Stepping into new territory: Three essays on agent-based computational economics and environmental economics

by

Jiaqi Ge

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:

Leigh Tesfatsion, Major Professor

Catherine Kling

Joseph Herriges

Steve Kautz

John Schroeter

Iowa State University

Ames, Iowa

2014

To my parents

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis. First and foremost, Dr. Leigh Tesfatsion for her guidance, patience and support throughout this research and the writing of this thesis. Her insights and words of encouragement have often inspired me and renewed my hopes for completing my graduate education. I would also like to thank my committee members for their efforts and contributions to this work: Dr. Steve Kautz, Dr. Catherine Kling, Dr. Joseph Herriges, and Dr. John Schroeter.

CHAPTER 1. OVERVIEW

This dissertation explores various new tools and methodologies in economic research to study real-world social and economic systems. I am interested in developing research tools tailor-made to solve important social and economic issues. I am also interested in designing tools that are flexible and adaptable enough to be used in future research. The tools explored in the dissertation include agent-based computational modeling, meta-analysis methodology, as well as other analytical and empirical methodologies. In addition to tool building, I am interested in studying the complex and evolving nature of real-world social and economic systems. The issues studied in the dissertation range from water protection to housing market crises. The dissertation aims to expand the available toolkit for economic research, to establish a connection between conventional and new tools and push the research frontier, to answer important research questions and to solve real-world problems. Each core chapter in the dissertation is a stand-alone study. These core chapters are summarized below.

Chapter 2 investigates the issue of water quality and develops a meta-analysis of non-market valuation studies of water quality improvement. After reviewing more than 100 studies, we estimate a valuation model for water quality improvement based on a data set that consisting of 332 valuations from 37 distinct studies. The valuation model is then used to predict the mean willingness to pay for water-quality improvement at a given site by households living in a given region. For instance, the willingness to pay by a typical household living in the state of Iowa for a water quality increase from 40 to 50 (out of 100) at a one-square-mile aquatic site, like the Spirit Lake in Iowa, is predicted to be \$137.52. The valuation model developed in this study is particularly convenient when a researcher wishes to evaluate the benefit of a project that aims at improving water quality, but a primary study is too costly or time consuming.

Chapter 3 develops a spatial agent-based computational model of housing market to help understand what caused US housing prices to rise and collapse during the years immediately preceding the U.S. financial crisis (2007-2009). We study key factors for their impacts on housing price volatility, such as speculation and lenient financing. The dynamic simulation findings for this housing market model demonstrate in concrete terms how lenient bank lending practices combined with speculation can lead to increased volatility in housing prices, including sharp rises followed by sudden collapses. The exploratory work in this paper could contribute to the understanding of housing price volatility and hence inform policy decisions affecting the housing market.

Chapter 4 adapts and extends the model developed in Chapter 3 in order to conduct a case study of the Washington D.C. housing market. It is an example of knowledge accumulation through model adaptation. Agent-based computational models are known for being flexible and adaptable. However, few studies have made full use of the flexibility and adaptability of existing models. This study fills the gap by adapting an existing conceptual model to a specific context. Using empirical evidence, we tailor the housing market model developed in Chapter 3 to the city of Washington D.C. We use property transaction data and demographic data in Washington D.C. neighborhoods to calibrate and validate the adapted model. We conclude that we need segregated or even individual-level data for the calibration and validation of agent-based models. We also need more empirically-based model specification. Specifically, for the analysis of housing market, it is insufficient to look only at city-wide aggregate house prices, we need to take into account both city-wide trends and intra-city differences.

The dissertation concludes with a chapter of summary remarks.

CHAPTER 2. HOW MUCH IS CLEAN WATER WORTH? VALUING WATER QUALITY USING A META-ANALYSIS

2.1 Introduction to Chapter 2

Since the passage of the Clean Water Act in 1972, protecting and improving water quality has been an important issue of U.S. national environmental policy. Numerous clean water projects and programs have been proposed since then. The cost of such a project depends on available technologies, but the benefit of it is, by a large extent, subject to judgment. Do people value clean water? This study attempts to find the answer in existing literature. Our research questions include: How much is clean water worth? What is peoples willingness to pay for a better water quality? Does willingness to pay for clean water estimates vary systematically by research methodologies, sample characteristics and site characteristics?

We conducted a meta-analysis on existing non-market valuations of water quality improvement [1]. We collected 332 valuations from 37 distinct studies for the analysis, after having examined more than 100 studies. Since different studies use different ways to quantify water quality, one of the challenges is to devise methods to convert different water quality indicators to a consistent water quality index.

For any given region, our estimated valuation model enabled us to predict the mean willingness to pay for water quality improvement in a site of a certain size. For example, an average household living in the 50,000 square-mile region of the site might be willing

[1] An introduction to meta-analysis is presented in 2.3

to pay \$115.14 for a water quality improvement of 5 points (e.g., from 40 to 45) in a one-square-mile aquatic site. The model is particularly useful when we need to evaluate the benefit of a project that aims at improving water quality, but collecting first-hand data is considered too costly or time consuming.

We will also test the null hypothesis that the three main approaches in non-market valuationthe hedonic model, the travel cost model, and the contingent valuation model have generated statistically consistent valuations in these data. Our test results reject the null hypothesis, and we conclude that, for any given site and water quality improvement, the hedonic model estimates are the highest, the travel cost model the second highest, and the contingent valuation model the third highest.

This report is structured as follows. First, we describe the three main approaches to valuing water quality improvement, followed by a brief introduction to meta-analysis as a research methodology. We then describe our approach to converting different water quality indicators from the various studies to a common metric which is essential for undertaking the meta-analysis. After that, we move on to willingness-to-pay function and data specifications. We then display and discuss the regression results. Next, we compare this study with other meta-analyses. Last, we present conclusions.

2.2 Three Main Approaches for Valuing Water Quality

Unlike most commodities, access to lakes, rivers, and streams is generally not traded in a market, so there is no market price for clean water. Three standard approaches to the valuation of non-market goods have been used to solve the problem. The first approach is the travel cost method. The idea is, although people do not pay direct fees to visit the aquatic site, they do spend time and other costs, such as cost of gasoline, to travel to the site. The opportunity cost of time and other costs are the price for access to clean water. Hence, we can use it to elicit the value of clean water.

The second approach is the hedonic method. This approach recognizes that housing prices depend on water access and water quality. A house on a lake or river is usually more expensive than a similar one not on an aquatic site. Likewise, a house on a very clean lake or river is usually more expensive than one on a not-so-attractive lake or river. Thus, the differences in the housing price reflect peoples valuation of clean water.

One common feature of the two approaches is that they both use actual behavioral data, be it people's visitation to a site or transaction in the property market. Both approaches indirectly infer people's valuation of clean water from their behavior. The third approach, however, is not based on what people do, but what people say they will do under certain scenarios. The third approach, the contingent valuation method, directly elicits the maximum willingness to pay for better water quality in a survey.

The first two approaches, the travel cost and the hedonic method are revealed preference methods, because economic values are indirectly "revealed" from behavior. The third approach, the contingent valuation method, is a stated preference method because people directly state their preference (in a survey for example). Researchers have also combined stated and revealed preference methods for the same sample. In short, all three approaches have been widely used and have become standard tools in the non-market valuation literature. Since valuations of clean water have important policy implications, one purpose of this paper is to test if the three approaches generate the same statistical valuation. And, since the methods are likely to be applied to different types of resources and different populations of users, we do not necessarily expect the values generated by them to be the same, but understanding the source of difference in valuations will be valuable.

2.3 What is Meta-Analysis and Why Use Meta-Analysis?

Meta-analysis is a research method that collects results from existing studies by independent researchers . It is widely used in psychology, epidemiology, sociology, educational research, and evidence-based medicine. More recently, it has become more common in economic research as well (Johnston et al. (2005)). It serves as a base to achieve one or more of three purposes: research synthesis, hypothesis testing, and benefit transfer.

The first purpose, research synthesis, is to provide a quantitative review of the existing literature. The second purpose, hypothesis testing, is to test hypothetical patterns that might exist in the results from existing studies. The third purpose, benefit transfer, is to construct a valuation model from estimates in existing studies. The valuation model can then be used to derive benefit estimates in different settings.

In this paper, we will conduct a meta-analysis that serves all three purposes. We will provide a quantitative summary of existing literature in non-market valuation of water quality changes in lakes and streams. We will test the null hypothesis that the three standard non-market valuation approaches generate equal valuations. We will also develop a valuation model of peoples willingness to pay for clean water. The valuation model can be used to perform benefit transfer on any site in any region.

There are several advantages of using a meta-analysis as opposed to collecting first-hand data. First, it takes much less money and time to do a meta-analysis. Projects usually come under strict time and budget constraints and benefit estimates may be needed promptly. For example, a contingent valuation (CV) survey can last a year or more from survey design, to implementation, to data analysis. Additionally, primary data collection can cost thousands of dollars, depending on the scale of the survey. In short, collecting primary data for the purpose of study is costly and time consuming, if possible at all. On the contrary, a meta-analysis can be done with a fraction of the time and money. Moreover, once a meta-analysis is done, the resulting valuation model can be used to evaluate other projects.

Second, there are certain things that can only be learned in the context of multiple studies. Since each study is a snapshot, we need to combine many such studies to be able to identify any underlying trends and patterns within the existing literature. If we want to study the similarities and differences between valuations from travel cost, hedonic, and contingent valuation models, we need to look at multiple papers that cover the three methodologies. This is an important question that cannot be answered by any single study using only one of the three standard approaches.

Third, the meta-analysis provides a quantitative review of the literature in a way that identifies how differences in study design, resource characteristics, and sample characteristics translate into different economic values. This information can help identify weakness in the literature and where future research should best be targeted.

There are, however, critics of the methodology who cite its potential loss of accuracy. We are aware of these concerns and discuss some caveats in the conclusions.

2.4 Conversions Between Water Quality Indicators

2.4.1 Water Quality Indicator

To compare the willingness to pay for water quality across a range of studies, we need to identify a common unit of water quality change. Each study produces a willingness to pay in dollar value, which is the dependent variable. On the right hand side, we have water quality improvement, and other factors such as site characteristics, sample characteristics, and research methodologies.

In the studies we have collected for the meta-analysis, there are three common ways to quantify water quality: secchi depth, water quality index, and other water attributes. To conduct a meta-analysis, we need to find a way to convert all three types of water quality indicators to a consistent scale. Since a water quality index taking the form of a score from 1 to 100 is common, we have decided to convert all indicators to the water quality index. Before we do that, we will explain the three water quality indicators in turn.

The first indicator, secchi depth, is the deepest level that a secchi disk (a circular black and white disk) is visible in the water. It is used to measure the water transparency. The higher the secchi depth, the more transparent the water is, and the better the water quality.

The second indicator, referred to as index, uses an index or ladder to quantify water quality. A score from 0 to 100 is an example of ladder, and good, fair, poor is another example. One commonly used ladder is the Resources For the Future (RFF) water quality ladder (Mitchell and Carson (1981)). The RFF ladder identifies water quality by its suitable recreational use. From high to low, water quality can be identified as drinkable, swimmable, fishable, boatable, suitable for outings, and not suitable for any activities, with each corresponding to a score of 95, 70, 50, 25, 15, and 5, respectively.

The third indicator, referred to as water attribute, uses one or more of the nine water attributes to measure water quality. Those attributes include pH value, phosphorus level, oxygen level, and nitrogen level. The water quality index can be derived from these water attributes using the now standard method developed in 1970 by the National Sanitation Foundation (Brown et al. (1970)).

The National Sanitation Foundation index ranges from 0 to 100 and reflects the composite influence of nine physical, chemical, and microbiological attributes of water quality (McClelland (1974)). The nine attributes are: dissolved oxygen, fecal coliform, pH, biochemical oxygen demand, temperature change, total phosphate, nitrate, turbidity, and total solids. Each attribute is given a different weight according to its importance. The National Sanitation Foundation water quality index is a weighted average of the quantile value (Q value) of the nine attributes. Specifically, the formula to construct the water quality index (WQI) is,

$$WQI = \prod_{i=1}^{9} {q_i}^{w_i} \tag{2.1}$$

where q_i is the quantile, or Q value of parameter i, and w_i is the weight for parameter i, $\sum_{i=1}^{9} w_i = 1$. The Q values are used instead of raw measurements so that the scale is consistent. Water quality parameters and weights are shown in Table 2.1.

Table 2.1: Water quality parameters and weights (complete)

Parameter	**Weight**
dissolved oxygen	0.17
fecal coliform	0.16
pH	0.11
biochemical oxygen demand	0.11
temperature change	0.10
total phosphate	0.10
nitrates	0.10
turbidity	0.08
total solids	0.07
Total	1.0

If all nine parameters are not available, a WQI can still be calculated based on parameters that are available. For example, if we only have M out of nine parameters ($M < 9$), WQI can be obtained by adjusting the weight for the available parameters proportionately, such as in the following equation,

$$WQI = \prod_{i=1}^{M} {q_i}^{{w_i}'} \tag{2.2}$$

where ${w_i}'$ is the adjusted weight for attribute i, ${w_i}' = \frac{w_i}{\sum_{i=1}^{M} w_i}$, and $\sum_{i=1}^{M} {w_i}' = 1$.

The National Sanitation Foundation water quality index enables us to convert water attributes (the third indicator) to a water quality index (the second indicator). However, the first indicator, secchi depth, is not a parameter in the water quality index, although it is understood that the secchi measure is directly related to the individual components of the WQI. Hence, there is no readily available conversion between secchi depth and water quality index, and we must establish one ourselves. However, a conversion between secchi depth and water quality index can only be found if we have information for both indices on the same water body, such as in the National Lakes Assessment (NLA) (Holdsworth (2011)).

The NLA is a survey conducted by the U.S. Environmental Protection Agency in 2007. It was designed to assess, without bias, the water quality of the nations lakes, ponds, and reservoirs. A total of 1,028 lakes were sampled from across the nation. Excluding missing data, the public data has a sample of 1094 observations over two years. The NLA data report five water attributes that are used in the National Sanitation Foundation water quality index: dissolved oxygen, total phosphate, nitrate, turbidity, and pH level, and it also provides secchi depth for each water body. We constructed the water quality index from the five attributes using the National Sanitation Foundation formula above. The available parameters and adjusted weights are presented in Table 2.2. The summary statistics of the NLA data: the Q values of dissolved oxygen, pH, total phosphate, nitrates and turbidiry, the secchi depth (in meter), and the constructed water quality index (wqi) are shown in Table A.1 in the appendix. Now that we have the water quality index and secchi depth for the same lake, we can establish a link between the two.

Table 2.2: Water quality parameters and weights (adjusted)

Parameter	**Weight (Adjusted)**
dissolved oxygen	0.30
pH	0.20
total phosphorous	0.18
nitrates	0.18
turbidity	0.14
Total	1.0

2.4.2 Use of Eureqa to Find the Conversion

Since there is no scientific basis for a specific functional form between the water quality index and secchi depth, we would like to use a tool to help identify the best functional form as well as the parameter values. Although a more complex structure often means a better model fit, we do not want the model to be too complex. The tool we used to make this trade-off is "Eureqa" (Dubčáková (2011)).

Eureqa (or Eureqa Formulize) is a scientific data mining software that searches for mathematical patterns and relationships hidden in data. Behind Eureqa is a method called symbolic regression (Zelinka et al. (2010)). The biggest difference between Eureqa and conventional regression is that Eureqa does not impose any prior structures or specific functional forms before the search, so it is very flexible. Eureqa also shows the complexity of each model, depending on the number of terms and order of terms, as well as mean absolute error or fitness of the model. We chose from a dozen candidate models provided by Eureqa to strike a balance between complexity and fitness, as shown in Figure 2.1.

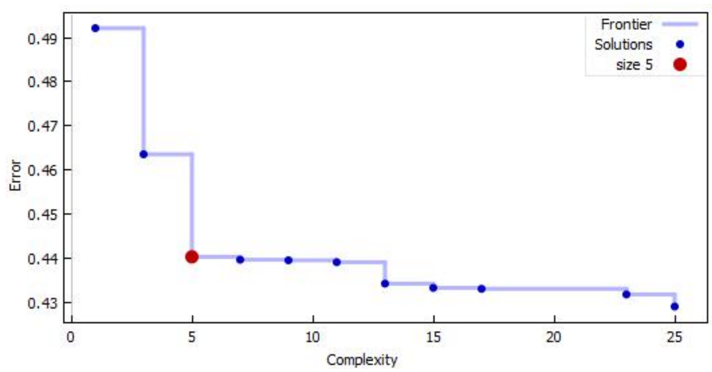

Figure 2.1: Complexity versus fitness provided by Eureqa

Each dot in Figure 2.1 represents a functional form and parameter estimates. The "Error" of the model is defined as mean absolute error between fitted and actual data and is plotted for each model on the vertical axis. The "Complexity", or size, of a function is defined as the number of operations in the function and is plotted on the horizontal axis.

The model we chose to use for the conversion between water quality index and secchi depth is the one identified by the red spot. We felt this model had a relatively simple form and good fitness. The slightly more complex models (the three blue dots to the right of the red dot) are not very different from the selected one, so the results should not be sensitive to our choice of the conversion. The model we chose is,

$$WQI = 78.9 + S + \frac{1.95}{0.06 - S^2} \tag{2.3}$$

where S is secchi depth and WQI stands for water quality index. The raw plots from NLA and fitted plots using Eq 2.2 are shown in Figure 2.2 (We truncate the data at secchi depth less than or equal to five meters to give a better visualization of model fit). There is a positive relationship between WQI and secchi depth as expected. The mapping from secchi depth to water quality index takes the shape as shown in Figure 2.2. When the secchi depth is small, i.e. when the water is not clear, a small increase in the secchi depth will result in a relatively large increase in water quality index. As the secchi depth becomes bigger, the curve flattens out, meaning that an increase in the secchi depth will not lead to as much of an increase in the water quality index.

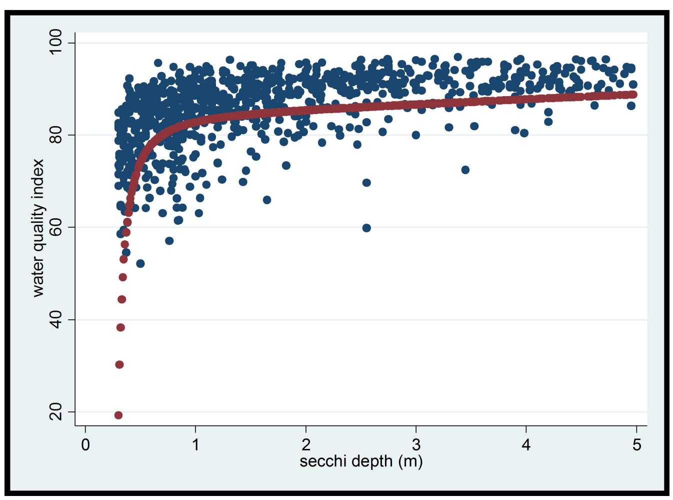

Figure 2.2: Mapping secchi depth to water quality index

With the use of Eureqa, we now have the means to convert secchi to the water quality index. Each observation in our data set has a water quality index that was either taken from the original study, or converted from other indicators. In the next section, we describe the data and estimation of the willingness to pay function.

2.5 WTP Function and Data Specification

Following Van Houtven et al. (2007), we define a WTP function to depend on initial water quality Q^0, the change in water quality ΔQ, and other characteristics (control variables) such as affected site, surveyed sample, and research methodology. A simple representation is

$$\text{WTP} = V(Q^0, \Delta Q; \text{control variables}) \tag{2.4}$$

To estimate Equation 2.4, we specify a linear regression model where the dependent variable is annual willingness to pay (WTP) per household in 2010 U.S. dollars. The independent variables include the initial water quality index (startingWQI) and the change in the water quality index (deltaWQI). We control for: (a) site characteristics, such as site size, type and location, (b) sample characteristics, such as income and sample regions, and (c) research methods, such as publication date, model approach, elicitation method, and water quality indicator. We use clustered robust regression, where each study is treated as a cluster (Williams (2000)).

Among the large body of papers on the valuation of aquatic resources, only those that meet three criteria are included in the meta data set. First, the studies must have considered changes in water quality. A paper that assesses the value of a lake without concerning any change in the lakes water quality, for example, will not be included. Second, even if the paper considers change in water quality, it must be measured in one of the three ways described in 2.4. In other words, it must be measured by secchi depth, the water quality index, or other water attributes; otherwise we cannot convert it to a consistent water quality index and include it in the data. For instance, some papers evaluate the value of implementing a specific management plan, but do not provide information on the water quality improvement. Another example is studies that use fish catch rate as the measure of interest. These papers focus more on the economic value of fishing or value of fishing as a recreational activity. Finally, we considered sites in the

United States only.

As a result, we have collected 332 observations from 37 distinct studies, including 25 journal articles, one book chapter, six government agency reports, one Masters dissertation, four PhD dissertations, and one working paper. Some explanatory variables such as publication year, income, and site type are shared by all estimates reported in a study; other variables, such as elicitation method are only applicable to a certain type of study, or only applicable to contingent valuation studies. Table 2.3 summarizes the primary studies and willingness-to-pay estimates included in this study; Table 2.4 lists all 37 studies used in the Meta-Analysis and provides some summary statistics about them; Table 2.5 contains the variable description; Table 2.6 is summary statistics of the variables. Finally Table A.2 lists all 332 willingness-to-pay estimates used in the meta-analysis.

