Modeling Household Residential Choice Using Multiple Imputation

By

Weixuan Li

B.E. in Urban and Regional Planning
B.A. in Economics
Peking University
Beijing, China (2011)

Submitted to the Department of Civil and Environmental Engineering and the Department of Urban Studies and Planning in partial fulfillment of the requirements for the degree of

Master of Science in Transportation
Master in City Planning

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2014

Author____

Department of Civil and Environmental Engineering
Department of Urban Studies and Planning
May 20, 2014

Certified by____

Associate Professor P. Christopher Zegras
Department of Urban Studies and Planning
Thesis Supervisor

Accepted By____

Associate Professor P. Christopher Zegras
Chair, MCP Committee
Department of Urban Studies and Planning

Accepted By____

Professor Heidi M. Nepf
Chair, Departmental Committee for Graduate Students
Department of Civil and Environmental Engineering

Modeling Household Residential Choice Using Multiple Imputation

By
Weixuan Li

Submitted to the Department of Civil and Environmental Engineering and the Department of Urban Studies and Planning on May 20, 2014 in partial fulfillment of the requirements for the degrees of Master of Science in Transportation and Master in City Planning

Abstract

Households are one of the core agents in the urban system. Household behavior plays a crucial role in urban system performance and can profoundly shape the urban landscape.

This thesis examines households' behavior in the housing market. Current integrated urban systems models provide few insights into the capability of random bidding models in simulating household residential choice behavior; rarely have random bidding models been applied in a micro-simulation context, due to insufficient data. Therefore, the main goal of this thesis is to explore a possible technique – Multiple Imputation – to integrate observations from dissimilar data sets to meet the data requirements of random bidding models of the housing market, and to test the capability of such a model.

The data used in this thesis come from two distinct data sets: 1) Singapore's Household Interview Travel Survey 2008, providing household demographic, socioeconomic and travel information; and 2) the Urban Redevelopment Authority's Real Estate Information System, which includes detailed descriptions of attributes of private dwellings that were purchased during 2008. Observations from these data sets, households and dwelling units, are firstly matched using Multiple Imputation; the resulting data are used to estimate a random bidding model using the bid-auction approach in a micro-simulation context – providing a rare example of a microscopic application of random bidding models.

This thesis validates the effectiveness of the Multiple Imputation method for matching observations from household and real estate data sets for estimating behavioral models. The estimation of the random bidding model shows that family structure is the most important factor shaping a household's willingness-to-pay for dwelling attributes. Households with children apparently more strongly consider the living environment for their children. Household income influences references as well, but not as much as family structure does. In general, households' willingness-to-pay increase with income level. The limitations of this thesis include the need to and means of grouping households and the few variables to represent dwelling and zonal attributes. Future research should aim to better represent dwelling units and their neighborhoods as well as incorporate more behavioral economics to better understand and predict household behavior in the housing market.

Thesis Supervisor: P. Christopher Zegras
Title: Associate Professor of Transportation and Urban Planning

Thesis Reader: Cristian A. Guevara
Title: Profesor Asociado, Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Chile

Acknowledgements

The writing process of this thesis has been an exciting and rewarding journey. I am grateful to many people who have been granting me support and encouragement throughout my time at MIT.

I would like to extend my deepest gratitude to both of my thesis advisors Prof. P. Christopher Zegras and Prof. Cristian A. Guevara. I greatly appreciate Chris led me into this research and provided generous supports. Both Angelo and Chris has been very supportive and offered me invaluable guidance and instruction.

I wish to thank "Future Urban Mobility" project for providing me such a wonderful opportunity to explore the field I wanted. I am truly thankful to all my colleagues in the project, especially Victor Rocco, Yi Zhu, Shan Jiang, Yunke Xiang, for their support and suggestions on my research. Also to all colleagues in the Long-term team and in the SMART center, I really enjoyed the time with you, in Singapore and MIT. I also want to express my sincere thanks to the Singapore-MIT Alliance for Research and Technology for sponsoring this project.

My MIT experience is invaluable to me. I am grateful to Prof. Nigel H. Willson, Prof. Harilaos N. Koutsopoulos, Prof. Moshe Ben-Akiva, Prof. Joseph Ferreria for all I have learned in and outside the classrooms. I am so fortunate to have such wonderful professors and amazing schoolmates, peers, and friends who I share my unforgettable memories with.

To my dearest mom and dad, thank you for being the best parents I can ever imagine.

Finally, to my beloved fiancée, Ning, you have always filled me with love. Thank you for your unfailing support, continuous encouragements, and a patient ear. I am so fortunate to have you.

(This research is supported by the Singapore National Research Foundation under the Singapore MIT Alliance for Research and Technology (SMART) Future Urban Mobility IRG. The views expressed herein are those of the author and are not necessarily those of the Singapore National Research Foundation.)

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

With ongoing urbanization, more and more people live in cities around the globe. According to the United Nations (United Nation, 2011), in the year 2010, the population living in cities exceeds the population living in rural areas for the first time in history. Cities are the places where many complex systems (like land use, housing, transportation and utilities) interact with each other and major economic and social activities take place.

Understanding urban system dynamics is extremely important for policy making, since city-scale interventions can be powerful, expensive, often irreversible, and relevant to the lives of millions of people. To represent and analyze this complexity, models of urban systems serve as important tools for policy analysis and decision support. It is not easy to understand the urban system because so many sub-systems like land use, transportation etc. interact with each other, together with a large number of heterogeneous, interacting agents.

Households are one of the core agents in the urban system. Household behavior plays a crucial role in urban system performance and can profoundly shape the urban landscape. Household choices of residence, work, shopping and entertainment condition individuals' travel patterns and a city's aggregate travel demand and will be reflected in urban land uses. A comprehensive micro-simulation approach, able to reflect household behaviors can, thus, be a potentially powerful tool to forecast future urban mobility system performance and policy and investment effectiveness.

A general framework to model household decisions and their interactions with other parts of the urban system can be framed in three interrelated stages, corresponding to urban development or land-use, travel demand, and transportation system performance (Figure 1-1), which correspond, respectively, to long-, medium-, and short-term decisions (years, months; days, hours; seconds, fractions of seconds). Depending on the level of detail and the scale of the analysis, this framework can be used to model aggregated flows of trips by geographical zones or disaggregated household and individual decisions that influence relevant demands, including mobility and lifestyle, activity/travel scheduling, and others (Ben-Akiva et al., 1996).

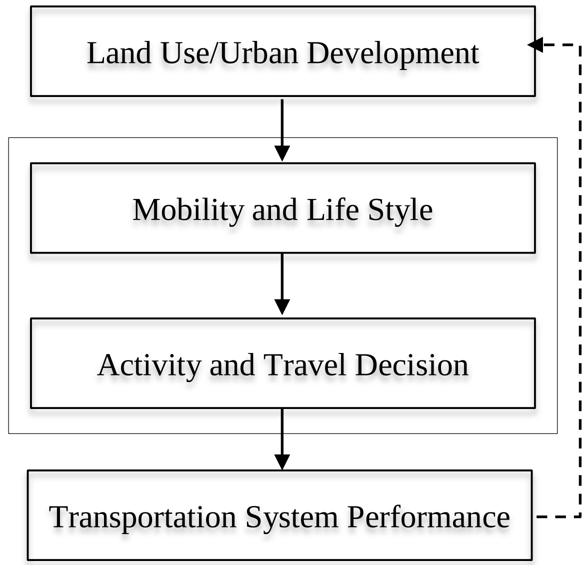


Figure 1-1 Urban System Modeling Framework

In this general framework, the land use/urban development system, representing long-term behaviors in the urban system, plays a key role. The land market itself is a complex and dynamic system with several core interacting agents like real estate developers, firms, households and governments. In this research, I only focus on the residential market, which represents a key long-term sub-market as it both conditions and is conditioned by a large share of total urban travel demand and supply. Households and real estate developers are the core agents in the residential market. This thesis focuses on households and, specifically, residential choice models.

1.2 Statement of the Problem

Current micro-simulation models of the urban system rely on hedonic pricing and residential location choice models to represent the housing market (Wegener and Spiekermann, 1996; Vencatasawmy et al., 1999; Holm et al., 2002; Ettema et al., 2007; Miller et al., 2005). Such an approach has two major drawbacks: it fails to properly account for the market clearance process and fails to model price evolution over time. In contrast, random bidding models of the housing market can account for price evolution and market clearance. However, such models have rarely been applied in a micro-simulation context and they rely on some arguably unrealistic assumptions due to modeling constraints, such as assuming all agents bid on all dwellings in the urban system.

Despite the potential theoretical superiority of the random bidding models, insufficient data also hamper the application of such models at a microscopic level. The data requirements are high and the ideal dataset does not exist. Thus, innovations for making the most of available datasets and/or collecting the appropriate datasets are necessary.

1.3 Thesis Objectives and Outline

This thesis aims to

1) Identify and address the challenges involved in estimating random bidding models of the housing market in a micro-simulation context, to fill the gap in the microscopic application such models
2) Explore possible techniques to enable the use of available data for the calibration and validation of random bidding models; such techniques will be demonstrated using data from Singapore
3) Identify the data needs to model the full housing market using the techniques developed, and develop the survey instruments that would satisfy the data needs.

This thesis is composed of six chapters including this introduction. Chapter 2 reviews the theoretical framework of residential choice and its two main modeling approaches and then analyzes the existing micro-simulation models and residential mobility models, comparing to SimMobility, the large-scale urban micro-simulation modeling project within which this thesis is embedded.

Chapter 3 details the research context by summarizing the economic, transportation and urban development background of Singapore and introduces the SimMobility framework upon which this research is built.

Chapter 4 introduces the currently available datasets in Singapore and briefly discusses their problems. In this chapter, I will also illustrate the matching technique for working with inconsistent data and the way such data available in Singapore can be used for the residential models.

Chapter 5 demonstrates an estimation of one of the random bidding models using the data from the previous chapter to investigate the taste preferences and willingness-to-pay for Singapore households.

Chapter 6 concludes the thesis, summarizing key research finding and identifying possible further research. Chapter 6 also proposes a survey instrument design to collect sufficient data for further model use.

Chapter 2

Residential Mobility and Location Choice

In the urban context, *residential choice*, especially *residential location choice*, can be described as a matching process, during which agents (like households, firms and developers) choose among location alternatives (like dwelling units or buildings). Meanwhile, *residential mobility* refers to the spatial movement of individuals and households between residential dwellings within an urban area. Modern mobility systems and some of their accompanying devices, like smart phones and smart cards, generate increasingly cheap and abundant data streams, making the study of human behavior at a micro level possible. The available datasets also make possible the micro-level modeling of residential mobility and location choices. At the micro level, location choices are modeled incorporating decision-making processes. This chapter first reviews the theoretical framework of residential choice, and the two main approaches to represent this behavior—the choice approach and the bid-rent approach—and their applications. Within this theoretical framework, this chapter further reviews the state of the art in residential location choice and residential mobility modeling.

2.1 Residential Location Choices: The Choice Approach

2.1.1 Random utility maximization (RUM)

Residential location models rest on a basic microeconomic framework. This framework assumes that the ultimate goal of a household's behavior is to maximize the combined utility for all of its members, given income constraints:

$$\max_{x,i} U(x, z_i) \tag{2.1}$$
$$s.t. px + r_i \leq I$$

Households maximize their utilities by choosing a vector of continuous goods x and a discret residential location, i, described by a set of attributes, z_i, where z_i includes both building environment attributes, and the time and goods associated with the activities that each household member performs (e.g., transportation, work, and other derived activities) (Martinez, 1992, 1996, 2000; Guevara-Cue, 2005). The utility is constrained by a household's income I, which means that the total amount of money spent on goods with price, p, and the price of location, r_i, must be less than or equal to the household's available income. Solving this problem of x while assuming constant income, the objective function can be rewritten, as Rosen (1974) proposes:

$$\max_{i} V(p, I - r_i, z_i) \tag{2.2}$$

The optimal solution to the objective function of the utility maximization problem is known as *the conditional indirect utility* V on location i. It is also assumed that the household's utility regarding a dwelling is composed of a conditional indirect utility function and a random component, which is referred to as the Random Utility assumption (Ben-Akiva and Lerman, 1985). Under the random utility assumption, households (h) have different preferences and vary in their tastes, and the location alternatives (i) have unobserved attributes. The random component is set to capture such differences in preferences, tastes, and unobserved attributes. Hence, we can rewrite the utility of (2.2) as $V_{ih} + \varepsilon_{ih}$, where ε is a stochastic error term accounting for taste heterogeneity and the unobserved attributes of locations. Based on the random utility assumption, the probability of a household h choosing location i can be expressed as the probability that location i provides the maximum utility for household h:

$$P(i|h) = Prob\{V_{ih} + \varepsilon_i > V_{jh} + \varepsilon_j \ \forall \ j \neq i\} \tag{2.3}$$

Assuming an Extreme Value distribution under the independent and identically distributed (iid) assumption for the error term of the utility function, the probability of household h choosing a location i can be derived as (McFadden, 1978):

$$P(i|h, S) = \frac{e^{\mu V_{ih}}}{\sum_{j \epsilon S} e^{\mu V_{jh}}} \tag{2.4}$$

where μ is a positive scale parameter and S is the set of available locations the household can choose from. The choice probability as a closed-form expression shown in (2.4) is also known as the Multinomial Logit (MNL) formulation and is the standard model for discrete choice, which has been widely applied to residential location choice models. The MNL model assumes Extreme Value distribution for the error term, while other assumptions of the distribution of the error terms lead to various choice models such as probit and nested logit. The probit model is inapplicable to the present research because it does not have a closed form and the choice probability is an integral. A review of the nested logit models follows.

2.1.2 Nested logit

According to McFadden (1978), the multinomial logit (MNL) form enables the consistent estimation of choice models from a subset of alternatives, where the property of independence from irrelevant alternatives (IIA) is assumed. However, the MNL model will lead to biased results if the IIA assumption is violated (e.g., the alternatives in the choice set are correlated). In such cases, a nested logit (NL) structure can be introduced to capture the correlations in the choice set (Ben-Akiva and Lerman, 1985).

In the NL model, the probability of a household choosing location i is then defined as:

$$P(i) = P(i|m) \cdot P(m) \tag{2.5}$$

where m is the nest and i is the location alternative, and $P(m)$ is the marginal probability of choosing nest m while $P(i|m)$ is the conditional probability of i being chosen, given the nest m. The conditional choice probability for the bottom-level choice of alternative i inside nest m is equivalent to the standard multinomial logit form:

$$P(i|m) = \frac{e^{\mu V_i}}{\sum_{j \epsilon S_m} e^{\mu V_j}} \tag{2.6}$$

where V_i represents the observable components of the utility function for each location alternative i, and μ is the associated positive scale parameter. The marginal choice probability of choosing nest m is:

$$P(m) = \frac{e^{\mu_m V'_m}}{\sum_{m' \epsilon M} e^{\mu_m V'_{m'}}} \tag{2.7}$$

where μ_m is the scale parameter for nest m, and V'_m is the logsum associated with nest m. The logsum represents the expected value of the maximum random utility of all alternatives in the nest m:

$$V'_m = \frac{\ln(\sum_{i \epsilon S_m} e^{\mu V_i})}{\mu} \tag{2.8}$$

The nested logit model relaxes the IIA assumption, with the correlations captured by the nests. This nested structure is likely similar to most actual housing markets, where dwelling units fit into different market segments (i.e., nests). Such a nested structure can be used to model a whole urban housing market, which will be discussed in Chapter 6.

2.1.3 Hybrid choice approach

Despite extensions like the NL model, the standard discrete choice model has been criticized for at least two main shortcomings. First, it is too simple to adequately model behavior, representing the decision-making process like a black box (McFadden, 1999). Second, it does not adequately capture taste heterogeneity, i.e., the fact that different people have different sensitivities to attributes of the same alternatives. The taste parameters are specified as constants that do not vary over individuals, and taste heterogeneity is represented by interacting socio-demographic variables with alternative attributes.

Walker (2001) introduced latent variables and latent classes in response to the criticisms of the standard discrete choice model. Walker's approach has been termed the Hybrid Choice model (HCM) (Abou-Zeid and Ben-Akiva, 2013). The aim of the HCM is to extend the standard discrete choice model to account for the effects of latent variables such as knowledge, perceptions, attitudes, choice sets, decision protocols, etc. on choice and to capture systematic taste heterogeneity. The HCM framework is shown in Figure 2-1.

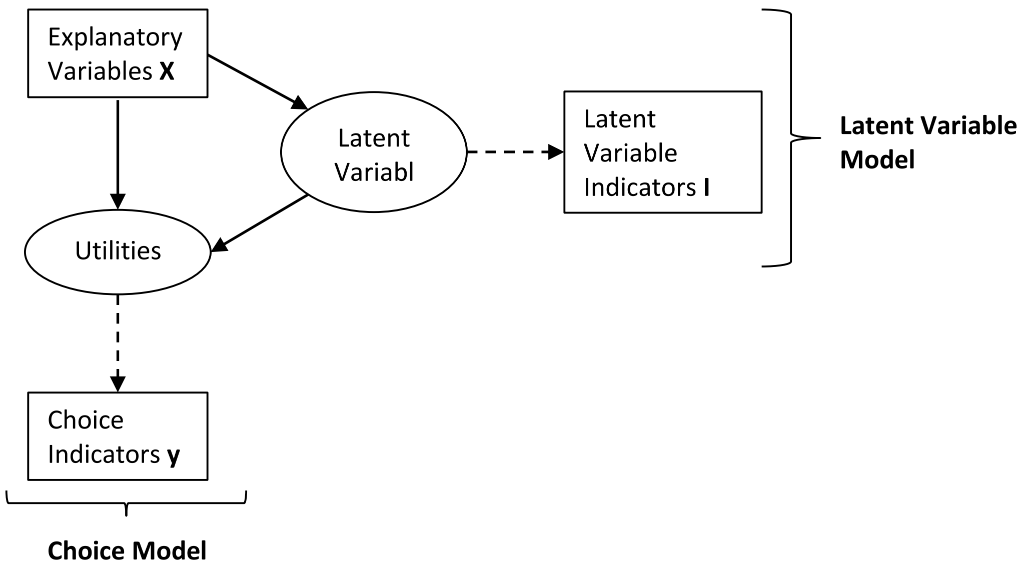


Figure 2-1 Integrated Choice and Latent Variable Model (Walker, 2001)

With HCM, choice sets are better represented as latent variables (Abou-Zeid and Ben-Akiva, 2013), because in addition to observed socio-demographic variables that determine the choice set (car availability, driver's license, etc., in the context of mode choice), the perceived availability of alternatives may depend on subjective factors like the individual's travel attitudes and perceptions of the attributes of the modes (Abou-Zeid and Ben-Akiva, 2013)..

Walker and Li (2007) apply the HCM approach in residential location choice incorporates lifestyle as a latent variable or latent class into the HCM framework. Incorporating lifestyle factors in household choice is a relatively new trend in the behavioral modeling field. Regarding location choices, Aeroe (2001) describes lifestyle as the "deep-rooted and embedded prevalent attitudes towards different types of residential areas." Walker and Li (2007) attempt to operationalize the lifestyle construct in household residential location choice with the first application of latent class choice. In their approach, which simultaneously estimates class membership and class-specific residential choice, the classes represent "lifestyles."

The idea of incorporating lifestyles and using latent class choice models to categorize households and predict their residential location choice is adapted in the SimMobility platform, which will be introduced in the next chapter.

2.2 Residential Location Choices: The Bid-Auction Approach

The choice approach to residential location focuses on the decision-making process from the individual household side of the market, while the bid-auction approach pays more attention to the interactions among households in the housing market. The housing market is assumed to function as an auction market where households bid according to their willingness to pay for a residential unit (Alonso, 1964). The household with the highest willingness to pay and placing the highest bid will get the unit at the price set as its bid in the auction process.

