Application of Hazard Based Model for Housing Location Based on Travel Distance to Work

Behnam Amini[1], Mahmood Javanmardi [2], Kouros Mohammadian[3]

Submit: 12.07.2014 *Accepted: 25.12.2014*

Abstract:

Residential location choice modeling is one of the areas in transportation planning that attempts to examine households location search behavior incorporating their trade-offs between housing quality, prices or rents, distance to work and other key factors. This brings up the need to come up with methods to logically allocate credible choice alternatives for individuals.This article attempts to provide a detailed study of this practice to develop a modeling framework that can replicate the choice process. In order to show the potential of the method, a decision criterion—maximum distance to work—is considered the potential attribute that the household evaluates for feasible housing alternatives. It is postulated that alternatives will only be included in the choice set if the maximum work distancesatisfies the household thresholds. This research explores the application of proportional hazard models in the housing search process. Some of the specifications of hazard-based models that are typically used on temporal data are examined on work distance. A log-logistic function is used for hazard base-line. The study has used the household travel behavior survey conducted by Chicago Metropolitan Agency for Planning (CMAP). Furthermore, several extensive land use and transportation related data sources are incorporated to complement the scope of the modeling results.

Keywords: Residential location, spatial structure, hazard-based models, commute distance

---

Corresponding author Email: behamin95@yahoo.com

1- Associate Professor, Imam Khomeini International University, Ghazvin, Iran

2- Ph.D. Candidate, University of Illinois at Chicago, Illinois, Chicago, USA

3- Professor, University of Illinois at Chicago, Illinois, Chicago, USA

1. Introduction

The spatial location decision-making process has become a topic of interest in many fields, including transportation, urban planning, psychology, and other relateddisciplines. Since the early introduction of the discrete choice paradigm [McFadden, 1978], the individual's alternative selection behavior has been primarily modeledusing the discrete choice modeling approach. However, the prediction potential ofa discrete choice model and the accuracy of its parameter estimates are highlydependent on the choice set composition. This study examines a behavioral method for housing location choice.

Previously, Thill and Horowitz [1991] discussed an approach to the context ofdestination choice, in which they used travel time for alternative screening andassumed that it was an unobservable random variable that did not depend on anyobservable attribute of travelers. In the current article, after a decision maker becameactive in the housing market, it was assumed that residential location choice process started with an alternative evaluation and screening practice. People scanned theiralternatives first, then filtered them based on their priorities, lifestyle, preferences, budget, and perceived utilities. Finally, among the filtered alternatives, the most desired option with the highest utility was chosen. While several factors affect the selection of housing alternatives and the spatial choice decision mechanism (e.g., property value, commute distance, school quality, safety, tax rate, etc.), in order to show the practicality of the approach, an influential factor such as commute distance, known as an essential variable in residential selection behavior, is considered in the screening process model of this study [Kim 1992; Van Ommeren et al. 1997; Rashidi et al. 2011].

The analogy between duration and commute distance (spatial duration) is justifiable from a mathematical perspective. Furthermore, estimation, specification, tests, and diagnostic issues can be easily handled using the duration modeling methodology. This analogy has been intermittently announced in some research areas with claims that the spatial duration model can be a useful toolbox for analyzing residential location search behavior [Odland and Ellis 1992; Diggle 1983; Boots and Getis 1988].Nonetheless, some conceptual difficulties in the interpretation of spatial duration models have not yet been sufficiently addressed in this emerging field. At the same time, the advantages of applying the longitudinal framework in the context of spatial duration have not been appropriately discussed.

This paper, in general, examines accessibility as a major factor of housing location and in particular focuses on the effect of distance to work along with the other variables such as income, housing price and security parameters. The same methodology could follow to deal with the other accessibility variables including distance to retail stores, schools and etc. In short, the major contribution of this article is the exploration of the viability of using proportional hazard formulation in this application. In addition, the utilization of a log-logistic function for hazard base-line will be examined.

