THE WASHINGTON, D.C. HOUSING AFFORDABILITY SIMULATOR

by

Shawn Bucholtz
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Computational Social Science

Committee:

________________________	Dr. Robert Axtell, Dissertation Director
________________________	Dr. Andrew Crooks, Committee Member
________________________	Dr. Katrin Anacker, Committee Member
________________________	Dr. William G. Kennedy, Committee Member
________________________	Dr. Jason Kinser, Acting Chair, Department of Computational and Data Sciences
________________________	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
________________________	Dr. Peggy Agouris, Dean, College of Science
Date: ________________	Fall Semester 2017 George Mason University Fairfax, VA

The Washington, D.C. Housing Affordability Simulator

A dissertation submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy at George Mason University

by

Shawn Bucholtz
Master of Science
University of Maryland, 2003
Bachelor of Science
Michigan State University, 2000

Director: Robert Axtell, Professor
Department of Computational and Data Science

Fall Semester 2017
George Mason University
Fairfax, VA

ACKNOWLEDGMENTS

First, I would like to thank my committee and the faculty in the Computational and Social Science program for their support throughout my years at GMU, especially over the past two years. I would like to thank Dawn Parker for her initial encouragement to pursue my doctorate at GMU. I would like to thank my colleagues at HUD for their encouragement, as well as their patience during my frequent absences. I would like to thank Peter Tatian at the Urban Institute for his generous assistance in acquiring and using the D.C. Preservation Catalog. Finally, and most importantly, I would like to thank my family and friends for their encouragement throughout the years as I pursued this degree. Their support kept my focused on my goal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ACRONYMS

2010 U.S. Synthetic Population	SynPop
Agent-based modeling	ABM
American Community Survey	ACS
American Housing Survey	AHS
Area median income	AMI
Building Permits Survey	BPS
Cellular automata	CA
Computable general equilibrium	CGE
Computer assisted mass appraisal	CAMA
Floor area ratio	FAR
Housing Choice Voucher Program	HCVP
Housing Production Trust Fund	BPS
Internal use file	IUF
Low Income Housing Tax Credit	LIHTC
Multinomial logit	MNL
Planned unit development	PUD
Public use microsample	PUMS
Washington, D.C. Housing Affordability Simulator	DCHAS
Washington, D.C. Office of the Chief Technology Officer	DCOCTO
Washington, D.C. Office of Tax Revenue	DCOTR
Washington, D.C. Preservation Catalog	DCPC
United States Department of Housing and Urban Development	HUD

ABSTRACT

THE WASHINGTON, D.C. HOUSING AFFORDABILITY SIMULATOR

Shawn J. Bucholtz, Ph. D.

George Mason University, 2017

Dissertation Director: Dr. Robert Axtell

This dissertation presents the Washington, D.C. Housing Affordability Simulator, or DCHAS. DCHAS is an empirical agent-based model of urban housing supply and demand, with a special emphasis on housing affordability and affordable housing production. DCHAS agents include households, landlords, developers and the local government. Past agent-based and microsimulation modeling efforts have demonstrated the importance of including agent heterogeneity and land markets in models of urban housing supply and demand. DCHAS builds upon this foundation and extends prior efforts by including six additional features important to on housing affordability and affordable housing production: agent variation appropriate to low-income households, explicit representation of Federal housing subsidies, explicit representation of affordable housing supply, rent control, zoning and regeneration of properties, and filtering and rehabilitation of housing units. DCHAS is calibrated to the population and housing stock as it existed in 2010. The behaviors of DCHAS's agent are parameterized with data from

2011 to 2015. Combining a 2010-based population and housing stock with agent behavior parameterized with data from 2011 to 2015, it is demonstrated that DCHAS reliably reproduces housing supply and demand outcomes observed in 2015. Then, DCHAS is used to simulate three housing supply and demand scenarios over the next ten years (2016 -2025). The principle contributions of this dissertation are to: (1) identify and explore six concepts critical to housing affordability in an urban environment; (2) demonstrate how to empirically represent these concepts through the use of administrative data sources, and (3) demonstrate how to build an empirically-based ABM that can be used to simulate housing affordability under different market conditions or housing policy scenarios.

CHAPTER 1: INTRODUCTION

Much like food and water, housing is necessary for survival. Housing is a good that can be produced and consumed at various locations, and various levels of quality. Consequently, there is variation in housing prices. Economic theory suggests that in a competitive market, housing supply and housing demand should be in long-run equilibrium, meaning sufficient housing will be supplied in varied locations and at different price-points to meet the range of household demand. However, this simple treatment of housing as if it were any other durable good masks an array of complexities within the housing market.

In many parts of the country, both urban and rural, there are housing affordability issues, especially for renters (Edmiston, 2016). In some cases, households pay more than 30 percent of their income for housing – an amount policy makers and housing advocates have historically deemed acceptable (Schwartz & Wilson, 2008). In other cases, households are living in substandard or overcrowded housing. Some households are living in areas with little or no economic opportunities, while others are living far away from employment centers, simply trading lower housing costs for higher commuting costs and/or longer commuting times. Taken together, these households represent a "shadow" demand for housing that is affordable and located in areas of higher opportunity, such as in Boston, Massachusetts or Washington, D.C. (Opportunity Index, 2016).

A natural response to an existing demand is to increase supply, whether it be the supply of market-rate housing or the supply of affordable housing, which is housing built specifically for households with incomes below a certain threshold, typically with non-market financial intervention. In some housing markets, especially those constrained by geography or regulations, the supply of *market-rate* housing has not kept pace with the demand, leading to increased prices and reduced affordability for many households. The effects of this type of market can be seen in cities such as San Francisco, CA; San Jose, CA; and Honolulu, HI (Lu, Stilwell, & Cannon, 2016). To be clear, the lack of supply of market-rate housing due to geographic constraints does not constitute a market failure. However, the lack of supply due to regulation (e.g., zoning, building codes) *could* be considered a market failure requiring intervention by whichever level of government is responsible for such regulation[1].

In many housing markets, the supply of *affordable* housing has not kept pace with the demand, as evidenced by long wait lists for placement into affordable housing (Hughes, 2014; Navarro, 2013). There are at least three reasons why affordable housing supply has not kept pace with demand. First, given current construction and operating costs, it is often not economically viable for developers to build affordable housing without government intervention in the form of tax credits or financing. Second,

[1] Alternatively, the case can be made that regulation is an expression of the market preferences of consumers who purchase housing or are impacted by housing development, meaning that it is not a market failure.

governments at all levels (Federal, state, and local) have not supplied enough intervention to entice builders to building affordable housing. Third, in certain areas, zoning regulations that place caps on the density of housing unit often prevent the construction of affordable housing at densities that make development financially viable.

Mirroring the rest of the United States, the Washington, D.C. metropolitan region has significant housing affordability issues. However, unlike many other parts of the U.S., the metro D.C. area did not experience a dramatic downturn in the real estate market during the late 2000s. The George Mason University Center for Regional Analysis published a report entitled, "Housing the Region's Future Workforce" (Sturtevant & Fuller, 2011), in which the principle finding was that the Washington, D.C. metropolitan area will need to add more than 730,000 additional housing units between now and 2032 to meet the expected demand. Perhaps more importantly, at least 25 percent of the owner-occupied units will need to be priced at $200,000, and 178,000 renter-occupied units will need to be priced below $1,250 per month in order to house the expected increase in low-income workers. At current land and construction costs, Sturtevant and Fuller (2011) concluded it will be difficult to build these units.

1.1. Why is Regional Housing Affordability a Complex Issue?

Although housing markets are influenced by national economic trends, housing markets operate at the regional and local levels. Characterizing regional housing markets may appear to be a straightforward task from an economist's perspective, as it is relatively simple to use aggregate demographic and economic data to estimate current and expected demand, as well as the number of households who can afford their homes.

Data points on current individual and household income, current rent or mortgage costs, and current utility costs, are collected by no fewer than three major federal surveys at national, metropolitan area, and local levels[2]. In fact, the U.S. Department of Housing and Urban Development's (HUD) American Housing Survey (AHS) collects data on income, rent, and utility costs, repair and remodeling costs, and expenditures on the receipt of housing assistance.

Unfortunately, relatively simple models based on aggregate data rarely capture the inherent complexities of the regional and local housing markets, including complexities that impact housing affordability. On the housing demand side, these complexities may include: regional in- and out-migration, demographic changes, household composition changes, household formation rates, changing preferences for housing types and locations, household expectations about future changes, and the calculation of what is affordable to a household. On the housing supply side, complexities may include: the impact of affordable housing programs such as public housing and project-based section 8; changing availability of financing; local regulatory constraints; neighborhood opposition to housing construction; as well as filtering of existing housing units into the stock of affordable units. Finally, addressing housing affordability issues requires planning at the local and regional levels. Planning, by its very nature, necessitates predicting the future, or at least predicting what could happen under various (and uncertain) market conditions and policy scenarios five to fifteen or

[2] Examples include the American Community Survey, the American Housing Survey, and the Survey of Income and Program Participation.

more years into the future. A model that is useful for planners must account for uncertainty in future predictions.

1.2. Traditional Tools of Regional and Urban Economists

The economy has been described as a dynamic adaptive system (Tesfatsion , 2006). It is comprised of individual consumers who supply labor and purchase goods and services, profit-maximizing firms who produce goods and services, and governments who produce services and set economic policies. These actors exist in a heterogeneous landscape and interact with each other at different spatial (national, regional, local, etc.) and temporal (daily, monthly, annual, etc.) scales. Representations of the economy range in complexity but are generally built on the principle of equilibrium in which market clearing conditions are satisfied (i.e., prices adjust such that excess demands are zero in all factor and good markets).

There are a handful of approaches to building regional or urban economic models that can be used to analyze housing supply and demand under various market conditions or policy scenarios.[3] A primary distinction in modeling approaches is whether the models are partial or general equilibrium. Partial equilibrium models are designed to study equilibrium in one sector of the economy, while general equilibrium models are designed to study the equilibrium in all sectors of the economy simultaneously. One class of general equilibrium models, called computable general equilibrium models (CGE), is widely used to study the impact of various housing policy changes (often referred to as supply or demand shocks) on sectors of the economy. For example, O'Connell (2007)

[3] For a historical perspective on how urban modeling has changed, see Batty (2008, 2012).

used a CGE model to study the impact of property tax policy changes on various sectors of Florida's economy, including the owner-occupied and rental housing markets. Euijune (2003) employed a CGE model to assess the economic effects of housing supply shocks on urban growth and income distribution in Seoul, Korea.

Despite their wide use, CGE models have shortcomings. Peters and Brassel (2000) offered a general discussion of the weaknesses of CGE models, which included issues regarding the lack of room for unambiguous improvement; assumptions of perfectly rational actors who possess perfect information; the lack of markets for environmental quality or social cohesion; convex production technologies that rule out increasing returns to scale; and a focus on stable equilibrium that will not shift over time.

Partial equilibrium models come in several types. Irwin and Wren (2014) provide an overview of three types of partial equilibrium models commonly used in the land use literature, which is closely related to the housing supply literature: structural models, reduced form models, and spatial simulation models. In structural models, the basic assumption is that housing demand and supply are in equilibrium such that the parameters of the supply and demand equations can be recovered. One advantage to structural models is that they can capture general equilibrium feedback of non-marginal changes. Reduced form models, while very common in the housing demand literature (e.g., hedonic models of housing prices), are also present in land use and housing supply literature. Irwin and Wren (2014) discuss several examples of reduced form models, noting that the leading drawback of reduced form models is that they cannot capture general equilibrium feedback from non-marginal changes.

Spatial simulation models, which include cellular automata, spatial equilibrium models, microsimulation models, and agent-based models (ABMs), may be used as stand-alone models or can be coupled with any of the aforementioned general or partial equilibrium models.

Cellular automata (CA) models have been widely and successfully used to study urban growth, although they often lack market interactions between buyers and sellers. Examples of CA modeling have been discussed in White and Engelen (1993), Wu (2002), Yang and Lo (2003), and Chaudhuri and Clarke (2008). Detailed reviews of CA applications can be found in Benenson and Torrens (2004) and Batty (2005). Birkin and Wu (2012) provide a review of microsimulation and hybrid ABM approaches, which combine microsimulation and ABM.

Several modeling frameworks have been combined to inform the various complexities in the housing market. Plantinga and Lewis (2014) provide an example of combining results from a reduced form econometric model of land use change with a spatial simulation framework to simulate future changes in a landscape. Savard (2014) presents an example of combining a CGE model with a micro-simulation model to investigate the impact of trade and fiscal policies on poverty and income distribution. Ahmed and O'Donoghue (2007) provide a literature overview of previous efforts to combine CGE and microsimulation models. Lastly, some spatial simulation models are built on the principle of spatial equilibrium, where price differences reflect locational differences, and firms and people are (re)located such that they are spatially inconsequential within the landscape (Irwin, 2010).

1.3. How Can ABMs Contribute to Urban Economic Analysis?

Urban economic models that include urban land modeling are derived from the monocentric city model developed by William Alonso (1964). Alonso's bid-rent model posits that households choose locations at a certain distance from a central business district so as to maximize the utility they get from the joint consumption of location (the place where their home is located) and a composite good representing all other goods, subject to budget constraints defined as income minus transportation costs. Applying a few assumptions to this model (supply equals demand at equilibrium and equal utility for all agents), the equilibrium land rent can be derived for any location, resulting in a set of rent gradients.

It is often the case that urban economic models include assumptions necessary to solve the model. Examples of such assumptions include homogenous agents, complete information and perfect foresight, no interactions among agents, and instantaneous equilibrium. Each of these assumptions comes with drawbacks, which have been frequently discussed in the urban economics literature. For instance, the assumption of homogenous agents may ignore important behavioral variation, leaning to model outcomes that do not align with empirical outcomes. Examples of discussions of the drawbacks include Epstein and Axtell (1996), Axtell (2005), Tesfatsion and Judd (2006), and Parker and Filatova (2008).

Simulation models, of which ABM are one type, can complement traditional models of housing demand and supply. Simulation has been described as a "third way of doing science," augmenting deductive and inductive reasoning as discovery methods (Axelrod, 1997; Macal & North, 2009). The main advantage of the ABM approach is the ability to deal with many of the complexities inherent in the regional or urban economy, including agent heterogeneity, spatial heterogeneity, spatial dynamics, lack of equilibrium, lack of rationality, and impact of market interactions among agents. Moreover, an ABM can account for uncertainties in the future by introducing stochasticity where appropriate.

As discussed in Guerrero and Axtel (2011), the ABM approach requires agentization of the actors within the system of study. They provide a basic systematic method for agentizing neoclassical economic models. Smajgl, Brown, Valbuena, and Huigen (2011) identify three necessary components for an ABM: agents, the environment in which they exist, and a network for agent interaction. Roundsevell, Robinson, and Murray-Rust (2012) provide additional perspectives on the process of creating agents, including how to scale ABMs in various ways.

Huang, Parker, Filatova, and Sun (2014) provide a review of urban land modeling that summarizes 51 examples of ABMs, concluding that agent heterogeneity and interactions among buyers and sellers (i.e., a land market) are important components of urban land models. Irwin (2010) suggests that general equilibrium is a desired factor in an ABM. Heppenstall, Malleson, and Crooks (2016) discuss the importance of calibration and validation in urban models, with a focus in the era of "big data."

1.4. Statement of the Problem

To understand the short and long-term implications of the housing affordability crisis, planners and policymakers at multiple levels of government must have tools to estimate affordability and to estimate the supply and demand for affordable housing under various market conditions and policy scenarios.

The traditional tools of regional and urban economists, including general equilibrium models, can be useful for estimating housing supply and/or demand under various market conditions or policy scenarios. However, estimating housing supply or demand over time and throughout space can be complex. Moreover, estimating supply and demand under conditions that do not necessarily reflect perfect markets in equilibrium - a characteristic of affordable housing supply - can further increase complexity. The presence of complexity opens the door for complementary or alternative approaches, and ABMs are one such approach.

1.5. Purpose of the Study

This study has three primary purposes: (1) to build an empirical ABM of housing supply and demand in an urban environment (Washington, D.C.), including agent heterogeneity and endogenous land markets, with a *specific emphasis* on housing affordability and the supply of and demand for affordable housing; (2) to use the ABM to simulate affordability and affordable housing supply and demand under current market conditions; and (3) to use the ABM to explore different policies aimed at improving housing affordability.

1.6. Theoretical Framework and Method

The ABM presented in this research is motivated by the neoclassical theory of individual utility maximization for housing consumers and firm profit maximization for landlords and market-rate housing developers. In the ABM, household members and households progress through a life cycle, which includes aging, income change, marriage, divorce, birth, and death. The current life cycle stage of household members impacts the household's demand for housing. However, the life cycle progression of household members is controlled by exogenous life cycle parameters that probabilistically trigger life cycle events. The parameters are based on empirical data such that the probabilistic life cycle events occur at a frequency consistent with past data. Similarly, a household member or household's decision to relocate, to form a new household and choice of tenure are exogenous parameters based on empirical data.

In the ABM, a household's choice of housing unit is motived by the model design in UrbanSim (Waddell, 2000, 2002, 2011), where individual consumers select from numerous housing choices by maximizing utility, subject to an empirically-supported housing budget constraint (share of income spent on housing). Households who own their housing units, but wish to sell, endogenously determine an ask price based on local market conditions, then enter limited negotiations with prospective buyers to reach a final selling price.

In the ABM, landlords have three roles. First, they are suppliers of multifamily housing and they engage in profit maximization, whereby they offer units for rent in a rental market where prices are endogenously determined based on local market conditions, including local vacancy rates. Second, landlords invest in rehabilitating their units, where the rate of rehabilitation is an exogenous parameter adjustable by the modeler. Third, landlords own land suitable for multifamily housing (vacant or occupied), and receive offers to purchase their land from market rate housing developers.

Offer prices for land are endogenously determined based on local market conditions, and landlords accept offers from housing developers.

Market rate housing developers are firms engaging in profit maximization, whereby they seek to develop housing in locations where returns to development are the highest. The developers form endogenous expectations regarding revenue, purchase land, and construct units. Following Magliocca et al. (2011), market rate developers are represented by a single representative market rate developer, called the Developer. The Developer can propose major zoning changes, permitting additional construction of units beyond what is allowed "by right" in the zoning code.

In the ABM, the local government serves two functions: to develop affordable housing, and to approve zoning changes. No attempt is made in this research to precisely model the *process* by which affordable housing construction or redevelopment occurs. Rather, the rate at which new affordable housing is constructed is a parameter in the model that is adjusted by the modeler, resulting in outcomes of interest to the modeler. Moreover, no attempt is made in this research to model the complex process of zoning approvals. Rather, the rate at which the local government approves major zoning changes is a parameter in the model that is adjusted by the modeler. Modeling the process of affordable housing development or the process of zoning approvals could be addressed in future work.

In the ABM, credit availability and interest rates are not included in the model. Credit is assumed to be available for all buyers of homes and all developers of housing.

The only macroeconomic factor shifting the demand for housing is net migration. In the model, net migration is an exogenous parameter that can be adjusted by the modeler.

The ABM's environment is the Washington, D.C. housing and land markets. The environment includes single-family houses, condominiums, and multifamily rental properties. The housing is sited on land, or in some cases, land does not have housing. The land is subject to zoning constraints which limits the amount of allowable housing.

1.7. Research Questions

This study presents an ABM of the city of Washington, D.C. housing market with specific emphasis on housing affordability and the supply of and demand for affordable housing. This study has four primary research questions:

(1) What are the critical concepts to include in an ABM of housing supply and demand in an urban environment that is characterized by agent heterogeneity and endogenous land markets, and which includes a specific emphasis on housing affordability and affordable housing?

(2) Can each of the important concepts be represented empirically, and if so, what are the best data sources?

(3) Does the ABM produce city-level estimates of housing affordability that are in global agreement with known indicators?

(4) What happens to city-level affordability under different housing policy scenarios?

1.8. Delimitations

There are at least four important delimitations with the ABM presented in this study. First, and most importantly, the ABM presented reflects the housing market in

Washington, D.C., which is functionally a submarket within the greater metropolitan area. In fact, nearby areas such as Silver Spring, MD; Arlington, VA; and Alexandria, VA, contain important segments of the "close in" urban portion of the greater Washington, D.C. metropolitan area housing market. The ABM presented in this study is not a complete representation of the entire metropolitan area housing market or the close in urban portion of the metropolitan area housing market. The principle reason for not modeling the entire metropolitan housing market or the close-in urban portion of the metropolitan housing market was the availability of data on affordable housing locations. If data becomes available, future work should include these areas.

Second, the homeownership rates for different cohorts of the population are exogenous to the model. There are no mechanisms in the model that would permit an endogenous increase in the homeownership rate for a cohort of the population. In fact, the overall homeownership rate can only increase and decrease based on the changes in the proportions of the cohorts of the underlying population. In other words, the homeownership rate for 35-65-year-old households does not change. However, if these households become a larger part of the population, they can result in the overall increase in homeownership rate.

Third, endogenous out-migration due to housing affordability is not reflected in the model. There are undoubtedly households who leave Washington, D.C. for affordability reasons. However, empirical data on the reasons households leave the region are not readily available.

Fourth, the land market is not fully endogenous. While the residential location selection and the development location are driven by past and current prices, the amount of new supply is controlled by the modeler. This decision was made so as to allow the modeler to investigate the impacts on affordability from different supply growth scenarios.

1.9. Organization of the Dissertation

Chapter 2 presents a literature review and begins with a general assessment of the factors affecting housing demand and supply, moving then to discuss six concepts related to housing affordability and affordable housing in Washington, D.C. The final two sections of Chapter 2 review various ABMs of housing demand and/or supply, including ABMs that reflect or closely reflect the six concepts related to housing affordability and affordable housing. Chapter 3 begins with a general discussion on creating an ABM, and then moves to present the methodology for the relevant ABM, named the D.C. Housing Affordability Simulator (DCHAS). Subsequent sections address the DCHAS agents (households, landlords, market rate developers, and affordable housing developers) and the DCHAS environment (properties and housing units). Chapter 4 presents the data sources used in DCHAS. Chapter 4 is organized similar to Chapter 3, discussing data sources used to create and parameterize the attributes and behavior of households, landlords, and developers, as well as the attributes of properties and housing units. Chapter 5 presents the initialization, verification, and validation of the DCHAS model. Chapter 6 presents and discusses the results of three alternative scenarios. Chapter 7 presents the conclusion, research contributions and areas of future work.

CHAPTER 2: LITERATURE REVIEW

This chapter presents a review of the relevant literature on housing demand, housing supply, and the use of ABMs for simulating housing market processes. Chapter 2 begins by discussing housing demand. In section 2.1, the individual (household-level) determinants of demand are reviewed. Section 2.2 discusses regional- and national-level determinants of demand. Section 2.3 discusses the use of "representative agents" in modeling housing demand, with an emphasis on affordability. In section 2.4 the determinants of supply of housing are reviewed, with an emphasis on supply of affordable housing. Section 2.5 discusses six additional issues affecting housing affordability and affordable housing in Washington, D.C. Section 2.6 provides a general literature review of existing ABMs of demand for and supply of housing, focusing on the features of ABMs that are the literature identifies as most important, including heterogenous agents, land markets and empirical representations. Finally, Section 2.7 provides a review of six agent-based models with features that address, or attempt to address, the six issues discussed in section 2.5.

2.1. Individual (Household) Determinants of Demand for Housing

When describing the determinants of the demand for housing, it is useful to begin with a household's demand determinants and then describe how household demands aggregate to a region-wide demand (Attanasio, Bottazzi, Low, Nesheim, & Wakefield, 2012). In the most general sense, a household's demand for housing is a function of the household's wealth and income, composition, preferences, access to credit, level of risk aversion, and expectations about future house prices. Household composition generally includes the size of the household, the age of the household members, the relationship of the household members to each other, and other characteristics such as sex or educational attainment. A house, unlike other durable goods, generally reflects a basket of services provided by the structure and the location on which the structure sits (Rosen, 1974). As such, a household's preferences for housing may reflect structural preferences as well as neighborhood preferences (Tiebout, 1956). Finally, the latter three elements of a household's demand for housing, access to credit, the level of risk aversion and expectations about future housing prices, influence a household's decision of whether to purchase or rent their housing (Attanasio et al. 2012).

In neoclassical economic theory, formalizations of housing demand begin with a utility function and/or demand function derived from the utility function. However, the specific formalization used varies. For instance, Kuismanen, Laakso, and Loikkanen (1999) present a simple demand function based only on household age. Paciorek (2013) presents a demand function that includes composition (age, sex, household relationships, educational attainment) and access to credit. Attanasio et al. (2012) present a demand function based on income, wealth, age, education, preferences for housing of varying sizes, and preferences for ownership. Fisher, Pollakowski, and Zabel (2009) present a utility function that includes both structural and locational characteristics, with the purpose of specifying a locally-varying housing price index equation.

2.2. Regional and National Determinants of Demand for Housing

The prior section included several household-level determinants of demand for housing. Aggregate demand is simply the sum of household demand within an area of interest. As suggested in Chapter 1, it is useful to explore housing demand and supply from a regional perspective, so aggregate demand is referred to as regional demand.

In addition to the household-level determinants of demand for housing, there are a handful of other determinants that operate at a regional or at a national level. Two regional determinants are in-migration and out-migration. These determinants are generally rooted in the responses of individuals to conditions in regional labor markets (Partridge & Rickman, 2007) or to exits from the labor market by retirees (Park & Hewings, 2009).

Two national-level determinants of the demand for housing are credit availability and terms (i.e., interest rate and mortgage insurance premiums) (Anenberg, Hizmo, Kung, & Molloy, 2015). Arguably, households have direct control over their level of wealth and income, as well as their credit score. Wealth, income and credit score influence whether a household can obtain a mortgage as well as the terms of the mortgage. However, households have no control over the minimum qualifying standards for receiving credit, which are driven in large part by federal regulations, such as the Ability-to-Repay and Qualified Mortgage rule (Consumer Financal Protection Bureau, 2016). Moreover, households have no control over the Federal Funds Rate, which, along with other macroeconomic factors, influences mortgage interest rates.

2.3. Estimating Regional Demand for Housing Using Representative Agents

Regional planners and other researchers require estimates of regional housing demand for various purposes. However, at least two factors make this task difficult. First, in practice, housing demand functions are not available for every (or perhaps any) household, so aggregating household-level housing demand functions is not feasible. Second, household variation in wealth, income, composition, preferences, access to credit, risk aversion, and expectations, when aggregated to a region, may translate into within-region variation that is important in housing demand estimation. In other words, the sources of variation should be ignored or “smoothed over” in economic models of housing demand.

Faced with these challenges, regional planners and other researchers often adopt a series of household types and assume that all households within a type behave in the same manner (i.e., have the same housing demand function). In economic models, these household types are referred to as representative agents. For instance, in O’Connell’s (2007) CGE model of the impact of property tax policy in Florida, households are represented as nine different types distinguished only by income. The strategy of creating household types has been employed in the study of regional demand for housing in Washington, D.C. In their report, “Housing the Region’s Future Workforce,” Sturtevant and Fuller (2011) estimate future demand for housing by creating household types based on county location, age, household size, household composition, and income.

