Households Long-Term Decision Making Process: Vehicle Transactions, Employment, and Residential Location Choices

Taha Hossein Rashidi

CATMUG Presentation

Department of Civil and Materials Engineering
University of Illinois at Chicago

August 2010
### Overview

- **Introduction and Motivation**
  - General Framework
  - Vehicle Transaction
  - Residential Relocation
  - Data

- **Hazard-Based Duration Models**
  - General Methodology
  - Different Parametric Models
  - Competing Hazard Model

- **Major Household Decisions**
  - Introduction
  - Formulation and Left Censorship
  - Model and Results

- **Housing Search Model**
  - Introduction
  - Choice Set Formation
  - Sample Selection Probability
  - Actual Choice Selection

- **Summary and Conclusion**
  - Conclusion
  - Future Directions
Introduction and Motivation
Hazard-Based Duration Models
Major Household Decisions
Housing Search Model
Summary and Conclusion

General Framework

- Land Use Data
- Population Synthesis
- Transportation System

Input

- External Variables:
  - Economic Variables
  - Market Factors

Home/Work Location Choice
Household Composition
Vehicle Ownership
Social Network Formation

Household Long-Term Context

Traffic Simulation
Activity/Travel Model

Taha Hossein Rashidi
University of Illinois at Chicago
General Framework

- Land Use Data
- Population Synthesis
- Transportation System

- Home/Work Location Choice
- Household Composition
- Vehicle Ownership
- Social Network Formation

- Household Long-Term Context

- Traffic Simulation
- Activity/Travel Model

Taha Hossein Rashidi  
University of Illinois at Chicago
General Framework

- Land Use Data
- Population Synthesis
- Transportation System

- Home/Work Location Choice
- Household Composition
- Vehicle Ownership
- Social Network Formation

- Household Long-Term Context

- Traffic Simulation
- Activity/Travel Model
General Framework

- Land Use Data
  - Home/Work Location Choice
- Population Synthesis
  - Household Composition
  - Vehicle Ownership
- Transportation System
  - Social Network Formation

Household Long-Term Context

Traffic Simulation

Activity/Travel Model

Long-Term Decision

Short Term Simulation
Introduction and Motivation
Hazard-Based Duration Models
Major Household Decisions
Housing Search Model
Summary and Conclusion

General Framework

Land Use Data
- Home/Work Location Choice

Population Synthesis
- Household Composition

Transportation System
- Vehicle Ownership
- Social Network Formation

Household Long-Term Context

Traffic Simulation
Activity/Travel Model

Long-Term Decision
Short Term Simulation

Taha Hossein Rashidi
University of Illinois at Chicago
Introduction and Motivation
Hazard-Based Duration Models
Major Household Decisions
Housing Search Model
Summary and Conclusion

General Framework

- Land Use Data
- Population Synthesis
- Transportation System

Regional or Local Freight Model

- Activity/Travel Model
- Traffic Simulation

Home/Work Location Choice
Household Composition
Vehicle Ownership
Social Network Formation
Residential Relocation
Vehicle Transaction
Data
Population Synthesis
System
Synthesis
Transportation
Household
Context
Long-Term Decision
Short Term Simulation

Taha Hossein Rashidi
University of Illinois at Chicago
Vehicle Transaction

• **Vehicle Ownership Models**
  
  – **Aggregate** Vehicle Ownership Models
    
    • Total number of vehicles in a zone during a period of time *(GDP, Fuel Price, etc)*
    
    • Watch a population cohort over time *(License holding behavior)*
    
    • Studying demand and supply of car market
  
  – **Disaggregate** Vehicle Ownership Models
    
    • Static Models
    
    • Dynamic Models
Vehicle Transaction

- Vehicle Ownership Models
  - Dynamic Models
    - Dynamic car-ownership models assume that no transaction will take place as long as the household maintains its utility level with respect to its vehicle fleet.
    - *Transaction timing* is the *central variable* in a vehicle transaction model or a dynamic vehicle ownership model.
    - Other vehicle attributes can be modeled conditional on vehicle transaction decision or even jointly.

General Framework

Vehicle Transaction

Residential Relocation Data

Vehicle

- Vehicle Transaction Type
- Vehicle Transaction Timing
- Vehicle Class
- Vehicle Model/Make
- Vehicle Vintage
- Vehicle Body Type
Vehicle Transaction

- **Vehicle Ownership Models**
  - Dynamic Models
    - A Joint vehicle transaction timing and type decision model is presented here
Residential Relocation

- Residential and Job Relocation Timing Decision
  - Residential and job search behaviors are commonly discussed together because of their close relationship.
  - This link between these two decisions, commute distance, has convinced the researchers to jointly model these two decisions.
  - Job search behavior is generally more complex than residential search behavior because more external agents such as the employer’s behavior, skill acquisition and existing job opportunities affect employment location opportunities.

