INTERNATIONAL JOURNAL OF SUSTAINABLE CONSTRUCTION ENGINEERING AND TECHNOLOGY
ISSN: 2180-3242 e-ISSN: 2600-7959

IJSCET

Vol. 15 No. 1 (2024) 41-47
https://publisher.uthm.edu.my/ojs/index.php/ijscet

Defining the Existence of Housing Submarkets for Terraced Properties in Johor Bahru

Wendy Lim Wen Xin[1*], Burhaida Burhan[1], Mohd Lizam Mohd Diah[1]

[1] *Department of Real Estate Management, Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, MALAYSIA*

*Corresponding Author: limwendywenxin@gmail.com
DOI: https://doi.org/10.30880/ijscet.2024.15.01.004

Article Info

Received: 28 November 2023
Accepted: 05 December 2023
Available online: 26 February 2024

Keywords

housing submarkets, housing attributes, principal component analysis, cluster analysis, spatial analysis

Abstract

Recent research on housing submarkets over the decades and the necessity to categorise heterogeneous property markets into distinct submarkets are deemed essential in the real estate field. Housing submarkets operate based on the concept of a group of dwellings that are substantially close substitutes to one another, but are relatively poor substitutes for dwellings in different submarkets. This mechanism primarily depends on spatial variations in housing characteristics related to structure, location, and neighbourhood. Housing submarkets are vital in enhancing the precision of housing pricing and analysing the dynamic shifts in the property markets. This study aimed to define and examine the presence of housing submarkets for the terraced properties in Johor Bahru with the Malaysian Valuation and Property Services Department (JPPH) housing transaction database. The database documented 63,036 housing transactions between 2009 and 2018 with thirteen independent structural and locational housing attributes. The study also recommended an objective data-driven methodology by integrating principal component analysis (PCA) and cluster analysis to delineate housing submarkets. Resultantly, the presence of housing submarkets was identified within the property market with the delineation of two specific submarket effects, namely the housing location cluster and housing quality cluster. The submarkets provided a comprehensive framework for segregating property traits into prospective clusters. Establishing housing submarkets may provide valuable insights, implicit cluster details, and a more reliable hedonic price model prediction. The outcomes could be employed by property valuers, analysts, and urban planners for a sound comprehension of current real estate circumstances.

1. Introduction

Property markets substantially facilitate national economies with their feedback relationship. Based on Chin and Chau (2003), the high demand for housing properties would catalyse multiple economic sub-sectoral advancement and property market expansion across urban and suburban areas (Navaneethan, Ismail and Mar Iman, 2009). Given that housing is uniquely heterogeneous following its type, building size, location, neighbourhood, and other relevant characteristics (Thevendran and Jarnell, 2019; Mohd Diah *et al.*, 2014; Yang, 2017), one rule does not apply to all due to subtle variances between each housing. In line with past studies, housing prices could be typically modelled and summarised with physical, spatial, and environmental elements using the hedonic price model (Bourassa, Hoesli and Peng, 2003; Chen *et al.*, 2009; Feng and Jones, 2016). Nevertheless, the model might reflect aggregation bias as housing markets are regarded as aggregates (Rosmera and Mohd Diah, 2016) with assumptions that the coefficient estimates hold constant without considering location-oriented differences (Watkins, 2001; Usman, Mohd Diah and Adekunle, 2020). For example, the hedonic price model would not tabulate the actual heterogeneity of housing market attributes following the spatial variance of housing prices. Likewise, Watkins (1999) elaborated that housing markets portray attributes of durability, heterogeneity, and spatial fixity. Adair, Berry and McGreal (1996) also elaborated on the improbability of considering housing markets in any geographical location as one entity.

As the necessity to address an optimal statistical method in modelling housing prices is often debatable in real estate studies, a thorough examination and sound understanding of current property market versatility and structure offer useful insights into and precision in modelling housing prices (Rosmera and Mohd Diah, 2016). Heterogeneous housing in spatial form could be differentiated by the following attributes: structural, locational, and neighbourhood. In Warren, Elliott and Staines (2017) and Keskin (2008), divergent housing characteristics require a holistic consideration of specific property components with market segmentation or housing submarket analysis.