Table 2.3: Summary statistics for studies included in the meta-analysis

	Studies(I=37) Number(Percent)	**WTP Estimates(N=332)** Number(Percent)
Type of Publication		
Peer-reviewed jornal/book	26(68%)	182(55%)
PhD/Mater's thesis	5(13%)	38(11%)
Other (working paper, report, etc.)	7(18%)	112(34%)
Year of Publication		
Before 1980	2(5%)	3(1%)
1980-1989	15(39%)	116(35%)
1990-1999	7(18%)	24(7%)
2000 or later	14(37%)	189(57%)
Type of Water Resource Affected		
Lake	13(34%)	168(51%)
River/stream	14(37%)	106(32%)
Estuary	9(24%)	42(13%)
Other	2(5%)	16(5%)
Regions Experiencing WQ Change		
Northeast	23(61%)	137(41%)
Midwest	11(29%)	166(50%)
South and West	2(5%)	18(5%)
Other	2(5%)	11(4%)
Valuation Method Used		
Contingent valuation and combined	22(58%)	153(46%)
Travel cost	6(16%)	52(16%)
Hedonic property value	10(26%)	127(38%)
WQ Indicator Used		
Ladder	18(47%)	137(41%)
Secchi depth	11(29%)	168(51%)
Water attributes	9(24%)	27(8%)

Table 2.4: List of studies used in the meta-analysis

	Author(s)	Affected Site	Methodology	Water Quality Indicator	#Est
1	Azevedo et al. (2001)	Clear Lake, IA	CV	Secchi	5
2	Bockstael et al. (1987)	Boston area beaches	TC	oil, turbidity, COD and fecal coliform	1
3	Bockstael et al. (1989)	Chesapeake Bay, beaches in Maryland	CV, TC	RFF Ladder	4
4	Boyle et al. (1999)	Lakes in Maine	Hedonic	Secchi	6
5	Boyle and Bouchard (2003)	Selected lakes in Vermont, New Hampshire, and Maine	Hedonic	Secchi	22
6	Brashares (1985)	Lakes in southeast Michigan	Hedonic	Secchi, turbidity, and fecal coliform	7
7	Carson and Mitchell (1993)	National lakes rivers and streams	CV	RFF Ladder	3
8	Croke et al. (1986)	Rivers in Chicago	CV	RFF Ladder	6

Table 2.4 Continued

9	Cronin (1982)	Potomac River, Washington D.C.	CV	dissolved oxygen, fecal coliform, nitrogen, phosphorus, turbidity and recreational activities	20
10	d'Arge and Shogren (1989)	Lake Okoboji, IA	Hedonic	Water quality index	2
11	Desvousges et al. (1987)	Monongahela River, PA	CV	RFF Ladder	16
12	Edwards (1984)	Salt pond lake, RI	CV	Score out of 100	3
13	Egan et al. (2009)	Lakes in Iowa	TC	Improve from current to west Okoboji lake	20
14	Epp and Al-Ani (1979)	Rivers in Rural Pennsylvania	Hedonic	Secchi	1
15	Farber and Griner (2000)	River in Western Penssylvania	CV	3-level scale	18
16	Gibbs et al. (2002)	Lakes in New Hampshire	Hedonic	Secchi	4

Table 2.4 Continued

17	Gramlich (1977)	Charles River, MA, and National lakes rivers and streams	CV	RFF Ladder	2
18	Greenley et al. (1981)	South Platte River, CO	CV	RFF Ladder	12
19	Huang (1986)	Selected lakes in Minnesota	TC	Water quality index	22
20	Johnston et al. (1999)	River in Rhode Island	CV	RFF Ladder	3
21	Krysel et al. (2003)	Mississippi Headwaters Region, MN	Hedonic	Secchi	74
22	Leggett and Bockstael (2000)	Chesapeake Bay	Hedonic	fecal coliform	1
23	Lipton et al. (2004)	Chesapeake Bay	CV	5-step improvement	1
24	Magat et al. (2000)	National lakes rivers and streams	CV	15% improvement	7
25	Mathews et al. (1999)	Minnesota River, MN	CV, Combined	Phosphorous	3
26	Michael et al. (1996)	Lakes in Maine	Hedonic	Secchi	6
27	Moore et al. (2011)	Green Bay, WI	CV	Secchi	24

Table 2.4 Continued

28	Mullen and Menz (1985)	Rivers in Adirondack, NY	TC	fishable	1
29	Randall et al. (2001)	Maumee River, OH	CV	Nitrate	3
30	Schuetz et al. (2001)	Maine's great ponds	CV	Secchi	3
31	Smith et al. (1983)	Monongahela River, PA	CV	RFF Ladder	2
32	Smith et al. (1986)	Monongahela River, PA	CV, TC	RFF Ladder	14
33	Steinnes (1992)	Lakes in northern Minnesota	Hedonic	Secchi	2
34	Stumborg et al. (2001)	Lake Mendota, WI	CV	Phosphorus	1
35	Wey (1990)	Great Salt Pond, RI	CV	1-6 scale	2
36	Whitehead (2005)	Neuse River, NC	Combined	RFF Ladder	6
37	Young (1984)	St. Albans Bay on Lake Champaign, VT	Hedonic	1-10 scale	2
				TOTAL:	332

Table 2.5: Description of dependent variables

Variable	**Description**
WTP	willingness to pay in 2010 dollars
D_NE	=1 if the affected water bodies are in the Northeast
D_lakeEstuary	=1 if the affected water bodies are lakes and estuaries
pubDate	publication year, 0=year 1977
D_inPerson	=1 if the survey was administered with an in-person interview
income	average household income in 2010 dollars
D_totalValue	=1 if the original study estimates total value
D_improvement	=1 if the change in water quality is an improvement
D_index	=1 if the water quality indicator used in the original study is an index
startingWQI	starting water quality index of affected water bodies
deltaWQI	change in water quality index of affected water bodies
D_CV	=1 if the original paper uses contingent valuation method
D_hedonic	=1 if the original paper uses hedonic method
D_openended	=1 if elicitation method is open-ended
D_bidding	=1 if elicitation method is iterative bidding
D_elitmtdOther	=1 if elicitation method is other
sitesize	the size of the affected water bodies in square miles
regionsize	the size of the sampling region in square miles

Publication date, pubDate, is the year that the study was published, or for unpublished studies, the date it first became available. The base year is 1977, in which pubDate equals 0. Publication date may differ from when the data is collected. We used the latter to convert the value to 2010 dollars. Income is the median income of the state where the sampled households live. The state-level median income is only a crude measure of sampled households mean income. Since not all studies in the meta-base have reported the mean income of their sample, we used the state median income as a proxy (DeNavas-Walt (2010)). The dummy variable for total value, D_totalValue, captures whether the valuations from the original studies include use value, non-use value only, or if they are total value (i.e. the sum of both use and non-use values). The variable D_totalValue takes the value 0 for all revealed preference papers, since use value is the only component that they can measure. Improvement dummy, D_improvement, equals 1 if the valuation is for water quality improvement and it equals 0 if the valuation is for avoiding water quality

Table 2.6: Summary statistics of data used in meta-analysis (N=332)

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
WTP	312.14	679.54	3.08	5491.65
D_NE	0.41	0.49	0	1
D_lakeEstuary	0.63	0.48	0	1
pubDate	19.64	9.63	0	34
D_inPerson	0.22	0.41	0	1
income	51582.9	6606.17	39701	69047
D_totalValue	0.29	0.45	0	1
D_improvement	0.75	0.43	0	1
D_index	0.41	0.49	0	1
startingWQI	61.2	26.62	5	92
deltaWQI	16.3	19.18	0.42	85
D_CV	0.44	0.5	0	1
D_hedonic	0.38	0.49	0	1
D_openended	0.13	0.34	0	1
D_bidding	0.1	0.3	0	1
D_elitmtdOther	0.12	0.33	0	1
D_dichotomous	0.11	0.31	0	1
sitesize	7908.13	43873.87	0.22	256481.23
regionsize	119851.56	648653.95	0.22	3794101

degradation. The three dummy variables D_openended, D_bidding, and D_elitmtdOther are elicitation methods in surveys. The default is dichotomous choice. The variable sitesize is the size of the affected water body, and the variable regionsize is the size of sampling region. We expect region size to have a negative effect on the willingness to pay, because peoples willingness to pay for a site depends on the accessibility of the site, and a bigger region means less accessibility on average.

2.6 Results

For most studies, more than one observation is included, and as a result, our data is naturally clustered. Observations from the same study may exhibit dependency not present in observations from different studies. One source of this dependence is the same observations or study from which the estimates are obtained. Other factors, such as author and journal effect, can also cause dependency. In short, observations from the same study may have different correlation structures than the ones from different studies. To take into account the clustered nature of the data, we use clustered robust regression where each study is a cluster instead of standard OLS (Williams (2000)).

Table 2.7 shows the clustered robust regression results with water quality index as the water quality indicator (the regression results with Secchi as the water quality indicator are shown in Table A.3 in the Appendix). The column labeled Pooled 1 is the regression results using the full data, with all explanatory variables; the column labeled Pooled 2 is the regression results from the pooled data, with all explanatory variables except site size and region size; the column labeled CV is the regression results from the CV papers only; and the column labeled Hedonic is the regression results from the hedonic papers only. Only 52 observations fell into the travel cost category, so a regression on that sub-sample was not estimated due to its small size.

The results on the full (pooled) data (columns 1 and 2) show that peoples willingness to pay does depend on the (absolute) level of change in the water quality index. For a 10-point change (out of 100) in the water quality index, an average household is willing to pay around $45. In addition, willingness to pay for given water quality improvement is higher for lakes and estuaries than for rivers. It is also higher when the survey is administered in person, or when water quality is indicated by secchi depth, as opposed to the water quality index. Moreover, people are willing to pay more for avoiding degradation than making an improvement; and people are also willing to pay more for an improvement in water bodies with bad initial condition than those with already good initial conditions, reflecting the declining marginal utility in water quality. The hedonic dummy is positive and significant, while the CV dummy is negative and significant. So the hedonic approach tends to produce larger valuations, followed by the travel cost approach (the default), which is followed by the contingent valuation approach.

Region and site size have significant impacts on the willingness to pay for water

Table 2.7: Clustered robust regression results

	Pooled 1	Pooled 2	CV	Hedonic
	(1)	(2)	(3)	(4)
D_NE	27.94	-2.76	-72.62***	21.31
	(83.13)	(72.63)	(26.76)	(110.34)
D_lakeEstuary	287.23**	274.01**	268.11***	1507.93**
	(112.97)	(124.26)	(59.78)	(631.09)
pubDate	4.69	3.75	-2.95	-60.41
	(5.54)	(5.50)	(2.27)	(41.69)
D_inPerson	284.09***	283.37**	133.32***	
	(109.18)	(110.66)	(30.77)	
income	-.01	-.01	.003	-.009
	(.01)	(.01)	(.003)	(.02)
D_totalValue	78.96	92.80*	104.49***	
	(56.64)	(55.98)	(31.01)	
D_improvement	-212.50*	-193.56*	12.91	-272.56
	(110.19)	(106.33)	(33.34)	(181.26)
D_ladder	-208.04*	-142.73	-141.76*	-13773.26***
	(109.72)	(98.50)	(76.35)	(569.44)
startingWQI	-2.67*	-1.89	-.37	121.04***
	(1.38)	(1.31)	(.58)	(13.78)
deltaWQI	4.48*	4.62*	1.67***	142.94***
	(2.40)	(2.42)	(.50)	(18.18)
D_CV	-277.26*	-123.59		
	(146.45)	(125.98)		
D_hedonic	217.88*	349.16**		
	(120.53)	(136.81)		
sitesize	.06**		.003	29.88***
	(.03)		(.02)	(.62)
regionsize	-.004**		-.0002	
	(.002)		(.001)	
N	332	332	146	127
r^2	.13	.12	.46	.51
F	12.41	11.25	29.47	.

[a] *p=.10 **p=.05 ***p=.01

[b] Pooled 1 is the pooled regression with site and region size.

[c] Pooled 2 is the pooled regression without site and region size.

quality improvement. Site size, the size of the affected water bodies, has a positive effect. The willingness to pay for a given water quality improvement in an aquatic site will be \$0.60 higher if the site size increases by 10 square miles. Region size, the size of sampling region, has a negative effect. The willingness to pay, on average, will be \$4 lower if we expand the sampling region by 1,000 square miles. We conjecture that this is because the further away a household lives from the site, the less accessible the site is to the household, and the less important the quality of the site is to the household. The pooled results are robust to the inclusion of region and site size.

Columns 3 and 4 show the regression results for the CV and hedonic papers, respectively. Compared with the general population or local residents, from which the sample of most CV studies are drawn, homeowners, the sample of virtually all hedonic models, are more responsive to water quality change in the site on which their houses sit. For hedonic papers, the region size and site size are highly correlated, because most properties are on the site, we therefore included only site size in the regression. On-site property values respond positively to site size. A Chow test rejects the null hypothesis that the groups share the same coefficients. As noted below, existent meta-analyses only include CV papers. Yet, our study suggests that valuations from CV studies are, on average, the smallest among the three approaches. As a result, benefit transfers based only on CV studies could be biased downward.

Table 2.8 shows the predicted annual willingness to pay per household (in 2010 dollars) for different levels of water quality improvement using the pooled results in Column 1 in Table 2.7 as a demonstration of the values that the meta regression generates. For example, for a small site that is only one square mile, (such as Little Spirit Lake in Iowa), a household living in 50,000 square mile area around the site is estimated to be willing to pay \$115.14 for a 5-point increase (from 40 to 45) in the water quality index. Naturally, willingness to pay is larger for a big site than for a small one, and is also larger for a 10-point increase in water quality than for a 5 points increase.

Table 2.8: WTP for water quality improvement

site type \ WQI change	40 to 45	40 to 50	70 to 75	70 to 80
small site (1 sq mi) (Little Spirit, IA)	115.14 (143.84)	137.52 (141.78)	35.12 (142.33)	57.50 (139.9)
medium site (100 sq mi) (Lake Winnibigoshish, MN)	121.46 (141.32)	143.85 (139.29)	41.44 (139.62)	63.83 (137.23)
big site (10,000 sq mi) (Great Lakes)	753.89 (210.71)	776.27 (213.64)	673.87 (197.95)	696.25 (200.84)

[a] standard error in parenthesis
[b] in 2010 dollars
[c] sample region: 50,000 square miles

2.7 Comparison With Other Meta Analyses on Aquatic Sites

Two other meta analyses on the valuation of water quality improvement have been completed; one is Van Houtven et al. (2007) and the other is Johnston et al. (2005). Our work differs from these in three important ways. First, neither Van Houtven et al. (2007) or Johnston et al. (2005) controlled for site size and region size in the meta-regression of their papers.

Second, both Van Houtven et al. (2007) and Johnston et al. (2005) limit their analysis to contingent valuation (CV) studies, so papers using hedonic and travel cost approaches are excluded. When doing meta-analysis, we make a trade-off between including more studies and having a bigger sample size and consistency across studies. Van Houtven et al. (2007) and Johnston et al. (2005) include only CV papers in their meta-database to ensure consistency across studies. However, there is no evidence that not including papers using the other two approaches is the optimal trade-off. One benefit is, of course, a larger sample size: we obtain 332 observations from 37 unique studies, compared with 81 observations from 34 studies in Johnston et al. (2005), and 131 observations from 18 studies in Van Houtven et al. (2007). Moreover, it enables us to compare valuations from different approaches. If the hedonic and travel cost studies systemically produce larger valuations than the CV papers, benefit transfer based only on CV papers does not fully use the knowledge in existing literature, and is likely to be biased downward.

Third, both Van Houtven et al. (2007) and Johnston et al. (2005) limit their input to only studies using the RFF water quality ladder. Although this ladder is often used, other indicators such as Secchi depth are also used by a large number of studies. Our study appears to be the first one to estimate a link between Secchi depth and the water quality index.

2.8 Concluding Remarks on Chapter 2

This study is an attempt to answer the important question: How much is clean water worth? We do so by developing and estimating a valuation model based on a meta-analysis on non-market valuations of water quality improvements. After reviewing more than 100 non-market valuation studies on aquatic sites, we have 332 valuations from 37 distinct existing studies in the meta-database. The valuation model estimated in this study can be used to predict the mean willingness to pay by households living in a given region for water quality improvement in a given site.

We first developed a link between water quality index and Secchi depth, based on national lakes assessment (NLA) data. Eureqa, a data-mining software, enabled us to search for a model that is extremely flexible in functional structure. What we found, through Eureqa, is a link between water quality index and Secchi depth that has a flexible, but relatively simple, function form and reasonably good model fit. We then used the link to convert Secchi depth to water quality index, so that all observations in the metadata set have consistent water quality measurement (i.e. water quality index), a key component in our model. This completes the data set for the meat-analysis.

We then used the completed data set to estimate a valuation model for water quality improvement. Some findings from estimation results included: for a 10-point (out of 100 points) additional change in water quality index, a households willingness to pay will increase by $45. Willingness to pay is higher for lakes and estuaries than for rivers. It is also higher if the survey in the original study is administered in person. Willingness to pay is lower for making improvement than for avoiding degradation. It is also lower if we started with an already good initial water quality condition, probably reflecting the decreasing marginal utility of water quality. We found that both size of the affected site and size of sampling region have a significant effect on willingness to pay. Site size has a positive effect and region size has a negative one, perhaps due to less accessibility of the site as the region became bigger.

The valuation model we estimated in this study enables us to predict the mean willingness to pay in a given region for water quality improvement in a site of certain size. For example, an average household living in a 50,000 square mile region around a given site is willing to pay$115.14 for a 5-point water quality improvement (from 40 to 45) in a one-square-mile aquatic site. This tool is particularly convenient and useful when we want to evaluate the benefit of a project that aims at improving water quality, but a primary study is regarded as too costly or time consuming. We also test the null hypothesis that the three main approaches in non-market valuationthe hedonic model, the travel cost model, and the contingent valuation modelgenerate consistent valuation estimates. Our test results reject the null hypothesis. We found that, among the three approaches, the hedonic model tends to produce the largest valuation, the second-largest were produced by the travel cost model, and the third-largest were produced by the contingent valuation model.

This study is different from other meta-analyses, to which two were paid particular attention in three important ways: (a) this paper includes studies using all three dominant approaches in non-market valuation, while others only include contingent valuation paper; (b) this paper includes studies using different water quality indicator, like Secchi depth, while others only include studies using RFF water quality ladder; and, (c) in this paper we controlled for size of the affected site and size of sampling region, while others have not, and found both to have a significant effect.

There are critics of using meta-analysis for benefit transfer, and some doubt its accuracy. Others argue that studies used in meta-analysis should be very restrictive to ensure consistency, and both are valid concerns. After all, hedonic, travel cost, and contingent valuation models are very different approaches to the same problem. However, if all three models have been standing side by side in the non-market valuation literature for decades, and if the valuations they produce have important policy implications, maybe it is worth the effort to examine them on the same plate, and have a valuation model that is inclusive of all three approaches.

CHAPTER 3. ENDOGENOUS RISE AND COLLAPSE OF HOUSING PRICES

3.1 Introduction to Chapter 3

The housing market is very unique. It has several characteristics that other commodity markets do not have. For example, it is highly leveraged, i.e. most houses are bought with mortgage loans, and in a typical mortgage contract, the house bought with the help of the loan is used as the collateral for the loan. Moreover, mortgage insurance (issued by government or private insurance company) is often required for mortgages with little (usually less 20%) upfront or down payment. Another unique feature is that a house's can be both a consumption and an investment good. As an investment good, return on housing in a boom market often beats that on stocks. The third unique feature is that the housing market is a thin market with relatively small trading volumes, and there is an extended time on market associated with the selling of a house. Interestingly, the trading volume and average time on market are closely correlated with the market cycles. Finally, a house cannot be separated from its location. There is no such thing as a standard house; Every house is unique because every location is unique. All these characteristics make a housing market different from other commodity or financial markets.

U.S. housing prices started to rise around year 2000. The average price of housing then almost doubled in just a few years. It kept rising until reaching its peak at the end of 2006. When housing prices collapsed, the consequences were devastating; Nearly

$11 trillion in household wealth has evaporated because of it. America went into the most severe recession since the big depression in the 1930s. Likewise, Japan's property price rose quickly in the late 1980s. It collapsed in the early 1990s, causing economic slow down in years after. Similar large-scale rises and falls in housing prices took place in European countries such as Norway, Spain and Ireland. Large house price volatilities are even more pronounced in emerging economies such as South Korea, Russia, China, India, each time causing large interruptions in the economy.

Although property makes up a significant portion of national economy (by one estimate, construction, the buying, selling and renting of properties and the imputed benefits to owner-occupiers account for around 15% of rich countries' GDP), so far less economic research has been done on the housing market than on the stock market, the bond market or the foreign-exchange market. There are many open questions remained in the area (Mayer (2011)). The need for a deeper understanding of the housing market is confirmed in the final report by the national commission on the causes of the recent financial and economic crisis in the United State (Commission (2011)). This study is an attempt to fill the gap.

This study makes two major contributions. First, it develops an agent-based computational model of the housing market. The model will take into account the unique characteristics and complications mentioned above. It is also flexible enough to be used as a platform by other researchers for future housing market analysis. For example, the model can be adapted and used to study a local housing market or to analyze a particular policy scenario. Second, this study attempts to answer the following research questions: will housing prices rise and collapse endogenously without an external shock? If so, under what conditions? Our answer is yes to the first question. To the second question, our answer is it will happen when banks engage in lenient financing and there is speculation in the market.

We first develop a simple analytical framework to demonstrate how housing prices can endogenously rise and fall. With simplified assumptions, we show the two market dynamics that together lead to cycles in the housing market. The simple framework helps us identify the two key treatment factors in the computational experiment: leniency and speculation. Since the analytical framework is based on simplified assumptions, we next develop a more flexible and complex agent-based model that incorporates a richer setting for housing markets. We then use the agent-based model to analyze the effects of speculation and lenient lending on housing price volatility, home-ownership rates, foreclosure rates, mortgage rates and the bank's expected profit.

Some of our key findings are as follows. First, we find that under certain circumstance, housing prices rise and collapse by themselves without any external shocks. We find that lenient financing or high leverage combined with speculation is responsible for the endogenous rise and fall of housing prices. Moreover, we find that it is possible to set the down payment rate at an optimal level and achieve both affordable housing and market stabilization. Finally, we find that, with lenient financing, banks and financial institutions will have incentives to adopt lenient financing even without the securitization of mortgage loans.

The paper is organized as follows. Section 3.2 presents a review of related literature. Section 3.3 presents the analytical framework of housing market. Section 3.4 presents the model logic of the agent-based housing market model. Section 3.5 presents the treatment factors and experimental design. Section 3.6 presents the results. Section 3.7 presents the concluding remarks. For more details of the agent-based model, the UML diagrams of the model's class structure and activity flow can be found in Appendix B.4; A detailed structural framework can be found in Appendix B.3.

3.2 Related Literature

There are different ways to model the housing market. Theoretically, Poterba (1984) developed an asset-market model of the housing market and estimated how changes in the expected inflation rate affect the real price of houses. Stein (1995) used a simple model of trade in the housing market to show that price volatility and trading volume can be explained by minimum down payment requirement through repeated buyers. Iacoviello (2005) developed a monetary business cycle model and found that collateral effects dramatically improve the response of aggregate demand to house price shocks. Silos (2007) constructed a business cycle model with heterogeneous agents to investigate the properties of the wealth distribution and the portfolio composition regarding housing and equity holdings. Finally, Sommervoll et al. (2010) developed a heterogeneous agent model illustrating the connection between adaptive expectations and housing market fluctuations.

Empirically, Case and Shiller (2003) measured the extent of housing bubble in the U.S. housing market long before it became obviously dangerous. Interestingly, in 1989 Mankiw and Weil (1989) predicted that real house price would fall substantially in the next two decades by looking at the historical relation between housing demand and housing prices. Glaeser et al. (2008) investigated housing supply and found that the price run-ups of the 1980s were almost exclusively experienced in cities where housing supply is more inelastic.