So job relocation is not studied in detail in this study.
Residential Relocation

- Residential and Job Relocation Timing Decision
- Timing Decision (Jointly)
Residential Relocation

• Housing Search Model

• Conditional on residential relocation timing

- The housing search process starts with an alternative formation and screening stage. At this level households evaluate all potential alternatives based on their lifestyle, preferences, and utilities to form a manageable choice set with a limited number of plausible alternatives.
Residential Relocation

- Housing Search Model
- Conditional on residential relocation timing

- A household specific choice set is drawn from the entire possible alternatives in the area based on the average household work distance to each alternative.
Residential Relocation

- Housing Search Model
- Conditional on residential relocation timing

Following the choice set formation step, a discrete choice model is utilized for modeling the final residential zone selection of the household.
Data

- **Major Datasets**

  **Household automobile ownership in Toronto**
  - Panel data
  - 900 households
  - 9 years period from 1990 to 1998
  - Decision Making Unit (Individuals within a household that make vehicle ownership decisions in conjunction with each other)

  **Puget Sound Transportation Panel Survey**
  - A longitudinal panel survey
  - 10 waves from 1989 to 2002 in Seattle
  - The last three waves of the PSTP are used in this study to estimate the parameters of the model.

- **Additional Data**

  The built-environment characteristics were borrowed from an adjunct survey of the PSTP (Housing Search and Interdependencies among the major household decisions).
  Land values and house prices are mainly obtained by county assessment departments (Housing Search).
Hazard-Based Duration Models, Methodology

- **Continuous Formulation**
  
  - The hazard function can be expressed as a function of the probability density function $f(t)$ and the cumulative distribution function $F(t)$, as shown in the following equation.

$$h(t) = \frac{f(t)}{1 - F(t)}$$

The hazard, $h(t)$, gives the rate at which events (such as purchasing a vehicle) are occurring at time $t$, given that the event has not occurred up to time $t$.

$$S(t) = \frac{f(t)}{h(t)}$$
Analysis on Different Parametric Hazard Models

- **Model Definitions**

  Eight parametric hazard formulations are developed and their goodness-of-fits are compared against each other.

![Diagram](image-url)
Analysis on Different Parametric Hazard Models

- Modeling Results
- Eight Models with alternative baseline hazard and heterogeneity

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Likelihood at Convergence</th>
<th>Number of Parameters</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWNH</td>
<td>-1280.84</td>
<td>26</td>
<td>1366.42</td>
</tr>
<tr>
<td>CWGH</td>
<td>-1280.76</td>
<td>28</td>
<td>1372.93</td>
</tr>
<tr>
<td>CLNH</td>
<td>-1280.84</td>
<td>28</td>
<td>1373.01</td>
</tr>
<tr>
<td>CLGH</td>
<td>-1284.08</td>
<td>31</td>
<td>1386.12</td>
</tr>
<tr>
<td>DWNH</td>
<td>-1292.17</td>
<td>26</td>
<td>1377.75</td>
</tr>
<tr>
<td>DWGH</td>
<td>-1290.64</td>
<td>28</td>
<td>1382.81</td>
</tr>
<tr>
<td>DLNH</td>
<td>-1291.97</td>
<td>28</td>
<td>1384.14</td>
</tr>
<tr>
<td>DLGH</td>
<td>-1291.67</td>
<td>31</td>
<td>1393.71</td>
</tr>
</tbody>
</table>

\[
\text{BIC} = - \ln(L_C) + 0.5 \ p \ \ln(N)
\]

\[\ln(L_C)\] is the log-likelihood value at convergence

\[p\] is the number of parameters

\[N\] is the number of samples

The continuous model with monotonic baseline hazard outperforms other models including unobserved heterogeneity. Generally, models provide better results. Lower BIC implies better fit.
## Analysis on Different Parametric Hazard Models

- **Modeling Results**
- **Eight Models with alternative baseline hazard and gamma distributions for heterogeneity**

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Likelihood at Convergence</th>
<th>Number of Parameters</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWNH</td>
<td>-1280.84</td>
<td>26</td>
<td>1366.42</td>
</tr>
<tr>
<td>CWGH</td>
<td>-1280.76</td>
<td>28</td>
<td>1372.93</td>
</tr>
<tr>
<td>CLNH</td>
<td>-1280.84</td>
<td>28</td>
<td>1373.01</td>
</tr>
<tr>
<td>CLGH</td>
<td>-1284.08</td>
<td>31</td>
<td>1386.12</td>
</tr>
<tr>
<td>DWNH</td>
<td>-1292.17</td>
<td>26</td>
<td>1377.75</td>
</tr>
<tr>
<td>DWGH</td>
<td>-1290.64</td>
<td>28</td>
<td>1382.81</td>
</tr>
<tr>
<td>DLNH</td>
<td>-1291.97</td>
<td>28</td>
<td>1384.14</td>
</tr>
<tr>
<td>DLGH</td>
<td>-1291.67</td>
<td>31</td>
<td>1393.71</td>
</tr>
</tbody>
</table>

\[
\text{BIC} = - \ln(L_C) + 0.5 \, p \, \ln(N)
\]

- Weibull models have smaller BIC values which implies that they provide better model fits.
- Shape parameters of log-logistic baseline hazards were all greater than one which means they were all non-monotonic.
### Analysis on Different Parametric Hazard Models

- **Modeling Results**
- **Eight Models with alternative baseline hazard and gamma distributions for heterogeneity**

