Spring 5-2017

An Agent-Based Model of Urban Sprawl: York and Cumberland Counties, Maine

Kaitlyn G. Lavallee
University of Maine

AN AGENT-BASED MODEL OF URBAN SPRAWL:

YORK AND CUMBERLAND COUNTIES, MAINE

by

Kaitlyn G. Lavallee

A Thesis Submitted in Partial Fulfillment
of the Requirements for a Degree with Honors
(Mathematics and Economics)

The Honors College

University of Maine

May 2017

Advisory Committee:
Richard Barringer, Research Professor Emeritus, Muskie School of Public Service
James Breece, Associate Professor of Economics, Advisor
Kathleen Ellis, Lecturer in English and Adjunct Assistant Professor in Honors
David Hiebeler, Associate Professor of Mathematics
Evan Richert, Town Planner, Orono

ABSTRACT

Urban sprawl is defined as the movement of populations towards the fringes of urban centers, leading to the conversion of rural land to suburban consumption. This expansion in the distribution of populations has many implications for local and state policymakers, business owners and consumers. In Maine, sprawl is particularly prevalent in Cumberland County and York County, where the state's population is the densest. The objective of this paper is to develop an agent-based model (ABM), which attempts to reflect the movement of households within these counties. These households make decisions sourced in microeconomic theory that are built into the model. Households seek locations that maximize utility, based on their income and time constraints. This model also incorporates a gravity model of migration to determine the likelihood that a household will migrate to another area when motivated by a higher income. Simulations of the model display characteristics of sprawl, including a decline in population density. Additionally, several policy simulations were conducted to demonstrate the effects on land use and projected population migration patterns. This model serves as a basis for future exploration and customization to forecast land-use trends, as well as the corollaries of potential economic policies or development

PREFACE

The model that was built to accompany this paper is experimental and serves as a basis for future exploration. This explorative study attempts to examine the strength of predictability utilizing economic and social motivations for migration (available housing, median gross rents, median household incomes, crime rates and median commute times) and the results of adjustments to these variables in potential policy scenarios. The scope of the model was limited to these five factors, however there is an abundance of additional incentives that can be added to future model adaptations as data is acquired or become available. With the inclusion of supplementary data, the land use and population distribution projections would substantially increase the model's accuracy and representation of the true trends of the area. This model is a foundation for customization, supplementation and application, based on the needs and inquiries of future research and analysis.

TABLE OF CONTENTS

TABLE OF FIGURES

INTRODUCTION AND LITERATURE SURVEY

To many people, living in a suburb provides the benefits of a rural and urban lifestyle without many of the drawbacks. Suburbs typically offer lower housing costs and lower crime than urban areas, as well as a variety of property options, while the broader employment opportunities, attractions and amenities of nearby cities are close enough to utilize. From 1980 to 2000, 869,000 acres of rural land in Maine became categorized as suburban, which is approximately the same amount of total land in Rhode Island (Katz, 2006). This development of rural land has continued and expanded throughout the state, although most prominently along the southern coast where the Maine population is most populous and densest, becoming sparser in more northern and western counties (Mattingly and Schaefer, 2012). This paper focuses on two southeastern coastal counties, Cumberland and York, and the behavior of households within these regions.

The conversion of land from rural to suburban usage, often surrounding a more urbanized area, is referred to as *urban sprawl*. Until the last few decades, sprawl was difficult to define and quantify. This is likely because sprawl was understood as a "combination of its causes (e.g., zoning and poor planning), characteristics (e.g., low-density development), and effects (e.g., traffic congestion and air pollution)" (Hess et al., 2001). The majority of modern interpretations and models of sprawl utilize population density through time to measure these development patterns. For instance, the population density of Portland, Maine is 1271.5 people per square mile, whereas the population density of Portland, Oregon is 4360.75 people per square mile (Census). Therefore, one can conclude that since Portland, Maine is less dense, the population distribution experiences greater sprawl within the city.

Because urban sprawl describes the growth and expansion of populations, time is an essential consideration. When examining sprawl purely through static measurements like population density, the dynamic interactions of the populations are ignored and areas can be easily be misclassified as sprawling when they are merely naturally and appropriately expanding areas of a city (Harvey and Clark, 1965). Including the factor of time into an analysis of sprawl also allows for identification of population distribution changes, comparisons to previous conditions, and opportunities for future projections. Population density can help identify patterns of sprawl in comparison to other areas, but comparing patterns of sprawl in one area over time can allow for more compelling results, understanding and accuracy. Therefore, even though Portland, Oregon has a higher population density than Portland, Maine, if Portland, Oregon is becoming less dense over time at a greater rate than Portland, Maine, then Portland, Oregon is believed to be sprawling at a greater rate (Hess et al., 2001).

Causes and Consequences

Although many factors contribute to urban sprawl nationally and internationally, Maine has several specific traits that encourage sprawl. Large property tax variations between older and newer towns, as well as escalating housing prices in central urban locations, have been encouraging residents to move further away from regional hubs into developing suburbs (Katz, 2006). Financially, moving toward the fringes of Maine urban areas is more affordable than living in cities like Portland or Lewiston. The development of rural areas is also favored to the redevelopment of older and historic structures. State and local regulations and rules, including zoning ordinances and building codes, involved in working with established buildings or areas have built up over time and become so strict and extensive that it is simpler and more affordable to develop rural sites (Richert, 1997). Additionally, as the Maine public continues to migrate away from urban centers, a growing number of citizens live in one town and work in another. The common disconnect between town management and absence of regional collaboration has led to disorganized and costly sprawl patterns (Katz, 2006). Although many Maine residents are living and working in separate towns within a region, based on personal preferences, careers, and the varying costs of living in each town, the frequent and past lack of regional planning has led to a disconnect in regional growth patterns, encouraging sprawling populations.

Urban sprawl has been consistently researched and considered in academia, news articles, policy decisions, and public discussion since the mid-twentieth century. Environmental concerns, high government and social costs, and even negative health impacts have been attributed to sprawl. As populations spread and rural areas are developed, social infrastructure is demanded in budding suburbs. Road maintenance, public safety, town management, and schools are required to support the developing area. Between 1996 and 2006, it is estimated that Maine spent $200 million constructing schools in new Maine suburbs that were required by the dispersing population (Katz, 2006). In addition to the financial expense, one of the most frequent criticisms of urban sprawl is the negative implication on the environment. The conversion of rural landscapes into developed suburbs and neighborhoods naturally results in the loss of farmland, natural habitats and forestry. Greater levels of air and water pollution have also been attributed to urban sprawl since the dispersed populations can encourage and require a higher usage of motor vehicles. Pollutants from these vehicles can impede plant and animal development, contaminate the air and water supplies, contribute to global warming and lead to human health complications (Bhatta, 2014). Where Maine's economy heavily relies on tourism involving natural attractions and wildlife, higher pollution and destruction of rural land could negatively influence this industry. Furthermore, there have been a variety of other researched negative consequences of urban sprawl, including lower physical activity and higher levels of obesity and chronic disease, since residents of sprawling suburban areas typically walk less and have a greater reliance on motor vehicles than their urban counterparts (McCann and Ewing, 2003).

When living among a more dense urban population, many amenities, including health services, social interests, restaurants, and shopping, are often available within walking distance. As households become more dispersed, to take advantage of these resources, either public transport or personal vehicles become a necessity. However, cars require an expensive initial cost and maintenance, which have led many to believe that the increasing reliance on automobiles due to sprawl is increasing social segregation. The wealthier community members are able to afford to live and commute from sprawling areas and suburbs, while the poor are concentrated near public transportation stops and downtown urban districts (Glaeser and Kahn, 2003). However, analysis of this potential result of sprawl has been mixed. Kahn (2001) found that more affordable housing is often available in suburban areas, which is potentially reducing the segregation in housing consumption.

Despite the potential consequences of sprawl, suburban residences remain attractive to many. The opportunity to live in single-family households and own sizable plots of land is not as abundantly available in urbanized areas. Suburbs can often provide a less population dense, safer, and quieter environment than cities. With the development of new suburban schools in up-and-coming Maine suburbs, families may seek out modern public education facilities that are inaccessible in Maine urban areas. As long as these suburban features remain available and demanded, sprawl will likely continue. Understanding how and where these populations will develop and spread is essential to appropriate land-use planning. The model I have built and will describe throughout this paper will allow for these projections, as well as the opportunity to simulate the effects of various land use policy measures. Providing the demanded suburban lifestyle while minimizing negative externalities can benefit Maine residents, government, and the environment.

An Overview of Agent-Based Modeling

Sprawl and land-use have been modeled for over a century, using a variety of modeling techniques. Traditional models often relied on uniform and rational agents (representations of individuals or groups) who were acting on spaces lacking geographical characteristics. This did not represent interactions between humans and the environment, human perceptions or feedbacks that alter patterns of sprawl (Brown and Robinson, 2006). Due to the limitations of traditional models, the model described in this paper was developed using the more contemporary approach of agent-based modeling (ABM).

Agent-based modeling is a method that allows for the creation, analysis, and simulation of the interactions between individual agents and the comprehensive effects that these interactions and agent decisions have on the overall environment (Gilbert, 2007). Since sprawl is driven by human behaviors and decisions, these can be appropriately incorporated into agent-based models to encourage accurate results. Additionally, it is necessary to ensure the assigned heterogeneous qualities of agents in agent-based models accurately reflect the “real world” characteristics (Brown and Robinson, 2006). If true agent attributes or behaviors are not represented in the model, the outcome of their interactions will not be accurate. The assigned characteristics and behaviors of the agents in this model will be further detailed in the “Materials and Methods” section.

Economic Drivers of Sprawl

The drivers of local, state, and international migration have been a heavily researched topic. Financial motivation in the form of income is a widely accepted variable that motivates relocation. Researchers have found that households often seek the greatest exchange between costs of land and costs of commuting, so wealthier households will live in suburban areas because their income elasticity of demand for housing and land consumption is greater then their income elasticity for their commuting costs (Wu, 2006). Therefore, the variation in incomes can allow for a wider selection of housing locations, or incomes can serve as a constraining factor and limit regional options. As previously described, Maine residents are encouraged to move away from urban centers towards suburban, and rural areas because housing prices and property taxes are more affordable in these areas (Katz, 2006). Since wealthier Maine households can better afford to remain in high-taxed urban areas or consider other suburban options, it is important to consider how their community needs and desires will affect these decisions.

