International Journal of Strategic Property Management
ISSN: 1648-715X / eISSN: 1648-9179
2023 Volume 27 Issue 6: 379–390
https://doi.org/10.3846/ijspm.2023.20498

EXPLORING SOME SPATIALLY CONSTRAINED DELINEATION METHODS IN SEGMENTING THE MALAYSIAN COMMERCIAL PROPERTY MARKET

Hamza USMAN [1], Mohd LIZAM [2*]

[1] Department of Estate Management and Valuation, Faculty of Environmental Technology, Abubakar Tafawa Balewa University, Bauchi, Nigeria
[2] Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia

Received 22 August 2023; accepted 29 November 2023

Abstract. This study delves into the property submarket in Kuala Lumpur and Selangor, Malaysia. The submarket is anticipated to be simple, uniform, and dense, making it highly influenced by neighbouring properties. However, traditional data-driven methods that overlook spatial contiguity disregard this density condition. To tackle this problem, the study investigates spatially constrained data-driven methods utilizing Principal Component Analysis (PCA) and cluster analysis. The findings reveal that spatially constrained methods outperform traditional methods by minimizing errors and enhancing model fit. Specifically, the two-step cluster method and k-means cluster method reduce errors by 6.96% and 7.22%, respectively, but at the cost of model fit by 11.23% and 13.94%. Conversely, the spatial k-means and spatial agglomerative hierarchical cluster methods reduce errors by 8.68% and 8.17%, respectively, while improving model fit by 7.1% and 6.35%. Hence, the study concludes that spatially constrained data-driven methods are more effective in differentiating commercial property submarkets than traditional methods.

Keywords: submarket, segmentation, delineation, commercial property market, spatial constraint, cluster analysis, Principal Component Analysis (PCA).

Introduction

The study of property market segmentation has been largely focused on the housing market due to the abundance of available data. This data has allowed researchers to extensively model housing market segmentation. Numerous studies that analyse the pricing of the housing market have verified the presence of market segmentation. Amédée-Manesme et al. (2017) have found that due to the extensive range of diversity in the urban housing market, it can be challenging for the hedonic method to precisely determine the value of a particular attribute in a housing package. The hedonic pricing function is typically employed to model property prices using multiple regression analysis. The model typically uses housing characteristics which include physical, neighbourhood and location as attributes that explain variations in the property prices (Mayer et al., 2019; Usman & Lizam, 2020; Usman et al., 2020a; Wu et al., 2020; Roubi & Ghazaly, 2007). Fundamental to the hedonic price theory, it assumes that property is a heterogeneous product whose price can be decomposed into individual property characteristics (Helbich et al., 2013; Rosen, 1974). The hedonic function assumes a spatial equilibrium of property demand and supply conditions. The implicit price of individual property characteristics is deemed constant and stationary across all spaces within the market.

As pointed earlier, the property market is characteristically heterogeneous, and the nature of heterogeneity includes structural and spatial. The source of heterogeneity could result from the consumers' choice of property characteristics, incomes, tastes and preferences. The property market's complexity and heterogeneous nature require its compartmentalisation into several submarkets. Properties within a submarket are theoretically homogenous and, therefore, close substitutes to one another. However, it may exhibit poor substitutes across other submarkets (Keskin & Watkins, 2017; Pryce, 2013). The submarket effect needs to be modelled to account for the heterogeneity. Conversely, using the traditional unitary equilibrium, such as the hedonic price model based on a one-price-for-all assumption for a structurally and spatially differentiated property market, is inadequate and inefficient. The premise of spatial independence, error independence and spatial equilibrium

*Corresponding author. E-mail: *lizam@uthm.edu.my*

is less likely to be valid in a heterogeneous property market (Beracha et al., 2018; Bourassa et al., 2007; Goodman & Thibodeau, 2007; Morawakage et al., 2023). The violation of the underlying assumptions makes the price modelling inefficient and biased. The nature of market equilibrium gives rise to the property submarket. The price function is continuous within the property submarket, while the property attributes' implicit price function is discontinuous at the submarket points (Baudry & Maslianskaia-Pautrel, 2016). Therefore, property market segmentation is required to model the property market efficiently.

Property market segmentation delineates large coverage of the property market into several submarkets. The properties within the respective virtual boundary are relatively similar based on a given characteristic, hence demonstrating constant implicit prices. According to Baudry and Maslianskaia-Pautrel (2016), "market segmentation occurs if and only if, at market equilibrium, a partition of the market, with homogenous groups of consumers within each part, emerges. The different elements of the partition are referred to as submarkets". The submarkets are, therefore, the distinct component of the whole market. Prior studies are in agreement that property market segmentation improves price prediction accuracy and reduces estimation bias (Chen et al., 2023; Kopczewska & Ćwiakowski, 2021). Nevertheless, no such agreement exists on how the submarkets can be operationalised. Broadly, there are two ways of operationalising property submarkets. The first method is the ad hoc procedure based on prior knowledge about the market (Usman et al., 2021). The second approach is using data to delineate submarkets statistically and empirically. The ad hoc method has been challenged as arbitrarily subjective. It is difficult to confidently declare the submarkets driven by this method as optimal (Bourassa et al., 1999).

Addressing critical issues with the data-driven methodology is imperative. Although cluster analysis is a common approach to grouping properties, it falls short in accounting for the spatial nature of property market data (Chen et al., 2021, 2023; Hu et al., 2020). The approach violates the assumption that nearer properties are more related than distant properties. The Tobler (1970) first law of geography states that "everything is related to everything else, but nearer things are more related than distant things". It is a well-established fact that properties in close proximity tend to belong to the same submarket, while those situated further away do not share the same market.

Similarly, a property submarket is expected to exhibit simplicity, similarity, and compactness, implying contiguity among properties within a submarket (Keskin & Watkins, 2017). The conventional data-driven approach goes against the need for a concise organization. This results in fragmented submarkets rather than consolidated ones. Accordingly, the relative advantage of the two approaches is combined into the spatial data-driven approach. The technique is data-driven with a constraint that it must be a neighbour to at least one of the submarket properties for a property to belong to a submarket. The motivation for this paper is, therefore, two folds. Firstly, the paper aims to empirically delineate the commercial property market in Kuala Lumpur and Selangor into homogeneous submarkets. As noted earlier, most past property market segmentation research was limited to the housing market (He, 2020; Le Gallo et al., 2020; Lisi, 2019), with only a handful of studies focusing on the commercial property market. It is crucial to note that most market segmentation research on commercial properties focuses on non-spatial methods within the office market segment. However, this paper takes a different approach by empirically modelling submarkets that include office lots, retail lots, and shop properties. Furthermore, the paper thoroughly investigates the effectiveness of using spatial data-driven methods to delineate the commercial property market. It is imperative that we acknowledge the value of this approach and consider its potential benefits for future research in the field.

