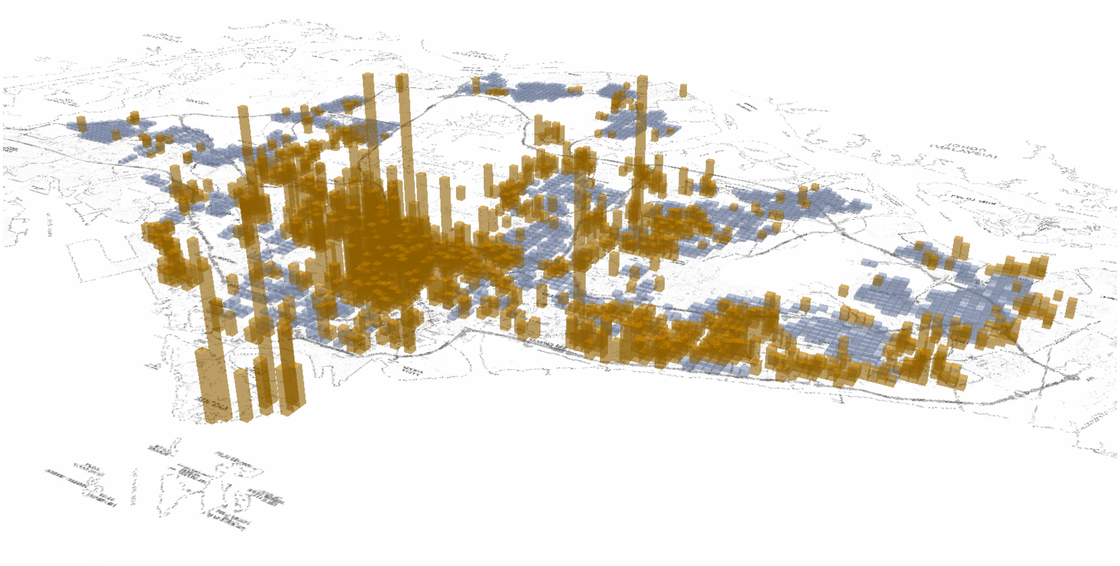


Modelling housing prices in Singapore applying spatial hedonic regression

Manuel Lehner

Supervision:
Prof. Dr. Kay W. Axhausen
Michael van Eggermond

Master of Science Thesis
Spatial Development and Infrastructure Systems **July 2011**

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems
FUTURE CITIES LABORATORY

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Acknowledgments

I would like to thank Professor Kay W. Axhausen for his guidance and inspiration and for giving me the opportunity to conduct my research at the Future Cities Laboratory. I thank my supervisor Michael van Eggermond, who supported me in various ways including data collection and proof-reading of the thesis. Moreover, I would like to thank Alexander Erath who gave valuable input to my work and warmly welcomed me with his family during my stay in Singapore. I further thank Michael Löchl - former member of Professor Axhausen's group at ETH Zürich - for providing me with his modelling scripts and useful recommendations.

I would like to express my gratitude to Nathalie Casas and thank her for the outstanding support over the last years. Moreover, I want to thank friends and colleagues who supported me during my studies at ETH Zürich: Patricia Lehner, Benjamin Belart, David Sorg, Reto Rieder, Thomas Rüfli, Jonas Stillhard and Daniel Köchli.

Master of Science Thesis

Modelling housing prices in Singapore applying spatial hedonic regression

Manuel Lehner
Habsburgstrasse 28
CH-8037 Zurich
lehnermanuel@hotmail.com

July 2011

Abstract

A hedonic housing price model for Singapore is developed in order to generate input data for agent-based modelling at the ETH Future Cities Laboratory. Around 110'000 asking and transaction price listings from online sources are used for the study. The sample contains observations from both the private and the HDB market and includes sale and rental housing units. Asking prices (expected preferences) are found to be up to 70% higher than transaction prices (revealed preferences). Spatial error models perform better than other modelling approaches. Unit prices are found to be mainly determined by the floor area, the distance to the central business district and the age. Depending on the market segment, between ten and twenty-five variables add significant explanatory power to the models. Housing preferences are found to vary between different market segments while model estimates show similar impacts of the most important price determinants. Therefore the price difference can be modeled with a constant. Spatial error models performed best and geographically weighted regression models point to spatially varying housing preferences.

Keywords

Singapore, housing prices, hedonic regression, asking prices, transaction prices

Preferred citation style

Lehner, M. (2011) Modelling housing prices in Singapore applying spatial hedonic regression, *Master of Science Thesis*, Institute for Transport Planning and Systems (IVT), ETH Zurich, Zurich.

Masterarbeit

Modellierung von Wohnungspreisen in Singapur unter Anwendung räumlicher Regressionstechniken

Manuel Lehner
Habsburgstrasse 28
CH-8037 Zurich
lehnermanuel@hotmail.com

July 2011

Zusammenfassung

In dieser Arbeit wird ein hedonisches Preismodell für Wohnimmobilien in Singapur entwickelt, welches für die Etablierung eines Agenten-basierten Simulationsmodells am ETH Future Cities Laboratory benötigt wird. Die Datenbasis umfasst rund 110'000 Beobachtungen von Angebots- und Transaktionspreisen. Sowohl der private als auch der öffentliche Immobiliensektor sind im Sample repräsentiert. Die besten Modellgüten werden mit räumlich autoregressiven Modellierungsansätzen erreicht. Die Schätzungen zeigen, dass Wohnungspreise in Singpur hauptsächlich durch die Wohnfläche, die Distanz zum "Central Business District" and durch das Alter determiniert sind. Abhängig vom Marktsegement sind zwischen zehn und fünfundzwanzig Variablen signifikant für die Erklärung der Preise. Angebotspreise sind bis zu 70% höher als Transaktionspreise, weisen aber eine sehr ähnliche Präferenzstruktur auf. Der Preisunterschied kann daher näherungsweise mit einer Konstante modelliert werden. Zusätzlich durchgeführte geographisch gewichtete Regressionen deuten auf räumlich variierende Wohungspräferenzen hin.

Schlagworte

Singapur, Hedonische Regressionen, Angebotspreise, Transaktionspreise

Bevorzugter Zitierstil

Lehner, M. (2011) Modellierung von Wohnungspreisen in Singapur unter Anwendung räumlicher Regressionstechniken, *Masterarbeit*, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Contents

List of Figures

List of Tables

1. Introduction

Residential location choice of households is considered to be a fundamental driver of land use, mobility behavior and social development. Housing prices are main determinants of location choice but also driven by market forces resulting from these decisions. To deal with this complexity, land-use and transport interaction models have been developed and applied in various cases. To establish such models significant efforts are needed for the housing price data generation and modelling. Researchers and practitioners have developed housing price models since the 1960's and there is a variety of studies available. The most widely used modelling approach is the hedonic regression method. Hedonic theory assumes that housing prices can be decomposed into measurable prices of specific housing characteristics.

As a part of the research at ETH Future Cities Laboratory (ETHZ, 2011) a land-use and transport interaction model will be established for the Singapore case. The aim is to advance research into the complexity of land transport which derives from the demands of managing, planning and optimizing the flow of people and goods at different time scales and in its interaction with all elements of the future city. Therefore, the software toolkit MATSim (MATSim-T, 2011) will be implemented and extended for the Singapore case. It is going to include the moves of firms in response to transport and land-use policies on a unit level. Furthermore, mobility behavior of households will be included as well. The aim of this thesis is therefore to develop a hedonic housing price model to generate price input data für MATSim.

This report is structured as follows: Section 2 provides a selection of hedonic modelling approaches and results of existing studies in Southeast Asia. It further describes the Singapore housing market based on scientific studies publicly available data. Section 3 includes a descriptive analysis of gathered data for the Singapore case. Based on this data Section 4 provides results of different hedonic models. Finally, Section 5 provides a comparison of housing preferences in different market segments and a discussion of possible reasons. It further includes recommendations for further research.

1.1 Goals and working steps

In the above mentioned context design and implementation of software agents representing households and firms as well as location and residential choice modelling will become major tasks of module VIII. Therefore, this thesis aims to **provide a hedonic housing price model** to generate island-wide input data for MATSim. In order to reach this goal, nine working steps (as shown in Table 1) were used to structure this study.

Table 1: Working process

#	Working step	Output	Resp.
1	Analysis of Southeast Asian housing price determinants	Desirable price determinants for Singapore case and impact expectations	ml
2	Analysis of Singapore housing market	Market segmentation and key figures for relevant market segments	ml
3	Data acquisition for relevant market segments	Dataset including desirable variables and geographic information (without locational variables)	FCL
4	Generation of locational variables and data segmentation	Datesets for each market segment including locational variables	ml
5	Descriptive statisitcs and comparison of dataset with housing market	Assessment of data representativeness compared to housing market key figures	ml
6	Estimation of OLS, SAR and GWR models	Significant variable set and parameter analysis for each modelling approach	ml
7	Comparison of modelling results for different market segments	Overview of hedonic preferences in different housing markets	ml
8	Comparison and assessment of different modelling approaches	Suggestion of a model to use for MATSim data acquisition	ml

Note: OLS = Ordinary least square models, SAR = Simultaneous autoregressive models, GWR = Geographically weighted regression models, FCL = ETH Future Cities Laboratory, ml = Manuel Lehner

2. Literature review

This chapter provides an overview of state of the art hedonic pricing models while focussing on the incorporation of spatial effects. Additionally existing studies for the area of Southeast Asia are analyzed in order to create a basis for later model specification for the Singapore case.

2.1 Modelling housing prices using hedonic regression

According to Fahrländer (2007) housing units are heterogenous goods due to their immobility and resulting locational differences. Additionally, most dwelling units are unique concerning technical and architectural qualities. In order to incorporate these heterogeneities into price estimations, hedonic theory can be applied. The object of the hedonic pricing approach is valuing specific goods characteristics depending on their utility for potential buyers. Sirmans *et al.* (2009) point out that a dwelling unit is made up of many characteristics, all of which may affect its value. The hedonic pricing approach is typically used to estimate the contribution of these individual characteristics to the total value of the unit. Lancaster (1966) applied hedonic theory in the field of real estate for the first time in the sixties. Löchl (2010) states that today - in the real estate environment - the approach is regularly used for property taxation and mortgage underwriting, but also for property price generation in land use and transport models.

There are reams of different studies applying the hedonic approach for real estate purposes. As a result of this empirical work there is an extensive list of attributes that scientists used for specifying their models. Different authors used different approaches to divide these attributes into categories. In a broad literature overview, Malpezzi (2002) identifies structural (describing the dwelling unit itself, such as size, number of rooms, age etc.), locational (depending on the absolute location within the study area, such as distance to central business district etc.), neighborhood (incorporating qualities of contiguous areas, such as availability of public schools, population density etc.), contract depending and time specific attributes. Sirmans *et al.* (2009) additionally mentions internal features (baths, fireplace, air conditioning, hardwood floors, basement etc.), external features (garage spaces, deck, pool, porch, carport, garage etc.), natural environmental features (lake view, lake front, ocean view etc.), public services (school district, percent of school district minority, public sewer), marketing, occupancy, and selling factors (assessor's quality, assessed condition, vacant, owner-occupied, time on the market etc.) and financing factors.

At its simplest, parameters of a hedonic equation can be estimated using ordinary least square models (OLS) as a regression of housing prices on housing characteristics (Malpezzi, 2002) where β is a vector of regression coefficients, X is a matrix with observations on explanatory characteristics and ε representing the error vector. It can be written as follows:

$$\begin{aligned} P &= \beta X + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{1}$$

2.1.1 Taking spatial effects into account

It is assumed that locational and neighborhood determinants do not necessarily take the entire range of spatial effects into account (Löchl, 2010). According to Anselin (1988) two major types of spatial effects can be identified: spatial dependence and spatial heterogeneity. He describes spatial dependence (also called spatial autocorrelation) as a functional relationship between what happens at one point in space and what happens elsewhere. Spatial heterogeneity on the other hand is supposed to appear when there is a lack of uniformity from the effects of space resulting in spatial heteroscedasticity or spatially varying parameters. In order to deal with the described spatial effects, various modelling approaches have been developed by the scientific community. According to Löchl (2010) the most popular approaches are the following:

- Expansion methods (Fotheringham and Pitts, 1995)
- Multi-level approaches (Jones, 1991)
- Spatial simultaneous autoregressive approaches (SAR) (Anselin, 1988)
- Geographically weighted regression (GWR) models (Fotheringham *et al.*, 2002)

Taking the findings of Löchl (2010) into account, for this thesis only SAR and GWR models will be adapted since they had high explanatory power for housing prices in the Zurich area. According to Kissling and Carl (2008) spatial simultaneous autoregressive models can be divided into three subgroups depending on where the autoregressive process is expected to occur. Spatial autoregressive lag models (SARlag) assume that an inherent spatial autocorrelation is present in the response variable. A SARlag model can be written as

$$\begin{aligned} P &= \rho W P + \beta X + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{2}$$

where P is a vector of housing prices, ρ is a spatial autocorrelation parameter, W is a $N \times N$ spatial weights matrix, β is a vector of regression coefficients, X is a matrix with observations on explanatory characteristics and ε representing the error vector (Löchl, 2010). If spatial dependence is assumed to appear in the disturbance process, an error vector u containing the spatial weights matrix is used. This leads to the so-called spatial error model (SARerr), which can be be written as

$$\begin{aligned} P &= \beta X + u \\ u &= \lambda W u + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{3}$$

where λ is a spatial autoregressive coefficient and W is the spatial weights matrix which appears now in the error term. If spatial autocorrelation is assumed to appear in both the explanatory and response processes, Kissling and Carl (2008) suggest to use the so-called spatial Durbin model (SARdurbin), which contains additionally a term $WX\gamma$ which describes the autoregression coefficient γ of the spatially lagged explanatory variables. The SARdurbin model can be written as follows:

$$\begin{aligned} P &= \rho W P + \beta X + W X \gamma + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{4}$$

Besides the above described SAR models, geographically weighted regression models (GWR) will be estimated for this research as well. Fotheringham *et al.* (2002) point out that GWR estimate linear regressions for every data point in space using overlapping samples of the data. Therefore distance-dependent weights are used. GWR essentially allow parameters to vary over space, which can lead to a increased understanding of varying relationships between variables across space (Löchl, 2010). A GWR model can be written as follows:

$$P_i = \beta_{i0} + \sum_k \beta_{ik} X_{ik} + \varepsilon_i$$
$$\varepsilon \sim N(0, \sigma^2 I_n) \tag{5}$$

where P_i represents the ith housing price observation (one certain data point in space), β_{i0} is the constant estimated for this observation, β_{ik} is the coefficient of the explanatory variable k and ε_i is the ith estimate of the error vector (Farber and Yeates, 2006). The estimation of β_i can be written as follows:

$$\beta_i = (X^T W_i X)^{-1} X^T W_i P \tag{6}$$

where β_i is the vector of estimated coefficients for observation i, X is the $N \times K$ matrix of explanatory variables, W_i is a diagonal distance-decay weight matrix customized for i's location relative to the surrounding observations and P is the vector of observed housing prices (Löchl, 2010).

2.2 Housing price determinants in Southeast Asia

In order to identify adequate and possibly significant variables for the Singapore case, this chapter explores existing hedonic studies for Southeast Asia and Hong Kong. Ten scientific studies concerning twelve areas have been selected as shown in Table 2.

Table 2: Existing studies for Southeast Asia and Hong Kong

Country	Area	Reference	Nr. in this thesis
Singapore	Whole city	Ong and Ho (2003)	[1]
Singapore	Hougang district	Sue and Wong (2010)	[2]
Singapore	Potong Pasir district	Sue and Wong (2010)	[3]
Vietnam	Ho Chi Minh City	Kim (2007)	[4]
Vietnam	Hanoi	Kim (2007)	[5]
Malaysia	Penang Island	Chin *et al.* (2004)	[6]
Thailand	Bankok	Chalermpong (2007)	[7]
Indonesia	Jakarta	Yusuf and Resosudarmo (2009)	[8]
China	Hong Kong	Wong *et al.* (2005)	[9]
China	Hong Kong	Tse and Love (2000)	[10]
China	Hong Kong	Tang and Chung (2010)	[11]
China	Hong Kong	Jim and Chen (2009)	[12]

2.2.1 Structural variables

Table 3 shows that structural variables can be subdivided into building and unit depending factors. According to Sirmans *et al.* (2009) a substantial part of real estate price variance can be explained with the variables ***age*** (appearing in 9 of 12 studies) and ***floor area*** (appearing in all 12 studies). In the analyzed studies these two factors have been used in almost all cases and turned out to be highly significant. The age of the property is normally used as a proxy for residential depreciation in terms of deterioration and obsolescence (Ong and Ho, 2003) and has a negative impact on the price in all cases. The area of a specific dwelling unit is - as intuitively expectable - positively related to the price because larger flats command higher prices.

The variable *number of rooms* also produced a significant positive impact on the price in almost all of the analyzed studies. It must be noted that in study [12] (Hong Kong), where the impact was negative, the number of rooms only incorporated bedrooms. The authors (Jim and Chen, 2009) argue, that the negative effect on price is related to a reduction of usable area in the apartment unit. Normally, multicollinearity appears in models including both area and number of rooms, while Ong and Ho (2003) point out, that the floor area can vary for flats with the same number of rooms but with different layout. On the other hand, Sue and Wong (2010) state that the interior characteristics of the Singaporean HDB flats are relatively homogeneous. This leads to the insight, that only one of the two variables should be included in models for Singapore.

Many of the analyzed works have pointed out the importance of the price determinant *floor level*. It indicates, on which floor a certain flat is located. Especially in a dense city context with a elevated proportion of high rising buildings, this variable integrates the third dimension into hedonic price models. Seven of the twelve analyzed studies used floor level as price determinant and all of them identified a significant positive impact. Preconditions for this positive correlations are adequate lift and sanitary infrastructure providing sufficient water pressure and ability of vertical movement. In order to incorporate negative effects of flats located on the base floor (noise, crime, external exposure etc.), Wong *et al.* (2005) added a second floor-level-related variable called *is on the base floor* to their model. This factor turned out to have a significant negative influence onto the price level.

The analyzed structural variables sets also included additional amenities like *availability of a car park*, *availability of a pool* or *availability of a garden*. Explanatory contribution of these factors strongly depends on the urban context. While a pool is considered to be important in warm and generally rich regions, garden-availability is possibly higher valued in cities with a small percentage of public green space. Surprisingly car parking was only included in one of the twelve studies (Tse and Love, 2000) and turns there out to have a significant positive impact on the flats prices.

Tang and Chung (2010) introduce a new measure for *spaciousness* in their study about interrelations of development intensity and housing prices in Hong Kong. They define spaciousness as the average dwelling unit percentage of public space within the condominium. They found, that buyers tend to pay more for a larger amount of internal and external housing space and that there is an ideal range for the total *number of units* within a condominium. Anther conclusion is that flats in low-rising buildings generate higher market prices than equal flats in high-riser. They argue, that a taller building necessitates a longer lift waiting time and a larger occupancy rate because the lifts have to serve more people.

Finally, some authors used structural variables concerning the specific spatial orientation of a flat. Wong *et al.* (2005) found that flats *facing a garden* were sold more expensively than flats without. Additionally they showed that flats *facing a street* or *a MRT depot* generate a significant price discount. Jim and Chen (2009) point out, that residential units with a view of streets, particularly those in the lower floors, would be affected by such negative impacts.

Table 3: Structural variables

	Singapore	Singapore	Singapore	HCMC	Hanoi	Penang	Bankok	Jakarta	Hong Kong	Hong Kong	Hong Kong	Hong Kong
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Building / Estate												
Age	* (-)	* (-)	* (-)			* (-)	* (-)		* (-)	* (-)	* (-)	* (-)
Availability of a car park										* (+)		
Availability of a pool						* (+)					(+)	* (+)
Availability of a garden						* (+)						
Number of units											* (+)	
Spaciousness											* (+)	
Is a low-riser									* (+)			
Unit												
Floor area	* (+)	* (+)	* (+)	* (+)	* (+)	* (-)	* (+)	* (+)	* (+)	* (+)	* (+)	* (+)
Number of rooms	* (+)	* (+)	* (+)					* (+)				* (-)
Floor level	* (+)	* (+)	* (+)			* (+)			* (+)		* (+)	* (+)
Facing a street									* (-)			* (-)
Facing a garden									* (+)			
Facing a MRT depot									* (-)			
Is on the base floor									* (-)			

Note: * Significant variables, (+) Estimated coefficient sign is positive, (-) Estimated coefficient sign is negative

2.2.2 Locational and neighborhood variables

As mentioned above, locational variables incorporate the geographical location information into hedonic models while neighborhood variables take aspects of contiguous areas into account. Table 4 gives an overview of variables used in the studies analyzed. Most of them included some kind of distance indicators like *distance to nearest bus station* or *distance to nearest industrial estate*. Where it must be noted that distance variables are included in many different ways. Ong and Ho (2003) for example used the straight-line distances between the resale flats and the urban amenities while Sue and Wong (2010) calculated dummies in order to define, whether a point of interest was located within a certain distance to the flat or not (proximity). The most important points of interests seem to be public transportation access points (bus stations, MRT stations) and working areas (central business district, industrial estates). Higher distances to these kinds of areas turned out to have a significant negative impact to property prices.

While distances to points of interest incorporate positive impacts of certain areas to flat prices, proximity factors were also used to include external effects of public infrastructure. Ong and Ho (2003) point out that the expected impact of *proximity to expressway* is negative as a result of the negative externalities arising from pollution and congestion caused by the nearby expressway. Yusuf and Resosudarmo (2009) used the *annual number of vehicles passing* within a certain distance to the flat as a proxy for the level of congestion.

There where only three neighborhood variables used in the studies analyzed. *High education ratio in district* turned out to have a significant positive impact to the price level in Ho Chi Minh City, Hanoi and Jakarta. Kim (2007) states that higher educational attainment is supposed to be correlated with higher incomes. The opposite effect can be assumed from a high *unemployment rate in the district*. Yusuf and Resosudarmo (2009) emphasize that both, high education rate and unemployment rate can be used as proxies for the general quality of the neighborhood. Tang and Chung (2010) finally measure the *popularity of the housing estate* as the share of transactions within the same housing estate and incorporate it to their model. It is found to have a positive effect on housing prices.

Table 4: Locational and neighborhood variables

	Singapore	Singapore	Singapore	HCMC	Hanoi	Penang	Bankok	Jakarta	Hong Kong	Hong Kong	Hong Kong	Hong Kong
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Distance to nearest bus station	* (-)	* (-)	(-)					* (-)				
Distance to nearest MRT station	* (-)	(-)	* (-)				* (-)					* (-)
Distance to nearest shopping centre	(-)					* (-)				(+)		
Distance to nearest industrial estate	* (+)	(+)	* (+)									
Distance to nearest primary school				* (-)		* (-)						
Distance to nearest seashore												* (-)
Distance to central business district				* (-)	* (-)	* (-)					* (-)	
Distance to arterial road							* (-)					
Distance to sport facilities										* (+)		
Building is at the urban fringe				* (+)	* (-)							
Proximity to expressway	* (-)											
Yearly number of vehicles passing								* (+)				
Sea view						* (+)						* (+)
Cemetery view						* (-)				* (-)		
High eduction ratio in district				* (+)	* (+)			* (+)				
Unemployment rate in district								* (-)				
Popularity of housing estate											(+)	

Note: * Significant variables, (+) Estimated coefficient sign is positive, (-) Estimated coefficient sign is negative

2.2.3 Specific variables used in Singapore

Eleven variables were only used for Singapore studies. Three of them concern the main upgrading program (MUP) and eight are including locational indicators. MUP are launched by the government in order to enhance HDB estates by upgrading of services, facade enhancement, inclusion of space-adding items within the flat, landscaping and other external works (Ong and Ho, 2003). In their study Ong and Ho defined three dummy variables to include MUP to their model. The dummies concern three stages of the MUP process. *MUP completed* was expected to have a positive price impact due to the enhancements of the program. But model estimations did not show a significant effect of this variable. The second MUP variable - *upgrading in progress* - was expected to have a negative price impact due to resulting inconvenience and pollution. Ong and Ho did not find a significant impact while Sue and Wong (2010) detected a substantial negative impact for the Potong Pasir district. Finally the variable *upgrading planned*, which was expected to have a positive price impact, did not turn out to be significant.

Sue and Wong (2010) additionally used locational variables to incorporate effects of the *proximity to good performance and good progress schools* and found contradictory but significant impacts. But they detected that flats *located in a PAP ward* (PAP: the ruling People Actions Party) tend to be significantly more expensive than flats in opposition constituencies.

Table 5: Specific variables used in Singapore

Nr. in this Thesis	[1]	[2]	[3]
Main Upgrading Program (MUP) completed	(+)		
Main Upgrading Program (MUP) in progress	(+)	(-)	* (-)
Main Upgrading Program (MUP) is planned	(+)		
Proximity to private housing (within 300m)	(+)		
Proximity to popular primary school (within 400m)	(-)		
Proximity to good performance school (within 1km)		* (+)	* (-)
Proximity to good performance school (within 1-2km)		(-)	* (-)
Proximity to good progress school (within 1km)		(+)	* (+)
Proximity to good progress school (within 1-2km)		* (+)	* (+)
Unit is located in a PAP ward		* (+)	* (+)

Note: * Significant variables, (+) Estimated coefficient sign is positive, (-) Estimated coefficient sign is negative

2.2.4 Comparison with determinants in Zurich

A special task of this thesis is to compare hedonic pricing results in Southeast Asia/Hong Kong with results of the Zurich area (see section 1.1). Two studies have been taken into account for the Zurich case: The PhD thesis of Löchl (2010) and a study of the Zurich cantonal bank (ZKB, 2004) produced for the whole Canton of Zurich. To compare variable impacts, the twelve studies for Southeast Asia have been summarized, resulting in a overall impact sign (see Table 6). Unfortunately only six variables appeared in more than one Southeast Asian study and at least in one of the Zurich studies (three structural and three locational variables). These main price determinants show the same impact direction for all studies. Therefore the existence of a location-independent price determinant set is assumed. These variables including general requirements of modern societies to housing such as availability of enough space, modern housing standard or proximity to working areas and mobility networks.

Table 6: Comparison of Southeast Asian and Zurich studies

	Southeast Asia and Hongkong					Zurich	
						[13]	[14]
	Used	Sign.	Sign (+)	Sign (-)	Overall	Sign	Sign
Floor area	12	12	12		+	+	+
Age	9	9	0	9	-	-	-
Availability of a pool	3	2	2	0	+		+
Distance to nearest MRT	5	4	0	4	-	-	
Distance to CBD	4	4	0	4	-	-	-
Sea view / Lake view	2	2	2	0	+	+	+

Note: CBD = central business district, Sign. = Number of cases, where a variable turned out significant

On the other hand there seem to be location-dependent price determinants which vary over macroscopic geographical space. For example the study of ZKB (2004) found considerable added value for dwelling units including floor heating, insulation glass windows, energy standards and evening solar exposure. Löchl (2010) additionally found a significant effect of the availability of a fireplace. Of course, nobody is willing to pay extra for such amenities in the humid and warm climate of Southeast Asia.