The remainder of the article is structured as follows. First, a brief literature review is presented and the study approach is discussed. Model derivation and the mathematical formulations of the system of equations are presented next. The data sets used in this study are then explained, and their key variables are discussed. Following that, experimental results of different steps of the parameter estimation process are presented. Conclusions are discussed in the final section.

2. Background

2.1 Residential Location

Residential location choice modeling is in the stage of disaggregate computational and econometric models. Researchers have used many different econometric discrete choice models to address this problem. Some studies have focused on single aspects of households ' concerns in residential location choice; for instance, commuting factors in residence choice [Clark and Withers 1999], accessibility to non-work activities [Ben-Akiva and Bowman, 1998], travel mode choice [Pinjari et al. 2008] or modeling challenges such as choice set formation in multinomial logit models (MNL) or non-MNLs [Guevara and Ben-Akiva, 2013]. In terms of different types of discrete choice models, researchers have applied various modeling structures.

A critical step in residential location choice problem through discrete choice modeling is the choice set formation which is associated with the level of geographic aggregation for the alternatives. Even though it is ideal to set parcels of land or buildings as the alternatives that households encounter in choosing housing location [Lee et al. 2010], computational and data availability issues compel researchers towards using more aggregate geographic levels such as neighborhoods or traffic analysis zones [Guo and Bhat, 2007]. However, the nature of this problem imposes a computational barrier which is the large number of alternatives even in the aggregate case. In order to avoid the infeasible or hardly achievable computational issue of large number of alternatives, researchers have tried different sampling methods to shrink the choice set. It is noteworthy to mention that there are also arguments other than the computational issue against the idea of universal choice set that question the knowledge of individuals about the entire choice possibilities [Fotheringham, 1988].

Auld and Mohammadian [2011] applied time space prism constraint to form the choice set for destination choice in their destination choice model of ADAPTS activity-based model. In order to assess the performance of sampling methods, Zolfaghari et al [2011] presented a comparison among several common sampling methods like importance sampling with bias correction, importance sampling without bias correction and random sampling. Langerudi et al [2014] proposed a two-step model for choice set formation using a weighted stratified sampling method. Then hazard-based models were applied to filter conditionalalternatives for prediction step of a previously estimated model. The results of prediction for this two-step model were compared with the results generated from a multinomial logit model.

2.2 Spatial Hazard-based Models

Cox discussed the basics of duration models as an analysis of exponentially distributedlifetimes [1959]. These statistical techniques were generalizedin another greatly referenced paper of him in 1972. In his later paper, he discussed asatisfying analysis method for failure times, such as the length of time a person isalive. Generally speaking, in a duration model, the time to reach failure, loss, orcensoring is observed for each individual in the population. Later, Cox andOakes published a book titled, Analysis of Survival Data in which they provideda comprehensive study about hazard models and related topics [1984]. This book is oneof themost important references in the duration modeling literature. Since theseearly discussions about hazard-based and duration models, an extensive amountof discussions and applications of these models have been presented [Han and Hausman 1990; Rashidi et al. 2011]. Nonetheless, it has been only recently that discussions about the conceptual equivalence between spatial duration models with their temporal counterparts can be found in the literature [Waldorf, 2003; Carruthers et al. 2009].

Commute distance, known as a key variable in housing search behavior, isutilized in this study to construct the choice set of the housing decision. It is wellknown that the distance between residence and job location has a significant impacton residential location choice behavior. The correlation between job and residentiallocations has been studied extensively in the literature, in most of which commutedistance is distinguished as the critical factor [Van Ommeren et al.1997, 1999; Rashidi et al. 2011]. However, there is a gap in the literature for studying the commute distance as the dependentvariable of a spatial duration model. This article attempts to address this missingresearch gap.

3. Model Formulation

In this section, an introduction to the parametric hazard-based models is presented. Cox, who pioneered the area of hazard models in 1959, presented the early versions of hazard models with the Weibull baseline hazard, which will be further discussed in this section. Since then, the Weibull function has been frequently used in the duration modeling context by many other researchers. It should be noted that in all of the formulations discussed in this section, duration can be replaced with distance without losing the generality.

