AN EMPIRICAL EVALUATION OF HEDONIC REGRESSION MODELS

Xiaolu Gao[a, *], Yasushi Asami[b], and Chang Jo Chung[c]

[a] JST Domestic Research Fellow, Dept. of Urban Research, National Research Institute for Land and Infrastructure Management. Tachihara 1, Tsukuba, Ibaraki 305-0802, JAPAN - gao-x925t@nilim.go.jp.
[b] Center for Spatial Information Science, the University of Tokyo. Hongo 7-3-1, Bunkyo, Tokyo 113-8656, JAPAN - asami@csis.u-tokyo.ac.jp.
[c] Spatial Data Analysis Lab., Geological Survey of Canada, Natural Resources Canada. 601 Booth Street, Ottawa, Ontario, CANADA K1A 0E8 - chung@gsc.nrcan.gc.ca.

Commission IV, WG IV/1

KEY WORDS*:* Empirical evaluation, Cross-validation, Prediction, Spatial model

ABSTRACT:

Statistical criteria are poor at evaluating spatial exploratory models of hedonic regression because they are heavily dependent on uncertain assumptions concerning spatial relationships. To resolve this problem, empirical evaluation methods are proposed in this paper. A simple linear model and three spatial models for a housing and land price dataset in Tokyo are studied for illustration. The prediction power of models is emphasized. With a cross-validation technique, the housing and land price at each sample point are predicted, then, numerical and graphical criteria are supposed using these predicted values and real observed prices. The methods not only help us to select suitable models for a dataset, but also provide an alternative for the significance test of concerned spatial relationships.

1. INTRODUCTION

Recently, a variety of hedonic regression models have been proposed in addition to a simple formed model to study the spatial nature of variables. All of them proposed to make use of the spatial characteristics of variables to improve models. Since they are often exploratory models, it is critical issue as how well a model is, whether the spatial characteristics identified by a model are convincible, and how to select the most suitable model for a dataset.

A difficulty in making such judgment is the lack of appropriate evaluation techniques. This paper sets out to address this issue.

Traditionally, statistical testing criteria such as *R*-square, *t*-statistic, *F* statistic, AIC (*Akaike Information Criterion*, a tradeoff of likelihood and the number of estimated parameters), and so on, are often used to evaluate regression models. However, models are always based upon such assumptions as linear assumption, normality, or non-collinearity, thus statistical testing methods inevitably depend on these assumptions. For exploratory models, this creates problems because assumptions are open for test.

To overcome this problem, other evaluation methods have to be developed.

2. ALTERNATIVE REGRESSION MODELS

In this paper, we use a housing and land price dataset in Tokyo for illustration. It covers the transaction price, structural attributes, environmental attributes, and *x*, *y* coordinates of 190 detached housing lots. See Appendix A for details of it.

2.1 A Simple Linear Regression Model

Equation (1) shows the simplest form of hedonic regression models:

$$y = a_0 + \sum a_k x_k + \varepsilon, \qquad (1)$$

where, y is unit price, x_k for k= 1, 2, …, m are independent variables, ε is an error term, and a_0 and a_k are parameters to be estimated.

An ordinary linear model in this form was developed in Gao and Asami (2001). It explains 75.6% of unit price, has 16 independent variables, all of which statistically significant at the level of F below 0.05. See Appendix B.

Since this model has not fully considered the impact of spatial location on unit price, to study the spatial characters of dataset with location data is thought of a way to improve it.

One of the spatial relationships being explored is spatial dependency. Another is spatial heterogeneity. The estimates of regression parameters in the presence of spatial dependency have been discussed by a number of literatures, *e.g*., Dubin (1992; 1998) and Can (1990; 1992), while, various localized modeling techniques were proposed to capture spatial heterogeneity (Casetti, 1972; Getis and Ord, 1992; Fotheringham and Brunsdon, 1999). In addition, some models are developed to detect both of them (Anselin, 1988; Can, 1992).

* Corresponding author.

In this paper, we choose three spatial regression models to explore the data.

2.2 A Spatial Dependency Model

In order to study spatial dependency effects, a *prior probability* method proposed by Switzer *et al.* (1982) is applied. The model had been used to process satellite classification maps. To evaluate the class at the center of a window using the data for that location and the prior probability estimates obtained from the nearby observations in the window area was shown to have increased classification accuracy. Similarly, the prior characteristic terms of sample lots, denoted by x_k' for k= 1, 2, …, m, are added to the simple linear model:

$$y = a_0 + \sum a_k x_k + \sum a_k' x_k' + \varepsilon. \quad (2)$$

For simplification, let x_k' take the value of x_k that is associated with the nearest neighbor of a sample. Appendix C gives the OLS estimates of this model, with an R-square of 0.801.