As for agent-based computational models [1], Markose et al. (2009) developed agent-based models to study the relationship between property market and the value of financial securities. Torrens (2001) developed a multi-leveled agent-based model for individual's housing choices, but without the issue of mortgage. Geanakoplos et al. (2012) proposed an agent-based model of the housing market with a sophisticated mortgage structure. Agent-based models of housing market tend to fall in two categories: those have a detailed

[1] See Appendix B.1 for an introduction to agent-based computational method.

spatial landscape but no financial sector, and those have a sophisticated financial sector but no spatial landscape. Our work will fill the gap: our model has both a spatial landscape and a financial sector.

Regarding the role of leverage in business cycles, Geanakoplos (2010) found that a small shock in one sector can cause wide-spread crises across sectors with independent payoffs because of the leverage constraints that connects all sectors. In another study by Thurner et al. (2012), the authors argued that it is leverage, not interest rate, that causes fat-tailed return distribution and clustered volatility. They show that even a small negative shock in the market will trigger a large price drop, because leveraged investors are forced to sell assets to stay within the leverage limit. Similarly, our study will show that high leverage in the housing market is responsible to cause the endogenous rise and collapse of housing prices.

3.3 An Analytical Housing Market Model

This section presents an analytical housing market model. Section 3.3.1 presents an overview of the model. Section 3.3.2 presents the model in summarized form. Section 3.3.3 presents the economic interpretation of the model. Finally, section 3.3.4 presents phase diagrams for the analytic model with special functional forms.

3.3.1 Overview of the Analytic Model

As previously stated, one unique characteristic of housing markets is that they are highly leveraged, meaning most houses are bought on mortgage loans. As a result, the cost of a house consists of two parts: the purchase price and the cost of financing. The simple analytic housing model includes the lending behavior of a mortgage-issuing bank, which is a function of the house price and mortgage rate, and the net demand of the non-speculative and the speculative house buyer, which again is a function of the house price and mortgage rate. There are two markets that we model: the mortgage market and the housing market, and two state variables: the house price at time t, $p(t)$), and the mortgage rate (including insurance premium) at time t, $m(t)$. The model setup is based on a non-arbitrage condition in the mortgage market and a market clearing condition in the housing market.

The analytic model is able to capture the reinforcing feedback loop between the housing market and the mortgage market. When the house price is high, mortgages backed up by houses are regarded as a safe asset, and the resulting mortgage rate is low. A low mortgage rate will in turn fuel demand for housing and push the price even higher. On the other hand, when the house price is low, mortgages backed up by houses become more risky and lenders require a higher mortgage rate to compensate for the increased riskiness, which will in turn suppress demand for housing and further lower the house price. We will analyze the role of down payments as a protection against price fluctuations.

We develop the analytic model to demonstrate the basic underlying mechanism that drives the house price to rise and collapse endogenously. We use it to help identify one of the key treatment factors in the model: the level of leniency, measured in terms of the down-payment requirement. We also use it to help set the stage for a more sophisticated and full-fledged modeling of the housing market. The analytic model is highly simplified and omits some important aspects and market participants of the housing market. To incorporate these complications we need a more flexible modeling tool. a model with a fuller set of market participants in a more realistic setting will be developed in a later section.

3.3.2 Analytic Model in Summarized Form

In this section we present the analytical model in summarized form. We start with the model equations. For all $t \geq 0$

$$0 = g(p(t), cr(t); \theta) \tag{3.1}$$

$$\dot{m}(t) = \frac{dm(t)}{dt} = S_1\left(m(t), p(t), cr(t); \theta\right) \tag{3.2}$$

$$\dot{p}(t) = \frac{dp(t)}{dt} = S_2\left(m(t), p(t), cr(t); \theta\right) \tag{3.3}$$

Time-t endogenous variables: For all $t \geq 0$,

- $\dot{m}(t)$: change in mortgage rate during period t
- $\dot{p}(t)$: change in house price during period t
- $cr(t)$: collateral rate at the beginning of period t

Time-t predetermined (state) variables: For all $t > 0$,

- $m(t)$: mortgage rate (including insurance premium) at the beginning of period t
- $p(t)$: house price at the beginning of period t

Exogenous variables and functional forms:

- $g : R^2 \to R$
- $S_1 : R^3 \to R$
- $S_2 : R^3 \to R$
- m_0: initial mortgage rate (including insurance premium)
- p_0: initial house price
- θ: other exogenous variables and functional forms, which will be explained in more detail in Section 3.3.3 along with admissibility restrictions

3.3.3 Economic Model Interpretation

In this section we provide an economic interpretation of the model. We will explain the exogenous variables and functional forms and the admissibility restrictions. We start with the mortgage market and then move on to the housing market. In the mortgage market, a lending contract requires collateral to protect the lender from loan default. Collateral in a typical mortgage loan contract is the house being acquired with the loan. Mortgages with lower than 20% upfront or down payment are required to have mortgage insurance either from the government or a private insurance provider. Collateral rate, $cr(t)$ is defined as the value of collateral over the value of loan, which has an upper limit at the market risk-free rate:

$$\begin{aligned} cr(t) &= \max\left\{\frac{p(t)}{(1-\text{down})\,p_0}, 1+r^F\right\} \\ &= \begin{cases} \frac{p(t)}{(1-\text{down})p_0} & \text{if } p(t) < (1+r^F)(1-\text{down})p_0 \\ 1+r^F & \text{if } p(t) \geq (1+r^F)(1-\text{down})p_0 \end{cases} \end{aligned} \tag{3.4}$$

where p_0 is the initial purchase price, down is the minimum down payment rate required by the bank, and r^F is the risk-free return rate. For simplicity, we assume that the loan value equals the purchase price minus down payment. Therefore an increasing relationship between $cr(t)$ and $p(t)$ exist only if $p(t)$ is below the cutoff at $p_0\left(1+r^F\right)(1-\text{down})$. When $p(t)$ is above the cutoff, $cr(t)$ is fixed at $1+r^F$. Equation 3.4 defines the relationship between the collateral rate $cr(t)$ and the house price $p(t)$, conditional on exogenous variables. Equation 3.4 can thus be expressed in the form of Equation 3.1.

We assume perfect competition among funds for mortgage loans. Moreover, because insurance is required for mortgages with little upfront or down payment, the absence of arbitrage opportunities requires that the expected return rate on mortgage loans equals the exogenous risk-free return rate,

$$0 = \text{prob}^d(cr(t)) \cdot cr(t) + \left(1 - \text{prob}^d(cr(t))\right) \cdot (1 + m(t) + \dot{m}(t)) - \left(1 + r^F\right) \tag{3.5}$$

where $\text{prob}^d(cr(t))$ is default probability function, which we assume to be decreasing in collateral rate, $\frac{d\text{prob}^d(cr(t))}{dcr(t)} < 0$, and $0 \leq \text{prob}^d(cr(t)) \leq 1$. We then substitute $cr(t)$ with the function of $p(t)$. Therefore Equations 3.4 and 3.5 together define $\dot{m}(t)$ as a function of $m(t)$, $p(t)$ and $cr(t)$, given exogenous variables. Thus, Equation 3.5 can be expressed in the form of Equation 3.2 .

We use a net supply function to represent the difference between housing supply and demand. We assume the net supply function is a natural log function that is increasing in price. Market clearing (at a positive price) requires that net supply equal zero,

$$0 = ln\left((1 + m(t))\left(p_t + \dot{p}(t)\right)\right) \tag{3.6}$$

Equation 3.6 implicitly defines $\dot{p}(t)$ locally as a function of $m(t)$ and $p(t)$. Thus, Equation 3.6 can be approximately expressed in the form of Equation 3.3.

Take the first partial derivatives of the function $\dot{m} = S_1(m, p, cr; \theta)$ with respect to m and p, to get:

$$\begin{aligned} \frac{\partial S_1}{\partial p} &= \begin{cases} 0 & \text{if } p \geq p_0\left(1 + r^F\right)(1 - \text{down}) \\ \frac{-\text{prob}^d(cr)(1-\text{prob}^d(cr)) + \text{prob}^d(cr)'(1+r^F-cr)}{\left(1-\text{prob}^d(cr)\right)^2} < 0 & \text{if } p < p_0\left(1 + r^F\right)(1 - \text{down}) \end{cases} \\ \frac{\partial S_1}{\partial m} &= -1 < 0 \end{aligned} \tag{3.7}$$

Similarly, take the first partial derivatives of the function $\dot{p} = S_2(m, p, cr; \theta)$, to get:

$$\begin{aligned} \frac{\partial S_2}{\partial p} &= -1 < 0 \\ \frac{\partial S_2}{\partial m} &= -\frac{1}{(1+m)^2} < 0 \end{aligned} \tag{3.8}$$

3.3.4 Phase Diagrams for a Special Case

The dynamic properties of solutions derived in the previous section apply to all exogenous variable values and functional forms that satisfy the admissibility restrictions. We now use the following functional forms and exogenous variable values to derive a phase diagram for the system:

$$\begin{aligned}
\text{prob}^d(cr) &= 1 - \frac{cr}{1+r^F} \\
r^F &= 0 \\
p_0 &= 1
\end{aligned} \tag{3.9}$$

For these specifications, Equation 3.5 and 3.6 reduce to the following forms, with $cr(t)$ substituted out using Equation 3.4,

$$\begin{aligned}
\dot{m}(t) &= \begin{cases} -m(t) & \text{if } p(t) \geq (1-\text{down}) \\ \frac{1-\text{down}}{p(t)} + \frac{p(t)}{1-\text{down}} - 2 - m(t) & \text{if } p(t) < (1-\text{down}) \end{cases} \\
\dot{p}(t) &= \frac{1}{1+m(t)} - p(t)
\end{aligned} \tag{3.10}$$

Therefore,

$$\begin{aligned}
\dot{m}(t) = 0 \Leftrightarrow m(t) &= \begin{cases} 0 & \text{if } p(t) \geq (1-\text{down}) \\ \frac{1-\text{down}}{p(t)} + \frac{p(t)}{1-\text{down}} - 2 & \text{if } p(t) < (1-\text{down}) \end{cases} \\
\dot{p}(t) = 0 \Leftrightarrow p(t) &= \frac{1}{1+m(t)}
\end{aligned} \tag{3.11}$$

The above system has a locally stable equilibrium at $\bar{m} = 0, \bar{p} = 1$. (See Appendix B.2 for a proof of local stability.)

Figure 3.2 shows two phase diagrams corresponding to down payment requirements of 20% and 0% respectively. The $\dot{p} = 0$ curves are the same in the two diagrams. When price is above the $\dot{p} = 0$ curve, it will go down; and when price is below the curve, it will go up. On the other hand, the $\dot{m} = 0$ curves are different in the two diagrams due to the different down payment requirements. There is a vertical portion on the left of the curve, whose length equals the level of the minimum down payment (down). The mortgage rate will only increase if the house price falls below the cutoff, 1-down; otherwise it stays at zero.

The phase diagrams in Figure 3.2 demonstrates two different types housing price fluctuations resulted from different down payment requirements. In the left diagram where

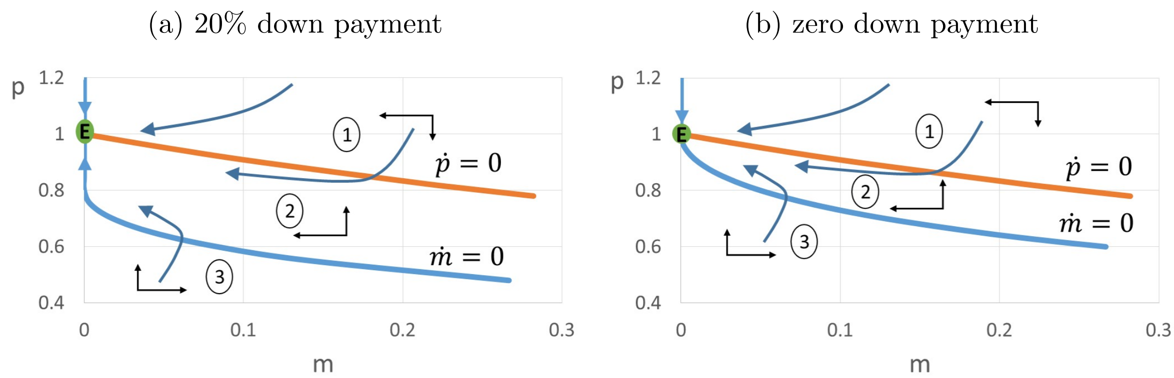

Figure 3.2: Phase diagrams for the mortgage rate (m) and housing price (p) under different down payments

down payment requirement is high, when the house price is slightly below equilibrium, it goes up back to the equilibrium price while remaining at the same mortgage rate. In the right diagram where the down payment requirement is low, however, when house price is slightly below equilibrium, it does not go straight up. Instead, it goes to the right where mortgage rate is higher and house price is lower. As a result, price fluctuation is more significant when down payment requirement is low, because mortgage rate is more sensitive to house price when little down payment is required.

We have shown in the above framework how the down payment requirement can play an essential role in housing market dynamics and cause housing prices to rise and collapse endogenously in the system. Knowing this is important because it helps us understand one of the underlying mechanisms that drives the system dynamics. It also helps us to identify one of the two key treatment factors in our computational model: the degree of leniency, in particular, the down payment requirement.

However, the above framework makes simplifying assumptions about the housing market in many ways. For example, actual housing markets do not always clear. It usually takes an extended period of time for a listed house to be sold. Secondly, we have not yet incorporated speculative demand, which also plays an important role in determining market price volatility as will be shown later. Thirdly, every house is different with its unique location properties. Location is an important determination of the value of a house, and price movements will in turn alter the attractiveness of the location. The analytic framework also ignores the typical long search period and negotiation process in the buying and selling of a property. Finally it does not account for the heterogeneity among buyers and sellers. All these complications are beyond the scope of our simple analytic model. For these reasons we next proceed to the development of a more flexible agent-based computational housing market model.

3.4 An Agent-Based Housing Market Model

For the sake of space, this section will only use a model diagram and a flow chart to demonstrate the workings of the model. For more details, the structural framework of the housing market model can be found in Appendix B.3; the UML diagrams of the model's class structure and activity flow can be found in Appendix B.4. As previously stated, the housing market is very unique: it is less liquid, spatial, highly leveraged, and can be both a consumption and an investment good. To this day, there are few models of housing market that take into account all of these complications. The agent-based model proposed in this paper can potentially capture all the unique characteristics mentioned above.

As previously stated, a housing market cannot be separated from the landscape in which it sits. In this model, the housing market sits on a two dimensional landscape that contains 25 regions: a downtown, suburbs and rural areas as shown in Figure 3.3. Two regions are neighbors if they share a common border. Each region is assigned an exogenous location quality, which is represented by the number in each region in Figure 3.3). The location quality captures exogenously factors that affect quality of life in that region, such as distance to natural sites and distance to downtown. In this model, we assume that location qualities are symmetric and suburbs have better natural qualities than the city center and rural areas. Apart from location quality, each region also has an endogenous neighborhood quality. Neighborhood quality captures endogenous factors that affect quality of life in that region, such as public facilities, public school quality and crime rate. It is endogenous because it depends on the residents living in that region, which are endogenous.

0.35	0.56	0.65	0.56	0.35
0.56	0.85	1.00	0.85	0.56
0.65	1.00	0.65	1.00	0.65
0.56	0.85	1.00	0.85	0.56
0.35	0.56	0.65	0.56	0.35

Figure 3.3: The 5x5 landscape

There are five types of market participants: the real estate agent, the developer, buyers, homeowners, and the bank. We further distinguish buyers and homeowners as investors and non-investors. Investors buyers buy a property in hope of profiting from house price appreciation. Regular buyers, on the other hand, obtain utility from living in the house. Each period in the model represents a month in real time. In each period, speculative and non-speculative buyers are created to enter the market in search of a house. Meanwhile, existing homeowners in the city decide whether to list their houses for sale. The number of non-speculative buyers generated in each period equals 5% of the total number of non-speculative homeowners, which is also the probability that a non-speculative homeowners will decide to sell the house for exogenous reasons (job move, divorce etc.). Therefore in this model we have a balanced city where the number of (non-speculative) buyers equals the number of (non-speculative) sellers. While the numbers of non-speculative buyers and seller are given, the numbers of speculative buyers and sellers are treatment factors thus will vary in the model. Figure 3.4 illustrated the model logic and the relationship between the five market participants.

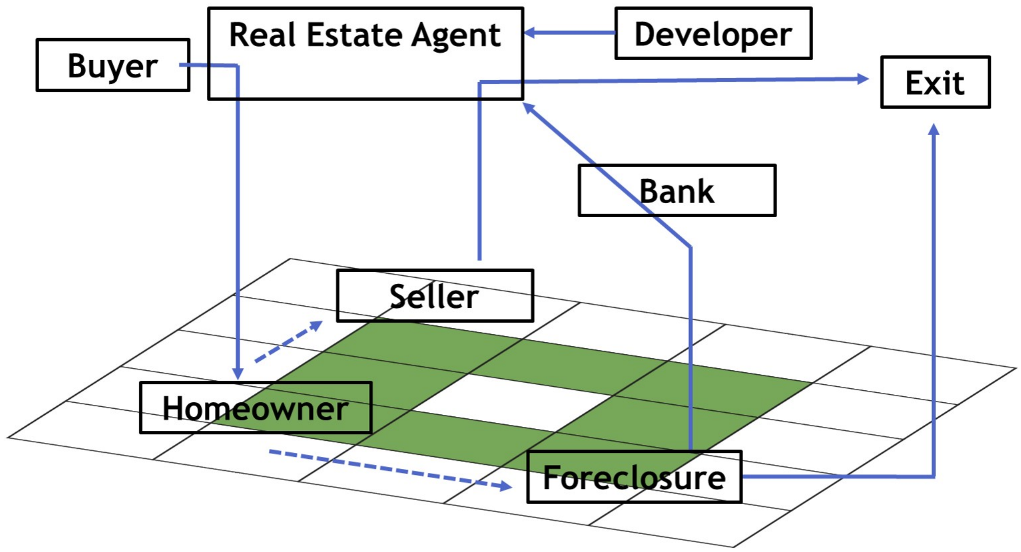

Figure 3.4: The housing model illustration

The model procedure can be described as sequences of steps:

Step1 At the beginning of period t (a period equals a month in real time),

The Real Estate Agent: Announces last period prices

The Bank: Announces mortgage rate

Step2 **The Developer**: Builds new houses in each region

A Homeowner: Decides whether to default, to list their property and submit an ask, or remain status quo and pay monthly payment

A Buyer: Searches for a house, decide whether to submit a bid, and to which region

Step3 **The Real Estate Agent**: Collects all the bids and asks, and settles the market price

Step4 **A Homeowner**: If her ask is accepted, sells her house, and exits the market

A Buyer: If her bid is accepted, buys a house, becomes a homeowner and a borrower from the bank

Step5 Update t to $t+1$. Go to Step 1.

A detailed structural framework of the housing market model is documented in Section B.3.

3.5 Treatment Factors and Experimental Design

In this section we will discuss the treatment factors in the model and the experimental design based on the treatment factors.

3.5.1 Treatment Factors: Leniency Index and Speculation Index

We identify two key treatment factors in the model that we think is central to the rise and collapse of housing prices: leniency index, denoted by L, and speculation index, denoted by S. Leniency index is defined as one minus the down payment rate. It measures the degree of leniency. We have shown in the analytical framework that low down payment requirement can lead to the rise and collapse of housing prices. Therefore we predicted that high leniency index will lead to large house price volatility.

Leniency index is closely related to the idea of leverage, defined as the ratio between debt and collateral. In fact, there is a one-to-one positive relationship between leniency index (L) and leverage,

$$\text{leverage} = \frac{1}{1-L} - 1 = \frac{1}{\text{down}} - 1 \tag{3.12}$$

When down payment requirement is 100% or leniency index is zero, leverage is also zero; when down payment requirement is 0% so leniency index is one, leverage is infinity. Because of this positive one-to-one relationship between leverage and leniency index, we can also say that we predict high leverage will lead to large house price volatility.

During the latest crisis, banks and financial institutions were found to be engaged in aggressive lending practice, which sometimes means zero down payment rate and zero mortgage payment in the first few years. The rationale behind banks' aggressive lending behavior has been the center of discussion and a focus has been made on the role of securitization. The reason is that securitization shifts the risks away from the initial lender thus reduces the lender's incentive to play safe. We agree that securitization is one of the reasons for aggressive lending, but we argue it is not the only reason. In this paper we will show that even without securitaztion, banks may still be willing to expand lending at the expense of increasing risks.

The second treatment factor, speculation index, is defined as the percentage of speculators in total homeowners. It measures the degree of speculation in the market. Speculators are the ones who buy or sell a property for profit rather than for consumption. They are trend-followers: they buy when the price is rising, and sell when the price is falling. As demonstrated in the analytic framework, we predict that more speculation will lead to large house price volatility. We should clarify that by making speculation a treatment factor, we control the number of potential speculators searching for profitable investment opportunities in the housing market. However, we do not directly control the number of speculators buying or selling in the market. Potential speculators will stay inactive if they fail to find any profitable opportunities.

The degree of speculation can differ by culture and the availability of alternative investment opportunities. For example in most East Asian cultures, it is a tradition for people to invest in properties with their spare money; while another culture has no such tradition. Also, in countries that lack a mature financial market, people are more prone to invest in the real estate market because there are fewer investment opportunities elsewhere. In short, the degree of speculation is affected by exogenous factors such as culture and the availability of alternative investment opportunities, thus we include it as a treatment factor in the model.

3.5.2 Experimental Design

Since we have two treatment factors, our experimental design is a two dimensional matrix. We also identify two key functional values in the model. The first one is house price volatility, defined as the standard deviation of housing prices in a region over its mean. The second functional value is the number of non-speculative homeowners at the end of simulation. It measures the degree of home-ownership. The higher the number of non-speculative homeowners, the higher the degree of home-ownership.

Leniency index is defined as one minus the minimum down payment rate by the bank. It is bounded between zero and one. Although it can take a value down to zero, meaning 100 % down payment rate, it is unusual that an up-front payment in full is required for buying a property. We believe a more realistic range for leniency index would be between 0.6 to 1, or between 40% to 0% down payment requirement. That will be the densely-sampled region for leniency index.

Speculation index is defined as the percentage of speculative buyers in total homeowners in each period. For example, in a city with 100,000 homeowners, a 0.05 speculation index means that in each month 5,000 speculative buyers are looking for investment opportunities. The higher the index, the more wide-spread speculation is in the housing market. Speculation index is bounded below at zero, but not bounded above. We put a very generous upper bound of 0.25 on speculation index, and we believe that a more realistic range for speculation index would be between 0 and 0.15. That will be the densely-sampled region for speculation index.