Market segmentation denotes property market disaggregation or delineation into novel submarkets: homogeneous within the submarket and heterogeneous in comparison to other groups (Wu and Sharma, 2012). Strategically, housing submarkets offer an accurate and specified housing price structure as the segmentation significantly revises the reliability of housing price prediction models and allocates explicit or precise forecasting for housing prices (Burhan, 2014). For example, Bourassa, Cantoni and Hoesli (2007) and Goodman and Thibodeau (2007) asserted that market segmentation optimally improves property price accuracy and prediction.

Various techniques were highlighted in past studies for housing submarket delineation. Previous classification methods denoted pre-defined boundary and data-driven techniques that adopt statistical approaches: PCA, cluster analysis, factor analysis, and spatial econometrics (Bourassa, Cantoni and Hoesli, 2007). As such, this research strived to provide an overview of housing submarkets, adequate techniques to delineate housing submarkets, and definitions of housing submarkets for the terraced properties in Johor Bahru.

2. Housing Submarkets

2.1 Definition

Technically, housing submarkets involve classifying similar data sample characteristics into novel submarkets or clusters. Such collective attributes share similar characteristics and differ in comparison to other groups. Likewise, Bourassa et al. (1999) denoted housing submarkets as "*a set of dwellings that are reasonably close substitutes of one another, but relatively poor substitutes for dwellings in other submarkets*". Bourassa, Hoesli and Peng (2003) added that housing submarkets incorporate substitutability and equilibrium concepts.

Palpably, housing submarkets demonstrate the substitutability patterns associated with housing prices and structural, locational, and neighbourhood attributes (Burhan, 2014). Substitutability essentially involves paired goods (housing) where an increase on one side catalysed an increase on the other side. Housing is deemed to be an optimal substitute in sharing similar structural attributes, such as property type and building size and age and locational attributes: situated close to one another or located within the same residential scheme.

Housing submarkets also adopt the equilibrium concept for property market delineation as housing price attributes are assumably constant across substitutes. Given the close substitute within submarkets, Goodman and Thibodeau (2007) defined housing price uniformity within the group while Leishman (2009) proposed that housing price could vary from other submarkets.

2.2 The Advantages of Housing Submarkets

As housing price differs in each submarket due to the spatial segmentation of property markets, it is vital to determine the presence of a submarket for accurate housing prices. Based on Wu *et al*. (2018), spatial variations in housing prices could be tabulated by determining the right housing market segmentation. Thus, housing submarket delineation provides a sound comprehension of neighbourhood effects to monitor versatile housing market shifts and establish a reliable housing price forecast in the market (Manganelli *et al*., 2014). Schnare and Struyk (1976) pioneered the discovery of housing submarkets and proved the presence of housing segmentation with Rosen's (1974) hedonic price model.

Following Goodman and Thibodeau (2003), the incorporation of housing submarkets and the hedonic price model alleviates aggregation bias. Specifically, aggregation bias occurs when homogeneously-distinct submarket groups are forcefully integrated with the entire market. The integration results in particular concerns involving incongruent parameter estimates and poor model fit.

As such, housing submarkets offer a better prediction for housing price structure as the segmentation process holistically examines the whole market, generates price fluctuations for every geographical location, and develops an optimal and accurate housing price forecast model (Gabrielli, Giuffrida and Trovato, 2017; Pryce, 2013).

2.3 The Methods to Delineate Housing Submarkets

Specific methods are implemented to empirically delineate housing submarkets by incorporating demand and supply factors and structural, spatial, and neighbourhood characteristics (Alkay, 2008). Based on extensive discussions from past studies, housing submarkets are typically derived from past categorisation or data-driven methods.

In line with past classifications, housing submarkets are segmented by spatial divisions or spatially-contiguous boundaries that are subsequently categorised based on distinct criteria (Wu and Sharma, 2012). For example, past studies examined housing submarkets with spatial division features: geographical boundaries (Bourassa, Hoesli and Peng, 2003), aggregated census blocks (Goodman and Thibodeau, 2003; 2007), zip or postal codes (Goodman and Thibodeau, 2003; 2007; Leishman, 2009), administrative areas and local government boundaries (Bourassa, Hoesli and Peng, 2003; Adair, Berry and McGreal, 1996), and physical features (Watkins, 2001). Additionally, Schnare and Struyk (1976) derived housing submarkets from past methods using socioeconomics and demographic components. As past categorisations could prove inadequate concerning precision, accuracy, and objectivity, this explicit method was derived from intuitive knowledge and some pre-defined views of present criteria (Watkins, 2001; Wu *et al*., 2018). As such, past classifications failed to sufficiently tabulate the housing submarket dynamics.