1. Property market segmentation

Segmenting the property market has been widely acknowledged as crucial for improving prediction accuracy, minimising estimation bias, and optimising model fit since the 20th century (Bourassa et al., 2007; Dale-Johnson, 1982; Schnare & Struyk, 1976; Usman et al., 2020a). However, it is worth noting that the methodology employed in segmenting the market into submarkets may present peculiar challenges (Bangura & Lee, 2020; Bourassa et al., 2003; Helbich et al., 2013; Usman et al., 2020b). The meaning and interpretation of "property submarket" can vary significantly. Palm (1978) defined it as a "collectivity of buyers and sellers with a distinct pattern of price-attribute valuations". Goodman and Thibodeau (1998) described submarkets as spatial boundaries where property price per unit of characteristics is constant and available for purchase. The definition considered the implicit prices of individual property characteristics to be stationary across space within the submarket. In the property market, imperfections such as high search costs, differences in information, varying demand among consumers, and immobility can lead to submarkets. These submarkets can be classified based on supply, demand, or a combination of both, depending on the underlying factors. The supply-based definition differentiates submarkets based on property structural and neighbourhood diverse configurations. On the other hand, the demand-based definition distinguishes properties into submarkets based on consumers' socioeconomic and demographic peculiarities. Both definitions are also incorporated in some researches (Goodman & Thibodeau, 2007; Helbich et al., 2013).

Although there is a common understanding of the need for segmenting the property market into submarkets (Chen et al., 2023), there is no consensus on the methodology for delineating the submarket. Based on the available definition, two broad approaches exist for modelling property submarkets. The first approach is the ad hoc delineation of the property market into submarkets a priori (Bourassa et al., 2003; Chen et al., 2009; Keskin & Watkins, 2017). The second approach involves empirical delineation of the property market using various approaches such as cluster analysis, artificial neural network, geostatistical methods, and spatial econometrics methods (Bourassa et al., 2007; Kauko et al., 2002; Keskin & Watkins, 2017; Sobrino, 2014; Li et al., 2018). The two approaches are further discussed in the following subsections.

1.1. Ad hoc property submarket delineation

The property market is typically delineated into submarkets based on a priori predefined boundaries such as administrative boundaries (Bourassa et al., 1999; Chen et al., 2009; Keskin & Watkins, 2017), school districts (Goodman & Thibodeau, 2003), expert-defined boundaries (Chen et al., 2009; Keskin & Watkins, 2017), property types (Xiao et al., 2016) and other socio-economic and demographic considerations. The ad hoc model of property price modelling is based on the Ordinary Least Squares (OLS) regression. Two methods have been utilised in many of the previous studies. One of the methods is using a fixed effect model, which modelled submarkets as additional explanatory variables using dichotomous variables, enabling the intercept to vary spatially. Thus, the interaction effect of the spatial submarket dummy with the predictors' covariates is used to control the heterogeneity in slope. The method improves model quality and is easier to implement and interpret (Bourassa et al., 2007; Lisi, 2019). The second method involves creating separate hedonic equations for each pre-defined submarket. This approach is considered important for delineating pre-defined submarkets (Usman et al., 2021; Yuan et al., 2020). This method compares the sum of squared residual in estimated equations to test the hypothesis of equal coefficients. If significant *F* statistics values are found among submarkets, it indicates the presence of distinct submarkets.

The earlier application of the a priori delineation was a work by Straszheim (1975), which delineated submarkets based on racial composition. Schnare and Struyk (1976) stratified the Boston property market spatially and structurally and found evidence of slight market segmentation. Palm (1978) later found market segmentation evidence based on real estate board jurisdiction and racial composition. Afterwards, different studies delineate submarkets a priori through several criteria such as neighbourhood condition, property type, zoning, school districts etc (Keskin, 2008; Goodman & Thibodeau, 2003; Watkins, 1999; Xiao et al., 2016; Chen et al., 2009; Inoue et al., 2018; Levkovich et al., 2018). It is evident that a significant focus of the applications is on the housing market, whereas only a few studies have been conducted on the commercial property market. A few studies in the commercial property market include Costa et al. (2016), found evidence of distinct office submarkets based on spatial consideration, physical characteristics and property type. Other studies that model the commercial property submarket a priori include Deryol (2019), Fell and Kousky (2015) and Raposo and Evangelista (2017). However, most of the submarkets modelling in the commercial property market are limited to the office segment.

1.2. Aspatial data-driven submarket delineation

Another way to identify property submarkets is by using statistical techniques to analyse multiple datasets and empirically determine the submarkets. This involves the use of different methods to examine the submarkets such as the Principal Component Analysis (PCA) (Watkins, 1999), cluster analysis (Alkan, 2015; Burhan, 2014; Chen et al., 2009; Keskin & Watkins, 2017), neural networks (Kauko et al., 2002). Other techniques include the generalised fused LASSO (Inoue et al., 2018), fused minimax concave penalty (fused-MCP) (Inoue et al., 2020), and multilevel models (Leishman et al., 2013). Dale-Johnson (1982) was one of the first to use a data-driven approach, specifically factor analysis, to identify 13 submarkets within the property market. Their study also included a Chow *F* test to determine if there was any evidence of market segmentation. In previous study by Bourassa et al. (1997), as well as later studies by Bourassa et al. (1999, 2003), they used a combination of PCA and cluster analysis to identify submarkets within the property market. They found that this method resulted in a slight improvement in performance compared to using an ad hoc approach. Burhan (2014) also combined PCA and cluster analysis to derive housing submarkets in Johor, Malaysia. With Geographical Information System's aid, the finding shows that the combined structural and spatial housing attributes effects best capture housing market dynamics. Shi (2015) used Fuzzy C-Means (FCM) clustering algorithm they termed "innovative" to derive property submarket. Gabrielli et al. (2017) also used FCM and compared it with the hard clustering algorithm of k-means and found improved performance associated with FCM. Most studies that used data-driven delineation are found in the housing market with limited commercial property market application.

The current method of submarket delineation, which relies on aspatial data analysis, fails to take into account the physical layout of properties. The allocation of properties to a submarket is based solely on the similarity of their physical attributes to other properties, without any consideration of their physical proximity. This can result in overlapping submarkets without clear boundaries, and may not accurately reflect the fact that nearby properties are often more related than those far apart. As a result, this approach lacks the ability to incorporate spatial contiguity, highlighting the need for a spatial data-driven methodology.

2. Spatial data-driven submarket delineation

Location is essentially the core tenet of real estate analysis. Common terminology in the real estate industry is the phrase "location, location, location", signifying the significance of location in the real estate market. The influence of location is also expected to be considered in submarket modelling. One of the requirements of a property submarket, in addition to simplicity and similarity, is compactness which implies contiguity among properties within a submarket (Keskin & Watkins, 2017). Traditional data-driven approaches don't meet compactness requirements, resulting in scattered submarkets rather than concentrated ones. A spatial data-driven method for submarket delineation is needed to address this issue. Real estate data are considered spatial data because neighbouring properties have a significant impact on each other, leading to a spatial dependence (Chun-Chang et al., 2020; Copiello, 2020; Morales et al., 2020; Usman et al., 2020b). Spatial cluster analysis is a crucial method that considers spatial dependence when analysing spatial data. Using spatial cluster techniques, it is essential to optimize data separation into distinct clusters, where neighbours fall within the same group. Besides property similarity, spatial proximity is also crucial in this analysis. In detail, Młodak (2020) and Zhu et al. (2020) have explored this technique, making it an essential tool for any spatial data analysis.

