MODELING RESIDENTIAL MOBILITY USING AMERICAN HOUSING SURVEY

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Hao Wang

August 2015

ABSTRACT

The relocation decisions by individual households collectively shape urban areas; hence we use the American Household Survey to empirically study these mobility decisions. We focus on both the decision to move as well as home location selection. We reconfirm the importance of several socio-demographic and macro-economic variables in these decisions as well as identify several other important variables that are relevant to these decision processes.

BIOGRAPHICAL SKETCH

I was born in a lovely city, Chang Chun, in the northeastern part of China. Luckily, I enrolled in the best middle school and high school in Chang Chun to acquire basic science knowledge. More luckily, I was taught to chase my dream freely while building good learning habits.

Soon after my admittance into Tongji University in Shanghai, I found science is a complicated but interesting thing, and the feeling of acquiring new knowledge is really exciting to me. Knowing that an academic career is for scholars with a strong will, I believe that this is the life that I wish to devote my time and talents to. I applied for Cornell University's Civil and Environment Engineering, in order to be better educated and work with the most talented people in the world. Luckily enough, I was admitted! More importantly, I am able to work with the most caring and erudite professor I have ever met, Linda K. Nozick, during my MS and Phd period! I have to say this is the best chance ever in my life throughout this entire 24 years. I can feel I am much closer to my dream.

ACKNOWLEDGMENTS

This is an incredible journey in Cornell in which I am endowed with enormous amount of support, love, encouragement and freedom throughout these two years study. I have been touched by so many ways, so many earnest professors and so many excellent friends.

First and foremost, I would like to offer my sincere gratitude to Professor Linda K. Nozick. She is an extraordinary supervisor both in academic and in life. Her happiness and endless energy touched me deeply in heart. She is so busy but can also be so happy. Whenever I met difficulty during the research, she was always there, and could always help me solve it, so without her help, I can’t even think about this research can be done. Her endless encouragement and constant faith are rocks when I encounter difficulty. I have a deep sense of pride to be instructed by her so closely.

I am also grateful to my Committee member Professor Ricardo Alvarez Daziano. He is the key person helping me solve technical problems when building models and the person who taught me lots of knowledge on discrete choice model and on why to explain people’s behavior in economics.

I am also thankful for the entire fellow graduate students I had worked with at Cornell, especially Yaze Chen, Cong Xu, Kecheng Xu, Yilin Wang, Mingze Niu, Peng Wang and all my friends in 470. It is happy to be with you guys! You are all excellent and brilliant!

Last but no least I would like to thank my parents for their tremendous love and support throughout my life.

TABLE OF CONTENTS

1. INTRODUCTION

People move for a variety of personal reasons: the birth of a child, marriage, a youngest child entering college, a new job, retirement, just to name a few. Hence, many people will experience the relocation processes at least a couple of times during their lives. These decisions, in the aggregate, have a substantial impact on the urban landscape; hence there has been substantial research across several disciplines to understand this behavior. It is not surprising, that comprehensive modeling environments of urban systems such as UrbanSim and ILUTE include a specific module to deal with this relocation process.

The relocation decisions can be conceptually separated into two parts: (1) the decision to move, which is defined as deciding whether or not to be active in the housing market, and (2) the location selection decision, which involves an evaluation of different alternative homes. Despite substantial research into aggregate methods to forecast land use, population and employment change, these models are not sufficiently effective so as to be able to generate detailed forecasts of urban development at different points in time (Habib, 2009). Thus, the research emphasis has shifted towards disaggregate and agent-based modeling to understand the evolution of the urban landscape. The core idea is to understand individual household's behaviors. Then, based on that understanding, to construct a simulation of the dynamic decisions of agents that represent individuals over time and geographic space so as to estimate changes in the urban form. In this vein, we first focus on the household residential mobility, including whether a household will relocate based on household characters and other factors, then analyses household preferences for home selection.

The key data set is the American Housing Survey (AHS), which is an important source of information on the housing stock and their occupants. It is the largest regular housing sample survey in the United States containing data covering almost 30 years across 60 metropolitan areas. The AHS has several advantages in the development of a residential mobility model: it follows the same houses over time, has information about the occupants, including occupants' subjective judgment about their neighborhood quality, can be linked to Census tract data and can be easily obtained by researchers.

The residential mobility model uses the 1996 and 2004 American Housing Survey (AHS) Metro dataset, which contains 9 different cities with over 10,000 households. The data of macro-economic factors comes from the U.S. Bureau of Labor Statistics. The housing choice model uses 1996 and 2004 AHS dataset of metropolitan statistical areas focused on Seattle and the 1990 United States Census as data sources.

This paper extends the literature on residential mobility in the US housing market by providing empirical evidence of the importance of several non-linear and an interaction term in the mobility decision. It also reconfirms the importance of several socio-demographic and macro-economic variables. In the home selection model we add an independent variable to understand the importance of racial homogeneity in the decision as well as modify the room stress independent variable suggested by Clark and Huang (2003).

The next section provides a review of the relevant literature. The third section presents the data and how that data is used to infer the statistical importance of variables like tenure, age and unemployment for homes that were not transacted during the study period. The fourth section gives the model formulation. The fifth section gives conclusions and next steps.

2. REVIEW OF LITERATURE

The literature on residential mobility and location choice are each extensive and interrelated. Although these two decisions are related, the exact correlation between them remains unclear. Also the factors that are believed to contribute to each decision are mostly different. Due to the complexity of these decisions, most of the research treats them separately. We first focus on residential mobility and then home selection.