Although one might argue that correlations between x_k and x_k' may lead to unstable estimates, the question is secondary to the concern whether spatial dependency is present. If the absence of spatial dependency effects were demonstrated, the model itself would be improper, let alone its estimates.

2.3 A Geographically Weighted Regression (GWR) Model

To identify spatial variations in relationships, a GWR model proposed in Fotheringham, *et al.* (1998) is adopted. The model extends traditional regression framework by allowing parameters to be estimated locally so that the model in Equation (1) is rewritten as

$$y_i = a_{0i} + \sum a_{ki} x_{ki} + \varepsilon_i, \quad (3)$$

where, a_{0i} and a_{ki} represent the values of a_0 and a_k at point i. In order to estimate the model, an observation is weighted in accordance with its proximity to point i. Let W_i be an $n \times n$ matrix whose diagonal elements w_{ij} denotes the *geographical weighting* of all observed data for point i, and the off-diagonal elements are zero. Data from observations close to i are weighted more than data from observations far away. Equation (4) gives the estimation of a_i:

$$\hat{a}_i = (X^T W_i X)^{-1} X^T W_i Y , \quad (4)$$

where, X and Y are the matrix of explanatory variables and unit prices, respectively.

A weighting function in Equation (5) is applied:

$$w_{ij} = \exp(-\frac{d_{ij}^2}{\beta^2}). \quad (5)$$

where, d_{ij} is the Euclidean distance between point i and j, β is a bandwidth. By minimizing (6), β is set to 1,250m.

$$\sum_i [y_i - \hat{y}_{\neq i}(\beta)]^2 . \quad (6)$$

Note that $\hat{y}_{\neq i}(\beta)$ is the estimates at i with samples near to i but not i.*

At the location of each of the 190 observations, regressions are run. The obtained localized parameter estimates exhibit a high degree of variability over space and demonstrate fairly complex spatial patterns of each variable.

The localized regressions result in high R-square fittings, beyond 0.95 at all sample points. Apparently, it is the outcome of localized regression, so these values are not comparable to that of the simple linear model.

2.4 A Spatial Dependency +GWR Model

In the following, we use a specification in (7) to investigate whether spatial dependency and heterogeneity effects are both present.

$$y_i = a_{0i} + \sum a_{ki} x_{ki} + \sum a_{ki}' x_{ki}' + \varepsilon_i, \quad (7)$$

Similar to the spatial dependency model, let x_{ki}' represent the corresponding values of the nearest sample to i.

In addition, regressions are localized at each sample point with a GWR technique, adopting a weighting scheme of (5), and the optimal value of β being 1650m. The regression yields the estimates of parameters at each sample points.

3. AN EMPIRICAL EVALUATION OF THE MODELS

Now, let us discuss how to select a best one from these models. The weakness of "black-box" statistical evaluation method is well aware of. For instance, models supported by high R-square fitting or satisfying significance level are frequently doubted because of weak assumptions or irrational estimates of parameters, thus a clear conclusion can hardly be reached with these statistical criteria. This is an unavoidable result of heavy dependency on statistical assumptions. The problem is especially distinct in spatial exploratory models for the assumptions of models are to a large degree uncertain. This prompts us to turn to an empirical evaluation approach. In particular, the methods being presented in this paper focus on the prediction power of models.

In the practice of hedonic regression, we often focus more on estimating unbiased coefficients to identify the marginal

---

* Suppose that point i itself is also included to estimate a_{0i} and a_{ki}. If β is very small, the weightings of all points except for i become negligible and the estimates will fluctuate wildly throughout space.

prices of explanatory variables, rather than simply see the prediction powers. Nonetheless, when spatial relationships are introduced to a simple ordinary model, we instinctively expect that this can improve the fitting of the model. Thus conversely, if it does not work better at predicting prices, we may just be satisfied with the simple model and think that its estimates are robust. From this viewpoint, the improvement of fittings can be been seen as an alternative to test the significance of spatial relationships, too. (Can, 1992; Dubin, 1992 and 1998; Brunsdon, *et al.*, 1999)

A cross-validation approach is used to carry out the idea, *i.e.*, to see how well the prices already being observed can be predicted with the rest samples. An application of a similar method can be seen in Bourassa *et al.* (2001), where, random 80% samples are modeled to predict the rest 20% samples, and the sum of square of error is used as a criterion to evaluate proposed model. However, we are concerned that, to estimate a hedonic regression model, it is important to keep as large a sample size as possible. Thus, we predict the price at every point with all the rest 189 samples.