Social Drivers of Sprawl

While there is a wide variety of personal motivations and preferences when citizens choose where to live, the data regarding the social incentives and priorities of living in suburban versus urban versus rural areas are not always readily available. For example, some people prefer to live in single-family homes on a secluded plot of land, while others prefer to live in apartments located close to the downtown hub of a city. Although currently unavailable, these preferences could be quantified through conducting regional surveys and added to this model or future models to better represent the populations' personal motivations.

Crime rates and commuting times were incorporated into the model to ensure that community or social-based incentives were represented. Cullen and Levitt (1996) found that a 10% rise in crime in an area correlates to a 1% decline in local population. As anticipated, higher crime rates would encourage out-migration to localities with lower crime rates. There have been debated results in research relating commuting time and sprawl. As populations spread out, places of employment become more decentralized. As a result, this can lead to a reduction in commuting times, however sprawl can also lead to greater road congestion and increase commuting times into the areas that are more concentrated with businesses and employment opportunities (Zolnik, 2011). Although sprawl can affect commuting times in a variety of ways, it is expected that households seek to minimize their commute times and that this would be an influential factor when choosing a housing location.

METHODS AND PROCESSES

Maine is demographically unique when compared to other New England states as well as other regions of the country. Maine's percentage change in population was 0.1% from 2010 to 2015, which is drastically lower than the overall United States' 4.1% population percentage change from 2010 to 2016 (Census). The relatively stagnant growth in Maine's population can be attributed to several factors, including Maine's having one of the nation's oldest populations. From 2012 to 2014, annual deaths have outweighed annual births in the state, slowing the rate of natural increase (births-deaths) (Maine Department of …and Vital Statistics, 2015). With a negative rate of natural increase, Maine's positive net migration is responsible for the slightly positive overall change in population. *Net migration* is defined as the difference between how many people are moving to and from each state. However, with a 0.1% growth rate, Maine's population is being approximately sustained, rather than grown. For this reason, the factors of natural increase and state net migration are not included in this model. Therefore, the total number of occupied households in the model is considered an exogenous variable. However, the total number of households (including vacant houses) is an endogenous variable that is expanded upon in "Patch Variables."

The agent-based model considers households as the only type of agent. These households are motivated to find a location that minimizes their distance from their starting location in accordance with a gravity model of migration (see "Gravity Model of Migration and Household Utility Maximization"). In addition, the households attempt to maximize utility, where utility is a function of median gross rent, median household income, median commute time, and crime rate. The households also have a budget constraint in which their annual rent cannot exceed 30% of their yearly income.[1] Each household has four attributes, including agent name, stay counter, job, and income, which will be further detailed in "Household Variables."

The model's landscape is created from loading a GIS shapefile containing polygons describing the town lines for each town in the counties. *Ticks* represent time in the model, where each tick is one year. The landscape is made up of patches with an area of four pixels. Each patch is approximately 1/10th of a square mile and each household is the size of two pixels. Initially, each household is created on its own patch; however, multiple households can share a single patch. Additionally, every patch contained within a town's borders is assigned the town's name. All patches contain seven other assigned variables that are unique for each town, including total number of households, number of occupied households, total area of the town, median household income, gross median rent, crime rate and median commute time.

Gravity Model of Migration and Household Utility Maximization

This model implements a combination of a gravity model and household utility maximization. The gravity model of migration attempts to relate the gravity law of spatial interaction to economic migration. The relationship is expressed in the following manner:

$$F = G\frac{P_i P_j}{D_{ij}^2}$$

[1] A common indicator of housing affordability is that housing expenditures do not exceed 30% of household income. This can be altered in future models adjustments.

where F is the demographic force, G is a constant, P_i is the population of location i, P_j is the population at location j, and D_{ij} is the distance between location i and j (Greenwood, 2005). The gravitation migration model has been modified to use other parameters instead of population, such as the GDP of two locations (Ramos, 2016). In this model, the income of the two locations will be utilized and an error term will be added to account for factors not within the scope of the model. The implementation of the gravity model can be expressed in the following manner:

$$F = C_m \frac{I_m I_h}{D_{hm}^2} - \varepsilon$$

where C_m is the crime rate of the prospective town, I_m is the gross median income of the prospective town, I_h is the current gross income of the household, and D_{hm} is the distance between the prospective patch in the town and the household. The error term, ε, is randomly generated from a normal distribution with a user selected mean and standard deviation. If the F term is greater than zero, the patch will then be subjected to a household utility optimization.

The household utility function used in this model is the neoclassical model of labor-leisure choice. This can be summarized by the following function:

$$U = f(C, L)$$

where C is the consumption of goods and L is the consumption of leisure (Borjas, 2016). A household will attempt to maximize both factors. For simplicity of the model, it is assumed that the consumption of goods is a function of median monthly rent, represented by the variable R, and household income, represented by I_h:

$$C = f(I_h, R)$$

In addition, it is assumed that all households work the same number of hours and that leisure is only a function of the span of time a household will wait before attempting to move, represented by W, and by minimizing the commute to work, represented by T:

$$L = f(W, T)$$

This can be combined into an overall utility function summarized in the following manner:

$$U = f(T, W, I_h, R)$$

The utility function is maximized in a piecewise function with preference for maximizing consumption of goods over consumption of leisure. Initially, each household is assigned a span of time that the household must wait before moving, which will be represented by W. A random number α, derived from a normal distribution with a mean and standard deviation of 1, is subtracted from W for each year within the model. A random number is used to represent factors not accounted for within the model that would cause a household to either move more quickly or stay at a given location for a longer period. This is summarized by the identity:

$$P = W - \alpha$$

where P is the countdown to initiating a job search. Once P is equal to or less than zero, the household is then offered a job at a patch with a higher median wage than the households current wage. The patch containing the job is then subjected to the gravity migration model. Assuming F is larger than zero, the household accepts the job and searches for a location to live. The household then seeks to find a patch within the average commuting time with the lowest rent. Households give preference to a town with lower median rent over a shorter work commute. The worker is then assigned a “job” and an “income” based on the procedures outlined in the section “Household Variables.”

In addition to the above stipulations, there is an income constraint that yearly rent cannot exceed 30% of yearly income.[2] This is summarized by the following constraint:

$$I_h \geq 36R$$

If the minimum rent within the average travel time does not follow the constraint, the household will be unable to accept the job. Regardless of whether the households accept the job and move, their W is reset.

Data Acquisition & Programs

The data sources and necessary adjustments are described in the following table for each of the 58 towns between the two counties. Data was collected for the 2010, because this was the earliest complete data set available. The model was built and executed utilizing Microsoft Excel and the modeling program NetLogo.

[2] A common indicator of housing affordability is that housing expenditures do not exceed 30% of household income. This can be altered in future models adjustments.

Variable	Data Adjustments	Source
Housing Units	The number of total, occupied and vacant households for each town in Cumberland County, ME and York County, ME was collected.	Census Quick Facts (2010)
Median Gross Rent	The median gross rent for each town in Cumberland County, ME and York County, ME was collected.	Census Quick Facts (2010)
Median Household Income	The median annual income for each town in Cumberland County, ME and York County, ME was collected.	Census Quick Facts (2010)
Median Commute Time	The aggregate median commute time for each town in Cumberland County, ME and York County, ME was collected. The aggregate median commute times were then converted to one-way commute times.	Census Quick Facts (2010)
Crime Rate	The Crime Rate is based on the occurrence of an Index Offense per 1,000 residents of the state. Local and county rates are based on their individual populations. The crime rates were collected for each town in Cumberland County, ME and York County, ME that was recorded.	Crime in Maine Reports (Theriault et al., 2010)

Table 1: Dependent and Explanatory Variables

Household Variables

Households are assigned a name based on the patch they are on. The stay-counter is set by the user and can be manipulated to indicate an integer value that represents on average how many years will pass before a household will consider moving. Initially, a household will be assigned a random number between 0 and the chosen integer value. This random number is generated from a discrete equal distribution function, meaning that all values between 0 and the chosen value have an equal likelihood of being selected. The randomly generated number is intended to simulate how it is unknown when the households last considered moving. Once a household has moved, the stay-counter reverts back to the user-selected value. In addition, every turn that that stay-counter is above zero, it is reduced by a random number. This random number is generated from a normal distribution with a mean of 1 and a standard deviation of 1. This distribution was chosen to represent variables not measured by this model and how some households may choose to remain in their current town, while others may move much sooner than average.

The "job" assigned to each household is a random number between 0 and 1, which will be used to generate a household income through each town's income function. The random number is generated from a continuous equal distribution function. This distribution is used, because the income function for each town is constructed in such a way that a randomly distributed variable best reflects the actual distribution of income for each town. To construct the income function, the percentage of people in a town that falls into each income interval was collected. The income intervals were then averaged to determine the mean and the percentages were converted into cumulative percentages. The data was then fit with a second-degree polynomial to most appropriately represent the income distribution for each town. An example of a town's income distribution is shown in the figure below.