There is a large body of literature on residential mobility: sociologists, economists, and geographers have all provided insights into the residential mobility process and its relationship to urban land use. Two particularly important contributions are Rossi (1955) and Brown and Moore (1970). Rossi (1955) shifted the research focus to the individual household and its motivation to change dwellings from a focus on aggregate patterns of residential mobility. Brown and Moore (1970) used “stress” to describe why households move. Stress arises in a household, as the needs of the individuals within the house hold changes. When reduce the stress, either they will search for new house or decide to stay.

Nearly all the research on residential mobility emphasizes household level characteristics. Household life cycle, age of householder, the educational level and income have shown high correlation with the mobility decision and also choice of a dwelling (Clark and Dieleman, 1996). The size, type, price and tenure of dwelling and its location with respect to workplaces are also found to be critical to decision process (Dieleman, 2001).

Dieleman (2001) identifies three attributes that are often motivators for mobility. First, there is a strong correlation between life cycle and rate of mobility. By far the most mobile segments of the population are young adults between 20 and 35. Second, the relationship between size of the present dwelling and the needs of the household drives mobility. Households in relatively large units are less like to move. Finally, life events such as family formation and dissolution are clearly correlated with the probability of relocation.

Most studies can be classified by different research areas, by cross-sectional or longitudinal data, and by different modelling methods. Different areas have different housing policy, which may result in a different housing market. Thus, many models are not generalization to all areas. For example British housing markets have a higher proportion of public housing and are more similar to the Dutch context than that in the United States (Clark and Huang, 2003). European and North American researches take different approaches in analyzing residential mobility. European researchers often treat the supply of housing as an exogenous factor, while North American researchers give primacy to market forces; supply-side factors are often endogenous to the models (Strassman, 2001). There is also a debate on the relative value of cross-sectional and longitudinal data to estimate the coefficient for the effect of room stress on mobility (Clark, 1992). Finally, there are many different ways to build a residential mobility model, just to name a few: Hazard-based duration models (Tatsiramos, 2006; Habib, 2009, etc.), discrete choice models (Habib and Miller, 2010; Clark and Huang, 2003; Kan, 1998; etc.), nested logit model that treats residential mobility and location choice together (Lee and Waddell, 2010).

There is also a substantial literature on home selection. The selection of a home, which includes its location, has a substantial influence on people's daily activities and travel behaviors. Most of the early literature in this area focuses on models for which the elemental decision is a zone choice and therefore assume that all dwellings within a zone are indistinguishable (Habib and Kockelman, 2008; Weisbrod et al., 1980). For our purposes, this is an important shortcoming.

More recently, the focus has shifted to models at the disaggregate level of individual dwelling units with unique aggregate characteristics (e.g. square feet, number of rooms, etc.). The analysis of location choice was largely enabled by the development of discrete choice models. The multinomial logit model (MNL) is the most widely used model to analyze people's home selection behaviors based on a random utility maximization (RUM) hypothesis. Gabriel and Rosenthal (1989) concluded white and black households' location choices are importantly and differently affected by their socioeconomic characteristics using this model. Waddell (1992) studied the influence of residential and workplace location choice on urban spatial structure by race. The advantages of MNL are properties that allow for closed form expression and consistent estimations within a subset of alternatives. However, MNL model impose two important constrains: the independent and identically distributed (IID) error structure and unobserved response homogeneity (Bhat and Guo, 2004). In the nearest decade, many researchers are trying to relax these two constrains in different ways. Two types of models are used most frequently, the nested logit model and mixed multinomial logit model. Lee and Waddell (2010) combined residential mobility and home selection behaviors together and built a nested logit model. Habib (2009) applied a mixed logit formulation and included current dwelling as reference point in evaluation of alternative dwellings.

The role of transportation on people's location choice has received substantial attention. Some studies emphasize the link between location choice and traffic modes (Eliasson and Mattsson, 2000). Others have studied the relations between workplace and location (Pinjari et al., 2008). Since the decision making frequently involves more than one individual, researchers found different patterns of selection in multi-person households (Plaut 2006).

Because the evolution of the urban environment is so critical and complex of a domain for decision-making two integrated modeling environments have been created. Both environments cover a broad range of issues in the evolution that impact the urban landscape. These environments are UrbanSim and the integrated land use, transportation and environment (ILUTE) model environments. These models are substantially more dynamic and microscopic than the approaches in earlier aggregate models (Wang, Waddell, and Outwater, 2011). Both are developed to deal with the relationship between transportation and urban land use. Both of them have modules for residential mobility, home selection and land price models. UrbanSim assumes that land prices are exogenous to home selection; hence price is determined through a hedonic price model external to location models (Habib, 2009), while ILUTE determines land price in conjunction with home selection through a bidding mechanism (Rosenfield, Chingcuanco and Miller, 2013). Both microsimulation models deal with residential mobility and home selection behaviors separately, although recently Lee and Waddell (2010) develop a nested logit model which combines residential mobility process and location choice together.

3. RESIDENTIAL MOBILITY MODEL

Data and Variables Used in Analysis

AHS is sponsored by the Department of Housing and Urban Development (HUD) and is conducted by the U.S. Census Bureau. The 1996 AHS metropolitan sample covers about 18,300 occupied housing units, in nine different cities: Atlanta, Cleveland, St. Louis, Seattle, Sacramento, Oklahoma, Memphis, Indianapolis and Hartford. The 2004 metro sample covers about 28,900 housing units in thirteen cities including those nine cities mentioned above. We focused on owner occupied housing units that are present in both 1996 and 2004 surveys and for which the data fields needed in the analysis were available. This yielded about 11,331 owned housing units. Among them, the total number of households for which the homes were transacted between 1996 and 2004 is 3,891 or about 34%.