The length of a spell for a subject (e.g., a household) is translated in the hazard formulation as a continuous random variable T with a cumulative distribution function (CDF), F(t) and probability density function (PDF), f(t) where t is the elapsed time since entry to the state at time 0. The survival function is defined as 1 -F(t) and is also known as the failure function. In the mathematical context, the failure function can be written as:

$$\Pr(T \le t) = F(t). \tag{1}$$

Therefore, the survival function can be written as:

$$\Pr(T > t) = 1 - F(t) = S(t). \quad (2)$$

PDF, which is the slope of the failure function, is

$$f(t) = \lim_{\Delta t \to 0} \frac{\Pr(t+\Delta t \geq T \geq t)}{\Delta t} = \frac{\partial F(t)}{\partial t} = -\frac{\partial S(t)}{\partial (t)} \quad (3)$$

Where, Δt is a very small infinitesimal interval of time. The hazard rate can be defined then as the probability of leaving in the interval,(t, t + Δt) , conditional on survival up to time t:

$$\lambda(t)dt = \Pr(t + \Delta t \geq T \geq t | T \geq t) = \frac{f(t)dt}{S(t)} = \frac{S'(t)dt}{S(t)} \quad (4)$$

Where, λ(t) is the probability of failure for individual i given that it has survived until time T,f(t)is failure PDF and S(t) is the survival function.
The survival function can be calculated using equation (4) as:

$$S(t) = exp\left[-\int_0^t \lambda(u)du\right] \quad (5)$$

When the impacts of person-specific covariates, like socio-demographic attributes, built-environment variables, and macroeconomic factors are included in a hazard model, it is called a proportional hazard model [Cox 1959], which can be formulated as:

$$\lambda_i(t) = \lambda_0(t) \exp(-\theta_x X_i) \quad (6)$$

Where, is the baseline hazard function, which only depends on t (but not X). It represents the pattern of duration dependence and is assumed to be common among all households. is the coefficient of the covariates. The exponential part of equation (6) scales the baseline hazard function depending on the covariates. The most important property of the proportional hazard is that absolute differences in X imply proportionate differences in the hazard values at a specific time t. In mathematical terms:

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)\exp(-\theta_x X_i)}{\lambda_0(t)\exp(-\theta_x X_j)} = \exp\left[-\theta_x(X_i - X_j)\right] \quad (7)$$

Consequently, the proportionate change in the hazard function can be shown by a unit change in covariate X. Unlike the nonnegative part for the covariates, which is always used in an exponential form, the baseline hazard part can take several shapes among which log-logistic is a well-known function [Cox, 1972]:

$$\lambda_i(t) = \frac{\frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}}{1+\left(\frac{t}{\alpha}\right)^{\beta}} \exp(-\theta_x X_i) \quad (8)$$

where and are scale and shapeparameters of log-logistic distribution, X denotes explanatory variablesis the vector of parameters. Using the same definitions, the survival function with log-logistic assumption for the baseline hazard can be shown as:

$$S_i(t) = \left(1 + \left(\frac{t}{\alpha}\right)^{\beta}\right)^{-\exp(-\theta_x X_i)} \quad (9)$$

In a mathematical language, the likelihood of failure in accepting a distance while examining different alternatives is equal to the hazard of failure to accept the alternative times the probability of surviving without accepting it. The likelihood function is used to estimate the model parameters.

4. Data

The geographic scope selected for this study is Chicago's 7 county metropolitan area in Northeastern Illinois including Cook, Du Page, Kane, Lake, Kendall, McHenry and Will counties which have a total household population of approximately 2.9 million households covering 1711 Traffic Analysis Zones (TAZ). This study has used the Travel Tracker Survey conducted by Chicago Metropolitan Agency for Planning (CMAP). The survey was designed for the purpose of regional travel demand modeling and included over 10,000 households, providing a detailed travel inventory for the members of each household as well as socio-demographic information. Furthermore, the exact coordinates of home and work location of the households were available to examine various aspects of transportation and land use accessibilities in GIS. The final sample for the study was truncated to about 6000 samples that contained the necessary information required for the analysis in this work noting that the primary distribution of the sample was preserved in this process. One of the barriers in this study was the static source of data, the fact that the previous housing locations of the households were not available to conceive the pattern behind their movement and as a result the self-selection bias could be a potential issue. However, that should not be a point of concern in this study as the focus of the paper is on choice set formation.