The strategy of creating household types has been employed in the study of demand for *affordable* housing. For instance, the Woodwell (2015) found that households with affordable housing needs can generally be broken into three broad categories: low- and very-low-income households; moderate-income workforce households; and special needs populations. Very-low-income households are defined as those earning less than 50 percent of the area median income (AMI); low-income households are defined as those earning between 50 and 80 percent of the area median income; moderate-income households are defined as earning between 80 and 115 percent of the AMI; special needs populations include, but are not limited to, the elderly, the disabled, veterans, and people with AIDS.

Another example of creating household types to study the demand for affordable housing is the HUD's "Worst Case Needs" report (2015). In their report, HUD defined a household type, *worst case needs*, as those with very low-incomes who do not receive government housing assistance, and who either pay more than one-half of their household income for rent, or who live in severely inadequate conditions, or both. Other notable housing reports, such as the "State of the Nation's Housing" (Joint Center for Housing Studies of Harvard University, 2014) follow a similar classification of household types.

2.4. Determinants of New Housing Supply

Whereas it is individuals who demand housing (either through purchase or rent), new housing is typically supplied by firms, often referred to as developers. Developers can bring agricultural land into residential use, convert commercial or industrial land into residential use, or increase the density of existing residential use through demolition and rebuilding. In the most general sense, an individual developer's determinants of housing supply are current housing prices, cost of land, cost of building materials, and the availability and cost of credit. As discussed in Hedberg and Krainer (2012) and Mayer and Somerville (1996), a developer's supply decision has been viewed in two ways in the housing supply literature. First, as a traditional variant of the classic firm investment problem where new supply depends on expected revenues, land costs, building costs, and credit costs. The second way is as a real option, where uncertainty in housing prices, land prices, and other input prices results in developers maximizing per-period profits by choosing the optimal time and type of construction (Murphy, 2015).

One factor that influences the aggregate supply of housing in a locality or region is the vacancy rate. Due to normal frictions in the housing market, there is a "built-in" vacancy rate. In fact, the long-term average rental vacancy rate in the U.S. is 7.3 percent (United States Bureau of the Census, 2015). All else being equal, fluctuations from the long-term average vacancy rate should be reflected in the price of existing housing - when the vacancy rate increase (or decrease), housing prices should increase (or decrease), often with a lag. However, from the perspective of an individual developer, big changes in vacancy rates may represent shocks to the housing market that not only impact current housing prices, but may alter *expectations* about future prices. One such supply shock is foreclosures, which increases not only the supply of existing homes, but also the future expected supply of existing homes. In fact, Hedberg and Krainer (2012) found that foreclosure rates were highly important in predicting a drop in the supply of new homes in the post-2006 housing bust. Moreover, they found evidence that foreclosures may have contributed to further uncertainty in markets, leading to additional supply decreases beyond what was predicted by inventory changes alone.

Two other factors that influence the aggregate supply of new housing in a locality or region are physical and regulatory constraints. As discussed in Glaeser, Gyourko, and Saks (2005) and Saiz (2010), physical constraints such as water and elevation changes reduce the amount of land available for housing, while regulatory constraints such as zoning influence the density of housing. Moreover, regulatory constraints may reduce the elasticity of supply by slowing down the housing development process or driving up development costs.

2.5. Additional Demand and Supply Concepts Specific to Affordability and Affordable Housing

The previous four sections described, in general terms, the determinants of demand and supply of new housing. As discussed in section 1.7, the first research question is:

> What are the critical concepts to include in an ABM of housing supply and demand in an urban environment that is characterized by agent heterogeneity and endogenous land markets, and which includes a specific emphasis on housing affordability and affordable housing?

There are at least six concepts that should be considered when building an ABM with emphasis on housing affordability and affordable housing demand and supply in Washington, DC. These six concepts are discussed in the next section sections.

2.5.1. Affordability and Housing Choice Variation for Low-Income Households

It is important to recognize how affordability, as traditionally measured, may not be appropriate to high-cost regions like Washington, D.C. or to low-income households. Housing is considered affordable to a household if the household spends no more than 30 percent of their income on housing costs – generally referred to as the 30 percent standard. This is the standard adopted by the Federal government when formulating housing subsidy policy. A household is said to face a moderate burden when it pays 30 to 50 percent of its income for housing. A severe burden is defined as paying more than 50 percent of income for housing costs.[4]

While the 30 percent standard may be useful in terms of its simplicity, it has been criticized for at least a few reasons. First, the measure fails to consider transportation costs, which are often the second biggest household budget item (Hass, Makarewicz, Benedict, Sanchez, & Dawkins, 2006). Second, the measure is not necessarily useful if it includes high-income households who choose to spend more than 30 percent of its income on housing, but have more than sufficient resources for food and clothing, with income remaining for savings and leisure. Similarly, the measure is not necessarily useful if it fails to consider a low-income household spending 29 percent of their income on housing, but with insufficient income left for food, clothing and other necessities. Lastly, some researchers have concluded that a poor renter household's demand for housing may be inelastic to income or price, suggesting that some poor renters are forced to spend large fractions of their income on housing out of necessity (Albouy, Ehrlich, & Liu, 2016).

Housing researchers have developed an alternative measure to the 30 percent standard, called the residual income approach (Stone, Burke, & Ralston, 2011; DiPasquale & Murray, 2017). In short, the residual income approach calculates whether a

[4] See Schwartz and Wilson (2008) for a thorough review of the history of this standard.

household has enough funds remaining to afford housing after considering expenditures necessary for living, including food, clothing, utilities, and transportation. The residual income approach does not suffer from the same issues associated with the 30 percent standard. However, in practice, it can be difficult to use the residual income approach to measure housing affordability due to the lack of household expenditure data (Stone, Burke, & Ralston, 2011).

Notwithstanding the issues cited above, an empirical ABM reflecting housing affordability must have agent variation appropriate to low-income populations, and the variation must be supported by empirical data. Given the existing data sources, including the American Community Survey (ACS) and the AHS, the one estimate that is widely available is the share of income spent on housing for various segments of the population, including low-income households.

2.5.2. Federal Housing Subsidies

The scope of federal housing subsidies in Washington, D.C is large. According to the Urban Institute (2017b), there are approximately 8,000 units of public housing, 10,500 households receiving Housing Choice Vouchers, and as many as 10,000 households are in privately-owned assisted housing (United States Department of Housing and Urban Development, 2017). Taken together, these units represent almost 10 percent of the total housing stock in Washington, D.C. and they house nearly 11 percent of all households, including 40 percent of all low-income households. Although the details of each subsidy program vary, the subsidies can generally be thought of as an income subsidy. However, federal housing subsidies are not an entitlement. Funding levels for housing subsidies are subject to budget fluctuations. Given the relative contribution of these housing units to the housing stock and their importance in housing low-income households, an ABM of housing affordability and affordable housing demand must consider the presence of federal housing subsidies.

2.5.3. Sources of Affordable Housing Supply

There are additional sources of affordable housing in Washington, D.C. and the stock of these units may change over time. Housing programs such as the Low-income Housing Tax Credit (LIHTC) program and Washington D.C.'s Housing Production Trust Fund (HPTF) play an important role in creating new affordable housing and preserving existing affordable housing. There are approximately 18,000 privately subsidized LIHTC housing units in Washington, D.C. (Urban Institute, 2017a). The HPTF was created in 1998 and has helped build or preserve nearly 10,000 affordable housing units (D.C. Fiscal Policy Institute, 2016). The current and future HPTF funding levels, projected to be $100 million per year, may have an impact on the new supply of affordable housing options or the preservation of the existing stock of affordable housing.

While programs such as LIHTC and HPTF create or preserve affordable housing, program requirements are such that many LIHTC properties will expire over the next decade, meaning that the developments can transition the properties too market-rate housing (Abt Associates, 2012). Depending on market conditions, existing affordable units may be lost. An ABM of housing affordability and affordable housing supply must consider the presence of these programs and the potential expiration of housing developments over time.

2.5.4. Rent Control in Washington, D.C.

Rent control impacts the affordability of the rental housing stock, and it applies to a large number of rental units in Washington, D.C. As of 2011, the Urban Institute estimated that there were nearly 80,000 rent-controlled units in Washington, D.C., representing approximately two-thirds of the entire rental housing stock (Tatian & Williams, 2011). From a housing demand perspective, rent control may preserve the affordability of units in high-cost areas. From a housing supply perspective, rent control effectively limits the amount of rent growth allowable, meaning that returns to owning rental property are lower than they might otherwise be, thereby reducing the incentive of landlords to make capital investments in their property. Lastly, among other exceptions to the policy, it is essential to note that rent control does not apply to units permitted after 1975.

2.5.5. Zoning and Regeneration of Existing Properties

It is important to recognize the roles of zoning and Washington, D.C.'s Comprehensive Plan in projecting future housing supply, including the demolition and regeneration of existing properties. The Comprehensive Plan is meant to guide the use, density, and design of buildings within Washington, D.C. (Washington, D.C. Office of Planning, 2017). Functionally, zoning is a regulatory constraint on the supply of housing. When a homeowner or developer seeks to build new housing, there are three options. First, if the proposed project is allowable within the current zoning code, the development is permitted "by right." Second, if minor changes (referred to as variances) are required, the homeowner or developer can seek relief from the Board of Zoning.

The third option for a homeowner or developer seeking to build housing is to apply for a Planned Unit Development (PUD). This option is necessary when the homeowner or developer requires a significant amount of relief from the zoning regulations, such as increasing the density of housing permitted on the lot. In short, a PUD requires the homeowner or developer to provide community benefits in exchange for the relief, such as neighborhood improvements or donations to neighborhood entities. The likelihood of a PUD approval is influenced by whether the PUD application is consistent with the Comprehensive Plan. An ABM of housing affordability and affordable housing supply must consider the new housing supply constraints due to zoning and the option to apply for a PUD.

2.5.6. Filtering and Rehabilitation

Whereas the production of new affordable housing increases the overall availability of housing that is affordable to low-income people, the existing privately-owned housing stock is still the dominant source of housing for low-income households. At a regional scale, the housing stock affordable for low-income households may be scattered throughout the region, or may follow a pattern, such as clusters and/or concentric circle patterns of the Burgess model (Burgess, 1924). At the housing unit scale, the privately-owned stock typically is not originally built to be affordable for low-income households. Rather, the housing unit becomes affordable over time. This could occur if the neighborhood becomes affordable or the housing unit becomes affordable, or both. As discussed in Rosenthal (2014), filtering is when homes built for higher income households deteriorate and become available for lower income households. Rosenthal

(2014) finds that the nation's housing stock filters down at a rate of 1.9 percent per year in real terms. This means that a 1-year old home is typically occupied by a household with an income 1.9 percent below a household who occupied the home when it was brand new. Rosenthal (2014) confirms filtering to be a viable source of long-run, market-based lower-income housing. An ABM of housing affordability must consider the filtering down of housing such that it becomes affordable for low-income households. Conversely, an ABM of housing affordability must also consider the rehabilitation of existing market rate housing units.

2.6. General Review of Agent-based Models of Housing Supply and Demand

The application of spatial ABMs to the study of urban housing markets, including urban residential choice and urban land markets, is mature to the point that a comprehensive literature review has been completed by Huang et al. (2014). They reviewed 51 spatial ABMs that fell into three general categories: (i) urban land-use models based on the classic theories of Schelling's segregation model (Schelling, 1969; Schelling, 1971) and its variations or the Von Thünen-Alonso model (Alonso, 1964) and its variations; (ii) urban land-use models reflecting different stages of the urbanization process, including gentrification, urban shrinkage, urban regeneration, and urban sprawl; and (iii) integrated agent-based and microsimulation models. The model in the third category are most closely aligned with this research effort.

Within the models in the third category, Huang et al. (2014) focused on three fundamental model features: agent heterogeneity, representation of land market processes, and methods for measuring model outputs. With respect to land market processes, they describe each model in terms of resource constraints (i.e., budget constraints), presence of competitive bidding, and endogenous relocation. The importance of modeling land market processes and its impact on land-use change, both spatially and quantitatively, is discussed in Parker, Brown, Filatova, Riolo, Robinson, and Sun (2011). Other reviews of land market processes include Ettema (2011), Magliocca, Safirova, McConnell, and Walls (2011), Filatova, Parker, and Van der Veen (2009) and Parker and Filatova (2008).

Huang et al. (2014) also described each of the models in their third category (integrated agent-based and microsimulation models) in terms of the data used to parameterize the model: artificial, semi-empirical, empirical, or both artificial and empirical. However, it is important to note that even within the set of models described as empirical, most of the models are hypothetical models in the sense that they do not operate in full empirical environments (landscapes) with model outcomes measured against observed outcomes in the real environment.

Building on her work with Huang et al. (2014) and others, Filatova (2015) provides an abbreviated review of ABMs used to evaluate urban housing markets, with emphasis on the use (or lack thereof) of empirical data. Filatova suggests that there is a lack of spatial ABMs that demonstrate "the feasibility of combining empirics and theory when designing micro-foundations of agents' behavior in spatial markets..." (page 398). She cites three advantages of connecting empirical data to ABMs: (i) increase trust of model stakeholders; (ii) data is used to filter the parameter set to resemble a realistic case; and (iii) data is used to examine the theoretical consequences of more realistic assumptions about agent behavior and interaction. She further cites several challenges for connecting empirical data to ABMs, including scaling observed behavioral data to large populations and finding the right type of data to match the design of the ABM – a common theme in all ABMs (Heppenstall, Malleson, & Crooks, 2016).

The model most closely related to the model presented in this research is UrbanSim. As discussed in Waddell (2000, 2002, 2011), UrbanSim is best described as a microsimulation modeling framework that can be used to simulate a number of urban economic processes, including residential location choice, firm location choice, real estate development, real estate price changes, and transportation and commuting choices. For residential location choice, UrbanSim uses a multinomial logit (MNL) specification, allowing comparisons of various housing alternatives based on empirical data. UrbanSim uses a pro forma real estate development process whereby developers evaluate the expected returns to real estate investment. UrbanSim also includes endogenous price changes.

2.7. Review of Agent-based Models with Concepts Relevant to Housing Affordability and Affordable Housing

The literature and literature reviews cited in section 2.6 provide a foundation on which to build an ABM of housing supply and demand in an urban environment. The foundation includes agent heterogeneity, a functioning land market, and parameterization of agents and the landscape using empirical data. This research is focused on expanding beyond this list by building an ABM with emphasis on housing affordability and affordable housing. As discussed in section 2.5, there are six concepts that should be considered when building an ABM that focuses on housing affordability and affordable housing: agent heterogeneity appropriate for lower income renters; housing subsidies; affordable housing construction; rent control; zoning and urban regeneration; and filtering and rehabilitation. This section reviews six ABMs that have one or more of the six aforementioned concepts that are relevant to modeling housing affordability and affordable housing development. The focus of this section of the literature review is to discuss how the features are implemented, rather than model results. Table 2.1 lists each model and which of the six concepts are represented.

Table 2.1: Six Models with Concepts Important to Housing Affordability and Affordable Housing Construction

Model	Low-income heterogeneity	Housing Subsidy	Affordable Construction	Rent Control	Zoning and Urban Regeneration	Filtering and Rehabilitation
Jackson, Forest, and Sengupta (2008)	Yes				Yes	
Magliocca, McConnell, Walls, and Safirova (2012)					Yes	
Jordan, Birkin, and Evans (2012)		Yes			Yes	
O'Sullivan (2002)					Yes	Yes
Torrens (2007)	Yes					
Bernard (1999)				Yes		Yes

Jackson, Forest, and Sengupta (2008) developed an empirical ABM to simulate demand-side residential dynamics in a gentrifying area of Boston. Their model contained two of the six features (agent heterogeneity appropriate for lower income renters and zoning and urban regeneration) that are relevant to modeling housing affordability and affordable housing development. First, they introduced agent heterogeneity in their ABM via four classes of agents: professionals, college students, non-professionals, and the elderly. Each agent class made their locational decision based on four factors: proximity to desired amenities (commercial center or college campus), affordability, and having at least one neighbor in the same agent class. Moreover, how each of the three factors entered an agent's location decision differed. For instance, professional agents placed more emphasis on proximity to desired amenities than on similarity of neighbors or ability to pay, while non-professional agents did not place emphasis on proximity to desired amenities.

Second, Jackson, Forest, and Sengupta (2008) modeled some aspects of urban regeneration, although zoning did not play a role. Their ABM included a loose representation of a land market in that land rent increases varied across their study area based on the type of agent occupying the parcel - land rent increased faster when professionals occupied the parcel. This resulted in less land available for non-professionals, who in turn moved out of the area. However, since there was no fully functioning land market, there was no increase in the supply of housing via redevelopment of land parcels.

Magliocca, McConnell, Walls, and Safirova (2012) presented an ABM of housing demand and supply on the urban fringe. Their model contained one of the six features (zoning and urban regeneration) that are relevant to modeling housing affordability and affordable housing development. Their model builds upon previous work (Magliocca et al. 2011) and included agent heterogeneity (via agent income and share of income devoted to housing expenditure) and land market representation. Furthermore, their model explicitly incorporated zoning into their landscape. The presence of zoning restrictions in one part of their study area changed the developer's profit function for that area, which in turn changed the timing and pattern of development across other parts of their study area.

Jordan, Birkin, and Evans (2012) presented an ABM of housing demand and supply with a focus on urban regeneration and its effect on public housing tenants. A substantially similar model was presented in Jordan, Birkin, and Evans (2014), which studied the effects of urban regeneration via new mixed-use developments on socioeconomic diversity. Both models contained two of the four features (agent heterogeneity appropriate for lower income renters and subsidized housing) that are relevant to modeling housing affordability and affordable housing development. With respect to agent heterogeneity appropriate for lower income renters, both models classified agents (households) based on socio-economic status, and this classification influenced the type of housing chosen by the household. Additionally, the authors developed a set of movement rules that reflected decisions made by owners and renters. For instance, households moved to areas where the racial and ethnic makeup was tolerable; where the house was of adequate size; where schools were accessible (if the household contained children); where neighborhood quality was better; and where transport routes were accessible. With respect to subsidized housing, their model categorizes housing by the type of home (detached, semi-detached, terrace, flat), size (number of rooms), and tenure. For tenure, a home could be on the private rental market, the public rental market (i.e., public housing), or available for ownership.

In addition to stratification of housing types, O'Sullivan (2002) presented an ABM that illustrates the rent gap hypothesis explanation of gentrification, first posited by Smith (1979). While O'Sullivan's model is relatively simple, he incorporated filtering and rehabilitation, as well as urban regeneration, although not tied to zoning. At each time step of the model, the current physical condition of the property was adjusted by subtracting a depreciation parameter. Properties occupied by renters depreciated faster than properties occupied by owners. Moreover, buyers decided to either occupy a new unit as is, rehabilitate their unit (increasing the physical condition score), or turn it into a rental property. Lastly, the model included a global assessment of the incomes and physical conditions of each property in a neighborhood (referred to as neighborhood status) which influenced the likely incomes of new buyers and tenants, providing a feedback mechanism that reflected aspects of urban regeneration.

Torrens (2007) developed an agent-based model of residential mobility that was later applied to the study of gentrification dynamics in Salt Lake City, UT (Torrens & Nara, 2007). Torrens (2007) incorporated agent heterogeneity appropriate for lower income renters. Households in the model were endowed with economic status, race and ethnicity, and a set of preferences for housing type. As households moved through their life-cycle (young to middle age to elderly), their preferences changed. For instance, young households have a preference to rent apartments, where senior households prefer to own their units.

Bernard (1999) developed a stylized ABM of a rental market to investigate the impacts of rent control on prices and quality. Renters in the market looked to maximize their utility for housing where the utility equation (additive) included location, quality, and price. Renters experienced a search cost which was reflected in the price of the apartment. Landlords responded to apartment vacancy (or lack thereof) by lowering (or raising) rental rates. Landlords posted vacancies on a list and renters seeking housing reviewed the list.

Apartments in the model suggested by Bernard (1999) were assigned a quality level, reflected as a number between 0 and infinity. The quality level of the apartment decayed over time such that the quality decreased by half each year. When a landlord raised rents on an apartment, the landlord also improved the quality level. As such, a rent control policy that placed a cap on the amount a landlord could increase the rent also implicitly placed a cap on the amount that quality could increase.

2.8. Summary

This chapter presented a literature review of the determinants of housing supply and demand, identified six additional demand and supply concepts specific to housing affordability and affordable housing, presented a general review of ABMs of housing supply and demand and concluded with a review of ABMs with one or more of the six aforementioned concepts. The principle conclusion from this chapter is that an ABM of housing supply and demand should include agent heterogeneity, land markets and empirical representations of agents and the landscape. Moreover, an ABM that focuses on housing affordability and affordable housing supply should include the six concepts identified as potentially being important.

CHAPTER 3: MODEL METHODS

Building on the conclusion of chapter 2, this chapter describes methodology used to construct the DCHAS model. Section 3.1 presents a discussion and general description of the model framework for establishing agents and their attributes and behaviors. Section 3.2 describes the attributes of DCHAS household members and households. Section 3.3 describes several household member and household behaviors, which are formalized as events. Section 3.4 discusses landlords and their behaviors. Section 3.5 describes market rate developers and their behaviors. Section 3.6 discusses the Washington, D.C. government and its behavior. Finally, section 3.7 describes the property attributes, while section 3.8 describes the housing unit attributes.

3.1. General Description of the D.C. Housing Affordability Simulator

As mentioned in Chapter 1, the purpose of this dissertation is to develop an empirically-based ABM of urban housing supply and demand, with a special emphasis on housing affordability and affordable housing. The ABM developed here is called DCHAS. As discussed in Smajgl et al. (2011), the ABM approach requires three features: agents, the environment in which they exist, and a network for agent interaction. The rest of the discussion in this section is based closely on the ABM development process discussed in Smajgl et al. (2011) and Rounsevell et al. (2012).

3.1.1. The Agents

To build agents, the first step is to determine agent classes, which are groups of agents with a similar functional role. Chapter 2 reviewed and discussed the existing literature on numerous ABMs focused on the housing market. Examples of agents from prior models include: households (seeking housing to fit their preferences and budgets; selling their homes); real estate agents (matching housing buyers and sellers); landlords (supplying housing); lenders (providing capital for housing construction and housing purchases); farmers (supplying land for housing to be built); land developers (buying land and building housing); and government (establishing zoning, providing subsidies).

In the DCHAS model, four agent classes are defined: households, landlords, developers, and local government. Although buying and selling of homes will be a part of the model, the focus of this ABM is not on the owner-occupied housing market. As such, real estate agents are not explicitly modeled. Farmers are not a part of this model, as Washington, D.C. is not characterized as an "urban fringe" area where land is currently in agricultural production that may transition to housing in the future. Finally, while lenders or other entities in the lending sector are an important part of housing market, they are not a focus of this model. As such, their functions are assumed to exist, and the availability and price of capital is treated as exogenous.

The second step in the process of building agents is to define attributes and behaviors, and the third step is to define typologies of agents. This is the point in which agent heterogeneity is introduced. Attributes can be constant (race and ethnicity, age) or variable (income); behaviors can be similarly expressed (Huang et al, 2014). Typologies may reflect the clustering of heterogeneity within agent attributes and/or agent behavior and are often used to simplify the description and function of the ABM. Like the modeler's choice of agent classes, the choice of agent attributes and behaviors, as well as the typologies based on attributes and behaviors, is often dictated by the research question for which the modeler is attempting to answer. Moreover, the DCHAS model is empirically-based, and the availability of empirical sources to specify attribute values and behavioral parameters contributes to which attributes and behaviors are defined.

3.1.2. The Environment

The urban environment in which the agents exist is the housing and land markets of Washington, D.C. It is important to note that Washington, D.C. is, in most respects, a submarket of the larger Washington, D.C. metropolitan region, which includes parts of Maryland, Virginia, and West Virginia. The agents exist within the housing landscape, which includes properties with housing units and vacant land. Households occupy housing units, and in some cases, own properties (land and housing units). Landlords own properties which may be vacant or may have rental housing units. Developers purchase and develop properties from landlords and construct new housing. The D.C. government determines the rate of construction of affordable housing. Moreover, the D.C. government approves zoning changes for the building of market-rate housing.

3.1.3. The Interaction Network

There are four interaction networks in DCHAS. The first interaction network is between owner-occupied housing sellers and buyers. The second is between renters seeking rental housing and landlords supplying rental housing. The third interaction network is between developers seeking to purchase properties and landlords selling their properties for (re)development. Housing and land market interactions occur every month. As such, the time step in DCHAS is monthly. Finally, household members interact with each other during the marriage event (discussed in section 3.3.2) and the formation portion of the mobility and formation event (discussed in section 3.3.3).

3.2. Household Members and Household Attributes

In DCHAS, households are the decision-making agent. However, households are composed of household members. A household's attributes are the aggregation of the attributes of the household's members. A household's typology is based its attributes.

3.2.1. Household Member Attributes and Typologies

In the DCHAS, households are composed of household members. Each household member has five attributes, as shown Table 3.1. The unique combination of household member attribute values is used to form household member typologies.

Table 3.1: Household Member Attributes and Attribute Values

Attribute	**Values**
Age	1. 0-4; 2. 5-14; 3. 15-24; 4. 25-34; 5. 35-44; 6. 45-54; 7. 55-64 8. 65-74; 9. 75-84; 10. 85+
Sex	Male or female
Race and Ethnicity	White or non-white
Relationship to householder	1. Householder 2. Spouse 3. Child or foster child 4. Other family member 5. Not related to householder
Income	Continuous value

3.2.2. Household Attributes and Typologies

In the DCHAS, the attributes of household members are aggregated to the household level, resulting in six attributes that determine a household's typology: household income, head of household's age, household type, number of people, presence of children, and housing tenure and subsidy type. The household typologies reflect the unique combinations of each possible value of the attributes. This typology strategy is motivated by Waddell (2000), but with the low-income category divided into two categories, thereby aligning the categories more closely with the common groupings used in housing affordability research (as discussed in section 2.3). Moreover, DCHAS includes housing subsidy type – an important consideration discussed in section 2.5.2. The attributes and their values are reflected in Table 3.2. Taken together, there are as many as 2,304 possible household typologies, although not all of these unique groupings actually have households.

Table 3.2: Household Attributes and Attribute Values

Attribute	**Values**
Household income	1. Very or extremely low-income (<= 50% of Area Median Income) 2. Low-income (50% - 80% of Area Median Income) 3. Middle-income (80% - 120% of Area Median Income) 4. High-income (> 120% of Area Median Income)
Head of household age	1. <35 2. 35-65 3. >65
Household type	1. One or more single persons 2. Other family household 3. Married couple
Number of people	1. 1 2. 2 3. 3

	4. 4 5. 5+
Presence of children (17 or younger)	Yes or no
Housing tenure and subsidy	1. Public Housing or Project-based Section 8. 2. Housing Choice Voucher Program (HCVP) 3. Non-subsidized renter 4. Owner

3.3. Household Events

In DCHAS, households exhibit behavior, hereinafter referred to as events. All household events in DCHAS are controlled by exogenous parameters, meaning the parameter values affect the household's outcomes in the model, but the parameter values themselves remain unaffected. The household events, when triggered, result in endogenous (within the model) changes to certain attributes of households as well endogenous changes to the attributes of properties such as housing price or rent.