The factors that affect people's mobility decisions can be categorized as follows: household characteristics, neighborhood characteristics and macro-economic factors. We describe each category below including the factors that are included. For household and housing unit characteristics, we consider age, income, education level, marital status, number of adults, is there at least one person over 60 years of age, unit square feet, tenure, bathrooms, and the number of cars.

Neighborhood characteristics analyzed are: noise, the presence of open spaces, neighborhood crime, shopping satisfaction, public transportation availability, community services, opinion of their neighborhood, litter.

The key macroeconomic factors considered are the unemployment rate in different years and the trend of unemployment rate change. These factors are summarized in Table 1.

TABLE 1: Summary of Variables

	Moved	Stayed

N	3891 (34%)	7440 (66%)
A. Household and housing units characteristics		
Age of householder		
20-30	106 (63%)	62 (37%)
30-40	866 (49%)	897 (51%)
40-50	1082(36%)	1965(64%)
50-60	727 (29%)	1816(71%)
60-70	434 (25%)	1269(75%)
70-80	320 (23%)	1064(77%)
80+	356 (49%)	367 (51%)
Household income in 1996		
<=30k	1281(33%)	2537(67%)
30-40k	430(33%)	867(67%)
40-50k	448(34%)	885(66%)
50-60k	394(35%)	736(65%)
60-70k	331(37%)	574(63%)
70-80k	260(35%)	473(65%)
80-90k	183(37%)	315(63%)
90-100k	198(35%)	372(65%)
>=100k	194(36%)	338(64%)
Educational level		
Without a bachelor degree	2378(32%)	5036(68%)
Bachelor degree and up	1513(38%)	2404(62%)
Marital status		
Married	3057(33%)	6099(67%)

Others	834 (38%)	1341(62%)
Adults		
1	879 (39%)	1352(61%)
2	2562(35%)	4712(65%)
>=3	450 (25%)	1376(75%)
Race		
White	3589(35%)	6710(65%)
Others	302(29%)	730(71%)
Senior over 60		
With	920 (29%)	2304(71%)
Without	2971(37%)	5136(63%)
Unit square feet		
<=1500	1455(35%)	2689(65%)
1500-2500	1792(35%)	3340(65%)
>=2500	644(31%)	1411(69%)
Tenure		
<=10	1862(43%)	2409(57%)
10-20	1194(34%)	2348(66%)
20-30	421 (24%)	1331(76%)
30-40	222 (21%)	834 (79%)
40-50	137 (25%)	397 (75%)
>=50	39 (30%)	90 (70%)
Number of bathrooms		
1	1287(31%)	2822(69%)
2	2253(37%)	3860(63%)

3 and up	350 (32%)	759(68%)
B. Neighborhood characteristics		
Noise that is bothersome		
Yes	198(42%)	268(58%)
No	3693(34%)	7172(66%)
Open space		
Yes	869(31%)	1955(69%)
No	3022(36%)	5485(64%)
C. Macro-economic factors		
Unemployment rate 1 year prior movement	NA	NA
(change by year and area)		
Unemployment rate change in last 2 years	NA	NA
(change by year and area)		

Modeling Issues

Weight in the Model

The AHS is a stratified survey and so requires the use of "weights" to estimate the total numbers of housing unit with a particular characteristic. Three kinds of weight are stored in AHS: The pure weight (PWT) which is the inverse of the probability of selection of the housing unit. The adjusted weight (WEIGHT) is the estimated number of households that a particular observation is believed to represent. The WGT90GEO variable gives users access to the weights used in the report.

Weights are assigned only at the housing level. For nearly all applications, the adjusted weight is the most appropriate weight to use. In this research, we use WEIGHT. Since we eliminated the observations with uncompleted data, we adjusted this weight as follows.

$$\sum_{j=1}^{9} \frac{\frac{11331}{9}}{\sum_{i} w_{ij}} \times w_{ij} = 11331$$

Randomly Generate Values for the Independent Variables for Those that did not Move

A binary logistic model is employed to model whether a household will move between 1996 and 2004. The dependent variable is coded as 1 if the household actually moved and 0 if the household did not move._It's easy to calculate the macro-economic variables, tenure and age for the households who actually moved, however, for those that did not move the process is more complicated. We randomly generate a year between 1996 and 2004 as the reference year for each housing unit that was not transacted. With that reference year we then compute the age and tenure and associated unemployment statistics for that particular housing unit. Notice that education level (Bachlorandup), race (white), the flag for household member(s) with an age above 60 (Over60years), marital status (Married), the presence of noise (Noise) and open spaces (Openspace) and the remaining city flag (Hartford) are categorical variables. The remainders of the independent variables are treated as continuous.