2.3 Singapore housing market

The Singapore model of housing has been described by Sock-Yong Phang in a very clear and comprehensible way (Phang, 2007). Among many other things, she described the main players and the finance process of the Singapore housing market.

> **Housing Development Board (HDB)**: *"[..] was set up as a statutory board in 1960 [..] to provide decent homes equipped with modern amenities for all those who needed them. [..] From 1964, the HDB began offering housing units for sale at below market prices, on 99-year leasehold basis, under its Home Ownership Scheme (HOS). The HDB was able to price its units below market prices mainly because HDB flats are built on state owned land [..]. [..] Singapore's large public housing sector is therefore in ownership terms, a largely privatized sector. However, ownership tenure of a HDB dwelling differs in many aspects from ownership of a private dwelling. Ownership rights are limited by numerous regulations concerning eligibility conditions for purchase, resale, subletting and housing loans."* (Phang, 2007, p. 21)
>
> **Central Provident Fund (CPF)**: *"[..] had been [..] established as a pension plan in 1955 by the colonial government to provide social security for the working population in Singapore. The scheme required contributions by both employers and employees, [..]. All employers are required to contribute monthly to the fund. The bulk of contributions can only be withdrawn for specific purposes (of which housing dominates), [..]. The CPF became an important institution for financing housing purchases from September 1968 when legislation was enacted to allow withdrawals from the fund to finance the purchase of housing sold by the HDB and subsequently sold by other public sector agencies as well."* (Phang, 2007, p. 21-22)

Additionally Phang mentions the following stakeholders playing a relevant role: **private real estate developers**, **the government**, **finance houses and commercial banks** and of course the **buyers and tenants of private and public housing**.

2.3.1 Public sector

The public housing market (HDB housing market) can be divided into two sectors: the public home-ownership market and the public rental market. According to Neo *et al.* (2003) the public home-ownership market in turn is divided into three subsectors: the *public new housing market*, the *HDB resale market* and the *HDB executive condominium market*. The public new housing sector contains new housing units developed and sold by the HDB to buyers who meet specified social, demographic and income criteria. The access to this submarket is strongly limited and the prices are highly subsidized. By contrast, the HDB resale sector generates higher housing prices than the new housing sector because prices are driven by market forces. Neo *et al.* (2003) point out that new and resale public housing markets target the low- and middle-income households while the HDB executive condominium market provides high-quality condominiums for upper- and middle-income households.

According to Phang (2007) the public rental market stands for the social housing sector in Singapore. It is regulated by the HDB and provides minimum standard housing for families with a monthly income below 1'500 Singapore Dollar. These households pay monthly rentals of ten to thirty percent of market rents depending on their current monthly income. Additionally a small share of HDB rental housing units is used for transitional families waiting for home ownership and for foreign workers in Singapore.

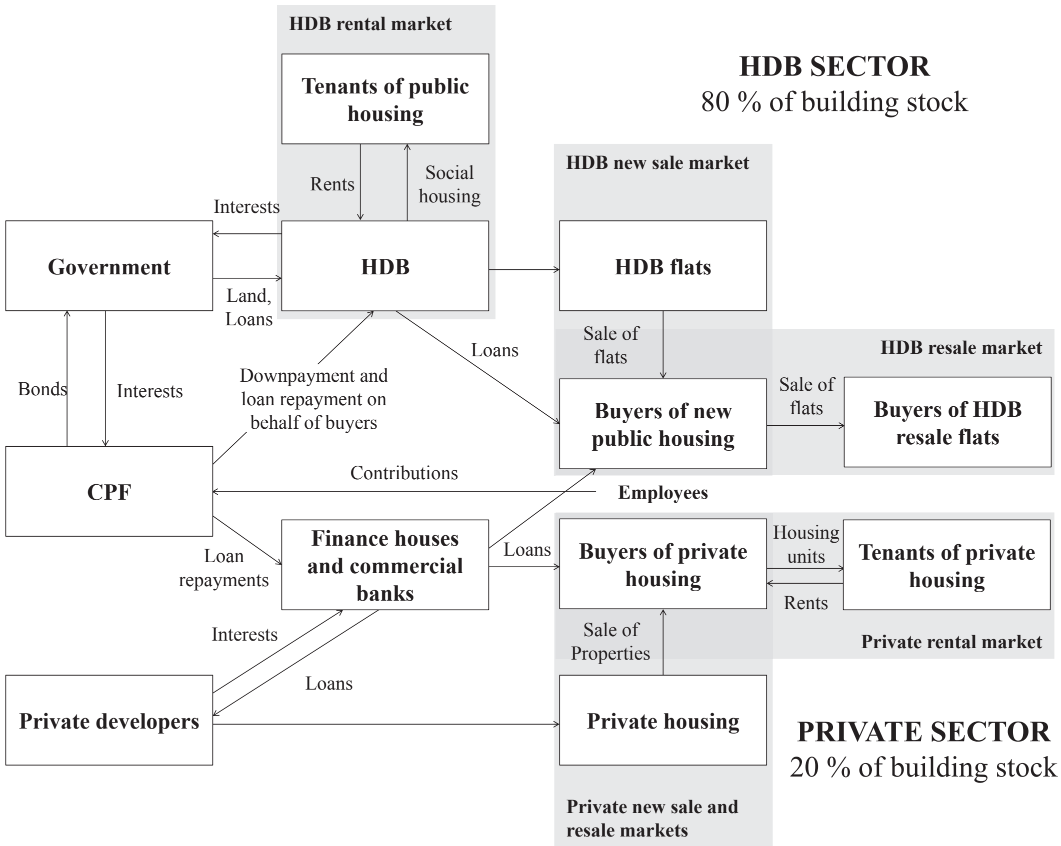
2.3.2 Private sector

The private housing sector in Singapore is a very open and deregulated market and has much higher housing prices. Sing *et al.* (2006) point out that private dwelling units are often better designed and the building quality is higher compared to the public sector. Most of the private flats are located in big condominiums and often equipped with full recreational facilities. The private market can be subdivided into the *private owner-occupier housing market* and the *private rental market*. The owner-occupier market mainly caters the richer part of Singaporean society. It incorporates around ten percent of the total number of housing units, with an increasing share (Neo *et al.*, 2003). Phang (2007) points out that foreign residents' demand in Singapore is limited to private flats and condominiums. The main reason for that is that foreigners need governmental approval to own private landed properties and private flats in buildings of less than six storeys. Within the rental housing market, rents are market determined and the sector basically caters to the expatriate population. As pointed out above, the Singaporean housing market incorporates several players and processes. Figure 1 aims to summarize the major stakeholders and processes in order to show the five relevant market segments.

2.3.3 Market players and processes

As shown in Figure 1 and described above, the Singapore housing market can roughly be divided into five market segments (grey boxes).

Figure 1: Market players and finance processes



Source: (Phang, 2007), modified

There is a chain of dependencies in the public sector within the circle of Employees/Buyers, CPF and HDB. Employees are obliged to contribute a share of their income to the CPF and have to fulfill restrictive requirements to become HDB buyers. If they are eligible to buy, they can enjoy the subsidized mortgage rates provided by the HDB public financing system (Neo *et al.*, 2003). The government provides the mortgage finance loan to the HDB at the prevailing CPF saving interest rate. Neo *et al.* also point out, that since the eighties the scheme was extended to allow for CPF withdrawals for mortgage payments to buy private housing.

2.3.4 Key figures

Table 7 shows the key figures of the important sectors and housing categories in Singapore. It stands out that HDB units represent more than three quarters of the entire housing stock in Singapore. But compared with the figures of 2003 - as found in Sing *et al.* (2006) - the HDB stock share slightly decreased from 79.5 to 77.5 percent. In absolute terms the HDB stock increased from around 815'000 in 2003 units to 885'000 units in 2010. The HDB flat stock is dominated by 3 to 5-room flats while executive flats and 1/2-room flats are of minor importance.

The private housing stock has increased as well for the last eight years. It contained around 210'000 units in 2003, but it includes around 260'000 units in 2010. Almost fifty percent of the private housing units are located in condominiums while other forms of private dwellings only represent small shares. The condominium share has grown from 8.36 percent in 2003 to 10.64 in 2010.

Comparing public and private markets concerning price structure and market activity shows the the huge gap between subsidized (regulated) and free (deregulated) markets. While the median price of a new private apartment is around 825'000 S$, HDB buyers get a 5-room flat for 350'000 S$. Due to fewer subsidizes on the HDB resale market, the median prices for resale HDB flats are higher than for new HDB flats.

Table 7: Housing market key figures

	New sale			Resale			Rentals		Stock	
Private sector	Price[a]	Price[a]	Share[a]	Price[a]	Price[a]	Share[a]	Rent[a]	Share[a]	Number[a]	Share[a]
	[S$'000]	[S$/sqm]	[%]	[S$'000]	[S$/sqm]	[%]	[S$/sqm*a]	[%]	[units]	[%]
Detached house	6'007	13'486	0.18	6'800	10'925	2.48	446.76	2.10	10'350	0.90
Semi-detached house	2'700	9'802	0.75	2'795	8'933	4.52	329.4	3.02	21'185	1.85
Terraced house	2'200	10'787	1.33	1'768	9'192	11.15	279.6	5.74	38'208	3.34
Apartment	826	14'324	30.08	1'010	9'800	26.42			66'638	5.82
Condominium	1'155	11'905	56.72	1'100	9'450	50.41			121'862	10.64
Executive condominium	774	7'717	10.94	800	6'677	5.02				
Non-landed property							417.36	89.14		
Subtotal			100.00			100.00		100.00	258'243	22.55
Public sector (HDB Data)	Price[b]		Share[c]	Price[c]		Share[c]	Rent[c]		Number[b]	
	[S$'000]		[%]	[S$'000]		[%]	[S$/mnt]		[units]	
Executive flat			1.28	548		8.11	2'300		65'077	5.68
5-room flat	350		15.61	460		24.01	2'150		209'765	18.31
4-room flat	240		72.47	385		36.42	2'000		340'069	29.69
3-room flat	155		9.55	300		29.76	1'700		220'770	19.28
2-room flat	90		1.08	234		1.70	1'300		30'210	2.64
1-room flat									21'217	1.85
Subtotal			100.00			100.00			887'108	77.45
Total									1'145'351	100.00

Sources:

[a] Computed based on median prices of all private residential properties transacted from 1 September 2010 to 28 February 2011, as captured in REALIS (URA, 2011), Stock: 4th quarter of 2010

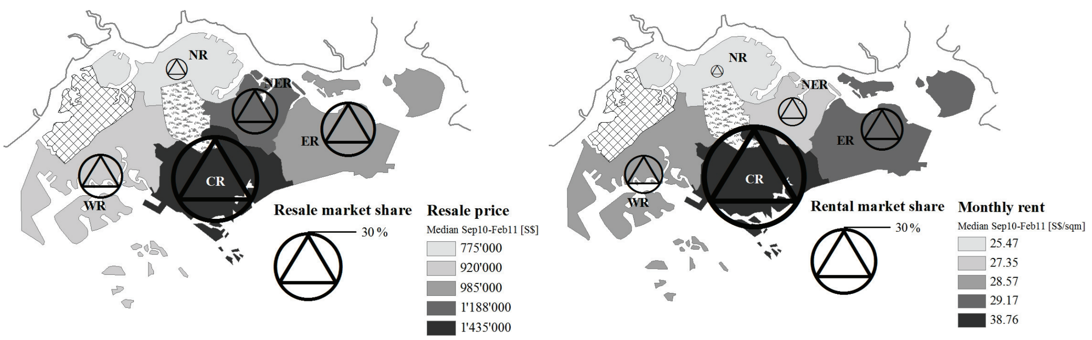
[b] HDB new flat prices are obtained from the HDB Annual Report 2009/2010 (Annex "Key statistics for FY 2009/2010") (HDB, 2011a), bandwidths are calculated to average values

[c] HDB resale and open market rental data for the 4th quarter of 2010 is obtained from the official public HDB website (HDB, 2011d)

2.3.5 Regional dependence of market shares and prices

Singapore is divided into five planning regions called central region (CR), east region (ER), north region (NR), north-east region (NER) and west region (WR). Figure 2 shows market shares and prices for private resale and rental markets on a regional basis. Market shares represent the proportion of transactions in a certain region to the entire number of transactions in the time between September 2010 and February 2011. Prices are median values for the same period. The maps clearly demonstrate the outstanding role of the CR while this dominance is stronger in the rental market than in the resale market. The median resale price in this area is almost twice as high as in the NR and the rental market transactions in the CR stand for around 65 percent of the entire number of transactions. The resale market in general is less fragmented than the rental market showing smaller relative price differences between ER, NER and WR and smoother distributions of market shares.

Figure 2: Regional price characteristics

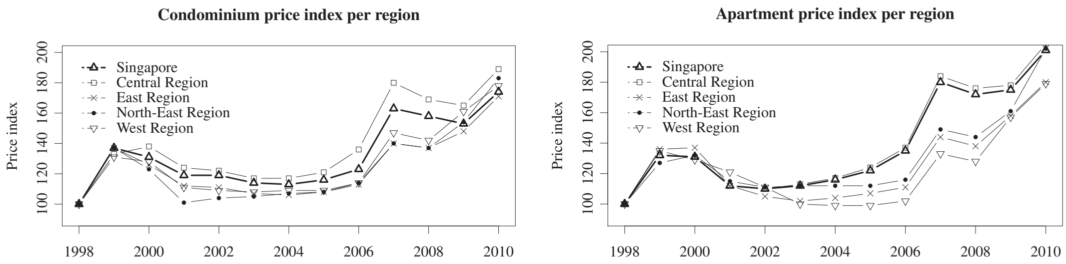


Source: Data obtained from REALIS (URA, 2011)

2.3.6 Medium term price development

Around eighty percent oft the transacted private units are condominiums and apartments (see Table 7). Figure 3 shows the medium term price development (indexed) of these two housing types from 1998 to 2010 while distinguishing between planning regions (unfortunately price index data for the NR is not available from REALIS). Both indices show the same general structure: A steep increase between 1998 and 1999, further a slight decrease respectively a stagnation from 2000 to 2006 and then another increase with a stabilization on a high level until today.

Figure 3: Temporal price development



Source: Data obtained from REALIS (URA, 2011)

Condominium price development reflects relatively good the spatial structure in Singapore. As expected, prices in the CR outperformed during bull markets compared to the rest of the island (in figure: "Singapore"). On the other hand the WR and the NER clearly underperformed during growth phases but almost catched up during the last bull markets between 2008 and 2010. For the apartment sector the price development shows a different trend. Prices in the CR outperformed much more clearly than in the condominium sector. ER and WR stayed clearly below average performance while the NER used to underperform until 2009 but then catched up with the CR by 2010.

2.4 Hypotheses and expectations

Table 8 shows what key determinants are gathered and what their expected price impact is. The literature review showed that floor area always turned out to have a strong positive impact on a housing unit's price. Positive price impacts are furthermore expected from the availability of amenities such as car park, pool, sport facilities as well as security services. It is also expected that HDB flats which have been subject to a main upgrading program (MUP) yield higher market prices than others. Flats with a freehold contract are expected to be more expensive than flats with a limited lease. It is furthermore expected, that older flats yield lower prices than newer ones. Negative price impacts are also expected from the distance to the CBD as well as from distances to other points of interest (access to public transportation, top schools, shopping malls and car parking). Finally it is assumed that earlier transactions lead to lower prices compared to recent ones because prices have dramatically increased in the last few months (see Section 2.3.6).

It is expected that housing markets (private/HDB, sale/rental) vary concerning price level, structural and locational characteristics as well as housing preferences.

Table 8: Expected impacts of key price determinants

Housing characteristics	Expected sign
Structural variables	
Floor area	+
Age	-
Floor level	+
Availability of amenities (pool, wellness, security etc.)	+
Main Upgrading Program (MUP) completed	+
Locational variables	
Distance to the CBD	-
Distance to nearest MRT and bus stations	-
Distance to nearest top/good performance school	-
Distance to nearest shopping mall	-
Distance to nearest public car parking	-
Contactual variables	
Freehold tenure	+
Duration since transaction	-

3. Data

Before carrying out statistical analysis and modelling, representative data has to obtained. This chapter therefore describes the data gathered in order to assess representativeness of the sample. The data is segmented and compared concerning prices, structural characteristics and locations. This chapter also aims to test, if the market segments as described by Phang (2007) are present in the data.

3.1 Asking and transaction listings

Data was gathered from several sources between February and May 2011 using automatic web robots developed by Michael van Eggermond (PHD student at Future Cities Laboratory). Two kinds of housing listings are obtained: asking data and transaction data. Asking prices are obtained from the commercial online portal Property Guru (All-property, 2011) which contains listings from the private sale and rental markets as well as from HDB resale and rental flats. Data on housing transactions was gathered from the URA real estate portal (URA, 2011) and from the HDB InfoWEB (HDB, 2011b). The transaction listings contain housing transactions between July 2010 and March 2011.

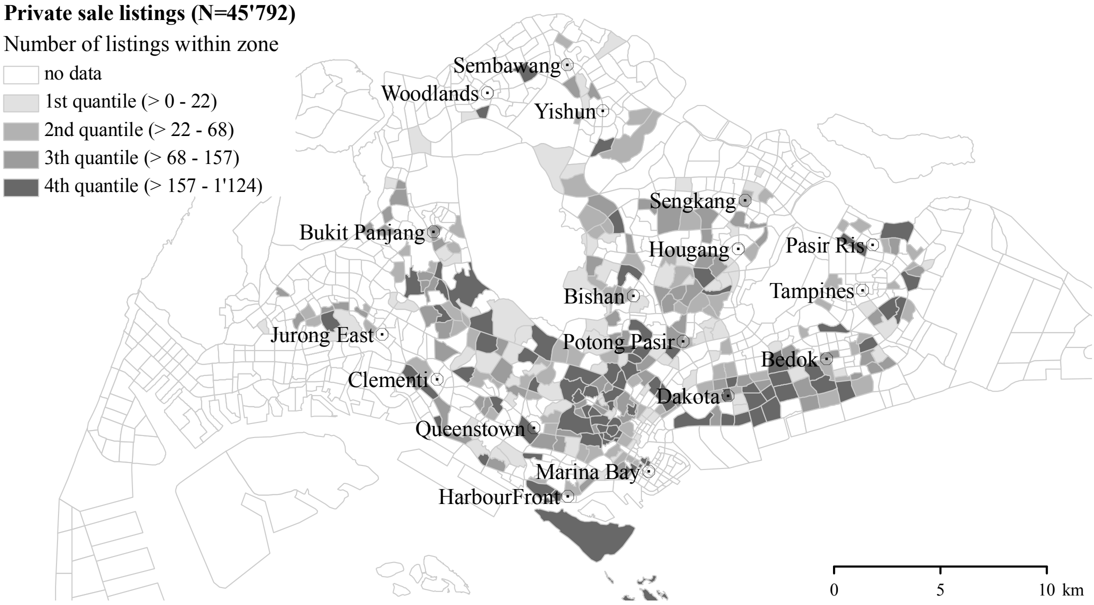
Table 9: Overview of the gathered data

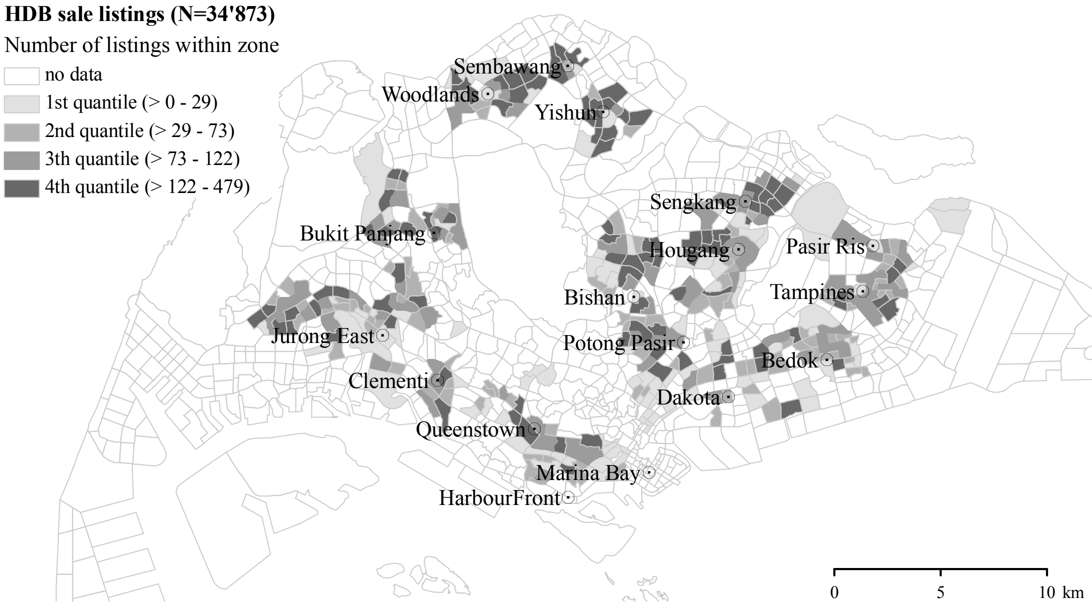
	Asking price data		**Transaction price data**		**Total**	
	N=	[%]	N=	[%]	N=	[%]
Private housing						
Sale	33’325	30.6	12’467	11.4	45’792	42.0
Rental	22’011	20.2	no data		22’011	20.2
Subtotal	55’336	50.8	12’467	11.4	67’803	62.2
HDB housing						
Sale	2’638	2.4	32’235	29.6	34’873	32.0
Rental	6’351	5.8	no data		6’351	5.8
Subtotal	8’989	8.2	32’235	29.6	41’224	37.8
Total						
Sale	35’963	33.0	44’702	41.0	80’665	74.0
Rental	28’362	26.0	no data		28’362	26.0
Total	64’325	59.0	44’702	41.0	**109’027**	100.0

Table 9 shows numbers and percentages of the different segments after exclusion of doublets, listings with missing attributes and outliers (as described in Section 4). Around 1% of the raw data is excluded. Roughly 110'000 observations can be used for further data analysis and model estimation. Asking price data is considered to represent *expected preferences* while transaction price data reflects *revealed preferences*. Besides the price, the collected listings include structural information such as floor area, construction year, number of rooms, floor level and availability of facilities (pool, wellness, car park, security services etc.). Additionally the observations contain an address or a building name making it possible to locate them in space. For geo-location a Google maps API (Google Maps, 2011) was used. Transaction data also includes information about the transaction date while HDB upgrading program data is obtained from the HDB website (HDB, 2011c).

Figure 4 shows the spatial distribution of the gathered sale data. Asking and transaction price listings are combined and aggregated to the planning zone level. The maps clearly show that private and public housing markets are spatially segregated. There are absolutely no HDB listings available in the whole Tanglin district between Queenstown and Orchard Road. The same region contains relatively numerous listings for the private housing market. In general the private market seems to be focused towards the city centre and the central business district. HDB listings on the other hand are particularly located where the big HDB towns are: Jurong East and Jurong West, Woodlands, Sengkang, Hougang, Bedok and Bukit Merah. The HDB data therefore shows a much more decentralized spatial structure than data of the private market. The north of the island (Woodlands, Sembawang, Yishun) is poorly covered by private market listings. Analogue maps for the rental markets are given in Appendix A.1 and show a similar spatial pattern.

Figure 4: Spatial distribution of gathered sale listings





3.2 Points of interest

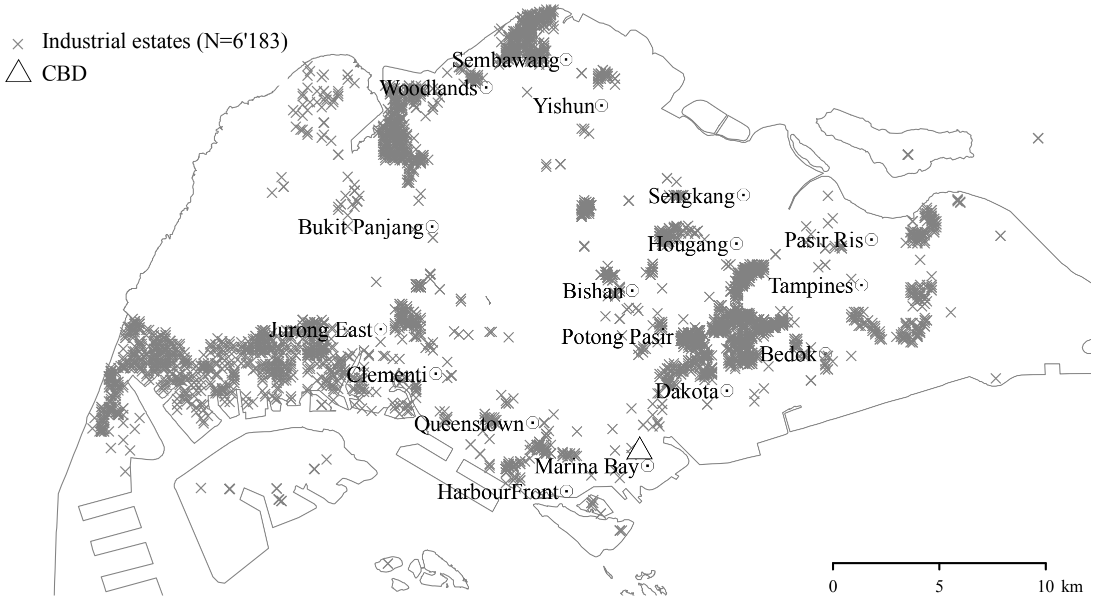
Distances to points of interest (POI) are used in hedonic price models to incorporate locational price effects. Data of POI is therefore gathered and geo-coded. Table 10 shows data sources of different POI and the variables which have been generated for model estimation.

