International Conference: Spatial Thinking and Geographic Information Science 2011

Addressing Structural Instability in Housing Market Segmentation of the Used Houses of Tokyo, Japan

Kazi Saiful Islama*, Yasushi Asamib

aUrban and Rural Planning Discipline, Khulna University, Khulna-9208, Bangladesh
bCenter for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi Chiba 277-8568, Japan

Abstract

Hedonic price is the most popular determinant of housing market segmentation. Nonetheless, hedonic regression is often criticized for not being able to tackle heterogeneity of hedonic price of urban scapes, which creates structural instability in the regression. Spatial switching regression is used in this paper to develop localized hedonic regression ensuring constancy of the parameters within each localized regression. Central Tokyo and the outskirts are clearly divided into two segments. This research provides a new dimension to the existing methods of housing market segmentation. Additionally, this research also improves the predictive capability and model fitting statistics of hedonic regression.

Keywords: Hedonic regression; structural instability; spatial switching regression; housing market segmentation;

1. Introduction

Housing as a complex market is often defined as a set of distinctive submarkets arising from structural and local attributes, and inelastic demand for short-run supply over a given period of time [1]. Emphasizing housing market segmentation, Goodman and Thibodeau [14] stated that proper understanding of housing market segmentation will likely increase the predicted accuracy of statistical models used to estimate house prices and will enable researchers to better model spatial and temporal variation in these prices. They have also pointed out that an accurate assignment of properties to submarkets will improve lenders’ and investors’ abilities to price the risk associated with financing homeownership and would also reduce the search cost for housing consumers [14].

Even after six decades of research, the definition and methods of housing market segmentation is not yet fixed. For a brief review of literatures of housing market segmentation please see [17]. It has been extensively documented that the analysis of house prices using hedonic modeling makes it possible to

*Corresponding author. Tel.: +0088-041-731244; fax: +0088-041-731244.
E-mail address: saiful_ku@yahoo.com.

© 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and/or peer-review under responsibility of Yasushi Asami

1877–0428 © 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of Yasushi Asami
Open access under CC BY-NC-ND license. doi:10.1016/j.sbspro.2011.07.021
estimate the marginal monetary contribution of property attributes and neighborhood externalities [26].

Hedonic price is the most popular determinant of housing market segmentation. Xu [28] portrayed HR as an “addictive” regression model, in which, price of housing units is determined by property specifics plus location attributes and no interactive effect between these two parameters [28]. Hannonen [15] recognized some of the elemental predicaments of hedonic regression (HR) that include spatiotemporal variability of land prices, the model specification dilemma, outlying and influential observations [15]. They have also avowed that the assumption of “constant marginal price for property specifics” is quite incongruous against the established theory that prices of housing attributes exhibit distinct spatial heterogeneity within housing markets [13,20]. Thus, addictive regression fails to permit marginal value of property specifics to vary spatially over the city, which may result in biased coefficients and a loss of explanatory power.

Spatial discrepancy of land price is the most important consideration of this paper that can be attributed to spatial heterogeneity and spatial dependency [10]. Spatial dependencies can be solved introducing variables of spatial relationship in the regression. Spatial heterogeneity is a more problematic issue. Hannonen [15] has also recommended narrowing the analyses into reasonably small submarkets, which homogenizes the data [15]. Because of heterogeneity, structural stability of HR is significantly hindered. Regime switching regression has been evolved to address this issue. Much of the Quandt’s works [23,24] are in line of switching regression. Scholars have attempted to include spatial aspect with the traditional regression to estimate localized parameters. As a result Geographically Weighted Regression (GWR) and some mixed models evolved. But according to [10] GWR and the mixed model are not significantly better than a simple linear model. Hence, there remains scope of further investigation into this matter.

In general, parameters of econometric models cannot be expected to remain constant [25]. Numerous authors have addressed this issue [7,22,27]. A number of test procedures for parameter instability is reviewed and proposed by [2,16] and Carrasco (2004). It is beyond the scope of this paper to review this issue. Localization of the models is well discussed by [6,9,11]. Based on the spatial distribution of the values of these parameters, spatial housing submarket can be delineated. Primary objective of this paper is to address structural instability issue in housing market segmentation and thus providing a new perspective of the definition of housing market segmentation. Spatial HR is used in this research to prepare spatial switching regression developed by [3,4]. Two non-spatial HRs are also developed to show the effect of segmentation on HR. The output of spatial switching regression is plotted and spatial housing market segmentation is collected.