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Likelihood at Convergence</th>
<th>Number of Parameters</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWNH</td>
<td>-1280.84</td>
<td>26</td>
<td>1366.42</td>
</tr>
<tr>
<td>CWGH</td>
<td>-1280.76</td>
<td>28</td>
<td>1372.93</td>
</tr>
<tr>
<td>CLNH</td>
<td>-1280.84</td>
<td>28</td>
<td>1373.01</td>
</tr>
<tr>
<td>CLGH</td>
<td>-1284.08</td>
<td>31</td>
<td>1386.12</td>
</tr>
<tr>
<td>DWNH</td>
<td>-1292.17</td>
<td>26</td>
<td>1377.76</td>
</tr>
<tr>
<td>DWGH</td>
<td>-1290.64</td>
<td>28</td>
<td>1382.81</td>
</tr>
<tr>
<td>DLNH</td>
<td>-1291.97</td>
<td>28</td>
<td>1384.14</td>
</tr>
<tr>
<td>DLGH</td>
<td>-1291.67</td>
<td>31</td>
<td>1393.71</td>
</tr>
</tbody>
</table>

\[
\text{BIC} = - \ln(L_C) + 0.5 \ p \ \ln(N)
\]

- \( \ln(L_C) \) is the log-likelihood value at convergence
- \( p \) is the number of parameters
- \( N \) is the number of samples

Including unobserved heterogeneity does not necessarily improve the goodness-of-fit.
Introduction

There is no true competing dynamic vehicle transaction decision model in the literature. The existing competing models in other fields such as economics have considered only two competing events types which should be included in the model to have accurate results.
Competing Hazard Model

- **Assumptions**

  It is assumed that the transactions occur in discrete time intervals.

  Acquisition or dispose did not happen and may happen after this time.

  Household Traded a Car

  March

  April

  End of year

  Time
Competing Hazard Model

- **Methodology**

Likelihood function for modeling the competing transaction type and timing for the case that trade transaction has been observed.

\[
P_{Tra}^t = P[Tra = 1, Acq = 0 \text{ and } Dis = 0 \text{ at } (t - 1, t) =
\]

\[
\int_{\delta_{t-1}}^{\infty} \int_{\delta_{t}}^{\infty} \int_{\delta_{t}}^{\infty} f(\epsilon_{Tra}, \epsilon_{Acq}, \epsilon_{Dis}) d\epsilon_{Dis} d\epsilon_{Acq} d\epsilon_{Tra}
\]

A trade transaction was observed

\(\delta_{t}\) is the logarithm of the integrated baseline hazard of failure type \(i\)

Copula Function is used to replace this joint function

Marginal Distributions Copula Function (Gumbel in this case)

\[
f(\epsilon_{Dis}, \epsilon_{Acq}, \epsilon_{Tra}) = f(\epsilon_{Dis}) f(\epsilon_{Acq}) f(\epsilon_{Tra}) \times c_{\theta}(F(\epsilon_{Dis}), F(\epsilon_{Acq}), F(\epsilon_{Tra}))
\]
Competing Hazard Model

- Results and Findings
  - Explanatory Variables

Individual’s attributes
  - Age and gender of the car owner

Attributes of the DMU
  - Income (log), Housing Tenure, No. of Vehicles, Former members, New members, Workers, Adults, Youths, and Children, Education status

Attributes of the vehicle
  - Age, Price, MPG, Weight, New/Used

Vehicle and market interaction:
  - Avg depreciation cost, Avg parking cost, Avg fuel cost
Competing Hazard Model

- Results and Findings
  - General Model Comparison Results

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Likelihood at Convergence</th>
<th>Number of Parameters</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gumble Copula with Weibull Baseline</td>
<td>-4026.71</td>
<td>30</td>
<td>4126.70</td>
</tr>
<tr>
<td>Gumble Copula with Log-logistic Baseline</td>
<td>-4059.96</td>
<td>33</td>
<td>4169.94</td>
</tr>
<tr>
<td>No Copula With Weibull Baseline</td>
<td>-4486.34</td>
<td>29</td>
<td>4582.99</td>
</tr>
<tr>
<td>No Copula with Log-logistic Baseline</td>
<td>-4188.13</td>
<td>32</td>
<td>4294.78</td>
</tr>
</tbody>
</table>

Weibull model outperforms the log-logistic model.

with-copula models considerably dominate the without-copula models. If the interdependencies among the transaction types are included in the modeling formulation, it is expected that the model better explain the household and individual’s behaviors.

\[ \text{BIC} = - \ln(L_C) + 0.5 \ p \ \ln(N) \]

\( \ln(L_C) \) is the log-likelihood value at convergence
\( p \) is the number of parameters
\( N \) is the number of samples

Taha Hossein Rashidi  University of Illinois at Chicago
Competing Hazard Model

• Results and Findings
  – Simulation results

- The estimated model of this study was found to replicate **77%** of the trade decisions, **96%** of the acquisition decisions and **96%** of the disposal decisions correctly.

- It was found that the shortage in accurately predicting trade decisions has been connected to almost the same amount of redundancy in overestimating the total number of disposal decision.
Introduction

• Major household decisions

This study introduce a disaggregate dynamic model for major household decisions on durable products.