The spatial cluster analysis follows a two-step procedure based on a relational constraint. The first is the pre-clustering stage, where objects with a minimum of one neighbour are determined. The second step involves clustering the pre-clustered objects and the objects with no neighbours into the final cluster (Młodak, 2020). Several spatial cluster analysis methods exist, such as spatial k-means, spatial hierarchical, Density-Based Spatial Clustering Association with Noise and others. This paper is limited to the spatial k-means and hierarchical clustering methods only. The few application of this property submarket analysis method were found in the housing market (Wu & Sharma, 2012; Wu et al., 2018). No such application was found in the commercial property market. This paper, therefore, attempts to fill this gap.

3. Data and methodology

3.1. Data

A total of 14,043 commercial property transaction records from the National Property Information Centre (NAPIC), Malaysia were obtained. NAPIC is a government agency that keeps records of property transactions made available for valuation and research purposes. The study covered the state of Kuala Lumpur and Selangor, which are among the developed states in Malaysia with a significant number of annual commercial property transactions. The data covered transactions recorded in Gombak, Hulu Langat, Hulu Selangor, Kelang, Kuala Langat, Kuala Lumpur, Kuala Selangor, Petaling, Sepang and Sabak Bernam districts for a period between 2014 and 2018. It covers transaction information related to shop/office buildings, retail and office property types. The data on property addresses were geocoded into

Table 1. Descriptive statistics of the models' variables

Variable	Variable definition	Variable type	Mean	Std. Dev.	Min	Max
LnPrice	Log of transaction price	Continuous	13.33904	1.056879	10.30895	16.08764
LnLotArea	Log plot of land area	Continuous	4.860776	0.562706	2.890372	8.46168
LnBLDArea	Log of building area	Continuous	5.177247	0.812751	2.890372	7.996148
LnAge	Log of age of the building	Continuous	2.325099	1.016369	0	4.110874
SqAge	Squared age of the building	Continuous	353.8646	513.3783	1	3721
ProConditi	Property condition	Ordinal	4.09592	0.704625	1	6
Unit	Property sold as a unit or whole	Dummy	0.217973	0.412884	0	1
Office	Property being office or otherwise	Dummy	0.309336	0.462236	0	1
LowH	Low Height – Properties with up to 2 Levels	Dummy	0.300862	0.458649	0	1
HighH	High Height – Properties with more than 4 levels	Dummy	0.283558	0.450741	0	1
Tenure	Property held as freehold or leasehold interest	Dummy	0.610625	0.487626	0	1
NeigQual	Neighbourhood quality	Ordinal	3.572527	0.629186	1	4
SecRural	Property located at secondary rural area	Dummy	0.037243	0.189363	0	1
PriRural	Property located at primary rural area	Dummy	0.120202	0.32521	0	1
PriCentral	Property located at primary central area	Dummy	0.20836	0.40615	0	1
LnCBD	Log of distance to the Central Business District	Continuous	9.540476	0.879787	2.486618	11.57661
LnCityCent	Log of distance to the nearest city centre	Continuous	8.45303	0.734717	4.121846	10.06075
LnTrainst	Log of distance to the nearest train station	Continuous	7.39255	1.214581	2.331732	11.2753
LnAirport	Log of distance to the Airport	Continuous	10.61422	0.348083	6.816922	11.84088
LnParks	Log of distance to the nearest park	Continuous	8.581832	0.998832	–4.15485	11.456
Y2014	Property sold in 2014	Dummy	0.019654	0.138813	0	1
Y2015	Property sold in 2015	Dummy	0.122481	0.327852	0	1
Y2017	Property sold in 2017	Dummy	0.315104	0.464574	0	1
Y2018	Property sold in 2018	Dummy	0.235064	0.424054	0	1

spatial coordinates using the MMQGIS plugin in QGIS using Google map services API. The data includes the structural characteristics, neighbourhood attributes and location factors. The descriptive statistics of the variables and the market-wide model are presented in Table 1.

Table 1 shows the descriptive statistics of the variables used in constructing the models. The model was a log-log transformed, linear model. All the continuous variables were log-transformed.

3.2. Empirical methods

This study uses four-step procedures: The first step used PCA to reduce the data's dimensionality in orthogonal factors based on which factor scores were generated. The second procedure based on the generated factor scores in PCA, the data were partitioned into submarkets using a two-step cluster: k-means, spatial k-means and spatial hierarchical agglomerative cluster algorithms. The third procedure modelled the derived clusters using the hedonic pricing model. The fourth procedure evaluates the existence of a submarket using the Chow test and weighted errors. The methods are discussed in the following sections.

3.2.1. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) reduces the original variable into a set of orthogonal factors. PCA requires a sample size to be adequate. With a sample of 14,043 observations, the sample size requirement is met (Mooi et al., 2018). Our findings are confirmed by the Kaiser-Meyer-Olkin (KMO) measure of sample adequacy, which surpasses the required threshold of 0.5. The initial challenge was selecting which variables to include in the PCA. We chose variables that had significant β values in the market-wide hedonic model, but we did not include time dummy variables. Overall, 15 variables were included, which cumulatively accounted for 64.71% of the variance and resulted in five retained factors. The proportion of variance explained by the retained factors is considered satisfactory as it's above the 50% required minimum (Mooi et al., 2018). Orthogonal rotation using varimax with Kaiser Normalisation was used for the rotation since the aim is to analyse market segmentation. Each variable is assigned to a given factor based on its highest loadings. The factors and their interpretations are provided in Table 2.

Table 2. Interpretation of factor solution

Factor	Variables	Definition
Factor1	LnCBD, LnCityCent, LnTrainst, LnParks, LnParks, Height	Distance factor
Factor2	LnAge, SqAge, ProConditi	Depreciation factor
Factor3	LnLotArea, LnBLDArea, Type	Physical factor
Factor4	AreaClass, LnAirport	Area location factor
Factor5	Tenure, NeigQual	Tenure factor

The factor scores were calculated using the regression scoring method for each case in the data set. These factor scores are z-standardized, with mean values near 0 and a standard deviation of 1. We then used these scores as input variables for the cluster analysis.

3.2.2. Cluster analysis methods

1) Two-step cluster

The two-step cluster analysis involved the pre-clustering stage and the clustering stage. The pre-cluster stage used a distance measure to cluster cases into several sub-clusters. The algorithm analyses each datum individually and decides whether it aligns with an existing cluster or necessitates the creation of a new one (Benassi et al., 2020; Fuerst & Marcato, 2012; Sobrino, 2014). The clustering stage uses a probabilistic method using maximum likelihood based on measures of fit such as AIC and BIC to cluster the subgroup into the optimal number of clusters (Benassi et al., 2020; Fuerst & Marcato, 2012). The optimal number of clusters was determined by comparing each cluster solution's information criteria (either AIC or BIC). To determine the optimal cluster, we look for the cluster number where the information criteria are minimised. This is achieved when the ratio of BIC or AIC ratio changes and the ratio of distance measures are maximised (Li, 2018). Using this procedure, 8 numbers of optimal clusters were derived. The Silhouette score analysis was also used to confirm the optimal number of clusters.