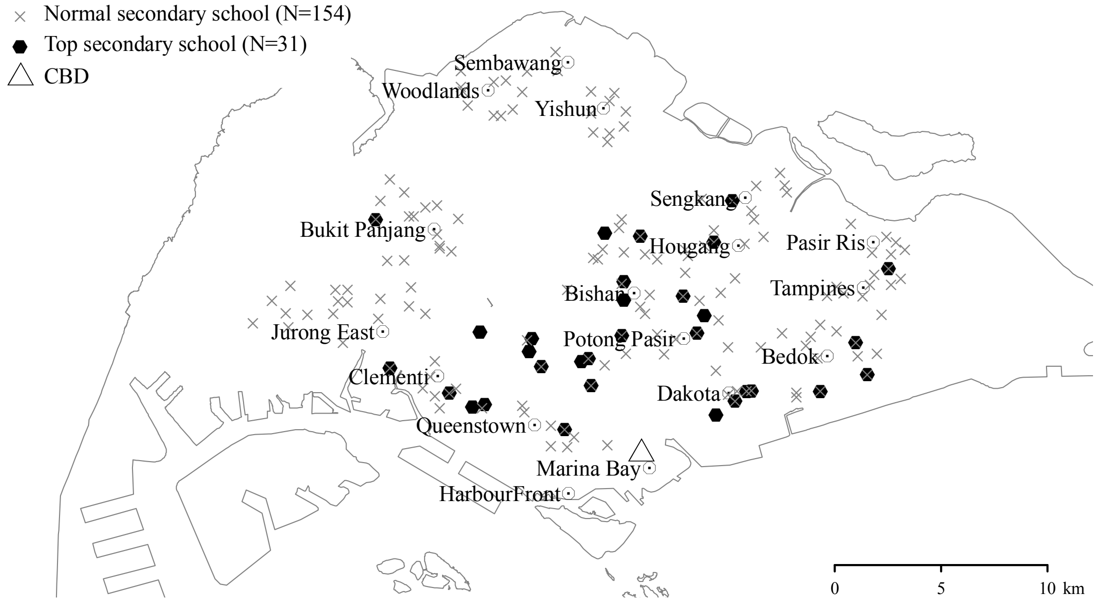
Table 10: POI data sources and variables

POI data	Source	Computed variables
Public transport	NAVTEQ (2011)	Distance to nearest bus station
		Number of lines at nearest station
		Distance to nearest MRT station
Private transport	Streetdirectory (2011)	Distance to nearest car park
		Distance to nearest multi storey car park
Shopping	Streetdirectory (2011)	Distance to nearest shopping mall
Daily supply	Streetdirectory (2011)	Distance to nearest food centre
	Giant (2011),NTUC (2011)	Distance to nearest supermarket
Work places	Streetdirectory (2011)	Distance to nearest industrial estate
		Distance to central business district
Education	Ministry of Education (2011)	Distance to nearest primary school
		Distance to nearest secondary school
		Distance to nearest junior college
	PAEXCO (2011)	Distance to nearest top primary school
		Distance to nearest top secondary school

As shown in Section 2.2.2 the distance to the central business district (CBD) turned out to be a very important variable in several studies. For Singapore, **Raffles Place** is the referenced point for the CBD. As an example, the spatial distribution of industrial estates and secondary schools are shown in Figure 5. Maps of the spatial distribution patterns of other POI are given in Appendix A.2. Euclidean distances to the nearest POIs are computed for every single listing in order to use them as locational variables in the regression models. The procedure is carried out in ArcGIS 9.3 (ESRI, 2011) using the free toolbox Hawth's Analysis Tools (Beyer, 2004).

Figure 5: Spatial distribution of industrial estates and secondary schools





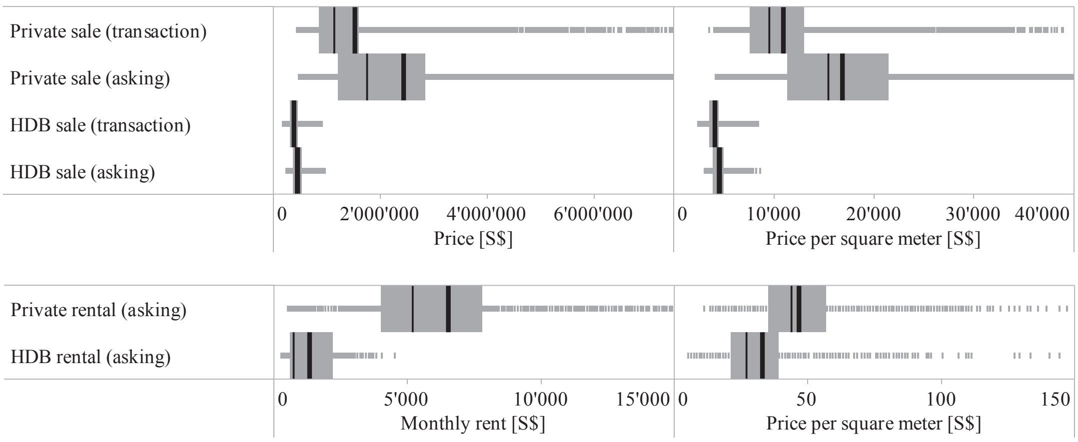
3.3 Comparison of market segments

This section aims to compare the different data segments (asking/transaction prices, private/HDB, sale/rental). The gathered listings are compared in terms of prices as well as structural and locational characteristics.

3.3.1 Price comparison

The box-plot style graphs in Figure 6 clearly show the price gaps between the different segments. The grey boxes represent fifty percent of the observations (2nd and 3rd quartile), the fat vertical line in the box represents the mean value and the thin one the median. The average asking price in the private sale market of 2’840’000 S$ is about sixty percent higher than the average unit price of the transaction data (1’524’000 S$). It stands out that there is no big difference between asking and transaction prices in the HDB sector (asking prices are around 14% higher).

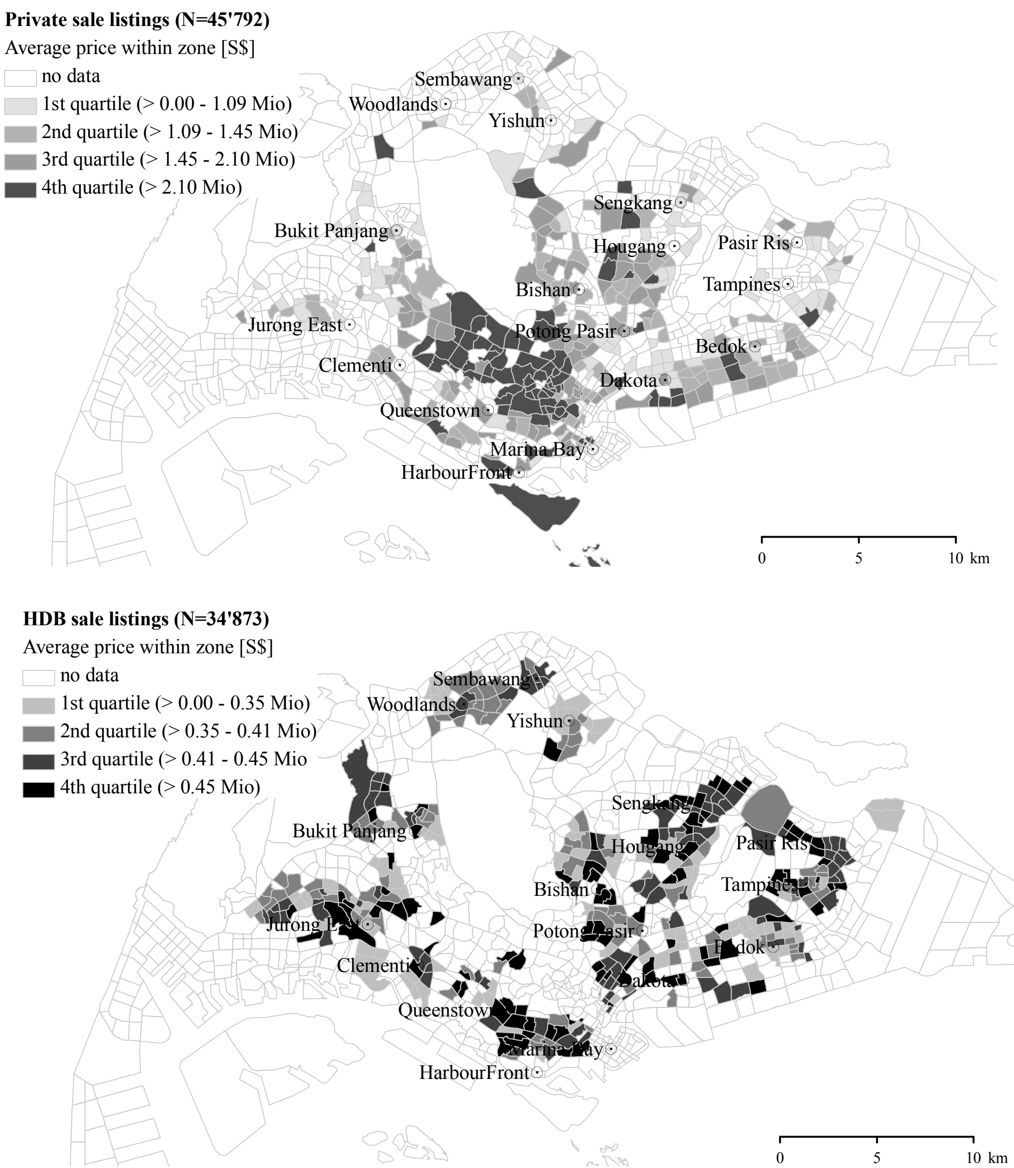
Figure 6: Comparison of unit prices and monthly rents



Grey box: 2nd and 3rd quartile | Vertical lines: bold=Mean, thin=Median

Figure 6 also shows price differences for square meter prices (right hand side). In general the differences show the same structure as the absolute prices. But it stands out that rental square meter prices vary less than absolute prices. This can be explained with a big difference in average floor area between these two market segments (see Figure 8). The plots further show that the price differences between private and HDB flats are very big, as expected. The average transaction price of a private flat (1'524'000 S$) is almost four times higher than the average HDB flat price (395'000 S$). The average flat price not only varies between market segments but also across geographical space. Figure 7 shows the spatial price structure for the private and HDB sale markets. The maps clearly show that private prices vary stronger over space and increase towards the city centre.

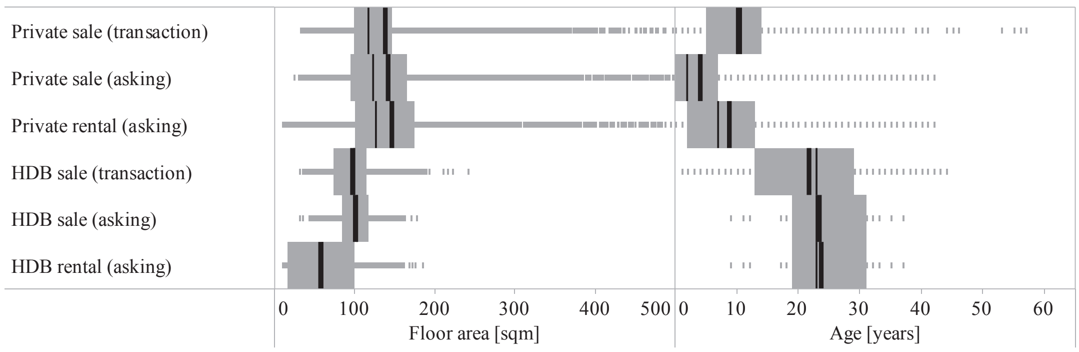
Figure 7: Spatial distribution of average prices in the sale markets



3.3.2 Structural comparison

Besides the price gaps there are structural differences between the data market segments as well. Figure 8 shows that private sale flats are about fifty percent larger than HDB flats. HDB rental flats - which are considered to reflect the highly subsidized social housing market (Phang, 2007) - are only half as large as HDB sale flats and almost three times smaller than private sale flats. It stands out that there is no significant difference between asking and transaction price data concerning the floor area. Figure 8 also shows that HDB flats are much older in general. The average age of a sold HDB flat is 27.7 years while the equivalent value in the private sale market is 10.5 years. The reason for this difference is probably that new HDB flats (public new housing market, see Section 2.3.3) do not appear in the HDB resale listings.

Figure 8: Comparison of floor area and age

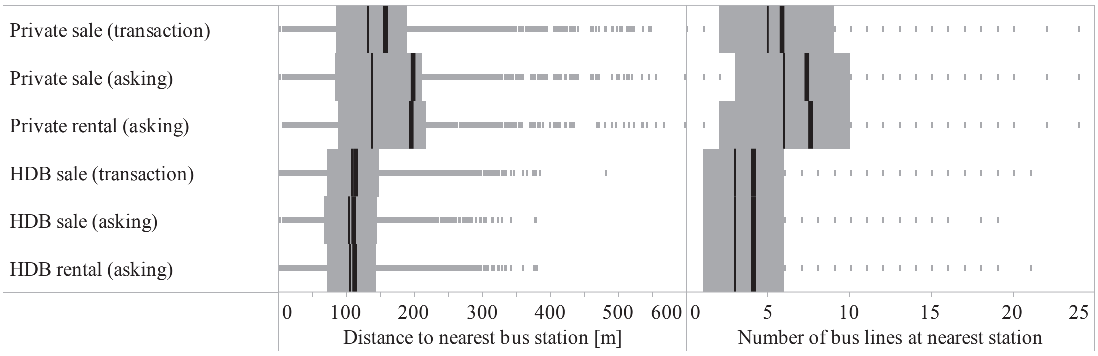


Grey box: 2nd and 3rd quartile | Vertical lines: bold=Mean, thin=Median

3.3.3 Locational comparison

Besides the above mentioned market differences there is also considerable locational variability between the segments. Figure 9 shows that HDB flats are in average much closer to the nearest bus stop than private flats. But these stops provide only around four bus lines on average while the average nearest stop to private flats provides around seven lines. The proximity to bus stops of HDB flats (110 meters in average) can be explained with the hight spatial concentration of HDB flats in HDB towns. These towns normally have centrally located access to public transportation. Both distances to bus stops and numbers of bus lines do not vary significantly between asking and transaction price data.

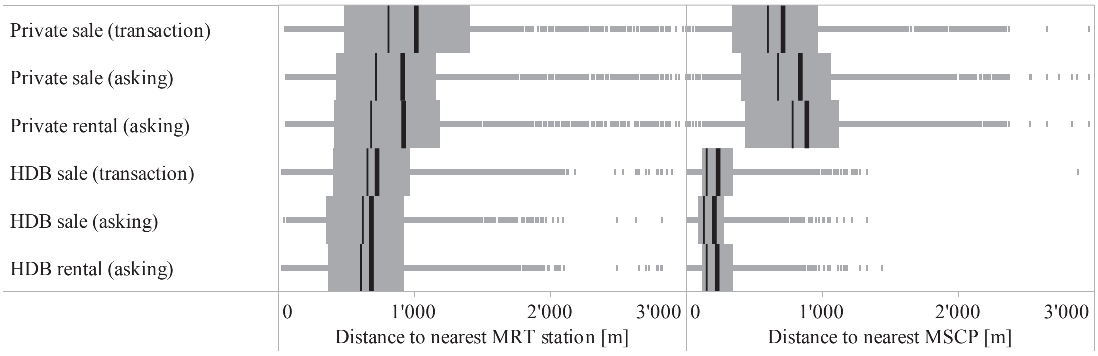
Figure 9: Comparison of distances to bus stop and number of bus lines



Grey box: 2nd and 3rd quartile | Vertical lines: bold=Mean, thin=Median

A similar segmentation can be seen in Figure 10 where average distances to nearest MRT stations and distances to nearest multi storey car parks (MSCP) are shown. HDB flats are closer to both MRT stations and MSCP. The proximity to MRT stations can be explained with the same arguments as the proximity to bus stops (see above). The proximity to MSCP becomes clear when looking at the spatial distribution of MSCP (see Figure 24 in Appendix A.2). Lots of the 783 gathered MSCP are located in HDB towns (especially around Woodlands, Sembawang, Bukit Panjang, Jurong West, and Sengkang) and very few of them are in the city centre. The differences between asking and transaction price listings are not very big in both cases.

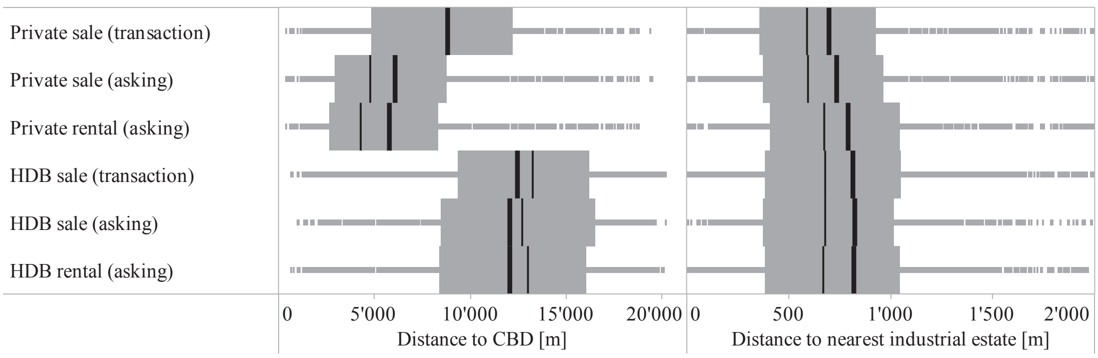
Figure 10: Comparison of distances to MRT stations and multi storey car parks



Grey box: 2nd and 3rd quartile | Vertical lines: bold=Mean, thin=Median

Figure 11 shows that private flats are closer to the central business district (CBD) than HDB flats as expected after plotting the data (see Figure 4). But it stands out that private asking price listings are in average around 30% closer to the CBD than private transaction price listings. In the HDB sector, there is no big difference between asking and transaction data in terms of distance to the CBD. The plots further show that the mean and spread of the distance to the nearest industrial estate are almost equal in all segments. This is surprising because the majority of the industrial estates is located in the peripheral areas in the north and south west of the island (see Figure 5). It was therefore expected that HDB flats would be closer to industrial estates on average.

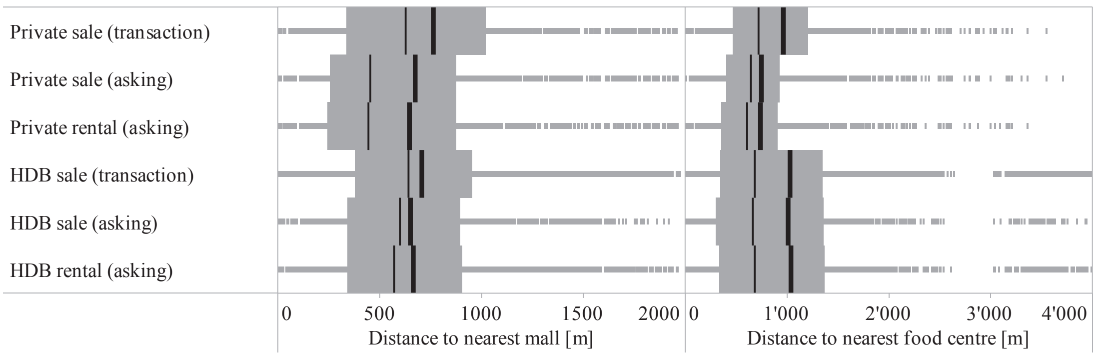
Figure 11: Comparison of distances to the CBD and industrial estates



Grey box: 2nd and 3rd quartile | Vertical lines: bold=Mean, thin=Median

Figure 12 shows that there is no notable difference in the proximity to shopping malls and food centres between private and public markets. But it stands out that private asking price listings are on average around twenty percent closer to food centres than all other market segments. Further analysis shows that transaction price listings are wider distributed than asking price listings (histograms are given in Appendix A.3). But the most numerous observations appear in the range between 401 and 600 meters in both the asking and the transaction price listings.

Figure 12: Comparison of distances to malls and food centres



Grey box: 2nd and 3rd quartile | Vertical lines: bold=Mean, thin=Median

3.4 Representativeness of the samples

Transaction listings include all housing transactions of the last six months and are considered to be highly representative. Transaction data is therefore used as reference to assess representativeness of the asking price data. Table 11 gives a summary review of the main differences between asking and transaction sales data. Rental data cannot be reviewed since there is no ex post data for the rental markets. The right hand column clearly shows that HDB asking data can be classified as very representative since structural and locational characteristics are very similar to transaction data. The price difference of 15% can be explained with the general price increase of the last months (as shown in Section 2.3.6).

Representativeness of the private sale asking listings must be questioned. As Table 11 shows, asking price flats are 60% younger, 30% closer to the CBD and 60% more expensive than transaction price flats. It is assumed that around 15-20% of the price difference can be explained with the general price increase as described above. And it is expected that newer flats which are closer to the CBD yield higher prices. But it stays an open question why the asking price data obviously only covers a part of the private sale market. It is assumed that some market players use different and/or informal trading platforms which cannot be captured on commercial real estate websites.

Comparison of the data further confirmed the assumption of two distinct markets, the private and the HDB sectors as identified by Phang (2007). These markets are clearly different with regard to prices, structural properties and locations.

Table 11: Comparison of asking and transaction data

Asking vs. transaction	Private market	HDB market
Price	+60%, larger spread	+15%, similar spread
Structure		
Floor area	+3%, similar spread	+3%, similar spread
Age	-60%, similar spread	+5%, smaller spread
Location		
Distance to the CBD	-30%, similar spread	-3%, similar spread
Distance to nearest bus station	+26%, similar spread	+2%, similar spread
Distance to nearest MRT station	-9%, similar spread	-6%, similar spread
Distance to nearest industrial estate	+5%, similar spread	+1%, similar spread

4. Model estimations

Different functional specifications of hedonic equations can be found in literature. Most studies used either semi-log or log-log specifications (Malpezzi, 2002). In semi-log models either the dependent variable or the explanatory variables are transformed. In log-log models both sides are logarithmized. Regression coefficients of semi-log models can be interpreted as the relative change of the dependent variable given a change of the explanatory variable. Log-log coefficients on the other hand can be interpreted as elasticities. Elasticities are approximately the change of the dependent variable in percent if the explanatory variable changes one percent. Different alternative model specifications have been used as well. Box and Cox (1964) introduced the so called Box-Cox transformation and others (for example Fahrländer (2006)) used nonparametric methods. In this thesis, only semi-log and log-log specifications are used due to their good economic interpretability and comparability.

4.1 Methodology

The following procedure was applied to all market segments (Models A-F) and was carried out with the environment for statistical computing R (R Development Core Team, 2011).

1. First variable selection using stepwise OLS regressions
2. Test for heteroscedasticity and change of variable transformation if necessary
3. Outlier exclusion using studentised residuals
4. OLS and SAR coefficient estimation and final variable selection
5. Test for spatial autocorrelation

After stepwise variable selection and heteroscedasticity tests potential outliers were identified. Observations with studentised residual exceeding 3 were excluded as proposed by Fotheringham *et al.* (2002). Variables with a variance inflation factor (VIF) exceeding 10 are excluded as well. The VIF provides an index that measures how much the variance of an estimated regression coefficient is increased because of collinearity (Studenmund, 2006). For estimation of the spatial error and durbin models (SAR models as described in Section 2.1.1) the R packages *sp* (Pebesma, 2011) and *spdep* (Bivand, 2011a) were used. Estimation of geographically weighted regressions were carried out with package *spgwr* (Bivand, 2011b). For the SAR models, a spatial weights matrix is used (see Section 2.1.1). A k-nearest-neighbors approach was applied where the weights are row standardized. K was chosen for every model by optimizing model fit as proposed by Gelfand *et al.* (2010). Model fit for comparison is measured by the sum of squared errors (SSE). SSE indicate better model fit if the test value is smaller.

Besides the model coefficients, standard errors (SE), coefficients for standardized variables (Scaled; centering is done by subtracting the column means of the variables from their corresponding columns) and t statistics (T stat.) are computed. The SE represents the average difference between the estimated coefficient and the true coefficient. Scaled coefficients are computed because they can be directly compared with each other. The higher the absolute value the stronger the impact of a variable. The t-test is used to examine the hypothesis that a regression coefficient is actually equal to zero. Higher t-values indicate a higher precision of the estimated parameter.

All models were tested for spatial autocorrelation with the Moran's I statistic (Cliff and Ord, 1981) and Lagrange multipliers tests (Anselin *et al.*, 1996). Moran's I can take positive and negative values from -1 to 1. A Moran's I of -1 indicates perfect dispersion while a test value of 1 assumes perfect spatial correlation. A zero value indicates a random spatial pattern (Fischer and Getis, 2010). According to Anselin *et al.* (1996) Lagrange multiplier tests indicate if spatial errors (LMerr) or spatial lags (LMlag) are present in the OLS model and give a good guide to decide which specification for the SAR models is the most appropriate. Higher LM values indicate stronger spatial dependence.

Different models are compared in order to show differences in housing preferences of different markets. To compare goodness of models in general, the Akaike's information criterion (AIC) is used. This index basically takes into account both the statistical goodness of fit and the number of parameters that have to be estimated to achieve this particular degree of fit (Sakamoto and Kitagawa, 1987). Lower AIC values indicate a better model specification.

4.2 Variable selection and descriptive statistics

Table 12 shows the six models which are chosen for estimation. Rental models are estimated with asking price data only due to unavailability of transaction/contractual price data. Combined models are used to estimate constants for the surcharge of asking prices and to take advantage of a bigger number of observations. The number of variables differs because of two reasons: Not all variables are available for each market segment and stepwise exclusion is based on a correlation analysis which differs from market to market. An overview of all variables available is given in Appendix A.4.

Table 12: Overview of estimated models

Name	Market	Data type	Number of observations	Number of variables[1]
Model A	Private sale	Asking and transaction	45’792	23
Model B	Private sale	Transaction	12’467	21
Model C	Private rental	Asking	22’011	16
Model D	HDB sale	Asking and transaction	34’873	13
Model E	HDB sale	Transaction	32’235	21
Model F	HDB rental	Asking	6’351	8

[1]Groups of dummies count as one variable

4.2.1 Model A: Private sale combined

Table 13 shows the descriptive statistics of the final variable selection in the private sale combined market. This data set contains more than 45’000 observations out of which around 73% are asking price listings (variable ASKING) and 27% transaction price listings. It stands out that the building stock represented in this data set is quite young: 47% of the flats have been built between 2001 and 2010 (BUI_0110) and less then 1% of the flats have been built before 1970 (BUI_5160 and BUI_6170). 71% of the observations are condominiums (variable CONDO) which is much more than in the existing building stock (see Section 2.3.4, where condominiums represent a total share of 47% of the private properties). Fifty percent of the observed flats further have a parking within the building and almost 90% have a swimming pool available.

