1-1-2010

Assessing Portland's Smart Growth: A Comprehensive Housing Supply and Location Choice Modeling Approach

Hongwei Dong
Portland State University

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Assessing Portland's Smart Growth: A Comprehensive Housing Supply and Location

Choice Modeling Approach

by

Hongwei Dong

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Urban Studies

Dissertation Committee:
John Gliebe, Chair
James Strathman
Anthony Rufolo
Gerard Mildner
Jiunn-Der Duh

Portland State University
©2010

ABSTRACT

There are extensive empirical studies on the impacts and effectiveness of Smart Growth policies; however, very few of them consider the perspective of individual decision makers and, to this author's knowledge, none have studied developers as location-aware decision-making agents. This study tries to fill this gap partially by assessing the impacts of Portland's smart growth policies on developers' location choice behavior with developer-based location choice models.

The dissertation has two purposes. By assessing the impacts of Smart Growth policies on individual home developer's location choice, it provides a micro- and behavioral foundation for the understanding of Smart Growth policies. As a bi-state metropolitan area located on the border between Oregon and Washington, the Portland region provides a unique environment that allows my research to examine whether home developers react to Smart Growth policies differently in the two states with different land use policy systems. The dissertation also aims to create a developer-based land development forecast model, which can be used as a scenario analysis tool for the Portland region's long-term land use and transportation planning. Besides the developer location choice model mentioned above, the components of this comprehensive developer-based land development model also include a time series regression model that predicts annual new housing supply in the region and a model that synthesizes housing projects in a forecast year.

The study shows that home developers in the Portland metropolitan area are sensitive to most Smart Growth policies that have been implemented in the region, but they react to

them differently across the border between Oregon and Washington. Single-family home (SFH) and multi-family home (MFH) developers show different preferences for location attributes. The most significant predictors of where a developer will choose to locate a project are the locations of previous projects. After controlling for all of the other factors discussed above, there remains a strong preference for developing SFH units outside of the UGB in both Oregon and Washington sides of the Portland metropolitan area.

Latent class models have been developed to detect taste variations among home developers in the SFH and MFH markets separately. Estimation results show clear taste variations across developers and housing projects with respect to site attributes in their location choice. With other variables in the segmentation model being the same, project size provides a better fit to the data than developer size, indicating that developers have taste variations among their different projects. Large size SFH projects developed by contractor-owners are more likely to be within the UGB and their locations tend to have higher residential density, housing diversity, transportation accessibility, road density, and land price. With most MFH projects within the UGB, estimation results show that large size MFH projects prefer the locations with higher residential density, housing diversity, mixed use, road density, land price, average household income, and proportion of young and middle age households.

The three-step new housing supply and location choice forecast model seem to be able to capture the basic trend of housing market and land development in the Portland region. Three different aggregate housing supply forecast models, an conditional time series regressive model, a unconditional time series regression model, and an auto-regression

integrated moving average (ARIMA) model were tested and their advantages and disadvantages were discussed. Both the SFH and MFH project synthesis models can simulate housing projects well for a forecast year.

Three location choice models were developed to allocate synthesized housing projects into space. The three models are characterized separately as: (1) assumed market homogeneity and atomization of development projects; (2) deterministic market segmentation and synthesis of projects by size; and (3) probabilistic market segmentation and synthesis of projects by size, using a latent class approach. Examination of forecast results shows that all three models can successfully capture the basic spatial pattern of housing development in the region; however, the spatial distribution of MFH development is lumpier and more unpredictable. While Models 2 and 3 are more sophisticated and make more sense from a theoretical perspective, they do not return better forecast results than Model 1 due to some practical issues. Models 2 and 3 would be expected to perform better when those practical issues are solved, at least partially, in future research.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Research Purposes

This dissertation project was inspired by two research gaps found by literature review. First, despite the popularity of Smart Growth policies, much of the current understanding of the effectuality of those policies lacks a micro-level behavioral foundation. There are some studies that use agent-based models to test the impacts of land use on individuals' residential location choice, but being household-based represent only the demand side of the land development market. The other side of the market, the provision of housing supply as represented by developers remains under-studied. At first glance, this is puzzling because real estate developers are the primary urban space producers and their perspective on Smart Growth policies is so critical to the policy's success.

Secondly, simulation of the housing market is one of the most important components of integrated land use-transport models. Most extant integrated models, however, have focused lopsidedly on the demand for dwelling units by households, with scant attention paid to the supply side of the market, individual home developers. As Hunt *et al*. (2005) put it, "across all integrated models, housing/floor space supply models are probably the least well developed of any component of the entire modeling system."

The purpose of this dissertation is to fill these two research gaps partially by developing new housing supply and location choice models for the Portland, Oregon metropolitan area. First, a series of dynamic developer-based location choice models are developed to examine the impacts of Smart Growth policies on home developers' location choice in

the metropolitan area and to detect the preference heterogeneity in developers' location choice. This is believed to be the first study that assesses the impacts of Smart Growth Policies within a developer-based location choice model. As such, it provides a microeconomic treatment of the perceived utility/disutility of various attributes of alternative project locations from the perspective of developers.

Secondly, this study proposes a comprehensive new housing supply and developer-based location choice model, which is rarely seen in existing integrated land use and transportation models. As such it provide a scenario planning tool for the long-term land use and transportation planning in the Portland metropolitan area.

1.2 Literature Review

1.2.1 Assessing the Effectiveness of Smart Growth Policies

There are two main themes in the literature that examines the effectiveness of Smart Growth policies. One is to assess whether Smart Growth policies have effectively prevented urban sprawl and shaped land use in the patterns that are promoted by Smart Growth advocates. Urban growth containment policies have been the focal point of these studies. The second theme is to study whether land use patterns that follow Smart Growth principles have effectively reduced car travel and shifted people's travel modes from car to public transit and non-motorized modes, thus reducing energy use and CO_2 emissions in the transportation sector. This study falls in the first theme.

A review of relevant literature finds three major approaches used by empirical studies to measure the impacts of Smart Growth policies on land use patterns. The first approach is

to use aggregate data at metropolitan level to examine the marginal effects of state growth management policies on land use while controlling other relevant variables. Multivariate regression is a commonly used econometric tool used in those studies. Explanatory variables are the variables representing the presence of growth management policies and other controlling variables measuring the socio-economic status, land use, public policy, and the availability of certain infrastructures in each metropolitan area at one or multiple time points. The dependent variables vary with the research questions raised by each study. Usually, they are the variables that can be used to measure land use characteristics such as density.

For example, using a panel data composed of 19 time periods (from 1980 to 1998) on each of 293 metropolitan statistical areas (MSAs), Dawkins and Nelson (2003) examined the effectiveness of state growth management programs in steering new residential development from the suburbs toward the central city by regressing the share of new residential building permits attracted by the primary central city on state growth management variables and other controlled variables. The conclusion was that primary central cities in states with state growth management programs had attracted marginally higher shares of residential development than primary central cities in other states. With a similar approach, Nelson and Peterman (2000) evaluated the impacts of growth management on economic performance, measured by the change in the relative share of total personal income in 182 moderate-size MSAs from 1972 to 1992, and found a positive association between the presence of growth management efforts and economic performance. Carruthers (2002) examined the impact of state growth management growth

3

on five dimensions of urban development (density, the spatial extent of urbanized land area, property value, public expenditure on infrastructure and population change) with a dataset composed of 283 metropolitan counties at four time points. A basic finding by this study is that state-based planning programs worked better when strong consistency and enforcement mechanisms were required by the states. To examine the impact of state growth management policies on non-central retail sprawl, Wassmer (2002) studied 54 western U.S. cities at three time points (1977, 1987, and 1997) by regressing retail sales in non-central places on urban containment policy variables and other controlling variables, and found that over time the most restrictive form of an urban growth boundary reduces the de-centralization of retail sales in metropolitan areas. Another regression based analysis by Wassmer (2006) used log square miles in 452 U.S. urbanized areas designated by the Census as a dependent variable to evaluate the effects of different types of urban containment policies implemented in different regions. Wassmer found that, while those different forms of local urban containment and statewide growth management policies are more or less achieving their goals, their outcomes are different. Only statewide growth management programs with vertically or horizontally integrated components were effective at reducing the sizes of urbanized areas. A study by Pendall (1999) found non-significant effects of urban growth boundaries on urban land use density. Nelson et al. (2004) assesses the effect of urban containment on the percent change in racial segregation change among U.S. metropolitan areas during 1990s and found that, while metropolitan areas with strong urban containment policies showed a higher percent decline in Caucasian/African American residential segregation, the effect

4

of urban containment policies are not statistically significant on segregation between Caucasian and other races.

The second approach used by empirical studies to evaluate the effectiveness of Smart Growth policies on land use is descriptive analysis in an area where Smart Growth policies have been implemented. The basic idea behind this approach is to measure land use characteristics such as density and mixed use at sub-area levels to determine whether land use patterns have changed since the implementation of Smart Growth policies in the area, or whether land use patterns in sub-areas are different when their relative locations to urban growth boundaries (UGBs) are different. Song and Knaap (2004) measured land use characteristics (density, mixed use, road connectivity) at the neighborhood level in Washington County, Oregon at several different time points, and found that neighborhoods in Washington County are becoming better internally connected, more pedestrian accessible, and denser but remain relatively homogeneous in land use. In a similar spirit, Song (2005) compared the changes of urban land use patterns in three areas (Portland, Oregon; Orange County, Florida; and Montgomery County, Maryland) over time and found that all the three areas had similar development patterns: the growth of neighborhoods were denser and internal road connectivity better, but remained poor in terms of external connectivity, accessibility, and mixed land use. Nelson and Moor (1993) divided the Portland region into several parts based on their relative locations to UGBs and compared the percentages of housing units located in each of them, concluding that most regional development had been directed to the UGB and away from resource lands. Weitz and Moore (1998) selected three small size metropolitan areas in Oregon (Florence,

McMinnville, and Medford), divided their areas within UGB into six subareas, and compared the changes of percents of housing units located in them over time, finding that more recent development inside UGBs tended to be more contiguous rather than dispersed, compared with older development. A study by Gordon and Vipond (2005) compared the densities in two communities in Markham, Ontario designed under two different design spirits: new urbanism and conventional suburban planning, concluding that New Urbanism can accommodate high gross densities and consume less land.

Compared to the empirical studies at aggregate level, descriptive analyses at sub-area level provide a spatial perspective at disaggregate level; however, the causal link between Smart Growth policies implemented in the region and the changes of land use pattern occurring in the region is still weak. Due to its descriptive nature, many relevant factors are not controlled and the impacts of those policies are not separable. To solve these problems, more complicated modeling approaches are needed. Compared to studies taking the first two approaches, there are fewer studies using modeling tools at disaggregate spatial level to examine the effectiveness of Smart Growth policies. An earlier example is a study by Jun (2004), which uses housing units at census-block group level as a linear function of the housing market and location attributes to evaluate the effectiveness of UGBs in the Portland metropolitan area, and found that Portland's UGBs had no statistically significant impact on new housing construction location. By estimating binary logit models of land conversion for selected counties in Maryland for both pre-smart-growth and post-smart-growth period, Shen and Zhang (2007) tested the effectiveness of Smart Growth policies in Maryland. It was reported that Maryland's

Smart Growth initiatives had been generally successful, but the effectiveness of those policies varied by county. Cho et al. (2006) used a heteroscedastic probit model to estimate the effects of an UGB on land development decisions in Knoxville and Knox County in Tennessee, finding that the UGB in the regions have successfully encouraged urban revitalization within the boundary and discouraged urban sprawl outside the UGB.

1.2.2 Housing Supply Models in Existing Land Use Models

The basic purpose of a housing supply model in an integrated land use model is to answer two questions: how many housing units will be produced in the forecast period and where they are. New housing supply models in existing integrated land use-transport models can be generally classified by one of two perspectives, "bottom-up" or "top-down."

In the bottom up perspective, the total housing supply in a metropolitan area is the collective outcome of individual land development decisions made for each land parcel, or other spatial unit, in the region. The assumption is that each land parcel may represent an independent housing supplier and that housing supply is determined at this disaggregate level. Prominent examples of the bottom-up approach are the MUSSA model developed by Martínez (1996, 2007) and UrbanSim by Waddell (2002). The theoretical foundations of these models are rooted in the classic bid-choice and hedonic approaches to valuation of land.

In the operative MUSSA model (Martínez, 1996), housing suppliers are the owners of each dwelling-zone. In a bid-rent framework, each owner decides the combination of dwelling-zone options to supply and chooses the highest bidder among many consumers to maximize their profit. At the same time, consumers bid for the dwelling units (or land lot)

that will maximize their consumer surplus. Finally, supply-demand equilibrium is achieved and equilibrium prices are the results of auctions and market clearing. Martínez (2007) recently proposed a new conceptual framework for the MUSSA model. This new conceptual real estate market model is still a static equilibrium model, but the idiosyncratic nature of suppliers' behavior was incorporated into the model, making the suppliers' market behavior more explicit. In UrbanSim (Waddel et al., 2003), developed gridcells are used as the basic unit of built space. A multinomial logit model was used to estimate the probability of a grid cell experiencing a development event, and if it does experience such an event, calculating the probability of different types of events. Based on the probabilities estimated for a grid cell, commitment of development is simulated using a Monte Carlo sampling process.

Land use models that take a "top down" perspective tend to be more aggregate in nature. One approach is rooted in industrial production theory, based on a spatial "input-output" model of a regional economy, and results in the consumption of land by residents. A recent example of this tradition is the PECAS model developed by Hunt and Abraham (2003). Another approach uses simultaneous equations to balancing demand and supply for housing, a contemporary example being the MetroScope model developed by Conder (2002) and applied to numerous studies in the Portland metropolitan area. A third top-down approach is the ILUTE model designed by Miller et al. (2006) in which housing supply is based on the thesis work by Haider (2003). The ILUTE model divides the housing supply process into two major parts: compute new housing stock at an aggregate level, using time-series analysis, and distribute new housing stock to zones at a

disaggregate level. Compared with the "bottom-up" perspective model, the "top-down" perspective models are more sensitive to macro-level economic factors and embed the notion of market clearing prices to balance supply and demand.

1.3 Contributions

This dissertation makes two major contributions to the field: the evaluation of Smart Growth policies from the developer's perspective, and a new housing supply model in which the location choice is explicitly developer-based.

In terms of evaluating the effectiveness of Smart Growth policies, this study contributes in several respects. First, as the literature review shows, there have been no studies evaluating the effectiveness of Smart Growth policies from an individual agent's perspective. To my knowledge, this is the first study that assesses the effectiveness of Smart Growth policies using developer-based location choice models. As such it provides a microeconomic treatment of the perceived utility/disutility of various attributes of alternative project locations from the perspective of developers. Secondly, whereas other research has focused on just one Smart Growth policy, principally the urban growth boundary (UGB), this study tests for the impacts of UGBs in conjunction with several other important policies in one comprehensive model, allowing us to isolate individual policy effects and control for exogenous factors. Thirdly, except for the study by Jun (2004), other research focusing on Portland as their case study have ignored the portion of the market in Clark County, Washington. In this study, I explicitly include project location alternatives in Clark County, allowing us to assess whether home developers respond differently to different state land use policy systems within the same

metropolitan market. Lastly, previous studies examined Portland's Smart Growth policies using data from before 2000. In the past decade, Metro has systematically collected land use and building permit data at a small spatial scale, providing us with reliable, detailed data from which to identify trends in the region since 2000, an important consideration given the promulgation of Washington State's Growth Management Act (GMA) in the mid-1990s and the recentness of it impacts.

A basic trend in the integrated land use modeling field is to develop agent-based housing supply model. However, due to the shortage of data and the complexity in housing supply market, to my knowledge, none of the housing supply models in existing integrated land use models are actually developer-based. Recent update of MUSSA by Martínez (2007) considers the idiosyncratic nature of housing suppliers, but the housing supplies in his model are still "dwelling-zone" owners, not developers in real world. The housing supply model in ILUTE (Miller et al., 2006) uses developers' housing projects as observation units to estimate their location choice models. However, in their forecast model, housing projects are atomized into individual housing units and each unit is assumed to be homogenous and independent. Thus, it is still not an agent-based model. In this dissertation, developer-based location choice models are developed, developers' taste heterogeneity is investigated, and the forecast capabilities of the models are tested and compared to non-agent based forecast models. This is believed to be the first housing supply model in which individual home developers' behavior and idiosyncratic nature are estimated and calibrated in an integrated land use model.

1.4 Structure

The main body of this dissertation is divided into two parts. The first part consists of three chapters that examine the impacts of Portland's Smart Growth policies on home developers' location choice behavior. Chapter 2 gives descriptive analyses of the study area, the region's home developers, their housing projects, spatial distribution of new housing development, and location attributes selected to explain developers' location choices. In Chapter 3, simple Multinomial logit (MNL) models are developed to assess whether Smart Growth Policies have different impacts on home developers' location choice in Oregon, compared with across the border in the State of Washington. Chapter 4 explores the preference heterogeneity in home developers' location choice by developing mixed multinomial logit (MMNL) models. The second part develops a comprehensive new housing supply and location choice model.

As indicated by the flowchart in Figure 1, the model consists of three sub-models: an aggregate housing supply forecast model, a housing project synthesis model, and a housing location choice model. Chapter 5 develops and compares three time series models that can be used to predict the total amount of new housing supply in the forecast year in the region. Chapter 6 synthesizes housing projects in the forecast year, which can be used as forecasting units in the location choice model. Chapter 7 proposes three different new housing location choice models and compares their forecast capabilities. The dissertation ends with a concluding Chapter 8 that summarizes its main accomplishments and limitations.

```
                                                    ┌─────────────────────────┐
                                                    │  Location choice model  │
                                                    │ (MNL model without market│
                                                    │      segmentation)       │
                                                    └─────────────────────────┘

┌─────────────────────────┐                         ┌─────────────────────────┐
│  Aggregate new housing  │                         │  Location choice model  │
│  supply forecast model  │                         │ (MNL model with exogenous│
│   (time series model)   │                         │   market segmentation)   │
└─────────────────────────┘                         └─────────────────────────┘
                  ┌─────────────────┐
                  │ Housing project │               ┌─────────────────────────┐
                  │ synthesis model │               │  Location choice model  │
                  └─────────────────┘               │   (LCM with endogenous  │
                                                    │   market segmentation)   │
                                                    └─────────────────────────┘
```
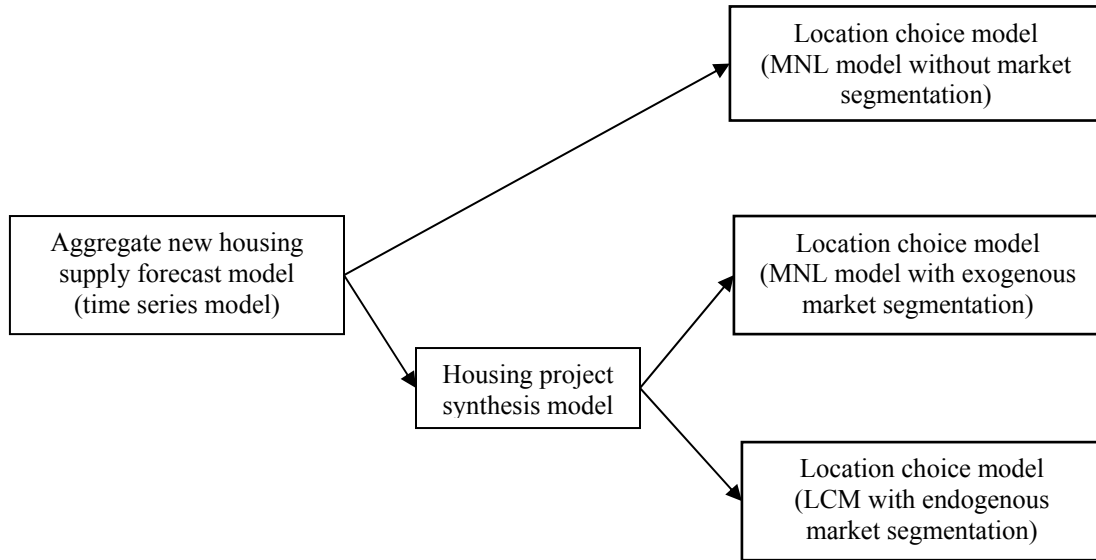
Figure 1 Comprehensive new housing supply and location choice model

PART I

IMPACTS OF SMART GROWTH POLICIES ON HOME DEVELOPERS

CHAPTER 2 DESCRIPTIVE ANALYSES

This chapter gives descriptive analyses on the study area, home developers and their housing developments in the region from 2000 to 2007, and explanatory variables selected to model home developers' location choice.

2.1 Study Area

Straddling the border between Oregon and Washington, the Portland metropolitan area is a bi-state metropolitan area composed of four counties: Multnomah, Washington, and Clackamas in Oregon, and Clark County in Washington (Figure 2). The total area of the region is about 3728.7 square miles. According to the Census estimation, the population in the metropolitan area was about 2.08 million in 2009. There are 41 cities in the region, and the city of Portland is the biggest one with a population of 566,143 in 2009. For transportation planning purposes, the four-county region is divided into 1998 Traffic Analysis Zones (TAZs), which I use as spatial units for the location choice models developed in this study. As indicated by Figure 2, the sizes of TAZs vary, but 86.4% are smaller than one square mile.

While the four counties in the region comprise the same housing market, the market functions under two different state land use systems. The three counties in Oregon are served by Metro, the only elected regional government in the United States. Metro includes Clark County, Washington (Vancouver area) on its 2040 Growth Concept maps (Metro, 1995), but has no regulatory powers there. The planning and zoning in Clark County is regulated by Washington state laws.
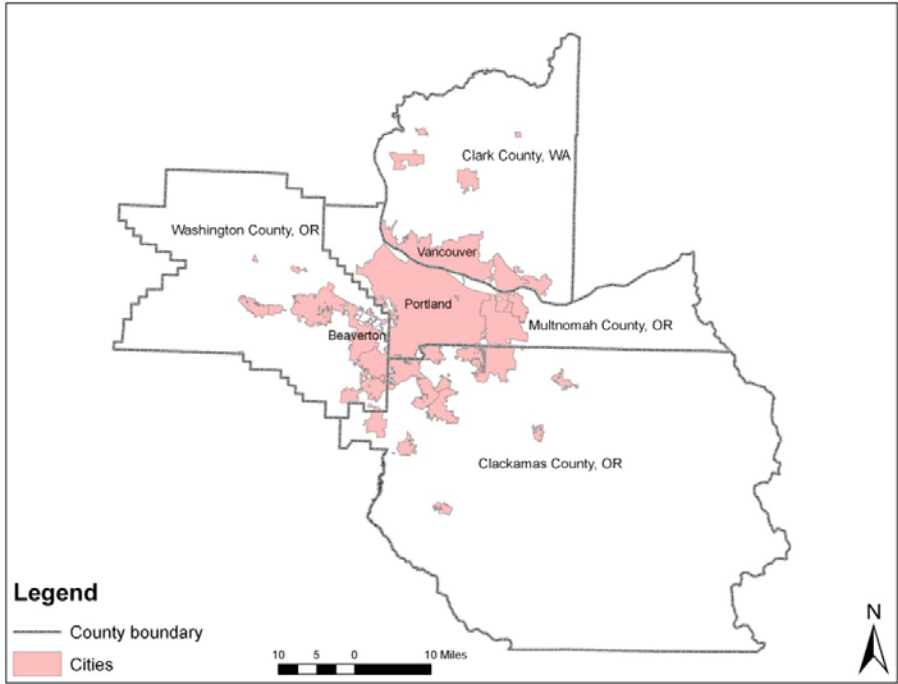
Figure 2 Counties and cities in the Portland metropolitan area

Data source: Metro, RLIS (2007)
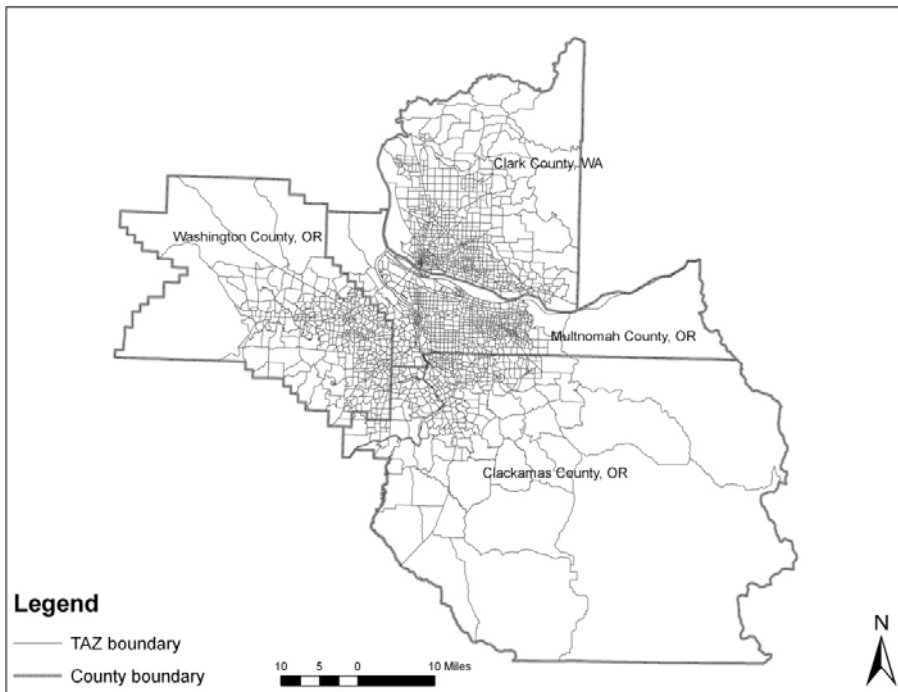
Figure 3 TAZs in the Portland metropolitan area

Data source: Metro, RLIS (2007)

2.2 Home Developers in the Portland Metropolitan Area

New housing development data in the study area used by this study was extracted from the geo-coded building permit data provided by Metro (2000-2007a), which includes all the housing permits issued in the region from 2000 to 2007. With this data, I am able to know the names of home developers, their contract types, specializations, development locations, and the number of housing units they developed.

In this study, a single family home (SFH)/multifamily home (MFH) project is defined as a cluster of new SFH/MFH units permitted for construction by a developer in one TAZ in a single year. The size of a project is defined as the number of housing units included in it. Since some developers have development activities in multi-years, the size of a developer is measured by dividing the total number of housing units they developed by the number of their active years. Table 1 shows the description of SFH and MFH developers and their projects in the region from 2000 to 2007.

2.2.1 SFH Developers

As shown by Table 1, from 2000 to 2007, there were 7,123 SFH developers in the region that collectively developed 21,237 SFH projects, producing 65,377 new SFH units. The average SFH developer size was 2.58 units and the average project size was 3.08 units.

To illustrate the size distribution of SFH developers, I categorized SFH developers into three groups based on their sizes: small developers who produced only one unit per year, medium size developers who produced more than one but less or equal to five units per year, and large size developers who produced more than five units per year. As Figure 4

indicates, small size SFH developers accounted for 67 percent of all SFH developers, the number of projects developed by them accounted for 27 percent of all SFH projects, but they contributed only 9 percent of all new housing units on the SFH market. Medium size SFH developers accounted for 27 percent of all SFH developers, their projects accounted for 40 percent of all SFH projects, and new housing units developed by them accounted for 20 percent of all new SFH units. Large size developers accounted for only 7 percent of all SFH developers, they built 33 percent of all SFH projects, but they produced 71 percent of all new SFH units on the market. Thus, the SFH market in the Portland region was dominated by large size developers, though their number is very few compared to the total number of SFH developers in the region.

The data also provides developers' contract information. Based on their contract types, I categorize developers in the region into two types: contractor-owners and contractors. A contractor-owner is a general contractor who owns the land with the intent to build and sell, while a contractor is a general contractor who is hired by an individual or company to build. The main difference between the two lies in that contractor-owners are real location decision makers, but contractors are not. Within 7,123 SFH developers, 35.1 percent of them were contractor-owners and 64.9 percent were pure contractors. The data shows that the mean size of SFH developers who were contractor-owners was much larger than SFH developers who were contractors (4.88 units vs 1.34 units). Contractor-owner type developers developed 95 percent of SFH projects whose sizes were larger than 5 units in those eight years.

Variable name	SFH	MFH
Total number of developers	7123	934
Number of projects	21237	1530
Average developer size (units)	2.58	23.6
Average project size (units)	3.08	24.05
Percent of devlopers who are contractor-owners	35.11	51.7
Percent of devlopers who are specialized on SFH	94.33	0.00
Percent of devlopers who are specialized on MFH	0.00	57.49

Table 1 Description of SFH and MFH developers and projects (2000-2007)

Data source: author's computation based on the building permit data from Metro (2000-2007a)
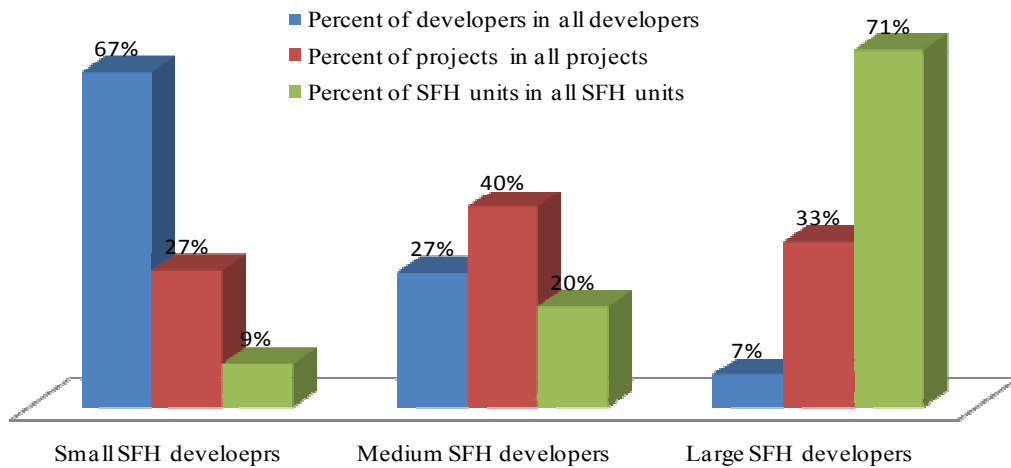
Figure 4 Breakdown of SFH developers, their projects and new housing units

Data source: author's computation based on the building permit data from Metro (2000-2007a)
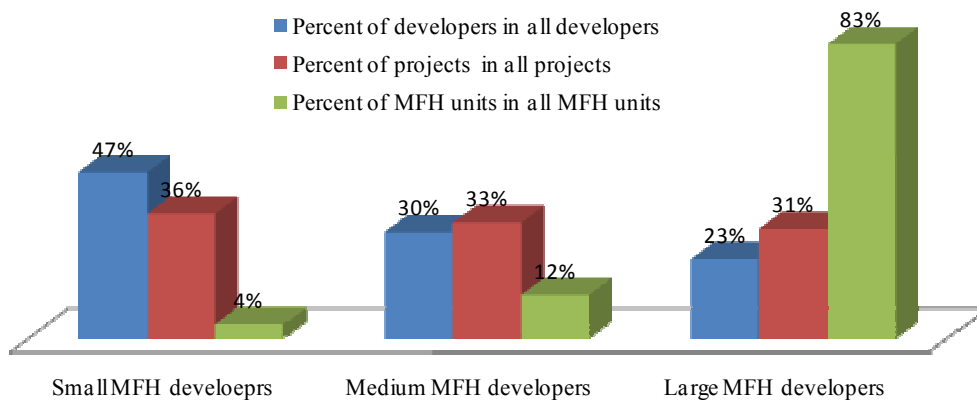
Figure 5 Breakdown of MFH developers, their projects and new housing units

Data source: author's computation based on the building permit data from Metro (2000-2007a)

Specialization is another developer attribute that is available from the data. The data shows that 94.3 percent of SFH developers concentrated on SFH projects and only 5.7 percent of them also developed MFH projects. The mean size of projects developed by SFH developers who were also active on the MFH market is larger than average SFH project size (5.98 units vs 3.08 units), indicating that SFH developers who developed both SFH and MFH tend to be large-scale developers. Not all SFH developers were active in all eight years: 67.4 percent were active only in one year, 32.6 percent were active in two or more years, and only 1.9 percent were active in all eight years.

2.2.2 MFH Developers

As indicated by Table 1, there were 934 MFH developers from 2000 to 2007 in the Portland metropolitan area. They collectively developed 1,530 MFH projects with 36,803 new MFH units. The average MFH developer size was 23.6 units and the average size of MFH projects was 24.1 units.

Similar to SFH developers, I categorized MFH developers into three groups: small MFH developers who produced only two to five units per year, medium size MFH developers who produced more than five but less or equal to twenty units per year, and large size MFH developers who produced more than twenty units per year. As indicated by Figure 5, small size MFH developers accounted for 47 percent of all MFH developers; the number of their projects accounted for 36 percent of all MFH projects; and they contributed 4 percent of total new MFH units in those eight years. In contrast, the proportion of large size MFH developers in all MFH developers was only 23 percent, but they produced 87 percent of all new MFH units on the market. Similar to the SFH market,

a small amount of large size developers dominated the MFH market in the Portland metropolitan area.

The data shows that 51.7 percent of MFH developers were contractor-owners, and 42.5% of them were also active on the SFH market in those eight years. The average size of the projects developed by MFH developers who were contractor-owners is very close to the average MFH project size (23.3 vs 24.1 units). The mean size of MFH projects by MFH developers who were active in both SFH and MFH markets was larger than the average MFH project size (30.1 vs 24.1 units).

Similar to SFH developers, not all MFH developer were active in consecutive years. About 21.6 percent were active in two or more years, and less than 1 percent were active in all eight years.

2.3 Spatial Distribution of New Housing Development (2000-2007)

To show the spatial distribution of new housing development in the region, new SFH and MFH units produced between 2000 and 2007 were aggregated to 1,998 TAZs. In Figures 6 and 7, each TAZ is symbolized by a circle whose size is proportional to the number of new SFH/MFH units in that TAZ.

