Identification of Geographical Segmentation of the Rental Apartment Market in the Tokyo Metropolitan Area

Ryo Inoue

Graduate School of Information Sciences, Tohoku University, 6-6-06 Aramaki-Aoba, Aoba, Sendai, Miyagi 980-8579, Japan
rinoue@tohoku.ac.jp
https://orcid.org/0000-0002-0106-9777

Rihoko Ishiyama

Graduate School of Information Sciences, Tohoku University, 6-6-06 Aramaki-Aoba, Aoba, Sendai, Miyagi 980-8579, Japan
rihoko.ishiyama.r5@dc.tohoku.ac.jp

Ayako Sugiura

Phronesis Inc., 1-14-9, Nishi-Shimbashi, Minato, Tokyo 105-0003, Japan
a.sugiura@phronesis.link

Abstract

It is often said that the real estate market is divided geographically in such a manner that the value of attributes of real estate properties is different for each area. This study proposes a new approach to the investigation of the geographical segmentation of the real estate market. We develop a price model with many regional explanatory variables, and implement the generalized fused lasso - a regression method for promoting sparsity - to extract the areas where the valuation standard is the same. The proposed method is applied to rental data of apartments in the Tokyo metropolitan area, and we find that the geographical segmentation displays hierarchal patterns. Specifically, we observe that the market is divided by wards, railway lines and stations, and neighbourhoods.

2012 ACM Subject Classification Applied computing → Economics

Keywords and phrases geographical market segmentations, rental housing market, sparse modelling, generalised fused lasso, Tokyo metropolitan area

Digital Object Identifier 10.4230/LIPIcs.GIScience.2018.32

Category Short Paper

Funding This study was supported by JSPS KAKENHI Grant Number JP18H01552 and the Tokyo Association of Real Estate Appraisers.

Acknowledgements The apartment rental data was provided to us by At Home Co, Ltd.

1 Introduction

The real estate market is segmented by many aspects, including consumer types, property types, and environmental factors. Above all, location plays a major part in market segmentation. People who prefer to live urban areas highly value accessibility to the city centre and proximity to convenient urban amenities, while people who prefer to live in suburbs value property size and proximity to green spaces. As a result, the value of attributes of real estate properties is different in each area.

Geographic market segmentation in the real estate market has attracted much research interest, and attempts have been made to understand the area where valuation standards are the same (see Goodman and Thibodeau (2003) [1]). Previous studies presume a division structure according to specific geographical units, such as school districts, postal districts, and census tracts. However, since the real estate market has a hierarchical division structure from municipality to neighbourhood levels, they might have failed to extract the actual condition of geographic segmentation.

This study proposes a new approach to the investigation of the geographical segmentation of the real estate market. We construct a real estate price model with many regional explanatory variables that depend on different spatial resolutions, and implement the generalized fused lasso - a regression method for promoting sparsity - to extract areas where the valuation standard is the same. The proposed method is applied to the rent data of apartments in the Tokyo metropolitan area to confirm the applicability of the proposed approach.

2 Generalized Fused Lasso

The generalized fused lasso is one method of sparse modelling, which is the solution of a constrained optimisation problem that selects the substantial parameters from among many candidates.

2.1 Lasso

Lasso [2] is a method that minimises the residual sum of squares subject to a constraint on the sum of the absolute values of regression coefficients (excluding the intercept). Hence, lasso gives a solution to the constrained optimisation problem

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X\beta}\|_2^2 \qquad \text{subject to} \quad \|\boldsymbol{\beta}\|_1 \leq t \tag{1}$$

where $\boldsymbol{y}$ is $n \times 1$ vector of the observations, $\boldsymbol{X}$ is an $n \times k$ matrix of explanatory variables, $\boldsymbol{\beta}$ is a $k \times 1$ regression coefficient vector, and t is the positive lasso regularisation parameter. Equation (1) is equivalent to

$$\min_{\boldsymbol{\beta}} \left[\frac{1}{2} \|\boldsymbol{y} - \boldsymbol{X\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|\right] \tag{2}$$

where λ is a Lagrange multiplier. The optimal values of λ or t are usually determined through cross-validation.