An ideal framework encompasses these decisions along with the specific components of each one of them in a joint structure. Nonetheless, here only the timing decisions are studied.
Introduction

- Framework

Job relocation decision is made at the individual level but it is influenced by other members' decisions.
Introduction

• Framework

Job relocation decisions and residential relocation decision are jointly modeled to incorporate the two-way impact between these decisions.
Introduction

• Framework
Model and Results

- Explanatory variables

**Individual’s attributes**
- Age

**Attributes of the household**
- Income (log), Housing Tenure, No. of Vehicles, former members, New members, Workers, Adults, Youths, and Children

**Economic Characteristics**
- **Gas Price Change** and **Unemployment Rate Change**

**Built Environment and Land-Use**
- Housing Units Density, Real Estate Jobs Density, Education Jobs Density, Total Job Density and Spatial Employment Population

**Activity Attributes**  
- Can be borrowed from an activity-based model
- Household Travel Time, Household Activity Time, Husband Travel Time and Husband Work Distance
Model and Results

- Baseline Hazard Analysis
- Weibull and log-logistic baseline hazard function for job relocation decisions

The log-logistic function shows a more rapidly increasing hazard rate for both the husband and wife during the first ten years followed by a decreasing rate for the husband and very little change for the wife after 10- to 15 years.

The Weibull model gives a steadily increasing rate in both the husband and wife hazard.

Therefore, both log-logistic and Weibull hazards give monotonically increasing patterns for the meaningful job relocation durations which is on average between three and four year in the case of the utilized data.
Model and Results

- Baseline Hazard Analysis
- Weibull and log-logistic baseline hazard function for residential relocation decisions

The BIC comparison between the log-logistic and Weibull baseline hazard functions prefers the log-logistic function for residential relocation decisions.

But the most prominent differences are in the first year.

However, for relocation durations greater than 1 year, the log-logistic hazard drops more rapidly than the Weibull hazard which suggests that household decision makers becoming increasingly resistant to change residential location over time.
Model and Results

- Baseline Hazard Analysis
- Weibull and log-logistic baseline hazard function for residential relocation decisions

Log-logistic baseline hazard provides a non-monotonic baseline hazard while the Weibull baseline hazard is monotonically increasing.

The log-logistic baseline hazard increases up to 2 years and then decreases which means people prefer not to make a transaction before two years and their willingness to make a transaction declines after the two year point.
## Model and Results

- Statistical analysis for difference scenarios

<table>
<thead>
<tr>
<th>Scenario ID</th>
<th>NumObs</th>
<th>NumHzPar</th>
<th>NumExpPar</th>
<th>LLConst</th>
<th>LLVal</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>757</td>
<td>4</td>
<td>39</td>
<td>-3074</td>
<td>-2887.84</td>
<td>3030.37</td>
</tr>
<tr>
<td>2</td>
<td>757</td>
<td>5</td>
<td>39</td>
<td>-3031</td>
<td>-2843.91</td>
<td>2989.76</td>
</tr>
<tr>
<td>3</td>
<td>757</td>
<td>6</td>
<td>39</td>
<td>-3069</td>
<td>-2887.60</td>
<td>3036.76</td>
</tr>
<tr>
<td>4</td>
<td>757</td>
<td>7</td>
<td>39</td>
<td>-3025</td>
<td>-2843.91</td>
<td>2996.39</td>
</tr>
<tr>
<td>5</td>
<td>757</td>
<td>5</td>
<td>39</td>
<td>-2985</td>
<td>-2861.86</td>
<td>3007.71</td>
</tr>
<tr>
<td>6</td>
<td>757</td>
<td>6</td>
<td>39</td>
<td>-2942</td>
<td>-2817.33</td>
<td>2966.49</td>
</tr>
<tr>
<td>7</td>
<td>757</td>
<td>7</td>
<td>39</td>
<td>-2980</td>
<td>-2861.50</td>
<td>3013.98</td>
</tr>
<tr>
<td>8</td>
<td>757</td>
<td>8</td>
<td>39</td>
<td>-2937</td>
<td>-2817.43</td>
<td>2973.22</td>
</tr>
</tbody>
</table>

**Notes:**
- **NumObs**: Number of Observations
- **NumHzPar**: Number of Hazard Function Parameters
- **NumExpPar**: Number of Parameters for Explanatory Variables
- **LLConst**: Likelihood With Only Constant
- **LLVal**: Likelihood at Convergence
- **BIC**: Bayesian Information Criterion
Introduction

- Study Framework

- Choice Set Formation
  - A manageable set of alternatives is selected from the universal choice set

- Selection Probability
  - Probability of being selected in the choice is calculated which is used for sampling correction purpose.

- Actual Choice Selection
  - The most desired alternative in the choice set of alternatives is selected

Household members form their choice sets by screening available alternatives and filtering them based on their priorities, and preferences.
Introduction

- Study Framework

Choice Set Formation
- A manageable set of alternatives is selected from the universal choice set

Selection Probability
- Probability of being selected in the choice is calculated which is used for sampling correction purpose.