The data shows that from 2000 to 2007, 74 percent of TAZs (1479 out of 1998) in the region attracted some SHF development. As shown in Figure 6, most TAZs where the number of new SFH units exceeded 150 in the eight years were in urban peripheral areas. Among the four counties in the region, Washington County, Oregon attracted the most
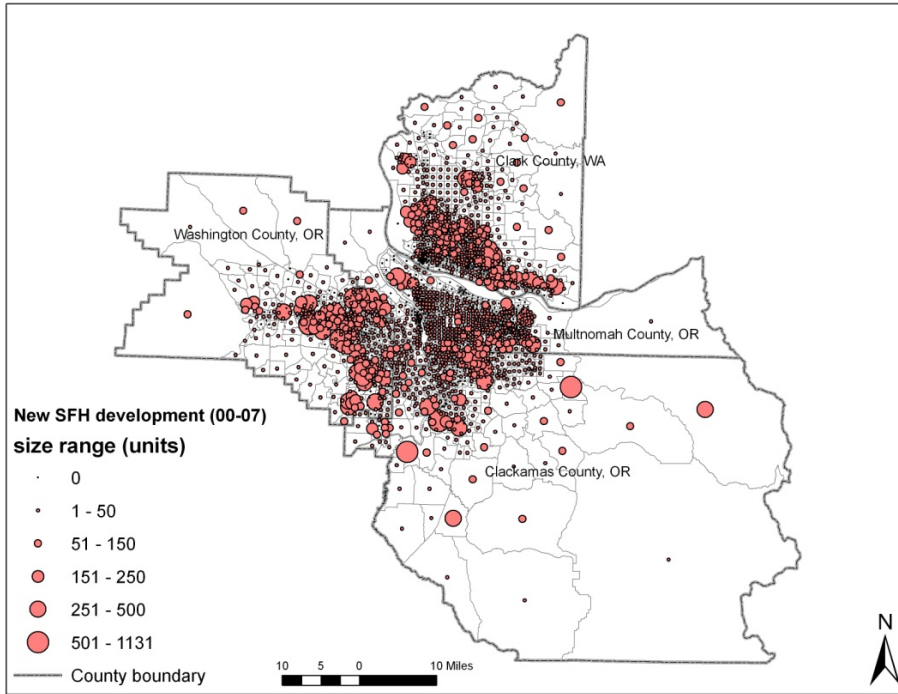
Figure 6 New SFH development in the Portland Metropolitan area (2000-2007)

Data source: author's computation based on the building permit data from Metro (2000-2007a)
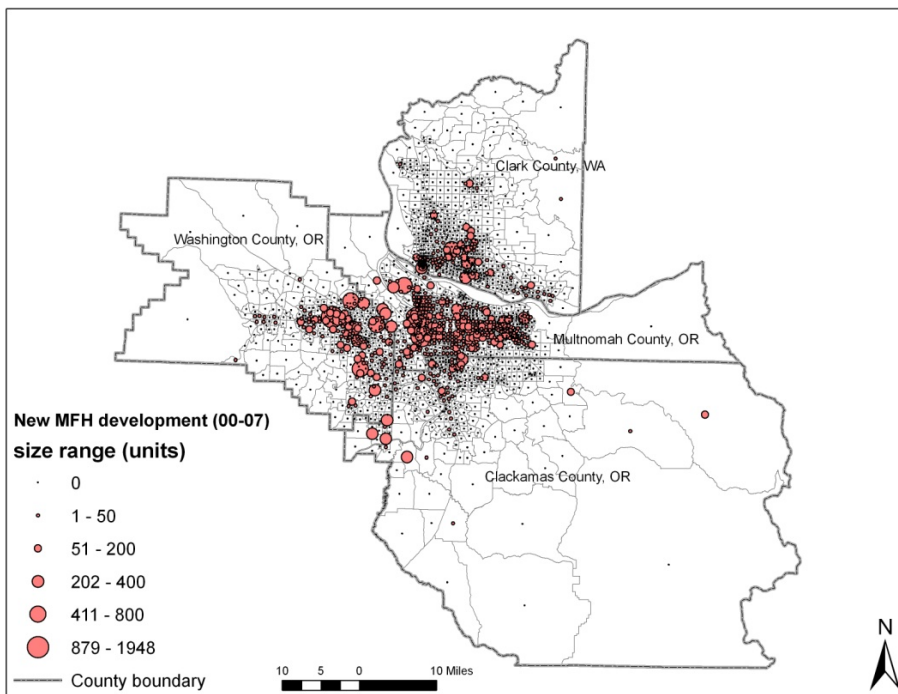
Figure 7 New MFH development in the Portland Metropolitan area (2000-2007)

Data source: author's computation based on the building permit data from Metro (2000-2007a)

new SFH developments; Clark County, Washington was second; and Multnomah County had the fewest. Most new SFH developments in Washington County, Oregon were concentrated in the peripheries of cities of Beaverton, Hillsboro, Tigard, Sherwood, and Northwest Portland. In Clark County, Washington, except for the new SFH developments in cities of Battle Ground and Ridgefield, most new SFH developments occurred in the far suburban areas of cities of Vancouver and Washougal. In Clackamas County, Oregon, the cities of Happy Valley and Oregon City attracted most new SFH developments during the study period. Multnomah County, Oregon, home to the city of Portland, had far fewer SFH developments than the other three counties. Most new SFH developments in Multnomah County were located in the peripheries of cities of Gresham, Troutdale, Fairview, and East Portland. There were also some large SFH projects on the south side of the Columbia Boulevard in West Portland.

Compared to SFH development, MFH development in the region from 2000 to 2007 was more concentrated in the built-up areas within city boundaries. As shown in Figure 7, Multnomah County, Oregon attracted the most new MFH developments and most of them were in the City of Portland. Washington County, Oregon was second to it, and most MFH developments in it were in the Cities of Hillsboro and Beaverton. In Clark County, Washington, MFH developments were mainly located in the City of Vancouver. Clackamas County, Oregon attracted the least MFH development during the study period.

2.4 Data and Explanatory Variables

My study utilized two sets of data provided by Metro and a dataset from the GIS department of Clark County, Washington to create explanatory variables for location

Variable name	Variable description	2000	2007
UGB dummy variables:			
UGB_IN	Within UGB (yes=1, no=0)	0.75	0.75
UGB_ON	Crossed UGB (yes=1, no=0)	0.05	0.03
UGB_OUT	Out of UGB (yes=1, no=0)	0.20	0.16
UGB_EXP	In or partially in UGB expansion areas (yes=1, no=0)	0.00	0.06
SFH net density:			
SFH_DEN	SFH net density (units/acre)	2.91	3.79
SFH net density dummy variables:			
SFHDEN_N	No SFH in the TAZ (yes=1, no=0)	0.13	0.12
SFHDEN_L	Low SFH density (yes=1, no=0)	0.22	0.20
SFHDEN_M	Medium SFH density (yes=1, no=0)	0.60	0.59
SFHDEN_H	High SFH density (yes=1, no=0)	0.05	0.08
MFH net density:			
MFH_DEN	MFH net density (units/acre)	12.17	12.39
MFH net density dummy variables:			
MFHDEN_N	No MFH in the TAZ (yes=1, no=0)	0.43	0.42
MFHDEN_L	Low MFH density (yes=1, no=0)	0.14	0.13
MFHDEN_M	Medium MFH density (yes=1, no=0)	0.25	0.26
MFHDEN_H	High MFH density (yes=1, no=0)	0.18	0.18
Housing diversity ratio:			
H_DIV	Housing diversity	9.57	10.68
Housing diversity dummy:			
H_MIX	MFH and SFH mixed (yes=1, no=0)	0.30	0.31
H_SFH	Dominated by SFH (yes=1, no=0)	0.55	0.54
H_MFH	Dominated by MFH (yes=1, no=0)	0.15	0.15
Mixed use ratio:			
MIX_U	Mixed use	1.23	1.25
Mixed use dummy variables:			
MIX_USE	Mixed use (yes=1, no=0)	0.32	0.22
MIX_RES	Dominated by residential (yes=1, no=0)	0.33	0.40
MIX_NON	Dominated by non-residential (yes=1, no=0)	0.36	0.38
Transportation accessibilities:			
AUTO_RET	Auto for retail employment	57.01	
AUTO_NRET	Auto for non-retail employment	272.09	
TRS_RET	Transit for retail employment	20.27	
TRS_NRET	Transit for non-retail employment	92.96	
Existing infrastructure:			
RD_DEN	Road density in TAZ (ft/acre)	112.67	114.01
Other location and site attributes controlled:			
LAND_SFH	Buildable land for SFH (acre)	9.00	7.49
LAND_MFH	Buildable land for MFH (acre)	2.72	3.06
LAND_VL	Average land value in TAZ ($/ft^2, in 2000 dollars)	5.45	7.41
Socio-economic characteristics:			
HINC	Median household income ($1000)	50.39	
AGE	Percent of household whose head is aged between 25-54	64.84	
Clark	The TAZ is in Clark County, Washington (yes=1, no=0)	0.23	0.23
N	Number of TAZs in the region	1998	1998

Table 2 Mean values of the TAZ attributes in 2001 and 2007

choice models. Metro's Regional Land Information System (RLIS) and a dataset from the GIS department of Clark County, Washington provided geographically-detailed land use, existing housing stock, zoning, and transportation information for each year of the study. From Metro's travel demand modeling system, I derived 2005 TAZ-to-TAZ morning peak two-hour period travel time matrices by auto and transit modes and used these to calculate accessibility measures, namely a zone's accessibility to employment (discussed below). The Census 2000 data at the census tract level (U.S. Census Bureau, 2000) and the County Business Patterns data at the ZIP code level (U.S. Census Bureau, 2000-2007) were also used to provide local socioeconomic variables.

The mean values of the TAZ attributes used to explain developers' location choice are presented in Table 2 for the years 2000 and 2007. For most of the variables, we can observe a change in their mean values over the eight-year span. Note that socioeconomic characteristics were available only for 2000 and travel accessibility variables were only available for 2005; therefore, the values for these variables were used for each year. Individual variable descriptions are provided below.

2.4.1 Variables Relevant to Smart Growth Policies

The UGB in the Portland metropolitan area represents a set of planning policies aiming at preserving agricultural and forest lands, limiting the outward extension of the public infrastructure and facilities, increasing residential density and diversity, encouraging mixed use, and reducing auto travel. These Smart Growth policies might be made explicitly or implicitly to support infill development within the UGB and to prevent urban sprawl outside of it. The data available for this study allows us to measure the

effects of some of those polices, such as those related to residential density and diversity, mixed land use, transportation accessibility, and utilization of existing infrastructure. An important consideration of this inquiry is that being inside or outside UGB can also capture other relevant factors that are concomitant to the UGB, such as the availability of buildable land and land price, two effects that were controlled by including them in our model specifications. What remains after controlling for the measurement of these direct policy outcomes and the concomitant factors that I can measure, is the UGB effect which I hope to capture through parameter estimates. This UGB effect represent policies and concomitant land use conditions and costs, such as development impact fees and taxes, and perhaps consumer demand that could not be measured directly and hence control. Thus, UGB dummy variables were specified to indicate whether a particular location alternative is within the UGB, within a designated expansion area, straddling the UGB border, or completely outside the UGB. If the area of the intersection between a TAZ and the UGB expansion area is larger than 30 acres, the TAZ was defined to be partially in the UGB expansion area. Between 2000 and 2007, the UGB for the three Oregon Counties was expanded three times, in 2002, 2004 and 2005, and the UGB in Clark County, Washington was expanded once, in 2004. Since it takes time for developers to respond to UGB expansion, each expansion is counted after the year it was designated. Figure 8 shows the UGBs in 2000 and their expansions between 2000 and 2007. Table 2 shows the proportion of TAZs in each of these categories and reflects the expansion from 2000 to 2007.
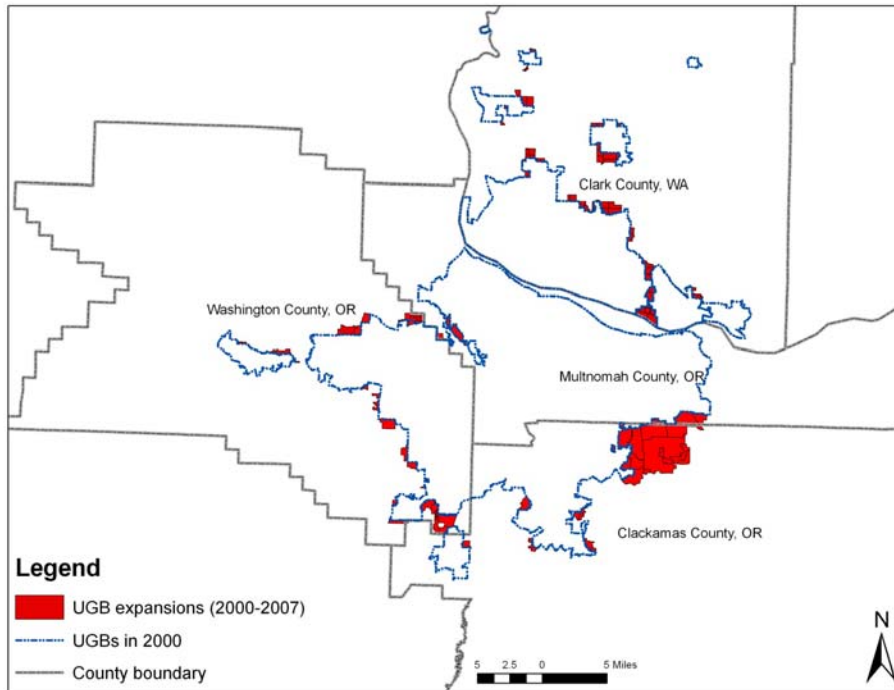
Figure 8 The UGBs in 2000 and their expansions between 2000 and 2007

Data source: Metro, RLIS (2007); GIS department at Clark County, Washington (2007)

Higher residential densities are generally thought to conserve land and provide critical mass for efficient transit service. In this study, separated TAZ-based measures of density for SFH and MFH were created. Density was calculated as the total number of units divided by the total land area of the properties they actually occupy. Based on the definition used in Portland-area zoning system, dummy variables were created to indicate the residential density in each TAZ as follows: (1) TAZs without any SFHs/MFHs; (2) Low-density (SFH less than 1 unit per acres, MFH less than 10 units per acre); (3) Medium-density (SFH 1 to 8 units per acre, MFH 10 to 20 units per acre); and (4) High-density (SFH greater than 8 units per acre, MFH greater than 20 units per acre). Table 2

shows that the proportions of the TAZs in each density group shifted slightly towards the higher density categories of the study period.

A mix of housing types is often advocated as means of providing affordable living options to various socio-economic strata and more equitable access to neighborhood resources, e.g., schools. To measure the housing diversity in each TAZ, the ratios of MFH units to SFH units in each TAZ were calculated and the TAZs in the region were categorized into three groups evenly based on that ratio. I defined a TAZ as being "dominated by SFH" if it had a ratio less than or equal to 0.1; "MFH and SFH mixed" if it had a ratio greater than 0.1 and less than 2.0, and "dominated by MFH" if it had a ratio larger than 2.0. Table 2 shows that the proportions of TAZ falling into these categories did not change significantly in the region between 2000 and 2007.

Smart Growth advocates mixing residential and commercial land uses to provide work and non-work activity opportunities within shorter travel distances, ideally fostering a sense of community and reducing auto travel. As is standard practice in transportation planning, I used employment as a proxy for the amount of commercial activity opportunities in a TAZ. I then created a "mixed use" index from the ratio of number of employees to the number of housing units in each TAZ. Yearly employee data came from the County Business Patterns data. Since these were available at ZIP code level, which is larger than the TAZ, the mixed use index for each TAZ was set to that of the ZIP code to which it belongs. In addition, the employee data from the ZIP code business pattern data does not distinguish by industry sector, so mixed use was defined broadly as a mixture of residential and non-residential uses. Based on this index, the TAZs are categorized into

three groups: (1) "dominated by residential" with an index equal to or less than 0.5;

"mixed use" with an index between 0.5 and 1.0; and TAZs "dominated by non-

residential" land uses with an index larger than 1.0. As Table 2 indicates, from 2000 to

2007, the percent of TAZs dominated by residence in the region decreased substantially

while the level of mixed use areas increased in the Portland metropolitan area.

Transportation accessibility was based on the modeled morning two-hour peak travel

times for pairs of TAZs, utilizing a static estimate of 2005 congested network travel times.

The Metro travel demand model also provides 2005 estimates of employment by TAZ

and by industry sector. These data were used to calculate four accessibility variables,

differentiated by auto and transit modes and by retail and non-retail employment. I

adapted the negative exponential travel impedance formula from Meyer and Miller (2001,

p 336):

$$Accessibility_{i,emp} = \left(\sum_{j=1}^{J} \exp(-\beta * ttime_{ij}) * employment_j\right) \qquad (1)$$

in which $Accessibility_{i,emp}$ measures the employment accessibility for TAZ i, β is

parameter indicating the sensitivity of trip making to travel time, $ttime_{ij}$ is the travel

time from TAZ i to TAZ j, and $employment_j$ is the number of jobs in TAZ j. In this

study, I used the reciprocal of the mean travel time for journey to work in the region to

represent the mean impedance effects of travel time. According to the American

Community Survey, in 2005 the mean commuter travel time in the Portland region was

24.4 minutes (U.S. Census Bureau, 2005), resulting in a β value of 0.041. We scaled the

calculations by 1000 when used in the model in order to produce coefficient values

comparable in magnitude to other variables, which has no effect on model fit or the significance of parameter estimates. As Table 2 shows, auto provides much better travel-time weighted access to retail and non-retail opportunities than transit in the region.

Housing development requires the support of the infrastructure such as road, water, and sewer systems. Haider and Miller (2004) use the length of road to measure the physical development in a zone. In this study, considering that the sizes of TAZs vary, the density of road was used to represent the level of infrastructure in a TAZ. The road density is calculated by dividing the total length of roadways in a TAZ over the area of the TAZ. While calculating the roadway length, only local streets were counted and highways and arterial roads were excluded. The number of road lanes was not considered. Table 2 shows that average regional road density increased slightly over the study period.

2.4.2 Land Supply, Price and Socioeconomic Attributes

In addition, to the potential impact of Smart Growth policies on the attractiveness of alternative locations for residential development, it is essential to control for more generally well-known attributes of development potential. Most new housing developments are still being built on vacant land, making the availability of buildable land a critical factor that developers must consider when choosing locations. For SFH development, the availability of buildable land in each TAZ is the area of vacant land zoned for low-density residential use and suitable for building houses. For MFH developments, the availability of buildable land in each TAZ is the area of the vacant land zoned for medium- and high-density residential and mixed use purposes. Clark County, Washington has a Vacant Buildable Lands Model (Clark County, 2010a) to

29

estimate the buildable land for the county. In Portland, the buildable vacant land for SFH

is defined by the vacant land zoned for low-density residential use, and the slope of the

land is less than 25%. There are two prominent limitations to defining buildable land

based on the aggregated acres of appropriately zoned vacant land. First, land

fragmentation and zoning change are not considered. Secondly, it omits the possibility of

redevelopment that might make more land available in built-up areas. Table 2 shows that

from 2000 to 2007 the amount of buildable land in the region zoned for SFH

development decreased while the amount of buildable land available to MFH increased.

We used the natural log of the buildable land area in each TAZ for model specifications,

which is mathematically necessary for consistency with the notion that, all else being

equal, a percentage change in the utility of more buildable land will result in the same

percentage change in the probability of choosing that alternative.

Land price is another important factor that developers will consider when they select

locations for their housing development. In this study, market prices for land in all TAZs

were not available, so I used as proxy the tax assessment value of residential land. For

each study year, this variable was calculated as the total value of residential land in each

TAZ by the total area of residential land, deflated by the annual consumer price index

(CPI) for the Portland Region (US Bureau of Labor Statistics, 2010). I note that the land

values used are assessed values, not transaction prices, thus they lag behind market rates

usually by a few years, if not more. Since tax assessed value was frozen in 1996 in

Portland, Oregon, and is only allowed to be increased by 3 percent per year, the gap

between the market value and the tax value became wider during the boom years. On one

hand, this makes the tax assessed land value a less accurate proxy for the market land value; on the other hand, this fact helps us to minimize the endogeneity problem in which the marginal utility of land values reflects the values of other attributes included in the model that contribute to land price. Table 2 shows that the average TAZ value for residential land in the Portland metropolitan region increased significantly from 2000 to 2007.

In order to assess the effects of existing area demographics, which might have some bearing on the attractiveness of potential project sites, I included the average household income and the percent of households whose heads were between the ages of 25 and 54 to represent the proportion of young and middle aged households in each census tract, as reported in the 2000 Census data, and applied these to each TAZ for all eight years.

2.4.3 Effects of Previous Choices

A developer's familiarity with an area based on past experience is likely to make it attractive for future work for reasons such as comfort with the local market, relationships with local contractors and government officials, and investments in local infrastructure. In addition, what appears to be multiple projects separated in time might actually be multiple phases of the same project. The permit data shows that 2,324 of the 7,123 developers receiving permits from 2000 to 2007 were active in more than one year (32.6 percent), with just 138 (1.9 percent) active in all eight years. In this study, I hypothesize that home developers tend to locate their new projects in the TAZs they had development activities in previous years or the TAZs that are geographically close to those TAZs. To test this hypothesis, variables were created to indicate whether TAZs were chosen by the

same developer in a previous year or were adjacent to TAZs chosen by the same developer in previous years. I limited consideration to lagged variables within the past two years in order to use as many observations as possible. For example, with the data available from 2000 to 2007 and using two-year lags, I can utilize the dependent variables of location choices from 2002 to 2007 (16,320 SFH projects and 1,138 MFH projects), which proved to be a sufficient sample. Accordingly, I created four dummy variables to represent the lagged effects for developers' location choice in year t. Auto-regressive variables AR(1) and AR(2) indicate that a TAZ in choice set was chosen by the same developer in years $t - 1$ and $t - 2$, respectively. Neighboring zone indicator variables NB(1) and NB(2) indicate that a TAZ in choice set shares boundary with the TAZ chosen by the same developer in years $t - 1$ and $t - 2$, respectively. Inclusion of these lagged effects in the model has two potential benefits. First, it can reflect the spatial attachment between developers and their old locations helping to account for potential spatial autocorrelation. Such attachment can be formed due to developers' familiarity with location attributes, local housing market, administrative process, local officials, etc. Secondly, it can capture some factors that make locations attractive to developers, but which were not measured by the available data, helping to control for repeated observations of the same developer that could lead to biased parameter estimates if otherwise ignored. It is assumed that these benefits are worth and outweigh any potential information lost due to using two fewer years of observations.

CHAPTER 3 ASSESSING THE IMPACTS OF SMART GROWTH POLICES ON HOUSING DEVELOPERS IN A BI-STATE METROPOLITAN AREA

3.1 Introduction

There are extensive empirical studies on the impacts of Smart Growth policies; however, very few of them consider the perspective of individual decision makers and, to our knowledge, none have studied developers as location-aware decision-making agents. Thus, much of the current understanding of the effectuality of Smart Growth policies lacks a micro-level behavioral foundation to capture the dynamics between growth policies and the key land-conversion agents. Literature review finds some studies using agent-based models to test the impacts of land use on individuals' residential location choices, but being household-based represent only the demand side of the land development market. The other side of the market, the provision of housing supply as represented by developers remains under-studied. At first glance, this is puzzling because real estate developers are the primary urban space producer and their perspective on Smart Growth policies is so critical to the policy's success. Hunt *et al*. (2005) summarize three reasons for this: lack of data, complexity of the supply process, and relatively few decision-makers involved in the process, all of which make developing statistically reliable models difficult.

This chapter fills this research gap by studying the impacts of Smart Growth policies on home developers' location choice behavior in the Portland metropolitan area, a bi-state metropolitan area straddling the border between Oregon and Washington. While numerous Smart Growth policies have been implemented in the Portland region, this

chapter focuses on the policies related to five important aspects of Smart Growth: urban growth containment, residential density and housing diversity, mixed use, transportation system accessibility, and utilization of existing infrastructure. There are three research questions this study tries to answer:

1) To what extent do Smart Growth policies influence home developers' location choice in the Portland metropolitan area?

2) Do Smart Growth Policies have different impacts on home developers' location choice in Oregon compared with across the border in the State of Washington?

3) Do SFH and MFH developers exhibit different preferences for their project location choices?

To answer those research questions, a set of developer location choice models were developed for SFH and MFH developers separately. The remainder of this chapter discusses the Portland metropolitan area and its Smart Growth policies; reviews some of the prior research that has examined the effectiveness of these policies; discusses the model design and estimation results; and concludes with a summary of important findings.

3.2 Smart Growth Policies in the Portland Metropolitan Area

While the four counties in the region comprise the same housing market, the market functions under two different state land use systems. The three counties in Oregon are served by Metro, the only elected regional government in the United States. According to its Charter (Metro, 1992), Metro has the power to require changes in local comprehensive plans to make them consistent with its functional plans. Oregon law also requires local comprehensive plans to be compatible with statewide planning goals (Abbott, 2002).

Metro includes Clark County, Washington (Vancouver area) on its 2040 Growth Concept maps, but has no regulatory powers there. The planning and zoning in Clark County is regulated by Washington state laws, which include the Growth Management Act (GMA), enacted in 1990 to guide planning for growth and development. Notably, while designing their growth management system, the Washington legislature deliberately avoided the top-down planning system found in Oregon (Weitz, 1999), and limited the power of the State to alter the content of local plans (Abbott, 1997).

Portland has an international reputation as a model for Smart Growth. Among many Smart Growth policies that have been implemented in Portland region, its UGB garners the most attention. Metro adopted the UGB for its 24 cities and parts of its three counties. Each jurisdiction in the region is required to make their plans and implement actions consistent with the UGB (Seltzer, 2004). The basic purpose of the UGB is to prevent urban sprawl by identifying and separating rural land from urban and "urbanizable" land, thereby providing orderly, efficient transitions from rural to urban use (Metro, 1995). In order to provide sufficient land for urban growth, Metro is required also to maintain a 20-year inventory of developable land within the UGB, which has resulted in several UGB expansions since its adoption in 1979.

In Clark County, Washington, as required by the GMA, urban growth areas (UGAs) were designated and urban growth is encouraged to be inside of them. In this study, the boundary of the UGAs in Clark County, Washington were considered to be its UGB. As a policy tool, the UGB represents a set of planning policies that regulate the land use outside and inside it. A comparison of the statewide planning goals in Oregon and

Washington show that both the states have the land use policies regulating the land use outside UGB by: 1) protecting agricultural and forest lands, environmentally critical areas, and open space; 2) restricting the establishment and extension of public facilities; and 3) controlling residential, commercial, and industrial land development on rural land. While the two states share similar goals in restricting land development outside UGB, the regulations on the land development outside UGB in Clark County, Washington tend to be less restrictive due to several reasons. First, Clark County has much more land outside UGB zoned for low density residential purpose than the three Oregon counties. To compare Clark County to the other three Oregon counties, the area of land zoned for residential, rural, mixed use and other purposes in Clark County are calculated based on the GIS data provided by Clark County, Washington. The area of land zoned for residential, rural, mixed use and other purposes in the three Oregon Counties are calculated based on the Regional Land Information System (RLIS) GIS data provided by Metro.

Calculation results show that based on 2005 zoning categories, Clark County, Washington has 21,132 acres of land outside UGB zoned for residential purposes, compared with a three-county total of just 7,572 acres of land zoned for residential purposes outside of UGBs on the Oregon side. Secondly, the proportion of land outside UGBs zoned as rural land is much higher in Clark County, Washington. Generally, compared to farm and forest lands, the restriction of land development on rural land is less strict, allowing low-density residential development on large land parcels. According to the Clark County 20-Year Comprehensive Growth Management Plan 2004-2024

(Clark County, 2010b), rural lands are the areas that lie outside of the UGB and do not include designated long-term resource lands (agriculture, forest or mineral resources). Based on the 2005 zoning categories, Clark County has about 331,000 acres of land outside the UGB of which 34 percent (about 111,138 acres) was designated as rural lands. In contrast, in the three Oregon counties there were 1711,717 acres of land outside the UGB of which only 7.2 percent (about 123,521) acres were zoned as rural lands. In addition, some rural areas in Clark County with small lot development patterns, natural features as boundaries, and access to arterials are designated as rural centers. In rural centers, both residential and commercial developments, and necessary public facilities and services are allowed to support the needs of rural residents and natural resource industries.

In addition to the land development regulations outside of it, the success of the UGB hinges upon policies that encourage mixed-use, high-density, and transit-oriented development (TOD) on the lands within it. Oregon Statewide Planning Goal 10 on housing requires the provision of a variety of densities and types of residences in each community (Oregon Department of Land Conservation and Development, 2010). The 2040 Growth Concept adopted by Metro in 1995 also calls for substantial amounts of the region's growth to occur in medium- to high-density mixed-use, walk-able urban centers and corridors linked by high-quality transit service (Metro, 1995). In Washington, Statewide Planning Goal 3 on affordable housing set by the GMA encourages the availability of affordable housing, promotes a variety of residential densities and housing types, and encourages preservation of existing housing stock (Washington State

Legislature, 1990). Accordingly, moderately smaller lot sizes, higher average, and provisions for a wide range of housing densities and types are encouraged in Clark County's 20-year Comprehensive Growth Management Plan 2004-2010 (Clark County, Washington, 2010).

While the statewide planning goals in both states encourage compact development and TOD within the UGB, the real effects of those state-level policies are highly dependent on the planning and land use regulations by local governments, especially their zoning. Based on the 2005 zoning categories, Clark County has 72,848 acres of land zoned for residential purpose, of which 11.6 percent (about 8,483 acres) is zoned for medium and high-density residential use. In the three Oregon counties, there are 143,217 acres of land zoned for residence, of which 16.8 percent (about 23,989 acres) are zoned for medium and high-density residential use. Thus, there is more land zoned for medium and high density residential use in the three Oregon counties than Clark County, Washington, both in total and proportionally. In addition, the 2005 zoning categories indicate that the three Oregon counties have much more land zoned for "mixed uses" (13,465 acres) than Clark County's mixed use zoning (1,464 acres).

In addition to its land use components, Metro's 2040 Growth Concept Plan (1995) emphasizes TOD to reduce car travel. To reduce the reliance on single-occupant automobile use, the plan promotes increased planning for alternative modes and street connectivity and encourages land use patterns throughout urban areas that make it more convenient for people to walk, bicycle, use transit, and generally drive less to meet their daily needs. Similar to their Oregon neighbor, the Washington State GMA sets it

transport goal as encouraging efficient, multi-modal transportation systems that are based on regional priorities and coordinated with county and city comprehensive plans; however, promotion of the TOD concept in Clark County came much later than in Oregon. Figure 9, below, shows the transit lines in the Portland metropolitan area. The Oregon side of the metro area has higher densities, a more fully developed bus transit system, and light rail and streetcar--modes not found in Clark County/Vancouver.
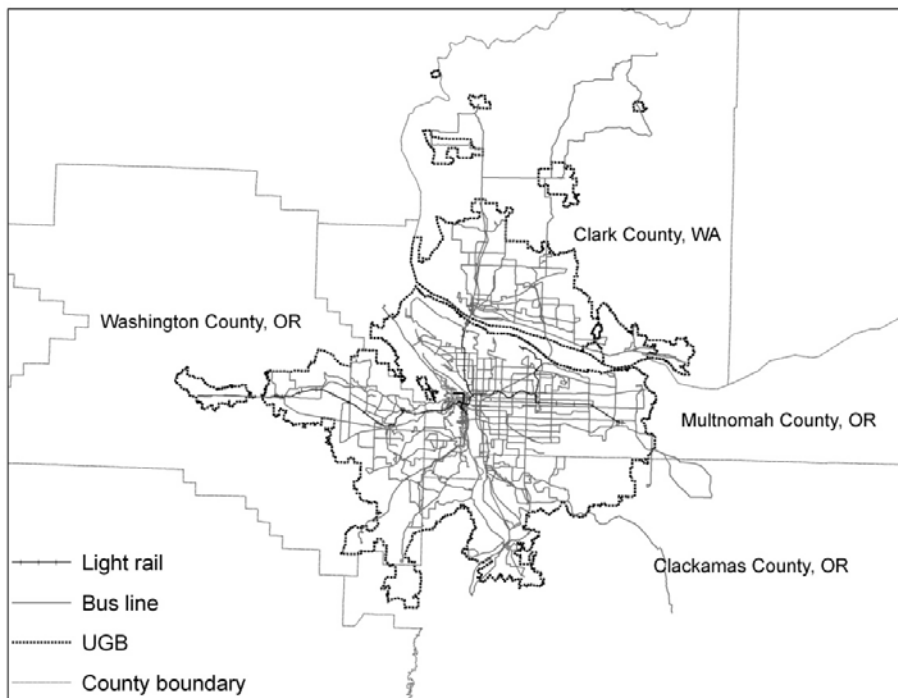
Figure 9 Transit lines in Portland metropolitan area in 2005

Data source: Metro, RLIS (2005)

3.3 Prior Examination of Portland's Smart Growth Policies

Previous studies examining the impacts of Portland's Smart Growth policies have focused on its UGB, questioning whether it has successfully prevented urban sprawl and

encouraged certain growth patterns such as higher land use density, more mixed use, and TOD.

Nelson and Moor (1993) divided the fore-mentioned three Oregon counties into four parts based on their relative locations to the Portland UGB. Comparing land use density in these four areas from 1985 to 1989, they concluded that development was being directed to the areas within the UGB but recognized that considerable development continued outside the UGB and efficient expansion of the UGB in the future may be jeopardized by low-density development along the boundary. Abbott's (2002) analysis of multiple data sources concluded that Portland's UGB had successfully contained new development within the UGB, increased land use density, and decreased the average new lot size. Further, he noted that almost all the new housing outside the UGB was located in Clark County, Washington, where traditional subdivisions of moderately priced houses were still available. Song and Knaap (2004) developed several quantitative measures of urban form and computed these for neighborhoods of varying age in Washington County, Oregon, part of the Metro region. Their findings indicated that single-family housing unit density, internal street connectivity, and pedestrian accessibility have increased at the neighborhood level, but that the mixing of land uses was limited.

While the studies reviewed above reported that Portland's UGB was at least partially achieving its goals, other researchers have come to contradictory conclusions. For example, Richardson and Gordon (2001) argued that Portland is actually statistically similar to Los Angeles in terms of suburbanization, decentralization, and public transit. Jun (2004) evaluated the impact of Portland's UGB on urban spatial form in three

different ways: comparing Portland with other metropolitan regions, comparing areas inside and outside the UGB, and developing a regression model. Jun's study confirms the argument by Richardson and Gordon that Portland has not been successful in sprawl control, citing that Portland experienced substantial population suburbanization from 1980 to 2000. Jun's study also suggested that Portland's UGB has had little effect on attracting new residential development into UGB and has actually had a significant impact on diverting new growth into Clark County, Washington.

There may be many reasons for the contradictory conclusions of previous studies, but data limitations and methodological inconsistency figure prominently. First, many studies fail to recognize that the UGB is only one of several Smart Growth policy instruments implemented in the Portland metropolitan area. Studies examining the impacts of Smart Growth policies other than the UGB are rare, such as densification and mixed-use development incentives, TOD, and road and sewerage extension policies. The success of the UGB is highly dependent on the support of these other planning policies as well as the existing built environment and exogenous socio-economic factors. In the extensive literature assessing the effectiveness of the UGB, very few of them control for those factors. Indeed, these studies also rely upon outcomes aggregated at a fairly coarse spatial level and lack a behavioral foundation rooted in economic theory.

Another gap in literature is that very few studies have examined the impacts of Smart Growth policies from developers' perspective. As major space producers in U.S. cities, real estate developers' response to such policies is critical to its success. Downs (2005) summarized three main groups that advocate Smart Growth practices: nongovernment

41

environmentalists, urban planners and other local public officials, and innovative private real estate developers. Compared to the other two groups, however, the preferences of developers toward Smart Growth polices is ambiguous. On one hand, it is reported that consumers are willing to pay a premium for housing with Smart Growth characteristics (Song & Knaap, 2003), and some developers have applied for permits to build projects following Smart Growth principles (Downs, 2005). On the other hand, O'Connell's (2009) recent survey over 202 U.S. cities on their planning and development officials suggested that developers and real estate interests were the most active opponents of Smart Growth policies.

3.4 Model Formulation

Discrete choice modeling techniques were used to reveal the compensatory tradeoffs that developers make when choosing sites for their housing projects among a set of alternative locations. Since each developer may have several housing projects, projects rather than developers were used as individual observations in the model. Thus the model represents the choice of location for a particular project, given the attributes of that project and its developer.