Model Results and Discussion

TABLE 2: Analysis of Maximum Likelihood Estimates

Variables	Estimate	Standard Error	Odds Ratio Estimates	Wald Chi-Square	Pr > ChiSq
Intercept	4.222	0.325	-	169.2	<.0001
Age	-0.181	0.011	-	292.2	<.0001
Income	0.007	0.001	-	30.7	<.0001

Bachelorandup	0.174	0.046	1.190	14.2	.0002
Fullbathrooms	0.678	0.174	-	15.1	<.0001
Adult	-0.211	0.047	0.810	20.4	<.0001
White	0.342	0.075	1.407	20.8	<.0001
Over60years	-0.378	0.086	0.685	19.5	<.0001
Married	-0.140	0.059	0.869	5.7	0.0167
Tenure	-0.039	0.006	-	38.9	<.0001
Tenure^2	0.0003	0.0001	-	6.7	0.0098
Age^2	0.002	0.000	-	280.2	<.0001
Fullbathroom^2	-0.153	0.047	-	10.6	0.0011
Income × Unitsf	-0.002	0.0004	-	23.5	<.0001
Noise	0.401	0.101	1.493	15.8	<.0001
Openspace	-0.208	0.049	0.812	18.1	<.0001
Unemploy_1yrprior	-0.052	0.020	0.949	7.0	0.0082
Unemploy_increase_2yr	0.166	0.019	1.181	75.9	<.0001
Hartford	-0.139	0.070	0.870	4.0	0.0461
A. Model Fit Statistics					
-2Log Likelihood (intercept only)				13657.01	
-2Log Likelihood (with covariates)				14601.38	
B. Testing Global Null Hypothesis				944.37	<.0001
Likelihood Ratio				917.33	<.0001
				842.39	<.0001

Score		
Wald		
	6.1256	0.6332
C. Goodness of fit test	0.1104	
Hosmer and Lemeshow		
Adjusted R-Square		

Table 2 shows the results of the model. According to the Testing Global Null Hypothesis statistic, the whole model has significantly statistical meaning. The adjusted R-Square is 0.1104. Hosmer and Lemeshow Goodness of fit test statistic equals 0.6332(significant if >0.05).

For all the variables in the table, according to their p value, they are significant at least at 0.05 level. Among them, Intercept, Age, Income, Adults, Race, over 60 years of age, Tenure, bathrooms, age square, income × unit square feet, noise, open space, unemployment changing trend are significant at 0.0001 level. Education level, Marital status, Tenure square, bathrooms square, unemployment rate at moving year, city indicator Hartford, are significant at 0.05 level.

The impact of age on mobility is not always one direction. Actually, for people over 80 years, we have a relative high mobility rate. Naturally, we could guess the death of senior people may be the main factor of the unusual increase on mobility. Thus, the coefficient of age is negative, and the coefficient of age square is positive. Age and age square, their combination can better describe the impact of age on mobility than only use one variable age to describe mobility.

Education level is also related to mobility probability. For those householders who received bachelor degree and up, the mobility is higher. As the number of adults in the household increase, the probability for that household to move is relative less. White people turn out to be more “active” than other races of people. The presence of senior over 60 years old will decrease the probability of move. Married householder with spouse present is less likely to move compared to others, like single, separated, divorced, etc. Tenure, the length of time in the house, also has a non-linear effect on mobility, like the impact of age. Number of full bathrooms turns out also has non-linear effect on mobility. From the data, the fraction of houses with more than three full bathrooms is very small, so the number of full bathrooms grouped to take value one, two, and three. The sign of coefficient of full bathroom is positive, and the sign of coefficient of full bathroom square is negative. This pattern suggests the turnover ratio for two full bathrooms is the highest. We can also interpret it as this: houses with two full bathrooms are relative more popular than those with only one or with three or more bathrooms.

Total income has a positive impact on mobility. However, the interaction between income and the unit square feet of the housing unit is also significant, suggesting that the effect of income on the odds of mobility depends on unit square feet and vice versa. The interaction suggests that the households that live in a relatively smaller house have a higher probability of moving than the households that live in relatively bigger houses, all other things equal.

The neighborhood characteristics considered are whether the neighborhood noise is bothersome and whether there are open spaces within ½ block of the house. The sign of these two variables is as expected; bothersome noise increases the probability the household will move whereas open spaces cause that probability to decline.

The key macroeconomic factors considered are the unemployment rate in the previous year to the relocation decision and the unemployment rate increase over the previous two years. Higher unemployment rates in the previous year tend to dampen relocation however, increases in unemployment over a couple of years tends to accelerate mobility. The empirical observation that that high unemployment dampens mobility has been observed in other studies including Liu, Miao and Zha(2013) and Gan and Zhang (2013).

We should also note that, different cities can have different housing markets and therefore the mobility of households can be different holding the other factors constant; hence a city indicator many be needed. In this study, we include a city indicator for Hartford; since the mobility in Hartford is the lowest of the nine cities considered, and that relative low mobility cannot be explain by the other variables included in this model.

Finally, bootstrapping (1000 replicates of samples of 11,331 households) was used to estimate the 95% confidence interval for the coefficients and those confidence intervals are given in Table 3. It is useful to notice that none of the confidence intervals includes zero; hence the estimate of the sign for each variable is quite stable.

TABLE 3: 95% Confidence Intervals for Coefficients.