Actual Choice Selection
- The most desired alternative in the choice set of alternatives is selected

Using the selection probabilities the sampling bias is adjusted. Then by using a discrete choice model the final residential location is selected.
Choice Set Formation

- Evaluation of the choice set formation models

<table>
<thead>
<tr>
<th>Random Draws</th>
<th>Truly Included Final Decision (1)</th>
<th>Average Choice Set Size (2)</th>
<th>Predictive Ability (1)/693 (%)</th>
<th>Set Size (2)/741 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>94</td>
<td>23</td>
<td>13.56%</td>
<td>3.10%</td>
</tr>
<tr>
<td>50</td>
<td>167</td>
<td>43</td>
<td>24.10%</td>
<td>5.80%</td>
</tr>
<tr>
<td>100</td>
<td>241</td>
<td>77</td>
<td>34.78%</td>
<td>10.39%</td>
</tr>
<tr>
<td>200</td>
<td>367</td>
<td>128</td>
<td>52.96%</td>
<td>17.27%</td>
</tr>
<tr>
<td>300</td>
<td>424</td>
<td>165</td>
<td>61.18%</td>
<td>22.27%</td>
</tr>
<tr>
<td>400</td>
<td>446</td>
<td>195</td>
<td>64.36%</td>
<td>26.32%</td>
</tr>
<tr>
<td>500</td>
<td>506</td>
<td>219</td>
<td>73.02%</td>
<td>29.55%</td>
</tr>
<tr>
<td>600</td>
<td>518</td>
<td>239</td>
<td>74.75%</td>
<td>32.25%</td>
</tr>
<tr>
<td>700</td>
<td>524</td>
<td>255</td>
<td>75.61%</td>
<td>34.41%</td>
</tr>
</tbody>
</table>

There are two important factors in evaluating a choice set generator algorithm: the predictive ability of the algorithm and size of the generated choice sets. Unfortunately, these two factors are negatively correlated.
Sample Selection Probability

- Probability Calculation

Out of the 824 Transportation Analysis Zones (TAZ) in the Seattle Metropolitan Area, 741 of them are included in the universal choice set available to the households from which they select their residential locations.

\[
f_i(\text{wd}) = \left[ \gamma \ \text{wd}^{-1} \ \exp(-\theta_{x} \ X_i) \right] \times \left[ e^{-\text{wd} \ \exp(-\theta_{x} \ X_i)} \right]
\]

It is assumed that depending on the individual household’s attributes, decision makers have some threshold for the maximum commute distance beyond which housing alternatives will not be attractive to the household.

In such cases the household will reject any alternative with the distance surpassing the threshold defined for the household.

The average distance to the household employed members' work locations

Using the desired work distance and actual work distance the probability of being selected is calculated for each alternative.
Actual Choice Selection

• Methodology

A discrete choice model is employed for residential location selection in which a sampling selection correction factor is included.

This sampling correction factor works like the LOGSUM variable of nested logit models and it accounts for the impact of choice set formation method into the discrete choice model.

\[ P_{ij} = \frac{e^{\mu V_{ij} - \ln C_{ij}}}{L \sum_{l=1}^{L} e^{\mu V_{il} - \ln C_{il}}} \]

Where

\[ C_{ij} = \frac{q_{ij}}{\sum_{k=1}^{K} q_{ik}} \]

Roughly speaking, \( q_{ij} \) represents exponential of subtraction between the most desired work distance and the alternative residential location distance to the household employed members’ work locations (actual work distance).

Where \( \mu \) is a scale parameter and \( V_{ij} \) is the deterministic utility, \( K \) is total number of alternatives (741) and \( L \) is the total number of alternatives in the choice subset. The \( C_{ij} \) alternative specific term corrects for sampling bias.
Actual Choice Selection

- **Explanatory Variables**

  Land use and built-environment variables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Name</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of total number of jobs*</td>
<td>Jobs</td>
<td>4.20</td>
</tr>
<tr>
<td>Log of total number of real estate, rental and leasing jobs**</td>
<td>Real</td>
<td>0.38</td>
</tr>
<tr>
<td>Log of total number of finance and insurance jobs**</td>
<td>Fina</td>
<td>0.43</td>
</tr>
<tr>
<td>Log of number of residential housing units</td>
<td>Unit</td>
<td>2.52</td>
</tr>
<tr>
<td>Log of Industrial square feet**</td>
<td>Indsqf</td>
<td>4.94</td>
</tr>
<tr>
<td>Log of manufacturing jobs-Neighbors**</td>
<td>Manu_N</td>
<td>3.04</td>
</tr>
<tr>
<td>Log of utility jobs-Neighbors**</td>
<td>Util_N</td>
<td>0.31</td>
</tr>
<tr>
<td>Log of total number of finance and insurance jobs-Neighbors**</td>
<td>Fina_N</td>
<td>2.65</td>
</tr>
<tr>
<td>Log of government square feet-Neighbors**</td>
<td>Govsqft_N</td>
<td>10.49</td>
</tr>
<tr>
<td>Log of number of children (&lt;16)/Area***</td>
<td>Child</td>
<td>6.14</td>
</tr>
<tr>
<td>Log of number of middle age (&lt;44 and &gt;35)/Area***</td>
<td>Midage</td>
<td>6.20</td>
</tr>
<tr>
<td>Log of number of seniors (&lt;75 and &gt;64)/Area***</td>
<td>Senior</td>
<td>5.15</td>
</tr>
</tbody>
</table>

* 450 meters by 450 meters gridcells

** 750 meters by 750 meters gridcells

*** TAZ

“Log” values are used instead of the actual value to address the arbitrary boundary issues that may happen in spatial locations search models.