Under the general microeconomic framework, willingness to pay can be derived from the utility maximization problem under income constraints. The maximum indirect utility, V, derived from equations 2.1 and 2.2, can be inverted in the dwelling-price variable (Jara-Díaz and Martínez, 1999):

$$r_i = I - V^{-1}(V, p, z_i) \tag{2.9}$$

Under the auction market assumption, the dwelling-price variable r_i can be treated as the willingness to pay for location i. Assuming r_i is the willingness to pay of household h, the household places the bid based on r_i. Therefore the bid function B can be presented as:

$$B_{hi} = I_h - V_h^{-1}(V_h, p, z_i) \tag{2.10}$$

According to Ellickson (1981), we can also write the bid (or bid-rent) B_{hi} as a function of location attributes: $B_{hi}(z_i)$. Taking the unobserved heterogeneity in preferences among households captured by a random term, ε_h, the bid can be written as $B_{hi}(z_i) + \varepsilon_h$. Therefore, the probability that household h occupies the residential dwelling i is the probability that household h places the highest bid for dwelling i, beating all other competing households:

$$P(h|z_i) = Prob\{B_{hi}(z_i) + \varepsilon_h > B'_{h'i}(z_i) + \varepsilon'_{h'}, \forall\, h' \neq h\} \tag{2.11}$$

If ε_h is iid Extreme Value distributed, the probability of household h winning dwelling i can be presented as:

$$P(h|z_i) = \frac{e^{\mu B_{hi}}}{\sum_{h^* \epsilon H} e^{\mu B_{h^* i}}} \tag{2.12}$$

Ellickson's method serves as an alternative to the hedonic price model (Rosen, 1974). The hedonic theory of urban housing market focuses on housing's underlying characteristics and treats a housing unit as a collection of its component attributes. In hedonic housing market models, the characteristics of housing determine its (hedonic) price, and this price is matched with households' bids to achieve market equilibrium (Mas-Colell, 1975). However, hedonic theory is biased in the bid price interpretation, unless it accounts for the fact that the current residents have the highest willingness-to-pay, as pointed by Lerman and Kern (1981). Ellickson tackles this problem and estimates the willingness to pay for housing attributes of different groups of households by likelihood maximization:

$$L = \prod_{i \in S} (\prod_{h \in H} P(h|z_i)^{y_{hi}}) \tag{2.13}$$

where y_{hi} is a binary indicator that assumes the value of one if household h is observed to be located in dwelling i and zero otherwise. In order to simplify the bid function in applying his model, Ellickson aggregates households into homogenous groups and estimates a linear-in-parameters bid function for each of them.

Ellickson's model is not, however, capable of identifying the scale parameter μ; it can only estimate relative parameters. Lerman and Kern (1983) extend Ellickson's model by including the observed highest bid from the last period into the probability density function of the bid. Denoting the highest bid or, say, price paid as P^* from last period, this information allows for specification of the following probability density:

$$P(h|z_i) = Prob\{B_{hi}(z_i) + \varepsilon_h = P \ and \ B_{hi}(z_i) + \varepsilon_h > B'_{h'i}(z_i) + \varepsilon'_{h'}, \forall \ h' \neq h\} \tag{2.14}$$

Assuming that the error terms are Extreme Value distributed, (2.13) can be written as:

$$P(h|z_i) = f_\varepsilon(P^* - B_{hi}(z_i)) \prod_{\forall \ h' \neq h} F_\varepsilon(P^* - B'_{h'i}(z_i)) \tag{2.15}$$

with the probability density function f_ε and cumulative density function F_ε given by:

$$f_\varepsilon = \mu \exp(-\mu\varepsilon) \exp(-exp(-\mu\varepsilon)) \tag{2.16}$$

and

$$F_{\varepsilon} = \exp(-exp(-\mu\varepsilon)) \tag{2.17}$$

Therefore, the new likelihood function to estimate the parameters of the willingness to pay is:

$$L = \prod_{i\in S}\big(-\mu exp(-\mu\big(P^{*} - B_{hi}(z_i)\big)\prod_{h'\in H}(exp(-\mu(P^{*} - B_{h'i}(z_i)))\ \big)^{y_{hi}} \tag{2.18}$$

where H is the total number of households participating in the auction and S is the total number of dwellings in the market. If the bid function in linear in its parameters, the parameters can be estimated consistently.

Lerman and Kern's approach solves the problem of under-determination in Ellickson's approach and generates absolute estimates of willingness to pay for location attributes.

Horowitz (1986) goes a step further by using the seller's asking price to truncate the distribution of the bids, bringing in a sequential bidding process including a comparison to a household's current dwelling, and calculating the probability for a household to place a bid for a certain dwelling. Assuming a household must choose between a dwelling i, and its current location c, if the individual household knows that the probability of being the highest bidder with a bid b is $P_i(b)$, the expected value of getting dwelling i can be expressed as:

$$E(V_i) = P_i(b)[V_h(b, z_i) - V_{hc}] + V_{hc} \tag{2.19}$$

where $V_h(b, z_i)$ is the value of dwelling i to household h based on the bid and the attributes of the dwelling, and V_{hc} of value household h's current dwelling. Since the utility of the current dwelling can also be written as a function of the willingness to pay and the dwelling attributes, then we have:

$$V_{hc} = V(B_{\,hc}(z_c)\ , z_c) \tag{2.20}$$

where $B_{\,hc}(z_c)$ is the price the household paid for its current dwelling. Therefore the utility function $V^{*}(b)$ for placing a bid b is:

$$V_h^{*}(b) = E(V_i) = P_i(b)[B_{hc}(z_c) - b] + V_{hc} \tag{2.21}$$

Maximizing (2.21), the optimum bid should accomplish:

$$b_{hi} = B_{hc}(z_i) - \frac{P_i(b)}{P_i'(b)} \tag{2.22}$$

Taking the unobserved heterogeneity in various preferences among households captured by a random term ε_h, the bid can be written as:

$$b_{hi} = B_c(z_i) - \frac{P(b)}{P'(b)} + \varepsilon_h \tag{2.23}$$

Therefore, the probability of household h having a bid equal to the observed price B_{hi}^* is given by:

$$P_{hi}(B_{hi}^*) = f_\varepsilon\,(b_{hi} - B_c - \frac{P(b)}{P'(b)}) \tag{2.24}$$

Let V_{hi}^* denote the value of $E(V)$ corresponding to household h, dwelling i and bid b_{hi}^* . Assuming that the household places the bid on the dwelling that maximizes its expected utility --$\max_{k\neq h} V_{ki}^* < V^*{}_{hi}$, we have:

$$V_{hi}^* = E(maxU) = P_{hi}(b)[B_c(z_i) - b] + V_c \tag{2.25}$$

If the variables $V^*{}_{ki}$ with $k \neq h$ constitute a sequence of iid random variables, then $\max_{k\neq h} V_{hi}^*$ follows a Gumbel distribution, giving us the probability of a household making an offer for the location i:

$$\text{Prob}\left(\max_{k\neq h} V_{ki}^* < V_{hi}\right) = \exp\left(exp\left(-\frac{V_{hi}^* - V_c - \mu}{\eta}\right)\right) \tag{2.26}$$

where μ and η are constant parameters whose values may depend on the observed attributes of household h .

Although Horowitz's approach and Lerman and Kern's approach are preferable to Ellickson's approach theoretically, implementing Lerman and Kern's approach and Horowitz's approach requires data that are, in general, either nonexistent or not easy to collect on a large scale.

Several empirical applications of Ellickson's and Lerman and Kern's method can be found. In general, the literature shows the bid-rent approach outperforms the hedonic approach and the bid-rent models provide information on household behavior by estimating the willingness to pay for different groups of agents. Gross (1988) finds that the bid-auction approach performs better than the hedonic models for predicting rents and marginal willingness-to-pay predictions, using Bogota, Colombia as a case study. Gross et al. (1990) find that the bid-rent model works well in estimating housing price if the problem of household group division can be resolved. Chattopadhyay (1998) uses Chicago as a case study and concludes that the bid-auction approach has the advantage of providing estimates of willingness to pay for different groups of agents, although it does not show an obvious advantage over the hedonic approach in estimating rents. Muto (2006), examining Tokyo, compares Lerman and Kern's method with Ellickson's approach, and finds systematic difference between land value and land use type. He also finds systematic deviations in the location choice results between using Lerman and Kern's approach and the observed location distribution in Tokyo, which may be caused by the correlation between prices of different dwellings. This deviation suggests that the Lerman and Kern approach sacrifices location forecast capabilities for absolute price estimation. All of these studies have their agents grouped into homogenous groups: Gross et al. (1990) use only one homogeneous group – the "baby condo"; Chattopadhyay (1998) categorized households into four income groups; and Muto (2006) used residential and commercial land use as two homogeneous groups. For each homogeneous group, they estimate group-specific parameters instead for of an individual, household or firm. Moreover, the estimation is done only over a small sample of locations, where detailed attributes and price information are available. None of the existing models has been applied at a microscopic level for the whole city, likely due to data constraints. In later chapters, I aim to fill the gap in city-scale microscopic applications of the bid-rent approach. Table 2-1 shows a summary of those empirical applications.

Table 2-1 Empirical application summary

Author	Gross (1988)	Gross et al. (1990)	Chattopadhyay (1998)	Muto (2006)
Setting	Bogota, Colombia	Baton Rouge, Louisiana, area	Chicago	Tokyo, Japan
Empirical date	1978	1983-1985	1989-1990	2000
Data source	Renter household survey data	Housing transaction data	Housing transaction data	Land prices
Comparison	Bid-rent model with the hedonic model	None	Bid-rent model with the hedonic model	Bid-rent models
Approaches	Ellickson's approach	Lerman and Kern's approach	Lerman and Kern's approach	Lerman and Kern's approach
Metrics compared	Four household types based on income (rich/poor) and size (large/small)	None, only included one type of housing	Four household types based on income (higher/lower than $40,000) and children(with/without children)	Two land use type(residential/commercial)
Findings	The bid-auction approach performs better than the hedonic models for predicting rents and marginal willingness-to-pay predictions	Bid-rent model works well in estimating housing price if the problem of household group division can be resolved	The bid-auction approach has the advantage of providing estimates of willingness to pay for different groups of agents, although it does not show an obvious advantage over the hedonic approach,	Lerman and Kern's estimation works, but also results in a systematic difference between land value and land use type.

2.3 Residential Mobility and Location Choice Models

Based on the theoretical approaches introduced in the previous sections, a number of simulation models have been developed; these models go beyond estimation and try to predict future residential mobility and location patterns on a large scale.

2.3.1 Microsimulation Models

Rossi (1955) shifted the residential mobility research focus from aggregate spatial patterns to behaviors at individual/household level. Micro-simulation modeling was first introduced by

Orcutt (1957). According to Mitton (2000),

> Microsimulation models use micro-data on persons (or households, or firms or other micro-units) and simulate the effect of changes in policy (or other changes) on each of these units. Differences before and after the changes can be analyzed at the micro-level; and also aggregated to show the overall effect of the changes. It is the dependence on individual information from the micro-data at every stage of analysis that distinguishes microsimulation models from other sorts of economic, statistical or descriptive models.

Microsimulation models of urban land use and transportation are considered state-of-the-art (Wegener, 1998).

Ben-Akiva and Bowman (1998) present a theoretical framework for microscopic modeling of residential, activity, and travel decisions for households and individuals, and the interactions between urban development, household decisions and transport system performance (Figure 2-2). In this thesis, I focus on long-term household decisions related to residential location.

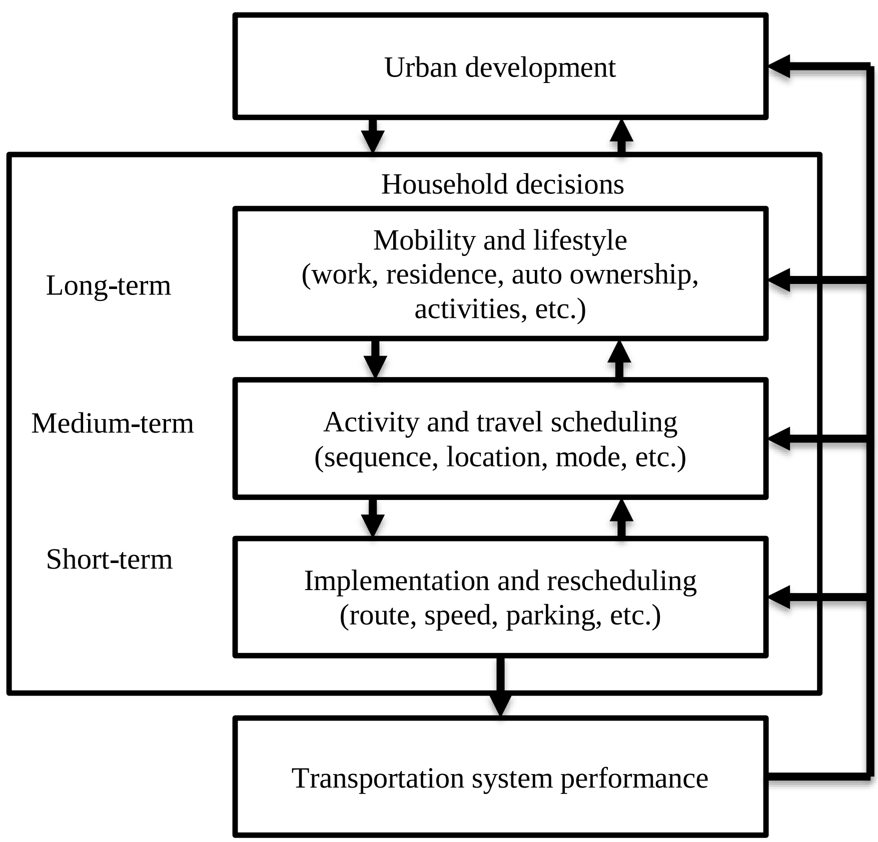


Figure 2-2 Households' decisions in the long-, mid-, and short-terms (source: adapted from Ben-Akiva & Bowman, 1998)

Researchers have long tried to estimate households' spatial movement between dwellings within an urban area. Existing residential mobility models tend to focus on modeling the process of households' moving decisions and choices.

2.3.2 Microsimulation Models/Modules for Residential Mobility

Brown and Moore (1970) divide the mobility process into two stages: the decision to move and the choice of location. People grow aware of the possibility to move and then make decisions to either relocate or stay. This awakening process is triggered by households' progression through their lifecycle (like marriages and deaths), according to Rossi (1980) and McCarthy (1976). Once they decide to relocate, people need to choose a dwelling with certain attributes that satisfy their needs. Studies have shown that changes in family size and household structure are major determinants of short-distance residential relocations (Mulder and Wagner, 1993; Clark and Huang, 2003). Other researchers argue that additional social and economic factors like education or changing of job also provoke moving activities.

Researchers have built several residential mobility microsimulation models and applied them to the context of different countries and regions all over the world:

- UrbanSim (Waddell, 2001, 2004; Waddell et al., 2003)

  UrbanSim is an open-source platform for urban models[1] and has been applied to multiple regions. UrbanSim models residential mobility in two of its core models: the demographic transition model and the household location-choice model. The demographic transition model models the formation and dissolution of households following a pre-determined, exogenous population distribution. Households' movement probabilities are calculated from historical data. Based on those demographic changes, the location-choice model is specified as a multinomial logit model. Location choices are made based on the housing characteristics, zonal accessibilities and neighborhood environment with the random sampling of all vacant dwellings. However, one critical limitation of UrbanSim is that the dwelling price is based on hedonic price theory, not allowing the model to capture the effect of market conditions like

[1] Source code is available at http://www.urbansim.org

supply and demand surplus and thus introducing biases (Hurtubia, 2012).

- ILUTE Modeling System (Miller et al., 2005)

The Integrated Land Use, Transportation, Environment (ILUTE) Modeling System aims to simulate the evolution of an entire urban region over an extended period of time using an agent-based, dynamic microsimulation model. The demographic changes of a household are modeled dynamically: the model not only intends to capture individual characteristics' changes over time, but also to model life-cycle events such as fertility, mortality, and household formation and dissolution. The ILUTE model tries to represent the housing market in three processes: a demand process, a supply process and a market clearing process, where a household may decide to become active in the market during the demand process and bid for a dwelling during the market clearing processes. The market clearing process simplifies housing market dynamics as a fixed-price process, in which the prices are either fixed by policy or adjusted globally. A simple operational model has been applied to Toronto, Canada. With a main focus on land use and residential markets, some other components in ILUTE, like a labor market model and firmographic model, are still under development. .

- IRPUD (Wegener and Spiekermann, 1996)

The IRPUD model is a simulation model of intraregional location and mobility decisions in a metropolitan area (Wegener and Spiekermann, 1996). It was initially designed and implemented since 1977 at the Institute of Spatial Planning of the University of Dortmund (IRPUD). The IRPUD model uses an ageing sub-model to capture demographic changes, and uses a housing market sub-model to simulate intraregional migration decisions of households as a searching process in the regional housing market. For both models, individuals and households are aggregated by type. The ageing sub-model updates the demographic changes with life-cycle events (like formation and dissolution of a household) first and then passes households to the housing market sub-model to predict their probabilities of moving and their location choices through a fixed price market clearance process. The operational model has been applied in Dortmund, Germany. IRPUD uses micro level data but aggregates to a zonal level for modeling. After such an aggregation, IRPUD models zonal level data, and thus is not considered as individual-level simulation model.

- SVERIGE (Vencatasawmy et al., 1999; Holm et al., 2002)

SVERIGE (System for Visualizing Economic and Regional Influences in Governing the Environment) is a national-level dynamic spatial microsimulation model of economic-demographic effects. It models the movement of individuals and households from one state to another in Sweden. The household structure in SVERIGE is predetermined by population synthesis (mortality, emigration and then fertility). Then life-cycle events are modeled in order of execution: education, marriage, leaving home, divorce. At last, the migration of those individuals is determined by logistic regression models with labor market adjustments. The operational model has been applied to Sweden. The SVERIGE model only focuses on modeling residential mobility on a large scale (interstate movement) and it only considers the life-cycle change from the household side. Therefore, SVERIGE ignores the complex housing market in their model, and do not include other parts of urban system like real estate developers into the model.

- SMILE (Ballas et al., 2005)

SMILE (Simulation Model for Irish Local Economy) is a static and dynamic population and spatial microsimulation model designed to capture the population changes in rural Ireland. The simulation first goes through a static process to create the base disaggregate spatial population (synthetic population) and then the dynamic process ages the population by evaluating three individual-level processes: fertility, mortality and migration. These three processes are determined by probabilities based on age, gender and location derived from census data. The operational model has been applied in Ireland. SMILE, with a very limited scope, only focuses on the population evolution for nation-level population projections over time and space. SMILE is capable of modeling spatial population changes but is not designed for modeling the urban system as a whole.

- PUMA (Ettema et al., 2007)

The PUMA (Predicting Urbanization with Multi-Agent) model is a multi-agent system working at the disaggregate level to represent land-use changes in a behavioral way. The model is composed of three modules: a land conversion module, a household module, and a firm module. The life-cycle events are embodied in the PUMA model to update the characteristics of each individual year by year. The residential choice model in PUMA is structured as a nested logit model. The upper-level decision is whether or not to search for another dwelling, which depends not only on the characteristics of the current dwelling, but also on the expected utility of moving to another dwelling; the second-level decision is to move or not, which also depends on characteristics of the current dwelling and on the expected utility of moving to another dwelling; the bottom-level decision is choice of one out of a set of available dwellings. The operational model has been applied in the northern part of the Dutch Randstad. The key assumption of the residential choice model in PUMA is that households try to increase their lifetime utility by moving to another dwelling. This assumption of maximizing lifetime utility when households make moving decisions is, however, unrealistic. Moreover, PUMA does not include the housing market into the operational model and only models the one-way impact of household choices on the dwelling supply.

Note that the PUMA model also tries to represent the awakening process when a household becomes aware of the fact that it can improve its utility by moving to another dwelling. The awakening process is inspired by: 1) push factors, which relate to changes in the household or in the living conditions (e.g. a change in household composition or finding a new job elsewhere); and 2) pull factors, which relate to the opportunity to find a better dwelling elsewhere. The push and pull factors are like market segregations to distinguish different consumers. But the PUMA model doesn't explore this part explicitly; this part is better explored in the LocSim model.

- LocSim (Oskamp, 1993, 1994a, 1994c)

LocSim (Local Simulation) aims at simulating developments in demographic structures, housing markets, and the interrelation between them, but only operates at the local level and for the short term. LocSim is comprised of three sub-models: the demographic model, which is similar to SVERIGE; the housing supply model, which simulates the *exogenous* changes on the supply side of the housing market; and the housing market model. The occurrence of events is based on probability distributions by age, gender and other characteristics. LocSim captures the reasons to move by three types of moves or moving desires indicated or triggered by demographic changes: 1) implied move, which is event-dependent, due to the occurrence of some demographic events that might imply a relocation (e.g. leave the parental home); 2) preferred move, which is state-dependent, inferred by either demographic events (e.g. change in household size may lead to dissatisfaction with the current housing situation) or dwelling characteristic changes that might trigger a relocation; and 3) forced move, due to unit demolishment or eviction. Forced moves are out of the control of households and thus cannot be postponed. The LocSim model is strong on modeling decision-making for household searching and moves, but it does not includes housing price and household income dynamics and cannot facilitate the calculation of spatial characteristics.

