# Hedonic Modeling of Singapore's Resale Public Housing Market

Jiakun Xu

*Under the supervision of*

Dr. Charles M. Becker

Dr. Tracy A. Falba

Department of Economics, Duke University

Durham, North Carolina, USA

2017

*Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.*

## Table of Contents

## Acknowledgements

I am incredibly thankful for the invaluable advice and support by my faculty advisor, Dr. Charles Becker, who made this piece of work both possible and enjoyable. I would like to thank my honors thesis course instructor, Dr. Tracy Falba, for her guidance through the stages of thesis writing. I am grateful to my course instructor, Victor Yifan Ye, for the discussions and consultations from which the concept of this paper was born. I would also like to thank my Writing Consultant, Lillian Xie, for working with me through an earlier draft. My heartfelt gratitude goes out to my close friends and supporters, Clement Lee and Grace Lim, for our discussions that have shaped some of the insights in this paper and for offering comments on my drafts.

In addition, I would like to acknowledge the following data sources and services:

1. The "Resale Flat Prices" dataset, accessed on September 20, 2016, from the Singapore government open data repository, which is made available under the terms of the Singapore Open Data License version 1.0 {https://data.gov.sg/open-data-licence}.
2. The "Train Station" geospatial dataset, accessed on November 5, 2016, from the Singapore Land Transport Authority DataMall, which is made available under the terms of the Singapore Open Data License version 1.0 {https://data.gov.sg/open-data-licence}.
3. Microsoft Bing Maps REST Services API, for providing geocoding and driving directions services.
4. The OpenStreetMap project, for providing map data and map imagery.
   © OpenStreetMap contributors.
5. The Kiasuparents community, for providing compiled statistics on the annual Primary One Registration Exercise from 2006 to 2015.
6. R code written by Shane Lynn and shared on the R-bloggers community, originally written for geocoding using the Google Maps API, and adapted for use in this paper with the Bing Maps REST Services API {https://www.r-bloggers.com/batch-geocoding-with-r-and-google-maps/}.

## Abstract

The large-scale, high-density public housing market in Singapore invites hedonic analysis, due to its homogeneity in structure quality across all neighborhoods. This paper builds a time-dummy hedonic regression model incorporating geospatial features for a large dataset of resale transactions from 2000 to 2016. Significant anticipatory price effects are found for new subway stations, which peak at two years before station opening. A hedonic price index suggests that affordability was a problem during the sustained period of property price inflation from 2011 to 2013. District-level analysis shows evidence of increasing rent gradients, wealth disparities, and "lottery" effects in asset growth. I discuss the potential contributions of these insights to wealth and equity considerations in public policy design.

*JEL codes:* C21, R3, R31, R38, R41
*Keywords:* hedonic models, housing, public transit, Singapore

## 1. Introduction

Property markets are unique in two ways. First, each housing unit itself is a unique combination of a wide range of attributes, such that no two units are exactly alike. Second, the pricing of property transactions reflects not only these attributes, but also location factors, neighborhood quality, as well as larger macroeconomic trends and expectations. In their most basic form, houses are necessities in the sense that they fulfill one of our most fundamental needs of shelter. Concurrently, at the other end of the spectrum, they are vehicles for economic investment, or even speculation. Urbanization has steadily driven up real estate prices in many cities, accruing positive wealth effects to its residents in the process. However, serious problems arise when these prices swing sharply in either direction, or when speculative bubbles burst, such as during the 2007–09 subprime mortgage crisis in the United States. Thus, housing prices have been closely monitored from a policy perspective in advanced economies. The relation between housing prices and the macroeconomy has also received growing interest in the economics literature (see, for example, Iacoviello, 2005; Mendicino and Punzi, 2014; Shi et al., 2014).

Yet, there is no universally accepted way of calculating property price indices. Three main methods currently dominate: average price, repeated-sales, and hedonic regression. An average price index, while widely used and straightforward in its calculation, is limited in its usefulness as it does not account for a changing mix of the quality of houses transacted. This is a major flaw if we have reason to believe that the quality of houses sold changes across the growth and decline stages of an economic cycle. The U.S. government uses a repeated-sales method index, introduced by Bailey et al. (1963) and developed in the seminal work by Case and Shiller (1987, 1989). However, due to the long purchase-resale cycles of real estate, particularly owner-occupied mass-market housing, the repeated-sales method only uses a small portion of available transaction data, and suffers from significant sample selection bias (Clapp et al., 1991). Furthermore, this method assumes constant quality over time, when in fact infrastructural developments can dramatically change the locational value of houses, particularly in dense urban environments. Deterioration through wear-and-tear and renewal through upgrading and renovation works can also change the quality of individual homes, buildings, or neighborhoods.

The hedonic approach, in contrast, uses all available transaction data to estimate a model that prices each property based on its individual attributes. In that sense, a building can be seen as a "bundle of goods" (Monson, 2009), comprising its physical, locational, and accessibility characteristics. The transaction price is then determined by the sum of the individual values or implicit prices of these characteristics (Rosen, 1974). The hedonic pricing model enables the distinction between price changes arising from individual characteristics (such as the building of new urban mass transport) and external macroeconomic events or policy changes. Intuitively, hedonic methods come closest to how property valuations are made by realtors, particularly when spatial and temporal autoregressive features are included. Furthermore, Diewert et al. (2008) note that while matched-item models are often used to measure inflation across time periods, hedonic indices are more appropriate for "product areas with a high turnover of differentiated models," such as real estate.

This paper develops and applies hedonic modeling to a high-density, urban public housing resale market in Singapore. Approximately 80% of households live in high-rise public housing apartments, with a 92% home ownership rate (Singapore Department of Statistics, 2016). These apartments, built by the Housing and Development Board (HDB), come in several sizes of largely homogeneous layouts, and adhere to stringent quality standards. New flats are sold directly by the HDB at subsidized prices, with a set of restrictions including a five-year minimum occupation period, but can be resold on an open market thereafter.[1]

The scale and nature of Singapore's public housing scheme invites extensive hedonic analysis. The homogeneity in structure quality and internal layouts helps to reduce unobserved variations in quality between houses that cannot be captured in transactions datasets without physically

[1] Certain eligibility conditions still apply on the resale market. In particular, ownership is restricted to Singapore citizens and permanent residents (PRs), or families with at least one Singapore citizen. There are also ethnic and PR quotas for each apartment block to ensure social integration. Source: HDB.

inspecting each unit.[2] Meanwhile, the high-rise nature of apartments provides a large set of transactions data for each address, thus enabling fine-tuned calibration of geospatial preferences.

Using HDB resale transactions data from January 1, 2000 to July 30, 2016, and incorporating geospatial data, this paper explores three main objectives: first, I develop a time dummy hedonic pricing model for public housing resale transactions in Singapore, using spatial modeling to determine implicit prices for a set of locational and accessibility attributes. Building on this model, I then examine the wealth effects of transport infrastructure and accessibility, through major events such as the opening of new mass transit lines. Finally, I construct a hedonic price index, and discuss the impacts of broader policy shifts and public projects on housing price appreciation, wealth effects, and rent gradients. I believe that this type of hedonic modeling contributes meaningfully to existing literature, and can be further developed as a useful policy evaluation tool in analyzing both the overall wealth impacts and effects on inequality between districts.

The next section provides a review of hedonic methods as applied to urban economics, and existing research on Singapore's housing market. Section 3 builds a theoretical framework for my model and summarizes the datasets and methods used. Section 4 presents the regression results and analysis. Finally, Section 5 concludes with a discussion on the possible extensions and implications of the hedonic model.

[2] Other unobserved variations in quality between houses remain, such as interior decoration and upkeep.

## 2. Literature Review

*2.1 Hedonic modeling in urban economics*

Hedonic methods are widely used in real estate analysis today. However, there is large variation in the complexity and specifications of hedonic models across applications. Two main approaches exist for hedonic models: hedonic imputation indices and time dummy hedonic indices. Diewert et al. (2008) present a formal analysis of the difference between the two approaches. In essence, the time dummy method constrains the quality adjustment parameters to be constant over time, while the hedonic imputation approach allows these parameters to change in each period, and uses an average of two sets of quality adjustments for each comparison of prices between two periods. One popular method in hedonic price index literature is a special case of the time dummy approach that uses "rolling windows," whereby hedonic parameters are fixed and estimated over two adjacent periods at a time, in an overlapping manner. However, Rambaldi and Rao (2011) demonstrate that the rolling window approach is no better than a basic time dummy model with fixed hedonic parameters across the entire sample period, and that "secular trends in prices" (rather than movements in the hedonic coefficients) dominated movements in housing prices. Therefore, within a relatively short time frame in the housing market and barring major changes in the population's wealth levels, we can consider the time dummy approach as a suitable base model by assuming that the structural characteristics influencing housing resale prices remain relatively stable.

However, a simple OLS regression model of property values suffers from two main violations of basic OLS assumptions: heterogeneity and autocorrelation. The typical mantra of real estate purchase decisions is "location, location, location." It is intuitively obvious that spatial factors play a crucial role in determining housing prices. Differences in intrinsic value exist between locations, resulting in spatial heterogeneity. District or region-based dummies can account for these differences to some extent, but spatial heterogeneity is continuous and dependent on a range of spatial factors that differ in the scale in which they vary (Chew, 2011).[3] Moreover, in determining property valuations, realtors often reference the transaction prices of similar houses that were sold recently in the same area. Thus, we can expect housing prices to exhibit both spatial and temporal autocorrelation. Spatial econometrics has been developed as a field to address the issues of spatial dependence, through highly influential work by Cliff and Ord (1981) (see Pace et al., 2009 for a discussion). Various methods have been used to correct for these issues in housing markets, including spatial autoregressive (SAR) and geographically-weighted regression (GWR) methods (see Brunsdon et al., 1996; Can, 1992; Kissling and Carl, 2008; Wilhelmsson, 2002).

### *2.2 Current research in Singapore*

The official HDB resale price index in Singapore was calculated using a stratified simple average until the third quarter of 2014, after which a stratified hedonic regression method was used (Housing and Development Board, 2016). Existing research using hedonic modeling methods in Singapore mainly focuses on the impact of specific attributes on housing prices, such as the proximity to popular primary schools (Wong, 2008), ethnic preferences (Wong, 2013), and political boundaries (Sue and Wong, 2010). Muhammad et al. (2005) construct a hedonic automated valuation model for HDB flats, and conclude that floor area and age are the most significant variables, while distances to the CBD and the nearest subway station are also important factors.

Focusing on public transport accessibility, Chew (2011) analyzes the opening of one major subway line, and demonstrates significant forward-looking behavior of housing prices in anticipation of future improved connectivity. This is in line with international literature (for examples, see Agostini and Palmucci, 2008 on Santiago's metro line 4; Bae et al., 2003 on Seoul's subway line 5; and McMillen and McDonald, 2004 on Chicago's Midway Line). While much of existing work studying the effects of new transit lines focuses on proximity to transit stations, Jing

[3] For example, regional dummies could roughly reflect varying distances to the city center, but do not capture varying distances to the nearest subway station.

and Liao (2016) develop a broader measure of accessibility by constructing a connectivity index for each subway station, incorporating quality adjustments for individual subway lines. They then use a difference-in-difference model to show that new subway lines, which improve connectivity across existing lines, have positive wealth impacts on property values for houses located near existing stations across the entire subway network. However, they do not factor in distances between the house and the nearest subway station—instead using an 800-meter radius to define the boundary between houses "near" subway stations and houses in the "control" group. Nevertheless, this finding has notable implications in cost-benefit analyses of government-funded infrastructural investments for subway system expansions.

In their work examining the private condominium market, Sun et al. (2005) demonstrate the existence of both spatial and temporal autocorrelation, which results in OLS estimates being inefficient. They develop two-order spatio-temporal autoregressive models that separately capture building and neighborhood effects, and show that a Bayesian estimation procedure can detect and correct heteroscedasticity to produce more robust coefficients. Van Eggermond et al. (2011) analyze both private and HDB resale and rental markets using OLS, SAR, and GWR methods. In line with other literature, they find that floor area and the distance to CBD are the most important factors, while floor level has an approximately linear effect.

In a broader market analysis, Ong et al. (2003) compare a hedonic price index to the published HDB resale (average) price index for transactions from 1997 to 2000. They examine trends in the quality of flats transacted and parameters of price determinants across positive and negative phases of the market cycle. Finally, in a more conceptual work, Jiang et al. (2015) develop a hybrid method for constructing housing price indices, using hedonic information and pairing of repeat sales at the building level (instead of the individual house level), to overcome the sample selection bias in traditional repeated-sales methods while controlling for unobserved hedonic information by using a semiparametric specification.

In this paper, I extend the broad market analysis of the type conducted by Ong et al. (2003) to the time period from 2000 to 2016. In contrast to previous studies, I attempt to build a more fine-tuned spatial model—in part enabled by the large dataset and time period studied. To account for the opening of multiple major new subway lines, I incorporate separate proximity and time effects in modeling forward-looking behavior of property values—thus extending Chew (2011)'s approach in both depth and scope. Finally, a comprehensive hedonic model allows for unbiased, district-level comparisons of quality-adjusted prices. This reveals changes in rent gradients and disparities in the rate of wealth appreciation over time. Through these discussions, I hope to demonstrate the role of hedonic models in public policy evaluation, including cost-benefit analyses for large public infrastructure projects (such as subway lines) and welfare analyses of public housing policies.

## 3. Empirical Design

*3.1 Theoretical framework*

The basic hedonic approach, based on influential work by Rosen (1974), models the price of a house ($P$) as a function of its individual attributes (such as age, floor level, distance to the city):

$$P = f(x_1, \dots, x_k) \tag{1}$$

Adapting from Ong et al. (2003) and Ye (2016), I use a time dummy hedonic price model, which includes a dummy variable for each time period $t$ belonging to the set of all time periods $\{1, \dots, T\}$. Thus, a linear regression model normalized to the first time period has the functional form

$$P_{it} = \boldsymbol{\delta}_0 + \sum_{t=2}^{T} \boldsymbol{\delta}_t D_t + \sum_{j=1}^{k} \boldsymbol{\beta}_j X_{ijt} + \varepsilon_{it} \tag{2}$$

where $P_{it}$ denotes the transaction time of house $i$ made in the time period $t$, $\boldsymbol{\delta}_t$ is the vector of estimated coefficients for each of the time dummies $D_t$, and $\boldsymbol{\beta}_j$ is the vector of estimated coefficients for each of the $k$ quality variables of transaction $i$ denoted by the vector $X_{ijt}$. Consequentially, if all quality variables are perfectly adjusted for, then the estimated values of $\boldsymbol{\delta}_t$ represent the time period effects and can be normalized into a price index for all periods.

It is important to address here two major criticisms of hedonic regression models. First, the heterogeneous nature of housing markets results in heteroscedasticity among hedonic residuals, as shown in previous work (Chew, 2011; Sun et al., 2005). Using log and other power transformations on both price and certain independent variables (such as distances and travel times) will reduce the severity of heteroscedasticity.

Second, due to the complex nature of property values, the large number of attributes, and the presence of unobserved quality factors that contribute to prices, it is impossible to specify a functional form *a priori,* based on theoretical models. The choice of any particular functional form or transformation has the potential of introducing model specification bias (Jiang et al., 2015).

According to Linneman (1980), when varying functional forms, estimations are “substantially more sensitive to changes in the specification of the dependent variable than to changes in the specification of independent variables.” Thus, I adopt an approach of specifying linear, logarithmic, or higher-order forms for independent variables, while testing different specifications of the dependent variable to find the most optimal power transformation.

Section 3.2 discusses the set of resale transactions used for my hedonic model analysis, and the physical quality adjustment variables used. The subsequent subsections detail the geospatial data used to model various locational and accessibility attributes, in order to account for spatial heterogeneity between different locations.

*3.2 Data summary*

The raw dataset, provided by the HDB, contains data for all resale HDB flat transactions between January 1, 2000 and July 31, 2016, totaling 450,872 entries.[4] Each record includes information on the month and year of transaction, the town where the apartment is located, the layout type (number of bedrooms) and flat model, street address, floor range (accurate to either a 3- or 5-floor range), floor area, the year of lease commencement, and the resale price in nominal Singapore dollars.[5] An indicative floor level is assigned to each transaction using the median value of the floor range. The date of lease commencement is usually up to one year after the construction is completed and keys are collected, thus, the age of the apartment at the time of transaction is approximated as *age = year of sale – lease_commence_date + 1*. A summary and description of these variables is presented in **Appendix 1**.

[4] The dataset was downloaded from the Government of Singapore’s open data portal at: https://data.gov.sg.

[5] All HDB apartments are built and sold with 99-year lease terms, after which they are returned to the government. Therefore, the year of lease commencement refers to the time which the apartment was initially built. Unlike privately-developed properties, where the lease commences from the time the developer acquires the land, the legal lease commencement date for HDB apartments occurs only after construction has completed.

Influenced by existing literature, the floor level is entered into the regression equation linearly, while the floor area is log-transformed. We can expect higher-order effects of apartment age, as it influences property values in multiple ways. Older apartments are generally valued less due to deterioration and shorter remaining leases. However, beyond certain thresholds, older estates are also eligible for various government-subsidized upgrading programs that make them more desirable. There have also been policy changes over time that have improved the quality of apartments at certain points in time, for example, the introduction and retrofitting of elevators that stop at every floor. Rigorous analysis by Ye & Becker (2016) for the subsidized apartments market in Hong Kong has shown significant effects of higher order terms for age up to the $6^{th}$ power. Thus, I test for the significance of multiple orders of the age variable. Finally, flat type (which corresponds to the number of bedrooms) and town (i.e. district) are also included using dummy variables.