Housing submarkets could also be delineated with an objective approach (data-driven methodology) which incorporates a wide array of housing attributes and categorises the dwellings based on specific attributes. Data-driven methodology enables pertinent figures and datasets to identify housing submarket structures. Scholars could also assess the most distinguished housing components from the large dataset (Wu *et al*., 2018). This objective technique seemed more precise and reflective on housing submarkets based on time and location. Following Wu and Sharma (2012), data-driven classification could expand temporal housing submarket dynamics as real estate data are periodically updated per quarter and annum.

In line with past research, the most extensively-utilised data-driven methodologies employed for housing submarket identification encompass factor analysis (Watkins, 1999), PCA (Bourassa, Cantoni and Hoesli, 2007; Bourassa *et al*., 1999), hierarchical clustering (Goodman and Thibodeau, 2003), and clustering (Hwang and Thill, 2009). Other data-driven methodologies include classification regression trees (Fan, Ong and Koh, 2006), neural networks (Kauko, 2004), and partitioning algorithms (Bourassa *et al*., 1999; Chen *et al*., 2009). Furthermore, Day (2003) attempted to assess housing submarkets by incorporating partitioning algorithms into the hierarchical model. The hedonic price model was subsequently utilised to assess the novelty of housing submarkets. Typically, this literature review asserted that housing submarket delineation strived to optimise housing price prediction albeit with distinct methods and case studies involving the superiority of one over another.

3. Methodology

This study aimed to develop and evaluate housing submarkets for the terraced properties in Johor Bahru with the aforementioned data-driven methodology by recognising the essentiality and benefits of housing segmentation. The statistically-robust methodology has been extensively-utilised in Bourassa, Hoesli and Peng (2003), Bourassa, Cantoni and Hoesli (2007), and Bourassa *et al*. (1999) and applies to various case studies for a sound understanding of housing submarket delineation complexities.

The study dataset obtained from JPPH encompassed 63,036 sale records of single-storey and double-storey terraced properties in Johor Bahru between 2009 and 2018. This housing transaction dataset encompassed 13 independent structural and locational housing attributes. Practically, this research implemented two primary statistical methods to obtain housing submarkets. Initially, PCA was performed to attain dimension-reduction and deduce a set of principal factors from original variables or datasets (Wu and Sharma, 2012). The principal components originating from PCA was subsequently employed in cluster analysis to identify the number of clusters and housing submarket composition. A two-step and *K-means* cluster analysis offer useful insights into housing submarket delineation.

$$\rho = \frac{\vec{E}}{J_c(T = \text{const.}) \cdot \left(P \cdot \left(\frac{\vec{E}}{E_c} \right)^m + (1 - P) \right)} \quad (1)$$

4. Analysis

4.1 Principal Component Analysis (PCA)

The PCA or dimension-reduction method is a statistical procedure developed to convert a set of highly-correlated predictors into groups that consist of significant and informative variables (Bourassa, Hoesli and Peng, 2003; Bourassa *et al.*, 1999). In Burhan (2014), the submarket strategy functions by determining and pooling similar data sample attributes into specific groups. The PCA distinguishes significant and vital predictors while omitting unnecessary loads from the predictor list.

The PCA in this research elicited a set of structural and locational attributes from 13 original variables in the JPPH housing transaction data. This process proves crucial to determine substantial and principal components within the dataset. The researcher performed PCA with SPSS twice until all the inserted structural and locational attributes attained the stopping-rule prerequisites. Specifically, this analysis obtained the aforementioned prerequisite when all the input variable values exceeded 0.500 in both anti-image matrix analyses. Although 13 input variables were initially incorporated into the analysis, not all of them were retained. Through PCA, the most significant components that conducted the most measurement out of the original dataset were retained. Resultantly, 8 variables were adequately delineated as principal study variables. Table 1 presents how the value of every component attribute is connected. The rotated factor loadings defining the retained variable-derived component relationship and how each variable was weighted for relevant components were duly documented.