Tables 2-2 and 2-3 summarize and compares these models.

Table 2-2 Comparisons of Operational Microsimulation Model Characteristics 1

	Model type	Estimation method	Decision-making basis	Theoretical foundation	Main focus
UrbanSim	Behavior model of decision-making units	Probability	Movement probabilities based on historical data.	Utility maximization and market clear	Decision-making between alternative dwellings
ILUTE	Integrated urban modeling system	Exogenous	Push/pull factors and exogenous policies	Utility maximization	
IRPUD (ILUMASS)	A simulation model of intraregional location and mobility decisions	Joint probability	The probability that the household searches in certain zone	Utility maximization and market clear	
SVERIGE	National-level time-geographic microsimulation	Probability	Life events and stepwise logistic regression of the head of household	Demographic microsimulation	Demographic-changes-triggered changes in demand for new dwellings and migration
SMILE	Spatial microsimulation model for demographic changes	Joint probability	Age, gender and county location	Spatial microsimulation (geographical information to micro-level data)	
PUMA	Multi-agent, behavior microsimulation model	Comparison between utilities	Utility of current dwelling and perception of the market	Utility maximization	Demographic-changes-triggered changes in demand for new dwellings and migration
LocSim	Local-levels, short-term, agent-based, simulation model (the probabilistic heuristic search model)	Capture demographic changes	Demographic changes, urgency and search intensity	Market clear	

Table 2-3 Comparisons of Operational Microsimulation Model Characteristics 2

	Choice Scale	Attributes Considered	Operational Model	Data Used
UrbanSim	Variant	Characteristics of household, income, age of household head, household size, presence of children, and mobility	Eugene-Springfield, Oregon etc.	Transportation home interview survey Census data The housing stock
ILUTE	Unit	Individual's age, gender, education, marriage status, length of marriage, driver's license, current dwelling, dwelling type, size, price and location	Toronto, Canada	Census data Household travel survey data Real estate transaction data Special-purpose survey data
IRPUD (ILUMASS)	Zone	Zonal individual's age, gender and nationality; Zonal household age of head, nationality, income, size; Zonal dwellings: type of building, tenure, quality, size; Regional employment, immigration	Dortmund, Germany	Housing census population data Employment census data
SVERIGE	100m×100m grid cells	Individual's age, marriage status, length of marriage, length of staying in current dwelling, and employment; Household structure, income, current dwelling; Dwelling type, size, price and location.	Sweden	Regional census data from Statistics Sweden (SCB)
SMILE	County	County level population by age, gender, marital status, employment status and location	Ireland	Irish Census data Vital Statistics data
PUMA	500m×500m grid cells	Individual's age, gender, income, education, marriage status, current dwelling, commute distance, dwelling type, price and location	The northern part of the Dutch Randstad	Census data The Dutch residential preferences survey (WBO) data
LocSim	Dwelling types	Individual age, gender, income, education, employment, marriage status; household structure, current dwelling; dwelling type, size, tenure, vacancy, price/rent	Enigma City, Netherland	10% sample data from the population register of a Dutch municipality The Dutch residential preferences survey (WBO) data

As we can see from the overview of the existing microsimulation models, the models tend to have limited scope or capabilities, such as:

- IRPUD: only capable of zonal level modeling;
- SVERIGE: only capable of regional mobility forecasting
- SMILE: only capable of projecting population evolution at a national level;
- LocSim: only capable of small scale and short term simulating.

The other residential mobility models, although simulating behavior at the individual level, tend to do a poor job at considering the interaction among the individuals or households in the housing market and fail to explicitly incorporate the housing market mechanism, which may weaken results of the models. In order to capture the interactions among individuals and households and to represent the housing market where different agents (individual/households, firms, and real estate developers) interact with each other, our team is developing a new platform called SimMoblity. Table 2-4 compares key features of approaches between SimMobility and UrbanSim, ILUTE and PUMA. The SimMobility platform will be introduced in detail in the next chapter.

Table 2-4 Objectives of SimMobility in Comparison to Select Operational Microsimulation Model Features and Capabilities

	Long-term SimMobility	UrbanSim	ILUTE	PUMA
Model Type	Discrete Choice	Discrete Choice	Discrete Choice	Discrete Choice
Awakening Process	Modeled	Not Modeled	Not Modeled	Modeled
Household Mobility	Disaggregate, Individual based	Aggregate, Historical data based	Disaggregate, Individual based	Disaggregate, Individual based
Household Residential Choice	Modeled, Choice among Dwellings	Modeled, Choice among Zones and Land use types	Modeled, Choice among Dwellings	Modeled, Choice among Grid Cells
Household Job Location Choice	Modeled, Disaggregate	Modeled, Aggregate	Not Modeled	Modeled, Aggregate
Household Classification	Disaggregate, 5-15 characteristics	Disaggregate, Income, Persons, Workers, Child	Disaggregate	Disaggregate, Income, Persons, Workers, Child
Household-Firm Interaction (through job markets)	Modeled, Two-way interaction	Not Modeled	Not Modeled	Modeled, One-way influence
Household-Developer Interaction (through housing markets)	Modeled, two-way interaction	Not Modeled	Modeled, two-way interaction	Modeled, one-way influence
Dwelling Classification	Disaggregate, 5-10 attributes	Not Modeled, Aggregate in Grid Cells	Disaggregate,3 attributes	Disaggregate,3 attributes
Housing Market	Modeled	Modeled	Modeled	Not Modeled
Dwelling Price Evolution	Modeled, based on bid-auction theory	Modeled, based on hedonic theory	Modeled, based on bid-auction theory	Not Modeled
Geographic Basis	MTZ, TAZ, Buildings	Grid Cells	Grid Cells,buildings	Grid Cells
Temporal Basis	Monthly, Dynamic	Annual, Dynamic	Annual, Dynamic	Annual, Dynamic
Activities	Modeled	Not Modeled	Modeled	Modeled, Only for Commuting
Accessibility Measurement	Activity-based, dynamic, feed from Medium-term model	Adapted from external models	Modeled	Activity-based

Chapter 3

SimMobility and the Housing Market in Singapore

SimMobility is the simulation laboratory under development by the Future Urban Mobility Interdisciplinary Research Group. The SimMobility platform tries to go beyond the existing microsimulation models to create an integrated simulation platform for various mobility-sensitive behavioral models. The objective of this chapter is to describe the SimMobility framework and the foundation upon which the present research is built. The first part of this chapter introduces the SimMobility framework, describes the Long-term SimMobility conceptual model in detail, and then presents the operational model for household location choice. The second part of this chapter provides an overview of the land use, transportation, and housing market in Singapore, which serves as the case study application of the operational model of household location choice.

3.1 Integrated Simulation Platform –SimMobility

The SimMobility platform tries to integrate various mobility-sensitive behavioral models with state-of-the-art simulators to predict impacts of, and interactions among, mobility demands, transportation networks and services, land development patterns, environmental impacts, and so on. Integration will make it possible to simulate the effects of a portfolio of technology, policy and investment options under alternative future scenarios.

SimMobility has three inter-related modeling time-frames—the short-term, medium-term and long-term (Figure 3-1). The short-term model takes the trip chains and activity schedules provided from the medium-term model, and simulates network performance, feeding the resulting performance parameters back to the medium-term model. The medium-term model, taking the locations of agents and other longer-term attributes provided from the long-term model, simulates daily travel demands, feeding accessibility estimates back to the long-term model. The role of the long-term model is to simulate longer-term behaviors, including location, lifestyle, and vehicle ownership choices. The following part of this chapter will focus on the Long-term SimMobility Model, within which this specific research fits.

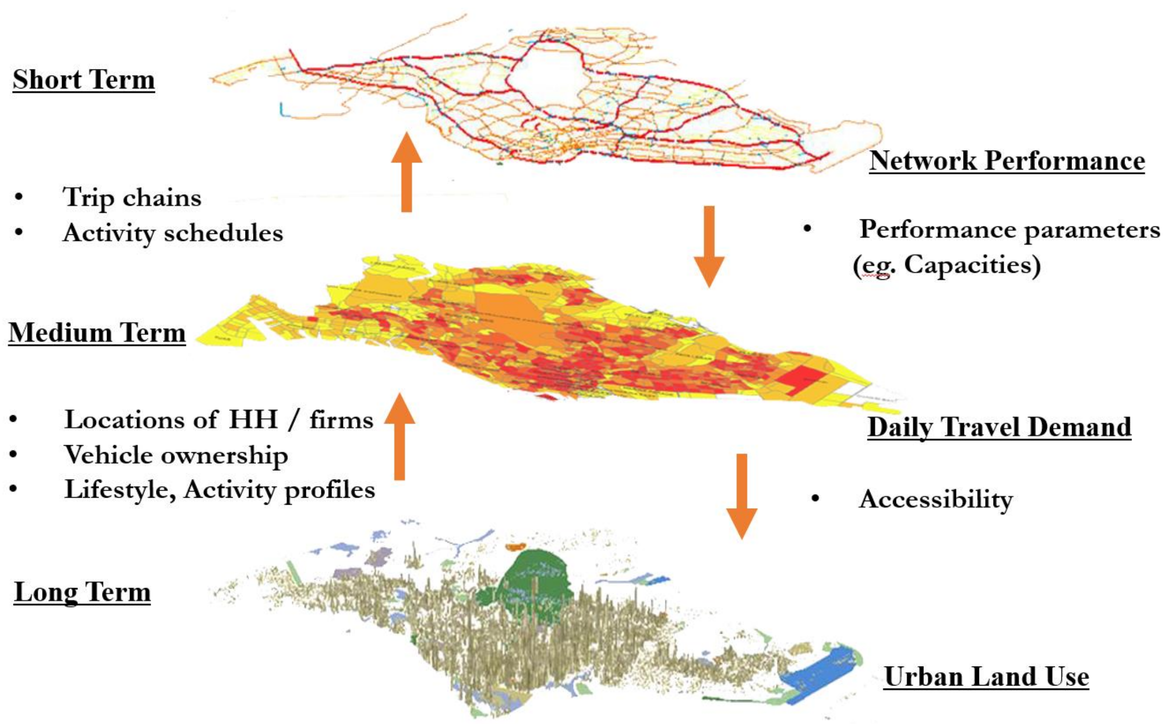


Figure 3-1 SimMobility Framework

3.2 Long-term SimMobility Conceptual Model

3.2.1 Objectives of the Long-term SimMobility Model

Changes in land use and urban mobility systems result from the needs and activities of individuals, households, firms and developers. The needs and activities of different agents are interdependent and their interactive behaviors are ultimately reflected in the changes in land use and mobility system. Therefore, understanding the interactions between agents is crucial in forecasting future urban land use and mobility systems.

The objective of developing the Long-term SimMobility Model is to capture those interactions in a theoretically and operationally sound way by applying and/or developing state-of -the-art models of individual and institutional choice behaviors.

In particular, the Long-term SimMobility Model aims to capture various processes that could independently or mutually influence land use and mobility systems:

- The evolution of individuals/households through demographic development, and residential and job choices. In particular, the residential choice determines the demand for residential buildings and land that influences real estate supply; the job choice, in turn, influences firms' decisions, hence affecting job supply.
- The evolution of firms through firmographic changes (like firm creation and demise) and their location decisions and job supply.
- The evolution of real estate developers in terms of real estate supply decisions and their reactions to housing and job market dynamics.
- The impacts on the above of policy interventions over time, and, ultimately, the implications for overall system performance.

Figure 3-2 illustrates the mutual interactions among those fundamental processes:

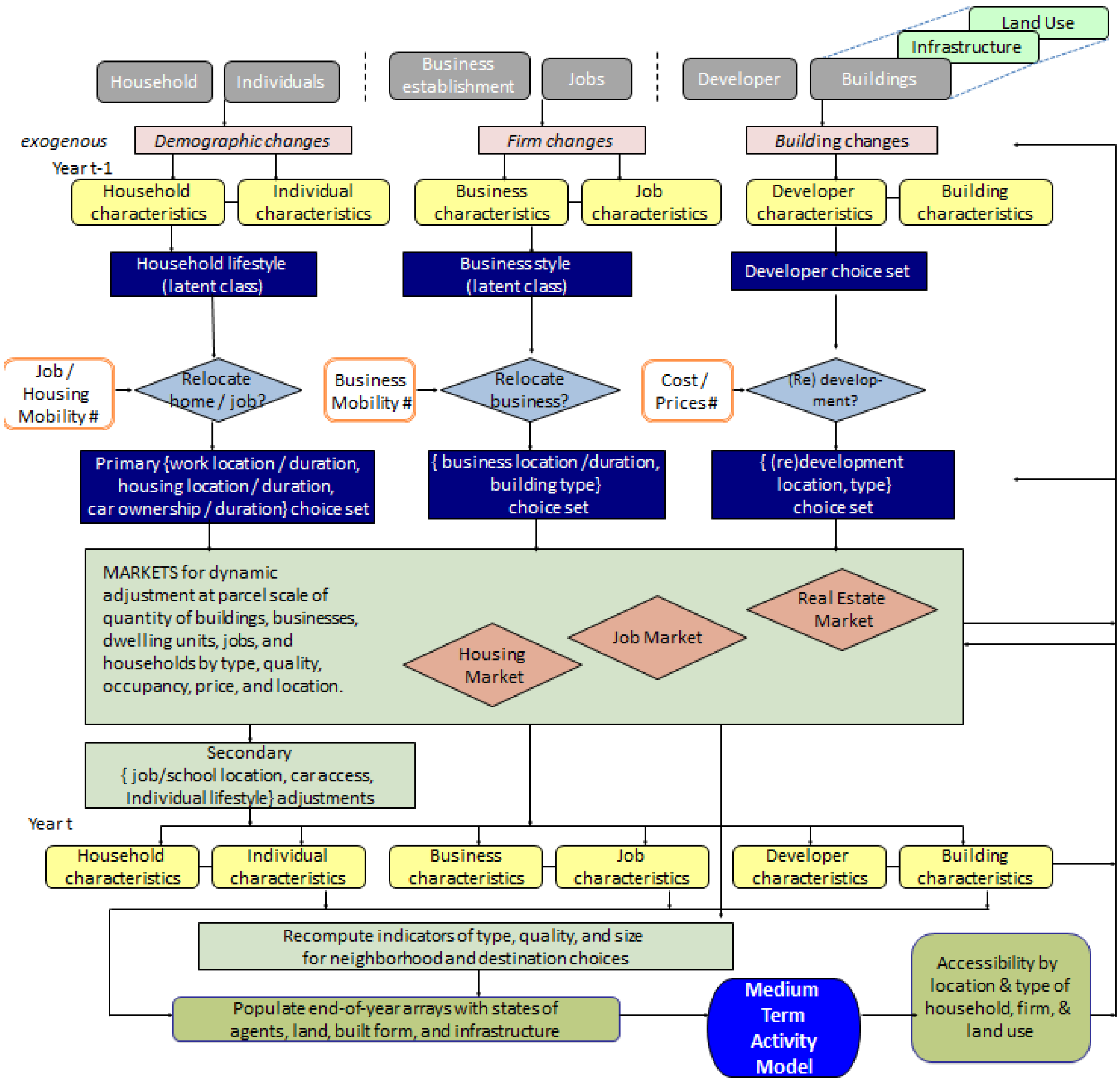


Figure 3-2 Long-term SimMobility Model Flow Diagram[2]

The Long-term SimMobility model consists of three main agents (households, firms and real estate developers) and three main markets (housing market, job markets and real estate market), and models the dynamics and interactions among these three agents in the three marketsThe microscopic population data used in the Long-term SimMobility model are synthesized through iterative proportional fitting (IPF) process (Zhu and Ferreira, 2014). The population data will be updated based on future scenarios. The daily activity and travel pattern are adopted from the

[2] Adapted from SimMobility poster: SimMobility – Next Generation Long-Term Simulator, 2012

medium-term SimMobility model that estimates detailed travel behavior in an activity-based approach.

3.2.2 Agents

The long-term SimMobility model consists of three main agents -- households, firms and real estate developers -- the main drivers of urban development. The emergence and evolution of these agents create demand and supply for housing and other facilities that shape the development of urban areas. The behaviors of these three agents depend on and interact with each other in three different markets.

a) Households/individuals

Households' demand for housing, jobs, schools and other facilities throughout their lives involves certain behaviors and activity patterns. The relevant household behaviors include, residential choice, work location choice, and auto ownership choice. The daily activity pattern can also indicate latent classes of individuals and households' life-styles, which in turn affect households' other behaviors and choices. Empirical evidence can be found in Jiang et al.'s paper, in which they clustered individuals into different life-style revealing daily activity patterns in Chicago (Jiang et al., 2013).

i. Housing (re)location choice

Regarding the (re)location behavior of households, the Long-term SimMobility Model operates under the assumption that the agents will try to maximize their utility provided by housing. The utility of a dwelling to a particular household is determined by various factors like household characteristics, job and school location, car ownership, dwelling attributes, and accessibility.

Changes in a household's perception of its current dwelling might result in the awareness that it could improve its utility by moving to another dwelling. In such a case, the household would start to search for available dwellings and enter the housing market as a potential buyer. If the household owns their current dwelling, it also would put that dwelling up for sale while looking for a new one. In this way, the household not only creates demand for housing but also increases supply as well.

ii. Job and work location choice

Within a household, individuals' working situations determine income and hence largely influences what kind of life it can afford to live. Among other aspects of the working situation, work location choice is constrained by firm location choice, which influences travel time to work and the possible lifestyle of an individual or household.

Work location choice is similar to housing location choice. The agents would choose one out of the available jobs, based on skill sets, and personal considerations about salary, location and other preferences. In terms of changing jobs, we assume that the agents compare the utility of their current job and the utility of another potential job with respect to salary, location, etc. and then leave their current job if the utility improvement exceeds the cost of changing jobs.

iii. Auto ownership choice

As noted before, auto ownership is one of the factors that influence a household's perception of the accessibility of a dwelling. The decision to purchase an automobile is also influenced by other aspects of accessibility, such as travel time to work, and accessible opportunities. Empirical study shows that accessibility and general travel costs (both monetary and time) also influenced auto ownership in Singapore (Xiang, 2014). Auto ownership choice relates to both residential choice and activity location choice, which together determine overall accessibility. Household-level accessibilities come from the medium-term model and are updated based on changes in the mobility and land use systems. Besides accessibility, household characteristics and lifestyle could also influence auto ownership decisions.

iv. Lifestyle

The concept of lifestyle attempts to capture the unobserved factors that influence and may be influenced by household behaviors. Different studies of travel behavior and residential location choice take different approaches to defining the scope of lifestyle and measuring it. Some define lifestyle as merely a pattern of behavior, while others focus on attitudes and preferences (Chen, 2013). In the long-term SimMobility model, lifestyle will be modeled as a latent class from the daily activity and travel pattern adopted from the medium-term model. Lifestyle here, in the long-term model, constitutes the long-term interplay between the physical urban system and individuals' behavior.

To conclude, the interdependent relationship of different household behaviors is shown in Figure 3-3:

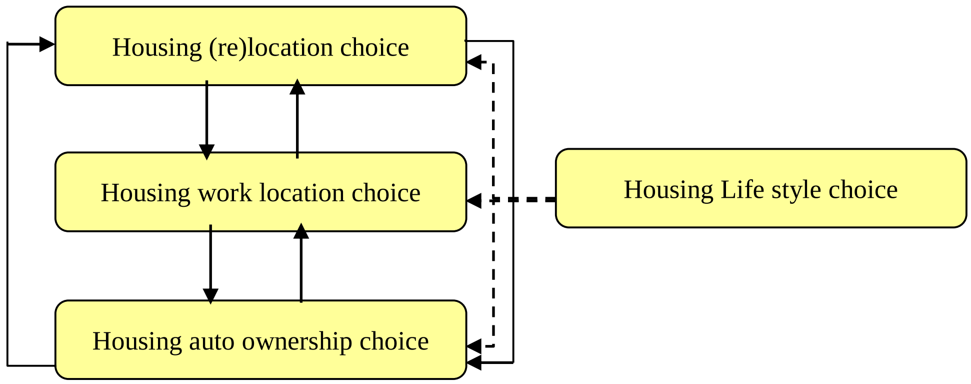


Figure 3-3 The Interdependent Relationship of Household Behaviors

b) Firms

By locating in certain places, firms affect the spatial distribution of jobs and other opportunities (e.g., shopping) and, thus, the usage of urban transportation systems. The emergence and evolution of firms change urban structure in the long run.