2. Characteristics of Data

This research is conducted on the housing market of Tokyo, Japan. Two sets of data is used in this research, (1) the used detached houses of 23 wards of Tokyo Metropolitan area, and (2) telephone data provided by Nippon Telegraph and Telephone (NTT) Corporation. A total of 7893 samples for used-detached houses with household characteristics variables (such as presence of veranda, stories of a building etc.) are used. This dataset also includes the coordinates depicting the spatial locations of the households. This is extremely essential for searching spatial variables. These datasets were provided by “At Home” as part of the agreement between Center for Spatial Information Science (CSIS) and At Home. Spatial variables representing the distance of the households from different land uses (i.e. train station, shopping mall etc.) are collected from addressed matched NTTtownpage data. Address matching service is provided by Centre for Spatial Information Science (CSIS) of the University of Tokyo. NTTtownpage data was also provided by CSIS.
Moran’s $I$ statistic for spatial autocorrelation is a measure of the overall clustering of the data [3, p. 101]. The value of $I$ is 0.332. As a rule of thumb, if this value in less than 1.96, the $H_0$ of no serial correlation is rejected. Lagrange multiplier (LM) is another test [3, p. 104] that depends on least square estimates and use spatial weight matrix that describes the spatial dependence of the error term. If the value of this test is below 6.635, it can be assumed that $H_0$ of no spatial correlation can be rejected. The value of LM is 2.3254. This implies that the dataset has spatial error dependence.

Given a set of weighted features, Local Moran’s $I$ (also known as Local Indicators of spatial association or LISA) statistic identifies clusters of features with values similar in magnitude. $Z$ scores are measures of standard deviation. $Z$-score with associated proper P-value depicts the strength of local relationship. For better understanding, please see [8]. Figure 1 shows the $Z$ score rendering for error terms. Error values are one way or another crowd together into two clusters. Higher heterogeneity leads to higher error and higher structural instability in the CBD areas. At the outskirts relatively same level of marginal utility may have led to same kinds of error level. Therefore these two parts of the cities are identified differently. Middle portion remains neutral for the error term.

It is proved that error terms are spatially correlated, which contravenes the usual assumption of least square methods. Spatial process model is thus applicable. Previously stated spatial autocorrelation or Moran’s $I$ statistic also supports this argument. From this point forward spatial econometrical approach is described.

### 3. Spatial Switching Regression

If regression analysis is based on a cross-section of contagious spatial units of observation or on a pooled set of cross sections, spatial dependence is likely to be present in the error terms [4]. Switching regression is originally developed on time-series data [23,24]. Spatial component is added later by [3,4]. In a sense, switching regression segments housing market based on the predictive capacity, spatial heterogeneity of the parameters, stability of the regression and homogeneity of the data pattern. The model also includes a spatial lag term that does not vary across regimes, meaning that the spatial process is assumed to be uniform over the space [12, p. 167].

Let a HR be –

$$y = X\beta + \epsilon \hspace{1cm} (I)$$

Here $y$ is a vector ($N$ by $I$) of land price observations, $X$ is a matrix ($N$ by $K$) of the exogenous variables depicting structural and locational characteristics including the usual constant, $\beta$ is a $K$ by $I$
vector of parameters associated with \( X \) and the disturbance is \( \varepsilon \). A simple switching regression that classifies the observations into one or the other can be expressed as-

\[
\begin{bmatrix}
y_i \\
y_j \\
\end{bmatrix} = \begin{bmatrix} X_i & 0 \\ 0 & X_j \end{bmatrix} \begin{bmatrix} \beta_i \\ \beta_j \end{bmatrix} + \begin{bmatrix} \varepsilon_i \\ \varepsilon_j \end{bmatrix}
\]