While overall inflation is an important factor in determining the actual wealth effects that arise from changes in property prices, adjusting transaction prices using inflation indices is a blunt method of correction, and the results are not easy to interpret due to an endogeneity problem. With housing being one of the biggest components of household expenditure, property prices have a strong influence on the reported inflation rates, such as the Consumer Price Index—particularly in Singapore, where the high population density means that property is rather expensive. This problem may be especially pertinent when trying to capture the wealth effects of major public projects that have the potential to influence property prices on a wide scale, such as the construction a new subway line. The use of nominal prices and individual time period dummies in my model represents a more fine-tuned approach, as the time dummies both measure and control for property market-specific price inflation.

### *3.3 Monocentric city model*

Anecdotal evidence and existing literature suggest that Singapore can be largely characterized by a monocentric city model with a single central business district (CBD). Combining mobile phone geolocation and public transit farecard data, Poonawala et al. (2016) show a dominant flow of traffic into and out of the Downtown Core during the weekday morning and evening peak hours, respectively. While the Singapore government has actively developed regional centers outside the Downtown Core, these are still in the early stages of development. Schläpfer (2014) demonstrates that the regional centers had a substantially smaller catchment area for visitors compared to the Downtown Core region, thus demonstrating significant monocentricity.

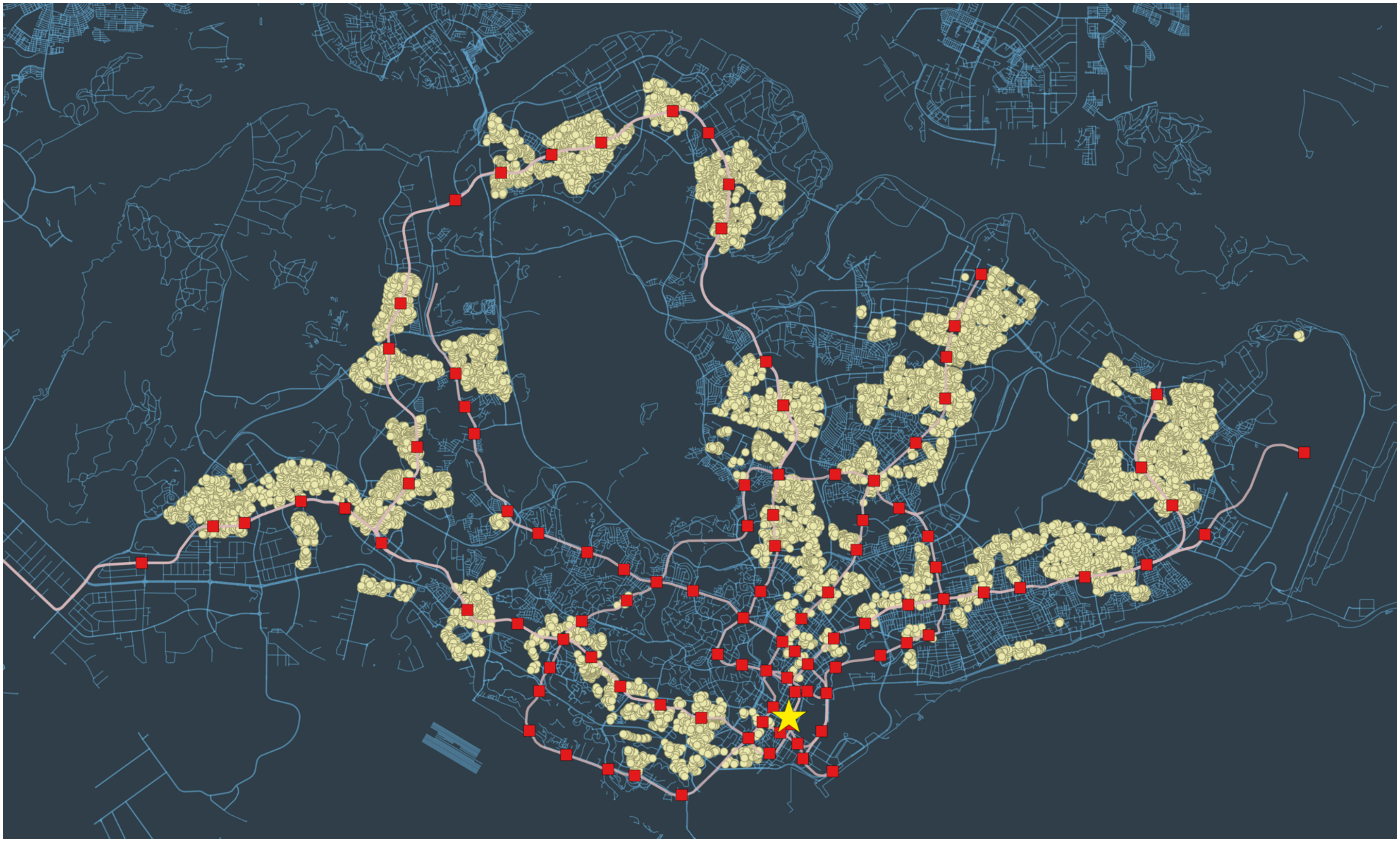
In such a monocentric city model, bid rents for land or real estate are expected to decrease as the distance (and correspondingly travel time) from the CBD increases, thus forming a rent gradient for housing prices. In other words, there will be a preference for centrally located houses that are closer to the CBD, as they provide greater accessibility to a wide range of destinations, including workplaces, shopping malls, restaurants, and points of interest. I adopt the monocentric city model in my spatial modeling by using a single CBD as the reference point in calculating distance and travel time parameters. In line with previous literature, the CBD is defined at Raffles Place Park, in the Downtown Core region (Deng et al., 2012; see **Figure 1** for an illustration).

The hedonic regression model captures this preference for centrally located houses by including variables that measure accessibility to the CBD from each location. From the full dataset of resale transactions, 8,607 unique addresses were extracted. The Bing Maps Representational State Transfer Application Programming Interface (API) was used to geocode each address, and calculate the driving distance to a central point in the CBD. While traffic congestion and actual driving times would also influence the perceived accessibility of each location, these are much more difficult to incorporate for a dataset that spans multiple years, as the level of congestion fluctuates over time as new towns are built and roads are widened. There is generally heavy peak-hour traffic but not crippling congestion across all parts of Singapore, thus, the effects of congestion can be expected to apply relatively evenly across all distances.[6] All distances are natural log-transformed in the regression equation.

---

[6] In part contributing to the mild congestion in Singapore is variable road pricing implemented by the government, where road tolls are regularly adjusted to maintain target average speeds on major roads and expressways. Given further historical data, the impacts of these changes in road pricing on property prices may be an interesting topic to explore in further research.

**Figure 1.** Location of MRT stations and HDB towns in Singapore.



**Legend:** Red boxes indicate MRT station locations, and light yellow dots plot the addresses of HDB flats. The yellow star (near the bottom, middle) locates the central CBD in the Downtown Core. Light blue lines indicate roads and territorial boundaries.

*3.4 Modeling public transit*

Less than half of Singaporean households own cars, as the government maintains high taxes and quota controls over the vehicle population, owing to the limited land available and a desire to avoid crippling traffic congestion. Thus, accessibility by public transit features heavily in the average Singaporean's daily commute, and, naturally, in determining the value of an HDB apartment. The subway system, known as the mass rapid transit (MRT), forms the backbone of the public transport system.[7] Since the first line opening in 1984, the government has continually invested in expanding the MRT network and planning far ahead into the future.[8] Because of the long lead times of MRT line construction, intuitively and empirically, housing prices are likely to exhibit forward-looking behavior in anticipation of increased connectivity afforded by future MRT station openings (Chew, 2011). McMillen and McDonald (2004) provide a simple theoretical basis for increases in home values ahead of station opening, due to expected increases in rent discounted to present prices. However, the largely owner-occupied nature of HDB housing suggests that price changes are more closely linked to the expected standard of living—which improves when new stations are built, but is also negatively affected during the construction phase. This may lead to nonlinear price effects leading up to new station opening (as seen in Chew, 2011).

Generally, HDB towns are built around MRT stations as the central transport node (see **Figure 1**). Thus, the straight-line distance to the nearest MRT station is used as one measure of public transit accessibility. This is calculated using MRT station locations obtained from OpenStreetMap and the Land Transport Authority.[9] First, all current and future stations are plotted on a map, and

---

[7] The average daily ridership on the MRT network was over 2.7 million trips per day in 2014. Source: Land Transport Authority.

[8] For example, the Northeast Line, the third line to be built, was opened in 2003. Initial planning for the line began in 1986, and final details for the line were announced in 1996. As of 2016, the government has announced plans for three new lines in addition to the current five, which will be completed over the next 15 years.

[9] The OpenStreetMap data more accurately reflected the locations of ground-level exits at each station, and included both current and future stations. However, it had one missing station, which was supplemented by data from the Land Transport Authority.

the nearest station is determined for each of the 8,607 unique addresses in the dataset. For instances where the nearest station was opened after January 2000 (i.e. earliest time period of the dataset), the most recently opened set of stations was removed, and the next nearest station that was opened at an earlier date was determined. This process was repeated for addresses that experienced successive station opening events. In other words, for each address in the dataset, the nearest MRT station at any point in time from 2000 to 2024 (when the last of the currently announced stations will open) was determined.

The distance to the nearest existing MRT station at the time of transaction is entered into the regression equation for all transactions. For houses where a future MRT station opening will reduce this distance, we can expect that its influence on price is relative to (i) *how much* nearer the future station is, and (ii) *how long* it is expected to take until the station opening, i.e. time discounting. There is a case to be made for a third effect: instantaneous price shocks at the time of announcement. However, plans for all new MRT lines have been announced at least 12 years in advance, and preliminary plans are usually shared with the public before station locations are finalized. Therefore, it is difficult to specify how and when an instantaneous price shock should occur. Instead, I assume that the combination of information diffusion and future discounting effects result in a gradually increasing price effect as the time interval to future station opening draws nearer, with an effect size of zero at an upper bound time interval of 12 years. Thus, in the regression equation, the effect is entered as an interaction term $\log(\varDelta d + 1) \times (144 - t)$, where $\Delta d$ is the distance reduction as a result of a future MRT station opening event (i.e. difference between distance to current nearest station and distance to future nearest station), and $t$ is the time interval (in months) between the transaction date and the expected station opening date.[10] Higher order terms for time interval are also tested to allow for nonlinear trends.

[10] The linear shift $\varDelta d + 1$ is used in the log-transformed term to avoid negative values, which would otherwise occur where $\varDelta d < 1$.

*3.5 Other geospatial qualities*

Enrollment in primary schools is performed in phases based on the place of residence, with priority given to children living within 1km of the school, followed by those living between 1–2km, and thereafter all other children.[11] There is much anecdotal evidence to suggest that parents exhibit a strong desire to enroll their children in top-performing primary schools, and the list of popular schools that tend to be oversubscribed during registration remains rather stable over successive years. Since the enrollment priority and ballot procedures are arbitrarily and uniformly defined, we can expect a discontinuous change in admission probability to a particular school across the 1km and 2km perimeters (Wong, 2008). A total of 29 (out of about 187) primary schools were identified as highly popular, based on their frequency of oversubscription during the annual Primary One Registration Exercise.[12] Popular primary schools were identified as schools that were frequently oversubscribed in Phase 2B of the Primary One Registration Exercise from 2006–2015. This phase is open to parents who have served a minimum of 40 hours as volunteers at the school, and precedes Phase 2C, which is the general registration phase for all Singapore citizens and permanent residents. I use separate dummy variables to identify houses that are located within 1km or between 1–2km from popular primary schools.

As a measure of local accessibility by driving, the distance to the nearest expressway or semi-expressway is included with a natural log transformation. Semi-expressways refer to arterial roads

---

[11] There are also certain priority schemes in place for children with siblings already studying in the school, or whose parents are alumni of the school.

[12] The total number of primary schools varies slightly from year to year due to mergers and new school openings (depending on the population of eligible students and building of new towns), but these events tend to affect the less popular schools—the list of popular schools is dominated by well-established schools with a rich and long history. The historical oversubscription data is compiled by a group of parents, from statistics released by the Ministry of Education over the time period.

Source: https://www.kiasuparents.com/kiasu/article/2016-p1-registration-oversubscription-risk/.

that are part of the outer ring road system, or the West Coast Highway.[13] In addition, I use the (natural log transformed) distances to the nearest hawker center and nearest park as measures of neighborhood amenities.[14] A summary and description of all variables included in the regression model is in **Appendix 2**.

### *3.6 A note on demographic attributes*

Variations in demographic attributes, such as household income and racial composition, across geographical space can also influence housing prices. For example, influential work by Courant (1978) demonstrates that racial prejudice in housing transactions cause market segmentation by race in a long-run equilibrium. While Singapore is a multi-racial society, albeit predominantly Chinese (comprising about 74% of the resident population), the presence of racial prejudice cannot be ruled out. However, the government has maintained an Ethnic Integration Policy since 1989, whereby all HDB flats, including both new sales and resale, are subject to ethnic quotas proportionate to the overall population's composition. Previous work by Wong (2013) shows that the policy has been largely successful in desegregating historically ethnically-concentrated neighborhoods, and estimates inverted U-shape preferences for neighborhood-level ethnic concentrations in line with the overall population mix. Thus, it is reasonable to assume an even racial distribution across neighborhoods.

Given the lack of variation in factors like race, security, and building and neighborhood quality across neighborhoods, it is also likely that the distribution of household income is highly correlated with other quality-adjustment variables. Thus, segregation by income levels occurs naturally as a result of price differences, and not due to inherent demographic differences across neighborhoods.

---

[13] These roads have been upgraded with viaducts, flyovers, and underpasses to allow smoother traffic flow and bypassing of intersections. In particular, the outer ring road system connects major arterial roads, expressways, and towns on the city fringe.

[14] Hawker centers are popular local cooked food centers, with a large collection of hawker stalls under one roof. They are generally co-located with wet markets in residential estates.

# 4. Data Analysis

*4.1 Regression model and diagnostics*

A regression model with a total of 248 variables is used to create the hedonic price index, including 198 time-period dummies, 25 district-specific dummies, and 25 quality-adjustment variables (see **Appendix 2** for full regression inputs and results). Higher-order terms for apartment age are significant and included up to the 7th power. Higher-order terms for the time interval to future MRT station opening are included up to the 6th power in the interaction terms with station proximity reduction.[15] A 1/3rd power transformation for the dependent variable, resale price, is selected for producing the best linear fit, and resulting in lower heteroscedasticity among residuals as compared to the natural log transformation commonly used in existing literature.

Outlier transactions and apartment types were identified and excluded. HDB regulations stipulate a five-year minimum occupation period, during which houses can only be sold under exceptional circumstances with approval from HDB on a case-by-case basis.[16] Thus, transactions where the age of the flat is less than four years can be considered as distressed sales, with an extremely small sample size of 431 observations (0.096%). These were removed from the dataset.[17] Moreover, 1-room (studio) and 2-room flats are generally bought by a different target group, such as the elderly after retirement or very low-income households.[18] These added up to another 4,799 transactions which were removed from the dataset.

---

[15] The 9th order age term and the 8th order of time interval interaction term exhibited high collinearity and could not be entered into the regression. The inclusion of the 8th order of age and 7th order of time interval resulted in inflated standard errors in the lower-order terms.

[16] This usually occurs when, for example, a couple has divorced and neither party is eligible to retain the flat (due to HDB regulations that require a family nucleus in order to buy a new flat). Thus, transactions that occur within the minimum occupation period are often distressed sales that may not reflect market prices.

[17] The inclusion of these distressed sales had no discernible impact on the quality-adjustment parameters. The sample size was also too small to determine conclusive trends for apartments aged between 0–4 years.

[18] This is evidenced by the different eligibility criteria for buying these flats, as well as the facilities provisioned in new 2-room flats, such as elderly-friendly bathrooms and emergency pull-cords (1-room flats are no longer being built).

Upon examination of the residuals resulting from a regression of the remaining sample, two apartment models ("terrace" and "adjoined flat") dominated the extreme outliers observed, with residuals greater than 4 standard deviations from the predicted mean.[19] These added up to 1,265 transactions and were excluded from further analysis. Thus, a total of 6,495 out of 450,872 transactions (1.44%) were removed, leaving 444,377 observations for the following analysis.

The regression model produces an $R^2$ value of 0.9477 (adjusted $R^2$ = 0.9477), suggesting that it is able to explain most of the observed variation in resale prices. The root mean square error (RMSE) value is 2.1197. In comparison, the standard deviation of price (adjusted with the 1/3rd power transformation) is 9.2704, and a model with only time period and district dummies yields an $R^2$ of 0.5155 and an RMSE of 6.4544. Bootstrapping (with 50 replications) is used to calculate standard errors. The estimated coefficients on the regression variables are all significant at the 99.9% confidence level, with the exception of six time-period dummies and the dummy variable for being located 1–2km from a popular primary school. To test for over-fitting, a 20-fold cross-validation test was performed, which produced consistent RMSE values and estimated coefficients, suggesting that the model specification is robust and has high predictive power for out-of-sample observations (see **Appendix 3**). Additionally, three shrinkage statistics that measure overfitting as well as out- and in-sample predictive bias generated values of 0.00 when calculated using ten iterations of five-fold cross-validation (Bilger and Manning, 2015).