Husslage et al. (2011) show that a design for computer experiments should satisfy two criteria: space-filling and non-collapsing. The first criteria, space-filling, means that the experiment should be designed to obtain information from the entire parameter space, which in our case is a 1 by 0.25 square. The second criteria, non-collapsing, means there should be no irrelevant parameters which value does not affect the function value. Our experimental design satisfies both criteria. Figure 3.5 shows our experimental design.

For each cell, we will run the model 20 times. For each run, we start from an empty landscape. Buyers are generated each period to start house hunting. As some buyers become homeowners the landscape is gradually filled. Each run is consisted of 500 periods, excluding 200 burnout periods. We then calculate mean and standard deviation for the two function values: price volatility and number of non-speculative homeowners for each cell.

S \ L	0	0.2	0.4	0.6	0.7	0.75	0.8	0.825	0.85	0.875	0.9	0.925	0.95	0.975	1
0.000															
0.025															
0.050															
0.075															
0.100															
0.150															
0.200															
0.250															

Densely Sampled Region

Figure 3.5: Experimental design for two treatments: speculation (s) and leniency (l)

3.6 Results

In this section we are going to show simulation results from the housing market model. In subsequent Section 6.1 we present illustrated results for single runs. In Section 6.2 will present the report on full experimental design results, and finally in Section 6.3 we compare historical and simulated data of the U.S. housing market.

3.6.1 Illustrative Results for Single Runs

In Section 3.6.1, we are going to show the house price, foreclosure rate, mortgage rate, and bank's expected return from one simulation run, with a specific speculation index and leniency index. We show simulation results for 500 periods, which represents 500 months or around 42 years in real time. Figure 3.7 shows the house price under low (0.8) and high (0.9) leniency index. We fix the speculation index at 0.05. The results show that with low leniency index, house price fluctuates slightly; while with high leniency index, house price fluctuates wildly. The results confirm our prediction that high leniency index leads to large price volatility.

(a) leniency index=0.8 (b) leniency index=0.9

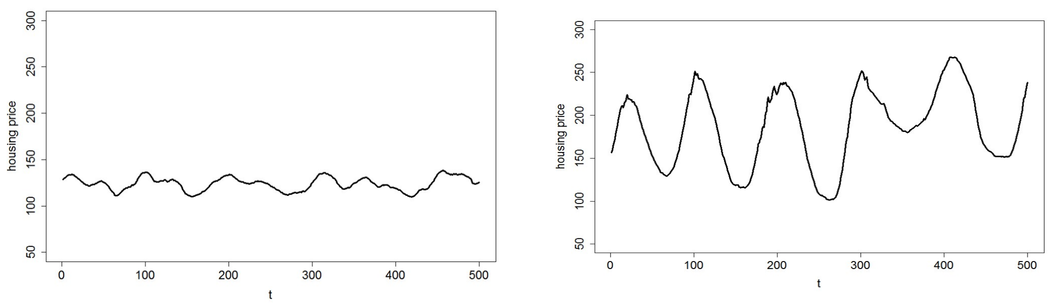

Figure 3.7: Housing price with low and high leniency index (S=0.05)

Figure 3.9 shows house price and the corresponding foreclosure rate, defined as the percentage of foreclosures in total homes. If house price dives so deep that the

(a) leniency index=0.8 (b) leniency index=0.9

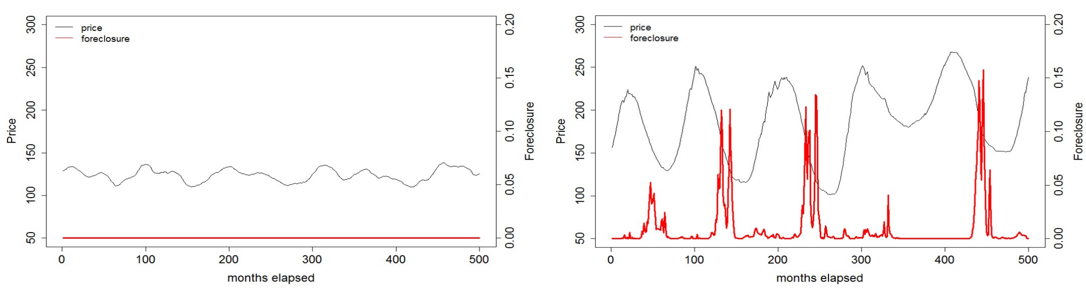

Figure 3.9: Foreclosure rate under low and high leniency index (S=0.05)

outstanding loan is more than the value of the house plus the default cost (so the house is not only under water, but deeply under water), the homeowner will choose to default. The results show that when leniency index is low, foreclosure is very rare; while when leniency index is high and when the price is falling quickly, foreclosure is wide-spread across the region (more than 10 %).

Figure 3.11 shows house price and mortgage rate under low and high leniency index. With low leniency index, mortgage rate is low and does not change much over time.

(a) leniency index=0.8 (b) leniency index=0.9

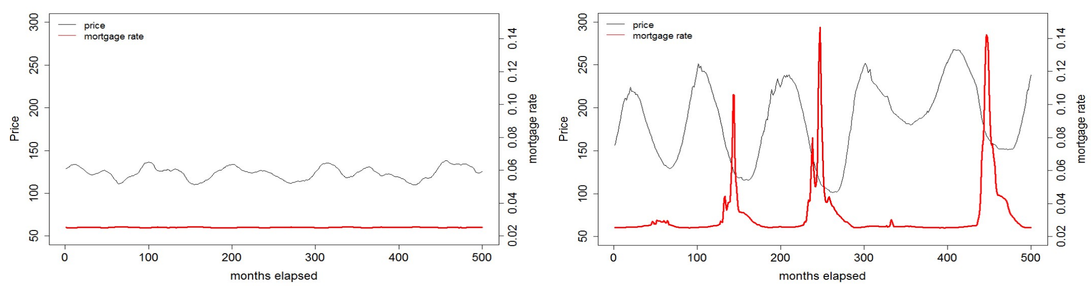

Figure 3.11: Mortgage rate under low and high leniency index (S=0.05)

With high leniency index, however, mortgage rate remains low when the house price is stabilized or in the rise, but jumps when house price starts to collapse. The reason is what we have shown in the analytical framework in Section 3.3. When a large down payment is required, a decrease in house price is unlikely to trigger the mortgage rate to increase in response. But when little or none down payment is required, a small decrease in house price will trigger the mortgage rate to increase to compensate for the depreciation in the value of collateral. An individual lender is regarded too small to affect the market price by raising mortgage rate. It rationally raises mortgage rate to protect itself from increased default risk. However, if every lender does so, demand for housing will be suppressed, which will in turn worsen the market condition and drag down house price.

Figure 3.13 shows house price and bank's expected profit under low and high leniency index. With low leniency index, bank's expected profit is very stable over time. With high leniency index, on the other hand, for most of the time bank's expected profit is higher than that under low leniency index. However, higher profits comes with a price. In

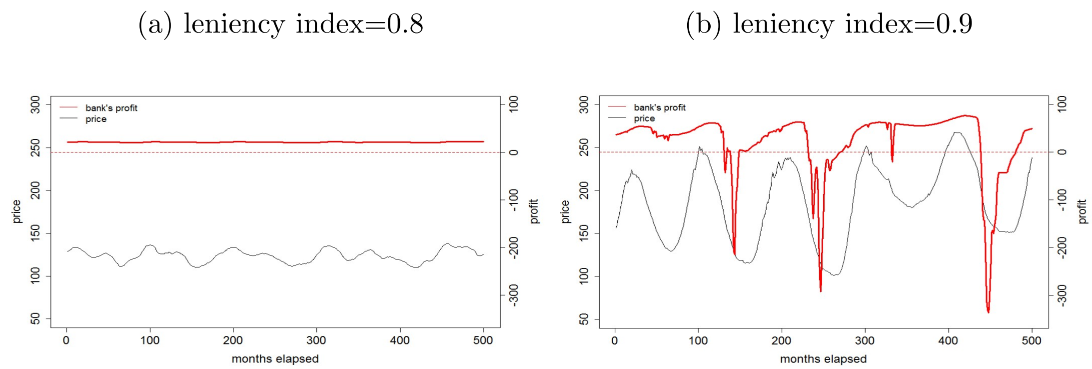

Figure 3.13: Housing price and bank's profit under low and high leniency index (S=0.05)

three occasions in the simulation over a period of over 42 years, the bank would suffer big losses. Since in the current model we do not explicitly model competitions among banks so there is only one representative bank, we assume that the bank is well-capitalized and will never go out of business.

If the bank is forward-looking long enough into the future, they might prefer to have a more prudent lending policy just to avoid big losses doomed in the future. However, the bank might not be all forward-looking: the average CEO tenure in the financial industry is only five years (Stuart (2004)). Compared with a five-year tenure, we see that in the simulation the first loss does not occur until after 10 years of high profits, and it only happens three times over a period of 42 years. At all other times, banks enjoy high profits which more than doubles that under a tight lending policy.

We now show that the simulation results for the second treatment factor, speculation index. Figure 3.15 shows house price without speculation (speculation index=0) and with speculation (speculation index=0.05). We fix the leniency index at 0.9. The results confirm our conjecture that speculation fueled with lenient financing will lead to large price volatilities in the real estate market.

Figure 3.17 shows the corresponding foreclosure rate without and with speculation. We found that foreclosure rate is high with speculation and is close to zero without speculation. However it does not mean that only speculative homeowners will default.

(a) speculation index=0 (b) speculation index=0.05

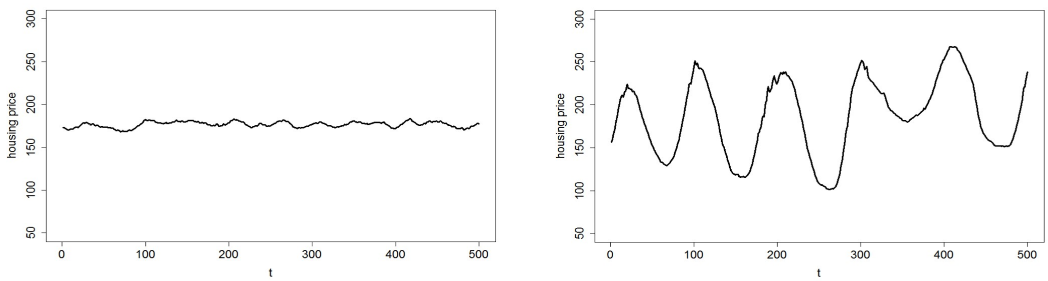

Figure 3.15: Housing price without and with speculation (L=0.9)

(a) speculation index=0 (b) speculation index=0.05

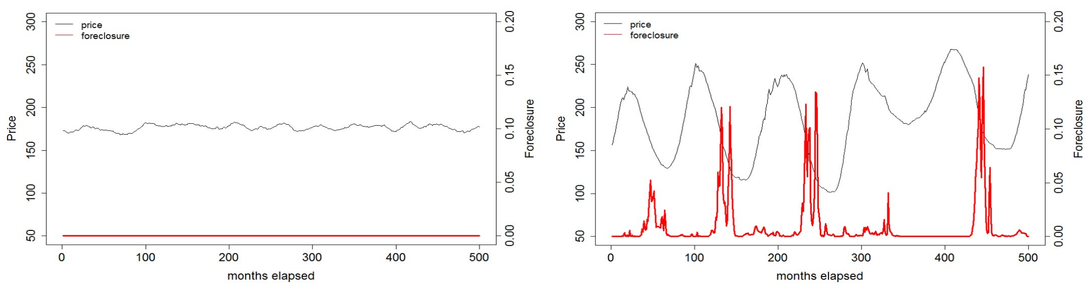

Figure 3.17: Foreclosure rate without and with speculation (L=0.9)

Both speculative and regular homeowners is more likely to default if house price becomes more volatile. The high foreclosure rate is a direct result of increased volatilities in house price, which is caused by the existence of speculators.

3.6.2 Report on Full Experimental Design Results

In Section 3.6.2 we shows results of the computational experiment, in which we systematically change the two treatment factors, leniency index and speculation index, and simulate the two functional values-house price volatility and number of non-speculative homeowners. For each combination of treatment factors, we run 20 simulations. We also run 100 simulations for five randomly selected cells and found that our results are robust to sample size: the difference between the mean values from 20 and 100 simulations is less than 5%. To save space, we only display the mean functional values in the densely-sampled region. The mean and standard deviation in the entire parameter space for both functional values can be found in Section B.5.

In Figure 3.18, we present a heat map of mean house price volatility in the more densely-sampled region. High-valued cells are darker and low-valued cells in lighter. The mean and standard deviation of house price volatility for the entire parameter space can be found in Figure B.3 in Section B.5.

S \ L	0.6	0.7	0.75	0.82	0.825	0.85	0.875	0.9	0.925	0.95	0.975	1
0.000	0.023	0.025	0.022	0.021	0.026	0.017	0.017	0.016	0.015	0.017	0.018	0.018
0.025	0.032	0.031	0.030	0.026	0.029	0.033	0.062	0.111	0.181	0.204	0.244	0.235
0.050	0.042	0.039	0.038	0.040	0.046	0.060	0.123	0.223	0.219	0.228	0.241	0.222
0.075	0.042	0.046	0.051	0.055	0.067	0.085	0.162	0.265	0.247	0.228	0.189	0.213
0.100	0.046	0.054	0.059	0.066	0.081	0.097	0.178	0.273	0.258	0.246	0.224	0.214
0.150	0.054	0.068	0.068	0.085	0.099	0.115	0.194	0.248	0.241	0.230	0.219	0.215

Figure 3.18: Housing price volatility, densely sampled region

Cells are in lighter shades in the northwestern corner of the heat map, where leverage is low and speculation is low, meaning low housing prices volatility and a stabilized market. Cells are in darker shades in the southeast corner of the heat map, where leverage is high and speculation is high, representing large house price volatility. We show that house price volatility is increasing with leverage and speculation. Lenient financing which allows high leverage will cause large cycles in the house price given that speculation exists in the market.

In Figure 3.19, we present a heat map of mean number of non-speculative homeowners in the more densely-sampled region. Again, high-valued cell is in darker shades and low-valued cell in lighter shades. The mean and standard deviation of the number of non-speculative homeowners in the entire parameter space can be found in Figure B.4 in Section B.5.

Cells are in darker shades in the northeast corner of the heat map, where speculation

S \ L	0.6	0.7	0.75	0.8	0.825	0.85	0.875	0.9	0.925	0.95	0.975	1
0.000	3482	6667	8098	9476	9966	10150	10474	10850	10841	10562	10346	10151
0.025	5146	6707	7635	8735	9346	9953	10491	11534	11886	11300	11322	8636
0.050	4881	6479	7430	8520	9226	9898	10446	10526	10522	10970	9577	9451
0.075	4695	6161	7168	8444	8984	9440	9390	9322	8659	9523	8884	9219
0.100	4398	5939	6919	7968	8537	8964	8100	6319	7194	7436	7022	7705
0.150	3875	5531	6409	7177	8205	8184	5707	4659	4214	4726	4585	4817

Figure 3.19: Number of non-speculative homeowners, densely sampled region

index is low and leniency index is high: home-ownership rate is highest when we have a lenient lending criteria but few speculative buyers. Moving from left to right, as leniency index increases, the number of non-speculative homeowners continue to increase until it is 0.85. After that, the gain in home-ownership is only marginal. Moving down, as the number of speculators increases, the number of non-speculative homeowners decreases. Speculators has driven up the price and crowded out some of the non-speculative homeowners.

Now we pick a row in the heat map of price volatility and fix speculation index at 0.05. In Figure 3.21 we show the relationship between leniency index and house price volatility, conditioning on speculation index equals 0.05. The result shows that there

(a) complete region (b) densely sampled region

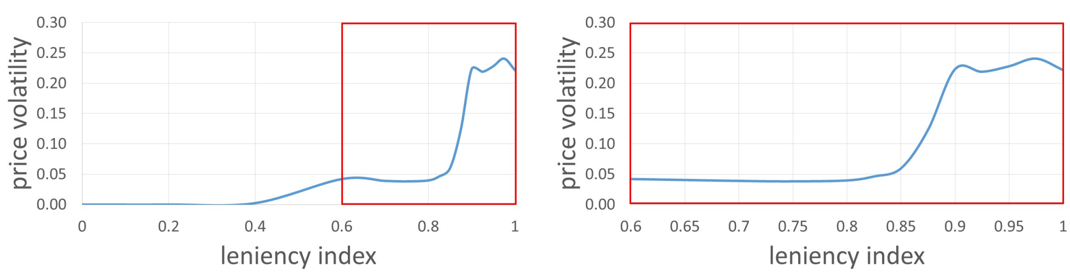

Figure 3.21: Leniency index and housing price volatility (S=0.05)

exists a positive relationship between leniency index and house price volatility. It also shows that the relationship is not necessarily linear. Volatility stays at a low level as long as leniency index is between 0 and 0.85, that is, down payment rate is higher than

15%; when down payment rate is less than 15%, price volatility increases quickly from less than 5% to more than 15% as down payment rate decreases from 15% to 10%. When down payment rate is further lowered to less than 10%, price volatility stays at a high level of above 15%, but does not further increase. This implies that we do not need to make down payment rate prohibitively high to prevent large house price volatilities. If we set the down payment rate at the right level, which is at about 15% in this case, we can both avoid large price volatility and have high home-ownership rate.

We then pick a column in the heat map at leniency index equals 0.9. In Figure 3.23 we show the relationship between speculation and house price volatility, conditioning on leniency index equals 0.9. Like we have predicted, speculation will lead to larger price

(a) complete region (b) densely sampled region

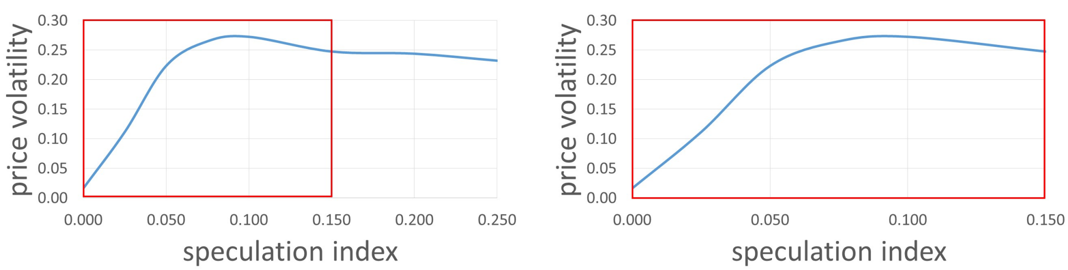

Figure 3.23: Speculation index and housing price volatility (L=0.9)

volatility. Speculation is a treatment factor in this model for the reasons we describe in Section 3.5: it is affected by things like culture or housing market history, which are exogenous in our model. However, Culture and housing market history is not all exogenous: people learn and adapt; culture can be formed. Non-investors can be turned into investors if they constantly see their friends making profit in the housing market. Slowly, the change in behavior is ingrained in the collective memory and becomes part of the culture. Since we now know that speculation will cause larger price volatility, it would be interesting to see what will happen if we endogenize it. However this is beyond the scope of this paper, so we will leave it for future study.

3.6.3 Comparison of Historical and Simulated Data

In this subsection, we demonstrate that our simulation results can resemble some characteristics of house price history. We make a qualitative comparison between the quarterly U.S. home price index (shi ()) between 1970 and 2012, and simulation results over the same time span. In the simulation, before the year 2000, down payment rate is set at a moderate 15% and market interest rate is set at 4%. At the beginning of January, 2000, we put introduce lenient financing into the system and decrease down payment from 15% to 4 %. We also decrease the market rate from 4% to 1%, to represent the decreased interest rate, due to the abundance of hot money in the U.S. at that time. Other than that one time shock, we let the model run without interference. Figure 3.25 shows a comparison between the history of U.S. home price index and simulated price index. We use the earliest price in 1970 as baseline.

(a) historical data**?**)) (b) simulated data

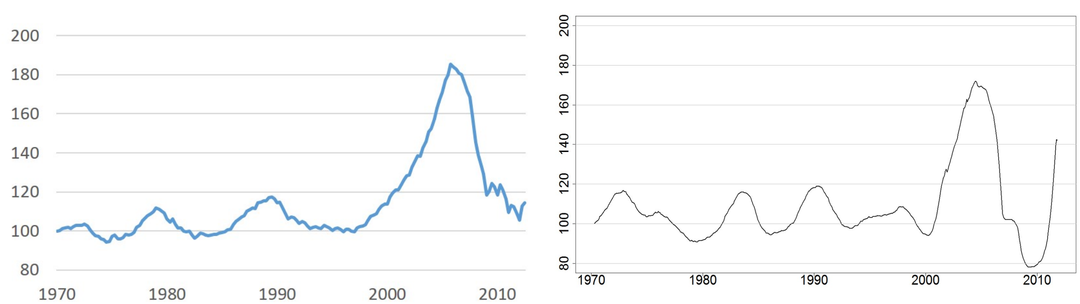

Figure 3.25: Comparison of historical and simulated data, 1970=100

The two graphs have shared some common patterns. When we began to deregulate the financial market and adopt more lenient lending criteria around the year 2000, house price starts to rise. Housing price keeps rising until all of sudden, without any obvious external reasons, it collapsed. Our simulation model has captured that important aspect of the reality. However, we are cautious not to overstate the implications of the two graphs' similarities. To actually validate the model using empirical data on the housing market, we need to carefully calibrate model parameters such as region characteristics, income, and market interest rate. All these parameter values will affect simulation results. Since the purpose of this paper is to provide a computational model of the housing market and use it to demonstrate the endogenous rise and collapse of housing prices, we will leave model validation for future research.

3.7 Concluding Remarks on Chapter 3

The recent crisis in the U.S. housing market has had devastating consequences. About four and half million American families have lost their homes to foreclosures or were on the edge of going foreclosure. America has since gone into the most severe recession since the big depression in the 1930s. The housing market receives less attention from researchers than the stock and foreign exchange market despite its heavy weight and influence in the economy. Even today we are still lacking in understanding of this important market. The housing market is very unique: it is less liquid, highly leveraged, spatial are both investment and consumption good. To this day, there are few housing market models that take into account all of these complications.

This study develops an agent-based spatial model of the housing market that incorporates all of the issues above. We attempt to answer the following questions: Will housing prices rise and collapse endogenously in a housing market? If so, under what conditions? Our results show that housing prices will rise and collapse endogenously without any external shock, although a shock may magnify the process. We find that the two necessary conditions for the endogenous rise and collapse of housing prices are low down payment requirement or leniency, and speculation. When banks are engaged in lenient financing and there is speculation in the market, the rise and collapse of housing prices are inevitable. Our simulation results also show that when housing prices start to fall, foreclosure is wide-spread and mortgage rate increases, all of which are similar to the observations in the recent U.S. housing market crisis.

