



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/186026/>

Version: Published Version

Article:

Keskin, B. (2022) Multilevel approach to the analysis of housing submarkets. *Regional Studies, Regional Science*, 9 (1). pp. 264-279. ISSN: 2168-1376

<https://doi.org/10.1080/21681376.2022.2067005>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Multilevel approach to the analysis of housing submarkets

Berna Keskin

To cite this article: Berna Keskin (2022) Multilevel approach to the analysis of housing submarkets, *Regional Studies, Regional Science*, 9:1, 264-279, DOI: [10.1080/21681376.2022.2067005](https://doi.org/10.1080/21681376.2022.2067005)

To link to this article: <https://doi.org/10.1080/21681376.2022.2067005>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 27 May 2022.



Submit your article to this journal [↗](#)



Article views: 303



View related articles [↗](#)



View Crossmark data [↗](#)

Multilevel approach to the analysis of housing submarkets

Berna Keskin

ABSTRACT

There is a vast literature that seeks to define and identify spatial submarkets in metropolitan housing systems. These tend to use one of three methods to delineate submarkets: a priori geographies, ad hoc subdivision and data-driven approaches to grouping units. Recently, analysts have increasingly used multilevel modelling strategies to analyse spatial segmentation in the housing market. Despite the increasing prevalence of multilevel approaches, there is no existing systematic analysis of which of these three main approaches to submarket definition has the greatest effectiveness when employed in a multilevel modelling framework. This paper addresses the gap in the literature by comparing the utility of these main approaches to submarket definition. It develops and evaluates three separate, distinct multilevel models of submarkets to a data set comprising 2175 transactions in the Istanbul housing market of Turkey, an emergent market context. The results show that multilevel models with a priori submarket dummy variable can predict price more accurately than the models with ad hoc subdivision or data-driven stratified submarkets. Similarly, test results indicate that multilevel models with neighbourhood submarket dummy variables (a priori) perform better than other models. These test results show that granular definition of submarkets tend to perform better in terms of predictive accuracy than less spatially granular models. The paper also suggests that real estate agents' views of submarket structures might be particularly useful as inputs into micro-modelling processes in contexts where datasets are thin.

ARTICLE HISTORY

Received 13 March 2021; Accepted 2 March 2022

KEYWORDS

Housing submarkets; multilevel modelling; housing prices; Istanbul

INTRODUCTION

A substantial literature argues that house price models should accommodate housing submarkets for conceptual and technical reasons (see Watkins, 2012, for a review). From a conceptual perspective, it is argued that submarkets are a function of the multiple equilibrium (or, in some cases, disequilibrium) nature of the market (Goodman & Thibodeau, 1998; Maclennan et al., 1987; Watkins, 2001). In practical terms, it has been shown that submarkets must be taken into account to avoid aggregation bias (Straszheim, 1975) and enhance the predictive performance of house price models (Leishman et al., 2013). The vast majority of papers that take account of submarkets do so by either incorporating them within hedonic models (Fletcher et al., 2000) or by estimating submarket-specific hedonic models (Watkins, 2001). Importantly, however,

CONTACT

(Corresponding author)  b.keskin@sheffield.ac.uk

Department of Urban Studies and Planning, University of Sheffield, Sheffield, UK

an increasingly voluminous series of studies have begun to advocate the inclusion of submarkets within a multilevel modelling framework (see Goodman & Thibodeau, 1998, for an early example; also Orford, 2002; Leishman, 2009; Keskin, 2010). Significantly, Leishman et al. (2013) conducted an experiment that compares the performance of different approaches to modelling submarkets and demonstrates that multilevel techniques can improve model accuracy. These findings emerged from a systematic comparison of different approaches. Specifically, they evaluated the use of a simple hedonic augmented with dummy variables to capture submarket effects, a system of submarket specific hedonic equations, and a multilevel model specification with submarket levels included by comparing the likelihood of predicting prices within 10% of the actual price, and the average size of the standard errors of the model estimates to assess the utility of the modelling approaches.

An immense number of studies illustrate alternative partitioning methods for submarkets in hedonic modelling analysis. The last three decades have seen a growing body of research on the submarket delineation that employ multilevel models as an important strand of this literature. These multilevel studies have also been used to test the performance of alternative submarket definitions. Unlike the broader literature, however, there is a lack of research comparing the effectiveness of methods of allocating dwellings with submarkets or of their performance when used in multilevel modelling of housing prices. This study attempts to fill this gap by testing different submarket delineation approaches in a multilevel modelling framework as part of a systematic comparison of different modelling procedures.

The empirical analysis focuses on the Istanbul housing market of Turkey. In doing so, the analysis throws up a challenge common to many studies conducted in emergent property market contexts in that the datasets are thin in terms of both the number of observations and the number of hedonic variables for which data exist. In this context, although there are already several studies examining submarket identification and delineation as a prior stage in hedonic house price modelling procedures, this paper seeks to be distinctive in two ways. First, by comparing different methods of identifying submarket units for analysis, the paper offers a systematic review of which submarket description performs best when used in a multilevel modelling setting. Such comparisons of multilevel models are rare in the literature. Second, with the emergent market data limitations in mind, the use of real estate experts' views as one means of identifying submarket units as a prior step to modelling is explored (see also Palm, 1978; Watkins, 2001). This is of interest because it offers a shortcut to identifying submarkets where data-driven approaches are limited due to the paucity of market information. This has arguably been one of the constraints on the greater use of multilevel approaches in transitional and emergent market contexts.

Thus, this paper seeks to empirically explore the utility of the multilevel modelling framework further by looking at how it works when used with the three major approaches to submarket delineation. Furthermore, by using the Istanbul housing market as a case study in which to undertake the predictive performance of the multilevel modelling, the paper is also able to offer comments on the relative merits of the use of experts' views as part of the analytical process where data limitations exist.

The remainder of this paper is organized as follows. The next section reviews the concept of housing market segmentation and explores how it leads to the formation of housing submarkets. It focuses on the delineation of submarkets and also provides a brief review of studies that accommodate submarkets in multilevel models of metropolitan house prices. The paper then summarizes the research design. It introduces the case study area, provides an overview of the data, and explains the rationale for using multilevel models and the method used to estimate the models. The paper then sets out the results and compares the model performance when each of the three main approaches to submarket delineation are used. The final section summarizes the findings and offers some concluding remarks.