Discrete choice models are usually derived under the assumption that a decision maker seeks to maximize his/her utility. As one cannot directly observed utility, random utility theory suggests that utility is a latent construct that can be measured indirectly as a function of the attributes of each alternative and the preferences of each decision maker for each attribute (Train, 2003). In this research, each developer chooses from J alternative TAZs for project i. The developer expects to obtain a certain level of utility

U_{ij} from each alternative TAZ j for project i. The utility is composed of two parts, the systematic portion V_{ij}, which can be measured and an unobserved random portion, ε_{ij}, an error term:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{2}$$

In this study, for each alternative TAZ j, there are four groups of alternative specific attributes: a vector of location-specific variables representing Portland's Smart Growth policies X_{sj}, a vector of TAZ site attributes X_{lj}, a vector of TAZ socio-demographic attributes X_{dj}, and lagged-effect variables X_{gj}, such as whether a TAZ or its neighbors had been previously chosen as a project site by the same developer. For project i, the function for the systematic utility portion is:

$$V_{ij} = \beta_s X_{sj} + \beta_l X_{lj} + \beta_d X_{dj} + \beta_g X_{gj} \tag{3}$$

where, β_s represents a parameter vector for Smart Growth policy variables, β_l represents the parameters for TAZ site variables, β_d represents the parameters for TAZ socio-demographic variables, and β_g represents the parameters for lagged effects. Assuming that the error components ε_{ij} are identically and independently distributed (IID) across the alternatives, and following an extreme value distribution, also known as a Gumbel distribution, the choice probability for alternative TAZ j may be written as:

$$\Pr(j'|i) = \frac{\exp{(\beta_s X_{sj} + \beta_l X_{lj} + \beta_d X_{dj} + \beta_g X_{gj})}}{\sum_{j=1}^{J} \exp{(\beta_s X_{sj} + \beta_l X_{lj} + \beta_d X_{dj} + \beta_g X_{gj})}} \tag{4}$$

which is the multinomial logit model (MNL) model. Equation (4) thus represents the probability of choosing location j' for project i from a set of alternative locations.

Baseline models using Equation (4) were estimated for each of the two market segments, SFH and MFH.

Each individual makes a choice from a set of alternatives available assumed to be available to them. In this study, there are 1998 TAZs in the region. It is assumed that each SFH unit consumes at least 0.1 acres of buildable vacant land, each MFH unit consumes at least 0.01 acres of buildable land, and only TAZs with enough buildable land were considered to be eligible candidates for the choice set. Thus, smaller size projects tend to have larger full choice set, and larger projects tend to have smaller full choice sets. With this constraint, it was still neither computationally feasible nor theoretically realistic to assume that developers would consider all the eligible TAZs as alternatives in the choice set for each project. Therefore, a random sample of 19 alternative TAZs from all the eligible TAZs, plus the chosen TAZ were used as the choice set for each project. Alternatives were sampled without replacement and without any type of importance sampling or stratification. According to MacFadden (1978) and Nerella and Bhat (2004), random sampling of alternatives provides consistent estimates even for small sample sizes when using a MNL model. One advantage gained by sampling alternatives from full choice set is that it helps avoid the spatial autocorrelation among alternative locations, who are likely to be proximate to each other if full choice sets are used.

To examine whether developers behave differently in Oregon and Washington, I specified a "bi-state" model in which all the variables were interacted with a dummy variable d_{clk}, which represents whether the TAZs are in Clark County, Washington. The equation for the models is:

$$\Pr(j'|i) = \frac{\exp\left(\beta_s X_{sj} + \beta_l X_{lj} + \beta_d X_{dj} + \beta_g X_{gj} + \beta_{ds} X_{sj} d_{clk} + \beta_{dl} X_{lj} d_{clk} + \beta_{dd} X_{dj} d_{clk} + \beta_{dg} X_{gj} d_{clk}\right)}{\sum_{j=1}^{J} \exp\left(\beta_s X_{sj} + \beta_l X_{lj} + \beta_d X_{dj} + \beta_g X_{gj} + \beta_{ds} X_{sj} d_{clk} + \beta_{dl} X_{lj} d_{clk} + \beta_{dd} X_{dj} d_{clk} + \beta_{dg} X_{gj} d_{clk}\right)} \quad (5)$$

where β_{d*} represents the parameters for all the interaction effects in the model. Models were estimated using maximum likelihood methods using NLOGIT software version 4.0 (Econometric Software, Inc., 2007).

My descriptive analysis revealed that the sizes of housing projects vary quite a bit. The distribution of project sizes tends to follow a basic power-law curve, with many projects of small size monotonically transitioning to very few projects of large size. In order to account for these differences and the relatively greater complexity and constraints faced by developers as the magnitude of a project increases, I estimated models using the project size in dwelling units as the observation weight. Consequently, the preferences of a developer with a project size of 100 units contributed the same information to the likelihood calculations as 100 projects in which a single home was constructed. Preliminary estimation results show much better goodness-of-fit measures for the weighted models compared with the un-weighted models and parameter estimates that were more consistent with our expectations. In this chapter, only the estimation results from the weighted models are reported. Explanatory variables are the TAZ attributes presented and discussed in Chapter 2 (see Table 2).

3.5 Model Results and Findings

To compare model goodness of fit, I calculate the adjusted pseudo R² ($\bar{\rho}^2$) for each model based on the formula:

$$\bar{\rho}^2 = 1 - \frac{(LL(\hat{\beta}) - N)}{LL(c)} \qquad (6)$$

where $LL(\hat{\beta})$ is the log likelihood value at convergence, N is the number of parameters, and $LL(0)$ is the log likelihood value with only the constant in the model.

As indicated by Table 3, SFH models returned much higher adjusted pseudo R² than MFH models, suggesting that SFH models fit in the data better. For both SFH and MFH sectors, the bi-state models show slightly better model goodness of fit than the base models.

3.5.1 Smart Growth Policies

Table 3 presents the estimation results from the base models and the bi-state models for SFH and MFH developers. The base model assesses to what extent SFH and MFH developer location choices are influenced by Smart Growth policies in the region while controlling for other site and socioeconomic variables and the lagged effects of past choices. In the bi-state model, all the variables are interacted with a dummy variable representing the TAZs in Clark County, Washington to indicate whether there are statistically significant differences across state lines.

UGB

As discussed above, estimated parameters on UGB dummy variables represent policies, land use conditions and development-related costs that are concomitant to the UGB but not included and controlled for elsewhere in the model. It should be noted that since most MFH developments are within the UGB, UGB dummy variables were not included in the MFH developer models and therefore only appear in the table for the SFH models.

	SFH				MFH			
	Base Model		*Bi-State Model*		*Base Model*		*Bi-State Model*	
Variables	*Coef*	*t-value*	*Coef*	*t-value*	*Coef*	*t-value*	*Coef*	*t-value*
Smart Growth policies examined and their corresponding variabels								
UGB:								
UGB_IN	--	--	--	--	--	--	--	--
UGB_ON	0.0648	1.47	-0.1564	-2.44	--	--	--	--
UGB_ON*Clark			0.3235	3.58			--	--
UGB_OUT	0.5622	10.29	0.6798	9.80	--	--	--	--
UGB_OUT*Clark			-0.4325	-3.51			--	--
UGB_EXP	0.0133	0.23	-0.0525	-0.65	--	--	--	--
UGB_EXP*Clark			0.0961	0.80			--	--
SFH net density:								
SFH_DEN_N	--	--	--	--	--	--	--	--
SFH_DEN_L	0.7452	8.36	0.7501	7.79	--	--	--	--
SFH_DEN_L*Clark			-0.6150	-2.92			--	--
SFH_DEN_M	1.4189	15.28	1.3476	13.71	--	--	--	--
SFH_DEN_M*Clark			-0.4861	-2.30			--	--
SFH_DEN_H	2.0555	18.13	2.0961	17.31	--	--	--	--
SFH_DEN_H*Clark			-1.9169	-5.59			--	--
MFH net density:								
MFH_DEN_N	--	--	--	--	--	--	--	--
MFH_DEN_L	--	--	--	--	-0.3423	-1.49	-0.2450	-0.90
MFH_DEN_L*Clark			--	--			-0.1521	-0.25
MFH_DEN_M	--	--	--	--	0.3055	1.35	0.3540	1.34
MFH_DEN_M*Clark			--	--			0.7150	1.17
MFH_DEN_H	--	--	--	--	0.7795	3.32	0.9104	3.36
MFH_DEN_H*Clark			--	--			0.3673	0.55
Housing diversity:								
H_MIX	--	--	--	--	--	--	--	--
H_SFH	-0.0309	-1.19	-0.1533	-4.85	-0.0167	-0.09	0.0212	0.10
H_SFH*Clark			0.3136	4.67			0.0499	0.09
H_MFH	-0.1562	-3.67	-0.1584	-3.35	0.4456	5.44	0.4885	5.52
H_MFH*Clark			0.3242	2.89			-0.5253	-2.04
Mixed use:								
MIX_USE	--	--	--	--	--	--	--	--
MIX_RES	0.0888	3.36	0.1541	4.57	0.4205	4.19	0.3532	3.26
MIX_RES*Clark			-0.1725	-3.10			0.1276	0.41
MIX_NON	0.0504	1.92	-0.1214	-3.75	-0.0039	-0.05	-0.0748	-0.79
MIX_NON*Clark			0.5640	9.45			0.3688	1.39
Accessibility:								
AUTO_RET	-0.0191	-3.43	-0.0313	-5.21	-0.0051	-0.25	0.0082	0.38
AUTO_RET*Clark			0.1206	4.90			-0.0564	-0.43
AUTO_NRET	-0.0001	-0.04	0.0011	0.85	0.0029	0.62	-0.0002	-0.03
AUTO_NRET*Clark			-0.0225	-4.28			0.0132	0.47
TRS_RET	0.0213	2.81	0.0199	2.39	-0.1072	-4.64	-0.1109	-4.50
TRE_RET*Clark			-0.0348	-1.05			0.0401	0.26
TRS_NRET	-0.0043	-2.58	-0.0050	-2.76	0.0241	4.75	0.0249	4.71
TRS_NRET*Clark			0.0101	1.37			-0.0093	-0.27
Existing infrastructure								
RD_DEN	0.0016	5.72	0.0027	7.09	-0.0019	-2.96	-0.0010	-1.47
RD_DEN*Clark			-0.0019	-3.15			-0.0038	-1.73

Table 3 Base model and bi-state model results

	SFH				MFH			
	Base Model		Bi-State Model		Base Model		Bi-State Model	
Variables	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value
Variables controlled while assessing smart growth policies								
Existing site attributes:								
LAND_SFH	0.1114	27.08	0.1197	23.71	--	--	--	--
LAND_SFH*Clark			-0.0313	-3.43			--	--
LAND_MFH	--	--	--	--	-0.2677	-15.88	-0.2038	-10.32
LAND_MFH*Clark			--	--			-0.2540	-6.57
LAND_VL	-0.0605	-11.27	-0.0447	-7.10	0.0160	3.94	0.0170	3.99
LAND_VL*Clark			-0.0847	-5.63			-0.0464	-0.80
Socio-Economic attributes:								
HINC	0.0019	2.07	-0.0007	-0.67	-0.0008	-0.26	0.0022	0.69
HINC*Clark			0.0180	7.88			-0.0006	-0.05
AGE	0.0096	6.29	0.0109	5.93	-0.0029	-0.64	0.0011	0.22
Age*Clark			-0.0163	-5.31			0.0023	0.18
Lagged effects								
AR(1)	4.6152	97.90	4.6158	96.94	5.1828	18.24	5.0662	17.79
NB(1)	1.3429	28.51	1.3301	28.26	3.4067	9.90	3.3323	9.64
AR(2)	1.9542	31.64	1.9470	31.56	2.4208	4.79	2.3429	4.67
NB(2)	1.2503	25.28	1.2250	24.71	1.3276	3.92	1.2324	3.60
Number of parameters	23		42		20		36	
log likelihood	-28208		-27916		-2441		-2409	
Null log likelihood	-48652		-48652		-3301		-3301	
Pseudo R sq	0.42		0.43		0.26		0.27	
Adjusted Pseudo R sq	0.42		0.43		0.25		0.26	
Weighting variable	project size		project size		project size		project size	
Sample size	16320		16320		1138		1138	

Table 3. (Continued)

Using TAZs within the UGB as a reference case, the estimation results suggest that, all else being equal, there is statistically significant preference for developing SFH outside of the UGB, and this effect is consistent for both states. As suggested above, after controlling for land value and availability, the lower cost of doing business outside of the UGB and pent up demand for more rural living would seem to hold sway in some project location decisions. In Oregon these would be limited to single dwelling units on very large lots (e.g., two acres or more). As suggested by the base model, there is no significant difference between the TAZs within the UGB and the TAZs on the UGB

boundary line. However, the bi-state model suggests that while there is a statistically significant propensity to avoid developing along the UGB boundaries in the three Oregon counties, new SFH developments in Clark County, Washington were more likely to be in the locations along the UGB boundaries. In the base and bi-state models, there were no significant effects attributed to a TAZ located in a UGB expansion area, meaning that these lands are no more attractive for SFH home development than the land originally contained within the UGB. This finding might be temporary, however, since all the UGB expansion areas counted by this study were drawn after 2000, most in 2004, and it takes time for developers to respond to the new designations.

Residential Density

The existing density of SFH in units per acre was found to have a significant, positive effect of the project location preference for SFH development. As shown in Table 3, relative to the reference case of zero existing SFH units, the magnitude of the coefficients estimated for the low-, medium- and high-density SFH indicator variables increases from lowest to highest in the base model, and these coefficient are statistically different from one another. In the bi-state model, however, there is a significant offsetting negative parameter for Clark County in all the density categories. The net effect is that, while positive, developers in Clark County are more likely to develop medium-density SFH projects than high-density SFH projects. These results may be attributed in part to increased profitability of developing SFHs at higher densities, up-zoning and policy-driven density bonuses, and the relative cost of land within the UGBs, which tends to be higher in the three Oregon counties compared with Clark, County.

In the MFH models, the estimation results show a similar albeit somewhat weaker propensity for locating new MFH projects in TAZs where there are currently high densities of MFH. The results are not significant statistically for the low- and medium-density categories, but significant for high-density category, which is also true in the Bi-state model. These results are consistent with the notion that, in both states, MFH development is being directed to areas that already have a lot of MFH housing.

Housing Diversity

The estimation results for the base model in Table 3 show that, relative to the reference case of "mixed SFH and MFH," developers are significantly less likely to develop SFH in TAZs in which MFH dominates. In the bi-state model, developers are significantly less likely to develop SFH in TAZs in which either SFH or MFH dominates. However, there is a significant, positive offsetting effect for Clark County, resulting in a different preference ordering. In the three Oregon Counties, the preference would be for developing in "mixed" housing zones, followed by "SFH dominated" and "MFH dominated." In Clark County, the net effect of the parameter estimates reveals a similar preference to "SFH dominated" and "MFH dominated" zones, with the "mixed housing" category being least preferred, all else being equal. The seeming ambiguity of the Clark County results might reflect a less mature growth management system in which there is an existing shortage of mixed housing areas. This is gradually changing as new SFH units are built in areas currently dominated by MFH.

The MFH base model shows that MFH developers have the preference to the TAZs dominated by MFHs, but they are indifferent between the TAZs with mixed housing and

the TAZs dominated by SFHs; however, the Bi-state model shows that this is only true

for developers in the three Oregon counties. MFH developers in Clark County,

Washington are also indifferent between the TAZs with mixed housing and the TAZs

dominated by SFHs, but there is a significant, negative parameter for the interaction with

"MFH dominated" dummy variable, which has a magnitude large enough to offset the

coefficient of the main effect. This indicates that in Clark County, Washington, MFH

developers are less likely to choose the location dominated by MFH.

Mixed Use

Downs (2005) argued that among many Smart Growth policies, providing for mixed land

uses is likely to be implemented since it does not incur strong opponents. Using TAZs

characterized as "mixed use" as the reference case, the estimation results shown in Table

3 suggest that SFH developers are more likely to choose the TAZs dominated by

residences but are indifferent between the land that is dominated by non-residential use

and the land that is mixed use. The SFH bi-state model shows SFH developers in the two

states have a different preference ordering. In the three Oregon counties, the preference

ordering is: TAZs dominated by residential use, followed by "mixed use," and then TAZs

dominated by non-residential uses. In Clark County, the SFH preference ordering favors

the TAZs dominated by non-residential use, followed by the TAZs with mixed use, and

finally the TAZs dominated by residential use. This seemingly surprising difference

between the two states may again reflect differences in maturity of their respective

growth management systems. Recent efforts to guide growth into a mixed used scenario

in Washington is in its early stages; therefore, lacking a great deal of existing mixed use

areas, new SFH is being built in areas that have heretofore been dominated by non-residential land uses. A future analysis might re-label these TAZs as "mixed use" as the residential housing stock catches up with non-residential land uses.

For MFH, the estimation results show a more prosaic outcome in which development of new MFH projects shows the strongest propensity towards TAZs that are currently dominated by residential uses, while exhibiting no significant secondary preference for locating in commercial-dominated zones versus mixed use. These outcomes are consistent between the two states.

Auto and Transit Accessibility

As shown in Table 3, the effects of four transportation-related accessibility variables were tested. These four measures were differentiated by mode, auto or transit, and by the object of access, retail or non-retail employment, as surrogates for employment or shopping opportunities, respectively. For SFH, the parameter estimates of the base model suggest an aversion to the areas with a high level of access to retail employment by car and non-retail employment by transit, and significant preference to locations with better accessibility to retail employment by transit; however, the bi-state model reveals that this is only true on the Oregon side of the border. SFH developers in Clark County show similar preference for transit accessibility to SFH developers in Oregon. In addition, Washington developers show preference for locations with better accessibility for retail employment by car and aversion to locations with better accessibility for non-retail employment by car. Anecdotally, greater Vancouver and Clark County would seem to have a higher proportion of strip commercial development and shopping areas close to

new residential subdivisions. Statistically, this finding is consistent with the above analysis on mixing of uses in which there is a significant preference for developing SFH projects in zones dominated by commercial uses.

The MFH model estimation results paint a quite different picture. As shown in Table 2, auto accessibility to either retail or non-retail employment has had no significant effect on the propensities of developers to site new MFH projects in both states. In contrast, accessibility by the transit mode is highly significant, with the base model showing a significant negative effect towards locating projects in areas with high transit access to retail and a significant positive effect of locating in TAZs with high access to non-retail employment. The bi-state model shows no significant difference between the two states. These results seem to reflect the vast conversion of former industrial lands along Portland's south waterfront and north of downtown in the now famous Pearl District to high-rise multi-family structures during the study period. Similar, smaller-scale conversions have also taken place in Vancouver, Washington near the waterfront. These areas have high transit accessibility and are close to downtown employment centers.

Existing Infrastructure

Using road density as a proxy for development-supportive existing infrastructure, I tested the notion that, Smart Growth policies encourage more efficient use of existing infrastructure. It is noted that while residential infrastructures such as sewage and water system are usually accompanied with road systems, road density is not always a perfect proxy. Another issue is that since the number of road lanes is not counted, road density

tends to be over-measured in the areas where the number of road lanes is small; while in the areas where the road lanes are large, road density is under-measured.

The base models show that this would seem to be the case for new SFH development but the reverse seems to be true for MFH development. The SFH bi-state model estimates show a significant negative offsetting effect for SFH developers in Clark County. This may reflect a more lenient policy framework in Clark County towards infrastructure extensions and lower development surcharges, compared with the three Oregon counties. In Oregon, state legislation requires consistency between transportation and land use plans and the Oregon Department of Transportation is involved in reviewing plan amendments, thus the adequate provision of infrastructure is assured.

3.5.2 Site Characteristics and Socio-Economic Variables Controlled

The supply of buildable land in a TAZ, expressed in the natural log in the model, is expected to be a primary attractor for new housing development. The estimation results show that this is evident in the SFH base model by a very strong, significant positive parameter estimate for SFH buildable land. In the SFH bi-state model, there is an offsetting negative effect in Clark County for SFH acreage, meaning that the marginal utility of each additional vacant square meter is somewhat less valuable. Given the weighting of projects by size, this may be interpreted in terms of the utilization rate of vacant land, which in Clark County means larger average SFH lot sizes, compared with the three Oregon Counties.

However, in both the MFH base model and bi-state model, Table 3 shows significant negative coefficients for buildable land in both states. Considering that all the alternatives

in the choice set were required to have enough buildable land when the choice set was determined, this suggests that for MFH development, as long as the locations have enough buildable land, MFH developers prefer to the TAZs with less amount of buildable land, which are more likely to be in the built-up areas. This is consistent with the descriptive analysis in Chapter 2 which showed that MFH developments were much less likely to occur in urban peripheries than SFH development.

As mentioned earlier, vacant land fragmentation and zoning change were not considered when measuring the acreage of buildable land in TAZs, which might have caused some bias in model estimation. To ignore land fragmentation tends to over-measure the amount of buildable land in the TAZs where vacant land is more fragmented, and those TAZs are more likely to be in built-up areas. This might lead to the underestimation of the importance of buildable land and the attractiveness of TAZs on urban peripheries. I also noticed that some housing developments, especially MFH developments in high density areas were built on developed land originally zoned for non-residential purpose, such as industrial or commercial. Ignoring zoning change could cause the overestimation of the importance of vacant land and underestimation the possibility of housing development in built-up areas where vacant land is scarce.

The estimation results also show that, all else being equal, developers of SFH prefer less expensive land, as suggested by the significant negative coefficient estimates show in Table 3. There is a significant additional, negative effect in Clark County shown in the bi-state model results, which suggest that developers there are even more cost averse. In contrast, there is significant positive effect of land value among developers of MFH in

both states. This might be due to the fact that MFH development is more likely to occur in high-density areas where land price tends to be higher and land price captures some positive factors for MFH development that are missing from the models. As noted above, land values used are assessed values, not transaction prices, thus they lag behind market rates usually by a few years, if not more. This fact helps us to avoid the endogeneity problem in which the marginal utility of land values reflects the values of other attributes included in the model that contribute to land price. Nevertheless, the gap between the assess land value and the market land value tends to under-measure the importance of land price in developers' location decision, and this is especially true for a housing market boom period.

As shown in Table 3, the two socio-economic variables show significantly positive parameters in the SFH based model. The results from the SFH bi-state model indicate that SFH developers in Clark County are more likely to choose TAZs with higher average household income than SFH developers in three Oregon counties, but they prefer to build in the TAZs with fewer young and middle-age households, ages 25 to 54. This might suggest a stronger trend in Clark County for developing SFH homes for wealthier, older households, a supposition that may be supported by the previous analysis indicating trends towards lower density SFH developments compared with the Oregon counties. The MFH models show that MFH developers are not sensitive to the average household income and the age structure in the TAZs in both states. Acknowledging that there might be loop causality between the socio-economic status of a TAZ and the housing

development in that TAZ, but the use of Census 2000 data for all the eight years of study period helps us avoid that problem.

3.5.3 Lagged Effects of Previous Choices

The estimation results on lagged effects for SFH and MFH developers are similar and quite intuitive. As shown in Table 3, if a developer was active within the previous two years, there is a significant positive effect of locating the current project in the same TAZ as the previous project, or in a neighboring TAZ. The effects lessen somewhat going from a one-year lag to a two-year lag, and are greater for the same zone (AR) than for a neighboring zone (NB). The magnitude and significance of these variables makes them the strongest predictors in the entire specification. This would seem to confirm that developers form attachments to the locations with which they are familiar, or are phasing projects with separate permits in consecutive years.

3.6 Conclusions

The developer project location choice models developed in this study show that home developers in the Portland metropolitan area are sensitive to most Smart Growth policies being implemented in the region, but they react to them differently across the border between Oregon and Washington. In addition, SFH developers and MFH developers show different preferences for location attributes.

In our models, the most significant predictors of where a developer will choose to locate a project are the locations of previous projects. For SFH development, the availability of appropriately zoned vacant land is another very significant positive factor, with a

significant difference between states in terms of the marginal value of vacant land, which seems to be higher in the Oregon portion of the region for SFH. This would seem to reflect larger average lot sizes for new SFH in Clark County, Washington. For MFH development, developers tend to choose the locations with less amount of buildable land when all the alternatives have enough buildable land for development. There is a strong propensity to save money on land prices for SFH developers. But the MFH models suggest that MFH developments are more likely to occur on land with high price, reflecting the fact that MFH development is more likely to be in high density location and higher land price might capture some positive factors that are missing from the model. In terms of demographics, there is a slight significant preference for developing SFH in Clark County in TAZs with higher incomes and with a lower proportion of young and middle-aged adult households, compared with across the border in Oregon.

In both states, developers seem to respond positively to Smart Growth policies that would encourage developing at higher densities, providing a mix of housing types, and encouraging mixed used development. On the Oregon side of the metropolitan area, the highest SFH densities are preferred, but on the Washington side there remains a stronger propensity to develop in medium-density areas. In addition, there is evidence that new SFH is built predominantly in TAZs that already contain a mix between SFH and MFH housing types or is dominated by SFH in Oregon, whereas in Clark County recent SFH developments have been more likely to be developed in either existing SFH or MFH-dominated zones, but not mixed housing. The interpretation is that Clark County contains fewer existing mixed housing types and that this reflects the gradual transition to mixed

housing types in the propensity to develop SFH in areas now dominated by MFH. Supporting this notion of a maturing growth management system is the significant positive preference estimated for Clark County to develop SFH in areas dominated by non-residential development, compared with Oregon-side preferences for avoiding non-residentially-dominated zones and favoring existing residential and mixed use zones in that order. Essentially, Clark County development is transitioning toward a more mixed use pattern through SFH development in commercial areas. This is also supported by a significant positive impact of automobile accessibility to retail employment in Clark County and an opposite negative impact of auto access to retail employment in the three Oregon counties.

In terms of MFH, in both states the development preferences seem to favor existing locations that are dominated by residential use, especially the locations with high MFH density. This confirms the existence of "state dependency" in residential development. At the same time, there is a significant impact in both states of developing MFH in TAZs that are highly accessible by transit to non-retail employment, which supports the Smart Growth ideals of fostering shorter, non-auto work commutes and reflects recent large-scale development of former industrial lands.

Finally, after controlling for all of the other factors discussed above, there remains a strong preference for developing SFH units outside of the UGB in both Oregon and Washington sides of the Portland metropolitan area. I hypothesize that this effect may reflect demand by a market segment with preferences for lower-cost rural living and may

be largely the collective outcome of many one-time private developers who build a single residence on a large lot.

To synthesize, there would seem to be evidence that Smart Growth policies have been implemented in ways that are achieving objectives of greater densification, mixes of housing types and land uses, and transportation system accessibility on both sides of the Portland Metropolitan area. At the same time, there exists what seems to be a latent demand by some consumers which motivates some developers to seek out lower-cost and lower density opportunities where permitted and, in the Portland Metro area, there are more of these opportunities in Clark County.

The models developed in this study implicitly assume taste homogeneity for location attribute preferences. The next chapter examines to what extent taste variation plays in location choice decisions.

CHAPTER 4 EXPLORING THE PREFERENCE HETEROGENEITY IN HOME DEVELOPERS' LOCATION CHOICE

4.1 Introduction

In Chapter 3, each type developers (SFH and MFH) are assumed to share the same tastes in their location choice, which is a very strong consumption. In this chapter, the preference heterogeneity in home developers' location choice is explored by developing mixed multinomial logit (MMNL) models for SFH and MFH developers. Compared to the MNL model, MMNL models are free from the restrictive independently and identically distributed (IID) assumption and able to accommodate taste variations across individuals. To my knowledge, this is the first study that explores the preference heterogeneity in home developers' location choice using MMNL models.

In the reminder of the chapter, I compare several MMNL models and give the formulation of the models selected to explore home developers' preference heterogeneity; discusses model estimation results; and conclude with summary of findings and their policy implications.

4.2 Methodology

4.2.1 Model Comparison

As in Chapter 3, I exogenously divide the housing market into SFH and MFH segments. Within each segment, to explore home developers' taste heterogeneity in their location choice, the following models have been tested and compared: 1) simple MNL models with partial exogenous market segmentation by having explanatory variables interacted

with dummy variables created based on developer and project attributes; 2) pure random parameter logit (RPL) models in which some parameters are set as randomly distributed; 3) RPL models with the heterogeneity in the means of random parameters considered by setting the means of random parameters as a function of developer and project attributes; and 4) latent class models with endogenous market segmentation.

I finally chose latent class models because it helped avoid the following issues associated with other types of models mentioned above. First, while MNL models with partial exogenous market segmentation can capture systematic heterogeneity, the housing market segmentation has to be determined in an ad hoc way, which is inherently arbitrary. In addition, when using more than one developer attribute to group developers into different market segments, the number of segments grew fast and the sample sizes of some segments became too small to return reasonable coefficients. Secondly, pure RPL models cannot capture the systematic variations in individual developers' preference, making it less useful for this study. Moreover, there is little agreement or governance as to selecting the distribution type of random parameters. Thirdly, compared to pure RPL models, RPL models with heterogeneity in the means of random parameters are able to capture systematic taste variation among developers, but the issue of selecting distribution types for random parameters remains. Finally, it takes much longer time for RPL models (pure and non-pure) to converge, making exploration of alternative specifications very time consuming.

As summarized by Bhat (1997), compared to the models mentioned above, the latent class model determines the market segmentation endogenously, and it shows the

following advantages: 1) it jointly determines the number of segments, the assignment of individuals to segments, and segment specific choice model parameters, which allows the use of many segmentation variables; 2) it obviates the need to determine the market segmentation in an ad hoc way. In contrast to RPL formulations, the LCM provides a more intuitive basis for describing heterogeneity.

4.2.2 Latent Class Model Formulation

As in Chapter 3, since each developer can have several projects in different locations, housing projects were used as observation units in the model.

Following Train (2003), latent class models in this study are derived as follows. Each developer faces choice among J alternative locations for his/her project i. The developer obtains a certain level of utility U_{ij} from each alternative location j for project i, and the utility is composed of two parts, the systematic portion V_{ij} and the error ε_{ij}:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{7}$$

In this study, for each alternative location j, there is a set of alternative specific location attributes X_{ij}. Assuming that the error ε_{ij} in utility function is independent and identically distributed (IID), the choice probability for alternative location j is:

$$\Pr(i,j) = \frac{\exp(\beta' X_{ij})}{\sum_{j=1}^{J} \exp(\beta' X_{ij})} \tag{8}$$

where β' denotes the parameter for each TAZ attribute.

Following Bhat (1997) and Greene and Hensher (2003), assuming that there are C relatively homogenous segments in the market, the choice probability for alternative location j chosen by a developer for project i in segment c is

$$Pr(i,j|c) = \frac{\exp(\beta_c' X_{ij})}{\sum_{j=1}^{J} \exp(\beta_c' X_{ij})}$$

(9)

where β_c' denotes the parameter for each TAZ attribute in segment c. A segmentation model that decides the probability of project i in segment c can be expressed as

$$H_{ic} = \frac{\exp(\theta_c z_i)}{\sum_{c=1}^{C} \exp(\theta_c z_i)}, \quad c = 1, \ldots, C,$$

(10)

where z_i denotes project and developer related attributes which are used to estimate the segmentation models and θ_c represents the parameters for those variables. Then, the unconditional (on segment membership) probability of project i choosing alternative TAZ j can be written from equation (9) and (10) as

$$P(i,j) = \sum_{c=1}^{C} H_{ic} Pr(i,j|c)$$

(11)

With the parameters in the segment-specific location choice models and the parameters in the segmentation model, one can estimate the mean values of developer and project related attributes in segment c ($\bar{z}_c$) as follows:

$$\bar{z}_c = \frac{\sum_n H_{ic} z_i}{\sum_n H_{ic}}$$

(12)

The means of location related attributes in segment c ($\bar{x}_c$) can also be calculated as follows:

$$\bar{x}_c = \frac{\sum_n H_{ic} x_j}{\sum_n H_{ic}}$$

(13)

4.2.3 Determination of Choice Sets

As mentioned in Chapter 3, McFadden (1978) proved that random sampling of alternatives provides consistent estimates even for small sample sizes in the MNL model. Unfortunately, in the non-MNL models, such consistency does not hold. The good news is that Nerella and Bhat (2004) found that increasing sample size for non-MNL models can dramatically improve the accuracy and, with very large sample sizes, the accuracy of random sampling is comparable to the accuracy of the MNL model. They recommended that when sampling alternative is the only realistic way to estimate non-MNL models, the analyst should consider a sample size no less than one fourth of the full choice set and preferably one half of the full choice set.

As in Chapter 3, it is assumed that each SFH unit consumes at least 0.1 acres of buildable vacant land, each MFH unit consumes at least 0.01 acres of buildable land, and only TAZs with enough buildable land were considered to be eligible candidates for the choice set. With this constraint, the full choice set for most projects is still too large to estimate the model. To achieve a realistic choice set size for the latent class model estimation, the full choice set faced by each housing project was reduced by assuming that all the eligible alternative TAZs must be in the same county as the chosen TAZ. This was based on the finding that less than 10 percent SFH developers in the region had projects in more than one county during the eight-year study period. With these two constraints, 99 alternative TAZs for each project were sampled without replacement, making the average choice set size closer to or larger than one-fourth of the full choice set for all projects. As mentioned in Chapter 3, an advantage gained by sampling alternatives is that it helps

avoid the spatial autocorrelation among alternatives, which are likely to be proximate to each other if full choice sets are used.

Explanatory variables in location choice models are the selected TAZ attributes presented in Chapter 2 (see Table 2). Different from Chapter 3, however, the model specifications developed in this chapter deliberately avoided the use of dummy variables when possible since the use of dummy variables in the market segmentation model tends to create small segments in which the sample size is too small to return reasonable coefficients. For the same reason, the two UGB dummy variables representing TAZs on the UGB line and TAZs in the UGB expansion areas were combined into one group called TAZs in UGB peripheral areas. Developer size, project size, developer's contract type, and specialization were the variables used to determine the market segmentations for SFH and MFH developers. Descriptive analysis on these variables was presented in Chapter 2.

4.3 Model Estimation Results and Findings

While exploring developers' taste heterogeneity in their location choice, the housing market was divided into SFH and MFH sectors and these two sectors were modeled separately. Within each sector, estimation results from three models are reported: a simple MNL model, a developer-based latent class model, and a project-based latent class model. The difference between the developer-based and the project-based models is that the former uses developer size in the segmentation model while the later uses project size. By using developer size in the segmentation model, it was assumed that developers make location choice decisions based on their own size and different projects by the same developer reflect the same tastes for location attributes. In contrast, by using project size

66

in the segmentation model, it was assumed that developers make location choice by projects and different projects developed by the same developer reflect different tastes for location attributes.

Latent class models for SFH and MFH sectors were estimated with number of segments $C = 2$, 3, and 4. The estimation results show that, for each sector, latent class models with more segments yield better model goodness of fit in terms of adjusted pseudo R^2; however, the latent class model with four segments returned some unreasonable parameters. Thus, the latent class models with three market segments for the SFH and MFH sectors were preferred as offering the best combination of fit and interpretability.

Model estimation results for the SFH sector are presented in Table 4. As indicated by Table 4, compared to the MNL base model, both latent class models show better goodness of fit in terms of adjusted pseudo R^2 (0.154 and 0.157 vs 0.136), meaning that segmenting the markets endogenously with the latent class model does make significant improvements. The results show that the developer-based and project-based models tell quite similar stories due to the positive correlation between developer size and project size. Compared to the developer-based latent class model, however, the project-based latent class model returns better goodness of fit in terms of adjusted pseudo R^2 (0.157 vs 0.154), indicating that project size is a better explanatory variable than developer size for determining between market segments. This is also an indication that SFH developers make location choice decisions by projects, and they show different tastes for their different projects. Due to the space limitation, in this study, only the results from the

project-based model are discussed. The mean values of the segment-specific project- and location-related attributes based on it are presented in Table 6.