2.2 Generalized fused lasso

Fused lasso [3] is a method to investigate the presence or absence of a difference between consecutive parameters. The optimisation problem of fused lasso imposes a new condition on the differences between consecutive parameters;

$$\min_{\boldsymbol{\beta}} \left[\frac{1}{2}\left(y_i - \sum_{j=1}^{k} \beta_j x_i^{(j)}\right)^2 + \lambda \sum_{j=1}^{k} |\beta_{j+1} - \beta_j| + \gamma\lambda \sum_{j=1}^{k} |\beta_j|\right] \tag{3}$$

where y_i, $x_i^{(j)}$, and β_j are components of $\boldsymbol{y}$, $\boldsymbol{X}$, and $\boldsymbol{\beta}$ respectively. The hyperparameter γ determines the weight between the two regularisation terms.

Table 1 Summary of variables.

Variable name	Mean	Standard deviation	Maximum	Minimum
Rent per square meter (yen / m^2)	3123.34	804.19	6999.50	1000
Apartment age (year)	21.63	11.99	69.17	0
Area of property (m^2)	36.33	19.03	440	10
Walking time to the nearest station (min)	6.75	4.18	60	0
Floor number	3.77	2.58	15	1
Number of rooms	1.41	0.66	8	1

Fused lasso estimates parameters whose difference to consecutive parameters tends to be zero; it can estimate common parameters. Generalised fused lasso [3] is a generalised form of fused lasso, in that it imposes constraints on differences between arbitrary neighbouring parameters. It is given by

$$\min_{\boldsymbol{\beta}} \left[\frac{1}{2} \left(y_i - \sum_{j=1}^{k} \beta_j x_i^{(j)} \right)^2 + \lambda \sum_{(m,n)\in E} |\beta_m - \beta_n| + \gamma\lambda \sum_{j=1}^{k} |\beta_j| \right] \tag{4}$$

where E is a set of combinations of neighbouring parameters.

2.3 The Application of generalised fused lasso in geographical analysis

Generalised fused lasso can be applied to geographical analysis. Wang and Rodriguez (2014) [4] estimate the regional divisions of incidence rate of pediatric cancer, for example. The regularisation term that is imposed on the difference between parameters of neighbouring districts enable the authors to estimate a common parameter for them if the difference is not significant.

This study applies generalised fused lasso to the apartment rent data in the Tokyo metropolitan area to investigate the regions where the pricing of real estate properties is the same among neighbouring districts. By setting the explanatory variables that represent regions to different spatial resolutions (i.e. from a municipality level to a neighbourhood level), the analysis could identify the geographical segmentation of the market that was different to previously determined regional divisions

3 Analysis of the Rental Apartment Market in the Tokyo Metropolitan Area

3.1 Apartment rental data

This study utilises apartment rent data in the Tokyo metropolitan area for the years 2015 and 2016. It was collected by At Home Co., Ltd. High-rise condominiums whose number of floors exceed 15 are excluded as their rents have a different pricing structure compared other apartments. Consequently, the total number of records used in this study is 270,605. The data have many property attributes; the natural logarithm of rent per square meter is used as the dependent variable, and the other attributes shown in Tables 1 and 2 are set as explanatory variables.

Table 2 Description of dummy variables.

Dummy name	Description	Number of variables
Railway line dummy	All railway lines are included, except dummies that are the same as some nearest station dummies	59
Nearest station dummy	All nearest stations that appear in data are included Reference: Heiwajima station	474
Cho dummy	All chos that appear in data are included Reference: Nansa-3	293

3.2 Apartment rent model

This study sets the following apartment rent model.

First, the five explanatory variables of apartment age, area of property, walking time to the nearest station, floor number, and number of rooms are used to estimate the ward (municipality)-level parameters. The Tokyo metropolitan area, which is the target area, of consists of 23 wards. As such, 23 parameters are estimated for these five factors.

Next, another three different levels of location factors that affect the market are considered in this study: railway lines, nearest railway stations, and "cho" (neighbourhood). These location factors are represented by dummy variables in this model.