The agent-based long-term SimMobility model aims to present firms' behavior in a behaviorally sound way with a focus on their location choices. Firms are represented as individual agents, similar to individuals/households, which go through firmographic events such as creation, business expansion, demise, etc. Among the firms' behaviors, the long-term model pays more attention to firms' location choice since the location choice connects businesses (and commercial zones) with individuals/households as employees and consumers, and determines the demand for business/commercial real estate and location of jobs and other opportunities.

A firm's location decision is based on the perceived profitability of a particular location, as influenced by various factors like the firm's characteristics, location accessibility to inputs and consumers, and other locational attributes that might influence profitability, such as agglomeration economies.

Firms' characteristics include business type, size and life stage (e.g., start-up, early growth, expansion, maturity), all of which may influence the location needs and employment size of a firm. A location's accessibility is result of transportation system performance and relative location of inputs and consumer markets. Other physical attributes include floor area, building environment and facilities (elevator, etc.). The range of potential agglomeration economies relate to proximity to other firms, public services, labor markets, etc.

Besides location choice, firms also interact with other agents (like households) in their employment size. Decisions about the number of employees a firm hires at a particular location determine the spatial distribution of jobs. The employment size of a firm can be derived from its firmographic stage (e.g., start-up, early growth, expansion, maturity) and its business size.

c) Real Estate Developers

Real estate developers determine the supply of new units for both residential and commercial use and have a profound influence on land use and urban development. SimMobility attempts to represent real estate developers as individual agents in the long-term model. Their behavior is basically modeled as (re)development and investment decisions, which are made mostly for profit, based on perceptions about residential and commercial real estate markets.

The real estate developer decides whether, when, where and what type of development to invest in to maximize profit, based on the perceived market demand from both households and firms. Besides the market perception of demand for new dwellings, the real estate developers also takes into consideration the demand and supply of the resale market when making development decisions. As profit maximizers, real estate developers would also react to the unit price changes, interest rate changes, etc.

In the Singapore context, the development of public housing provided by the Housing and Development Board (known as HDB) is a distinctive case, as discussed further below. The demand for HDB is relatively predictable since the required applications for HDB reveal the demand for HDB new sales. Moreover, the highly regulated HDB new sale market is different from the private new sale/resale market and the less regulated HDB resale market; thus we might expect a less elastic supply for HDB new sales than for the supply of other kinds of dwellings since the government might not react to the market promptly.

3.2.3 Interactions

In the long-term SimMobility model, the interactions between the agents' behaviors occur in three main markets – for housing, jobs and non-residential land uses, real estate. The housing market captures the interactions between households and real estate developers in terms of buying and selling dwelling units. The real estate market determines the supply of dwellings provided by real estate developers, based on the demand represented by households' behavior; the non-residential market captures the interactions between firms and households and firms and real estate developers.

Households' residential choices and real estate developers' (re)development choices are mutually dependent with each other in determining housing demand and supply. In the housing market, households make residential decisions based on their own perceived utility improvement from an alternative dwelling, but their choice set is actually limited to the supply and the availability of dwellings, which is determined by real estate developers and other homeowners. Households who decide to relocate from units they currently own, will also put their current dwellings up for sale, which, in turn, affects supply in the resale housing market and might, in turn, affects new sale markets. Households' behavior in the housing market, which reveals the demand for housing, will also, in turn, affect real estate developers' (re)development decisions and will change housing supply in the next time period.

Interactions among households also take place in the housing market. For example, if one dwelling unit maximizes the utility for multiple individuals/households, they will compete with each other through their willingness-to-pay (derived from their perceived utility) to get this particular dwelling. The factors that influence a households' willingness-to-pay may include its own characteristics (including lifestyle), work and school locations, and car ownership. Both work location and car ownership are influenced by households' desired and needed accessibilities and firms' location choices.

Firm-real estate developer interaction is similar to household-real estate developer interaction. Firms' location choices are also interdependent with real estate developers' (re)development choices, which determine demand for and supply of commercial/business real estate. The competition mechanism among firms in the location market also resembles that for households. The firm-real estate developer interaction determines firm location and further determines the available job and other destination location choice sets for households.

Individual/households and firms interact in the job markets, demand for other activities (shopping, recreation, etc.) and households' job choices and other regular/preferred travel destinations likewise are interdependent with residential choice. A firm follows its own growth path, and at one time period, the firm's location and its employment size are fixed. Given this, an individual's job choice influences and is influenced by residential choice since the work and dwelling location together affect the utilities of the job and the dwelling and will further affect both choices. Firms' location choices also influence the spatial distribution of numerous other opportunities, which will affect the accessibility of dwellings and further affect residential choice.

The interactions and interdependent relationships between the behaviors of households, firms and real estate developers are summarized in Figure 3-4:

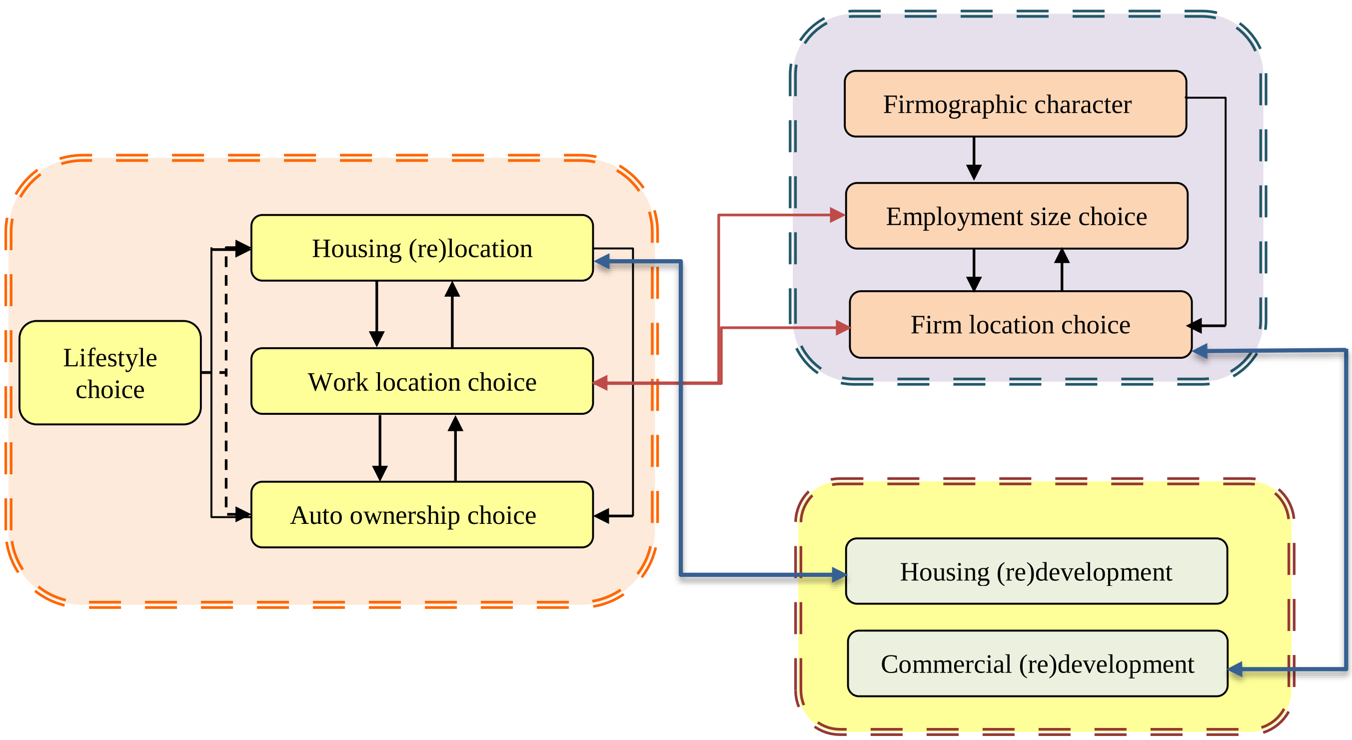


Figure 3-4 The Interdependent Relationship of Behaviors of Households, Firms and Developers

3.3 Singapore's Housing Market

Economic and population growth have pushed up Singapore's living expenses. In 2009, the *Economist Intelligence Unit* ranked Singapore the tenth-most expensive city in the world in which to live—the third in Asia, after Tokyo and Osaka. One of the most substantial increases in the cost of living in Singapore has been in housing. Facing the sharp increase in housing prices, the Singapore government plans to increase the land supply for residential use and increase the supply of new homes (Ministry of National Development of Singapore, 2013).

Singapore represents a unique housing market with a dominant public sector, different from most countries or regions that have been studied using micro-simulation models. In particular, Singapore is one of the few countries in the world that practices an integrated housing sector policy, in which planning, urban policy and government objectives guide the real estate development (Phang, 2001).

The Singapore housing market is characterized by the coexistence of a dominant public sector and a small, growing private sector with relatively higher-quality housing (Bardhana et al., 2003). The majority of the residential housing developments in Singapore are publicly governed and developed. This public housing (or "HDB") in Singapore is managed by the Housing and Development Board (HDB). About 82% of Singaporeans live in HDBs (Table 3-1).

Table 3-1 Households & Housing Statistics (2012)

Number of Resident Households	1,152,000
Average Household Size	3.53
Home Ownership Rate	90.1%
Resident Households by Type of Dwelling	
HDB	81.6%
Condominiums & Other Apartments	12.1%
Landed Properties	6%

Source: Department of Statistics, Singapore (2013)

Public housing in Singapore is highly regulated, not only in terms of price, but also for eligibility of buyers. There are a number of eligibility conditions in order for a flat to be purchased. A buyer must be a Singaporean citizen, or Permanent Resident (P.R.) and be 21 years of age with a family. Non-citizens and singles below 35 years old are not allowed to purchase new HDB flats. Other requirements concern household status, minimum time requirements between purchases, income, and other special requirements.[3]

[3] Details of eligibility conditions can be found at:
http://www.hdb.gov.sg/fi10/fi10321p.nsf/w/BuyingNewFlatEligibilitytobuynewHDBflat?OpenDocument

Although not as strictly regulated as the new sale market, the resale market for HDB flats is also subject to many restrictions. Existing flat owners are allowed to sell their flats on the open market to any eligible buyer at a mutually agreed price. While the HDB does not regulate these prices, the buyer and seller must declare the true resale price to the HDB. In addition, most flat owners may only sell their flat if they have met the Minimum Occupation Period (MOP) requirement which is the minimum duration of physically occupying the HDB flat, introduced to help reduce speculative activities. Buyers are also subject to a set of eligibility conditions.

Besides pricing and eligibility restrictions, HDB also maintains a quota system of ethnicities through the Ethnic Integration Policy (EIP)[4]. By ensuring that each block of units is sold to families from ethnicities roughly comparable to the national average, it seeks to avoid physical racial segregation and formation of ethnic enclaves common in other multi-racial societies. A study has shown that EIP has succeeded in reducing the intensity of the ethnic enclaves while increasing social integration (Sim et al., 2003).

Private apartments, condominiums, and landed properties are the most common forms of private residential properties in Singapore. Landed residential properties (houses) in Singapore can only be owned by foreigners who have obtained specific approval from the government. This approval is granted on a case-by-case basis[5]. There are also executive condominiums, which are built by HDB but are automatically converted into private properties 10 years from the date of construction. The private apartments and condominiums are relatively popular since they are categorized as non-restricted property, are not subject to the eligibility conditions; thus, they are relatively expensive. The private housing market for apartments and condominiums can be seen as a free market.

Given that the public sector is highly regulated, the interaction between the two markets is mainly affected by households' income, loan interest rates and the public housing resale price (Bardhana et al., 2003). Singapore also exhibits a special type of housing-related mobility, where its public homeowners continually seek to upgrade to private housing (Tu et al., 2005).

[4] Ethnic Integration Policy & SPR Quota
http://www.hdb.gov.sg/fi10/fi10322p.nsf/w/SellFlatEthnicIntegrationPolicy_EIP?OpenDocument

[5] Details of foreign ownership restrictions can be found at Singapore Land Authority website:
http://www.sla.gov.sg/htm/ser/ser0306.htm

Chapter 4

The Treatment for Data Inconsistency: A Matching Approach

In order to model household residential behavior as part of the Long-term SimMobility Model, I use Singapore and its private housing market as a case study. As mentioned in the previous chapters, the lack of sufficient data has hindered the application of more advanced models of household residential location behaviors at a microscopic level. Although more data-rich than many other places around the world, the Singapore case also suffers from insufficient data. This chapter will walk through the available data sources and their key statistics, explaining the problems and challenges of using the data sets at hand. The second part of this chapter will introduce a method to treat the imperfect data sets and then apply it to the case of Singapore.

4.1 Singapore Data Sources

As mentioned in chapter 3, I use only Singapore's private housing market as a case study. The data come from: 1) Singapore's Household Interview Travel Survey 2008 (HITS), conducted by the Land Transport Authority (LTA), which provides household demographic, socioeconomic and travel information; and 2) the Urban Redevelopment Authority's (URA) Real Estate Information System (REALIS), which tracks real estate transactions for private properties. REALIS includes detailed descriptions of attributes of private dwellings purchased during 2008. The observations in the two data sets, HITS' households and REALIS' private dwelling unit transactions lack any direct connections between them.

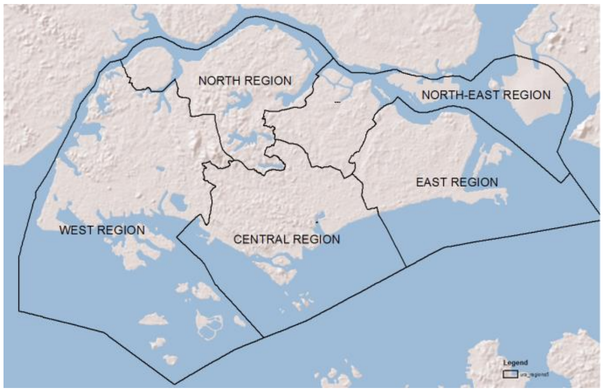
4.1.1 Spatial units

The spatial units used by the Urban Redevelopment Authority (URA) and LTA in Singapore for planning purposes include the following six levels (see Table 4-1).

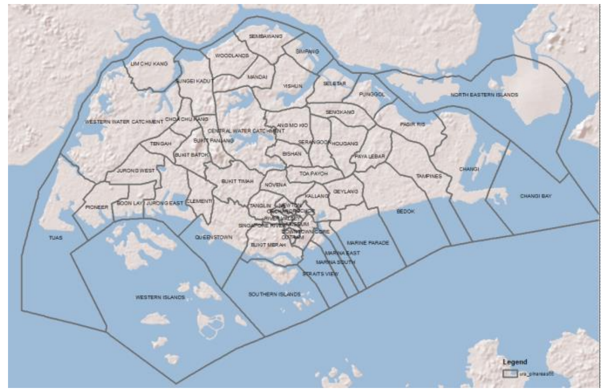
Table 4-1 Spatial units used in Singapore

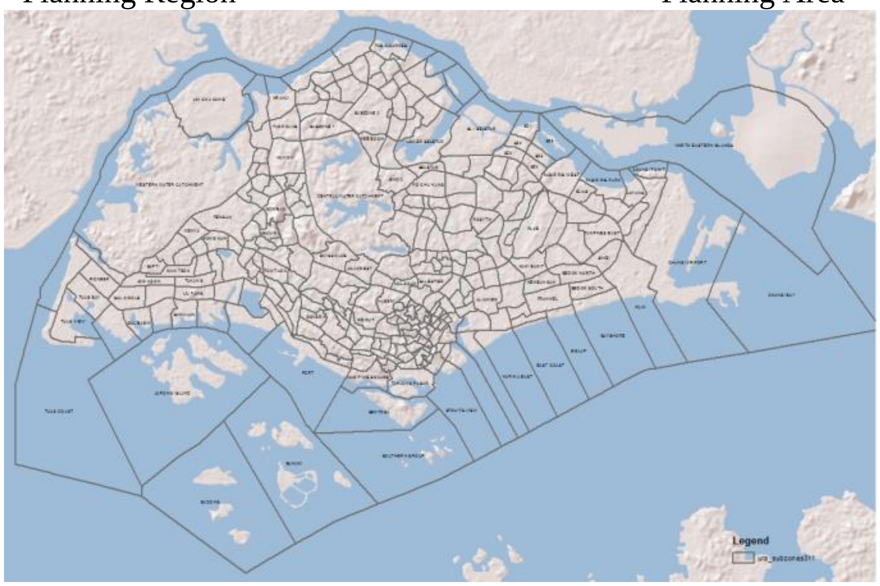
Spatial Level	Count	Min	Max	Mean	st.dev
Planning Region (km^2)	5	157.90	445.00	233.68	113.92
Planning Area (km^2)	55	0.85	251.31	18.37	33.97
Planning Subzone (km^2)	311	0.040	103.03	4.08	11.24
MTZ (km^2)	514	0.027	54.04	1.65	3.93
Transportation Analysis Zone (km^2)	1,092	0.014	129	0.959	5.85
Postcodes	>12,000	--	--	--	--

The planning region, area, and subzone are mainly used for land use planning purposes, and the MTZ is a substitute for the planning subzone at a smaller scale. The Transportation Analysis Zone (TAZ) is basically used for transportation modeling. Postcodes in Singapore relate directly to buildings, which means that each building has a unique postcode. Thus, with a postcode, a building can be located precisely. The boundary of each spatial unit is shown in Figure 4-1.

