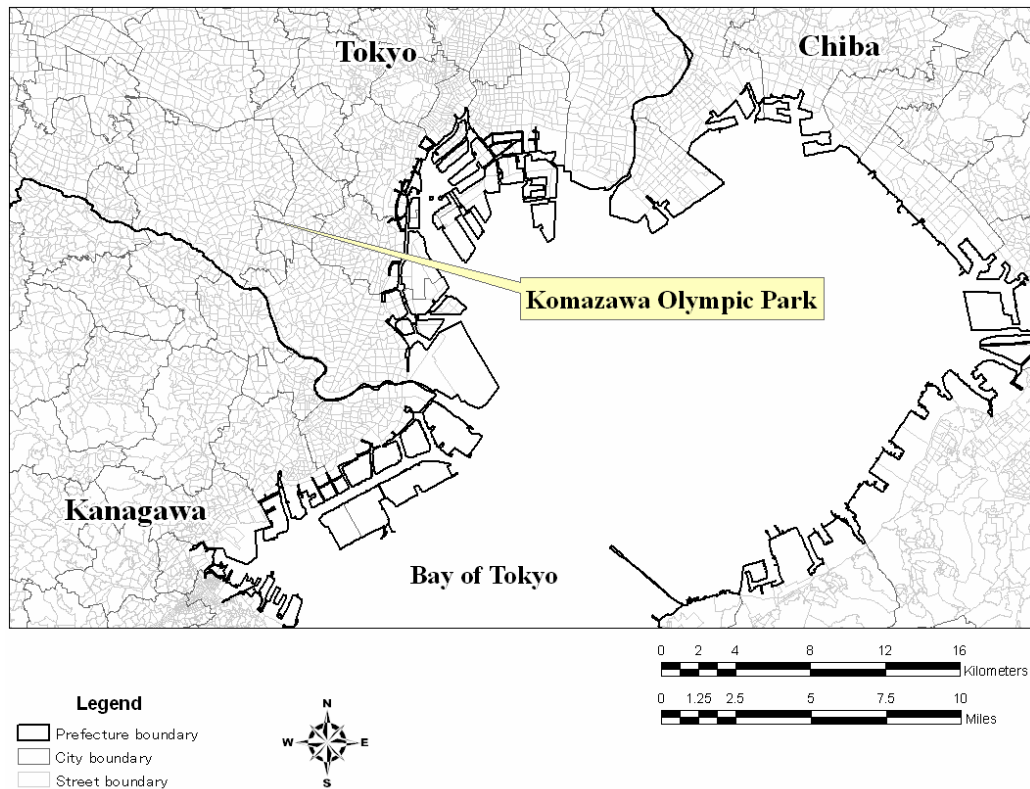

Figure 1.Study area

Questionnaire contains four different scenarios of the hypothetical dog runs to be developed in existing urban parks. Scenarios are different in number of attributes, including (1) field size, (2) ground material, (3) presence of safeguards, (4) availability of water fountain and lighting equipments, (5) distance from respondent, and (6) entrance fee. Questionnaire asked respondents to choose one of four alternative scenarios including status quo. After four sets of hypothetical questions, questionnaire asked characteristics of respondent (e.g., gender, age, dog type, housing facility, and seven-digit postal code) and current usage of dog runs currently available in Tokyo. A total of 4 choice experiments was asked to each respondent to choose most desirable development scenario among 4 alternatives including status quo (no further development).

Of sample size of 400 respondents in Tokyo and its suburbs, 177 respondents in Tokyo

and its neighboring prefectures (Kanawaga and Tokyo) completed and returned questionnaires (44% response rate). After removing incomplete choice experiments, a total of 639 observations was collected. Although exact addresses of respondents are not known, their approximate spatial locations were derived from seven-digit zip code asked in questionnaires. Using this information, GIS operations were then performed to calculate centroid (geographical center) of each seven-digit zip code boundary parcel and defined as approximate location of each respondent (see figure 2 for boundaries in Tokyo).

Figure 2. Seven-digit zip code boundaries in Tokyo

3.   Empirical Results

Table 1 reports estimated coefficients for CL and GWCL models. Overall, the models fit the data reasonably well. Results from the CL model indicate the most of the independent variables have expected signs with statistical significance at 5 percent or greater. In addition, the model correctly predicts 74.7 percent of actual choice. Table 2 indicates that we can benefit from estimating the model locally, rather than globally. Estimated coefficients for the GWCL model

are different over space, assigning unique values on each location in space. The mean coefficients of most attributes are very close to coefficients from the CL model. They have spatial variation. The GWCL model correctly predicts more than 80 percent of actual choice behaviors. Furthermore, the reduction in the Akaike Information Criterion (AIC) from 671.47 in the CL model to 613.50 in the GWCL model indicates that the GWCL has better model fit after accounting for the difference in degree of freedom. A Monte Carlo significant test suggests that spatial variations in coefficients are significant for ASC (alternative-specific constant), AREA (size of dogrun), and DIST (distance to dogrun from respondent's resident).