Variables	95% confident interval
intercept	(3.570, 4.877)
age	(-0.2023, -0.1593)
income	(0.0046, 0.0095)
bachelorandup	(0.0794, 0.2685)
fullbathrooms	(0.3360, 1.0251)
adult	(-0.3043, -0.1185)
white	(0.1920, 0.4916)
over60years	(-0.5654, -0.2010)
married	(-0.2540, -0.0215)
tenure	(-0.0511, -0.0268)
tenuresquare	(0.0001, 0.0005)
agesquare	(0.0014, 0.0018)
fullbathsquare	(-0.2469, -0.0617)
incomeunitsf	(-0.0030, -0.0012)
noise	(0.1950, 0.6096)
openspace	(-0.3050, -0.1124)
unemploy_1yrprior	(-0.0925, -0.0125)
unemploy_increase_2yrs	(0.1273, 0.2042)
Hartford	(-0.2707, -0.0123)

4. HOUSING CHOICE MODEL

Data and Variables used in Analysis

To understand the second component of the relocation decision, we again use AHS in conjunction with the 1990 Census data. Via the unique "CONTRAL" number, the housing information is linked with the characteristics of the home buyer. Label "ZONE" is used to link to the 1990 Census data. We use social explorer to collect data from Census tract as of 1990. We use the 1990 Census because the AHS geography uses the tract boundaries from 1990, and some tracts have been changed in 2000 Census.

Three categories of variables are considered and summarized in Table 4. In the analysis that follows, we deflate the income (from 2004 dollars) to the year at which the household actually moved.

TABLE 4: Descriptive Analysis of Explanatory Variables

Variable Name	Definition	Mean	SD
Household Level			
ZINC22	Household income(2004 dollars)	9.857	9.098
HHAGE	Age of Householder	43.621	13.377
BACHELOR	Whether householder has a bachelor degree	0.499	0.501
MASTER	Whether householder has a master degree	0.163	0.369
SONGLE	Whether householder is single	0.140	0.348
MARRIED	Whether householder is married	0.657	0.475
HHSEX	Sex of householder(male=1, female=0)	0.539	0.499
CARS	Number of cars	1.468	0.904
PER	Number of persons	2.755	1.359
ZADULT	Number of adults	1.942	0.708
children	Number of children	0.813	1.074
expected rooms	Expected number of rooms	3.461	1.037
hhCaucasian	Whether householder's race is Caucasian	0.878	0.328
hhAA	Whether householder's race is African American	0.013	0.115
hhAsian	Whether householder's race is Asian	0.065	0.246
Housing characteristics			
LPRICE2	List price of the house($$$$$)	24.411	15.515

UNITSF2	Unit square feet of the house(in hundred)	17.971	6.418
unitprice	Price per hundred square feet	143.888	108.508
BATHS	Number of bathrooms	1.786	0.622
BEDRMS	Number of bedrooms	3.038	0.846
ROOMS	Number of rooms	6.606	1.597
BUILT2	Year the house was built	1967.145	0.220
METRO	Whether in central of the city	0.214	0.410
SCHOOLYES	Elementary school within 1 mile	0.265	0.442
NEWTRN	Public transportation available	0.800	0.401
Zonal level			
population density	Population density	702.685	862.182
HCaucasian	Caucasian households percentage	0.913	0.062
HAA	African American households percentage	0.027	0.036
HAsian	Asian household percentage	0.043	0.026
dropoutrate	Dropout rate	0.090	0.027
Unemployed	Unemployment rate	0.039	0.009
renter	Renter percentage	0.358	0.098
poverty	Poverty percentage	0.060	0.032
poorly	Percentage of living poorly	0.070	0.031
struggling	Percentage of living struggling	0.114	0.030
doingok	Percentage of doing ok	0.815	0.056

Discrete Choice Model

Choice Set Formation

Choice set formation for those households that moved from 1996 to 2004 is a critical element of modeling home selection. For each homeowner that moved in during this period, effectively all homes that were transacted were plausible options. McFadden (1978) demonstrated that consistent estimation of multinomial logistic models can be achieved through a random sampling of the alternatives. We use this conclusion to build the choice set based on random sampling from the AHS survey. Aside from the actual dwelling that was purchased, 11 randomly selected dwellings are added to form the choice set. Thus, each household faces a choice set composed of 12 dwellings including the one that was purchased. This implies that the choice set for each household is unique.

Mixed logit model: Parameter Heterogeneity

The household must choose one alternative among 12 different dwellings. A level of utility is defined for each alternative. The individual is supposed to choose the alternative with the highest level of utility. The utility can be expressed as:

$$U_{ij} = x_{ij}\beta + \varepsilon_{ij}$$
$$\varepsilon_{ij} \sim EV1(0,\lambda)$$

Where β is allowed to vary randomly among households, $\beta = \beta^m + \beta^s\eta$. β^m and β^s are fixed mean and scale parameters. The stochastic component η can be assumed to be standard normal. Alternatively, we can also assume a log-normal distribution or other distributions. In this research, we assume η follows a standard normal distribution. Thus, we can specify the distribution of the parameter, β, as follows:

$$\beta \sim N(\beta^*, \ \sigma_\beta^2)$$

The individual *i* will choose alternative *j* in his own choice set if and only if for $\forall k \neq l, U_j \geq U_k$. This leads to the log-likelihood function:

$$\ln L = \sum_i \sum_j y_{ij} \ln P_i(j)$$

$$P_i(j) = \frac{\exp(x_{ij}^{'}\beta_i \mid \beta_i)}{\sum_{k=1}^{J} \exp(x_{ij}^{'}\beta_i \mid \beta_i)}$$

Since β_i is assumed to be normally distributed, the choice probabilities do not have a closed form expression. The estimation was conducted using the SAS procedure MDC.

Table 5 gives the attributes that are statistically significant in the resultant Mixed Multinomial Logit model. That model is presented in Table 6.