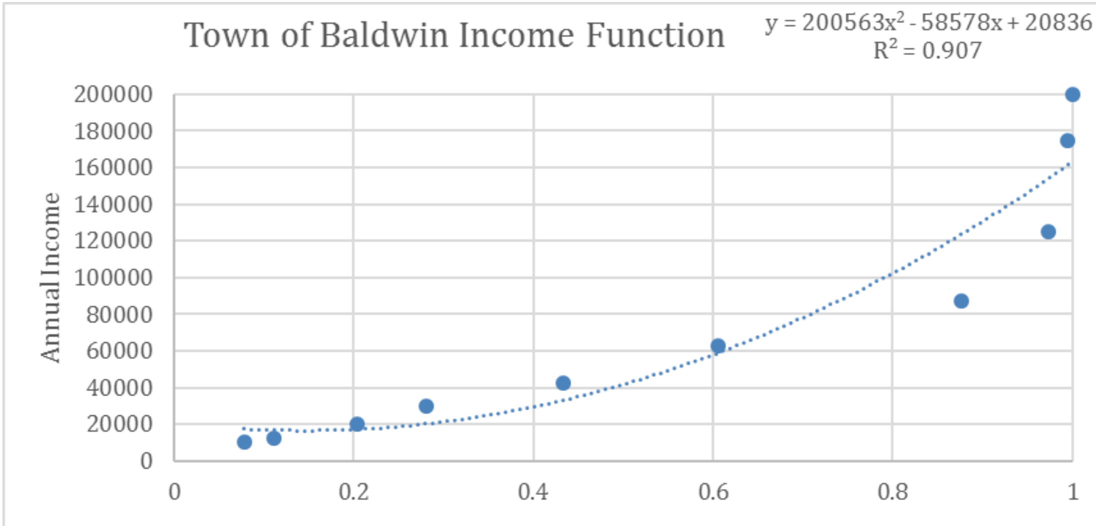

Figure 1: Example Income Regression

<u>Patch Variables</u>

The name of the patch, the number of total and the number of occupied households is initially assigned when the GIS shapefile is loaded into NetLogo. Each time that a household moves, the number of occupied households in each town is updated. Similarly, the number of total households changes by the following correlation.

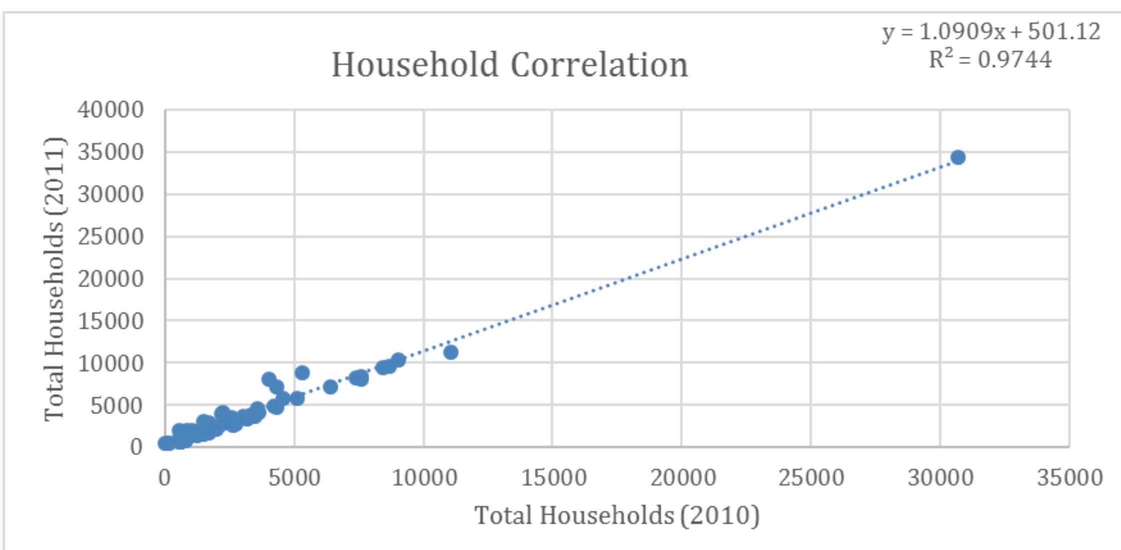

Figure 2: Linear Equation Describing Household Growth

The R-squared value demonstrates that the number of total households can explain 97% of the variance for the number of total households in the previous year. This is logical, as an increase in demand for households will cause an influx in the number of households being built.

The assigned area of each town remains consistent throughout the model. However, the median household incomes are initially set, and once the model is run, the median household incomes are updated based on the profiles of the new full set of households in each town. Median gross rent is considered an exogenous variable, meaning this variable does not change throughout the course of the model. Several attempts were made to explain the variation in rent within the scope of the model, however none were successful.[3]

The "crime rate" is an endogenous variable that is a function of occupied households per square mile. The correlation is described in the following figure.

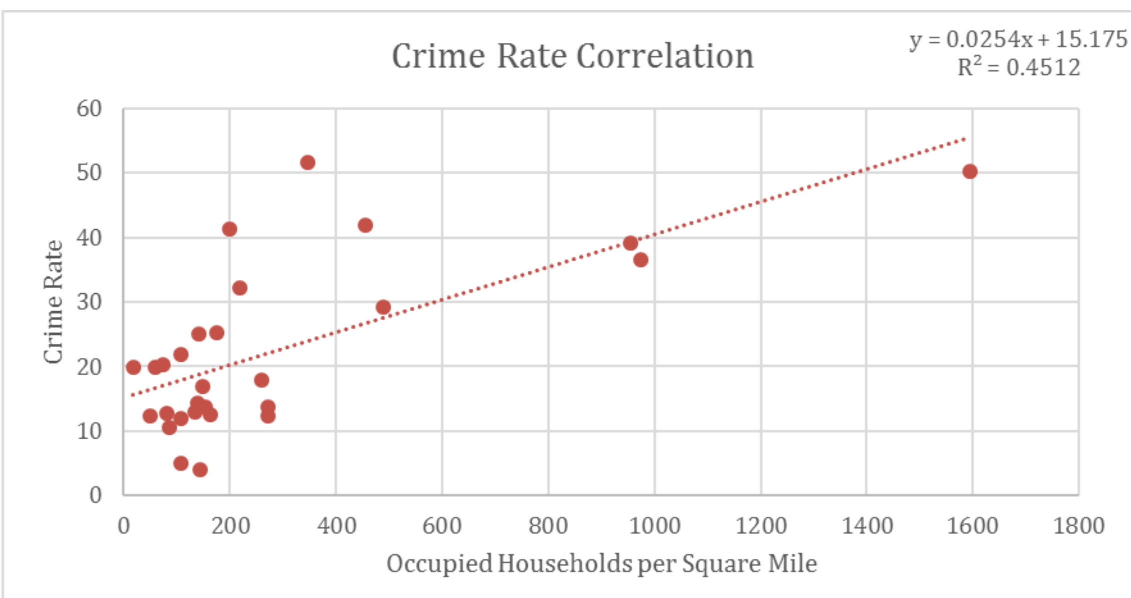

Figure 3: Linear Equation Describing Crime Rate With Respect To Occupied Household Density

[3] See Figure 9 and Figure 10 in the appendix, which illustrate the poor correlation between the variation in median gross rent and "Vacant Houses Per Square Mile" and "Percent of Houses Vacant."

As mentioned above, each year the number of occupied households for each town is updated. Since crime rate is a function of occupied households per square mile, the updated number of occupied households is used to derive the towns' estimated crime rates. Despite the mediocre R-squared value of 0.45, this is the only correlation for crime that can be derived within the scope of the model.

RESULTS

Model Validation

The model is extended for five ticks, simulating five years of population migration. This simulation was repeated five times, and the data was then aggregated to determine the mean, which can be seen for Cumberland and York counties in figures 11 and 13 in the appendix. The aggregate data compiled from this simulation was then compared to the 2015 census data to determine the model's accuracy. In figure 4 below, the shade of the patches represents the number of occupied households within each town, where the darker shade indicates a greater number of residents.[4] The model was found to have overestimated the population growth in many of the towns. Potential causes for this overestimation will be expanded upon in "Recommendations for Future Manipulation." This was particularly pronounced in Portland, Biddeford, Brunswick, Old Orchard Beach, and Sanford.

[4] The shade of the patches was chosen to indicate overall household population rather than household population density because the variation of density in the majority of the area was nominal.

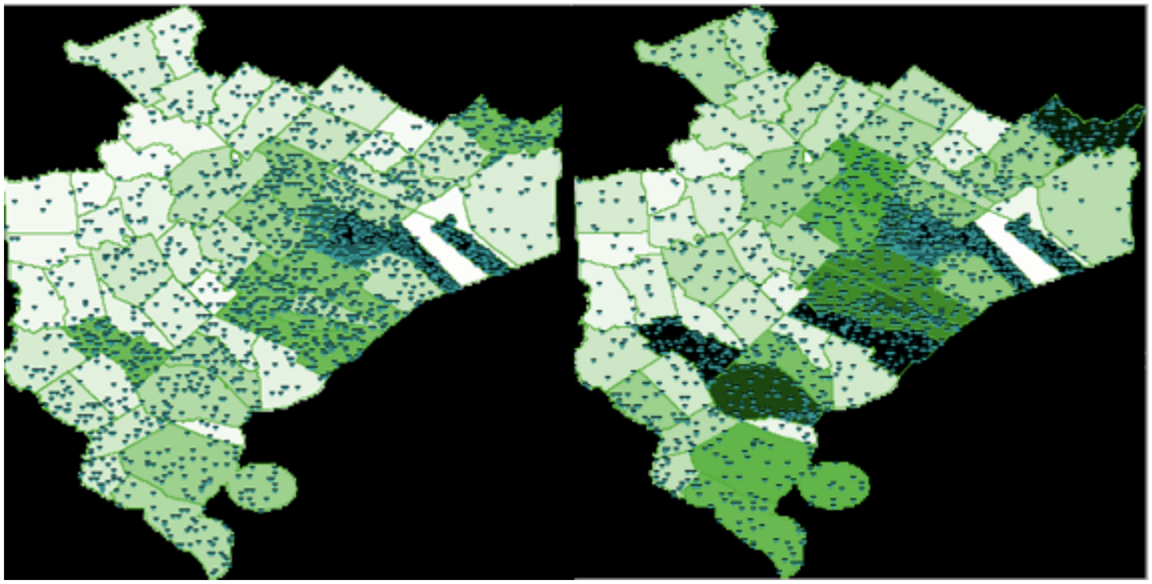

Figure 4: Actual Migration (Left) vs. Tested Migration (Right)

However, the modeled population, on average, did not deviate from the 2015 Census population by more than 20%. There are several reasons why the model may have overestimated the town population growth. As previously noted, the error term, ε, is selected from a normal distribution. In the initial model specification, the normal distribution has a mean of 300 and a standard deviation of 100. This specification may have been too low, allowing too many households to migrate to a different job. The mean and standard deviation are adjusted in "Policy Simulations" to determine the effects on the number of households that move each year. Other factors that could have led to imprecision in the model are the excluded potential social drivers, other financial variables and consumer preferences. This will be further detailed in "Recommendations for Future Manipulation."