The paper has focused on zonal level residential choice; therefore, TAZs were selected as the zonal level of geography and detailed TAZ level built environment attributes were attained from a number of different data sources. Land Use Inventory of Chicago 2005 was used to extract accessibility measures to land use categories such as Urban Mix, Shopping Malls, and Office. Urban Mix land use category includes retail trade, but neither in shopping malls, office campuses, single structure offices nor hotels according to the Land Use Inventory definition. It basically includes retail trade services such as general merchandise, food, vehicular, eating and drinking places, etc.

CMAP also provided aggregate data for property values in TAZs. A procedure was implemented to obtain average property values for a housing unit based on total housing units in TAZs. Assessment factors were extracted from county assessors' website and applied to adjust the property values.

The other source of data provided by CMAP was the number of jobs in various employment categories which could represent the regional job opportunities. Moreover, census data was used to access zonal demographic information like racial composition. Even though direct TAZ attributes could not be made through Census data, Census tract attributes were assigned to TAZs through spatial join in GIS. By utilizing GIS tools, transportation accessibility and distance to the nearest available transit opportunities were calculated with the help of various transit shapefiles of urban and suburban rail and bus transit system in Chicago region.

Finally, school quality data was extracted from Public School Ranking website (2001) and the index is based on SAT exam score. The crime data factor was accessed through Police Department website, the comprehensive publicly available data (2010) for all the crimes occurred as well as their exact location. A number of main crime types such as homicide, assault, sex offense were selected and simply added to represent the crime level of the TAZs and it was proportioned to the highest number of crimes observed among the TAZs to come up with an index between 0 and 1. Most of the TAZs were ranked in the range under 0.1.

Table 1 demonstrates the summary statistics of the variables used in this study. The first section shows the household level variables and the second part represents the zonal attributes. The mean of the crime index is 0.04, i.e. a considerable number of the TAZs are in the safe zone with few harsh crime occurrences with respect to the worst TAZ in terms of crime frequency. The distance to transit stations are shown in natural logarithm scale and the main variables from which they are extracted were in ft scale. Correlation analysis of the data showed a considerable association between number of cars and distance to rail stations as well as distance to urban mix and malls. Using the clustering GIS tools, racial clustering among Asian and Black families issue was also another factor that must be considered in residential location choice problems.

Table 1. Descriptive Analysis of the Explanatory Variables

5. Calibration of the Hazard Model

As stated earlier, it is still a gap in the literature on the possibility of utilizing a different methodology for housing location. It is always desirable to estimate a reliable model that has capability of prediction over different household choice criteria. In order to combine statistical and behavioral soundness, a log-logistic hazard-based models was developed to justify the behavioral choice of the households for residential location. The hazard-based models have the potential to constrain household choices based on a criteria, such as maximum distance to work thresholds. Based on household socioeconomic attributes, households are assumed to have particular tolerance level for distance threshold which is obtained with the help of hazard model.

Log-logistic distribution was selected for probability distribution of distance to work for households. Typically, the scale parameter in log-logistic distribution specifies the role of socioeconomic characteristics of a household in the probability distribution of distance to work.

Scale parameter: $\alpha = e^{(\theta_0 - \hat{\theta}\hat{x})}$ (10)