Here, two regimes represented by \( i \) and \( j \) correspond to the subset of observations for which the regression model follows a different set of coefficients. The notations of (2) are the same as (1) but are divided into two regressions representing \( j \) and \( i \) subsets of observations. This is the basic formation of non-spatial switching regression [3,4,21]. Total number of observations, \( N \) consist of \( N_i \) in subset \( i \) and \( N_j \) in subset \( j \), such that \( N = N_i + N_j \). For the sake of efficient estimation of \( \beta_i \) and \( \beta_j \), both \( N_i \) and \( N_j \) should be large enough. Spatial error dependence directly refers to regime classification, i.e., spatial error will vary between the regimes. The general spatial model of simplest form is

\[
y = \rho Wy + X \beta + \varepsilon
\]

where, \( \varepsilon = \lambda W \varepsilon + \mu \)

with \( \mu \sim N(0, \Omega) \), and the diagonal elements of the error covariance matrix, \( \Omega = h_i(z, \alpha) \) where \( h_i > 0 \).

Here, \( \rho \) is the coefficient of spatially lagged dependent variable; \( W \) is the weight matrix (\( N \) by \( N \)). \( \lambda \) is the coefficient in a spatial autoregressive structure for the error, \( \varepsilon \). The diagonal elements of the diagonal covariance matrix \( \Omega \) allow for heteroskedasticity as a function of \( P+1 \) exogenous variables \( z \), which include a constant term. The \( P \) parameter \( \alpha \) are associated with the non-constant terms, such that \( \alpha = 0 \), it follows that \( H = \sigma^2 \).

There are number of ways of dealing with spatial error model based spatially switching regression [3,4,21]. Spatially switching regression in the simplest form in line with (2) is

\[
\begin{bmatrix}
y_i \\
y_j \\
\end{bmatrix} = \rho \begin{bmatrix} W_{ii} & W_{ij} \\ W_{ji} & W_{jj} \end{bmatrix} \begin{bmatrix} y_i \\
y_j \\
\end{bmatrix} + \begin{bmatrix} X_i & 0 \\ 0 & X_j \end{bmatrix} \begin{bmatrix} \beta_i \\ \beta_j \end{bmatrix} + \begin{bmatrix} \varepsilon_i \\ \varepsilon_j \end{bmatrix}
\]

The typical form of error dependence in spatial autoregressive structure in (4) can be introduced in the following ways

\[
\begin{bmatrix}
\varepsilon_i \\
\varepsilon_j \\
\end{bmatrix} = \begin{bmatrix} \lambda_i W_{i} & \lambda_j W_{ij} \\ \lambda_i W_{ji} & \lambda_j W_{jj} \end{bmatrix} \begin{bmatrix} \varepsilon_i \\
\varepsilon_j \\
\end{bmatrix} + \begin{bmatrix} \mu_i \\ \mu_j \end{bmatrix}
\]

Here \( \mu_i \) and \( \mu_j \) within each regime of (5) follow a spatial process reflected in spatial weight matrices \( W_i \) and \( W_j \) of dimension \( N_i \) and \( N_j \) respectively. Consequently, spatial weights have a direct relationship to the spatial structure of the instability [4]. \( \lambda_i \) and \( \lambda_j \) are the coefficients in spatial autoregressive structure for the disturbance \( \varepsilon_i \) and \( \varepsilon_j \). The error variance-covariance matrix is,

\[
E[\mu \mu] = \Omega = \begin{bmatrix} \sigma_i^2 I_i & 0 \\ 0 & \sigma_j^2 I_j \end{bmatrix}
\]

where, \( I_i \) and \( I_j \) are identity matrices of dimension \( N_i \) and \( N_j \) respectively.