<u>Policy Simulations</u>

Although three types of policies will be explored to determine their effect on migration, similar simulations can be extended, adapted, and created for individualized needs and locations. The first policy to be explored is the creation of several housing developments in Biddeford. During initial tests, it was found that many households attempted to live in Biddeford due to its proximity to Portland and comparably inexpensive rent. However, due to housing availability restrictions, the households were unable to move to this location. The number of available houses was increased by 100% to determine the effect on migration for Biddeford and the surrounding area. While doubling the number of total households is not necessarily a realistic increase, this is intended to showcase the effects of housing developments in the suburban fringes of urban areas, where housing options are often more affordable.

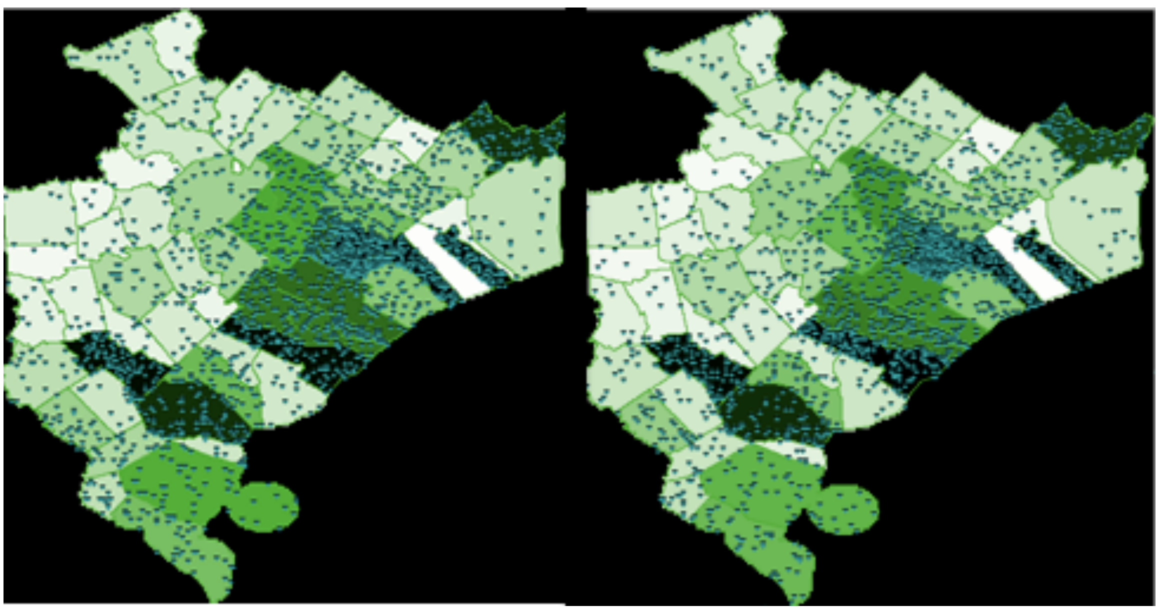

Figure 5: 9500 Houses (Left) vs. 18000 Houses in Biddeford (Right)

Biddeford's location nearby towns with higher median household incomes and its comparatively low gross median rent leads it to be a highly desired living location, as estimated by this model. This results in the model predicting that there will be less than 100 vacant homes in Biddeford. Although this indicates that an increase in available housing would greatly attract new residents, the extent of the attraction is likely inflated since gross median rent was not an endogenous variable, as previously explained. Therefore, the increased demand for housing was not inflating the gross median rent, leading to higher than expected housing consumption. Furthermore, the inclusion of additional consumer preferences would also prevent unrealistic growth in Biddeford's household population.

The second policy that was examined was a reduction in the mean and standard deviation of the normal distribution from which the error term, ε, is selected. A decrease in this value would represent any situation that would lower the barrier of households moving from locations with lower median incomes to towns with higher median incomes. This could be a result of many different policies, including investment into workforce training, subsidies for higher education, or the opening of large companies that would offer more employment opportunities to local residents. In the context of the model, decreasing the value of epsilon should allow for a higher percentage of the population to move in any given year. As seen below in figure 6, multiple towns on the right side have lighter shades, indicating that more households have sprawled outward in comparison to the model on the left side.

Figure 6: Mean = 300, Std. dv = 200 (Left) vs. Mean = 100, Std. dv = 100 (Right)

The final policy simulation explored the effects of increasing the median household income in a specified town. This could be the result of an increase in college-educated workers or an increase in higher compensating or new employment opportunities in the area. To implement a higher average household income in a town, the income function for the town must be modified. Dayton was chosen due to its proximity to Portland and its relatively low number of occupied households. The first term in Dayton's income function was increased from 221966 to 330000. As seen below in Figure 7, this has a substantial effect on the residents' average incomes.

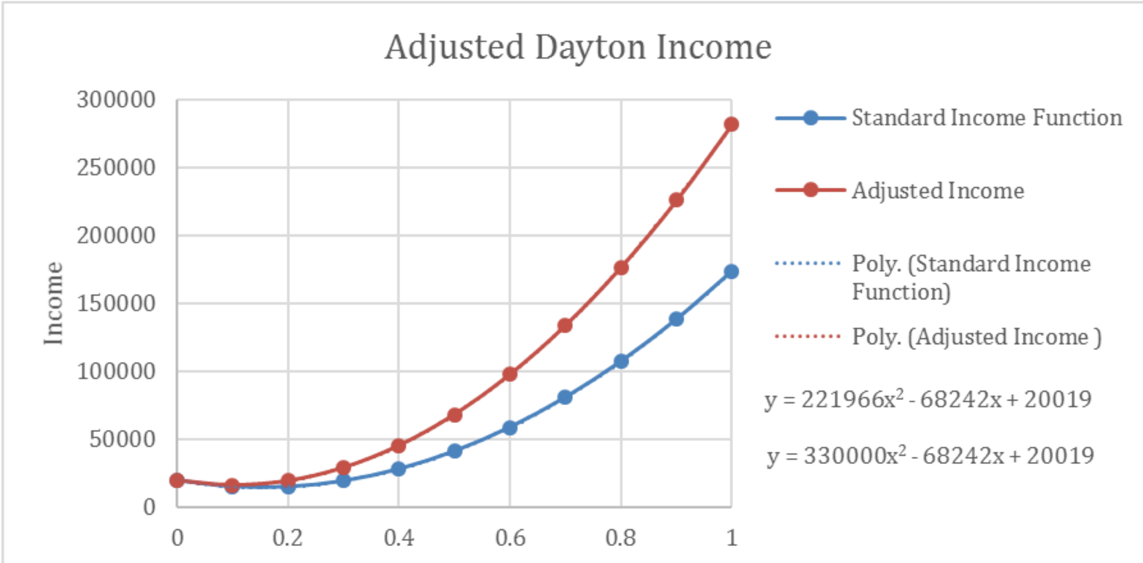

Figure 7: Polynomial Describing Dayton Income

Figure 8 demonstrates that despite an increase in median income of Dayton, there was not an increase in household population. This could be a realistic outcome if there are not many jobs available in the area or residents of other towns are not attracted to the town for other aspects of the locality. However, in the model, the locations of potential new jobs were randomly chosen from all patches with a higher median income than the household's current income. This randomization was used to account for the many unobservable factors that affect whether a household chooses to accept a position in the new location.

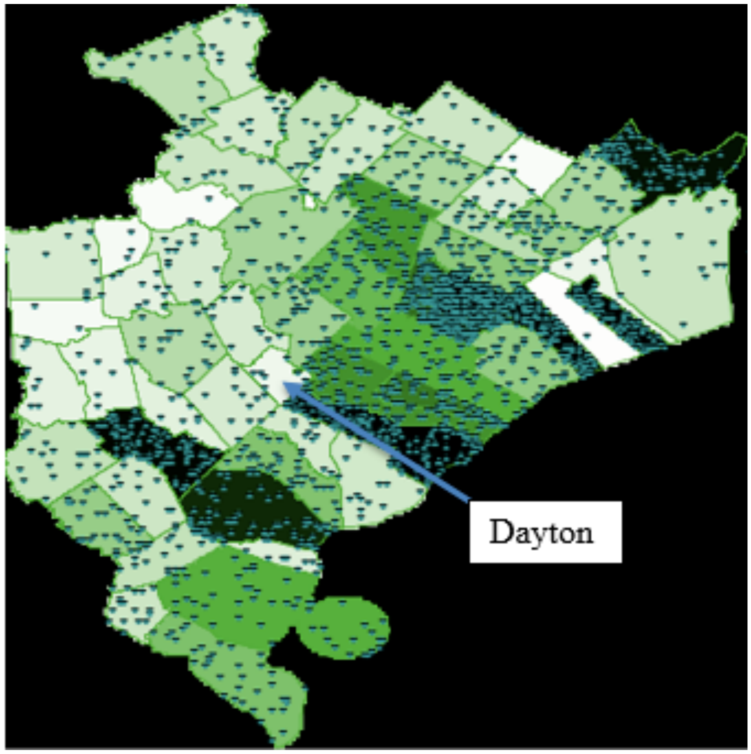

Figure 8: Dayton with Increase Income

Overall, while these policy models may not be entirely accurate, they still allow insight into how potential policy changes can influence urban sprawl. For example, while it may be unrealistic that Biddeford would experience little to no housing vacancies, it still demonstrates that the construction of high density housing projects can be used to maintain low gross median rent while allowing for households to work in higher median income areas.

Recommendations for Future Manipulation

The next step for this model would be to convert gross median rent from an exogenous variable to an endogenous variable. Additionally, it may be beneficial to incorporate the total number of occupied households in a town into the town's income function. Many of the issues that arose in this model were seemingly due to both gross median rent and median household income not adjusting to changes in population.