Using the probability density functions, a likelihood

variable Name	Definition	Min	Mean	Max	SD
Household Level					
NPerson	Number of Persons	1.0		8.0	
HHIncome	Household Income ($)	10000	2.27	100000	1.27
NWorker	Number of Workers	0	66104	5.0	30902
NStudent	Number of Students	0	1.25	6.0	0.89
Ncar	Number of Cars	0	0.52	4.0	0.93
LStay	(Length of Stay in Current Location (years	1.0	1.64	9.0	0.98
Ndriver	Number of Drivers	0	4.1	6.0	1.12
NChild	Number of Children	0	1.66	6.0	0.81
AgeHead	Age of Household Head	18	0.47	99	0.91
Zonal Level	(Crime Level Index between 0(low) and 1 (high		54.6		16.02
CrimeIn	School Quality Index between 0 and 100	0		1.0	
SchoolIn	Log Distance to the nearest Metra station	0	0.04	97	0.13
LogDisMetra	(Suburban Rail)	4.76	30.29	11.64	19.41
	Log Distance to the nearest CTA rail station		9.231		0.892
LogDisCTARail	(intra-urban rail)	5.0		12.0	
	Number of CTA Bus Stops per sq miles		8.216		1.146
CTABusstops	(Intra-urban Bus)	0		269	
	Log Distance to the nearest PACE stop		62.7		49.1
LogDisPACE	(Suburban Bus)	1.02		11.33	
	Percentage of White people in a TAZ		7.867		1.760
ZWhite	Percentage of Black people in a TAZ	0.003		0.96	
ZBlack	Percentage of Asian people in a TAZ	0.003	0.70	0.97	0.23
ZAsain	Average Housing (unit) Market Value in a TAZ	0.0003	0.14	0.4	0.23
AvgValue	Average Total Employment in a TAZ	10000	0.05	4092000	0.06
AvgEmp	(Average Distance to UrbanMix Land (ft	11	333510	19484	713396
DistUrbanMix	Average Distance to Malls	0	2433.3	25583	4627.7
DistMall	Average Distance to Office Land	600	2361	35000	3056
DistOffice	(Log number of jobs within 30 minutes drive (am peak	0	10381	47500	14091
lnTAZJOB30		6.8	9498	17.4	11818
			12.402		1.013

function can be written to estimate thehousehold-specific coefficients according to:

$$£_d = \prod_{i=1}^{n} f_i(d) \quad (11)$$

Table 2 shows the result of the hazard model developed for distance to work. For distance to work model, income, car ownership and age of household's head member are the key determinant of the hazard equation. It is noteworthy to remind that the covariates in the model are formulated with a negative sign. To interpret that, for example, income parameter has come out positive meaning that its effect on the hazard function is negative. Therefore, higher income reduces the hazard rate i.e. affluent households are probabilistically inclined to housing further away from their work location. Car ownership is the next determinant in making longer distances to work location possible. On the other hand, age has negative association with distance which means households with elderly head members tend to locate closer to their work location.

Log likelihood function for constant model along with log likelihood measure for the MLE is given. The likelihood ratio statistic equals 31. In order to assert that the model with household-specific variables works better than a model with just constants (just parametersand), the null hypothesis of the constants must be tested. Based on Wilks theorem, the likelihood ratio statistic must have a chi-square distribution with 3 (=5-2) degrees of freedom. The likelihood ratio statistic for this model is 31 which is placed at the very end of the right tale of this chi-square distribution letting us to reject the null in favor of the model with 99% significance level.

6. Results

In order to try a new method for housing location, an intuitive hazard-based approach was implemented for predicting the location choice of a hold-out sample. The performance of the predictions is explored by comparing the results to the actual location choices of the hold-out sample.The intuitive hazard-based approach is comprised of a log-logistic hazard model for acceptable housing distance to work as explained in the previous section. This model was estimated to be applied in limiting the choice set of individuals for more behavioral and realistic choices relative to household socioeconomic conditions. In other words, the model introduces household-specific hazard and probability density function for acceptable housing distance to work. Based on the probability density function, one can filter the less probable choices through various filtering methods such as cutoffs through cumulative density function (CDF). The filtering used in this method was based on %80 CDF for distance to work displayed as the following:

Table 2. Log-logistic Hazard Model

	Distance to Work Model	
Parameter	Estimate	t-value
	-0.56	-4.01
	0.35	21.57
Income (×1/1000)	0.001	1.97
Number of Cars	0.025	1.78
Household Age of the head	-0.004	-4.16
Summary Statistics		
Number of Observations	6047	
log likelihood function for null hypothesis (just ,)	-8623	
log likelihood function	-8608	
-2((0)-())	31	