Table 13: Descriptive statistics of private sale combined data (Model A)

N = 45'792

Variable	Description	T[1]	Min	Max	Mean	S.D.[3]
Dependent variables						
PRICE	Price [S$]	C	427'000	118.12 Mio.	2.19 Mio.	2.13 Mio.
SQMPR	Price per sqm [S$]	C	3'451.12	51'136.32	15'250.87	7'091.05
Structural explanatory variables						
SIZE	Floor area in square meter	C	24.00	7'219.00	140.95	100.92
BUI_5160	Built between 1951 and 1960	D	0.00	1.00	0.00	
BUI_6170	Built between 1911 and 1970	D	0.00	1.00	0.00	
BUI_7180	Built between 1971 and 1980	D	0.00	1.00	0.02	
BUI_8190	Built between 1981 and 1990	D	0.00	1.00	0.04	
BUI_9100	Built between 1991 and 2000	D	0.00	1.00	0.20	
BUI_0110	Built between 2001 and 2010	D	0.00	1.00	0.47	
BUI_1220	Planned for after 2011	D	0.00	1.00	0.19	
CONDO	Building is a condominium	D	0.00	1.00	0.71	
PARKING	Availability of a car park	D	0.00	1.00	0.50	
WELL	Availability of wellness	D	0.00	1.00	0.51	
SEC	Availability of security	D	0.00	1.00	0.53	
POOL	Availability of a pool	D	0.00	1.00	0.89	
GARD	Availability of a garden	D	0.00	1.00	0.67	
Locational explanatory variables						
CBD	Distance to CBD [m]	C	348.00	19'458.00	6'833.57	4'533.44
INDUS	Distance to an industrial estate [m]	C	1.00	3'581.00	725.12	495.67
MSCP	Distance to a MSCP[2] [m]	C	1.00	3'860.00	800.19	581.62
BUSLINES	Number of bus lines at nearest stop	C	0.00	33.00	6.99	5.43
PRIM	Distance to a primary school [m]	C	13.00	3'988.00	770.16	548.46
MALL	Distance to a mall [m]	C	1.00	4'006.00	697.92	603.77
TOPPRIM	Distance to a top primary school [m]	C	14.00	9'107.00	2'801.68	1'826.55
FOOD	Distance to a food centre [m]	C	1.00	4'956.00	807.40	620.56
BUS_26	Nearest bus stop within 200-600 m	D	0.00	1.00	0.23	
MRT	Distance to an MRT station [m]	C	45.00	3'856.00	941.31	704.63
SECOND	Distance to a secondary school [m]	C	1.00	5'106.00	973.19	704.02
SUPERM	Distance to a supermarket [m]	C	86.00	8'284.00	1'792.23	1'311.24
TOPSEC	Distance to a top secondary school [m]	C	1.00	3'426.00	587.03	453.51
Contractual and data source dependent variables						
ASKING	Is an asking price	D	0.00	1.00	0.73	
FREE	Contract is freehold	D	0.00	1.00	0.51	

[1]Type: C=continuous variable, D=dummy variable; [2]MSCP=multi storey car park; [3]S.D.=standard deviation

4.2.2 Model B: Private sale transaction

The descriptive statistics of the private sale transaction market are given in Table 14. The mean price per square meter is with 10'955 S$ around 16 % higher than in the reference data of the last six months (see Section 2.3.4). This can be explained with the general price increase of the last months. Fifty percent of the flats are located between the first and the fifth floor level (FLO0105) and another 23% between the sixth and the tenth (FLO0610). Less than 1% of the listed flats are located higher than on the 36th floor (FLO3640 and FLO41UP).

Only five percent of the observed transactions have been made in the second quarter of 2011. The other three quarters are almost equally represented with 36% in third quarter 2010 (Q0310), 33% in the forth quarter 2010 (Q0410) and around 27% of the first three months of 2011 (Q0111). Around 35% of the observed flats were bought by individuals previously living in HDB estates (FROMPUB).

Table 14: Descriptive statistics of private sale transaction data (Model B)

N = 12'467						
Variable	Description	T[1]	Min	Max	Mean	S.D.[3]
Dependent variables						
PRICE	Transaction price [S$]	C	0.43 Mio.	118.12 Mio.	1.52 Mio.	2.25 Mio.
SQMPR	Transaction price per square meter [S$]	C	3'451.00	47'016.00	10'955.25	4'770.15
Structural explanatory variables						
SIZE	Floor area in square meter	C	32.00	7'219.00	137.68	144.40
CONDO	Building is a condominium	D	0.00	1.00	0.60	
YEAR	Construction year of the building	C	1'954.00	2'011.00	2'000.54	7.65
PARKING	Availability of a car park	D	0.00	1.00	0.62	
FLO0105	Floor level 1-5	D	0.00	1.00	0.50	
FLO0610	Floor level 6-10	D	0.00	1.00	0.23	
FLO1115	Floor level 11-15	D	0.00	1.00	0.13	
FLO1620	Floor level 16-20	D	0.00	1.00	0.07	
FLO2125	Floor level 21-25	D	0.00	1.00	0.03	
FLO2630	Floor level 26-30	D	0.00	1.00	0.02	
FLO3135	Floor level 31-35	D	0.00	1.00	0.01	
FLO3640	Floor level 36-40	D	0.00	1.00	0.00	
FLO41UP	Floor level > 40	D	0.00	1.00	0.00	
WELL	Availability of wellness	D	0.00	1.00	0.41	
POOL	Availability of a pool	D	0.00	1.00	0.87	
Locational explanatory variables						
CBD	Distance to CBD [m]	C	348.00	19'337.00	8'810.95	4'602.24
INDUS	Distance to an industrial estate [m]	C	1.00	2'886.00	698.02	477.79
MSCP	Distance to a MSCP[2] [m]	C	1.00	3'379.00	712.10	504.85
TOPPRIM	Distance to a top primary school [m]	C	86.00	6'999.00	1'971.74	1'379.27
PRIM	Distance to a primary school [m]	C	17.00	3'481.00	730.22	455.22
TOPSEC	Distance to a top secondary school [m]	C	41.00	9'103.00	1'743.95	1'459.52
BUSLINES	Number of bus lines at nearest stop	C	1.00	32.00	5.87	4.63
SECOND	Distance to a secondary school [m]	C	1.00	4'666.00	837.52	563.30
BUS_26	Nearest bus stop in 200-600 meters	D	0.00	1.00	0.22	
MALL	Distance to a mall [m]	C	1.00	3'826.00	765.21	560.63
MRT	Distance to an MRT station [m]	C	45.00	3'740.00	1'010.65	709.51
Contractual variables						
FREE	Contract is freehold	D	0.00	1.00	0.46	
FROMPUB	Buyer lived in a HDB flat before	D	0.00	1.00	0.35	
Q0310	Transaction in 3rd quarter of 2010	D	0.00	1.00	0.36	
Q0410	Transcation in 4th quarter of 2010	D	0.00	1.00	0.33	
Q0111	Transaction in 1st quarter 2011	D	0.00	1.00	0.27	
Q0211	Transaction in 2nd quarter 2011	D	0.00	1.00	0.05	

[1]Type: C=continuous variable, D=dummy variable; [2] MSCP=multi storey car park; [3]S.D.=standard deviation

4.2.3 Model C: Private rental asking

The descriptive statistics of the final variable selection for the private rental market is shown in Table 15. The average monthly rent per square meter of 46.74 S$ is around thirty percent higher than the reference price of the last six months (see Section 2.3.4). This was expected due to the general higher asking prices, as already described above.

Table 15: Descriptive statistics of private rental asking data (Model C)

N = 22'011

Variable	Description	T[1]	Min	Max	Mean	S.D.[3]
Dependent variables						
PRICE	Asking monthly rent [S$]	C	1800.00	46'000.00	6'655.87	3'986.55
SQMPR	Asking monthly rent per square meter [S$]	C	11.00	200.00	46.36	15.46
Structural explanatory variables						
SIZE	Floor area in square meter	C	12.00	957.00	148.92	80.08
YEAR	Construction year of the building	C	1'969.00	2'013.00	2'002.42	8.41
CONDO	Building is a condominium	D	0.00	1.00	0.71	
GARD	Availability of a garden	D	0.00	1.00	0.67	
WELL	Availability of wellness	D	0.00	1.00	0.50	
Locational explanatory variables						
CBD	Distance to CBD [m]	C	348.00	18'737.00	5'752.69	4'326.96
INDUS	Distance to an industrial estate [m]	C	1.00	4'156.00	792.27	541.53
BUS	Distance to a bus stop [m]	C	6.00	2'502.00	197.10	298.11
MSCP	Distance to an MSCP[2] [m]	C	1.00	3'860.00	890.95	604.63
BUSLINES	Number of bus lines at a bus stop	C	1.00	33.00	7.60	6.21
MRT	Distance to an MRT station [m]	C	52.00	3'891.00	919.29	734.07
SUPERM	Distance to a supermarket [m]	C	1.00	3'426.00	593.69	485.41
SECOND	Distance to a secondary school [m]	C	1.00	5'106.00	1'087.41	732.84
FOOD	Distance to a food centre [m]	C	1.00	4'871.00	738.24	562.14
PRIM	Distance to a primary school [m]	C	17.00	3'988.00	856.86	568.20
Contractual variables						
FREE	Contract is freehold	D	0.00	1.00	0.52	

[1]Type: C=continuous variable, D=dummy variable; [2] MSCP=multi storey car park; [3]S.D.=standard deviation;
[4]SAP=Special Assistance Plan

4.2.4 Model D: HDB sale combined

The descriptive statistics of the final variable selection for the HDB sale combined model is shown in Table 16. Only 8% of the observations are asking listings (variable ASKING). The mean price of 398'000 S$ is in the same range as the reference price of 388'000 S$ (see Section 2.3.4). But it has to be stated that asking prices are higher in general and the effect is small due to the small share of asking price listings within this data set. The HDB observations are around ninety eight square meters large (SIZE) in average and twenty two years old.

Table 16: Descriptive statistics of HDB sale combined data (Model D)

N = 34'873						
Variable	Description	T[1]	Min	Max	Mean	S.D.[3]
Dependent variables						
PRICE	Price [S$]	C	0.166 Mio.	0.950 Mio.	0.398 Mio.	0.106 Mio.
SQMPR	Price per square meter [S$]	C	2'349.32	8'522.73	4'151.25	774.61
Structural explanatory variables						
SIZE	Floor area in square meter	C	31.00	241.00	97.69	25.28
YEAR	Construction year of the building	C	1'967.00	2'010.00	1'989.14	9.62
Locational explanatory variables						
CBD	Distance to CBD [m]	C	608.00	20'134.00	12'428.31	4'562.76
MRT	Distance to an MRT station [m]	C	17.00	3'469.00	722.44	420.92
TOPSEC	Distance to a top secondary school [m]	C	10.00	9'237.00	2'917.85	2'498.83
TOPPRIM	Distance to a top primary school [m]	C	15.00	8'418.00	2'396.57	1'674.68
BUSLINES	Number of bus lines at nearest stop	C	1.00	31.00	4.09	3.43
MSCP	Distance to an MSCP[2] [m]	C	1.00	3'136.00	233.12	213.63
SUPERM	Distance to a supermarket [m]	C	1.00	3'346.00	393.38	221.42
MALL	Distance to a mall [m]	C	1.00	3'440.00	704.09	419.72
SECOND	Distance to a secondary school [m]	C	1.00	3'426.00	460.41	283.30
INDUS	Distance to an industrial estate [m]	C	1.00	3'318.00	817.10	601.03
Data source dependent variables						
ASKING	Is an asking price	D	0.00	1.00	0.08	

[1]Type: C=continuous variable, D=dummy variable; [2] MSCP=multi storey car park; [3]S.D.=standard deviation

4.2.5 Model E: HDB sale transaction

Table 17 shows the final variable selection for the HDB sale transaction market and the descriptive statistics. The mean price of 394'000 S$ is almost identical with the reference price of 388'000 S$ (see Section 2.3.4). 27% of the observations are qualified as model A flats (IS_MODA) and 18% are new generation flats (IS_NGEN). Detailed specifications of HDB flat types can be found at the HDB InfoWEB online portal HDB (2011b).

14% of the listed flats have been subject to a main upgrading program (MUP). MUP are launched by the government in order to enhance HDB estates by upgrading services, facade enhancement, inclusion of space-adding items within the flat, landscaping and other external works (Ong and Ho, 2003). 26 % of the flats have been passing an interim upgrading program (IUP). According to HDB (2011b) IUP focus on the improvement of functional building items such as linkways, wall painting and letter boxes. The IUP budget is up to 2'400 S$ per flat and is fully funded by the government.

Almost half of all observations (48%) have been subject to a lift upgrading program (LUP). The LUP aims to achieve direct lift access for all flats (HDB, 2011b). Finally, around 5% of the listed flats were improved by a home improvement program (HIP). HIP are considered to solve problems related to ageing flats, such as spalling concrete and ceiling leaks. Only flats built before 1986 and which have not undergone an MUP are eligible for HIP (HDB, 2011b). HDB flats are categorized into different flat types such as apartment, maisonette, terrace, standard etc.

Table 17: Descriptive statistics of HDB sale transaction data (Model E)

N = 32'235

Variable	Description	T[1]	Min	Max	Mean	S.D.[3]
Dependent variables						
PRICE	Transaction price [S$]	C	0.166 Mio.	0.900 Mio.	0.394 Mio.	0.103 Mio.
SQMPR	Transaction price per sqm [S$]	C	2'349.00	8'356.00	4'115.34	747.25
Structural explanatory variables						
SIZE	Floor area in square meter	C	31.00	241.00	97.48	25.34
YEAR	Construction year of the building	C	1'967.00	2'010.00	1'989.27	9.78
FLO0105	Floor level 1-5	D	0.00	1.00	0.41	
FLO0610	Floor level 6-10	D	0.00	1.00	0.36	
FLO1115	Floor level 11-15	D	0.00	1.00	0.18	
FLO1620	Floor level 16-20	D	0.00	1.00	0.04	
FLO2125	Floor level 21-25	D	0.00	1.00	0.01	
FLO2630	Floor level 26-30	D	0.00	1.00	0.00	
FLO3135	Floor level 31-35	D	0.00	1.00	0.00	
FLO3640	Floor level 36-40	D	0.00	1.00	0.00	
MUP	Main upgrading program mentioned	D	0.00	1.00	0.14	
IUP	Interim upgrading program mentioned	D	0.00	1.00	0.26	
LUP	Lift upgrading program mentioned	D	0.00	1.00	0.48	
HIP	Home improvement mentioned	D	0.00	1.00	0.05	
IS_MAIS	Flat type: maisonette	D	0.00	1.00	0.03	
IS_AP	Flat type: apartment	D	0.00	1.00	0.04	
IS_NGEN	Flat type: new generation	D	0.00	1.00	0.18	
IS_MODA	Flat type: model A	D	0.00	1.00	0.27	
IS_SIMPL	Flat type: simplified	D	0.00	1.00	0.06	
Locational explanatory variables						
CBD	Distance to CBD [m]	C	608.00	20'134.00	12'458.94	4'525.73
MRT	Distance to an MRT station [m]	C	17.00	3'469.00	725.55	421.52
TOPSEC	Distance to a top secondary school [m]	C	10.00	9'237.00	2'910.87	2'484.57
TOPPRIM	Distance to a top primary school [m]	C	15.00	8'418.00	2'399.91	1'681.46
BUSLINES	Number of bus lines at nearest stop	C	1.00	31.00	4.09	3.43
MSCP	Distance to an MSCP[2] [m]	C	1.00	3'136.00	235.31	214.81
SUPERM	Distance to a supermarket [m]	C	1.00	3'346.00	393.97	221.37
INDUS	Distance to an industrial estate [m]	C	1.00	3'318.00	816.33	597.92
Contractual variables						
Q0210	Transaction in 2nd quarter of 2010	D	0.00	1.00	0.21	
Q0310	Transaction in 3th quarter of 2010	D	0.00	1.00	0.32	
Q0410	Transaction in 4th quarter of 2010	D	0.00	1.00	0.22	
Q0111	Transaction in 1st quarter of 2011	D	0.00	1.00	0.18	
Q0211	Transaction in 2nd quarter of 2011	D	0.00	1.00	0.08	

[1]Type: C=continuous variable, D=dummy variable; [2]MSCP=multi storey car park; [3]S.D.=standard deviation

4.2.6 Model F: HDB rental asking

The descriptive statistics of the final variable selection for the HDB rental asking model is shown in Table 18. As shown later in Section 4.3.6, only a few variables can be used for price modelling of the HDB rental market due to heteroscedasticity and multicollinearity. The average monthly rent of 1'340 S$ (PRICE) and the small average number of bedrooms of only 1.7 (NOBED) indicate that the HDB rental market really represents the social housing sector, as discussed by Phang (2007). Aside from that, the HDB rental observations are quite far away from the CBD with an average distance of more than twelve kilometers (variable CBD).

Table 18: Descriptive statistics of HDB rental asking data (Model F)

N = 6'351

Variable	Description	T[1]	Min	Max	Mean	S.D.[2]
Dependent variables						
PRICE	Asking monthly rent [S$]	C	300.00	4'500.00	1'316.33	887.54
SQMPR	Asking monthly rent per square meter [S$]	C	5.00	356.00	33.28	25.15
Structural explanatory variables						
NOBED	Number of bedrooms	C	1.00	3.00	1.65	0.86
NOBATH	Number of bathrooms	C	1.00	5.00	1.43	0.53
Locational explanatory variables						
CBD	Distance to CBD [m]	C	608.00	20'037.00	12'051.76	4'895.72
FOOD	Distance to a food centre [m]	C	1.00	5'159.00	1'034.21	1'039.93
SECOND	Distance to a secondary school [m]	C	10.00	2'570.00	494.85	322.36
MRT	Distance to a MRT station [m]	C	17.00	2'807.00	686.70	422.90
INDUS	Distance to a industrial estate [m]	C	1.00	3'303.00	820.26	619.10
BUS_26	Nearest bus stop within 200-600 meters	D	0.00	1.00	0.08	

[1]Type: C=continuous variable, D=dummy variable; [2]S.D.=standard deviation

4.3 Estimation results

In this chapter the final coefficient estimations are presented. Every model A to F has been optimized based on the modelling procedure shown in Section 4.1. OLS, SARerr and SARdurbin estimations are shown for all models.

4.3.1 Model A: Private sale combined

Model A is specified as a log-log model with the logarithm of the unit price as dependent variable. It includes 45’792 observations and 23 explanatory variables. Table 19 shows the estimated OLS coefficients and important model diagnostics. The floor area (SIZE) turned out to have an extremely large influence on the price. Since coefficients can be interpreted as sensitivities, an increase of 10% in size would cause a 9.7% increase of the price. The variables reflecting the construction periods (BUI_5160 - BUI_1220) all turned out to have a negative price impact. This was expected since the reference period reflects brand new buildings of 2011. In general older buildings show lower prices except of flats constructed in the Sixties (BUI_6170). It stands out that both availability of a parking (PARKING) and availability of a pool (POOL) turned out to cause negative price impacts. But it must be emphasized that the contribution of these variables is very low in general.

The locational coefficients show that the distance to the CBD is very important. The log-log coefficient shows that a 10% longer distance to the CBD causes a price discount of around 3.2%. It is surprising that longer distances to primary and secondary schools (PRIM and SECOND) as well as to food centres (FOOD) and supermarkets (SUPERM) affect higher prices. On the other hand, longer distances to top schools (TOPPRIM and TOPSEC) cause lower prices in this model.

Spatial autocorrelation is assumed to be present in model A since the Moran’s I test value is very high and highly significant. Lagrange multiplier tests indicate both spatial errors and spatial lags and are significant on the 0.01 level as well. The normal Q-Q plot on the left side of Figure27 in Appendix A.5 compares the quantiles of the empirical error distribution with the quantiles of a normal distribution. It shows that residuals can be assumed as normally distributed. The Tukey-Anscombe plot on the right side plots estimated residuals against fitted values. There is no systematic pattern observable so that the hypothesis of linearity can be accepted.

Table 20 compares OLS coefficients with the estimated SAR coefficients of model A. Details of the SAR estimations are given in Table 34 of Appendix A.6. The spatial weights matrix was generated using eight nearest neighbors since it produced the best results in terms of model fit measured by AIC. Both SAR models considerably increase model fit compared to the OLS model (lower SSE and AIC). Spatial autocorrelation is not longer present since the Moran's I test value becomes much smaller and is not significant anymore. As expected, coefficients of the SARerr model are not very different from the OLS estimates since the weights matrix only affects the error term. On the other hand SARdurbin coefficients are sometimes very different and do even take different signs because spatial weights are incorporated in the explanatory term. SARdurbin coefficients are therefore economically difficult to interpret.

The results show that model fit of SARerr and SARdurbin models are almost identical. This verifies the Lagrange multiplier tests which detect spatial dependence mainly in the error term. It is therefore concluded that the **SARerr estimate is the best specification for model A**.

Table 19: Model A: OLS estimation

Dependent: log(PRICE)							*N* = 45'792
Explanatory variable	Exp.	Coeff.	SE	Scaled	T stat.	Sign.	VIF
Constant		11.098					
Structural variables							
log(SIZE)	+	0.970	0.002	0.712	412.977	***	1.239
BUI_5160	-	-0.213	0.075	-0.004	-2.857	***	1.005
BUI_6170	-	-0.101	0.019	-0.008	-5.257	***	1.059
BUI_7180	-	-0.214	0.009	-0.042	-23.740	***	1.283
BUI_8190	-	-0.190	0.007	-0.055	-28.204	***	1.594
BUI_9100	-	-0.188	0.005	-0.119	-39.476	***	3.774
BUI_0110	-	-0.059	0.004	-0.046	-14.792	***	4.082
BUI_1220	-	-0.014	0.004	-0.009	-3.242	***	2.965
CONDO	+	0.135	0.002	0.096	54.528	***	1.296
PARKING	+	-0.077	0.003	-0.061	-28.627	***	1.882
WELL	+	0.025	0.002	0.020	11.169	***	1.342
SEC	+	0.025	0.003	0.020	9.164	***	1.991
POOL	+	-0.022	0.004	-0.011	-5.370	***	1.636
GARD	+	0.006	0.003	0.004	2.175	**	1.479
Locational variables							
log(CBD)	-	-0.316	0.002	-0.414	-170.566	***	2.450
log(INDUS)	+	0.078	0.001	0.101	60.562	***	1.157
log(MSCP)	+	0.098	0.002	0.120	55.583	***	1.928
log(BUSLINES)	+	0.006	0.000	0.052	30.162	***	1.256
log(PRIM)	-	0.047	0.002	0.054	26.951	***	1.657
log(MALL)	-	-0.029	0.001	-0.046	-22.160	***	1.805
log(TOPPRIM)	-	-0.027	0.002	-0.033	-16.749	***	1.578
log(FOOD)	-	0.015	0.001	0.022	11.737	***	1.449
BUS_26	+	0.028	0.002	0.019	11.522	***	1.126
log(MRT)	-	-0.017	0.002	-0.021	-10.280	***	1.703
log(SECOND)	-	0.007	0.002	0.008	4.062	***	1.758
log(SUPERM)	-	0.006	0.002	0.007	3.524	***	1.722
log(TOPSEC)	-	-0.005	0.001	-0.006	-3.302	***	1.466
Contractual and data source dependent variables							
ASKING	+	0.137	0.003	0.096	50.211	***	1.534
FREE	+	0.113	0.002	0.089	49.078	***	1.366
Adjusted R-square				0.890			
Sum of squared errors (SSE)				2'027.1			
Akaike Information Criterion (AIC)				-12'743.3			
Moran's I				0.739	***		
Robust LMerr				144'351.1	***		
Robust LMlag				1'682.2	***		

Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 20: Model A: Comparison of OLS and SAR coefficients

				N = 45'792
Dependent: log(PRICE)		**OLS**	**SARerr**	**SARdurbin**
Explanatory variable	Exp.	Coefficient	Coefficient	Coefficient
Constant		11.098	11.822	0.987
Lambda/Rho			0.92	0.904
Structural variables				
log(SIZE)	+	0.970	0.894	0.893
BUI_5160	-	-0.213	-0.233	-0.222
BUI_6170	-	-0.101	**-0.007**	**0.003**
BUI_7180	-	-0.214	-0.136	-0.126
BUI_8190	-	-0.190	-0.092	-0.078
BUI_9100	-	-0.188	-0.107	-0.096
BUI_0110	-	-0.059	-0.035	-0.028
BUI_1220	-	-0.014	0.028	0.038
CONDO	+	0.135	0.022	0.018
PARKING	+	-0.077	-0.040	-0.031
WELL	+	0.025	0.056	0.046
SEC	+	0.025	**0.006**	**0.006**
POOL	+	-0.022	**0.007**	0.006
GARD	+	0.006	**-0.007**	-0.016
Locational variables				
log(CBD)	-	-0.316	-0.334	0.340
log(INDUS)	+	0.078	0.037	-0.016
log(MSCP)	+	0.098	0.100	0.067
log(BUSLINES)	+	0.006	0.001	-0.002
log(PRIM)	-	0.047	0.084	0.064
log(MALL)	-	-0.029	-0.017	-0.031
log(TOPPRIM)	-	-0.027	-0.035	-0.063
log(FOOD)	-	0.015	0.061	0.073
BUS_26	+	0.028	0.029	0.008
log(MRT)	-	-0.017	-0.023	-0.031
log(SECOND)	-	0.007	-0.036	-0.053
log(SUPERM)	-	0.006	-0.019	-0.006
log(TOPSEC)	-	-0.005	**0.002**	**0.001**
Contractual and data source dependent variables				
ASKING	+	0.137	0.090	0.093
FREE	+	0.113	0.129	0.112
SSE		2'027.1	492.6	490.23
AIC		-12'743.3	-73'014.6	-73'798.6
Moran's I		0.739	-0.073	-0.069
Bold: Not significant at 0.1 level				

4.3.2 Model B: Private sale transaction

As well as model A, model B is specified as a log-log model with the logarithm of the unit price as dependent variable. It includes 12'467 observations and 21 explanatory variables. Table 21 shows the estimated OLS coefficients and important model diagnostics. As already seen in model A, the floor area is the most dominant price determinant as well. Since Model B observations do not include unfinished buildings, the construction year (YEAR) is included as continuous variable and turned out to have a positive impact: newer flats are assumed to yield higher prices. The estimated parameter of 15.683 can be interpreted as follows: a flat from 2010 yields an around 15% higher price than a flat from 1990 since the construction year is around 1% higher and parameters can be interpreted as sensitivities. Coefficients of the dummy variables FLO0610 to FLO41UP show the surcharge of different floor levels compared to the lowest level (levels 1-5). Flats between 16th and 25th floor (FLO1620 and FLO2125) are considered to yield the highest prices.

It is estimated that a 10% longer distance to the CBD causes a price discount of around 2.8%. As in model A longer distances to primary and secondary schools (PRIM and SECOND) have a positive price impact while longer distances to top schools (TOPPRIM and TOPSEC) cause lower prices. The estimated coefficient of the variable BUSLINES can be interpreted as follows: 200% more bus lines at the nearest stop (for example 6 instead of 2) cause a 2.2% ($0.011 \cdot 200$) higher price. It is surprising that the distance to the nearest MRT station turns out to be insignificant.