The researcher then summarised the principal components in table 2 based on the derived solution and its suitability outcome for every variable in table 1.

Table 1 *Rotated Component Matrix*

	Component		
	1	2	3
Subdistrict	.033	.089	.899
Building_Size	.802	.168	.161
No_Bedroom	.653	-.025	.322
Property_Type	.917	-.070	-.078
No_Storey	.926	-.041	-.060
Building_Age	-.473	.457	-.460
Area_Classification	.061	.813	-.143
Area_Category	.000	.782	.330

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization [a]
a. Rotation converged in 5 iterations.

4.2 Cluster Analysis

The PCA serves to elicit a reduced set of orthogonal variables from the original dataset while cluster analysis subsequently identifies the most adequate housing submarket composition. Cluster analysis is employed to classify observations or clusters following similar dataset patterns. This analysis incorporated similar observations to develop homogeneity within the cluster and heterogeneity between every categorised group (Yang, 2017). Summarily, cluster analysis strives to optimise statistically-generated housing submarket outcomes.

The eight variables identified in PCA (see table 1) were utilised to produce an optimal number of housing submarkets. Two cluster analysis forms were implemented in this study: two-step and *K-means* clustering analysis. Two-step clustering, which aimed to determine the optimal number of housing submarket clusters, delineated two final clusters with 39,550 observations in the first cluster and 23,486 observations in the second counterpart. This technique divides the entire dataset into two steps, which begin with pre-clustering the observations into several sub-clusters first and are followed by grouping the result of the first step into a desirable number of housing submarket clusters. Overall, the two-step clustering model reflected good cluster quality.

Following the optimal number of housing submarket clusters elicited from two-step clustering analysis, the *K-means* counterpart subsequently assigned the dataset observations to its best-fit cluster. The *K-means* cluster analysis efficiently processed large data encompassing 63,036 observations based on the two preliminary clusters following two-step clustering analysis. The *K-means* clustering analysis outcomes are presented in table 3.

Table 2 *Principal Components Derived*

Component	Variables	Definitions of Components
1st Component	Building Size No. of Bedroom Property Type Number of Storey Building Age	**Housing Quality** Structural attributes, such as building size and number of storeys and bedrooms are comparatively different for each property type. Building age also significantly distinguishes housing prices. Past studies implied that older homes reflect larger building sizes while recently-built homes are smaller following limited residential land use.
2nd Component	Area Classification Area Category	**Housing Location** Locational attributes and house categories (rural, secondary rural, secondary central, primary rural, and primary central areas) determine the quality of housing locations as a substantial characteristic in housing price analysis.
3rd Component	Subdistrict	**Housing Subdistricts** Sub-district attributes may reflect on area development. Sub-districts with low developmental and economic activities may reflect comparatively lower housing prices.

Table 3 *Results of K-Means Cluster Analysis*

	Cluster	
	1	**2**
REGR factor score 1 for analysis 2	-.38317	.29168
REGR factor score 2 for analysis 2	.87993	-.66983
REGR factor score 3 for analysis 2	.15136	-.11522

The final cluster centers from the *K-means* cluster analysis were computed as the mean value for each final-cluster variable. The final cluster centers aimed to highlight the attributes of typical observations for each cluster. In table 3, the factor score for Component 1 (housing quality) proved far from Cluster 1 (-3.8317) while Component 2 (housing location) was closest to Cluster 1 (0.87993). Contrarily, the factor score for Component 1 was closest to Cluster 2 (0.29168) while Component 2 was farthest from Cluster 2 (-0.66983).

Based on table 3, Component 1 (housing quality) denoted the principal component for Cluster 2 while Component 2 (housing location) implied the principal component for Cluster 1. The principal component in every cluster would reflect immediate cluster reactions to the nearest individual component.