2) K-means cluster

K-means is within the family of partitioning cluster methods. After initialisation, k-means reassigns objects to other clusters to minimise the within-cluster variation (Małkowska & Uhruska, 2019; Mooi et al., 2018). One of the major advantages of k-means clustering is that it handles large data sets efficiently (Mooi et al., 2018). To accurately segment the 14,403 commercial property transactions in the dataset, it is highly recommended to use the k-means method for optimal analysis. According to Bourassa et al. (1997), the k-means cluster algorithm is effective for segmenting the housing market. Unlike other clustering methods, such as hierarchical and two-step methods, the k-means method requires the number of clusters to be determined in advance. The process of determining the appropriate number of clusters may appear somewhat subjective, leaving some uncertainty regarding the effectiveness of using predetermined values for achieving optimal cluster solutions. A study conducted by Bourassa et al. (1997) suggests that the optimal number of clusters should be determined statistically. The k-means method was applied in their research with the optimal number of clusters obtained from a two-step cluster analysis. As per the study, eight clusters were considered to be the most appropriate for the task at hand.

3) Spatial k-means cluster

The spatial k-means cluster analysis algorithm effectively incorporates a spatial constraint to reinforce that cluster members are spatial neighbours, setting it apart from traditional k-means (Młodak, 2020). To group objects into clusters, the algorithm first uses their centroids to pre-cluster them. It then assigns objects with at least one neighbour to clusters based on the minimax criterion. The spatial k-means cluster analysis performed well in the property market segmentation (Wu & Sharma, 2012; Wu et al., 2018). To analyze the data, we used standardized scores from PCA and conducted a spatial k-means cluster analysis. Through a two-step cluster analysis, we determined that there were eight optimal clusters. In the analysis, we included the centroids X and Y and applied a constraint using KNN spatial weight matrix.

4) Spatial agglomerative hierarchical cluster

The spatial agglomerative hierarchical cluster analysis was used using Ward's method. Ward's approach minimises the sum of squares of any pair of clusters that could be formed at each stage using variance analysis to calculate the similarity (Bourassa et al., 1997). It begins by considering each object as a cluster and combines the singletons clusters in successive steps until it becomes a single cluster with all the objects. The algorithm merges each pair of clusters whose combination maximises the sum of squares within-group error at each successive step. The spatial Ward's cluster is obtained by restricting the merger such that a cluster contains the same neighbours' (Młodak, 2020). Thus, the hierarchical spatial method considers both the objects' attributes and characteristics and their spatial relationships (Gnat, 2019). The method is specified using the number of optimal clusters earlier determined.

3.2.3. Model

The estimates for the models were obtained through a hedonic pricing approach specified using log-log linear OLS regression. We initiated the process by establishing the market-wide model as a fundamental benchmark for evaluating the effectiveness of various delineation techniques. To estimate this model, the following market-wide Equation (1) was estimated.

$$\ln P_i = \beta_i + \sum_k \beta_{ki} \ln X_{ki} + \sum_l \beta_{li} dX_{li} + \sum_n \beta_{ni} Td_{ni} + \varepsilon, \quad (1)$$

where: $\ln P_i$ is the $n \times 1$ vector of commercial property prices; β_{ki}, β_{li}, β_{ni} are regression coefficients of logarithmically transformed continuous commercial property attributes, dummy commercial property attributes and time dummies; $\ln X_{ki}$, dX_{li}, and Td_{ni} are $i \times k$, $i \times l$ and $i \times n$ vectors logarithmically transformed continuous commercial property attributes, dummy commercial property attributes and time dummies respectively where i is the number of observations; ε is the error term which is assumed to be identically and independently distributed (i.i.d.). The derived submarkets were modelled using a separate hedonic equation for each as specified in Equation (2).

$$\ln P_{ij} = \beta_{ij} + \sum_{kj} \beta_{kij} \ln X_{kij} + \sum_{lj} \beta_{lij} dX_{lij} + \sum_{nj} \beta_{nij} Td_{nij} + \varepsilon, \quad (2)$$

where: $\ln P_{ij}$ is the price of commercial property i in commercial property submarket j. For other property attributes and parameters, i is associated with particular commercial property observation in a submarket j. j is the data-driven submarket.

3.2.4. Submarket identification

The existence of submarkets in the commercial property market was tested using two established procedures – the Chow F test and the models' weighted errors (Dale-Johnson, 1982; Schnare & Struyk, 1976; Xiao et al., 2016). The Chow test tests the hypotheses of model equality across the submarkets. The Chow test was computed as in Equation (3).

$$F = \frac{\left[SSR_c - \left(SSR_1 + SSR_2\right)\right]}{\left(SSR_1 + SSR_2\right)} \cdot \frac{(N_1 + N_2 - 2k)}{k}, \quad (3)$$

where: SSR_c, SSR_1, and SSR_2 represent the sum of squared residuals for the market-wide model and individual models respectively; N_1 and N_2 represent the number of observations in respective models. Similarly, the evidence of submarket existence was identified using a "common sense" test. The common-sense test requires the weighted error of the submarket models be at least 5% lower than that of the market-wide model (Dale-Johnson, 1982; Xiao et al., 2016). The weighted error was computed as in Equation (4).

$$RMSE_w = \frac{N_1 + K_1 - 1}{\sum\left(N_j + K_j - 1\right)}\left(RMSE_1\right) + \frac{N_2 + K_2 - 1}{\sum\left(N_j + K_j - 1\right)}\left(RMSE_2\right) + \ldots + \frac{N_j + K_j - 1}{\sum\left(N_j + K_j - 1\right)}\left(RMSE_j\right), \quad (4)$$

where: N_j is the number of observations in jth submarket; and there are j submarkets. The delineation methods' performance was compared using RMSE, R^2, Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE).

4. Results and discussion

The data-driven submarket delineation was carried out using PCA and cluster analysis for both the spatial and the aspatial methods. Before delineating the property, the variables were orthogonally reduced into five factors based on an eigenvalue of one using PCA. The five factors, distance factor, depreciation factor, physical factor, area location factor, and tenure factor (see Table 2), were used to generate factor scores used for the cluster analysis to partition the data into potential submarkets. Eight distinct clusters were derived using the four different cluster algorithms. The spatial distribution of the clusters is presented in Figure 1.

The resulting cluster configuration is depicted in this figure, showcasing the outcomes of four different methods: two-step, k-means, spatial k-means, and spatial agglomerative hierarchical cluster. The two-step approach yielded eight clusters, ranging from 995 to 2998 observations.