: maximum acceptable distance to work

The choice of percentage cut-off is very crucial and extreme cut-offs could eliminate a large portion of alternatives. On the other hand, very conservative cut-offs might not represent the real choice set formation behavior. Having said that, range of %80 coverage was found as an acceptable balance not to mention the significant need for research in this arena. Since the hold-out sample was small including approximately 2000 households, comparing zonal re-location of the households to their actual zone residence was not rational; therefore, to rationalize the process the result of the zonal predictions were aggregated to sub-counties in suburban counties and neighborhoods within city limits. For representing the performance of the predictions, looking for the re-location zone assigned to each household to be exactly the same as their actual zone, is neither feasible nor logical; however, comparing the similarities between them is an acceptable comparison strategy. Consequently, Figure 1 is displayed to show the distribution of median income through prediction method and its comparison to the actual distribution geographic profile. Comparison between the actual sample distribution and the simulated hazard-based predictions are hardly achievable by just looking at the figure. As a result, to facilitate the comparison, root mean squared error (RMSE), and average relative error between the prediction result and the actual profile were calculated to show the scale of the induced errors. Table 3 shows the RMSE and relative error for certain household socioeconomic variables aggregated for sub-county and neighborhood level of geography.

RMSE and relative error are calculated based on the formulations below:

$$RMSE = \sqrt{\frac{1}{n}\sum(\hat{X}_i - X_i)^2} \qquad (12)$$

i : Each single sub-county or neighborhood within Chicago seven county area that contain assigned samples in actual and predicted conditions

X_i: An actual household attribute X_i:A predicted household attribute

Relative Error for non-zero attributes: $\hat{\eta} = \frac{(\hat{X}_i - X_i)}{X_i}$ The table shows that for most of the attributes the relative error decreases when hazard model is used for prediction choice set. One of the key attributes is income distribution throughout the geography which is the best determinant for prediction. The results show that the error in income distribution is %5 less when an intuitive approach is used for prediction. Moreover, the error in most of the other variables including number of persons, students, cars, children, income and distance to work has decreased. Even though one example might not be sufficient to get to a broad conclusion, the result brings about the differences between the effectiveness of choice set formation for model prediction vs. estimation. It is the fact that estimation is the statistical process that requires maximum information and a broad choice of alternatives while prediction is the intuitive process that should get along with common sense.

7. Conclusions and Future Directions

This study presented a behavioral model for residential location choice problem. In a housing search process, one can consider an approach in which alternatives are

Table 3. RMSE and relative error

Variable	RMSE	
number of persons	0.83	0.21
Household income $	19231	0.18
Number of workers	0.48	- -
Number of students	0.81	- -
Number of cars	0.63	- -
Number of drivers	0.52	- -
Number of children	0.77	- -
Household Age of Head	7.16	0.09