However, as rent had no correlation with the density of either occupied or vacant households, it is possible that consumer choice and location of properties had the largest effect on gross median rent. In addition, gross median rent could potentially be dependent on the density of vacant homes as a piecewise function, where towns, such as Portland, Biddeford, Brunswick, Scarborough, and Old Orchard, with high population density and demand for housing would be affected by the density of vacant homes whereas smaller towns would not be since the demand for housing is much lower.

In addition, income could be examined to determine the effect of an increasing population on a town's median household income. It would be expected that as the supply of workers increase, the median household income should decrease. However, much like gross median rent, this correlation may not exist or it may only exist in towns were the job market is competitive.

The inclusion of additional social migration motivators would be a beneficial addition to the model, should this data become acquired or made available. This could account for the desire of proximity to established social networks of friends or family, the appeal of superior school districts if the household contains children or the attractiveness of housing location based on individual preferences (i.e. single-family homes versus apartment buildings, walking distance to town amenities, etc.).

Although outside the scope of this model, other potential opportunities for further research could include the implementation of road layouts, current housing and other structures and geological features to help project more specifically where a sprawling population will most likely build businesses, homes and schools within each of the specific towns.

CONCLUSIONS

Although there are opportunities for expansion and development, this experimental agent-based model provides valuable information and a constructive foundation for modeling urban sprawl in York County and Cumberland County, Maine. The modeled household populations for each town, on average, did not deviate from the 2015 Census population by more than 20%. This demonstrates that the households' decision-making process, which reflected the microeconomic theory examined in the literature review and processes description, can predict a general trend of population fluctuations. However, population growth was not accurately projected in the model, likely due to potentially significant variables that were omitted from the model as a result of insufficient data availability. Additionally, the interaction between median gross rent and population density was unexpectedly negligible, so the standard dynamic nature of the median gross rent variable was not included within the model, impacting migration patterns of the households.

Based on the policy case studies, the model demonstrates potential methods of influence on population distribution patterns. Understanding the effects of policies on basic economic variables, such as housing costs or median incomes, can help determine and project the intended and unintended land use results in the short and long term. Therefore, policy-makers and town planners can potentially alter the coefficients, exogenous and endogenous variables to forecast the effects on urban sprawl. For example, as shown in the Biddeford case study, the construction of high density suburban housing developments that maintain inexpensive rent and allow households to utilize the amenities of the nearby urban area, can influence the pattern of migrations, while also considering the demands of the local market. Although this is an expected result, the model has the potential to provide an analytical and visual resource when researching policy results, particularly with the addition of local data.

BIBLIOGRAPHY

Bhatta, B. *Analysis of urban growth and sprawl from remote sensing data*. Heidelberg: Springer, 2014. Web. 14 Apr. 2017.

Borjas, George J. Labor economics. 6th ed. Singapore: McGraw-Hill Education, 2016. Print.

Brown, Daniel G., and Derek T. Robinson. "Effects of Heterogeneity in Residential Preferences on an Agent-Based Model of Urban Sprawl." *Ecology and Society* 11.1 (2006): n. pag. Web. 15 Apr. 2017.

Cullen, Julie Berry, and Steven D. Levitt. "Crime, Urban Flight, and the Consequences for Cities." *The Review of Economics and Statistics* 81.2 (1996): n. pag. Web. 17 Apr. 2017.

Gilbert, Nigel. *Agent-based models*. Thousand Oaks, Calif: Sage Publications, 2007. Print.

Glaeser, Edward, and Matthew Kahn. "Sprawl and Urban Growth." *Handbook of Urban and Regional Economics* IV (2003): n. pag. Web. 15 Apr. 2017.

Harvey, Robert O., and W. A. V. Clark. "The Nature and Economics of Urban Sprawl." *Land Economics* 41.1 (1965): 1. Web. 14 Apr. 2017.

Hess, George R., Salinda S. Daley, Becky K. Dennison, Sharon R. Lubkin, Robert P. McGuinn, Venessa Z. Morin, Kevin M. Porter, Rick E. Savage, Wade G. Shelton, Chris M. Snow, and Beth M. Were. "Just What is Sprawl, Anyway?" North Carolina State University, 28 May 2001. Web. 14 Apr. 2017.

Greenwood, Michael. "Modeling Migration." Encyclopedia of Social Measurement 2 (2005): n. pag. Web. 18 Apr. 2017.

Katz, Bruce. "Charting Maine’s Future: Delivering the Next Economy in Maine." Brookings, 20 Aug. 2006. Web. 14 Apr. 2017.

Kahn, Matthew E. "Does sprawl reduce the black/white housing consumption gap?" *Housing Policy Debate* 12.1 (2001): n. pag. Web. 15 Apr. 2017.

Maine Department of Health and Human Services, Maine Center for Disease Control and Prevention, Data, Research, and Vital Statistics. "Total Live Births by Residence County and Year." *Maine Resident Data, 2003-2014*. Maine Vital Statistics Data Files, 2015. Web. 12 Apr. 2017.

Maine Department of Health and Human Services, Maine Center for Disease Control and Prevention, Data, Research, and Vital Statistics. "Deaths by Residence County and Year." *Maine Resident Data, 2003-2014*. Maine Vital Statistics Data Files, 2015. Web. 12 Apr. 2017.

"Maine." *QuickFacts*. US Census Bureau, 2016. Web. 12 Apr. 2017.

Mattingly, Marybeth J., and Andrew P. Schaefer. "Report 1: A Demographic Profile of Maine." *Highlighting the Distribution of Vulnerable Populations* (2012). Carsey Institute at the University of New Hampshire. Web. 12 Apr. 2017.

McCann, B. A., and R. Ewing. "MEASURING THE HEALTH EFFECTS OF SPRAWL: A NATIONAL ANALYSIS OF PHYSICAL ACTIVITY, OBESITY AND CHRONIC DISEASE." *Smart Growth America Surface Transportation Policy Project* (2003): n. pag. Web. 15 Apr. 2017.

Ramos, Raul. "Gravity models: A tool for migration analysis." *IZA World of Labor* (2016): n. pag. Web. 18 Apr. 2017.

Richert, Evan D. "The Cost of Sprawl." Maine State Planning Office, 1997. Web. 15 Apr. 2017.

Theriault, Jackie, Jonathan Wilson, Jessica Cummings, Stephen McCausland, Steven Francoeur, Michael Swanholm, and Robert Williams. "CRIME IN MAINE." State of Maine Department of Public Safety, 2010. Web. 16 Apr. 2017.

"United States." *QuickFacts*. US Census Bureau, 2016. Web. 12 Apr. 2017.

Wu, Junjie. "Environmental amenities, urban sprawl, and community characteristics." *Journal of Environmental Economics and Management* 52.2 (2006): 527-47. Web. 17 Apr. 2017.

Zolnik, Edmund J. "The Effect of Sprawl on Private-Vehicle Commuting Outcomes." *Environment and Planning A* 43.8 (2011): 1875-893. Web. 17 Apr. 2017.

APPENDIX A – SUPPLEMENTAL GRAPHS AND FIGURES

Appendix A includes various figures that were described and cited throughout the paper, as well as supplementary graphs to aid in the understanding of the results. Figures 9 and 10 are the results of computed correlations that were implemented into the NetLogo simulation detailed in Appendix B – Model Coding with Annotations. Figures 11 through 14 are graphs concerning the validation of the model, as referenced in the "Results" section of the paper. Finally, figures 15 and 16 are the results from implementing the policies of raising housing availability and lowering housing prices, which were discussed in "Policy Simulations."