Model estimation results for the MFH sector are shown in Table 5. Compared to the MNL base model, the latent class models improves the adjusted pseudo R² from 0.074 to 0.114, indicating that endogenous MFH market segmentation explains MFH developers' location choice significantly better. Different from the SFH sector, however, the developer-based model and the project-based model for the MFH sector return the same model goodness of fit. One possible explanation is that, compared to SFH developers, MFH developers tend to have fewer projects, and the correlation between developer size and project size is higher for the MFH sector than the SFH sector (0.771 Vs 0.539). Similar to the SFH sector, only the results from the MFH project-based model are discussed below. The mean values of the project- and location-related attributes based on it are also presented in Table 6.

4.3.1 Class Probability Models

A latent class model estimates two sub-models simultaneously: the segment-specific choice model based on equation (9) and the segmentation model by equation (10). The segmentation model is the model that determines the probability that each project belongs to each segment. As Table 4 shows, the sizes of segment 1, 2, and 3 in the SFH project-based model are 33.0 percent, 25.2 percent, and 41.8 percent respectively. Table 6 shows that Segment 2 has the largest average project size (8.49 units), segment 1 is second (1.47 units), and Segment 3 has the smallest average project size (1.07 units). About 90 percent SFH projects in Segment 2 were developed by contractor-owners, much higher

	Base Model		Latent Class Model(developer-based)						Latent Class Model(project-based)					
			Segment 1		Segment 2		Segment 3		Segment 1		Segment 2		Segment 3	
Variables	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value
			Segment-specific location choice model						Segment-specific location choice model					
UGB dummy variables:														
Within the UGB	--	--	--	--	--	--	--	--	--	--	--	--	--	--
In UGB peripheral areas	0.1799	5.53	0.0721	1.16	0.0605	1.51	1.0207	15.15	0.0015	0.03	-0.1587	-2.99	0.9999	18.38
Outside the UGB	1.4210	36.30	-0.3369	-1.60	0.9205	18.91	2.5549	39.13	-0.0562	-0.30	-0.1316	-1.71	2.4178	45.11
SFH net density	0.0319	4.67	-0.0147	-0.76	0.0720	10.16	0.0561	5.17	-0.0010	-0.05	0.1177	14.60	0.0324	3.42
Housing diversity	-0.0061	-5.09	-0.1400	-4.66	-0.0004	-0.75	-0.0525	-6.15	-0.1535	-5.25	0.0009	1.65	-0.0332	-7.27
Mixed use	0.0134	3.41	-0.2035	-5.05	-0.0245	-3.74	0.0170	6.63	-0.1632	-4.49	-0.1843	-8.09	0.0171	7.23
Accessibility:														
Auto for retail emp	-0.0132	-2.32	0.0224	1.92	-0.0124	-1.65	-0.0751	-8.48	0.0238	2.12	-0.0205	-1.97	-0.0734	-9.91
Auto for non-retail emp	-0.0009	-0.78	-0.0070	-2.84	0.0014	0.87	0.0099	5.15	-0.0070	-2.95	0.0023	1.05	0.0109	6.80
Transit for retail emp	0.0031	0.49	-0.0131	-0.89	0.0171	2.19	0.0554	5.48	-0.0170	-1.19	0.0038	0.36	0.0691	8.19
Transit for non-retail emp	0.0010	0.70	0.0035	1.10	-0.0031	-1.82	-0.0051	-2.24	0.0044	1.43	-0.0013	-0.54	-0.0092	-4.89
Road density	0.0006	2.26	0.0016	2.50	0.0011	4.16	0.0016	3.82	0.0007	1.21	0.0021	6.05	0.0014	4.22
Buildable land for SFH	0.1871	48.05	0.7591	29.80	0.1337	27.85	-0.0057	-0.93	0.6889	25.47	0.1635	23.89	-0.0050	-0.95
Land value (deflated, in 2000$)	0.0101	2.30	0.1065	9.51	-0.0086	-1.60	-0.0371	-4.60	0.0926	8.66	-0.0445	-5.96	-0.0088	-1.38
Average household income	0.0074	10.49	0.0109	6.70	-0.0001	-0.16	-0.0131	-10.22	0.0118	7.71	-0.0097	-7.60	-0.0074	-7.10
% of household head aged 25-54	0.0082	6.34	-0.0023	-0.79	0.0071	4.68	0.0357	16.61	-0.0015	-0.53	0.0073	3.65	0.0274	15.50
Lagged effects:														
AR(1)	3.3043	100.14	3.2118	141.43	3.2118	141.43	3.2118	141.43	3.2243	136.46	3.2243	136.46	3.2243	136.46
NB(1)	1.0632	28.67	1.0272	40.56	1.0272	40.56	1.0272	40.56	1.0315	38.90	1.0315	38.90	1.0315	38.90
AR(2)	1.1949	26.84	1.1614	38.99	1.1614	38.99	1.1614	38.99	1.2075	38.81	1.2075	38.81	1.2075	38.81
NB(2)	0.6356	15.68	0.5954	21.64	0.5954	21.64	0.5954	21.64	0.6081	21.17	0.6081	21.17	0.6081	21.17
			Segmentation model						Segmentation model					
Constant	--	--	-1.0448	-4.35	-0.8203	-2.17	--	--	-2.8468	-17.43	-2.3874	-8.78	--	--
Developer size	--	--	0.3673	8.98	0.7723	12.29	--	--	--	--	--	--	--	--
Project size	--	--	--	--			--	--	1.1255	13.35	1.8597	11.99	--	--
Contract type	--	--	1.0976	15.56	0.6608	2.48	--	--	1.0175	17.26	0.5560	2.15	--	--
Specialization	--	--	-0.3145	-1.36	-2.2793	-6.96	--	--	0.6789	5.58	-2.2169	-9.13	--	--
Class size (%)	--	--	28.3		42.6		29.2		33.0		25.2		41.8	
Number of parameters	18		54						54					
Log likelihood at convergence	-64598		-63219						-62963					
Null log likelihood	-74793		-74793						-74793					
Psuedo R^2	0.136		0.155						0.158					
Adjusted psuedo R^2	0.136		0.154						0.157					
Sample size	16320		16320						16320					

Table 4 Base and latent class model results for the SFH sector

Variables	Base Model		Latent Class Model(developer-based)						Latent Class Model(project-based)					
			Segment 1		Segment 2		Segment 3		Segment 1		Segment 2		Segment 3	
	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value
			Segment-specific location choice model						**Segment-specific location choice model**					
UGB dummy variables:														
Within the UGB	--	--	--	--	--	--	--	--	--	--	--	--	--	--
In UGB peripheral areas	1.5150	7.06	0.4258	0.94	1.4987	4.08	2.7704	4.60	0.4213	0.89	0.0162	0.03	3.3803	10.18
Outside the UGB	0.5100	2.81	-0.3028	-0.49	0.8376	5.00	-0.3851	-0.57	-0.2628	-0.42	0.0955	0.44	0.9291	3.05
MFH net density:	0.0048	2.77	0.0023	0.37	0.0090	4.11	0.0047	1.40	0.0052	0.86	0.0032	1.54	0.0126	4.02
Housing diversity:	0.0001	0.36	-0.0025	-1.16	-0.0091	-5.60	0.0005	1.30	-0.0026	-0.96	-0.0001	-0.05	-0.0029	-3.37
Mixed use:	-0.1803	-4.22	-0.1022	-0.72	-0.2683	-6.18	-0.1238	-1.22	-0.1760	-1.17	-0.1541	-2.88	-0.3756	-6.08
Accessibility:														
Auto for retail emp	-0.0974	-3.45	-0.0394	-0.42	-0.1496	-5.57	0.1633	1.79	-0.0542	-0.55	-0.1731	-5.14	-0.0356	-0.83
Auto for non-retail emp	0.0235	4.07	0.0062	0.33	0.0390	7.17	-0.0298	-1.43	0.0098	0.51	0.0319	4.48	0.0235	2.72
Transit for retail emp	0.0887	3.25	-0.0262	-0.31	0.1756	6.71	-0.2623	-3.12	0.0086	0.10	-0.2088	-5.93	0.2709	6.82
Transit for non-retail emp	-0.0196	-3.34	0.0081	0.45	-0.0397	-7.08	0.0654	3.39	0.0009	0.05	0.0370	5.03	-0.0567	-6.65
Road density:	0.0012	2.07	0.0016	0.78	0.0028	4.81	0.0032	1.90	0.0020	0.97	0.0031	3.84	0.0029	3.66
Buildable land for MFH	-0.1876	-10.25	-1.1648	-9.85	0.2598	11.19	0.5311	6.40	-1.2141	-9.50	0.1949	5.40	0.2864	10.14
Land value (deflated, in 2000$)	-0.0209	-3.05	-0.0387	-1.52	-0.0331	-3.53	0.0418	2.94	-0.0498	-1.91	-0.0072	-0.95	-0.0549	-3.59
Average household income	-0.0173	-5.83	-0.0086	-0.97	-0.0273	-9.30	0.0139	1.62	-0.0144	-1.53	-0.0076	-2.11	-0.0343	-7.91
% of household head aged 25-54	0.0070	1.52	0.0102	0.82	0.0052	1.27	-0.0172	-1.20	0.0107	0.81	-0.0012	-0.18	0.0213	3.96
Lagged effects:														
AR(1)	4.1808	26.04	4.0082	33.99	4.0082	33.99	4.0082	33.99	4.0854	33.60	4.0854	33.60	4.0854	33.60
NB(1)	1.1210	4.83	1.1174	7.10	1.1174	7.10	1.1174	7.10	1.0576	6.71	1.0576	6.71	1.0576	6.71
AR(2)	2.1939	8.09	2.3242	12.60	2.3242	12.60	2.3242	12.60	2.3520	12.92	2.3520	12.92	2.3520	12.92
NB(2)	1.0106	3.95	0.8512	4.66	0.8512	4.66	0.8512	4.66	0.7499	4.12	0.7499	4.12	0.7499	4.12
			Segmentation model						**Segmentation model**					
Constant	--	--	3.5970	3.32	5.3696	4.97	--	--	-1.2952	-4.30	-0.7950	-2.52	--	--
Developer size	--	--	-0.0242	-5.30	-0.0312	-6.78	--	--	--	--	--	--	--	--
Project size	--	--	--	--	--	--	--	--	0.0086	2.56	0.0118	3.48	--	--
Contract type	--	--	0.8602	1.03	0.8016	0.97	--	--	-0.0485	-0.18	-0.3322	-1.09	--	--
Specialization	--	--	-0.7632	-0.86	-1.5073	-1.71	--	--	0.6715	2.51	0.3562	1.16	--	--
Class size (%)	--	--	22.8		69.0		8.2		22.3		28.5		49.2	
Number of parameters	18		54						54					
Log likelihood at convergence	-4713		-4472						-4472					
Null log likelihood	-5110		-5110						-5110					
Psuedo R^2	0.078		0.125						0.125					
Adjusted psuedo R^2	0.074		0.114						0.114					
Sample size	1138		1138						1138					

Table 5 Base and latent class model results for the MFH sector

	SFH				MFH			
Variable	Segment 1	Segment 2	Segment 3	Overall market	Segment 1	Segment 2	Segment 3	Overall market
	Project attributes							
Project size	1.47	8.49	1.07	3.08	28.49	35.85	14.17	24.05
Developer is contractor-owner	0.77	0.90	0.51	0.70	--	--	--	--
Developer is specialized on SFH	0.92	0.49	0.88	0.79	0.49	0.44	0.37	0.42
	Site and socio-economic variables							
UGB:								
Within UGB	0.84	0.83	0.33	0.63	0.86	0.94	0.96	0.93
On UGB periphery	0.14	0.08	0.07	0.10	0.03	0.04	0.04	0.03
Outside UGB	0.02	0.08	0.61	0.27	0.11	0.03	0.01	0.04
SFH net density	3.11	4.00	2.10	2.93	--	--	--	--
MFH net density	--	--	--	--	25.81	29.38	19.87	24.13
Housing diversity	0.24	2.73	0.35	0.91	30.10	42.38	6.75	22.95
Mixed use	0.70	0.83	1.18	0.93	1.18	1.26	0.89	1.07
Accessibility:								
car for retail emp	55.59	56.11	48.36	52.80	60.30	60.63	67.53	63.79
car for non-retail emp	264.05	266.78	227.89	250.16	291.07	304.55	315.34	306.58
transit for retail emp	15.84	19.76	13.65	15.94	26.89	27.70	33.16	30.07
transit for non-retail emp	71.82	89.19	61.02	71.83	126.25	139.73	143.44	138.45
Buildable land for SFH	54.18	30.72	8.27	29.77	--	--	--	--
Buildable land for MFH	--	--	--	--	0.35	11.78	8.25	7.57
Average residential land value	4.41	4.82	2.55	3.76	8.43	9.59	6.41	7.84
Road density	109.43	123.94	75.13	99.25	155.52	172.62	151.64	158.97
Average household income	63.44	53.02	55.14	57.47	44.62	45.47	40.14	42.78
Age struture	67.22	66.67	65.50	66.39	64.33	65.65	63.81	64.49
Class size	7334	5343	8560	21237	342	472	716	1530

Table 6 Mean values of the project- and location-related attributes

than the 70 percent average level. About 77 percent SFH projects in segment 1 were

developed by contractor-owners, still higher than the 51 percent for the projects in

segment 3. In contrast to the projects in segment 1 and 3, much more projects in segment

2 were developed by developers who also had MFH developments. In summary, SFH

projects in segment 2 are more likely to be the large projects built by contractor-owner

type developers who were active on both SFH and MFH markets. In contrast, SFH

projects in Segment 3 tend to be small projects built by contractors who are focused on

SFH.

As indicated by Table 5, the sizes of the three segments for the MFH sector are 22.3 percent, 28.5 percent, and 49.2 percent respectively. Different from the SFH sector, contract type is not a significant variable in the segmentation model for the MFH sector. Table 6 shows that the average project size in Segment 2 is the largest and the mean size of the projects in Segment 3 is the smallest. About 44 percent of the MFH projects in Segment 2 are built by developers who are focused on MFH, slightly higher than the average 42 percent for the overall MFH sector. Compared to other segments, the MFH projects in Segment 3 are more likely to be built by developers who also had SFH developments.

4.3.2 Segment-Specific Location Choice Models

As shown in Table 4 and 5, both the SFH and MFH project-based latent class models give segment-specific parameters with quite different signs and magnitudes, indicating that there are clear differences in intrinsic preferences and sensitivities to location characteristics among the segments in each market sector. A comparison of parameters across segments provides important qualitative information regarding the characteristics of each segment versus other segments. The mean values of location attributes calculated based on them shown in Table 6 provide similar information, but in a more intuitive way.

As indicated by Table 6, about 61 percent of SFH projects in segment 3 are located outside the UGB, much higher than the larger projects in segment 1 and 2 (2 percent and 8 percent respectively). It also shows that SFH projects in segment 3 are more likely to be built by contractors. These are consistent with expectations since SFH projects outside the UGB tend to be on rural land with owners who are more likely to hire contractors to

72

build houses for their own use. In contrast, the projects in Segment 1 with an average size is slightly larger than the projects in Segment 3 are more likely to be developed by contractor-owners and are less likely to be outside the UGB. Instead, they are more likely to in the UGB peripheral areas. This suggests that small-size SFH projects built by contractor type developers are much more likely to be outside of the UGB. However, this is not the case for MFH projects. As Table 6 shows, overall, 93 percent of the MFH projects are within UGB. It is MFH projects in Segment 1, with medium average size, that are more likely to be outside the UGB than the small MFH projects in Segment 3 and the large MFH projects in Segment 2.

SFH and MFH developers share similar taste variation on residential density and housing diversity: larger projects tend to be located in the TAZs with higher residential density and housing diversity. For the MFH sector, larger projects are also more likely to be in the locations with higher mixed-use levels. For the SFH sector, the projects in segment 3 with the smallest mean size are more likely to be in the TAZs with the highest mixed use level. This might be due to the fact that many small size SFH projects are in Clark County, Washington, where the employment is more evenly distributed spatially, making the average level of mixed use relatively higher for TAZs in suburban areas, compared with the Oregon Counties.

SFH developers' preference for transportation accessibility as illustrated in Table 6 is also consistent with expectations that larger-size SFH projects are more likely to be located in TAZs with better transportation accessibility by car and transit for both retail and non-retail employment purposes. This is also consistent with the finding that large-size projects are more likely to be within the UGB. In contrast, for MFH projects, the

projects in Segment 3 with the smallest average size are located in the TAZs with the best

transportation accessibility for both retail and non-retail employment by car and transit.

The projects in Segment 1, with medium average size, are most likely to be located in the

TAZs with the worst transportation accessibility, which is consistent with finding that

MFH projects in Segment 2 are more likely to be outside the UGB.

Model results show that the relationship between the sizes of projects and the amounts of

buildable land in their locations is not linear. For the SFH sector, the projects in Segment

3 with the smallest average size, tend to be located in the TAZs with the least amount of

buildable land, which is consistent with expectations. However, though the average size

of projects in Segment 2 is much larger than Segment 1, the average amount of buildable

vacant land in the TAZs for SFH projects in Segment 2 is much less than the TAZs for

the projects in Segment 1. For the MFH sector, the average project size in Segment 2 is

larger than Segment 1, and the projects in Segment 2 are more likely to be located in the

TAZs with more buildable land. However, it is the MFH projects in Segment 3 with the

smallest size that are mostly likely to be located in the TAZs with the largest amount of

buildable land.

Table 6 shows that for both the SFH and MFH sectors, larger size projects are more

likely to be located in the TAZs with higher land value. It also indicates that larger size

SFH and MFH projects are more likely to be in the TAZs with higher road density. These

two findings are consistent with each other because the land with more physical

development tends to have higher land price. It might also reflect the fact large size

developers have deeper pocket than smaller ones and are more competitive bidders in the

land market.

The TAZs in which the SFH projects in Segment 1 are located have the highest average household income and proportion of the households with young and middle age household heads. Given the finding that the SFH projects in Segment 1 are more likely to be in the TAZs in the UGB peripheral areas with the most buildable land, moderate residential density and mixed use, and lowest housing diversity, the SFH projects in Segment 1 might be the large size single-family detached homes built for higher income, young and middle age households with kids. It is showed that the large MFH projects in segment 1 are located in the TAZs with the highest average household income and proportion of the households with young and middle age household heads.

Since small size developers tend to be active in only one year, the estimation of lagged effects returns unreasonable parameters for the segments in which most developers are small. To solve this problem, the lagged effect variables were excluded from the segment-specific location models and constrained to be equal for all segments, making their coefficients represent the mean lagged effects for all the projects on SFH/MFH sectors. The estimation results show that the lagged effects for SFH and MFH developers are quite similar and consistent with the findings in Chapter 3. The four lagged effect variables are positive and very significant for both the SFH and MFH models. The magnitude of the parameter for AR(1) is so large that its influence on developers' location choice in the next year is almost deterministic. The effects lessen somewhat going from a one-year lag to a two-year lag, and are greater for the same zone (AR) than for a neighboring zone (NB). This confirms the findings from Chapter 3 that developers form attachments to the locations with which they are familiar, or are phasing projects with separate permits in consecutive years.

4.4 Conclusions

Latent class models have been developed to detect taste variations among home developers in the SFH and MFH markets separately. The project attributes such as developer size, project size, developer contract type, and developer specialization are used to segment the housing projects into three market segments endogenously. It is found that there are clear taste variations across developers and housing projects with respect to site attributes in their location choice. Compared to the simple MNL base model, the latent class models provided better fit to the data and offered more intuitively appealing results. It also shows that with other variables in the segmentation model being the same, project size provides a better fit to the data than developer size, indicating that developers have taste variations among their different projects.

The segment-specific location choice model shows that large size SFH projects developed by contractor-owners are more likely to be within the UGB and their locations tend to have higher residential density, housing diversity, transportation accessibility, road density, and land price. At the same time, the results show that smaller-size projects for SFH display a preference for building outside the UGB and in areas featured by older adults. With most MFH projects within the UGB, estimation results show that large size MFH projects prefer the locations with higher residential density, housing diversity, mixed use, road density, land price, average household income, and proportions of young and middle age household. However, different from SFH projects, it is the small MFH projects that are more likely to be located in the locations with the best transportation accessibility, most likely the result of urban infill and redevelopment.

Some policy implications can be drawn from this research. First, this study shows that large size housing projects are more likely to be located within the UGB and in locations with compact development characteristics such as higher residential density, housing diversity, mixed use, and existing physical development. This confirms the conclusion by Downs (2005) that large size developers are more likely to develop innovative housing projects following the Smart Growth principles.

Secondly, the descriptive analysis shows that the housing market in the Portland region is dominated by a small number of large developers. Model estimation indicates that developers with different attributes show clearly different intrinsic preferences in their location choice. Thus, it is important to treat developers and projects differently based on their sizes and other attributes when they are used as forecasting units to predict their location choice. More attention should be paid to large size developers and projects.

PART II

**COMPREHENSIVE NEW HOUSING SUPPLY AND LOCATION CHOICE
FORECAST MODEL**

CHAPTER 5 AGGREGATE NEW HOUSING SUPPLY MODEL

5.1 Introduction

The design of the housing supply and location choice model proposed by this study follows three basic steps:

    1) aggregate housing supply forecast;

    2) project synthesis;

    3) location choice.

This chapter presents the first step of the model. Chapter 6 and Chapter 7 discuss steps 2 and 3 respectively.

In this chapter, time series analysis techniques are used to analyze the housing market in the Portland metropolitan area. These are fully dynamic structural models, with changes in housing prices and other supply side cost shifters as predictors. Based on this analysis, three new housing supply forecast models are proposed.

It should be noted that, in this chapter, the MFH sector is segmented further into two sub-sectors: attached house (ATH) with two to four units in the structure, and apartment and condo (APT) with five or more in the structure. This stratification was done because these three market segmentation might have different sensitivity to market factors. When the forecast results are used for the next two steps, the forecast numbers of housing units in these two sub-sectors are combined as the predicted total new MFH supply.

5.2 Literature Review

In existing housing supply research literature, two basic approaches have been widely used: reduced form equations and structural equations (Dipasquale, 1999). Both of these two approaches estimate new housing supply by aggregate data. In the reduced form estimation, new housing supply is a function of prices and factors from both supply and demand sides, while in the structural equation models new housing supply is directly estimated as a function of price and cost shifters from the supply side only. Here, the term "new housing supply" is used to refer to new housing starts on the market, and to differentiate it from the term "housing supply" which sometimes is used to refer to the whole housing stock in the market.

Generally, the null hypothesis tested by reduced form estimations is that the long-run supply of new residential construction is perfectly elastic, and that the long-run equilibrium of the housing supply is entirely determined by factors on the demand side. Representative research using reduced form equations all reported the finding of a horizontal supply curve (Follain, 1979; Leeuw and Ekanem, 1971; Stover, 1986). The findings from these studies concluded that, in the long term, housing price is determined by input prices and long-run equilibrium housing supply is determined by demand.

Since the late 1980s, with the availability of more sophisticated time series analysis techniques, more structural dynamic supply models have been developed to estimate new housing supply by house price and cost shifters on the supply side only. It is argued that demand factors can be reflected by the house price in the model. There are two main theoretical underpinnings for these structural models: the investment theory and the urban

spatial/growth theory (Dipasquale, 1999). The main difference between these two approaches lies in their treatment of the land market. Those studies based on investment literature tend to ignore land as an input to production, while those based on urban spatial/growth theory explicitly include the land market in their models (Dipasquale, 1999). The papers by Poterba (1984) and Topel & Rosen (1988) are the two representatives for the literature using investment theory to estimate new housing supply. Both of these studies report significant impacts of housing price on new housing investment. The study by Topel and Rosen also found that the current housing price is not sufficient for new housing supply estimation, and builders have to form an expectation about future prices to determine their investment. This indicates the necessity of the dynamic model in new housing supply estimation.

Compared to the investment literature, studies based on spatial theory view the housing supply as more closely intertwined with the land market. An example is the work by DiPasquale and Wheaton (1994). According to DiPasquale and Wheaton, different from other production factors, the level of the land price is determined by the stock of the housing, not the flow or the level of building activity. With the increase of housing price, the level of the construction increases to obtain excess profit. However, as the housing stock grows, land prices rise and eventually absorb the excess returns. Construction then declines back to its normal level. Housing price levels generate new construction only if those prices dictate a long-run equilibrium stock that is higher than the current level. The construction activity reflects the dynamic adjustment process as the current stock moves to its long-run equilibrium level.

Mayer and Somerville (2000) also develop an empirical model of housing supply derived from urban growth theory. Instead of using the level of housing prices and other cost shifters, they argue that changes to housing prices and costs are more appropriate for predicting the housing starts. Their reasons can be summarized by three arguments. First, housing starts are a flow variable, thus should be estimated by other flow variables such as the change of price instead of the level of price. The level of price and cost predict the housing stock instead of new housing starts. Secondly, their historical trend analysis shows that the housing starts curve is more consistent with the change of price instead of the price level. Thirdly, treating housing starts as a function of housing price change is also consistent with the time series properties of housing stocks and price, because both the housing starts and housing price change are stationary variables while the housing price is non-stationary. In order to reflect the dynamic properties of the housing market, they use lagged housing prices as proxies for developers' future price expectations. Their model results show the significance of up to three lags of price changes and one lag of interest rate changes.

While the above literature provides important findings on the nature of the housing supply process at the aggregate level, there remain deficiencies in the analysis done to date. First, most of the existing housing supply research uses national data, assuming a single national housing market which might not reflect any local reality. This is an even more serious problem for the models based on the urban growth theory. By using national data, they are assuming the same urban form across different cities and states. Although Mayer and Somerville (2000) use national data in their study, they admit that "…the model is most appropriate for a single city..." Second, much of the housing supply

82

research focuses on single family housing market or the investment of the whole housing market, ignoring the differences between different housing sectors. Third, although recent studies use sophisticated time series analysis techniques, many early papers using aggregate time series data failed to carefully examine and correct the assumptions that econometric time series analysis tend to violate, making their research results questionable.

5.3 Data and Methodology

Table 7 presents the descriptive statistics for all the data used in this study. Quarterly data from 1996:Q1 to 2008:Q4 were used. Here, the term "1996:Q1" is used to represent the first quarter of 1996. Quarterly numbers of housing units in the permits (unit permits) issued by municipal authorities are used as a proxy for the new housing starts since actual housing starts data is not available at county or metropolitan area level. The unit permit data for each housing sector is from the monthly new privately owned residential building permits data (estimates with computation) reported by the U.S. Census Bureau at county and place level. The quarterly numbers of permitted units for SFH, ATH, and APT from 1996:Q1 to 2008:Q1 are plotted in Figure 10.

As indicated by the Table 7, from 1996:Q1 to 2008:Q4, the average numbers of units for single family houses, attached houses, and apartments and condos are 2,319, 167, and 1,022, respectively, accounting for about 66 percent, 5 percent, and 29 percent in the whole new housing market. The numbers of new housing units for the three housing types vary significantly over the cycle, with total unit permits ranging from 890 to 5,599. Most lower numbers occurred in 2007 and 2008.

Variable	Mean	Std.	Minimum	Maximum
Unit permits for single family houses				
Level	2329	566	502	3087
Changes	-43	398	-855	791
Unit permits for multi-family houses	167	77	28	325
Unit permits for apartments and condos	1022	542	339	2986
Unit permits for all housing types	3518	926	890	5599
Real new home sale price ($1000)				
Level	196.99	34.55	152.03	263.62
Changes	1.48	6.42	-12.08	19.17
New home sale time (days)				
Level	67.02	16.46	36.68	132.49
Changes	1.41	9.79	-14.17	32.08
Real prime interest rate	5.58	1.62	2.20	8.15
ENR construction cost index	3233	128	3072	3517
Non-construction employment in Oregon (000, NSA)				
Level	1527	66	1368	1650
Changes	5	25	-40	40
Average gas retail price (unleaded)				
Level	149.63	46.40	89.24	279.57
Changes	1.09	22.73	-110.17	48.12

Table 7 Descriptive statistics of time series model variables (1996:Q1-2008:Q4)
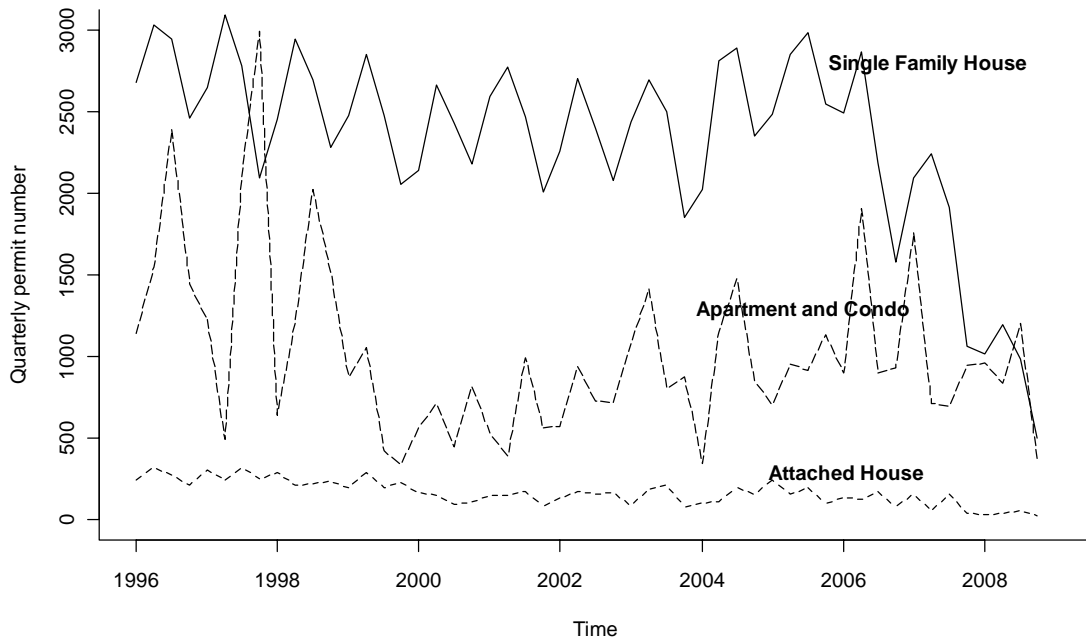
Figure 10 Quarterly permit unit numbers for SFH, ATH, and APT (1996:Q1-2008:Q4)

Data source: author's computation based the building permit data from U.S. Census Bureau (1996-2008)

One issue related to the use of Census housing permit data is that, to what extent this data is consistent with the real housing start numbers. To answer this question, U.S. Census Bureau (2009) gives their estimates of the relationship of new housing permits, starts, and completions on their website based on a summary of data from 1999 to 2004. The estimates were derived by either comparing published estimates for recent years or by tabulating unpublished data. For the most part, they are rough approximations, and measurements of their sampling errors have not been calculated, subject to sampling variability as well as non-sampling error. Their estimates are the housing starts are generally 2.5 percent less than the permits by total units, 2.5 percent greater than the permits in single family units, and 22.5 percent less than permits in multifamily units; the housing completions are generally 4.0 percent less than starts in total units, 3.5 percent less than starts in single family units, and 7.5 percent less than starts in multifamily units.

The quarterly new home sale price data are converted from the monthly average housing price data collected by the private company "Regional Multiple List Service (RMLS)". RMLS relies on their Realtor subscribers to collect the monthly housing price data for the Portland metropolitan area. This price data is not quality controlled. It is the average number of all housing types for both new and existing homes on the market. Real housing price is calculated by deflating the house price data with CPI index. As indicated by Table 7, real house prices in the Portland metropolitan area increased by an average of $1,480 per quarter (0.7 percent of the mean price level) between 1996:1 and 2008:4, with declines as large as -$12,080 in 2008:4 and increases as high as $19,170 in 2006:2.

Besides local housing prices, RMLS also reports the monthly market sale time for the Portland metropolitan area. The market sale time measures the average number of days it

takes for homes (new and existing) to sell. As Table 7 shows, the average market sale time is 67 days, ranging from 37 days in 2005:3 to 132 days in 2008:4.

The real prime interest rate is used to measure the capital cost to builders. The prime interest rate is the interest rate charged by banks for short-term loans to their most creditworthy customers. Only a small percentage of customers qualify for the prime rate, and thus it always serves as a basis for other, higher risk loans. The rate is almost always the same amongst major banks.

The construction cost for builders are measured by the cost index reported by Engineering News-Record (ENR). The construction cost index reported by ENR has four basic cost components: common labor, steel, lumber, and cement. Their proportions in the index are 80 percent, 13 percent, 6 percent, and 1 percent respectively. While this cost index typically measures the costs for nonresidential buildings, this study uses it as a proxy for the residential construction cost. The data is a national cost index based on the 20-city averages. Considering the national single market for construction materials, it is believed that this cost index can measure the local construction cost movement appropriately. Compared to the variables discussed above, the variation for the construction cost index is pretty small, ranging from 3,072 in 2001:1 to 3,517 in 2008:4.

As revealed by literature review, land price data can be very useful for the estimation of the housing starts. However, land price data has been notoriously difficult to collect, and this is the case for the Portland metropolitan area. This study does not include the land price variable in the model.

The changes of total employment (excluding construction) and real retail prices for unleaded gas price in Oregon are used as instrumental variables for house price changes to test for endogeneity and identification problems in the model. Since the endogeneity hypothesis is rejected, the results of the models using these two variables to test endogeneity are not reported. The issues of endogeneity and identification problems will be discussed in detail in the section discussing empirical estimation results.

When an OLS regression model is used to estimate time series data, both dependent and independent variables must be stationary. Otherwise the residuals of the OLS regression tend to be highly correlated, causing a spurious regression problem, which can yield erroneous estimation results. A common tool used to test the stationary nature of the variable is the Augmented Dickey-Fuller (ADF) unit root tests. The ADF test results for all the variables used in this study are presented in Table 8. Considering that the unit permit data might display a seasonal pattern, ADF tests with both seasonal and non-seasonal deterministic components were conducted. The results show that with the exception of the unit permit data for single family house, the unit permit data for attached house, apartment and condo and the new housing market as a whole are all stationary. This suggests the use of differences for the single family unit permit data and the level data for other housing types and the whole market. The unit root test shows that both the real housing price and market sale time data are not stationary but their changes are, suggesting the use of their changes instead of levels in the regression models. For the real prime interest rate and construction cost index, both their changes and levels are stationary.