It is important to note that for the migration, life cycle, mobility, formation and tenure choice events described in sections 3.3.1 through 3.3.4, respectively, different modeling strrategies are used to implement the events. Inherent in each event parameter or set of parameters is a global frequency or rate of the event occurring and an agent-level likelihood the event occurs for that agent, given the agent's attributes. For instance, the global marriage rate may be 2 percent, but the likelihood that agent *A* is selected to get married depends upon the attributes of agent *A*. How this "global frequency/agent-level likelihood" strategy is implemented varies by event, and is described in the sections below.

3.3.1. Household Member In- and Out-Migration Events and Parameters

Section 2.2 discussed in- and out-migration as important determinants of housing demand. Migration of household members into a region increases the demand for housing, while migration of household members out of a region has the opposite effect. Both types of migration are included in DCHAS.

The global frequency of in-migration of household members is represented by the parameter *InMigrationCount* which specifies the number of household members that will migrate in to Washington, D.C. in each time step. This parameter can be altered by the modeler, thereby allowing assessment of different in-migration rate scenarios. The agent-level likelihood that a household member migrates in to Washington, D.C. is based on the typologies of household members that have recently migrated to Washington, D.C.

The out-migration of household members is controlled by two parameters. The global frequency of out-migration is represented by the parameter *OutMigrationCount* which specifies precisely the number of household members chosen to migrate out. The *OutMigrationCount* parameter can be altered by the modeler, thereby allowing assessment of different household member loss scenarios.

The agent-level likelihood that a household member migrates out of Washington, D.C., is controlled by the parameter *OutMigrationProb,* which is a household-specific value. *OutMigrationProb* is derived from a simple logistic regression model that predicts the likelihood of moving out of Washington, D.C., as a function of the householder member's age, household type and presence of children five years old or younger. For married couple households or other family households, *OutMigrationProb* is applied to the head of the household. If the head of the household is chosen to out-migrate, the entire household out-migrates. In contrast, for non-family households, *OutMigrationProb* is applied to each household member. If a household member is chosen to migrate out, the other members of the household remain. Putting it all together, in each time step, the overall number of household members selected to out-migrate (controlled by *OutMigrationCount*) is set. Then, individual household members or households are chosen to out-migrate with a probability equal to *OutMigrationProb*. It is important to note that housing prices are absent from the determination of *OutMigrationCount* and the out-migration probability specification. This means DCHAS household members do not endogenously respond to changes (increases) in housing prices by leaving Washington, D.C.

3.3.2. Household Member Life Cycle Events and Parameters

In DCHAS, household members progress through a life cycle, which includes aging and up to six life cycle events. Exogenous life cycle parameters, best characterized as *hazard rates*, are used to trigger household member life cycle events. The global frequency of each life cycle event and the agent-level likelihood of the event occurring are *jointly* implemented by having the parameters vary by the typology of the household members, coupled with random selection of an agent with a typology. For instance, if the marriage rate for household members of a specific typology is four percent per year, then a random four out of each 100 household members in that typology will be chosen to be married each year. The parameter values are based on empirical data from prior years, and remain static throughout model runs. Table 3.3 lists the six life cycle events, the formal parameter(s) for each event, and how the parameter values varies by household member typology. All parameter values are rates, except for the *FirstJobIncome18* and *FirstJobIncome25*, which are continuous values.

Table 3.3: Household Member Life Cycle Events. Parameters, and Variation in Parameter Values

Life cycle Event	**Parameter Name**	**Parameter Value Varies By…**
Birth	*BirthRate*	Race/ethnicity, age, income and marital status
Death	*DeathRate*	Age
Marriage	*MarriageRate*	Race/ethnicity, age, gender and income
Divorce	*DivorceRate*	Race/ethnicity, age, gender and income
First job	*FirstJobIncome18; First JobIncome25*	Race/ethnicity, age, gender
Income change	*IncomeChangeRate*	Race/ethnicity, age, gender

The birth event happens to female household members. The event parameter, *BirthRate*, is an annual probability that a female gives birth. At each time step, females are selected to give birth and the female's child becomes a member of the female's household. The parameter values vary by race/ethnicity, age, income and marital status of the female – the four attributes available for each female household member.

The death event happens to all household members. The event parameter, *DeathRate*, is an annual probability of death. At each time step, household members are selected to die. The death of a household member reduces the size of the household, possibly the income of the household. If the death occurs to a head of household, a new head of household is chosen. If the death occurs to a single-person household, the household is removed. The *DeathRate* parameter values vary only by age, reflecting what data on death rates is available. It is acknowledged that other factors may contribute to death rates.

The marriage event happens to single people who are 18 years old or older. The event parameter, *MarriageRate*, is an annual probability of a person getting married. The rate varies by race/ethnicity, age, gender and income, reflecting the attributes of persons available in DCHAS. It is acknowledged that other factors may contribute to marriage rates but are not currently available in DCHAS.

In DCHAS, marriage happens only between men and women[5]. At each time step, unmarried household members are selected to be married. An equal number of men and women are randomly selected, then paired together based on attributes. The pairing is relatively simple (a sorted list based on the aforementioned attributes) and it is acknowledged that there are other ABMs of the marriage process that could produce results more consistent with the empirical data[6].

The divorce event happens to married couples. The event parameter, *DivorceRate*, is an annual probability of a married couple getting divorced. The rate varies by race/ethnicity, age, gender and income, reflecting the attributes of persons available in DCHAS. It is acknowledged that other factors may contribute to divorce rates but are not currently available in DCHAS. When a married couple divorces, one member of the couple is forced to move (as discussed in the section 3.3.3).

---

[5] Same-sex marriages were not legalized until June 26, 2015. The Bureau of the Census did not designate same-sex couple survey respondents who were legally married as married until the 2017. As a result, it was not possible to empirically estimate marriage rates for same-sex couples. This is a shortcoming of DCHAS that can be rectified in the future when data becomes available.

[6] For a good example, see the "Wedding Ring" model (Billari, Prskawetz, Diaz, & Fent, 2007).

The first job event occurs twice. The first occurrence is when a household member reaches age 18. The household member is given a job with a modest annual income based on the parameter *FirstJobIncome18*. The second occurrence is when a household member reaches the age of 25, based on the parameter *FirstJobIncome25*. The second occurrence reflects the transition of a person into a full-time work after competition of (what is assumed) any schooling or job training that occurs between the ages of 18 and 24.

Functionally, both parameters are cumulative distributions of income, varying by race/ethnicity, age and gender. Each person who reaches the age of 18 (or 25) is given a random number from the uniform distribution. This number is then used to select a corresponding value from the *FirstJobIncome18* or *FirstJobIncome25* cumulative distribution. There is no formal representation of education attendance or attainment in DCHAS due to the lack of this information in the data source used to create DCHAS persons (discussed in section 4.1).

The income change event occurs for all persons within income. The event parameter, *IncomeChangeRate*, is an annual percent change in income for a person. The rate varies by race/ethnicity, age and gender, reflecting the attributes of persons available in DCHAS. The implementation represents a smooth income growth or decline. There is no mechanism in DCHAS for a household member to increase their income faster than the *IncomeChangeRate* for that member's typology. It is acknowledged that other attributes (i.e., education level) could contribute to the rate of income change, but are not currently available in DCHAS.

3.3.3. Household Mobility and Formation Events and Parameters

In general, a household's relocation decision involves four components: the decision to move (mobility), the decision to live with or without roommates (for single people) (formation), the decision to rent or own (tenure choice) and the choice of location. In practice, these decisions may be made independently or interdependently. A typical neoclassical formulation of the relocation decision is to assume that households are constantly evaluating the utility of their current location relative to all other locations, and choosing to relocate if they find an alternative location that yields greater utility than their current location (Ettema, 2011).[7]

In DCHAS, households are consumers of housing units (as owners and renters) and sellers of housing units (as owners). As consumers of housing units, households face four decisions: whether to relocate, whether to live alone or with others (for single persons), whether to be renters or owners and where to relocate. Those decisions are reflected in three events: the mobility and formation event, the tenure choice event, and the location selection event.

In DCHAS, the decision to move is not of importance to the modeling effort. As such, the mobility portion of the mobility and formation event is controlled by an exogenous parameter *MoveRate*, similar to the formulation presented in Jordan, Birken, and Evans (2012). The global frequency and the agent-level likelihood of the purchasing a home are *jointly* implemented by having *MoveRate* vary by the typology of the household members. For unmarried household members, *MoveRate* varies by age,

[7] For a historical perspective on how relocation decisions have been represented in the literature, see Dieleman (2001).

income and gender, reflecting the attributes of persons available in DCHAS. For other family or married couple households, *MoveRate* varies by age of householder, household income, and presence of a child younger than five years old. Both specification of *MoveRate* reflect the attributes of household members and households available in DCHAS. It is acknowledged that other attributes could contribute to the probability of moving, but are not currently available in DCHAS.

Before discussing the household formation portion of the mobility and formation event, it is useful to discuss the immediate impact on household formation from the events mentioned in sections 3.3.1 and 3.3.2. Table 3.4 summarizes the immediate impact on household formation from the aforementioned events. Any new household created from a life cycle event is automatically given a *MoveRate* value equal to 1, meaning the household must go through the processes described in sections 3.3.4 and 3.3.5.

Table 3.4: Household Formation for Each Household Member Life Cycle Events

Life cycle Event	**Immediate Impact on Household Formation**
In-Migration	Creates new households and new household members to move into existing non-family households
Out-Migration	Removes entire existing households or members from existing non-family households
Birth	No net impact. Births occur within existing households.
Death	A household with one person would be removed. A household with more than one person would remain a household
Marriage	If two household members living by themselves get married, then one of the households is removed. In all other cases, the number of households remains unchanged.
Divorce	A member of an existing household is forced to move. If that member decides to live by themself, then a new household is created. If that member moved into a non-family household, then there is no net change in the number of households.
First job	Does not have an immediate impact the number of households.

Income change	Does not have an immediate impact the number of households.

In DCHAS, additional household formation occurs when an unmarried household member living decides to move into their own housing unit. This formation event is controlled by an exogenous parameter *FormationRate*. The global frequency and the agent-level likelihood of the purchasing a home are *jointly* implemented by having *FormationRate* vary by the typology of the household members. *FormationRate* varies by age, income and gender, reflecting the attributes of persons available in DCHAS. When married and other family households are selected to move, they move as an entire household. As such, there is no additional household formation (or loss) resulting from moves of these household types.

Finally, DCHAS includes a non-market mechanism for placing household members who must move, but are not picked to form their own household, as tenants into existing household. This mechanism applies to in-migrants, divorcees and household member picked to move via the application of *MoveRate.* The placement is based on whether the household has a free bedroom available, as well as whether the household is of the same or similar household income group, householder age group, and race. It is important to note this mechanism is non-market, meaning prices and income relative to prices do not factor into the placement process.

3.3.4. Household Tenure Choice Event and Parameter

When an *entire* household decides to move, they must make a choice to be renters or owners. Some households who are currently renters may choose to purchase a home, and vice versa. Household *preference* for owning versus renting varies with the life cycle, as is obvious in any analysis of homeownership rates by age. Household wealth influences the ability, and perhaps the preference, of a household to purchase a home. Unfortunately, the primary sources of household data used in DCHAS do not include household wealth.

In DCHAS, the decision to rent or own is not of importance to the modeling effort – DCHAS focuses on the low-income population, which are predominantly renters. Therefore, in DCHAS, the decision to purchase or rent a home is determined by an exogenous parameter, *PurchaseRate*. The global frequency and the agent-level likelihood of purchasing a home are *jointly* implemented by having *PurchaseRate* vary by the typology of the household members, coupled with random selection of an agent with a typology. The parameter values are based on empirical data. The net result is that the overall homeownership rate for the typology is maintained.

3.3.5. Owner Household Appraisal Event and Parameters

Households who own their home but have decided to move must sell their existing home. To sell, these households must list the property on the market, which requires establishing a sale price. The first step in a typical process of setting the price of a housing unit is to obtain an appraisal. An appraiser estimates the potential sale price of a home by reviewing the sale prices of comparable housing units that have sold recently, then applying statistical measurements, such as a median or average. To make recently sold housing units comparable to the housing unit about to enter the market, the appraiser typically adjusts the value of the comparable housing units to capture any differences, such as the number of bedrooms, number of bathrooms, square footage, and age of the home.

For buyers, the process is largely the same. Buyers obtain an appraisal either on their own or as a condition of receiving financing through a lender.

In DCHAS, buyer and seller households initiate the appraisal event. The actions of an appraiser are performed as a feature of the model – appraisers are not an agent in the model. Current appraised price is predicted based on a hedonic price model with a spatial auto-regressive term, where the weights reflect the recency of sale for units in the same neighborhood. Formally, this is expressed as

$$P_l = \delta Z_l + \rho W P \tag{3.1}$$

where δ and ρ are parameters to be estimated, Z_l is an array of housing and characteristics describing housing unit l; P_l is the price of housing unit l; P is a vector of all recent sale prices; and W is a spatial weight matrix (Anselin, 1988). The appraisal process is performed each month for each housing unit that enters the for-sale market. When a housing unit is sold, it becomes a data point for all future appraisal processes for other housing units. In other words, it becomes part of the updated vector of prices, P, which then acts through the spatial weight matrix W and the parameter ρ to form new predictions of P_l.

3.3.6. Household Location Selection Event and Parameters

Following Waddell (2010) and Lee and Waddell (2010), a DCHAS household's location decisions are based on calculating the utility for various housing alternatives, then choosing the option with the highest utility.[8] Formally, as specified by Lee and Waddell (2010), the utility of location l for household n, is expressed as

$$V_l = \alpha X_l + \beta (Y_n - P_l) + \gamma H_n X_l \tag{3.2}$$

where α, β, and γ are parameters to be estimated; X_l is an array of housing and neighborhood characteristics describing housing alternative l; Y_n is the household annual income; P_l is the housing price in annualized rents; and H_n is an array of attributes for household n. The interpretation of this utility expression is that when a household makes a residential location decision, they consider their own characteristics (e.g., income and size) and the characteristics of the available housing options (e.g., price, number of bedrooms, median tract income, distance to metro station). The interaction terms (Y_n - P_l and H_nX_l) permit the household attributes to enter the model with the location characteristics. As noted by Lee and Waddell (2010), the household attributes cannot be specified by themselves in the utility function because they do not vary across the alternatives and there would be no way to estimate coefficients for such variables. Finally, it must be noted that the X_l varies based on whether the household is evaluating rental or owner-occupied options.

[8] The underlying economic model is called the Random Utility Model. For further discussion, see McFadden (1986).

Not all housing options, or even a household's highest utility option, are feasible for the household. As evidenced by survey data on income and housing expenditures, households typically do not spend more than a certain percentage of their income on housing. This percentage varies between owners and renters, and at various income levels.[9] The utility formulation in equation 3.2 endogenously captures the share of income devoted to housing among households. However, to ensure that households do not spend too much income on housing, the set of alternatives from which a household chooses is constrained by a threshold representing the maximum share of their income, *IncomeShare,* that can be devoted to housing costs, where housing costs include the annual rent or annual mortgage payments, in addition to the annual cost of utilities. The *IncomeShare* parameter varies by household typology and is based on empirical estimates. Moreover, as discussed in section 3.2.2, some renter households are subsidized. Functionally, this means that they only pay one-third of their income towards housing, with the government paying the difference between the renter's contribution and the market rent.

Finally, the mobility and formation event discussed in section 3.3.3 treats a single person's decision to live by themselves as exogenous to market conditions (i.e., not dependent on housing price). DCHAS was designed in the manner to have better control over the formation rate, which is a vital part of the demographic process. An alternative specification, and one what may have better alignment with neoclassic economic theory, would be to encapsulate a single person's decision to live by themselves within the utility

[9] In neoclassical economics, this is formalized as a preference for the bundle of all other goods.

calculation in equation 3.2. For instance, a single person could evaluate the utility of several rental housing options where a characteristic of a rental unit is whether the unit currently has a roommate present. The rental rate for the unit would reflect the presence of the roommate, who could be assumed to be contributing to the overall rent of the unit. In effect, all single persons would be treated as their own household, while all rental units with two or more bedrooms would be treated as though they provide two single-bed units. This is certainly a subject that could be addressed in future work.

3.3.7. Owner Household Price Negotiation and Market Clearing Event

Hedonic models of housing prices may not necessarily produce a current market equilibrium price because the process does not capture the underlying supply or demand for housing in a market at a given time (Epple, 1987). Unobservable characteristics of the buyers and sellers, the number of other homes currently for sale, and the number of buyers currently looking for a home could all potentially impact the market-clearing equilibrium price.

One important feature of the modern information-rich world is the amount of information readily available about current transactions in a housing market. For instance, real estate websites such as Redfin© can easily be queried to extract the number of housing units currently on the market (an indicator of supply) and the price and average length of time recently sold housing units were on the market (an indicator of demand). Unfortunately, it is difficult to globally capture historical data on the number of for-sale housing units or time-on-market data for *all* properties using Redfin or other sites[10]. As such, these indicators of supply and demand cannot be used as an explanatory variable in the hedonic price equation specified in equation 3.1.

The characteristics of buyers and sellers are less apparent from an empirical perspective. Specifically, what is captured by the housing transaction is the price at which the two parties agreed. This price may not necessarily reflect either the buyer's willingness to pay or the seller's willingness to accept. Willingness to pay and willingness to accept estimates would only be available from a buyer's or seller's fully specified utility function. They are not available from the random utility framework underlying the household's location decision process described in section 3.3.6.

Notwithstanding the aforementioned issues, the hedonic price equation (equation 3.1) used in DCHAS (and described in section 3.3.5) does attempt to reflect the underlying supply and demand through two mechanisms. First, recent sales are weighted more heavily via the spatial autoregressive term, meaning that the appraised price will be more current. Second, new sales from the most recent month become data points for appraising for-sale housing units in the current month, meaning that the appraised price will be more current.

As discussed in Waddell (2010), there are numerous plausible ways to represent the price negotiation and market-clearing process in a simulation framework. DCHAS uses a relatively simple approach, described in the following sequential steps:

1. DCHAS creates a list of households seeking to purchase housing and a list of

[10] This assertion is based on the author's experience. It is highly likely real estate websites have this data available for internal use and they may be willing to provide it to the public upon request.

households seeking to sell their housing unit at price P_l.

2. Each buyer household calculates their utility for all housing units available for sale at their appraisal price (P_l), then ranks all housing units according to their utility (1= highest utility).

3. For each for-sale housing unit, DCHAS calculates how many buyers have that unit ranked number one.

4. For housing units with only one buyer, the seller sets the asking price equal to the appraisal price and the purchase is made. These transactions are final.

5. For housing units with more than one buyer ranking the housing units as number one, the seller forms an asking price based on the appraisal price plus two percent[11].

6. Each remaining buyer household calculates their utility at the new asking prices, as well as their utility for all housing units not sold in step 4.

7. Repeat steps 3 through 6 until all buyers have made a purchase. The market clears when all buyers have purchased a housing unit.

Of course, steps 1 through 7 could lead to unsold housing units. If a housing unit remains on the market for more than three months, the original appraisal price is decreased by two percent each month until the unit has sold[12].

To recap, equation 3.2 is used to estimate the utility of various housing locations, and is based on the appraised price estimates (P_l) of housing units currently for sale.

---

[11] This number could be set lower or higher. Two percent is chosen based solely on the author's intuition as to the typical "bidding war" price increase rate when more than one buyer seeks to purchase the housing unit.

[12] This number could be set lower or higher. Two percent is chosen based solely on the author's intuition as to the typical price reduction rate when a property has not sold. In DCHAS, units for sale are not removed from the market.

Equation 3.1 continually produces updated appraised price estimates (P_l) based in part on recent sales prices (P), while asking price increases occur when more than one buyer is interested in the housing unit. The market clears when price increases result in only one remaining buyer. Housing units that remain on the market for more than three months are reduced in price until they are sold. All recent sales cycle back into the price vector (P). The net result of these features is that DCHAS housing location decisions, owner housing price formation, and market clearing is fully endogenous.

3.3.8. Renter Household Bid Rent Formation

Section 3.3.5 discussed buyer's bid price formation for owner-occupied housing. Renter households seeking to rent a new housing unit must also perform a similar calculation. Anecdotally, it is known that renters can and do negotiate with landlords over rental price. However, there is little easily accessible empirical data available on initial asking rents, the rent negotiation process, or incentives offered to renters, such as free amenities or one month free rent.

Because of the lack of empirical data, DCHAS does not include a renter's bid rent formation process or a negotiation process between renters and landlords. Instead, it is assumed that renters are price takers in that they must pay the rental rate offered by the landlord.

3.4. Landlords Types and Events

Generally speaking, property owners who are landlords (i.e., providing housing to others) are responsible for generating rental income, making decisions about capital investment (property improvement), and making decisions about whether to keep or sell their property. Moreover, landlords may realize property value gains through land value appreciation over time.

There is much variation in landlord sizes and motivations.[13] For instance, single family and small multifamily landlords (1-4 units) may be motivated by cash flow, by appreciation, or by both. Data from the 2012 Rental Housing Finance Survey (United States Department of Housing and Urban Development and United States Bureau of the Census, 2012) show that single family and small multifamily landlords do not generate large rental returns, and many are sole proprietors who actively manage their own properties. In contrast, commercial landlords who own large multifamily properties are often motivated by positive cash flow, resulting in returns to investors. Commercial landlords typically have professional management companies who are expert at optimizing rents so as to ensure low vacancy rates, as well as making capital investment decisions.

3.4.1. Landlord Types

Although variation in landlord types and motivation exists, in the DCHAS model there are only three types of landlords: market rate landlords who are subject to rent control, market rate landlords who are not subject to rent control, and a single landlord that oversees public housing and project-based section 8 properties. The market rate landlords are assumed to be profit maximizing firms.

[13] For an interesting profit analysis of landlord types and motivations, see Mallach (2016).

3.4.2. Market Rate Landlord's Rent Price Setting Event

As discussed in section 2.5.4, about two-thirds of rental units in Washington, D.C. are subject to rent control, which caps the annual rate of increase on rents. Rent control is not applicable to units owned by an individual who does not own more than five units. Moreover, rent control is not applicable to rental units built after 1975.

Whereas the seller of a housing unit (and their appraiser) has access to a rich set of information about recent housing sales, landlords likely have access to the rental rates for many recent rental transactions. Both small and large landlords have access to rental units that are advertised through mediums such as Rent.com® or Craigslist® and some may subscribe to services from rental data providers, such as REIS®. Moreover, it is widely known that rent optimization software is used by large landlords[14]. While a full explanation of this software (and its data sources) is beyond the scope of this analysis, this software likely includes data from recent rental transactions reported by other subscribers to the software.

There is no available empirical evidence to suggest that small landlords are failing to form asking rents that are less optimal (profit maximizing) than large landlords. As such, all landlords in DCHAS are assumed to follow the same process to form initial asking rents. Formally, this is given by the equation

$$R_l = \mu X_l + \sigma WR \tag{3.3}$$

[14] One example is RentPush.com®.

where μ and σ are parameters to be estimated; X_l is an array of housing and neighborhood characteristics describing the rental housing unit l; R_l is the rental price of a unit; R is a vector of current rentals, and W is a spatial weight matrix (Anselin, 1988). The spatial weight matrix W gives equal weight to all currently rented units (and their rental rates) within the neighborhood. As such, this formulation reflects the average rental price for the neighborhood, adjusted for characteristics of the units.

Of course, the initial asking rent may not reflect the rental vacancy rate. Landlords are assumed to be profit maximizers, meaning they adjust prices based on vacancy rates. Formally, the adjustment is

$$R_{ask} = R_l * (1 - \varepsilon) \tag{3.4}$$

where ε equals the neighborhood vacancy rate less the D.C. vacancy rate. The neighborhood rental vacancy rate is the current rental vacancy rate for only the neighborhood in which the property is located, while the D.C. vacancy rate is the vacancy rate for the entire District. This adjustment is relatively simple because there is no available empirical evidence on the rate at which landlords gradually reduce asking rent until a unit is rented. The rationale for this type of adjustment is based on solid empirical evidence that there is an underlying rental vacancy rate in the rental market at the national level, regional level, and both within and outside of metropolitan areas (United States Bureau of the Census, 2017). This rental vacancy rate may somewhat fluctuate over time, but is always present. A discussion of precisely why rental housing markets have an underlying vacancy rate is beyond the scope of the analysis. However, it is a fact that profit-maximizing rental housing developers build rental housing in markets where there have historically been excess rental housing units available to households seeking to rent.

Landlords who are subject to rent control perform the same calculation. However, the asking rent is simply the rent they are allowed to charge under rent control regulations, unless R_{ask} is less than the rent allowable under rent control, at which point the landlord would charge R_{ask}.

3.4.3. Rental Market Clearing Event

All landlords post their vacant rental units on the rental list and engage in rental market transactions with renter households (discussed in section 3.3.8) until every renter seeking housing has found a unit they can afford. DCHAS uses a relatively simple approach for rental market transactions and market clearing, described in the following sequential steps:

1. DCHAS creates a list of households seeking to rent housing and a list of landlords with available units to rent at rental price R_{ask}.
2. Each renter household calculates their utility for all housing units available for rent at their asking rent (R_{ask}), then ranks all rental housing units according to their utility (1=highest utility).
3. For each available rental housing unit, DCHAS calculates how many renters have that unit ranked number one.
4. For housing units with only one potential renter, the renter rents at R_{ask}. This

transaction is final.

5. For housing units with more than one renter household ranking the rental housing units as number one, a single random renter household is chosen to receive the rental unit as rental price R_{ask}

6. Renters re-rank each remaining rental units.

7. Repeat steps 5 through 6 until all renter households have rented a unit. The market clears when all renter households have rented a housing unit.

To recap, equations 3.3 and 3.4 continually produce updated rent estimates based on recent rentals. Rent adjustments occur based on a relative vacancy rate. The market clears when all renters have found a unit they can afford. The result is that DCHAS rent formation is fully endogenous.

3.4.4. Market Rate Landlord's Capital Improvement Spending Event and Parameter

Landlords make (or chose not to make) capital improvements to their units and capital improvements to units change their effective age, which in turn allow a landlord to charge an additional rent premium. The 2012 Rental Housing Finance Survey includes statistics on the capital improvement expenditures for properties of various sizes (United States Department of Housing and Urban Development and United States Bureau of the Census, 2012). However, the estimates are national and may not necessarily reflect behavior of D.C. landlords.

Capital improvements are important to DCHAS because the downward filtering of units increases the available rental housing stock for low-income renters. However, rather than model the complex investment behavior of landlords, DCHAS includes a global parameter, *PropInvestRate,* which is adjustable by the modeler. *PropInvestRate* reflects the share of a landlord's units for which the landlord performs capital improvements, thereby reducing the effective age of the unit to 0. For instance, a *PropInvestRate* of one percent means that, annually, one percent of the landlord's units have their effective age changed to 0.

3.4.5. Market Rate Landlord's Land Price Setting

In DCHAS, a simplifying assumption is made that landlords are price takers and will sell their existing property (which may include existing multifamily structures) to the Developer when an offer is received. Section 3.5.4 provides further explanation.