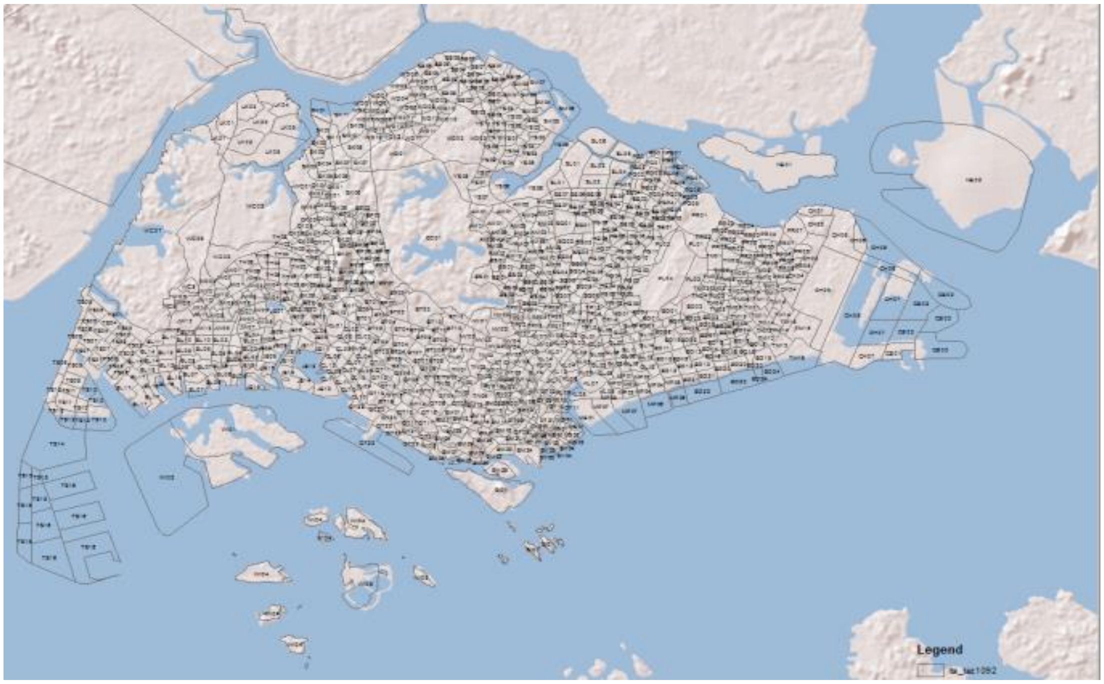


Planning Region

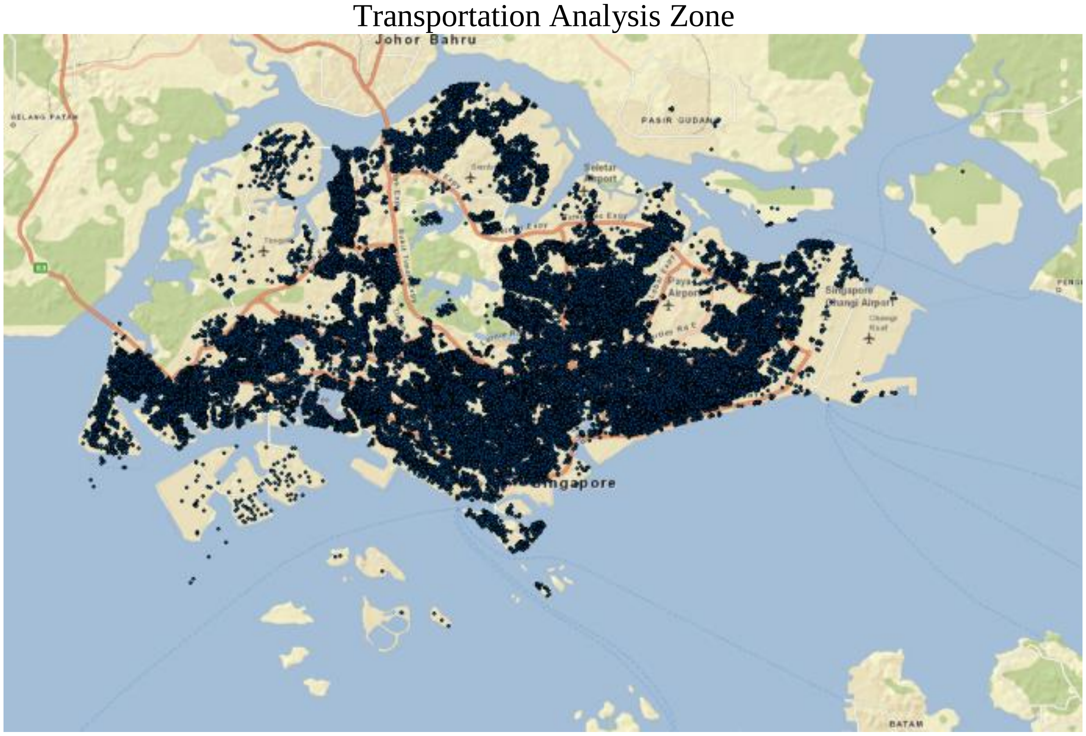

Planning Area


Planning Subzone


MTZ



Transportation Analysis Zone



Postcodes

Figure 4-1 Boundaries of Spatial Units

4.1.2 Household Interview Travel Survey 2008 (HITS 2008)

HITS 2008 data provide essential information on the individual characteristics for 1% of Singapore's households in 2008. HITS 2008 contains 1221 households that lived in private dwellings in 2008. Table 4-2 shows the fields from HITS 2008 that could be useful in our household residential choice model and the statistics of the key household variables.

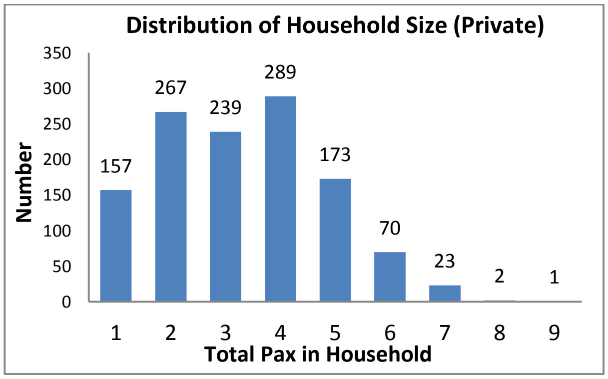
Table 4-2 HITS Data Description

Description	**Data type**	**Min**	**Max**	**Mean**	**Median**	**std.dev.**
Household Postcode	Discrete	--	--	--	--	--
Household dwelling type	Text	--	--	--	--	--
Household Size	Discrete	1	9	3.30	3	1.53
Vehicle availability	Categorical	0	1	0.71	1	0.45
Number of bike HH owns	Discrete	0	8	0.27	0	0.90

Description	**Data type**	**Category**	**Share**
Number of children in the HH*	Discrete	0	58.23%
		1	16.05%
		2	18.35%
		3	6.80%
		4	0.57%
Household Income per Person	Categorical	No Income	11.63%
		$1-$1000	11.38%
		$1001-$1499	7.37%
		$1500-$1999	9.75%
		$2000-$2499	8.76%
		$2500-$2999	7.78%
		$3000-$3999	11.88%
		$4000-$4999	4.34%
		$5000-$5999	6.80%
		$6000-$6999	2.05%
		$7000-$7999	1.15%
		$8000 and above	3.11%

***** The number of children each household has is calculated from the individual age (from 0-19 years old are viewed as children here)

The size of the households varies from one to nine. The detailed distribution of household size is shown in Figure 4-2. The median household size is 3 people and 78% of the households have four or fewer people. The household size distribution is similar between households living in HDB and private dwelling. If looking into different HDB types by number of rooms (Figure 4-3), we can see the distributions of household size vary among different HDB sizes; such variance of distributions might be due to the household size requirement when applying for different type of HDB.



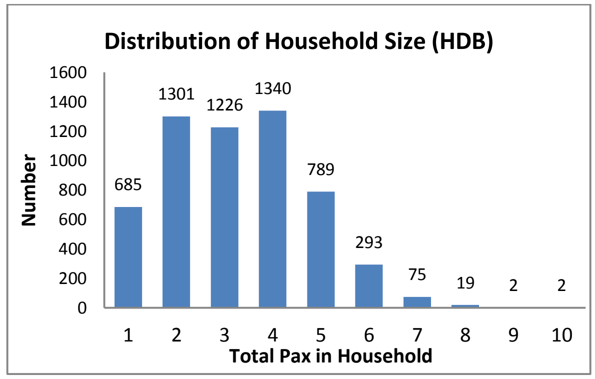


Figure 4-2 Distribution of Household Size in HITS

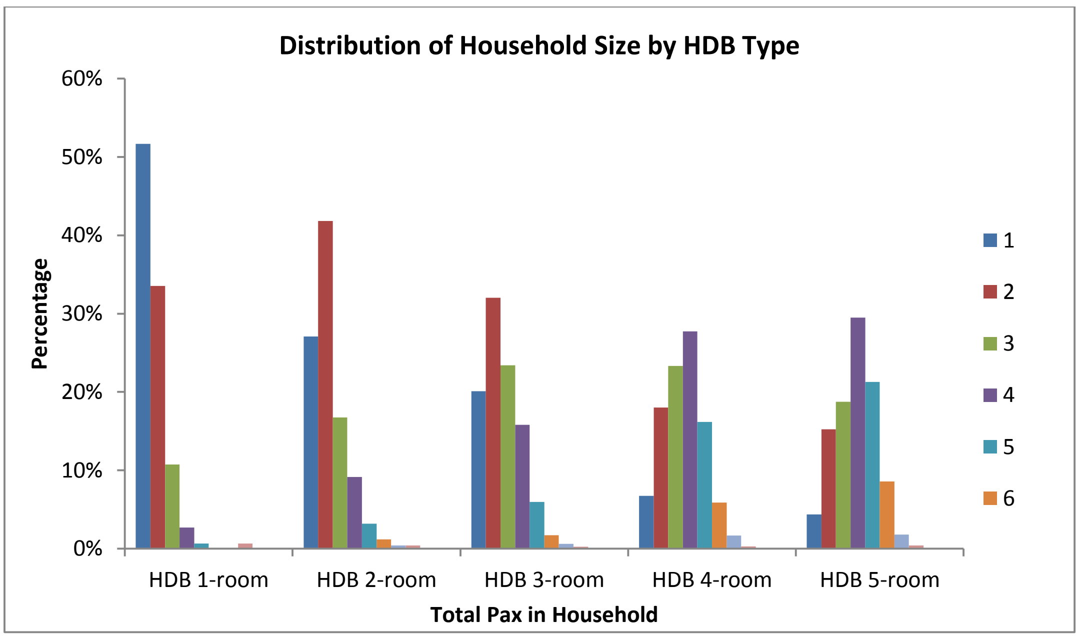
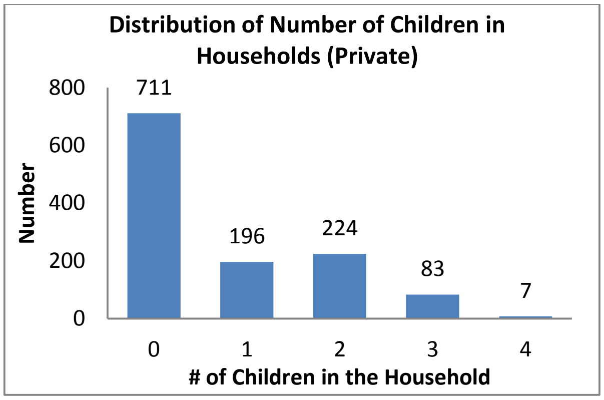


Figure 4-3 Household Size Distribution by HDB Type

Among all the households in HITS 2008 living in private dwellings, 58.2% of the households have no children, and the average number of children a household is 0.75. Only 7.4% have more than three children and 34.4% of the households have one or two children in their households. When comparing the households living in HDB and private dwelling, we find the distribution of the number of children in households who live in HDB is similar to those who live in private dwelling.



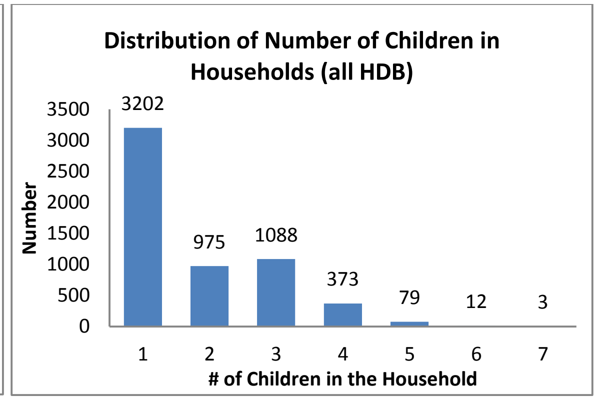
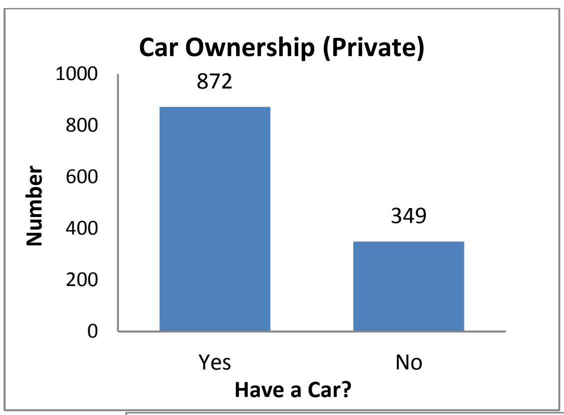
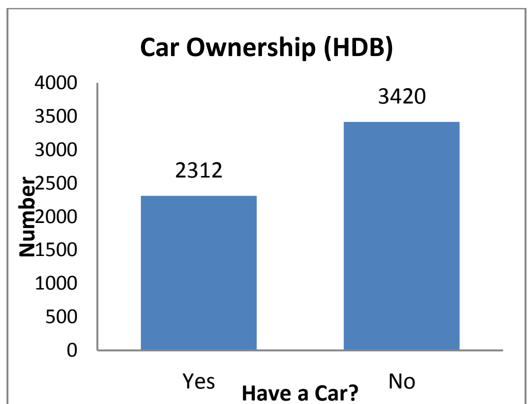


Figure 4-4 Distribution of Number of Children in HITS

HITS 2008 also provides information about car ownership, but does not specify the number of cars, rather only if households have access to a car. According to HITS 2008, 72% of the households living in private dwellings have access to a car, while the number falls to 40.3% for the households living in HDB; the share of HDB households with car access grows with an increase in households' dwelling unit size (number of rooms in HDB).





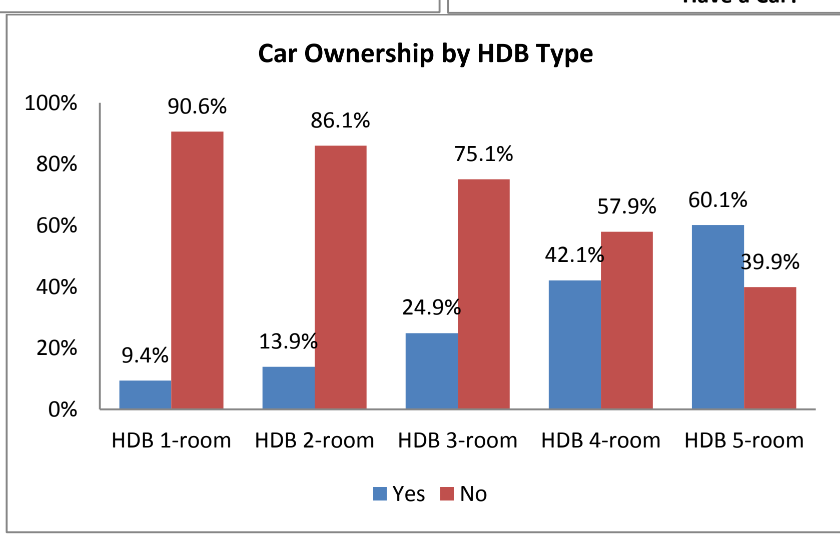


Figure 4-5 Car Ownership in HITS

HITS 2008 provides individual household members' income in categories and I aggregated these values into household income using the high end of the categories, since people likely report their income conservatively. I then calculated household income per capita – dividing the aggregate household income by the number of household members. For households living in private dwellings, 70.6% households earn more than S$4,000 per capita per month. In general, incomes are much higher for households living in private dwellings than in HDB. For households living in HDBs, a higher proportion of high-income households live in larger units (more rooms).

In general, the distribution of household size and the number of children are similar for households living in HDB and private dwellings. However, households living in private dwellings tend to have higher income and are more likely to own a car.

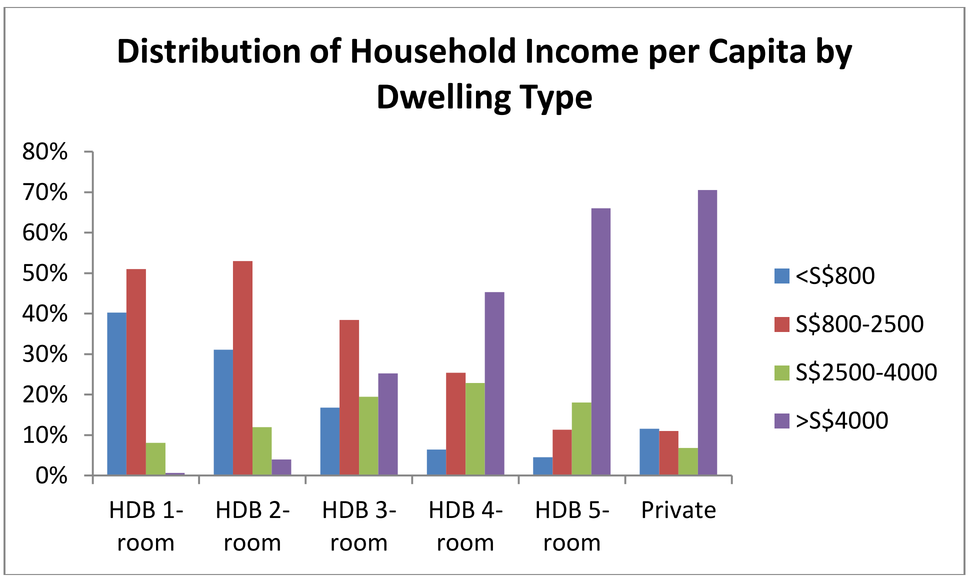


Figure 4-6 Distribution of Household Income per Capita by Dwelling Type

HITS 2008 also provides the postcode of the household's dwelling unit, which is critical in connecting with the REALIS data. However, HITS does not include further details about the dwelling, which is the main barrier to using both HITS 2008 and REALIS as a consistent data set.

4.1.3 Real Estate Information System (REALIS)

The Real Estate Information System (REALIS) includes detailed descriptions of attributes of private dwellings that were purchased during 2008, including the address, floor area, price, etc.

The address and postcode allow for precisely locating each transaction. This spatial resolution also allows for the calculation of other location-specific attributes, like distance to the CBD area, and neighborhood attributes like school enrollment and total employment (available at, and thus, measured at the MTZ level. The MTZ level school enrollment and employment data are provided by Singapore's Urban Redevelopment Authority. Tables 4-3 shows the fields from REALIS and the basic statistics of the dwelling attributes that will be used in our model.

Table 4-3 REALIS Data Description

Data	Data type	Min	Max	Mean	Median	std.dev
Postcode	Discrete	--	--	--	--	--
Floor area (m^2)	Continuous	29	4,880	148.19	118	146.68
Transaction price (S$)	Continuous	225,000	43,200,000	1,313,070	935,000	1,622,472
New sale or resale	Dummy	0	1	0.30	0	0.46
Distance to CBD (m)	Continuous	528	4,880	8276.76	7,659	4,487.81
School enrollment	Continuous	0	27,800	2,568.47	2300	3,038.90
Employment	Continuous	0	85,720	5841.73	4,200	6,988.11

4.1.4 Problems with the existing data sources

As we can see from the available data sets described above, household survey data with household characteristics (HITS) and housing transaction data with attributes of dwellings (REALIS) are available only as separate data sets (a problem not found only in Singapore). However, consistent household characteristics and dwelling attributes are required for the residential location choice model. Given the data sets available for model estimation, household characteristics from HITS 2008 need to be matched with dwelling attributes from REALIS transaction records in 2008. In the two data sets, however, no direct connection for matching exists, and the only connection in the Singapore data sets is the postcode. A postcode in Singapore generally corresponds to a particular building, which improves the likelihood of accurate matching. However, if we match the two data sets with postcodes, we will lose more than half of the household records – only 565 out of 1233 households left, and millions of transaction records are lost. Besides the loss of data, this "naïve" kind of matching also creates biases, because this kind of arbitrary matching loses the variance of attributes of the dwellings in the same building (with the same postcode, also). In consequence, the next section will explore using imputation to match the HITS and REALIS data sets.

4.2 Matching Method

4.2.1 Missing data imputation

Since we cannot directly match households to transactions across the available datasets and "naïve" approaches to matching (by postcode only) are subject to large potential error, a more sophisticated approach is necessary. One approach is to view the information of the characteristics of households who live in certain dwellings as missing at random. Obviously, this special case of missing all household information cannot be treated using the traditional treatment, -- list-wise deletion-- which removes observations with missing value(s). Rather than removing variables or observations with missing data, an alternative is to fill in or "impute" missing values. This approach keeps the full sample size, which avoids unnecessary data loss. The simplest approach is to impute missing values based on the observed data for this variable from other observations in this data set, which can only be applied to a small fraction of missing data and will be severely biased if applied to a data set with a large fraction of missing data. Such single imputation cannot satisfy our needs in matching the two data sets. The next section will explore a more advanced way to treat missing data while keeping its estimates unbiased.

4.2.2 Multiple imputation

Multiple imputation has been used to address the missing data problem in many travel surveys, which are prone to non-response (Bush 2003). Multiple imputation replaces each missing value by a vector composed of $M \geq 2$ possible values. The M values are ordered in the sense that the first components of the vectors for the missing values are used to create one complete data set, the second components of the vectors are used to create the second completed data set, and so on; each completed data set is analyzed using standard complete-data methods (Little and Rubin 1987).

Figure 4-7 depicts a multiply-imputed data set.