Estimated coefficients on most of the quality-adjustment variables, including floor area, floor height, and flat type are of the expected signs. Different combinations of spatial quality-adjustment variables are used to check for robustness of the estimated coefficients. Distances to the CBD, nearest MRT station, hawker center, and park, as well as the dummy for being within 1km of a popular primary school, are of the expected signs and robust to different model specifications. The

[19] The observation that these apartment models are outliers make intuitive sense. "Terrace" units were built for a short period of time prior to 1960, by the predecessor of HDB, and resemble private, landed, semi-detached houses. "Adjoined flat" refers to adjacent apartment-style units that have been converted into a single large unit, by removing the dividing wall. These two apartment models are rare, and comprise 0.28% of the resale transactions observed over the entire time period studied.

dummy for being located 1–2km from a popular primary school is sensitive to model specification and is insignificant in most models. **Appendix 2** lists the full regression results for several model specifications tested.

Influenced by previous work, I also test an alternative functional form for accessibility-related variables, using a distance-decay parameter $e^{-\lambda d}$ with $\lambda$ calibrated between 0.04 to 0.2 (Jing and Liao, 2016). This transformation did not substantially influence the magnitude and trends of the estimates as presented in the following sections.

Some heteroscedasticity is observed in a bias towards positive residuals for apartments with resale prices above \$600,000, and towards negative residuals for apartments below \$200,000 (all prices in Singapore Dollars). A plausible explanation for this could be that high- and low-priced apartments are more likely to have building- or apartment-specific characteristics contributing to their prices that are not captured in the regression model.[20] Residual plots against the explanatory variables, including time period, apartment age, district, floor area, and distance to the CBD, show no significant biases across the range of each variable. **Appendix 3** details the residual plots and outcomes of the cross-validation test. In general, the hedonic model is a consistent and robust predictor of apartment resale prices across different kinds of apartments and locations.

*4.2 Implicit prices of quality-adjustment attributes*

To enable a more intuitive understanding of the influence of specific quality-adjustment variables, the following analysis expresses the partial effect of any particular variable on the resale price in percentage terms, evaluated at the average transaction price of S\$313,198.60 (approximately USD 224,380 as of March 2017). Given the linear model specification, the absolute magnitude of the effect size of any single variable (on the adjusted price) is fixed across the entire

[20] For example, low-priced apartments could suffer from poor upkeep, while high-priced apartments could come with quality furnishing and/or a desirable view (which could depend on which way a unit faces within an apartment block as well as its floor height).

price range of apartments—therefore, the effect size in percentage terms naturally varies across each transaction. Nevertheless, this way of evaluating effects at the average price provides a straightforward illustration of the influence of particular quality-adjustment variables for a "typical" apartment transacted.

Using the method described above, every one-story increase in floor level has an effect size of approximately $2,418, or 0.77% of the mean transaction price. This is in line with the range for story-related price variation of new flats sold directly by the HDB, based on published prices.[21] Being located within 1km of a popular primary school has a significant and positive influence on resale price, with an effect size of $3,906 or 1.25% at the mean transaction price. There is an insignificant result for houses located between the 1–2km boundaries, at 0.01% of the overall transaction price.[22] Interestingly, there is a positive effect on resale price associated with being located further away from an expressway or semi-expressway.[23] This could be attributed to expressways generally running along the outskirts of each HDB town and the low overall reliance on personal cars as a form of transport. Furthermore, proximity to a busy expressway could yield disutility in the form of noise pollution. This estimate is in line with previous literature (Ong et al., 2003).

Apartment age has a large effect on resale price, as shown in **Figure 2**. An apartment that is 50 years old can be expected to sell for $99,925 less than a similar apartment that is four years old, or an effect size of approximately 32% when evaluated at the average transaction price. This large effect is reasonable when considering the fact that all HDB flats are sold with 99-year lease terms—thus, a 50-year old apartment would have less than half of its lease remaining.

---

[21] The sale prices of new flats currently available for selection at any point in time is published on the HDB's website, under various sales launches.

[22] In contrast, in an analysis of houses located near the 1km and 2km perimeters, Wong (2008) found that being located within 1km and between 1–2km of a good performance primary school increased resale prices by 1.9% and 1.3%, respectively.

[23] One might suspect that the positive coefficient on the distance to expressways could be influenced by collinearity with distance to MRT stations (which tend to be located toward the center of HDB towns). However, this is not the case in our dataset.

**Figure 2.** Influence of apartment age on resale price (effect = 0% at median age of 19 years).


There is a plateau and slight reversal of the overall downward trend, for apartments between 20 to 40 years old. This could be explained by positive utility associated with being in mature estates, where these older apartments are located.[24] While proximity to central transport nodes are controlled for via the distance to the nearest MRT station term, other types of amenities tend to be more developed and evenly spread across mature estates—for example, childcare centers, supermarkets, hawker centers, and neighborhood malls and restaurants. Though it is possible to attempt to control for all of these factors, the list of amenities that people take into consideration when buying a resale apartment may be large and varied. This could be studied as an extension of the current model; however, it must be noted that including all these variables in the regression equation may introduce instability in the estimation results. Given the Singapore government's carefully planned development and expansion of all HDB towns over time, it may be reasonable to assume that the overall level of amenities is captured in age effects. This is supported by the observation that estimated coefficients on district-specific dummies are lower across the board for non-mature estates than those for mature estates. For apartments older than 40 years, the government has imposed restrictions on the usage of money from the Central Provident Fund

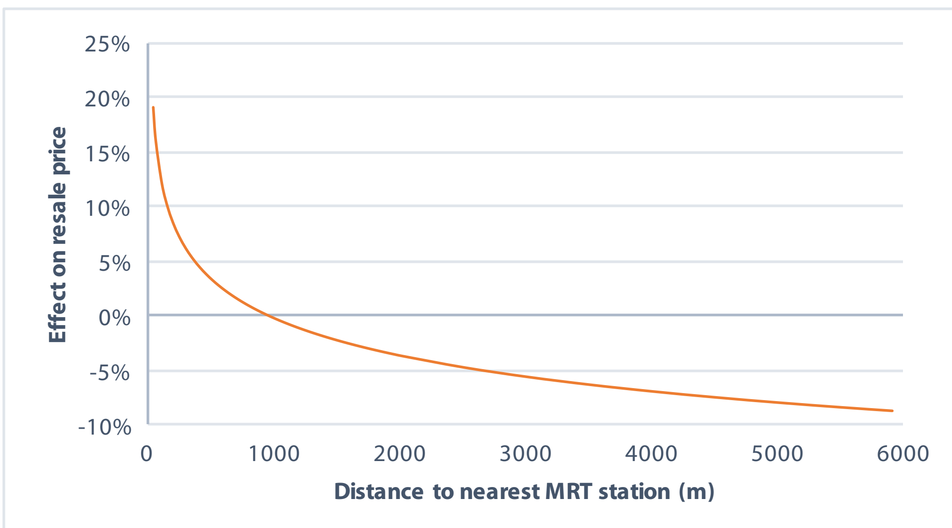
[24] According to HDB's classification, mature estates are those that are more than 20 years old.

(CPF, a national mandatory savings account) to finance such transactions, likely contributing to the observed decline in prices.

### *4.3 Public transit characteristics*

**Figure 3** shows the existing station effects on resale price, based on the logarithmic form specified for distance. There is a sharp drop in resale price as the distance to the nearest MRT station increases from 30m (the minimum distance in our dataset) to 1km, with an effect size of 20% of the mean transaction price. The effect is much more gradual as this distance increases from 1km to 6km (the upper bound distance in our dataset), with each additional kilometer increase beyond 2km contributing to a smaller than 2.5% effect on transaction price.

**Figure 3.** Influence of proximity to existing MRT station on resale price.




The large value associated with proximity to MRT stations is in line with the heavy reliance on public transport among Singapore residents. Being located 1km away from an MRT station would therefore require people to take feeder bus services before transferring into the MRT network. Given Singapore's compact size and fast journeys afforded by the MRT system, this could significantly increase total travel times. It is also reasonable to believe that households located more than 2 or 3km away from an MRT station are likely to rely on other modes of transport, such as public buses—thus, the price elasticity of an additional 1km distance to an MRT station would be much smaller.

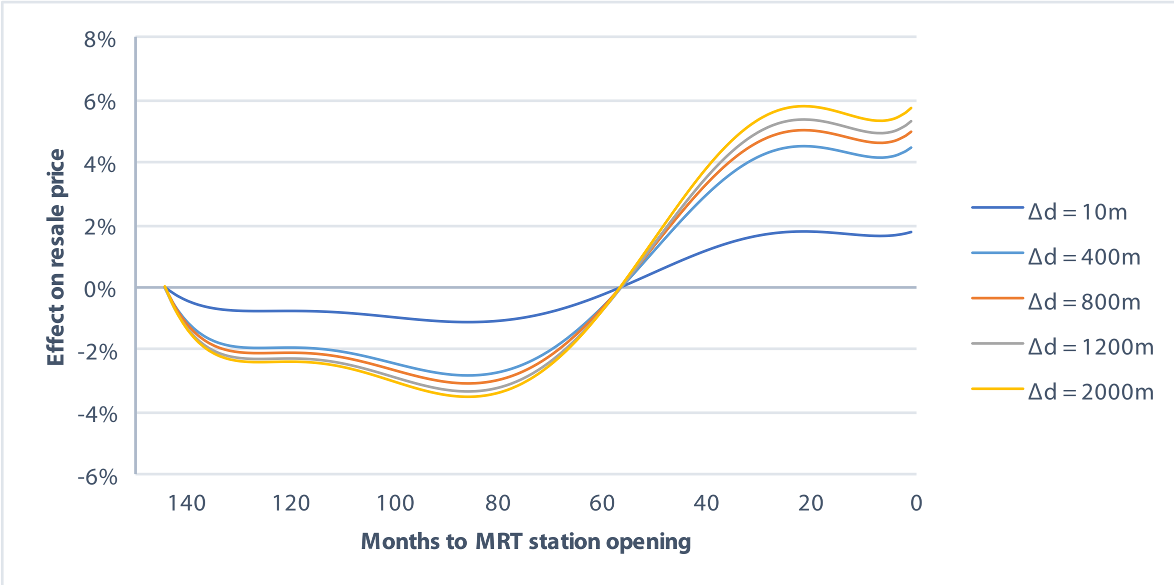
In total, 80,554 transactions occurred where the apartment experienced a future MRT station opening event within 144 months of the transaction date. The estimated price effects of a future MRT station depend on both $\Delta d$, how much nearer the future station is compared to the nearest existing station, and $t$, the time interval between the resale transaction and future station opening. **Figure 4** plots the price effects from 144 months to 1 month prior to station opening, for six illustrative values of $\Delta d$ (all within the range of the dataset).[25] There is significant forward-looking behavior in resale prices, which extends to about 4 years in advance of an MRT station opening. This effect increases as the station opening draws nearer, and reaches a peak at about 2 years prior to the opening. This is similar to the effect found in a previous study of non-landed private residential properties located along the Northeast Line, which opened in 2003. As Chew (2011) suggested in that study, the price increase can be attributed to a combination of anticipatory effects of improved connectivity and a decrease in construction disturbances.[26]

There is a significant positive effect (of nearly 2% of the mean resale price) even when the estimated reduction in distance afforded by a new MRT station is only 10m. While the difference in walking distance to the new station would be negligible, the price difference likely reflects the increased connectivity provided by the opening of a second MRT station nearby—providing direct access to another MRT line.

[25] The mean $\Delta d$ in the dataset is 1,038m, with a median of 746m.

[26] It is reasonable to expect that heavy construction concludes two years prior to station opening. The remaining two years would be required for installing interior fittings, as well as extensive system testing.

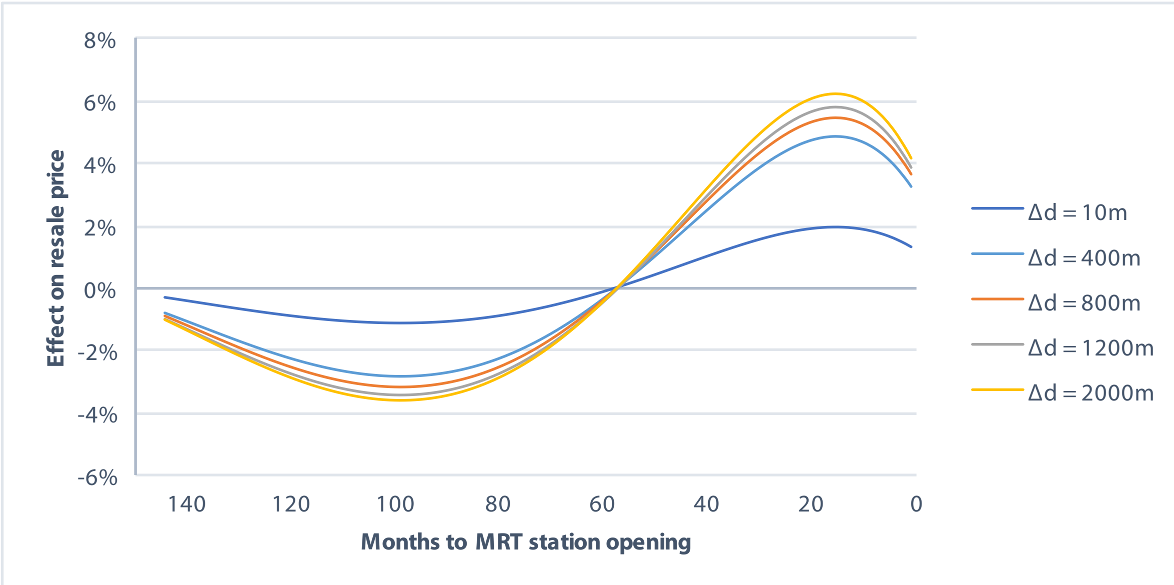
**Figure 4.** Future MRT station effects on resale price.[27]




Notably, the regression model predicts a negative influence on price when the station opening is between 5–12 years into the future. This can at least partially be explained by construction disturbances, which can include noise, dust, repeated re-alignment of roads, and traffic congestion. Indeed, construction for new MRT lines usually commence as much as seven to ten years before the line becomes operational. The initial stages of construction also involve the most disruptive types of work, such as drilling and tunneling. Since resale buyers take possession of their new house immediately, but will not experience the benefits of a new MRT station until many years later, the disutility from construction disturbances likely outweighs the (discounted) utility of future improved connectivity. It may also be difficult for home buyers to ascertain if they will stay at the same house long enough to reap the benefits of the new MRT station. Another explanation could be a readjustment of prices to their base levels, following an initial spike after the official announcement of new lines and station locations, which is generally made at approximately or slightly more than 12 years prior to station opening.

[27] All the lines cross zero (on the vertical axis) at the same point since I use a single equation to estimate the anticipatory price effect as an interaction between $\Delta d$ and $t$ (and higher order terms of $t$). Therefore, when simulated values of $\Delta d$ are chosen to generate the graph, they all cross zero at the value of $t$ where the products of estimated coefficients and the respective terms of $t$ cross from negative to positive.

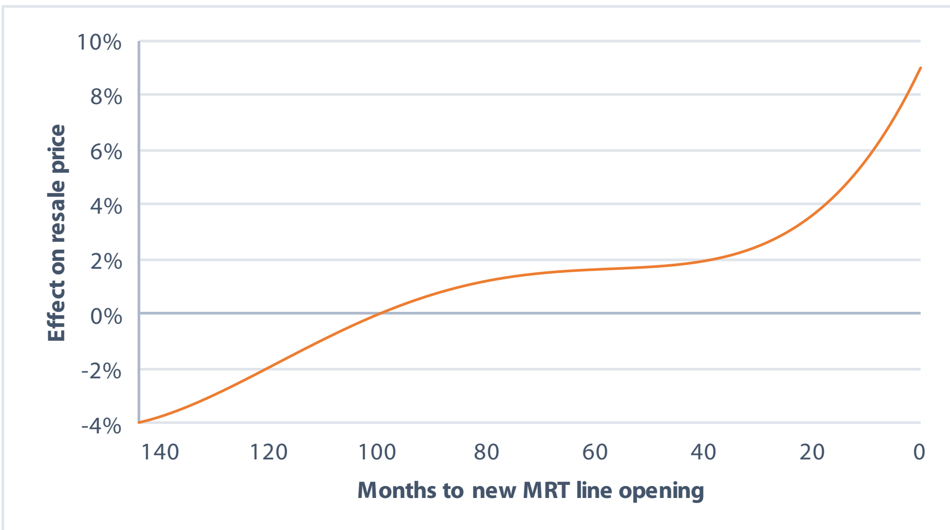
**Figure 5.** Future MRT station effects with alternative model specification (upper bound time interval = 180 months).




One could suspect that the chosen model specification, which arbitrarily limits the effect size to zero at 12 years before station opening, could have influenced the estimation outcome. To verify the results, I test alternative model specifications, with upper-bound time intervals of 15 and 24 years. Both models produce similar trends in estimation outcomes (see **Figure 5**).

Finally, an important limitation of the model is that it only measures changes in distance to the nearest MRT station; thus, it does not account for new lines that open at an existing station, which can substantially improve connectivity. I test for this "connectivity effect" in an alternative specification, where I include terms for the time interval to a new MRT line opening at the existing nearest station. This simplified test finds substantial positive price effects leading up to an existing MRT station gaining a new line (**Figure 6**), which can be explained as follows: the addition of a new line implies the conversion of a station into an interchange station, resulting in a substantial increase of traffic, therefore attracting new commercial developments or rejuvenation of existing developments in the area.[28] This subsequently raises the amenity level of nearby residential neighborhoods, and contributes to housing price appreciation. In an alternative test, I also find (somewhat smaller) positive price effects on houses where the nearest MRT station is adjacent to a new interchange station. However, these results (and particularly the numerical estimates) should be viewed as a preliminary exploration, as my models do not fully incorporate improvements to connectivity *after* a new line has opened (see Section 5 for a more detailed discussion).