Contractual variables are showing the expected price impacts: freehold contracts (FREE) yield higher prices and flats bought by buyers who have lived in a HDB flat before (FROMPUB) are considered to be cheaper than others. Finally the estimation shows that the transaction date is relevant for the price: recently transacted flats are more expensive than flats sold earlier.

OLS model diagnostics clearly show, that spatial autocorrelation is present in model B since the Moran's I test value is very high and highly significant. Lagrange multiplier tests indicate particularly spatial errors but both the LMerr and the LMlag test are significant on the 0.01 level. The normal Q-Q plot on the left side of Figure 27 in Appendix A.5 shows that the residuals don't perfectly follow a normal distribution. But the error distribution does not show a clear structure making it possible to improve it by variable transformation. Since semi-log estimation showed similar problems and lower model fit, the model specification is anyhow tolerated. The Tukey-Anscombe plot on the right side shows no systematic pattern so that the assumption on linearity is acceptable.

Table 22 compares OLS coefficients with the additionally estimated SAR coefficients of model B. Details of the SAR estimations are given in Table 35 of Appendix A.6. Eight nearest neighbors were used for generating the SAR spatial weights matrix since it produced the best results in terms of model fit measured by AIC. OLS and SAR estimates show a very similar structure in terms of relative impacts and signs. But SAR models turn out with a much better model fit measured by SSE and AIC.

It stands out that the estimated coefficients for the distance to the nearest MRT station were not significant in all three estimations. And the distance to the nearest secondary school (SECOND) is not significant in both SAR estimations. The number of bus lines (BUSLINES) is insignificant in the SARerr model and the distance to the nearest top secondary school (TOPSEC) is not significant in the SARdurbin model. Since model fit of the SAR models is nearly identical, the **SARerr estimate is considered the best specification for model B** due to its better economic interpretability.

Table 21: Model B: OLS estimation

Dependent: log(PRICE)							N = 12'467
Explanatory variable	Exp.	Coeff.	SE	Scaled	T stat.	Sign.	VIF
Constant		-107.657					
Structural variables							
log(SIZE)	+	0.934	0.005	0.730	198.612	***	1.236
CONDO	+	0.123	0.004	0.115	30.063	***	1.346
log(YEAR)	+	15.683	0.539	0.115	29.112	***	1.435
PARKING	+	-0.068	0.004	-0.064	-16.463	***	1.376
FLO0610	+	0.048	0.004	0.039	10.735	***	1.212
FLO1115	+	0.074	0.006	0.048	13.129	***	1.200
FLO1620	+	0.116	0.007	0.057	16.137	***	1.139
FLO2125	+	0.165	0.010	0.054	15.739	***	1.076
FLO2630	+	0.164	0.014	0.041	12.113	***	1.072
FLO3135	+	0.266	0.022	0.041	12.249	***	1.041
FLO3640	+	0.225	0.031	0.025	7.327	***	1.037
FLO41UP	+	0.167	0.028	0.021	5.978	***	1.115
WELL	+	0.028	0.004	0.026	6.970	***	1.292
POOL	+	-0.032	0.007	-0.020	-4.770	***	1.647
Locational variables							
log(CBD)	-	-0.288	0.003	-0.380	-82.606	***	1.940
log(INDUS)	+	0.047	0.002	0.079	22.592	***	1.109
log(MSCP)	+	0.054	0.003	0.083	19.954	***	1.576
log(TOPPRIM)	-	-0.041	0.003	-0.060	-14.567	***	1.548
log(PRIM)	-	0.028	0.003	0.035	8.496	***	1.511
log(TOPSEC)	-	-0.020	0.002	-0.031	-8.027	***	1.342
log(BUSLINES)	+	0.011	0.002	0.020	5.301	***	1.260
log(SECOND)	-	0.025	0.003	0.032	7.535	***	1.640
BUS_26	+	0.030	0.004	0.024	6.891	***	1.092
log(MALL)	-	-0.012	0.002	-0.021	-5.544	***	1.318
log(MRT)	-	0.001	0.003	0.002	0.488		1.604
Contractual variables							
FREE	+	0.133	0.004	0.127	33.606	***	1.315
FROMPUB	-	-0.040	0.004	-0.036	-10.364	***	1.128
Q0310	-	-0.081	0.008	-0.075	-9.662	***	5.466
Q0410	-	-0.043	0.008	-0.039	-5.130	***	5.293
Q0111	-	-0.013	0.009	-0.011	-1.519		4.845
Adjusted R-square				0.864			
Sum of squared errors (SSE)				460.6			
Akaike Information Criterion (AIC)				-5675.8			
Moran's I				0.703	***		
Robust LMerr				29283.7	***		
Robust LMlag				446.7	***		
Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$							

Table 22: Model B: Comparison of OLS and SAR coefficients

				N = 12'467
Dependent: log(PRICE)		**OLS**	**SARerr**	**SARdurbin**
Explanatory variable	Exp.	Coefficient	Coefficient	Coefficient
Constant		-107.657	-67.074	-20.686
Lambda / Rho			0.883	0.864
Structural variables				
log(SIZE)	+	0.934	0.862	0.858
CONDO	+	0.123	0.066	0.055
log(YEAR)	+	15.683	10.414	9.933
PARKING	+	-0.068	-0.070	-0.073
FLO0610	+	0.116	0.065	0.066
FLO1115	+	0.165	0.104	0.103
FLO1620	+	0.074	0.047	0.046
FLO2125	+	0.164	0.131	0.124
FLO2630	+	0.266	0.139	0.139
FLO3135	+	0.048	0.029	0.030
FLO3640	+	0.028	0.020	0.018
FLO41UP	+	0.225	0.113	0.121
WELL	+	0.167	0.190	0.203
POOL	+	-0.032	-0.028	-0.027
Locational variables				
log(CBD)	-	-0.288	-0.255	0.340
log(INDUS)	+	0.047	0.033	0.018
log(MSCP)	+	0.054	0.043	0.044
log(TOPPRIM)	-	-0.041	-0.047	0.004
log(PRIM)	-	0.028	0.039	0.030
log(TOPSEC)	-	-0.020	-0.032	**-0.017**
log(BUSLINES)	+	0.011	**-0.003**	-0.012
log(SECOND)	-	0.025	**0.010**	**-0.013**
BUS_26	+	0.030	0.041	0.038
log(MALL)	-	-0.012	-0.028	-0.032
log(MRT)	-	**0.001**	**-0.001**	**-0.014**
Contractual and data source dependent variables				
FREE	+	0.133	0.152	0.143
FROMPUB	-	-0.081	-0.078	-0.078
Q0310	-	-0.043	-0.051	-0.050
Q0410	-	-0.040	-0.007	-0.009
Q0111	-	**-0.013**	-0.015	-0.016
SSE		460.6	127.3	125.9
AIC		-5'675.8	-19'517.1	-19'760.3
Moran's I		0.703	-0.014	-0.012

Bold: Not significant at 0.1 level

4.3.3 Model C: Private rental asking

Model C is specified as a log-log model with the logarithm of the monthly rent as dependent variable. It includes 22'011 observations and 16 explanatory variables. Observations with a monthly rent of below 1'800 S$ (433 observations, 1.9%) were excluded because they are considered to belong to a kind of sub-market. As Figure 13 shows, this small segment represents small and low-cost flats and is clearly isolated from the rest of the observations. The exclusion was executed since first OLS estimations pointed to serious problems with heteroscedasticity.

Figure 13: Model C: Rent plotted against floor area and rent per square meter

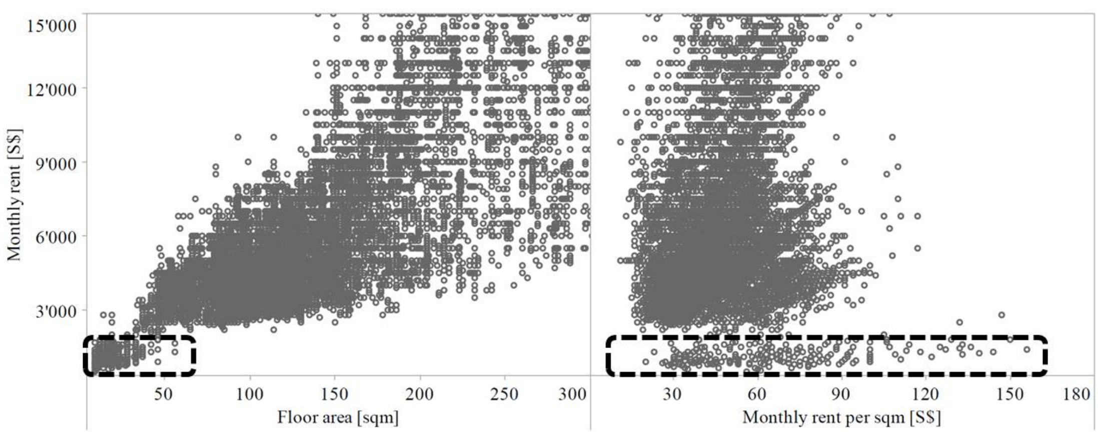


Table 23 shows the estimated OLS coefficients and important model diagnostics. The coefficients of the scaled explanatory variables show that floor area (SIZE) causes the strongest price impact. The construction year (YEAR) also turned out to have a considerable influence. It can be interpreted as follows: a flat from 2000 yields an around 21% higher price than a flat from 1980 since the construction year is around 1% higher and parameters can be interpreted as sensitivities. It is surprising that the availability of a garden (GARD) causes a significant negative price impact. But it must be emphasized that the impact of GARD is very small (around a hundred times smaller than the impact of the floor area).

Spatial autocorrelation is assumed to be present in model C since the Moran's I test value is very high and highly significant. Lagrange multiplier tests indicate spatial errors and spatial lags and both the LMerr and the LMlag test are significant on the 0.01 level. The normal Q-Q plot on the left side of Figure 27 in Appendix A.5 shows that the residuals follow pretty good the normal distribution. The Tukey-Anscombe plot on the right side of the same figure shows no pattern so that the hypothesis of linearity is accepted.

Table 23: Model C: OLS estimation

Dependent: log(PRICE)							N = 22'011
Explanatory variable	Exp.	Coeff.	SE	Scaled	T stat.	Sign.	VIF
Constant		-170.984					
Structural variables							
log(SIZE)	+	0.825	0.003	0.788	267.771	***	1.294
log(YEAR)	+	23.242	0.347	0.200	66.976	***	1.336
CONDO	+	0.073	0.003	0.068	24.326	***	1.159
GARD	+	-0.020	0.003	-0.019	-6.842	***	1.176
WELL	+	0.007	0.003	0.007	2.502	**	1.337
Locational variables							
log(CBD)	-	-0.212	0.002	-0.381	-106.651	***	1.903
log(INDUS)	+	0.064	0.002	0.103	36.542	***	1.194
log(BUS)	-	0.046	0.002	0.086	28.709	***	1.344
log(MSCP)	+	0.067	0.002	0.103	28.461	***	1.968
log(BUSLINES)	+	0.040	0.002	0.076	25.420	***	1.322
log(MRT)	-	-0.031	0.002	-0.052	-15.292	***	1.719
log(SUPERM)	-	-0.027	0.002	-0.042	-13.056	***	1.511
log(SECOND)	-	-0.025	0.002	-0.036	-11.235	***	1.501
log(FOOD)	-	-0.019	0.002	-0.033	-10.871	***	1.379
log(PRIM)	-	0.024	0.002	0.033	10.739	***	1.441
Contractual variables							
FREE	+	0.032	0.003	0.032	10.974	***	1.306
Adjusted R-square				0.853			
Sum of squared errors (SSE)				773.6			
Akaike Information Criterion (AIC)				-11'198.2			
Moran's I				0.615	***		
Robust LMerr				44'314.3	***		
Robust LMlag				839.7	***		
Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$							

Table 24 compares OLS coefficients with the additionally estimated SAR coefficients of model C. Details of the SAR estimations are given in Table 36 of Appendix A.6. Eight nearest neighbors were used for generating the SAR spatial weights matrix since it produced the best results in terms of model fit measured by AIC. OLS and SAR estimates show a very similar structure in terms of relative impacts and signs. But SAR models turn out with a much better model fit measured by SSE and AIC. Since model fit of the SAR models is very similar, the **SARerr model is considered the best specification for model C** due to its better economic interpretability.

Table 24: Model C: Comparison of OLS and SAR coefficients

				N = 22'011
Dependent: log(PRICE)		**OLS**	**SARerr**	**SARdurbin**
Explanatory variable	Exp.	Coefficient	Coefficient	Coefficient
Constant		-170.984	-164.750	-28.867
Lambda / Rho			0.869	0.845
Structural variables				
log(SIZE)	+	0.825	0.741	0.738
log(YEAR)	+	23.242	22.509	22.258
CONDO	+	0.073	0.022	0.012
GARD	+	-0.020	**0.000**	0.019
WELL	+	0.007	0.028	0.025
Locational variables				
log(CBD)	-	-0.212	-0.205	0.023
log(INDUS)	+	0.064	0.032	-0.008
log(BUS)	-	0.046	0.022	-0.030
log(MSCP)	+	0.067	0.043	0.023
log(BUSLINES)	+	0.040	**0.002**	-0.013
log(MRT)	-	-0.031	**-0.010**	-0.021
log(SUPERM)	-	-0.027	-0.017	-0.004
log(SECOND)	-	-0.025	-0.024	-0.039
log(FOOD)	-	-0.019	-0.015	-0.004
log(PRIM)	-	0.024	0.023	0.050
Contractual variables				
FREE	+	0.032	0.041	0.035
SSE		773.6	308.4	304.2
AIC		-11'198.2	-29'404.8	-29'858.3
Moran's I		0.615	-0.043	-0.036

Bold: Not significant at 0.1 level

4.3.4 Model D: HDB sale combined

Model D is specified as a log-log model with the logarithm of the unit price as dependent variable. It includes 34'873 observations and 13 explanatory variables. Table 23 shows the estimated OLS coefficients and important model diagnostics. As expected the floor area (SIZE) and the construction year (YEAR) have a positive effect on the price. Longer distances to the CBD, MRT stations and top schools (TOPSEC and TOPPRIM) cause lower prices and a higher number of bus lines at the nearest stop (BUSLINES) has a positive price impact. It is surprising that longer distances to shopping malls (MALL) cause higher housing prices. It must be mentioned, however, that the relative impact is very small.

OLS model diagnostics clearly show, that spatial autocorrelation is present in model D since the Moran's I test value is high and significant. Lagrange multiplier tests indicate predominantly spatial errors and only secondary spatial lags since the test value of the LMerr is much higher than the LMlag value. But both test are significant on the 0.01 level. The normal Q-Q plot on the left side of Figure 28 in Appendix A.5 shows that the residuals can be assumed to be normally distributed. The Tukey-Anscombe plot on the right side of the same figure shows no pattern so that the hypothesis of linearity is accepted.

Table 26 compares OLS coefficients with the additionally estimated SAR coefficients of model C. Details of the SAR estimations are given in Table 33 of Appendix A.6. Fifteen nearest neighbors were used for generating the SAR spatial weights matrix since it produced the best results in terms of model fit measured by AIC. OLS and SAR estimates show a very similar structure in terms of relative impacts and signs. But SAR models turn out with a much better model fit measured by SSE and AIC. Since model fit of the SAR models is very similar and autocorrelation appeared mainly in the error term, the **SARerr estimation is considered the best specification for model D**.

Table 25: Model D: OLS estimation

Dependent: log(PRICE)							N = 34'873
Explanatory variable	Exp.	Coeff.	SE	Scaled	T stat.	Sign.	VIF
Constant		-61.705					
Structural variables							
log(SIZE)	+	0.789	0.003	0.830	295.115	***	1.534
log(YEAR)	+	9.681	0.170	0.184	56.796	***	2.029
Locational variables							
log(CBD)	-	-0.212	0.002	-0.402	-127.380	***	1.936
log(MRT)	-	-0.040	0.001	-0.105	-42.256	***	1.200
log(TOPSEC)	-	-0.020	0.001	-0.072	-26.625	***	1.438
log(TOPPRIM)	-	-0.020	0.001	-0.066	-26.409	***	1.222
log(BUSLINES)	+	0.013	0.001	0.042	17.801	***	1.098
log(MSCP)	+	-0.007	0.000	-0.035	-13.748	***	1.223
log(SUPERM)	-	-0.009	0.001	-0.025	-10.219	***	1.118
log(MALL)	-	0.005	0.001	0.015	6.055	***	1.218
log(SECOND)	-	0.005	0.001	0.013	5.335	***	1.155
log(INDUS)	+	-0.003	0.001	-0.011	-4.206	***	1.306
Data source dependent variables							
ASKING	+	0.101	0.002	0.105	45.942	***	1.014
Adjusted R-square				0.820			
Sum of squared errors (SSE)				407.6			
Akaike Information Criterion (AIC)				-56'158.9			
Moran's I				0.623	***		
Robust LMerr				101'695.6	***		
Robust LMlag				89.2	***		

Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 26: Model D: Comparison of OLS and SAR coefficients

				N = 34'873
Dependent: log(PRICE)		**OLS**	**SARerr**	**SARdurbin**
Explanatory variable	Exp.	Coefficient	Coefficient	Coefficient
Constant		-61.705	-62.333	-10.433
Lamda / Rho			0.839	0.834
Structural variables				
log(SIZE)	+	0.789	0.813	0.815
log(YEAR)	+	9.681	9.756	9.850
Locational variables				
log(CBD)	-	-0.212	-0.227	-0.053
log(MRT)	-	-0.040	-0.027	0.000
log(TOPSEC)	-	-0.020	-0.019	-0.011
log(TOPPRIM)	-	-0.020	-0.023	-0.029
log(BUSLINES)	+	0.013	**0.001**	-0.002
log(MSCP)	+	-0.007	-0.003	-0.003
log(SUPERM)	-	-0.009	-0.005	-0.005
log(MALL)	-	0.005	**-0.002**	-0.005
log(SECOND)	-	0.005	**0.001**	**-0.003**
log(INDUS)	+	-0.003	0.006	0.015
Data source dependent variables				
ASKING	+	0.101	0.093	0.093
SSE		407.6	175.3	175.1
AIC		-56'158.9	-80'461.4	-80'555.1
Moran's I		0.623	0.020	0.021

Bold: Not significant at 0.1 level

4.3.5 Model E: HDB sale transaction

Model E is specified as a log-log model with the logarithm of the unit price as dependent variable. It includes 32'235 observations and 21 explanatory variables. Table 25 shows the estimated OLS coefficients and important model diagnostics. As expected the floor area (SIZE) and the construction year (YEAR) have a positive effect on the price. Dummy variables for the floor level (FLOOR0610 to FLOOR3640) show that there is a positive price impact for all floor levels compared to flats between the first and the fifth storey. Based on the parameter estimate, flats between the 11th and 25th yield the highest premium. The results further show that HDB upgrading programs cause higher housing prices. Main upgrading programs (MUP) and interim upgrading programs (IUP) are considered to have a stronger positive impact than lift upgrading programs (LUP) and home improvement programs (HIP). Longer distances to the CBD, MRT stations and top schools (TOPSEC and TOPPRIM) cause lower prices and a higher number of bus lines at the nearest stop (BUSLINES) has a positive price impact. Further the estimation shows that the transaction date is very relevant for the price: recently transacted flats are more expensive than flats sold earlier (Q0210 to Q0111).

The fact that the Moran's I test value is very high and highly significant indicates that spatial autocorrelation is present. Lagrange multiplier tests indicate spatial errors but no spatial lags since the robust LMlag test is not significant on the 0.01 level. The normal Q-Q plot on the left side of Figure 28 in Appendix A.5 that the residuals follow pretty good the normal distribution. The Tukey-Anscombe plot on the right side of the same figure shows no pattern so that the hypothesis of linearity is accepted.

A comparison of OLS and SAR coefficients is shown in Table 28. Details of the SAR estimations are given in Table 37 of Appendix A.6. Ten nearest neighbors were used for generating the SAR spatial weights matrix since it produced the best results in terms of model fit measured by AIC. OLS and SAR estimates show a very similar structure in terms of relative impacts and signs. SAR models however turn out with a much better model fit measured by SSE and AIC. Since model fit of the SAR models is very similar, the **SARerr estimation is considered the best specification for model E** due to its better economical interpretability.

Table 27: Model E: OLS estimation

Dependent: log(PRICE)							N = 32'235
Explonatory variable	Exp.	Coeff.	SE	Scaled	T stat.	Sign.	VIF
Constant		-63.043					
Structural variables							
log(SIZE)	+	0.760	0.003	0.809	257.954	***	2.423
log(YEAR)	+	13.789	0.240	0.269	57.369	***	5.393
FLOOR0610	+	0.044	0.001	0.084	38.015	***	1.213
FLOOR1115	+	0.067	0.001	0.102	46.199	***	1.211
FLOOR1620	+	0.128	0.003	0.096	45.221	***	1.107
FLOOR2125	+	0.205	0.005	0.093	45.217	***	1.052
FLOOR2630	+	0.232	0.009	0.056	27.267	***	1.038
FLOOR3135	+	0.320	0.031	0.021	10.473	***	1.005
FLOOR3640	+	0.357	0.028	0.026	12.909	***	1.007
MUP	+	0.069	0.003	0.094	25.281	***	3.409
IUP	+	0.033	0.002	0.057	20.534	***	1.900
LUP	+	0.009	0.002	0.017	5.047	***	2.786
HIP	+	0.011	0.002	0.010	4.506	***	1.127
IS_MAIS	+	0.057	0.003	0.039	17.218	***	1.237
IS_AP	+	0.046	0.003	0.037	16.666	***	1.226
IS_NEWGEN	+	0.015	0.002	0.023	8.228	***	1.902
IS_MODA	-	-0.006	0.001	-0.010	-4.345	***	1.430
IS_SIMPL	-	0.006	0.003	0.006	2.357	**	1.391
Locational variables							
log(CBD)	-	-0.194	0.001	-0.366	-133.110	***	1.865
log(MRT)	-	-0.035	0.001	-0.094	-42.555	***	1.195
log(TOPSEC)	-	-0.020	0.001	-0.073	-30.609	***	1.416
log(TOPPRIM)	-	-0.019	0.001	-0.064	-28.640	***	1.229
log(BUSLINES)	+	0.012	0.001	0.041	19.232	***	1.109
log(MSCP)	+	-0.005	0.000	-0.027	-10.955	***	1.448
log(SUPERM)	-	-0.004	0.001	-0.010	-5.005	***	1.039
log(INDUS)	+	-0.001	0.001	-0.005	-2.063	**	1.316
Contractual variables							
Q0210	-	-0.106	0.002	-0.170	-50.913	***	2.743
Q0310	-	-0.069	0.002	-0.127	-34.992	***	3.252
Q0410	-	-0.037	0.002	-0.060	-17.667	***	2.803
Q0111	-	-0.017	0.002	-0.026	-8.041	***	2.546
Adjusted R-square				0.869			
Sum of squared errors (SSE)				268.7			
Akaike Information Criterion (AIC)				-62769.5			
Moran's I				0.679	***		
Robust LMerr				84989.3	***		
Robust LMlag				1.2			

Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 28: Model E: Comparison of OLS and SAR coefficients

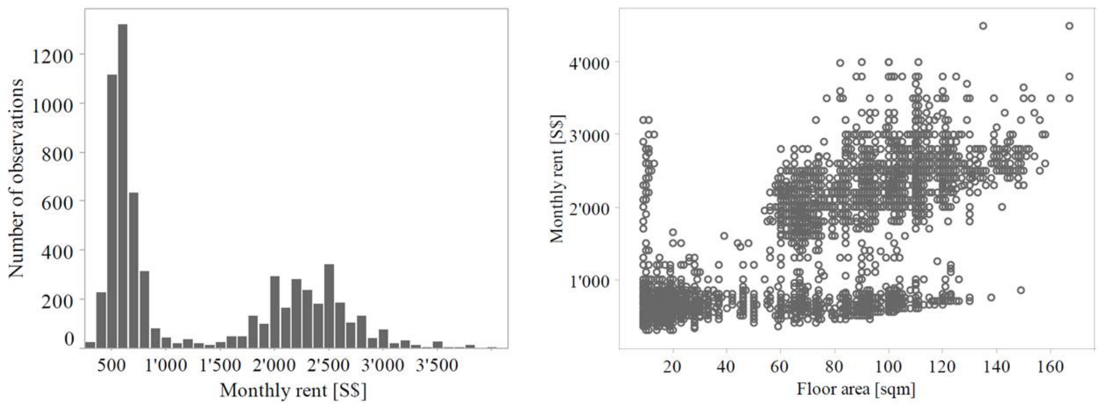
				N = 32’235
Dependent: log(PRICE)		**OLS**	**SARerr**	**SARdurbin**
Explanatory variable	Exp.	Coefficient	Coefficient	Coefficient
Constant		-92.979	-85.111	-14.026
Lamda / Rho			0.851	0.835
Structural variables				
log(SIZE)	+	0.760	0.760	0.759
log(YEAR)	+	13.789	12.773	13.196
FLOOR0610	+	0.044	0.045	0.045
FLOOR1115	+	0.067	0.065	0.065
FLOOR1620	+	0.128	0.098	0.098
FLOOR2125	+	0.205	0.142	0.140
FLOOR2630	+	0.232	0.153	0.152
FLOOR3135	+	0.320	0.178	0.173
FLOOR3640	+	0.357	0.204	0.198
MUP	+	0.069	0.016	0.009
IUP	+	0.033	0.018	0.015
LUP	+	0.009	-0.017	-0.018
HIP	+	0.011	0.006	0.005
IS_MAIS	+	0.057	0.049	0.049
IS_AP	+	0.046	0.047	0.047
IS_NEWGEN	+	0.015	-0.013	-0.017
IS_MODA	-	-0.006	-0.006	-0.007
IS_SIMPL	-	0.006	0.012	0.013
Locational variables				
log(CBD)	-	-0.194	-0.214	-0.032
log(MRT)	-	-0.035	-0.026	-0.014
log(TOPSEC)	-	-0.020	-0.021	-0.026
log(TOPPRIM)	-	-0.019	-0.024	**-0.016**
log(BUSLINES)	+	0.012	0.002	-0.003
log(MSCP)	+	-0.005	**0.000**	**0.000**
log(SUPERM)	-	-0.004	**-0.002**	**-0.001**
log(INDUS)	+	-0.001	0.004	0.014
Contractual				
Q0210	-	-0.106	-0.107	-0.107
Q0310	-	-0.069	-0.069	-0.069
Q0410	-	-0.037	-0.035	-0.036
Q0111	-	-0.017	-0.018	-0.018
SSE		268.7	85.6	85.4
AIC		-62’769.5	-92’702.0	-93’065.5
Moran’s I		0.679	0.022	0.022

Bold: Not significant at 0.1 level

4.3.6 Model F: HDB rental asking

In contrast to all other models, Model F is specified as a semi-log model with the logarithm of the monthly rent as dependent variable. It includes 6'351 observations and only 8 explanatory variables since the majority of the available variables was not significant. Figure 28 in Appendix A.5 clearly shows serious problems with heteroscedasticity. These problems are caused by a sub-market segmentation, as shown in Figure 14. It is assumed that these two segments represent the **room rental** and the **apartment rental market**. But the segments cannot clearly be subdivided since a price-size-overlap is assumed. The scatter plot at the right side shows that prices of the two segments are differently related to the floor area. These are characterized by different gradients. Further analysis did not point to a spatial segmentation, as shown in Appendix A.8.