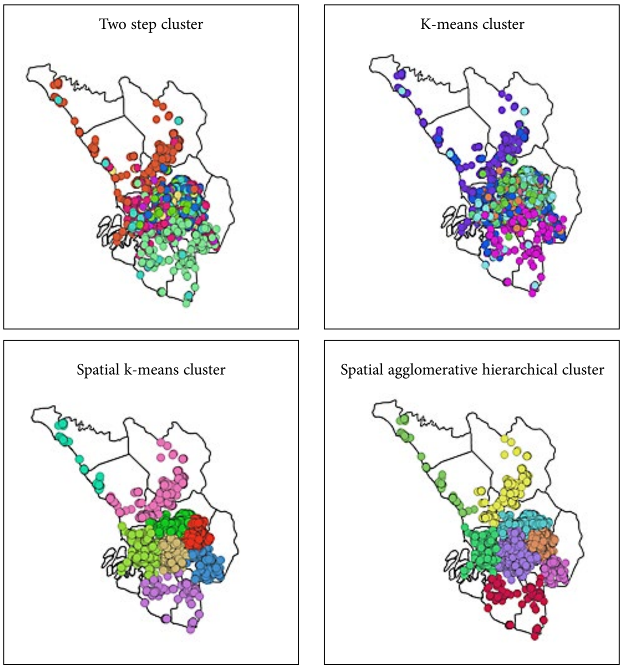

Figure 1. Submarket delineations

Nonetheless, upon visual examination, it is clear that the properties within each non-spatial cluster methods are not compact but rather dispersed across space, a trend also noticed in the k-means clustering method. K-means and wards methods were employed to create compact spatial clusters with sufficient data to identify submarkets for hedonic pricing modelling. The performance of each method was evaluated to compare to the overall market model. The "common sense test" was performed using weighted Root Mean Squared Error (RMSE), weighted Mean Percentage Error (MAE) and weighted Mean Absolute Percentage Error (MAPE) in addition to R^2 relative to the market-wide model. The performance evaluation was conducted using the holdout dataset.

4.1. Market-wide model

The market-wide model was estimated by utilizing multiple regression with Ordinary Least Squares (OLS). The log-log linear model was employed as model allows for the logarithmic transformation of both sides of the equation. Using dummy variables was more efficient than using a linear specification (Soguel et al., 2008). The log of all continuous variables was taken to ensure the linearity and normality of the variables. The results indicate that all parameters have significant coefficients with expected signs. The diagnostic tests include the coefficient of determination and standard error. The model generated an R^2 value of 0.653, indicating that the explanatory variables explain around 65.3% of the variance in commercial property prices. The reported R^2 is within the range of most R^2s reported in commercial property price modelling literature (Ke et al., 2017; Seo, 2016). A series of diagnostic tests were conducted to check the normality of the error, collinearity in the explanatory variables, and the model's information lost through AIC and BIC. The model produced a standard error of 0.62312. The AIC and BIC produced by the model were 26591.3 and 26772.5, respectively. Collinearity issues were observed in shop offices, medium height, secondary central and year 2016 variables which were subsequently removed from the model estimation.

4.2. Two-steps cluster submarkets model

Eight (8) optimal clusters were derived using the two-step cluster algorithm. Separate hedonic models were estimated for each of the derived clusters. The models' fitness was evaluated using R^2, the AIC and BIC, and the standard error of their estimates. The result showed that only three of the eight models indicate an R^2 value greater than the market-wide model. The remaining models exhibit R^2 values below 0.653. However, all the AIC and BIC of the models were greatly reduced. The standard errors of all the submarkets were significantly reduced relative to the market-wide model except for two submarkets with RMSE values above 0.62312. Most of the coefficients of the variables were significant and with the expected signs. The existence of the submarket was tested using the Chow and common sense tests. The Chow test revealed significant *F* statistics for all the submarket pairs, indicating the submarket's distinctiveness. The weighted RMSE was found to be 0.579, which is 6.96% lower than the market-wide model, thereby passing the common sense test.

4.3. K-means cluster submarkets model

The evidence of submarket existence using k-means cluster algorithms was tested by estimating separate hedonic models for the eight derived clusters. The estimated models' R^2 ranges between 0.486 and 0.700. The RMSE also range between 0.50329 and 0.68524. The result showed that the weighted RMSE for the submarket models derived using the k-means method was reduced by 7.22% relative to the market-wide model. The two-step cluster reduced the RMSE by 6.96%. The results clearly demonstrate that implementing the k-means method results in superior prediction accuracy as compared to the two-step cluster method for the given model. Bourassa et al. (1997) found that the k-means method performs better than the hierarchical method in improving the derived property submarkets' price prediction accuracy. However, the number of clusters was a priori-defined, although with the aid of a two-step optimal number of clusters. Chen et al. (2009) also found the superiority of the k-means method above other aspatial data-driven market segmentation methods. The superior performance of the k-means may be related to the sample size used. The k-means algorithm performs well with larger sample sizes and is highly robust (Benassi et al., 2020; Mooi et al., 2018). The Chow test result also revealed that all the delineated submarkets using k-means methods are distinct.

4.4. Spatial k-means cluster submarket model

The result of the 8 separate models based on the spatial k-means cluster method showed that all the models have R^2 values greater than the market-wide model except for two models with R^2 values of 0.628 and 0.388. These submarkets' exceptionally lower R^2 value may be due to a relatively smaller sample size. All the submarket models have their AIC and BIC greatly reduced. The RMSE of all the submarkets were significantly reduced relative to the market-wide model except for two submarkets with RMSE above 0.62312. Commercial property segmentation using the spatial k-means cluster method improved the model fit and substantially reduced the standard error. The weighted R^2 value for the 8 models was 0.6992 indicating a 7.1% in model fit over the market-wide mode. The weighted RMSE for the eight models was 0.569 indicating an 8.68% reduction in the error relative to the market-wide model. Property submarket exists when segmentation results in more than 5% in standard error (Dale-Johnson, 1982). The Chow test was also significant for all the submarket pairs. Therefore, the eight-driven spatial clusters using the spatial k-means clustering constitute distinct commercial property submarkets.

4.5. Spatial agglomerative hierarchical cluster submarket model

The spatial agglomerative hierarchical cluster models showed that all models have R^2 values greater than the market-wide model except for two submarkets with R^2 values of 0.419 and 0.636. All the submarket models have their AIC and BIC greatly reduced. The RMSE of the submarkets were significantly reduced relative to the market-wide model except for two with values above 0.62312. Commercial property segmentation using the spatial agglomerative hierarchical clustering method improved the model fit and substantially reduced the RMSE. The weighted R^2 value for the 8 models was 0.6944 indicating a 6.35% in model fit over the market-wide mode. The weighted RMSE for the eight models was 0.572 indicating an 8.17% reduction in the error relative to the market-wide model. Property submarket exists when segmentation results in more than 5% in RMSE (Dale-Johnson, 1982). The Chow test for the submarkets pairs was significant. Thus, the submarkets are distinct.

4.6. Comparison of data-driven submarket delineation methods

The relative performance of submarket delineation using the different cluster algorithms was compared using the coefficient of determination (R^2), the weighted RMSE, the weighted MAE, and the weighted MAPE. The comparison is presented in Table 3.