3.4.6. Subsidized Housing Landlord's Tenant Selection Event

In DCHAS, all public housing and project-based section 8 properties are managed by a single landlord. When a subsidized unit becomes available, the subsidized housing landlord takes two actions. First, the subsidized housing landlord fills an empty subsidized unit by selecting a low-income renter household who is currently seeking market-rate rental housing to occupy the subsidized rental unit. Second, the subsidized housing landlord calculates the household's share of the rent based on the Fair Market Rent for each unit and the household's income, where Fair Market Rents are determined by HUD. The subsidized household pays one-third of their income to the subsidized housing landlord.

3.5. Market Rate Developer Events

In a neoclassical economic context, developers are profit maximizing firms who choose to develop housing units in places where expected returns are the highest, given input and output prices. Their decisions are often formalized in a production function that equates output (a unit of housing) to units of input (labor hours, materials, land, permitting costs, etc.).

Estimating the parameters of a housing production function is notoriously difficult (Epple, Gordon, & Sieg, 2010; Combes & Duranton, 2016; Ahlfeldt & McMillen, 2013). One difficulty, which is especially prevalent in urban areas, is that some land is more productive than other land in the sense that it can accommodate more units of housing (more square footage) per acre of land, commonly known as a floor-to-area ratio (FAR). There are a few reasons why this feature of land is present. First, zoning regulations often place a cap on FAR and/or a cap on height (these are functionally similar). Second, physical features such as slope or soil type limit the FAR, although physical limits vary across the landscape. Regardless of the reason, the productivity of land for housing impacts the input choices of a developer, which creates an endogeneity problem when estimating production function parameters. To be sure, statistical techniques, such as Instrumental Variables, have been used to address the endogeneity issue, but this technique requires a suitable instrument (i.e., data) which may not be available (Epple, Gordon, & Sieg, 2010).

Another difficulty arises because multifamily properties, like single family properties, are heterogeneous in quality; even one-bedroom apartments of similar size can vary in quality from property to property. In other words, not every square foot of housing produced is the same. As such, it is difficult to determine what constitutes a "unit" of housing production for multifamily properties when quality is heterogeneous.

The type of data necessary to determine quality (and hence, a unit of quantity), such as detailed characteristics of the housing units or prices of the construction materials, is not often available for a large cross-section of newly-developed units.

3.5.1. The Impact of Zoning on Market Rate Developers

As mentioned before, there is a major constraint to building that impacts the expected profits to development: zoning restrictions. In D.C., a developer is permitted by right to construct the maximum number of units allowed under the zoning code. For some zoning codes, this translates into no more than two units. For other zoning codes, the maximum number of units is not specified, but expressed as a maximum FAR, which can then be used to estimate the maximum number of units.

When developers in Washington, D.C. desire to build more housing units than is allowable by the current zoning code (i.e., an increase in the FAR), they apply for a PUD. The PUD application process is not automatic and developers have to adhere to certain restrictions that still cap the number of units they can build, expressed as a new maximum FAR. Moreover, PUD developers are required to offer community benefits, such as donations to local entities or improvements to neighboring properties. Finally, PUD developers are required to dedicate a certain number of units in their development as affordable.

Many PUD applications are approved by the D.C. Zoning Board and, anecdotally, there is a pattern to where PUDs are approved: on big properties located near subway stops and commercial areas[15].

3.5.2. Simplifying Assumptions

The actions of market rate developers are not of central importance in the DCHAS model. Moreover, as discussed in the section 3.5, the type of data necessary to estimate the parameters of a production function is not readily available. As such, DCHAS simplifies the development process into two decisions: how much market-rate development will occur, and where it will be located. To control the amount of development, DCHAS includes an exogenous market-rate development parameter, *DevelopmentRate*, that allows the modeler to conduct scenario analysis using different development rates.

Following Magliocca et al. (2011), market rate developers in DCHAS are represented by a single representative market rate developer, hereinafter referred to as the Developer. The Developer makes the decision concerning where new development will occur. New development can be accomplished through demolition and redevelopment of properties with existing units, or through the development of vacant properties.

Following Waddell (2002), the Developer performs a pro forma analysis whereby he seeks to place housing in locations where the profits to development are the highest, as defined by the following profit function

[15] Of course, this patter could simply reflect where developers chose submit PUD applications or where developers believe PUD applications are likely to be approved.

$$\pi_i = ER_i - PC_i - CC_i \tag{3.5}$$

where ER is the expected revenue for property i; PC is the cost to acquire property i; and CC is the construction cost to build on property i. Because the expected revenue and property costs are based on current property and housing unit sales, the decision of where to locate development is endogenous to the model.

To simplify the Developer's profit function, four assumptions are made. First, the Developer builds the maximum number of units permitted "by right" in the zoning code, or if permitted by the Local Government, the maximum number of units permitted under a PUD as discussed in section 3.5.1. This in turn dictates the type of building constructed (single family/townhouse, low rise, or high rise). Second, the Developer chooses a standard mix of unit sizes, expressed as the number of bedrooms, with the unit size mix based on empirical data. Third, for each property developed, the Developer chooses whether the units will be rented as part of a multi-family rental property or sold as single-family units, including condos, based on a simple rule that maintains the overall level of homes that are rental versus owner-occupied.

Finally, as discussed in section 3.5.1, can submit a PUD application to increase the maximum number of units allowable. As such, the fourth assumption is the Developer performs the all the actions discussed in sections 3.5.3 through 3.5.6 based on the number of units the developer would be allowed to build under a PUD.

3.5.3. The Market Rate Developer's Expected Revenue Calculation Event

In DCHAS, the expected revenue for a property (ER_i) is equal to the non-discounted sum of appraised prices estimates in equation 3.2, and the sum of the appraised prices becomes the final sale price. Although some market-rate development is multifamily housing, the final sale price is a good proxy for the discounted stream of future rents that would be earned if the unit is rented.

3.5.4. The Market Rate Developer's Property Cost Appraisal Event

It can reasonably be assumed that multifamily landlords are informed enough to understand what their property is worth to developers, given its current land use. Similarly, it can reasonably be assumed that developers are informed enough to calculate the present value of a landlord's property, given current rental rates, current property condition, current costs, and current land use. Moreover, economic theory suggests that the current value of a property reflects that capitalized stream of revenue generated by the property under its current land use.

In DCHAS, the Developer estimates a current appraised price for a vacant property or a multifamily property based on a hedonic price model with a spatial auto-regressive term, where the weights reflect the recency of sale for other properties. Formally, this is expressed as

$$Y_i = \lambda M_i + \nu W \tag{3.6}$$

where λ and ν are parameters to be estimated; M_l is an array of characteristics about the property i, including distance to transportation and allowable number of units; Y_i is the price of property i; Y is a vector of sales prices of properties; and W is a spatial weight matrix (Anselin, 1988). When a property is sold, it becomes a data point for all future appraisal processes for other properties.

3.5.5. The Market Rate Developer's Construction Costs Calculation Event

Construction costs per square foot vary by building type (townhouse, low-rise multifamily, high rise multifamily) and quality (high or average). Formally, total construction costs for a property are

$$CC_i = \ gross\ building\ area * cost\ per\ square\ foot \qquad (3.7)$$

3.5.6. The Market Rate Development Event and Parameter

Given the modeler's choice of *DevelopmentRate,* the Developer then calculates the profit from all potential development sites, ranks the profit potential of all sites, then chooses the sites with the greatest profit potential. The sites are chosen sequentially until the total number of units to be developed equals the *DevelopmentRate*. A simplifying assumption is made that construction of the units occurs directly after the development location decision has been made. This assumption is purely for modeling convenience and is not meant to reflect the time it takes to construct properties.

To recap, expected revenue (*ER*) is the sum of appraised prices estimated in equation 3.2, which itself is based on recent sales. Property costs (PC) are estimated based on recent sales. As such, the Developer's profit function is endogenously determined.

3.6. D.C. Government Events

In the DCHAS model, the D.C. government has two roles. The first is to develop affordable housing. The second role is to approve zoning variances.

3.6.1. The Affordable Housing Development Event

Projecting affordable housing development is incredibly difficult. First, affordable housing development depends heavily on the availability of LIHTC funding. LIHTC funding is provided by the federal government and subject to fluctuations due to annual appropriations. Second, there are typically numerous other smaller funding streams used to produce affordable housing, including the HPTF, which is controlled by the Washington, D.C. government. Lastly, affordable housing developers include both for-profit developers and non-profit developers, and their respective decision-making processes are likely different.

Rather than create a model of the behavior (i.e., the decision-making process) of affordable housing developers, the DCHAS model includes two exogenous parameters that can be adjusted by the modeler. The first is called *SubUnitInvestment,* which reflects the total dollars invested in rehabilitating and retaining expiring LIHTC units, or constructing new units. The second adjustable parameter is *RehabShare,* which reflects the share of *SubUnitInvestment* set aside for rehabilitating or retaining expiring LIHTC units versus constructing new units. The set-up allows for a DCHAS model that provides insight into D.C.'s state of subsidized housing quality and need, with the added ability to adjust the parameters based on the modeler's judgement.

With respect to where new subsidized units would be located, two simplifying assumptions are made. First, only properties that are currently owned by the D.C. government or by non-profit entities are eligible to receive subsidized housing. Second, the number of units built on those properties reflect similar housing density in the neighborhood.

3.6.2. Approval of Zoning Changes (PUD Approval Event)

In DCHAS, the D.C. government approves (or does not approve) zoning variances and other zoning changes that allows developers to build more units than the by right limit. This representation of local government is abstract and glosses over numerous entities involved in the decision to approve zoning variances or zoning changes, including the Advisory Neighborhood Commissions and the Zoning Board.

Functionally, DCHAS allows the Developer to submit a PUD application to the D.C. government. The D.C. government that decides whether to approve the PUD application or reject the application. If an application is approved, the Developer is permitted to construct the number of units allowable under PUD. If the application is rejected, the Developer simply build the number of unit allowable under current zoning. To implement this decision framework, DCHAS includes a parameter, *PUDUnitsApproved,* that is adjustable by the modeler. The baseline number of PUD units approved is based on empirical data from past PUD approvals. However, the modeler can increase or decrease the number of units approved. The PUD units are a subset of the overall *DevelopmentRate*, meaning that if the modeler specifies *DevelopmentRate* to be 20,000 units per year, and *PUDUnitsApproved* to be 5,000 units per year, then 15,000 non-PUD units will be built.

3.7. Landscape: Property Attributes

Housing units are a necessary part of an ABM of the housing market. However, it is important to recognize that housing units are part of residential properties composed of land and residential structures containing housing units. In the case of multifamily properties, decisions are made by landlords at the property level.

As discussed in Chapter 2, there are a handful of property features that are relevant to modeling housing affordability and affordable housing production. First, Washington, D.C. has a significant number of subsidized housing units occupied by low-income households. These properties are part of the housing stock, but not truly part of the private housing market. It is essential to identify the properties that are subsidized, as well as to understand when the LIHTC tax credit period will expire, which translates into a potential loss in affordable housing units unless new credits are received to retain the units.

Second, Washington, D.C. has rent control policies that apply to certain multifamily properties. Residential properties permitted after 1976 are not subject to rent control, as are residential properties owned by a person who owns three or fewer rental housing units.

Third, as discussed in Chapter 2 and previous sections in this chapter, zoning impacts the number of units permitted on a property, which directly impacts the expected revenue, and hence, the likelihood of development. However, exceptions to the zoning code to increase the number of units allowed may be granted through the PUD process. The PUD process is time consuming and developers are unlikely to seek approval for a PUD unless the likelihood of approval is high. Each property has an expected number of units permissible under PUD, based on an analysis of prior PUD approvals.

Table 3.4 below lists the attributes that are defined for each property.

Table 3.4: Attributes of Properties in the DCHAS Model

Property Attribute	**Description**
Ownership type	A categorical variable identifying the type of property based on ownership. Categories are: HUD subsidized rental, LIHTC rental, single-family, small multifamily, large multifamily, and vacant land/commercial/industrial.
Neighborhood	The neighborhood where the property is located, as defined by the Washington, D.C. government.
Subsidized expiration year	The year in which the LIHTC tax credit period expires.
Rent control flag	Flag indicating whether the property is under rent control.
By right maximum units	The maximum number of unit permitted under current zoning.
PUD maximum units	The expected maximum number of units permitted under a PUD.
Sale Price	The sale price of the most recent sale
Appraised Price	The current appraised price

3.8. Landscape: Housing Unit Attributes

Housing units are a part of properties. One possible ABM development strategy is to simply add “number of units” as an attribute of a property. However, there are some advantages to separating housing units from properties when building an ABM that explicitly includes, and places emphasis on, multifamily housing. These include, but are not limited to:

1. Multifamily properties most often include housing units of different sizes (as denoted by the number of bedrooms).

2. Units in multifamily properties are available for rent at different times.

3. Units in multifamily properties may rent for different rates, reflecting upgrades or remodeling of specific units.

As discussed in section 2.5.6, non-subsidized, low-income renters tend to occupy an older rental housing stock that has filtered down. However, as discussed in section 3.4.4, DCHAS landlords can rehabilitate units, thereby allowing landlords to charge higher rents. Knowing the age of each housing unit is vital to determining its market rent. DCHAS uses a concept known as *effective* age. Effective age is the age of the housing unit, but adjusted for major remodeling or gut rehabilitation. Data on the effective age of a housing unit is available (discussed in section 4.6.1), thereby permitting the use of this concept within DCHAS.

In addition to the housing unit attributes discussed above, DCHAS housing units have an estimated utility cost, which becomes part of the overall housing cost of a household occupying the unit. Table 3.5 below lists the attributes that are defined for each housing unit.

Table 3.5: Attributes of Housing Units in the DCHAS Model

Housing Unit Attribute	**Description**
Bedrooms	The number of bedrooms in the housing unit.
Effective age	The effective age of the housing unit.

Current market rental rate or mortgage cost	The current market rental rate or the mortgage cost based on recent sale price (annual).
Rent controlled rental rate	The actual permitted rental rate (annual).
Utility cost	The estimated cost of utilities (annual).

Finally, as mentioned in section 3.5, the Developer creates new housing units on properties. For properties that currently have units, the Developer will demolish the existing units. Similarity, the D.C. government rehabilitates and retains, or creates new affordable housing. In DCHAS, this is reflected by new housing units entering the housing stock or previously subsidized units transitioning to the market-rate housing stock.

CHAPTER 4: DATA

This chapter describes the data sources used to create the attributes of the agents and the housing stock in the DCHAS model, as well as the data sources used to parameterize the 17 agent events. To the maximum extent possible, the sections of this chapter align with sections in Chapter 3.

In the most general sense, the DCHAS model includes simulated households that reside in the housing stock. The attributes and behaviors (events) of the simulated households are derived from various survey sources or administrative data sources. The housing stock is simulated, but can be described as a realistic representation of the actual housing stock because most of the attributes are derived from tax assessment data. Some attributes of the housing stock are simulated using survey data.

As mentioned in Chapter 3, the DCHAS model is initialized for 2010, meaning it starts with conditions as they existed in 2010 and simulates going forward. As such, the base year for the household members and housing stock data sources mentioned below is 2010 or a year as close as possible to 2010. In contrast, the data sources used to parameterize the events are from 2011 to 2015, or as close as possible to that period.

This modeling strategy was implemented for a specific reason: a model with realistic representations of households and the housing stock as of 2010, subject to events based on data from 2011 to 2015, should produce simulated outcomes in 2015 that are in general agreement with conditions in 2015. If the model (DCHAS) can produce simulated outcomes consistent with conditions in 2015, then there is a stronger likelihood it can be used to simulate "out of sample" outcomes in year after 2016.

Finally, it is important to note that the initial placement of 2010 simulated households into a 2010 simulated housing stock is not covered in chapters 3 or 4. The discussion of that process is described in detail in Chapter 5 (section 5.2).

4.1. Household Member Attributes Data Sources

Household member data comes from the 2010 U.S. Synthetic Population Version 1 (Wheaton, 2010 U.S. Synthetic Population Ver. 1, 2014), hereinafter referred to as SynPop. SynPop is based on the 2010 Decennial Census Summary File 1 and the 2007-2011 ACS.[16] Synthetic household members in SynPop are endowed with various characteristics. These characteristics permit a straightforward determination of the four attributes of household members: age, sex, race, ethnicity and relationship to householder. In SynPop, income is simulated for the household. A simple process is used to distribute household income to members of the household[17].

SynPop does not indicate which households are owners or renters, nor what type of renter subsidy is received, if any. To add this information to SynPop, a special unpublished "HUD/ACS administrative match" version of the 5-year 2015 ACS Internal Use File (IUF) is used. The U.S. Bureau of the Census and HUD created this special version of the 5-year 2015 ACS IUF by matching HUD administrative records to the

[16] For additional information, see (Wheaton, 2014).
[17] In short, household income was distributed among the members of the household age 18 and older in proportion to their share of income in the overall population.

ACS households using addresses and person-level information. The result is a 5-year 2015 ACS IUF file with a flag indicating whether the ACS household receives HUD assistance and what type (public housing/project-based section 8, or voucher).

The special HUD/ACS administrative match version of the 5-year 2015 ACS IUF data was used to estimate the parameters of a MNL regression model that predicts the type of tenure (owner, nonsubsidized renter, public housing/project-based section 8, or HCVP) based on a vector of characteristics available in both SynPop and ACS. These include age of the householder, race and ethnicity of the householder, income, and family size. Once the parameters of the model are estimated using the ACS data, they are then applied to the SynPop data set to predict the likelihood a SynPop household is a member of each of the four types. Then, the SynPop households with the highest likelihood of receiving HUD assistance, by type, are flagged as such, up until the total number of households flagged as receiving HUD assistance, by type, matches a 2010 independent counts, by Ward, available from HUD's *Picture of Subsidized Housing* (United States Department of Housing and Urban Development, 2017). Then, the remaining households are assigned as non-subsidized renters up until the total number of renters matched the 2010 Decennial Census, by Ward[18]. Finally, all other households are assigned as owners. The independent counts of households by tenure (for the whole of DC) are available in Appendix A, Table A1. The parameter estimates of the tenure assignment model are available in Appendix A, Table A2.

[18] Washington, D.C. is divided geographically into eight Wards.

4.2. Households Event Data Sources

4.2.1. Migration Event Parameter Values

The DCHAS model parameters *InMigrationCount* and *OutMigrationCount* reflect the number of household members moving into and out of Washington, D.C., respectively. These two parameters are adjustable by the modeler. However, DCHAS includes a baseline number based on prior year net migration rates from 2010 through 2015. This number comes from the U.S. Bureau of Census's Population Estimates Program (United States Bureau of the Census, 2016a). The annual value for *InMigrationCount* is 64,930. The annual value for *OutMigrationCount* is 57,710.

To ensure an empirically accurate representation of the *type* of household members migrating into Washington, D.C., DCHAS uses a "donor" household member data set. This data set includes representative household members that have recently migrated to Washington, D.C., and is derived from the 5-year 2015 ACS public use microsample (PUMS) for Washington, D.C. (United States Bureau of the Census, 2016b)[19]. At each time step in the DCHAS, a fixed number of household members (5,411) are randomly selected from the donor set to be migrated into Washington, D.C.

To ensure an empirically accurate representation of the *type* of household members migrating out of Washington, D.C., DCHAS includes a parameter *OutMigrationProb*, reflecting the agent-level likelihood that a household members will move out of area. As discussed in section 3.3.1, the probability is derived from a simple logistic regression model expressing likelihood of out-migration as a function of the share

[19] The ACS PUMS data has a flag indicating if a household recently moved into Washington, D.C., from outside of the Washington, D.C., region.

of household type, householder age and the presence of children younger than five years old. The data used to estimate the parameters of the logistic regression model comes from the 5-year 2015 ACS PUMS for Washington, D.C. (United States Bureau of the Census, 2016b)[20]. At each time step in the DCHAS a fixed number of household members are selected from to leave Washington, D.C, and the probability that a household member is selected is based on *OutMigrationProb*. The parameters of this model are available in Appendix A, Table A3.

4.2.2. Household Member Life Cycle Event Parameter Values

As described in section 3.3.2, household members age and experience up to six life cycle events: getting first job, income growth, marriage, divorce, birth, and death. The parameters controlling these events vary by attributes of the household member.

For *FirstJobIncome, IncomeGrowthRate, MarriageRate, DivorceRate, and BirthRate,* the source of data is the 5-year 2015 ACS PUMS for Washington, D.C. (United States Bureau of the Census, 2016b). The 5-year PUMS includes data collected between 2011 and 2015. As such, an estimate from the 5-year PUMS reflects an average of the past five years. The marriage and divorce rates for Washington, D.C. are calculated following the methods described by Lewis and Krieder (2015). Due to the large number of data points, the values of the six parameters, which vary by attributes of the household member or household, are available upon request to the author.

For *DeathProb*, the source of data is the 2008-2012 Mortality Report published by D.C. government (Center for Policy, Planning, and Evaluation, 2016). The report

[20] The ACS PUMS data has a flag indicating if a household recently moved out of Washington, D.C.

includes death rates by age. The death rates by age are available in Appendix A, Table A4.

4.2.3. Household Mobility and Formation Event Parameter Values

For unmarried household members, the DCHAS model parameter *MoveRate* is based on the average number of household members that moved in the last year. The source of data is the 5-year 2015 ACS PUMS for Washington, D.C. However, the calculation excludes any household member that was recently divorced or migrated into Washington, D.C. As mentioned in section 3.3.3, DCHAS household members experiencing divorce or in-migration are automatically required to move. As such, *MoveRate* for unmarried household members is the average number of household members that moved in the last year for (what is assumed) reasons other than divorce or in-migration. For unmarried household members, *MoveRate* varies by personal income and age. For married and other family households, the *MoveRate* values vary by household income, the householder's age, and whether the household has a child under the age of five. The *MoveRate* values are available upon request to the author.

The DCHAS model parameter *FormationRate* is based on the average number of unmarried household members that moved in the last year *and* moved into their own housing units. The source of data is the 5-year 2015 ACS PUMS for Washington, D.C. The *FormationRate* values vary by age and personal income and are available upon request to the author.

4.2.4. Household Tenure Choice Event Parameter Value

In DCHAS, the decision to purchase or rent a home is determined by an exogenous parameter, *PurchaseProb*. Each household typology has its own *PurchaseProb* value. The values are based on the average number of households who owned their home between 2011 and 2015. The source of data is the 5-year 2015 ACS PUMS for Washington, D.C. The *PurchaseProb* values by household typology are available upon request to the author.

4.2.5. Buyer and Seller Appraisal (Hedonic Price) Event Data Sources

In DCHAS, both buyers and sellers are assumed to conduct appraisals of their property. The appraisal process is a feature of the model and occurs monthly for all vacant for-sale properties as well as any property newly entering the market. The appraisal process is formalized through a hedonic price equation (equation 3.2) which includes a vector of characteristics about the housing unit and neighborhood, as well as a spatial autoregressive parameter.

The initial prices and the characteristics of the housing units are derived from the D.C. Office of Tax and Revenue's (DCOTR) Computer Assisted Mass Appraisal (CAMA) data (District of Columbia Office of Tax and Revenue, 2016), which is described in section 4.6.1. The housing characteristics included in the hedonic model are: effective age of the structure, square footage, number of bedrooms, number of bathrooms, and housing structure type (single family attached, single family detached, condominium unit in a multifamily building). The neighborhood characteristics include median tract income and distance to the nearest metro station. Median tract income is derived from the 5-year 2015 ACS IUF for Washington, D.C. Distance to the nearest metro station is estimated using the physical location of CAMA housing units and the physical location of metro stations, which is available from DCOTR. The spatial weight matrix is based on the recency of sales, which is derived from the sales date. The parameter estimates for the hedonic price model are available in Appendix A, Table A5.

The housing and neighborhood characteristics were chosen for several reasons. First, and most importantly, the housing characteristics represent what is available in the CAMA data set. Second, the author conducted a limited qualitive review of recent appraisal documents to determine which characteristics of housing units and neighborhoods were formally captured in the appraisal process. Finally, median tract income is included as it is a common proxy for observable and unobservable neighborhood characteristics.

4.2.6. Household's Location Selection Event Data

In DCHAS, a household's location decision is determined by choosing the location with the highest utility, subject to a budget constraint. The utility formulation in equation 3.1 includes housing and neighborhood characteristics. The household characteristics are income, size, and the race of the householder. The characteristics of the available housing options are price; whether the housing unit is a single-family home; the number of bedrooms; and distance to the nearest metro station. The parameters of this model are estimated using the 5-year 2015 ACS IUF [21] for Washington, D.C. The

[21] It is important to note that the IUF is a somewhat different data source than the 5-year 2015 ACS PUMS. The IUF includes additional location information that permits determining median tract income and distance to the nearest metro station. This resource is only available to researchers with Census Special Sworn Status. However, the parameter estimates of the model (equation 3.1) are available to the public.

parameter estimates are available in Appendix A, Table A6A (for owners) and Table A6B (for renters).

The housing and neighborhood characteristics were chosen specifically because they are available for both the 2015 ACS and the CAMA data. In short, the parameter estimates from equation 3.1, which are estimated using the 2015 ACS data, are then applied to the housing units in the CAMA data, thereby allowing a household to estimate the utility for each simulated housing unit. This technique is only possible if the explanatory variables are available in both data sets.

4.2.7. Household's Income Share Parameter Value

In DCHAS, households cannot spend more than a certain percentage of their income on housing costs, which include both rent or mortgage and utilities, as denoted by the parameter *IncomeShare*. For simplicity sake, this value is set at the 95th quantile of percent of income devoted to housing, based on empirical estimates. Data on the 95th quantile comes from the 5-year 2015 ACS PUMS. The *IncomeShare* values by household typology are available upon request to the author.

4.3. Landlord Attribute and Event Data Sources

4.3.1. Rent Hedonic Model Data

In DCHAS, the landlords set rents for market-rate units (rents for LIHTC units are discussed in section 4.5.8). Initial rents for market-rate units are set via a hedonic rent model (equation 3.4) which includes a vector of characteristics about the rental housing unit and neighborhood, as well as a spatial autoregressive parameter. The source of rents, housing attributes, and neighborhood attributes for the hedonic price model is the 5-year

2015 ACS IUF for Washington, D.C. The housing attributes include structure type (dummy variable for single family), number of bedrooms, and age. The neighborhood attributes include the distance to the nearest metro station and median tract income.

These characteristics were chosen because they are available in both the ACS and the CAMA data, which means that the parameter estimates from the ACS-based hedonic rent model can be used to predict rents for all CAMA housing units. The parameter estimates from the model are available in Appendix A, Table A7.

The hedonic rent model (equation 3.4) includes a spatial autoregressive parameter and spatial weight matrix. The spatial weight matrix gives equal weight to all current rent values within the neighborhood of the rental housing unit. Moreover, rents are adjusted based on neighborhood vacancy (equation 3.5). Both equations assume the existence of a neighborhood. The neighborhood designation comes from the CAMA data, as discussed in section 4.5.1.

4.3.2. Property Investment Rate Parameter Estimates

DCHAS includes a global parameter *PropInvestRate* which is adjustable by the modeler. *PropInvestRate* reflects the share of a landlord's units for which the landlord performs capital improvements, thereby reducing the effective age of the unit to 0. The baseline value for *PropInvestRate* (6 percent) comes from the 2015 Rental Housing Finance Survey (United States Department of Housing and Urban Development and United States Bureau of the Census, 2017).