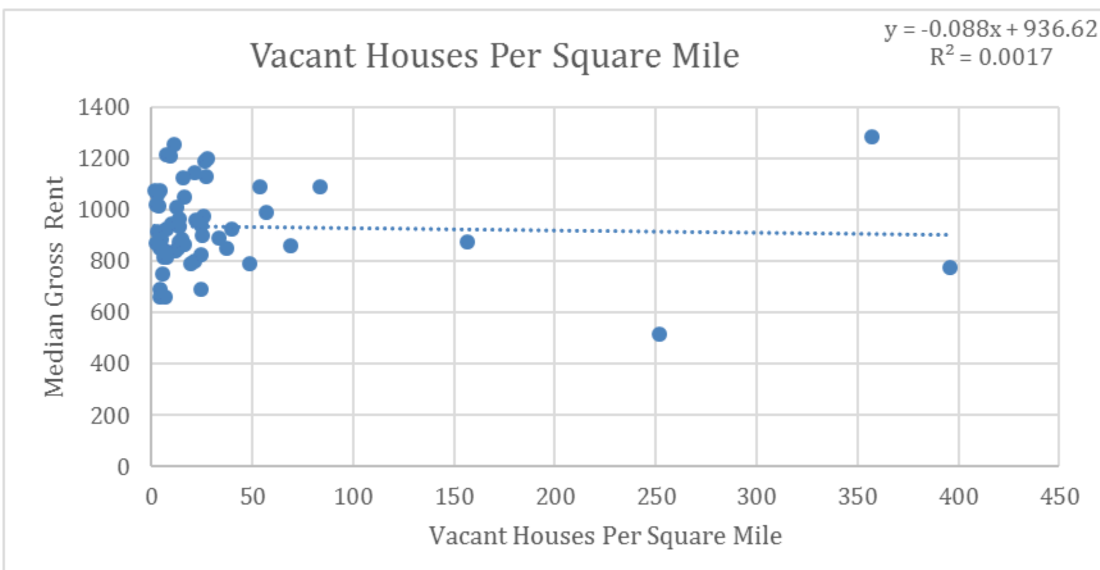

Figure 9: Correlation between Median Gross Rent and Density of Vacant Houses

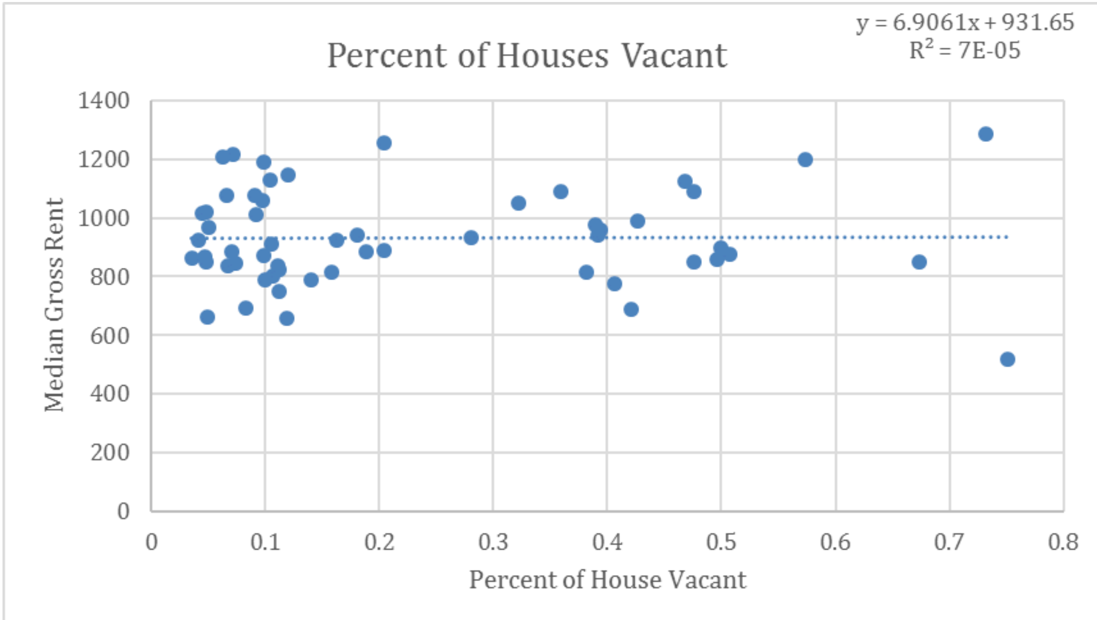

Figure 10: Correlation between Median Gross Rent and Percent of Houses Vacant

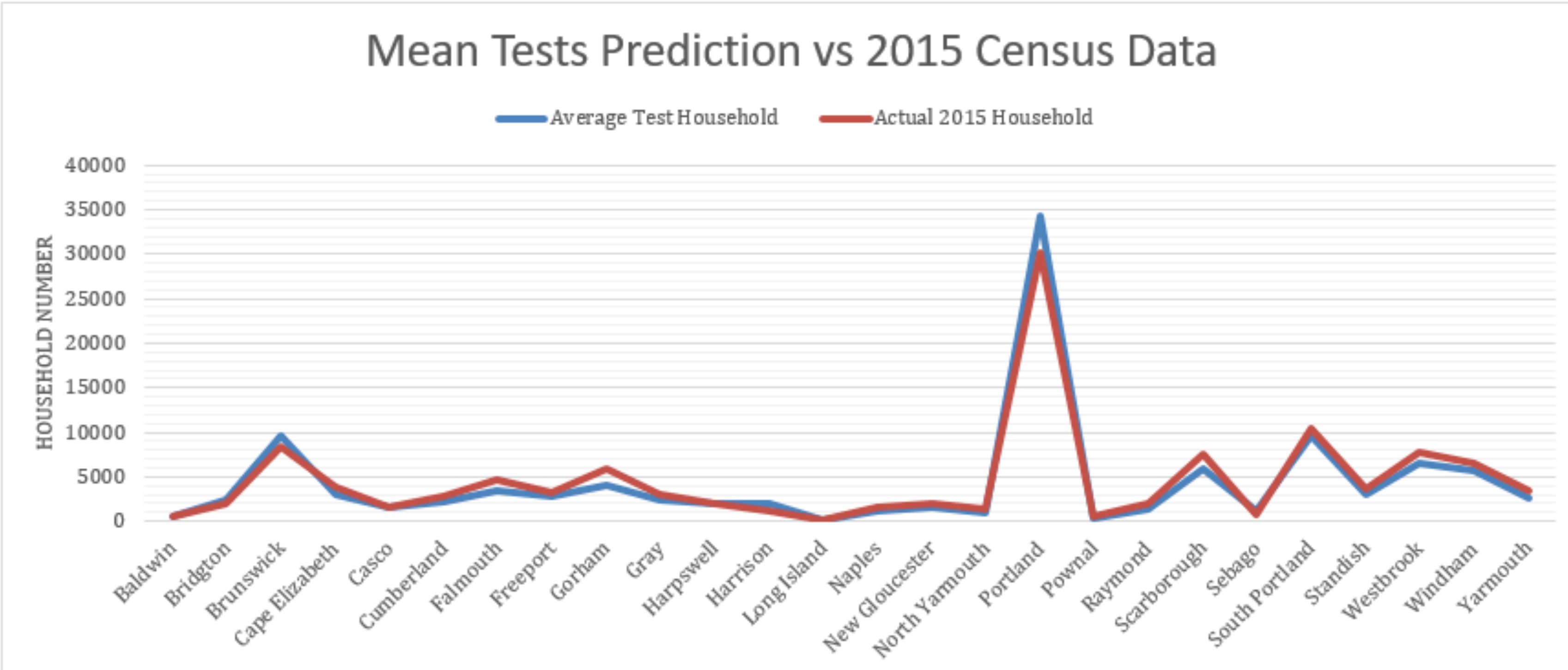

Figure 11: Cumberland Mean Test Prediction vs. 2015 Census Data

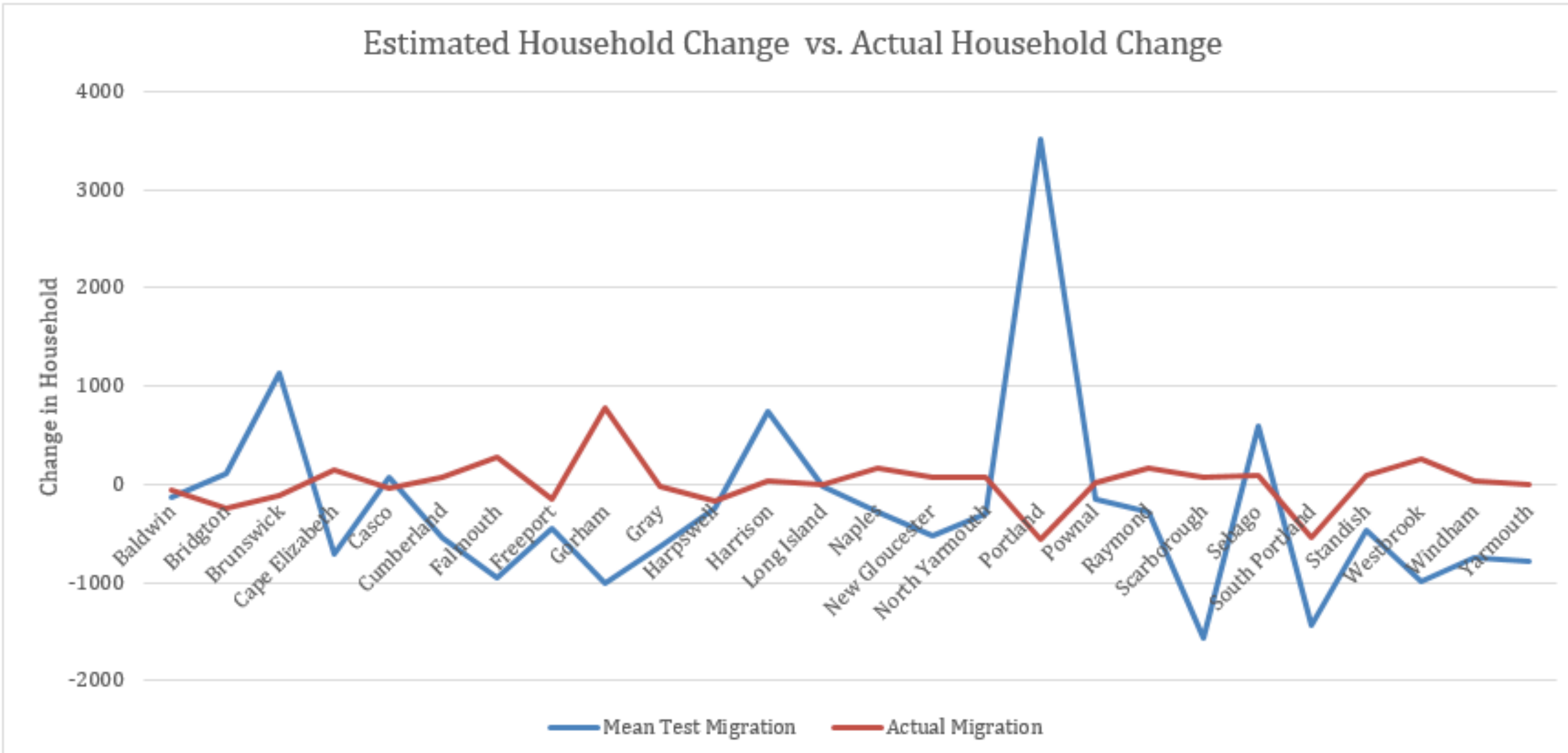

Figure 12: Cumberland Mean Test Migration vs. Actual Migration

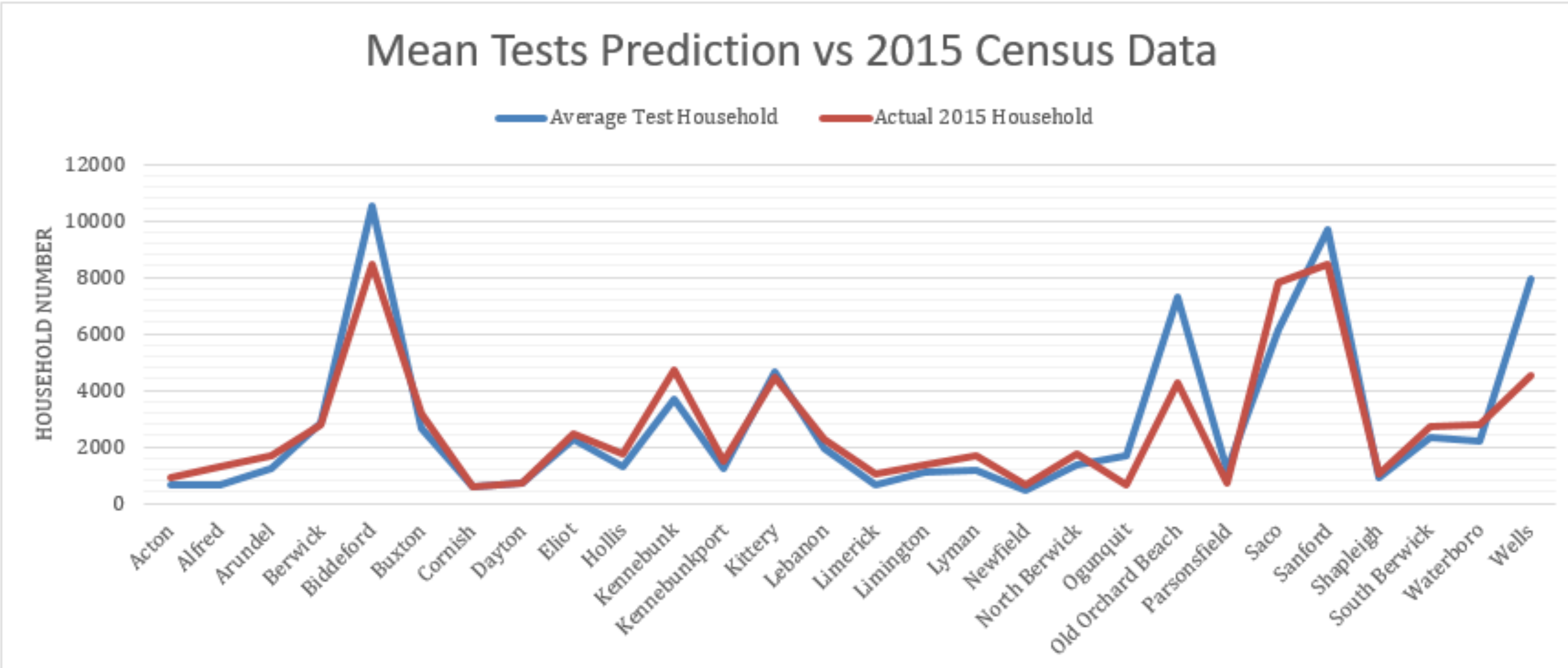

Figure 13: York Mean Test vs. 2015 Census Data

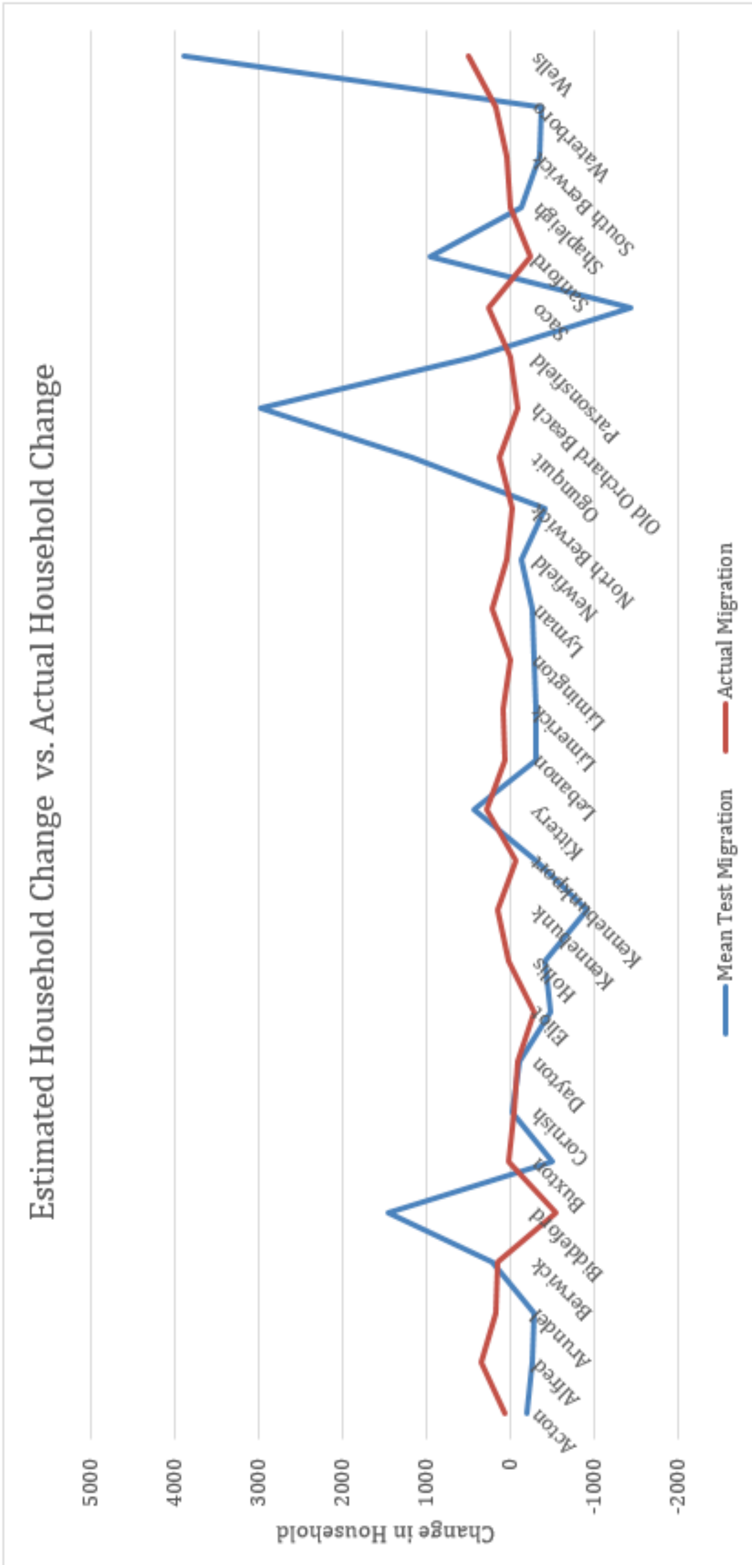

Figure 14: York Mean Test Migration vs. Actual Migration

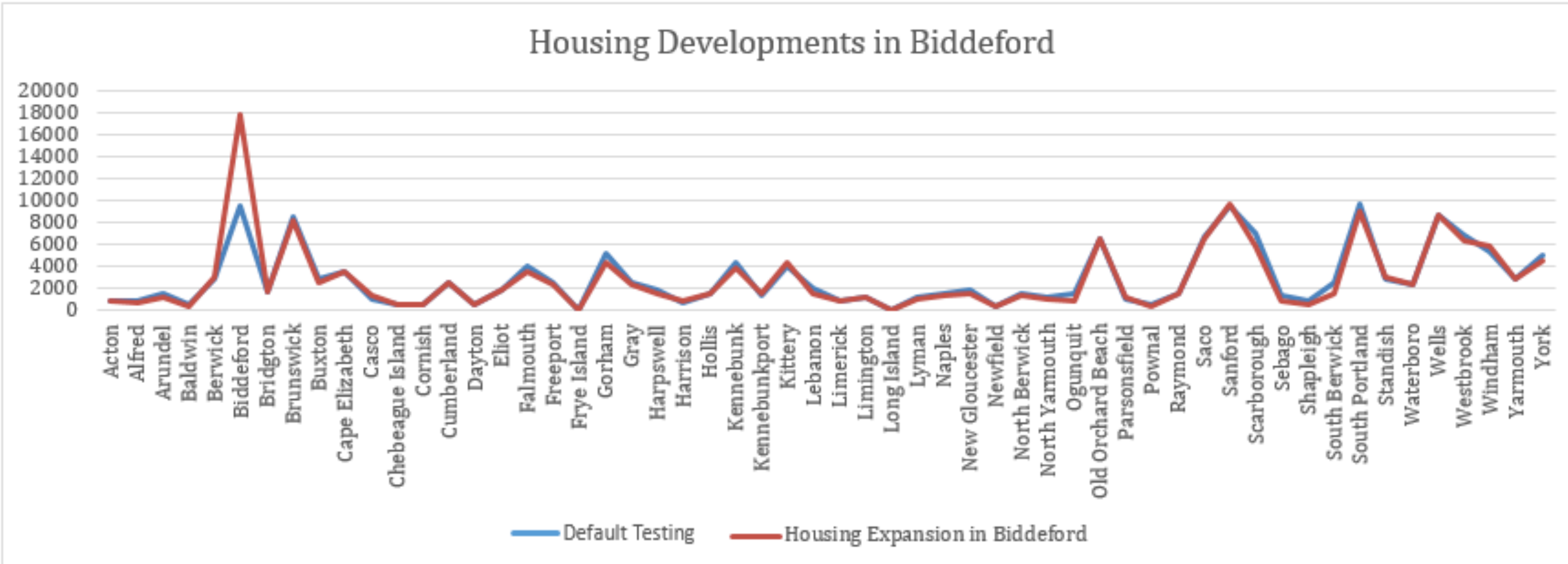

Figure 15: 9000 Housing Developments In Biddeford

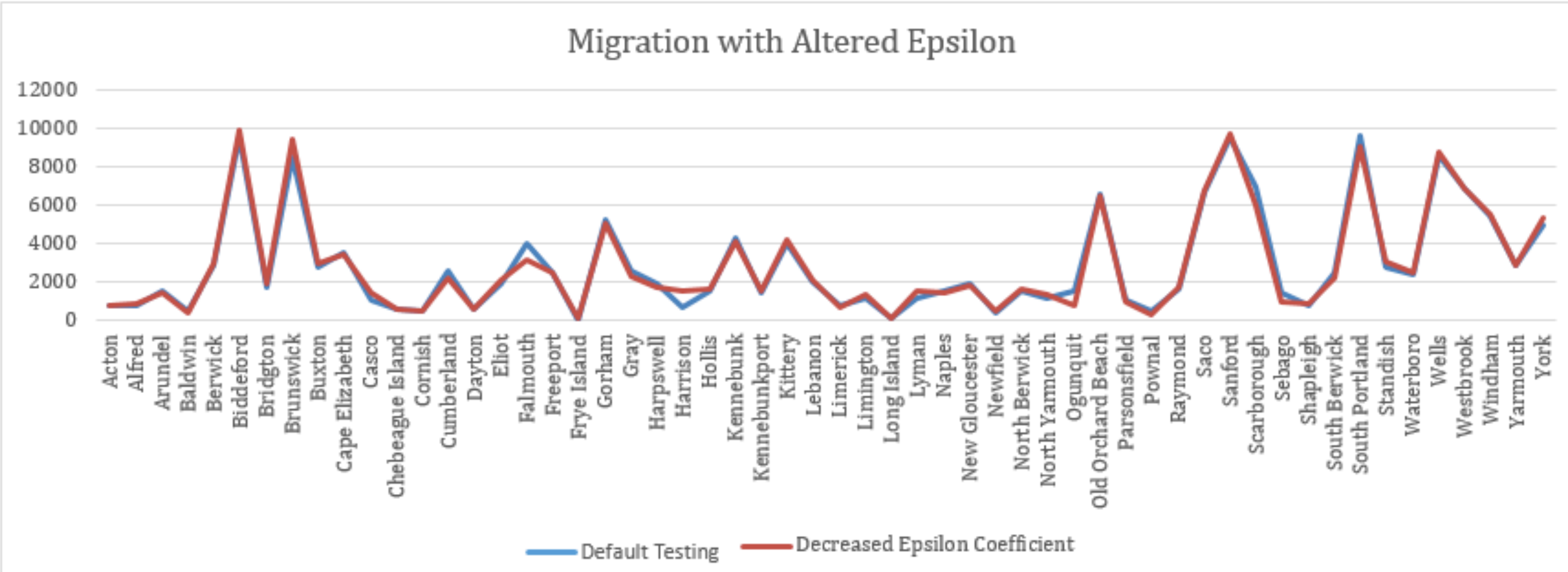

Figure 16: Default Testing vs. Decreased Epsilon Coefficient

APPENDIX B – MODEL CODING WITH ANNOTATIONS

Appendix B summarizes the code implemented in NetLogo for the agent-based model described throughout this paper. The data is parsed from three data files. This includes a shapefile containing the polygons of each town in Cumberland and York County, Maine, as well as their names. The second and third data files are CSV files that can be altered by the user to create simulations of different policies or trends. These files are entitled "Town_Names" and "Income_coefficients."

```
extensions [
  gis
  csv
]
globals [           ;;Creates the Global Variables
  Counties-dataset
  Coeff
  Town_names
  Towns
  Occupied
  Potential_Job
  alpha
  Potential_Living
  beta

]
patches-own [        ;;Creates Patch Variables
  Name
  Total_Household
  Occupied_Household
  Town_size
  Travel_time
  Median_Income
  Median_Rent
  crime_rate
]

turtles-own [ turtle-name ]
breed [ houses house ]
```

```
houses-own [          ;;Creates agent variables
 stay-counter
 Job_Patch
 Job_Location
 Job
 Income
]

to setup             ;;Loads up GIS data, Calls Draw-GIS and Setup-World Functions
 clear-all
 file-close-all
 reset-ticks
 set Counties-dataset gis:load-dataset "Data/SHP_1/Cumberland_and_York.shp"
 Draw-GIS
 setup-world
end

to spawn-agents       ;; Calls spawn-household function
  spawn-household
end

to setup-info
 setup-patches
 reset-ticks
end

to setup-world        ;; Sets up patch size and world size (in patches)
 set-patch-size 2
 resize-world -100 100 -100 100
end

to Draw-GIS          ;; Draws a green line at each town border, and gives patches
town-name
 gis:set-drawing-color green
 gis:draw Counties-dataset 1
 gis:set-world-envelope-ds gis:envelope-of Counties-dataset
 gis:apply-coverage counties-dataset "NAME10" Name
end

to Spawn-Household    ;; Spawn household is called by the user
 ct              ;; clears turtles
 reset-ticks
 file-open "Data/Town_Name.csv"  ;; reads csv file line by line
 while [ not file-at-end?] [
 let data csv:from-row file-read-line
```

```
 ask n-of ((item 1 data) / 100) patches with [name = item 0 data][sprout-houses 1]]