TABLE 5: Variable Definition and Explanation

Variable	Definition
School	Present of school within 1 mile of the house
Children	Household with children under 16 years old.
Pricehigh	List price of the house/ annual household income>4.0
Pricetlow	List price of the house/ annual household income<1.6
RoomStresshigh	Room stress for the household >-0.1
RoomStresslow	Room stress for the household <-0.6
Roomstress	expect rooms/actual rooms - 1
Incomeaboveage	Income above average income
hhasian	Race of householder is Asian
Hasian	Percentage of Asian population in the zone

TABLE 6: Mixed Multinomial Logit Estimates

Variables	Estimate	Standard Error	t Value	Pr > |t|
Pricehigh	-1.117	0.174	-6.40	<.0001
Pricelow	-1.279	0.212	-6.01	<.0001
RoomStresshigh	-0.881	0.413	-2.13	<.0001
RoomStresslow	-0.761	0.175	-4.34	0.0042
Hhasian0×Hasian	-0.899	2.137	-0.42	0.6742
Hhasian1×Hasian	21.568	7.279	2.96	0.0029
Children1×School0	-0.712	0.1613	-4.42	<.0001
Children1×School1	0.712	-	-	-
Children0×School0	8.713	-	-	-
Children0×School1	-8.713	-	-	-
Incomeaboveave0×Roomstress	0.283	0.615	0.46	0.646
Incomeaboveave1×Roomstress	-4.089	0.870	-4.66	<.0001

A. Model Fit Statistics	
Log Likelihood Null	-1116
Log Likelihood	-908.9
B. Goodness of fit test	
Estrella	0.6389
Adjusted Estrella	0.6146
McFadden's LRI	0.1853

The heterogeneity of all impendent variables was found to be not statistically significant. Hence it model is equivalent to the multinomial logit.

Figures 1 and 2 gives the LRI measure as a function of the cut off values adopted for room-stress high and low as well as the ratio of price of the house to income, respectively. LRI appears to be a convex function of the break points for room stress, with the values selected yielding the largest value of LRI. In contrast, the LRI as a function of the break points for the ratio of the home price to income appears to be quite flat but with reasonable values, based on this sensitivity analysis, at 1.6 and 4.0, for the low and high break points, respectively.

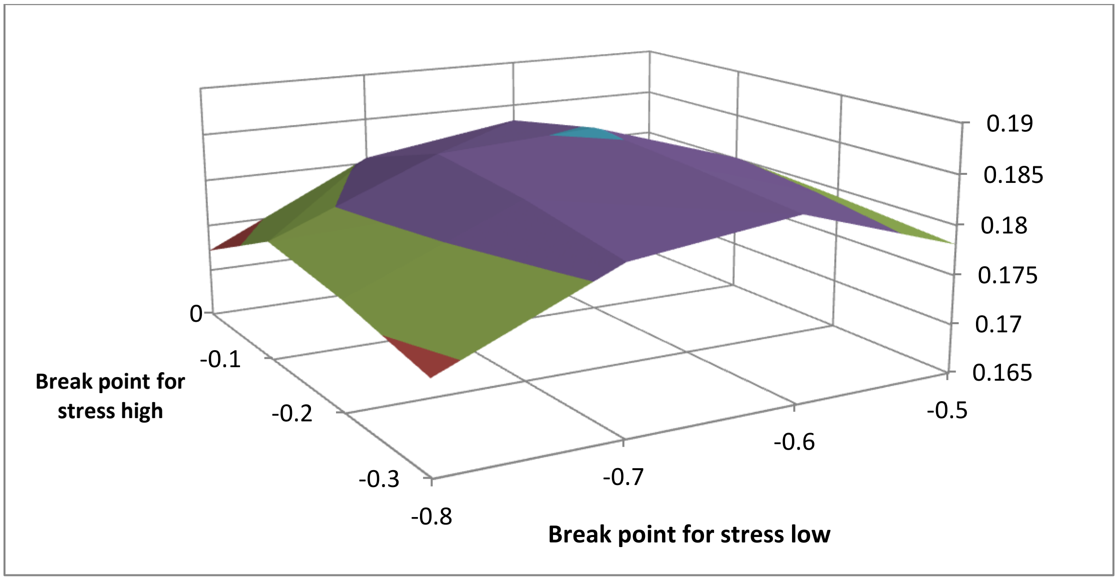

FIGURE 1: McFadden's LRI as a Function of the Room-Stress Cut-off Values.