HOUSING SEGMENTATION AND FORM OF SUBMARKETS

Since housing is a multifunctional composite concept, all kinds of investments, actions, interventions and policies about housing affect the built environment as well as the socio-economic environment. This differentiation has an impact on both supply and demand side of the market, which also leads to segmentation in the housing market. Housing submarkets – arising from housing market segmentation – are crucial to understanding the operation and structure of the housing system, which is mainly shaped by the interaction of demand and supply dynamics (Maclennan & Tu, 1987).

The housing market activity involves the exchange of dwellings (or housing units) rights between buyers and sellers. This process determines the individual price of housing unit as well as the quantity that will be exchanged and the market level price structure. It is therefore important to understand the factors that influence and underpin demand and supply. The demand side of the housing market is composed of diverse consumer groups that vary based on their socio-economic background, household composition, cultural background, lifestyle preferences and tastes. On the other hand, the supply side can be viewed as consisting of products groups constituting of clusters of similar dwellings with variation between groups reflecting differences in dwelling size, type, quality of construction and location. All these variances, as well as heterogeneous demand, and differentiated supply interaction create complexity in the market structure, which also underpins spatial price differentiation.

In this context, Jones and Watkins (2009) point out that, in theoretical terms, submarket existence can be shown to be consistent with both disequilibrium and multiple equilibria processes. In the former case, a mismatch between demand and supply leads to disequilibrium, which is one of the aforementioned explanations of housing market segmentation. Disequilibrium can result from heterogeneous factors including financial conditions, individuals' preferences, information search costs, as well as from exogenous shocks such as pandemics or wars that change the dynamics of the market processes towards disequilibrium.

According to another explanation for the existence of submarkets – multiple equilibria – each segment has its own equilibrium price; therefore, within each segment there is an equilibrium that is determined by the balance between supply and demand forces (Goodman, 1981). This assumption dominates most of the housing market studies since they acknowledge the existence of housing price differences among submarkets. Thus, most of the studies in the housing market literature assume that a housing market consists of a set of linked submarkets that exhibit supply and demand imbalances, resulting in spatial differences within the market (Watkins, 1998; Orford, 2000). This raises the question of whether a single/overall market analysis – in other words fitting a general model – could be sufficiently reliable in modelling.

In this study, a multiple equilibrium perspective is adopted. It is accepted that the existence of price differentials among market segments is based on the assumption of equilibrium within submarkets. The equilibrium of housing demand and supply in a submarket is achieved by changes to the housing stock and/or the turnover rate in the existing stock. The equilibrium house price will be determined by the sum of the value of individual physical attributes, local amenities and neighbourhood quality typical within the prevalent product group in the market segment.

Definition of housing submarkets

Many studies in the housing literature have suggested that the housing market system is best analysed as a collection of 'functionally independent geographic submarkets differentiated by the characteristics of their housing units and/or the locations of the submarkets' (Rothenberg et al., 1991, p. 63). Although there is a consensus on the theoretical existence of submarkets,

there is insufficient agreement on how to delineate the submarkets in practice. As Watkins (2001) states, there is no single definition of housing submarket, and the precise definition of submarkets has always been contested.

For many scholars, Grigsby's approach on substitutability is considered the basis of submarket definition. Grigsby (1963) explains substitutability in terms of optimization of preferences within a price limitation. Substitutability requires home buyers to be indifferent between the entire bundle of structural, locational and neighbourhood quality attributes that characterize the competing housing units (Watkins, 2001). It has been revealed in many housing studies that both spatial and structural features are significant in leading to the emergence of housing submarkets and shaping their structure (Adair et al., 1996; Maclennan & Tu, 1996; Watkins, 2001; Kauko, 2002; Bourassa et al., 2007). As Evans (1995, pp. 6–7; added emphasis) pointed out:

the buyer is purchasing a property which is a bundle of characteristics. So, in the case of a house, the purchaser buys a location relative, say, to shops and workplaces, fertility in the sense of the quality of environment, also a house where attributes of the house – such as central heating, number of bathrooms, size and number of rooms – *cannot be detached and sold separately*.

Since submarkets are formed by a complex bundle of structural and spatial attributes, analysis through segments/submarkets can provide a better understanding of housing market dynamics than might be possible from an analysis at the aggregate level. Therefore, delineating submarkets and understanding their dynamics is a critical step in analysing housing markets.

Delineating submarkets in practice

Three main approaches are used in the housing market literature to delineate submarkets: a priori geographies, ad hoc subdivision groupings and data-driven classifications. The first approach, which is often labelled a priori submarket delineation, typically uses predefined geographical areas or spatial contiguous boundaries such as administrative, postcode areas, census divisions and school catchment areas. This approach, based on predefined administrative spatial areas, has been widely used by scholars such as Maclennan et al. (1987), Adair et al. (1996), Goodman and Thibodeau (1998), Jones (2002), and Goodman and Thibodeau (2003) in housing market studies. This classification is simple to use and technically accessible, yet it cannot fully capture the dynamics of housing submarket due to its fixed structure.

The second approach is based on the ad hoc allocation of spatial boundaries by market experts such as real estate agents and valuers. Palm (1978) argued that submarkets that are defined by real estate agencies showed better performance than those determined by data or a priori classification. Similarly, Michaels and Smith (1990) used an expert-defined submarket model that yielded the best performance in housing price estimation when compared with alternative a priori submarket constructions. Bourassa et al. (2003) compared a set of submarkets based on geographical areas defined by real estate agents and appraisers with a set of statistically generated submarkets. They also found out that price predictions are more accurate when based on the housing segmentation defined by real estate appraisers than when based on statistical techniques. However, the use of this approach to establish submarket classifications has criticized for its lack of rigour (Costello et al., 2019).

The third common method is a data-driven approach that uses statistical techniques to delineate submarkets. Based on a given set of variables that capture structural and spatial characteristics, the data-driven techniques assign housing units into groups that are statistically shown to be relatively homogeneous within themselves and heterogeneous when compared with each other. As noted above, the nature of housing segmentation particularly depends on homogeneity within submarkets, while heterogeneity between submarkets is also expected. In this

context, maximization of cluster homogeneity is the key principle used to operationalize Grigsby's notion of substitutability. Several studies use statistical methods when defining submarkets, such as cluster analysis (Bourassa et al., 1999), principal component analysis (Watkins, 1998; Bourassa et al., 1999, 2003) and, less commonly, factor analysis or neural network analysis (Kauko, 2003).