**Figure 6.** "Connectivity effect" of a future MRT line opening through an existing station.




The significant wealth effects associated with proximity to both current and future MRT stations have important implications for public policy decisions. Given more data on the number of housing units at each location, the full capitalized value of government investments in mass public transit infrastructure can be estimated—which contributes to cost-benefit evaluations in justifying such investments. Anticipated future wealth effects arising from subsidized public

[28] Major suburban malls have opened next to MRT stations that became interchange stations with the addition of a new line: for example, NEX opened in November 2010 following the opening of the Circle Line through Serangoon station in May 2009; the Star Vista opened in September 2012, after the Circle Line through Buona Vista station commenced operations in October 2011.

housing ownership schemes are also relevant in equity-focused policymaking.[29] For example, new flats sold directly by the HDB are priced according to market factors (with a discount applied). However, the majority of these sales occur under the "Build-to-Order" scheme, where flats are priced and sold before they are built, with an approximately 5-year lead time to completion. If that time interval coincides with the major construction phase of a future MRT line, for example, buyers may experience large positive wealth effects immediately upon receiving their flats. The 5-year minimum occupation period for new flats further increase the potential for inequality-inducing wealth effects. Even if a new MRT station is planned for completion 12 years later (from the time of pricing and sale of a new HDB flat), it would be just two years to the MRT station opening by the time the flat can be put up for resale—and the buyer would experience the full wealth effects, as predicted by the hedonic regression model. Given that future MRT stations are announced at least 12 years in advance (and planned much earlier), it is entirely possible to adjust the pricing of new HDB flats to account for future MRT connectivity.[30]

### *4.4 Hedonic price index*

A hedonic price index is constructed by using the estimated values of each hedonic variable to predict the transaction price of a "typical" house in each period. This "typical" house is assigned the average value observed in the dataset of each quality-adjustment characteristic.[31] For comparability with the official resale price index published by the HDB, I calculate the hedonic price index quarterly using an average of the predicted price over the three time periods in each

---

[29] The Singapore government has revealed that equity is an important consideration in pricing public housing projects, as it has announced that it is considering imposing stricter rules on flats located in the central region, which have seen massive capital gains over the past decade.

[30] One caveat is that HDB prices have always been set at a discount relative to market prices of comparable resale flats in the area. Thus, it is also important to consider public reaction if prices are adjusted to equalize future wealth gains.

[31] Thus, the "typical" house has approximately the following characteristics: 19 years old, "4-room" layout (3 bedrooms, with a living and dining room), floor area of 97sqm, located on the 8th floor, 18km from the CBD by driving, 964m from the nearest MRT station, 1.08km from the nearest highway, 1.28km from the nearest hawker center, and 367m from the nearest park. It is not affected by future MRT lines.

quarter, with the base period set to the first quarter of 2009.[32] Both indices are shown in **Figure 7**. The Singapore property market experienced rapid price appreciation from the beginning of 2007 until the first half of 2013, when the government imposed a slew of property cooling measures. Notably, the strong property market was mostly unaffected by the global financial crisis of 2007–09, except for a slight weakening in the first quarter of 2009.

On first examination, the hedonic price index appears to trend below HDB's average-price index for much of the first two-thirds of the period studied. However, the first quarter of 2009 may not be the best base period to use, as it happens to be a point of divergence between the two indices. **Figure 8** shows an adjusted hedonic price index, which has been scaled to match the HDB price index in the first period, with shaded bars (in light blue, measured on the right-side axis) showing the difference between the two indices (numerical values can be found in **Appendix 4**). With this adjustment, the two indices track each other closely for the first half of the period studied, until the first quarter of 2008. The hedonic index does not exhibit the brief but noticeable dip that is reflected in the average-price index in the first quarter of 2009. This suggests that the dip was influenced by lower-quality sales, possibly linked to the effects of the global financial crisis. In the subsequent growth period, up until its peak in mid-2013, the hedonic price index diverges and increases even faster than the average-price index. The sustained price growth over the five-year period since 2007 could have significantly decreased affordability of HDB flats, leading to a shift towards smaller and lower-quality units being transacted. This result contrasts with earlier findings by Ong et al. (2003), where the hedonic price index exhibits lower volatility (both upwards and downwards) than the average-price index for an earlier time period from 1997 to 2000—however, the price movements were of significantly smaller magnitudes during that period.

More importantly, this brief and qualitative discussion shows that the hedonic price index, which is quality-adjusted, can either avoid false volatility or reveal the true growth momentum in the measurement of overall price levels—an important insight in macroeconomic policy decisions,

[32] The officially published HDB resale price index is calculated using a stratified average price method up till the third quarter of 2014, and a stratified hedonic regression method from the fourth quarter of 2014.

such as the implementation and lifting of cooling measures. When used in tandem with average price indices, it can also show short-term shifts in the quality of apartments transacted, which can provide further insights into affordability and market sentiment trends, as alluded to in the above discussion.

**Figure 7.** Hedonic price index and the published HDB resale price index (base period = 1Q 2009).

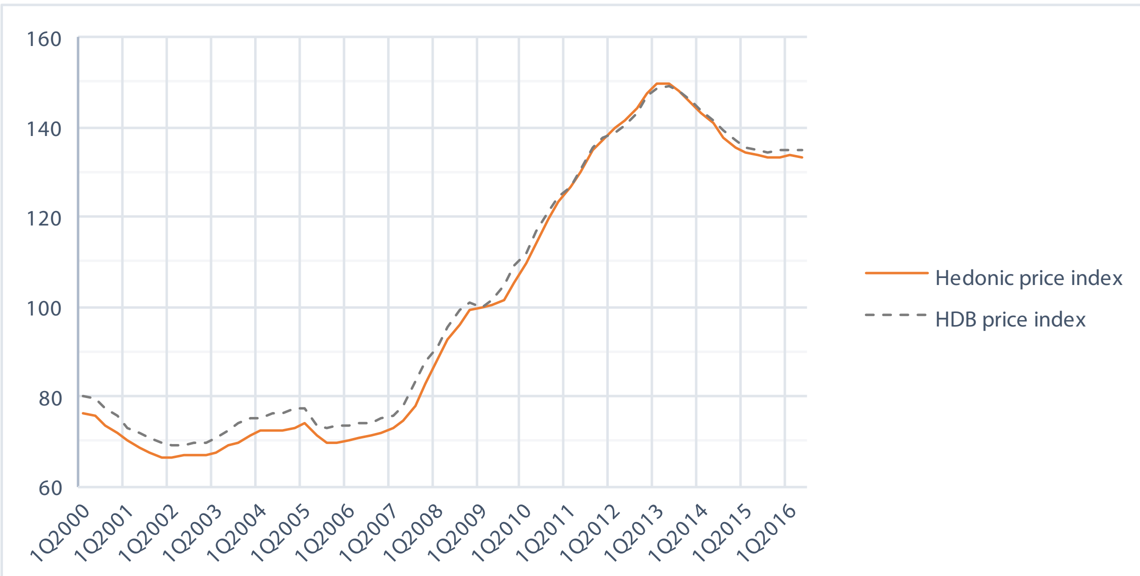
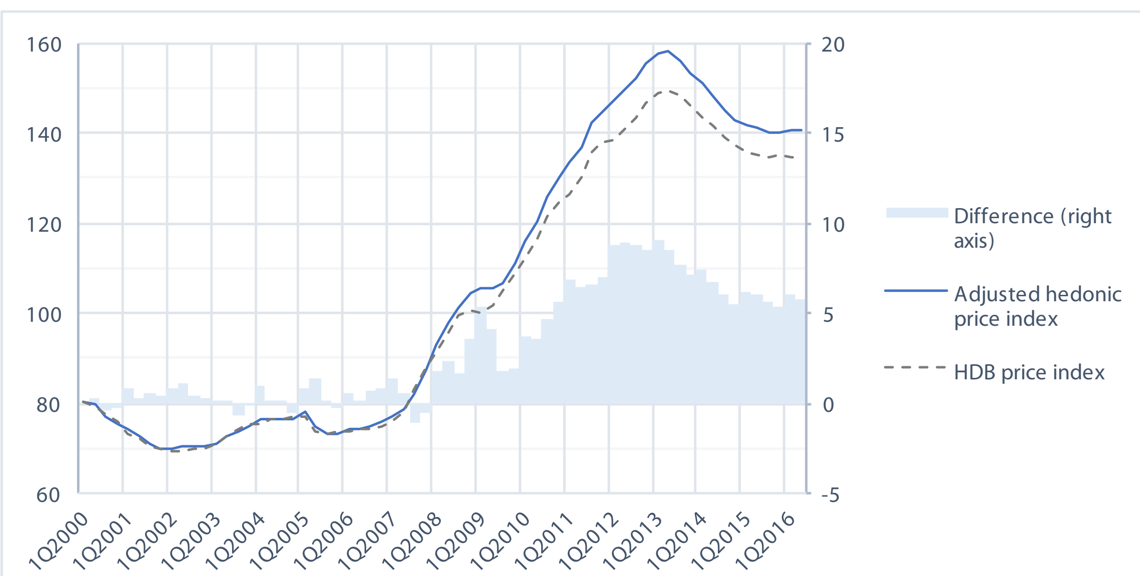



**Figure 8.** Adjusted hedonic price index and the published HDB resale price index.



### *4.5 Differential wealth appreciation based on public transit accessibility*

The hedonic price index methodology, as described in the previous section, allows for further comparison of housing price appreciation between different types of houses, by running separate hedonic regressions on each group of transactions. This enables an analysis of the *distribution* of wealth effects arising from public housing ownership.

First, I examine differences in price trends for houses "near" and "not near" the subway network—defined using a threshold radius from the nearest MRT station. The threshold is initially set at 800 meters, while values between 400 and 1,200 meters are also tested.[33] For this section, I consider only short-term anticipatory effects, that is, a house is considered "near" an MRT station if there is a future MRT station opening inside the threshold radius within 24 months from the transaction date.[34] **Figure 9** shows that houses located near MRT stations appreciate significantly faster than houses located further away (nearly 20 percentage points over the entire period). Houses near MRT stations are also better able to maintain their value during periods of market decline. Similar effects are observed at all threshold distances between 400 and 1,200 meters. To eliminate the possibility of uncaptured effects of new MRT developments, I test an alternative model that excludes all houses near new MRT lines, which results in similar observations.

I suggest two explanations for the differential rates of price appreciation:

(1) other commercial developments that occur near MRT stations (which are usually also the town centers), such as the building of new shopping malls, serve to further increase the amenity level of houses in the surrounding area;

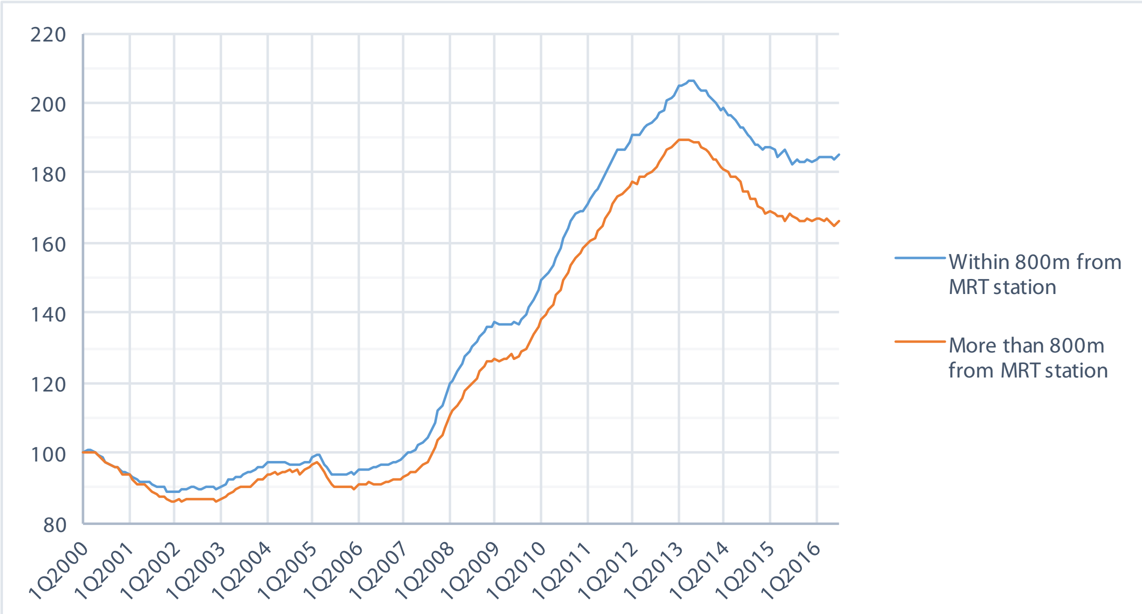
(2) improvements made to the MRT network over time, such as the building of new lines in other areas, improve the value of being close to an MRT station independent of the walking distance to the station.

---

[33] These distances are intended to represent "walking distance" thresholds and cover the range between 5–15 minutes of walking.

[34] This is sufficient to capture the peak anticipatory effects, as shown in the results from section 4.3.

**Figure 9.** Hedonic price indices by proximity to MRT station (base period = 1Q 2000).




### *4.6 Differential wealth appreciation across districts*

Next, I construct a separate price index for each district (**Figure 10**). Since grouping transactions by district introduces severe biases and heteroscedasticity in the spatial attributes (such as the distance to CBD and proximity to popular primary schools), I fix the coefficients on the quality-adjustment variables using estimates obtained from the overall (pooled) regression, and allow only the time dummies to vary across districts.[35] This analysis shows substantial variations in the growth of housing values across districts. To visualize the spatial patterns, I create a map combining recent resale prices and historical trends in price growth. In **Figure 11**, each circle represents a public housing block, with its size indicating the predicted resale price of a "typical" apartment at that address in July 2016 (the most recent period). The color intensity represents the predicted rate of price appreciation in that district, over the period from 2000 to 2016 (equivalent to the hedonic price indices shown in **Figure 10**).

---

[35] Three districts (Bukit Timah, Central Area, and Marine Parade) are not full-scale HDB towns, and are excluded from this district-wise analysis due to the small number of HDB flats built in these districts.

There are strong spatial correlations—houses in well-developed districts located near the CBD not only cost more, but have gained more in value over the period analyzed. These include: Queenstown (also the oldest HDB town) and Bukit Merah—where the price index increased by 118–130% over the 16-year period from January 2000 to July 2016. In contrast, districts with the least gains are located far from the CBD. They include newer, non-mature estates Hougang, Choa Chu Kang, and Bukit Panjang, as well as mature estates such as Pasir Ris and Tampines—with growth rates of 52–64% over the same period. Among districts that are further from the CBD, price growth appears to be concentrated in areas that have seen rapid expansion (i.e., the northeastern district of Punggol) or active development by the government into regional hubs (i.e., Jurong Lake District in the west, and Woodlands in the north). This implies that government policies in urban development have substantial effects on property values, albeit not as much as a good location does. Furthermore, it suggests that the comparatively rich, who can afford more centrally-located houses, have seen their asset values grow faster over time.