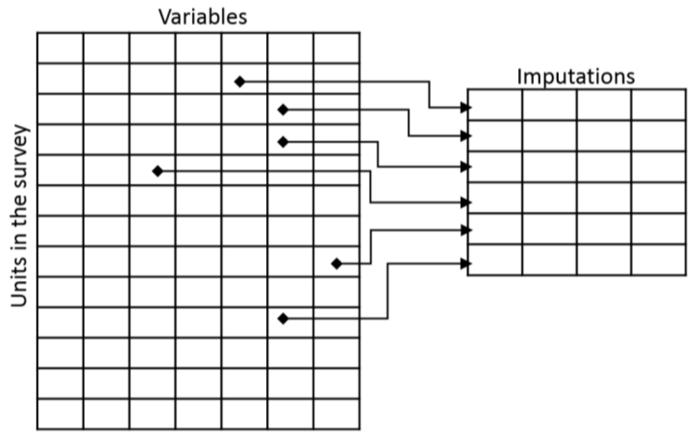


Figure 4-7 Data set with M imputations for each missing datum (Rubin, 1988)

M repetitions yield M completed data sets, each of which can be analyzed by standard complete-data methods as if it were a real data set. The results from M repetitions of M vectors during multiple imputation are combined to create one inference. If $\theta_l, U_l \ (l = 1 \ldots . M)$ denote parameters and variance calculated from imputation *l*, then the final estimation of θ is:

$$\bar{\theta} = \sum_{l=1}^{M} \frac{\widehat{\theta_l}}{M} . \tag{4.1}$$

The variability associated with this estimate has two components: the average within-imputation variance,

$$\bar{U} = \sum_{l=1}^{M} \frac{\widehat{U_l}}{M} \ , \tag{4.2}$$

and the between-imputation component,

$$B = \frac{\sum(\widehat{\theta_l} - \bar{\theta})^2}{M-1} . \tag{4.3}$$

The total variability is:

$$T = \bar{U} + (1 + M^{-1})B \ . \tag{4.4}$$

The significance test is a *t* distribution:

$$(\theta - \bar{\theta})T^{-\frac{1}{2}} \sim t_V , \tag{4.5}$$

with degrees of freedom:

$$V = (M-1)\left(1+\frac{(1+M^{-1})B}{\bar{U}}\right)^2 \tag{4.6}$$

In contrast to single imputation (e.g., mean imputation and regression imputation), multiple imputation accounts for the uncertainty attributable to missing data in the estimated variances. Furthermore, multiple imputation does not suffer from the inconsistency and inefficiency that besets simple case deletion of incomplete observations. Finally, multiple imputation is relatively simple to implement when compared to full maximum likelihood estimation, where the missing data values and model parameters are estimated simultaneously (Bush, 2003).

4.3 Treatment for the Singapore Data: A Matching Approach

In order to create a sufficient data set for the bidding models, I apply Multiple Imputation to deal with the missing household information for the transaction records. Multiple Imputation takes the households that live in the same spatial unit as a possible household pool. It randomly selects one to match with the dwelling, repeats for *M (M>2)* times, and calculates the parameters based on results from all the imputations.

4.3.1 Spatial unit selection

The choice of spatial unit determines the size of the possible household pool. For multiple imputation using the available data, the ideal would be a spatial unit that could not only locate both households and dwellings with adequate precision but also guarantee the size of household pool can represent the distribution of households that are living in that spatial area. A larger spatial level like a planning region or planning area will lose the desirable precision; the smallest spatial level, on the other hand, will have too few households living inside the spatial area to get a plausible distribution. In both cases, the random selection of households out of a too big/too small pool might generate biases. Thus, the MTZ is selected to ensure both accuracy of location and a reasonable size of possible household pool.

4.3.2 Scale parameter setting

For different MTZs, the available household information from Singapore's Household Interview

Travel Survey data varies: in average, 9.18 household observations in HITS with a standard deviation of 10.86. Those variations may lead to some variance in the size of the possible household pool after matching. In order to account for the differences in the size of household pool, three scale parameters are applied for different pool sizes, and one of them is normalized to 1. These scale parameters will be estimated simultaneously in the model (see chapter 5).

Table 4-4 Scale Parameter Definition

μ	Definition
μ_1	The size of possible household pool is (0-10] (Normalize to 1)
μ_2	The size of possible household pool is (10-25]
μ_3	The size of possible household pool is more than 25

As tested, the estimated parameters are not sensitive to the number of imputations M when $M >=10$; hence, ten imputations are conducted.

4.3.3 Matching results

After matching the HITS and REALIS data at the MTZ level, 9495 dwellings find their household pair, following the distribution of household characteristics in the MTZ. As described before, the matching process will be replicated during each imputation, to approximate the true distribution of the characteristics of households living in the same MTZ. The matching results will apply to the household residential choice model for each imputation. The next chapter will introduce the household residential choice model in detail, using the data set generated through the matching processes described above.

Chapter 5

Modeling the Housing Market in Singapore: A Bid-rent Approach

The Multiple Imputation method helps us to match the two available data sets in Singapore, as described in the previous chapter. This makes it possible to test the bid-auction approach to residential location choice modeling at a microscopic level. In this chapter, Ellickson's approach, which has been outlined in chapter 2, will be estimated. Following Ellickson's specification, this model will first divide households into eight approximately homogeneous groups and then estimate, for each group, their willingness to pay for each dwelling attribute. The model estimation is presented, followed by a discussion of the implications, limitations and potential extensions.

5.1 Method

The model will be estimated using Ellickson's specification in order to get the best possible maximum bid model. As noted in chapter 2, although there are more advanced model specifications like Horowitz's approach and Lerman and Kern's approach, implementing such approaches is very difficult and sometimes leads to biased results (Muto, 2006). Therefore, in this thesis, I will explore Ellickson's approach in the case of the Singapore private housing market. Further discussion of modeling the entire market can be found in chapter 6.

According to Ellickson (1981), and as described in chapter 2, the probability that household h occupies the residential dwelling i is the probability that household h places the highest bid for dwelling i, beating all other competing households:

$$P(h|z_i) = Prob\{B_{hi}(z_i) + \varepsilon_h > B'_{h'i}(z_i) + \varepsilon'_{h'}, \forall\, h' \neq h\} \tag{5.1}$$

If ε_h follows an Extreme Value distribution and is independent and identically distributed (iid), as Lerman and Kern (1983) pointed out, the means of ε_h, which is $ln|N_h|$, depends on the sizes of respective groups N_h (N_h denotes the number of members of group h). Therefore, probability of household h winning dwelling i can be presented as:

$$P(h|z_i) = \frac{e^{\mu B_{hi} + ln|N_h|}}{\sum_{h^* \epsilon H} e^{\mu V_{h^*i} + ln|N_{h^*}|}} \tag{5.2}$$

For different groups of households, the willingness to pay for housing attributes can be estimated from likelihood maximization:

$$L = \prod_{i \in S} \left(\prod_{h \in H} P(h|z_i)^{y_{hi}} \right) \tag{5.3}$$

where y_{hi} is a binary indicator that assumes the value of one if household h is observed to be located in dwelling i and zero otherwise. In order to simplify the bid function, Ellickson's approach also aggregates households into homogenous groups and estimates a linear-in-parameter bid function for each household group.

5.2 Model specification

We expect that income and family structure might influence a household's willingness to pay for different attributes of a dwelling. We suspect that higher income could imply a higher willingness-to-pay in general. Also, we assume that preferences for a dwelling will vary not only with income, but also with family structure, especially having children or not.

5.2.1 Explanatory Variables

The variables included in this model can be grouped into three types: 1) dwelling characteristics, including floor area, sale type of dwellings; 2) zonal characteristics, including accessibility, population density, school enrollment; and 3) household characteristics, including income, household size, structure and ethnicity.

Dwelling characteristics

Households' willingness to pay varies with dwelling characteristics, as shown by the original hedonic model (Rosen, 1974). In this model, I include floor area and sale type of dwellings which are available from REALIS data.

My hypotheses are the households' willingness-to-pay for floor area may increase with income and households with children are willing to pay more for a larger dwelling. For the willingness-to-pay for resale or new sale dwelling, I assume income and family structure might have mixed influences on households' preference.

Zonal characteristics

A gravity-based measure of accessibility from the home location represents the potential ease of accessing opportunities across the island. I follow Xiang's calculation of accessibility and opportunity measures: six types of opportunities were used in the calculation: manufacturing, office, retail, hotel, port & airport, and education institutions (Xiang, 2014). Singapore's Urban Redevelopment Authority (URA) provided the data.

In this thesis, we adopted the impedance function from the 2008 trip distribution model provided by Land Transport Agency[6]. According to LTA, the bell shaped $f(C_{ijPV})$ grows as C_{ij} increases and reaches its peak at $C_{ij} = 35$. After that, $f(C_{ijPV})$ decreases as C_{ij} increases.

[6]LTA has estimated the following impedance function for home-based work trips by private vehicle from its 2008 travel survey:

I assume that households want to live close to job centers, but do not want to live too close to job centers. In other words, they want to keep a reasonable distance with job centers to avoid negative side of job centers while remaining a reasonable proximity. A high accessibility value implies ease of access to these opportunities. However, the accessibility can only measure linear preference. As we can see from the bell-curved impedance function, a single accessibility cannot capture the bell-shaped $f(C_{ijPV})$, and thus another variable –accessibility square– is introduced to capture nonlinear preference. Therefore, I assume that the parameter estimates are expected to be positive for accessibility and to be negative for accessibility square.

Population density is calculated using the total number of residents in one MTZ over the area of the MTZ. I assume households are more willing to live in MTZ with higher population density and such preferences may increase if the households have children, since higher population density in Singapore often means more facilities in the area.

School enrollment includes total enrollment for all the schools in that MTZ, provided by URA. I assume that households with children prefer to live in a MTZ with higher school enrollment (as a proxy for school quality) than would those without children. The influence of income and family structure might have mixed impacts, both positively and negatively, on households' willingness to pay.

Household characteristics

Household's own characteristics also influence their willingness-to-pay for certain dwellings in an environment that is similar or different to their own characteristics. I use four variables to capture such differences/similarities between household characteristics and zonal characteristics: income difference, household size similarity, family structure similarity and ethnic similarity.

---

for $0 < c_{ij} < 8, f(C_{ijPV})=0$
for $8 < c_{ij} < 35, f\left(C_{ijPV}\right) = 287 - 287(C_{ijPV} - 168)/(360 - 168)$
for $35 < c_{ij} < 168, f\left(C_{ijPV}\right) = 0.04040723(C_{ijPV} - 6)^4 - 3.4860416\left(C_{ijPV} - 6\right)^3 + 117.32876\left(C_{ijPV} - 6\right)^2 - 2026.192\left(C_{ijPV} - 6\right) + 19874.2$
for $c_{ij} > 168, f\left(C_{ijPV}\right) = 0.00002639(C_{ijPV} - 32)^4 - 0.01031252\left(C_{ijPV} - 32\right)^3 + 1.516426\left(C_{ijPV} - 32\right)^2 - 106.161\left(C_{ijPV} - 32\right) + 3589.5$

- Income difference is measured by the absolute value of the gap between household income and the MTZ-level average household income. The parameters can be interpreted as the preference of living in a MTZ that the income level is similar from their own. I assume that households prefer to live in the MTZ with higher income level. If such places are beyond their affordability, they may choose MTZs with similar income level as their own.
- The household size similarity is measured by the share of households having the same size in the same MTZ. The parameters can be interpreted as the willingness-to-pay for increasing by 1% of households sharing the same size. My hypothesis is households are willing to live in a MTZ having more households of their size.
- The family structure difference is represented by a dummy variable, measured by the relative similarity between the household and the other households in the MTZ regarding children. If the MTZ has more than 50% of households with children and the individual household has children, the value of family structure difference is 1, otherwise, 0; if a household has no children and the MTZ has more than 50% of households with children, then the value of family structure difference is 0, otherwise, 1. I assume households prefer to live in the MTZ with a similar family structure.
- Ethnic similarity is defined as the percentage of households' in the MTZ having the same ethnicity as the household; the parameters thus represent the willingness-to-pay for living in a MTZ with more people of the same ethnicity. Households may want to pay more for more residents sharing the same ethnicity of their own.

Model estimation will help identify the possible impacts of different household characteristics on willingness to pay for different dwelling attributes.

5.2.2 Household Group Definition

Following Ellickson (1981), households enter a bidding process as homogeneous household groups. I divide the households into eight groups based on 1) the average income per household member, and 2) the household structure (having children or not). Based on Average Monthly Household Income per Household Member data in 2008 from the Department of Statistics Singapore, we divide the income into four groups: low income (<20^{th} percentile), middle income (20^{th} -60^{th} percentile), high income (60^{th} -80^{th} percentile), and very high income (top 20^{th}

percentile). The income groups are denoted in lower case (*l* for low, *m* for middle, *h* for high, *vh* for very high income). Family structure is denoted by C for having children and NC for not having children.

Table 5-1 Household Group Definitions

Household Group	Definition
Low Income without Children (Base group)	Monthly average individual income lower than 20^{th} percentile, Household without children
Low Income with Children (lC)	Monthly average individual income lower than 20^{th} percentile, Household with children
Middle Income without Children (mNC)	Monthly average individual income between 20^{th} -60^{th} percentile, Household without children
Middle Income with Children (mC)	Monthly average individual income between 20^{th} -60^{th} percentile, Household with children
High Income without Children (hNC)	Monthly average individual income between 60^{th} -80^{th} percentile, Household without children
High Income with Children (hC)	Monthly average individual income between 60^{th} -80^{th} percentile, Household with children
Very High Income without Children (vhNC)	Monthly average individual income higher than 80^{th} percentile, Household without children
Very High Income with Children (vhC)	Monthly household income higher than 80^{th} percentile, Household with children

5.2.3 Utility Function

Every household group has a utility function. That is to say, its preference for a given dwelling can be aggregated into household groups:

$$
\begin{aligned}
U_i = ASC_i + \beta_{AREA_i} Area\ +\ \beta_{ACC_i} Accessibility + \beta_{ACC2_i} Accessibility^2 +\ \beta_{SCH_i} School \\
+ \beta_{DEN_i} Density + \beta_{RESALE_i} Resale \\
+ \beta_{INCDF_i} IncomeDifference + \beta_{HHDF_i} HouseholdSize \\
+ \beta_{CHILDF_i} FamilyStructure + \beta_{ETHDF_i} EthnicDifference + \varepsilon_i \\
(i = 1,2, \dots .8 \text{ for all the household groups})
\end{aligned}
$$

5.3 Model Estimation and Results Interpretation

I estimated the model using Ellickson's approach and using multiple imputation to match the two data sets. The final model results are shown in Table 5-2.

Table 5-2 Model estimation results

		β	s.e.	t-test	
lC	ASC_lC	-3.712470	1.453197	-2.554692	**
	area_lC	0.025484	0.134297	0.189755	
	acc_lC	0.607453	0.715158	0.849397	
	sch_lC	0.005562	0.002930	1.898357	**
	density_lC	0.133191	0.085543	1.557005	*
	resale_lC	0.005436	0.162162	0.033525	
	incdiff_lC	-0.178596	0.133745	-1.335350	
	hhsizediff_lC	-3.650379	1.687898	-2.162678	**
	childdiff_lC	3.604090	0.216191	16.670852	**
	ethnicdiff_lC	0.800911	0.289117	2.770193	**
mNC	ASC_mNC	1.075360	0.921082	1.167496	
	area_mNC	-0.232384	0.114978	-2.021112	**
	acc_mNC	0.352218	0.422297	0.834054	
	sch_mNC	-0.000110	0.003009	-0.036649	
	density_mNC	0.124174	0.063535	1.954420	**
	resale_mNC	0.044267	0.105972	0.417724	
	incdiff_mNC	-2.381531	0.102299	-23.280208	**
	hhsizediff_mNC	2.102574	1.061224	1.981273	**
	childdiff_mNC	-0.272878	0.235619	-1.158131	
	ethnicdiff_mNC	0.738219	0.194637	3.792794	**
mC	ASC_mC	0.196393	0.793560	0.247484	
	area_mC	0.182863	0.106456	1.717733	*
	acc_mC	-0.995912	0.362848	-2.744710	**
	sch_mC	0.007644	0.002396	3.190833	**
	density_mC	0.075730	0.060072	1.260647	
	resale_mC	0.091375	0.126968	0.719668	
	incdiff_mC	-2.310932	0.138737	-16.656935	**
	hhsizediff_mC	4.960453	1.088230	4.558278	**
	childdiff_mC	3.341382	0.211840	15.773145	**
	ethnicdiff_mC	0.692676	0.216211	3.203710	**
hNC	ASC_hNC	3.329799	0.658862	5.053868	**
	area_hNC	-0.108671	0.092639	-1.173052	
	acc_hNC	-1.265126	0.279657	-4.523850	**
	sch_hNC	0.005612	0.002312	2.427957	**
	density_hNC	-0.087168	0.065775	-1.325240	
	resale_hNC	-0.141184	0.093588	-1.508568	*
	incdiff_hNC	-0.043340	0.090054	-0.481264	
	hhsizediff_hNC	-1.236237	0.966652	-1.278885	
	childdiff_hNC	0.151111	0.184601	0.818578	

	ethnicdiff_hNC	-0.081134	0.180876	-0.448561	
hC	ASC_hC	-0.628503	0.713538	-0.880827	
	area_hC	0.175118	0.108857	1.608699	*
	acc_hC	-1.785930	0.336619	-5.305499	**
	sch_hC	0.008476	0.002375	3.568626	**
	density_hC	0.281913	0.065723	4.289383	**
	resale_hC	0.175271	0.107167	1.635497	*
	incdiff_hC	0.103117	0.097049	1.062521	
	hhsizedif_hC	4.808999	1.040331	4.622567	**
	childdiff_hC	3.150454	0.177323	17.766786	**
	ethnicdiff_hC	-0.104198	0.187010	-0.557178	
vhNC	ASC_vhNC	-10.871426	1.545065	-7.036227	**
	area_vhNC	-0.147866	0.193140	-0.765592	
	acc_vhNC	1.027316	0.593850	1.729926	**
	sch_vhNC	0.006537	0.003847	1.699457	**
	density_vhNC	-0.127448	0.101174	-1.259691	
	resale_vhNC	0.569672	0.214885	2.651060	**
	incdiff_vhNC	3.919014	0.189247	20.708476	**
	hhsizediff_vhNC	6.778743	1.674711	4.047710	**
	childdiff_vhNC	-0.803125	0.293941	-2.732268	**
	ethnicdiff_vhNC	-0.982157	0.284769	-3.448960	**
vhC	ASC_vhC	-12.029118	1.615176	-7.447557	**
	area_vhC	0.185501	0.206009	0.900451	
	acc_vhC	0.279650	0.670536	0.417054	
	sch_vhC	0.001725	0.005042	0.342049	
	density_vhC	0.744973	0.089341	8.338538	**
	resale_vhC	0.334851	0.189936	1.762965	*
	incdiff_vhC	2.459605	0.193003	12.743864	**
	hhsizediff_vhC	5.875044	1.757903	3.342075	**
	childdiff_vhC	3.354383	0.222800	15.055611	**
	ethnicdiff_vhC	1.758649	0.414417	4.243675	**
	mu2	1.563186	0.064800	24.123385	**
	mu3	1.017078	0.059719	17.031000	**
N	9495				
M	10				
rho2	0.4938247				
Loglik	9514.0101				

** 0.05 * 0.1

The total sample size counts all the matchings by MTZ after each imputation. The average sample sizes for all household groups over the 10 imputations are listed below.

Table 5-3 Household Group Average Sample Sizes

Group	Sample size	Group	Sample size
lC	503	hC	1281
lNC	1143	hNC	1037
mC	1892	vhC	395
mNC	1564	vhNC	1681

In the following discussion, the variables are grouped into three sections: 1) dwelling characteristics, including floor area, sale type of dwellings; 2) zonal characteristics, including accessibility, population density, school enrollment; and 3) household characteristics, including income, household size, structure and ethnicity. The estimated parameters are listed by variables in each section, and all the parameters will be compared with the household group of low income, without children (lNC). Comparisons are also made between all other household groups.

Dwelling characteristics

- **Floor area**

As we can see from Table 5-4, in general, households with children want to pay more for larger dwellings, consistent with our *a priori* assumption. For households with children, the willingness to pay for floor area changes slightly with income. For households without children, the income relationship with dwelling size largely disappears relative to the base group (low income, without children), except for the middle-income families, who may prefer other attributes over floor area and want to arrange their limited income wisely on other attributes.

Table 5-4 Floor area estimation results for Singapore

Households without Children					Households with Children			
	β	s.e.	t-test			β	s.e.	t-test
area_lNC	0				area_lC	0.02548	0.13430	0.18976
area_mNC	**-0.23238**	**0.11498**	**-2.02111**		**area_mC**	**0.18286**	**0.10646**	**1.71773**
area_hNC	-0.10867	0.09264	-1.17305		**area_hC**	**0.17512**	**0.10886**	**1.60870**
area_vhNC	-0.14787	0.19314	-0.76559		**area_vhC**	**0.18550**	**0.20601**	**0.90045**

- **Resale**

Regarding resale units, only very-high-income household groups and high-income households with children are willing to pay more for a resale dwelling, compared to a new sale one. I suspect that new sale or resale might not be a direct factor that affects people's willingness to pay. Rather, resale might be a proxy for other factors like location or neighborhood (more established, more prestigious). And resales have a higher likelihood of being in older neighborhoods. Such neighborhoods may have better 'reputations', and thus the resale is an indicator of buying into existing prestige.