The improvements in quantitative research methods and increased availability of large micro-datasets (with geocodes, characteristics and price information) have contributed to conceptual and methodological debate about housing market analysis (Watkins, 2012). The increased accessibility to the more detailed and granular data enables one to delineate submarkets by using different methods that allows to collaborate with hard and soft boundaries. There are some studies using absolute location – $\{x, y\}$ coordinates – to capture the impact of local and non-local externalities on properties (Pavlov, 2000; Clapp and Wang, 2006; Fik et al., 2003). By doing so, researchers could overcome with the obstacles of using a priori defined submarkets with hard boundaries which could 'mask significant value discontinuities' (Fik et al., 2003, p. 625). As an example, Bayer et al. (2007) applied a boundary discontinuity design (BDD) to address the endogeneity of school and neighbourhood characteristics in a heterogenous residential choice model. They identified elementary school attendance zone boundaries by implementing BDD and assigned each block to one of the school zones. Therefore, by combining a priori and data-driven approach, the study made a significant contribution in defining submarket boundaries and analysing the fixed effects of the submarkets whilst taking endogeneity into account.

Elsewhere in the literature, new modes of technology in online housing services and changes in search behaviour have been recognized by housing market studies (Dunning et al., 2019). Advances in collection of data and accessibility have opened up some innovative approaches to delineating the housing submarkets. One such pioneering study by Rae (2015) investigated the performance of 'user-generated search areas' as a form of submarkets and his findings show that this approach could contribute to the analysis of housing markets by providing a deeper, alternative understanding of how housing submarkets are formed and evolve. The influence of broader conceptual and methodological developments has started to become evident in housing submarket studies (see below).

Assessing submarket variation in multilevel models

To have a better understanding of the operation and structure of housing market through models, the conceptualization of submarkets is a fundamental step in the analysis. A number of studies have investigated the dynamics of the housing market by incorporating submarkets into multilevel models over the last three decades. The nature of the data used in housing studies that have a hierarchical structure with levels such as the housing unit is located in a neighbourhood/district/local authority/region. Multilevel models enable the analysis of the housing market at several levels simultaneously, rather than at single level of the data individually. This method is equipped to analyse multiple levels of data and is able to capture relationships between the independent and dependent variables that can vary between places (spatial units). Several housing market studies apply multilevel model specifically to explore spatial dependencies in both micro- (individual) and macro-levels (contextual level).

In one of the earliest housing market studies to employ multilevel models, Jones and Bullen (1993) employed an a priori approach to segmentation to investigate empirically the determinants of housing prices. They used a two-level approach: individual level (housing unit) and submarket level (local authority areas). The individual-level, property characteristics of housing units are examined within the submarket levels which were based on local authority areas in Southern England and London boroughs between 1980 and 1987. The study makes an innovative methodological contribution since it recognizes the distinction between compositional

and contextual effects and provides an improvement in the empirical examination of housing prices by being sensitive to drivers that vary across both space and time.

Bullen et al. (1997) built on this innovation when they investigated the variations in residential prices in England and Wales by using travel time-to-work areas (TTWAs) – as determined by the Office for Statistics (ONS) derived from census data – as an alternative form of a priori definition of submarkets. Results from fitting a multilevel model with the TTWA-driven submarkets indicates that complex heterogeneity can be addressed better at both housing unit and submarket levels. Multilevel modelling appeared in this study to help overcome the problems arising from heterogeneity and spatial autocorrelation and thus offer what is arguably a better contextualized methodology.

Orford (2000, 2002) used a multilevel approach that allowed the compositional effects of the housing stock and the contextual effects of a priori determined submarkets in Cardiff, Wales. Both structural and locational attributes were taken into account and analysed at submarkets level, where the segments were based upon the 26 community units used by the city council for administrative purposes, and also at street level. The multilevel model allowed the impact of submarkets on price differentials to be clearly displayed in ‘an holistic view of the housing market, one that is more comparable to conceptual than the standard single-level specification’ (Orford, 2000, p. 1670). It was also emphasized that a clear theoretical understanding is needed about how spatiality enters in the analytical process (Orford, 2002)

Leishman et al. (2013) focused on spatially segmented submarkets based on real estate agents’ definitions and postcode areas to compare different modelling strategies for Perth, Western Australia. They estimated a series of submarket-specific hedonic models and multilevel models with different spatial levels and compared their performance. The study found that more granular submarket specification (based on postcodes) increased analytical efficiency and decreased the occurrence of spatial autocorrelation.

Keskin et al. (2017) adopted a priori definition of submarkets and used administrative neighbourhood boundaries in their study of changes in the pattern of housing prices in Istanbul, followed by earthquake activity in the wider region. The results revealed discrete spatial impacts on housing prices and showed that the impact of earthquake varied widely at the neighbourhood level. These findings were consistent with Leishman et al.’s (2013) argument that the more granular structure of submarkets tends to be more effective when it comes to the accurate prediction and measurement of house prices and modelling their spatial distribution.

Most recently, Alas (2020) explored the spatial of housing prices by implementing a further variant on the a priori approach to the definition of submarkets. The models were constructed by using submarkets based on administrative district boundaries, neighbourhood boundaries and street level in Istanbul. The study concluded that multilevel models based on smaller geographical spatial units are capable of increasing predictive efficiency in assessing housing price differences.

Most of these previous studies have adopted a two-level approach by using submarkets as a nested (second/contextual) level. They also usually compared the performance or multilevel model against hedonic model. Some of the studies tested the effectiveness of models employing a priori and ad hoc approach in submarket description. This study extends this literature by testing more approaches in submarket delineation, namely ad hoc, a priori and data-driven methods.

RESEARCH DESIGN

The focus of this paper is to investigate the performance of different submarket delineations in modelling the dynamics of a local housing market. The empirical analysis underlying it uses multilevel modelling in order to examine the performance of different types of submarket

delineation. To compare the efficacy of the different approaches in submarket definition, three separate multilevel models are analysed on the basis of a priori, expert-based (ad hoc) and data-driven submarket allocation.