Table 3 showed the diagnostics of the various methods used in data-driven submarket modelling. The submarkets were empirically modelled using data both with spatial and without spatial constraints. The aspatial methods are the two-step cluster and k-means cluster methods. The methods severed the model fit by 11.23% and 13.94%, reduced the RMSE by 6.96% and 7.22%, decreased the MAE by 9.01% and 9.04%, and have MAPE of 3.3% and 3.4% which are 9.10% and 9.02% lower than that of the base model, respectively. Although the methods have reduced errors and achieved evidence of submarket with more than 5% error reductions (Dale-Johnson, 1982; Xiao et al., 2016), the models' fits were severely hampered. The spatial data-driven methods improved the fit and accuracy of submarket modelling. Unlike the aspatial techniques, the spatial k-means and spatial agglomerative hierarchical cluster methods improved the model fit by 7.10% and 6.35%, reduced the RMSE by 8.68% and 8.17%, diminished the MAE by 11.32% and 10.72%, and have MAPE of 3.26% and 3.28% which are 11.34% and 10.78% lower than that of the base model respectively. The spatial data-driven submarket models performed substantially better than the aspatial methods and satisfied all submarket existence requirements. This result confirms the efficacy of spatial data-driven submarket methods (Gnat, 2019; Hayles, 2006; Wu & Sharma, 2012; Wu et al., 2018).

Table 3. Performance of data-driven submarket delineation method

Method	R^2		R.M.S.E.		MAE		MAPE	
	Value	Improvement	Value	Reduction	Value	Reduction	Value	Reduction
Market-wide model	*0.6530*		*0.62312*		*0.47958*		*3.67*	
Aspatial data-driven submarket delineation approach								
Two-step	0.5790	–11.23%	0.57900	6.96%	0.43638	9.01%	3.33%	9.10%
K-means	0.5620	–13.94%	0.57800	7.22%	0.43621	9.04%	3.34%	9.02%
Spatial data-driven submarket delineation approach								
Spatial k-means	0.6992	7.10%	0.56900	8.68%	0.42528	11.32%	3.26%	11.34%
Spatial agglomerative hierarchical	0.6944	6.35%	0.57200	8.17%	0.42819	10.72%	3.28%	10.78%

Conclusions

This research explored the potential of spatially constrained data-driven submarket methods for delineating the commercial property market into spatially contagious submarkets. Accordingly, four submarket delineation methods were developed by combining Principal Component Analysis (PCA) and cluster analysis with and without spatial constraints. The methods were applied to the commercial property market in Kuala Lumpur and Selangor. The PCA was used to reduce the dimensionality of the data sets of orthogonal factors. Five factors, defined as distance factor, depreciation factor, physical factor, area location factor, and tenure factor, were used to generate factor scores subsequently used for the cluster analyses to partition the data into potential submarkets. The separate hedonic model was estimated for each defined submarket whose existence is checked by the Chow test and the "common sense" test, comparing the error reduction relative to the market-wide model. Results show that submarkets defined by the aspatial methods reduce error but hampered model fit. The two-step and k-means cluster methods reduced error by 6.96% and 7.22% and severed the model fit by 11.23% and 13.94%, respectively. The spatially constrained methods reduced the error and improved model fit. The spatial k-means and the spatial agglomerative hierarchical cluster methods reduced the error by 8.68% and 8.17%, improving the model fit by 7.1% and 6.35%, respectively.

In the commercial real estate market segmentation, price is a reference point for a range of attributes. While it is evident in past studies that real estate market segmentation, particularly for the housing market, the analysis focuses on tax purposes, a similar analysis may also be extended to understand the clustering effect of businesses being in close proximity to each other. The outcome of the data-driven model ought to be applied in other contexts, such as the assessment of criminal activities and environmental externalities, which not only focuses on locational price variation but also other factors that could lead to demand for commercial real estate.

From a policy implication perspective, the identification of commercial property submarkets may unravel significant discoveries about the locational peculiarity of the respective submarket. For example, the impact of agglomeration economies (typically related to the geographical clustering of the economy) can be explored through the understanding of how commercial real estate is segmented through its distinctive locational feature based on specific attributes such as building age, the mixture of commercial activities, building types, etc. In that sense, the data-driven approach allows additional information to be seamlessly fitted into the model. As more data becomes available as a result of advancements in data extraction technology, this fits the idea of evaluating the impact of agglomeration economies on a specific commercial real estate submarket.

Implicit in the segmentation model is its ability to capture variation in the geographical characteristics of the respective submarket. This specific information is essential feedback to the policymaker or city manager to devise a practical urban planning policy to either close the gap in the property value among the commercial real estate submarket or to identify a specific value proposition that makes each commercial submarket unique as a result of agglomeration economies or distinct environmental externalities. For example, a specific commercial location that consists of a large number of old buildings is a distinct physical characteristic of the location. This may result in a lower rental rate that can be charged and subsequently lower property value. However, the location may also be located close to a historical site, which makes it an area that attracts tourists. Therefore, the local council or city manager, based on the given information, may devise policies that make the owner of the old buildings refurbish or repurpose the building, create demand for its use, and hence increase the property value.

This research provides a new interpretation of commercial real estate segmentation through the lens of geographical clustering on economic activities. The result showed that the spatially constrained data-driven methods performed better than the aspatial methods in empirically delineating commercial property submarkets. The results have significant implications in commercial submarket modelling. It indicates that the spatial methods allow for compactness and provide a basis for identifying spatially contagious submarkets, which may be further extended to model agglomeration impact. The model's data-driven approach allows for replication by either fitting it with new data or testing it in other commercial real estate markets. The research is limited to cross-sectional considerations. Further works that include spatiotemporal dynamics of the commercial property market that reveal the variations of determining factors over time and space are recommended.

Acknowledgements

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 (vot Q120).

Author contributions

Both authors equally contributed to the draft paper. Mohd Lizam conceived the study and was responsible for data collection. Hamza Usman was responsible for data analysis and interpretation. Both authors were responsible for refining the paper to its present state.

Disclosure statement

We declare that we do not have any competing financial, professional, or personal interests from other parties.

References

Alkan, L. (2015). Housing market differentiation: the cases of Yenimahalle and Çankaya in Ankara. *International Journal of Strategic Property Management*, *19*(1), 13–26. https://doi.org/10.3846/1648715X.2014.1000429

Amédée-Manesme, C. O., Baroni, M., Barthélémy, F., & Des Rosiers, F. (2017). Market heterogeneity and the determinants of Paris apartment prices: a quantile regression approach. *Urban Studies*, *54*(14), 3260–3280. https://doi.org/10.1177/0042098016665955

Bangura, M., & Lee, C. L. (2020). House price diffusion of housing submarkets in Greater Sydney. *Housing Studies*, *35*(6), 1110–1141. https://doi.org/10.1080/02673037.2019.1648772

Baudry, M., & Maslianskaia-Pautrel, M. (2016). Revisiting the hedonic price method in the presence of market segmentation. *Environmental Economics and Policy Studies*, *18*(4), 527–555. https://doi.org/10.1007/s10018-015-0122-5

Benassi, M., Garofalo, S., Ambrosini, F., Sant'Angelo, R. P., Raggini, R., De Paoli, G., Ravani, C., Giovagnoli, S., Orsoni, M., & Piraccini, G. (2020). Using two-step cluster analysis and latent class cluster analysis to classify the cognitive heterogeneity of cross-diagnostic psychiatric inpatients. *Frontiers in Psychology*, *11*, 1–11. https://doi.org/10.3389/fpsyg.2020.01085

Beracha, E., Hardin III, W. G., & Skiba, H. M. (2018). Real estate market segmentation: hotels as exemplar. *Journal of Real Estate Finance and Economics*, *56*, 252–273. https://doi.org/10.1007/s11146-017-9598-z

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial dependence, housing submarkets, and house price prediction. *Journal of Real Estate Finance and Economics*, *35*, 143–160. https://doi.org/10.1007/s11146-007-9036-8

Bourassa, S. C., Hamelink, F., Hoesli, M., & Macgregor, B. D. (1999). Defining housing submarkets. *Journal of Housing Economics*, *183*, 160–183. https://doi.org/10.1006/jhec.1999.0246

Bourassa, S. C., Hoesli, M., & Peng, V. S. (2003). Do housing submarkets really matter? *Journal of Housing Economics*, *12*, 12–28. https://doi.org/10.1016/S1051-1377(03)00003-2

Bourassa, S., Hoesli, M., & MacGregor, R. D. (1997). *Defining residential submarkets: evidence from Sydney and Melbourne* (Working Papers). Hautes Etudes Commerciales, Universite de Geneve.