Figure 14: Model F: Sub-market segments



Estimated OLS coefficients and important model diagnostics are given in Appendix A.7. Number of bedrooms (NOBED) was included since the segmented data led to poor estimates for the floor area. The reason is probably that in the room rental market the floor area of the entire flat is announced but the price is related to a single bedroom. Locational variables turned out to have a very small impact in this market, coefficients are below/above zero at the forth and fifth decimal point. It is surprising that the variable BUS_26 has a negative impact since it was expected to be positive. The Moran's I test value indicates little spatial autocorrelation for model F. Lagrange multiplier tests point to spatial errors but no spatial lags since the test value of LMlag is very small and not significant at the 0.1 level.

Table 29 compares OLS coefficients with the additionally estimated SAR coefficients of model F. Details of the SAR estimations are given in Table 38 of Appendix A.6. Eight nearest neighbors were used for generating the SAR spatial weights matrix since it produced the best results in terms of model fit measured by AIC. OLS and SAR estimates show a very similar structure in terms of relative impacts and signs. But SAR models turn out with better model fit measured by SSE and AIC. Since LMlag tests were not significant and SAR models show similar results, the **SARerr estimation is considered the best specification for model D**.

Table 29: Model F: Comparison of OLS and SAR coefficients

				N = 6'351
Dependent: log(PRICE)		**OLS**	**SARerr**	**SARdurbin**
Explanatory variable	Exp.	Coefficient	Coefficient	Coefficient
Constant		6.004	6.007	3.397
Lamda / Rho			0.438	0.436
Structural variables				
NOBED	+	0.710	0.711	0.712
NOBATH	+	0.033	0.034	0.035
Locational variables				
CBD	-	0.000	0.000	0.000
FOOD	-	0.000	0.000	0.000
SECOND	-	0.000	0.000	0.000
MRT	-	0.000	0.000	**0.000**
INDUS	+	0.000	0.000	0.000
BUS_26	+	-0.031	-0.031	-0.030
SSE		310.2	269.3	268.5
AIC		-1'130.6	-1'606.2	-1'610.5
Moran's I		0.204	0.039	0.039

Bold: Not significant at 0.1 level

4.4 Spatially varying coefficients

OLS models estimations showed that the spatial autocorrelation is considerable. SAR models solve this problem by introducing spatial weights matrices into the equations. Since autocorrelation mainly appeared in the error term, it is assumed that either spatial dependence is present in the "unobserved" variables or housing preferences vary across space. To explore spatial variation of hedonic coefficients, geographically weighted regressions were applied to the variable selection of model B (private sale transaction). An AIC optimization method was used to determine the bandwidth. This bandwidth is used to construct a separate equation for every feature in the data set incorporating the dependent and explanatory variables of features falling within it (Bivand, 2011b). As proposed by Fotheringham *et al.* (2002) a Gaussian scheme was used as geographical weighting function. The process returned an optimized bandwidth of around 980 meters.

Table 30 compares estimated GWR and OLS coefficients. Since 12'467 individual coefficient vectors for each data point have been computed, the table shows mean, standard deviation (S.D.) and median of the estimates. The results show that average GWR coefficients for structural variables do not vary significantly. But high standard deviations point to considerable variation across the observations. The GWR model performed better than the OLS estimation measured by adjusted R-square.

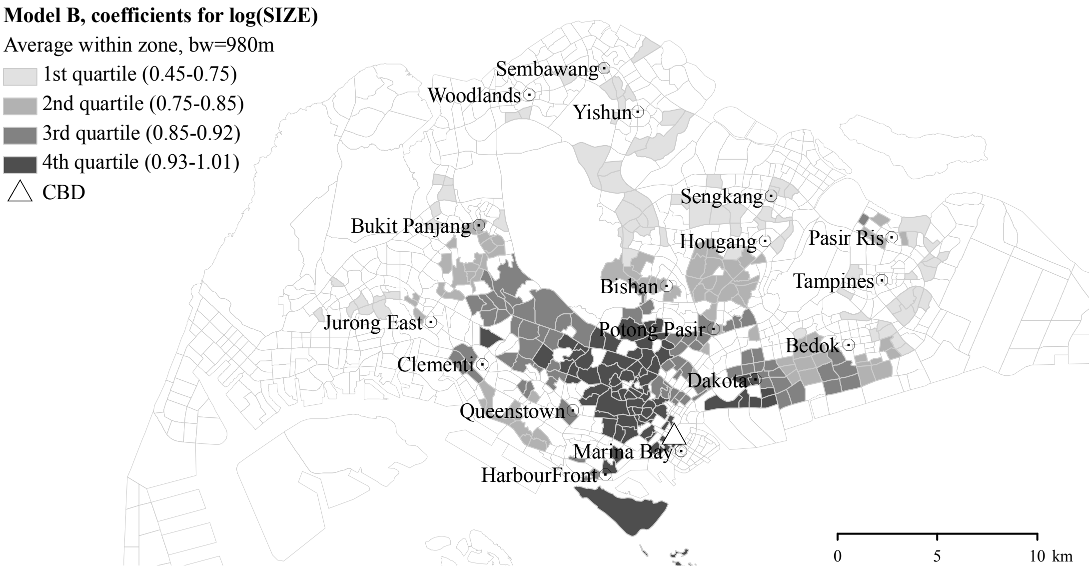
Figure 15 shows the spatial pattern of the two most important structural variables floor area (SIZE) and constructions year (YEAR). Estimates are aggregated to the planning zone level and classified into four classes representing quartiles. The floor area turned out with a strong positive impact in all cases and has a small scattering (0.45 - 1.01). The spatial distribution indicates a decrease in relative importance of floor area towards the CBD. In the dark colored areas a doubling of a flats size is expected to cause a doubled price (sensitivity $\cong 1.0$). In contrast, the preference variation concerning the construction year does not follow such a clear spatial pattern. Zones appearing in the 4th quartile of the parameter range are however spatially clustered. A strong preference for young flats (high coefficient values for the construction year) is therefore assumed close to the CBD, in Bukit Panjang, at the northeastern border of the big nature reserve as well as around Tampines.

Table 30: Model B: OLS estimation

Dependent: log(PRICE)					N=12'467
OLS model			**GWR model**		
Explanatory variable	Exp.	Estimate	Coeff. mean	Coeff. S.D.	Coeff. median
Constant		-107.657	-89.866	68.411	-85.993
Structural variables					
log(SIZE)	+	0.934	0.838	0.120	0.849
CONDO	+	0.123	0.077	0.114	0.080
log(YEAR)	+	15.683	13.193	8.668	13.118
PARKING	+	-0.068	-0.069	0.057	-0.063
FLO0610	+	0.048	0.037	0.026	0.034
FLO1115	+	0.074	0.060	0.042	0.062
FLO1620	+	0.116	0.086	0.072	0.089
FLO2125	+	0.165	0.177	0.695	0.140
FLO2630	+	0.164	0.212	0.399	0.165
FLO3135	+	0.266	0.289	0.489	0.257
FLO3640	+	0.225	0.153	0.900	0.228
FLO41UP	+	0.167	-0.030	4.231	0.311
WELL	+	0.028	0.034	0.045	0.032
POOL	+	-0.032	-0.097	0.175	-0.059
Locational variables					
log(CBD)	-	-0.288	0.084	4.781	-0.308
log(INDUS)	+	0.047	0.010	0.052	0.014
log(MSCP)	+	0.054	0.049	0.080	0.046
log(TOPPRIM)	-	-0.041	-0.004	0.146	-0.009
log(PRIM)	-	0.028	-0.001	0.104	0.002
log(TOPSEC)	-	-0.020	-0.200	1.783	-0.053
log(BUSLINES)	+	0.011	0.001	0.012	0.000
log(SECOND)	-	0.025	0.020	0.080	0.019
BUS_26	+	0.030	0.029	0.121	0.032
log(MALL)	-	-0.012	-0.009	0.073	-0.006
log(MRT)	-	0.001	-0.032	0.117	-0.025
Contractual and data source dependent variables					
FREE	+	0.133	0.139	0.091	0.144
FROMPUB	-	-0.040	-0.014	0.018	-0.010
Q0310	-	-0.081	-0.077	0.025	-0.078
Q0410	-	-0.043	-0.052	0.024	-0.053
Q0111	-	-0.013	-0.016	0.023	-0.018
Adjusted R-square		0.864	0.937	0.021	0.941

Bold: Not significant at 0.1 level

Figure 15: Model B: Spatial variation of structural coefficients



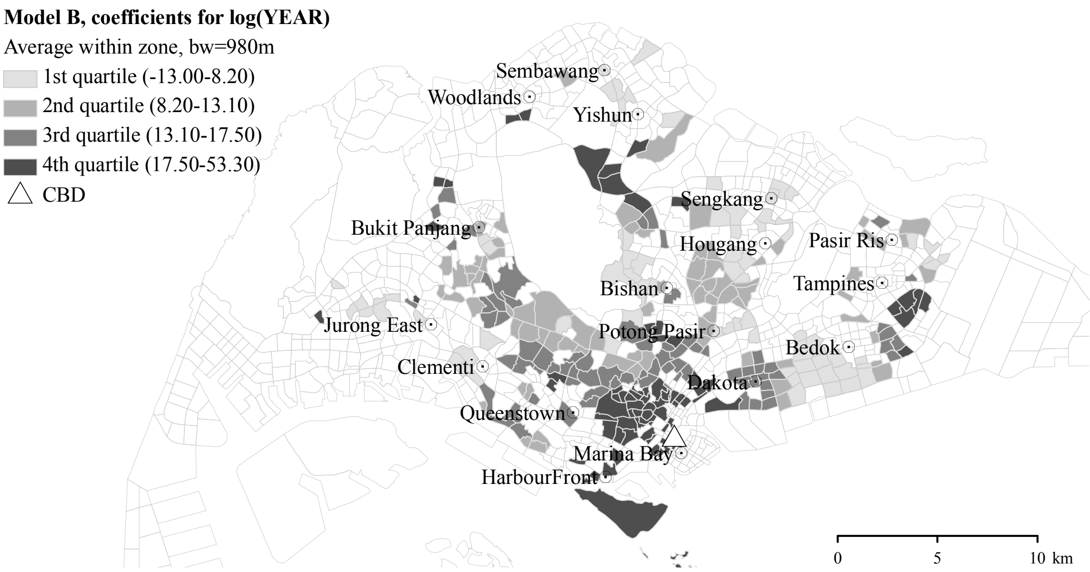
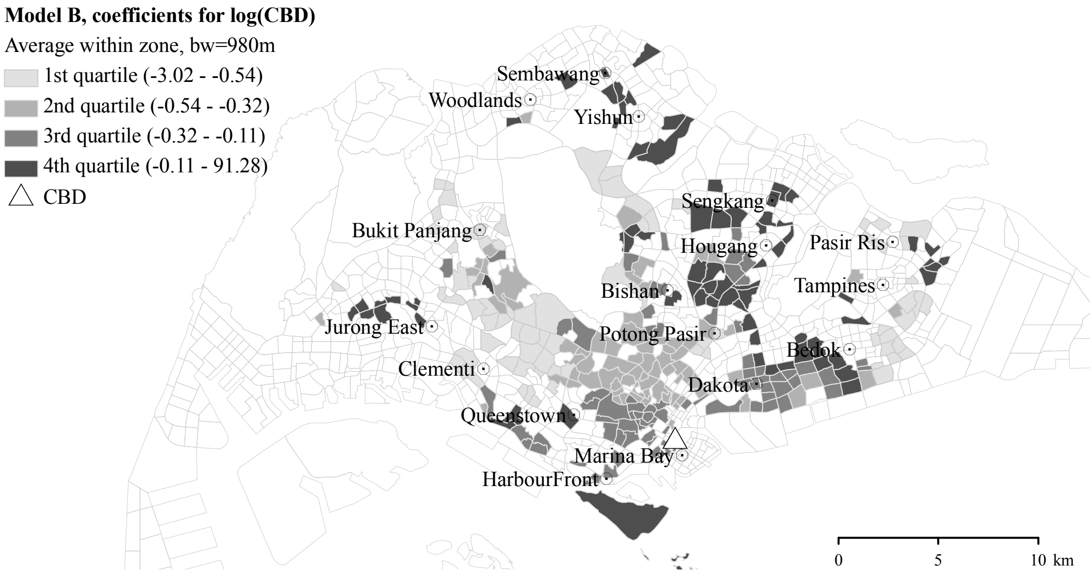
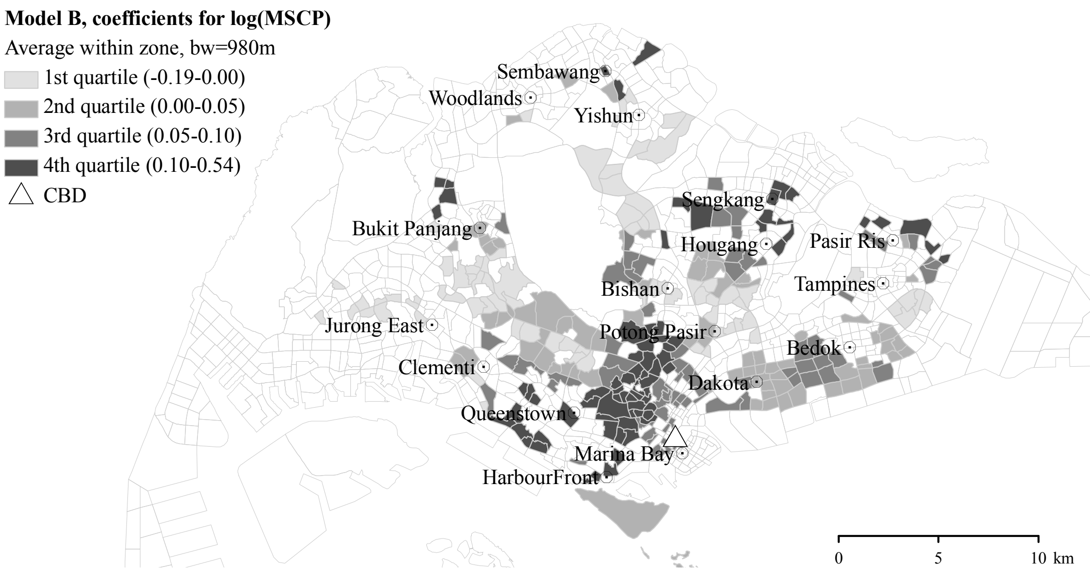


Figure 16 shows spatially varying coefficients of important locational variables. The first map indicates that the price impact of the distance to the CBD varies strongly over space. The coefficients vary between -3.02 and +91.28. It must be emphasized that only four out of 307 zones turned out with a average coefficient of above twenty. These four zones are located in the very North of the island in more than 15 kilometers air distance from the CBD. In 56 zones (18%) a longer distance to the CBD is estimated to have a positive price impact in average. Darker colors represent higher coefficients and therefore a lower importance of the proximity to the CBD. The map shows that the variable is not very important for flats that are very close to the CBD or close to regional sub-centres. On the other hand longer distances to the CBD cause lower prices for flats located between Holland Road and the nature reserve since coefficients turned out to be negative there.

The second map shows coefficients for the variable MSCP (distance to the nearest multi storey car park). Darker colors again represent higher coefficients and therefore a lower importance of the proximity to MSCP. The coefficients vary between -0.2 and +0.5. The lowest quartile exactly represents all zones, where the average coefficient is negative. In these zones proximity to MSCP causes higher housing prices. In all other zones (2nd, 3rd and 4th quartile) the average coefficients are positive and the proximity to MSCP is therefore considered to have a negative price impact. It stands out that the coefficients turned out with a positive sign in all central zones between Queenstown, Marina Bay and Potong Pasir. This is surprising since there are very few MSCP within this zone, as shown in Figure 24 of Appendix A.2. Around Jorong East, the proximity to MSCP is expected to have a positive price impact while all other sub-centres turned out with varying results.

Figure 16: Model B: Spatial variation of locational coefficients





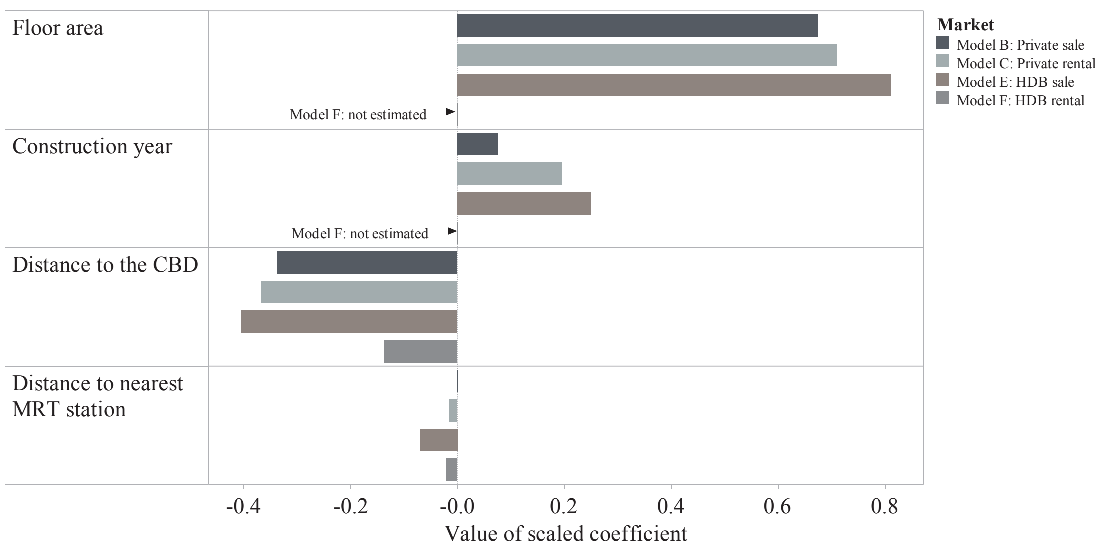
5. Discussion

This chapter aims to compare modelling results of different housing markets in order to point out the structure of housing preferences in Singapore. Further the estimation results are compared to insights gained through the literature review. The chapter concludes with recommendations for further research at Future Cities Laboratory.

5.1 Housing preferences in Singapore

In this chapter, housing preferences are compared among different market segments using the estimated model coefficients of the best models (SARerr specification in all cases). Models A and D are excluded for this comparison since these were carried out to estimate a constant parameter for the asking price premium. Models B and E represent the sale markets and Models C and F the rental markets. It must be emphasized that the rental models show expected preferences (asking data) and the sale models represent revealed preferences (transaction data). All coefficients used for comparison are scaled (variables are normalized before the estimation). Figure 17 compares scaled coefficients of selected explanatory variables.

Figure 17: Comparison of scaled estimates of selected variables

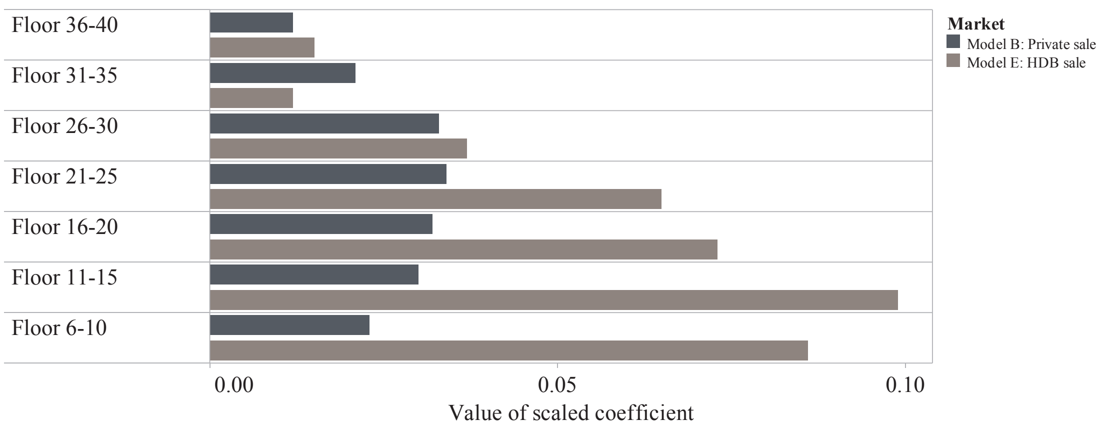


The aim of Figure 17 is to visualize the absolute differences between the coefficients. The comparison clearly shows the strong dominance of the floor area and a relatively high importance of the distance to the CBD in all markets. But it also illustrates the marginal importance of the majority of the selected variables. Table 31 provides an overview of the scaled coefficients of all variables appearing in two models at least. The variables are sorted by height of Model B scaled estimates.

5.1.1 Structural housing preferences

The two most important structural variables - the floor area and the construction year - have already been compared across markets in Figure 17. It stands out that the coefficients for the construction year indicate a much lower impact in the private sale market than in private rental and HDB sale markets. Further the availability of wellness facilities (WELL) turned out to have a significant but very small impact in both private markets.

Figure 18: Comparison of scaled estimates of different floor levels



Another important structural variable is the floor level. It has been discretized with dummy variables for both the private and the HDB sale markets. Figure 18 shows a comparison of scaled coefficients estimating price impacts of 5-storey-ranges compared to the lowest floor levels (floor 1-5). Obviously the preferences in private and HDB markets show different structures. In the HDB market the relatively big price premium for flats between the sixth and the twentieth floor decrease rapidly for higher floors. Flats located higher than floor 30 do not yield considerably higher prices than flats on

Table 31: Comparison of scaled coefficients

	Sale		Rental	
	Private	HDB	Private	HDB
Name	Model B	Model E	Model C	Model F
Specification	SARerr	SARerr	SARerr	SARerr
Number of observations	12’467	32’235	22’011	6’351
Dependent variable	log(PRICE)	log(PRICE)	log(PRICE)	log(PRICE)
Structural variables				
NOBED				0.912
log(SIZE)	0.674	0.810	0.708	
log(YEAR)	0.076	0.249	0.194	
CONDO	0.062		0.020	
FLO0610	0.023	0.086		
FLO1115	0.030	0.099		
FLO1620	0.032	0.073		
FLO2125	0.034	0.065		
FLO2630	0.033	0.037		
FLO3135	0.021	0.012		
FLO3640	0.012	0.015		
WELL	0.019		0.029	
Locational variables				
log(CBD)	-0.337	-0.405	-0.368	-0.138[1]
log(INDUS)	0.054	0.012	0.051	-0.018[1]
log(MRT)	**-0.001**	-0.069	**-0.016**	-0.021[1]
log(MSCP)	0.065	**0.000**	0.067	
log(BUSLINES)	**-0.006**	0.006	**0.005**	
log(SECOND)	**0.013**		-0.035	0.019[1]
log(TOPPRIM)	-0.069	-0.079		
log(TOPSEC)	-0.049	-0.079		
log(PRIM)	0.049		0.032	
BUS_26	0.033			-0.012
log(SUPERM)		-0.005	-0.027	
log(FOOD)		-0.260		-0.045[1]
Contractual variables				
FREE	0.146		0.042	
Q0310	-0.071	-0.127		
Q0410	-0.045	-0.058		
Q0111	-0.013	-0.027		
AIC	-3’293.7	-3’956.1	2’139.5	3’426.0

Bold: Not significant at the 0.1 level, [1]no transformation the lowest floors. The structure in the private sale market is different. Estimates do not vary between different floor levels. But high located flats still yield a price premium compared to flats between the first and the fifth floor.

5.1.2 Locational housing preferences

The most important locational variable - the distance to the CBD - has already been compared across markets in Figure 17. But variation concerning locational preferences can also be seen when looking at other variables. Figure 19 compares scaled coefficients of three locational variables, which where all selected for models B, C and E. The number of bus lines at the nearest station and the distance to the nearest multi storey car park were not included in model F.

Private flats yield significantly higher prices when the the distance to the nearest multi storey car park increases. The HDB sale market on the other hand is not sensitive to this variable (scaled coefficient $\sim$ 0.00). It stands out that a longer distance the nearest industrial estate causes higher prices in all markets except of the HDB rental market. An increasing number of bus lines at the nearest station has a negative price impact in the private sale market while the effect is positive in the two compared markets.

Figure 19: Comparison of scaled estimates of distances to transportation

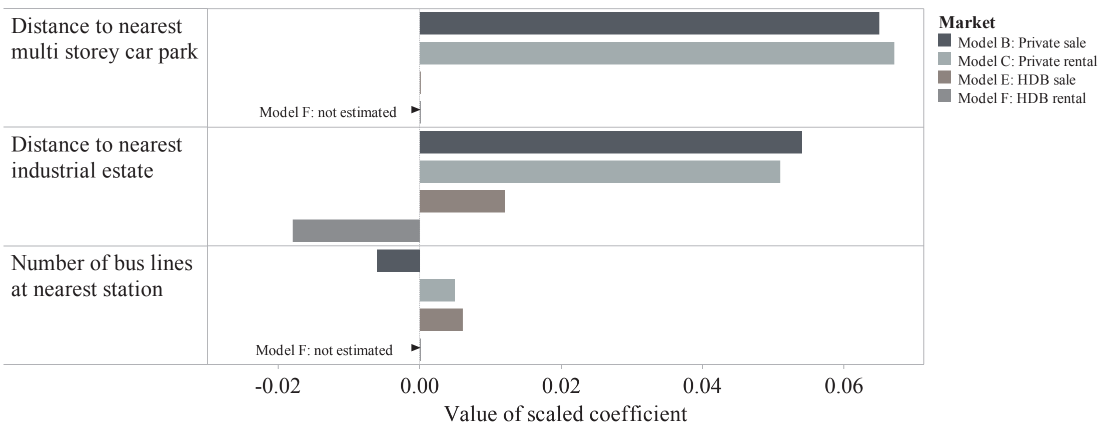
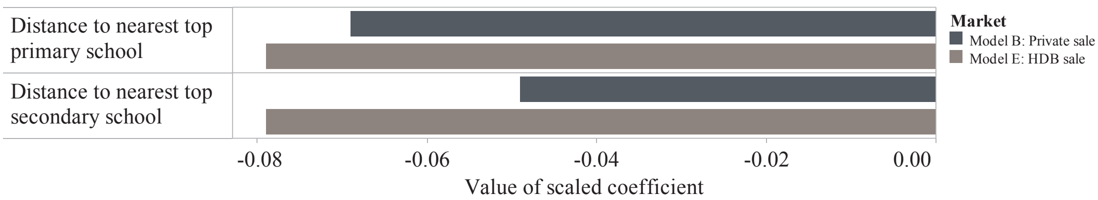


Figure 20 shows that short distances to top schools lead to higher housing prices. Private housing prices are slightly less sensitive to the proximity to top secondary schools than HDB prices.