In addition, we find that a bank may opt for lenient lending practices even without the securitization of mortgage loans. Although the securitization of mortgage loans will lead to more aggressive lending from the banks, we believe this is only part of the underlying reasons. Our simulation results show that, even without securitization of mortgage loans, banks may be willing to adopt a lenient lending policy, knowing they may face losses in the future, for the benefit of higher-than-normal profits for extended period of time. It is especially true when the average CEO tenures in the financial sector is much shorter than the average length of boom period, and a CEO's payoff depends only on short-run profits. Since the securitization of mortgage loans is not the only reason for banks' aggressive lending behavior, simply restricting securitization alone is no guarantee that a housing market crisis will not come back in the future. A more fundamental reform of financial institutions is needed.

We also find that policies can be made so that it encourages home-ownership and achieves housing market stability at the same time. One of the reasons for government sponsored entities (GSEs) like Frannie Mae and Freddie Mac to lower lending criteria is to extend lending to low income household and encourage home-ownership. However, our simulation results show that at 15% down payment rate, home-ownership rate reaches its high: further lowering down payment requirement will no longer increase home-ownership rate. Rather, it will cause the housing prices to rise so high that it hinders rather than helps home-ownership among low income households. In other words, lowering lending criteria can make housing less affordable, not more. We find that it is possible to achieve both affordable housing and market stability by setting the down payment requirement at an optimal level, which is 15% in our simulation.

Finally, we show that our agent-based model is able to generate price patterns similar to those in the U.S. housing market. We believe our agent-based model has incorporate the key elements in the housing market and is able to capture the dynamics that give rise to large house price volatility. However, we do not claim that our model is able to match real world data on a value base. Neither is data replication the purpose of this study. The empirical validation of the model is beyond the scope of this study, and we will leave it for future research.

To sum up, we have shown that housing prices can rise and collapse by itself without any external economic shock. The two necessary conditions for the endogenous rise and collapse of housing prices are lenient financing and speculation. We find that banks have incentives to engage in aggressive lending even without the securitization of mortgage loans. We also find that we can achieve both affordable housing and housing market stability by setting the down payment requirement at an optimal level. Moreover, we show that the agent-based model is able to generate price patterns similar to those in the U.S. housing market. Future work includes the empirical validation of the model and the endogenization of speculative behavior.

CHAPTER 4. CASE STUDY OF THE WASHINGTON D.C. HOUSING MARKET

4.1 Introduction to Chapter 4

Since the outbreak of the sub-prime crises in late 2006, the causes and implications of the crises in the U.S. housing market has been the center of public discussions and a heated subject of academic research. Among the discussions and analysis revolving around the housing market crises, most look at national housing market and home price indices, some look at inter-city differences, but very few studies has investigated in intra-city variations. However, even within the same city and housing market, home prices in different regions could exhibit price patterns that are very different than national or regional home price indices. We believe it is equally important to look at the housing market on a more segregated, micro level, because it is in the micro-environment decision makers make a decision. A decision maker does not make a decision on property purchase based on national or regional home price index. Rather she bases her decision on the price of a particular home. This paper thus takes one step further: we segregate regional home price indices, establish local home price indices within a city, and model individual decision maker's behavior based on segregated local information.

However, it is challenging to simultaneously model multiple related intra-city housing market. The agent-based housing market model developed in Chapter 3 is among the first to account for intra-city heterogeneity in the regional housing market and correlations between neighborhoods in a city. Chapter 3 develops a theoretical, spatial agent-based model where individual buyers make discrete choices among the 25 regions in the hypothetical city. This study applies the theoretical model developed in Chapter 3 and adapt it to the Washington D.C. housing market. It attempts to build on an existing theoretical model and adapt it into a specific context and test the model's credibility using empirical evidence within the context.

This study is also an experiment of knowledge accumulation in agent-based modeling, which means learning and experimenting with existing models rather than develop a new model from scratch. Being able to developing upon existing models, rather than having to build new models for a specific problems each time could be an important advantage of agent-based models, as they are known for being flexible and adaptable. An existing model can also be integrated in a larger model and become part of a more complex system, which is another form of knowledge accumulation with agent-based models. We will go into more details in Section 4.2.

4.2 From Concept to Context: Development and Evaluation of Agent-Based Models

Agent-based models can range from being conceptual and theoretical to highly contextual and applied. A pure conceptual model does not have an immediate real-world context. It's objective is not to replicate a real-world system at a value level, but experiment with a concept or prove a theory with minimum assumptions. As a result a conceptual model is usually parsimonious and simplistic in its assumptions. A guideline to develop this type of proof-of-concept model is the "KISS" principle: keep it simple, stupid (KISS). Many early agent-based models in social sciences fall in this category. Classic examples include the Sugarscape model by Epstein (1996) and the zero intelligence trader model by Gode and Sunder (1993). Such abstract, non-contextual models are important for us to understand a complex system. By isolating and focusing on a few most prominent elements in the system, researchers are able to observe and analyze directly the role each part plays in the system and how it interacts with each other.

Opposite to a conceptual model, a contextual model tries to replicate a particular target system in the real world or match empirical data on a value level. To do so, the model has to be inclusive the critical components in the real-world system and make realistic assumptions. Since One of the model's objectives is to simulate as accurately as possible the running of the target system and provide practical advice to its operation, the models are usually large-scaled and highly-detailed, and have a large number of parameters and many modules (Carley (1996)). Examples include the "AMES" test bed for the power market (Li and Tesfatsion (2009)) and the agent-based simulation of crime systems at the level of individual houses and offenders (Malleson (2012)). A model specifically developed for a particular social system can be very useful, too. For example, the simulation of crime systems can be developed to make real-time predictions for occurrences of burglary in all neighborhoods in Leeds, UK. Such predictions cannot be made with a general, proof-of-concept model.

If we draw a line with the most conceptual and theoretical models at one end and detailed and contextual models at the other, most agent-based models lie in between the two extremes and scatter along the line. Borrowing the idea of Technology Readiness Level (TRL) (Heslop et al. (2001)) from engineering, we can sort agent-based models in social science by the level of abstraction. In this paper we show an example of developing multiple related agent-based models and move along the line of abstraction. Based on a simplistic and general model, we add more context to the model and calibrate it using empirical information. The opposite can be also be done, meaning researchers start from a model built for a particular system and generalize it later. In agent-based modeling, this type of model accumulation is rarely observed, which is almost surprising because one of the advantage of agent-based models is they are more flexible and adaptable than equation-based models. This paper shows how we can build on existing framework and adapt it to specific setting or context as an effort to explore the full potential of a computational model.

Another issue regarding agent-based models is model calibration and validation, the latter means using empirical data to demonstrate the model's level of credibility. Although agent-based computational models are now widely used in many fields, empirical validation of agent-based models remains an unsettled area (LeBaron (2000)). Most models inspired by real-world observations have not gone beyond a "proof of concept" (Janssen and Ostrom (2006)). Since then researchers have devised different techniques for calibration and validation of different types of agent-based computational models.

For conceptual models, since the aim is to explain observed phenomenon or proof/disproof a theory, pattern-match is often used as a way to validate such models. To establish an agent-based model's credibility, researchers test whether the model can generate some stylized facts or observed phenomenon in the real world from interactions between individual interactions. Examples include right-skewed wealth distribution (Epstein (1996)), right-skewed firm size and growth rate distribution (Axtell (1999)) and fat tails and clustered volatility in investment returns (Thurner et al. (2012)). In most cases, aggregated distributional data is used to validate a model.

In addition, since conceptual models tend to be parsimonious, it is sometime possible to estimate a few parameters in the model, using indirect inference estimation or least-squared type of estimation. Examples are Boswijk et al. (2007), Bianchi et al. (2007) and Gilli and Winker (2003). The estimated optimal parameters are then checked against feasible ranges or known relationships between the parameters. To use indirect inference to estimate parameters in agent-based models, researchers first have to identify some key characteristics or moments in order to construct an objective function. In some cases it is straightforward which characteristics should be targeted, but not so in other cases, depending on the research questions and objective of the study. Researchers also need to specify the few parameters to be estimated, while trying to calibrate all other parameters.

At the other end, contextual models has to go through rigorous model calibration that is based on detailed information about the parameters and the running of the target real-world system, so as to replicate it. Often times on-site investigation of the target system needs to be conducted to obtain such information. Due to the system's large scale and complex nature, collaboration among researchers are necessary to carry out the task of developing and validating the model. Moreover, most context model has a large number of parameters, it is hard to estimate the parameters using techniques mentioned before. Rather, parameters in an contextual model are calibrated using data, expert knowledge on business practice, or treated as control variables.

For most agent-based models in between, empirical validation of agent-based models depends on the data available. Bianchi et al. (2005) has empirically validated a macroeconomic model using a comprehensive data of 6422 Italian firms from 1996 to 2001. For each firm and year, they have acquired reliable data on equities, long term debts and loans, short term debts, total capital, gearing ratio, solvency ratio, debt ratio, number of employees, cost of employees and revenues. Geanakoplos et al. (2012), in an effort to calibrate and validate an agent-based housing market model, have collected data on every housing unit in the greater Washington D.C. area, and on every individual in the area by race, income, wealth, age, marital status, household position, and so on. In general, the data requirement for empirical validation for agent-based economic model is high: data at individual firm or household level is often required for model calibration and validation.

This paper adapts a conceptual agent-based model to a specific context, and uses empirical evidence to calibrate model parameters and test the model's credibility. This paper present an example of knowledge accumulation through developing applications based on an existing model to take full advantage of the flexibility and adaptability of agent-based models. The paper also shows that when developing agent-based models along the line of abstraction, different validation techniques and data are required for model calibration and validation. In particular, we will use the agent-based model of housing market developed in Chapter 3 and adapt it to the city of Washington D.C. We will use property transaction data and demographic data in Washington D.C. neighborhoods to calibrate and validate the adapted housing model.

4.3 The Washington D.C. Housing Market

4.3.1 Overview of the Washington D.C. Housing Market

Washington D.C. has been chosen as the object of our case study mainly due to the plenty readily available data of the area and its well-defined neighborhood clusters. In 2012 Washington D.C. has a population of about 632 thousand. Like many large cities in the United States, it has also experienced decades of decline in population and economic influence; Its population dropped from 802,178 in 1950 to 572,000. However since the second half of the 1990s the city has had a dramatic renaissance driven by the fast growth of the private sectors, especially in the high-value, high-tech services sector (Tatian (2008)). Because the fast growth of the city since the late 1990 and a national trend of housing price appreciation starting 2000, property value in Washington D.C. has enjoyed a double-digit annual appreciation that lasts through the first half of the 2000s. Then in late 2005, housing price in Washington D.C. collapsed from its peak when a nationwide housing market crises took place. Figure 4.1 shows the Washington D.C. housing price index from late 80s to 2013.

4.3.2 Intra-City Spatial Variation

The office of planning in Washington D.C. divides the city into 39 neighboring clusters. Each cluster is made up of three to five neighborhoods. Neighborhood clusters are being used by the D.C. government for budgeting, planning, service delivery, and

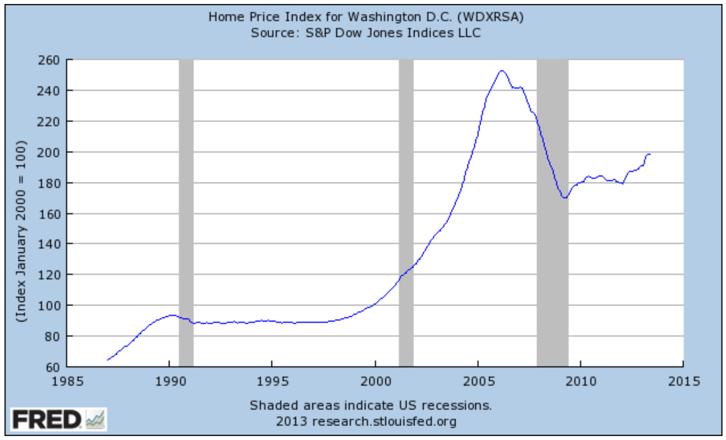

Figure 4.1: The Washington D.C. housing price index (Source: S&P Dow Jones Indices LLC)

analysis purposes. Properties in each cluster are generally considered to belong to the same local real estate market. Clusters within Washington D.C. have different landscape and demographic characteristics. Property values also vary largely by clusters. Figure 4.3 shows the 39 neighborhood clusters in Washington D.C., as well as the median sales price in 1999 and average family income in 2000 in each cluster.

(a) Median Property Price

(b) Average Family Income

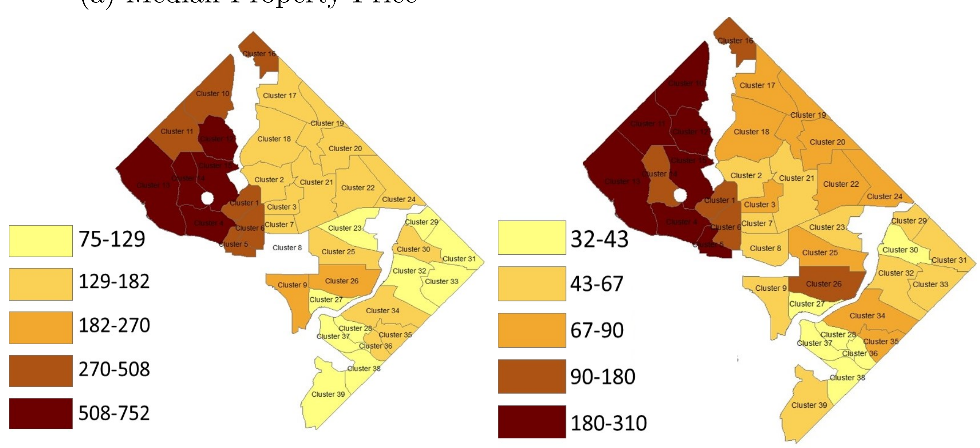

Figure 4.3: Property price and average family income across neighborhood clusters in Washington D.C. in 1999 and 2000 respectively

As we can see, median property price and average family income are both higher in the western part of the city. Moreover, the inequality across neighborhood clusters in D.C. is substantial; In 2000, the average family income in cluster 15 is 311,035 while it is only 31,957 in cluster 37; As for property price, in 1999, the median sales price in cluster 13 is 752,000 while it is only 75,000 in cluster 37. The highest average family income is ten times the lowest one across clusters.

From the graphs we can see that housing prices highly correlate with average family income across clusters, which is not surprising. This selection process is self-enforcing: on the one hand, high property prices in some neighborhoods prevent median to low income families from moving into those neighborhoods; on the other hand, neighborhoods dominated with high income families are most likely able to provide better local amenities and services. Properties are also more likely to be well-maintained. All these will make the neighborhood more attractive and keep the price high.

When large changes take place in a city's housing market, a city's demographic landscape might also change. Even in the midst of a universal rise of property prices, the extent and order of the house price appreciation vary across neighborhoods. For example when property prices goes up in some neighborhoods, neighborhoods once dominated by middle class families are no longer affordable to them. Instead, middle income families move into neighborhoods once dominated by low income family, changing the demographics and characters of the neighborhoods. Those revived neighborhoods then attract more middle income families, and property prices start to rise there. The rise in the housing price spreads across neighborhoods like a ripple. This study will look at the impact of the rise and collapse of housing price on inequalities across neighborhood clusters and changes in neighborhood demographics in Washington D.C.

4.3.3 Local Housing Price Indices

Houses are not standard products: they differ largely in various attributes like the number of bedrooms and bathrooms. The huge heterogeneity in houses makes it difficult to construct house price index and compare housing prices across time. There are two approaches to tackle the problem: the hedonic price approach and the repeated sales approach. The hedonic price approach tries to account for all attributes affecting the house value for which data is available, such as the number of bedrooms and bathrooms, lot size and age. However, there is always the problem of unobservable variables when using the hedonic price approach: not all attributes affecting the value of a house can be accounted for. In the case of house price, one possible and very important unobservable variable determining the property value is the location of the house.

The second approach, the repeated sales approach uses information on properties being traded more than once throughout the time. The change in sales price of the same property is used to elicit house price index overtime. The advantage of the repeated sales approach is that it can control for unobservable variables such as location, as long as they are time-invariant. The disadvantage is this approach has to throw away a large amount of information because it only uses repeated sales data. In the housing literature, both approaches are used, depending on data availability and research objective. The Case-Shiller Home Price Indices (Case and Shiller (1989)), the leading measures of U.S. residential real estate prices, use the repeated sales approach and is now standard practice in the housing research literature.

In this study I will use the repeated sales approach to estimate cluster-level, quarterly house price indices (HPIs) for the 39 neighborhood clusters in Washington D.C.. The data I use to estimate the local HPIs is the real estate transaction data in Washington D.C. from 2004 to 2012, which is public data on the Washington D.C. government website (http://dc.gov/DC/). For each real estate transaction, the data includes the address of the property, transaction price and date. It does not, however, have information on the property's attributes, such as number of bedrooms, bathrooms and lot size. The absence of data on property attributes prevents me from estimating a hedonic pricing model. Therefore I will use the repeated sales approach to estimate quarterly HPIs for the 39 neighborhood clusters in Washington D.C..

The original dataset contains 54142 transactions of 31816 unique residential properties (condos and houses) in Washington D.C. between January 2004 and January 2013. Of the 31816 properties, 1350 had been transacted at least twice, which leaves us with 11029 transactions or data points from which we will estimate HPIs for the 39 neighborhood clusters in Washington D.C..

The repeated sales approach models individual housing prices as the sum of a market index, a dispersion around the market index arising from a log normal diffusion process and a white noise (Calhoun (1996)).

$$ln(p_t^i) = \beta_t + H_t^i + N_t^i \tag{4.1}$$

where β_t is the market price index at time t, H_t^i is a Gaussian random walk of property i at time t, and N_t^i is the white noise of property i at time t. The total percentage change in price for property i transacted in time period t and s $(t > s)$ therefore is,

$$\begin{aligned}
\Delta V_{ts}^i &= ln(p_t^i) - ln(p_s^i) \\
&= \beta_t - \beta_s + H_t^i - H_s^i + N_t^i - N_s^i \\
&= \sum_{\tau=0}^{T} \beta_\tau D_\tau^i + \varepsilon_{ts}^i \qquad (4.2)
\end{aligned}$$

where $D_\tau^i = 1$ if $\tau = t$, $D_\tau^i = -1$ if $\tau = s$ and $D_\tau^i = 0$ otherwise. $\varepsilon_{ts}^i = H_t^i - H^i + N_t^i - N_s^i$. Since the dispersion around the price index, $H_t^i - H_s^i$, is a Gaussian random walk process, it is assumed that the diffusion between time period t and s is an increasing function of difference in transaction time, $t - s$.

$$E[{\varepsilon_{ts}^i}^2] = c_1(t-s) + c_2(t-s)^2$$

We then use a two-stage, general least squares procedure to estimate house price index, $\beta_1, ..., \beta_3 5$ along with diffusion parameters c_1 and c_2. Out of the 39 neighborhood clusters in Washington D.C., HPIs for 20 regions are identifiable. HPIs for the rest of the regions are unidentifiable due to the small sample size. The estimated HPIs for the 20 clusters in Washington D.C. are shown as the solid curves in Figure 4.6.

4.4 Model Calibration

4.4.1 Location Quality Calibration

As previously stated, this paper adapts the agent-based housing market model developed in Chapter 3. In Chapter 3, however, we use a landscape that is a symmetric, five by five square with a downtown in the middle, a circle of suburban areas and a circle of rural areas. In this study, we adapt the original landscape to that of Washington D.C..

Apart from the layout of the landscape, we also need to estimate the location quality in each region in the landscape, since it is not observable. A region's location quality index is a composite index including all external factors that affect the quality of life in that region. It can include natural factors, such as a region's closeness to a natural site or vulnerability to flooding; it can also include historic factors, such as a region's historic and cultural characters. In short, anything that affect the attractiveness of a region and is not determined by the region's residents or housing market conditions is captured by the location quality index.

In the original landscape in Chapter 3, location qualities are assumed to be symmetric, with the suburban area as more desired neighborhoods than downtown and rural areas. When we modify the basic landscape to that of Washington D.C., we need to determine the location quality index for each neighborhood cluster in Washington D.C. As previously stated, location quality index includes all factors external to the housing market that affects the quality of living in that region. As a result, to construct such indices requires a large amount of information. In this paper, we use demographic data and real estate transaction data to elicit the perceived location quality for each region.

The idea is as follows. Home buyers incorporate their perceived location qualities, among other factors like house price and neighborhood quality, when they choose across regions in Washington D.C. The hedonic house price in a region reflects people valuation of living in that region. We regress local housing prices on demographic variables such as average family income, crime rate, unemployment rate and percentage of poor persons from year 2000 and 2005. We use demographic data as an control for endogenous neighborhood qualities which is determined by the residents living in a region; and we use the residuals from the regression as an indicator for the location quality perceived by home buyers. The estimation results for location quality can be found in Table C.1. We also display the calibrated location quality in the 39 regions in Washington D.C. in Figure 4.4.

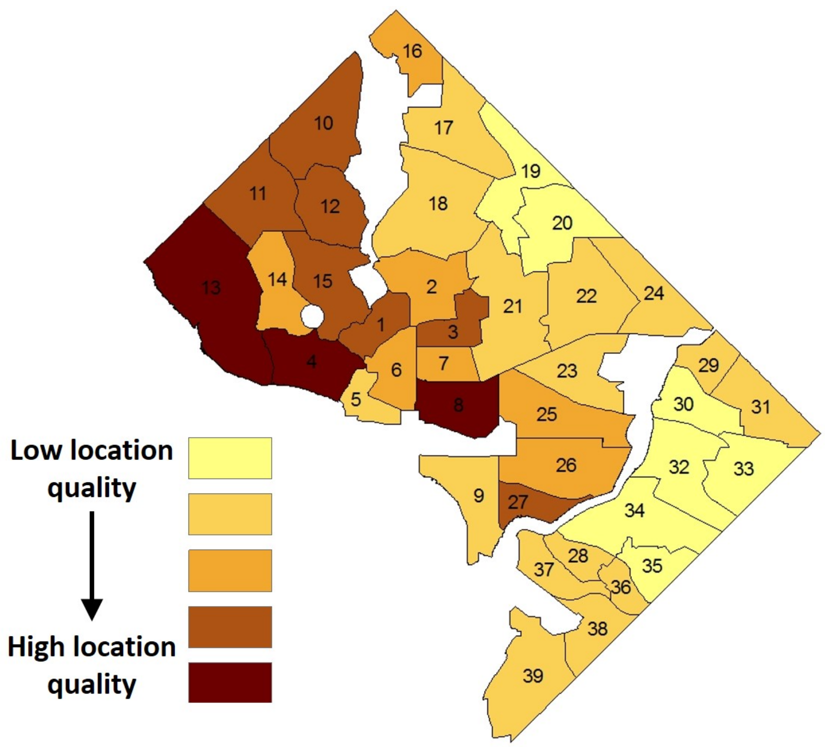

Figure 4.4: Location quality of 39 neighborhood clusters in Washington D.C.