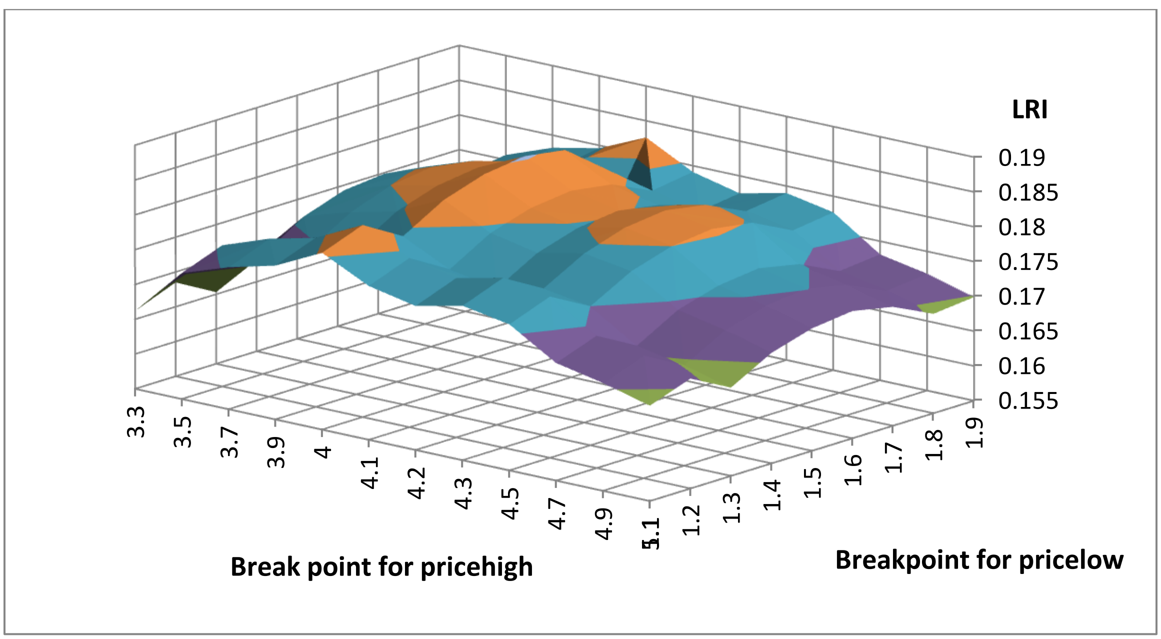

FIGURE 2: McFadden's LRI as a Function of the Cut-off Values for the Ratio of the Home Price to Income.

Variables Specification and Result Discussion

It is useful to notice that Roomstress is statistically significant. Using a standard model developed for the Panel Study of Income Dynamics (PSID) (MS Hill 1992) in the United States, we calculate the required number of rooms for each household, based on family size and composition, and compare it with the actual number of rooms.

For each head of household two rooms are allocated with or without a spouse. Then, one room is added for each additional adult aged 18 or over; one room is added for every two boys under 18, and one room for every two girls under 18. If there is an odd number of girls and an odd number of boys, then those under 10 years of age are paired regardless of sex. If there are an odd number of children then the numbers are rounded up (Clark 1992). Hence Roomstress calculates the mismatch between actual housing space and required housing space as given in the equation below:

$$Roomstress = required\ rooms / actual\ rooms - 1$$

Notice that Roomstress can be positive or negative. We translate this independent variable into two variables for modeling: RoomStresshigh and RoomStresslow. RoomStresshigh is set to one if Roomstress is greater than -0.1 and is zero otherwise. Similarly, RoomStresslow is set to one if Roomstress is less than -0.6 and is zero otherwise. If either of these conditions are true for a homebuyer and home combination, that home is less likely to be purchased.

Income, as expected, has a significant impact on the home selected. Pricehigh and Pricelow are each binary variables. Pricehigh takes on a value of one if the ratio of the price of the home and the annual household income is greater than 4.0 and is zero otherwise. Similarly, Pricelow takes on a value of one if the ratio of the price of the home and the annual household income is less than 1.6 and is zero otherwise.

Both have negative impact on utility. That is, if the home is relatively expensive for the prospective buyer or is too inexpensive in comparison to their means, the home is less likely to be purchased.

Langerudi et al. (2014), also used this measure but with break points of 5.0 and 1.6. Suhaida et al.(2011), finds that housing markets are rated as "affordable" at or below 3 times gross annual household income, "moderately unaffordable" at or below 4 times income, "seriously unaffordable" at or below 5 times income and above 5, as "severely unaffordable". Based on this dataset values of 4.0 and 1.6 provide the best fit as measured by the R-square value.

There is a statistically significant interaction between income and Roomstress. We compute the average income across the dataset and mark those buyers with an income that is above average (Incomeaboveave). The product of this independent variable with Roomstress is statistically significant. For buyers with incomes that are higher than the average this product implies that as Roomstress rises the home becomes increasingly relatively less desirable. However, for incomes that are lower than the average, this product implies the as Roomstress increases the home becomes increasingly desirable.

An interesting phenomena is that Asian buyers prefer homes that are in areas with relatively higher Asian populations (hhasian × Hasian). Finally, it is useful to notice that the distance from home to school is of concern for households with children present (Children1×School). That is, houses within one mile of a school are more preferable then those that are located farther from schools for buyers with children.

5. CONCLUSION

People's relocation decisions affect land use and urban form hence it is a critical area of study. Based on the assumption that the relocation decision can be conceptually separated into two sequential decision processes; the mobility decision process and the home selection decision process, we model these two behaviors separately.

In the mobility decision model, household characteristics, neighborhood characteristics and macro-economic factors were determined to be important aspects that helps to explain the behavior. In general, older households, ones with more adults, ones contain people over 60 year old, marital couple, and those in houses which have open spaces nearby are less likely to move. In addition, income level, education level, race and the unemployment rates are also found to be significant factors that affect the mobility decision. Income and the square foot of the existing home are found to be correlated in affecting household behavior. This is the first paper that has suggested the relevance of using the product of income and square feet.

In the analysis of home selection preference, although a mixed logit model was consider so as to provide for household heterogeneity, no significant heterogeneity was observed. Household level characters, zonal level characters and housing characters are considered in the model. “Room stress” is used to better reflect the relationship between household’s size and dwelling space. Housing affordability, which is defined as the ratio of the housing price and the household income, is found to be a significant factor and a reasonable range for housing affordability is found to be between 1.6 and 4.0. Household race and the presence of children are also found to be correlated with people’s home selection preference.