4.4. Developer Attributes and Event Data Sources

In DCHAS, the amount of new market-rate housing development is controlled by a parameter *DevelopmentRate* that is adjustable by the modeler. Once the rate is set, DCHAS's Developer is then left to determine where the development is located. This decision is based on a pro forma analysis of the expected profit (expected revenue minus property and construction costs) from each potential development location, and locations with the highest profit potential are chosen.

4.4.1. Development Rate Parameter Value

While the value for *DevelopmentRate* is set by the modeler, the baseline rate is based on the average number of new units constructed between 2011 and 2015. This information was extracted from the Bureau of the Census's Building Permits Survey (BPS) (United States Bureau of the Census, 2017). Table 4.1 shows the number of units built between 2011 and 2015, and the average number of units built during that time period.

Table 4.1: Total and Average Number of Housing Units Constructed in Washington, D.C., by Structure Type

Year	**In 1-unit Structures**	**In 2-4 unit Structures**	**In 5+ unit Structures**	**Total**
2011	227	100	4,285	4,612
2012	271	68	3,484	3,823
2013	333	112	2,810	3,255
2014	288	56	3,845	4,189
2015	255	113	4,588	4,956
Average 2011- 2015	275	90	3,802	4,167

4.4.2. Expected Revenue

The expected revenue from constructing new units, whether they are intended for rent or for sale, is based on the appraised price of new units. The methodology and data are the same as is described in section 4.2.5. However, the expected revenue calculation is made twice. The first calculation is based on the allowable number of units under the current zoning. The second calculation is based on the allowable number of units under a PUD.

4.4.3. Property Cost (Land Price) Hedonic Model Data

The cost to acquire existing properties, whether they are vacant or currently have housing units, is estimated using a hedonic price model of recent sales, as specified in equation 3.2. However, the dependent variable is price per square foot of land. The dependent variable is constructed by using the recent sales price, but calculating the portion of the recent sales attributed to the value of land. This calculation is performed by using the ratio of the assessed value of the land relative to the assessed value of the entire property, which includes the existing structures. The characteristics of the property include the distance to the nearest metro station. The spatial weight matrix is based on the recency of sales. CAMA data is used to estimate the parameters of the hedonic model. The parameter estimates from the model are available in Appendix A, Table A8.

4.4.4. Construction Costs Parameter Values

Construction costs come from 2014 estimates published by RSMeans® and are based on square footage and construction type. The construction types (I, II, and III) generally reflect high-rise, medium-rise, and single family, respectively. To ensure that construction costs match the year of construction, they are inflated or deflated based on the RSMeans® Construction Cost Index. The construction cost estimates are available in Appendix A, Table A9. As with the expected revenue calculation, the construction cost calculation is performed twice. First, the construction costs are calculated based on the number of units allowable under current zoning. Second, construction costs are calculated based on the number of unit allowable under a PUD.

4.5. D.C. Government Attribute and Event Data Sources

In DCHAS, the D.C. government invests in subsidized housing and approves PUD zoning changes.

4.5.1. Subsidized Unit Investment Parameter Values

DCHAS includes a parameter, *SubUnitInvestment*, which reflects the total dollars invested in rehabilitating or retaining expiring LIHTC units, or the construction of new units. DCHAS also includes a parameter, *RehabShare* which reflects the share of *SubUnitInvestment* set aside for rehabilitating or retaining expiring LIHTC units versus constructing new units. The parameter is adjustable by the modeler.

The baseline *SubUnitInvestment* and *RehabShare* values come from the D.C. Office of Chief Technology Officer's (OCTO) "10 x 20" database of affordable housing developments. This database includes D.C.'s investment levels as well as the number of units produced and rehabilitated. The default value of investment is $195 million per year. The default value of *RehabShare* is 55 percent. Finally, the investment cost per unit is $121,500, meaning that 1,605 units per year can be built or rehabilitated, given the investment level and per unit cost.

4.5.2. Planned Unit Development Units Approved Parameter Value

Whereas the baseline rate of overall development (*DevelopmentRate)* comes from the Bureau of Census's BPS, this data source does not indicate whether those units were by right development or part of a PUD. The CAMA data system includes a table that lists all the approved PUDs by year and, for some projects, number of units permitted. This data is used to estimate a baseline annual value of *PUDUnitsApproved,* which is 500.

4.6. Properties and Housing Units Data Sources

4.6.1. Market Rate Properties and Housing Units

Market rate properties are derived from the DCOTR's CAMA system. DCOTR annually publishes three CAMA data sets: Residential, Condominium, and Commercial. Each CAMA data set has one record per structure on a parcel, where parcels are uniquely identified by their value in the field "SSL." Each CAMA data set was extracted as of March 31, 2015, from DC.gov. Adjacent parcels with the same owner were aggregated to single properties.

The Residential CAMA data set generally contains properties with single-family residential structures, meaning that they are 1-unit properties. Nearly all single-family detached houses and row or townhouses are part of the Residential CAMA data set. The Condominium CAMA data set contains properties within condominium projects, and includes one record for each condominium unit. Structurally, condominium projects may resemble row houses, townhouses, or multifamily buildings, but where housing units are deeded at the individual unit level. For instance, a structure that appears to be a row house, but is split into three individually-owned units belonging to a condominium owner's association, would appear in the Condominium CAMA data set. As such, each condominium unit is considered a 1-unit property. Finally, the Commercial CAMA data set contains properties with commercial structures, including privately-owned apartment buildings.

The CAMA data sets contain numerous pieces of information used to create properties and the housing units within those properties, as well as the attributes of properties and housing units. First, the CAMA data sets contain information about the land parcel, including precise location, current use, current zoning and lot size. Second, the CAMA data sets contain information about each structure on the property, including size (gross building area), the number of units, the number of bedrooms, the number of bathrooms, and the year built. Third, the CAMA data sets contain information about the assessed value of the land and the structure(s). Finally, the CAMA data sets contain the sale date and sale price of the most recent sales.

The CAMA system, like other similar systems, is designed to facilitate the property tax assessment and collection functions of local governments. These systems are not necessarily designed to be current inventories of all housing units or track every important feature of individual housing units. As such, numerous procedures, hereinafter referred to as "cleaning," are required to transform the CAMA data into a "snapshot" of all housing units, which then becomes the housing units in the DCHAS model. These procedures are described in Appendix B. Table 4.2 contains the attributes extracted from each CAMA data set.

Table 4.2: Attributes of the CAMA Data sets

CAMA Attribute	Residential	Condominium	Commercial
Ownership type[22]	Yes	Yes	Yes
Current zoning	Yes	Yes	Yes
Neighborhood	Yes	Yes	Yes
Lot size	Yes		Yes
Gross building/unit size	Yes	Yes	Yes
Number of bedrooms	Yes	Yes	No
Number of bathrooms	Yes	Yes	No
Number of units	Yes (=1)	Yes (=1)	Yes (>=1)
Year built/age	Yes	Yes	Yes
Owner name	Yes	Yes	Yes
Year sold	Yes	Yes	Yes
Sale price	Yes	Yes	Yes
Is a PUD?	Yes	Yes	Yes

4.6.2. Determining Current Cost for Owner-occupied Units

For owner-occupied housing units, housing costs typically include mortgage payment, property taxes, and homeowner's insurance. The CAMA data contains annual property tax amounts for each property, so CAMA information is used to impute annual property taxes for future years. Although sales prices are available in the CAMA data set, there is no readily available data source linking individual CAMA properties to mortgage amount or interest rates. As such, the CAMA data cannot be used to precisely determine mortgage cost. However, a tax history dataset from CoreLogic© was available, and it includes both recent sales price and mortgage amount. This data set is used to estimate mortgage cost as a function of sales price for recently sold properties. On average, the value of the mortgage is 95 percent of the sale price. Lastly, homeowner's insurance costs are not available in the CAMA data set.

[22] One-unit properties in the Residential, Condominium, and Commercial CAMA data sets are all assumed to be single-family owned. All other properties are assumed to be multifamily. This determination is made without regard to what the physical structure resembles.

Finally, some homeowners do not have a mortgage, as they either paid cash for their home or have paid off their mortgage balance. In DCHAS, a share of homeowners has their mortgage cost set to $0. This share comes from the 5-year 2015 ACS PUMS and varies by household typology. The values are available upon request to author.

4.6.3. Determining Rent Control Status

As discussed in Chapters 2 and 3, many privately-owned rental units in Washington, D.C. are subject to rent control. To determine which of the privately-owned units in each CAMA data set are subject to rent control, the following procedures were applied to create a rent control flag:

1. All properties with year built prior to 1976 were given a rent control flag value = "Yes."

2. Each CAMA data set was queried by owner name, then all owner names were combined into a list. The list was queried to determine if the owner owned five or more units. If the owner owned five or more units, each CAMA property owned by that owner was given a rent control flag value = "Yes."

4.6.4. Estimating the Number of Bedrooms for Commercial CAMA Units

As mentioned in section 4.6.1, the Residential and Condominium CAMA data sets generally include one-unit properties, each with a bedroom count. However, the Commercial CAMA data set, which includes properties with one or more apartment buildings, does not include total number of bedrooms on the property or by building.

The number of bedrooms for each rental housing unit from the Commercial CAMA records was imputed using the 5-year 2015 ACS IUF data. In short, all housing units from the 5-year 2015 ACS IUF data set that were not receiving HUD assistance and were in structures with five or more units were used to determine the overall share of units that were studio, one-bedroom, two-bedroom, three-bedroom, and 4 or more bedrooms. These shares were then applied at the property level. The share estimates are available in Appendix A, Table A10.

4.6.5. Estimating the Utility Costs for Each CAMA Unit

Utility costs add to the overall housing cost for a unit. Utility costs are known to vary by household size, the age of structure, the type of heating, and the size of the structure. For DCHAS, utility costs are estimated using data from the 5-year 2015 ACS PUMS. To capture a portion of the known variation in utility costs, average utility cost estimates are calculated by the age of structure (pre-1990, post-1990), size (number of bedrooms), and household size. The estimates are available upon request to the author.

4.6.6. Determining a Property's Maximum Allowable Units by Right

The CAMA data set includes the zoning type for each property. CAMA data is used to estimate the maximum allowable by right units for each zoning type. This is achieved by taking the average number of units per square foot of lot size for each zoning type, then applying that average to all existing properties. However, two considerations are made. First, only properties with buildings constructed in the past 20 years are used in the estimation. This consideration helps to remove properties that do not conform to the current zoning code. Second, properties that were constructed under a PUD are removed, as they are, by definition, not conforming to the current zoning code. The average values are available in Appendix A, Table A11.

4.6.7. Determining a Property's Maximum Allowable Units via Planned Unit Development

In DCHAS, it is assumed the Developer maximizes expected profits by building as many units on a property as the by right development will allow, or can elect to build as many units as a PUD will allow. The Developer choses which properties to build new units based on where expected profits are the highest.

Not all properties are good candidates for a PUD. In fact, a visual examination of PUD projects reveals that they tend to be located on large properties near metro stations. There is undoubtedly endogeneity in the selection of properties on which developers seek to build a PUD. Simply put, developers are not going to submit a PUD application to the D.C. government unless there is a good chance the PUD will be approved, as the application process is costly and time consuming, and developers are aware of what the D.C. government is likely to approve, and more importantly, not approve. As such, there are few PUD applications that are outright rejected, consequently, there is little data on rejections that can be used as a source of variance in a model predicting which properties are likely to be approved. However, by relying only on the characteristics of previously approved PUD properties, it is statistically possible to predict where PUD applications are likely to be approved.

In DCHAS, predictions of the likelihood that a property would have a PUD application approved are made by estimating parameters of a logistic regression model. The dependent variable is whether a property is a PUD. The explanatory variables are property size and distance to a metro station. CAMA data is used to estimate the parameters of the model. The parameter estimates from this model are available in Appendix A, Table A12.

The allowable increase in building area per square foot of land approved under a PUD is calculated from CAMA data. Within the CAMA data, properties that were approved under a PUD are identified, as well as their prior zoning and the new zoning. Using this information, the average increase in building area per square foot of land for PUD projects can be calculated, conditioned on thee prior zoning of the property. The average increase, by zoning type, is listed in Appendix A, Table A11.

4.6.8. Affordable Housing Properties

Affordable housing units are derived from the D.C. Preservation Catalog (DCPC) (Urban Institute, 2017a). The DCPC contains public housing developments, privately-owned developments with project-based rental assistance contracts, and privately-owned developments with LIHTC restrictions. The DCPC contains several pieces of information about each affordable housing development, including the number of units, the type of subsidy or program, and the subsidy expiration date. The DCPC also contains the precise location of each development and the CAMA parcel number(s) on which the development is sited.

It was assumed that the DCPC is the most accurate inventory of affordable housing developments available. However, the CAMA data sets include all residential properties, including publicly and privately owned affordable housing developments. As such, simply *appending* the DCPC developments to the properties in the CAMA data sets would result in duplicate representations of these properties. Fortunately, the process of *merging* DCPC data with the CAMA data sets was a simple process because both the CAMA and DCPC data sets included parcel numbers. The merging process did not create any new properties per se. Rather, the merging process resulted in edits or additions of subsidy information to the properties already contained in the CAMA data sets. Specifically, the property type becomes subsidized and the number of units is set equal to DCPC counts, and a subsidy expiration year is added.

For affordable housing units in public housing developments or privately-owned developments with project-based rental assistance contracts, the occupying household pays one-third of their income as rent, with HUD paying the other two-thirds. As such, no rent needs to be established. For affordable housing units in LIHTC developments, rents are calculated as one-third of 60 percent of AMI, adjusted by household size. AMI values by household size were extracted from HUD's Income Limits website (United States Department of Housing and Urban Development, 2017).

4.6.9. Estimating Bedroom Counts for Affordable Housing Units

The bedroom counts for affordable housing units are, generally speaking, public information available from either HUD or the D.C. Housing Authority. However, matching public information to the DCPC data would be a time-consuming process. As such, for DCHAS, bedroom counts for affordable units were imputed using the average bedroom counts across all properties. This information is available from HUD's *Picture of Subsidized Housing* (United States Department of Housing and Urban Development, 2017). The average values are available in Appendix A, Table A10.

4.7. Final List of Model Attributes, Event Parameters, and Data Sources

Tables 4.3 through 4.8 provide a list of the attributes and event parameters, data sources for the attributes and event parameters and the attribute or parameter estimation technique (if applicable). There is one Table each for households, landlords, the Developer, the D.C. government, properties, and housing units, respectively.

In these tables, the attributes are those that are either fixed throughout the course of the simulation or are changed by events occurring in the model. For instance, a household's relationship structure can only be changed by a probabilistically triggered life cycle event controlled by marriage rate, divorce rate, birth rates, or death rates. Non-adjustable event parameters are fixed at the start of the simulation and do not change throughout the simulation. Adjustable event parameters are adjustable by the modeler at the start of the simulation, but remain the same throughout the simulation.

Table 4.3: Household Member Attributes and Household Event Parameters, Source, and Estimation Technique

Attribute or Event Parameter	**Source of Data**	**Estimation Technique**
Attributes		
Age, sex, race, relationship, and income	SynPop	Synthetic population
HUD subsidy type	5-year 2015 ACS/HUD IUF	MNL regression with parameter estimates applied to SynPop households
Non-adjustable Event Parameters		
OutMigrationProb	5-year 2015 ACS PUMS	Logistic regression with parameter estimates applied to SynPop households
Income at first job, income change rate, marriage rate,	5-year 2015 ACS PUMS	Averages by typology

divorce rate, birth rate, and cast-off rate		
DeathRate	D.C. Government	Average
MoveRate	5-year 2015 ACS PUMS	Average by typology
FormationRate	5-year 2015 ACS PUMS	Average by typology
PurchaseRate	5-year 2015 ACS PUMS	Average by typology
IncomeShare	5-year 2015 ACS PUMS	Simple quantile by typology
Appraisal (δ, ρ)	CAMA	Spatial autoregressive model, with parameter estimates applied to CAMA properties
Location decision (α, β, γ)	5-year 2015 ACS IUF	MNL logistic regression, with parameter estimates applied to CAMA housing units
Adjustable Event Parameters		
InMigrationCount	Census Population Estimates Program	Average
OutMigrationCount	Census Population Estimates Program	Average

Table 4.4: Landlord Attributes and Event Parameters, Source, and Estimation Technique

Attribute or Event Parameter	**Source of Data**	**Estimation Technique**
Adjustable Event Parameters		
PropInvestmentRate	2012 RHFS	Average
Non-adjustable Event Parameters		
Rental price estimation (μ, σ)	5-year 2015 ACS IUF	Spatial autoregressive model with parameter estimates applied to CAMA housing units

Table 4.5: Developer Attributes and Model Parameters, Source, and Estimation Technique

Attribute or Event Parameter	**Source of Data**	**Estimation Technique**
Adjustable Event Parameters		
DevelopmentRate	BPS	Exact count
Non-adjustable Event Parameters		
Expected revenue estimation (same as appraisal process)	CAMA	See Appraisal in Household Table
Property cost estimation (λ, ν)	CAMA	Spatial autoregressive model, with parameter estimates applied to CAMA properties
Construction costs estimation	RSMeans®	Average

Table 4.6: D.C. Government Attributes and Model Parameters, Source, and Estimation Technique

Attribute or Event Parameter	**Source of Data**	**Estimation Technique**
Adjustable Event Parameters		
SubUnitInvestment	DC OCTO's 10x20 database	Exact count
RehabShare	DC OCTO's 10x20 database	Exact count
PUDUnitsApproved	CAMA	Exact count

Table 4.7: Property Attributes and Model Parameters, Source, and Estimation Technique

Attribute or behavior parameter	Source of data	Estimation technique
Attributes		
Market rate properties location, type, and rent control flag	CAMA	Exact
Market rate properties by right maximum units	CAMA	Average
Market rate properties PUD maximum units	CAMA	Logistic regression model with parameter estimates applied to CAMA properties
Sale Price	CAMA	Exact

Table 4.8: Housing Unit Attributes and Model Parameters, Source, and Estimation Technique

Attribute or behavior parameter	Source of data	Estimation technique
Attributes		
Bedrooms	CAMA or 2011 ACS/HUD special file	Exact count for CAMA; Averages for ACS/HUD special file
Share with \$0 mortgage cost	5-year 2015 ACS PUMS	Average
Effective age	CAMA	Exact value
Rental rate	5-year 2015 ACS PUMS	Spatial autoregressive model, with parameter estimates applied to CAMA properties
Mortgage cost	5-year 2015 ACS IUF and CoreLogic®	Linear regression
Utility cost	5-year ACS PUMS	Average
Homeowner's insurance cost	5-year ACS PUMS	Linear regression

4.8. Final List of Events

Table 4.9 provides a list of the 17 events in DCHAS and which agent or agents participate in the event. Operationally, the events occur in the order they appear in the Table. The first two events each of sub-events.

Table 4.9: DCHAS Events

Event	Participating Agents
1. Migration (two sub-events)	Households
2. Life cycle (six sub-events)	Household members
3. Mobility and Formation (two sub-events)	Household members and households
4. Tenure selection	Households
5. Subsidized renter selection and placement	Subsidized housing landlord
6. Appraisal	Households
7. Location selection	Households
8. Price negotiation and market clearing	Households
9. Rental price setting	Market rate landlords
10. Rental market clearing	Market rate landlords and households
11. Capital investment	Market rate landlords
12. Expected revenue calculation	The Developer
13. Property cost appraisal	The Developer
14. Construction cost calculation	The Developer
15. PUD approval	D.C. Government
16. Market rate development	The Developer
17. Affordable housing development	D.C. Government

CHAPTER 5: MODEL INITALIZATION, CALIBRATION, VERIFICATION AND VALIDATION

This purpose of this chapter is to describe how DCHAS is initialized and calibrated to the year 2010 and to describe the verification and validation steps undertaken to demonstrate that DCHAS's 17 events function as intended, reproducing results than match known outcomes in 2015.

Section 5.1 explains the initialization and calibration to the year 2010. The initialization and calibration phase includes household members and households (section 5.1.2), affordable housing units (section 5.1.3), market rate housing units (section 5.1.4), as well as estimating initial market-rate rents (section 5.1.5)

DCHAS events 3 (mobility) through 17 (affordable housing development) assume that households are occupying housing units. Because the households and housing units come from a different data sources, there is no "built-in" initial relationship between households and housing units. As such, section 5.2 explains how DCHAS households are initially placed into housing units, including subsidized renters (section 5.2.2), owners (section 5.2.3), and market-rate renters (section 5.2.4). Section 5.3 describes the programs process for creating DCHAS event parameter estimates, including the names of the programs (code) used to produce the estimates.

With households placed into housing units, and with DCHAS event parameters estimates, simulation can proceed. Section 5.4 describes the verification and validation process and results for the first two DCHAS events, migration and life cycle. A significant amount of attention is devoted to the migration and life cycle events for two reasons. First, they represent DCHAS's most complicated processes, from an operational perspective. Second, the migration and life cycle events can operate without having to "turn on" subsequent events (events 3 through 17) in the model. Section 5.4.1 described the general operation of the migration and life cycle events; sections 5.4.2 and 5.4.3 describe verification steps and results; section 5.4.4 describes validation steps and results.

Section 5.5 is devoted to verification and validation of the mobility, formation, and tenure assignment events. Since these events are not operationally complicated, a joint verification and validation test and results are presented. Section 5.6 is devoted to the verification and validation of the remaining events, while section 5.7 presents conclusions about initialization, calibration, verification, and validation.

DCHAS is programmed using SAS. References to SAS programs that accomplish the tasks described in this chapter are included in appropriate places. The SAS programs referenced in the chapter are available upon request to the author.

5.1. Initialization and Calibration of Households and Housing Units

5.1.1. Initialization and Calibrating Households Members and Households

The initialization of household members and households is relatively straightforward. As discussed in section 4.1.2, DCHAS household members and households are derived from the SynPop data set (Wheaton, 2010 U.S. Synthetic Population Ver. 1, 2014), which itself is based on the 2010 Decennial Census population count. It is generally accepted by Federal statistical agencies that the 2010 Decennial

Census counts of people and housing units are accurate. Therefore, the count of household members and households in the SynPop data set is accepted as the best available representation of 2010. The SAS programs initializing household members and households are called “Prepare GIS Data.sas” and “Household Initialization.sas.”

The second step to initialize household members and households is to determine which households are owners, non-subsidized renters, or renters receiving a Federal housing subsidy, and from which program (public housing/project-based section 8 or HCVP), including calibrating the household counts to independent totals from HUD and the 2010 Decennial Census. This process was discussed in section 4.1.2. The SAS program accomplishing this step is called “Tenure Initialization.sas.”

5.1.2. Initializing and Calibrating Affordable Housing Units in the DCPC Data Set

As mentioned in section 4.5.8, the source of affordable housing properties is the DCPC data set, which includes information on the number of housing units and subsidy type. The DCPC is assumed to be the most comprehensive accounting of affordable properties available[23]. The initialization of affordable housing units to 2010 requires several steps.

The first step is to initialize the properties to 2010. The DCPC data used in the analysis is “as of” February 2017. However, DCPC properties with an original “in service” date after 2010 are removed. The second step is to determine which affordable housing properties are HUD-subsidized (public housing or project-based section 8) and

[23] The assumption is based on the author’s expert knowledge of sources of affordable housing data, as well as knowledge of the process to create DCPS.

which are LIHTC, and in the process of making the determination, to calibrate the HUD-subsidized unit counts to independent totals. In most cases the HUD-subsidized determination is straightforward. However, the total units in DCPC properties identified as HUD-subsidized is less than a 2010 independent count of HUD-assisted households in public housing or project-based section 8 from HUD's Picture of Subsidized Housing (available in Appendix A, Table A1). To calibrate the DCPC data to 2010 HUD-assisted households counts, a random number of DCPC properties are designated as HUD-subsidized until the total unit count in HUD-subsidized DCPC properties matched the independent 2010 HUD-assisted households count[24]. The final HUD-subsidized unit count is 22,859.

It is assumed that all properties not designated as HUD-subsidized are designated as LIHTC. In fact, the DCPC data designates most of these properties as LIHTC properties. For the few that are not HUD-subsidized and not LIHTC (these properties are receiving subsidies directly from the D.C. government), it is assumed that their rents are the same as LIHTC rents (described in section 4.6.8). The final LIHTC count is 12,660.

Finally, the DCPC data, initialized and calibrated to 2010, is likely a slight undercount of the total number of affordable housing units. The DCPS data set does not include information enabling determination of affordable housing properties taken out of service between 2010 and 2017. An alternative data source for affordable properties taken out of service could not be located.

---

[24] The assumption in DCHAS is that the number of HUD-assisted households equals the number of HUD-subsidized units. In pratice, there are more HUD-subsidized units than HUD-assisted households, and those units are temporarily vacant. Like the market-rate housing market, the affordable housing market has a small, but naturally occurring, vacancy rate.

5.1.3. Initializing and Calibrating Total Market-Rate Housing Units in the CAMA data set

According to the 2010 Decennial Census, there were 296,719 housing units (occupied and vacant) in Washington, D.C. as of April 2010. Initialization and calibration of affordable housing units to the year 2010 yielded 35,519 units. Therefore, the target count of market-rate units in 2010 is 261,200.

While Appendix B contains details of the CAMA cleaning process, there are two important steps to mention regarding initializing the CAMA data. In the first step, all properties constructed after 2010 are removed. In the second step, properties in the residential portion of the CAMA data set with more than one kitchen (as denoted by the variable *Kitchen*) are subdivided into two or more units, equal to the number of kitchens. This step resulted in the creation of more than 23,000 housing units. The net result of the two steps is 274,854 CAMA housing units, or 13,654 more units that the target count of 261,200 market-rate units.

Calibration of the CAMA data set's total housing units to the 2010 Decennial Census is conducted by Ward. The calibration focuses on removing units in properties with two or more units. It is assumed that the unit count for properties with one unit is, in all likelihood, correct. Therefore, the additional 13,654 units are assumed spread among the properties with two or more units. To calibrate to housing unit totals by Ward, random properties with two or more units are selected and their unit count is reduced by one. The process is repeated until the target number of market-rate housing units (261,200) is reached.

Table 5.1 below shows the final housing unit counts by ownership type. The SAS program accomplishing the initialization and calibration of affordable and market-rate housing units is called "Housing Unit Initialization.sas."

Table 5.1: Housing Unit Count by Ownership Type

	Housing Unit Count
Tenure Type	
Public Housing or Project-based Section 8	22,859
Low-income Housing Tax Credit	12,660
Single-family	136,432
Small multifamily (2-4 units)	33,651
Large multifamily (5+ units)	91,117
Total	296,719

5.1.4. Initializing 2010 Rents

Section 3.4.2 described the model landlords use to form asking rents, while section 4.3.1 describes the data source used to estimate the parameters of the model. The model specification includes a spatial autoregressive term, which itself requires existing rent values be available for all units before the parameter on the autoregressive term can be estimated. This is typically not an issue in a model of housing prices because most houses have a recent sale price available in the CAMA data. However, CAMA data does not include rents.