Variable	Deterministic Components[1]	Significant lags order	Estimated Theta	ADF t-value	Reject unit root[2]
Single-Family housing unit permit					
Level	(0,0,(2,3,4))	9, 12	0.065	1.690	
Level	(0,0,(0))	N/A	-0.025	-1.093	
Differences	(1,1,(2,3,4))	N/A	-0.903	-6.016	***
Differences	(1,1,(0))	N/A	-0.985	-6.738	***
Attached housing unit permit					
Level	(1,1,(2,3,4))	N/A	-0.755	-4.620	***
Level	(1,1,(0))	8, 9	-0.801	-4.970	***
Apt-Condo housing unit permit					
Level	(1,1,(2,3,4))	N/A	-0.748	-5.138	***
Level	(1,1,(0))	N/A	-0.766	-5.423	***
Total housing unit permit					
Level	(0,0,(2,3,4))	N/A	-0.062	-1.129	
Level	(1,1,(0))	4	-0.549	-3.806	**
Real housing price					
Level	(0,0,(0))	4, 10	0.005	1.093	
Differences	(0,0,(0))	4, 5, 6, 7, 8, 9, 10	-0.988	-5.561	***
Real prime interest rate					
Level	(1,1,(0))	1, 3, 5, 7, 9, 11	-0.447	-3.720	**
Differences	(1,1,(0))	N/A	-0.760	-5.318	***
New home sale time					
Level	(1,1,(0))	12	0.253	3.432	
Differences	(1,1,(0))	N/A	-0.952	-6.177	***
Real Construction Cost Index					
Level	(1,1,(0))	4	-0.132	-2.350	*
Differences	(1,0,(0))	N/A	-1.255	-8.892	***
Non-constrution employment in Oregon					
Level	(1,1,(0))	1	-0.031	-2.125	
Changes	(0,0,(0))	N/A	-0.223	-2.163	**
Average real gas retail price (unleaded)					
Level	(0,0,(0))	N/A	-0.381	-3.257	*
Changes	(1,1,(0))	1	-2.007	-7.528	***

1. Deterministic components are intercept, time trend, and the quarterly seasonal dummies.
2. ***significant at 1% level; **significant at 5% level; *significant at 10%.

Table 8 Unit root test for variables

The equations used by this study to estimate new housing supply follow Mayer and Somerville (2000) with some modifications. Mayer and Somerville used the changes in house price, the changes to the supply side cost shifters and their lags to estimate the new housing starts. This study adds the changes in house prices, while other cost shifters are represented by their levels (not changes). Mayer and Somerville also use the level of the

market sale time to represent the market conditions; however, this study uses differences in market sale time, recognizing that market sale time levels are not stationary in the data set. Based on this descriptive data analysis, equation (14) is used to estimate new single family housing supply and equation (15) is used to estimate new housing supply for other market segments and the whole housing market:

$$\Delta S_t^{SF} = f(\Delta S_{t-1}^{SF}, \Delta S_{t-j}^{SF}, \Delta p_t, \dots, \Delta p_{t-j}, \Delta m_t, \Delta m_{t-j}, r_t, r_{t-j}, c_t, c_{t-j}) \qquad (14)$$

$$S_t^{Oth} = f(\Delta S_{t-1}^{Oth}, \Delta S_{t-j}^{Oth}, \Delta p_t, \dots, \Delta p_{t-j}, \Delta m_t, \Delta m_{t-j}, r_t, r_{t-j}, c_t, c_{t-j}) \qquad (15)$$

where ΔS_t^{SF} represents the changes of new housing supply for single family housing market, S_t^{Oth} represents new housing supply for attached housing market, apartment and condo market, and the total housing supply for the all housing types, Δp represents the current and lagged changes in house prices, Δm represents the change in sale time and its lags, Δr represents the capital cost to builders measured by real prime interest rate and its lags, and c represents the construction cost and its lags measured by the ENR construction cost index. Since quarterly data is used in this study, the dummy variables for the four quarters are also included in the equations with the first quarter used as the reference group. A time trend variable is also in included in the model. In order to sweep out possible autocorrelation and to identify full dynamic models, the lags of the dependent and independent variables are included in some models. The number of lags j for each variable is explored by the model specification in order to identify the full dynamic model.

5.4 Empirical Estimation Results

Using the variables and equations discussed above, the new housing supply models are estimated for each housing segments and the entire housing market. For each housing market segment, different numbers of lags for each variable have been tested to identify full dynamic models. Considering that quarterly data used by the study might bring seasonal autocorrelation to the model, AR(4) is used for each market segment. In order to save the degree of freedom for the estimation, the quarterly dummy for ATH model and the whole market model is dropped since it is not significant. Season adjustment was tried on the permit numbers to remove the seasonal effects in the market, but the estimation shows that seasonally-adjusted-unit permit data did not make a big difference from the unadjusted data.

Considering that the price variable in the model might be correlated with some variables omitted from the model, causing an endogeneity problem, two-stage least square (2SLS) models were tried for each market segment and the Durbin-Wu-Hausman test is used to test the existence of endogeneity problem in the model. Instruments for current changes in real housing prices are the current changes in non-construction employment and real gas price in Oregon. Lagged changes in construction cost index are used as the instrument for the current changes in construction cost index. The results from the Durbin-Wu-Hausman test for each model reject the endogeneity hypothesis for both housing prices and construction cost index. The possibility of identification problems can be ruled out from the theoretical perspective. Identification problem arise when simultaneous equations are used to estimate current supply by current price because supply and price tend to influence each other in both directions. In this study, however,

new housing supply is measured by unit permits issued by government authorities, rather than the new housing supply observed from the market. Due to the lag time between permit issuance and fully constructed housing units becoming available for sale, it is reasonable to assume that at least one year will pass before permitted units begin to affect market prices. Therefore, it is reasonable to believe that the lagged housing prices influence unit permit applications, not vice versa. This is also true for the construction cost variable in the model. Given the exclusion of the endogeneity and identification problems, only the estimation results from the OLS models are reported in Table 9. Since the dependent variable for the new single family housing supply model uses the changes, the estimation results from this model might not be directly comparable to the results from the other three models.

As Table 9 indicates, changes in housing prices have significant impacts on the new housing supply for all markets segment, but the magnitudes of the impacts and the number of lags differ. For the single family housing market, only the current changes in housing prices significantly impact the new single family housing supply, suggesting that the single family housing market responds rapidly to the market change. For example, with a $1,000 increase in the current housing price change, the change of the new single family housing supply in current quarter will be increased by about 14 units. The impacts of housing price changes on new attached housing supply are relatively minor. The first and third lags in housing price changes are significant at 10 percent level, suggesting that it takes more time for the attached house market to respond to the price signal. $1,000 increases in housing price changes at $t-1$ and $t-3$ quarters can lead to the increase in new attached housing supply by 2.69 and 1.99 units for current quarter t. For the

apartment and condo market, only the housing price changes at $t - 3$ and $t - 4$ quarters are significant, suggesting that the apartment and condo market is the slowest to respond to price changes. $1,000 increases in housing price changes at $t - 3$ and $t - 4$ quarters can lead to the increase of the new apartment and condo supply by 47 and 49 units for quarter t. The impacts of the housing price changes on the whole market appear to be the mixed effects from the three market segments. Changes in housing prices at current, $t - 3$ and $t - 4$ quarters are significant. $1,000 increases in housing price changes at current, $t - 3$ and $t - 4$ quarters can lead to the increase of the total new housing supply by 27, 52 and 56 units respectively for quarter t.

Several studies (Topel and Rosen, 1988; DiPasquale and Wheaton,1994; Mayer and Somerville, 2000) reported the significant role played by the lagged market sale time, which is used to represent the non-price signals for market conditions. Due to the non-stationary nature of the market sale time data in this study, the changes in the market sale time rather than the level data are used. Both the current changes to the market sale time and its lags have been tried for each model. Only its first lag is found significant for the whole housing market, showing that builders do respond to non-price market signals. A one day increase in the changes of market sale time decreases the total new housing supply by 33 units. The inclusion of this variable tends to reduce both the magnitude and significance of the coefficient for the current change in housing price. However, for all housing market segments, neither the current changes in market sale time nor its lags are significant.

Both the changes and the level of real prime interest rate and their lags are tested in the model. Although none of them are significant, their signs are negative, which is consistent with expectations. Both the changes and the level of construction cost index and their lags are tested in the model. For single family house and apartment they are not significant, which is consistent with the findings from other empirical housing supply studies. The level of the construction cost index in the single family house model and the model for the whole market are significant, but their signs are positive, which is counter to expectations and therefore difficult to interpret.

Dummy variables for the four quarters and time trend are included in each model to account for the potential seasonal pattern and time trend in the housing market. The estimation results show that only single family house and apartment markets demonstrate a significant seasonal pattern. Attached house and the whole market do not exhibit significant seasonal patterns, but demonstrate significant negative time trends, as opposed to the single family house and apartment markets. To remove the seasonal pattern from the model, the seasonal differences for single family housing unit permits and real housing prices were tried, but their ADF tests fail to reject the unit root hypothesis.

The calculation of the new housing supply elasticity estimated by this study depends on the length of the time over which the elasticity is calculated. For comparison with the elasticities reported by previous studies, the elasticity for the whole housing market are reported here. According to the model estimated for the whole housing market, a one-time change in housing prices influences new housing supply in current quarter, next $t + 3$ quarter and $t + 4$ quarter. Afterwards, the new housing supply returns to its previous level. A 1 percent increase in housing price in current quarter causes a 0.51

	SFH	ATH	APT	ALL
Variable	(Equation 1)	(Equation 2)	(Equation 2)	(Equation 2)
Price change at time t ($1000)	13.98**	1.63	-2.84	26.88*
	(2.03)	(1.36)	(-0.16)	(1.78)
Price change at time t-1 ($1000)	4.40	2.69*	-22.33	15.80
	(0.63)	(1.93)	(-1.31)	(0.85)
Price change at time t-2 ($1000)	-10.03	-1.68	-29.99	-3.40
	(-1.25)	(-1.45)	(-1.48)	(-0.21)
Price change at time t-3 ($1000)	8.65	1.99*	47.45**	51.74***
	(1.05)	(1.74)	(2.34)	(3.21)
Price change at time t-4 ($1000)	1.10	1.45	49.39**	56.01***
	(0.13)	(0.99)	(2.32)	(2.62)
Market sale time change at time t-1 (day)	-4.01	-0.83	-27.02	-33.18***
	(-0.59)	(-0.87)	(-1.51)	(-2.59)
Real prime interest rate at time t (%)	-30.63	-3.96	-35.76	-85.08
	(-1.51)	(-0.89)	(-0.66)	(-1.51)
ENR construction cost index at time t	-0.05	0.34***	1.44	2.38**
	(-0.14)	(3.42)	(1.17)	(2.08)
Time Trend	-6.04	-7.59***	-18.00	-55.01***
	(-1.65)	(-4.36)	(-1.41)	(-3.44)
2nd quarter dummy (Yes=1)	-69.85		77.67	
	(-0.40)		(0.20)	
3rd quarter dummy (Yes=1)	-708.22***		408.86	
	(-2.78)		(0.76)	
4th quarter dummy (Yes=1)	-672.84***		876.88**	
	(-3.65)		(2.51)	
AR(1)	-0.03	-0.11	0.19	0.17
	(-0.15)	(-0.64)	(1.10)	(1.11)
AR(2)	-0.29	0.34**	-0.18	-0.16
	(-1.64)	(2.41)	(-1.15)	(-1.06)
AR(3)	-0.20	-0.28	0.23	0.08
	(-1.22)	(-1.90)	(1.50)	(0.52)
AR(4)	-0.36*	-0.33*	-0.02	-0.10
	(-2.00)	(-1.96)	(-0.15)	(-0.70)
Constant	777.03	-637.53**	-3631.31	-2437.27
	(0.71)	(-2.59)	(-1.08)	(-0.84)
AR(1) rho	0.54	0.80	-0.04	-0.63
	(0.98)	(0.86)	(-0.08)	(-1.53)
AR(2) rho	-0.07	-1.11	0.44	0.28
	(-0.14)	(-1.54)	(1.35)	(0.72)
AR(3) rho	-0.14	-0.17	0.30	0.58
	(-0.38)	(-0.39)	(0.93)	(1.30)
AR(4) rho	-0.33	-0.43	0.26	0.45
	(-0.85)	(-1.20)	(0.91)	(1.15)
Adjusted R Sq	0.82	0.72	0.24	0.70
N	47	47	47	47

*t value in parentheses. ***significant at 1% level; **significant at 5% level;*
significant at 10% level.

Table 9 Time series regression results (1996:Q1-2008:Q4)

percent increase in new housing supply at current quarter t, a 0.96 percent increase in quarter $t + 3$ and another 1.1 percent in quarter $t + 4$. Over five quarters, a 1 percent increase in housing price causes a 2.56 percent increase in new housing supply. This is slightly smaller than the elasticities estimated by Mayer and Somerville (1988), but very close to the estimates by Topel and Rosen (1994) and DiPasquale and Wheaton (2000).

Comparison of the goodness of fit for the four models in Table 9 show that the model for single family house market based on equation (14) achieves the highest adjusted R square (0.82). The adjusted R squares for multifamily house and the whole market are around 0.70, and the adjusted R square for the apartment and condo market is very low (0.24). This indicates that factors other than those specified in the model influence the apartment and condo market in the Portland metropolitan area. The use of permit data in this study further makes the apartment and condo market difficult to predict. When applying for apartment and condo, one apartment or condo permits, builders might include many housing units, making the unit permit data for this sector lumpy and difficult to predict.

5.5 Forecast Models

Based on the analysis discussed above, three different models for forecasting the new housing supply for the Portland metropolitan area are proposed: a conditional regression forecast model, an unconditional regression forecast model, and an ARIMA (autoregressive integrated moving average) forecast model. The first two forecast models are time series regression models developed from the empirical estimations discussed above. They differ in that, the conditional forecast model requires knowledge of the predictor values for future years, whereas the unconditional forecast model does not.

Conversely, the ARIMA forecast model predicts new housing supply based on its past values only.

As the empirical estimation results discussed above indicate, changes in housing price and its lags are the most important predictors for the new housing supply in the Portland metropolitan area. The estimated capital and construction cost variables were not significant, or had the wrong signs, and were consequently omitted from the forecast models. The difficulty of future year prediction is another reason to exclude these cost variables from the models. Potential seasonal pattern and time trend for new housing supply are both included in the forecast models.

The conditional forecast model is developed based on data from the dataset. First, the data from 1996:Q1 to 2006:Q4 were used to estimate forecast models for each housing market segment and the whole housing market segment. In order to make sure that the predicted numbers of unit permits are always positive in the models for attached house, apartment and condo and the whole housing market, the natural log of permitted units was used as the dependent variables. The results are presented in Table 10.

Next, the price values for the period 2007:Q1 to 2008:Q4 were used to predict the new housing supply for the eight quarters during that period. This is also called an out-of-sample forecast. Last, the predicted new housing supply values are compared with the actual new housing supply values and the comparison is presented in Figure 11. The four graphs (a), (b), (c), and (d) in Figure 11 represent the single family, multifamily, apartment markets and the whole market respectively. As Figure 11 shows, the forecast model for the whole housing market performs very well, but it underestimates the

Variable	SFH (equation 1)	ATH (equation 2) (take log for DV)	APT (equation 2) (take log for DV)	All (equation 2) (take log for DV)
Price change at time t ($1000)	14.4140	0.0232	0.0327	0.0134
	(1.86)	(1.77)	(1.59)	(2.22)
Price change at time t-1 ($1000)	0.6336	0.0236	0.0012	0.0069
	(0.08)	(1.71)	(0.05)	(1.00)
Price change at time t-2 ($1000)	-15.0056	0.0012	-0.0053	-0.0021
	(-1.83)	(0.09)	(-0.23)	(-0.30)
Price change at time t-3 ($1000)	-1.6596	0.0166	0.0322	0.0134
	(-0.20)	(1.17)	(1.35)	(2.05)
Price change at time t-4 ($1000)	-0.5174	0.0098	0.0521	0.0176
	(-0.06)	(0.69)	(2.28)	(2.64)
2nd quarter dummy (Yes=1)	8.3030	-0.1572	-0.2335	-0.0008
	(0.06)	(-0.62)	(-0.58)	(-0.01)
3rd quarter dummy (Yes=1)	-530.0744	0.0232	0.4165	0.1110
	(-3.32)	(0.09)	(0.99)	(0.90)
4th quarter dummy (Yes=1)	-567.8542	0.0033	0.8109	0.1434
	(-4.24)	(0.01)	(2.32)	(1.39)
Time Trend	-0.6157	-0.0279	-0.0128	-0.0090
	(-0.17)	(-4.57)	(-1.17)	(-2.71)
AR(1)			0.2042	0.2419
			(1.18)	(1.52)
AR(2)			-0.1607	
			(-0.92)	
AR(3)			0.2087	
			(1.16)	
Intercept	275.5255	5.6467	4.8636	6.2451
	(2.25)	(27.33)	(2.58)	(4.67)
Adjusted R sq	0.8176	0.3902	0.2160	0.5417
t value in parentheses.				

**Table 10** Forecast regressions (1996:Q1-2006:Q4)

housing supply plunge in the last quarter of 2008. The forecast model for the single family housing market predicts the four quarters in 2007 well, but does not fully predict the big fall in new housing supply in 2008. The attached house model predicts the annual total accurately, but misses its seasonal pattern. Compared to the other three models, the performance of the model for the apartment market is poor, which is not surprising given its extremely low adjusted R square in the estimation model.
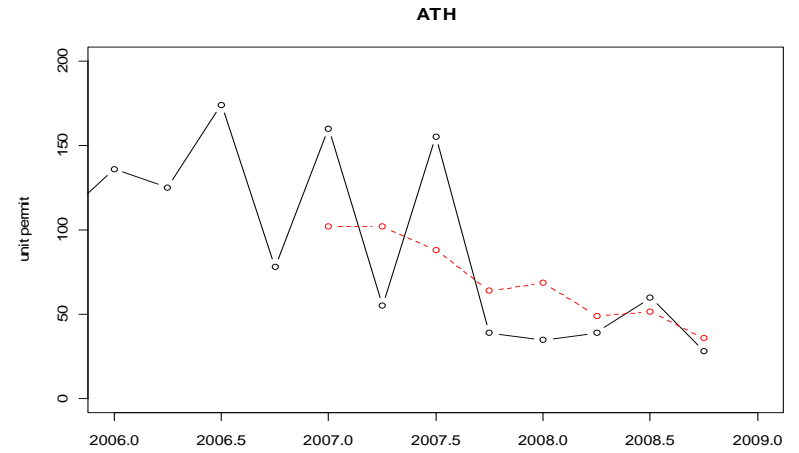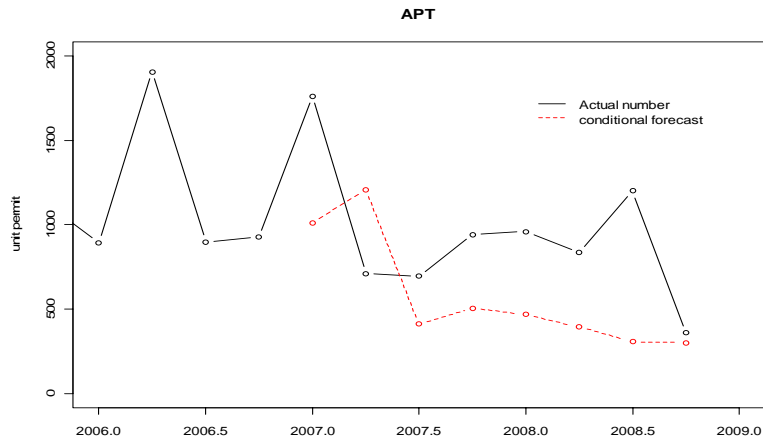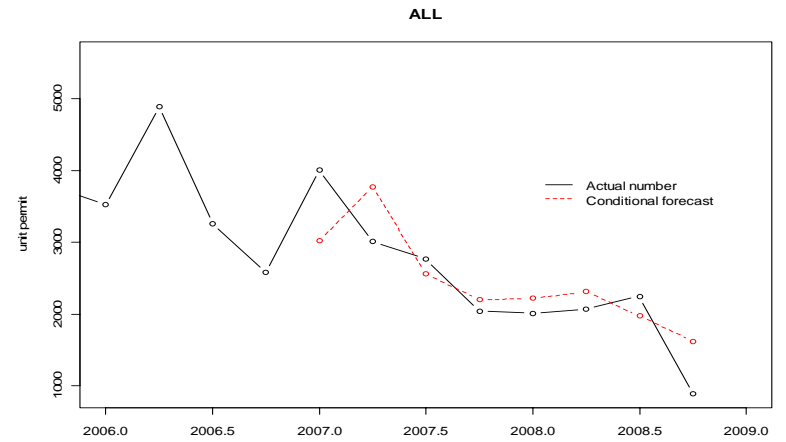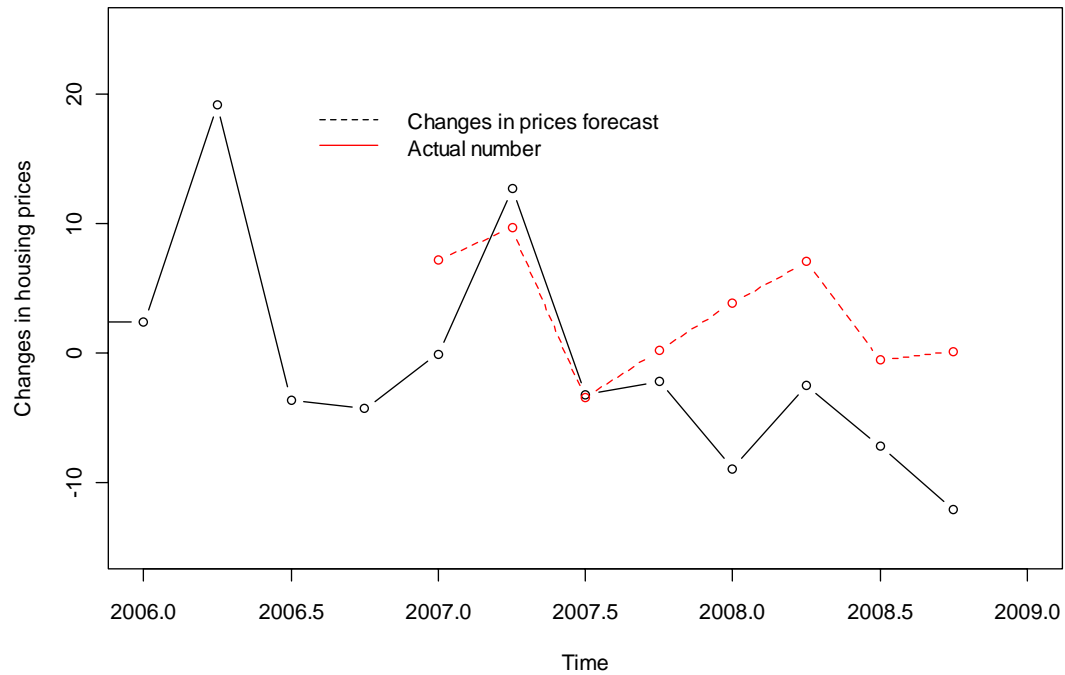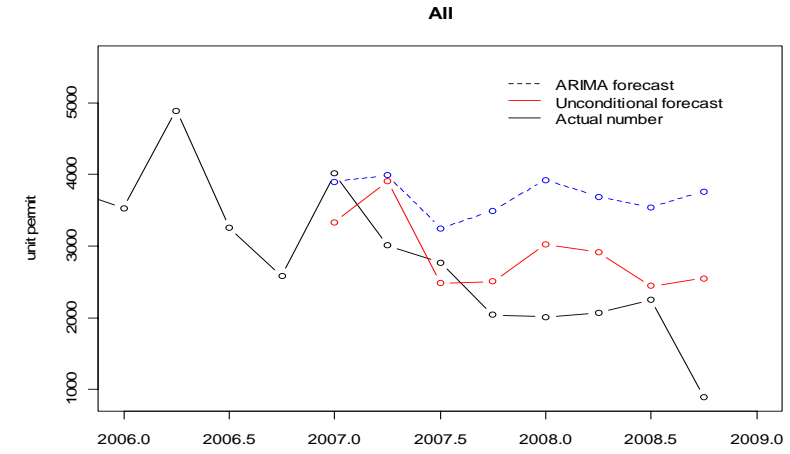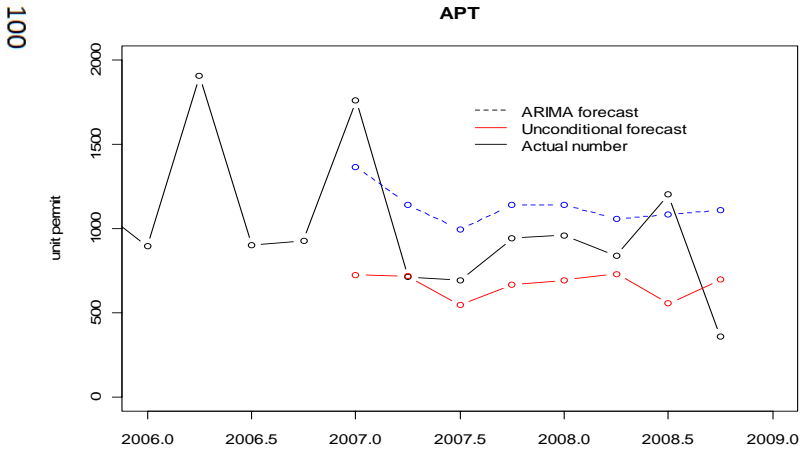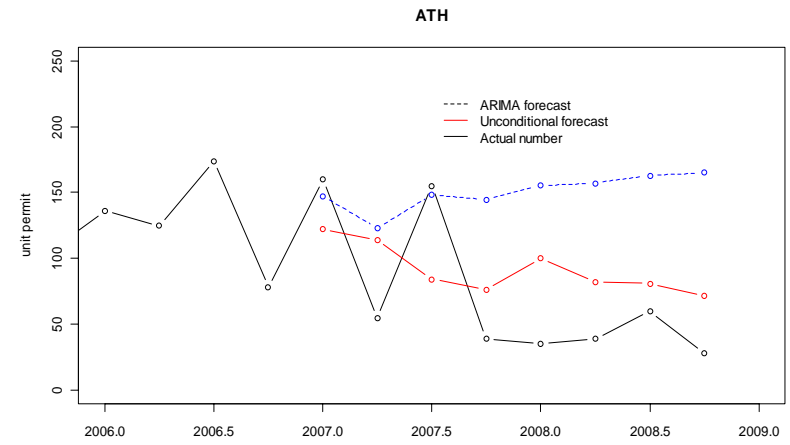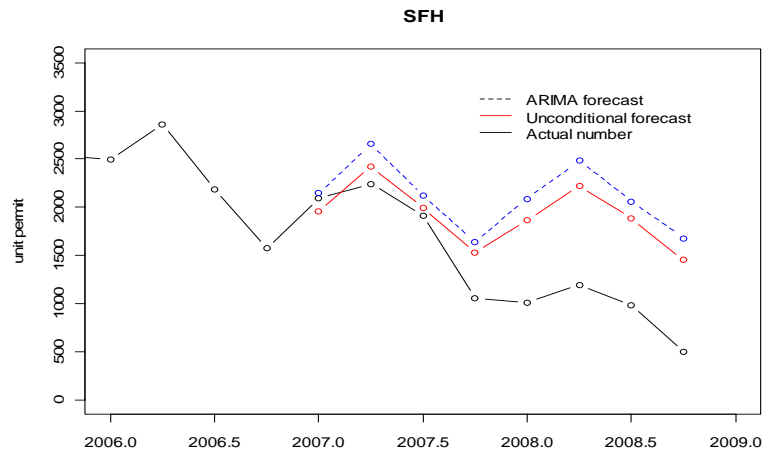
Figure 11 Conditional time series regression forecast versus actual (2007:1-2008:4)

Figure 12 Changes in prices forecast versus actual number (2007:1-2008:4)

	SFH	**ATH**	**APT**	**All**
Model Type	ARIMA(2,1,0)	ARIMA(2,0,0)	ARIMA(3,0,0)	ARIMA(3,0,0)
AR(1)	0.01	0.24	0.33	0.52
	(0.12)	(2.02)	(2.54)	(3.93)
AR(2)	-0.85	0.59	-0.18	-0.24
	(-11.69)	(4.64)	(-1.28)	(-1.68)
AR(3)			0.38	0.46
			(2.79)	(3.33)
Intercept/mean	-26.45	178.71	1055.10	3646.66
	(-1.50)	(5.01)	(7.22)	(11.64)
Log likelihood	-321.03	-255.53	-365.17	-378.42
AIC	650.06	519.06	740.34	766.83
t value in parentheses				

Table 11 ARIMA model estimation (1996:Q1-2006:Q4)

Figure 13 Unconditional time series regression forecast, ARIMA based forecast, and actual number (2007:Q1-2008:Q4)

The unconditional time series regression forecast model is developed in three steps. First, similar to the conditional forecast model, the data from 1996:Q1 to 2006:Q4 were used to estimate the housing supply model. Second, the historical real housing price data from 1996:Q1 to 2006:Q4 were used to predict values for the eight quarters in 2007 and 2008 by the ARIMA model. The estimated real housing prices for 2007 and 2008 and their actual numbers are displayed in Figure 12. Third, the new housing supply values for the eight quarters of 2007 and 2008 were predicted using the model estimated from the first step and the real housing price values estimated from the second step.

The ARIMA based forecast model is relatively straightforward. First, the unit permit data from 1996:Q1 to 2006:Q4 were used to estimate the ARIMA model for each housing market segment and the whole housing market. The results are presented in Table 11. Next these data were used to predict the new housing supply in 2007 and 2008. The actual housing supply number and the predicted numbers from the unconditional time series regression forecast model and the ARIMA forecast model are compared in Figure 14.

The four graphs (a), (b), (c), and (d) in Figure 14 represent the single family, multifamily, apartment markets and the whole market respectively. Since the data for 2009 are not yet available, it is difficult compare the two forecast models for that year. For 2008, however, the unconditional time series regression forecast model performs better for all three market segments and the whole market than the ARIMA forecast model.

The three forecast models proposed above all have advantages and disadvantages. The conditional time series regression forecast model predicts more accurately, but is

conditional on knowing the real housing price values for future years, which might not be the case in the land use modeling practice. For certain scenario analyses, where the real housing prices in future years are provided by an expert panel to test different scenarios, this model can be useful. The unconditional time series regression forecast model does not require knowledge of the housing prices in future years. Instead, the model forecasts based on the historical real housing price projected by forecast model. Compared with the ARIMA model, an advantage shared by both conditional and unconditional forecast models is the reflection of the real market dynamic process and the sensitivity to price signals from the housing market. The ARIMA model requires the least amount of data, but reveals nothing about other cost shifters in the housing market. Its prediction is entirely dependent on its past movement path.

5.6 Conclusion

In this chapter, fully dynamic structural models are developed to predict new housing starts for three market segments in the Portland metropolitan area. To my knowledge, the models developed in this research are unique among published sources for its use of local time-series data for building permit predictions. In addition, the models presented here have been rigorously tested and refined to avoid problems of non-stationarity, endogeneity and autocorrelation.

These results indicate the significant role played by the changes in real housing prices, but different housing sectors respond to the price signal at different paces. The single-family housing sector shows the fastest reaction; new starts adjust to a market shock within one quarter. For the attached housing, and the apartment and condo sector, it takes

102

four and five quarters respectively to adjust to a market shock. Over five quarters, a 1 percent increase in real housing price causes a 2.56 percent increase in the total new housing supply in the Portland metropolitan area. As a non-price signal for market conditions, the sale time for the houses on market is significant in new housing supply for the whole housing market, but not for any one of its three market segments. The estimations revealed that the capital cost and the construction cost to builders were not significant or had counter-intuitive signs. Based on this analysis, three new housing supply forecast models are proposed and their forecasting capabilities were compared to the empirical data. The advantages and disadvantages of these three models were discussed. When applied to land use modeling practice, data availability and purpose of the forecast should guide model choice.
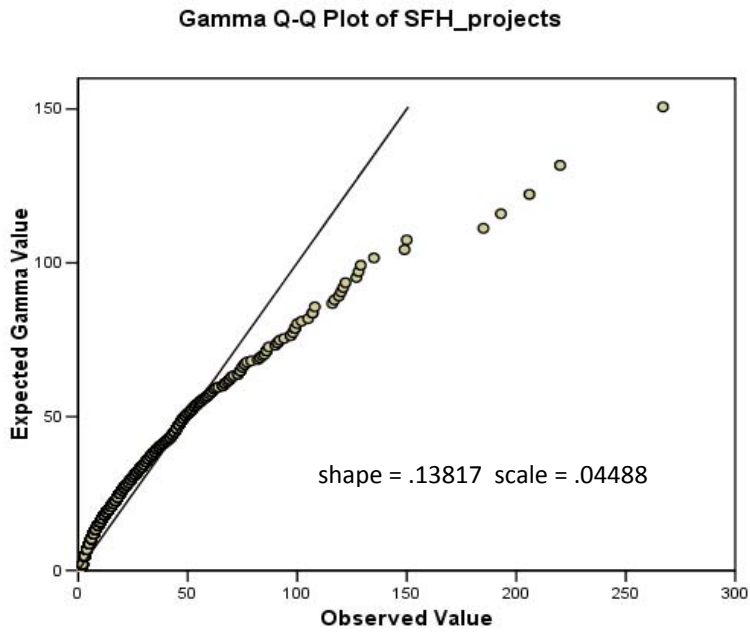
CHAPTER 6 PROJECT SYNTHESIS MODEL

In this chapter, SFH and MFH project synthesis models are developed based on the housing permit data from 2000 to 2006. The 2007 data was hold out to measure the performance of these models. The purpose of synthesizing SFH and MFH projects is to provide a basis for market segmentation in a forecasting context.

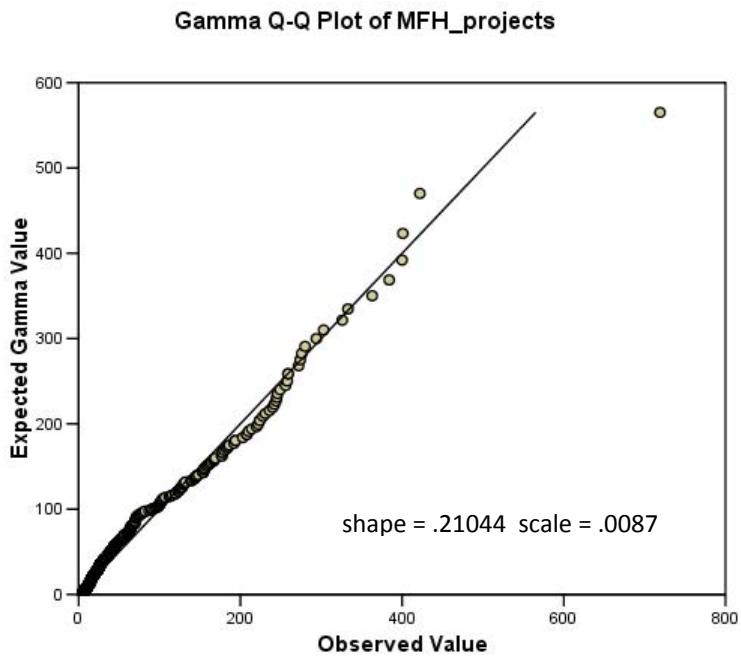## 6.1 Size Distributions of SFH and MFH Projects

Basic descriptive statistics of home developers and housing projects in the Portland region from 2000 to 2007 are discussed in Chapter 2. In this chapter, I focus on the size distributions of SFH and MFH projects, which are important inputs to the location choice models described in Chapter 7. This synthesis is needed to convert the forecast new housing starts, discussed in Chapter 5 into forecast projects by size and type.