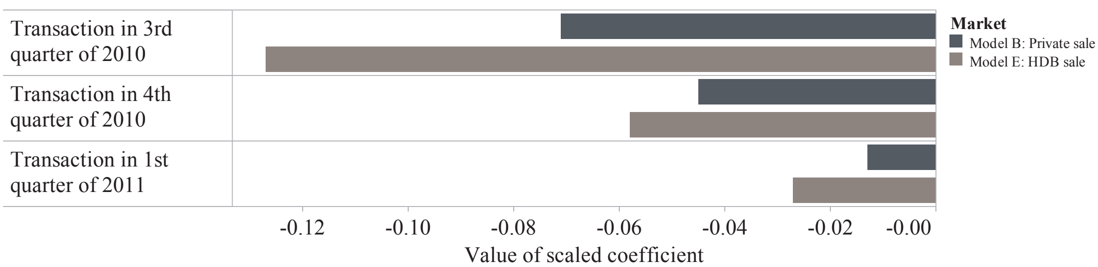
Figure 20: Comparison of scaled estimates of distances to top schools



5.1.3 Temporal effects

Figure 21 clearly shows that housing in both the private sale and the HDB sale markets becomes more expensive. The three dummy variables of former quarters compare prices with the second quarter of 2001. All coefficients turned out with negative signs and the negative becomes stronger the farther the transactions are in the past.

Figure 21: Comparison of scaled estimates of transaction time



5.2 Conclusions

Expectations have been formulated in Section 2.4 and were leading trough the whole empirical analysis and modelling process. The general hypothesis of the existence of a clear market segmentation was verified. As pointed out in a descriptive market comparison in Section 3.3, HDB and private markets differ dramatically. Average sale and rental prices in the private market are four times higher than in the HDB market. **Private flats are found to be more expensive, larger, younger and closer to the CBD than HDB flats**. This is valid for both private and rental markets. It is further found that **asking prices are much higher than transaction prices** in the private sale market. The tremendous gap between of around sixty percent is surprising and raises questions about reliability of asking data. On the other hand prices reflect market expectations and are therefore an indication of expected housing preferences. In the HDB sale market, however, asking and transaction prices do not differ significantly.

Hedonic models generally produced expected coefficient signs. **The floor area and the distance to the CBD are found to be the dominant price determinants** in all housing markets. But the results also revealed market-specific differences in housing preferences. Model results show that HDB and rental prices stronger rely on the availability of public transportation than sale prices. Both proximity to MRT stations and number of available bus lines turned out to have a positive impact for private rental and HDB flats. It is interesting that asking prices are sensitive to the proximity to MRT stations while this variable is not significant for transaction prices. It is further found that the floor level of a flat causes different price impacts in different markets. While HDB flats are estimated to yield the highest prices between the 11th and 15th floor, private flats are most expensive between the twenty first and the thirtieth floor. Further locational preference differences are found concerning distances to industrial estates and schools. Transactions in the latest quarter are estimated to yield significant higher prices than transactions from the 3rd and 4th quarter 2010.

Geographically weighted regression (GWR) shows **spatially varying housing preferences**. Floor area is found to have a significant stronger price impact in the city centre than on the outskirts. GWR further show that prices of remote flats are more sensitive to the proximity to private transportation facilities such as multi storey car parks. The GWR estimations clearly point to the existence of spatially segregated housing markets.

5.3 Recommendations

The following recommendations are directed to researchers of the FCL module VIII, who will use hedonic prices as input for agent based modelling. Concerning the gathered data it is suggested to verify asking prices with other sources since the prices are supposed to be too high. It is therefore recommended to work with transaction data for the moment. It is proposed to **base further work on spatial error models** (SARerr) since this specification solved the problem with autocorrelation in the OLS models. SARerr estimates performed with very good model fit and mainly produced expected coefficient signs. It is recommended to improve the SARerr models regarding the following aspects:

- **Replacement of distances to points of interest with travel times**. It is assumed that travel times would better represent pros and cons of a flats location. It is recommended to incorporate travel times combining all modes to represent whole trips (e.g. foot walk/bus ride/MRT ride/foot walk). Alternatively a high resolved **accessability index** could be generated and incorporated instead.
- Price distribution maps visually point to a dependence of housing prices to **population and/or workplace density**. Since population and workplace data was not available for this thesis it is recommended to incorporate it as soon as it is available.
- Inclusion of further variables representing **living quality** such as solar exposure, temperature and daylight availability.
- HDB rental market: Data analysis showed that HDB rental data possibly contains room rental listing as well as flat rental adds. Since these two sub-markets cannot be clearly distinguished with the available data it is recommended to **gather HDB rental data from different sources**. Solving this problem is not priority since HDB rental contract conclusions represent only around 3.5% of the total number of housing transactions/conclusions within the last six months.

Finally it is recommended to make **further efforts on the analysis of spatially varying housing preferences**. First GWR estimations point to the existence of spatially segregated markets with specific housing preferences. It is expected that further work on this topic can lead to a better understanding of the urban structure of Singapore.

References

Allproperty (2011) Property Guru, webpage, `http://www.propertyguru.com.sg/`. Accessed from February to March 2011.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

Anselin, L., A. K. Bera, R. Florax and M. J. Yoon (1996) Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics*, **26** (1) 77–104.

Beyer (2004) Hawth's Analysis Tools for ArcGIS, software, `http://www.spatialecology.com/htools`.

Bivand, R. (2011a) R package spdep: version 0.5-35, software, `http://cran.r-project.org/web/packages/spdep/index.html`.

Bivand, R. (2011b) R package spgwr: version 0.6-10, software, `http://cran.r-project.org/web/packages/spgwr/index.html`.

Box, G. E. P. and D. R. Cox (1964) An analysis of transformations, *Journal of the Royal Statistical Society*, **26** (2) 211–252.

Chalermpong, S. (2007) Rail transit and residential land use in developing countries: Hedonic study of residential property prices in Bangkok, Thailand, *Transportation Research Record*, (2038) 111–119.

Chin, T.-L., K. W. Chau and F. F. Ng (2004) The impact of the Asian financial crisis on the pricing of condominiums in Malaysia, *Journal of Real Estate Literature*, **12** (1) 33–50.

Cliff, A. D. and J. K. Ord (1981) *Spatial processes: models and applications*, Pion, London.

ESRI (2011) ArcGIS Desktop: Release 9.3, software, `http://www.esri.com/software/arcgis/arcinfo/index.html`.

ETHZ (2011) ETH Future Cities Laboratory, webpage, `http://www.futurecities.ethz.ch`. Accessed on 13/03/2011.

Fahrländer, S. S. (2006) Semiparametric construction of spatial generalized hedonic models for private properties, *Swiss Journal of Economics and Statistics*, **142** (4) 501–528.

Fahrländer, S. S. (2007) *Hedonische Immobilienbewertung: Eine empirische Untersuchung der Schweizer Märkte für Wohneigentum 1985 bis 2005*, m-press, München.

Farber, S. and M. Yeates (2006) A comparison of localized regression models in a hedonic house price context, *Canadian Journal of Regional Science*, **29** (3) 405–420.

Fischer, M. M. and A. Getis (eds.) (2010) *Handbook of Applied Spatial Analysis. Software Tools, Methods and Applications*, Springer, Heidelberg.

Fotheringham, A. S., C. Brunsdon and M. Charlton (2002) *Geographically Weighted Regression - The Analysis of Spatially varying Relationships*, John Wiley & Sons, Chichester.

Fotheringham, A. S. and T. C. Pitts (1995) Directional variation in distance decay, *Environment and Planning A*, **27** (5) 715–729.

Gelfand, A. E., P. J. Diggle, M. Fuentes and P. Guttorp (eds.) (2010) *Handbook of spatial statistics*, CRC Press, Boca Raton.

Giant (2011) Giant hypermarket, webpage, `http://www.gianthypermarket.com.sg`. Accessed on 10/04/2011.

Google Maps (2011) Google Geocoding API, Google Maps API Web Services, webpage, `http://www.geopostcodes.com`. Accessed on 10/04/2011.

HDB (2011a) HDB Annual Report 2009/2010, webpage, `http://www88.hdb.gov.sg/eBook/AnnualReport/Statistics2010.html`. Accessed on 29/03/2011.

HDB (2011b) HDB InfoWEB, webpage, `http://services2.hdb.gov.sg/webapp/BB33RTIS/BB33PReslTrans.jsp`. Accessed on 29/03/2011.

HDB (2011c) HDB Upgrading Programmes, webpage, `http://www.hdb.gov.sg/fi10/fi10328p.nsf/w/UpgradeWhatsLUP?OpenDocument`. Accessed on 25/04/2011.

HDB (2011d) HDB website, webpage, `http://www.hdb.gov.sg`. Accessed on 29/03/2011.

Jim, C. Y. and W. Y. Chen (2009) Value of scenic views: Hedonic assessment of private housing in Hong Kong, *Landscape and Urban Planning*, **91** (4) 226–234.

Jones, K. (1991) Specifying and estimating multi-level models for geograpical research, *Transactions of the Institute of Brithish Geographers*, **16** (2) 148–159.

Kim, A. M. (2007) North versus south: The impact of social norms in the market pricing of private property rights in Vietnam, *World Development*, **35** (12) 2079–2095.

Kissling, W. D. and G. Carl (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models, *Global Ecology and Biogeography*, **17** (1) 59–71.

Lancaster, K. J. (1966) A new approach to consumer theory, *Journal of Political Economy*, **74** (3) 132–157.

Löchl, M. (2010) Application of spatial analysis methods for understanding geographic variation of prices, Ph.D. Thesis, ETH Zurich.

Malpezzi, S. (2002) Hedonic pricing models: A selective and applied review, in T. O'Sullivan and K. Gibb (eds.) *Housing Economics and Public Policy*, Blackwell Science, Oxford.

MATSim-T (2011) Multi Agent Transportation Simulation Toolkit, webpage, `http://www.matsim.org`. Accessed on 03/04/2011.

Ministry of Education (2011) MOE Singapore, webpage, `http://www.moe.gov.sg`. Accessed on 05/04/2011.

NAVTEQ (2011) NAVTEQ Maps and Traffic, webpage, `http://www.navteq.com`. Accessed on 29/02/2011.

Neo, P. H., N. J. Lee and S. E. Ong (2003) Government policies and household mobility behaviour in Singapore, *Urban Studies*, **40** (13) 2643–2660.

NTUC (2011) webpage, `http://www.fairprice.com.sg`. Accessed on 10/04/2011.

Ong, S. E. and K. H. D. a. Ho (2003) A constant-quality price index for resale public housing flats in Singapore, *Urban Studies*, **40** (13) 2705–2729.

PAEXCO (2011) Directory of top schools in Singapore, webpage, `http://singapore-top-schools.blogspot.com`. Accessed on 12/05/2011.

Pebesma, E. (2011) R package sp: version 0.9-82, software, `http://cran.r-project.org/web/packages/sp/index.html`.

Phang, S.-Y. (2007) The Singapore model of housing and the welfare state, *Research Collection School of Economics Singapore*, (596) 15–46.

R Development Core Team (2011) A language and environment for statistical computing, software, `http://www.r-project.org`.

Sakamoto, Y. and G. Kitagawa (eds.) (1987) *Akaike information criterion statistics*, Kluwer Academic Publishers, Norwell.

Sing, T.-F., I.-C. Tsai and M.-C. Chen (2006) Price dynamics in public and private housing markets in Singapore, *Journal of Housing Economics*, **15** (4) 305–320.

Sirmans, S. G., D. A. Macpherson and E. N. Zietz (2009) The composition of hedonic pricing models, **13** (1) 1–44.

Streetdirectory (2011) Streetdirectory Singapore, webpage, `http://www.streetdirectory.com`. Accessed on 10/04/2011.

Studenmund, A. H. (2006) *Using econometrics: a practical guide*, 5. edn., Addison-Wesley, Boston.

Sue, E. D. W. and W.-K. Wong (2010) The political economy of housing prices: Hedonic pricing with regression discontinuity, *Journal of Housing Economics*, **19** (2) 133–144.

Tang, B.-s. and Y. Y. Chung (2010) Space and scale: A study of development intensity and housing price in Hong Kong, *Landscape and Urban Planning*, **96** (3) 172–182.

Tse, R. Y. C. and P. E. D. Love (2000) Measuring residential property values in Hong Kong, *Property Management*, **18** (5) 366–374.

URA (2011) Real estate informations system REALIS, webpage, `http://www.ura.gov.sg/realis/`. Accessed on 29/03/2011.

Wong, S. K., K. W. Chau, Y. Yau and A. K. C. Cheung (2005) Property price, floor level, and building density, paper presented at the *6th International Conference on Tall Buildings*, December 2005.

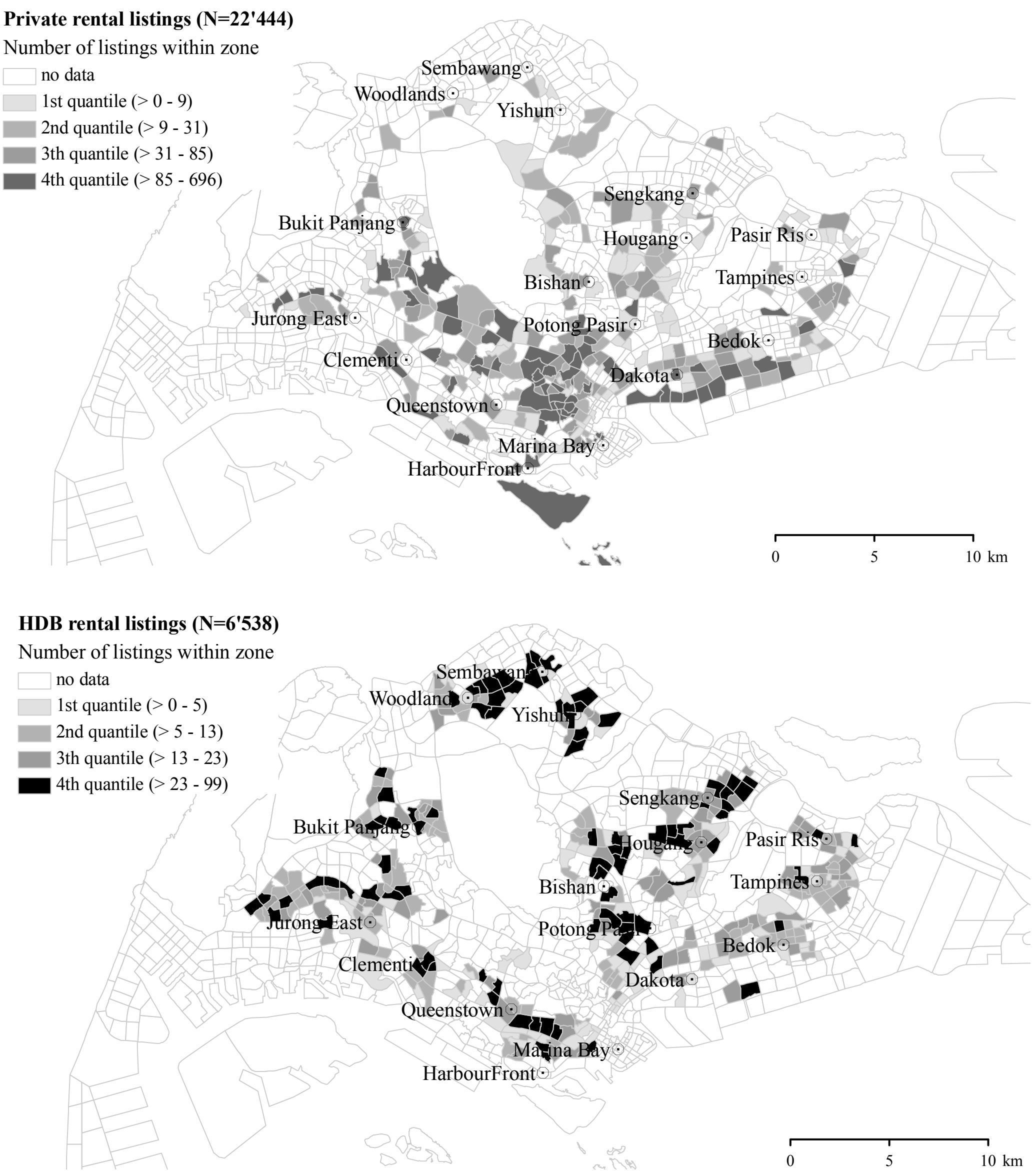
Yusuf, A. A. and B. P. Resosudarmo (2009) Does clean air matter in developing countries' megacities? A hedonic price analysis of the Jakarta housing market, Indonesia, *Ecological Economics*, **68** (5) 1398–1407.

ZKB (2004) Preise, Mieten und Renditen, webpage, `https://www.zkb.ch/etc/ml/repository/prospekte_und_broschueren/corporate/studien/studie_immomarkt_transparent_pdf.File.pdf`.

A. Appendix

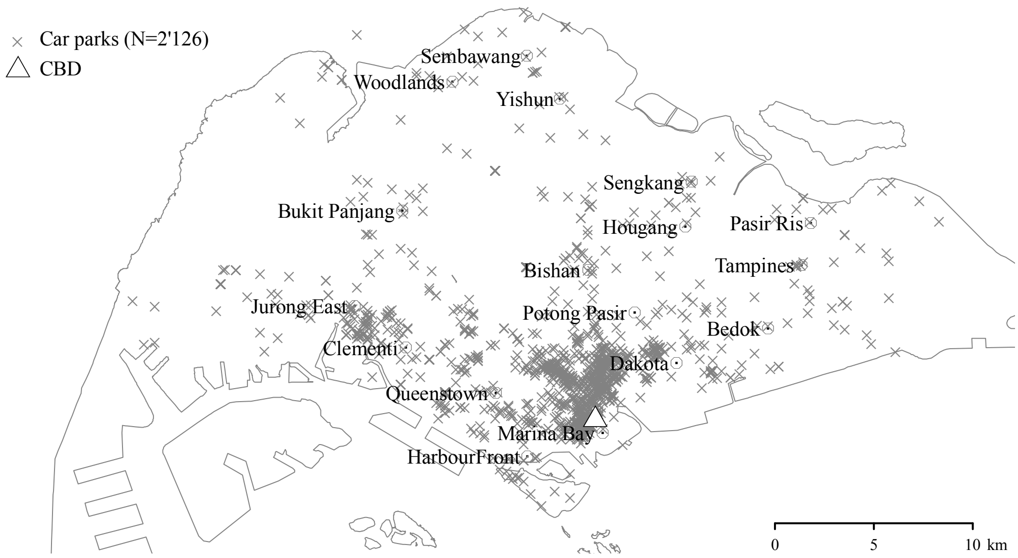
A.1 Spatial distribution of gathered rental listings

Figure 22: Spatial distribution of gathered rental data



A.2 Spatial distribution of points of interest

Figure 23: Spatial distribution of car parks and food centres



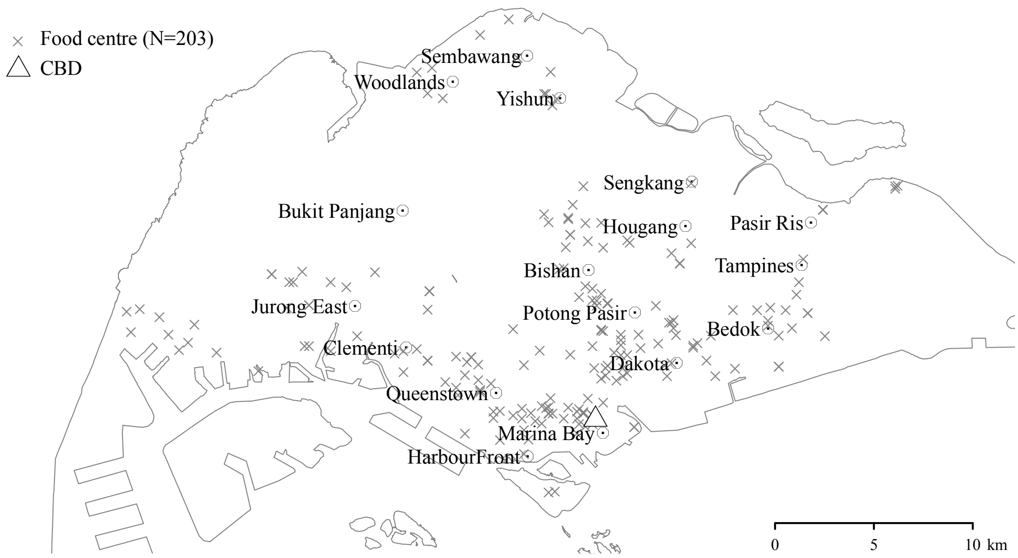
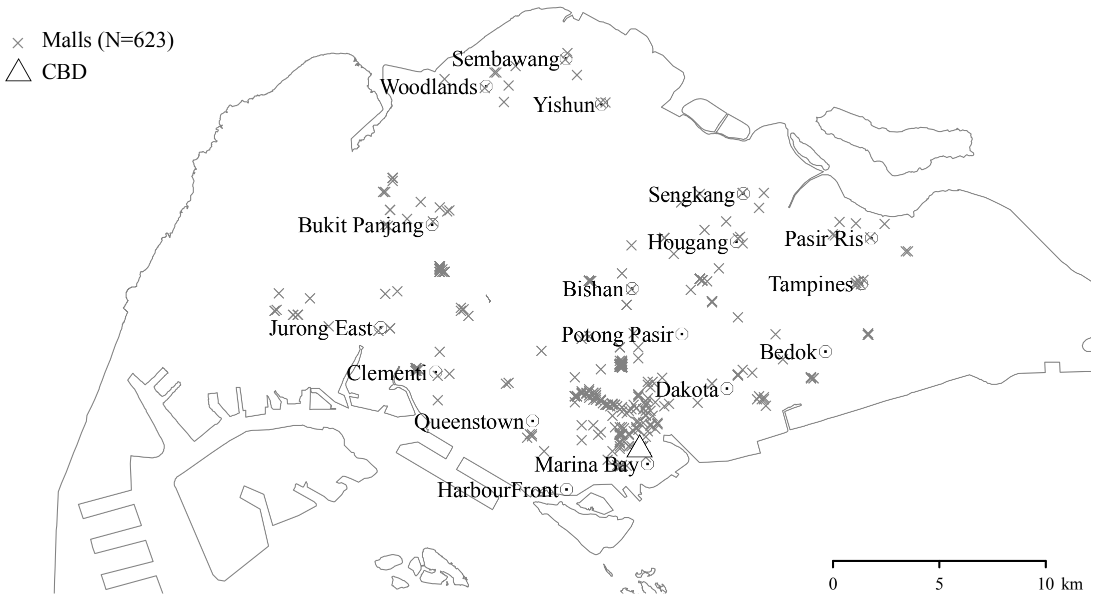