Multilevel modelling in housing studies

The multilevel model approach has been adopted in housing studies in the last three decades with increasing attention since it has the capacity to analyse individual-level characteristics embedded in different spatial levels. This method provides an analysis of the data with a hierarchical structure by using combinations of individual- and group-level independent variables for finding the impact of contextual effects.

As reviewed in the previous section, multilevel modelling is widely applied in housing market studies due to the technical advantages that help to overcome the limitations of the conventional models. Hedonic modelling, a conventional model widely used in housing market studies, is based on fixed effects of the independent variables, whereas multilevel models focus on both fixed effects and random effects of the group level. Furthermore, it is assumed that the effects of housing unit attributes are the same across submarkets in hedonic models; for example, the effect of age of the building is the same within the market regardless of the submarkets/neighbourhood. In comparison multilevel modelling enables the relationship to vary spatially.

Although hedonic modelling is a useful tool for understanding housing markets, technical constraints such as spatial autocorrelation, spatial heterogeneity, ecological fallacy and atomic fallacy led commentators to observe robustness problems in modelling (Malpezzi, 2003). Creating a dummy variable for submarkets is a commonly suggested method to overcome these problems. However, the degrees of freedom (DoF), which is needed to analyse cross-level interaction, is still problematic since DoF is calculated based on individual data rather than group data. Multilevel models overcome these spatial heterogeneity-related problems by testing the individual-level attributes on the housing prices within different clusters such as neighbourhoods or submarkets. Thus, by using multilevel models, it is possible to overcome the atomic fallacy problem that emerges from the generalization of housing unit-level data to submarket-level data and also to overcome the ecological fallacy problem arising due to the practice of generalizing submarket characteristics information to the housing unit level. As a substitute to fitting a general model that assumes that the relationship between the independent and dependent variables is constant everywhere, multilevel models allow this relationship to vary spatially within the study area (Owen et al., 2016). These cross-level interactions help to explain the impact of housing unit variables on housing prices according to the different contexts in submarket levels. Thus, in other words, multilevel modelling enables one to distinguish the impact of housing unit-level variables and geographical/group-level (such as neighbourhood, submarket, district and region) indicators. It then becomes possible to estimate the random part and the proportion of the variance at different levels. This method answers the question of ‘how big is the random part?’ occurring due to the variance (1) between-group levels (such as neighbourhoods) and (2) within-group level but between housing units.

The two-level multilevel model involves of individual housing units, i , nested in area j (hereby submarket j). Leishman et al. (2013) express this as:

$$P_i = \alpha + \sum_k \delta_k Y_i + \sum_j \eta_{0j} + \sum_j \sum_k \eta_{kj} Y_i + \epsilon_{ij}$$

where P_i represents the price of the i th dwelling in the j th spatial area; Y_i are physical and neighbourhood housing attributes for the i th dwelling; η_{0j} is the random intercept for the j th spatial area; η_{kj} are random slope parameters for the k attributes, specific to the j th spatial area; and ϵ_{ij} is an error term or residual.

In parallel with the widespread use of the multilevel approach, some methodological studies have criticized the technical handicaps of this method, especially on the identification of group level. If the geographical presumptions associated with multilevel modelling are unreliable, then this approach cannot provide solid foundations for a methodology that aims to identify contextual effects and differences (Owen et al., 2016). The spatial design of the group level incorporated into multilevel models is crucial since using the same data will generate different results depending on the method of aggregating individual units into spatial units (Gehlke & Biehl, 1934, cited in Owen et al., 2016).

This brings the question: What is the best fit for a group level? To include submarkets – as group level – that are based on different approaches in multilevel models is a way to assess the performance of multilevel modelling. But how could we define the best fit for determining the group level? As reviewed in the previous section, the main approaches in the housing market literature to delineating submarkets are a priori classification, ad hoc and data-driven stratification. This study compares the utility of these three main approaches to group level definition as a prior step to modelling house prices. In this respect, the spatial determinants of housing prices can be analysed via multilevel models, which are based on different submarket descriptions.

Case study and data

The analysis in this paper used data from the Istanbul housing market. Istanbul has experienced a very rapid urbanization and has become one of the most populous cities globally. By 2018, the population of Istanbul reached 15.1 million, an increase from 7.3 million in 1990 (TUIK, 2021a). Having doubled in population over the last three decades, Istanbul has also undergone a polycentric development due to globalization. With 31% of the country's gross domestic product (GDP) held by the city, it has become a financial and trade centre (TUIK, 2021b). The major changes in the service sector and the expansion of transportation infrastructure played essential roles in the multicentre development of Istanbul (Arslanlı, 2020). Moreover, with the approval of the urban regeneration legislation in 2004, many areas in Istanbul have been designated transformative zones (Kuyucu, 2020) in order to improve the quality of housing stock due to earthquake risk. Decentralization of central business districts (CBDs) and industrial areas and the change in the accessibility pattern and urban transformation within the city has led to variation in housing prices (Koramaz & Dokmeci, 2012). The combined effects of these processes have produced an urban structure with high levels of spatial disaggregation and significant socio-economic segmentation (Kisar Koramaz, 2014).

The housing stock in Istanbul can be categorized under three broad groups: 'urban housing, suburban housing and residences' (Ozus et al., 2007, p. 346). Urban housing areas are located in the historic districts, usually in planned inner-city areas, yet not with enough infrastructure such as green areas and recreational areas. The higher land prices in urban areas have prompted developers to construct suburban housing projects in peripheral areas that were dominated with squatter settlements. These peripheral areas have become an attractive alternative to urban housing not only due to the escalation of land prices in the inner city but also as a means of meeting demand from middle-to-high-income groups who want to live in those areas where the risk of earthquake is relatively smaller (Onder et al., 2004; Keskin et al., 2017). A variety of factors, including urban transformation initiatives and modifications to plans to reflect changing market conditions, have contributed to an increase in property prices in these suburban housing areas (Ozkan & Turk, 2016). Presently, most of the existing housing stock consists of multi-family apartment buildings and residential projects located primarily in the suburbs. The other main source of change to the stock takes the form of residential developments located within the inner-city area, usually on brownfield locations. These are targeted at a specific demographic group that prefers urban living but is also motivated by a preference for convenience-based living and 'premier' locations (for details, see Ozus et al., 2007).