The first three variables in the table represent the employment densities in the area.

The next two variables relate to the residential and industrial land use in the zones.
Actual Choice Selection

- **Explanatory Variables**

Land use and built-environment Variables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Name</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of total number of jobs*</td>
<td>Jobs</td>
<td>4.20</td>
</tr>
<tr>
<td>Log of total number of real estate, rental and leasing jobs**</td>
<td>Real</td>
<td>0.38</td>
</tr>
<tr>
<td>Log of total number of finance and insurance jobs**</td>
<td>Fina</td>
<td>0.43</td>
</tr>
<tr>
<td>Log of number of residential housing units</td>
<td>Unit</td>
<td>2.52</td>
</tr>
<tr>
<td>Log of Industrial square feet**</td>
<td>Indsqf</td>
<td>4.94</td>
</tr>
<tr>
<td>Log of manufacturing jobs-Neighbors**</td>
<td>Manu_N</td>
<td>3.04</td>
</tr>
<tr>
<td>Log of utility jobs-Neighbors**</td>
<td>Util_N</td>
<td>0.31</td>
</tr>
<tr>
<td>Log of total number of finance and insurance jobs-Neighbors**</td>
<td>Fina_N</td>
<td>2.65</td>
</tr>
<tr>
<td>Log of government square feet-Neighbors**</td>
<td>Govsqft_N</td>
<td>10.49</td>
</tr>
<tr>
<td>Log of number of children (&lt;16)/Area***</td>
<td>Child</td>
<td>6.14</td>
</tr>
<tr>
<td>Log of number of middle age (&lt;44 and &gt;35)/Area***</td>
<td>Midage</td>
<td>6.20</td>
</tr>
<tr>
<td>Log of number of seniors (&lt;75 and &gt;64)/Area***</td>
<td>Senior</td>
<td>5.15</td>
</tr>
</tbody>
</table>

* 450 meters by 450 meters gridcells
** 750 meters by 750 meters gridcells
*** TAZ

The next four explanatory variables are included in the model to account for spatial dependency between contiguous zones.

These four variables represent the land use conditions in the zones surrounding the zone under consideration.

“Log” values are used instead of the actual value to address the arbitrary boundary issues that may happen in spatial locations search models.
Actual Choice Selection

- **Explanatory Variables**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Name</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of total number of jobs*</td>
<td>Jobs</td>
<td>4.20</td>
</tr>
<tr>
<td>Log of total number of real estate, rental and leasing jobs**</td>
<td>Real</td>
<td>0.38</td>
</tr>
<tr>
<td>Log of total number of finance and insurance jobs**</td>
<td>Fina</td>
<td>0.43</td>
</tr>
<tr>
<td>Log of number of residential housing units</td>
<td>Unit</td>
<td>2.52</td>
</tr>
<tr>
<td>Log of Industrial square feet**</td>
<td>Indsqf</td>
<td>4.94</td>
</tr>
<tr>
<td>Log of manufacturing jobs-Neighbors**</td>
<td>Manu_N</td>
<td>3.04</td>
</tr>
<tr>
<td>Log of utility jobs-Neighbors**</td>
<td>Util_N</td>
<td>0.31</td>
</tr>
<tr>
<td>Log of total number of finance and insurance jobs-Neighbor**</td>
<td>Fina_N</td>
<td>2.65</td>
</tr>
<tr>
<td>Log of government square feet-Neighbors**</td>
<td>Govsqft_N</td>
<td>10.49</td>
</tr>
<tr>
<td>Log of number of children (&lt;16)/Area***</td>
<td>Child</td>
<td>6.14</td>
</tr>
<tr>
<td>Log of number of middle age (&lt;44 and &gt;35)/Area***</td>
<td>Midage</td>
<td>6.20</td>
</tr>
<tr>
<td>Log of number of seniors (&lt;75 and &gt;64)/Area***</td>
<td>Senior</td>
<td>5.15</td>
</tr>
</tbody>
</table>

* 450 meters by 450 meters gridcells
** 750 meters by 750 meters gridcells
*** TAZ

“Log” values are used instead of the actual value to address the arbitrary boundary issues that may happen in spatial locations search models.

Population density was also included in the model, as was density of children and seniors in a TAZ, which can imply whether a TAZ is family oriented or not.
Actual Choice Selection

- **Explanatory Variables**

  **Monetary-Related Variables**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Name</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute difference between average zonal income and HHId income (X$100,000) ***</td>
<td>DiffInc</td>
<td>0.23</td>
</tr>
<tr>
<td>Poor X (Log of absolute difference between average zonal land value and the average land value of the zone in which HHId lives) ***</td>
<td>PoorLandVal</td>
<td>1.44</td>
</tr>
<tr>
<td>Middle X Log of absolute difference between average zonal land value and the average land value of the zone in which HHId lives ***</td>
<td>MiddleLandVal</td>
<td>7.68</td>
</tr>
<tr>
<td>Rich X Log of absolute difference between average zonal land value and the average land value of the zone in which HHId lives ***</td>
<td>RichLandVal</td>
<td>2.09</td>
</tr>
<tr>
<td>Transit percentage usage X binary variable for decrease in gas price ***</td>
<td>TransitDec</td>
<td>0.07</td>
</tr>
</tbody>
</table>