Table 4 *ANOVA for K-Means Cluster Analysis*

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
REGR factor score 1 for analysis 2	7045.048	1	.888	63036	7931.379	.000
REGR factor score 2 for analysis 2	37153.472	1	.411	63036	90486.618	.000
REGR factor score 3 for analysis 2	1099.370	1	.983	63036	1118.866	.000

Based on table 4, the significant level for every component's factor score in *K-means* was satisfactory and significant (p = .000). Two potential housing submarkets (housing quality and housing location clusters) were delineated with PCA and cluster analysis. Based on the study outcomes, a significant and clear presence of housing submarkets was identified for the terraced properties in Johor Bahru within the property market. Extensive analytical discussions on the results proved advantageous for further empirical assessments or research on specific clusters.

5. Conclusion

This study primarily aimed to identify and examine the presence of housing submarkets for the terraced properties in Johor Bahru based on the JPPH housing transaction database. Specific methods were performed with SPSS to achieve the research goals: correlation matrix analysis, Principal Component Analysis, and two-step and *K-means* cluster analyses. The data-driven techniques extensively discussed and justified housing submarket delineation. Eight substantial variables were retained and developed into three primary principal components (housing quality and housing location and housing sub-district area) from the 13 variables derived from the JPPH data. The variables were employed to pre-cluster the observations into several sub-clusters. The sub-clusters (previously delineated in the analysis) mitigated potential bias from the initially-determined number of clusters. The *K-means* cluster analysis subsequently classified pre-cluster observations into final clusters for good model fit.

Summarily, this research asserted the presence of housing submarkets for terraced properties in Johor Bahru with statistical techniques. As housing prices differed in terms of location and its attributes, they could be divided into distinct housing submarkets. The submarkets offered a holistic framework to segregate property attributes into potential clusters. Housing submarket delineation may provide useful insights, implicit cluster details, and a better hedonic price model prediction. Although this research produced outcomes that would benefit all real estate authorities, several limitations were encountered in the number of components. As such, future research may be performed on multiple variable types that were not included in this study: detailed neighbourhood and environmental elements. As the current dataset did not encompass neighbourhood attributes, future studies could consider substitutable variables to assess the presence of residential housing submarkets and optimise the model.

Acknowledgement

Authors would also like to thank the Ministry of Higher Education Malaysia for supporting this research under Fundamental Research Grant Scheme Vot No. FRGS/1/2019/SS08/UTHM/02/1 and partially sponsored by Universiti Tun Hussein Onn Malaysia. In addition, the authors gratefully acknowledge National Information Centre (NAPIC) and Valuation and Property Services Department (JPPH) for providing the data used in this research.

References

Adair, A.S., Berry, J.N. & McGreal, W.S. (1996). Hedonic Modelling, Housing Submarkets and Residential Valuation. Journal of Property Research, 13, pp. 67-83

Alkay, E. (2008). Housing Submarkets in Istanbul. *International Real Estate Review*, *11*, pp. 113-127

Bourassa, S.C. *et al.* (1999). Defining Housing Submarkets. *Journal of Housing Economics*, *8*, pp. 160-183

Bourassa, S.C., Cantoni, E. & Hoesli, M. (2007). Spatial Dependence, Housing Submarkets and House Price Prediction. *Journal of Real Estate Finance Economics*, *35*, pp. 143-160

Bourassa, S.C., Hoesli, M. & Peng, V.S. (2003). Do Housing Submarkets Really Matter? *Journal of Housing Economics*, *12*, pp. 12-28

Burhan, B. (2014). Spatial Mechanism of Hedonic Price Functions for Housing Submarket Analysis. Saga University: Ph.D. Thesis

Chen, Z. *et al*. (2009). Forecasting Housing Prices under Different Market Segmentation Assumptions. *Urban Studies*, *46*, pp. 167-187

Chin, T.L. & Chau, K.W. (2003). A Critical Review of the Literature on the Hedonic Pricing Model and its Application to the Residential Market in Penang. *International Journal for Housing Science and Its Applications*, *27*, pp. 145- 165

Day, B. (2003). Submarket Identification in Property Markets: A Hedonic Housing Price Model for Glasgow. CSERGE Working Paper EDM 03-09 (Norwich: The Centre for Social and Economic Research on the Global Environment)

Fan, G.Z., Ong, S.E. & Koh, H.C. (2006). Determinants of House Price: A Decision Tree Approach. *Urban Studies*, *43*, pp. 2301-2315