Taken together, these ongoing changes in the urban transformation process mean that Istanbul's housing market has a dynamic structure with clear differentiation in terms of product group locations and consumer group preferences for different localities. As such, the city provides an ideal case study location for multilevel analysis of the housing system and its submarket structure.

Data set

Different data sources have been used to compile the database for the analysis in this study. The primary dataset comprises 2175 single-family house transactions sold in November 2006 and April 2007. These were obtained from two major real estates' websites. This dataset compiles housing unit transactions and characteristics from 283 neighbourhoods in 32 districts. Additionally, the second dataset consists of socio-economic characteristics, neighborhood quality characteristics, and location characteristics gathered from a survey that was conducted by the Istanbul Greater Municipality in 2006. This dataset (secondary data) provides information about the socio-economic structure of the neighbourhoods and the satisfaction of the inhabitants of the city. In addition to the socio-economic and neighbourhood quality characteristics, an important locational characteristic – earthquake risk data – is obtained from the JICA (2002) report on disaster prevention/mitigation plan in Istanbul. This is matched to the geocoded dwellings in the core dataset. As a whole, the data used in the study can be categorized into

Table 1. Variables used in the modelling.

Variables	Minimum	Maximum	Average	SD	Unit	Explanation
<i>Housing unit characteristics (individual level)</i>						
Price	34,013.60	8,000,000	251,083	382,467	US\$	Price of the housing unit
Age	0	150	12	14.6	Years	Age of the housing unit
Floor Area	45	1920	170	123	m ²	
Low Storey	0	1			Dummy variable	Housing unit being located in a building of fewer than five stories
Gated Community	0	1			Dummy variable	If the housing unit is located in a vertical/horizontal gated community
<i>Submarket level characteristics (group/nested level)</i>						
Neighbourhood Satisfaction	1	7	5.03	1.21	Likert Scale	Satisfaction score for the neighbourhood (1 = very poor to 7 = excellent)
Income	333	4444	1072	811	US\$	Average income of inhabitants in the neighbourhoods
Living Period in Istanbul	1	46	13.4	6.2	Years	Time the inhabitants have lived in Istanbul
Earthquake Risk	1.33	18.27	5.34	4.10	%	Estimated percentage of the buildings that will be highly damaged in a possible earthquake

four groups of variables: housing unit characteristics; socio-economic characteristics, neighbourhood quality characteristics and location characteristics (Table 1). It should be noted that the number of variables available in this study is thin relative to other studies of this sort. This is a common challenge in studies of emergent markets and, even after multicollinearity issues are taken into account, the number of variables available is still comparable with those used in the baseline hedonic models in several key submarket studies (e.g., Fletcher et al., 2000; Watkins, 2001).

In order to delineate the submarkets in Istanbul, three main approaches explained in the previous section are applied:

- *A priori (administrative neighborhood boundary) submarket delineation*: the administrative neighborhood boundaries are used as the basis for our a priori submarket delineation criterion. When this research was conducted, there were 946 neighbourhoods in Istanbul. However, as many of them do not contain market housing, the data are drawn from 283 neighbourhoods.
- *Ad hoc (expert based) submarket delineation*: 10 semi-structured interviews were held with real estate experts, and they were asked to draw the submarkets on the map that displayed all the administrative boundaries of neighbourhoods in Istanbul. The submarkets are shaped by overlapped boundaries and five submarkets were delineated by consensus from the experts.
- *Data-driven (cluster analysis) submarket delineation*: hierarchical cluster analysis is applied to derive submarket clusters. A total of 12 submarkets are designed by taking into characteristics such as housing prices, floor area, age of the building and number of rooms, income of households, living period of inhabitants in Istanbul, neighbourhood quality, and satisfaction from the public transportation facilities. In line with previous studies, the risk of an earthquake is also treated as an input for the cluster analysis (Onder et al., 2004; Naoi et al., 2009).

MODEL RESULTS

As noted, this study uses three different techniques that are most common in the literature. In this section, the results of three multilevel models of housing price based on the different submarket formulations are presented. First, the results of the intra-class correlation (ICC) values of different submarket delineation are discussed (Table 2). These are used to fit variance component models of housing prices without the explanatory variables for three nested levels: expert, cluster analysis and neighbourhood.

According to the ICC analysis, around 27% of the variance in housing price differences is due to differences across both ad hoc (expert defined) and data-driven submarkets, with the remaining 73% attributable to housing unit level. Unlike the experts-defined or data-driven categorizations, 49% of the variance in housing price is related to housing unit characteristics in the a priori submarket categorization, whereas 51% of the variance comes from submarket (neighbourhood) level.

Table 2. Residual intraclass correlation (ICC).

Level	ICC	SE
A priori submarket	0.510	0.318
Ad hoc (experts defined) submarket	0.267	0.126
Data-driven submarket	0.270	0.113

Table 3. A priori model (neighbourhood) results.

Model 3: Neighbourhood		
Number of groups = 283		
Wald Chi ² = 2627.68		
Prob > Chi ² = 0.00		
Variables	Coefficient	SE
Floor Area	1.094**	0.024
Age	-0.011	0.009
Low Storey	0.041**	0.013
Gated Community	0.097**	0.013
Neighbourhood SatisfactionAge	0.096	0.117
Income	0.186**	0.053
Living Period in Istanbul	0.285**	0.071
Earthquake Risk	-0.077**	0.035

Table 3 displays the results of the model based on the a priori submarket classification and gives the results for the factors that determine the housing prices. This model has the most granular structure with 283 administratively determined neighbourhoods. Floor area of the housing unit, with a positive relationship with price, has the greatest impact on the housing prices. Floor area is followed in terms of scale of influence by average living period in Istanbul, income of the inhabitants living in the neighbourhood, housing unit being located in a vertical/horizontal gated community, and then low storey, respectively, again with a positive impact on the housing prices. Lastly, in addition to those explanatory variables and in line with other studies of Istanbul prices, earthquake risk has a positive significant factor in housing prices.