3.1 Numerical Cross-validation Criteria

First, consider some numerical criteria, for instance, $\Sigma(y_i - \hat{y}_{\neq i})^2$ and $\Sigma|y_i - \hat{y}_{\neq i}|$ *, where, y_i and $\hat{y}_{\neq i}$ denote the observed and the predicted unit price at point i, respectively.

Table 2 gives their calculations. The values of all the three spatial models are slightly lower than that of the simple linear model. However, the falloffs are too small to suggest any critically large improvements.

Model	Numerical criteria	
	$\Sigma(y_i - \hat{y}_{\neq i})^2$	$\Sigma\lvert y_i - \hat{y}_{\neq i}\rvert$
Simple linear model	1.96	15.29
Spatial dependency model	1.92	15.11
GWR model	1.81	14.60
Spatial dependency +GWR model	1.76	14.33

Table 2. A comparison of models by numerical criteria

Centering on prediction error, more numerical criteria could be raised. Seriously though, numerical criteria just show a general aspect of model. Thus, some graphical criteria are further employed in order to illustrate broader and more complicated aspects of the relationships between model and data.

3.2 The Distribution of Observation and Prediction

On a scatter plot of observed unit price (y_i) and predicted unit price ($\hat{y}_{\neq i}$), the prediction power of a model is indicated by the range occupied by the scatter points, in particular, the short breadth of this range. Obviously, all points being on 45° diagonal line indicates a perfect prediction, and higher prediction power is suggested by how slim and how close the range is to the 45° line.

Following the idea, the three spatial models are compared with the simple linear model in Fig. 1~3. The 95% confidence regions of the scatter points are plotted, respectively. As results, again, all the spatial models are no much better than the simple linear model for this dataset.

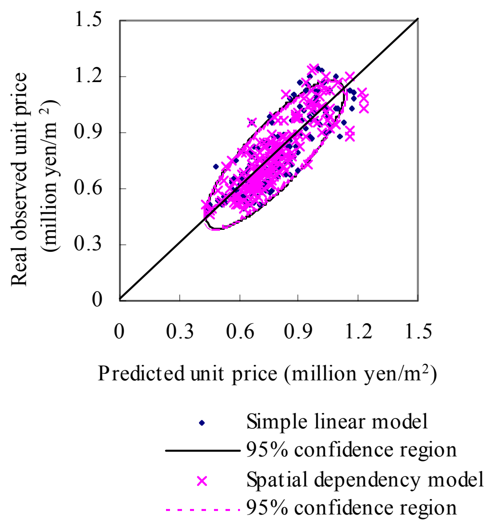

Figure 1. Simple linear model and spatial dependency model

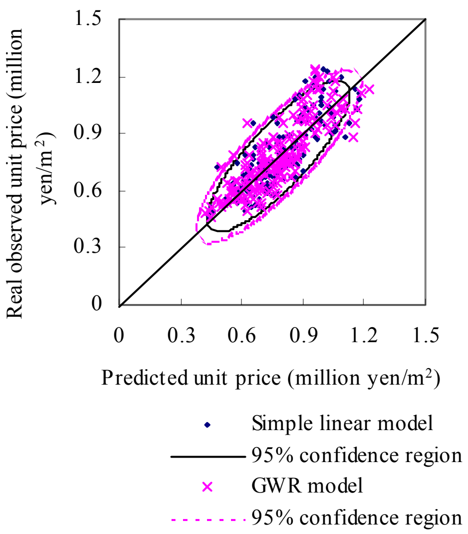

Figure 2. Simple linear model and GWR model

* This is deemed more robust than $\Sigma(y_i - \hat{y}_{\neq i})^2$.