* 450 meters by 450 meters gridcells
** 750 meters by 750 meters gridcells
*** TAZ

Households look for zones which are more similar to their socio-demographic attributes.
## Actual Choice Selection

### Explanatory Variables

#### Monetary-Related Variables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Name</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute difference between average zonal income and HHld income (X100,000) ***</td>
<td>DiffInc</td>
<td>0.23</td>
</tr>
<tr>
<td>Poor X (Log of absolute difference between average zonal land value and the average land value of the zone in which HHld lives) ***</td>
<td>PoorLandVal</td>
<td>1.44</td>
</tr>
<tr>
<td>Middle X Log of absolute difference between average zonal land value and the average land value of the zone in which HHld lives ***</td>
<td>MiddleLandVal</td>
<td>7.68</td>
</tr>
<tr>
<td>Rich X Log of absolute difference between average zonal land value and the average land value of the zone in which HHld lives ***</td>
<td>RichLandVal</td>
<td>2.09</td>
</tr>
<tr>
<td>Transit percentage usage X binary variable for decrease in gas price ***</td>
<td>TransitDec</td>
<td>0.07</td>
</tr>
</tbody>
</table>

* 450 meters by 450 meters gridcells
** 750 meters by 750 meters gridcells
*** TAZ

Households with less than 25,000 annual income are called poor, household with annual income greater than 75,000 are called rich and others are called middle.

The land value for each zone is not used directly in the model, however, it is transformed into the log of the absolute value of the difference in land value for each TAZ from the current residential location.
Actual Choice Selection

- Explanatory Variables

<table>
<thead>
<tr>
<th>Monetary-Related Variables</th>
<th>Parameter</th>
<th>Name</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absolute difference between average zonal income and HHld income(X100,000)***</td>
<td>DiffInc</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Poor X (Log of absolute difference between average zonal land value and the average land value of the zone in which HHld lives)***</td>
<td>PoorLandVal</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>Middle X Log of absolute difference between average zonal land value and the average land value of the zone in which HHld lives***</td>
<td>MiddleLandVal</td>
<td>7.68</td>
</tr>
<tr>
<td></td>
<td>Rich X Log of absolute difference between average zonal land value and the average land value of the zone in which HHld lives***</td>
<td>RichLandVal</td>
<td>2.09</td>
</tr>
<tr>
<td></td>
<td>Transit percentage usage X binary variable for decrease in gas price***</td>
<td>TransitDec</td>
<td>0.07</td>
</tr>
</tbody>
</table>

* 450 meters by 450 meters gridcells
** 750 meters by 750 meters gridcells
*** TAZ

The percentage of transit users in a zone is interacted with a variable which indicates a decrease in the gas price (in real terms) between waves, on the thought that transit oriented areas may be less attractive in an environment of declining fuel prices.
Actual Choice Selection

- **Results and Findings**
- **Interpretation of some of the variables**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimation</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoorLandVal</td>
<td>-0.268</td>
<td>-7.25</td>
</tr>
<tr>
<td>Middle LandVal</td>
<td>-0.2923</td>
<td>-19.25</td>
</tr>
<tr>
<td>RichLandVal</td>
<td>-0.2968</td>
<td>-10.58</td>
</tr>
<tr>
<td>Correction Factor</td>
<td>0.3259</td>
<td>11.44</td>
</tr>
<tr>
<td>DiffInc</td>
<td>-0.8824</td>
<td>-2.2</td>
</tr>
<tr>
<td>Child</td>
<td>0.2621</td>
<td>2.61</td>
</tr>
<tr>
<td>Midage</td>
<td>-0.2246</td>
<td>-1.84</td>
</tr>
<tr>
<td>Senior</td>
<td>-0.1574</td>
<td>-2.72</td>
</tr>
<tr>
<td>TransitDec</td>
<td>-2.5705</td>
<td>-2.45</td>
</tr>
<tr>
<td>Log_Likelihood at Convergence</td>
<td>-2370</td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>592.95</td>
<td></td>
</tr>
</tbody>
</table>

It can be interpreted from the **negative sign** of LandVal parameters that zones with greater difference from the land value of the current residential zone become **less attractive** to the household since they become either less affordable or too affordable, i.e. **not having the desired amenities, quality, etc. the household is accustomed to.**
### Actual Choice Selection

- **Results and Findings**
- **Interpretation of some of the variables**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimation</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoorLandVal</td>
<td>-0.268</td>
<td>-7.25</td>
</tr>
<tr>
<td>Middle LandVal</td>
<td>-0.2923</td>
<td>-19.25</td>
</tr>
<tr>
<td>RichLandVal</td>
<td>-0.2968</td>
<td>-10.58</td>
</tr>
<tr>
<td>Correction Factor</td>
<td>0.3259</td>
<td>11.44</td>
</tr>
<tr>
<td>DiffInc</td>
<td>-0.8824</td>
<td>-2.2</td>
</tr>
<tr>
<td>Child</td>
<td>0.2621</td>
<td>2.61</td>
</tr>
<tr>
<td>Midage</td>
<td>-0.2246</td>
<td>-1.84</td>
</tr>
<tr>
<td>Senior</td>
<td>-0.1574</td>
<td>-2.72</td>
</tr>
<tr>
<td>TransitDec</td>
<td>-2.5705</td>
<td>-2.45</td>
</tr>
<tr>
<td>Log_Likelihood at Convergence</td>
<td>-2370</td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>592.95</td>
<td></td>
</tr>
</tbody>
</table>

Analyzing different sample sizes showed no more than 42% difference on average between the presented results in table.