Table 4 presents the ad hoc (experts' defined) submarket model results. Similar to model 1 (a priori based on neighbourhood boundaries) (Table 3), this implies that floor area, living period in Istanbul, being located in a gated community, income and being located in a low-storey building have a statistically significant positive impact on housing prices have a statistically significant positive effect on housing prices.

Table 5 displays the results of the model based on the data-driven submarket structure and summarizes the results for the factors that are affecting the housing prices. Floor area of the housing unit, with a positive relation with price, has the greatest impact on the housing prices. In terms of scale of impact, floor area is followed by inhabitants' average living period in Istanbul

Table 4. Ad hoc (experts defined) submarket model results.

Model 1: Experts		
Number of groups = 5		
Wald Chi ² = 2158.85		
Prob > Chi ² = 0.00		
Variables	Coefficient	SE
Living Area	1.028**	0.026
Age of the Building	0.002	0.094
Low Storey	0.033**	0.014
Gated Community	0.055**	0.014
Neighbourhood SatisfactionAge	0.0869	0.066
Income	0.047**	0.023
Living Period in Istanbul	0.179**	0.035
Earthquake Risk	-0.007	0.018

Table 5. Data-driven (cluster analysis) submarket model results.

Model 2: Data-driven (cluster analysis) model		
Number of groups: 12		
Wald Chi ² = 2879.62		
Prob > Chi ² = 0.00		
Variables	Coefficient	SE
Floor Area	1.085**	0.029
Age	0.053**	0.098
Low Storey	0.024	0.015
Gated Community	0.065**	0.015
Neighbourhood Satisfaction	0.053	0.073
Income	0.222**	0.037
Living Period in Istanbul	0.258**	0.047
Earthquake Risk	-0.091	0.020
cons	1.829	

and income of the inhabitants living in the neighbourhood, respectively, which again have a positive impact on the housing prices. Aside from those explanatory variables, housing units located in vertical and horizontal gated communities and in low-storey buildings are significant positive factors affecting housing prices. Unlike the expert-based model, the earthquake risk does have an impact on the housing prices. In this case, as might be expected on the basis of prior knowledge, it is negatively related with housing prices.

The variables included in Tables 3–5 are selected on the basis of their significance level. *P*- and *Z*-values are not presented, but they are available from the author upon request. Each specification also generated random effect estimates. As they are not used in the testing regime, they are not reported here, but again they are available from the author upon request.

Overview of the results

Prediction accuracy (PA) and root mean square error (RMSE) tests have been adopted in large number of studies where they are comparing the performance of different submarket delineations. These tests assess the accuracy in terms of the proportion of estimated prices that fall within a narrow tolerable range of the actual price. Examples of such tests can be seen in Watkins (2001), Bourassa et al. (2003), Goodman and Thibodeau (2003), Keskin et al. (2017) as well as a review of around 30 cases in Watkins (2012). Leishman et al. (2013) compared the performance of the submarkets, delineated by real estate professionals, with a priori delineated postcodes. Building on this approach, PA and RMSE tests are applied on models (Table 6) that are derived from real estate experts' opinion, a priori administrative neighbourhood boundaries as well as data-driven delineated submarkets.

As can be seen from Table 5, the predictive performance for a priori submarket delineation, which consists of smaller spatial units, exceed all other submarket approaches examined in this

Table 6. Prediction accuracy and standard error reduction for alternative submarket delineation.

Model	% of cases predicted within 20% accuracy	% of cases predicted within 10% accuracy	Root mean square error (RMSE)
Ad hoc (expert defined) submarket	0.35	0.18	321,691
Data-driven submarket	0.38	0.19	316,844
A priori submarket	0.39	0.21	315,931

study. It is also evident that the spatial pattern of prediction errors is less concentrated. The results show that multilevel models with granular submarket delineation at group level have the capacity to improve predictive power and reduce spatial dependence. According to the test results, the model with a priori submarkets has the strongest predictive performance with more than 39% of cases within 20% accuracy and 21% within 10% accuracy. This is followed by data-driven, with 38% of cases within 20% accuracy and ad hoc (experts' opinion) submarket with 35% of cases within 20% accuracy. Furthermore, the RMSE test results show consistency with PA results since the difference between actual price and estimated values are least in a priori model which is followed again with data-driven and ad hoc submarket models.

CONCLUSIONS

Numerous previous market studies have argued that the complexity and dynamics of the housing system can be analysed more effectively and accurately by taking account of its segmented spatial structure. This implication of this assertion is that submarkets should be accommodated in modelling procedures and that the data used should ideally have a hierarchical and/or nested composition. Despite general agreement on this broad point, there remains a lack of consensus about how best to determine submarket delineations in practice, and the matter has become the subject of several recent empirical investigations.

The literature is also unclear on what form econometric models should take to best undertake such an investigation. The methods used range from simple hedonic regressions that include submarket dummies, systems of submarket specific regressions equations, and the use of neural networks and cellular automata approaches. Several recent studies have also stimulated a resurgence in interest in the use of multilevel modelling. The rationale behind employing multilevel model is that the technique allows readily for the examination of within and between submarket variability. Despite advocacy for greater use of multilevel models, there are two constraints on their use. First, it is not clear from the literature how analysts should identify submarket units for inclusion in models. Can these be captured by including spatial administrative boundaries at different levels, or does prior work need to be done to identify functional submarkets? Second, do multilevel methods have utility in study areas, such in emergent market contexts, where datasets might be thin.

This paper attempts to contribute to the multilevel modelling strand of the submarket literature by addressing both of these issues. It addressed the former issue by using multilevel methods to compare the performance of some of the different submarket descriptions used in the literature. The empirical analysis presented here is based on an experiment that tests different approaches to variants of the three most common approaches used to construct housing submarkets: an a priori classification which is based on the use of administrative neighbourhood boundaries; a set of partitions based on experts' views of submarket composition; and a data-driven approach which groups dwellings based on cluster analysis. The study employs data from the Istanbul housing market. Of course, the use of expert opinion in this way has been advocated as a means of addressing data limitations (Watkins, 2001) and helps us explore the second constraint.