This matrix is particularly significant because the classifications into different regimes are based on varying degrees of complexity of this matrix. This is a special form of heteroskedasticity where the error variance is different in each regime. In other words, each subset has different error variance and covariance matrix. According to [4], the model can be estimated in two ways, (1) through iterative technique using estimated generalized least square (EGLS), or (2) in a maximum likelihood framework. Nevertheless, he also argued that when spatial dependence is present in the error term, simple iterative technique is no longer valid. This is largely due to the two dimensional nature of the dependence in space
(simultaneity). Moreover, EGLS approach is unrealistic since \( \lambda \) is unknown and need to be estimated by means of maximum likelihood techniques. The corresponding log-likelihood function is

\[
L = -\left( \frac{N}{2} \right) \ln \left( \sigma_j^2 \right) - \left( \frac{N}{2} \right) \ln \left| I - \lambda W \right| - \frac{1}{2} \left( y_i - X_i \beta_j \right) \left( I - \lambda W \right)^{-1} \left( y_i - X_i \beta_j \right) (I - \lambda W) \\
- \frac{1}{2} \left( y_j - X_j \beta_j \right) \left( I - \lambda W \right)^{-1} \left( y_j - X_j \beta_j \right) (I - \lambda W)
\]

and

\[
\Omega^{-1} = \begin{bmatrix} (\sigma_j)^2 I & 0 \\ 0 & (\sigma_j)^2 I \end{bmatrix}
\]

Here, the Jacobian, \( \ln|I - \rho W| = \sum \ln(I - \rho W_i) \), where \( W_i \) are the Eigen values of \( W \) [3, p.184]. Following Páez, Uchida et al (2001), preliminary estimate for \( \rho \) is obtained from non-linear optimization of the log-likelihood function ignoring spatial heterogeneity in the covariance structure \( \sigma_j^2 = \sigma^2 \) (i.e., ). Then estimate for the variance parameters is obtained conditional on the autocorrelation parameter and these in turn are used to calculate a new value of \( \rho \) [21]. Spatially switching regression is already a well-developed technique. For more detail exploration please go through [3][4]. HR based housing market segmentation primarily assumes that, hedonic price resembles the people’s willingness to pay for the property attributes and neighborhood externalities. Therefore, housing market segmentation based on hedonic price; do resemble the homogeneity of the household characteristics.

4. Empirical result

Fig. 2. (a) First order raw standardized contiguity matrix (Sparse formation) of used- detached houses (nz means number of zeros); (b) Second order raw standardized contiguity matrix (Sparse formation) of used-detached houses (nz means number of zeros)
In this research sparse matrix is used. Figures 2 and 3 show the first and second order sparse weight matrix. For the simplicity of this research, raw standardized contiguity matrix is used throughout the whole process. It is to be noted here that second order raw standardized contiguity matrix, $W^2 = W \cdot W$.

Least square method is the starting point for all spatial process models. Table 1 represents the least square estimate (non-spatial) for used-detached houses. Variables at the beginning capital or small “nd” or “ND” followed by some numbers represents the distance from some sorts of land use. “ND” and “nd” mean nearest distance. A list of these variables is abbreviated at the appendix. For better understanding, coding is kept as it is in NTTtownpage data.

Hedonic price is positively affected by land uses of luxurious amenities like cosmetics, decorative and craft product's shop, photographs and design shops. Some shops of daily essence like food chain shops (Grains, noodles, breads, confectionaries and spice, drinks etc.), western and Chinese restaurants and fisheries also affect these houses positively. Used-detached houses are also allergic to the land uses that make the area chaotic through gathering many people. For example, travel agencies, hotels, educational institutes, religious institutes, press or broadcasting and communication companies, dry cleaning, barber shops and public bath etc. Walking time to the nearest station and travel time by bus are negatively related to hedonic prices.

Most of the explanatory variables are statistically significant as indicated by t-statistic and associated marginal probability. Both $\rho$ and $\lambda$ are also statistically significant and positive. The estimated value for the parameter, $\lambda$, is close to 0 (zero). According to LeSage, this is consistent with the idea that any second order effects are small [19]. It is to be noted here that general spatial model uses non-linear optimization. And often the estimate of the non-linear optimization does not reflect global optima.