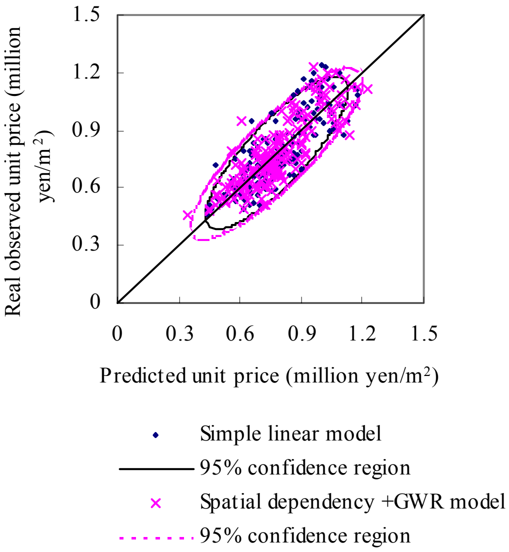

Figure 3. Simple linear model and spatial dependency +GWR model

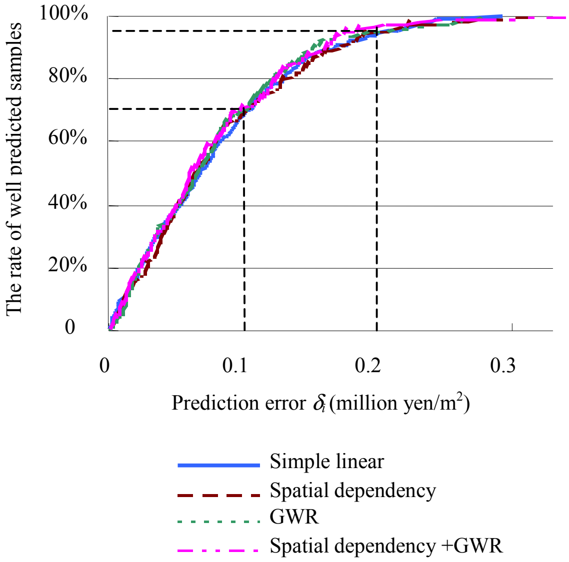

Figure 5. A comparison of prediction rate curves

3.3 Prediction Rate Curve

We further evaluate the models by the proportion of well-predicted samples with respect to prediction error, which is referred to as prediction rate.

To do this, let $\delta_i = \left| y_i - \hat{y}_{\neq i} \right|$ and sort δ_i. Then plot the ratio of well-predicted samples with sorted δ_i as horizontal axis. By this way, the prediction rate curve of a model is obtained.

The introduction of this criterion allows us to capture the comprehensive performance of a model. In particular, it is helpful when we select models under a given level of prediction error. Consider two models having prediction rate curves shown in Fig. 4. Highly possibly, the sums of prediction errors of them are close. At tolerance level ξ_1, 70% of samples are correctly predicted by model 1, while only 60% can be predicted by model 2. However, if tolerance ξ_2 is acceptable, more samples are correctly predicted by model 2.

From an empirical point of view, a dataset always has some poor-to-predict observations. Thus, the performance of the prediction curves within a tolerable error scope may be thought of more important.

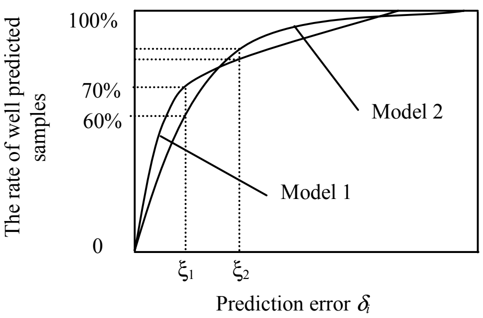

Figure 4. Evaluating models with prediction rate curves

Prediction rate curve actually indicates the density of the scatter points in Fig. 1~3 with respect to their distance to the 45° diagonal line. To illustrate, let us see the prediction rate curves of the simple linear model and the spatial models in Fig. 5.

The four curves twist together for most of the span. This implies that their performances are fairly alike. All of them predict the unit price of about 70% of samples if a tolerance of 0.1 million yen/m^2 is accepted. Within a tolerance of 0.2 million yen/m^2, about 95% are correctly predicted. With the prediction error being larger than 0.2 million yen/m^2, only about 5% of the sample can be predicted.

On the other hand, it might have been noted that the tail of the simple linear model winds up a bit earlier than the other models. This is able to explain why this model looks better than the others in Fig. 1~3. Nevertheless, we can still draw the conclusion that they are similar, allowing that the tail part is not very important.