Table 5-5 Resale estimation results for Singapore

Households without Children					Households with Children			
	β	s.e.	t-test			β	s.e.	t-test
resale_lNC	0.00000				resale_lC	0.00544	0.16216	0.03353
resale_mNC	0.04427	0.10597	0.41772		resale_mC	0.09137	0.12697	0.71967
resale_hNC	**-0.14118**	**0.09359**	**-1.50857**		**resale_hC**	**0.17527**	**0.10717**	**1.63550**
resale_vhNC	**0.56967**	**0.21488**	**2.65106**		**resale_vhC**	**0.33485**	**0.18994**	**1.76297**

Zonal characteristics

- **Accessibility**

Tables 5-6, 5-7 show the estimated parameters for the two accessibility variables – the accessibility and accessibility square to capture the bell shaped general cost function.

Table 5-6 Accessibility estimation results for Singapore

Households without Children					Households with Children			
	β	s.e.	t-test			β	s.e.	t-test
acc_lNC	0				acc_lC	0.607453	0.715158	0.849397
acc_mNC	0.352218	0.422297	0.834054		**acc_mC**	**-0.99591**	**0.362848**	**-2.74471**
acc_hNC	**-1.26513**	**0.279657**	**-4.52385**		**acc_hC**	**-1.78593**	**0.336619**	**-5.3055**
acc_vhNC	**1.027316**	**0.59385**	**1.729926**		acc_vhC	0.27965	0.670536	0.417054

Table 5-7 Accessibility square estimation results for Singapore

Households without Children					Households without Children			
	β	s.e.	t-test			β	s.e.	t-test
acc2_lNC	0.00000				acc2_lC	-0.03563	0.08243	-0.43227
acc2_mNC	-0.06298	0.05305	-1.18713		**acc2_mC**	**0.11280**	**0.04915**	**2.29506**
acc2_hNC	**0.13802**	**0.03582**	**3.85267**		**acc2_hC**	**0.24165**	**0.04348**	**5.55769**
acc2_vhNC	**-0.18657**	**0.07679**	**-2.42948**		acc2_vhC	-0.00660	0.08512	-0.07749

As we can see from Table 5-6, for some household groups, the parameters for accessibility are negative, which means they do not want to live too close to job centers. For the same household groups, in Table 5-7, their parameters are positive, which means they still want to live close to job center, but not too close. In other words, they want to keep a reasonable distance with job centers to avoid negative side of job centers while remaining a reasonable proximity. This willingness-to-pay for relative accessibility is more obvious for households with children, who care more about the living environment for their children. One exception is very high income households without children, they are willing to pay for high accessibility, since they do not have those concerns that bother families with children, and thus may value accessibility more than those families with children.

- **School enrollment**

As we can see from Table 5-8, in general, households with children want to pay more for higher school enrollment in the MTZ with an exception of very-high-income households, who might consider private schools. Looking at the households having children, they are willing to pay more for a higher school enrollment MTZ as income grows. Households without children shows the similar income effect since the school enrollment may also associate with other characteristics in the school districts (e.g., pedestrian safety).

Table 5-8 School enrollment estimation results for Singapore

Households without Children				Households with Children			
	β	s.e.	t-test		β	s.e.	t-test
sch_lNC	0			**sch_lC**	**0.005562**	**0.00293**	**1.898357**
sch_mNC	-0.00011	0.003009	-0.03665	**sch_mC**	**0.007644**	**0.002396**	**3.190833**
sch_hNC	**0.005612**	**0.002312**	**2.427957**	**sch_hC**	**0.008476**	**0.002375**	**3.568626**
sch_vhNC	**0.006537**	**0.003847**	**1.699457**	sch_vhC	0.001725	0.005042	0.342049

- **Population density**

The estimation results (Table 5-9) suggest that households with children care more about population density than households without children, and, in general, they have a higher willingness to pay for a denser MTZ, an effect that apparently increases with income.

Table 5-9 Population density estimation results for Singapore

Households without Children				Households with Children			
	β	s.e.	t-test		β	s.e.	t-test
density_lNC	0.00000			**density_lC**	**0.13319**	**0.08554**	**1.55701**
density_mNC	**0.12417**	**0.06353**	**1.95442**	density_mC	0.07573	0.06007	1.26065
density_hNC	-0.08717	0.06578	-1.32524	**density_hC**	**0.28191**	**0.06572**	**4.28938**
density_vhNC	-0.12745	0.10117	-1.25969	**density_vhC**	**0.74497**	**0.08934**	**8.33854**

For most households groups with children in Singapore apparently prefer to live in a denser area might because density may be a proxy for other zonal characteristics such as the presence of amenities (such as play grounds) usually associated with real estate developments, either private or public. As for middle-income families without children, they also are willing to pay more for living in a denser area, maybe because that they also want to take advantage of amenities and facilities that they cannot afford with limited income. .

Household characteristics

- **Income difference**

As we can see from the results, there is no major difference between households with and without children. Very-high-income households prefer to live in a MTZ where the average income is different to their own, but we do not know pay more for places where the income is even higher (social aspirants) or lower (possibly because they have nowhere else to go). The middle-income households, with negative coefficient, prefer to live in a MTZ where the average income is similar to their own.

Table 5-10 Income difference estimation results for Singapore

Households without Children				Households with Children			
	β	s.e.	t-test		β	s.e.	t-test
incdiff_lNC	0.00000			incdiff_lC	-0.17860	0.13374	-1.33535
incdiff_mNC	**-2.38153**	**0.10230**	**-23.28021**	**incdiff_mC**	**-2.31093**	**0.13874**	**-16.65693**
incdiff_hNC	-0.04334	0.09005	-0.48126	incdiff_hC	0.10312	0.09705	1.06252
incdiff_vhNC	**3.91901**	**0.18925**	**20.70848**	**incdiff_vhC**	**2.45960**	**0.19300**	**12.74386**

- **Household size**

The household size parameters can be interpreted as the willingness-to-pay for increasing 1% of households sharing the same size. We can observe from Table 5-11 that the willingness-to-pay for a similar household size varies with family structure and income: households with children like to pay more for having more households of the same size and such willingness-to-pay grows slightly with income (and the very-high-income households like to pay most for more households of their size); for household without children, their preferences vary case by case.

Table 5-11 Household size estimation results for Singapore

Households without Children				Households with Children			
	β	s.e.	t-test		β	s.e.	t-test
hhsizediff_lNC	0.00000			**hhsizediff_lC**	**-3.65038**	**1.68790**	**-2.16268**
hhsizediff_mNC	**2.10257**	**1.06122**	**1.98127**	**hhsizediff_mC**	**4.96045**	**1.08823**	**4.55828**
hhsizediff_hNC	-1.23624	0.96665	-1.27888	**hhsizedif_hC**	**4.80900**	**1.04033**	**4.62257**
hhsizediff_vhNC	**6.77874**	**1.67471**	**4.04771**	**hhsizediff_vhC**	**5.87504**	**1.75790**	**3.34207**

- **Family structure similarity**

As we can observe from the results, in general, households with children have similar and significant higher willingness-to-pay for living in a children-oriented MTZ.

Table 5-12 Family structure difference estimation results for Singapore

Households without Children				Households with Children			
	β	s.e.	t-test		β	s.e.	t-test
childdiff_lNC	0.00000			**childdiff_lC**	**3.60409**	**0.21619**	**16.67085**
childdiff_mNC	-0.27288	0.23562	-1.15813	**childdiff_mC**	**3.34138**	**0.21184**	**15.77315**
childdiff_hNC	0.15111	0.18460	0.81858	**childdiff_hC**	**3.15045**	**0.17732**	**17.76679**
childdiff_vhNC	**-0.80312**	**0.29394**	**-2.73227**	**childdiff_vhC**	**3.35438**	**0.22280**	**15.05561**

- **Ethnic Similarity**

Ethnic similarity is defined as the percentage of households' in the MTZ having the same ethnicity as the household; the parameters thus represent the willingness-to-pay for living in a MTZ with more people of the same ethnicity. Middle-, low-, very high-income households with children and middle-income household without children all seem to prefer more similar ethnicities than low-income household without children. Very high-income households without children have less relative preference for similar ethnicity in their MTZs.

Table 5-13 Ethnicity estimation results for Singapore

Households without Children				Households with Children			
	β	s.e.	t-test		β	s.e.	t-test
ethnicdiff_lNC	0.00000			**ethnicdiff_lC**	**0.80091**	**0.28912**	**2.77019**
ethnicdiff_mNC	**0.73822**	**0.19464**	**3.79279**	**ethnicdiff_mC**	**0.69268**	**0.21621**	**3.20371**
ethnicdiff_hNC	-0.08113	0.18088	-0.44856	ethnicdiff_hC	-0.10420	0.18701	-0.55718
ethnicdiff_vhNC	**-0.98216**	**0.28477**	**-3.44896**	**ethnicdiff_vhC**	**1.75865**	**0.41442**	**4.24368**

Chapter 6

Conclusions and Discussion

The previous chapters reviewed the available methods regarding household residential choice, applied Multiple Imputation to deal with inconsistent data sets, and then tested Ellickson's model specification under the bid-rent approach. This chapter concludes.

6.1 Conclusions

Household behavior plays a crucial role in urban system performance and can profoundly shape the urban landscape. This thesis examined households' behavior in the housing market. Such behavior models rest on a basic microeconomic framework. This framework assumes that the ultimate goal of a household's behavior is to maximize the combined utility for all of its members, given income. Random bidding models of the housing market can account for price evolution and market clearance, and thus are superior to hedonic models. However, current integrated urban systems models provide few insights into the capability of random bidding models for simulating household residential choice behavior; rarely have random bidding models been applied in a micro-simulation context, due to insufficient data. Therefore, this thesis explored a possible technique – Multiple Imputation – to integrate observations from dissimilar data sets to meet the data requirements of random bidding models of the housing market, and to test the capability of such a model.

The major conclusions drawn from this research cover two areas: the matching technique for inconsistent data sets and the variations of willingness-to-pay for different households groups.

First of all, regarding the matching technique for inconsistent data sets, the Multiple Imputation method is shown to be a feasible tool for household residential choice. The Multiple Imputation method does not require micro-level connections across the available data sets, but connects observations across the data sets through the shared spatial unit, which is large enough to capture the variations in households and/or dwelling characteristics, yet small enough to provide some spatial resolution of relevance (planning zone). The application of this method retained most of the records. The approach could be generalized for inconsistent data treatments in other contexts.

Secondly, as we expected, household income and family structure both have impacts, and act together to determine households' willingness-to-pay for certain dwelling and location attributes. In general, family structure shapes households' willingness-to-pay: households with children, unsurprisingly, tend to have a higher willingness to pay for a larger dwelling in a MTZ with higher school enrollments, higher population density and more households that are similar to them in terms of income, family structure and ethnicity.

- Flour area: the willingness to pay for floor area changes slightly with income for households with children. For households without children, the income relationship with dwelling size largely disappears relative to the base group (low income, without children), except for the middle-income households without children.
- Sale type: regarding resale units, we cannot easily conclude that some household groups have certain preferences for dwellings in the resale market. Resale dwellings may well be proxies for other unobservable attributes related to the neighborhood or even the dwelling unit quality. For example, for a city like Singapore, newly developed real estate is often located in newly developed areas, which might be attractive in some ways and unattractive in other ways (e.g., with poor public transit access) that our model does not adequately capture.
- Accessibility: middle- and high-income households with children and high-income households without children want to live in places with high accessibility, but not “too high.” One exception is very high income households without children, who are willing to pay for high accessibility.
- School enrollment: in general, households with children want to pay more for higher school enrollment in the MTZ with an exception of very-high-income households, who might consider private schools. Looking at the households with children, they are willing to pay more for a higher school enrollment MTZ as income grows. Households without children show a similar income effect.
- Population density: results suggest that households with children care more about population density than households without children, and, in general, they have a higher willingness to pay for a denser MTZ, an effect that apparently increases with income.
- Income difference: there is no major difference between households with and without children. Very-high-income households prefer to live in a MTZ where the average income is different to their own. The middle-income households, with negative coefficient, prefer to live in a MTZ where the average income is similar to their own.
- Household size: the willingness-to-pay for a similar household size varies with family structure and income: households with children like to pay more for having more households of the same size and such willingness-to-pay grows slightly with income; for household without children, their preferences vary.

- Family structure similarity: households with children have similar and significantly higher willingness-to-pay for living in a children-oriented MTZ.
- Ethnic Similarity: middle-, low-, very high-income households with children and middle-income household without children all seem to prefer more similar ethnicities than low-income household without children. Very high-income households without children have less relative preference for similar ethnicity in their MTZs.

This model shows the feasibility of using Multiple Imputation to match inconsistent data sets with a careful choice of matching unit. This model can be further improved by adding more variables that describe the dwelling as well as zonal characteristics more accurately.

6.2 Implications

Several direct policy and planning implications of the household residential choice findings and the modeling techniques can be identified.

- Improvement of data quality and compatibility

In the age of increasingly abundant big data, compatibility of data from different sources remains a challenge. Agencies tend to collect data serving certain purposes only (e.g., travel behaviour, real estate sales), and often such data cannot easily be adapted for other purposes and/or made compatible with other data sets. Such a myopic data collection strategy not only wastes resources but also impedes research requiring comprehensive data sets. Collaborations within and among agencies can dramatically improve data quality and compatibility, and subsequent analysis, and agencies can work together to collect the mutually beneficial data with limited resources.

- Dealing with inconsistent data sets

To my knowledge, this research represents the first time the Multiple Imputation method has been used to match inconsistent data sets and is shown to be a feasible tool to integrate inconsistent data sets for household residential choice modeling. The Multiple Imputation method could be a possible tool to match different data sets for other modeling uses as well. This method can help other researches that are constrained by inconsistent data to deal with their data problem. But this method requires connections between data sets that need to be matched.

- Matching unit choice

In this study, Multiple Imputation takes the households that live in the same spatial unit – MTZ– as a possible household pool. The alternative spatial units are either too big (like a planning area), or too small (like the postcode) to generate a possible household pool. The choice of matching unit needs to consider the sizes of the household pools generated from the selected spatial unit; the household pool sizes need to be large enough to represent the household characteristics in that spatial unit. If the unit is too small, say postcode level in this study, there might be few or even no observation to sample. With the household pools large enough for each spatial unit to represent the distribution of households living in that spatial unit, the smaller the spatial unit we choose, the more precise this match can be.

6.3 Limitations and future directions

Although this research has made contributions to the treatment of inconsistent data sets and the microscopic estimation of one of the bid-rent models, some relevant research questions remain unanswered and should be addressed in future research.

1) Grouping households

This thesis uses Ellickson's approach, in which households are grouped. Such a grouping could be problematic and lose some important household-specific features. In this thesis, households are grouped by income and with or without children only, which can be improved in future research by more comprehensive group criteria. Grouping households is also a limitation for Ellison's approach, thus a more advanced approach without grouping would solve this problem.

2) Missing part of the actual housing market (HDB).

The question pertains to modeling the household residential choice in a restricted market. Due to data constraints, this thesis only focused on the private housing market in Singapore, which can be seen as a relatively free market. However, Singapore's full housing market is mixed, with the major share provided as a form of "public" housing (HDB).

Future research could work on modelling the entire urban housing market. In such a market, a household that considers moving will first face the choice to buy a new dwelling or just rent one. If they decide to rent, then they face the choice of what kind of dwelling they would like to rent – a private or public dwelling. If they decide to buy, then the choices become: buying a dwelling from the new sale or resale market, and buying a public or private dwelling. Such choices that a household faces when deciding to move to a new dwelling can be framed as a nested structure, as shown in Figure 6-1.

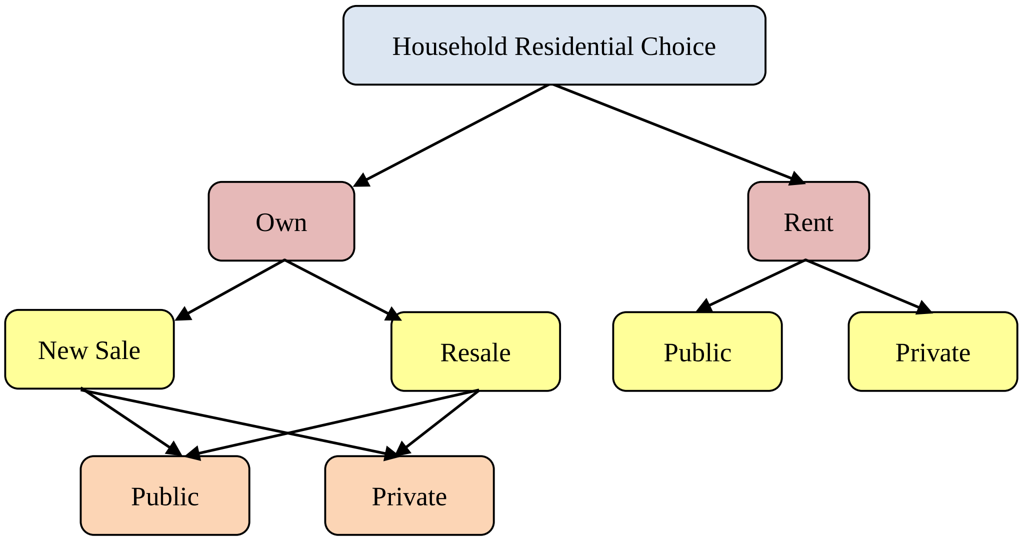


Figure 6-1 Household Residential Choice Scheme

Incorporating that market, and its interaction with Singapore's private housing market is an area for future research. Appendix 2 shows a proposed survey design that would help collect sufficient data for future research towards this end.

3) Missing attributes of dwelling

Due to data constraints, this study only includes two dwelling attributes – floor area and sale type, which is rough in presenting the characteristics of a dwelling. Some attributes like the construction year, floor level, and the availability of other facilities (e.g., elevator) can help represent the dwelling better. In future research, more variables depicting the dwelling characteristics could be included to describe a dwelling more accurately and more distinctively in the housing market.

4) Missing a range of attributes of the zones

This study only includes three zonal attributes (population density, accessibility, school enrollment, and the similarity between households' and zonal characteristics) in the model. Some of the current measures are weak. The accessibility measure only considers six types of job-related opportunity. However, in the reality, accessing many other kinds of opportunities, like school or social places, might be equally important to households when deciding where to live. Moreover, the measures of similarity between households' and zonal characteristics used in this thesis is also roughly proxied, and thus more comprehensive measures are recommended for future research.

Many attributes that may influence choices are missing. Attributes like green space, amenities, prestige of community/neighborhood and others can represent the neighborhood or community more accurately, and might be important to households when they make residential decisions. Future research may consider to include more zonal attributes to describe the zonal characteristics more accurately and thus could model more variables that influence choices.

In addition to the limitations to and potential improvements of this study, there are several areas in which similar studies in the future can make progress.

The evaluation of location can be improved. More comprehensive measurement of location, like accessibility, can be approved for a realistic description. Parts of the location attributes are related only to the building environments (like distance to grocery store/hospital) and the other parts are more activity-based and highly personalized attributes (like distance to *my* workplace and distance to *my child's* school). Therefore, personalized location measurement could expand our understanding of building environment relationships with behaviors, and also connects transportation planning more closely to land use and building environments.

Finally, future research might introduce more social, cognitive, and emotional factors in modelling the household's behavior. For example, future research can try to relax the rationality assumption behind the utility maximization assumption. Households might not be rational enough to maximize their utility all the time. They might be subject to some biases. Possession bias, for example, could make households overestimate the relative utility of living in their current dwelling when making moving decision (Kahneman, 2011). Households may also act differently when they make residential choice under risk, according to Prospect theory (Kahneman, 1979). Incorporating such behaviour economics theory and research can help better understand household behavior in the housing market and help predict household's behavior in the housing market more accurately.

Appendix 1

Ideal dataset description

In order to fully model the household residential choices in the housing market, more comprehensive data are required. Table A1-1 outlines the ideal data set that would support the estimation of microscopic household residential choice models. The ideal data set would include three main parts (see Table A1-1): household characteristics, the household member characteristics, and the dwelling attributes. Ideally, the data set for dwelling attributes would include the detailed bargaining process, including the evolution of the asking price from the seller and the bid from the household.