To form initial 2010 rents for all DCHAS market-rate housing units, the model described in section 3.4.2 is used, but without the spatial autoregressive term. Then, the parameter estimates from the model are applied to DCHAS housing units such that all housing units are given a predicted rent. Then, the predicted rents are used to estimate the parameters of the full model specification in section 3.4.2, which includes the spatial autoregressive term.

For LIHTC units, the asking rents are established based on LIHTC program procedures. In short, the asking rents are based on one-third of 60 percent of the 2010 area median income, adjusted for the number of bedrooms. The SAS program initializing 2010 rents is called "Rent Initialization.sas."

5.2. Initial Placement of Households into Housing Units

Chapter 3 discussed how *existing* households make their decision when to move, whether to be owners or renters, and where to move. Then, Chapter 4 discussed the data sources used to estimate the event decision parameters for existing households. However, chapters 3 and 4 did not discuss how the DCHAS households are *initially* placed into DCHAS housing units. This section describes the initial placement process.

5.2.1. Initial Placement Process: Subsidized Households

The first step in the initial placement process is to place public housing or project-based section 8 households into public housing or project-based section 8 units, respectively. By design, the placement is straightforward because the number of public housing or project-based section 8 households exactly matched the number of public housing or project-based section 8 units[25]. The placement occurred by Ward and is designed to ensure that larger households are placed into larger units.

[25] In practice, there are vacant subsidized units. However, in Washington, D.C., there is a long waiting list of households requesting a subsidized unit. Whatever subsidized vacancy rate exists is simply an artifact of the administration of the program. As such, for ease of modeling, it was assumed that all subsidized units are occupied.

The second step in the initial placement process is to place HCVP households into rental units. It is well-known that there is a significant number of HCVP holders who occupy LIHTC units. This is often referred to as “subsidy layering” and is a perfectly legal use of LIHTC units. However, HUD does not have an accurate count of the number of HCVP recipients occupying LIHTC units because they do not administer the program. One source of information that can be used to create an estimate is the 2012 Rental Housing Finance Survey, which shows that more than two-thirds of LIHTC properties have more than 50 percent of their units occupied by a HCVP household. To reflect subsidy layering, it is assumed that 60 percent of the LIHTC units are occupied by HCVP households. Operationally, about 7,600 HCVP households are placed into a LIHTC unit. This placement occurred by Ward and bedroom size.

To place the remaining HCVP households into non-LIHTC units, two assumptions are made. First, HCVP can be placed only into units where the market rent (established in section 5.1.4) is less than HUD’s Fair Market Rent (FMR) for the unit (based on bedroom count). Second, HCVP households cannot overcrowd a unit, meaning that they cannot be placed in to a unit with an insufficient number of bedrooms. The remaining HCVP households are randomly placed into units meeting the two conditions, by Ward. The SAS program performing the initial placement of subsidized tenants is called “Subsidized Household Initial Placement.sas.”

The third step in the initial placement phase is to place very low-income and low-income renters in to the remaining LIHTC units. To do this, a random non-subsidized renter household is chosen, their bedroom needs are measured, and they are placed into an empty LIHTC unit. This process is repeated until all LIHTC units have a renter. The SAS program performing the initial placement of very low income and low-income renters in to empty LIHTC units is called "LowInc to LIHTC Initial Placement.sas."

5.2.2. Initial Placement Process: Owners

The fourth step in the initial placement process is to place owner households into single-family owned housing units. As discussed in Chapter 4, single-family owned housing units are individually-deeded, and include detached and attached housing structures with one unit, as well as condominiums in buildings that structurally resemble multifamily buildings. By design, owner households cannot occupy units in small or large multifamily properties because those units are, by definition, rentals.

The model described in section 3.3.6 is appropriate for determining a new residential location for an existing household (i.e., a household already placed in a housing unit) when there is a small number of possible choices (homes for sale) for the household to relocate. However, it is not necessarily an appropriate framework for initially placing over 100,000 simulated households into housing units because it is too computationally intensive. To get a sense of why, consider the scenario where parameter estimates from the model in equation 3.1 are available. To place a DCHAS household into a housing unit, the process is:

1. Select a DCHAS household.
2. Select a sample of *n* DCHAS housing units from the available universe of housing units, where *n* is large enough to ensure an adequate cross-section of options.
3. Apply the parameter estimate to the combination of the DCHAS household placed into

the DCHAS housing unit.

4. Select the housing unit with the highest utility (probability).

5. Remove this housing unit from the available universe.

6. Repeat this process 112,000 times (once for each DCHAS household).

Another option is to randomly place DCHAS households in to housing units, perhaps based on Census tracts. A third option, and the option selected for DCHAS, is to use a more generalized model *motivated* by the random utility framework, but with some simplifying modifications. As discussed in chapter 4, the 5-year 2015 ACS is a survey of households and their housing units, so it is a good source of data for modeling how different types of households select different types of housing units. Thus, the third step in the initial placement process - the placement of owners - proceeds as follows,

1. Categorize all 5-year 2015 ACS housing units into one of 54 unique categories (HTYPE) based on four variables: *Value* (3 quantiles), *Distance to Metro* (2 quantiles), *Bedrooms* (1, 2, 3+) and *Tract Median Household Income* (3 quantiles).

2. Categorize all DCHAS housing units using the same categories.

3. Using the 5-year 2015 ACS, estimate the parameters of a MNL model with the following specification,

$$HTYPE = Household\ Income + Household\ Size + Presence\ of\ Children \\ + Age\ of\ Householder$$

4. Apply the parameter estimates from the above MNL to DCHAS households to predict their probability of (preference for) being in each of the 54 types of housing units (called

HTYPE). This is feasible because DCHAS includes each of the independent variables in step 3.

5. By Ward, place each DCHAS household into a DCHAS housing unit based on their highest preferred *HTYPE* until there are no more units of that *HTYPE* available within the Ward.

6. Repeat step 5 for unplaced household based on their second, third, etc., preference for *HTYPE*, until all households are placed.

The SAS program performing the initial placement of owner households is called "Owner Household Initial Placement.sas."

5.2.3. Initial Placement Process: Market-rate Renters

The fifth step in the initial placement process is to place the remaining renters in to housing units. The initial placement of market-rate renters in to housing units proceeds in nearly an identical fashion as the initial placement of owners described in section 5.2.2. However, there are two notable differences. First, whereas owner households are restricted to being placed in single-family owned housing units, renters are placed in any type of remaining unit. Second, with respect to the four housing characteristics that contribute to the formation of *HTYPE*, the V*alue* characteristics in the owner initial placement process is replaced with *Rent* for the renter placement process. As discussed in section 5.1.4, each housing unit in DCHAS is given an estimated rent. The SAS program performing the initial placement of market-rate renter households is called "Market-rate Renter Household Initial Placement.sas."

5.2.4. A Note About Stochasticity in the Initialization Steps

In the initialization and calibration steps described in sections 5.1 and 5.2, there are a handful of instances where randomization is introduced. For instance, a random number of subsidized properties had their subsidy type changed from LIHTC to HUD-subsidized. It is important to note that in each instance of randomization, the randomization is designed to be fully replicable. In other words, if the initialization and calibration steps are run multiple times, they will always yield the same set of households with the same characteristics; the same set of properties and housing units with the same set of characteristics; and the same households being placed into the same housing units.

The purpose of designing DCHAS in this fashion is to reduce the amount of stochasticity resulting from initialization steps that are not important to the results of the full model.

5.3. Event Parameter Estimation

Before DCHAS events can be executed, several event parameter estimates must be created. The creation of event parameter estimates proceeds in a series of nine steps, as follows:

1. Calculate the six life cycle event's parameter estimates. The SAS program accomplishing this step is called "Life Cycle Parameter Estimation.sas."

2. Calculate the migration event parameter estimates and in-migrants "donor" dataset. The SAS program accomplishing this step is called "Migration Parameter Estimation.sas."

3. Calculate the mobility, tenure, and formation parameter estimates. The parameters are estimated in the SAS program "Mobility, Formation and Tenure Selection Parameter

Estimation.sas."

4. Calculate the single-family appraisal model parameter estimates. The parameters are estimated in the SAS program "Single-family Appraisal Model Parameter Estimation.sas."

5. Calculate the rental appraisal model parameter estimates. The parameters are estimated in the SAS program "Rental Appraisal Model Parameter Estimation.sas."

6. Calculate the location choice model parameter estimates. The parameters are estimated in the SAS program "Location Choice Model Parameter Estimation.sas."

7. Calculate the appraisal model parameter estimates. The parameters are estimated in the SAS program "Land Appraisal Model Parameter Estimation.sas."

8. Calculate the average building size per square foot of lot, by zoning type parameter estimates. The parameters are estimated in the SAS program "Average Building Size by Zoning Type Estimation.sas."

9. Calculate the PUD probability model parameter estimates. The parameters are estimated in the SAS program "PUD Probability Model Parameter Estimation.sas."

5.4. Verification and Validation of Migration and Life Cycle Events

With properly initialized DCHAS households (described in section 5.1) placed in to DCHAS housing units (described in section 5.2), the migration (#1) and life cycle (#2) events are initiated to verify that they work as intended. This is accomplished without invoking any other DCHAS event. The SAS program for running the migration and life cycle events is called "Migration and Life cycle Events.sas."

5.4.1. DCHAS Migration and Life Cycle Event Sequence

The base year 2010 household members and households experience simulated migration and life cycle events based on the in and out-migration parameters described in 4.2.1 and the six life cycle parameters described in 4.2.2. For purposes of modeling, the sub-events within the migration and life cycle events occur in a sequence during a simulated month. The order of migration and life cycle sub-events is:

1. Out-migration
2. In-migration
3. Birth
4. Death
5. Marriage
6. Divorce
7. First job income
8. Income growth
9. Aging

5.4.2. Verification of Migration and Life Cycle Events: Distributional Equivalence

As discussed in Rand and Wilensky (2006) and Axelrod (1997), replication is the most important aspect of verification. Axelrod (1997) described three replication criteria: numerical identity, distributional equivalence, and relational alignment. Because DCHAS is not based on an analytical (equation-based) model with a known solution, only the latter two criteria are relevant.

Distributional equivalence occurs when two models produce results that are equivalent, where equivalency can be defined based on a single statistic or a statistical distribution. While there is no benchmark model for which to compare to DCHAS, this criterion is still useful for verifying that DCHAS migration and life cycle events are functioning correctly by comparing several runs of DCHAS. This is possible because the parameters that control the migration and life cycle events are exogenous to the model and do not change. Therefore, repeated runs of DCHAS without interference from the other 15 events should produce migration and life cycle outcomes that are distributionally equivalent to prior runs.

To understand why this is the case, recall that migration counts as well as the birth, death, marriage, and divorce rates are all based on point statistics calculated using data from 2011 through 2015. Although household members are randomly chosen to experience a migration or life cycle event (e.g., death), the total *share* of household members experiencing the event is based on the respective rates. For instance, if a household member typology has a death rate of one percent, then one percent of the members of that typology will be selected to die in every time step. As such, the total share of household members experiencing the event will not change as the demographic processes are repeated numerous times. The only source of stochasticity introduced into the migration and life cycle events comes from the random selection of a certain percentage of household members to experience the event.

To test this claim, five runs of the DCHAS migration and life cycle events are made and each run is for 36 time steps (months). Each run uses the default rates for in-migration (64,930 households per year) and out-migration (57,710 households per year). The results in Table 5.2 below reveal a nearly identical number of household members (less than 0.2 percent difference between the lowest and highest values), as well as an identical share of householder members who are male and share of members who are less than five years old. These results suggest that, from a computational perspective, DCHAS operates as intended.

Table 5.2: Distributional Equivalence Analysis of DCHAS Migration and Life cycle Events

	Total household members	**Percent of household members that are male**	**Total household members that are less than five years old**
Replication #1	598,500	46.5	6.1
Replication #2	597,600	46.5	6.1
Replication #3	598,000	46.5	6.1
Replication #4	597,300	46.5	6.1
Replication #5	597,900	46.5	6.1
Average	**597,900**	**46.5**	**6.1**

5.4.3. Verification of Demographic Events: Relational Alignment

Relational alignment occurs when two models show the same relationship between input and output. For instance, if an increase in *X* leads to an increase in *Y* in one model, the same relationship should hold for another model that is purported to be the same as the first model. In the case of DCHAS's migration and life cycle events, a change in one of the input variables (e.g., birth rates) should lead to a change in an output variable (e.g., total household members), and the direction and magnitude of this change should be consistent across successive runs of the model.

To test this claim, four of the events that directly impact the number of household members (in-migration, out-migration, birth, and death) are individually altered by doubling their respective event rates. For each event, three replication runs are made with the altered (doubled) rate for 36 time steps, and the total household members generated by the three runs are compared to a baseline total number of household members, which is the average number of household members from Table 5.2.

The results, presented in Table 5.3 reveal consistent directional change in each of the three replication runs, compared to the baseline. This suggests that, from a computational perspective, the four events that directly impact number of household members are functioning as intended.

Table 5.3: Relational Alignment Analysis of DCHAS Migration, Birth, and Death Events

	Double Rate of In-Migration	**Double Rate of Out-Migration**	**Double Rate of Birth**	**Double Rate of Death**
	Total household members			
Baseline run* (default rate)	597,900	597,900	597,900	597,900
Run #1	757,800	456,900	618,200	586,200
Run #2	757,800	456,400	621,200	587,400
Run #3	757,800	457,400	620,400	568,000

*The baseline run is the average from five replications generated from the distributional equivalence analysis.

The two other life cycle events, marriage and divorce, do not *directly* impact the number of household members[26]. However, marriage and divorce impact the number of households. To verify that these two events are operating as intended, the same "double the rate" strategy is applied, but the outcome of interest is the number of households, rather than the number of household members. As shown in Table 3.4, a doubling of the marriage rate should lead to a slight decrease in the number of households because the source of household members for *some* marriage is two single-person households. A doubling of the divorce rate should lead to an increase in the number of households.

To test this claim, marriage and divorce are individually altered by doubling their respective event rates. For each event, three replication runs are made with the altered (doubled) rate for 36 time steps, and the total number of households generated by the three runs are compared to a baseline total number of households, which is the average number of households from the runs that produced Table 5.2.

The results presented in Table 5.4 reveal consistent directional change in each of the three replication runs, compared to the baseline. This suggests that the marriage and divorce life cycle events are functioning as intended.

Table 5.4: Relational Alignment Analysis of DCHAS Marriage, Divorce, and Cast Off Sub-events

	Double Rate of Marriage	**Double Rate of Divorce**
	Total households	

[26] Marriage and divorce do *indirectl*y impact the number of household members because women who are married have a higher birth rate than single women.

Baseline run (default rate) *	277,600	277,600
Run #1	276,000	284,000
Run #2	276,700	284,000
Run #3	276,500	284,000

5.4.4. Validation and Calibration of Migration and Life Cycle Events

Results described in sections 5.4.2 and 5.4.3 suggest that, from a computational perspective, the DCHAS migration and life cycle events are functioning as intended. Repeated runs of the migration and life cycle events produce the same results, which is expected. However, the question remains as to whether the results are consistent with statistical evidence. In the case of DCHAS, the in- and out-migration event rates are derived from the United States Bureau of Census Population Estimates Program and the life cycle event rates (except for death) are derived from the 5-year 2015 ACS PUMS. Thus, it stands to reason that a correctly operating DCHAS should produce results for model runs through 2015 that align with estimates from the Census Population Estimates Program and the 2015 ACS.

To test this claim, four key statistics are evaluated: total household members, total percent of females that are married, median age and median income. DCHAS is run for 63 time steps. This number of time steps is chosen specifically because it represents the number of months between the official 2010 Decennial Census population count (April 2010) from which DCHAS is initialized and the official population estimate for July 2015 produced by the Bureau of Census.

Several initial runs of migration and lifecycle events were made to investigate the results and perform comparisons to independent estimates. The initial runs revealed some issues. First, DCHAS was not producing enough household members. Second, DCHAS was not producing enough married females. To overcome these two issues, three calibration adjustments were performed. First, the birth rates from the ACS were increased by 10 percent. Second, the marriage rate was increased by 20 percent. Third, the income growth rate was reduced by five percent.

Table 5.5 compares the average results from five calibrated runs of DCHAS with 2015 estimates from the relevant data source. The results show the DCHAS migration and life cycle events accurately reproduce total number of household members and the median age. The model produces fewer married females than the independent estimate from the ACS. However, this could be explained by the fact that DCHAS's population data source, SynPop, does not capture what are referred to as subfamilies. An example of a household with a subfamily is a husband and wife who share a home with the wife's parents. In SynPop, it is easy to identify if the householder (the husband) is married, but far more difficult to identify that the two older adults, who are his wife's parents, are in fact married. Finally, the median person income from the simulations is higher than the independent estimate from the ACS. However, as shown in Table 5.6, this doesn't seem to cause household income to be higher than observed statistical outcomes.

Table 5.5: Validation of Migration and Life Cycle Events

Metric	**DCHAS Estimate**	**2015 Population Estimates Program or ACS Estimate**
Total number of household members	628,500 – 629,400	630,556[27]
Percent of females >15 years old that are married	25.7 – 25.9	29.0[28] (1.2)
Median age	34	33.8[29] (0.2)
Median person income (includes all persons)	41,200 – 41,400	37,000[30] (1,115)

5.5. Verification and Validation of Mobility and Tenure Events

The mobility and formation (#3) and tenure selection (#4) events are driven by the exogenous parameters *MoveRate, FormationRate* and *PurchaseRate*. Like the migration and life cycle events, mobility, formation and tenure selection can occur without needing to turn on the rest of the relocation process events. This is because the mobility, formation and tenure events are controlled by exogenous parameters and are not influenced by actual housing location or other market forces. Moreover, because the operations of these two events are relatively simple, the verification and validation are performed jointly by comparing the results of five DCHAS runs against observed outcomes based on a small number of metrics. The SAS program implementing the

[27] The SynPop household members (not in group quarters) total (560,416) was less than the final enumerated total from the 2010 Decennial (561,702). Because the 2010 Decennial population counts are the basis for future estimates, the official 2015 Population Estimates population count, less people in group quarters (632,033) was deflated by the ratio of 2010 SynPop to 2010 Decennial (99.7%) to arrive at 630,556.

[28] American Fact Finder Table S1201 for 2015 1-year ACS Data

[29] American Fact Finder Table S0101 for 2015 1-year ACS Data

[30] Author's analysis of 1-year 2015 ACS PUMS. The estimate is based on all persons 18 years or older and includes persons with $0 or less than $0 income.

mobility, formation and tenure events is called “Mobility, Formation and Tenure Selection Events.sas.”

Several initial runs of mobility, formation, and tenure events were made to investigate the results and perform comparisons to independent estimates. The initial runs revealed that the model was producing too many households, as well as too few owner households. To overcome this issue, the formation rates derived from the ACS were reduced by 50 percent while the purchase probability was increased by 10%.

Table 5.6 presents results for five metrics from five DCHAS calibrated runs of 63 time steps (for the same reasons discussed in section 5.4.4) and compares the results to estimates of the same metrics derived from the 2015 ACS. The results show very close agreement between DCHAS and the 2015 ACS. It can safely be concluded that DCHAS mobility, formation and tenure choice events are functioning as intended.

Table 5.6: Validation Mobility, Formation and Tenure Events

Metric	**DCHAS estimate for 2015**	**2015 ACS Estimate**
Number of households	281,600 – 282,000	281,787[31] (3,030)
Median household income ($)	75,700 – 76,000	75,628[31] (2,493)
Percent of people that have moved in past year	18.0 - 18.1	20.8[32] (1.1)
Percent of households that been in their home less than five years	53.7 – 53.8	41.5[33]
Percent of households that own their home	39.5 - 39.6	39.9[34] (1.3)

[31] American Fact Finder Table B19013 for 2015 1-year ACS Data
[32] American Fact Finder Table B07003 for 2015 1-year ACS Data
[33] American Fact Finder Table B25083 for 2015 1-year ACS Data
[34] American Fact Finder Table B25003 for 2015 1-year ACS Data

5.6. Verification and Validation of the Demand Events

5.6.1. Overview

Sections 5.4 and 5.5 discussed the first four events, each of which could be subject to verification and validation tests without running the full set of DCHAS events. The next set of events are the housing demand events, beginning with the subsidized renter selection and placement event (#5) and ending with the rental market clearing event (#10). These events are not *individually* subject to verification because these events represent the full complement of demand-side transactions that result in endogenous changes in housing prices and rents. However, the *cumulative* outcomes of these five events are subject to verification tests.

The cumulative results of the demand event, by themselves, cannot be subject to validation test against observed outcomes because observed market outcomes are the culmination of both demand and supply events. However, validation can be performed against a hypothetical market outcome: no additional construction of housing. All else being equal, the default DCHAS in-migration, out-migration, and formational event parameters will result in an increasing number of households. This increase translates to an increase in demand and an increase in demand without a change in supply should result in price increases and reduced vacancy rates. The SAS program for running the migration and life cycle events is called "Demand Events.sas."

5.6.2. Verification of Demand Events

The demand events result in endogenous changes to sales prices and rents. To verify that the demand events are working as intended, a distributional equivalence is performed, similar to the analysis used for the migration and life cycle events.

For the distributional equivalence analysis, five runs are made with the default rates of in- and out-migration. Each run is 12 time steps (months). For each run, the outcomes of interest are median price, median rents, and share of renter households spending more than 30 percent of their income on rent. The first two are chosen because they indicate the events for appraisal, price adjustments, exchange and market clearing are working correctly. If repeated runs on DCHAS under the same market conditions produce substantial variation in the median price or rents, it would suggest that one or more of the events are not working correctly. The later outcome is chosen because it signifies the location select event (i.e., the utility calculation) is working correctly. If households are randomly choosing housing units without regard to preference or income, then it should manifest itself in variation in the share of income devoted to housing costs.

Table 5.7 presents the results of the distributional equivalence analysis. The results show good consistency through successive runs on the model. This suggests DCHAS demand events are functioning as intended.

Table 5.7: Distributional Equivalence Analysis of DCHAS Market Demand Events

	Median Housing Sales Price for Houses Sold in Past 12 Months	**Median Rental Rate for Market Rate Units**	**Share of Renter Households Paying More than 30 Percent of Income for Housing**
Replication #1	378,500	1,187	45.5

Replication #2	380,500	1,195	45.4
Replication #3	377,000	1,174	45.8
Replication #4	380,000	1,175	46.0
Replication #5	376,000	1,193	45.8
Average	378,400	1,185	46.0

5.6.3. Validation of the Demand Events

The validation of the demand event follows the same process as the verification of the demand events, with one slight tweak. Instead of running DCHAS for 12 time steps, DCHAS is run for 63 time steps. As with the verification of demand events discussed in the previous section, the validation of the demand events is based on a hypothetical scenario of no additional housing construction, but continued population grown. Whereas running DCHAS for 12 months without additional housing construction and continued population growth may have a modest effect on prices, running DCHAS for 63 time steps should produce observable changes in demand, and hence, change in prices.

Table 5.8 presents the results of the validation test for four metrics: median sales price, median rental rate, share of renter households paying more than 30 percent of their income for housing and vacancy rate at the end of the year. The table includes one row for each year. The results show growth in sale price, rents, and shares of renter households paying more than 30 percent of their income on housing, as well as a decrease in vacancy rate. These results suggest DCHAS market demand events are functioning as intended.

Table 5.8: Validation of DCHAS Market Demand Events

	Median Housing Sales Price for Houses Sold This Year	Median Rental Rate for Market Rate Units this Year	Share of Renter Households Paying More than 30 Percent of Income for Housing This Year	Vacancy Rate as of End of Year
2011	378,100	1,180	45.7	9.9
2012	394,500	1,290	48.0	9.0
2013	410,500	1,400	50.1	8.2
2014	438,000	1,480	52.4	7.3
2015	459,000	1,550	54.0	6.3

5.7. Verification and Validation of Full Set of Events

5.7.1. Overview

Sections 5.4 through 5.6 demonstrate that events 1 through 10 function as intended and produce outcomes that are consistent with either observed outcomes or expectations. The remaining events 11 through 17 are market supply events. To test if the market supply events function as intended and produce outcomes consistent with observed outcomes, the full set of events is subject to verification and validation. Unless otherwise noted, the verification and validation tests are performed using the default DCHAS values specified in Table 5.9. The SAS program for running the market supply events is called "Supply Events.sas."

Table 5.9: Default Values for the DCHAS

Parameter	Default Value
InMigrationCount	64,930 households per year
OutMigrationCount	57,710 households per year
PropInvestmentRate	6 percent per year
DevelopmentRate	4,170 units per year
SubUnitInvestment	$195 million per year
RehabShare	55 percent

PUDUnitsApproved	500 units per year

5.7.2. Distributional Equivalence Analysis of Full Model

For the distributional equivalence analysis, DCHAS is run using the default parameters for 36 time steps. The two outcomes of interest are median housing price for homes sold between the first and last time step and median rental rate for market-rate units as of the last time step. The results presented in Table 5.10 show consistent values throughout successive runs. The results suggest the full DCHAS model is operating as intended.

Table 5.10: Distributional Equivalence Analysis of Full Model

	Median Housing Price for Homes Sold	**Median Rental Rate for Market Rate Units**
Replication #1	429,500	1,270
Replication #2	431,000	1,240
Replication #3	431,000	1,270
Replication #4	432,500	1,250
Replication #5	430,000	1,290
Average	431,000	1,260

5.7.3. Relational Alignment Analysis of Full Model

For the relational alignment analysis, any one of DCHAS's seven adjustable event parameters are candidates to be varied. However, only one event parameter is tested: the parameter *DevelopmentRate* is doubled. All else being equal, a doubling of the development rate translates into greater supply of housing, which should lead to a decrease in median housing price and median rental rate, while leading to an increase in overall vacancy rate.

The relational alignment analysis is performed by running DCHAS for 36 time steps. The outcomes of interest are the median price of home sold between the first and last time step and the median rental rate as of the last time step. The outcomes are compared to the average outcomes presented in Table 5.10.

The results of the relational alignment analysis are presented in Table 5.11. The results are consistent with prior expectations: an increase in the supply of housing leads to a decrease in the price (or rental rate) of housing. The results provided further evidence that the full DCHAS is functioning as intended.

Table 5.11: Relational Alignment Analysis of Doubling the Development Rate

	Median Housing Price for Homes Sold	**Median Rental Rate for Market Rate Units**
Baseline run	431,000	1,260
Run #1	417,000	1,180
Run #2	414,500	1,150
Run #3	415,500	1,150

5.7.4. Validation of Full Model

The final step is to validate DCHAS results against actual outcomes from independent data sources. Recall that the initial rental rate event rates are derived from the 5-year 2015 ACS, so it stands to reason that DCHAS results for model runs through 2015 should match rental rate estimates from the ACS. The initial housing prices and the appraisal model estimates are from CAMA, which includes actual market transactions, so it stands to reason that DCHAS results for model runs through 2015 should match recent sales prices in the CAMA data.