From the above empirical tests, it seems proper to say that, for this dataset, the proposed spatial dependency model, GWR model, and the mixed model of the two are not significantly better than a simple linear model. In other words, the simple linear model is fairly satisfactory. Moreover, we should not stick to the spatial relationships revealed by the above spatial models.

The failure of these three spatial models might either result from inappreciable spatial relationships in the sample area, or result from inappropriate model specifications. In the latter case, alternative specifications might be constructed and tested with similar empirical techniques as described above.

4. CONCLUSION

In the sense that empirical criteria are not dependent on statistical assumptions, they fit for any dataset and their application is not restricted to a certain kind of models.

Nonetheless, these methods are especially outstanding to evaluate spatial exploratory models of hedonic regression, where assumptions concerning the presence of spatial relationships are uncertain. In such models, it is quite reasonable to believe that the additional consideration on spatial relationships should improve the prediction power of an ordinary model. Accordingly, we can simply focus on the prediction power of models. This is plain and straightforward. These tests also provide an alternative for the significance test of spatial relationships. If a spatial model does not outperform an ordinary model, we may just think that the estimates of the latter are robust enough.

Much extension can be made based on the empirical evaluation concept. For example, we can judge the contribution of a variable by comparing the prediction powers of models with and without the variable, or look for an optimal tolerance level for a dataset, and so forth. The empirical approach might become effective alternatives for statistical evaluation methods.

REFERENCES

Anselin, L., 1988. Lagrange multiplier test diagnostics for spatial dependence and heterogeneity. *Geographical analysis*, 20, pp. 1-17.

Bourassa, S. C., M. Hoesli, and V. S. Peng, 2001. Do housing submarkets really matter? *AsRes Sixth Annual Conference*, August 2001, Tokyo, Japan.

Brunsdon, C., A. S. Fotheringham, and M. Charlton, 1999. Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, 39(3), pp. 497-524.

Can, A. 1990. The measurement of neighborhood dynamics in urban house prices. *Economic Geography*, 66, pp. 254-272.

Can, A., 1992. Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, 22, pp. 453-474.

Casetti, E., 1972. Generating models by the expansion method: applications to geographic research. *Geographic Analysis*, 4, pp. 81-91.

Dubin, R. A., 1992. Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics*, 22, pp. 433-452.

Dubin R. A., 1998. Spatial autocorrelation: a primer. *Journal of Housing Economics*, 7, pp. 304-327.

Fotheringham, A. S. and C. Brunsdon, 1999. Local forms of spatial analysis. *Geographical Analysis*, 31(4), pp. 341-358.

Fotheringham, A. S., M. E. Charlton, and C. Brunsdon, 1998. Geographical weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30, pp. 1905-1927.

Gao, X. and Y. Asami, 2001. The external effects of local attributes on living environment in detached residential blocks in Tokyo. *Urban Studies*, 38(3), pp. 487-505.

Getis, A. and K. Ord, 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24, pp. 189-206.

Switzer, P., W. S. Kowalik, and R. J. P. Lyon, 1982. A prior probability method for smoothing discriminant analysis classification maps. *Mathematical Geology*, 14(5), pp. 433-444.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the comments by Atsuyuki Okabe and Yukio Sadahiro. The present study is supported by the Grant-in-Aid for Scientific Research by the Ministry of Education, Culture, Sports, Science and Technology.

APPENDIX A. A DESCRIPTION OF DATASET

The database used in this paper was developed in Gao and Asami (2001). It involves 190 detached housing lots in western Tokyo. Sample area is small, about 110 hectare. Sample properties were drawn from the Oct. 1996 to Sep. 1997 issues of *Weekly Housing Information* magazine, which provides transaction price and basic information of land and houses in Japan.

In addition to the information provided by the magazine, details of sample properties, such as road, sunshine, distances to public facilities, and neighborhood environments, were obtained through site survey, GIS applications, and so on. The dataset also includes the location information of sample lots, represented by the x and y coordinates of their center point.