Table A1-1 Ideal Household Information Data Description

Name	Notes
Household (HH) information	
HH size	
HH income	
Number of children in the HH	
Number of school-aged children in the HH	
Car ownership	
Bike ownership	
Name of school the children attend	(Singapore-specific)
Travel time to school	
Travel mode to school	
Number of working people in the household	
Average travel time for HH to work	
Travel mode to work	
Alternative dwelling considered before purchase	Other dwellings that this household bid for
Household Member Information	
HH member	
Ethnicity of member of HH	
HH member 's age	
HH member 's gender	
HH member 's occupation	
HH member’s working hours	
HH member 's individual income	
HH member 's travel time to work	
HH member 's work place	Postcode level

Dwelling information	
Dwelling type	HDB/Private/Land property
Dwelling ownership	Own (HDB new sale/ HDB resale/Private new sale/ Private resale), Rent (whole unit/room)
Dwelling value (bid paid by the winner)	Property value/Rent (monthly)
Dwelling size/floor area	(Approx.: # of bedrooms)
Purchase date (year)	For own only
Length of stay	
Dwelling's construction year	
Dwelling address	(Postcode for Singapore)
Asking price of the seller	Changes over time
Other bid received by the seller for this dwelling	
Bid rejected by the seller	

As we can see from the ideal data set description, it might not be feasible to collect some of the critical information in the situation in Singapore of 2013. Hence, in order to estimate a more advanced model like Lerman and Kern's approach, a minimum requirement for a sufficient data set is proposed in Table A1-2. The minimum data set retains the core information for estimating Lerman and Kern's model, while dropping the asking price and bid evolutions. It should be feasible to collect the minimum data set at a micro level at this time, and a questionnaire designed to collect such a data set is shown in Appendix 2.

Table A1-2 Minimum Singapore Data Description

Name	Notes
Household information	
HH size	
HH income	
Number of children in the HH	
Number of school-aged children in the HH	
Car ownership	
Household Head Information	
HH head	
HH head's age	
HH head's gender	
HH head's occupation	
HH head's travel time to work	
Travel mode to work	
Dwelling information	
Dwelling type	HDB/Private/Land property
Dwelling ownership	Own (HDB new sale/ HDB resale/Private new sale/ Private resale), Rent (whole unit/room)
Dwelling value	Property value/Rent (monthly)
Dwelling size/floor area	(Approx.: # of bedroom)
Purchase date (year)	For own only
Length of stay	
Dwelling address	(Postcode for Singapore)

Appendix 2

Survey Instrument Design

This proposed survey instrument is designed to collect sufficient information on households' residential behavior and location choice in a timely fashion. Data collected from this survey would enable the estimation of microscopic household residential choice models. The survey instrument design proposed in the following part uses Singapore as an example.

Sampling frame

The sampling frame for residential choice in Singapore is targeted at the households that:

- have at least one legal resident of Singapore:
  Singapore Citizen
  Permanent Residents (PR)
- occupy a whole housing unit (exclude the cases of multiple households in one dwelling unit)
- are eligible or were eligible to purchase HDB to simulate the case of free choice

Such a sampling frame will guarantee the long-term residency as well as the eligibility for both HDB and private housing of the households who intend to move. Similar to HITS, "the household" here refers to a person living alone or a group of two or more persons living together in the same unit and sharing cooking or other arrangements for essential living. (For households comprising family members, this includes any family member who lives in the same dwelling but might not eat with them.).

- Each lodger who pays for the room or living space only and not for food is treated as a separate household.
- Relatives of the family, including married children, who eat with the family but do not live with the family, are excluded.

In order to represent all agents in the whole choice scheme of Singapore's housing market, as shown in Figure 6-1, the households sampled for this research would include residents:

- If owner:
  - HDB new sale
  - HDB resale
  - Private new sale

    - Private resale
- If renter:
    - HDB
    - Private

The next section will discuss the detailed sampling approach to cover each set of choices. There are two types of potential populations: 1) recent movers; and 2) all resident households. If the research is focusing on recent household mobility, within a short time period with a limited budget, the recent mover population is a better target; while the whole resident population can serve better for research with more resources.

Sampling method

A stratified sampling approach is proposed, with the strata defined based on the type of dwelling each household occupies. Based on the choice scheme, the strata include: HDB new sale owners, HDB resale owners, private new sale owners, private resale owners, HDB renters, and private renters. This ensures the independence and mutual exclusiveness of the strata.

The sample size allocated to each stratum follows a proportionate allocation based on the market segmentation of each housing purchase type. By doing so, the samples in each stratum are with the same probability of selection. If we let n_g denote a sample size for stratum g and n denote sample size for all strata, then n_g/n is the same as the proportion of elements in the population $W_g = N_g/N$, where N_g denotes the number of population elements in stratum g.

We weight stratum results by the population proportions W_g, which are based on the latest data from SingStat and the quarterly housing market data from the Urban Redevelopment Authority (URA) and HDB. There are two ways to calculate the weights based on the population targets. If the population is recent movers, then the weights will be calculated based on the recent transactions. The weights for the recent mover strata are shown in Table A2-1:

Table A2-1 Transaction-based weights

	HDB	Private	Seed	**Updated**
New sale	26.0%	5.8%	31.8%	**31.8%**
Resale	24.1%	9.7%	33.9%	**33.9%**
Subletting	26.1%	8.2%[7]	34.3%	**34.3%**
Seed	76.2%	23.8%		
Updated	**76.2%**	**23.8%**		

The proportion for each segment is calculated based on the transaction records for HDB and private housing in 2012 (see Table A2-2 and A2-3).

Table A2-2 2012 Singapore Public Housing Market Statistics

HDB 2012	**# of transactions/approval**
New sale	27000
Resale	25094
Subletting	27129

Table A2-3 2012 Singapore Private Housing Market Statistics

	New Sale		**Sub Sale**		**Resale**	
Private 2012	Core Central	Rest of Central	Core Central	Rest of Central	Core Central	Rest of Central
1Q/2012	138	1113	77	179	591	692
2Q/2012	444	1221	122	299	1316	1126
3Q/2012	703	846	118	250	1578	1129
4Q/2012	610	987	106	199	1457	877
Total	**6062**		**1350**		**8766**	

The new sale, resale and subletting are roughly 1:1:1 for HDB and 2:2:1 for private housing.

The other set of weights is calculated if the targeted population is all residents. The weights for the whole resident population are shown in Table A2-4:

[7] There is no private renting rate available; the proportion of renting rate in private apartments and condos is inferred from the advertisement numbers from Property Guru.

Table A2-4 Population-based weights

	HDB	Private	Seed	**Updated**
New Sale	40.0%	7.1%	90.1%[8]	**47.1%**
Resale	35.0%	8.0%		**43.0%**
Rent	6.6%	3.3%	9.9%	**9.9%**
Seed	81.6%[9]	18.4%		
Updated	**81.6%**	**18.4%**		

The proportion for each segment is calculated using Iterative Proportional Fitting (IPF) and also shown in Tables A2-2 and A2-3.

This survey can be seen as a probability sample that is developed using a sampling protocol based on the choice probability revealed from the recent transactions or whole resident population.

In general, using all residents as the population is not as cost-effective as the recent mover population, but is a more convenient random sample, while the recent mover population is hard to identify in many cases. The choice of population will depend greatly on the resources and the convenience of the sample that a future research group might have.

Sample size

Sufficient sample size for each household in the estimation process can be estimated in the following way:

$$Total\ Sample\ Size\ N_S = \frac{Z_{\frac{\alpha}{2}}^2 \left(\frac{\sigma}{\mu}\right)^2}{d^2}$$

d: Allowable error

1-α: Desired confidence

σ^2: Population variability

μ: Population mean

$$Sample\ Size\ of\ Strata\ g\ N_{sg} = w_g N_s$$

w_g: the proportion of group *g* in population (see previous table)

Given budget constraint:

$$N_{sg} = \frac{w_g \sigma_g / \sqrt{c_g}}{\sum_{g' \neq g}\left(w_{g'} \sigma_{g'} / \sqrt{c_{g'}}\right)} N_s$$

c_g : the unit cost of data collection in stratum *g*

[8] The total proportion of owning a dwelling unit comes from the Home Ownership Rate from SingStat.

[9] These percentages come from the latest data of Household and Housing from SingStat.

σ_g: the standard deviation of stratum g

Survey Instrument

Two alternatives are explained below. The choice of survey instrument will highly depend on the resources and the target population that a future research group might have.

1) Interview-based household survey

In order to ensure adequate response rate and validity of the survey results, the questionnaire is proposed to be conducted by means of face-to-face interviews with at least one eligible member of sample households, who can provide information regarding their household socio-economic and demographic characteristics and the attributes of dwellings they occupy.

The reliability of the survey can be controlled by interviewers, and makes it easier to manage the proportions of the strata. The number of interviews can be controlled to be as small as the minimum sample size.

Conducting the interview-based survey with limited resources, I propose to sample residents living in one planning area with both an HDB cluster and private residential buildings. With very limited samples, controlling the residential location decreases the variance of the building environment. Controlling the residential location, we only control the place attributes, and it won't affect the accessibility for households since the travel behavior and pattern, especially the commuting time are asked in the survey.

Controlling location is also consistent with the two-stage model structure which first models the residential location choice and then passes the competing households to the bidding process. Controlling location is actually fixing the first stage and go directly to the second.

2) Email/web- based household survey

The email/web-based survey is to contact transaction makers for HDB and private housing as well as the HDB resale applicants as a supplement to HDB, REALIS transaction records.

The advantages of this method include:

a. it is easier to conduct and comparably inexpensive if we have the support from HDB and REALIS;

b. We have control over the transaction dates, which make the price/value comparable among the dwellings.
c. The sample size can be much larger than the interview-based survey

The disadvantages include:

a. The sample will be restricted to the available transaction and might not represent the whole population
b. We have no control over the proportion of the strata in the sample. The size of certain strata might not be sufficient for our estimation.
c. The email/ web might not access to every transaction makers
d. The response rate might be very low

Therefore, the choice of survey instrument, again, will highly depend on the resources and the target population that future research group might have.

This survey design tries to use stratified sampling to cover all agents in the whole choice scheme of Singapore's housing market. This survey can help researchers collect data that can be used to represent and model Singapore's entire housing market in the future.

Reference

Aeroe, T. (2001). Residential Preferences, Choice of Housing, and Lifestyle. PhD dissertation (English summary), Aalborg University

Alonso, W. (1964). Location and Land Use: Toward a General Theory of Land Rent, Harvard University Press, Cambridge,Massachusetts.

Ballas, D., Clarke, GP., and Wiemers, E. (2005). Building a Dynamic Spatial Microsimulation Model for Ireland, *Population, Space and Place, 11,* 157-172.

Bardhana, A. D., Dattaa, R., Edelsteina, R. H. and Kimb, L. S. (2003). A Tale of Two Sectors: Upward Mobility and the Private Housing Market in Singapore, *Journal of Housing Economics Volume 12, Issue 2, June 2003, Pages* 83–105.

Ben-Akiva M.E. and Lerman, S. R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand, The MIT press, Cambridge MA.

Ben-Akiva, M. Bowman, J. and Gopinath, D. (1996). Travel Demand Model System for the Information Era. *Transportation 23:* 241-266.

Ben-Akiva, Moshe, & Bowman, J. L. (1998). Integration of an Activity-based Model System and a Residential Location Model. *Urban Studies, 35(7),* 1131–1153

Brown, L.A. and Moore, E.G. (1970). The Intra-Urban Migration Process: A Perspective, *Geografiska Annaler, 52B,* 368-381.

Bush, S. (2003). Forecasting 65+ Travel: An Integration of Cohort Analysis and Travel Demand Modeling. PhD Dissertation, Department of Urban Studies and Planning, MIT, Cambridge, MA

Chattopadhyay, S. (1998). An Empirical Investigation into the Performance of Ellickson's random Bidding Model, with an Application to Air Quality Valuation, *Journal of Urban Economics 43(2):* 292 – 314.

Clark, W.A.V., and Huang, Y. (2003). The Life Course and Residential Mobility in British Housing Markets, *Environment and Planning A, Vol. 35,* pp. 323-339.

Chen, Y. (2013) Neighborhood design and the energy efficiency of urban lifestyle in China : treating residence and mobility as lifestyle bundle. PhD Dissertation, Massachusetts Institute of Technology.

Ellickson, B. (1981). An Alternative Test of The Hedonic Theory of Housing Markets, *Journal of Urban Economy 9,* 56-79.

Ettema, D., de Jong, K., Timmermans, H. and Bakema, A. (2005). PUMA Multi Agent Modeling of Urban Systems, in*: 45th Congress of the European Regional Science Association, Vrije Universiteit, Amsterdam*.

Gross, D. J. (1988). Estimating Willingness To Pay For Housing Characteristics: An Application of the Ellickson Bid-Rent Model, *Journal of Urban Economics 24(1):* 95 – 112.

Gross, D. J., Sirmans, C. and Benjamin, J. D. (1990). An Empirical Evaluation of The Probabilistic Bid-Rent Model: The case of homogenous households, *Regional Science and Urban Economics 20(1):* 103 – 110.

Guevara, C. A. (2005). Addressing Endogeneity in Residential Location Models. PhD Dissertation, Massachusetts Institute of Technology.

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2), 73-76.

Holm, E., Holme, K., Makila, K., Kauppi, M.M. and Mertvik, G. (2002). The SVERIGE Spatial Microsimulation Model: Content, Validation, and Example Applications, In: *GERUM kulturgeografi. Umea, Sweden: CERUM Kulturgeografi, Kulturgeografiska Institutionen/SMC, Umea University.*

Hurtubia, R. and Bierlaire, M. (2012). Estimation of Bid Functions for Location Choice and Price Modeling With a Latent Variable Approach, *Technical Report TRANSP-OR 120206, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.*

Horowitz, J. L. (1986). Bidding Models of Housing Markets. *Journal of Urban Economy: 2,* 168-190.

Jara-Díaz, S. R. andMartínez, F. J. (1999). On The Specification of Indirect Utility and Willingness to Pay for Discrete Residential Location Models, *Journal of Regional Science 39(4):* 675–688.

Jiang, S., J. Ferreira, and M. González. (2013). Inferring Human Activities by Mining Spatiotemporal Destinations in Trajectory. The ACM SIGKDD International Workshop on Urban Computing

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.

Lee, B.H.Y, Waddell, P. (2010). Residential Mobility and Location Choice: a Nested Logit Model with Sampling of Alternatives, *Transportation 37:*587-601

Lerman, S.R. and Kern, C. R. (1983). Hedonic Theory, Bid Rents, and Willingness to Pay: Some Extensions of Ellickson's Results, *Journal of Urban Economy:* 13, 358-363

Little, R.J.A., and Rubin, D.B. (1987). Statistical Analysis with Missing Data. New York" John Wiley and Sons.

Mas-Colell, A. (1975) A Model of Equilibrium with Differentiated Commodities, *Journal of Mathematical Economy, 2,* 263-295.

Martinez, F. J. (1992). The bid-choice land use model: an integrated economic framework. *Environment and Planning A*, *24*(6), 871-885.

Martínez, F. (1996). Mussa: Land use model for santiago city, *Transportation Research Record: Journal of the Transportation Research Board* 1552(1): 126–134.

Martinez, F. J. (2000). Towards a land use and transport interaction framework. *Handbook of transport modelling*, *1*, 393-407.

McCarthy, K.F. (1976). The Household Life Cycle and Housing Choices, *Papers of the Regional Science Association, Vol. 37,* pp. 55-80.

McFadden, D. (1978). Modeling the Choice of Residential Location, A. Karlqvist (ed.), Spatial Interaction Theory and Residential Location, North-Holland, Amsterdam, pp. 75–96.

McFadden, D. (1999) “Rationality for Economists?” *Journal of Risk and Uncertainty, Vol. 19, No. 1-3, pp.* 73-105.

Miller, E.J., Salvini, P. (2005) ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems, *Networks and Spatial Economics 5, Netherlands. pp* 217-234.

Ministry of National Development of Singapore. (2013). A High Quality Living Environment for All Singaporeans: Land Use Plan to Support Singapore's Future Population, Government of Singapore, Ministry of National Development

Mitton, L., Sutherland, H., Weeks, M. (Eds) (2000) Microsimulation Modelling for Policy Analysis: Challenges and Innovations, Cambridge University Press, Cambridge.

Moore, T. and Thorsnes, P. (1994) The Transportation/Land Use Connection, American Planning Association, Planning Advisory Service, Report Number 448/449, Chicago.

Mulder, C.H. and Wagner, M. (1993) Migration and Marriage in the Life Course: A Method for Studying Synchronized Events, *European Journal of Population, Vol. 9, pp*. 55-76.

Muto, S. (2006). Estimation of the Bid Rent Function with the Usage Decision Model, *Journal of Urban Economics 60(1):* 33 – 49.

Orcutt, G.H. (1957). A New Type of Socio-economic System, *Review of Economics and Statistics, 39,* 116-123.

Oskamp, A. (1993). Simulating Housing Search; A Model of Individual Decision and Acceptance Rules, Paper presented at the 8th European Theoretical and Quantitative Geographical Colloquium, Budapest, 12-16 September 1993.

Oskamp, A. (1994). A Probabilistic Heuristic Housing Search Model, paper prepared for the Summer School 'Choice Processes in Demography, 6-20 July, Wassenaar, The Netherlands.

Oskamp, A. (1994). LOCSIM: A Probabilistic Model of Choice Heuristics, *Netherlands Journal of Housing and the Built Environment, vol. 9, pp.* 285-309

Phang, Sock Yong (2001). Housing Policy, Wealth Formation and the Singapore Economy. *Housing Studies, Vol.16, No.4,* 443-459.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, The *Journal of Political Economy 82(1):* 34 – 55.

Rossi, P.H. [1955] (1980) Why Families Move, Sage Publications, Beverly Hills and London.

Sim L L,Yu S M, Han S S, (2003). Public Housing and Ethnic Integration in Singapore. *Habitat International 27* 293–307

Y. Tu, L. Kwee, B.K.P. Yuen. (2005). An Empirical Analysis of Singapore Households' Mobility Behavior from Public Ownership to Private Ownership. *Habitat Int., 29 (3) pp.* 511–525

United Nations (2011). Population Distribution, Urbanization, InternalMigration and Development: An International Perspective, United Nations, Department of Economic and Social Affairs, Population Division.

Vencatasawmy, C.P., Holm, E., Rephann, T., Esko, J., Swan, N., Ohman, M., Astrom, M., Alfredsson, E., Holme, K. and Siikavaara, J. (1999). Building a Spatial Microsimulation Model, *11th European Colloquium on Quantitative and Theoretical Geography in Durham, England.*

Waddell, P. (2001). Between Politics and Planning: UrbanSim as a Decision Support System for Metropolitan Planning, in Building urban planning support systems: combining information, models, visualization, Klosterman, R. and R. Brail (eds), Center for Urban Policy Research.

Waddell, P. (2004). Urban Simulation at the Intra-Metropolitan Scale: Model Design and Application in the Central Puget Sound Region, Paper presented at The State of the Art in Regional and Urban Modeling, Seville, Spain.

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M. and Ulfarsson, G. (2003). Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim, Preprint of an article that appeared in *Networks and Spatial Economics, Vol. 3, No.1, pp*. 43-67.

Walker, J. (2001). Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables. PhD Dissertation, Massachusetts Institute of Technology.

Walker, J., & Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems, 9(1),* 77–101.

Wegener, M. (1995). Current and Future Land Use Models, Proceedings of Land Use Modelling Conference, Arlington (TX).

Wegener, M., Spiekermann, K. (1996). The Potential of Microsimulation for Urban Models, In: Clarke, G. (ed.): Microsimulation for Urban and Regional Policy Analysis, *European Research in Regional Science 6, London: Pion,* 146-163.

Wegener, M. (1998). Applied Models of Urban Landuse, Transport and Environment State of the Art and Future Developments. Springer Verlag, Berlin.

Xiang, Y. (2014). Dynamics of Vehicle Ownership in Singapore. Master thesis, Massachusetts Institute of Technology.

Zhu, Y. & Ferreira, J. (2014) Synthetic Population Generation for Land Use and Transportation Microsimulation with Multiple Limited Data Sets. Paper presented at Transportation Research Board 2014.