Figure 24: Spatial distribution of malls and multi storey car parks



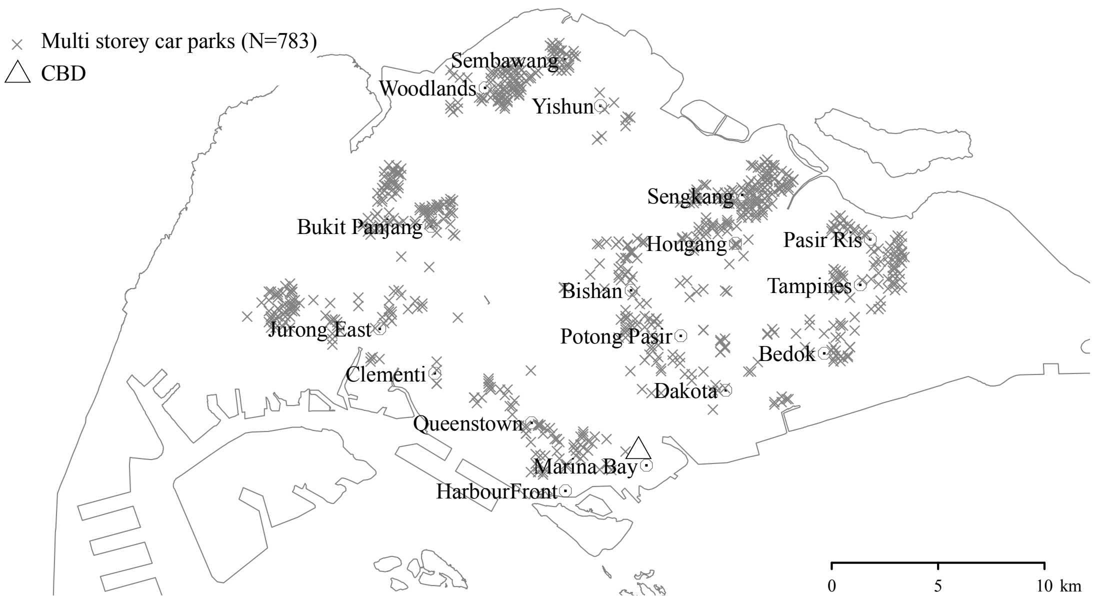
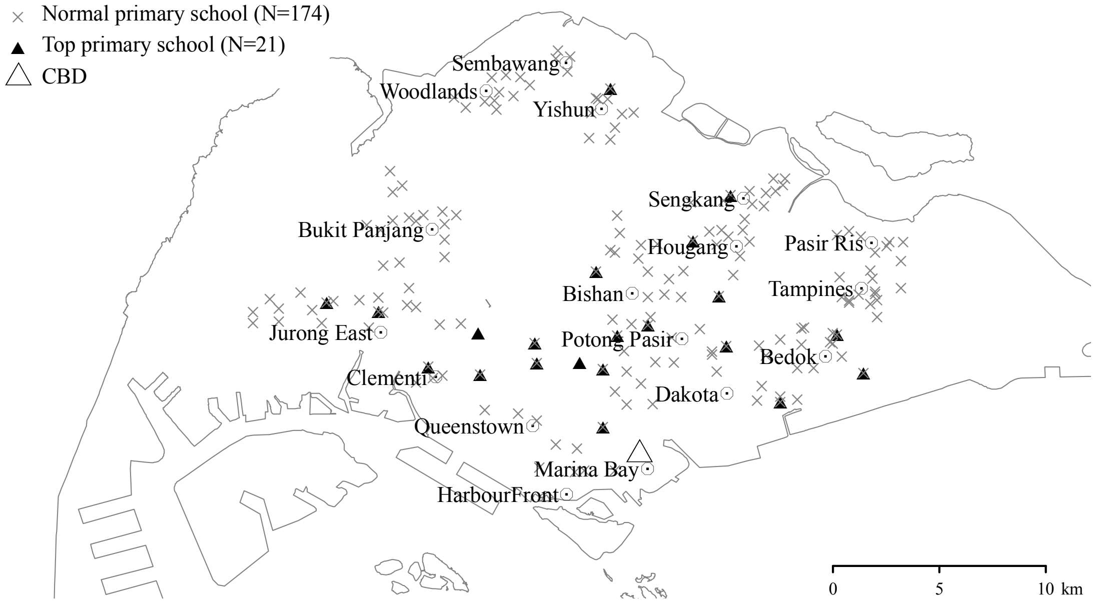
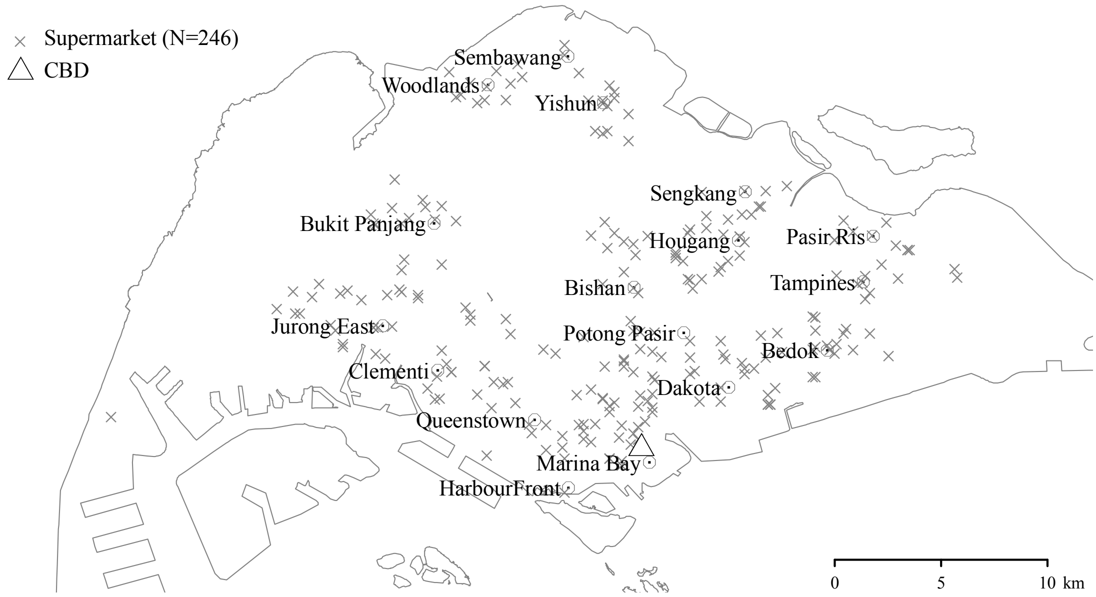


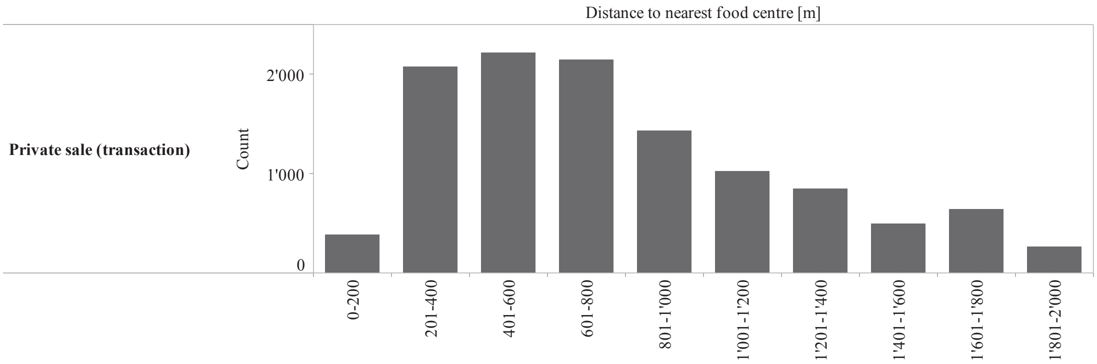
Figure 25: Spatial distribution of primary schools and supermarkets

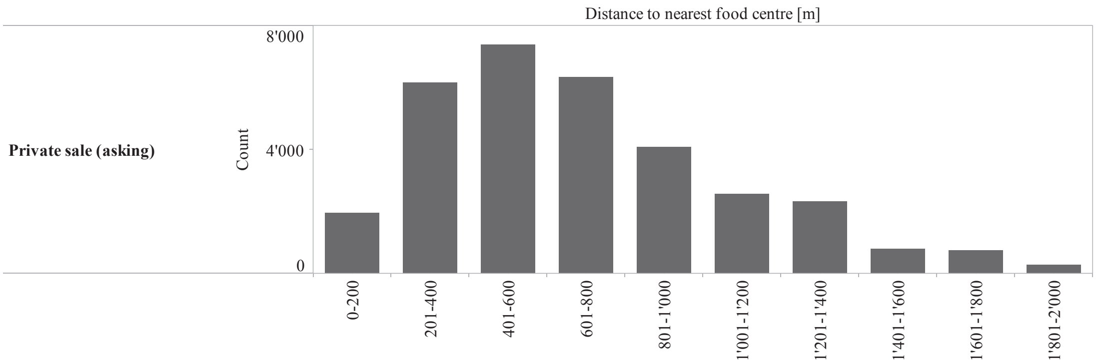




A.3 Distance to nearest food centre

Figure 26: Histograms of distance to nearest food centre





A.4 Overview of variables

Table 32: Overview of available variables

Variable	Description	Type
Structural explanatory variables		
CONDO	Building is a condominium	D
YEAR	Construction year of the building	D
FLOOR	Dummies for floor level ranges	D
PARKING/POOL/SEC/WELL	Dummies for availability of amenities	D
MUP/IUP/LUP/HIP	Dummies for HDB upgrading programs	D
MAIS/AP/NGEN/MODA/SIMPL	Dummies for types of HDB flats	D
SIZE	Floor area in square meter	C
NOBATH	Number of bathrooms	C
NOBED	Number of bedrooms	C
Locational explanatory variables		
BUS	Distance to nearest bus stop [m]	C
CARPARK	Distance to nearest car park [m]	C
FOOD	Distance to nearest food centre [m]	C
INDUS	Distance to nearest industrial estate [m]	C
MALL	Distance to nearest mall [m]	C
MRT	Distance to nearest MRT station [m]	C
MSCP	Distance to nearest multi storey car park [m]	C
PRIM	Distance to nearest primary school [m]	C
SECOND	Distance to nearest secondary school [m]	C
SUPERM	Distance to nearest supermarket [m]	C
TOPPRIM	Distance to nearest top primary school [m]	C
TOPSEC	Distance to nearest top secondary school [m]	C
CBD	Distance to the CBD [m]	C
BUSLINES	Number of bus lines at nearest bus station	C
Contractual and data source dependent variables		
FREE	Contract is freehold	D
ASKING	Is an asking price (from Property Guru)	D
FROMPUB	Buyer lived in a HDB flat before	D
QXXXX	Dummies for date of transaction (quarter)	D

A.5 OLS diagnostic plots

Figure 27: Model A-C: Normal Q-Q and Tukey-Anscombe plots

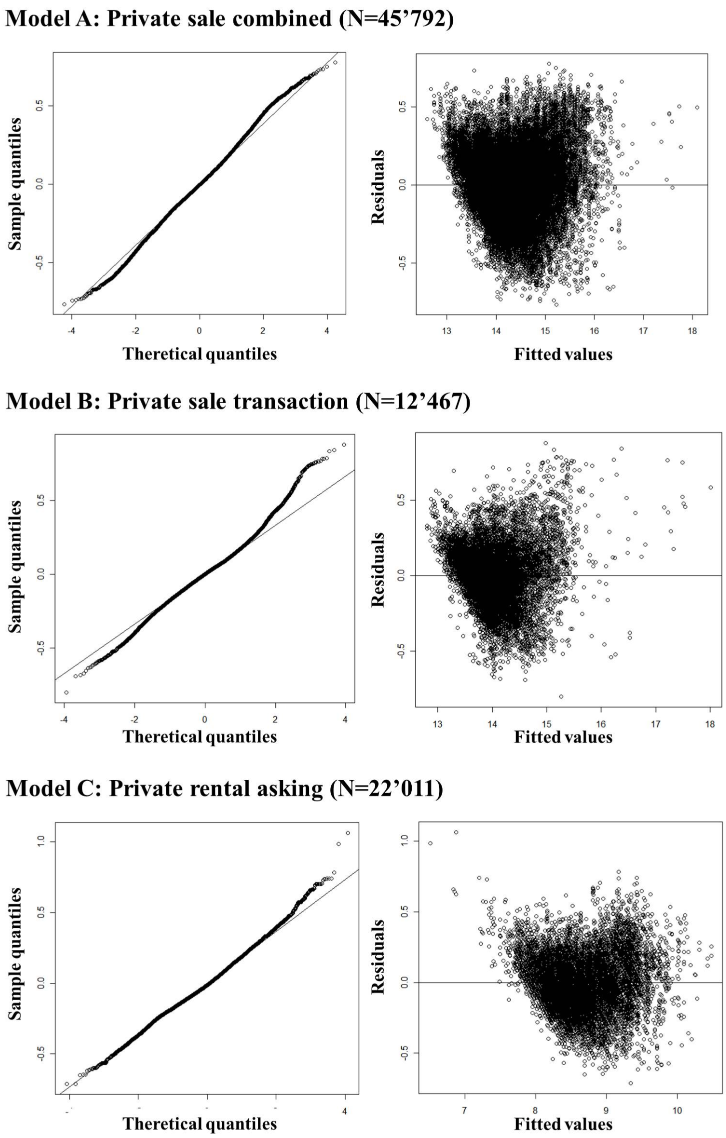
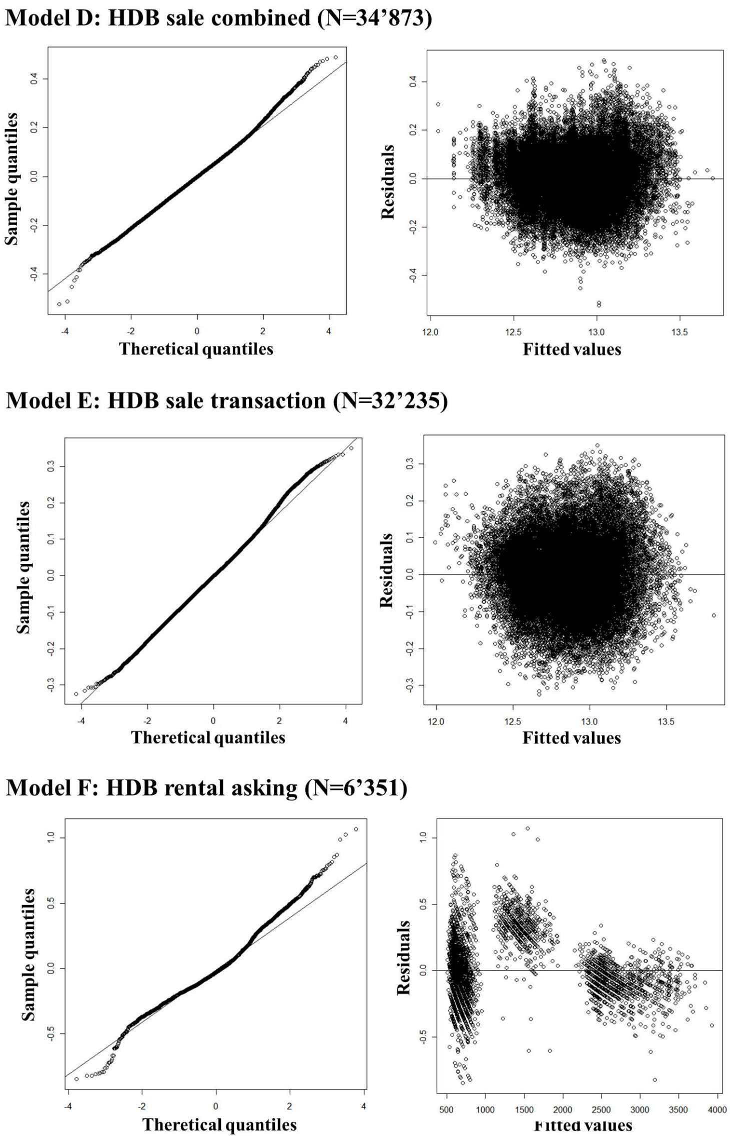


Figure 28: Model D-F: Normal Q-Q and Tukey-Anscombe plots



A.6 SAR estimations

Table 33: Model D: SAR estimations

Dependent: log(PRICE)							N = 34’873
		SARerr model (k=15)			**SARdurbin model (k=15)**		
Explanatory variable	Exp.	Estimate	Scaled	Sign.	Estimate	Scaled	Sign.
Constant		-62.333			-10.433		
Lambda / Rho		0.839			0.834		
Structural variables							
log(SIZE)	+	0.813	0.856	***	0.815	0.857	***
log(YEAR)	+	9.756	0.185	***	9.850	0.187	***
Locational variables							
log(CBD)	-	-0.227	-0.431	***	-0.053	-0.101	**
log(MRT)	-	-0.027	-0.071	***	0.000	0.000	***
log(TOPSEC)	-	-0.019	-0.070	***	-0.011	-0.042	*
log(TOPPRIM)	-	-0.023	-0.075	***	-0.029	-0.094	***
log(BUSLINES)	+	0.001	0.003		-0.002	-0.008	**
log(MSCP)	+	-0.003	-0.017	***	-0.003	-0.015	***
log(SUPERM)	-	-0.005	-0.014	***	-0.005	-0.013	**
log(MALL)	-	-0.002	-0.005		-0.005	-0.014	***
log(SECOND)	-	0.001	0.003		-0.003	-0.006	
log(INDUS)	+	0.006	0.020	***	0.015	0.047	***
Data source dependent variables							
ASKING	+	0.093	0.097	***	0.093	0.097	***
SSE		175.3			175.1		
AIC		-80’461.4			-80’555.1		
Moran’s I		0.020			0.021		

Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 34: Model A: SAR estimations

Dependent: log(PRICE)							N = 45'792
		SARerr model (*k*=8)			**SARdurbin model (*k*=8)**		
Explonatory variable	Exp.	Estimate	Scaled	Sign.	Estimate	Scaled	Sign.
Constant		11.822			0.987		
Lambda / Rho		0.920			0.904		
Structural variables							
log(SIZE)	+	0.894	0.657	***	0.893	0.656	***
BUI_5160	-	-0.233	-0.005	***	-0.222	-0.005	***
BUI_6170	-	-0.007	-0.001		0.003	0.000	
BUI_7180	-	-0.136	-0.026	***	-0.126	-0.025	***
BUI_8190	-	-0.092	-0.027	***	-0.078	-0.023	***
BUI_9100	-	-0.107	-0.068	***	-0.096	-0.061	***
BUI_0110	-	-0.035	-0.027	***	-0.028	-0.022	***
BUI_1220	-	0.028	0.017	***	0.038	0.023	***
CONDO	+	0.022	0.016	***	0.018	0.013	***
PARKING	+	-0.040	-0.032	***	-0.031	-0.024	***
WELL	+	0.056	0.044	***	0.046	0.036	***
SEC	+	0.006	0.005		0.006	0.005	
POOL	+	0.007	0.003		0.006	0.003	*
GARD	+	-0.007	-0.005		-0.016	-0.012	***
Locational variables							
log(CBD)	-	-0.334	-0.438	***	0.340	0.446	***
log(INDUS)	+	0.037	0.047	***	-0.016	-0.021	**
log(MSCP)	+	0.100	0.123	***	0.067	0.082	***
log(BUSLINES)	+	0.001	0.011	*	-0.002	-0.019	**
log(PRIM)	-	0.084	0.096	***	0.064	0.073	***
log(MALL)	-	-0.017	-0.027	***	-0.031	-0.050	***
log(TOPPRIM)	-	-0.035	-0.043	***	-0.063	-0.078	**
log(FOOD)	-	0.061	0.088	***	0.073	0.106	***
BUS_26	+	0.029	0.020	***	0.008	0.005	***
log(MRT)	-	-0.023	-0.028	***	-0.031	-0.039	**
log(SECOND)	-	-0.036	-0.040	***	-0.053	-0.060	***
log(SUPERM)	-	-0.019	-0.023	***	-0.006	-0.007	***
log(TOPSEC)	-	0.002	0.002		0.001	0.001	
Contractual and data source dependent variables							
ASKING	+	0.090	0.063	***	0.093	0.065	***
FREE	+	0.129	0.102	***	0.112	0.088	***
SSE		492.6			490.23		
AIC		-73'014.6			-73'798.6		
Moran's I		-0.073			-0.069		

Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 35: Model B: SAR estimations

Dependent: log(PRICE)							N = 12'467
		SARerr model (*k*=8)			**SARdurbin model (*k*=8)**		
Explonatory variable	Exp.	Estimate	Scaled	Sign.	Estimate	Scaled	Sign.
Constant		-67.074			-20.686		
Lambda / Rho		0.883			0.864		
Structural variables							
log(SIZE)	+	0.862	0.674	***	0.858	0.671	***
CONDO	+	0.066	0.062	***	0.055	0.052	***
log(YEAR)	+	10.414	0.076	***	9.933	0.073	***
PARKING	+	-0.070	-0.065	***	-0.073	-0.068	***
FLO0610	+	0.029	0.023	***	0.030	0.024	***
FLO1115	+	0.047	0.030	***	0.046	0.030	***
FLO1620	+	0.065	0.032	***	0.066	0.032	***
FLO2125	+	0.104	0.034	***	0.103	0.034	***
FLO2630	+	0.131	0.033	***	0.124	0.031	***
FLO3135	+	0.139	0.021	***	0.139	0.022	***
FLO3640	+	0.113	0.012	***	0.121	0.013	***
FLO41UP	+	0.190	0.024	***	0.203	0.025	***
WELL	+	0.020	0.019	***	0.018	0.017	***
POOL	+	-0.028	-0.018	***	-0.027	-0.017	***
Locational variables							
log(CBD)	-	-0.255	-0.337	***	0.340	0.450	***
log(INDUS)	+	0.033	0.054	***	0.018	0.031	***
log(MSCP)	+	0.043	0.065	***	0.044	0.068	***
log(TOPPRIM)	-	-0.047	-0.069	***	0.004	0.006	***
log(PRIM)	-	0.039	0.049	***	0.030	0.037	**
log(TOPSEC)	-	-0.032	-0.049	***	-0.017	-0.026	
log(BUSLINES)	+	-0.003	-0.006		-0.012	-0.021	**
log(SECOND)	-	0.010	0.013		-0.013	-0.018	
BUS_26	+	0.041	0.033	***	0.038	0.031	***
log(MALL)	-	-0.028	-0.049	***	-0.032	-0.055	***
log(MRT)	-	-0.001	-0.001		-0.014	-0.022	
Contractual variables							
FREE	+	0.152	0.146	***	0.143	0.137	***
FROMPUB	-	-0.007	-0.007	***	-0.009	-0.008	***
Q0310	-	-0.078	-0.071	***	-0.078	-0.072	***
Q0410	-	-0.051	-0.045	***	-0.050	-0.045	***
Q0111	-	-0.015	-0.013	***	-0.016	-0.013	***
SSE		127.3			125.9		
AIC		-19'517.1			-19'760.3		
Moran's I		-0.014			-0.012		

Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 36: Model C: SAR estimations

Dependent: log(PRICE)							N = 22'011
		SARerr model (*k*=8)			**SARdurbin model (*k*=8)**		
Explonatory variable	Exp.	Estimate	Scaled	Sign.	Estimate	Scaled	Sign.
Constant		-164.750			-28.867		
Lambda / Rho		0.869			0.845		
Structural variables							
log(SIZE)	+	0.741	0.708	***	0.738	0.705	***
log(YEAR)	+	22.509	0.194	***	22.258	0.192	***
CONDO	+	0.022	0.020	***	0.012	0.011	***
GARD	+	0.000	0.000		0.019	0.019	***
WELL	+	0.028	0.029	***	0.025	0.025	***
Locational variables							
log(CBD)	-	-0.205	-0.368	***	0.023	0.041	*
log(INDUS)	+	0.032	0.051	***	-0.008	-0.012	***
log(BUS)	-	0.022	0.041	***	-0.030	-0.057	**
log(MSCP)	+	0.043	0.067	***	0.023	0.035	**
log(BUSLINES)	+	0.002	0.005		-0.013	-0.024	*
log(MRT)	-	-0.010	-0.016		-0.021	-0.035	***
log(SUPERM)	-	-0.017	-0.027	***	-0.004	-0.006	***
log(SECOND)	-	-0.024	-0.035	***	-0.039	-0.057	***
log(FOOD)	-	-0.015	-0.026	***	-0.004	-0.006	***
log(PRIM)	-	0.023	0.032	***	0.050	0.070	***
Contractual variables							
FREE	+	0.041	0.042	***	0.035	0.035	***
SSE		308.4			304.2		
AIC		-29'404.8			-29858		
MoransI		-0.043			-0.036		

Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 37: Model E: SAR estimations

Dependent: log(PRICE)							N = 32'235
		SARerr model (k=10)			**SARdurbin model (k=10)**		
Explonatory variable	Exp.	Estimate	Scaled	Sign.	Estimate	Scaled	Sign.
Constant		-85.111			-14.026		
Lambda / Rho		0.851			0.835		
Structural variables							
log(SIZE)	+	0.760	0.810	***	0.759	0.809	***
log(YEAR)	+	12.773	0.249	***	13.196	0.257	***
FLOOR0610	+	0.045	0.086	***	0.045	0.086	***
FLOOR1115	+	0.065	0.099	***	0.065	0.099	***
FLOOR1620	+	0.098	0.073	***	0.098	0.073	***
FLOOR2125	+	0.142	0.065	***	0.140	0.064	***
FLOOR2630	+	0.153	0.037	***	0.152	0.037	***
FLOOR3135	+	0.178	0.012	***	0.173	0.011	***
FLOOR3640	+	0.204	0.015	***	0.198	0.014	***
MUP	+	0.016	0.022	***	0.009	0.013	***
IUP	+	0.018	0.032	***	0.015	0.026	***
LUP	+	-0.017	-0.033	***	-0.018	-0.036	***
HIP	+	0.006	0.005	*	0.005	0.004	***
IS_MAIS	+	0.049	0.033	***	0.049	0.033	***
IS_AP	+	0.047	0.038	***	0.047	0.038	***
IS_NEWGEN	+	-0.013	-0.020	***	-0.017	-0.025	***
IS_MODA	-	-0.006	-0.011	***	-0.007	-0.011	***
IS_SIMPL	-	0.012	0.011	***	0.013	0.012	***
Locational variables							
log(CBD)	-	-0.214	-0.405	***	-0.032	-0.060	***
log(MRT)	-	-0.026	-0.069	***	-0.014	-0.037	**
log(TOPSEC)	-	-0.021	-0.079	***	-0.026	-0.098	***
log(TOPPRIM)	-	-0.024	-0.079	***	-0.016	-0.053	
log(BUSLINES)	+	0.002	0.006	*	-0.003	-0.009	**
log(MSCP)	+	0.000	0.000		0.000	0.001	
log(SUPERM)	-	-0.002	-0.005		-0.001	-0.004	
log(INDUS)	+	0.004	0.012	**	0.014	0.044	***
Contractual variables							
Q0210	-	-0.107	-0.171	***	-0.107	-0.172	***
Q0310	-	-0.069	-0.127	***	-0.069	-0.128	***
Q0410	-	-0.035	-0.058	***	-0.036	-0.058	***
Q0111	-	-0.018	-0.027	***	-0.018	-0.028	***
SSE		85.6			85.4		
AIC		-92'702.0			-93'065.5		
Moran's I		0.022			0.022		

Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 38: Model F: SAR estimations

Dependent: log(PRICE)							N = 6'351
		SARerr model (k=8)			**SARdurbin model (k=8)**		
Explonatory variable	Exp.	Estimate	Scaled	Sign.	Estimate	Scaled	Sign.
Constant		6.007			3.397		
Lambda / Rho		0.438			0.436		
Structural variables							
NOBED	+	0.711	0.912	***	0.712	0.913	***
NOBATH	+	0.034	0.027	***	0.035	0.028	***
Locational variables							
CBD	-	0.000	-0.138	***	0.000	-0.025	***
FOOD	-	0.000	-0.045	***	0.000	-0.033	***
SECOND	-	0.000	0.019	***	0.000	-0.049	**
MRT	-	0.000	-0.021	***	0.000	0.006	
INDUS	+	0.000	-0.018	**	0.000	0.104	**
BUS_26	+	-0.031	-0.012	**	-0.030	-0.012	**
SSE		269.3			268.5		
AIC		-1'606.2			-1'610.5		
Moran's I		0.039			0.039		

Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$

A.7 Model F: OLS estimations

Table 39: Model F: OLS estimation

Dependent: log(PRICE)							N = 6'351
Explanatory variable	Exp.	Coeff.	SE	Scaled	T stat.	Sign.	VIF
Constant		6.004					
Structural variables							
NOBED	+	0.710	0.004	0.910	199.797	***	1.219
NOBATH	+	0.033	0.006	0.026	5.732	***	1.212
Locational variables							
CBD	-	0.000	0.000	-0.134	-26.709	***	1.488
FOOD	-	0.000	0.000	-0.045	-9.868	***	1.224
SECOND	-	0.000	0.000	0.024	5.386	***	1.195
MRT	-	0.000	0.000	-0.022	-5.065	***	1.076
INDUS	+	0.000	0.000	-0.020	-4.594	***	1.154
BUS_26	+	-0.031	0.011	-0.012	-2.941	***	1.010
Adjusted R-square				0.892			
Sum of squared errors (SSE)				310.2			
Akaike Information Criterion (AIC)				-1'130.553			
Moran's I				0.204	***		
Robust LMerr				1'268.5	***		
Robust LMlag				1.2			

Probability of rejecting H0 = *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$

A.8 Model F: Spatial price pattern

Figure 29: Model F: Monthly rent spatial pattern