To test this claim, DCHAS in run five times and each run is for 63 time steps. Four key statistics are evaluated: median rental rate, median sales price, percent of renter households paying more than 30 percent of their income for housing, and vacancy rate. Table 5.12 compares the results from DCHAS with 2015 estimates from the relevant data source. The results from DCHAS are presented as ranges so as to convey the amount of variation in the results. The results show the DCHAS produces median rental rates in 2015 that are slightly higher than the independent estimate from the ACS. Related to this, DCHAS produces estimates of the share of renter households paying more than 30 percent of their income for housing that are slightly higher than the independent estimate from the ACS. However, neither of the two DCHAS estimates are far outside the range of the independent estimates.

With respect to median sales price estimates, DCHAS produces estimates that are about 10 percent below the 2015 independent estimate. This suggests that either DCHAS's sales price appraisal process is not producing sales prices increases or the negotiation processes are not capturing sales price increases, perhaps due to less demand. Whatever the cause, it is important to note this issue does not impact the result of interest, which is share of renters paying more than 30 percent of their income for housing. However, this issue merits future investigation.

Table 5.12: Validation of Full Model

	2015 DCHAS estimate range	2015 Independent Estimate
Median rental rate	$1,410 ($15)	$1,327[35] ($42)
Median sale price	$457,000 ($454,500 - $459,000)	$508,000[36]
Share of renter households paying more than 30 percent of their income for housing	47.3% (0.5%)	45.6%[37] (1.8%)
Vacancy rate	8.2% (0.3%)	9.0%[38] (1.0%)

5.8. Final List of SAS Programs and Event Operational Order

Tables 5.13 through 5.16 contains the final list of the SAS programs that constitute DCHAS. Operationally, DCHAS performs the events in the order they appear in each table. All DCHAS programs and datasets are available at: https://dataverse.unc.edu/dataverse/DCHAS.

Table 5.13: SAS Programs Constituting DCHAS Model Initialization

SAS Program Name	Input Datasets	DCHAS SAS Output Datasets
Helper Functions	None	None
Prepare GIS Data	-Raw Data\GIS\CAMAGIS.dbf -Raw Data\GIS\DCPCGIS.dbf -Raw Data\Other\CensusTenure.xlsx -Raw Data\Other\HCVP.xlsx	-PropGIS -WardCounts
Household Initialization	-Raw Data\SynPop\Households.txt -Raw Data\SynPop\People.txt -PropGIS	-People
Housing Unit Initialization	-Raw Data\CAMA\Residential.dbf -Raw Data\CAMA\Commerical.dbf -Raw Data\CAMA\Condo.dbf	-Units -Properties

35 American Fact Finder Table B25064 for 2015 1-year ACS data.
36 Author's analysis of CAMA data for 2015. CAMA records all sales transactions. As such, there is no variance associated with the estimate.
37 American Fact Finder Table GCT2515 for 2015 1-year ACS data.
38 American Fact Finder Table C25004 for 2015 1-year ACS data.

	-Raw Data\CAMA\PUD.dbf -Raw Data\CAMA\Zoning.dbf -Raw Data\CAMA\Owner_Points.dbf -Raw Data\Other\Data Exploration.xlsx -Raw Data\Other\Subsidy DB.xlsx -Raw Data\Other\CensusTenure.xlsx -Raw Data\DCPC\parcel.csv -Raw Data\DCPC\project.csv -Raw Data\DCPC\subsidy.csv	
Tenure Initialization	-Raw Data\ACS IUF\ACSHUD_5yr.sas7bat* -People	-HUDAssistanceModel -InitialTenure
Rent Initialization	-Raw Data\ACS IUF\ACSHUD_5yr.sas7bat* -Raw Data\GIS\Blks10_met.dbf	-InitialRentModel -InitialRents

*This is an internal dataset that cannot be made publicly available. However, the resulting parameter estimates file, HUDAssistanceModel, is publicly available.

Table 5.14: SAS Programs Constituting DCHAS Initial Placement

SAS Program Name	**Input Datasets**	**DCHAS SAS Output Datasets**
Subsidized Household Initial Placement	People InitialTenure Units InitialRents	SubPlaced
VLI and LI into LIHTC Initial Placement	People InitialTenure Units InitialRents SubPlaced	LowInc2LIHTCPlaced
Owner Household Initial Placement	-Raw Data\ACS IUF\ACSHUD_5yr.sas7bat* People -Raw Data\GIS\Blks10_met.dbf People Units SubPlaced LowInc2LIHTCPlaced	InitialOwnerPlacementModel OwnerPlaced
Market-rate Renter Household Initial Placement	-Raw Data\ACS IUF\ACSHUD_5yr.sas7bat** People -Raw Data\GIS\Blks10_met.dbf People Units SubPlaced LowInc2LIHTCPlaced OwnerPlaced	InitialRenterPlacementModel RenterPlaced

*This is an internal dataset that cannot be made publicly available. However, the resulting parameter estimates file (InitialOwnerPlacementModel) is publicly available.
**This is an internal dataset that cannot be made publicly available. However, the resulting parameter estimates file (InitialRenterPlacementModel) is publicly available.

Table 5.15: SAS Programs Constituting DCHAS Parameter Estimation

SAS Program Name	Input Datasets	DCHAS SAS Output Datasets
Life Cycle Parameter Estimation	2015 5-yr ACS PUMS for DC	MarriageProb DivorceProb BirthProb DeathProb FirstJobIncome18 FirstJobIncome25 IncomeGrowthRate
Migration Parameter Estimation	2015 5-yr ACS PUMS for US	OutMigrationModel InMigrationPool
Mobility, Formation and Tenure Selection Parameter Estimation	2015 5-yr ACS PUMS for DC	MoveProb4Families MoveProb4Singles FormationProb PurchaseProb
Single Family Appraisal Model Parameter Estimation	-Intermediate CAMA data sets produced in the model initialization phase PropGIS -Raw Data\Other\Subsidy DB.xlsx -Raw Data\DCPC\parcel.csv	SFAppraisalModel
Rental Appraisal Model Parameter Estimation	Units InitialRents	RentalAppraisalModel
Location Choice Model Parameter Estimation	-Units -People -SubPlaced -LowInc2LIHTCPlaced -OwnerPlaced -RenterPlaced	OwnerChoiceModel RenterChoiceModel
Land Appraisal Model Parameter Estimation	-Intermediate CAMA data sets produced in the model initialization phase PropGIS -Raw Data\Other\Subsidy DB.xlsx -Raw Data\DCPC\parcel.csv	LandAppriasalModel
Average Building Size by Zoning Type Estimation	-Intermediate CAMA data sets produced in the model initialization phase PropGIS -Raw Data\Other\Subsidy DB.xlsx -Raw Data\DCPC\parcel.csv	ZoningMetrics
PUD Probability Model Parameter Estimation	ZoningMetrics -Intermediate CAMA data sets produced in the model initialization phase PropGIS -Raw Data\Other\Subsidy DB.xlsx -Raw Data\DCPC\parcel.csv	PUDMetrics PUDModel

*This is an internal dataset that cannot be made publicly available. However, the resulting parameter estimates files (OwnerChoiceModel and RenterChoiceModel) are publicly available.

Table 5.16: SAS Programs Constituting DCHAS Parameter Estimation

SAS Program Name	Input Datasets/Notes
Migration and Life Cycle Events	All DCHAS SAS Output Datasets
Mobility, Formation and Tenure Selection Events	
Demand Events	
Supply Events	
DCHAS	This is a main wrapper program for the four events.

5.9. Conclusions from Initialization, Calibration, Verification and Validation

DCHAS is initialized to the year 2010 using detailed Decennial Census data and detailed housing stock data. The parameters that control the events occurring in DCHAS are derived from survey data from 2011 through 2015. The results presented in this chapter demonstrate that when DCHAS uses the default parameter values derived from survey data, it produces simulated outcomes that closely track real world outcomes found in the survey data for 2015. Moreover, adjustment to the parameter values produce outcomes that are in line with what should be expected, given the methodology used to develop each of DCHAS's events. As such, it appears the DCHAS is functioning as intended and represents a simplified version of the land and housing market in Washington, D.C. Therefore, it can reasonably be expected that DCHAS can be used to simulate future years of the Washington, D.C., land and housing market under various scenarios.

CHAPTER 6: RESULTS AND DISCUSSION OF FOUR SCENARIOS

In this chapter, DCHAS is used to simulate three housing supply and demand scenarios over 10 years (2016 – 2025). Each scenario is intended to be a realistic portrayal of what might happen in the land and housing market. In each scenario, DCHAS is initially run from 2010 to 2015 using the default parameters so as to create a population of household members and households. Then, the 2015 population is simulated forward 10 years (2016 – 2025) with parameters specific to the scenario. Each scenario is run five times to produce an average estimate.

Section 6.1 presents the results of the "status quo" scenario, which reflects a continuation of trends observed between 2011 and 2015. Then, section 6.2 presents the results of a "construction boom" scenario where beginning in 2016, the number of new units constructed is increased 25 percent over the average rate observed between 2011 and 2015. Section 6.3 presents the results of a scenario where there is a substantial increase in investment in affordable housing. Starting in 2016, the level of investment is doubled relative to averages observed between 2011 and 2015.

For each scenario, the general outcome of interest is housing affordability, as measured by the share of households paying more than 30 percent of their income for housing. However, to provide the appropriate context to this one outcome, six other outcomes are also discussed: household growth, new market-rate unit construction, new subsidized unit construction, median income growth, median housing price growth, and median rent growth.

6.1. Scenario 1: Status Quo

In the status quo scenario, the seven adjustable event parameters are held at their default values, which reflect averages between 2011 and 2015. Table 6.1 shows the values of the adjustable parameters.

Table 6.1: Default Values for the Adjustable Parameters in Scenario 1

Parameter	**Default Value**
InMigrationCount	64,930 household members per year
OutMigrationCount	57,710 household members per year
PropInvestmentRate	6 percent per year
DevelopmentRate	4,170 units per year
SubUnitInvestment	$195 million per year
RehabShare	55 percent
PUDUnitsApproved	500 units per year

Table 6.2 shows the results of the scenario. Running DCHAS with the default parameter values in Table 6.1 for 10 years, reveals that in 2025, it can be expected that 7.9 percent of households will pay more than 30 percent of their income for housing costs. This finding is not surprising. As discussed in chapter 5 and shown in Table 5.13, approximately 8.2 percent of households paid more than 30 percent of their income for housing costs as of 2015. Table 6.2 shows that the Washington, D.C. household growth rate is about 3,550 households per year, while the total new unit development rate is about 4,170 units per year. Thus, on net, the demand for housing is increasing slightly

slower than supply, which can be expected to lead to a modest decline in the growth of housing prices relative to the growth in income. As Table 6.2 shows, the growth in median rent between 2016 and 2025 is approximately 4.1 percent per year, while the income growth rate is 4.4 percent. As a result, slightly fewer households pay more than 30 percent of their income for housing, relative to 2015.

Table 6.2: Outcomes from Scenario 1

Outcome	**Simulated Value (average results from 5 runs)**
Annual household growth households)	3,550
Annual new market-rate unit construction (units)	3,300
Annual subsidized unit construction (units)	870
Median income growth rate (percent)	4.6
Median market-rate housing price growth (percent)	1.8
Median market-rate rent growth (percent)	4.1
Share of renter households paying more than 30 percent of their income for housing (percent)	7.9

6.2. Scenario 2: Market Rate Construction Boom

In the market rate construction scenario, the six of seven adjustable event parameters are held at their default values, which reflect averages between 2011 and 2015. However, the *DevelopmentRate* parameter value is increased to 5,210 units per year starting in 2016 and the *PUDUnitsApproved* is increased to 2,000 units per year starting in 2016. Table 6.3 shows the values of the adjustable parameters.

Table 6.3: Default Values for the Adjustable Parameters in Scenario 2

Parameter	**Default Value**

InMigrationCount	64,930 household members per year
OutMigrationCount	57,710 household members per year
PropInvestmentRate	6 percent per year
DevelopmentRate	5,210 units per year in 2016
SubUnitInvestment	$195 million per year
RehabShare	55 percent
PUDUnitsApproved	2,000 units per year starting in 2016

Table 6.3 shows the results of the scenario. Running DCHAS with the parameter values in Table 6.3 for 10 years reveals that in 2025, it can be expected that 7.6 percent of household will pay more than 30 percent of their income for housing costs, or a 0.6 percent reduction over the 2015 values of 8.2 percent in Table 5.13. Table 6.4 shows the Washington, D.C. household growth rate is about 3,550 households per year between 2016 and 2025, while the total new unit development rate is 5,210 units per year during the same period. Thus, on net, the supply for housing is increasing faster than demand. In this scenario, it can be expected that housing prices and rent growth will be slower than income growth. In fact, as Table 6.4 shows, the growth in median rent between 2016 and 2025 is approximately 3.8 percent per year, while the income growth rate is 4.4 percent per year. As a result, 7.6 percent renter households paid more than 30 percent of their income for housing – an improvement over the 8.2 percent in 2015.

Table 6.4: Outcomes from Scenario 2 from 2016- 2025

Parameter	**Simulated value**
Annual household growth	3,550
Annual new market-rate unit construction	4,340
Annual subsidized unit construction	870
Median income growth rate	4.6

Median market-rate housing price growth	1.5
Median market-rate rent growth	3.8
Share of renter households paying more than 30 percent of their income for housing	7.6

6.3. Scenario 3: Affordable Housing Investment Boom

In the affordable housing investment boom scenario, the six of seven adjustable event parameters are held at their default values, which reflect averages between 2011 and 2015. However, the *SubUnitInvestment* parameter value is more than doubled to $400 million per year. Table 6.5 shows the values of the adjustable parameters.

Table 6.5: Default Values for the Adjustable Parameters in Scenario 3

Parameter	**Default Value**
InMigrationCount	64,930 household members per year
OutMigrationCount	57,710 household members per year
PropInvestmentRate	6 percent per year
DevelopmentRate	4,170 units per year
SubUnitInvestment	$400 million per year
RehabShare	55 percent
PUDUnitsApproved	500 units per year

Table 6.6 shows the results of the scenario. Running DCHAS with the parameter values in Table 6.5, for 10 years, reveals that in 2025, it can be expected that 8.0 percent of household will pay more than 30 percent of their income for housing costs, or a slight reduction over the 2015 value of 8.2 percent in the previous chapter in Table 5.13. Table 6.6 shows the Washington, D.C. household growth rate is about 3,550 households per year, while the total new unit development rate is about 4,170 units per year, with 1,740

of that amount being subsidized. Thus, on net, the supply for housing is increasing slightly faster than demand, but more of the supply is devoted towards low-income households. In this scenario, market-rate housing prices increase about 2.5 percent per year, while market-rate rents increase about 5 percent per year. Although market-rate rents are growing faster than income, the increase in affordable housing units, which by definition reduce the number of households paying more than 30 percent of their income for rent, almost offsets the number of renter households in market-rate rental housing who are now paying more than 30 percent of their income on rent. The net result is 8.0 percent of renter households paid more than 30 percent of their income for housing, relative to 2015. A policy implication from this scenario is that an affordable housing construction boom needs to be *in addition to* regular rates of market-rate housing development.

Table 6.6: Outcomes from Scenario 3

Parameter	**Simulated value**
Annual household growth	3,550
Annual new market-rate unit construction	2,430
Annual subsidized unit construction	1,740
Median income growth rate	4.6
Median market-rate housing price growth	2.5
Median market-rate rent growth	5.0
Share of renter households paying more than 30 percent of their income for housing	8.0

CHAPTER 7: CONCLUSION

This study had three primary purposes: (1) to build an empirical ABM of housing supply and demand in an urban environment, including agent heterogeneity and endogenous land markets, with a specific emphasis on housing affordability and the supply of and demand for affordable housing; (2) to use the ABM to simulate affordability and affordable housing supply and demand under current market conditions; and (3) to use the ABM to explore different policies aimed at improving housing affordability.

Related to these three purposes were four research questions:

(1) What are the critical concepts to include in an ABM of housing supply and demand in an urban environment that is characterized by agent heterogeneity and endogenous land markets, and which includes a specific emphasis on housing affordability and affordable housing?

(2) Can each of the important concepts be represented empirically, and if so, what are the best data sources?

(3) Does the ABM produce city-level estimates of housing affordability that are in global agreement with known indicators?

(4) What happens to city-level affordability under different housing policy scenarios?

This final chapter discusses the extent to which the four research questions were answered by the research and development of DCHAS, discusses aspects of the research question that were not answered, and suggests future directions for this line of research.

7.1. Review of Research Question 1

The first research question asked what are the critical concepts to include in an ABM of housing supply and demand that focuses on housing affordability and affordable housing. As discussed in section 2.6, prior modeling efforts recognized the importance of agent heterogeneity and endogenous land and housing markets. This research effort accepted the premise that those concepts are important to include in an ABM of housing supply and demand, and aspired to determine whether additional concepts are also important. Section 2.5 introduced and discussed six concepts that were believed to be important to modeling housing affordability. Each of these concepts was built into DCHAS.

The research effort did not independently test whether each of these concepts was statistically important to simulated outcomes over the next ten years (2016-2025). To do so would require developing a counter-factual environment where each concept is "turned off", so to speak. Scenario four was one such counter-factual (affordable housing investment was decreased) but it was the only one. For example, rent control could be turned off by allowing rents to increase at whatever rate the market supported.

It is reasonable to conclude that research question 1 was at least partially answered by introducing and discussing the concepts, including how many household or housing units are impacted by the concept. It is acknowledged that any of these concepts may not actually be statistically important to simulated outcomes. Future research could test whether each of the concepts are statistically important to simulated outcomes.

7.2. Review of Research Question 2

The second research question asked whether the concepts critical to modeling housing affordability could be represented empirically, and if so, what are the best data sources to represent the concept. In short, each of the six critical concepts were represented empirically, either by using survey data, or by using local, state, or Federal administrative data. In fact, one of the principle contributions of this research to the body of literature on ABMs of housing supply and demand was demonstrating use of administrative data sources to empirically parameterize both the agents and the environment. Below are four examples.

First, a special version of the 5-year 2011 – 2015 ACS linked to HUD rental subsidy administrative data was used to estimate the parameters of a model that identified which DCHAS agents were receiving a federal housing subsidy. Using this data source improved the empirical representation of two critical concepts: creating agents with variation appropriate to low-income populations and reflecting federal housing subsidies.

Second, a data set of federally subsidized properties, the DCPC data set, was used to: (1) better inventory and locate subsidized properties within the DCHAS environment; (b) to calibrate the number of subsidized housing units; and (c) to correctly determine subsidy expiration date, which facilitates developing more empirically realistic scenarios of the impacts of affordable housing investments. Using this data source improves the empirical representation of two critical concepts: reflecting federal housing subsidies and reflecting sources of affordable housing supply.

Third, the Washington, D.C. government's CAMA data set was the principle source of data used to create the housing stock. Attributes in the CAMA data, including sales price, year built, number bedrooms, and lot size, were included as attributes of the housing stock in DCHAS. Using the CAMA data improved the empirical representation of the land and housing markets by using unit counts and recent sales prices. The CAMA data also improved the empirical representation of two critical concepts: rent control and filtering and rehabilitation.

Fourth, the Washington, D.C. government's zoning and PUD data set was used to estimate parameters for two events in DCHAS: The Developer's expected revenue calculation and PUD approval event. These data sets improved the empirical representation of the one of the critical concepts: zoning and regeneration of existing properties.

7.3. Review of Research Question 3

The third research question asked whether the ABM could reproduce city-level estimates of housing affordability that were in global agreement with known indicators. As demonstrated in section 5.7.4, when running DCHAS for the period of 2010 through 2015, DCHAS reproduced estimates of population counts, housing stock, housing price changes, and housing affordability, that were in general agreement with survey data from 2015.

It is important to acknowledge that DCHAS was constructed with heavy reliance on statistical relationships based on prior data, so it should be expected to reproduce conditions that existed in 2015. For instance, the entire set of migration, life cycle, mobility, formation, and tenure choice events were driven exclusively by statistical data from 2011 to 2015. As such, there was every reason to expect DCHAS to reproduce population estimates. The same was true for housing production, as the creation of new housing units was tightly controlled by a few parameters that reflected actual production values. Moreover, even though the series of events representing relocation (appraisal, location selection, and price/rent negotiation and market clearing) were based on two underlying economic models (the hedonic demand and the random utility framework), the parameters of these models were estimated using actual transactions from 2011 to 2015. These two events, coupled with a simple price negotiation and market clearing event that assumes a two percent premium for properties in high demand, reproduced price and rent changes that generally tracked observed data from the same period.

7.4. Review of Research Question 4

The fourth research question asked what happens to city-level affordability under different housing policy scenarios. Scenarios 6.1 through 6.3, provided some insight into the impact of various scenarios on housing affordability over 10 years (2016 – 2025). Scenario 1 showed that affordability issues will continue to persist under the current levels of household formation and housing construction. Scenarios 2 and 3 showed that substantial increases in housing production could alleviate some of the affordability issues, but that increases in affordable housing supply must be in addition to increases in market-rate housing supply.

It is acknowledged there are numerous other scenarios that should be explored to help paint a more complete picture of the impact of various housing policies on housing affordability. For instance, DCHAS could be used to determine what rate of market rate and affordable housing construction would need to occur to reduce the number of low-income households paying more than 30 percent of the income for housing costs.

7.5. Concluding Thoughts

The goals of this research were to: (1) build an empirical ABM of housing supply and demand in an urban environment, including agent heterogeneity and endogenous land markets, with a specific emphasis on housing affordability and the supply of and demand for affordable housing; (2) show that the ABM works well; and (3) use the ABM to investigate affordability under various policies. Each of the goals were achieved. The three principal contributions of this research were to: (1) identify and explore six concepts critical to housing affordability in an urban environment; (2) demonstrate how to empirically represent these concepts through the use of administrative data sources, and (3) demonstrate how to build an empirically-based ABM that can be used to simulate housing affordability under different market conditions or housing policy scenarios.

There are numerous opportunities for future work to improve DCHAS. First, DCHAS could be extended to include the entire Washington, D.C. region. As discussed in section 1.8, this delimitation of DCHAS is due to the lack of readily-available data on subsidized housing locations. Second, DCHAS could benefit from an improved specification for a household's choice to rent or own. The current specification is based purely on past statistical relationships and does not account for endogenous changes in prices created within DCHAS. Third, DCHAS could benefit from a more empirically accurate representation of the marriage process, including accounting for same-sex marriage a cohabitation. Fourth, DCHAS could benefit from an improved household formation process. Currently, if a household member is selected to move, but not in to their own housing unit, DCHAS places them into an existing housing unit with an open bedroom, occupied by a household whose members share similar characteristics (race, income age). The process needs to be more empirically-grounded, and at a minimum, consider the cost of renting a bedroom. Fifth, DCHAS could benefit from a better representation of housing developers, rather than the current implementation of a single developer. Sixth, the DCHAS housing sale mechanism is not producing prices that align with empirical results. This issue merits further investigation, but does not impact the measure of affordability for renters. Finally, DCHAS could benefit for a land market that uses prices to determine the *amount* of new housing supply. In the current implementation of the land market, prices only impact the location of development, not the amount.

APPENDIX

Appendix A: DCHAS Parameter Values and Model-based Parameter Estimates

Table A1: 2010 Counts of Households by Tenure Type

	Household Count
Tenure Type	
Public Housing or Project-based Section 8	22,859
Housing Choice Voucher Program	13,400
Non-subsidized renters	118,383
Owners	112,043
Total Occupied	266,685
Vacant	30,034
Total	296,719

Table A2: HUD Assistance Model Parameter Estimates

Parameter	**Tenure**[39]	**Estimate**	**Standard Error**	**Wald Chi Square**	**Pr > ChiSq**
Intercept	PH/MF	-0.296	0.027	125	<.0001
Intercept	Non-sub. renter	0.593	0.010	3346	<.0001
Intercept	Voucher	-1.082	0.037	838	<.0001
Household income	PH/MF	0.000	0.000	12,840	<.0001
Household income	Non-sub. renter	0.000	0.000	18,886	<.0001
Household income	Voucher	0.000	0.000	8,158	<.0001
Non-elderly household	PH/MF	0.339	0.010	1,185	<.0001
Non-elderly household	Non-sub. renter	0.798	0.006	18,001	<.0001
Non-elderly household	Voucher	0.610	0.014	2,003	<.0001
Household size = 1	PH/MF	-0.356	0.016	509	<.0001
Household size = 1	Non-sub. renter	0.035	0.009	16	<.0001
Household size = 1	Voucher	-0.585	0.019	993	<.0001
Household size = 2	PH/MF	-0.131	0.018	51	<.0001
Household size = 2	Non-sub. renter	0.151	0.009	291	<.0001
Household size = 2	Voucher	-0.430	0.023	362	<.0001
Household size = 3	PH/MF	0.187	0.022	70	<.0001
Household size = 3	Non-sub. renter	0.016	0.012	2	0.1873
Household size = 3	Voucher	0.310	0.024	165	<.0001
Household size = 4	PH/MF	-0.060	0.028	5	0.031
Household size = 4	Non-sub. renter	-0.207	0.014	216	<.0001
Household size = 4	Voucher	0.011	0.030	0.1	0.7127
White	PH/MF	-1.064	0.021	2,662	<.0001
White	Non-sub. renter	0.133	0.005	719	<.0001
White	Voucher	-1.249	0.032	1569.4	<.0001

[39] The reference group for the dependent variable (Tenure) is "owners." The reference group for household size is 5 or more.

Table A3: Out-Migration (*OutMigrationProb*) Model Parameter Estimates

Parameter*	**Estimate**	**Standard Error**	**Wald Chi Square**	**Pr > ChiSq**
Intercept	-2.94	0.025	13,954	<.0001
Household type = 1	0.24	0.011	454	<.0001
Household type = 2	-0.52	0.015	1,169	<.0001
Person age group = 3	1.89	0.029	4,325	<.0001
Person age group = 4	1.15	0.024	2,336	<.0001
Person age group = 5	0.41	0.026	238	<.0001
Person age group = 6	0.08	0.029	7	0.008
Person age group = 7	-0.07	0.030	5	0.0186
Person age group = 8	-0.74	0.040	338	<.0001
Person age group = 9	-0.57	0.048	140	<.0001
Household with young children	-0.19	0.014	174	<.0001

*The reference groups are household type = 3 and person age group = 10.