The apartment rent model is given by

$$\begin{aligned} y_i = &\beta_0 + \sum_{p_w \in P_{ward}} \sum_{w \in Ward} \beta^{ward}_{p_w w} x^{ward}_{i p_w w} + \sum_{l \in Line} \beta^{line}_i d^{line}_{il} \\ &+ \sum_{s \in Station} \beta^{station}_s d^{station}_{is} + \sum_{c \in Cho} \beta^{cho}_c d^{cho}_{ic} + \epsilon_i \end{aligned} \tag{5}$$

where β_0 denotes the intercept of the regression, $\beta^{ward}_{p_w w}$ denotes the ward-level regression coefficient for the explanatory variable p_w in ward w, β^{line}_i denotes the regression coefficient of the railway line dummy variable l, $\beta^{station}_s$ denotes the regression coefficient of the nearest station dummy variable s, β^{cho}_c denotes the regression coefficient of the cho dummy variable c, P_{ward} denotes a set of ward-level explanatory variables, $Ward$ denotes a set of wards in the target area, $Line$ denotes a set of railway lines, $Station$ denotes a set of railway stations, and Cho denotes a set of chos. Note that a station and a cho whose average rent per square meter are selected as the reference and dummy variables respectively, are not set for that station and cho.

The regularisation terms that impose weights on the differences between parameters of adjacent regions are set for ward-level parameters and parameters of cho dummies. If the differences between parameters of adjacent wards and chos are not significant, the common parameters would be estimated. The optimisation problem for this analysis is given by

$$\begin{aligned} \min_{\boldsymbol{\beta}} \Bigg[&\frac{1}{2} \sum_{i \in Trans} \Big(y_i - \sum_{p \in P} \beta_p x_{ip} \Big)^2 + \lambda \sum_{p_w \in P_{ward}} \sum_{(a,b) \in Neighbor_{ward}} |\beta^{ward}_{p_w a} - \beta^{ward}_{p_w b}| \\ &+ \lambda \sum_{(c,d) \in Neighbor_{cho}} |\beta^{cho}_c - \beta^{cho}_d| + \gamma \lambda \sum_{p \in P} |\beta_p| \Bigg] \end{aligned} \tag{6}$$

where $Trans$ is the set of all properties, $Neighbor_{ward}$ is a set of 55 combinations of adjacent wards, $Neighbor_{cho}$ is a set of 5006 combinations of adjacent chos, and λ and γ are the regularization parameters.

When solving Equation (6), numeric explanatory variables are standardised.

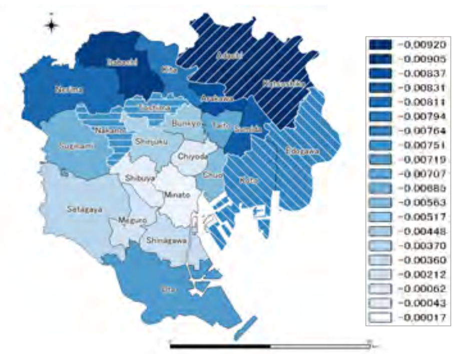

Figure 1 Parameters of property areas.

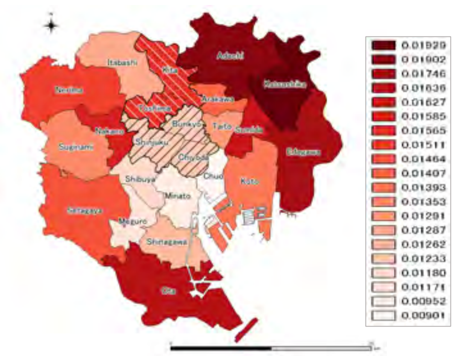

Figure 2 Parameters of floor numbers.

3.3 Results

Four settings of 0.001, 0.1, 1, and 10 for γ are tested, and the estimation with minimum AIC (Akaike's Information Criterion) value is selected. Consequently, when $\gamma = 1$, the model with 673 parameters was adopted. The adjusted coefficient of determination was 0.758.

Figures from 1 to 6 show the estimated parameters. Figures 1 and 2 indicate the parameters of property areas and floor numbers. The shaded parts represent the areas with common parameters. They indicate that similar valuations for apartment attributes occur in some wards. Figures 3 and 4 illustrate that the railway lines and stations in the south-western area are valued higher than those in the north-eastern area. Above all, the apartment rents in Minato and Shibuya wards are high in central Tokyo.

Cho is set as the smallest geographical unit in this study. Figure 5 shows that many parameters are estimated to be zero. The proposed approach suceeds in selecting substantial parameters from many among candidates and reveals that apartment rents are locally homogeneous in most of areas. However, many non-zero parameters are estimated in the Minato and Shibuya wards. Figure 6 shows the Hiroo and Shirokane districts. The thick green lines indicate the ranges where the estimated parameters of cho dummies are the same. The Hiroo and Shirokane districts are famous for being two of the most exclusive residential districts in Tokyo. The results confirm that the cho-level local geographical segmentation occurs in these areas. Rent formation around Hiroo station is fragmented; different levels of rent are formed depending on the direction of properties from the station.