Table 1. The estimated coefficients for the CL and GWCL models[+]

	CL			GWCL			
Variable	Coefficient	Std. Error	Sig.[++]	Mean	Min.	Max.	Std. Dev.
ASC	-1.0625 ***	0.1719	**	-1.0530	-0.9401	-1.3113	0.0595
AREA	0.0003 ***	0.000118	**	0.0003	0.0004	0.0002	0.0000
LAWN	1.4585 ***	0.3641		1.3964	2.0307	1.2617	0.1067
CHIP	0.7515 **	0.3826	*	0.6203	1.1289	0.5058	0.1151
SAND	-1.2120 **	0.5072	*	-1.1949	-1.0611	-1.7245	0.1100
CLAY	-0.2485	0.4993		-0.1728	0.1497	-0.3872	0.0914
EVERYDAY	0.7359 **	0.3066		0.6641	1.2514	0.6056	0.0527
WEEKEND	0.5962 *	0.3481		0.6269	1.0094	0.4083	0.0387
MANY	0.9741 ***	0.3078		0.9738	1.4180	0.7571	0.0804
FEW	0.2013	0.3157		0.2435	0.3259	0.0359	0.0518
WATER	0.5354 **	0.2662	*	0.5654	0.6947	0.2459	0.0710
DIST	-0.0009 ***	0.000201	**	-0.0010	-0.0007	-0.0011	0.0001
FEE	-0.0038 ***	0.000645	*	-0.0037	-0.0035	-0.0044	0.0001
n	639			639			
AIC	671.47			613.50			
% correctly predicted	74.70%			80.28%			

+ One, two, and three asterisks indicate statistical significance at 10%, 5%, and 1% levels.
++ Monte Calro significance test for spatial non-stationary

Table 2 shows the estimated marginal willingness to pay (MWTP) for each dog-run attributes. From the CL model, MWTP for AREA and DIST are estimated to be 0.08 yen per square meter and -0.238 yen per meter, respectively[2]. Although the mean MWTP's of these attributes in the GWCL are very close (0.083 yen and 0.294, respectively), the model predicts the MWTP's are significantly different over space. Figure 3 shows spatial distribution of MWTP for

AREA. The MWTP is estimated to be relatively high in respondents further away from dog run in Komazawa Olympic park. In particular, the MWTP tends to be high for respondents in population dense areas. The highest MWTP is observed for an area nearest to downtown of Tokyo. This is expected because in population dense areas, individuals tend to have less space for their dogs. They thus should have higher WTP for dogrun with greater fields.