One of the basic criticisms of the popular HR method is the assumption of homogeneity of dataset upon which it is applied. For urban scape, this assumption is never valid. Accordingly, regression coefficient does not remain constant over space. Housing market of Tokyo thus first need to be segmented based on the locally constant but regionally variable coefficient. This is called structural instability [3, p. 119]. According to [3, p. 119], structural instability is often expressed by changing functional forms or varying parameter. Switching regression addresses this issue. Spatial switching regression often called spatial regime model, in which the coefficients and constant term take on different values depending on the regime and the coefficients of each regime are jointly estimated. The model also includes spatial lag term that does not vary across regime, meaning that spatial process is assumed to be uniform across the entire city [12, p. 167]. Localized HR result is stated in table 2 and 3. Evidently, localized HRs performs better than global regression. Both R-squared and adjusted R-squared improved in localized regression. Standard error of the estimate also reduces. All these are quite predictable but the spatial distribution of the spatial switching regression result is quite perceptible.

This model doesn’t comply with the conventional definition of housing market segmentation. Conventionally, a submarket is defined as a set of dwellings that are reasonably close substitutes for one another, but relatively poor substitutes for dwellings in other submarkets. Depending on the substitutes and the level of aggregation (or disaggregation), definition of housing submarket varies [5,17,18]. In this model, housing submarket is defined based on the structural stability of HR. This segmentation improves the performance of the HR. That means the predictive power of HR improves within the submarket. As hedonic price is a function of the property attributes and neighborhood externalities, it is quite reasonable to segmentize housing market based on HR.

5. Conclusion

Because of measurability, simplicity and robustness, HR based approach is the most popular method for housing market segmentation. HR based housing market segmentation is often criticized for not being able to handle heterogeneity properly. It is also criticized for not being able to deal with topographically based segmentation. The model presented in this paper addresses these issues.
Spatial process models have travelled a long way to come to its current state. Housing market is highly influenced by the human nature. Human nature again responds to their ability, spatial characteristics of the housing unit, characteristics of the houses itself and so many other factors. Presumably, residents of the used-detached houses are old aged and rich. As these houses are old and infrastructures at the outskirts of Tokyo was not well developed at the time when these houses were built, these housing units are not locationally optimized.

Fig. 3. Spatial distribution of the spatial switching regression result of used-detached houses. Red dots depict a housing submarket.
### Table 2: Localized non-spatial HR for Central Business District (red dots at figure 3)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>t-statistics</th>
<th>t-probability</th>
<th>Variables</th>
<th>Coefficient</th>
<th>t-statistics</th>
<th>t-probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>841254.843</td>
<td>81.269</td>
<td>.000</td>
<td>Age_bldg_Y</td>
<td>-6720.252</td>
<td>-34.353</td>
<td>.000</td>
</tr>
<tr>
<td>ND623</td>
<td>-160.506</td>
<td>-17.267</td>
<td>.000</td>
<td>time_by_bu</td>
<td>-12760.60</td>
<td>-17.524</td>
<td>.000</td>
</tr>
<tr>
<td>ND1133</td>
<td>-81.974</td>
<td>-10.807</td>
<td>.000</td>
<td>walking_ti</td>
<td>-4360.291</td>
<td>-9.018</td>
<td>.000</td>
</tr>
<tr>
<td>ND143</td>
<td>-136.687</td>
<td>-9.554</td>
<td>.000</td>
<td>ND211</td>
<td>64.847</td>
<td>11.779</td>
<td>.000</td>
</tr>
<tr>
<td>land_area</td>
<td>-1026.146</td>
<td>-25.854</td>
<td>.000</td>
<td>building_a</td>
<td>1055.156</td>
<td>24.116</td>
<td>.000</td>
</tr>
<tr>
<td>ND1280</td>
<td>69.965</td>
<td>5.445</td>
<td>.000</td>
<td>ND930</td>
<td>-78.734</td>
<td>-8.113</td>
<td>.000</td>
</tr>
<tr>
<td>ND142</td>
<td>30.39</td>
<td>3.898</td>
<td>.000</td>
<td>ND1120</td>
<td>-172.611</td>
<td>-7.781</td>
<td>.000</td>
</tr>
<tr>
<td>ND232</td>
<td>73.482</td>
<td>3.176</td>
<td>.002</td>
<td>ND212</td>
<td>94.704</td>
<td>4.268</td>
<td>.000</td>
</tr>
<tr>
<td>ND223</td>
<td>-79.755</td>
<td>-5.282</td>
<td>.000</td>
<td>ND212</td>
<td>94.704</td>
<td>4.268</td>
<td>.000</td>
</tr>
<tr>
<td>ND221</td>
<td>74.652</td>
<td>2.991</td>
<td>.003</td>
<td>ND1250</td>
<td>-40.069</td>
<td>-3.537</td>
<td>.000</td>
</tr>
<tr>
<td>ND1220</td>
<td>71.872</td>
<td>4.439</td>
<td>.000</td>
<td>ND330</td>
<td>-46.406</td>
<td>-4.008</td>
<td>.000</td>
</tr>
<tr>
<td>ND1131</td>
<td>-46.343</td>
<td>-3.999</td>
<td>.000</td>
<td>ND830</td>
<td>59.677</td>
<td>3.032</td>
<td>.000</td>
</tr>
<tr>
<td>ND830</td>
<td>59.677</td>
<td>3.032</td>
<td>.002</td>
<td>ND1140</td>
<td>-84.552</td>
<td>-4.086</td>
<td>.000</td>
</tr>
<tr>
<td>ND430</td>
<td>21.938</td>
<td>3.377</td>
<td>.001</td>
<td>ND1132</td>
<td>40.356</td>
<td>2.674</td>
<td>.008</td>
</tr>
<tr>
<td>ND224</td>
<td>78.112</td>
<td>2.661</td>
<td>.008</td>
<td>ND1110</td>
<td>-24.736</td>
<td>-3.101</td>
<td>.002</td>
</tr>
<tr>
<td>ND520</td>
<td>29.278</td>
<td>2.405</td>
<td>.016</td>
<td>ND440</td>
<td>33.618</td>
<td>2.275</td>
<td>.023</td>
</tr>
<tr>
<td>ND840</td>
<td>24.167</td>
<td>2.216</td>
<td>.027</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: Localized non-spatial HR of the outskirts