Feng, Y. & Jones, K. (2016). Comparing Two Neighbourhood Classifications: A Multilevel Analysis of London Property Price 2011-2014. *Proceeding of the 22nd Annual Pacific-Rim Real Estate Conference (Queensland, Australia)*

Gabrielli, L., Giuffrida, S. & Trovato, M.R. (2017). Gaps and Overlaps of Urban Housing Sub-market: Hard Clustering and Fuzzy Clustering Approaches Appraisal: From Theory to Practice. Italy: Springer

Goodman, A.C. & Thibodeau, T.G. (2003). Housing Market Segmentation and Hedonic Prediction Accuracy. *Journal of Housing Economics*, *12*, pp. 181-201

Goodman, A.C. & Thibodeau, T.G. (2007). The Spatial Proximity of Metropolitan Area Housing Submarkets. *Journal of Housing Economics*, *35*, pp. 209-232

Hwang, S. & Thill, J.C. (2009). Delineating Urban Housing Submarkets with Fuzzy Clustering. *Environment and Planning B: Planning and Design*, *36*, pp. 865-882

Kauko, T. (2004). A Comparative Perspective on Urban Spatial Housing Market Structure: Some More Evidence of Local Sub-markets based on a Neural Network Classification of Amsterdam. *Urban Studies*, *41*, pp. 2555-2579

Keskin, B. (2008). Hedonic Analysis of Price in the Istanbul Housing Market. *International Journal of Strategic Property Management*, *12*, pp. 125-138

Leishman, C. (2009). Spatial Change and the Structure of Urban Housing Sub-markets. *Housing Studies*, *24*, pp. 563-585

Manganelli, B. *et al*. (2014). Using Geographically Weighted Regression for Housing Market Segmentation. *Journal of Business Intelligence and Data Mining*, *9*, pp. 161-177

Mohd Diah, M.L. *et al*. (2014). Comparison of Empirical Models in Malaysia House Price Index Construction. *Proceeding of the 20th Annual Pacific-Rim Real Estate Society Conference (Christchurch)*

Navaneethan, R., Ismail, S. & Mar Iman, A.H. (2009). Testing for the Existence of Housing Sub-markets in Penang, Malaysia. *Malaysian Journal of Real Estate*, 4, pp. 52-70

Pryce, G. (2013). Housing Submarkets and the Lattice of Substitution. *Urban Studies*, *50*, pp. 2682-2699

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*, pp. 34-55

Rosmera, N.A. & Mohd Diah, M.L. (2016). Housing Market Segmentation and the Spatially Varying House Prices. *The Social Sciences*, *11*, pp. 2712-2719

Schnare, A.B. & Struyk, R.J. (1976). Segmentation in Urban Housing Markets. *Journal of Urban Economics*, *3*, pp. 146-166

Thevendran, M. & Jarnell, D. (2019). Solutions to Housing Affordability Issues. The Star Online. Retrieved May 26, 2023, from https://www.thestar.com.my/business/business-news/2019/10/05/solutions-to-housing-affordability-issues

Usman, H., Mohd Diah, M.L. & Adekunle, M.U. (2020). Property Price Modelling, Market Segmentation and Submarket Classifications: A Review. *Journal of Real Estate Management and Valuation*, *28*, pp. 24-35

Warren, C.M.J., Elliott, P. & Staines, J. (2017). The Impacts of Historic Districts on residential Property Land Values in Australia. *International Journal of Housing Markets and Analysis*, *10*, pp. 66-80

Watkins, C.A. (1999). Property Valuation and the Structure of Urban Housing Markets. *Journal of Property Investment and Finance*, *17*, pp. 157-175

Watkins, C.A. (2001). The Definition and Identification of Housing Submarkets. *Journal of Environment and Planning*, *33*, pp. 2235-2253

Wu, C.S. & Sharma, R. (2012). Housing Submarket Classification: The Role of Spatial Contiguity. *Journal of Applied Geography*, *32*, pp. 746-756

Wu, C. *et al*. (2018). Modified Data-Driven Framework for Housing Market Segmentation. *Journal of Urban Planning Development*, *144*, pp. 1-15

Yang, X. (2017). *Urban Morphology and Housing Market*. Singapore: Springer Geography