Since most house price applications such as tax assessment, valuation and marketing require high levels of model accuracy, this experiment focuses on predictive accuracy as the basis from which to determine which submarket definition performs best in these models. Based on their precision, the performance of these models is compared in terms of the proportion of predicted prices falling within 10% and 20% range of the actual price. The a priori (administrative neighbourhood boundaries) formulation generated the greatest proportion of estimates within 10% and 20% of actual price with 39% and 21%, respectively, in these intervals. Furthermore, the RMSE test results (which provide an alternative assessment of accuracy) indicate that the a

priori classification has the smallest difference between actual and predicted standard error values. In other words, the categorization based on administrative neighbourhood boundaries has the largest reduction in terms of error. These test results show, as with previous studies in this field, that the most granular definition of submarkets deliver a better performance than those that use larger spatial units. Interestingly, in the context of markets where data constraints persist, the analysis suggests that real estate experts' views of submarket structures might be particularly useful as inputs into micro-modelling processes in contexts where datasets are thin.

ACKNOWLEDGEMENTS

I would like to thank the anonymous reviewers, the Critical Friends reading group (University of Sheffield), Ozlem Celik, Gwilym Owen & Craig Watkins for their valuable and insightful comments on earlier drafts.

DATA AVAILABILITY

The data that support the findings of this study are available from the author upon reasonable request.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author.

REFERENCES

- Adair, A. S., Berry, J., & McGreal, W. S. (1996). Hedonic modelling, housing submarkets and residential valuation. *Journal of Property Research*, 13(1), 67–83. <https://doi.org/10.1080/095999196368899>
- Alas, B. A. (2020). Multilevel Analysis of housing submarkets defined by The Municipal boundaries and By The street connections in the Metropolitan Area: Istanbul. *Journal of Housing and the Built Environment*, 35(4), 1201–1217. <https://doi.org/10.1007/s10901-020-09735-7>
- Arslanlı, K. Y. (2020). Analysis of house prices: A hedonic model proposal for Istanbul metropolitan area. *Journal of Design for Resilience in Architecture and Planning*, 1(1), 57–68. <https://doi.org/10.47818/DRArch.2020.v1i1004>
- Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4), 588–638. <https://doi.org/10.1086/522381>
- Bourassa, S., Cantoni, E., & Hoesli, M. (2007). Spatial dependence, housing submarkets, and house price prediction. *Journal of Real Estate Finance and Economics*, 35(2), 143–160. <https://doi.org/10.1007/s11146-007-9036-8>
- Bourassa, S., Hamelink, F., Hoesli, M., & MacGregor, B. (1999). Defining housing submarkets. *Journal of Housing Economics*, 8(2), 160–183. <https://doi.org/10.1006/jhec.1999.0246>
- Bourassa, S., Hoesli, M., & Peng, V. (2003). Do housing submarkets really matter? *Journal of Housing Economics*, 12(1), 12–28. [https://doi.org/10.1016/S1051-1377\(03\)00003-2](https://doi.org/10.1016/S1051-1377(03)00003-2)
- Bullen, N., Jones, K., & Duncan, C. (1997). Modelling complexity: Analysing between-individual and between-place variation – A multilevel tutorial. *Environment and Planning A: Economy and Space*, 29(4), 585–609. <https://doi.org/10.1068/a290585>
- Clapp, J. M., & Wang, Y. (2006). Defining neighborhood boundaries: Are census tracts obsolete? *Journal of Urban Economics*, 59(2), 259–284. <http://doi.org/10.1016/j.jue.2005.10.003>
- Costello, G., Leishman, C., Rowley, S., & Watkins, C. (2019). Drivers of spatial change in urban housing submarkets. *The Geographical Journal*, 185(4), 432–446. <https://doi.org/10.1111/geoj.12303>

- Dunning, R., Levy, D., Watkins, C., & Young, G. (2019). Technological change and estate agents' practices in the changing nature of housing transactions. *Housing Studies*, 34(5), 849–867. <https://doi.org/10.1080/02673037.2018.1487041>
- Evans, A. W. (1995). The property market: Ninety per cent efficient?. *Urban Studies*, 32(1), 5–29. <https://doi.org/10.1080/00420989550013194>
- Fik, T. J., Ling, D. C., & Mulligan, G. F. (2003). Modeling spatial variation in housing prices: A variable interaction approach. *Real Estate Economics*, 31(4), 623–646. <http://doi.org/10.1046/j.1080-8620.2003.00079.x>
- Fletcher, M., Gallimore, P., & Mangan, J. (2000). The modelling of housing submarkets. *Journal of Property Investment & Finance*, 18(4), 473–487. <https://doi.org/10.1108/14635780010345436>
- Goodman, A. C. (1981). Housing submarkets within urban areas: Definitions and evidence. *Journal of Regional Science*, 21(2), 175–185. <https://doi.org/10.1111/j.1467-9787.1981.tb00693.x>
- Goodman, A. C., & Thibodeau, T. (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](https://doi.org/10.1016/S1051-1377(03)00031-7)
- Goodman, A. C., & Thibodeau, T. G. (1998). Housing market segmentation. *Journal of Housing Economics*, 7(2), 121–143. <https://doi.org/10.1006/jhec.1998.0229>
- Grigsby, W. G. (1963). *Housing markets and public policy*. University of Pennsylvania Press.
- JICA. (2002). A Disaster Prevention/Mitigation Plan for Istanbul. Report to Istanbul Municipal Authority, Istanbul. Available at: http://www.ibb.gov.tr/tr-TR/SubSites/DepremSite/PublishingImages/JICA_ENG.pdf (retrieved 06 October 2008).
- Jones, C. (2002). The definition of housing market areas and strategic planning. *Urban Studies*, 39(3), 549–564. <https://doi.org/10.1080/00420980220112829>
- Jones, C., & Watkins, C. (2009). *Housing markets and Planning policy*. Wiley-Blackwell.
- Jones, K., & Bullen, N. (1993). A Multilevel Analysis of the variations in Domestic Property prices – Southern England, 1980–87. *Urban Studies*, 30(8), 1409–1426. <https://doi.org/10.1080/00420989320081341>
- Kauko, T. (2002). *Modelling the locational determinants of house prices: Neural network and value tree approaches* [Doctoral dissertation]. Utrecht University. Available at: <https://dspace.library.uu.nl/bitstream/handle/1874/555/inhoud.htm;sequence=13> (retrieved 27 January 2020).
- Kauko, T. (2003). Residential Property value and locational externalities. *Journal of Property Investment and Finance*, 21(3), 250–270. <https://doi.org/10.1108/14635780310481676>
- Keskin, B. (2010). *Alternative approaches to modelling housing market segmentation: Evidence from Istanbul* [Unpublished doctoral dissertation]. University of Sheffield.
- Keskin, B., Dunning, R., & Watkins, C. (2017). Modelling The impact of earthquake activity on real estate values: A multilevel approach?. *Journal of European Real Estate Research*, 10(1), 73–90. <https://doi.org/10.1108/JERER-03-2016-0014>
- Kısar Koramaz, E. (2014). The spatial context of Social integration. *Social Indicators Research*, 119(1), 49–71. <https://doi.org/10.1007/s11205-013-0481-x>
- Koramaz, K., & Dokmeci, V. (2012). Spatial determinants of housing price values in Istanbul. *European Planning Studies*, 20(7), 1221–1237. <https://doi.org/10.1080/09654313.2012.673569>
- Kuyucu, T. (2020). The great Failure: The roles of institutional conflict and social movements in the Failure of regeneration initiatives in Istanbul. *Urban Affairs Review*, 58(1), 129–163. <https://doi.org/10.1177/1078087420957736>
- Leishman, C. (2009). Spatial change and the structure of urban housing submarket. *Housing Studies*, 24(5), 563–585. <https://doi.org/10.1080/02673030903082310>
- Leishman, C., Costello, G., & Rowley, S. (2013). The predictive performance of multilevel models of housing submarkets: A comparative analysis. *Urban Studies*, 50(6), 1201–1220. <https://doi.org/10.1177/0042098012466603>
- Maclennan, D., Munro, M., & Wood, G. (1987). Housing choice and the structure of urban housing markets. In B. Turner, J. Kemeny, & L. A. Lundquist (Eds.), *Between state and market housing in the post-industrial Era* (pp. 26–52). Gothenburg.