To explore their probability distributions, the sizes of all SFH and MFH projects in the region are treated as continuous variables, and the QQ plot function in the statistical software "SPSS" is used to test their distribution. The tests show that Gamma distribution fits the size distributions of SFH and MFH projects best among many probability distributions that have been tried. The gamma distribution is a two-parameter family of continuous probability distributions. It has a shape parameter k and a scale parameter θ. The equation defining the probability density function of a gamma-distributed random variable x is:

$$f(x, k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)} \text{ for } x \geq 0 \text{ and } k, \theta > 0 \tag{16}$$

Gamma Q-Q Plot of SFH_projects

shape = .13817 scale = .04488

Figure 14 Gamma Q-Q plot for the size distribution of SFH projects (2000-2007)

Data source: author's computation based on the building permit data from Metro (2000-2007a)

Gamma Q-Q Plot of MFH_projects

shape = .21044 scale = .0087

Figure 15 Gamma Q-Q plot for the size distribution of MFH projects (2000-2007)

Data source: author's computation based on the building permit data from Metro (2000-2007a)

Here, the random variable x represents project size. Figure 14 and 15 shows the Gamma Q-Q plots for the sizes of SFH and MFH projects from 2000 to 2007 in the region.

6.2 Synthesizing SFH and MFH Projects

To synthesize SFH and MFH projects in a forecast year based on their size distributions in previous years, the following information is needed:

- Total amount of new SFH/MFH units in the forecast year;

- Size distributions of SFH and MFH projects (specifically, the shape and scale parameters of the Gamma distribution);

- Minimum and maximum sizes of SFH/MFH projects in the forecast year;

- Total number of SFH/MFH projects in the forecast year.

Total amount of the housing units in the forecast year can be estimated by the time series model developed in Chapter 5. Since the size distributions of SFH and MFH projects are stable across years, the shape and scale parameters of the Gamma distribution can be estimated based on housing project sizes in previous years. According to their definitions, the minimum sizes for SFH and MFH projects are 1 unit and 2 units respectively. The largest sizes for SFH and MFH projects in previous years can be used as the maximum sizes for SFH and MFH projects in the forecast year. With these parameters, one can generate many sets of projects with different number of projects in it. In order to control the number of synthesized projects in the forecast year and make it more realistic, the numbers of SFH and MFH projects in the forecast year are estimated by dividing the total number of new SFH/MFH units by their mean sizes in previous years. In order to make

the synthesis models converge quickly, a tolerance number is set for the total number of

SFH/MFH projects in the forecast year. In this study, the tolerance is set as ± 5. For

example, if the estimated number of projects is*n*, the model will converge when a set of

projects is generated in which the number of projects falls into the range n ± 5.

Since synthesized housing projects are generated randomly, the model results will not be

exactly the same every time the model is run. However, since each set of projects

synthesized by the same model is subject to the same constraints, such as the total

number of housing units, minimum and maximum project sizes, probability distribution,

and number of the projects, they tend to be very similar to each other.

6.2.1 SFH Project Synthesis Model

Figure 16 compares SFH project size distributions in each year from 2000 to 2007. From

the figure, we can see that SFH project size distributions are quite stable across years,

establishing a foundation for SFH project synthesis in a forecast year based on their

probability distribution in previous years.
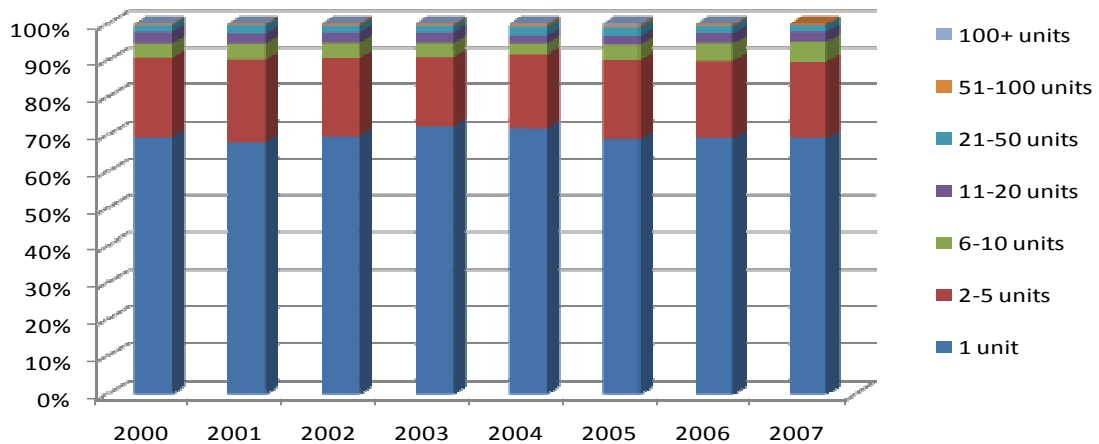
Figure 16 Size distribution of SFH projects (2000-2007)
Data source: author's computation based on the building permit data from Metro (2000-2007a)

Table 12 shows three proposed synthesis models that use 2000-2006 SFH project data to synthesize SFH projects in 2007. In Model 1, the total number of new SFH units in 2007 is the observed number showed by 2007 housing permit data. Gamma distribution parameters were estimated based on SFH projects from 2000 to 2006 and the plot is displayed in Figure 17. The minimum and maximum SFH project sizes in 2007 are the minimum and maximum project sizes revealed by the descriptive analysis on SFH project data from 2000 to 2006. The total number of SFH projects in 2007 is estimated by dividing the total number of housing units in 2007 by the mean SFH development project size from 2000 to 2006. Since there is a sudden reduction of the mean project size in 2007, the model tends to underestimate the number of housing projects in 2007 (2244 estimated vs 2530 observed).

Model 2 is similar to Model 1, but it only synthesizes SFH projects with 2 or more units. SFH projects with only 1 unit are assumed to account for 70 percent of all SFH projects in 2007, which is based on the observation of their proportions in all SFH projects from 2000 to 2006 shown in Figure 16. Model 3 makes the same assumption, but in Model 3 SFH project sizes are transformed into natural log.

The size distributions of SFH projects synthesized by the three models are showed in Figure 18. The observed size distribution of SFH projects in 2007 is also shown in Figure 18 as a benchmark to measure the performance of the three SFH project synthesis models. As indicated by Figure 18, compared to the size distribution of observed SFH projects in 2007, Model 1 tends to overestimate SFH projects with 1 unit and underestimate the SFH projects with 2 units. As mentioned above, in Model 2, only the SFH projects with 2 or

	Total	Gamma distribution		Project size		Estimated No. of
	SFH units	Shape	Scale	Min	Max	projects and tolerance
SFH Project Synthesis Model 1	7002	0.133	0.043	1	267	2244±5
SFH Project Synthesis Model 2[1]	5431	0.31	0.039	2	267	673±5
SFH Project Synthesis Model 3[1]	5431	2.541[2]	1.712[2]	2	267	673±5
1. projects with size 1 is not simulated and assumed to account for 70% of the total number of SFH projects						
2. the shape and scale parameters for simulation model 3 are etimated based on the data in natural log						

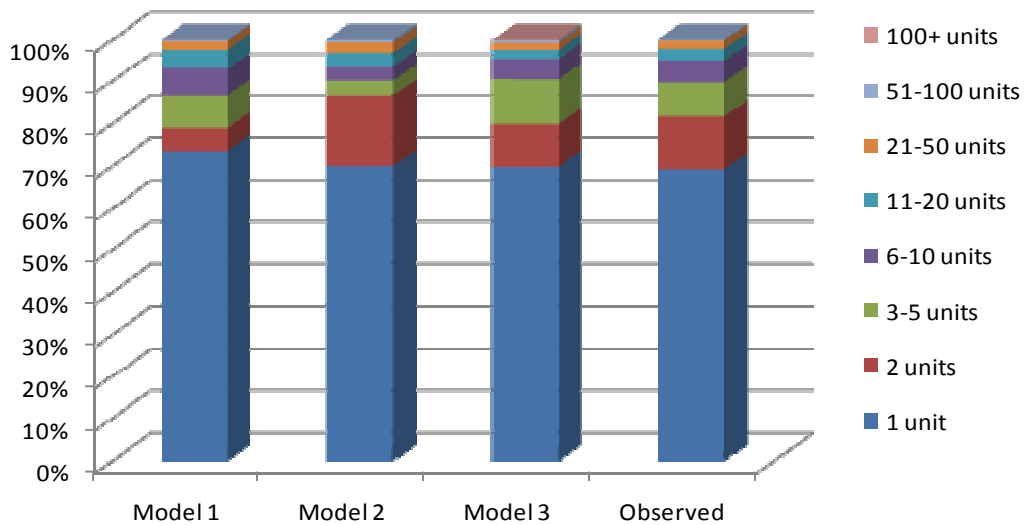**Table 12** SFH project synthesis models

Data source: author's computation based on the building permit data from Metro (2000-2006a)
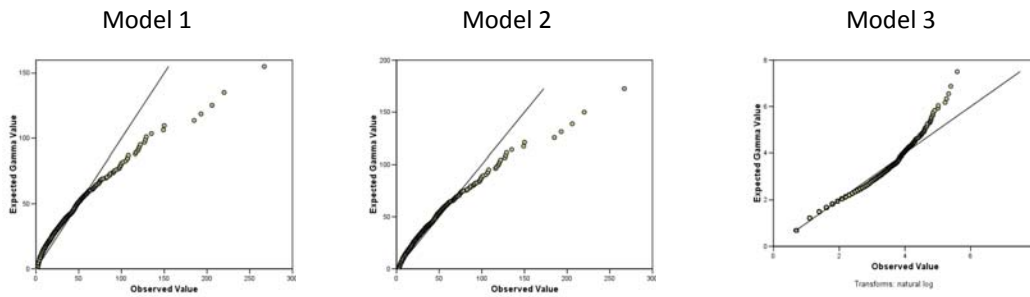
Figure 17 Gamma distribution of SFH project sizes for the three models (2000-2006)

Data source: author's computation based on the building permit data from Metro (2000-2006a)

Figure 18 Size distributions of synthesized and observed SFH projects in 2007

more units are synthesized. Figure 18 shows that Model 2 significantly overestimates the number of SFH projects with 2 units. Compared to Models 1 and 2, size distribution of SFH projects synthesized by Model 3 is much closer to the observed SFH projects in 2007, indicating that this model has the best performance in the three models. Thus this model is selected as the final model and SFH projects synthesized by this model are used as forecasting units for the SFH location choice models in Chapter 7.

6.2.2 MFH Project Synthesis Model

Figure 19 shows the size distributions of MFH projects in each year from 2000 to 2007 in the region. Compared to the size distribution of SFH projects, the size distribution of MFH projects is lumpier across years, making the synthesis of MFH projects for a forecast year based on their size distributions in previous years more difficult.

Descriptive analysis shows that there were only 9 MFH projects whose sizes were larger than 300 units from 2000 to 2006, so they are treated as outliers and not counted while calculating the mean size of MFH projects and estimating Gamma distribution parameters in the MFH project synthesis models. The maximum project size in the forecast year is set to 300 units.

Table 13 shows the three synthesis models that use 2000-2006 MFH project data to synthesize MFH projects in 2007. Model 1 is the base model. The total number of new MFH units is the observed number in 2007. The number of MFH projects is calculated by dividing the total new MFH unit in 2007 by the mean size of MFH projects from 2000 to 2006. Similar to SFH projects, the mean size of MFH projects suddenly dropped in 2007,

making the estimated number of MFH projects in 2007 smaller than the observed number (111 vs 139).

Model 2 is different from Model 1 in that MFH project sizes were transformed into natural log form while estimating the shape and scale parameters for the Gamma distribution. Model 3 also uses log transformation, but different from Model 2, MFH projects with 2 units are not synthesized in Model 3. Their proportion in the total number of MFH projects in 2007 was assumed to be 25 percent, as observed from previous years in Figure 19.

Figure 21 compares the size distributions of MFH projects synthesized by the three models and observed in 2007. As the figure shows, the size distribution of MFH projects synthesized by Model 3 is closest to the size distribution of observed MFH projects in 2007. Thus Model 3 is selected as the MFH project synthesis model and MFH projects synthesized by Model 3 are used as forecasting units for the MFH location choice models in Chapter 7.
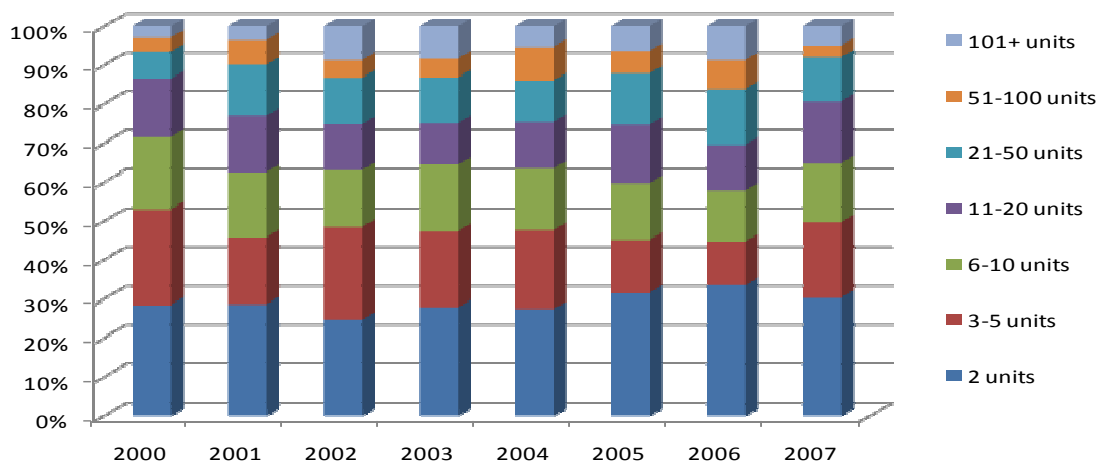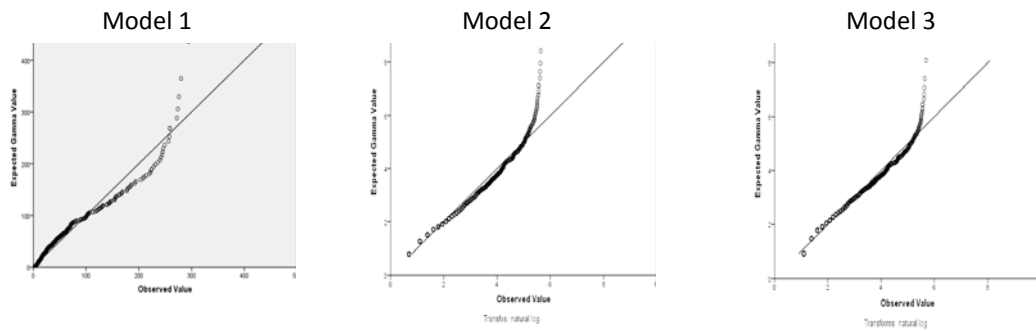

Figure 19 Size distribution of MFH projects (2000-2007)

Data source: author's computation based on the building permit data from Metro (2000-2007a)

	Total	Gamma distribution		Project size		Estimated No. of
	MFH units	Shape	Scale	Min	Max	projects and tolerance
MFH Project Synthesis Model 1	2458	0.262	0.012	2	300	111±5
MFH Project Synthesis Model 2[1]	2458	2.445	1.184	2	300	111±5
MFH Project Synthesis Model 3[1]	2402	4.988	1.903	3	300	83±5
1. the shape and scale parameters for simulation model 2 and 3 are etimated based on the data in natural log						
2. projects with size 2 is not simulated and assumed to account for 25% of the total number of MFH projects						

Table 13 MFH project synthesis models

Figure 20 Gamma distribution of MFH project sizes for the three models (2000-2006)

Data source: author's computation based on the building permit data from Metro (2000-2006a)
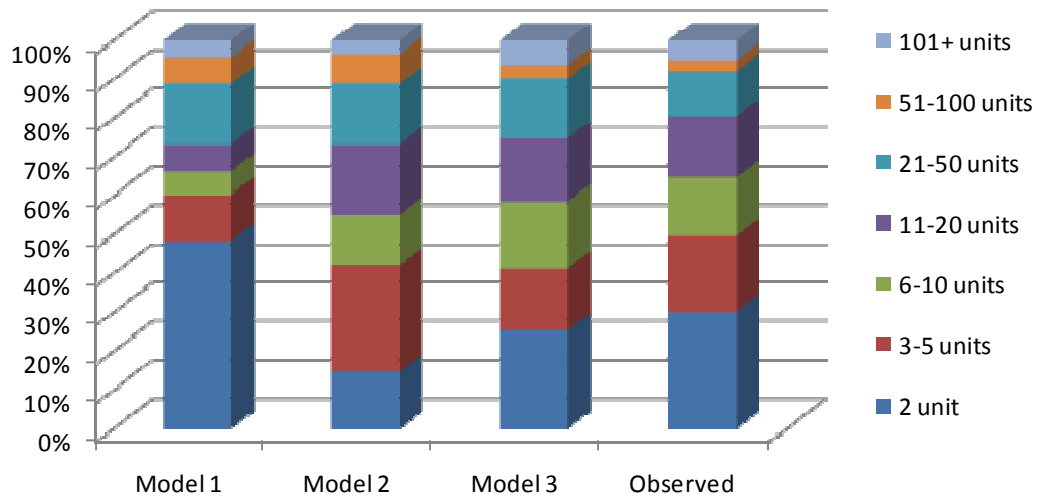
Figure 21 Size distributions of synthesized and observed MFH projects in 2007

112

6.3 Summary

SFH and MFH project synthesis models were developed to convert new housing start forecast into projects by size and type. Models based on a Gamma distribution provided the best fit to the data, following a log transformation. Additionally, projects with just one housing unit are synthesized separately.

CHAPTER 7 FORECASTING THE LOCATION OF NEW HOUSING: A COMPARISON OF THREE APPROACHES

7.1 Proposed New Housing Location Choice Models

The comprehensive new housing supply and location choice model proposed by this study follows a top-down perspective. The highlighted boxes in Figure 22 reflect three alternative paths to allocating housing units in space, using location choice sub-models, the subject of this Chapter.
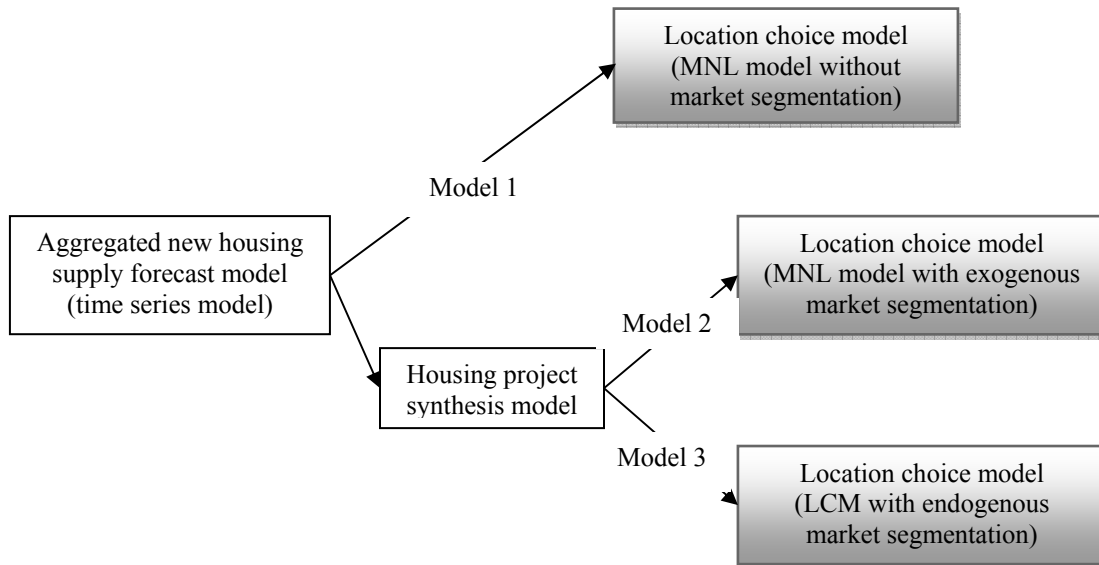
Figure 22 Three proposed location choice models within the large model system

As indicated by Figure 22, the main differences among the three location choice models lie in whether and how they handle developer taste heterogeneity and decision variables.

Model 1 is a MNL model and developer heterogeneity is ignored. The housing market is assumed to be homogenous and each housing unit is treated the same and distributed across the region as an individual unit. Thus, I essentially atomize development projects, ignoring the size of multi-unit housing projects when forecasting location choices. Model 1 is similar to the housing supply model in ILUTE (2006). The difference is that when estimating Model 1, project size was used as the weighting variable which, as discussed below, resulted in a significantly better fit to the data.

Different from Model 1, Models 2 and 3 require housing projects synthesized in Chapter 6, thus providing a more realistic depiction of land requirements for multi-unit projects. As suggested by the findings from Chapter 4, developers with different sizes have different attribute preferences when choosing locations for their projects. In addition, availability of vacant land is considered to accommodate the entire housing project, when evaluating the utility of project location alternatives.

Model 2 is also a MNL model, but differs from the first model option in that housing projects are categorized into three size groups: small SFH projects consisting of just one housing unit and small MFH projects consisting of two to five housing units; medium-size SFH projects of 2-5 units and medium-size MFH projects with six to twenty units; and large-size SFH projects of 6 or more units and large-size MFH projects with twenty more units. The single-house (small) developers are designed to represent custom-built homes and private owner/developers. Projects in the medium range are intended to represent the majority of multi-unit projects, whereas larger projects are much less common, but account for the majority of new homes, as described in the next section. In

115

Model 2, systematic heterogeneity is modeled through endogenous deterministic segmentation of predictor variables by these project size indicator variables.

In Model 3, market segmentation is determined endogenously and stochastically using a latent class model approach which has been discussed in Chapter 4. In the latent class models developed in this chapter, project size is used as the class membership variable. After testing for the existence of two, three and four latent classes, a three-class specification was chosen based on model fit and interpretability.

The simulation of the location choice for new SFH and MFH projects follows two steps. First, the SFH/MFH location choice models are estimated with the data from 2000 to 2006, and the 2007 data is set aside as a holdout sample. Secondly, new housing in 2007 is distributed into zones based on the coefficients from model estimation results and simulation results are compared to observed spatial distribution of new housing indicated by the 2007 holdout sample to test the forecasting effectiveness of the models.

Descriptions of the predictor variables used in models may be found in Table 2 in Chapter 2. Model estimations, including the determination of choice sets and the use of predictor variables, are based on the experiences and findings from Chapter 3 and Chapter 4. One major difference between the models estimated in this chapter and the models developed in previous chapters is the measurement of lagged effects. In previous chapters, lagged effects are measured by the dummy variables indicating if a developer had development in a TAZ or its neighbor TAZs in previous years, however, in the forecast model, information about developers' activities in previous years is not predicted. As a proxy, I measure lagged effects from locations' perspective. If a location had SFH

or MFH development in the previous year, then the lag indicator variable equals 1, and equals 0 otherwise.

7.2 Model Estimation Results

Each model was tested in an iterative way and only significant coefficients were retained in the final models for the forecast purpose. Estimation results for SFH models are presented in Tables 14 and 15, and estimation results for MFH models are presented in Tables 16 and 17.

As indicated by Table 14 and 16, for both SFH and MFH, Models 1 has a better goodness of fit than Models 2 in terms of adjusted pseudo R^2 (0.142 vs. 0.133 for SFH, 0.181 vs. 0.088 for MFH), despite having nearly half the number of estimated parameters, because it was estimated using project size as a weighting factor. Accordingly, a large project with 100 units contributed the same amount of information to the likelihood computations as 100 projects with just one unit. The theoretical justification being that developers of larger projects have many more constraints to overcome in their choices, not only in terms of land assembly but also development costs and project complexity. Strictly speaking, however, for both SFH and MFH, the fit of Models 1 and 2 should not be compared directly due to the difference in estimation methods. The adjusted pseudo R^2 for SFH and MFH Models 3, the latent class models are even smaller, 0.094 for SFH Model 3 and 0.074 for SFH Model 3; however, these results are not comparable to Models 1 and 2 because of the different choice set sampling procedure and number of alternatives.

| | Model 1 | | Model 2 | | | | | |
| | | | Main effects | | Mid-size projects | | Large projects | |
Variables	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value
Location relative to the UGB:								
Within the UGB	--	--	--	--	--	--	--	--
On the UGB periphery	-0.2568	-6.03	0.1614	3.39	-0.1489	-1.96	--	--
Clark County dummy	0.4962	8.13	0.1974	3.37	--	--	--	--
Outside the UGB	0.2170	3.78	1.1707	23.65	--	--	-0.6370	-3.84
Clark County dummy	-0.3098	-2.88	--	--	-0.9983	-6.01	-3.9799	-5.39
Buildable vacant land for SFH:	0.1265	28.75	0.0818	20.33	0.0786	9.39	0.0625	4.06
Clark County dummy	-0.0252	-3.10	--	--	--	--	-0.0684	-2.83
Deflated land value (in 2000$)	-0.0246	-4.73	-0.0404	-7.47	--	--	--	--
Clark County dummy	-0.0715	-5.75	--	--	--	--	--	--
Accessibility:								
Auto for retail employment	-0.0884	-19.09	0.0148	3.23	--	--	--	--
Clark County dummy	0.2320	11.10	0.0469	3.25	--	--	--	--
Auto for non-retail emp	0.0121	11.74	-0.0065	-6.48	--	--	--	--
Clark County dummy	-0.0445	-9.92	-0.0074	-2.41	--	--	--	--
Transit for retail emp	0.0627	9.31	-0.0172	-2.81	--	--	--	--
Clark County dummy	-0.2049	-7.17	--	--	--	--	--	--
Transit for non-retail emp	-0.0162	-11.05	0.0045	3.34	--	--	--	--
Clark County dummy	0.0485	7.65	--	--	--	--	--	--
SFH net density dummy:								
No SFH	--	--	--	--	--	--	--	--
Low SFH density	0.8840	9.93	0.2671	4.07	0.7573	2.23	--	--
Clark County dummy	-1.8288	-12.18	--	--	1.0251	6.38	--	--
Medium SFH density	1.5783	17.54	0.5863	8.24	2.0507	6.23	1.3585	9.02
Clark County dummy	-1.7535	-12.45	--	--	--	--	-0.6846	-4.06
High SFH density	2.6705	25.23	1.0228	10.37	2.2095	6.39	1.8551	8.11
Clark County dummy	-2.3100	-10.68	-0.9547	-4.14	--	--	--	--
Housing diversity dummy:								
Mixed housing	--	--	--	--	--	--	--	--
Dominated by SFH	-0.1760	-7.11	-0.1844	-6.90	0.2212	4.54	--	--
Clark County dummy	0.5576	9.91	0.1596	3.65	--	--	--	--
Dominated by MFH	0.1846	4.95	-0.6442	-10.74	0.3168	2.85	1.0455	9.33
Clark County dummy	0.3009	3.15	--	--	--	--	--	--
Mixed use dummy:								
Mixed use:	--	--	--	--	--	--	--	--
Dominated by residence	0.2710	11.32	0.3026	10.50	0.1383	2.79	-0.2083	-2.77
Clark County dummy	-0.3467	-8.10	-0.5183	-12.19	--	--	0.4080	2.91
Dominated by non-residence	--	--	-0.0662	-2.24	0.2193	3.99	--	--
Clark County dummy	0.3417	8.24	-0.2810	-5.10	--	--	0.8351	6.68
Road density	--	--	0.0022	7.18	--	--	-0.0044	-5.74
Clark County dummy	--	--	-0.0045	-10.73	--	--	0.0065	5.37
Household income	-0.0037	-4.42	0.0134	16.40	-0.0110	-6.80	-0.0152	-6.53
Clark County dummy	0.0067	4.07	--	--	--	--	--	--
% of household head aged 25-54	0.0125	9.74	-0.0057	-3.67	0.0113	3.61	0.0340	7.95
Clark County dummy	--	--	--	--	--	--	--	--
Lagged effect from previous year	1.3890	56.80	1.3467	52.03	0.1383	2.28	--	--
No. of parameters (K)	33		61					
Log likelihood at convergence	-41837		-42262					
Log likelihood with constant only	-48795		-48795					
Psuedo R^2	0.143		0.134					
Adjusted Psuedo R^2	0.142		0.133					
Weighting variable	Project size		N/A					
Sample size	16367		16367					

Table 14 Estimation results from SFH Models 1 and 2 (2000-2006)

Variables	Segment 1		Segment 2		Segment 3	
	Coef	t-value	Coef	t-value	Coef	t-value
Segment-specific location choice model						
Location relative to the UGB:						
Within the UGB	--	--	--	--	--	--
On the UGB periphey	0.5302	7.83	-0.7618	-15.20	0.8299	12.96
Outside the UGB	-0.1405	-0.61	0.0566	0.92	1.4304	21.55
Buildable land for SFH	0.5894	18.69	0.2544	39.36	-0.0467	-7.31
Clark County Dummy	-0.1842	-4.59	0.0630	6.11	-0.3080	-16.49
Land value (deflated, in 2000$)	0.1916	11.96	-0.0459	-5.44	-0.0256	-3.00
Clark County Dummy	-0.1317	-3.89	0.0411	2.33	-0.1271	-3.61
Accessibility:						
Auto for retail emp	0.2007	11.55	-0.0427	-4.12	-0.1529	-18.06
Clark County Dummy	-0.2817	-4.77	0.1379	4.82	0.2075	8.25
Auto for non-retail emp	-0.0309	-9.59	-0.0024	-1.09	0.0248	13.74
Clark County Dummy	0.0407	3.28	-0.0195	-3.18	-0.0366	-6.75
Transit for retail emp	0.0950	5.45	-0.0895	-8.57	0.1061	11.28
Clark County Dummy	-0.3041	-3.64	0.0020	0.06	0.1213	3.42
Transit for non-retail emp	0.0676	3.67	0.0022	0.27	-0.0285	-3.54
Clark County Dummy	-0.0232	-6.10	0.0207	8.71	-0.0195	-9.30
SFH net density	-0.1393	-4.76	0.1451	17.22	0.0578	4.87
Clark County Dummy	-0.0952	-1.93	-0.0551	-3.45	-0.1185	-3.93
Housing diversity	-0.1195	-2.83	0.0028	7.66	-0.0373	-5.44
Clark County Dummy	-8.6220	-3.74	0.0477	10.51	-0.0252	-0.82
Mixed use	0.2536	17.42	-0.4783	-18.16	-0.1699	-9.04
Clark County Dummy	-0.5700	-4.40	0.8255	16.38	0.1487	4.65
Road density	0.0015	1.53	0.0000	-0.04	0.0027	6.00
Clark County Dummy	0.0040	2.71	-0.0017	-2.79	0.0001	0.11
Average household income	0.0303	11.37	-0.0277	-20.05	-0.0039	-3.05
Clark County Dummy	0.0108	2.15	0.0252	10.28	-0.0004	-0.14
% of household head aged 25-54	0.0018	0.39	0.0202	8.86	0.0065	2.89
Clark County Dummy	-0.0417	-5.21	0.0048	1.25	0.0141	3.10
Lagged effect from previous year	1.7283	18.29	1.0773	35.92	1.2254	43.20
Segmentation model						
Constant	-2.1194	-17.61	-2.2897	-19.28	--	--
Project size	1.5005	13.94	1.6280	15.09	--	--
Class size (%)	28.5		35.3		36.2	
No. of parameters (K)	85					
Log likelihood at convergence	-67839					
Log likelihood with constant only	-74998					
Psuedo R^2	0.095					
Adjusted Psuedo R^2	0.094					
Weighting variable	N/A					
Sample size	16367					

Table 15 Estimation results from SFH Model 3: latent class model (2000-2006)

Variables	MFH Model 1		MFH Model 2 Main effects (Small size)		Interaction with medium size dummy		Interaction with large size dummy	
	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value
Buildable land for MFH	-0.2616	-17.57	-0.1711	-12.03	--	--	-0.0766	-2.39
Average residential land value	0.0162	4.32	-0.0143	-1.96	--	--	0.0325	3.02
Accessibility:								
Auto for retail emp	-0.0468	-2.53	0.0240	1.84	-0.0380	-1.85	-0.0649	-2.92
Auto for non-retail emp	0.0110	2.62	-0.0046	-1.63	0.0084	1.94	0.0145	3.08
Transit for retail emp	-0.0566	-2.76	--	--	--	--	--	--
Transit for non-retail emp	0.0132	2.92	--	--	--	--	--	--
MFH density:								
No MFH in TAZ	--	--	--	--	--	--	--	--
Low MFH density	--	--	0.8229	3.97	-0.3820	-1.57	--	--
Medium MFH density	--	--	0.7767	3.91	--	--	--	--
High MFH density	0.3975	5.08	0.9118	4.30	--	--	--	--
Housing diversity:								
MFH and SFH mixed	--	--	--	--	--	--	--	--
Dominated by SFH	--	--	-0.3059	-2.24	--	--	--	--
Dominated by MFH	0.7002	9.63	-0.9002	-6.76	0.6251	3.30	1.3206	6.93
Mixed use:								
Mixed use:	--	--	--	--	--	--	--	--
Dominated by residentical	0.3082	3.74	--	--	--	--	--	--
Dominated by non-residential			--	--	--	--	--	--
Road density	-0.0016	-2.72	0.0015	1.78	-0.0040	-3.06	-0.0027	-1.90
Average household income	0.0056	2.14	-0.0227	-5.10	0.0208	3.25	0.0339	5.41
Lagged effect from previous year	1.1992	17.78	1.0699	16.65	--	--	--	--
No. of parameters (K)	12		25					
Log likelihood at convergence	-2814		-3121					
Log likelihood with constant only	-3450		-3450					
Psuedo R^2	0.184		0.095					
Adjusted Psuedo R^2	0.181		0.088					
Weighting variable	project size		N/A					
Sample size	1191		1191					

Table 16 Estimation results from MFH Models 1 and 2 (2000-2006)

Variables	Segment 1		Segment 2		Segment 3	
	Coef	t-value	Coef	t-value	Coef	t-value
Segment-specific location choice model						
Buildable land for MFH	-1.1003	-11.00	0.2743	11.64	-0.0631	-1.10
Average residential land value	-0.0383	-2.04	-0.0262	-3.05	-0.0795	-1.46
Accessibility:						
Auto for retail emp	-0.0239	-0.29	-0.0689	-2.56	-0.7016	-6.87
Auto for non-retail emp	0.0031	0.19	0.0203	3.64	0.1371	6.16
Transit for retail emp	-0.0270	-0.36	0.0991	3.81	0.3503	3.79
Transit for non-retail emp	0.0085	0.53	-0.0214	-3.79	-0.0900	-4.21
MFH density	0.0036	0.79	0.0075	3.52	0.0476	3.73
Housing diversity	0.0009	1.05	-0.0075	-5.34	0.0081	7.91
Mixed use	-0.0314	-0.28	-0.2240	-5.52	-3.3827	-7.26
Road density	0.0025	1.46	0.0026	4.46	-0.0123	-2.96
Average household income	-0.0023	-0.32	-0.0226	-7.72	-0.0182	-1.25
% of household head aged 25-54	-0.0007	-0.07	0.0027	0.68	0.0912	3.18
Lagged effect from previous year	0.4706	2.27	0.9234	16.51	3.9684	9.83
Segmentation model						
Constant	2.5030	6.63	3.8228	10.53	--	--
Project size	-0.0081	-3.17	-0.0170	-5.63	--	--
Class size (%)	25		68.6		6.4	
No. of parameters (K)	43					
Log likelihood at convergence	-4687					
Log likelihood with constant only	-5110					
Psuedo R^2	0.083					
Adjusted Psuedo R^2	0.074					
Weighting variable	N/A					
Sample size	1191					

Table 17 Estimation results from MFH Model 3: latent class model (2000-2006)

Since the focus of this chapter is to compare the models' forecast capabilities, detailed discussion of estimation results that has been undertaken in Chapter 3 and 4 is omitted. It is sufficient to mention here, that all three models are specified with the same basic set of explanatory variables for the same observations, with specification differences as the results of different forms of market segmentation, which in Models 2 and 3 resulted in considerably more estimated parameters.