Burhan, B. B. (2014). *Spatial mechanism of hedonic price functions for housing submarket analysis* [Doctoral dissertation, Saga University]. Japan.

Chen, J. H., Ji, T., Su, M. C., Wei, H. H., Azzizi, V. T., & Hsu, S. C. (2021). Swarm-inspired data-driven approach for housing market segmentation: a case study of Taipei city. *Journal of Housing and the Built Environment*, *36*(4), 1787–1811. https://doi.org/10.1007/s10901-021-09824-1

Chen, M., Chun, Y., & Griffith, D. A. (2023). Delineating housing submarkets using space – time house sales data: spatially constrained data-driven approaches. *Journal of Risk and Financial Management*, *16*, 291. https://doi.org/10.3390/jrfm16060291

Chen, Z., Cho, S.-H., Poudyal, N., & Roberts, R. K. (2009). Forecasting housing prices under different market segmentation assumptions. *Urban Studies*, *46*(1), 167–187. https://doi.org/10.1177/0042098008098641

Chun-Chang, L., Chi-Ming, L., & Hui-Chuan, H. (2020). The impact of a mass rapid transit system on neighborhood housing prices: an application of difference-in-difference and spatial econometrics. *Real Estate Management and Valuation*, *28*(1), 28–40. https://doi.org/10.2478/remav-2020-0003

Copiello, S. (2020). Spatial dependence of housing values in Northeastern Italy. *Cities*, *96*, 102444. https://doi.org/10.1016/j.cities.2019.102444

Costa, O., Fuerst, F., & Mendes-Da-Silva, W. (2016, August). Office market segmentation in emerging markets: a study of Sao Paulo. *SSRN*. https://doi.org/10.2139/ssrn.2831615

Dale-Johnson, D. (1982). An alternative approach to housing market segmentation using hedonic price data. *Journal of Urban Economics*, *11*(3), 311–332. https://doi.org/10.1016/0094-1190(82)90078-X

Deryol, E. (2019). A hedonic analysis of price movements in commercial properties in the retail sector. In *International Conference on Real Estate Statistics* (pp. 1–10). Luxembourg.

Fell, H., & Kousky, C. (2015). The value of levee protection to commercial properties. *Ecological Economics*, *119*, 181–188. https://doi.org/10.1016/j.ecolecon.2015.08.019

Fuerst, F., & Marcato, G. (2012). Re-thinking commercial real estate market segmentation. *SSRN*, *44*(0). https://doi.org/10.2139/ssrn.1692953

Gabrielli, L., Giuffrida, S., & Trovato, M. R. (2017). Gaps and overlaps of urban housing sub-market: hard clustering and fuzzy clustering approaches. In *Appraisal: from theory to practice* (pp. 203–219). Springer. https://doi.org/10.1007/978-3-319-49676-4

Gnat, S. (2019). Spatial weight matrix impact on real estate hierarchical clustering in the process of mass valuation. *Oeconomia Copernicana*, *10*(1), 131–151. https://doi.org/10.24136/oc.2019.007

Goodman, A. C., & Thibodeau, T. G. (1998). Housing market segmentation. *Journal of Housing Economics*, *7*, 121–143. https://doi.org/10.1006/jhec.1998.0229

Goodman, A. C., & Thibodeau, T. G. (2003). Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics*, *12*(3), 181–201. https://doi.org/10.1016/S1051-1377(03)00031-7

Goodman, A. C., & Thibodeau, T. G. (2007). The spatial proximity of metropolitan area housing submarkets. *Real Estate Economics*, *35*(2), 209–232. https://doi.org/10.1111/j.1540-6229.2007.00188.x

Hayles, K. (2006). The use of GIS and cluster analysis to enhance property valuation modelling in rural Victoria. *Journal of Spatial Science*, *51*(2), 19–31. https://doi.org/10.1080/14498596.2006.9635078

He, S. Y. (2020). Regional impact of rail network accessibility on residential property price: modelling spatial heterogeneous capitalisation effects in Hong Kong. *Transportation Research Part A: Policy and Practice*, *135*, 244–263. https://doi.org/10.1016/j.tra.2020.01.025

Helbich, M., Brunauer, W., Hagenauer, J., & Leitner, M. (2013). Data-driven regionalization of housing markets. *Annals of the Association of American Geographers*, *103*(4), 871–889. https://doi.org/10.1080/00045608.2012.707587

Hu, J., Xiong, X., Cai, Y., & Yuan, F. (2020). The ripple effect and spatiotemporal dynamics of intra-urban housing prices at the submarket level in Shanghai, China. *Sustainability*, *12*(12), 6–11. https://doi.org/10.3390/su12125073

Inoue, R., Ishiyama, R., & Sugiura, A. (2018). Identification of geographical segmentation of the rental apartment market in the Tokyo metropolitan area. In *10th International Conference on Geographic Information Science* (*GISience 2018*) (pp. 1–6). Germany.

Inoue, R., Ishiyama, R., & Sugiura, A. (2020). Identifying local differences with fused-MCP: an apartment rental market case study on geographical segmentation detection. *Japanese Journal of Statistics and Data Science*, *3*, 183–214. https://doi.org/10.1007/s42081-019-00070-y

Kauko, T., Hooimeijer, P., & Hakfoort, J. (2002). Capturing housing market segmentation: an alternative approach based on neural network modelling. *Housing Studies*, *17*(6), 875–894. https://doi.org/10.1080/02673030215999

Ke, Q., Sieracki, K., & White, M. (2017). A spatial analysis of the central London office market. In *24th Annual European Real Estate Society Conference* (pp. 1–16). Netherlands.