Table 2. The estimated MWTP for dog-run attributes

|	CL	GWCL			
Variable	Coefficient	Mean	Mix.	Max.	Std. Dev.
AREA	0.080	0.083	0.051	0.094	0.004
LAWN	380.809	376.793	343.877	459.652	17.629
CHIP	196.214	166.997	141.056	291.176	25.899
SAND	-316.449	-322.325	-415.050	-291.154	20.480
CLAY	-64.883	-47.082	-105.987	35.841	24.587
EVERYDAY	192.141	179.554	137.077	360.221	15.945
WEEKEND	155.666	169.603	92.425	290.554	12.947
MANY	254.334	263.048	217.921	362.513	19.463
FEW	52.559	66.119	9.126	92.322	14.873
WATER	139.791	153.235	55.653	188.154	21.359
DISTANCE	-0.238	-0.266	-0.294	-0.149	0.021
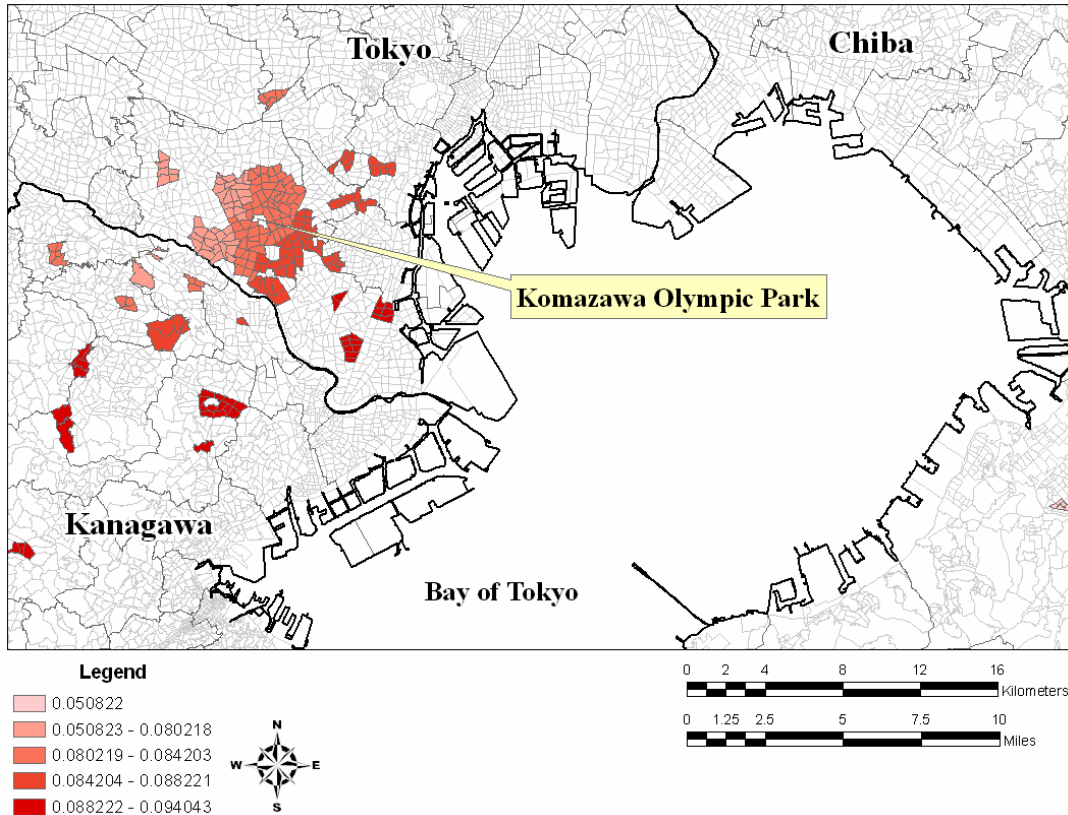
Figure 3. Spatial distribution of the MWTP for AREA

As figure 4 depicts, a very similar spatial pattern is found in MWTP for DIST. In general, respondents further away from dogrun in the park. The MWTP is particularly high for respondents locating in areas close to Tokyo downtown. Figure also shows that respondents in Kanagawa tend to have smaller MWTP although their location is relatively further away from dogrun in the park. This can be explained by the fact that alternative dogrun is relatively more available than respondents in Tokyo downtown area. Overall, spatial patterns of MWTP for attributes of dogrun facilities are reasonable and consistent with our intuition. More accurate and detailed patterns can be derived if the data includes more observations in different locations.
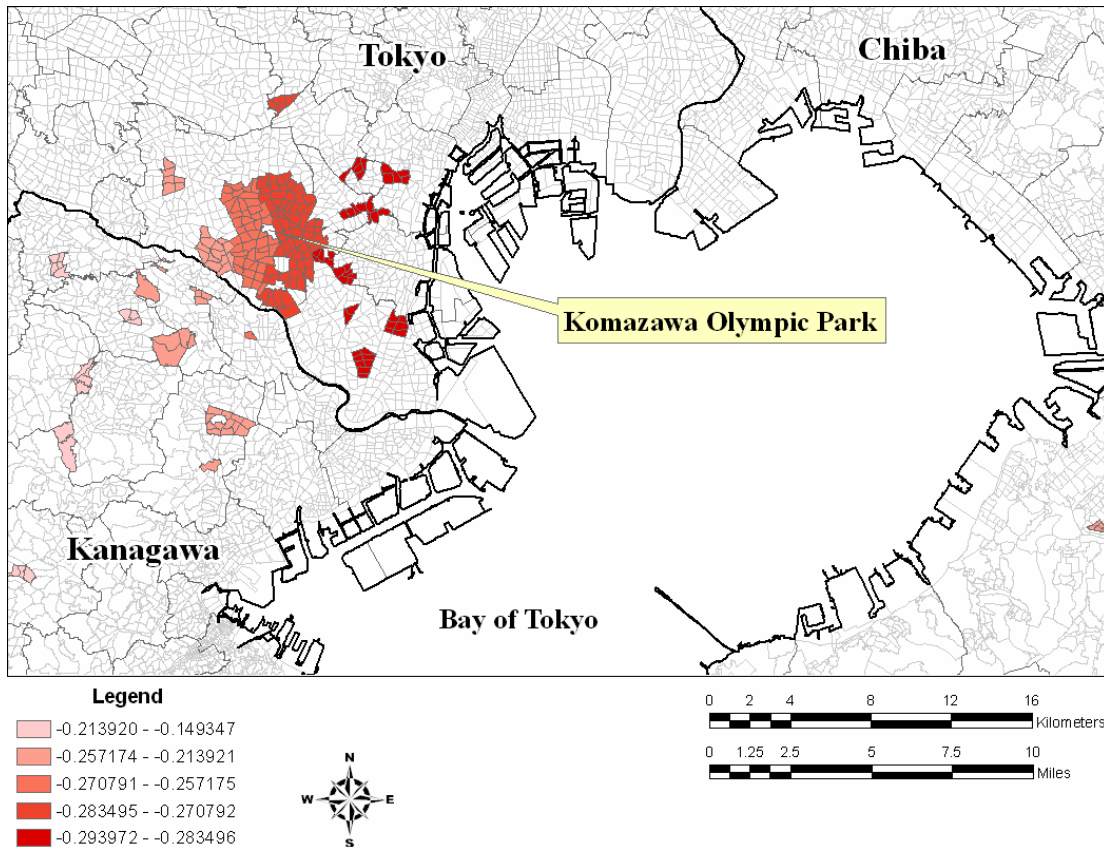
Figure 4. Spatial distribution of the MWTP for DIST