4.4.2 Calibration for Other Variables

Apart from location quality, we will also need to calibrate income distribution, population growth and in some cases mortgage rate. We use annual income data in Washington D.C. to calibrate the income attribute of new buyers. We assume log-normal income distribution and infer the standard deviation of the income distribution from data on percentage of poor persons and poverty line. We then consecutively draw from the calibrated income distribution and assign the income attribute for new buyers.

Second, we need to calibrate population growth in Washington D.C. Unlike the balanced city in Chapter 3 where the number of incoming new buyers equals the that of sellers, Washington D.C. is expanding for the first time after decades of declining. It is important that we capture the city's revival and growth in the midst of a universal property price appreciation across the country. The total change of population in Washington D.C. is -5.7% between 1990 and 2000; it is 5.2% between 2000 and 2010. We will calibrate the number of new buyers and sellers to reflect population change in Washington D.C. In the first decade of simulation, the number of incoming new buyers is 0.57% less than the number sellers; while in the second decade, it is 0.52% more than the number of sellers.

As for mortgage rate, in Chapter 3 mortgage rate is endogenously determined by risk-free return, history of foreclosure, history of house price and minimum down payment. In this study, since we have data on mortgage rate in Washington D.C. housing market, we will first use empirical data on mortgage rate as model input. Later, we will endogenize mortgage rate and compare the simulated and historical mortgage rate.

The fourth variable that we need to calibrate is the construction cost of the developer. Each of the 39 neighborhood clusters in Washington D.C. is different in size of area, population, number of houses, and regulatory restrictions. Therefore the construction cost to build a house will also be different across neighborhood clusters. In Chapter 3, construction cost has three parts: fixed cost, variable cost and congestion cost.

$$TC_t^g = Ah_t + \frac{1}{2}Bh_t^2 + C^g H_t^g \tag{4.3}$$

Where TC_t^g is total construction cost at time t in cluster g, h_t is houses built in period t and H_t is the total number of houses in region g at period t. A and B are universal fixed and variable cost parameters, including costs of materials and labor; while C_g is cluster-specific congestion parameter, which is used to capture the differences in construction cost across neighborhoods resulting from land scarcity or regulation. It is the cluster-specific congestion parameter, C_g. that we need to calibrate for the Washington D.C. housing market.

We use the the historical housing prices and the number of houses in each cluster to calibrate the cluster-specific congestion parameter, C_g. Between 1990 and 2000, the house price in Washington D.C. is stabilized (see Figure 4.1). Moreover, not many new houses had been built over that period of time. Population in each cluster had been static and slightly declining. We thus assume that average prices between 1990 and 2000 equals the congestion cost, so developers have not incentive to build new houses and the housing prices are static.

$$\bar{p}^g = C^g H_t^g \Rightarrow C^g = \bar{p}^g / H_t^g \tag{4.4}$$

A neighborhood with a larger number of houses and lower housing prices is assumed to have lower construction cost and vice versa.

The initialization and calibration of other selected variables are listed in Table 4.1.

4.5 Results

The preliminary results are shown in Figure 4.6. The dashed curves are simulated house price index using the Agent-based housing market model, and the sold curves are estimated house price index from empirical data.

Table 4.1: Model initialization and calibration

Variable		**Value**
annual depreciation rate	(0,1)	0.01
capital gain tax rate	(0,1)	from data
construction cost	R_+	estimated from data
initial housing price	R_+	from data
maximum debt to income ratio	R_+	0.4
median household income	R_+	from data
mortgage rate	percent	from data
net population growth rate	percent	from data
population in each region	I_+	from data
property tax rate	(0,1)	from data
rent-to-value ratio	R_+	0.05
transaction cost as % of housing value	(0,1)	0.05

4.6 Concluding Remarks on Chapter 4

This study is an experiment to adapt an existing conceptual agent-based model to a specific context. In particular, we build on the agent-based housing market model developed by in Chapter 3 and develop an application to the Washington D.C. housing market. We address the following research questions in the paper: How to explore the full potential of an existing agent-based model and take advantage of the flexibility and adaptability of agent-based models in general? How to move from conceptual to contextual models and empirically calibrate and validate the models? Finally, how to study intra-city variations in housing market dynamics?

To address the above issues, we have constructed a cluster-level house price index for 20 out of 39 clusters in Washington D.C. We have also adapted and calibrated the theoretical model with demographic and housing data in Washington D.C. For example, we regress housing transaction data on demographic variables and use residuals as the indicator for location quality across regions. We also calibrate construction cost in different regions using housing price and number of houses in each region.

The preliminary results are unclear. The fit at the cluster-level is relatively poor.

The reasons for the poor performance are manifold. First, there is limited data available for the estimation of cluster-level HPIs. Only a small proportion of properties has been transacted over the targeted period, which could introduce significant noises in the estimation of cluster-level HPIs. Second, further improvements can be made on model calibration. Individual-level demographic data such as household income and wealth could be employed to better calibrate household agents in the computational model. The calibration of location quality for Washington D.C. can be supplemented data on average house attributes across clusters. Third, the theoretical model previously developed in Chapter 3 still makes many simplifications in the behavior rules of the market participants. Human beings are more complex than computational robots. The model may not be ready for cluster-level value prediction yet.

This study is just a start of an experimental process of moving from a conceptual, proof-of-concept type of social simulation model to a contextual, empirically-sound one. This exercise poses questions that social simulation modelers rarely ask before: what data is needed for different types of model calibration and validation? Is data requirement different for agent-based social simulation than equation-based social simulation? How to start from a simple thought experiment and develop it into a model with predicting power? Finally, how to take full advantage of the flexibility and adaptability of agent-based computational models? These questions are often overlooked and the answers are certainly not there. Get ready for a journey of adventure.

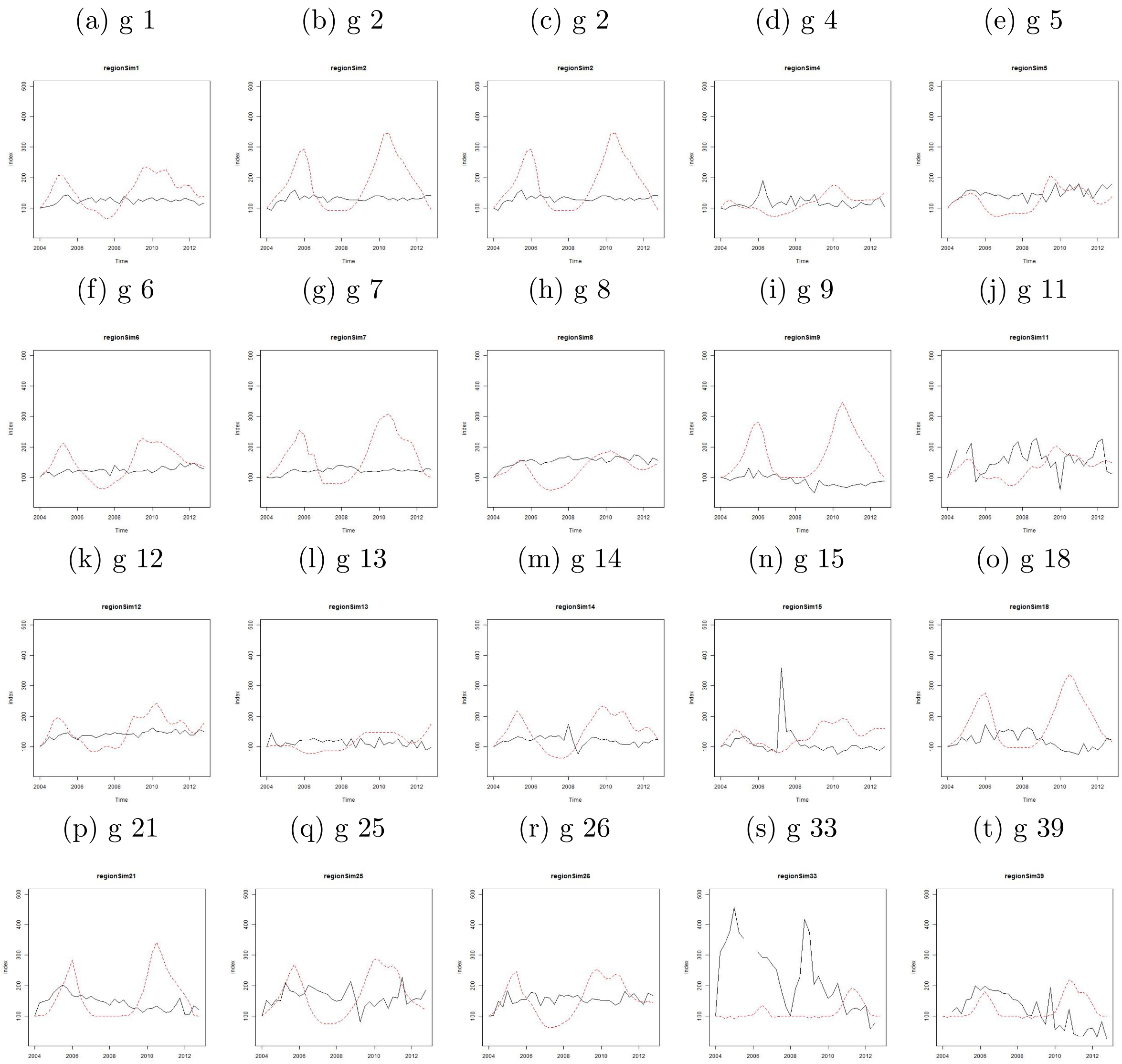

Figure 4.6: Comparison of simulated and estimated housing price indices in D.C.

CHAPTER 5. GENERAL CONCLUDING REMARKS

This dissertation covers topics ranging from housing markets to water protection. The three independent studies included in the dissertation have two things in common: they all explore innovative tools and methodologies in economic research, such as agent-based computational modeling and meta-analysis, and they all use those new tools in an attempt to solve important issues in the real world. Chapter 2 looks at the issue of water protection and develops a meta-analysis based on non-market valuation studies of water quality improvement. Chapter 3 develops a spatial agent-based computational model of a housing market to help understand the causes of the rise and collapse of US housing prices during the years immediately preceding the U.S. financial crisis (2007-2009). Chapter 4 adapts and extends the model developed in Chapter 3 to conducts a case study of the Washington D.C. housing market using empirical evidence. The dissertation presents alternative approaches to economic issues and at the same time raises new questions.

In Chapter 2, we ask an important question in natural resource management: How much value is clean water worth? We try to find the answer by estimating a valuation model based on a meta-analysis of non-market valuation studies of water quality improvements. After reviewing more than 100 non-market valuation studies on aquatic sites, we have collected 332 valuations from 37 distinct existing studies in the meta-database. The valuation model estimated in this study can be used to predict the mean willingness to pay for water quality improvement in a given site by households living in a given region. We also test the null hypothesis that the three main approaches in non-market valuation — the hedonic model, the travel cost model, and the contingent valuation model — generate consistent valuation estimates. Our test results reject the null hypothesis. We conclude that the hedonic model tends to produce the largest valuation, which is followed by the travel cost model, which is followed by the contingent valuation model.

In Chapter 3, our agent-based computational simulation shows that housing prices can rise and collapse by itself without any external economic shock. The two necessary conditions for the endogenous rise and collapse of housing prices are lenient financing and speculation. We find that banks have incentives to engage in aggressive lending even without the securitization of mortgage loans. We also find that we can achieve both affordable housing and housing market stability by setting the down payment requirement at an optimal level. Moreover, we show that the agent-based model is able to generate price patterns similar to those in the U.S. housing market.

In Chapter 4, we show an example of moving from a conceptual, proof-of-concept type of social simulation model to a contextual, empirically-sound one. In particular, we tailor the agent-based housing market model developed in Chapter 2 and to the city of Washington D.C. using empirical evidence. Chapter 3 poses questions that social simulation modelers have only recently begun to explore in depth: what is the data requirement for calibration and validation for different types of agent-based models? Is data requirement different for agent-based models than equation-based models? How to start from an abstract thought experiment and develop it into a model with predicting power? Finally, what steps should we take to take full advantage of the flexibility and adaptability of agent-based computational models? These questions are yet to be answered. We believe that we are already in a still largely unknown territory where more research effort is needed and the payoff can also be rewarding.

In summary, this dissertation studies a range of issues that are empirically important and complex in nature. We explore tools and research methodologies that are relatively new in economic research to tackle those issues. We find the new tools have provided us with new perspectives and insights. We thus feel the urgent need to expand the existing economic toolkit and explore alternative approaches in economic research, especially since the breakout of the recent economic crisis in 2009 when most standard economic models were unable to account for the situation. We have achieved promising results from our exploration of the new areas, but we have also encountered challenges as we venture into new territory. We believe this is exactly why we need to do it and where more research effort should be directed that will pay off. Let us be bold and embrace the new era.

APPENDIX A. CHAPTER 2 APPENDIX

Table A.1: Summary statistics of water attributes from National Lakes Assessment

Parameter	Mean	Std. Dev.	Min.	Max.
dissolved oxygen	84.99	17.05	7.11	99
pH	70.49	17.61	14.4	93
nitrates	97.67	1.57	63.85	97.98
total phosphorous	95.3	11.03	12.54	98.98
turbidity	81.39	16.02	18.2	96.62
wqi	85.73	8.25	42.21	96.97
secchi (m)	2.26	2.55	0.11	36.71
N		1094		

Table A.2: List of all estimates used in the meta analysis

Study	Author	# estimates	WTP Estimate in 2010$
1	Azevedo et al. (2001)	5	137.28
			749.76
			112.20
			726.00
			561.00
2	Bockstael et al. (1987)	1	55.38
3	Bockstael et al. (1989)	4	263.78
			82.84
			618.18
			128.53

Table A.2 Continued

4	Boyle et al. (1999)	6	1010.30
			3401.80
			295.55
			1992.96
			288.64
			3669.88
5	Boyle and Bouchard (2003)	22	97.72
			114.27
			636.77
			499.20
			218.20
			409.22
			92.37
			113.65
			457.99
			543.54
			315.37
			402.03
			786.77
			993.23
			372.14
			439.05
			198.71
			257.75
			157.62

Table A.2 Continued

			189.67
			570.17
			741.08
6	Brashares (1985)	7	185.42
			79.74
			4.06
			153.62
			41.40
			4.30
			93.05
7	Carson and Mitchell (1993)	3	211.11
			158.90
			177.06
8	Croke et al. (1986)	6	85.99
			87.32
			98.76
			64.64
			73.85
			91.06
9	Cronin (1982)	20	67.03
			51.56
			85.99
			78.43
			81.27
			86.39

Table A.2 Continued

			58.76
			42.54
			34.93
			59.47
			84.11
			57.29
			96.58
			81.27
			56.68
			31.28
			73.21
			77.22
			41.88
			83.60
10	d'Arge and Shogren (1989)	2	1722.73
			911.78
11	Desvousges et al. (1987)	16	72.21
			39.59
			21.66
			62.50
			142.93
			91.88
			46.81
			149.90
			61.01

Table A.2 Continued

			43.82
			30.88
			77.69
			126.99
			72.96
			31.13
			106.82
12	Edwards (1984)	3	167.86
			189.66
			368.42
13	Egan et al. (2009)	20	311.19
			256.47
			285.53
			237.99
			84.80
			19.24
			8.85
			14.63
			12.70
			3.77
			187.36
			154.05
			133.04
			185.80
			110.77

Table A.2 Continued

			13.41
			7.77
			6.82
			12.37
			11.58
14	Epp and Al-Ani (1979)	1	166.69
15	Farber and Griner (2000)	18	54.40
			51.14
			38.08
			72.96
			73.43
			55.18
			74.79
			70.96
			51.34
			95.38
			96.73
			79.31
			128.71
			125.02
			108.15
			157.19
			160.79
			132.65
16	Gibbs et al. (2002)	4	94.48

Table A.2 Continued

			456.11
			328.65
			826.51
17	Gramlich (1977)	2	155.75
			282.69
18	Greenley et al. (1981)	12	63.04
			134.68
			199.72
			441.58
			103.16
			198.68
			294.52
			591.59
			74.03
			152.24
			225.59
			482.50
19	Huang (1986)	22	11.07
			16.94
			13.72
			8.98
			5.69
			5.80
			6.60
			5.19

Table A.2 Continued

			5.19
			6.04
			5.29
			7.59
			7.94
			5.78
			16.31
			8.74
			4.32
			3.17
			3.22
			4.96
			5.80
			6.49
20	Johnston et al. (1999)	3	19.84
			27.69
			54.82
21	Krysel et al. (2003)	74	1778.88
			79.08
			225.00
			217.16
			130.37
			55.29
			73.36
			68.28

Table A.2 Continued

40.76
322.78
20.11
200.32
89.01
451.17
39.58
24.74
18.87
347.39
51.69
48.45
26.48
67.32
4032.48
84.92
134.92
242.35
206.33
287.02
576.76
340.05
31.04
1276.19
4371.92

Table A.2 Continued

205.89
95.59
273.07
192.54
4212.44
105.23
527.47
276.09
295.64
70.12
90.25
87.21
53.54
435.34
29.02
245.70
151.78
586.40
48.27
36.82
24.99
426.92
95.39
60.00
33.48

Table A.2 Continued

			87.74
			5656.40
			98.87
			172.65
			350.68
			300.30
			336.06
			782.15
			498.45
			36.95
			1833.69
			5629.00
			439.80
			134.52
			384.81
			414.82
22	Leggett and Bockstael (2000)	1 395.49	
23	Lipton et al. (2004)	1	72.18
24	Magat et al. (2000)	7	316.18
			150.86
			233.52
			157.39
			141.97
			126.57
			674.47

Table A.2 Continued

25	Mathews et al. (1999)	3	19.84
			27.69
			54.82
26	Michael et al. (1996)	6	841.14
			1464.21
			443.39
			847.67
			703.25
			1014.30
27	Moore et al. (2011)	24	508.19
			600.59
			582.80
			457.83
			115.46
			247.70
			130.69
			53.19
			409.55
			450.91
			465.49
			363.37
			161.88
			224.56
			120.59
			46.81

Table A.2 Continued

			382.95
			520.77
			807.63
			421.67
			89.14
			144.32
			245.57
28	Mullen and Menz (1985)	1	79.20
29	Randall et al. (2001)	3	89.08
			41.92
			61.57
30	Schuetz et al. (2001)	3	5.07
			11.37
			17.04
31	Smith et al. (1983)	2	16.90
			35.49
32	Smith et al. (1986)	14	52.74
			76.89
			10.48
			50.12
			520.83
			1269.00
			17.83
			77.64
			127.44

Table A.2 Continued

			26.22
			121.39
			1153.62
			2974.06
			71.86
33	Steinnes (1992)	2	18.85
			21.96
34	Stumborg et al. (2001)	1	452.52
35	Wey (1990)	2	70.14
			63.23
36	Whitehead (2005)	6	103.71
			105.26
			104.21
			16.86
			45.43
			45.26
37	Young (1984)	2	587.50
			525.00

Table A.3: Clustered robust regression results (with secchi as the indicator)

	Pooled 1	Pooled 2	CV	Hedonic
	(1)	(2)	(3)	(4)
D-NE	81.73	57.37	-24.51	-269.67
	(72.56)	(62.51)	(22.70)	(372.93)
D-lakeEstuary	215.00*	215.83	252.66***	408.70
	(125.38)	(132.34)	(44.16)	(561.91)
pubDate	7.40	6.41	-3.62*	-66.50**
	(4.97)	(4.88)	(1.96)	(32.41)
D-inPerson	281.32**	273.50**	83.38***	
	(112.14)	(113.95)	(25.06)	
income	-.01	-.01	.002	-.02
	(.01)	(.01)	(.003)	(.02)
D-totalValue	75.51	92.14*	108.42***	
	(59.20)	(54.56)	(25.04)	
D-improvement	-246.40**	-228.94**	-73.10***	-355.87**
	(114.22)	(110.01)	(23.87)	(159.39)
D-ladder	-100.75	-53.25	-129.79**	-11466.95***
	(101.22)	(102.46)	(59.20)	(3048.98)
startingSecchi	-87.88**	-87.30**	-86.65***	-91.98**
	(36.92)	(35.92)	(16.62)	(35.76)
deltaSecchi	2.81	2.99	-.24	2262.46***
	(2.28)	(2.22)	(1.25)	(579.30)
D-CV	-218.02	-94.74		
	(137.41)	(114.70)		
D-hedonic	424.70***	537.73***		
	(144.22)	(152.51)		
sitesize	.06*		-.006	24.27***
	(.03)		(.02)	(6.17)
regionsize	-.004*		.0004	
	(.002)		(.001)	
N	332	332	146	127
r_2	.14	.13	.47	.49
F	23.31	18.52	68.46	.

[a] *p=.10 **p=.05 ***p=.01

[b] secchi is used as the water quality indicator instead of water quality index

APPENDIX B. CHAPTER 3 APPENDIX

B.1 An Introduction to Agent-based Computational Modeling

Agent-based computational modeling is a research methodology that simulates natural, operational and social systems as automated computational agents interacting in a virtual environment. It sees the world as composed of individual agents, be it electronics or human beings, and the observed phenomenon the outcomes of those agents' interaction. As a result, agent-based models are capable of modeling systems that are complex and heterogeneous. Agent-based computational model is a major methodology in fields like engineering, national defense, and epidemiology. In areas like power market, it has become the dominant research methodology. In social sciences such as sociology and economics, agent-based computational models are increasingly used to study complex social or economic systems.

Agent-based computational economics (ACE) is a field in economics that studies the dynamic economic systems as virtual worlds of interacting agents (Tesfatsion ()). The defining characteristic of ACE models is their constructive grounding in the interactions of agents (Tesfatsion and Judd (2006)). In ACE models researchers obtain high-level aggregate outcomes from individual agent's interactions on the ground. In other words, researchers are "growing economies from the bottom up" (Tesfatsion (2002)). Because of that, agent-based modeler is able to relax many of the assumptions in equation-based models, and adopt a more flexible and/or realistic model setting.

B.2 Proof of Local Stability for the Dynamic Housing Market

The non-linear system in Section 3.3 is as follows,

$$\dot{m}(t) = \begin{cases} -m(t) & \text{if } p(t) \geq (1 - \text{down}) \\ \frac{1-\text{down}}{p_t} + \frac{p_t}{1-\text{down}} - 2 - m(t) & \text{if } p(t) < (1 - \text{down}) \end{cases}$$

$$\dot{p}(t) = \frac{1}{1 + m(t)} - p(t) \tag{B.1}$$

The above system has a equilibrium at $(\bar{m} = 0, \bar{p} = 1)$. The system can be linearized at the equilibrium as follows,

$$\begin{pmatrix} \dot{m}(t) \\ \dot{p}(t) \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} m(t) \\ p(t) - 1 \end{pmatrix} \tag{B.2}$$

The eigenvalues of the Jacobian matrix of the linearized system at equilibriums are both -1, which means the equilibrium is locally stable.