7.3 Location Choice Forecast Procedures

The location choice forecast applies the model estimation results in Tables 14-17. The first step is to determine the full choice set available to each project. To be consistent with the model estimation, for each observation unit only the TAZs with enough buildable land are eligible as candidate TAZs in their choice set, assuming that each SFH unit consumes at least 0.1 acre and each MFH unit consumes at least 0.01 acre buildable land. In SFH/MFH Models 1, each forecasting unit is a housing unit, and the total number of the housing units is the total amount of the new supply indicated by the permit data. Since each housing unit is treated exactly the same, they shares the same full choice set. In SFH/MFH Models 2 and 3, the forecasting units are the synthesized projects and the choice sets faced by them differ according to size.

The second major step is to calculate the utility and probability of each candidate TAZ in the choice set, based on the coefficients shown in Tables 14-17. In SFH/MFH Model 1, all the forecasting units share the same coefficients and choice sets, so the probabilities for candidate TAZs to be chosen by each of them are the same. In SFH/MFH Model 2, the synthesized projects are categorized into three size groups and the synthesized

projects in each size group share the same coefficients. In SFH/MFH Model 3, the latent class approach, the assignment of each synthesized project to the three market segments is determined by the segmentation model. The utilities and probabilities for each project to be in each segment are calculated first. Then each project is randomly assigned to one segment based on the probabilities. The probabilities of candidate TAZs in the choice set to be chosen by each project are calculated based on the segment-specific location choice model.

In the third step, the forecasting units are allocated to a TAZ in their choice sets based on the probabilities calculated and using Monte Carlo draws. With each housing unit or synthesized project located into a TAZ, the amount of the buildable land in that TAZ is reduced by the amount of land consumed. Once the buildable land in a TAZ is exhausted, the TAZ is no longer eligible for the unplaced housing units or projects.

After locating all housing units and synthesized projects into TAZs, the total number of new SFH/MFH units in each TAZ is calculated by aggregating all SFH/MFH housing units allocated to it. Since the sizes of most TAZs are very small, it would not be realistic to expect the models to predict the amount of the new housing unit accurately at that level. Thus, the forecast results from 1,998 TAZs were aggregated into 214 original census tracts in the region. Original census tracts in the region are obtained by aggregating subdivisions of census tracts into one. For example, census tracts 3.01 and 3.02 are aggregated into one original tract 3.

7.4 SFH Forecast Results Comparison

Observed new SFH units and the forecasted new SFH units in each original census tract are represented by graduated circles in Figures 22, 23, 24, 25, respectively. All three models are able to predict the basic spatial distribution of SFH development in 2007, but some are closer to the observed pattern than others.

In order to compare the forecast results from the three models, I calculate root mean square errors (RMSEs) for each model using the formula

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Forecast_i - Observed_i)^2}{N}}$$ (17)

where $Forecast_i$ is the number of the housing units in original census tract i predicted by the model, $Observed_i$ is the number of the housing units in each original census tract i observed from the housing permit data, and N is the total number of the original census tracts. The RMSEs for SFH Model 1, 2, and 3 are 39.3, 66.8, and 49.9, respectively, indicating that overall SFH Model 1 has the most accurate forecasts and SFH Model 2 is least accurate.

To make the comparison more intuitive and obtain a spatial perspective, the differences between the number of the housing units predicted by the three SFH forecast models and the observed number in each original census tract are illustrated in Figures 27, 28, and 29 separately. Triangles represent the census tracts where the numbers of the SFH housing units are underestimated by the forecast models, and circles represent the census tracts where the forecast models over-estimate SFH development. The legend on each map indicates ranges of the differences represented by the symbols. Figure 30 shows major

cities and freeways in the region. Overall, the comparison between these three maps tells a similar story; SFH Model 1 is closest to the observed situation, and SFH Model 2 shows the greatest deviation.

As the figures indicates, a common problem with all three SFH models is that they tend to under-estimate SFH development in Washington County, Oregon, mainly in an urban peripheral area. The permit data shows that only five census tracts (Tracts 315, 316, 317, 318, and 319) in the peripheries of Cities of Beaverton and Tigard produced more than 10 percent of new SFH units in the entire metropolitan area, each year from 2000 to 2007. Thus, this area is an outlier in the region in terms of SFH development, and the predictor variables included in the models fail to capture the rapid growth there.

As Figure 27 indicates, SFH Model 1 under-estimates SFH development in Washington and Clackamas Counties, and over-estimates SFH development in Clark and Multnomah Counties. Multnomah County is home to the City of Portland, and the TAZs in that area tend to have much less buildable vacant land. SFH Model 1 over-allocates development to Multnomah County because individual housing units are used as the forecasting units for the location choice forecast.

Since each housing unit requires only a small amount of vacant land, densely urbanized TAZs are more likely to be included in choice sets than they are in SFH Models 2 and 3. Compared to the other two models, SFH Model 2 tends to over-estimate SFH development in the UGB peripheral areas in Clark and Clackamas Counties, where vast amounts of buildable vacant land are available. At the same time, SFH Model 2 under-predicts development in areas within the UGB, and this is especially true in Washington
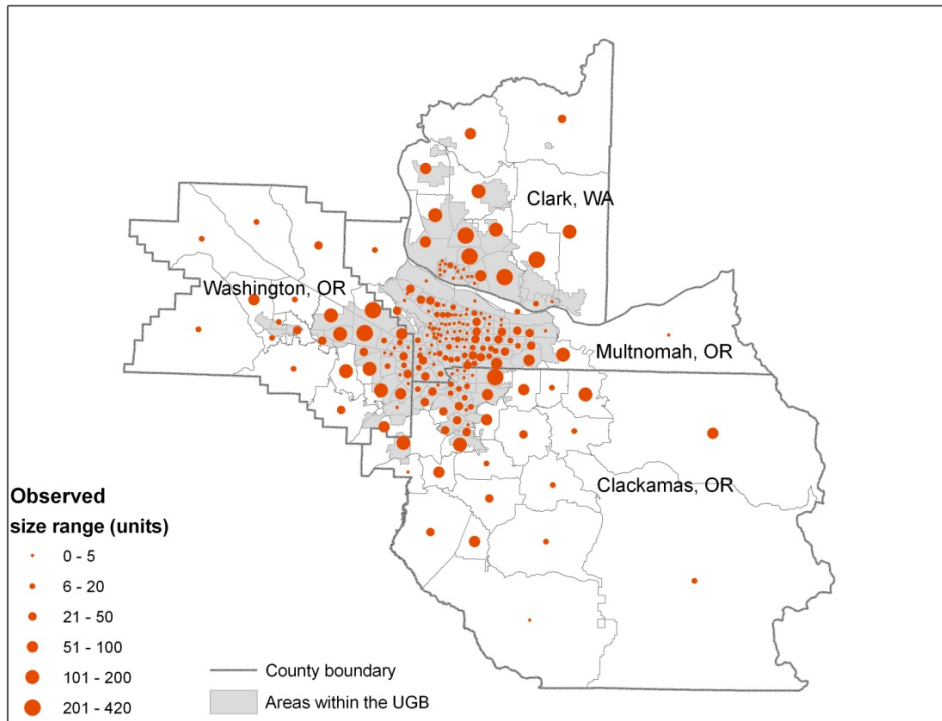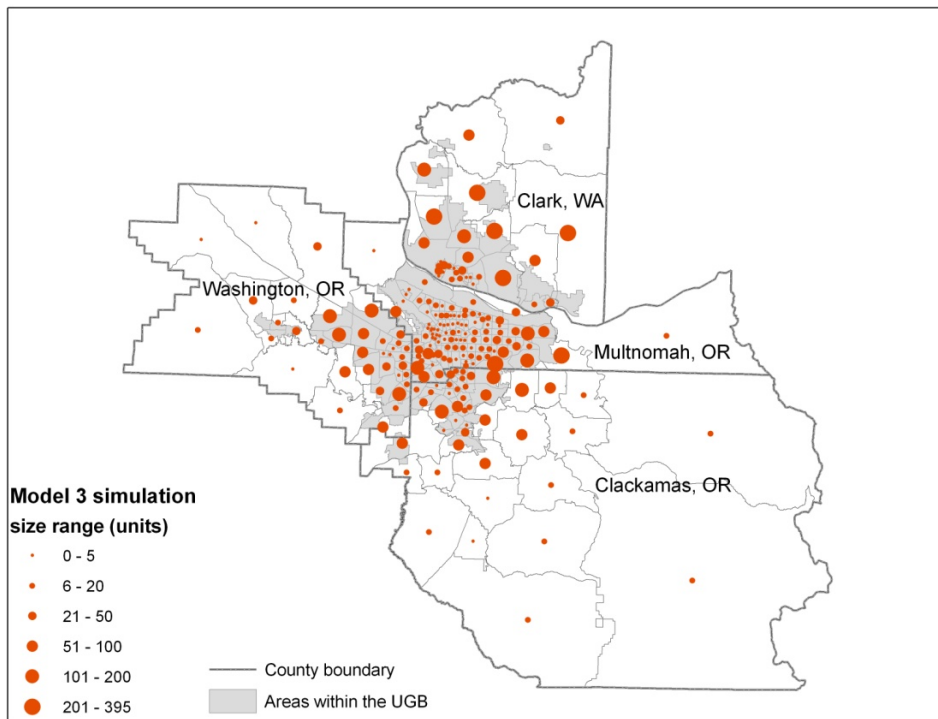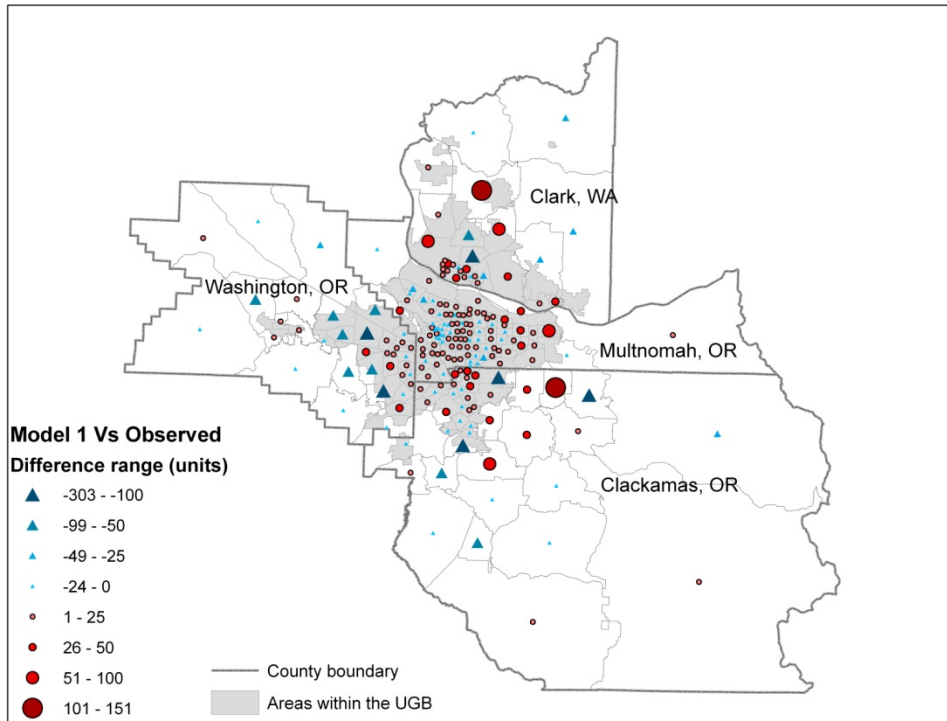
Figure 23 Observed SFH development in 2007
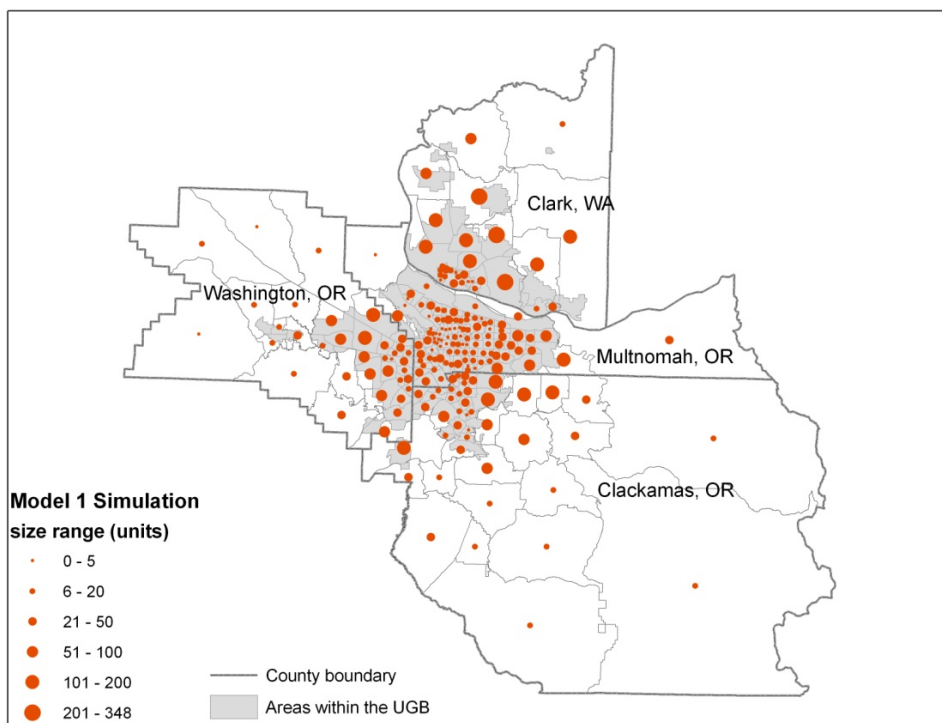
Figure 24 Predicted SFH development by SFH Model 1 in 2007
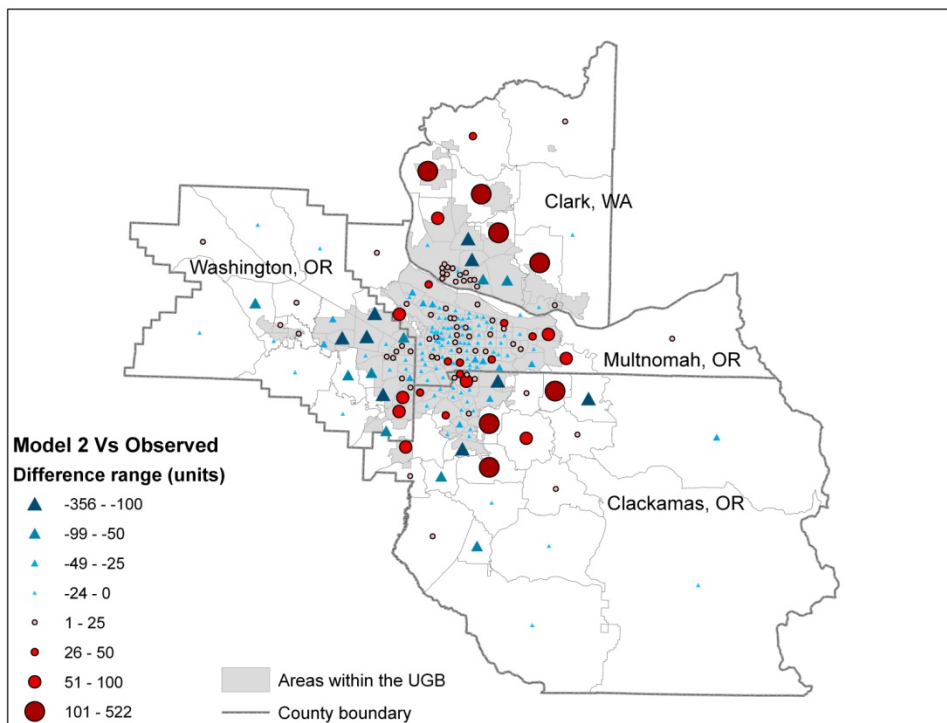
126
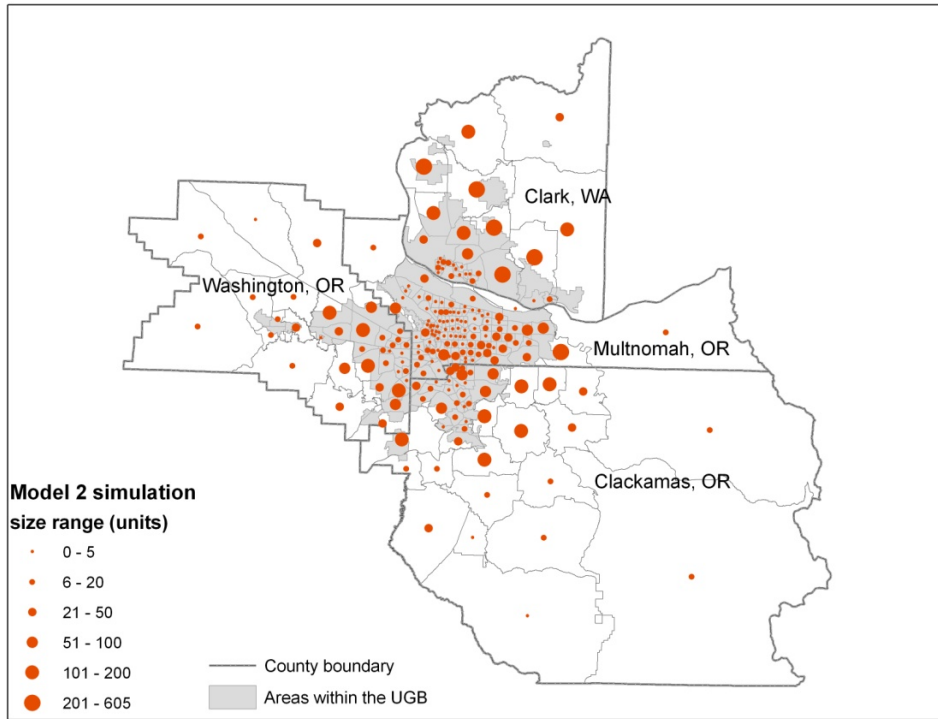
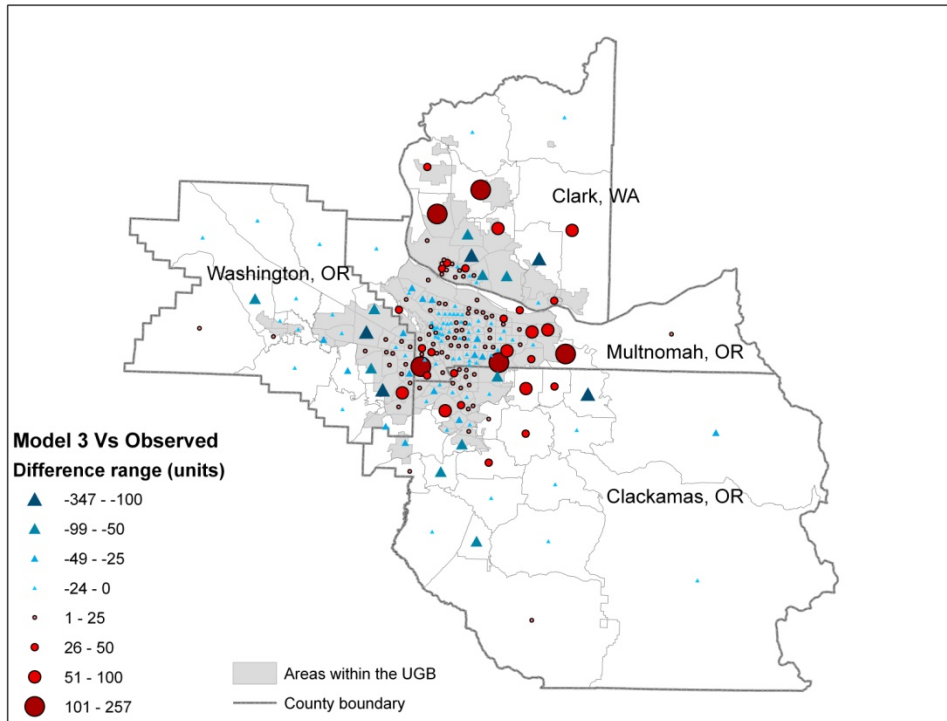**Figure 25** Predicted SFH development by SFH Model 2 in 2007

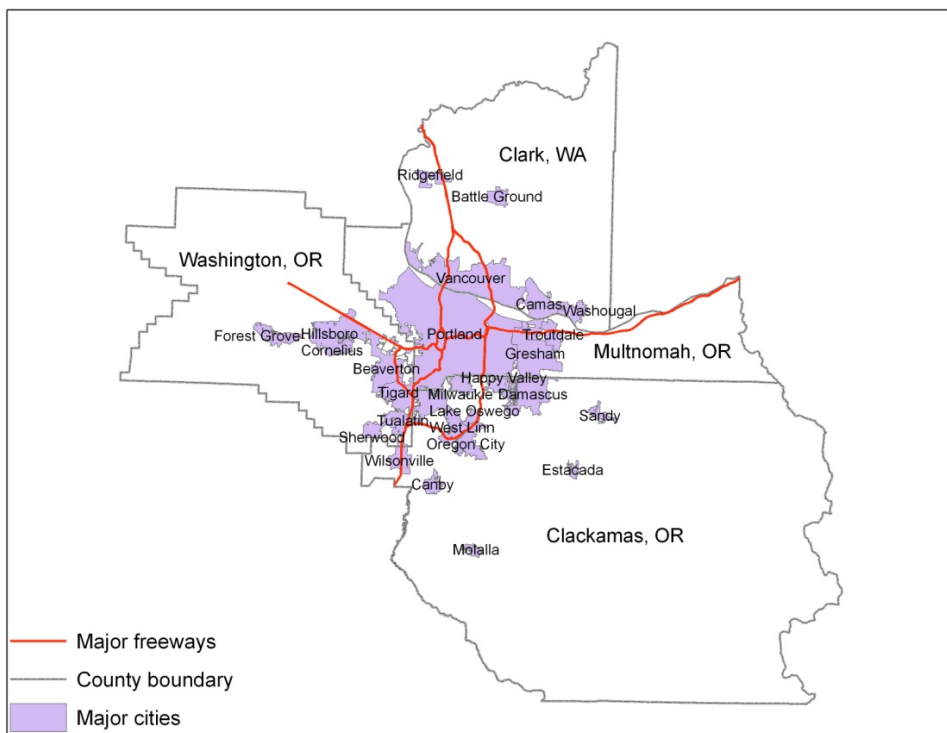**Figure 26** Predicted SFH development by SFH Model 3 in 2007

Figure 27 SFH Model 1 forecast vs observed in 2007

Figure 28 SFH Model 2 forecast vs observed in 2007

Figure 29 SFH Model 3 forecast vs observed in 2007

**Figure 30** Major cities and freeways in the region

129

and Clark Counties. This might indicate that the importance of buildable vacant land has been over-emphasized in the specification of the model. The reason could be due to the omitting of redevelopment, land fragmentation and zoning change while calculating the amount of buildable land in each TAZ. Another more important consideration is the exogenous market segmentation in SFH Model 2 in which the synthesized projects with 6 or more units are categorized as "large-size" projects, such that all projects within this group, whether they have 6 units or 106 units, share the same coefficients estimated by the model. The wide range of project sizes in the large-size project group makes the forecasts inaccurate and slightly unstable. One way to solve this problem would be to create more segments for the larger-size projects, but the number of large-size SFH projects in the Portland housing market is too small to support further segmentation.

Compared with the other two models, SFH Model 3 predicted SFH development most accurately in Washington and Clackamas Counties, but it tends to over-estimate SFH development in Multnomah County, especially in South and Southeast Portland. SFH Model 3 shares a similar issue with SFH Model 2, the difficulty of forecasting locations for a small number of very large projects. Model 3 differs from SFH Model 2, however, because market segmentation was determined endogenously and forecasting allocates the large-size projects into market segments in a probabilistic way. This results in a heterogeneous treatment of preferences. The comparison between Figures 28 and 29 indicate that this difference helps SFH Model 3 returns better forecast result than SFH Model 2, but it still not quite as accurate as SFH Model 1 overall.

An additional issue with SFH Models 2 and 3 is the use of dummy variables. When certain dummy variables were estimated for SFH Models 2 and 3, it was common to obtain non-significant results because it was often the case that projects within a particular market segment, even a latent one, would happen to be located in only the TAZs defined by the dummy variable. For example, when most of the projects in one market segment are located in the TAZs outside the UGB, the UGB dummy variable becomes a problem while estimating the model. In SFH Model 2, this problem can be solved by controlling the market segmentation; however, in SFH Model 3, the latent class model, market segmentation is determined probabilistically and is difficult to control. So in SFH Model 3, I avoided the use of dummy variables.

7.5 MFH Forecast Results Comparison

Figures 31, 32, 33, and 34 shows observed new MFH units and forecasted new SFH units in each original census tract represented by graduated circles. As Figure 31 indicates, most new MFH units in 2007 are in Multnomah County, which is home to the city of Portland. The results from the three MFH forecast models capture this basic trend. However, compared to SFH development, the spatial distribution of MFH development in the region is lumpier, and thus more difficult to predict. Compared to the forecast results by the SFH models, the results from the MFH forecast model show much greater deviations from the observed situation, and are less consistent with each other. To quantify the differences between the forecast results from the three MFH models and the observed situation, RMSEs are calculated for each of them based on equation (17) at the
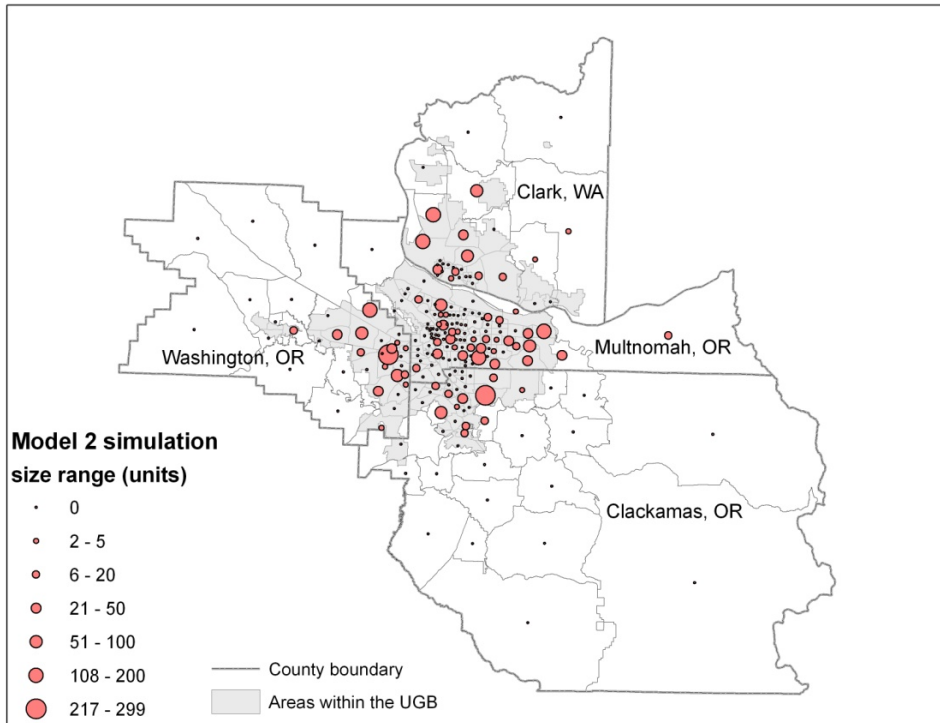
131
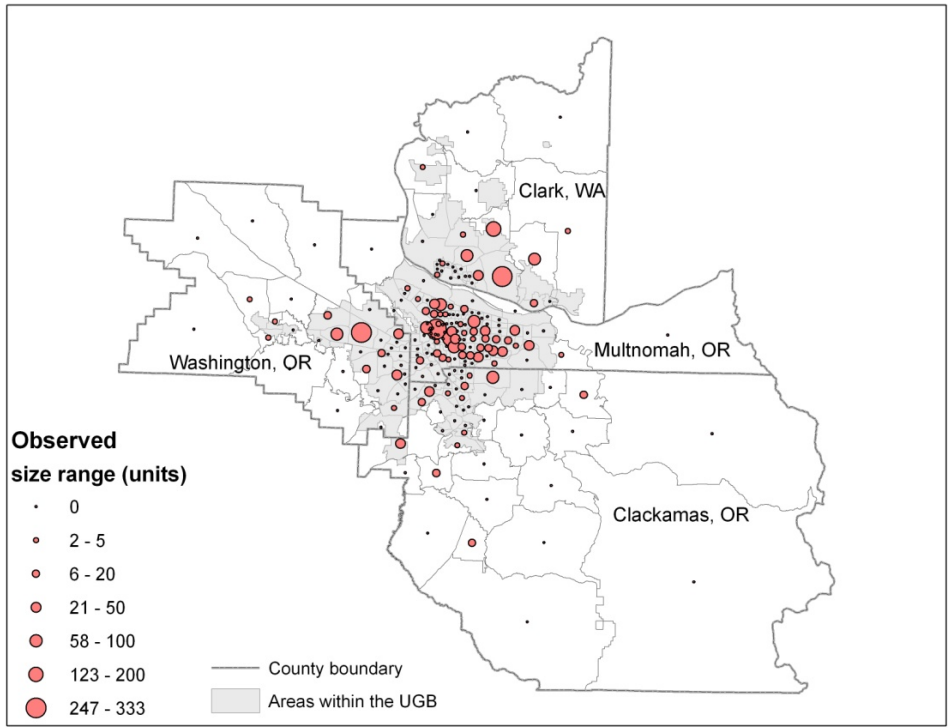
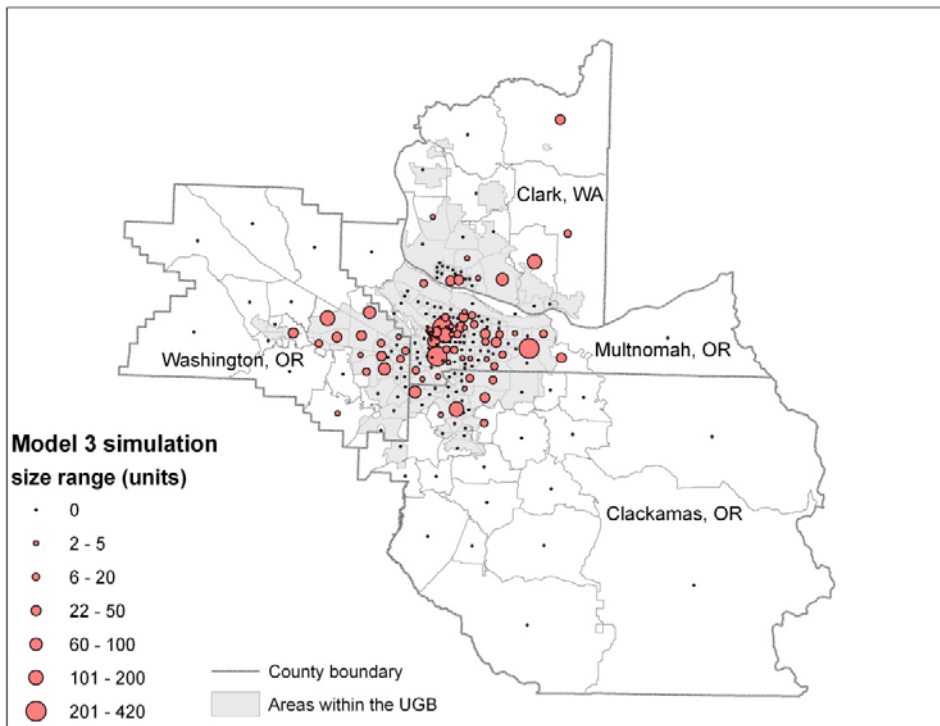**Figure 31** Observed MFH development in 2007
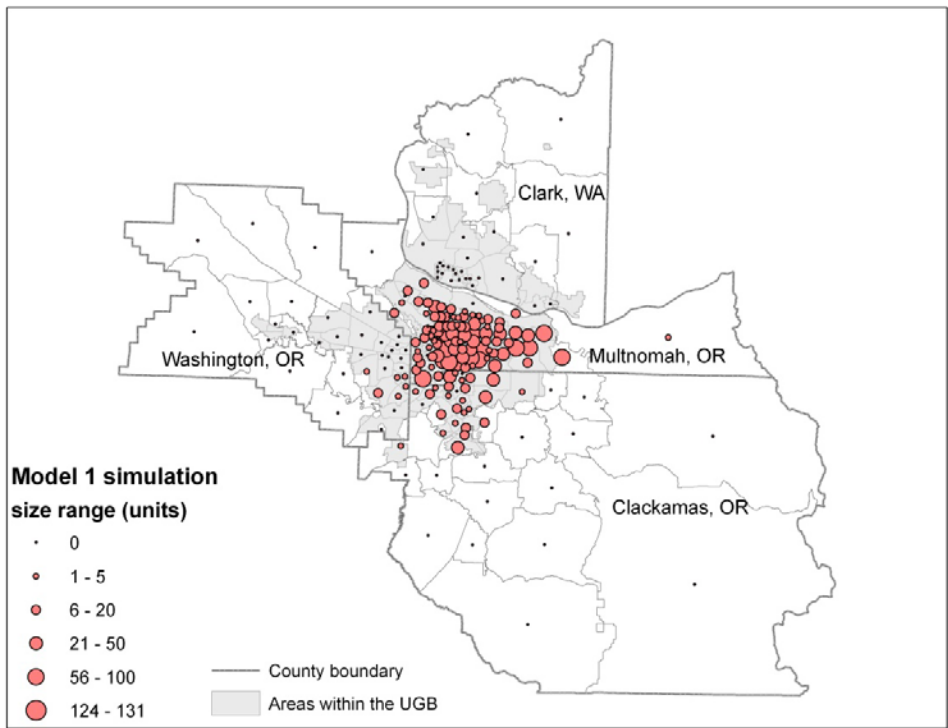
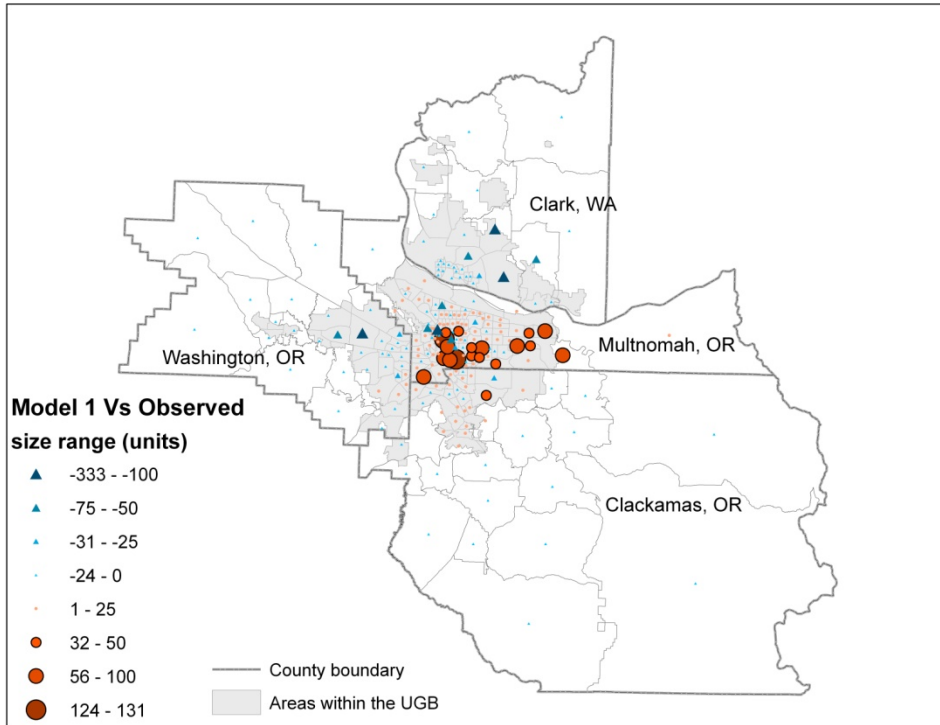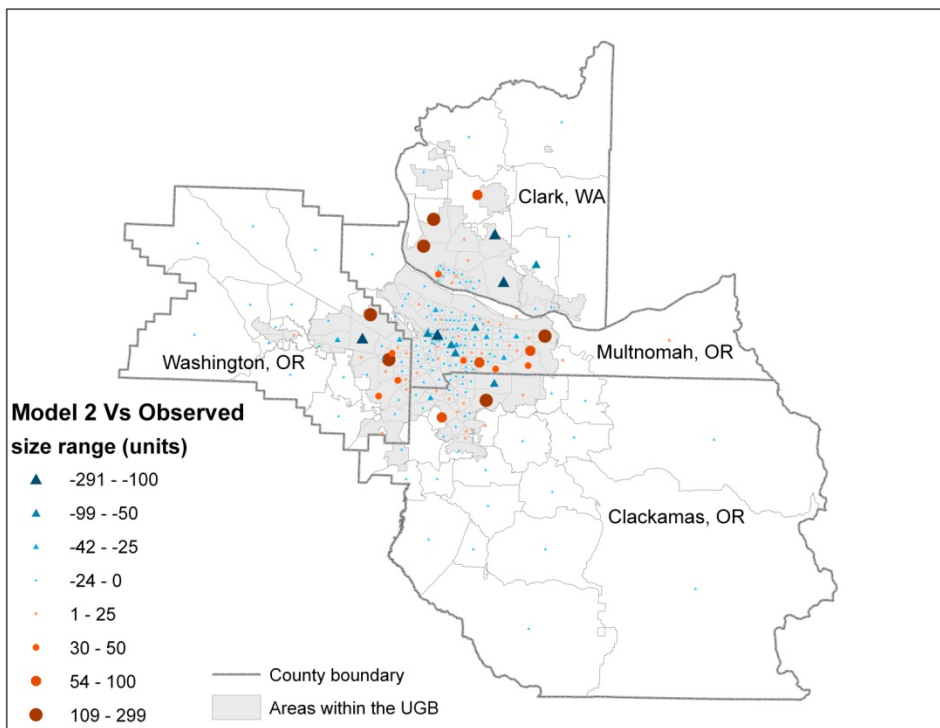**Figure 32** Predicted MFH development by MFH Model 1 in 2007