4. Summary and Conclusions

This study developed spatially-explicit choice model and investigate its validity and applicability in CA studies. This objective is achieved by applying locally-regressed geographically weighted regression (GWR) and GIS to survey data on hypothetical dogrun facilities (off-leash dog area) in urban recreational parks in Tokyo, Japan. Our results show that spatially-explicit conditional logit model developed in this study outperforms traditional model in terms of data fit and prediction accuracy. Our results also show that marginal willingness-to-pay for various attributes of dogrun facilities has significant spatial variation. Analytical procedure developed in this study can reveal spatially-varying individual preferences on attributes of urban park amenities, and facilitates area-specific decision makings in urban park planning.

This study can be improved in terms of its validity and effectiveness if data includes more variation. As mentioned in prior section, spatial variation of the estimated parameters (and thus

MTWP) can be more detailed and accurate if the data includes greater observations in different locations. Further studies are needed to facilitate application of geographically weighted discrete choice modeling to conjoint analysis to derive spatially varying local valuation of urban park amenities.

Notes

[1] For this analysis, MATLAB version R2007b is used. To derive spatial weights matrix and perform parameter estimation, two optional toolboxes (optimization and statistics) are also used.

[2] At the time of survey, one Japanese yen (JPY) is equivalent to 0.009 U.S. dollars and 0.006 Euro, respectively.

Reference

Bitter, C., G.F. Mulligan, D. Sandy. "Incorporating Spatial Variation in Housing Attribute Prices: A Comparison of Geographically Weighted Regression and the Spatial Expansion Method," *Journal of Geographical Systems*. 9:7-27, 2007.

Brunsdon C., A.S. Fotheringham, and M.E. Charlton. 1996. "Geographically weighted regression: A method for exploring spatial nonstationarity". *Geographical Analysis*. 28: 281-98.

Cho, S-H, Z. Chen, S.T. Yen, and B.C. English. "Spatial Variation of Output-input Elasticities: Evidence from Chinese County-level Agricultural Production Data," *Papers in Regional Science*, 86(1): 139-157. 2007.

Cleveland, S.W. 1979. "Robust Locally Weighted Regression and Smoothing   catterplots." *Journal of American Statistical Association* 74:829–836.

Fotheringham, A.S., C. Brunsdon, and M. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, West Sussex: John Wiley & Sons.

LeSage, J.P. 2004. "A Family of Geographically Weighted Regression." In L. Anselin, R. J. G. M. Florax, and S. J. Rey, eds. *Advances in Spatial Econometrics*. London: Springer, pp. 241–264.

Louviere, J.J., T.C. Eagle, and S.H. Cohen. "Conjoint Analysis: Methods, Myths and Much more," *CenSoC Working Paper*. No.05-001, 2005.

McMillen, D.P. "One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach," *Journal of Urban Economics*, 40:100-124. 1996.

Partridge, M.D. and D.S. Rickman. "Persistent Pockets of Extreme American Poverty and Job Growth: Is There a Place-based Policy Role?" *Journal of Agricultural and Resource Economics*, 32(1):201-224. 2007.

Yoshida, K. and Y. Kawase, "A Choice Modeling approach to the Improvement of Dogrun Facilities in Tokyo Metropolitan Parks," *Under review*.