B.3 The Structural Framework of the Housing Market

In this section We are going to present the structural framework of housing market that we used to develop programming code and generate simulation results. Before we proceed, we list below some key notations which will be used through out the rest of the section.

B.3.1 Classification of Variables

Indices and Index Sets

- t: index for discrete time points, month t=[t, t+1)
- B_t: set of buyers at time t, indexed by b
- G: set of 25 regions, g=0,1,2,...,24

- H_t: set of homeowners at time t, indexed by h
- $A_t = B_t \cup H_t$: set of all buyers and homeowners (decision-making private agent) at time t, indexed by i
- C_t: set of bank's clients at time t, $C_t \subseteq A_t$, indexed by i

Exogenous Variable

- c_1, c_2, c_3: parameters for construction cost for developer
- DC^h: default cost of homeowner h
- LocQ^g: location quality of region g
- income^i: monthly income attribute of agent i
- M: loan maturity measured in months set by bank set at 20 year or 240 months for all loans at current study
- MaxDTI: maximum debt-to-income ratio permitted by the bank
- MaxTOM: maximum time-on-market permitted for any Buyer b
- down: minimum down payment rate required by the bank
- N^g: the set of neighboring (bordering) regions for region g
- r_t^F: annual risk-free return rate at time t
- r^H annual rate of cost of holding a house
- T^{Dev}: number of periods it takes the developer to build a house
- ω^i: savings attribute of agent i

Endogenous Variable

- $\overline{co_t}$: mean collateral rate of all clients of the bank at time t
- co_t^i: collateral rate of client i at time t
- $\mathrm{DTI}_t^i(g)$: debt-to-income ratio of agent i in at time t if agent i buys a property in region g
- $ER_t^i(g)$: annual return rate expected by speculative agent i at time t from the purchase and reselling of a house in region g in a year
- $EU_t^b(g)$: monthly utility from living in a house in region g expected by regular buyer b at time t
- F_t^b: the set of regions considered by buyer b for purchase of a house at time t, including the null region
- $L_t^i(g)$: loan value of agent i at time t if agent i buys a property in region g
- m_t: annual mortgage rate required by bank at time t
- M_t^i: remaining loan maturity of agent i at time t
- $mp_t^i(g)$: monthly mortgage payment of agent i at time t if agent i buys a property in region g
- NbhdQ_t^g: neighborhood quality of region g at time t
- p_t^g: housing price at the beginning of month t in region g
- prob_t^D: default probability of the lending pool for month t estimated by the bank
- TOM_t^i: time-on-market of agent i at time t

B.3.2 The Mortgage Contract

A mortgage contract between the bank and agent i for a property in region g at time τ is made up of the following components,

- eligibility requirements:
  - maximum debt-to-income ratio, MaxDTI
  - minimum down payment rate, down
- value of loan for agent i at time τ in region g, $L^i_\tau(g)$
- loan maturity in months, M
- annual mortgage rate set by bank for agent i at time τ, m^i_τ
- collateral for the loan, which is a property in region g

Loan maturity is set to be 240 periods or months at the current study. In each period the remaining loan maturity of agent i is deducted by one.

$$M^i_\tau = M = 240 \tag{B.3}$$
$$M^i_t = \begin{cases} 0 & \text{if } t < \tau + M \\ M^i_{t-1} - 1 & \text{if } \tau < t \leq \tau + M \\ 0 & \text{if } t > \tau + M \end{cases}$$

Agent i's mortgage rate equals bank's mortgage rate at time τ, m_τ, and it does not change throughout the mortgage maturity. Agent i's Outstanding loan in region g at time t, $L^i_t(g)$, is defined in the following equation,

$$L^i_\tau(g) = (1 - \text{down}) \cdot p^g_\tau \tag{B.4}$$
$$L^i_t(g) = \begin{cases} 0 & \text{if } t < \tau \\ (1 + m_\tau/12) \cdot \big(L^i_{t-1} - mp^i_t(g)\big) & \text{if } \tau \leq t \leq \tau + M \\ 0 & \text{if } t > \tau + M \end{cases}$$

where p^g_τ is housing price in region g at time τ, and $mp^i_t(g)$ is agent i's monthly payment in region g at time t. Agent i's monthly payment at time t for a property in region g, $mp^i_t(g)$, is a function of the initial loan value $L^i_\tau(g)$, mortgage rate m_τ, total loan maturity M, and time t:

$$mp^i_t(g) = \begin{cases} 0 & \text{if } t < \tau \\ mp(L^i_\tau(g), m_\tau) = \frac{(m_\tau/12)\cdot L^i_\tau(g)}{1-(m_\tau/12)^M} & \text{if } \tau \leq t \leq \tau + M \\ 0 & \text{if } t > \tau + M \end{cases} \tag{B.5}$$

Agent i is eligible for a mortgage contract in region g at time τ if and only if all of the following conditions are met:

$$p^g_\tau - L^i_{\tau^i}(g) = \text{down} \cdot p^g_\tau \leq \omega_i \tag{B.6}$$

$$\frac{mp^i_\tau(g)}{\text{income}^i} \equiv DTI^i_\tau(g) \leq \text{MaxDTI} \tag{B.7}$$

where income_i is agent i's monthly income and ω_i is agent i's savings. Both are attributes of agent i and remain the same throughout time. The first condition says that the minimum down payment has to be less than or equal to savings. The second condition says that the debt-to-income ratio has to be less than or equal to the maximum debt-to-income ratio.

B.3.3 Bank

We assume perfect competition among banks (funds) for the loans. The market arbitrage condition thus require that the expected return on the loans equals the exogenous market return, r^F_t,

$$\text{prob}^D_t \cdot \overline{co_t} + (1 - \text{prob}^D_t) \cdot (1 + r_t) = 1 + r^F_t \tag{B.8}$$

where prob^D_t is the default probability estimated by bank, $\overline{co_t}$ is the mean collateral rate, m_t is the mortgage rate, and r^F_t is risk-free return rate at time t. prob^D_t is defined as follows,

$$\text{prob}_t^D = \frac{\sum_{i\in C} I(\text{client } c \text{ went into foreclosure})}{\sum_{i\in C} 1} \tag{B.9}$$

where C is the set of all clients borrowing from the bank, I is the indicator function, and $\overline{co_t}$ is the mean collateral rate, which is defined below,

$$\overline{co_t} = \frac{\sum_{i\in C} co_t^i \cdot L_t^i}{\sum_{i\in I} L_t^i} \tag{B.10}$$

where co_t^i is collateral rate of client i, which is defined below,

$$co_t^i = \max\left\{\frac{p_t^i}{L_t^i}, 1 + m_{\tau(i)}\right\} \tag{B.11}$$

where $p_t^{i,g(i)}$ is value of client i's property at time t, L_t^i is agent i's outstanding loan value at time t and $m_{\tau(i)}$ is the mortgage rate client i is paying when client i took out the loan in period $\tau(i)$. Bank's mortgage rate at period t, r_t thus is,

$$m_t = \frac{1 + r_t^F - \text{prob}_t^D \cdot \overline{co_t}}{1 - \text{prob}_t^D} - 1 \tag{B.12}$$

B.3.4 Regular Buyer

For regular buyer b, its objective is,

$$\max_{g\in F^b} \quad U_t^b(g) = U\left(\text{NbhdQ}_t^g, \text{LocQ}^g, \text{Composite}_t^{b,g};\ \boldsymbol{\beta}^b\right) \tag{B.13}$$

where F^b is the feasible choice set of regions for buyer b. It includes the null option, that is, the option of not buying. We assume U^b(null)=0. NbhdQ^g is the neighborhood qualify in region g, LocQ^g is the location quality in region g. Composite$_t^{b,g}$ is the composite good buyer b can consume if buyer b buys a property in region g at time t. $\boldsymbol{\beta}_b$ is buyer-specific preference parameter vector in the utility function. Composite$_t^{b,g}$ is defined as,

$$\text{Composite}_t^{b,g} = \left(1 - DTI_t^{b,g}\right) \cdot \text{income}^b \tag{B.14}$$

where income^b is buyer b's monthly income, and $DTI_t^{b,g}$ is buyer b's debt to income ratio if she purchases a property in region g. Composite$_t^{b,g}$ thus is the composite good buyer b can afford after paying the monthly payment for mortgage.

During month t, regular buyer b searches each region in the feasible choice set, and chooses the region that gives her the highest expected utility. Call this chosen region g^*. Regular buyer b's bid at period t for a house in region g^* is,

$$\text{bid}_t^{b,g^*} = (1 - \eta + \delta \cdot \text{TOM}_t^b) \cdot P_{t-1}^{g^*} \tag{B.15}$$

where η is a buyer's initial discount on current price, assumed to be 10% in the study. δ is the percentage increase in price each period the buyer's bid is not accepted, assumed to be 2% in the study. $P_{t-1}^{g^*}$ is the last period price in buyer b's chosen region g^*. TOM_t^b is buyer b's time-on-market at period t. The bigger the time-on-market, the higher the bidding price. For one more period on market, the bidding price is increased by δ percent.

Once the price is settled by the real estate agent, regular buyer b's action in period t, a_t^b is defined below.

$$a_t^b = \begin{cases} \text{buy a house and become a homeowner} & \text{if bid}_t^{b,g^*} \geq P_t^{g^*} \\ \text{enter period t+1, TOM}_{t+1}^b = \text{TOM}_t^b + 1 & \text{if bid}_t^{b,g^*} < P_t^{g^*} \text{ and TOM}_t^b < \text{MaxTOM} \\ \text{exit market} & \text{if bid}_t^{b,g^*} < P_t^{g^*} \text{ and TOM}_t^b \geq \text{MaxTOM} \end{cases}$$

where MaxTom is the maximum time on market. A buyer will continue to increase her bid until the time she has been on the market exceeds the maximum waiting period. Then the buyer will leave the housing market.

B.3.5 Speculative Buyer

We define $ER_t^b(g)$ as speculative buyer b's expected annual return rate of on the purchase of a house in region g in period t and resell the house in a year. $\forall b(g) \neq \text{null}$, $ER_t^b(g)$ equals,

$$ER_t^b(g) = (1 - \omega^b)\frac{p_t^g}{p_{t-12}^g} - m_t - r^H \tag{B.16}$$

where ω^b is the discount rate speculative buyer b assign to price appreciation, $\omega^b \in (0, 1)$. The expected net return also depends on the mortgage rate at time t, m_t, and the annual cost rate of holding a house as percentage of property value, r^H. The latter equals the sum of property tax rate, transaction cost rate, depreciation rate, subtracted by rental income as percentage of property value.

For a speculative buyer, her objective is to maximize net return rate. Speculative buyer b's objective function thus is,

$$\max_{g \in F^b} \quad ER_t^b(g) \tag{B.17}$$

where F^b is the feasible choice set of regions for speculative buyer b, including the option of not buying a house or g=null. We assume $ER(\text{null}) = m_t^F \cdot \text{down}$, because investing in the housing market gives a buyer access to the credit market to which is otherwise unaccessible to her. Because of this leverage effect, returns on the real estate investment is magnified.

During month t, speculative buyer b searches each region in the feasible choice set, and chooses the region that gives her the highest expected return. Call that chosen region g^*. Regular buyer b's bid at period t for a house in region g^* is,

$$\text{bid}_t^{b,g^*} = \begin{cases} \big(1 + ER_t^b(g^*)/12\big) \cdot P_{t-1}^{g^*} & \text{if } g^* \neq \text{null} \\ 0 & \text{if } g^* = \text{null} \end{cases} \tag{B.18}$$

Once the price is settled by the real estate agent, For any speculative buyer b, her action at time t, a_t^b is,

$$a_t^b = \begin{cases} \text{buy a house and become a homeowner} & \text{if } \text{bid}_t^{b,g^*} \geq P_t^{g^*} \\ \text{exit market} & \text{if } \text{bid}_t^{b,g^*} < P_t^{g^*} \end{cases}$$

B.3.6 Regular Homeowner

Once a regular buyer b has succeeded in the purchase of a house in her chosen region g^*, she becomes a regular homeowner h in region $g(b)$. At the beginning of each period, homeowner h can hold, list, or default on her property. A regular homeowner h's action at period t, a_t^h, thus is,

$$a_t^h = \begin{cases} \text{default} & \text{if } P_{t-1}^{g^*} + \mathrm{DC}^h < L_t^h \\ \text{list property, ask} = (1+\eta-\delta\cdot\mathrm{TOM}_t^h)P_{t-1}^{g^*} & \text{if } h \text{ does not default and} \\ & \text{job move occurs} \\ \text{hold property} & \text{if } h \text{ does not default and} \\ & \text{job move does not occur} \end{cases}$$

where L_t^h is homeowner h's outstanding loan at time t and DC^h is the default cost of homeowner h. η is the targeted margin over current price, which is set to be 10% at current study. δ is the percentage decrease in price each period a seller's bid is not accepted, which is set to be 2% at current study. P_{t-1}^{h,g^*} is the price in the region in period $t-1$. TOM_t^h is seller h's time one market at period t.

A regular homeowner will default if the price of her property and the default costs fall below the value of the outstanding loan. Default costs are costs associated with default and foreclosure, such as legal cost, loss of credibility, and mental stress. Default cost of homeowner h is assumed to be proportional to the monthly income of homeowner h at the current study. Moreover, there is a small probability (assumed to be 5% in this study) that any homeowner will want to sell the house for exogenous reasons such as job move or divorce. We call the rare event "job move". When job move occurs, homeowner h will list her property. Otherwise she will hold the property.

B.3.7 Speculative Homeowner

For speculative homeowner h, her objective is,

$$\max_{\text{default, hold, list}} ER_t^{h,g^*} \tag{B.19}$$

where R_t^{h,g^*} is speculative buyer h's expected return rate on the property she currently owns in region g^* at time t.

Speculative buyer h's action at period t thus is,

$$a_t^h = \begin{cases} \text{default} & \text{if } P_{t-1}^{g*} + \mathrm{DC}^h < L_t^h \\ \text{list the house, ask} = \left(1 + ER_t^{h,g^*}/12\right) P_{t-1}^{g^*} & \text{if } h \text{ does not default and} \\ & ER_t^{h,g^*} < ER(\text{null}) \\ \text{hold the property} & \text{if } h \text{ does not default and} \\ & ER_t^{h,g^*} \geq ER(\text{null}) \end{cases}$$

Unlike a regular homeowner who lists her house for exogenous reasons, a speculative homeowner will list her house if and only if the expected return on the property is negative. Like a regular homeowner, a speculative homeowner will default if the price of her property and the default costs fall below the value of the outstanding loan.

B.3.8 Developer

The total cost of building q houses in region g at the beginning of period t, $TC_t^g(q)$, is assumed to be a quadratic function in q:

$$TC_t^g(q_t^g) = c_1 h + \frac{c_2}{2}\left(q_t^g\right)^2 + \frac{c_3}{2}\left(Q_{t-1}^g\right)^2 \tag{B.20}$$

where c_1, c_2, and c_3 are positive parameters for construction cost. Q_{t-1}^g is total number houses already exist on region g at period $t-1$; while q_t^g is the number of houses just finished being constructed in region g at time t. We assume that the more houses exist in the region, the more expensive to build new constructions due to land scarcity, regulations, etc. Q_t^g for all regions $g \in G$ is updated at the beginning of period t,

$$Q_0^g = 0 \tag{B.21}$$

$$Q_t^g = Q_{t-1}^g + q_t^g \ \forall t > 0$$

The marginal cost, $MC_t^g(q_t^g)$, thus is

$$MC_t^g(q_t^g) = c_1 + c_2 q_t^g + c_3 Q_{t-1}^g \tag{B.22}$$

The developer's supply in region g in period t, q_t^g, is set where price equals marginal cost T^{Dev} periods ago, because the decision was made T^{Dev} periods ago, where T^{Dev} is the number of periods it takes for the developer to build houses.

$$P^g_{t-T^{Dev}} = MC^g_t\left(q^g_t\right) = c_1 + c_2 q^g_t + c_3 Q^g_{t-T^{Dev}} \tag{B.23}$$

Hence q_t^g equals,

$$q^g_t = \frac{P^g_{t-T^{Dev}} - c_1 - c_3 Q^g_{t-T^{Dev}}}{c_2} - \text{qStock}^g_t \tag{B.24}$$

where stock^g_t is housing stock in region g at period t. stock^g_t is updated at the beginning of period t,

$$\begin{aligned} \text{qStock}^g_0 &= 0 \\ \text{qStock}^g_t &= \text{qStock}^g_{t-1} + q^g_t - \text{qSold}^g_t \ \forall t > 0 \end{aligned} \tag{B.25}$$

where sold^g_t is the number of houses sold in region g during period t, which equals

$$\text{qSold}^g_t = \min\left\{q^g_t, \frac{p^g_t - c_1 - c_3 Q^g_{t-T^{Dev}}}{c_2}\right\} \tag{B.26}$$

The asking price for a newly constructed house j at period t, $\text{ask}^{j,g}_t$, is its marginal construction cost,

$$\begin{aligned} \text{ask}^{j,g}_t &= c_1 + c_2 \cdot j + c_3 H^g_t \\ \forall j &= 1, 2, ..., q^g_t \end{aligned} \tag{B.27}$$

Existing housing stocks in region g at time t are listed at $c_1 + c_3 H^g_{t-1}$.

B.4 UML Presentation of the Housing Model

Figure B.1 is a class diagram of the model. It summarizes the model's class structure and demonstrates relationships between different types of agents.

Figure B.2 is an activity diagram of the housing model. It summarizes how market participants interact at each point. A red line represents an information flow, and a black line represents an action flow.

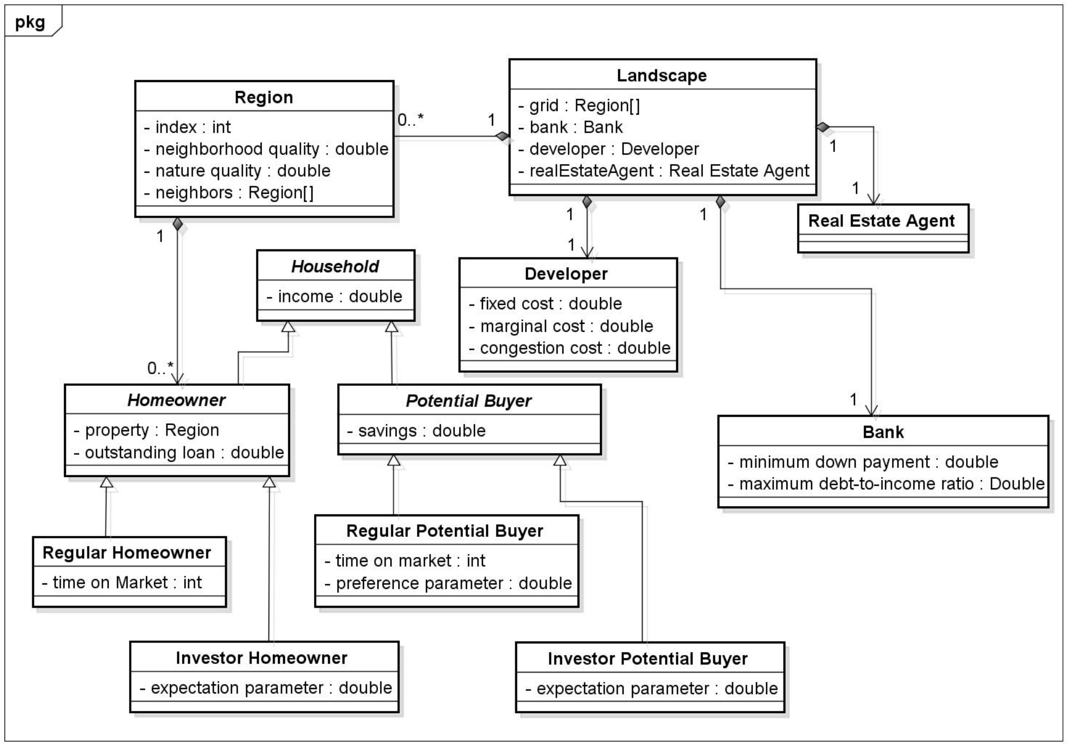

Figure B.1: Class diagram for the housing model

B.5 Results from Computational Experiments

Housing price volatility for the complete parameter space are shown in Figure B.3. In the red rectangle is the more densely-sampled region. Number of non-speculative homeowners for the complete parameter space are shown in Figure B.4. In the red rectangle is the more densely-sampled region.