 ;; ask in netlogo fuctions essentially as a for loop through the called variable
 ;; Asks (households in town/100) of patches with name of same town to spawn
 ;; household. The patch are selected at random
 file-close

 ask n-of (1) patches with [name = "Frye Island"][sprout-houses 1]

ask houses [set size 2]       ;; sets house size to two pixels
ask houses [set color cyan]
ask houses [set shape "house"] ;; gives households a house icon
ask turtles[ set turtle-name name ;; Gives each household the name of the town they
reside in

]
 ; ask houses [set stay-counter wait-before-seeking ]
end

to setup-patches ;; function that sets up patch variables

 ask turtles [set Job (random-float 1)] ;; assigns turtles a random number between 0
and 1

 ask houses ;; creates stay counter for each house
 [ set stay-counter (random wait-before-seeking)]

file-open "Data/Town_Name.csv"
 while [ not file-at-end?] [
 let data csv:from-row file-read-line
 ask patches with [name = item 0 data]
 ;; sets each patch with the following data for each town
 [set Median_rent (item 4 data)
  set Town_size (item 3 data)
  set Total_household (item 2 data)
  set Occupied_Household (item 1 data)
  set Crime_rate ( (0.0254 * (Total_Household / Town_size) + 15.175 ))
  set Travel_time (item 5 data)
 ]]
 file-close

 ask patches
 [( if (Median_rent = 0)
   [set Median_rent 99999])]
 ;; Sets patches not in maine to rent of 99999, this prevents
 ;; households traveling outside of the bounds
```

```
file-open "Data/Income_Coeffecients.csv"
  while [ not file-at-end?] [
  let data csv:from-row file-read-line
  ask turtles with [turtle-name = item 0 data]
  [set Income ((item 1 data) * (Job ^ 2) + (item 2 data) * Job + (item 3 data)) ]
  ;; Gives each household an income based on the "job"
  ask patches with [name = item 0 data]
  [set Median_Income (median [ Income ] of turtles with [turtle-name = item 0 data])
  ;; Each town finds the median income of turtles on all patches with their name
matching the town name
   set Crime_rate ( (0.0254 * (Occupied_Household / Town_size) + 15.175 ))]
  ;; sets crime rate based on the number of occupied household density
  ]
  file-close


ask patches
   [ ifelse (Total_Household > 0)
   [set pcolor scale-color green (Occupied_Household ) 20000 1]
   [set pcolor black]]

;; Colors each patch based on the total number of occupied households.
end

to go ;; Is initiated by user

 csv:to-file "Data/Results.csv" [ (list name occupied_household) ] of patches
;; Writes the name and occupied households of each town to a csv file.

(if (ticks = 5)
  [stop])
 ;; Stops simulation after 5 ticks

ask turtles
 [ ifelse (stay-counter <= 0)
   [ set Potential_Job one-of patches with [median_income > ([income] of myself)]
    (if (Potential_Job != nobody)
     [ask Potential_Job
       [set alpha (([crime_rate] of Potential_Job) * (median_income - ([income] of
myself)) / (distance myself + 1)^2)]
     ])