Table A4: Death Rates

	Annual Probability of Death (percent)
Age Group	
1. 0-4	0.2
2. 5-14	<0.1
3. 15-24	0.1
4. 25-34	0.1
5. 35-44	0.2
6. 45-54	0.6
7. 55-64	1.2
8. 65-74	2.0
9. 75-84	4.5
10. 85+	12.5

Table A5: The Sales Price Hedonic Model Parameter Estimates

Parameter	Estimate	Standard Error	t-value	Pr > F
Intercept	-210,620	18,343	-11.48	<.0001
Age of structure	-22,234	206	-10.81	<.0001
Distance to nearest metro stop	-111	7	-15.80	<.0001
Lot size	-3.6	2.27	-1.59	0.1110
Unit size	440	9	50.68	<.0001
Bedrooms	-28,137	5,640	-4.99	<.0001
Unit is a condo	78,297	11,832	6.62	<.0001
ρ value of spatial weight matrix	0.78620	0.01	40.87	<.0001

Table A6A: The Household Location Selection Model Parameter Estimates for Owner Households

Parameter	Estimate	Standard Error	Wald Chi Square	Pr > ChiSq
Unit is a single-family home	-0.4587	0.0689	44.4	<.0001
Unit is less than 10 years old	-0.3932	0.0589	44.5	<.0001
Distance to nearest metro stop	0.0001	0.0001	0.8	0.3805
Size of household minus number of bedrooms	-0.1799	0.0273	43.5	<.0001
Income after housing costs	0.0000	0.0000	0.1	0.7075
Householder race * unit is a single-family home	-0.0636	0.0949	0.4	0.5029
Householder age * distance to nearest metro stop	0.0000	0.0001	0.1	0.7868

Table A6B: The Household Location Selection Model Parameter Estimates for Renter Households

Parameter	Estimate	Standard Error	Wald Chi Square	Pr > ChiSq
Unit is a single-family home	0.7327	0.0537	186	<.0001
Unit is less than 10 years old	-0.6736	0.0383	309	<.0001
Distance to nearest metro stop	-0.0003	0.0001	20	<.0001
Size of household minus number of bedrooms	0.0237	0.0039	37	<.0001
Income after housing costs	0.0001	0.0000	48	<.0001
Householder race * unit is a single-family home	-0.4758	0.0762	39	<.0001
Householder age * distance to nearest metro stop	0.0002	0.0000	43	<.0001

Table A7: The Rent Hedonic Model Parameter Estimates

Parameter	Estimate	Standard Error	t-value	Pr > F
Intercept	172	2	106	<.0001
Age of structure	-5.2	0.01	-363	<.0001
Distance to nearest metro stop	-0.09	0.0005	-190	<.0001
Bedrooms	193	0.4	543	<.0001
Unit in a single-family home	180	1	178	<.0001
Unit in a building with 10 or more apartments	13.9	0.93	15	<.0001
σ value of spatial weight matrix	0.71	0.001	714	<.0001

Table A8: The Land Price Per Square Foot (Property Cost) Hedonic Model Parameter Estimates

Parameter	Estimate	Standard Error	t-value	Pr > F
Intercept	84	30	2.8	0.0049
Distance to nearest metro stop	-0.07	.02	-3.2	0.0013
λ value of spatial weight matrix	0.94	.03	31	<.0001

Table A9: RSMeans Construction Cost Estimates for 2014

Construction Type	Cost per Square Foot
Single-family	$98
Low-rise multifamily	$183
High-rise multifamily	$223

Table A10: Bedroom Shares for Subsidized and Commercial CAMA Rental Properties

	Subsidized (any size)	**Small Buildings (5-9 units)**	**Medium Buildings (10-49 units)**	**Large Buildings (50 or more units)**
Bedroom Count	*Share of units (percent)*			
Studio/1 bedroom	43	51	58	80
2 bedrooms	32	39	36	20
3 bedrooms	24	8	6	2
4 or more bedrooms	0	2	1	0

Table A11: Median Square Foot of Building Per Land Area and Square Foot of Units, by Zoning Type

Zoning Type	Median Building Square Feet Per Square Foot of Land Area	Median Unit Size (square feet)	Allowable Increase in Building Square Feet Per Square Foot of Land Under a PUD
C-1	0.88	1,040	0.56
C-2-A	1.97	879	2.00
C-2-B	4.10	928	0.05
C-2-C	7.53	880	0.00
C-3-A	3.12	873	0.12
C-3-B	1.19	1,157	6.21
C-3-C	7.40	945	0.00
C-4	9.80	793	0.00
C-M-1	1.15	1,644	2.38
C-M-2	1.21	684	1.36
CR	3.30	888	2.68
M	0.94	1,314	4.06
R-1-A	0.32	3,372	1.37
R-1-B	0.39	2,008	0.89
R-2	0.47	1,216	0.37
R-3	0.83	1,332	0.63
R-4	1.10	1,188	0.95
R-5-A	0.82	824	1.68
R-5-B	1.71	898	0.83
R-5-C	2.26	666	1.50
R-5-D	2.06	890	2.16
R-5-E	4.86	752	0.37
SP-1	2.93	984	2.78
SP-2	5.17	946	0.20
W-1	3.35	3,050	0.00
W-2	3.22	1,700	0.00

Table A12: Planned Unit Development Location Model Parameter Estimates

Parameter	Estimate	Standard Error	Wald Chi Square	Pr > ChiSq
Intercept	-2.6	.0004	44,206,707	<.0001
Distance to nearest metro stop	-0.0006	<0.0000	2,917,313	<.0001

Appendix B: Computer Assisted Mass Appraisal (CAMA) Cleaning Procedures

B.1. Residential CAMA Cleaning Procedures

The Residential CAMA data set contained numerous fields. However, the only pieces of information necessary for the DCHAS model were the following: SSL, type of parcel (residential structure or vacant residential land), number of bedrooms (count), number of units (count), year built or estimated year built of structure (year), year sold (year), sale price (dollars), current zoning, and owner name. The cleaning procedure applied to the Residential CAMA data set are listed in Table B1. In this Table, native CAMA fields are in bold.

Table B1: Cleaning Procedures for the Residential CAMA Data Set

Variable	**Cleaning procedure**	**Reason**
Property Type	Records with **USECD**[40] value 1;2;3;11 to 29; or 38, were categorized as "single-family." Records with **USECD** values 91 to 97, or 191 to 197, were categorized as "vacant land." Records with other **USECD** values not listed above were deleted.	These **USECD** values reflect residential uses or vacant lots intended for residential use. Other **USECD** values reflected commercial or industrial uses, or group quarters, such as dormitories.
Bedrooms	Based on the value in **BEDROOMS**. Where missing, impute a value based on the following equation: Bedrooms = 2.02 + **GROSS BUILDING AREA *** .00078)	Bedrooms are a necessary feature of housing units. Where missing, they must be imputed. The imputation equation is based on an ordinary least square regression of number of bedrooms on gross building areas.

[40] For a full listing of USECD values, see: http://app.cfo.dc.gov/services/tax/property/pdf/usecodes.pdf

Year built	Remove records with **ACTUAL YEAR BUILT** > 2010.	The DCHAS model is initialized with housing units as they existed in 2010.
Number of units	Set equal to **KITCHENS**, which reflects the number of kitchens.	Some homes that visually appear to be single-family structures, and are considered single-family structures for tax assessment purposes, are small multifamily properties composed of a row house subdivided into more than one housing unit. The subdivision into more than one unit may not be recorded by DCOTR. The author's expert opinion is the number of kitchens serves as a close proxy for number of units.

B.2. Condominium CAMA Cleaning Procedures

The data set contained numerous fields. However, the only pieces of information necessary for the DCHAS model were the following: SSL, number of bedrooms (count), year built or estimated year built of structure (years), year sold (year), sale price (dollars), current zoning, and owner name[41]. The only notable cleaning procedure applied to the Condominium CAMA data set was to remove any unit where the CAMA field "COMPLEX" was greater than 3179. The CAMA field COMPLEX represented the condominium project number assigned by DCOTR. Analysis revealed these condominium projects were built after 2010.

[41] By design, the Condominium CAMA data set contains only parcels that have residential structures (no vacant parcels) and each record reflects only one housing unit.

B.3. Commercial CAMA Cleaning Procedures

The Commercial CAMA data set contained numerous fields. However, the only pieces of information necessary for the DCHAS model were the following: SSL, type of parcel (residential structure, commercial structure, vacant land), number of units (count), year built or estimated year built of structure (years), year sold (year), sale price (dollars), current zoning, and owner name. The cleaning procedure applied to the Commercial CAMA data set are listed in Table 4.2. In this Table, native CAMA fields are in bold.

Table B2: Cleaning Procedures for the Commercial CAMA Data Set

Variable	**Cleaning procedure**
Property type	Records with the following **USECD** values were categorized as "market rate multifamily rental": 1, 11, 12, 13, 15,18,19, 21, 22, 23,25,26, 27, 28, 29, 117, 126, 127, 216, 217, and 316. Records with **USECD** values 91 to 97, or 191 to 197, were categorized as "vacant land." Records with **USECD** values 31 to 39, and 88, were categorized deleted. These USECD values reflected group quarters. Records with other **USECD** values not listed above were categorized as "commercial/industrial."
Number of units	Each property with a **USECD** value of one of the residential structures mentioned above, but having **NUM_UNITS** = 0, was individually investigated to determine the correct number of units. Most investigations included either web searches for the apartment complex name or review of other DC government records, including Master Address Record Address Points data set.
Year built	Remove records with **ACTUAL YEAR BUILT** > 2010.

REFERENCES

Abt Associates. (2012). *What happens to low-income housing tax credit properties at year 15 and beyond?* Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Ahlfeldt, G. M., & McMillen, D. P. (2013). New estimates of the elasticity of substitution of land for capital. *Working paper, London School of Economics*.

Ahmed, V., & O'Donoghue, C. (2007). *CGE-microsimulation modelling: a survey.* Planning Commission Pakistan and TEAGASC Ireland.

Albouy, D., Ehrlich, G., & Liu, Y. (2016). *Housing Demand, Cost-of-Living Inequality, and the Affordability Crisis.* National Bureau of Economic Research Working Paper No. 22816.

Alonso, W. (1964). *Location and Land Use.* Cambridge, MA: Harvard University Press.

Anenberg, E., Hizmo, A., Kung, E., & Molloy, R. (2015). *The effect of mortgage credit availability on house prices and construction: Evidence from a frontier estimation approach.* Working Paper.

Anselin, L. (1988). *Spatial econometrics: Methods and models.* Dordrecht, NL: Kluwer Academic Publishers.

Attanasio, O. P., Bottazzi, R., Low, H. W., Nesheim, L., & Wakefield, M. (2012). Modelling the demand for housing over the life cycle. *Review of Economic Dynamics, 15*(1), 1-18.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution, 41*, 203-226.

Axtell, R. (2005). The complexity of exchange. *Economic Journal, 115(504)*, F193-F210.

Batty, M. (2005). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals.* Cambridge, MA: The MIT Press.

Batty, M. (2008). Fifty Years of Urban Modelling: Macro Statics to. In S. Albeverio, D. Andrey, P. Giordano, & A. Vancheri, *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach.* Heidelberg, DE: Physica-Verlag.

Batty, M. (2012). A Generic Framework for Computational Spatial Modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty, *Agent-based Models of Geographical Systems* (pp. 19-50). New York, NY: Springer.

Benenson, I., & Torrens, P. M. (2004). *Geosimulation: Automata-Based Modelling of Urban Phenomena;.* London, UK: John Wiley & Sons.

Bernard, R. N. (1999). *Using adaptive agent-based simulation models to assist planners in policy development: The case of rent control.* Santa Fe Institute Working Paper: 1999-07-052.

Billari, F. C., Prskawetz, A., Diaz, B., & Fent, T. (2007). The ”Wedding-Ring”: An agent-based marriage model based on social interaction. *Demographic Research, 17*, 59-82.

Birkin, M., & Wu, B. (2012). A review of microsimulation and hybrid agent-based approaches. In A. J. Heppenstall , A. T. Crooks, L. M. See, & M. Batty, *Agent-Based Models* (pp. 51-68). Dordrecht Heidelberg London New York: Springer.

Burgess, E. W. (1924). The Growth of the City: An Introduction to a Research Project. *Publications of the American Sociological Society, 18*, 85-97.

Center for Policy, Planning, and Evaluation. (2016). *2008-2012 Mortality Report.* Washington, D.C.: District of Columbia Department of Health. Retrieved from https://doh.dc.gov/sites/default/files/dc/sites/doh/publication/attachments/2012%20MORTALITY%20REPORT%20FINAL.pdf

Chaudhuri, G., & Clarke, K. C. (2008). A Decade of Cellular Urban Modeling with SLEUTH: Unresolved Issues and Problems. In R. K. Brail, *Planning Support Systems for Cities and Regions* (p. Chapter 3). Cambridge, MA: Lincoln Institute of Land Policy.

Combes, P.-P., & Duranton, G. (2016). The production function for housing: Evidence from France. *CEPR Discussion Paper No. DP11669*.

Consumer Financal Protection Bureau. (2016). *Ability-to-Repay and Qualified Mortgage Rule.* Washington, D.C.: Consumer Financal Protection Bureau. Retrieved from http://files.consumerfinance.gov/f/201603_cfpb_atr-qm_small-entity-compliance-guide.pdf

D.C. Fiscal Policy Institute. (2016). *The Housing Production Trust Fund.* Washington, D.C. Retrieved from www.dcfpi.org/wp-content/uploads/2016/04/16-04-HPTF-Brief.pdf

Dieleman, F. (2001). Modelling residential mobility; a review of recent trends in research. *Journal of Housing and the Built Environment, 16*, 249–265.

DiPasquale, D., & Murray, M. P. (2017). The shifting demand for housing by american renters and its impact on household budgets: 1940–2010. *Journal of Regional Science, 57*(1), 3-27.

District of Columbia Office of Tax and Revenue. (2016). *Computer Assisted Mass Appriasal.* Retrieved from OpenData.DC.Gov: http://opendata.dc.gov/datasets/

Edmiston, K. D. (2016). *Residential rent affordability across U.S. Metropolitan Areas.* Federal Reserve Bank of Kansas City.

Epple, D. (1987). Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products. *Journal of Political Economy, 95(1)*, 59–80.

Epple, D., Gordon, B., & Sieg, H. (2010). A new approach to estimating the production function. *American Economic Review, 100*, 905–924.

Epstein, J. M., & Axtell, R. (1996). *Gorwing Artificial Societies: Social Science from the Bottom Up.* Washington, DC, Brookings and Cambridge, MA, The MIT Press.

Ettema, D. (2011). A multi-agent model of urban processes: Modelling relocation processes and price setting in housing markets. *Computers, Environment and Urban Systems, 35*, 1-11.

Euijune, J. (2003). Growth and distributional impacts of urban housing supply: An application of urban land use and a CGE model for Seoul. *Review of Urban and Regional Development Studies, 51(1)*, 66–81.

Filatova, T. (2015). Empircal agent-based land market: Integrating adaptive economic behavior in urban land-use models. *Computers, Environment and Urban Systems, 54*, 397-413.

Filatova, T., Parker, D., & Van der Veen, A. (2009). Agent-based urban land markets: agent's pricing behavior, land prices and urban land use change. *Journal of Artificial Societies and Social Simulation, 12 (1)*.

Fisher, L., Pollakowski, H., & Zabel, J. (2009). Amenity-based housing affordability indexes. *Real Estate Economics, 37*(4), 705-746.

Glaeser, E., Gyourko, J., & Saks, R. (2005). Why have housing prices gone up? *American Economic Review, 95*, 329-333.

Guerrero, O. A., & Axtell, R. L. (2011). Using agentization for exploring firm and labor dynamics. In S. Osinga, G. Hofstede, & T. Verwaart, *Emergent Results of Artificial Economics* (pp. 139-150). Springer-Verlag Berlin Heidelberg.

Hass, P. M., Makarewicz, C., Benedict, A., Sanchez, T. W., & Dawkins, C. J. (2006). *Housing & transportation cost trade-offs and burdens of working households in 28 metros.* Chicago, IL and Blacksburg, VA: Center for Neighborhood Technology.

Hedberg, W., & Krainer, J. (2012). Housing supply and foreclosures. *Federal Reserve Bank of San Francisco Working Paper Series 2012-20*.

Heppenstall, A., Malleson, N., & Crooks, A. (2016). Space, the final frontier: How good are agent-based models at simulating individuals and space in cities? *Systems, 4(1)*, 9-27.

Huang, Q., Parker, D. C., Filatova, T., & Sun, S. (2014). A review of urban residential choice models using agent-based modeling. *Environment and Planning B, 41*, 661-689.

Hughes, S. A. (2014, May 15). D.C. Housing Authority Will Contact All 72,000 Applicants On Its Wait List. *DCist*.

Irwin, E. G. (2010). New directions for urban economic models of land use change: Incorporating spatial dynamics and heterogeneity. *Journal of Regional Science, 50*(1), 65-91.

Irwin, E. G., & Wren, D. H. (2014). An assessment of empirical methods for modeling land use. In J. M. Duke, & J. Wu, *The Oxford Handbook of Land Economics.* Oxford Univ. Press.

Jackson, J., Forest, B., & Sengupta, R. (2008). Agent-Based Simulation of urban residential dynamics and land rent change in a gentrifying area of boston. *Transactions in GIS, 12 (4)*, 475–491.

Joint Center for Housing Studies of Harvard University. (2014). *State of the nation's housing.* Cambridge, MA: Harvard University.

Jordan, R., Birken, M., & Evans, A. (2014). An agent-based model of residential mobility Assessing the impacts of urban regeneration policy in the EASEL district. *Computers, Environment and Urban Systems, 48*, 49-63.

Jordan, R., Birkin, M., & Evans, A. (2012). Agent-based modelling of residential mobility, housing choice and regeneration. In A. J. Heppenstall, A. T. Crooks, L.

M. See, & M. Batty, *Agent-Based Models of Geographical Systems* (pp. 511-524). Dordrecht: Springer.

Kuismanen, M., Laakso, S., & Loikkanen, H. A. (1999). Demographic factors and the demand for housing in the Helsinki Metropolitan Area. *Government Institute for Economc Reesarch, Discussion Papers, 191*.

Lee, B. H., & Waddell, P. (2010). Residential mobility and location choice: a nested logit model with sampling of alternatives. *Transportation, 37*, 587–601.

Lewis, J. M., & Krieder, R. M. (2015). *Remarriage in the United States.* Washington, D.C.: United States Bureau of the Census.

Lu, W., Stilwell, V., & Cannon, C. (2016). *The most and least affordable places to buy your first home.* Bloomberg.

Macal, C. M., & North, M. J. (2009). Agent-based modeling and simulation. In M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, & R. G. Ingalls (Ed.), *Proceedings of the 2009 Winter Simulation Conference*, (pp. 86-98). Austin, TX.

Magliocca, N., McConnell, V., Walls, M., & Safirova, E. (2012). Zoning on the urban fringe: Results from a new approach to modeling land and housing markets. *Regional Science and Urban Economics, 42*, 198-210.

Magliocca, N., Safirova, E., McConell, V., & Walls, M. (2011). An economic agent-based model of coupled housing and land markets (CHALMS). *Computers, Environment and Urban Systems, 35*, 183-191.

Mallach, A. (2016). *Modeling single family investor behavior: Las Vegas and Detroit.* Federal Reserve Bank of Philadelphia. Retrieved 2017, from http://docplayer.net/14227111-Modeling-single-family-investor-behavior-las-vegas-and-detroit.html

Mayer, C. J., & Somerville, C. (1996). Regional housing supply and credit constraints. *New England Economic Review*, 39-51.

McFadden, D. (1986). The choice theory approach to market research. *Marketing Science, 5 (Fall)*, 275-97.

Murphy, A. (2015). *A dynamic model of housing supply.* Working Paper.

Navarro, M. (2013, July 23). 227,000 Names on List Vie for Rare Vacancies in City's Public Housing. *New York Times*.

O'Sullivan, D. (2002). Toward micro-scale spatial modeling of gentrification. *Journal of Geographical Systems, 4*, 251-274.

O'Connell, M. (2007). A general equilibrium analysis of property tax in Florida. Unpublished dissertation.

Opportunity Index. (2016). Retrieved from Opportunity Index: How opportunity measures up in your community: http://opportunityindex.org/

Paciorek, A. (2013). *The long and short of household formation.* Finance and Economics Discussion Series 2013-26, Board of Governors of the Federal Reserve System.

Park, S., & Hewings, G. J. (2009). Immigration, aging, and the regional economy. *Cityscape, 11*(3), 59-80.

Parker, D. C., & Filatova, T. (2008). A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Computers, Environment and Urban Systems, 32 (6)*, 454-463.

Parker, D. C., Brown, D. G., Filatova, T., Riolo, R. L., Robinson, D. T., & Shipeng, S. (2011). Do land markets matter? A modeling ontology and experimental design to test the effects of land markets for an agent-based model of ex-urban residential land-use change. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty, *Agent-based models of geographical systems* (pp. 525-542). Dordrecht: Springer.

Partridge, M. D., & Rickman, D. S. (1998). Regional computable general equilibrium modeling: A survey and critical appraisal. *International Regional Science Review, 21*(3), 205-248.

Partridge, M. D., & Rickman, D. S. (2007). Persistent pockets of extreme American poverty and job growth: Is there a plced-based policy role? *Journal of Agricultural and Resource Economics, 32*(1), 201-224.

Peters, I., & Brassel, K. H. (2000). Integrating computable general equilbrium models and multi-agent systems - why and how. In H. S. Sarjoughian, F. E. Cellier, M. M. Marefat, & J. W. Rozenblit, *AI. Simulation and Planning in High Autonomy Systems* (pp. 27-35). Simulation Councils, Inc.

Plantinga, A. J., & Lewis, D. J. (2014). Landscape simulations with econometric-based land-use models. In *Oxford Handbook of Land Economics* (p. Chapter 15). New York: Oxford University Press.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Public Economics, 82*, 34-55.

Rosenthal, S. S. (2014). Are private markets and filtering a viable source of low-income housing? Estimates from a “repeat income” model. *American Economic Review, 104*(2), 687-706.

Rounsevell, M., Robinson, D., & Murray-Rust, D. (2012). From actors to agents in socio-ecological systems models. *Philosophical Transactions of the Royal Society B, 367*, 259-269.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics*, 1253-1296.

Savard, L. (2014). Poverty and income distribution in a CGE-household micro-simulation model: Top-down/bottom up approach. *Cahiers de Recherche*.

Schelling, T. C. (1969). Models of segregation. *American Economic Review, 59*(2), 488–493.

Schelling, T. C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology, 1*(2), 143-186.

Schwartz , M., & Wilson, E. (2008). *Who Can Afford To Live in a Home?: A look at data from the 2006 American Community Survey.* Working Paper.

Smajgl, A., Brown, D. G., Valbuena, D., & Huigen, M. (2011). Empirical characterisation of agent behaviours in socio-ecological systems. *Environmental Modelling and Software, 26*(7), 837-844.

Smith, N. (1979). Toward a Theory of Gentrification: A Back to the City Movement by Capital, not People. *Journal of the American Planning Association, 45:4*, 538-548.

Stone, M. E., Burke, T., & Ralston, L. (2011). *The residual income approach to housing affordability: The theory and the practice.* Boston, MA: University of Massachusetts Boston.

Sturtevant, L., & Fuller, S. S. (2011). *Housing the region's future workforce: Policy challenges for local jurisdictions.* Arlington, VA: George Mason University.

Tatian, P., & Williams, A. (2011). *A rent control report for the District of Columbia.* Washington, DC: The Urban Institute.

Tesfatsion, L. (2006). Agent-based computational economics: A constructive approach to economic theory. In L. Testfatsion, & K. L. Judd, *Handbook of Computational Economics* (pp. 831–880).

Tesfatsion, L., & Judd, K. L. (2006). *Handbook of Computational Economics Volumn 2: Agent-Based Computation Economics.* Elsvier B.V.

Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy, 64(5)*, 416-424.

Torrens, P. M. (2007). A geographic automata model of residential mobility. *Environment and Planning B: Planning and Design, 34*, 200-222.

Torrens, P. M., & Nara, A. (2007). Modeling gentrification dynamics:. *Computers, Environment and Urban Systems, 31*, 337-361.

United States Bureau of the Census. (2015). Table 1. quarterly rental vacancy rates: 1956 to 2015.

United States Bureau of the Census. (2016a). *County Population Totals Datasets: 2010-2016.* Retrieved from Population Estimates Program: https://www.census.gov/data/datasets/2016/demo/popest/counties-total.html

United States Bureau of the Census. (2016b). *Public Use Microsample. 2011 – 2015 American Community Survey.* Retrieved from http://ftp2.census.gov/>

United States Bureau of the Census. (2017). *Building Permits Survey State Annual Estimates.* Retrieved from https://www.census.gov/construction/bps/stateannual.html

United States Bureau of the Census. (2017). *Historical Table 1.* Retrieved from Housing Vacancies and Homeownership: https://www.census.gov/housing/hvs/data/histtabs.html

United States Department of Housing and Urban Development. (2015). *Worst case houing needs 2015: Report to Congress.* Office of Policy Development and Research.

United States Department of Housing and Urban Development. (2017, August). Retrieved from Income Limits: https://www.huduser.gov/portal/datasets/il.html

United States Department of Housing and Urban Development. (2017). *Picture of Subsidized Households*. Retrieved from https://www.huduser.gov/portal/datasets/assthsg.html

United States Department of Housing and Urban Development and United States Bureau of the Census. (2012). Rental Housing Finance Survey summary tables. Retrieved from https://www.census.gov/programs-surveys/rhfs.html

United States Department of Housing and Urban Development and United States Bureau of the Census. (2016). *American Housing Survey Public Use Files, 2011 and 2013.* Retrieved from https://www.census.gov/programs-surveys/ahs.html

United States Department of Housing and Urban Development and United States Bureau of the Census. (2017). *2015 Rental Housing Finance Survey Summary Tables.* Retrieved from https://www.census.gov/programs-surveys/rhfs.html

Urban Institute. (2017a). *DC Preservation Catelog*.

Urban Institute. (2017b). *Our Changing City*. Retrieved from http://apps.urban.org/features/OurChangingCity/housing/index.html.

Waddell, P. (2000). A behavioral simulation model for metropolitan policy analysis and planning: residential location and housing market components of UrbanSim. *Environment and Planning B: Planning and Design, 27*, 247 ^ 263.

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

Waddell, P. (2010). Modeling Residential Location in UrbanSim. In F. Pagliara, J. Preston, & D. Simmonds, *Residential Location Choice: Models and Applications* (pp. 165-180). Verlag Berlin Heidelberg: Springer.

Waddell, P. (2011). Integrated land use and transportation planning and modelling: addressing challenges in research and practice. *Transport Reviews, 31(2)*, 209-229.

Washington, D.C. Office of Planning. (2017). *Ameding the DC comprehensive plan*. Retrieved from http://plandc.dc.gov/

Wheaton, W. D. (2014). *2010 U.S. Synthetic Population Ver. 1.* RTI International.

Wheaton, W. D. (2014). *U.S. Synthetic Population Database 2010: Quick Start Guide.* RTI International. Retrieved from https://www.epimodels.org/midasdocs/SynthPop/2010_synth_pop_ver1_quickstart.pdf

White, R., & Engelen, G. (1993). Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land use patterns. *Environment and Planning A, 25*, 1175-1199.

Woodwell, J. (2015). *Mind the gap: A high-level review of the need for - and suppy of - affordable multifamily rental housing.* Mortgage Bankers Association. Mortgage Bankers Association.

Wu, F. (2002). Calibration of stochastic cellular automata: the application to rural-urban land conversion. *International Journal of Geographical Information Science, 16*, 795-818.

Yang, X., & Lo, C. (2003). Modeling urban growth and landscape change in the Atlanta metropolitan area. *International Journal of Geographical Information Science, 17*, 463-488.

BIOGRAPHY

Shawn J. Bucholtz received his Bachelor of Science in Public Resource Management from Michigan State University in 2000. He went on to receive his Master of Science in Agriculture and Resource Economics from the University of Maryland in 2003. He is currently employed as the Director of the Housing and Demographic Analysis Division at the United States Department of Housing and Urban Development.