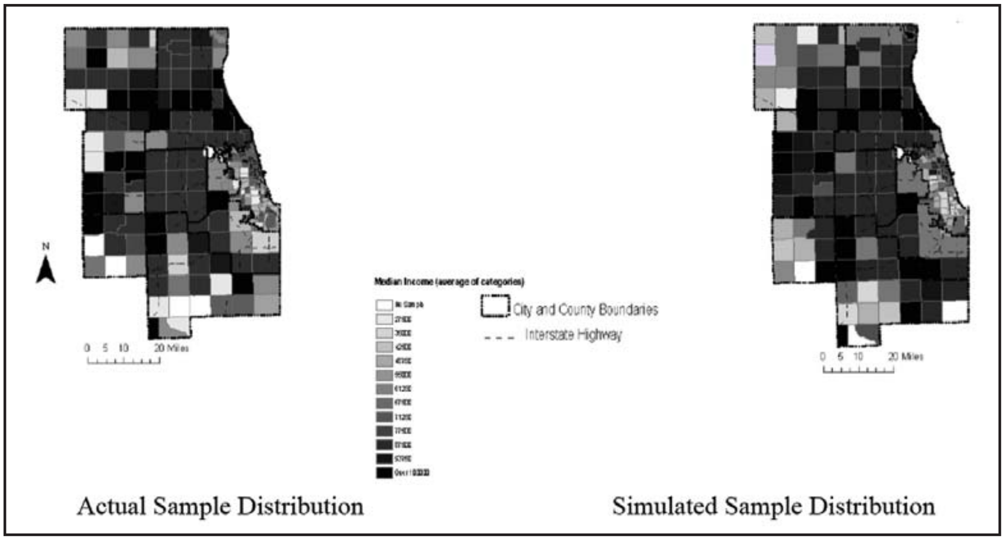

Figure 1. Comparison of median income distribution profile

evaluated and screened based on household priorities, lifestyle, and preferences and for each alternative. The probability of being selected in the choice set is estimated. Following that, the alternative with the highest utility can be selected using traditional choice models.

The housing screening process of this study is modeled using the maximum distance to work as a continuous variables. The analogy between distance and duration implied the application of hazard-based formulation to model the willingness of the employed household members to accept a commute distance. By contributing to the literature of the recently introduced spatial hazard models, this study explored and discussed the application of spatial log-logistic hazard-based models for housing search behavior modeling

The Chicago Metropolitan Agency for Planning (CMAP) data was used in this study for the modeling practice, along with other sources of data, such as built-environment, land-use, and economic factors. Many household socio-demographic attributes and several landuse indicators were tested in the modeling process. Further improvements to the model include investigating the importance of variables (beyond work distance) on housing search choice set formation. These improvements remain as future research tasks. It should also be noted that the application of the proposed modeling framework is not limited to the housing search problem.Such a framework can be used in other contexts in which a large number of alternatives should be evaluated. For instance, in the case of activity location choice (e.g. shopping), a similar approach can be used.

For future work, it is valuable to improve the prediction potential with more realistic alternative sampling approaches in prediction step to account for more detailed household socioeconomic conditions along with housing supply information to avoid over-concentration of households in certain geographic locations. Interestingly, the hazard-model that was used for the prediction step of this model eliminated a small portion (%20) of improbable alternatives through the tail of distance to work probability density functions; however, due to the few number of observations used to predict the model, the issue of over-concentration was not a significant problem. However, for large number of observations, fine consideration of housing supply and capacity constraints must be implemented to balance out the prediction results as the prediction choice set strategy becomes more behavioral.

8. References

-Auld, J. and Mohammadian, A. (2011)”Planning constrained destination choice in the ADAPTS activity-based model”, Paper presented at the 90th Annual Transportation Research Board Meeting, Washington DC.

-Ben-Akiva, M. E. and Bowman, J. L. (1998) “Integration of an activity-based model system and a residential location model", Urban Stud. 35, pp. 1131-1153.

-Ben-Akiva, M. E. and Lerman, S.R. (1985) "Discrete choice analysis: theory and application to travel demand", The MIT Press, Cambridge.

-Ben-Akiva, M. and Watanatada, T. (1981)" Application of a continuous spatial choice logit model, Structural analysis of discrete data with econometric applications". MIT Press, Cambridge, Mass.

-Boots, B. N. and Getis, A. (1988) "Point pattern analysis", Newbury Park, CA: Sage.

-Cox, D. R. (1959)" The analysis of exponentially distributed life-time with two types of failures", Journal of Royal Statistical Society 21B: pp. 411–421.

-Cox, D. R. (1972)" Regression models and life-tables" Journal of Royal Statistical Society 34B: pp.187-220.

-Cox, D. R. and Oakes, D. (1984) "Analysis of survival data", London, UK: Chapman and Hall.

-Clark, W.A.V. and Withers, S. D.(1999) "Changing jobs and changing houses: mobility outcomes of employment transitions". J. Regional Sci. 39, pp. 653-673

-Carruthers, J. I., Lewis, S. Knaap, G. J. and R. N. Renner. (2009) "Coming undone: a spatial
hazard analysis of urban form in American metropolitan areas.'' Papers in Regional Science 89:pp.65–88.