This is the first paper to use the AHS survey for this purpose. There are a wide collection of papers that stem from the analysis of this rich dataset include the development of price indices for housing, quantification of the extent of discrimination in housing prices, how housing prices are affected by changes in access to transportation, and a quantification of the trade-offs in communing distances between household members in the housing selection decision. This paper has extended the domains of application for this dataset to mobility decision-making. In order to support this extension, values for some of the independent variable for those households which did not move during the 8 year time period had to be established. This was done by randomly selecting a year for each household that did not move for which the decision not to move was made. Using this year, the age, home tenure and unemployment related independent values were selected.

One important shortcoming of using the AHS survey for this research is that this dataset only contains commute travel time for the home selected. When modeling home selection, this factor is very important and in this dataset there is not ability to compute what the commute would have been for other homes in the synthesized choice set. This is unfortunate because an independent variable of that nature is likely to be highly significant.

REFERENCES

Bhat, Chandra R., and Jessica Guo. "A mixed spatially correlated logit model: formulation and application to residential choice modeling." *Transportation Research Part B: Methodological* 38.2 (2004): 147-168.

Brown, Lawrence A., and Eric G. Moore. "The intra-urban migration process: a perspective." *Geografiska Annaler. Series B, Human Geography* 52.1 (1970): 1-13.

Clark, William AV. "Comparing cross-sectional and longitudinal analyses of residential mobility and migration." *Environment and Planning A* 24.9 (1992): 1291-1302.

Clark, WA William AV, and Frans M. Dieleman. *Households and housing: Choice and outcomes in the housing market*. Transaction Publishers, 1996.

Clark, William AV, and Youqin Huang. "The life course and residential mobility in British housing markets." *Environment and Planning A* 35.2 (2003): 323-340.

Dieleman, Frans M. "Modelling residential mobility; a review of recent trends in research." *Journal of housing and the built environment* 16.3-4 (2001): 249-265.

Eliasson, Jonas, and Lars-Göran Mattsson. "A model for integrated analysis of household location and travel choices." *Transportation Research Part A: Policy and Practice* 34.5 (2000): 375-394.

Gabriel, Stuart A., and Stuart S. Rosenthal. "Household location and race: Estimates of a multinomial logit model." *The Review of Economics and Statistics* (1989): 240-249.

Gan, Li, and Qinghua Zhang. *Market Thickness and the Impact of Unemployment on Housing Market Outcomes*. No. w19564. National Bureau of Economic Research, 2013.

Habib, Khandker M. Nurul, and K. Kockelman. "Modeling choice of residential location and home type: recent movers in Austin Texas." *87th Annual Meeting of the Transportation Research Board, Washington, DC*. 2008.

Habib, Muhammad Ahsanul, and Eric J. Miller. "Reference-dependent residential location choice model within a relocation context." *Transportation Research Record: Journal of the Transportation Research Board* 2133.1 (2009): 92-99.

Habib, Muhammad Ahsanul. *Microsimulating residential mobility and location choice processes within an integrated land use and transportation modelling system*. Diss. University of Toronto, 2009.

Hill, Martha S. "The Panel Study of Income Dynamics: a users guide." (1992).

Kan, Kamhon. "Expected and unexpected residential mobility." *Journal of Urban Economics* 45.1 (1999): 72-96.

Langerudi, Mehran Fasihozaman, et al. "Choice Set Imputation." *Transportation Research Record: Journal of the Transportation Research Board* 2429.1 (2014): 79-89.

Lee, Brian HY, and Paul Waddell. "Residential mobility and location choice: a nested logit model with sampling of alternatives." *Transportation* 37.4 (2010): 587-601.

Liu, Zheng, Jianjun Miao, and Tao Zha. *Land prices and unemployment*. No. w19382. National Bureau of Economic Research, 2013.

McFadden, D. "Modeling the Choice of Residential Location," in Spatial Interaction Theory and Planning Models, ed. by A. Larlqvist, L. Lundqvist, F. Snickars, and J. Weibull, North-Holland." (1978).

Pinjari, Abdul Rawoof, et al. "Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions." *Transportation* 38.6 (2011): 933-958.

Plaut, Pnina O. "The intra-household choices regarding commuting and housing." *Transportation Research Part A: Policy and Practice* 40.7 (2006): 561-571.

Rosenfield, Adam, Franco Chingcuanco, and Eric J. Miller. "Agent-based Housing Market Microsimulation for Integrated Land Use, Transportation, Environment Model System." *Procedia Computer Science* 19 (2013): 841-846.

Rossi, Peter Henry. *Why families move: A study in the social psychology of urban residential mobility*. Glencoe: Free Press, 1955.

Strassmann, W. Paul. "Residential mobility: contrasting approaches in Europe and the United States." *Housing Studies* 16.1 (2001): 7-20.

Suhaida, M. S., et al. "Housing affordability: a conceptual overview for house price index." *Procedia Engineering* 20 (2011): 346-353.

Tatsiramos, Konstantinos. "Residential mobility and housing adjustment of older households in Europe." (2006).

Waddell, Paul. "A multinomial logit model of race and urban structure." *Urban Geography* 13.2 (1992): 127-141.

Wang, Liming, Paul Waddell, and Maren L. Outwater. "Incremental Integration of Land Use and Activity-Based Travel Modeling." *Transportation Research Record: Journal of the Transportation Research Board* 2255.1 (2011): 1-10.

Weisbrod, Glen, Steven R. Lerman, and Moshe Ben-Akiva. "Tradeoffs in residential location decisions: Transportation versus other factors." *Transport Policy and Decision Making* 1.1 (1980): 13-26.