**Figure 10.** Hedonic price indices by district (base period = 1Q 2000).

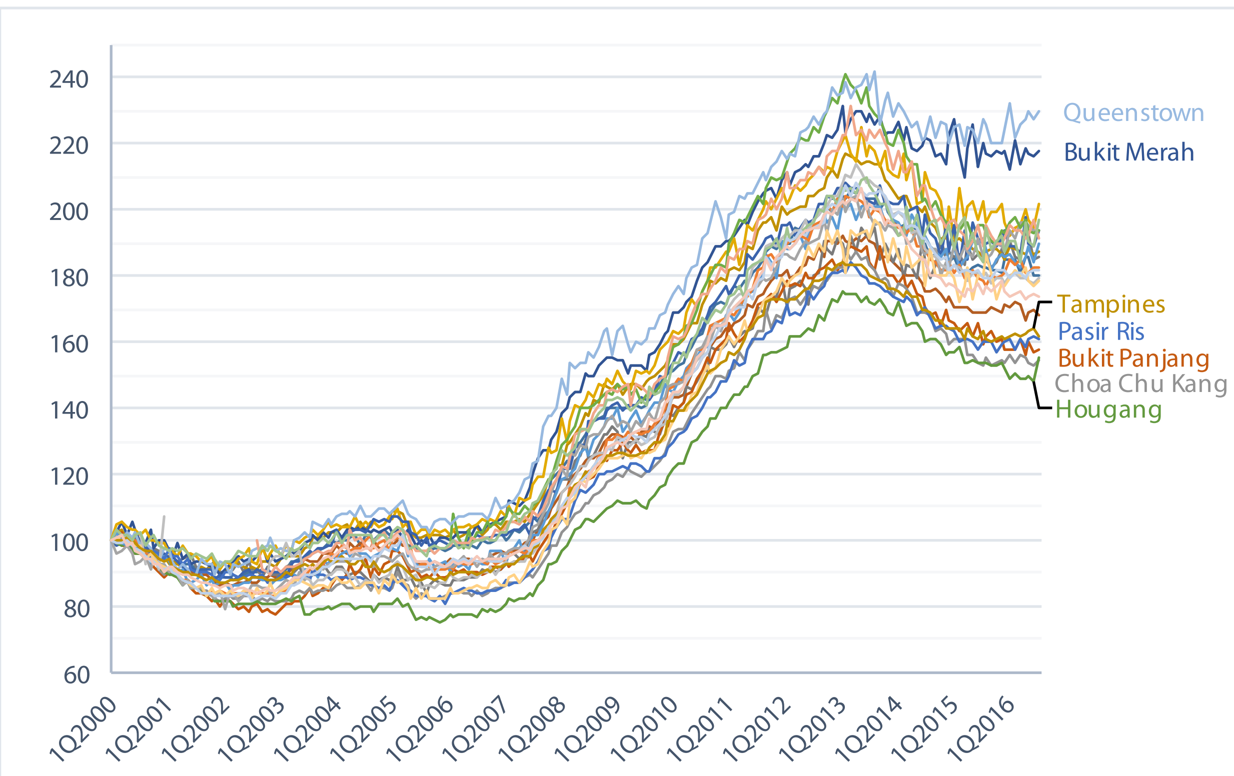
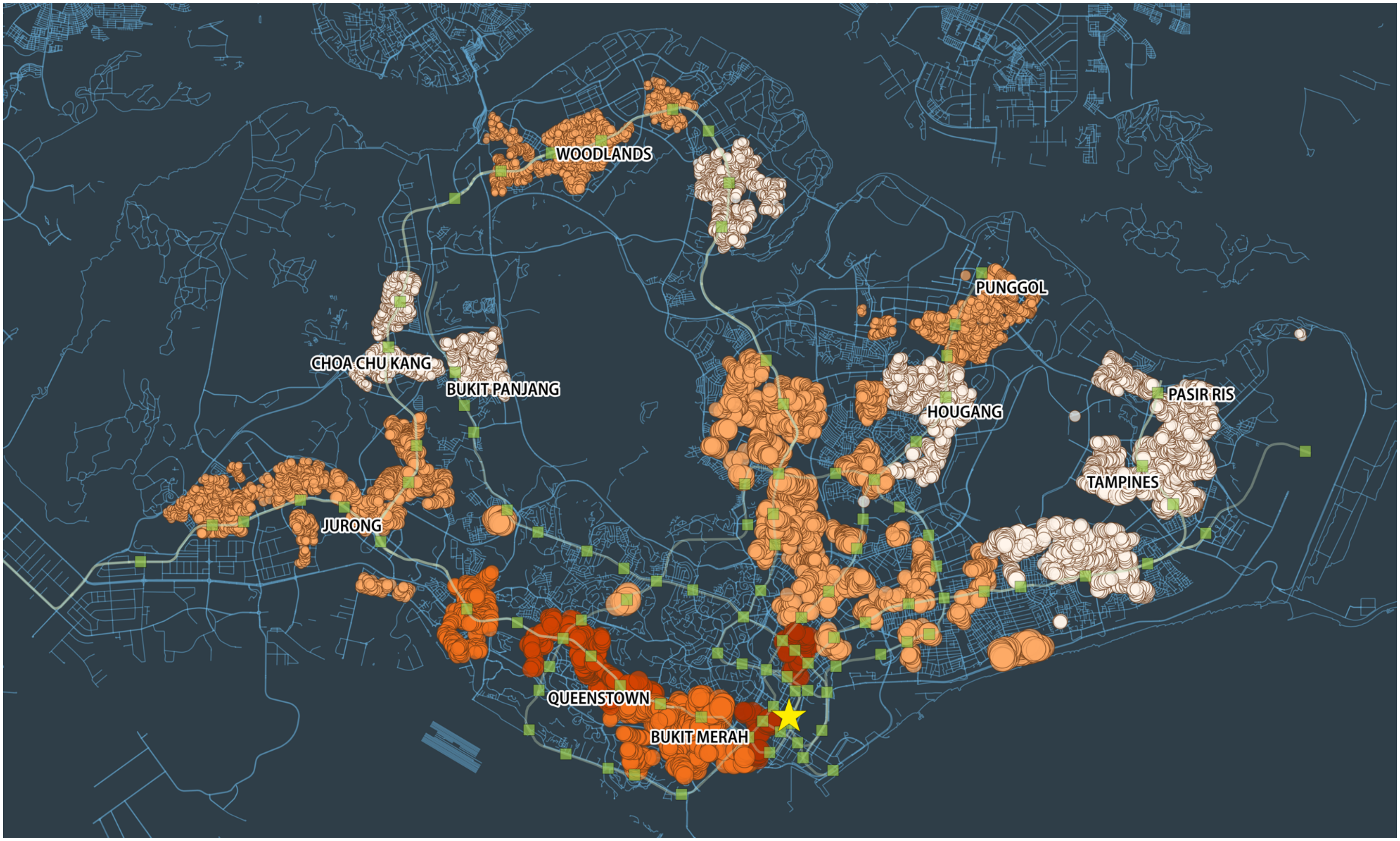


**Figure 11.** Predicted prices and predicted price growth over 2000–2016.

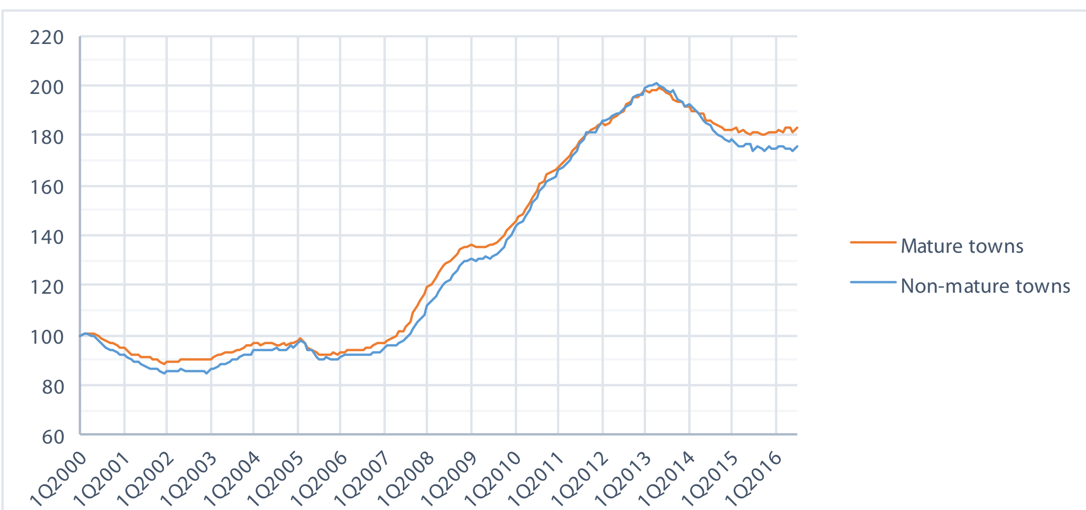



**Legend:** Circles indicate public housing blocks, with the size of circle corresponding to predicted resale price (in July 2016), and the color intensity indicating district-level price index growth over 2000–2016. Green boxes indicate MRT stations and the yellow star shows the location of the CBD. Selected housing estates are labeled.

The results presented in this section deserve further analysis, to determine if the different growth rates in the quality-adjusted price indices across districts are due to weaknesses in the model (i.e., uncaptured quality changes), or are truly suggestive of increasing disparities in wealth accumulation through public housing ownership. There are no obvious differences in the rate of price growth when districts are grouped according to the official classification of mature and non-mature estates, which suggests that variations in local amenity levels do not explain much of the observed disparities—thus lending some support to the credibility of my model (see **Figure 12**).[36]

New flats offered for sale by the government are often oversubscribed, with buyers selected through balloting. This has given rise to a "lottery effect," where those who are successful in the ballot for popular flats stand to reap large windfall profits. The Singapore government previously recognized this in the context of a special HDB housing project developed in the downtown core region, and expressed a desire to curb the inequalities arising from public housing lotteries (Wong, 2016). However, my analysis shows that large wealth disparities exist across other districts as well, and the "lottery effect" may be much more widespread than the few headline cases reported.

**Figure 12.** Hedonic price indices in mature and non-mature estates (base period = 1Q 2000).




[36] Note that there are significant differences in the actual resale prices between mature and non-mature estates, as one may expect (see results in Appendix 2, Table A2.3). However, the analysis here focuses on differences in the rates of growth in housing prices—alternatively, the rates of *return on investment* across districts.

## 5. Conclusion

This paper builds a hedonic regression model for the high-volume, high-density urban public housing resale market in Singapore. The set of physical and geospatial attributes specified is able to account for most of the observed variation in prices, while there will always be some unobservable characteristics (such as interior quality and the specific circumstances of each sale). The results are found to be robust to changes in model specification and out-sample cross-validation, supporting the hypothesis that the hedonic approach is especially relevant in mass-market public housing projects, where large numbers of apartments are built to similar quality specifications—for which Singapore is a prime example.

The hedonic pricing model is used to analyze the wealth effects of large-scale, publicly-funded projects—in this case, the building of new subway lines. It is also used to create a hedonic price index, and the policy implications have been briefly discussed. My analysis provides a case for more extensive applications in policy decisions, where cost-benefit analyses, equity implications, and housing affordability are of particular concern. It has also been suggested that hedonic pricing can be used as automated valuation models, which could bring time and cost savings to property firms, as well as serve as an input for government land sales pricing and/or new flat sales by the HDB (Muhammad et al., 2005).

The current model could be strengthened by more extensive geospatial modeling, to include the availability of neighborhood amenities, such as shopping malls, supermarkets and bus interchanges. Measures of public transit connectivity, which incorporate total transit times, would strengthen the modeling of public transit effects. Jing and Liao (2016) attempt this by generating a connectivity index for each MRT station, based on travel times and line quality to every other station in the network. While this serves as a measure of connectivity to the median station, it does not necessarily reflect the connectivity needs of most residents, since travel patterns tend to converge towards particular places of interest (Schläpfer, 2014). Furthermore, Jing and Liao (2016)'s model is limited by their exclusion of walking distances to the nearest MRT station—which has a substantial effect on housing prices, as shown in this paper. I suggest that future efforts incorporate the use of transit data analytics to determine the most common routes, destinations, travel times, and even different modes of transport (such as feeder bus services bridging the home and the MRT network). This could then form an empirical basis for the creation of a sophisticated connectivity index.

Finally, more complex econometric methods, such as geographically-weighted regression and mixed effects models, can provide greater flexibility and allow for heterogeneity in the values of individual parameters over time and space. For example, apartments located near the coastline may experience a higher value associated with increased floor level due to the sea view—which has particularly strong value in Chinese *feng shui*.

To my knowledge, this is the first application of hedonic methods that attempts to incorporate a comprehensive set of quality-adjustment parameters, including spatial and accessibility attributes, for all public housing resale transactions in Singapore in the time period studied. The insights gained on wealth and policy implications warrant further attention, both for Singapore and in other markets.

## References


Agostini, C.A., Palmucci, G.A., 2008. The Anticipated Capitalisation Effect of a New Metro Line on Housing Prices*. Fiscal Studies 29, 233–256. doi:10.1111/j.1475-5890.2008.00074.x

Bae, C.-H.C., Jun, M.-J., Park, H., 2003. The impact of Seoul's subway Line 5 on residential property values. Transport Policy 10, 85–94. doi:10.1016/S0967-070X(02)00048-3

Bailey, M.J., Muth, R.F., Nourse, H.O., 1963. A Regression Method for Real Estate Price Index Construction. Journal of the American Statistical Association 58, 933–942. doi:10.1080/01621459.1963.10480679

Bilger, M., Manning, W.G., 2015. Measuring Overfitting in Nonlinear Models: A New Method and an Application to Health Expenditures. Health Econ. 24, 75–85. doi:10.1002/hec.3003

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. Geographical Analysis 28, 281–298. doi:10.1111/j.1538-4632.1996.tb00936.x

Can, A., 1992. Specification and estimation of hedonic housing price models. Regional Science and Urban Economics, Special Issue Space and Applied Econometrics 22, 453–474. doi:10.1016/0166-0462(92)90039-4

Case, K.E., Shiller, R.J., 1989. The Efficiency Of The Market For Single-Family Homes. The American Economic Review 79, 125.

Case, K.E., Shiller, R.J., 1987. Prices of single-family homes since 1970: new indexes for four cities. New England Economic Review 45–56.

Chew, K.T., 2011. Forward-looking Behaviour In Singapore's Private Housing Market: The Impact of The North-East Transit Line On The Housing Price Gradient (Thesis).

Clapp, J., Giaccotto, C., Tirtiroglu, D., 1991. Housing Price Indexes Based on All Transactions Compared to Repeat Subsamples. Journal of the American Real Estate & Urban Economics Association 19, 270–285.

Cliff, A.D., Ord, J.K., 1981. Spatial processes: models & applications. Pion, London.

Courant, P.N., 1978. Racial prejudice in a search model of the urban housing market. Journal of Urban Economics 5, 329–345. doi:10.1016/0094-1190(78)90014-1

Deng, Y., McMillen, D.P., Sing, T.F., 2012. Private residential price indices in Singapore: A matching approach. Regional Science and Urban Economics, Special Section on Asian Real Estate Market 42, 485–494. doi:10.1016/j.regsciurbeco.2011.06.004

Diewert, E., Heravi, S., Silver, M., 2008. Hedonic imputation versus time dummy hedonic indexes. National Bureau of Economic Research.

Housing and Development Board, 2016. HDB Resale Price Index [WWW Document]. data.gov.sg. URL https://data.gov.sg/dataset/hdb-resale-price-index/resource/52e93430-01b7-4de0-80df-bc83d0afed40/view/14c47d07-1395-4661-8466-728abce27f5f (accessed 10.14.16).

Iacoviello, M., 2005. House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle. American Economic Review 95, 739–764. doi:10.1257/0002828054201477

Jiang, L., Phillips, P.C.B., Yu, J., 2015. New methodology for constructing real estate price indices applied to the Singapore residential market. J. Bank Financ. 61, S121–S131. doi:10.1016/j.jbankfin.2015.08.026

Jing, K., Liao, W., 2016. Public Transport Connectivity and Housing Price Premium: Evidence from Singapore. NUS Institute of Real Estate Studies Working Paper.

Kissling, W.D., Carl, G., 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. Global Ecology and Biogeography 17, 59–71. doi:10.1111/j.1466-8238.2007.00334.x

Linneman, P., 1980. Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing-Market. J. Urban Econ. 8, 47–68. doi:10.1016/0094-1190(80)90055-8

McMillen, D.P., McDonald, J., 2004. Reaction of House Prices to a New Rapid Transit Line: Chicago's Midway Line, 1983–1999. Real Estate Economics 32, 463–486. doi:10.1111/j.1080-8620.2004.00099.x

Mendicino, C., Punzi, M.T., 2014. House prices, capital inflows and macroprudential policy. Journal of Banking & Finance 49, 337–355. doi:10.1016/j.jbankfin.2014.06.007

Monson, M., 2009. Valuation using hedonic pricing models. Cornell Real Estate Review 7, 10.

Muhammad, F.I., Fook, J.C., Kheng, H.E., 2005. Automated valuation model: an application to the public housing resale market in Singapore. Property Management 23, 357–373. doi:10.1108/02637470510631492

Ong, S.E., Ho, K.H.D., Lim, C.H., 2003. A constant-quality price index for resale public housing flats in Singapore. Urban Studies 40, 2705–2729. doi:10.1080/0042098032000146858

Pace, R.K., LeSage, J., Zhu, S., 2009. Impact of Cliff and Ord on the Housing and Real Estate Literature. Geographical Analysis 41, 418–424. doi:10.1111/j.1538-4632.2009.00770.x

Poonawala, H., Kolar, V., Blandin, S., Wynter, L., Sahu, S., 2016. Singapore in Motion: Insights on Public Transport Service Level Through Farecard and Mobile Data Analytics. ACM Press, pp. 589–598. doi:10.1145/2939672.2939723

Rambaldi, A.N., Rao, D.P., others, 2011. Hedonic predicted house price indices using time-varying hedonic models with spatial autocorrelation. School of Economics, University of Queensland.

Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy 82, 34–55. doi:10.1086/260169

Schläpfer, M., 2014. How polycentric are our cities?, in: Decoding the City: Urbanism in the Age of Big Data. Birkhäuser, Germany, pp. 160–167.

Shi, S., Jou, J.-B., Tripe, D., 2014. Can interest rates really control house prices? Effectiveness and implications for macroprudential policy. Journal of Banking & Finance 47, 15–28. doi:10.1016/j.jbankfin.2014.06.012

Singapore Department of Statistics, 2016. General Household Survey 2015.

Sue, E.D.W., Wong, W.-K., 2010. The political economy of housing prices: Hedonic pricing with regression discontinuity. J. Hous. Econ. 19, 145–156. doi:10.1016/j.jhe.2010.04.004

Sun, H., Tu, Y., Yu, S.M., 2005. A spatio-temporal autoregressive model for multi-unit residential market analysis. J. Real Estate Financ. Econ. 31, 155–187. doi:10.1007/s11146-005-1370-0 van Eggermond, M. a. B., Lehner, M., Erath, A., 2011. Modeling Hedonic Prices in Singapore. Hong Kong Soc Transportation Studies Ltd, Hong Kong.

Wilhelmsson, M., 2002. Spatial models in real estate economics. Housing, Theory and Society 19, 92–101.

Wong, M., 2013. Estimating Ethnic Preferences Using Ethnic Housing Quotas in Singapore. The Review of Economic Studies 80, 1178–1214. doi:10.1093/restud/rdt002

Wong, S.Y., 2016. Plan to mitigate “lottery effect” of downtown flats. The Straits Times.

Wong, W.-K., 2008. Parental Valuation of Priority Admission in Elementary Schools: The Effects of Academic Reputation and Choices. NUS Department of Economics Working Paper.

Ye, V.Y., 2016. Analyzing the subsidized Hong Kong real estate market: A case for the hedonic index approach. Unpublished manuscript.