-Diggle, P. (1983) "Statistical analysis of spatial point patterns", New York, NY: Academic Press.

-Fotheringham, A. (1988) "Consumer store choice and choice set definition", Marketing Science, 7, pp. 299-310.

-Guevara, C.A. and Ben-Akiva, M. E. (2013) "Sampling of alternatives in multivariate extreme value (MEV) models", Transportation Research Part B: Methodological, pp. 31-52.

-Guo, J.Y. and Bhat, C. R. (2007)" Operationalizing the concept of neighborhood: Application to residential location choice analysis", J. Transp. Geogr. 15, pp.31-45.

-Habib, M. A. and Miller, E. J. (2008) "Modeling residential mobility and spatial search behaviorestimation of continuoustime hazard and discrete-time panel logit models for residential mobility", In: Transportation Research Board 87th Annual Meeting, Washington, DC.

-Han, A., and J. A. Hausman. (1990) "Flexible parametric estimation of duration andcompeting risk models." Journal of Applied Econometrics 5:pp.1–28.

-Langerudi, M. F., M. Javanmardi, A. Mohammadian, A.and Sriraj, P. S.(2014) "Choice set imputation: two-step weighted stratified and hazard-based approach", Transportation Research Record, Journal of Transportation Research Board, No. 2429,1. pp. 79-89.

-Langerudi, M. F., Javanmardi, M., Mohammadian, A., Sriraj, P. S. and Amini, B. (2013) "Choice set formation of housing location: A comparative analysis", in West Virginia University, 46th annual meeting.

-Lee, B. H.Y., Waddell, P. A., Wang, L. and Pendyala, R. M. (2010)" Operationalizing time-space prism accessibility in a building-level residential choice model: empirical results from the Puget Sound region". Environmental .Planning , .A 42.

-Kim, S. (1992) "Search, hedonic prices and housing demand", The Review of Economics and Statistics 74: pp.503–508.

-Lerman, S. R. (1984) "Recent advances in disaggregate demand modeling", In Transportation Planning Models, edited by M. Florian. Amsterdam, The Netherlands: North-Holland.

-McFadden, D. (1974) "Conditional logit analysis on the temporal stability of disaggregate travel demand models", Transportation Research Part B 16: pp.263–278.

-McFadden, D. (1978)" Modelling the choice of residential location." In: Karlqvist, A., Lundqvist, L., Snickars, F.,Weibull, J. (eds.), Spatial Interaction Theory and Planning Models, pp. 75-96.

-Odland, J. and Ellis, M. (1992) “Variations in the spatial pattern of settlement locations: an analysis based on proportional hazards models”, Geographical Analysis, 24: pp.97–109.

-Pinjari, A. R., Pendyala, R. M., Bhat, C. R. and Waddell, P. A. (2008) “Modeling the choice continuum: Integrated model of residential location automobile ownership bicycle ownership and commute tour mode choice decisions”, Transportation Research Board 87th Annual Meeting, Washington, DC

-Rashidi, T. H., Mohamamdian, A. and Koppel man, F. S. (2011) “Modeling interdependencies between vehicle transaction, residential relocation and job change”, Transportation 38: pp. 909–932.

-Thill, J. C., and Horowitz, J. L. (1991) “Estimating a destination-choice model from a choice-based sample with limited information”, Geographical Analysis 23: pp.298–315.

-Van Ommeren, J., P. Rietveld, and P. Nijkamp. (1997)” Commuting: In Search of Jobs and Residences.” Journal of Urban Economics 42: pp.402–21.

-Waldorf, B. S. (2003). “Spatial point patterns in a longitudinal framework.”International
Regional Science Review 26: pp.269–88.

-Zheng, J.,Guo, J. (2008)” Destination Choice Model Incorporating Choice Set Formation”,In: Transportation Research Board 87th Annual Meeting, Washington, DC

-Zolfaghari A., Sivakumar A., PolakJ. (2011),” Choice set formation in residential location choice modelling “, In: International Choice Modelling Conference