   (if ((random-normal 300 200) <= alpha)
    ;; This implements the gravity migration function as shown in the text.
    [(if (Potential_Job != nobody)
     [ask Potential_Job
       [ set Potential_Living min-one-of patches in-radius Travel_time [median_rent]]
```

```
    ;; Selects a random patch within the town with the lowest rent inside a circle
    ;; with the radius of travel time.
      ]
    )]
   )
   (if (Potential_Living != 0) and (([Total_household] of Potential_Living -
[Occupied_Household] of Potential_Living) > 0)
    and ([name] of Potential_Living != "Frye Island")
   and (income * 0.3 >= (median_rent) and ((([crime_rate] of Potential_Living)/
([crime_rate] of myself)) >= (random-normal 0 2 )))
  ;;This implements the rent requirementas well as the requirement that there is
enough housing to\
  ;;support additional households

    [move-to Potential_Living
     ask patches with [name = [name] of Potential_living]
       [set occupied_Household occupied_household + 100]

     print [name] of Potential_Living
    ]
  ;; Updates the number of households in each town
  file-close

    ])
   ]


  [ set stay-counter stay-counter - (round(random-normal 1 1))]
 ]
  ;; If the stay counter was above 0, the random number from a normal distribution
  ;; with a mean and standard deviation of 1 is subtracted from the stay counter



Print "1"

ask turtles
  [set Job (random-normal 0.5 0.2)  ]
;; this resets the "job" of the turtles that moved

file-open "Data/Income_Coeffecients.csv"
  while [ not file-at-end?] [
  let data csv:from-row file-read-line
  ask turtles with [(name) = item 0 data]
  [set Income ((item 1 data) * (Job ^ 2) + (item 2 data) * Job + (item 3 data)) ]
```

```
;;sets the income of the turtles based on their town and their "job"

Print "2"

ask patches with [name = item 0 data]
  [set Median_Income (median [ Income ] of turtles with [turtle-name = item 0 data])
   set Crime_rate ( (0.0254 * (Occupied_Household / Town_size) + 15.175 ))
   set Total_Household ( round (1.0051 * Occupied_Household + 7.9273 ))
  ]]
  file-close
;; Updates patches with new median income and crime rate. The number of total
households
;; is increased based on the number of occupied households.


Print "3"

ask patches
   [ ifelse (Total_Household > 0)
   [set pcolor scale-color green (Occupied_Household ) 20000 1]
   [set pcolor black]]

;; updates each town with colors based on the number of occupied households in
each town

tick
end
```

AUTHOR'S BIOGRAPHY

Kaitlyn G. Lavallee was born in Brunswick, Maine on September 2, 1995. She grew up in Sabattus, Maine and graduated from Oak Hill High School in 2013. Kaitlyn has a double major in mathematics (BA) and economics (BS). She is also a member of the National Society of Collegiate Scholars, Omicron Delta Epsilon and Phi Beta Kappa.

Upon graduation, Kaitlyn has accepted a full-time position as an Energy Analyst with Competitive Energy Services LLC, located in Portland, Maine.