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>t-statistics</th>
<th>t-probability</th>
<th>Variables</th>
<th>Coefficient</th>
<th>t-statistics</th>
<th>t-probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1350000.000</td>
<td>15.923</td>
<td>0.000</td>
<td>ND223</td>
<td>-1527.750</td>
<td>-5.832</td>
<td>0.000</td>
</tr>
<tr>
<td>building_a</td>
<td>1179.888</td>
<td>4.017</td>
<td>0.000</td>
<td>ND1120</td>
<td>2558.890</td>
<td>6.779</td>
<td>0.000</td>
</tr>
<tr>
<td>land_area</td>
<td>-1283.255</td>
<td>-3.003</td>
<td>0.003</td>
<td>walking_ti</td>
<td>-30918.618</td>
<td>-5.211</td>
<td>0.000</td>
</tr>
<tr>
<td>ND840</td>
<td>3493.482</td>
<td>18.115</td>
<td>0.000</td>
<td>ND1131</td>
<td>861.546</td>
<td>4.110</td>
<td>0.000</td>
</tr>
<tr>
<td>ND1110</td>
<td>-896.287</td>
<td>-5.860</td>
<td>0.000</td>
<td>NEAR_DIST</td>
<td>850.005</td>
<td>2.413</td>
<td>0.016</td>
</tr>
<tr>
<td>ND142</td>
<td>-1144.951</td>
<td>-6.356</td>
<td>0.000</td>
<td>ND143</td>
<td>-832.510</td>
<td>-2.728</td>
<td>0.007</td>
</tr>
<tr>
<td>ND1140</td>
<td>-4206.322</td>
<td>-10.603</td>
<td>0.000</td>
<td>ND1132</td>
<td>-513.973</td>
<td>-2.164</td>
<td>0.031</td>
</tr>
<tr>
<td>ND440</td>
<td>1105.041</td>
<td>5.711</td>
<td>0.000</td>
<td>ND1250</td>
<td>526.417</td>
<td>2.500</td>
<td>0.013</td>
</tr>
<tr>
<td>ND842</td>
<td>1945.064</td>
<td>6.607</td>
<td>0.000</td>
<td>ND221</td>
<td>1041.372</td>
<td>2.939</td>
<td>0.004</td>
</tr>
<tr>
<td>ND623</td>
<td>-1409.088</td>
<td>-6.554</td>
<td>0.000</td>
<td>ND232</td>
<td>-1086.809</td>
<td>-2.369</td>
<td>0.019</td>
</tr>
</tbody>
</table>
One of the sub-markets of the used-detached houses is clearly developed around the old part of Tokyo and the regions that have better spatial images (Chuo ku, Chiodaku, Shibuya ku, Minato ku, Meguro ku, Shinagawa ku, Setagaya ku). According to some of the residents of Tokyo, riding some of the railway lines is prestigious for them. Those railway lines primarily pass through these wards. These areas are also famous for highest living environment. This area is also perceived as culturally rich.