4 Conclusion

This study proposed a new approach to investigate the geographic segmentation of the real estate market. The approach consists of the price model with many regional parameters to represent the difference of price formation by region. Parameter estimation was performed by generalized fused lasso to extract substantial parameters (impose sparsity) and to search for common parameters in adjacent regions. The applicability of the approach is examined by the analysis of geographical segmentations of the rental apartment market in the Tokyo metropolitan area.

The estimated results confirmed the applicability of the proposed approach and revealed the following facts. Several adjacent wards had the same valuations for apartment attributes, the valuation on railway lines and stations was high in the south-western area, and cho-level geographic segmentation was observed, especially in the Minato and Shibuya wards.

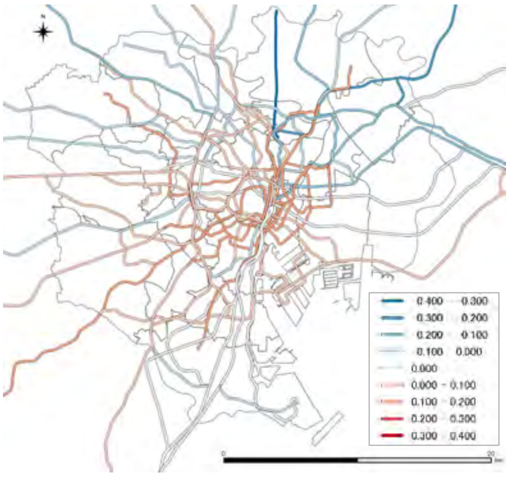

Figure 3 Parameters of railway lines.

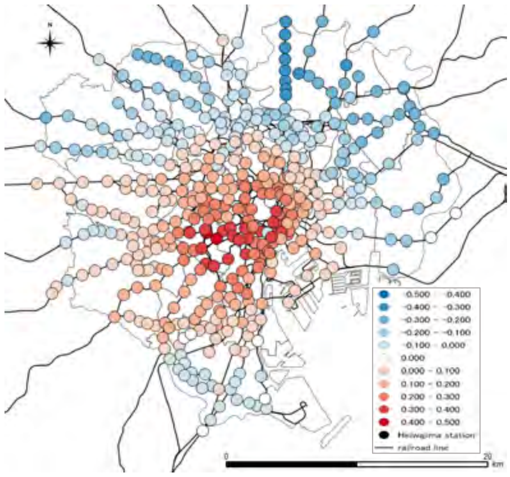

Figure 4 Parameters of railway stations.

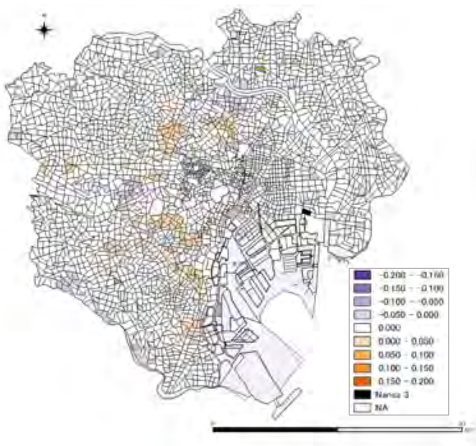

Figure 5 Parameters for chos.

Figure 6 Parameters for chos around Hiroo and Shirokane.

References

1 Allen C. Goodman and Thomas G.Thibodeau. Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics*, 12(3):181–201, 2003. `doi:10.1016/S1051-1377(03)00031-7`.

2 Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Methodological)*, 58(1):267–288, 1996. `doi:10.1111/j.1467-9868.2011.00771.x`.

3 Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fusedlasso. *Journal of the Royal Statistical Society, Series B (Methodological)*, 67(1):91–108, 2005. `doi:10.1111/j.1467-9868.2005.00490.x`.

4 Hao Wang and Abel Rodríguez. Identifying pediatric cancer clusters in florida using loglinear models and generalized lasso penalties. *Statistics and Public Policy*, 1(1):86–96, 2014. `doi:10.1080/2330443X.2014.960120`.