Ye, V.Y., Becker, C., 2016. The (Literally) Steepest Slope: Spatial, Temporal, and Elevation Variance Gradients in Urban Spatial Modelling. Economic Research Initiatives at Duke (ERID) Working Paper No. 202. doi:10.2139/ssrn.2706709

## Appendix 1. Summary Statistics

**Table A1.** Descriptive statistics of resale transactions dataset.

| Variable | Description | Mean (SD) | Minimum | Maximum |
|---|---|---|---|---|
| resale_price | Nominal selling price in Singapore Dollars | 312,077.4 (132,457.1) | 28000 | 1,088,888 |
| date | Year and month of transaction | | 2000-01 | 2016-08 |
| town | Name of the HDB town or estate where flat is located, with 26 possible values | | | |
| block | Address field #1, indicating the block number | | | |
| street_name | Address field #2, indicating the street | | | |
| storey_range | Numeric range of either 3 or 5 floors encompassing the floor level of the flat transacted | | 01 to 03 | 49 to 51 |
| storey | Floor level, approximated using the median of *storey_range* | 7.559 (4.624) | 2 | 50 |
| floor_area_sqm | Floor area in square meters | 96.566 (25.441) | 28 | 297 |
| flat_type | Type of flat, corresponding to the number of bedrooms and layout, with 7 possible values: 1- to 5-room, executive, or multi-generation† | | | |
| flat_model | Flat model within each layout type, which has evolved over several iterations over time | | | |
| lease_commence_date | Year of lease commencement, usually 1 year after temporary occupation and collection of keys to a new flat | 1988.338 (9.388) | 1966 | 2013 |
| age | Approximate age of flat at the time of transaction, in years, calculated as *year of sale – lease_commence_date + 1.* | 19.455 (9.885) | −1 | 50 |

† According to HDB's convention, the common living room is counted as a room. Hence, 1-room refers to studio apartments, 2-room flats have one bedroom, and so on. 5-room flats have three bedrooms and an optional study room. Executive flats are larger versions of 5-room flats, some split across two floors. Multi-generation flats are similar to 5-room flats attached to a studio apartment, and are meant for, as the name suggests, young families living with their parents.

**Figure A1.1.** Transaction volume and average resale price over time.

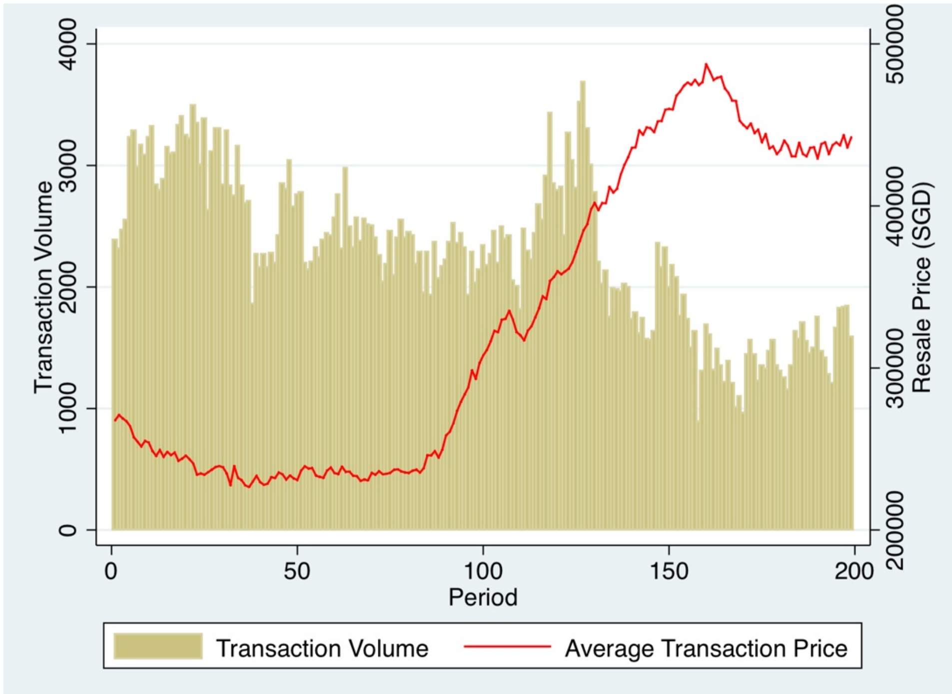
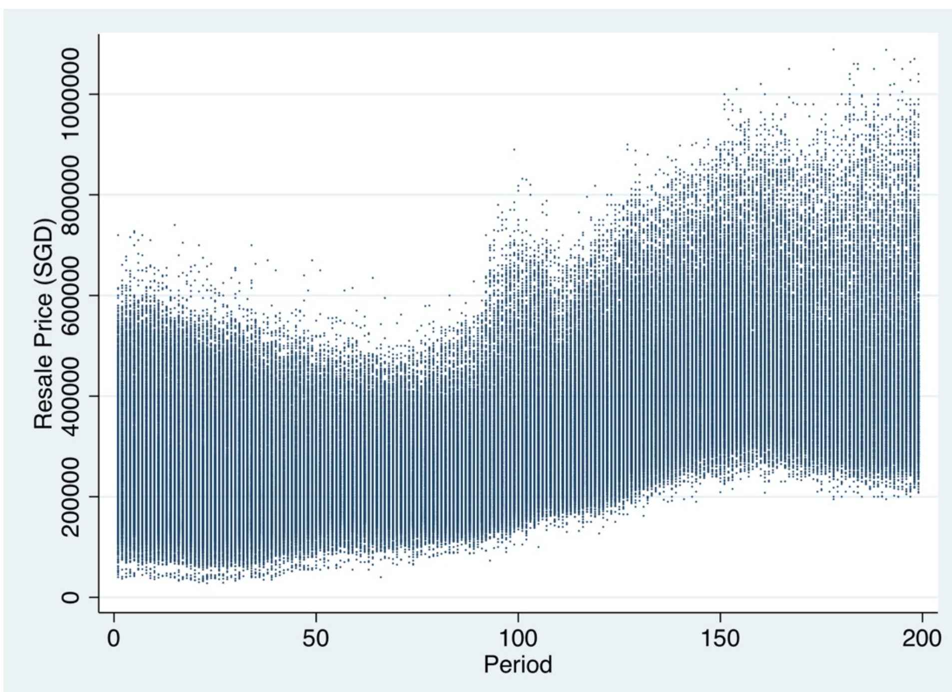



**Figure A1.2.** Scatter plot of transaction prices over time.



## Appendix 2. Regression Results

**Table A2.1.** Variables entered into regression equation.

| Variable name | Description |
|---|---|
| *Independent variable* | |
| price13 | Transaction price with a 1/3rd power transformation |
| | |
| *Dependent variables (base model)* | |
| period | Factor variable for time periods (values: 1–199) |
| townid | Factor variable for district (values: 1–26) |
| age | Approximate age of apartment at time of transaction, in years |
| age***X*** | Higher order terms of *age* (range of ***X***: 2–7) |
| lg _area | Natural log of floor area (in sq. meters) |
| storey | Floor level |
| rm4 | Dummy for flat type being 4-room |
| rm5 | Dummy for flat type being 5-room |
| exec | Dummy for flat type being Executive |
| multigen | Dummy for flat type being Multi-Generation |
| lg_ddistToCBD | Natural log of driving distance to CBD (in meters) |
| lg_dtoMRT | Natural log of linear distance to the nearest MRT station (in meters) |
| int_futureMRT | Interaction term: $\log(\Delta d) \times (144 - t)$ to measure change in distance and time interval to the opening of a future MRT station |
| int_futureMRT***X*** | Interaction term with higher order terms of $(144 - t)$ (range of ***X***: 2–6) |
| primarysch1 | Dummy for being located < 1km of a popular primary school |
| primarysch2 | Dummy for being located 1–2km from a popular primary school |
| lg_distToHighway | Natural log of linear distance to the nearest expressway or semi-expressway (measured in meters) |
| lg_distToHawker | Natural log of linear distance to the nearest hawker center |
| lg_distToPark | Natural log of linear distance to the nearest park or green space |
| | |
| *Dependent variables (alternative specifications)* | |
| decay04_ddistToCBD, decay04_dtoMRT, decay04_int_futureMRT, decay04_int_futureMRT***X***, decay04_distToHighway, decay04_distToHawker, decay04_distToPark | Alternative specification of distance-related variables, transformed using distance-decay parameter $e^{-\lambda d}$ with $\lambda = 0.04$ (also tested with values between 0.04 and 0.2) |
| newMRTline_open | Time interval (in months) to the opening of a new MRT line at the existing nearest MRT station (if none, value is zero) |
| newMRTline_open***X*** | Higher order terms of *newMRTline_open* (range of ***X***: 2–4) |

**Table A2.2.** Selected regression results showing estimated coefficients on quality-adjustment variables (with time period and district-specific dummies truncated).

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| age | 1.217*** | 1.482*** | 1.403*** | 1.397*** | 1.397*** | 1.379*** | 1.297*** |
| | (0.0802) | (0.0996) | (0.101) | (0.0962) | (0.0962) | (0.0963) | (0.0968) |
| age2 | -0.278*** | -0.321*** | -0.305*** | -0.304*** | -0.304*** | -0.300*** | -0.287*** |
| | (0.0143) | (0.0174) | (0.0177) | (0.0167) | (0.0167) | (0.0167) | (0.0168) |
| age3 | 0.0223*** | 0.0259*** | 0.0244*** | 0.0243*** | 0.0243*** | 0.0239*** | 0.0228*** |
| | (0.00127) | (0.00153) | (0.00156) | (0.00145) | (0.00145) | (0.00146) | (0.00147) |
| age4 | -0.000950*** | -0.00111*** | -0.00104*** | -0.00104*** | -0.00104*** | -0.00102*** | -0.000968*** |
| | (0.0000626) | (0.0000737) | (0.0000758) | (0.0000696) | (0.0000696) | (0.0000698) | (0.0000703) |
| age5 | 0.0000229*** | 0.0000270*** | 0.0000253*** | 0.0000252*** | 0.0000252*** | 0.0000246*** | 0.0000234*** |
| | (0.00000171) | (0.00000198) | (0.00000205) | (0.00000186) | (0.00000186) | (0.00000186) | (0.00000188) |
| age6 | -0.000000294*** | -0.000000347*** | -0.000000327*** | -0.000000327*** | -0.000000327*** | -0.000000318*** | -0.000000301*** |
| | (2.43e-08) | (2.78e-08) | (2.89e-08) | (2.58e-08) | (2.58e-08) | (2.59e-08) | (2.61e-08) |
| age7 | 1.54e-09*** | 1.83e-09*** | 1.74e-09*** | 1.74e-09*** | 1.74e-09*** | 1.69e-09*** | 1.60e-09*** |
| | (1.40e-10) | (1.58e-10) | (1.65e-10) | (1.46e-10) | (1.46e-10) | (1.46e-10) | (1.47e-10) |
| lg_area | 19.20*** | 19.20*** | 18.71*** | 18.05*** | 18.05*** | 18.06*** | 17.93*** |
| | (0.0573) | (0.0573) | (0.0617) | (0.0469) | (0.0469) | (0.0469) | (0.0468) |
| storey | 0.173*** | 0.173*** | 0.172*** | 0.174*** | 0.174*** | 0.175*** | 0.176*** |
| | (0.000794) | (0.000795) | (0.000790) | (0.000776) | (0.000775) | (0.000776) | (0.000780) |
| rm4 | 0.427*** | 0.426*** | 0.591*** | 0.833*** | 0.833*** | 0.817*** | 0.839*** |
| | (0.0211) | (0.0211) | (0.0221) | (0.0174) | (0.0174) | (0.0174) | (0.0174) |
| rm5 | 1.623*** | 1.622*** | 1.900*** | 2.295*** | 2.295*** | 2.273*** | 2.313*** |
| | (0.0333) | (0.0333) | (0.0352) | (0.0272) | (0.0273) | (0.0273) | (0.0273) |
| exec | 3.716*** | 3.715*** | 4.086*** | 4.614*** | 4.614*** | 4.599*** | 4.683*** |
| | (0.0452) | (0.0452) | (0.0478) | (0.0375) | (0.0375) | (0.0375) | (0.0375) |
| multigen | 5.197*** | 5.195*** | 5.608*** | 6.198*** | 6.197*** | 6.219*** | 6.169*** |
| | (0.181) | (0.181) | (0.181) | (0.178) | (0.178) | (0.176) | (0.178) |
| lg_ddistToCBD | -2.694*** | -2.696*** | -2.864*** | -2.901*** | -2.901*** | -2.940*** | -2.872*** |
| | (0.0358) | (0.0359) | (0.0350) | (0.0345) | (0.0346) | (0.0346) | (0.0345) |
| lg_dtoMRT | -1.588*** | -1.589*** | -1.605*** | -1.621*** | -1.621*** | -1.638*** | -1.682*** |
| | (0.00697) | (0.00697) | (0.00694) | (0.00681) | (0.00680) | (0.00676) | (0.00673) |
| int_futureMRT | -0.0130*** | -0.0130*** | -0.0135*** | -0.0127*** | -0.0127*** | -0.0132*** | -0.0122*** |
| | (0.000929) | (0.000930) | (0.000926) | (0.000917) | (0.000917) | (0.000917) | (0.000925) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| int_futureMRT2 | 0.000806*** | 0.000808*** | 0.000859*** | 0.000826*** | 0.000826*** | 0.000855*** | 0.000797*** |
| | (0.0000725) | (0.0000725) | (0.0000723) | (0.0000716) | (0.0000717) | (0.0000717) | (0.0000723) |
| int_futureMRT3 | -0.0000237*** | -0.0000238*** | -0.0000253*** | -0.0000247*** | -0.0000247*** | -0.0000254*** | -0.0000243*** |
| | (0.00000210) | (0.00000210) | (0.00000209) | (0.00000207) | (0.00000207) | (0.00000207) | (0.00000209) |
| int_futureMRT4 | 0.000000332*** | 0.000000332*** | 0.000000352*** | 0.000000345*** | 0.000000345*** | 0.000000354*** | 0.000000345*** |
| | (2.80e-08) | (2.80e-08) | (2.80e-08) | (2.77e-08) | (2.77e-08) | (2.77e-08) | (2.81e-08) |
| int_futureMRT5 | -2.10e-09*** | -2.10e-09*** | -2.21e-09*** | -2.17e-09*** | -2.17e-09*** | -2.23e-09*** | -2.20e-09*** |
| | (1.75e-10) | (1.75e-10) | (1.75e-10) | (1.74e-10) | (1.74e-10) | (1.74e-10) | (1.76e-10) |
| int_futureMRT6 | 4.89e-12*** | 4.89e-12*** | 5.12e-12*** | 5.03e-12*** | 5.03e-12*** | 5.16e-12*** | 5.14e-12*** |
| | (4.15e-13) | (4.16e-13) | (4.15e-13) | (4.11e-13) | (4.11e-13) | (4.12e-13) | (4.17e-13) |
| primarysch1 | 0.295*** | 0.294*** | 0.291*** | 0.281*** | 0.280*** | 0.266*** | |
| | (0.0131) | (0.0131) | (0.0131) | (0.0129) | (0.00902) | (0.00899) | |
| primarysch2 | 0.000676 | 0.000319 | -0.00446 | 0.00116 | | | |
| | (0.0115) | (0.0115) | (0.0115) | (0.0113) | | | |
| lg_distToHighway | 0.271*** | 0.271*** | 0.257*** | 0.266*** | 0.266*** | 0.272*** | |
| | (0.00494) | (0.00494) | (0.00488) | (0.00478) | (0.00473) | (0.00473) | |
| lg_distToHawker | -0.106*** | -0.106*** | -0.109*** | -0.138*** | -0.138*** | | |
| | (0.00724) | (0.00724) | (0.00718) | (0.00690) | (0.00683) | | |
| lg_distToPark | -0.0596*** | -0.0595*** | -0.0556*** | -0.0467*** | -0.0467*** | | |
| | (0.00416) | (0.00417) | (0.00415) | (0.00406) | (0.00406) | | |
| constant | 13.39*** | 12.77*** | 16.81*** | 20.12*** | 20.12*** | 19.44*** | 21.81*** |
| | (0.486) | (0.504) | (0.501) | (0.455) | (0.455) | (0.455) | (0.447) |
| Observations | 450872 | 450441 | 445642 | 444377 | 444377 | 444377 | 444377 |
| $R^2$ | 0.946 | 0.946 | 0.946 | 0.948 | 0.948 | 0.948 | 0.947 |

Standard errors in parenthesis.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