There lies a scope of applying this model in other megacities to come to a generalization. However for Tokyo, this model successfully identifies the central Tokyo and its periphery. There are a few controversies regarding the optimum number of submarkets for a specific data set, which is not done in this paper. The fact is till today no method is developed to identify optimum number housing submarkets. Spatial switching regression is developed only based on the general spatial model. There is a scope of developing the same for other models (for example discrete choice models). If it is possible to do so, the prediction accuracy of the already developed models used for housing market segmentation can be judged.

Reference


**Appendix A.**

List of spatial distance variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Abbreviations</th>
<th>Variables</th>
<th>Abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>nd621</td>
<td>Iron and steel</td>
<td>nd1120</td>
<td>Travel agencies, inns, and hotels</td>
</tr>
<tr>
<td>nd622</td>
<td>Nonferrous metal</td>
<td>nd1131</td>
<td>Hobbies, amusements, and related industries</td>
</tr>
<tr>
<td>nd623</td>
<td>Metal products</td>
<td>nd1132</td>
<td>Sports facilities and related industries</td>
</tr>
<tr>
<td>nd710</td>
<td>Chemical industry, excluding medicine</td>
<td>nd1133</td>
<td>Education</td>
</tr>
<tr>
<td>nd720</td>
<td>production made from rubber or plastic</td>
<td>nd1140</td>
<td>Other life related services, religion, arts</td>
</tr>
<tr>
<td>nd811</td>
<td>general machinery</td>
<td>nd1210</td>
<td>Cleaning and security industries</td>
</tr>
<tr>
<td>nd812</td>
<td>machinery used for processing foods or for agriculture</td>
<td>nd1220</td>
<td>Lease, rental, maintenance, repair industries</td>
</tr>
<tr>
<td>nd813</td>
<td>machinery used for environment or safety or health &amp; sanitation</td>
<td>nd1230</td>
<td>Logistics and warehouses</td>
</tr>
<tr>
<td>nd821</td>
<td>machinery of sound, communication or computation</td>
<td>nd1240</td>
<td>Finance, insurance, security</td>
</tr>
<tr>
<td>nd822</td>
<td>other electric machinery</td>
<td>nd1250</td>
<td>Broadcasting, communication, and press</td>
</tr>
<tr>
<td>nd830</td>
<td>transport machines</td>
<td>nd1260</td>
<td>Staff agencies, supply services</td>
</tr>
<tr>
<td>nd840</td>
<td>textile machines or precision instruments</td>
<td>nd1270</td>
<td>Information, surveying, and advertisement</td>
</tr>
<tr>
<td>nd910</td>
<td>distributors or market for a variety of goods</td>
<td>nd1280</td>
<td>Photographs and designs</td>
</tr>
<tr>
<td>nd920</td>
<td>retailers of a variety o goods</td>
<td>nd1290</td>
<td>Professional services (consultants)</td>
</tr>
<tr>
<td>nd930</td>
<td>companies of recycling resources or used goods</td>
<td>nd1310</td>
<td>Organization, associations, and unions</td>
</tr>
<tr>
<td>nd1010</td>
<td>Hospital and clinics</td>
<td>nd1320</td>
<td>Public facilities and institutions</td>
</tr>
<tr>
<td>nd1020</td>
<td>Medicine and medical machines and tools</td>
<td>nd1400</td>
<td>Government office</td>
</tr>
<tr>
<td>nd1110</td>
<td>Dry cleaning, barbers, public bath</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>