Keskin, B. (2008). Hedonic analysis of price in the Istanbul housing market. *International Journal of Strategic Property Management*, *12*(2), 125–138. https://doi.org/10.3846/1648-715X.2008.12.125-138

Keskin, B., & Watkins, C. (2017). Defining spatial housing submarkets: exploring the case for expert delineated boundaries. *Urban Studies*, *54*(6), 1446–1462. https://doi.org/10.1177/0042098015620351

Kopczewska, K., & Ćwiakowski, P. (2021). Spatio-temporal stability of housing submarkets. Tracking spatial location of clusters of geographically weighted regression estimates of price determinants. *Land Use Policy*, *103*, 105292. https://doi.org/10.1016/j.landusepol.2021.105292

Le Gallo, J., López, F. A., & Chasco, C. (2020). Testing for spatial group-wise heteroskedasticity in spatial autocorrelation regression models: Lagrange multiplier scan tests. *Annals of Regional Science*, *64*(2), 287–312. https://doi.org/10.1007/s00168-019-00919-w

Leishman, C., Costello, G., Rowley, S., & Watkins, C. (2013). The predictive performance of multilevel models of housing sub-markets: a comparative analysis. *Urban Studies, 50*(6), 1201–1220. https://doi.org/10.1177/0042098012466603

Levkovich, O., Rouwendal, J., & Brugman, L. (2018). Spatial planning and segmentation of the land market: the case of the Netherlands. *Land Economics*, *94*(1), 137–154. https://doi.org/10.3368/le.94.1.137

Li, K. J., Zhou, Y., Shrestha, A., & Liu, G. W. (2018). A cluster analysis of real estate business models in China. In K. Chau, I. Chan, W. Lu, & C. Webster (Eds.), *Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate* (pp. 1–9). Springer. https://doi.org/10.1007/978-981-10-6190-5_1

Lisi, G. (2019). Property valuation: the hedonic pricing model – location and housing submarkets. *Journal of Property Investment* & *Finance*, *37*(6), 589–596. https://doi.org/10.1108/JPIF-07-2019-0093

Małkowska, A., & Uhruska, M. (2019). Towards specialization or extension? Searching for valuation services models using cluster analysis. *Real Estate Management and Valuation*, *27*(4), 27–38. https://doi.org/10.2478/remav-2019-0033

Mayer, M., Bourassa, S. C., Hoesli, M., & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research*, *12*(1), 134–150. https://doi.org/10.1108/JERER-08-2018-0035

Młodak, A. (2020). k-means, ward and probabilistic distance-based clustering methods with contiguity constraint. *Journal of Classification*, *38*, 313–352. https://doi.org/10.1007/s00357-020-09370-5

Mooi, E., Sarstedt, M., & Mooi-Reci, I. (2018). *Market research: the process*, *data*, *and methods using Stata*. Springer Nature Singapore Pte Ltd. https://doi.org/10.1007/978-981-10-5218-7

Morales, J., Stein, A., Flacke, J., & Zevenbergen, J. (2020). Predictive land value modelling in Guatemala City using a geostatistical approach and Space Syntax. *International Journal of Geographical Information Science*, *34*(7), 1451–1474. https://doi.org/10.1080/13658816.2020.1725014

Morawakage, P. S., Earl, G., Liu, B., Roca, E., & Omura, A. (2023). Housing risk and returns in submarkets with spatial dependence and heterogeneity. *Journal of Real Estate Finance and Economics*, *67*, 695–734. https://doi.org/10.1007/s11146-021-09877-7

Palm, R. (1978). Spatial segmentation of the urban housing market. *Economic Geography*, *54*(3), 210–221. https://doi.org/10.2307/142835

Pryce, G. (2013). Housing submarkets and the lattice of substitution. *Urban Studies*, *50*(13), 2682–2699. https://doi.org/10.1177/0042098013482502

Raposo, I. G., & Evangelista, R. (2017). A transactions-based commercial property price index for Portugal. *Financial Stability Papers*, *3*, 1–25.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55. https://doi.org/10.1086/260169

Roubi, S., & Ghazaly, A. (2007). Pricing inter-neighbourhood variation: a case study on the rental apartment market in Greater Cairo. *Property Management*, *25*(1), 68–79. https://doi.org/10.1108/02637470710723263

Schnare, A. B., & Struyk, R. J. (1976). Segmentation in urban housing markets. *Journal of Urban Economics*, *3*(2), 146–166. https://doi.org/10.1016/0094-1190(76)90050-4

Seo, K. (2016). *Impacts of transportation investment on real property values: an analysis with spatial hedonic price models* [Doctoral dissertation]. http://search.proquest.com.ezaccess.library.uitm.edu.my/docview/1793940515?accountid=42518

Shi, D., Guan, J., Zurada, J., & Levitan, A. S. (2015). An innovative clustering approach to market segmentation for improved price prediction. *Journal of International Technology and Information Management*, *24*(1), 15–32. https://doi.org/10.58729/1941-6679.1033

Sobrino, J. (2014). Housing prices and submarkets in Mexico City: a hedonic assessment. *Estudios Económicos*, *29*(1), 57–84.

Soguel, N., Martin, M., & Tangerini, A. (2008). The impact of housing market segmentation between tourists and residents on the hedonic price. *Swiss Journal of Economics and Statistics*, *144*(4), 655–678. https://doi.org/10.1007/BF03399270

Straszheim, M. R. (1975). Housing-market compartmentalization and housing prices. In *An econometric analysis of the urban housing market* (pp. 28–77). NBER.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240. https://doi.org/10.1126/science.ns-13.332.462

Usman, H., & Lizam, M. (2020, August 10–14). Empirical modelling of commercial property market location submarket using hedonic price model in Malaysia. In *Proceedings of the 5th NA International Conference on Industrial Engineering and Operations Management* (pp. 3396–3406). Detroit, Michigan, USA.

Usman, H., Lizam, M., & Adekunle, M. U. (2020a). Property price modelling, market segmentation and submarket classifications: a review. *Real Estate Management and Valuation*, *28*(3), 24–35. https://doi.org/10.1515/remav-2020-0021

Usman, H., Lizam, M., & Burhan, B. (2021). A priori spatial segmentation of commercial property market using hedonic price modelling. *Real Estate Management and Valuation*, *29*(2), 16–28. https://doi.org/10.2478/remav-2021-0010

Usman, H., Lizam, M., & Burhan, B. B. (2020b). Explicit location modelling of commercial property market using spatial econometric approaches: a review. In *10th International Real Estate Research Symposium* (*IRERS 2020*). Selangor, Malaysia.

Watkins, C. (1999). Property valuation and the structure of urban housing markets. *Journal of Property Investment* & *Finance*, *17*(2), 157–175. https://doi.org/10.1108/14635789910258543

Wu, C., & Sharma, R. (2012). Housing submarket classification: the role of spatial contiguity. *Applied Geography*, *32*(2), 746–756. https://doi.org/10.1016/j.apgeog.2011.08.011

Wu, C., Ye, X., Ren, F., & Du, Q. (2018). Modified data-driven framework for housing market segmentation. *Journal of Urban Planning Development*, *144*(4), 04018036. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000473

Wu, Y., Wei, Y. D., & Li, H. (2020). Analyzing spatial heterogeneity of housing prices using large datasets. *Applied Spatial Analysis and Policy*, *13*(1), 223–256. https://doi.org/10.1007/s12061-019-09301-x

Xiao, Y., Webster, C., & Orford, S. (2016). Can street segments indexed for accessibility form the basis for housing submarket delineation? *Housing Studies*, *31*(7), 829–851. https://doi.org/10.1080/02673037.2016.1150433

Yuan, F., Wei, D. Y., & Wu, J. (2020). Amenity effects of urban facilities on housing prices in China: accessibility, scarcity, and urban spaces. *Cities*, *96*, 102433. https://doi.org/10.1016/j.cities.2019.102433