**Table A2.2** ***(continued).***

| | (8) | (9) | (10) | (11) | (12) | (13) | (14) |
|---|---|---|---|---|---|---|---|
| age | 1.297*** | 1.379*** | 0.995*** | 2.448*** | 0.595*** | 2.051*** | 1.436*** |
| | (0.0968) | (0.0962) | (0.102) | (0.131) | (0.143) | (0.181) | (0.134) |
| age2 | -0.287*** | -0.299*** | -0.240*** | -0.479*** | -0.178*** | -0.388*** | -0.324*** |
| | (0.0168) | (0.0167) | (0.0177) | (0.0226) | (0.0254) | (0.0296) | (0.0247) |
| age3 | 0.0228*** | 0.0237*** | 0.0191*** | 0.0384*** | 0.0148*** | 0.0289*** | 0.0267*** |
| | (0.00147) | (0.00146) | (0.00154) | (0.00195) | (0.00228) | (0.00245) | (0.00229) |
| age4 | -0.000968*** | -0.000999*** | -0.000813*** | -0.00166*** | -0.000646*** | -0.00116*** | -0.00115*** |
| | (0.0000703) | (0.0000698) | (0.0000735) | (0.0000926) | (0.000113) | (0.000112) | (0.000117) |
| age5 | 0.0000234*** | 0.0000240*** | 0.0000198*** | 0.0000403*** | 0.0000162*** | 0.0000265*** | 0.0000279*** |
| | (0.00000188) | (0.00000186) | (0.00000196) | (0.00000244) | (0.00000312) | (0.00000286) | (0.00000331) |
| age6 | -0.000000301*** | -0.000000308*** | -0.000000258*** | -0.000000516*** | -0.000000217*** | -0.000000327*** | -0.000000352*** |
| | (2.61e-08) | (2.59e-08) | (2.73e-08) | (3.36e-08) | (4.49e-08) | (3.84e-08) | (4.87e-08) |
| age7 | 1.60e-09*** | 1.62e-09*** | 1.39e-09*** | 2.70e-09*** | 1.20e-09*** | 1.67e-09*** | 1.81e-09*** |
| | (1.47e-10) | (1.46e-10) | (1.54e-10) | (1.87e-10) | (2.63e-10) | (2.09e-10) | (2.89e-10) |
| lg_area | 17.93*** | 18.03*** | 17.86*** | 18.58*** | 17.92*** | 17.42*** | 17.85*** |
| | (0.0468) | (0.0468) | (0.0491) | (0.0651) | (0.0659) | (0.0689) | (0.0632) |
| storey | 0.176*** | 0.174*** | 0.175*** | 0.177*** | 0.158*** | 0.185*** | 0.156*** |
| | (0.000780) | (0.000775) | (0.000810) | (0.00105) | (0.00107) | (0.00114) | (0.000968) |
| rm4 | 0.839*** | 0.837*** | 0.841*** | 0.910*** | 0.665*** | 1.555*** | 0.192*** |
| | (0.0174) | (0.0174) | (0.0183) | (0.0244) | (0.0241) | (0.0246) | (0.0238) |
| rm5 | 2.313*** | 2.304*** | 2.357*** | 2.243*** | 2.214*** | 3.702*** | 1.406*** |
| | (0.0273) | (0.0272) | (0.0286) | (0.0382) | (0.0376) | (0.0407) | (0.0363) |
| exec | 4.683*** | 4.627*** | 4.714*** | 4.289*** | 4.708*** | 6.007*** | 3.858*** |
| | (0.0375) | (0.0375) | (0.0392) | (0.0526) | (0.0516) | (0.0547) | (0.0503) |
| multigen | 6.169*** | 6.244*** | 6.006*** | 4.869*** | 6.930*** | 8.098*** | 5.737*** |
| | (0.178) | (0.177) | (0.178) | (0.243) | (0.240) | (0.267) | (0.225) |
| lg_ddistToCBD | -2.872*** | -2.882*** | | -3.251*** | -1.384*** | -2.545*** | -2.472*** |
| | (0.0345) | (0.0345) | | (0.0428) | (0.0572) | (0.0422) | (0.0569) |
| lg_dtoMRT | -1.682*** | -1.616*** | | -1.436*** | -0.970*** | -1.790*** | -1.623*** |
| | (0.00673) | (0.00682) | | (0.0112) | (0.0195) | (0.0107) | (0.00864) |
| int_futureMRT | -0.0122*** | -0.0125*** | | -0.0165*** | -0.0126*** | -0.00563*** | -0.00264 |
| | (0.000925) | (0.000916) | | (0.00311) | (0.000952) | (0.00119) | (0.00159) |
| int_futureMRT2 | 0.000797*** | 0.000820*** | | 0.000887*** | 0.000829*** | -0.0000617 | 0.000539*** |
| | (0.0000723) | (0.0000716) | | (0.000243) | (0.0000741) | (0.0000966) | (0.000118) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| int_futureMRT3 | -0.0000243*** | -0.0000246*** | | -0.0000199** | -0.0000246*** | 0.00000309 | -0.0000214*** |
| | (0.00000209) | (0.00000207) | | (0.00000694) | (0.00000215) | (0.00000286) | (0.00000327) |
| int_futureMRT4 | 0.000000345*** | 0.000000345*** | | 0.000000217* | 0.000000343*** | -1.08e-08 | 0.000000327*** |
| | (2.81e-08) | (2.77e-08) | | (9.07e-08) | (2.90e-08) | (3.89e-08) | (4.25e-08) |
| int_futureMRT5 | -2.20e-09*** | -2.17e-09*** | | -1.11e-09* | -2.18e-09*** | -1.79e-10 | -2.10e-09*** |
| | (1.76e-10) | (1.73e-10) | | (5.51e-10) | (1.83e-10) | (2.47e-10) | (2.60e-10) |
| int_futureMRT6 | 5.14e-12*** | 5.04e-12*** | | 2.17e-12 | 5.10e-12*** | 9.81e-13 | 4.81e-12*** |
| | (4.17e-13) | (4.11e-13) | | (1.26e-12) | (4.38e-13) | (5.90e-13) | (6.06e-13) |
| primarysch1 | | 0.287*** | 0.266*** | 0.721*** | 0.302*** | 0.0401* | 0.611*** |
| | | (0.0129) | (0.0140) | (0.0184) | (0.0190) | (0.0201) | (0.0167) |
| primarysch2 | | -0.000502 | -0.0449*** | 0.240*** | 0.135*** | 0.103*** | 0.101*** |
| | | (0.0113) | (0.0121) | (0.0146) | (0.0184) | (0.0190) | (0.0132) |
| lg_distToHighway | | 0.265*** | | 0.171*** | 0.243*** | 0.453*** | 0.0280*** |
| | | (0.00478) | | (0.00767) | (0.00652) | (0.00630) | (0.00714) |
| lg_distToHawker | | -0.144*** | | -0.0979*** | -0.165*** | -0.177*** | -0.0653*** |
| | | (0.00690) | | (0.00921) | (0.0106) | (0.00947) | (0.00921) |
| lg_distToPark | | -0.0444*** | | -0.134*** | 0.0107 | -0.162*** | 0.0155** |
| | | (0.00406) | | (0.00562) | (0.00588) | (0.00625) | (0.00512) |
| newMRTline_open | | -0.0788*** | | | | | |
| | | (0.0114) | | | | | |
| newMRTline_open2 | | 0.00202*** | | | | | |
| | | (0.000290) | | | | | |
| newMRTline_open3 | | -0.0000165*** | | | | | |
| | | (0.00000238) | | | | | |
| newMRTline_open4 | | 4.50e-08*** | | | | | |
| | | (6.31e-09) | | | | | |
| decay04_ddistToCBD | | | 9.847*** | | | | |
| | | | (0.128) | | | | |
| decay04_dtoMRT | | | 27.69*** | | | | |
| | | | (0.248) | | | | |
| decay04_int_futureMRT | | | -0.00205 | | | | |
| | | | (0.00163) | | | | |
| decay04_int_futureMRT2 | | | -0.000666*** | | | | |
| | | | (0.0000472) | | | | |
| decay04_int_futureMRT3 | | | 0.00000836*** | | | | |
| | | | (0.000000443) | | | | |
| decay04_int_futureMRT4 | | | -2.49e-08*** | | | | |
| | | | (1.32e-09) | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| decay04_distToHighway | | | -2.344*** | | | | |
| | | | (0.193) | | | | |
| decay04_distToHawker | | | 8.840*** | | | | |
| | | | (0.247) | | | | |
| decay04_distToPark | | | 14.71*** | | | | |
| | | | (0.405) | | | | |
| constant | 21.81*** | 20.03*** | -68.89*** | 18.24*** | 3.453*** | 18.35*** | 16.68*** |
| | (0.447) | (0.454) | (0.588) | (0.601) | (0.721) | (0.661) | (0.696) |
| Observations | 444377 | 444377 | 444377 | 231962 | 212415 | 195336 | 249041 |
| $R^2$ | 0.947 | 0.948 | 0.943 | 0.952 | 0.949 | 0.953 | 0.950 |

Standard errors in parenthesis.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

**Notes:** Regression model (1) is performed on the full dataset; model (2) excludes “distressed sale” transactions where the apartment age is less than four years; model (3) further excludes outlier apartment types, i.e. 1-room and 2-room flats; models (4) to (13) further exclude outlier apartment models, i.e. “terrace” and “adjoined flat”; model (9) uses the alternative distance-decay parameter transformation of distance-related variables; model (10) includes only apartments located within 800 meters from an MRT station that is operational at, or within 24 months from, the transaction date; model (11) includes only apartments located more than 800 meters from an MRT station; model (12) includes only apartments in mature estates; model (13) includes only non-mature estates. Heteroscedasticity-consistent standard errors are presented.

**Table A2.3.** Regression results showing estimated coefficients on district-specific dummies.

| | (1) | | *(cont'd)* | (1) | |
|---|---|---|---|---|---|
| | Estimate | SE | | Estimate | SE |
| *Mature estates*† | | | *Non-mature estates* | | |
| Ang Mo Kio | 0.000 | | Bukit Batok | -1.893*** | (0.0254) |
| Bedok | 0.149*** | (0.0198) | Bukit Panjang | -1.856*** | (0.0315) |
| Bishan | 2.474*** | (0.0274) | Choa Chu Kang | -3.609*** | (0.0317) |
| Bukit Merah | 0.274*** | (0.0388) | Hougang | -2.113*** | (0.0212) |
| Bukit Timah | 6.801*** | (0.0669) | Jurong East | -1.554*** | (0.0265) |
| Central Area | -0.691*** | (0.0620) | Jurong West | -3.735*** | (0.0234) |
| Clementi | 1.114*** | (0.0244) | Punggol | -3.169*** | (0.0344) |
| Geylang | -1.171*** | (0.0291) | Sembawang | -5.917*** | (0.0361) |
| Kallang/Whampoa | -1.158*** | (0.0330) | Sengkang | -4.158*** | (0.0258) |
| Marine Parade | 7.116*** | (0.0414) | Woodlands | -4.688*** | (0.0295) |
| Pasir Ris | -1.064*** | (0.0280) | Yishun | -2.801*** | (0.0257) |
| Queenstown | 1.306*** | (0.0270) | | | |
| Serangoon | -0.703*** | (0.0260) | | | |
| Tampines | 0.503*** | (0.0223) | | | |
| Toa Payoh | 0.360*** | (0.0258) | | | |

Standard errors in parenthesis.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
† HDB defines mature estates as those that are more than 20 years old.