APPENDIX B. THE ESTIMATES OF A SIMPLE LINEAR MODEL

	Coefficient (million yen/m^2)/unit	Std. Coefficient	t-statistic	Sig. level
Constant	0.9115		9.165	.000
ActualFAR	0.1276	0.182	3.215	.002
T.station	-0.0157	- 0.420	- 9.612	.000
W.road	0.0209	0.137	2.855	.005
Residual.b.age/S	0.5686	0.401	6.419	.000
Landscape	-0.1726	- 0.400	-8.463	.000
T.Shinjuku	-0.0168	- 0.310	-6.596	.000
Frontage	0.0058	0.113	2.383	.018
Goodpavement	0.0420	0.114	2.798	.006
Parkinglot	0.0382	0.153	3.536	.001
B.quality1	0.0575	0.150	3.507	.001
Sunshine/S	0.9476	0.125	2.669	.008
Con.greenery/S	21.4547	0.268	3.138	.002
Con.greenery	-0.1956	-0.257	-2.968	.003
Mix-use3/S	-17.4766	-0.381	-2.438	.016
Mix-use3	0.2384	0.404	2.635	.009
Tree1	0.0335	0.085	1.992	.048

R-square: 0.756, adjusted R-square: 0.734.
See Appendix D for the definition of variable names.

Statistical tests showed that the coefficients were quite stable, suggesting that the model does not have serious multi-collinearity problems. In addition, the model was shown to be superior to some other simple formed models such as log linear regression model.

APPENDIX C. THE ESTIMATES OF SPATIAL DEPENDENCY MODEL

	Coefficient (million yen/m^2)/unit	Std. Coefficient	*t*-statistic	Sig. level
Constant	0.8720		7.464	.000
ActualFAR	0.0949	0.136	2.073	.040
T.station	-0.0117	-0.311	-3.532	.001
W.road	0.0048	0.069	1.154	.250
Residual.b.age/S	0.6610	0.467	6.910	.000
Landscape	-0.4670	-1.082	-4.453	.000
T.Shinjuku	-0.0091	-0.168	-2.363	.019
Goodpavement	0.0400	0.108	2.269	.025
Frontage	0.0779	0.151	3.102	.002
Parkinglot	0.0296	0.118	2.559	.011
B.quality1	0.0686	0.179	3.637	.000
Sunshine/S	0.6310	0.083	1.773	.078
Con.greenery/S	34.4830	0.431	4.172	.000
Con.greenery	-0.2760	-0.363	-3.925	.000
Mix-use3/S	-13.7990	-0.301	-1.879	.062
Mix-use3	0.2070	0.351	2.147	.033
Tree1	0.0209	0.053	1.122	.264
NN-actualFAR	0.0299	0.044	0.680	.497
NN-t.station	-0.0033	-0.085	-0.978	.330
NN-w.road	0.0099	0.068	1.026	.306
NN-residual.b.age/S	-0.1080	-0.074	-1.127	.262
NN-landscape	0.2780	0.650	2.629	.009
NN-t.Shinjuku	-0.0074	-0.138	-1.941	.054
NN-goodpavement	0.0242	0.065	1.338	.183
NN-frontage	0.0015	0.031	0.572	.568
NN-parkinglot	0.0080	0.032	0.725	.469
NN-b.quality1	-0.0266	-0.070	-1.433	.154
NN-sunshine/S	0.6290	0.073	1.582	.116
NN-con.greenery/S	-13.7210	-0.163	-1.435	.153
NN-con.greenery	-0.0011	-0.001	-0.014	.989
NN-mix-use3/S	-8.8210	-0.174	-0.787	.432
NN-mix-use3	0.0914	0.145	0.650	.517
NN-tree1	0.0069	0.017	0.324	.746

NN- indicates the prior characteristic terms estimated from the nearest neighbor of sample lots.
R-square: 0.801, adjusted *R*-square: 0.761.
See Appendix D for the definition of variable names.

APPENDIX D. THE DEFINITION OF VARIABLES

Variable name	Meaning (unit)
ActualFAR	Building floor area / lot area (ratio)
Con.greenery	Adjacent to public green space, if true, 1, otherwise 0
Frontage	The frontage of a lot (m)
Goodpavement	The pavement of the road in front of a lot being good, if true, 1, otherwise 0
Landscape	Within designated landscape areas, if true, 1, otherwise 0
Mix-use3	Intensive mixed land use, if true, 1, otherwise, 0
Parkinglot	The count of parking lots (number)
Residual.b.age	Residual building age (year)
S	The size of a lot (m^2)
Sunshine	Sunshine duration (hour)
Tree1	Greenery in the district being good, if true, 1, otherwise 0
T.Shinjuku	Time distance to central city area (minute)
T.station	Time distance to the nearest station (minute)
W.road	The width of the road in front of a lot (m)