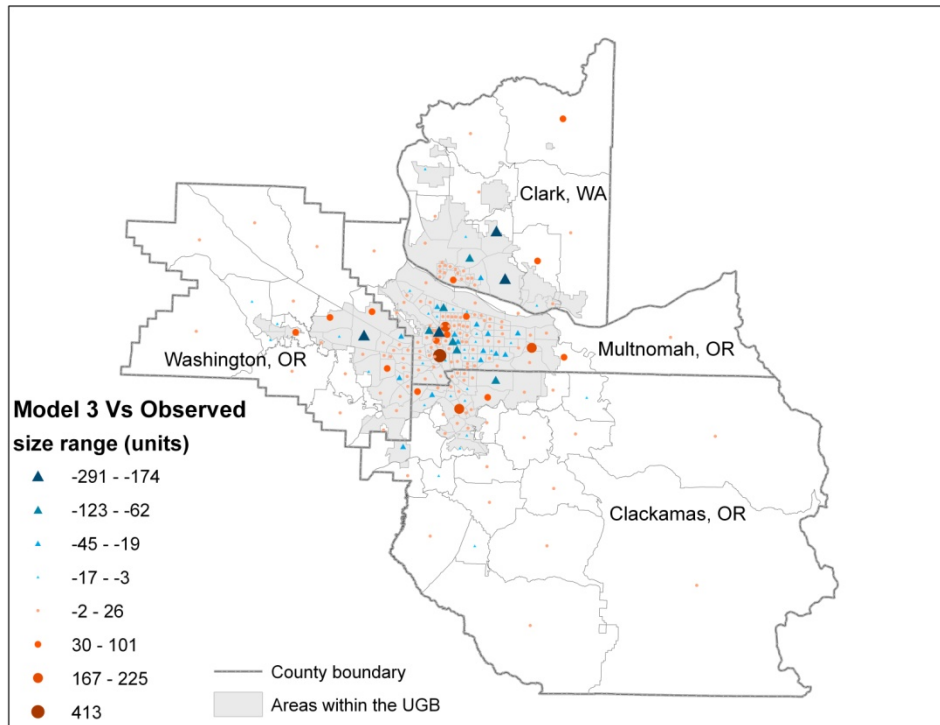**Figure 33** Predicted MFH development by MFH Model 2 in 2007

Figure 34 Predicted MFH development by MFH Model 3 in 2007

Figure 35 MFH Model 1 forecast vs observed in 2007

Figure 36 MFH Model 2 forecast vs observed in 2007

Figure 37 MFH Model 3 forecast vs observed in 2007

census tract level. The RMSEs for MFH Models 1, 2, and 3 are 43.19, 49.82, and 52.98 respectively. This suggests that MFH model 1 is closest to the observed situation, and MFH Model 3 shows the greatest deviation.

The differences between the number of the housing units predicted by the three MFH forecast models and the observed number in each original census tract are displayed in Figures 35, 36 and 37 separately. Again, triangles represent the census tracts where the numbers of new MFH housing units are underestimated by the forecast models, and circles represent the census tracts where the forecast models over-estimate MFH development. The legend on each map indicates ranges of the differences represented by the symbols.

As indicated by Figure 35, MFH Model 1 over-estimates new MFH development in Multnomah County, but under-estimates Washington County and Clark County. A main reason is that individual housing units are used as forecasting units for MFH Model 1. Since each MFH unit requires only a small amount of vacant land, densely urbanized TAZs in Multnomah County are much more likely to be included in choice sets than they are in Models 2 and 3. The large number of small size TAZs with small amount of the buildable land for new MFH in the Portland City makes the area dominate the whole metropolitan area in MFH Model 1.

Both Figure 32 and Figure 36 show that the spatial distribution of MFH development predicted by MFH Model 2 is more evenly distributed than the observed situation. Overall, it tends to overestimate the MFH development in peripheral area but underestimate built-up areas. In MFH model 2, the forecasting units in the model are synthesized projects, in which sizes vary a lot. By model design, for large size MFH projects, only the TAZs with large amount of buildable land are eligible as candidate locations and those TAZs tend to be the TAZs in urban peripheral area. Thus, large size projects are more likely to be located into TAZs in urban peripheral area by MFH model 2. This suggests that the role of buildable land availability has been overestimated in the specification of MFH Model 2.

Similar to MFH Model 2, MFH Model 3 uses synthesized MFH projects as forecasting units. Therefore, it shares the same challenge faced by MFH Model 2; the forecast of a small number of large projects across a large number of location alternatives, which is always subject to large errors. The difference is that the taste variation for projects with

136

different sizes is determined endogenously in MFH Model 3. As Figure 37 indicates, MFH model 3 tends to underestimate new MFH development in Clark County. It shows both underestimation and overestimation in Multnomah County. Interestingly, both the underestimated and overestimated areas in Multnomah County are concentrated in the Portland central city area. Thus, if the forecast results by MFH Model 3 are aggregated at larger spatial scales than the original census tracts, the underestimation and overestimation will cancel each other out, making the forecast results look better. This might suggest that TAZs, even census tracts are still too small to expect accurate forecast results for MFH developments.

The market segmentation for MFH projects also makes the use of dummy variables a problem. Since the number of the MFH projects is much smaller, this issue is more serious for the MFH market segmentation models, because identification and significance problems are more likely to be arise during estimation.

7.6 Lessons Learned

In this chapter, three location choice models are specified and compared for SFH and MFH sectors separately. While all the models use housing projects as observations to estimate models, they make different assumptions in terms of developers' taste heterogeneity in their location choice. In SFH/MFH Models 1, developers are assumed to have homogenous attribute preferences for location alternatives, regardless of project size; however, these utility function parameters were estimated using a weighted maximum likelihood procedure in which project size served as estimation weight. Since each housing project is viewed as a cluster of independent housing units, each project was

atomized when forecasting, allocating each housing unit individually. This seemed to over-predict infill development in more urbanized parts of the region, but yielded more accurate forecasts overall.

In SFH/MFH Models 2 and 3, projects are synthesized and projects with different sizes are assumed to be subject to different sets of attribute preferences representing different types of developers and project contexts. In SFH/MFH Models 2, developers' taste heterogeneity is represented systematically by assigning projects with different sizes into three market segments exogenously. In SFH/MFH Models 3, developers' taste heterogeneity is modeled endogenously by a latent class model. When forecasting both Models 2 and 3, the project is kept intact, requiring that candidate locations have sufficient buildable land to accommodate the entire project.

A comparison between forecasted and observed new SFH housing units in the Portland metropolitan area in 2007 showed that all three SFH models can capture the basic spatial pattern of SFH development in the region. Compared to SFH development, the spatial distribution of MFH development is lumpier and more difficult to forecast. The three MFH models show more deviations from the observed situation. While SFH/MFH Models 2 and 3 are more complex and make more sense from a theoretical perspective, they are not more accurate overall than SFH/MFH Models 1. There are several practical issues that have limited the performance of SFH/MFH Models 2 and 3.

First, the available information about developer attributes is somewhat difficult to use for forecasting purposes, which makes difficult the exploration of developers' location choice taste variation. One would expect to see SFH/MFH Models 2 and 3 produce

better forecasts when the data at individual developer level is richer and understanding of their market behavior is better.

More importantly, model estimation showed that the most significant predictors of where a developer will choose to locate a new housing project are the locations of their previous projects. One prominent advantage of developer-based models such as Models 2 and 3 is that they are able to take into account such the information; however, this study found it difficult to synthesize both developers and their projects. In addition, the data analysis shows that only a small number of very large developers were active in each year in the study period. Thus, if developers in a forecast year are to be synthesized based on developers in base years, one has to decide which developers will continue to be active, which developers will exit market, and how many and what types of developers will be formed on the housing market in the forecast year. All these require much more data than is available, and more studies in developers' market behavior. In this study, for Models 2 and 3, only housing project sizes were synthesized. Their developers and the locations they chose in previous years were missed, and this dramatically limited their forecast capability.

Moreover, this study revealed that market segmentation limits the use of some key dummy variables in the model when most observations in one market segment fall into one group defined by a dummy variable. This issue is especially true for the latent class model where market segmentation is determined endogenously through the estimation process and cannot be controlled directly by the modeler. Since the number of MFH

projects is always much smaller than the SFH projects in U.S. cities, this problem is more serious for MFH forecast models.

 A confounding feature of the Portland housing market is that a small number of large-size projects contribute to a large proportion of the total housing supply, which is likely to be the case in other markets as well. Their large market share makes large-size projects crucial to the forecast, but their small sample size means makes it difficult to estimate significant parameters for this group. In addition, the physics of forecasting a small number of large projects across a large number of location alternatives is challenging and will thus always be subject to large errors. Again, this issue is more challenging for MFH forecast models since the number of MFH projects is smaller, but the average size of the MFH projects is much larger. While constraining choice based on buildable vacant land is helpful, it is not necessarily a cure. Indeed, there is evidence that developers are able to assemble enough land for large projects through rezoning, which implies that some areas thought to not have enough buildable land may indeed be feasible candidates for projects.

Finally, using revealed developer location choices through permit data, one does not observed the choice sets considered by developers. Sampling a subset for the eligible alternative zones in the region offers a practical way out; however, consistency does not hold when sampling alternatives is applied to the non-IID models. Nerella and Bhat (2004) suggest a sample size no less than one fourth of the full choice set and preferably one half of the full choice set in sampling alternatives for the non-MNL models. When

the study area is a large metropolitan area or the spatial scale is very fine, one fourth of

the full choice set might still be too large to be practical.

CHAPTER 8 CONCLUSIONS

A comprehensive new housing supply and location choice model has been developed for

the Portland bi-state metropolitan area and the impacts of Smart Growth policies on home

developers were examined. The purposes are to provide a scenario planning tool for the

region's long-term land use and transportation planning and a micro- and behavioral

foundation for the understanding of Smart Growth Policies from developers' perspective.

The dissertation has reached most of its objectives, but more work needs to be done if in

order to have a better understanding of developer's housing supply and location choice

behaviors. The remainder of this chapter summarizes its main accomplishment, discusses

its shortcomings, and sets out some avenues for future research.

8.1 Smart Growth Policies and Home Developers

One major contribution of this dissertation is that it provides a developer perspective for

the understanding of Smart Growth policies at a micro- and behavioral level. This study

is believed to be the first one that measures the effectiveness of Smart Growth policies in

two neighboring states within a comprehensive developer-based location choice model.

8.1.1 General Evaluation

The assessment of the impacts of Smart Growth policies on home developers in the

Portland bi-state metropolitan area starts out with general location choice models for SFH

and MFH developers. The general models find that home developers in the metropolitan

area respond positively to Smart Growth policies that would encourage developing at

higher densities, providing a mix of housing types, and encouraging mixed used development. SFH and MFH developers show distinct preferences for location attributes, but the most significant predictors of where they would choose to locate a project are the same: the locations of their previous projects. After controlling for all of the other variables in the model, there remains a strong preference for developing SFH units outside of the UGB in both Oregon and Washington sides of the Portland metropolitan area, which indicates that some portion of the market remains strong for lower density dispersed living.

8.1.2 Smart Growth Policies in a Bi-State Context

Acknowledging that the Portland metropolitan area is a bi-state metropolitan area straddling on the border between Oregon and Washington and the two states have different land use systems, bi-state models were developed to examine whether Smart Growth policies have different impacts on SFH and MFH developers' location choice in Oregon compared with across the border in the State of Washington.

The model results show that home developers on the two sides of the state border do react to the Smart Growth policies differently. On the Oregon side of the metropolitan area, the highest SFH densities are preferred, but on the Washington side there remains a stronger propensity to develop in medium-density areas. I found evidence that new SFH is built predominantly in TAZs that already contain a mix between SFH and MFH housing types or is dominated by SFH in Oregon, whereas in Clark County recent SFH developments have been more likely to be developed in either existing SFH or MFH-dominated zones, but not mixed housing. The interpretation is that Clark County contains fewer existing

mixed housing types and that this is witness to the gradual transition to mixed housing types in the propensity to develop SFH in areas now dominated by MFH. Supporting this notion of a maturing growth management system is the significant positive preference estimated for Clark County to develop SFH in areas dominated by non-residential development, compared with Oregon-side preferences for avoiding non-residentially-dominated zones and favoring existing residential and mixed use zones in that order. Essentially, Clark County development is transitioning toward a more mixed use pattern through SFH development in commercial areas. This is also supported by a significant positive impact of automobile accessibility to retail employment in Clark County and an opposite negative impact of auto access to retail employment in the three Oregon counties.

In terms of MFH, in both states the development preferences seem to favor existing locations that are dominated by residential use, especially the locations with high MFH density. At the same time, there was a significant impact in both states of developing MFH in TAZs that are highly accessible by transit to non-retail employment, which clearly supports the Smart Growth ideals of fostering shorter, non-auto work commutes and reflects recent large-scale development of former industrial lands.

8.1.3 Home Developers' Preference Heterogeneity

Latent class models have been developed to detect taste variations among home developers in the SFH and MFH markets separately. The project attributes such as developer size, project size, developer contract type, and developer specialization are used to segment the housing projects into three market segments endogenously. It is

found that there are clear taste variations across developers and housing projects with respect to site attributes in their location choice. Among the four variables used in the segmentation model, project size is the most significant one, and it provides a better model fit to the data than developer size, indicating that developers show taste variations among their different projects.

The segment-specific location choice models show that large size SFH projects developed by contractor-owners are more likely to be within the UGB and their locations tend to have higher residential density, housing diversity, transportation accessibility, road density, and land price. With most MFH projects within the UGB, estimation results show that large size MFH projects prefer the locations with higher residential density, housing diversity, mixed use, road density, land price, average household income, and proportion of young and middle age household. Thus, a policy implication that can be drawn is that large size housing developers and projects are more likely to follow Smart Growth principles and more attention needs to be paid to small scale housing developments, especially those on the rural land.

8.2 Comprehensive New Housing Supply and Location Choice Model

Another major contribution this dissertation is the development of a set of developer-based new housing location choice models, which are rarely seen in existing integrated land use and transportation models. Time series regression models that predict annual new housing supply in the region and models that synthesize housing projects in the forecast year are developed to support the developer-based location choice models. Thus, the comprehensive new housing supply and location choice model proposed by this study

follows a top-down perspective with three major steps: aggregate housing supply forecast, project synthesis, and location choice.

8.2.1 Aggregate Housing Supply Forecast

Three aggregate housing supply forecast models were developed and compared: a conditional time series regression forecast model, an unconditional time series regression forecast model, and a simple ARIMA forecast model.

The three forecast models have advantages and disadvantages. The conditional time series regression forecast model predicts more accurately, but is conditional on knowing the real housing price values for future years, which might not be the case in the land use modeling practice. For certain scenario analyses, where the real housing prices in future years are provided by an expert panel to test different scenarios, this model can be useful. The unconditional time series regression forecast model does not require knowledge of the housing prices in future years. Instead, the model forecasts based on the historical real housing price data and uses projections from the ARIMA forecast model. Compared with the ARIMA model, an advantage shared by both conditional and unconditional forecast models is the reflection of the real market dynamic process and the sensitivity to price signals from the housing market. The ARIMA model requires the least amount of data, but reveals nothing about other cost shifters in the housing market. Its prediction is entirely dependent on its past movement path.

8.2.2 Project Synthesis Model

SFH and MFH projects are synthesized based on their size distributions in previous years. The tests show that Gamma distribution fits the size distributions of SFH and MFH projects best among many probability distributions that have been tried. Besides their size distributions in previous years, the following constraints are set to make synthesized projects closer to observed projects in the forecast year: total amount of new SFH/MFH units in the forecast year estimated by the aggregate housing supply models; minimum and maximum sizes of SFH/MFH projects in the forecast year, and the total number of SFH/MFH projects in the forecast year.

The model tests found that the SFH project synthesis model can predict better if it only synthesizes SFH projects with two or more units and transforms SFH project sizes into natural log, assuming that the proportion of SFH projects with only 1 unit in all SFH projects in the forecast year is the same as their proportion in previous years. Similarly, for MFH project synthesis model, synthesized projects are closer to observed ones when the model synthesizes projects with three or more units and transforms project sizes into natural log, assuming that the proportion of MFH projects with only 2 units in all MFH projects in the forecast year is the same as their proportion in previous years.

8.2.3 Location Choice Models

Three location choice models are proposed and compared. The first one is a MNL model and developer heterogeneity is ignored. The market is assumed to be homogenous and each housing unit is treated the same and distributed across the region as an individual unit; however, these utility function parameters were estimated using a weighted

maximum likelihood procedure in which project size served as estimation weight. Since each housing project is viewed as a cluster of independent housing units, each project is atomized when forecasting, allocating each housing unit individually. This seemed to over-predict infill development in more urbanized parts of the region, but yielded more accurate forecasts overall.

The other two models require synthesized housing projects, and projects with different sizes are assumed to be subject to different sets of attribute preferences representing different types of developers and project contexts, thus providing a more realistic depiction of land requirements for multi-unit projects. In the second model, developers' taste heterogeneity is represented systematically by assigning projects with different sizes into three market segments exogenously. In the third model, developers' taste heterogeneity is modeled endogenously by latent class models. When forecasting these two models, the project is kept intact, requiring that candidate locations have sufficient buildable land to accommodate the entire project. While these two models are more complex and make more sense from a theoretical perspective, they are not more accurate overall than the first one for both SFH and MFH market sectors. There are several practical issues that have limited their performance.

First, the available information about developer attributes is somewhat difficult to use for forecasting purposes, which makes difficult the exploration of developers' location choice taste variation. Moreover, it was found that market segmentation limits the use of some key dummy variables in the model when most observations in one market segment fall into one group defined by a dummy variable. In addition, the physics of forecasting a

small number of large projects across a large number of location alternatives is challenging and will thus always be subject to large errors.

8.3 Limitation and Future Research

Despite its accomplishments, this dissertation has its limitations. A few of them are discussed below, with some elements of solution for future research.

8.3.1 Data Constraint

I am fortunate to have the data support from Metro; however, land use models have been notoriously known for its data hungry nature and this study is not an exception. The impacts of data constraint on this study can be summarized into two aspects.

First, some important variables are missed from the models due to unavailable data. For example, it is well known that housing price and vacancy rates are important factors that influence developers' location choice decision, however, since the data for them are not available at zone level, they are missing from the models. One would expect to see the explanatory power of developer location choice model to be significantly improved when the data on these two variables are available at small spatial scales, if the endogeneity problems caused by them are appropriately handled.

Local governments' policy and tax incentives have a profound impact on the local housing market. Some of them were controlled in this study, such as zoning and the UGB, but many others are missing from the models. Future research can make improvements by including relevant variables in the model.

Second, some variables used for model estimations in this study are subject to measurement error. This study uses numerous sources of data to create variables needed for model estimation, but none of them were collected intentionally for building land use models. This not only creates a huge amount of data process work, but also produces some measurement errors while transforming the data into the variables required by this study. A good example is the measurement of buildable land for MFH development. In this study, I used the amount of vacant land zoned for medium and high-density residential use or mixed use in a TAZ to measure the availability of buildable land for MFH development in that TAZ. However, in many TAZs, vacant land is fragmented, and a simple aggregation could overestimate the ability to accommodate large-size projects. Another issue is that some MFH projects were located on developed land originally zoned for non-residential purposes. But there is no way to count in those zoning changes in the model because they were negotiated case by case and not showed in publically available data.

The use of proxy variables can also lead to measurement errors. In this study, some variables were used as proxies for the real variables of interest, which were not available. For example, road density was used as a proxy for existing infrastructure. In many cases residential infrastructures such as sewage and water system are accompanied with road systems, but road density is not always an appropriate variable to represent the availability and density of infrastructure in the locations.

8.3.2 Discrete Choice Modeling

Discrete choice modeling has become a commonly used tool to simulate residential location choice. However, all the residential location choice models, using either households (demand side) or developers (supply side) as observation units, share a similar problem: real alternative locations considered by the decision makers are not observed. This problem could be be partially solved by conducting surveys asking developers to state their considered choice sets, but it could be extremely difficult to accurately represent all the home buyers or developers in the whole metropolitan area.

Furthermore, while random sampling of alternatives provides statistically consistent estimates even for small sample sizes in the MNL model and helps avoid spatial autocorrelation issue, such consistency does not hold for non-MNL models. Nerella and Bhat (2004) suggest a sample size no less than one fourth of the full choice set and preferably one half of the full choice set in sampling alternatives for the non-MNL models. When the study area is a large metropolitan area or the spatial scale is very fine, one fourth of the full choice set might still be too large to be practical.

Discrete location choice models developed in this study assume that all developers already have an investment plan and what they need to do is to select a location from a set of alternative locations. In reality, different developers might follow different paths to look for locations for their projects. Some might already have an investment project and need to decide a location for that project, but there are other developers who already own the land and what they look for is the right investment project for the land. Even for developers who are searching locations for their projects, they might not just choose one

location by comparing a set of alternative locations simultaneously. Instead, they might reduce their search scope step by step, or search locations one-by-one until they get the right one for their projects. Future research can make contributions by studying those different location behavior patterns of developers.

Lastly, while the statistical models can show us significant factors that are influencing developers' location choices, causality is still not fully understood. Conducting in-depth interviews and structural survey on housing developers would be helpful to understanding how developers view Smart Growth policies and what role those policies play when they choose project locations.

8.3.3 Housing Market Volatility

The housing markets in U.S. cities have followed boom-and-bust cycles. The volatility of the housing market has made the forecast of new housing supply challenging, and this will be even especially true when the forecast is a long-term one. This study already shows that changes to housing prices is the most significant predictor for new housing supply, but forecasting housing price is difficult. New housing supply is also strongly influenced by long-term demographic trends, economic policy changes, and economic cycles, but the later two are also difficult to predict. At the micro-level, our understanding of how housing suppliers make decisions and view the market is still poor. These considerations call for more data collection and research at both macro-, and especially at micro-levels to help us understand housing developers better.

References

Abbott, C. (1997). The Portland region: where city and suburbs talk to each other and often agree. *Housing Policy Debate*, 8(1), pp. 11-51.

Abbott, C. (2002). Planning a sustainable city: the promise and performance of Portland's urban growth boundary. In G.D. Squires (Ed.), *Urban Sprawl: Causes, consequences, and policy responses* (pp. 207-235). Washington, DC: Urban Institute Press.

Bhat, C.R. (1997). Endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science*, 31(1), pp. 34-48.

Carruthers, J.I. (2002) The impacts of state growth management programmes: a comparative analysis. *Urban Studies*, 39(11), pp. 1959-1982.

Cho, S.H., Chen, Z., Yen, S.T., and Eastwood, D.B. (2006). Estimating effects of an urban growth boundary on land development. *Journal of Agricultural and Applied Economics*, 38(2), pp. 287-298.

Clark County. (2010a). Vacant Buildable Lands Model for Clark County, Washington. http://gis.clark.wa.gov/applications/gishome/reports/?pid=vblm, (accessed on 02 Mar 2010).

Clark County. (2010b). 20-year comprehensive growth management plan 2004-2024. http://www.co.clark.wa.us/longrangeplan/review/documents/CompPlan_2009-Amendments.pdf, (accessed on 02 Mar 2010).

Conder, S., and Lawton, K. (2002). Alternative futures for integrated transportation and land use models contrasted with trend-delphi models. *Transport Research Record: Journal of the Transportation Research Board,* 1805, Transportation Research Board of the National Academies, Washington, D.C., pp. 99-107.

Darkins, C.J., and Nelson, A.C. (2003). State growth management programs and central-city revitalization. *Journal of the American Planning Association*, 69(4), pp. 381-396.

de Palma, A., Motamedi, K., Picard, N. and P. Waddell. (2005). A model of residential location choice with endogenous housing prices and traffic for the Paris region. *European Transport*, 31, pp. 67-82.

DiPasquale, D., and Wheaton, W. (1994) Housing Market Dynamics and the Future of Housing Prices. *Journal of Urban Economics*, 35, pp. 1-27.

DiPasquale, D. (1999). Why don't we know more about housing supply? *Journal of Real Estate Finance and Economics*, 18(1), pp. 9-23.

Downs, A. (2005). Smart growth: why we discuss it more than we do it. *Journal of the American Planning Association*, 71(4), pp. 367-378.

Follain, J. R. (1979). The price elasticity of the long-run supply of new housing construction. *Land Economics*, 55(2), pp. 190-199.

GIS department of Clark County, Washington (2000-2007). Clark County GIS, CD ROM. Clark County, Washington: Author.

Gordon, D. and Vipond S. (2005) Gross density and new urbanism. *Journal of the American Planning Association*, 71(1), pp. 41-54.

Greene, W.H. & Hensher, D.A. (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B*, 37, pp. 681-698.

Haider, M. (2003). Spatio-Temporal Modeling of Housing Starts in the Greater Toronto Area. Ph.D. thesis, University of Toronto, pp. 313-344.

Haider, M. and Miller, E.J. (2004). Modeling location choices of home builders in the great Toronto area. *Transportation Research Record*, 1898, pp. 148-156.

Hunt, J. D., and Abraham, J. E. (2003). Design and application of the PECAS land use modelling System. Presented at 8th Computers in Urban Planning and Urban Management Conference, Sendai, Japan.

Hunt, J. D., Kriger, D. S., and Miller, E. J. (2005). Current operational urban land-use transport modeling frameworks: a review. *Transport Review*, 25(3), pp. 329-376.

Jun, M.J. (2004). The effects of Portland's urban growth boundary on urban development patterns and commuting. *Urban Studies*, 41, pp. 1333-1348.

Leeuw, F. D., and Ekanem, N. F. (1971). The supply of rental housing. *The American Economic Review*, 61(5), pp. 806-817.

Oregon Department of Land Conservation and Development. (2010). Oregon's 19 Statewide Planning Goals. http://www.lcd.state.or.us/LCD/goals.shtml (accessed on 03 Mar 2010).

Martinez, F. (1996). MUSSA: land use model for Santiago city. *Transportation Research Record: Journal of the Transportation Research Board*, 1552, Transportation Research Board of the National Academies, Washington, D.C., pp.126-134.

Martinez, F. and Henriquez, R. (2007) A random bidding and supply land use equilibrium model. *Transportation Research Part B*, 41, pp. 632-651.

Mayer, C., and Somerville, C. T.. Residential Construction: Using the Urban Growth Model to Estimate Housing Supply. In *Journal of Urban Economics*, 48, 2000, pp. 85-109.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record* 673, TRB, National Research Council, Washington DC., pp. 72-77.

Metro. (1992). Charter. As adopted in November 1992 and amended in November 2000. http://www.oregonmetro.gov/index.cfm/go/by.web/id=629 (accessed on 02 Mar 2010).

Metro. (1995). 2040 Growth Concept. http://www.oregonmetro.gov/index.cfm/go/by.web/id=29882 (accessed on 03 Mar 2010)

Metro. (2000-2007a). Geo-coded building permit data. Portland, OR: Author.

Metro. (2000-2007b). Regional land information system (RLIS). Portland, OR: Author.

Meyer, M.D. and Miller E.J. (2001). Urban transportation planning, international edition 2001. New York, NY: McGraw-Hill Companies, Inc.

Miller, E. J., Farooq, B., and Habib, M. A. Microsimulating (2006). Spatial market: conceptual design & implementation of a housing market model. CD-ROM. Urban Transportation Research & Advancement Centre, University of Toronto, Toronto, Canada, Vol. V, pp. 58-62.

Nelson, A.C. and Moore T. (1993). Assessing urban growth management: the case of Portland, Oregon, the USA's largest urban growth boundary. *Land Use Policy*, Oct., pp. 293-302.

Nelson, A.C. and Peterman, D.R. (2000) Does growth management matter? The effect of growth management on economic performance. *Journal of Planning Education and Research*, 19, pp. 277-285.

Nelson, A.C., Sanchez, T.W., and Dawkins, C.J. (2004). The effect of urban containment and mandatory housing elements on racial segregation in US metropolitan areas, 1990-2000. *Journal of Urban Affairs*, 26(3), pp. 339-350.

Nerella, S. and Bhat, C.R. (2004). Numerical analysis of effect of sampling of alternatives in discrete choice models. *Transportation Research Record* 1894, TRB, National Research Council, Washington DC., pp. 11-19.

NLOGIT Version 4.0, Econometric Software, Inc., Plainview NY, 2007.

O'Connell, L. (2009). The impact of local supporters on smart growth policy adoption. *Journal of the American Planning Association*, 75(3), pp. 281-291.

Pendall, R. (1999). Do land-use controls cause sprawls? Environment *and Planning B: Planning and Design*, 26, pp. 555-571.

Poterba, J. M. (1984). Tax subsidies to owner-occupied housing: an asset-market approach. *The Quarterly Journal of Economics*, 99(Nov.), pp. 729-752.

Richardson, H.W., and Gordon, P. (2001). Sustainable Portland? A critique, and the Los angeles counterpoint. Paper presented at the ACSP Conference, Cleveland, Ohio, November 10, 2001.

Seltzer, E. (2004). It is not an experiment: Regional planning at Metro, 1990 to present. In C. Ozawa (Ed.) *The Portland edge: Challenges and successes in growing communities* (pp. 35-60). Washington, DC: Island Press.

Shen, Q. and Zhang, F. (2007) Land-use changes in a pro-smart-growth state: Maryland, USA. *Environment and Planning A*, 39, pp. 1457-1477.

Song, Y. (2005). Smart growth and urban development pattern: a comparative study. *International Regional Science Review*, 28(2), pp. 239-265.

Song, Y. and Knaap, G.J. (2003). New urbanism and housing values: A disaggregate assessment. *Journal of Urban Economics*, 54, 218 - 238.

Song, Y. and Knaap, G.J. (2004). Measuring urban form: Is Portland winning the war on sprawl? *Journal of the American Planning Association*, 70(2), pp. 210-225.

Stover, M. E. (1986). The price elasticity of the supply of single-family detached urban housing. *Journal of Urban Economics*, 20, pp. 331-340.

Topel, R., and S. Rosen. (1988). Housing investment in the United States. *Journal of Political Economy*, 96(4), pp. 718-740.

Train, K.E. (2003). Discrete choice methods with simulation. New York, NY: Cambridge University Press.

U.S. Bureau of Labor Statistics (2010). Consumer price index for Portland-Salem area.

U.S. Census Bureau (1996-2008). Building permit data by county. US Department of Commerce, Census of Bureau, Washington, DC.

U.S. Census Bureau (2009). Relationship between building permits, housing starts, and housing completions. US Department of Commerce, Census of Bureau, Washington, DC. http://www.census.gov/const/www/nrcdatarelationships.html (Accessed on 18 April, 2009)

U.S. Census Bureau (2000). Census 2000. US Department of Commerce, Census of Bureau, Washington, DC.

U.S. Census Bureau (2005). 2005 American Community Survey. US Department of Commerce, Census of Bureau, Washington, DC.

U.S. Census Bureau (2000-2007). 2000-2007 ZIP Code Business Patterns. US Department of Commerce, Census of Bureau, Washington, DC.

Washington State Legislature (1990). Growth Management Act. http://apps.leg.wa.gov/RCW/default.aspx?cite=36.70A. (Accessed on 12 May 2010)

Waddell, P. (2002). UrbanSim: modeling urban development for land use, transportation and environmental planning. *Journal of the American Planning Association*, 68(3), pp. 297-314.

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. *Networks and Spatial Economics*, 3(1), pp. 43--67.

Wassmer, R.W. (2002). Fiscalisation of land use, urban growth boundaries and non-central retail sprawl in the western United States. *Urban Studies*, 39(8), pp. 1307-1327.

Wassmer, R.W. (2006). The influence of local urban containment policies and statewide growth management on the size of United States urban areas. *Journal of Regional Science*, 46(1), pp. 25-65.

Weitz, J. (1999). Sprawl busting: state programs to guide growth. Chicago, IL: Planners Press.

Weitz, J and Moor, T. (1998). Development inside urban growth boundaries: Oregon's empirical evidence of contiguous urban form. *Journal of the American Planning Association*, 64(4), pp. 424-440.