But if a random sample is drawn for each household the parameter estimations are at least 300% of what is presented in table.

Therefore, it can be concluded that the utilized correction factor can give consistent parameter estimates.
Actual Choice Selection

- Results and Findings
- Interpretation of some of the variables

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimation</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoorLandVal</td>
<td>-0.268</td>
<td>-7.25</td>
</tr>
<tr>
<td>Middle LandVal</td>
<td>-0.2923</td>
<td>-19.25</td>
</tr>
<tr>
<td>RichLandVal</td>
<td>-0.2968</td>
<td>-10.58</td>
</tr>
<tr>
<td>Correction Factor</td>
<td>0.3259</td>
<td>11.44</td>
</tr>
<tr>
<td>DiffInc</td>
<td>-0.8824</td>
<td>-2.2</td>
</tr>
<tr>
<td>Child</td>
<td>0.2621</td>
<td>2.61</td>
</tr>
<tr>
<td>Midage</td>
<td>-0.2246</td>
<td>-1.84</td>
</tr>
<tr>
<td>Senior</td>
<td>-0.1574</td>
<td>-2.72</td>
</tr>
<tr>
<td>TransitDec</td>
<td>-2.5705</td>
<td>-2.45</td>
</tr>
</tbody>
</table>

The findings of this study also confirm the intuitive result that zones with greater differences in income relative to the households’ income are less attractive.
Actual Choice Selection

- **Results and Findings**
- **Interpretation of some of the variables**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimation</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoorLandVal</td>
<td>-0.268</td>
<td>-7.25</td>
</tr>
<tr>
<td>Middle LandVal</td>
<td>-0.2923</td>
<td>-19.25</td>
</tr>
<tr>
<td>RichLandVal</td>
<td>-0.2968</td>
<td>-10.58</td>
</tr>
<tr>
<td>Correction Factor</td>
<td>0.3259</td>
<td>11.44</td>
</tr>
<tr>
<td>DiffInc</td>
<td>-0.8824</td>
<td>-2.2</td>
</tr>
<tr>
<td>Child</td>
<td>0.2621</td>
<td>2.61</td>
</tr>
<tr>
<td>Midage</td>
<td>-0.2246</td>
<td>-1.84</td>
</tr>
<tr>
<td>Senior</td>
<td>-0.1574</td>
<td>-2.72</td>
</tr>
<tr>
<td>TransitDec</td>
<td>-2.5705</td>
<td>-2.45</td>
</tr>
</tbody>
</table>

As fuel costs become *less* of an issue to families, they appear to *stop* focusing as much on transit-oriented zones.

Taha Hossein Rashidi  
University of Illinois at Chicago
Summary and Conclusion

• Hazard-Based Duration Models

Contributions

1- Several specifications of hazard-based models were discussed and analyzed.

2- A competing hazard model for three vehicle transaction types was introduced.

Findings

If accuracy of the models is a concern, it is recommended that for each case study, different specifications of the hazard-based are tested because there is not a general and universal rule of the superiority of one specification over others.

For example,

- there is no guarantee that including unobserved heterogeneity improves the modeling results,

- or there is non-monotonic baseline hazards do not necessarily outperform the monotonic formulations.
Summary and Conclusion

• Hazard-Based Duration Models

Contributions

1- Several specifications of hazard-based models were discussed and analyzed.

2- A competing hazard model for three vehicle transaction types was introduced.

Findings

It was found that the competing hazard model considerably outperform the non-competing model.

Through a simulation analysis, it was shown that the presented formulation appropriately captures the competing behavior between the transaction types.
Summary and Conclusion

- Residential Location Search

Contributions

1- The interdependencies among the major household decisions was modeled.

2- A behavioral housing search model was introduced.

Findings

It was shown that job and residential relocation causality is a two way relationship where job relocation impedes residential relocation while residential relocation can trigger job relocation for both wife and husband.

Additionally, it was found that job relocation does not accelerate a household vehicle transaction decision. Furthermore, change in household residential location has a positive effect on the transaction decision.
Summary and Conclusion

• Residential Location Search

Contributions

1. The interdependencies among the major household decision was modeled.

2. A behavioral two-stage housing search model was introduced.

Findings

An interesting advantage of the choice set formation method was that, most of the time; it includes the final decision of the decision maker in the choice set.

Sampling bias resulted from selecting a subset of alternatives in the discrete choice model was eliminated by including a sampling correction factor. It was found that the sampling correction factor efficiently results in consistent parameter estimations.
Future Directions

• Some of the major improvements to the presented models

• Further analysis on non-parametric hazard models
  This dissertation only studied parametric hazard models

• Improving the competing hazard model with considering other types of copula functions
  This dissertation only considered Gumbel copula

• Studying the impact of other major household decisions such as school location search
  Other than residential relocation, job relocation and vehicle transaction timing decisions

• Including the error correlation among the alternatives in the housing search model
Thank you