## Appendix 3. Regression Diagnostics

**Figure A3.1.** Residual scatter plot over transaction price.

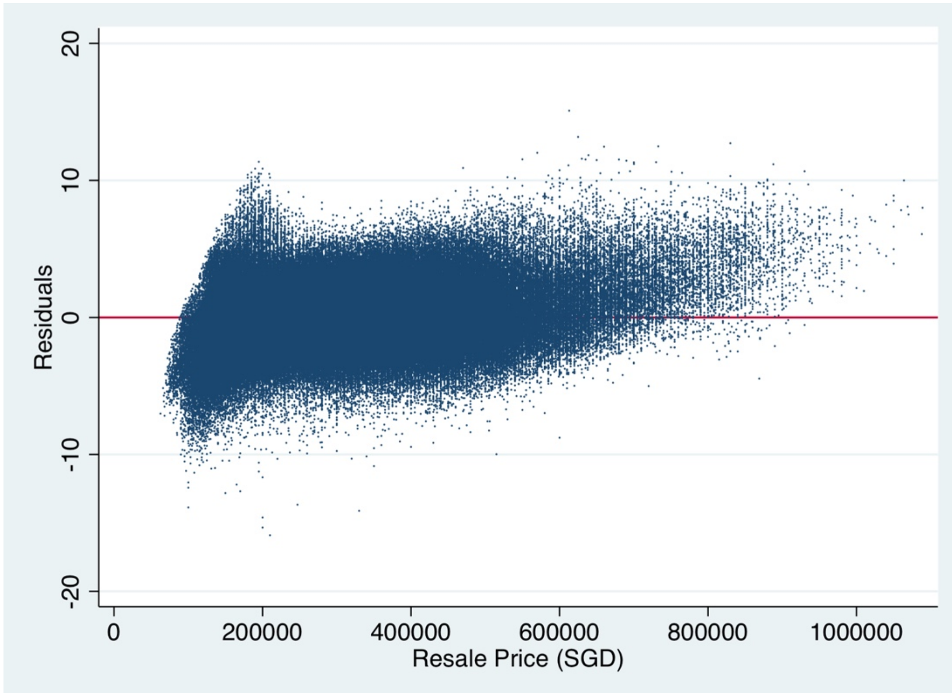



**Figure A3.2.** Residual scatter plot over time period.

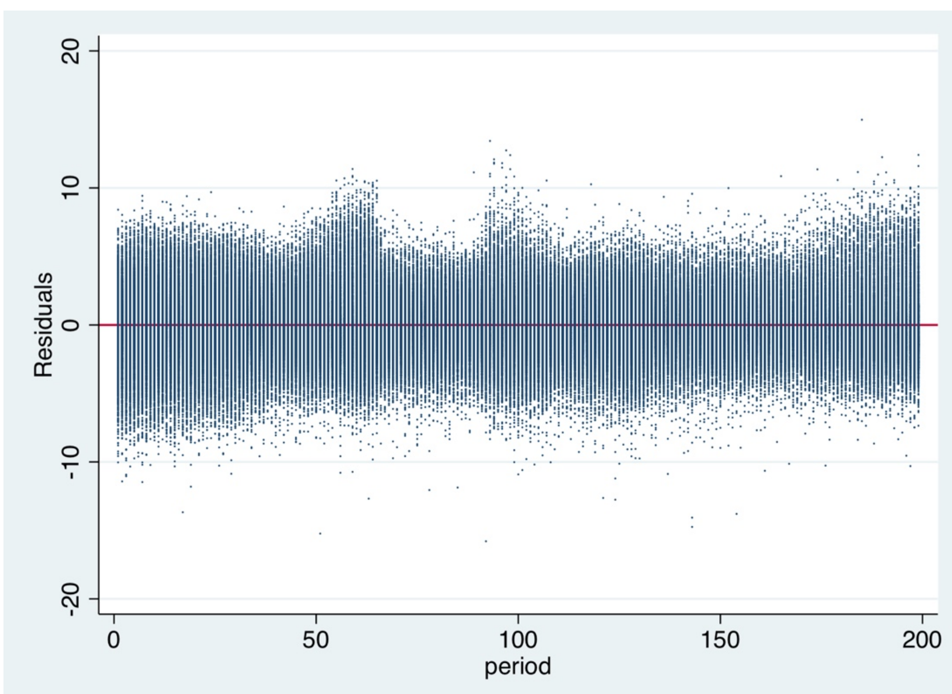


**Figure A3.3.** Residual scatter plot over district (HDB town).

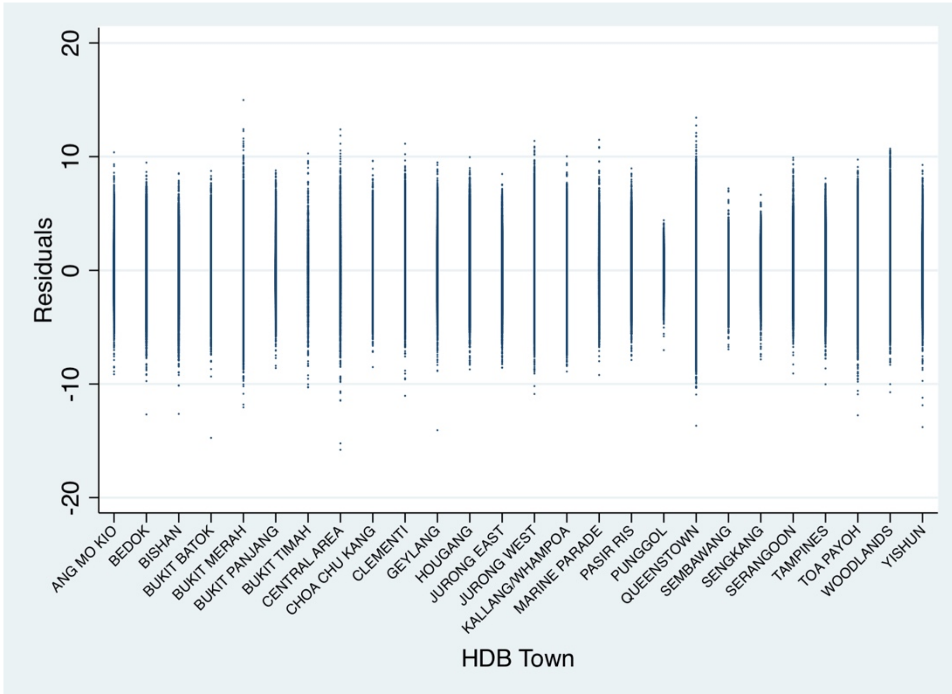



**Figure A3.4.** Residual scatter plot over floor area.

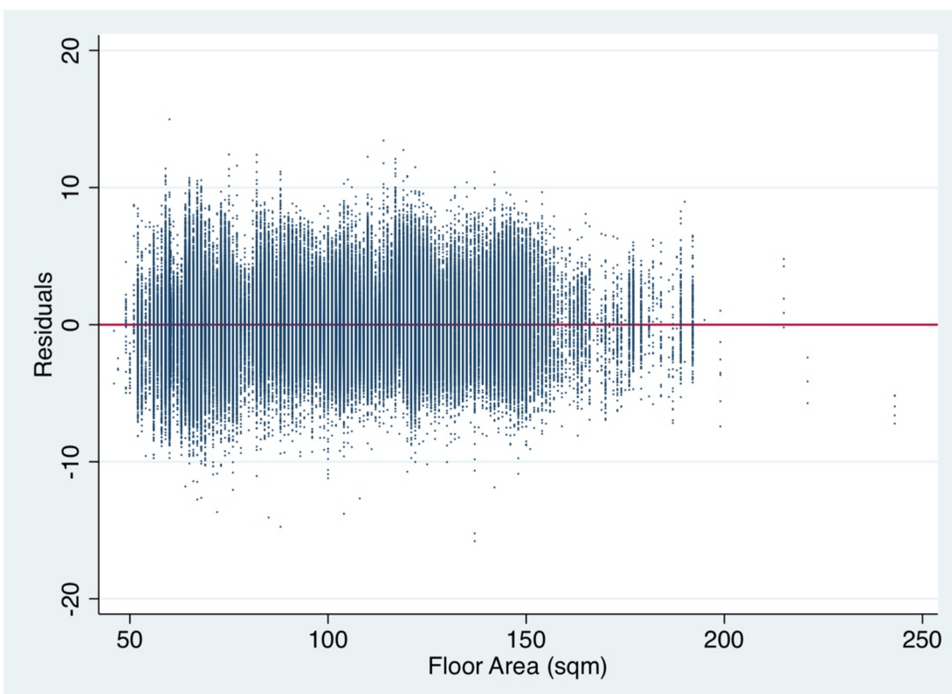


**Figure A3.5.** Residual scatter plot over distance to the CBD.

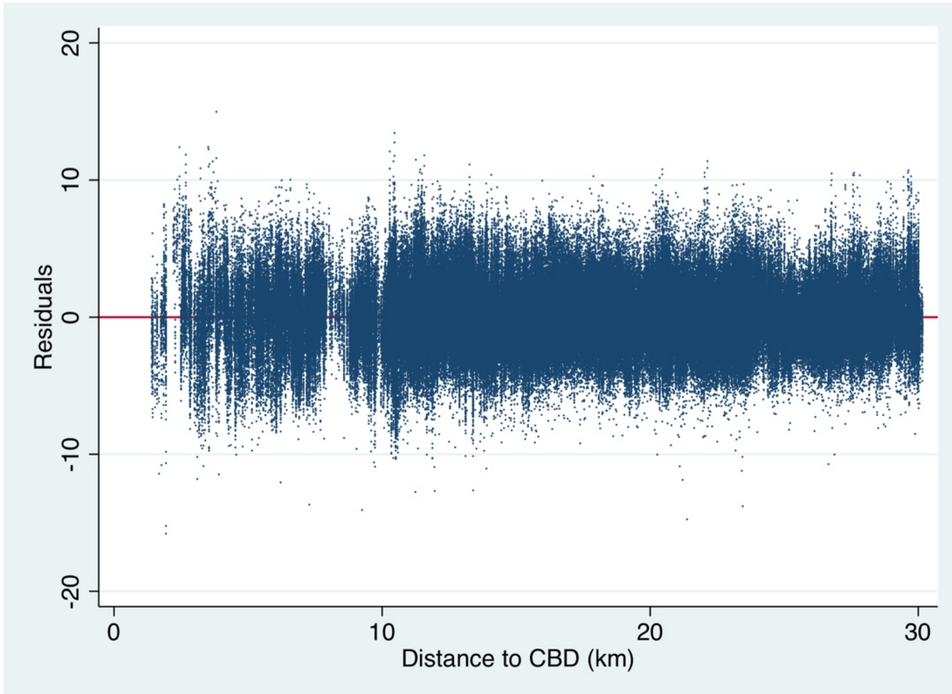
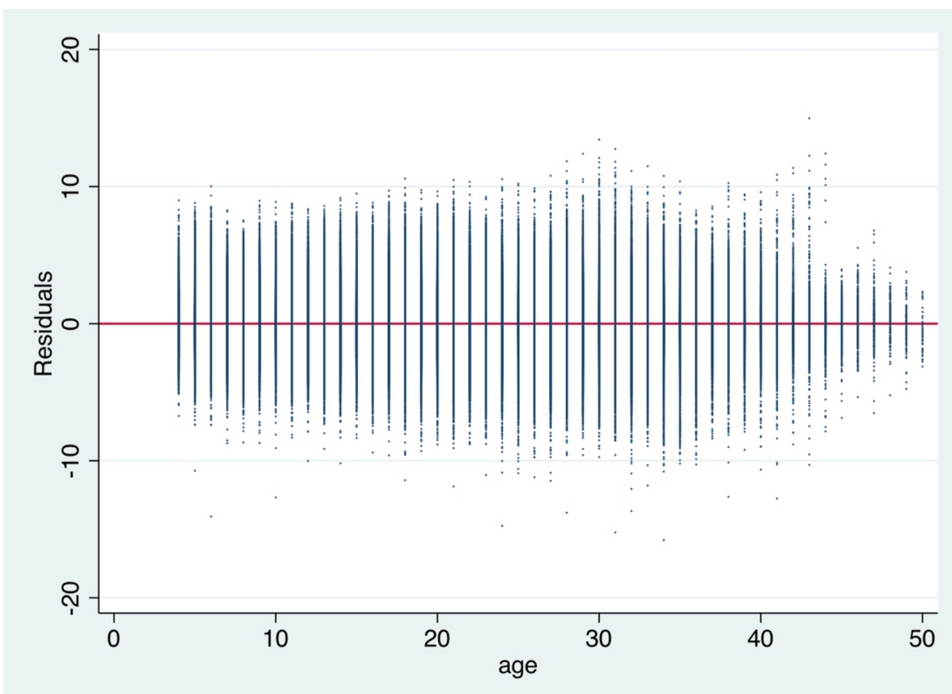



**Figure A3.6.** Residual scatter plot over apartment age.




Note: The lower variance of residuals for apartments older than 44 years could be attributed to the fewer number of transactions of such old apartments within the dataset.

**Figure A3.7.** Comparison of residual scatter plots over fitted values for 1/3rd power transformation (left) and natural log-transformation (right) of resale price.

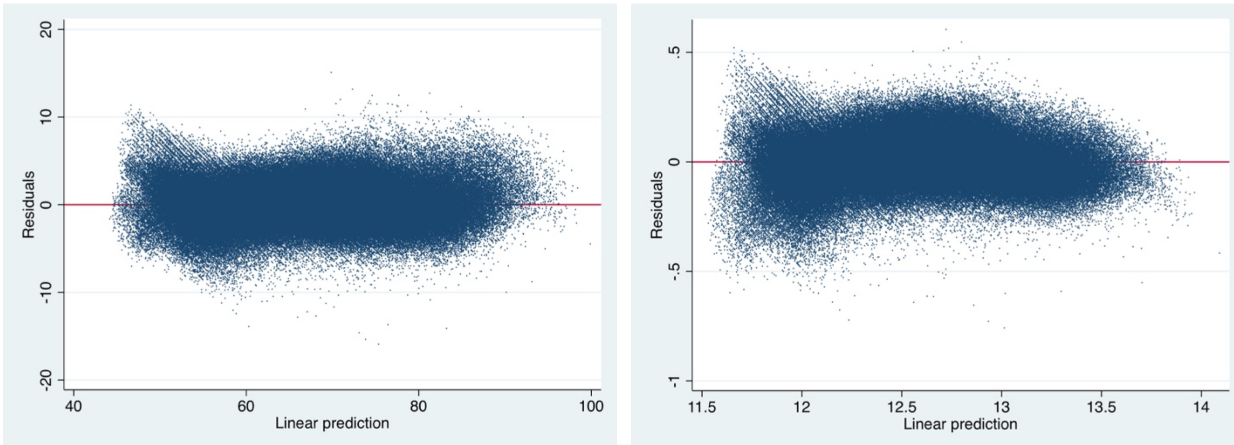


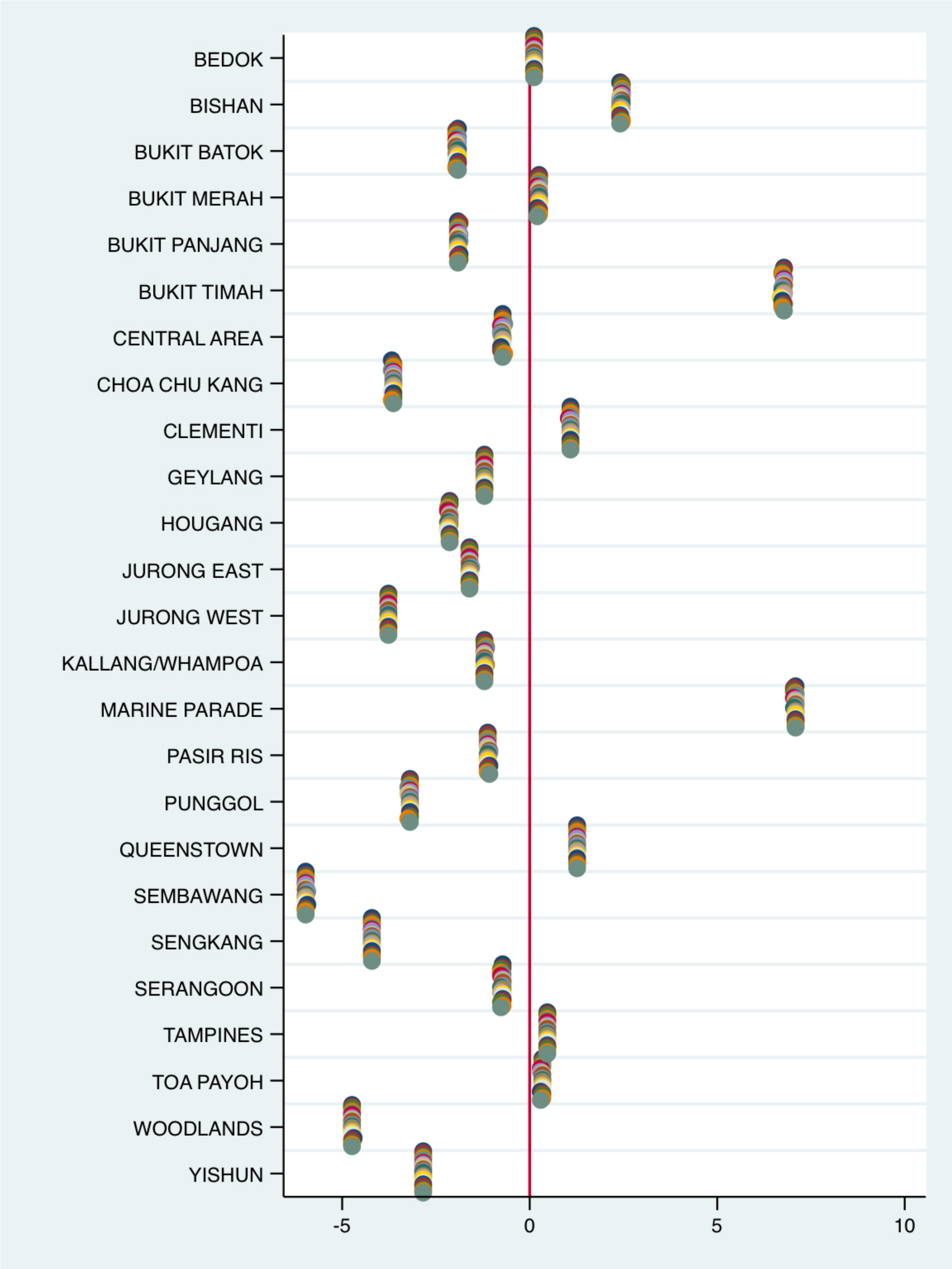

**Table A3.1.** RMSE values of a 20-fold cross-validation test.

| | **RMSE** | *(cont'd)* | **RMSE** |
|---|---|---|---|
| Estimate 1 | 2.127647 | Estimate 11 | 2.148130 |
| Estimate 2 | 2.119066 | Estimate 12 | 2.095337 |
| Estimate 3 | 2.102065 | Estimate 13 | 2.127638 |
| Estimate 4 | 2.125165 | Estimate 14 | 2.108700 |
| Estimate 5 | 2.112558 | Estimate 15 | 2.090798 |
| Estimate 6 | 2.119407 | Estimate 16 | 2.125521 |
| Estimate 7 | 2.108347 | Estimate 17 | 2.138117 |
| Estimate 8 | 2.115677 | Estimate 18 | 2.116284 |
| Estimate 9 | 2.120469 | Estimate 19 | 2.135642 |
| Estimate 10 | 2.129858 | Estimate 20 | 2.141261 |
| **Overall regression RMSE (full sample)** | | | **2.1197** |

**Figure A3.8.** Plots of estimated coefficients from a 20-fold cross-validation test.

**Figure A3.8.** *(cont'd)*

**Figure A3.8.** *(cont'd)*



## Appendix 4. Price Indices

**Table A4.** Hedonic price index, scaled hedonic price index, and the official HDB resale price index.

| Quarter | Hedonic price index | Hedonic price index (scaled)[†] | HDB price index | Quarter *(cont'd)* | Hedonic price index | Hedonic price index (scaled)[†] | HDB price index |
|---|---|---|---|---|---|---|---|
| 1Q2000 | 76.2 | 80.3 | 80.3 | 1Q2009 | 100.0 | 75.6 | 100.0 |
| 2Q2000 | 75.5 | 79.5 | 79.3 | 2Q2009 | 100.2 | 74.0 | 101.4 |
| 3Q2000 | 73.3 | 77.2 | 77.6 | 3Q2009 | 101.4 | 72.4 | 105.0 |
| 4Q2000 | 71.8 | 75.6 | 75.8 | 4Q2009 | 105.3 | 71.1 | 109.0 |
| 1Q2001 | 70.3 | 74.0 | 73.2 | 1Q2010 | 109.9 | 70.1 | 112.1 |
| 2Q2001 | 68.8 | 72.4 | 72.1 | 2Q2010 | 114.1 | 69.9 | 116.6 |
| 3Q2001 | 67.5 | 71.1 | 70.6 | 3Q2010 | 119.6 | 70.3 | 121.3 |
| 4Q2001 | 66.5 | 70.1 | 69.6 | 4Q2010 | 123.5 | 70.2 | 124.4 |
| 1Q2002 | 66.4 | 69.9 | 69.1 | 1Q2011 | 126.5 | 70.2 | 126.4 |
| 2Q2002 | 66.8 | 70.3 | 69.2 | 2Q2011 | 129.9 | 71.1 | 130.4 |
| 3Q2002 | 66.7 | 70.2 | 69.9 | 3Q2011 | 134.8 | 72.7 | 135.4 |
| 4Q2002 | 66.7 | 70.2 | 69.9 | 4Q2011 | 137.3 | 73.6 | 137.7 |
| 1Q2003 | 67.5 | 71.1 | 71.0 | 1Q2012 | 139.8 | 75.0 | 138.5 |
| 2Q2003 | 69.0 | 72.7 | 72.5 | 2Q2012 | 141.7 | 76.2 | 140.3 |
| 3Q2003 | 69.9 | 73.6 | 74.3 | 3Q2012 | 144.2 | 76.3 | 143.1 |
| 4Q2003 | 71.2 | 75.0 | 75.1 | 4Q2012 | 147.4 | 76.4 | 146.7 |
| 1Q2004 | 72.4 | 76.2 | 75.3 | 1Q2013 | 149.7 | 76.6 | 148.6 |
| 2Q2004 | 72.4 | 76.3 | 76.2 | 2Q2013 | 149.9 | 78.0 | 149.4 |
| 3Q2004 | 72.5 | 76.4 | 76.3 | 3Q2013 | 148.0 | 74.8 | 148.1 |
| 4Q2004 | 72.7 | 76.6 | 77.1 | 4Q2013 | 145.2 | 73.3 | 145.8 |
| 1Q2005 | 74.1 | 78.0 | 77.2 | 1Q2014 | 143.3 | 73.2 | 143.5 |
| 2Q2005 | 71.1 | 74.8 | 73.5 | 2Q2014 | 140.7 | 74.2 | 141.5 |
| 3Q2005 | 69.6 | 73.3 | 73.2 | 3Q2014 | 137.8 | 74.5 | 139.1 |
| 4Q2005 | 69.5 | 73.2 | 73.5 | 4Q2014 | 135.3 | 74.9 | 137.0 |
| 1Q2006 | 70.5 | 74.2 | 73.6 | 1Q2015 | 134.6 | 75.7 | 135.6 |
| 2Q2006 | 70.7 | 74.5 | 74.3 | 2Q2015 | 133.9 | 77.1 | 135.0 |
| 3Q2006 | 71.1 | 74.9 | 74.2 | 3Q2015 | 133.2 | 78.7 | 134.6 |
| 4Q2006 | 71.9 | 75.7 | 74.9 | 4Q2015 | 133.1 | 82.1 | 134.8 |
| 1Q2007 | 73.2 | 77.1 | 75.8 | 1Q2016 | 133.6 | 87.5 | 134.7 |
| 2Q2007 | 74.7 | 78.7 | 78.1 | 2Q2016 | 133.4 | 93.1 | 134.7 |
| 3Q2007 | 77.9 | 82.1 | 83.2 | | | | |
| 4Q2007 | 83.1 | 87.5 | 88.0 | | | | |
| 1Q2008 | 88.4 | 93.1 | 91.3 | | | | |
| 2Q2008 | 92.8 | 80.3 | 95.4 | | | | |
| 3Q2008 | 95.9 | 79.5 | 99.4 | | | | |
| 4Q2008 | 99.1 | 77.2 | 100.8 | | | | |

[†] The hedonic price index is scaled to match the HDB price index in the first period (i.e., 1Q2000).