- Maclennan, D., & Tu, Y. (1996). Economic perspectives on the structure of local housing markets. *Housing Studies*, 11(3), 387–406. <https://doi.org/10.1080/02673039608720864>
- Malpezzi, S. (2003). Hedonic pricing models: A selective and applied review. In A. O'Sullivan & K. O. Gibb (Eds.), *Housing Economics and Public Policy* (pp. 67–89). Blackwell.
- Michaels, R., & Smith, V. K. (1990). Market segmentation and valuing amenities with hedonic models: The case of hazardous waste sites. *Journal of Urban Economics*, 28(2), 223–242. [https://doi.org/10.1016/0094-1190\(90\)90052-O](https://doi.org/10.1016/0094-1190(90)90052-O)
- Naoi, M., Seko, M., & Sumita, K. (2009). Earthquake Risk and housing prices in Japan: Evidence before and after massive earthquakes. *Regional Science and Urban Economics*, 39(6), 658–669. <https://doi.org/10.1016/j.regsciurbeco.2009.08.002>
- Onder, Z., Dokmeci, V., & Keskin, B. (2004). The impact of public perception of earthquake risk on Istanbul's housing market. *Journal of Real Estate Literature*, 12(2), 181–194. <https://doi.org/10.1080/10835547.2004.12090136>
- Orford, S. (2000). Modelling spatial structures in Local Housing Market dynamics: A multilevel perspective. *Urban Studies*, 37(9), 1643–1671. <https://doi.org/10.1080/00420980020080301>
- Orford, S. (2002). Valuing locational externalities: A GIS and Multilevel Modelling approach. *Environment and Planning B: Planning and Design*, 29(1), 105–127. <https://doi.org/10.1068/b2780>
- Owen, G., Harris, R., & Jones, K. (2016). Under examination: Multilevel models, Geography and health research. *Progress in Human Geography*, 40(3), 394–412. <https://doi.org/10.1177/0309132515580814>
- Ozkan, H. A., & Turk, S. S. (2016). Emergence, formation and outcomes of flexibility in Turkish Planning practice". *Development Planning Review*, 38(1), 25–53. <https://doi.org/10.3828/idpr.2016.2>
- Ozus, E., Dokmeci, V., Kiroglu, G., & Egdemir, G. (2007). Spatial Analysis of residential prices. *Istanbul European Planning Studies*, 15(5), 707–721. <https://doi.org/10.1080/09654310701214085>
- Palm, R. (1978). Spatial segmentation of the urban housing market. *Economic Geography*, 54(3), 210–221. <https://doi.org/10.2307/142835>
- Pavlov, A. D. (2000). Space-varying regression coefficients: A semi-parametric approach applied to real estate markets. *Real Estate Economics*, 28(2), 249–283. <http://doi.org/10.1111/reec.2000.28.issue-2>
- Rae, A. (2015). Online housing search and The Geography of submarkets. *Housing Studies*, 30(3), 453–472. <https://doi.org/10.1080/02673037.2014.974142>
- Rothenberg, J., Galster, G., Butler, R., & Pitkin, J. (1991). *The maze of urban housing markets: Theory, evidence and policy*. University of Chicago Press.
- Straszheim, M. (1975). *An econometric analysis of the urban housing market*. National Bureau of Economic Research.
- Turkish Statistical Institution (TUIK). (2021a). <https://data.tuik.gov.tr/Bulten/Index?p=Adrese-Dayal%C4%B1-N%C3%BCfus-Kay%C4%B1t-Sistemi-Sonu%C3%A7lar%C4%B1-2020-37210&dil=1> retrieved at 12.01.2021.
- Turkish Statistical Institution (TUIK). (2021b). <https://data.tuik.gov.tr/Bulten/Index?p=Quarterly-Gross-Domestic-Product-Quarter-IV:-October-December,-2020-37180&dil=2> retrieved at 12.01.2021.
- Watkins, C. (1998). Are New entrants to the Residential Property market informationally disadvantaged? *Journal of Property Research*, 15(11), 57–70. <https://doi.org/10.1080/095999198368509>
- Watkins, C. (2001). The definition and identification of housing submarkets. *Environment and Planning A*, 33(12), 2235–2253. <https://doi.org/10.1068/a34162>
- Watkins, C. (2012). Housing submarkets. In S. Smith (Ed.), *International encyclopaedia of housing and home* (pp. 39–44). Elsevier.