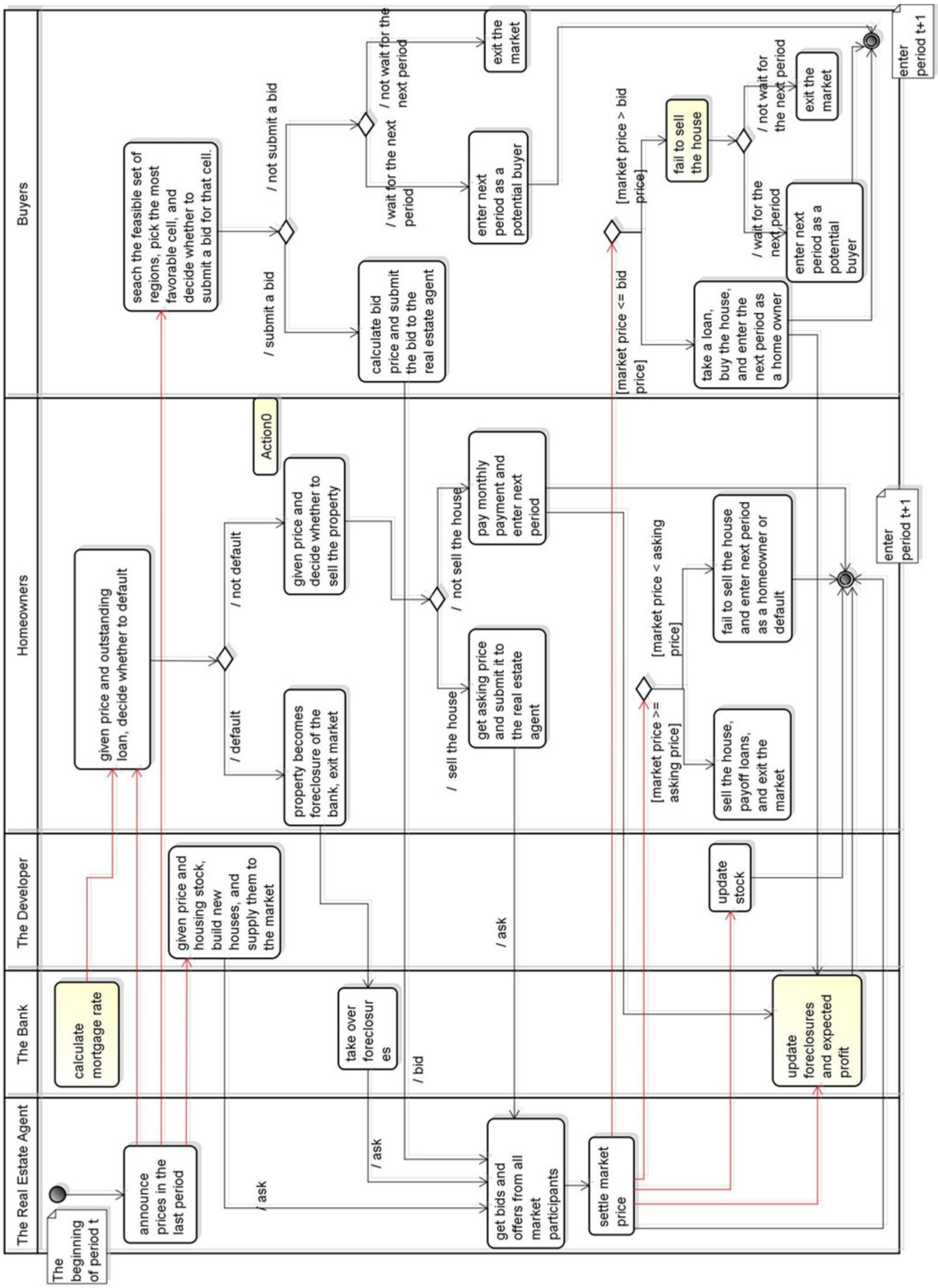

Figure B.2: Activity diagram for the housing model

S \ L	0.000	0.200	0.400	0.600	0.700	0.750	0.800	0.825	0.850	0.875	0.900	0.925	0.950	0.975	1.000
0.000	0.000	0.000	0.003	0.023	0.025	0.022	0.021	0.026	0.017	0.017	0.016	0.015	0.017	0.018	0.018
	(0.0000)	(0.0000)	(0.0044)	(0.0038)	(0.0067)	(0.0022)	(0.0026)	(0.0037)	(0.0020)	(0.0024)	(0.0022)	(0.0016)	(0.0023)	(0.0025)	(0.0021)
0.025	0.000	0.000	0.003	0.032	0.031	0.030	0.026	0.029	0.033	0.062	0.111	0.181	0.204	0.244	0.235
	(0.0000)	(0.0000)	(0.0038)	(0.0038)	(0.0043)	(0.0038)	(0.0029)	(0.0029)	(0.0050)	(0.0093)	(0.0452)	(0.0648)	(0.0568)	(0.0350)	(0.0593)
0.050	0.000	0.000	0.002	0.042	0.039	0.038	0.040	0.046	0.060	0.123	0.223	0.219	0.228	0.241	0.222
	(0.0000)	(0.0000)	(0.0033)	(0.0037)	(0.0056)	(0.0055)	(0.0053)	(0.0067)	(0.0105)	(0.0227)	(0.0309)	(0.0327)	(0.0218)	(0.0211)	(0.0235)
0.075	0.000	0.000	0.003	0.042	0.046	0.051	0.055	0.067	0.085	0.162	0.265	0.247	0.228	0.189	0.213
	(0.0000)	(0.0000)	(0.0045)	(0.0042)	(0.0057)	(0.0070)	(0.0068)	(0.0082)	(0.0090)	(0.0368)	(0.0248)	(0.0295)	(0.0198)	(0.0759)	(0.0527)
0.100	0.000	0.000	0.002	0.046	0.054	0.059	0.066	0.081	0.097	0.178	0.273	0.258	0.246	0.224	0.214
	(0.0000)	(0.0000)	(0.0034)	(0.0035)	(0.0065)	(0.0066)	(0.0081)	(0.0077)	(0.0113)	(0.0360)	(0.0236)	(0.0207)	(0.0323)	(0.0781)	(0.0745)
0.150	0.000	0.000	0.002	0.054	0.068	0.068	0.085	0.099	0.115	0.194	0.248	0.241	0.230	0.219	0.215
	(0.0000)	(0.0000)	(0.0044)	(0.0051)	(0.0062)	(0.0061)	(0.0087)	(0.0122)	(0.0153)	(0.0368)	(0.0277)	(0.0291)	(0.0289)	(0.0618)	(0.0551)
0.200	0.000	0.000	0.003	0.059	0.077	0.081	0.085	0.101	0.128	0.204	0.244	0.198	0.212	0.189	0.193
	(0.0000)	(0.0000)	(0.0041)	(0.0041)	(0.0110)	(0.0094)	(0.0114)	(0.0093)	(0.0191)	(0.0380)	(0.0316)	(0.0321)	(0.0346)	(0.0511)	(0.0268)
0.250	0.000	0.000	0.004	0.066	0.073	0.082	0.090	0.100	0.118	0.183	0.232	0.203	0.172	0.189	0.187
	(0.0000)	(0.0000)	(0.0050)	(0.0087)	(0.0064)	(0.0096)	(0.0084)	(0.0132)	(0.0211)	(0.0342)	(0.0393)	(0.0543)	(0.0582)	(0.0550)	(0.0326)

Figure B.3: Housing price volatility, complete range (standard deviation in parenthesis)

S \ L	0	0.2	0.4	0.6	0.7	0.75	0.8	0.825	0.85	0.875	0.9	0.925	0.95	0.975	1
0.000	0	0	400	3482	6667	8098	9476	9966	10150	10474	10850	10841	10562	10346	10151
	(0)	(0)	(31)	(77)	(54)	(20)	(27)	(28)	(20)	(22)	(16)	(12)	(15)	(16)	(17)
0.025	0	0	432	5146	6707	7635	8735	9346	9953	10491	11534	11886	11300	11322	8636
	(0)	(0)	(32)	(20)	(19)	(26)	(30)	(44)	(61)	(84)	(171)	(241)	(317)	(366)	(629)
0.050	0	0	454	4881	6479	7430	8520	9226	9898	10446	10526	10522	10970	9577	9451
	(0)	(0)	(46)	(28)	(30)	(32)	(55)	(73)	(99)	(202)	(545)	(493)	(508)	(563)	(696)
0.075	0	0	451	4695	6161	7168	8444	8984	9440	9390	9322	8659	9523	8884	9219
	(0)	(0)	(49)	(21)	(38)	(56)	(73)	(95)	(147)	(177)	(500)	(659)	(406)	(682)	(471)
0.100	0	0	481	4398	5939	6919	7968	8537	8964	8100	6319	7194	7436	7022	7705
	(0)	(0)	(40)	(36)	(53)	(71)	(116)	(120)	(163)	(354)	(460)	(412)	(328)	(522)	(581)
0.150	0	0	459	3875	5531	6409	7177	8205	8184	5707	4659	4214	4726	4585	4817
	(0)	(0)	(37)	(39)	(62)	(74)	(97)	(180)	(237)	(295)	(361)	(235)	(360)	(317)	(384)
0.200	0	0	432	3259	4927	5637	6529	6847	6586	4887	3677	2809	3180	3297	3589
	(0)	(0)	(34)	(77)	(96)	(107)	(86)	(217)	(246)	(249)	(268)	(195)	(295)	(329)	(279)
0.250	0	0	441	2442	4004	5086	5712	5962	5448	4346	2958	2616	2392	3137	2677
	(0)	(0)	(26)	(61)	(85)	(161)	(191)	(228)	(283)	(334)	(281)	(292)	(178)	(258)	(246)

Figure B.4: Number of non-speculative homeowners, complete range (standard deviation in parenthesis)

APPENDIX C. CHAPTER 4 APPENDIX

Table C.1: Estimation results: location quality

Variable	Coefficient	(Std. Err.)
year 2004	-23.900*	(11.092)
percentage of poor persons	-0.972	(2.130)
percentage of unemployed persons	-2.153	(1.333)
percentage of car holders	1.613	(2.376)
violent crime rate	-7.929†	(4.556)
property crime rate	-0.585	(1.097)
income	0.015	(0.507)
g2	-120.416†	(64.297)
g3	-26.897	(59.674)
g4	390.166**	(94.732)
g5	-255.442**	(80.924)
g6	-113.255	(78.215)
g7	-117.049*	(56.057)
g8	251.139**	(91.090)
g9	-221.470**	(59.893)
g10	24.818	(66.764)
g11	45.275	(81.524)
g12	15.534	(53.853)

Table C.1 Continued

g13	172.720†	(96.802)
g14	-64.235	(54.740)
g15	68.930	(81.608)
g16	-69.333	(75.478)
g17	-307.008**	(65.923)
g18	-245.612**	(63.065)
g19	-399.252**	(74.908)
g20	-362.322**	(83.456)
g21	-242.452**	(64.549)
g22	-252.411**	(66.726)
g23	-211.315*	(86.855)
g24	-276.366**	(67.364)
g25	-174.406**	(57.551)
g26	-152.912**	(50.076)
g27	16.102	(78.736)
g28	-282.143**	(84.442)
g29	-265.643**	(88.539)
g30	-336.178**	(78.763)
g31	-311.511**	(77.416)
g32	-368.057**	(74.484)
g33	-331.506**	(76.146)
g34	-382.641**	(67.460)
g35	-399.415**	(64.155)
g36	-293.426**	(92.092)
g37	-283.434**	(92.792)

Table C.1 Continued

g38	-308.869**	(83.287)
g39	-301.451**	(84.631)
Intercept	677.721**	(192.289)

BIBLIOGRAPHY

S&P/Case-Shiller home price index. http://us.spindices.com/indices/real-estate/sp-case-shiller-20-city-composite-home-price-index. Accessed: 2013-11-12.

Axtell, R. (1999). The emergence of firms in a population of agents. Technical report.

Azevedo, C., Herriges, J. A., and Kling, C. L. (2001). *Valuing preservation and improvements of water quality in Clear Lake*. Center for Agricultural and Rural Development, Iowa State University Iowa, USA.

Bianchi, C., Cirillo, P., Gallegati, M., and Vagliasindi, P. A. (2005). Validation and calibration in ace models: An investigation on the cats model.

Bianchi, C., Cirillo, P., Gallegati, M., and Vagliasindi, P. A. (2007). Validating and calibrating agent-based models: a case study. *Computational Economics*, 30(3):245–264.

Bockstael, N. E., Hanemann, W. M., and Kling, C. L. (1987). Estimating the value of water quality improvements in a recreational demand framework. *Water Resources Research*, 23(5):951–960.

Bockstael, N. E., McConnell, K. E., Strand, I. E., et al. (1989). Measuring the benefits of improvements in water quality: The chesapeake bay. *Marine Resource Economics*, 6(1):1–18.

Boswijk, H. P., Hommes, C. H., and Manzan, S. (2007). Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control*, 31(6):1938–1970.

Boyle, K. and Bouchard, R. (2003). Water quality effects on property prices in northern new england. *Lake Line*, 23(3):2427.

Boyle, K. J., Poor, P. J., and Taylor, L. O. (1999). Estimating the demand for protecting freshwater lakes from eutrophication. *American journal of agricultural economics*, pages 1118–1122.

Brashares, E. (1985). Estimating the instream value of lake water quality in southeast michigan. *Diss. University of Michigan.*

Brown, R. M., McClelland, N. I., Deininger, R. A., and Tozer, R. G. (1970). A water quality index- do we dare.

Calhoun, C. A. (1996). Ofheo house price indexes: Hpi technical description. *Office of Federal Housing Enterprise Oversight.*

Carley, K. M. (1996). Validating computational models. *Paper available at http://www. casos. cs. cmu. edu/publications/papers. php.*

Carson, R. T. and Mitchell, R. C. (1993). The value of clean water: the public's willingness to pay for boatable, fishable, and swimmable quality water. *Water resources research*, 29(7):2445–2454.

Case, K. E. and Shiller, R. J. (1989). The efficiency of the market for single-family homes.

Case, K. E. and Shiller, R. J. (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity*, 2003(2):299–362.

Commission, F. C. I. (2011). *Financial crisis inquiry report: final report of the national commission on the causes of the financial and economic crisis in the United States*. US Government Printing Office.

Croke, K., Fabian, R., and Brenniman, G. (1986). Estimating the value of improved water quality in an urban river system. *Journal of Environmental Systems*, 16(1):13–24.

Cronin, F. J. (1982). Valuing nonmarket goods through contingent markets. Technical report, Pacific Northwest Lab., Richland, WA (USA).

d'Arge, R. C. and Shogren, J. F. (1989). Okoboji experiment: Comparing non-market valuation techniques in an unusually well-defined market for water quality. *Ecological economics*, 1(3):251–259.

DeNavas-Walt, C. (2010). *Income, Poverty, and Health Insurance Coverage in the United States (2005)*. DIANE Publishing.

Desvousges, W. H., Smith, V. K., and Fisher, A. (1987). Option price estimates for water quality improvements: a contingent valuation study for the monongahela river. *Journal of Environmental Economics and Management*, 14(3):248–267.

Dubčáková, R. (2011). Eureqa: software review. *Genetic programming and evolvable machines*, 12(2):173–178.

Edwards, S. F. (1984). Analysis of the non-market benefits of protecting salt pond water quality in southern rhode island: an application of the hedonic price and contingent valuation techniques. Technical report, Rhode Island Univ., Kingston (USA).

Egan, K. J., Herriges, J. A., Kling, C. L., and Downing, J. A. (2009). Valuing water quality as a function of water quality measures. *American Journal of Agricultural Economics*, 91(1):106–123.

Epp, D. J. and Al-Ani, K. S. (1979). The effect of water quality on rural nonfarm residential property values. *American Journal of Agricultural Economics*, 61(3):529–534.

Epstein, J. M. (1996). *Growing artificial societies: social science from the bottom up.* Brookings Institution Press.

Farber, S. and Griner, B. (2000). Valuing watershed quality improvements using conjoint analysis. *Ecological Economics*, 34(1):63–76.

Geanakoplos, J. (2010). The leverage cycle. In *NBER Macroeconomics Annual 2009, Volume 24*, pages 1–65. University of Chicago Press.

Geanakoplos, J., Axtell, R., Farmer, J. D., Howitt, P., Conlee, B., Goldstein, J., Hendrey, M., Palmer, N. M., and Yang, C.-Y. (2012). Getting at systemic risk via an agent-based model of the housing market. *The American Economic Review*, 102(3):53–58.

Gibbs, J. P., Halstead, J. M., Boyle, K. J., and Huang, J.-C. (2002). An hedonic analysis of the effects of lake water clarity on new hampshire lakefront properties. *Agricultural and Resource Economics Review*, 31(1):39–46.

Gilli, M. and Winker, P. (2003). A global optimization heuristic for estimating agent based models. *Computational Statistics & Data Analysis*, 42(3):299–312.

Glaeser, E. L., Gyourko, J., and Saiz, A. (2008). Housing supply and housing bubbles. *Journal of Urban Economics*, 64(2):198–217.

Gode, D. K. and Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of political economy*, 101(1):119.

Gramlich, F. W. (1977). The demand for clean water: the case of the charles river. *National Tax Journal*, pages 183–194.

Greenley, D. A., Walsh, R. G., and Young, R. A. (1981). Option value: empirical evidence from a case study of recreation and water quality. *The Quarterly Journal of Economics*, 96(4):657–673.

Heslop, L. A., McGregor, E., and Griffith, M. (2001). Development of a technology readiness assessment measure: The cloverleaf model of technology transfer. *The Journal of Technology Transfer*, 26(4):369–384.

Holdsworth, S. (2011). *National Lakes Assessment: A Collaborative Survey of the Nation's Lakes*. DIANE Publishing.

Huang, C.-H. (1986). Recreation benefits of water quality improvement in selected lakes in minnesota. Technical report, Minnesota Univ., Minneapolis (USA).

Husslage, B. G. M., Rennen, G., van Dam, E. R., and den Hertog, D. (2011). Space-filling latin hypercube designs for computer experiments. *Optimization & Engineering*, 12(4):611–630.

Iacoviello, M. (2005). House prices, borrowing constraints, and monetary policy in the business cycle. *The American Economic Review*, 95(3):739–764.

Janssen, M. A. and Ostrom, E. (2006). Empirically based, agent-based models. *Ecology & Society*, 11(2).

Johnston, R. J., Besedin, E. Y., Iovanna, R., Miller, C. J., Wardwell, R. F., and Ranson, M. H. (2005). Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: A meta-analysis. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 53(2-3):221–248.

Johnston, R. J., Swallow, S. K., and Weaver, T. F. (1999). Estimating willingness to pay and resource tradeoffs with different payment mechanisms: an evaluation of a funding guarantee for watershed management. *Journal of Environmental Economics and Management*, 38(1):97–120.

Krysel, C., Boyer, E. M., Parson, C., and Welle, P. (2003). *Lakeshore property values and water quality: Evidence from property sales in the Mississippi Headwaters Region.* Mississippi Headwaters Board Walker, MN.

LeBaron, B. (2000). Agent-based computational finance: Suggested readings and early research. *Journal of Economic Dynamics and Control*, 24(5):679–702.

Leggett, C. G. and Bockstael, N. E. (2000). Evidence of the effects of water quality on residential land prices. *Journal of Environmental Economics and Management*, 39(2):121–144.

Li, H. and Tesfatsion, L. (2009). Development of open source software for power market research: The ames test bed. *Journal of Energy Markets*, 2(2):111–128.

Lipton, D. et al. (2004). The value of improved water quality to chesapeake bay boaters. *Marine Resource Economics*, 19(2):265–270.

Magat, W. A., Huber, J., Viscusi, W. K., and Bell, J. (2000). An iterative choice approach to valuing clean lakes, rivers, and streams. *Journal of Risk and Uncertainty*, 21(1):7–43.

Malleson, N. (2012). Using agent-based models to simulate crime. In *Agent-based models of geographical systems*, pages 411–434. Springer.

Mankiw, N. and Weil, D. N. (1989). The baby boom, the baby bust, and the housing market. *Regional Science and Urban Economics*, 19(2):235 – 258.

Markose, S., Giansante, S., Gatkowski, M., and Shaghaghi, A. R. (2009). Too interconnected to fail: Financial contagion and systemic risk in network model of cds and other credit enhancement obligations of us banks. *analysis*.

Mathews, L. G., Homans, F. R., and Easter, K. W. (1999). Reducing phosphorus pollution in the minnesota river: How much is it worth? Technical report, University of Minnesota, Department of Applied Economics.

Mayer, C. (2011). Housing bubbles: A survey. *Annu. Rev. Econ.*, 3(1):559–577.

McClelland, N. I. (1974). Water quality index application in the kansas river basin.

Michael, H. J., Boyle, K. J., and Bouchard, R. (1996). Water quality affects property prices: A case study of selected maine lakes.

Mitchell, R. C. and Carson, R. T. (1981). An experiment in determining willingness to pay for national water quality improvements. *draft report to the US Environmental Protection Agency, Washington, DC*.

Moore, R., Provencher, B., and Bishop, R. C. (2011). Valuing a spatially variable environmental resource: reducing non-point-source pollution in green bay, wisconsin. *Land Economics*, 87(1):45–59.

Mullen, J. K. and Menz, F. C. (1985). The effect of acidification damages on the economic value of the adirondack fishery to new york anglers. *American Journal of Agricultural Economics*, 67(1):112–119.

Poterba, J. M. (1984). Tax subsidies to owner-occupied housing: an asset-market approach. *The Quarterly Journal of Economics*, 99(4):729–752.

Randall, A., DeZoysa, D., and Yu, S. (2001). Ground water, surface water, and wetlands valuation in ohio. *The Economic Value of Water Quality*, pages 83–99.

Schuetz, J. F., Boyle, K., and Bouchard, R. (2001). Mr421: The effects of water clarity on economic values and economic impacts of recreational uses of maine's great ponds.

Silos, P. (2007). Housing, portfolio choice and the macroeconomy. *Journal of Economic Dynamics and Control*, 31(8):2774–2801.

Smith, V. K., Desvousges, W. H., and Fisher, A. (1986). A comparison of direct and indirect methods for estimating environmental benefits. *American Journal of Agricultural Economics*, 68(2):280–290.

Smith, V. K., Desvousges, W. H., and McGivney, M. P. (1983). Estimating water quality benefits: An econometric analysis. *Southern Economic Journal*, pages 422–437.

Sommervoll, D. E., Borgersen, T.-A., and Wennemo, T. (2010). Endogenous housing market cycles. *Journal Of Banking & Finance*, 34(3):557–567.

Stein, J. C. (1995). Prices and trading volume in the housing market: A model with down-payment effects. *The Quarterly Journal of Economics*, 110(2):379–406.

Steinnes, D. N. (1992). Measuring the economic value of water quality. *The Annals of Regional Science*, 26(2):171–176.

Stuart, S. (2004). 2004 CEO study: A statistical snapshot of leading CEOs. http://bus.wisc.edu/~/media/bus/mba/why%20wisconsin/statistical_snapshot_of_leading_ceos_relb3.ashx.

Stumborg, B. E., Baerenklau, K. A., and Bishop, R. C. (2001). Nonpoint source pollution and present values: a contingent valuation study of lake mendota. *Review of Agricultural Economics*, 23(1):120–132.

Tatian, P. A. (2008). *State of Washington, DC's Neighborhoods*. Urban Institute.

Tesfatsion, L. Agent-based computational economics. http://www2.econ.iastate.edu/tesfatsi/ace.htm. Accessed: 2013-11-12.

Tesfatsion, L. (2002). Agent-based computational economics: Growing economies from the bottom up. *Artificial life*, 8(1):55–82.

Tesfatsion, L. and Judd, K. L. (2006). *Handbook of Computational Economics: Agent-Based Computational Economics*, volume 2. Access Online via Elsevier.

Thurner, S., Farmer, J. D., and Geanakoplos, J. (2012). Leverage causes fat tails and clustered volatility. *Quantitative Finance*, 12(5):695–707.

Torrens, P. M. (2001). New tools for simulating housing choices. Technical report, Berkeley Program on Housing and Urban Policy.

Van Houtven, G., Powers, J., and Pattanayak, S. K. (2007). Valuing water quality improvements in the united states using meta-analysis: Is the glass half-full or half-empty for national policy analysis? *Resource and Energy Economics*, 29(3):206–228.

Wey, K. (1990). A social welfare analysis of congestion and water quality of great salt pond. *Block Island, Rhode Island*.

Whitehead, J. C. (2005). Combining willingness to pay and behavior data with limited information. *Resource and Energy Economics*, 27(2):143–155.

Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics*, 56(2):645–646.

Young, C. E. (1984). Perceived water quality and the value of seasonal homes1. *JAWRA Journal of the American Water Resources Association*, 20(2):163–166.

Zelinka, I., Celikovskỳ, S., Richter, H., and Chen, G. (2010). *Evolutionary algorithms and chaotic systems*, volume 267. Springer.