Latent Population Segments with Homogenous Residential and Work Location Choice Preferences

Mishra S., Golias M., Paleti, R., Sarker, A., and Balan, L

Presentation at TNMUG
April 28, 2016

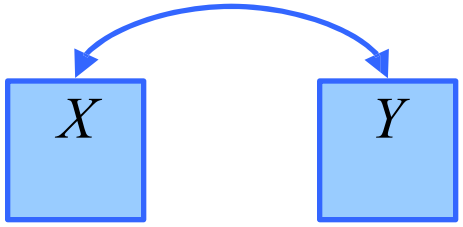
Outline

- Introduction
- Overview of Latent Class (LC) Models and Transportation Applications
- Project Framework and Data
- Preliminary Results
- Next Steps

Project Overview

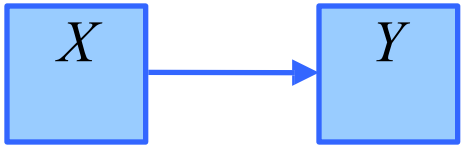
- One year project (Jan 2016 ~Dec 2016)
- Sponsored by TDOT
- Project Manager
  - Chin-Cheng Chen

Causation versus latent

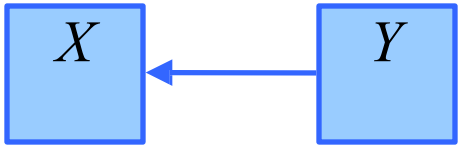


We observe a correlation between two variables. Why?

1. X causes Y ?

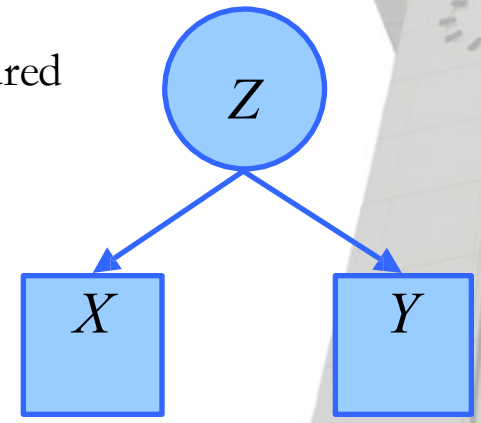


2. Y causes X ?



3. Reciprocal causation ($X \leftrightarrow Y$) ?

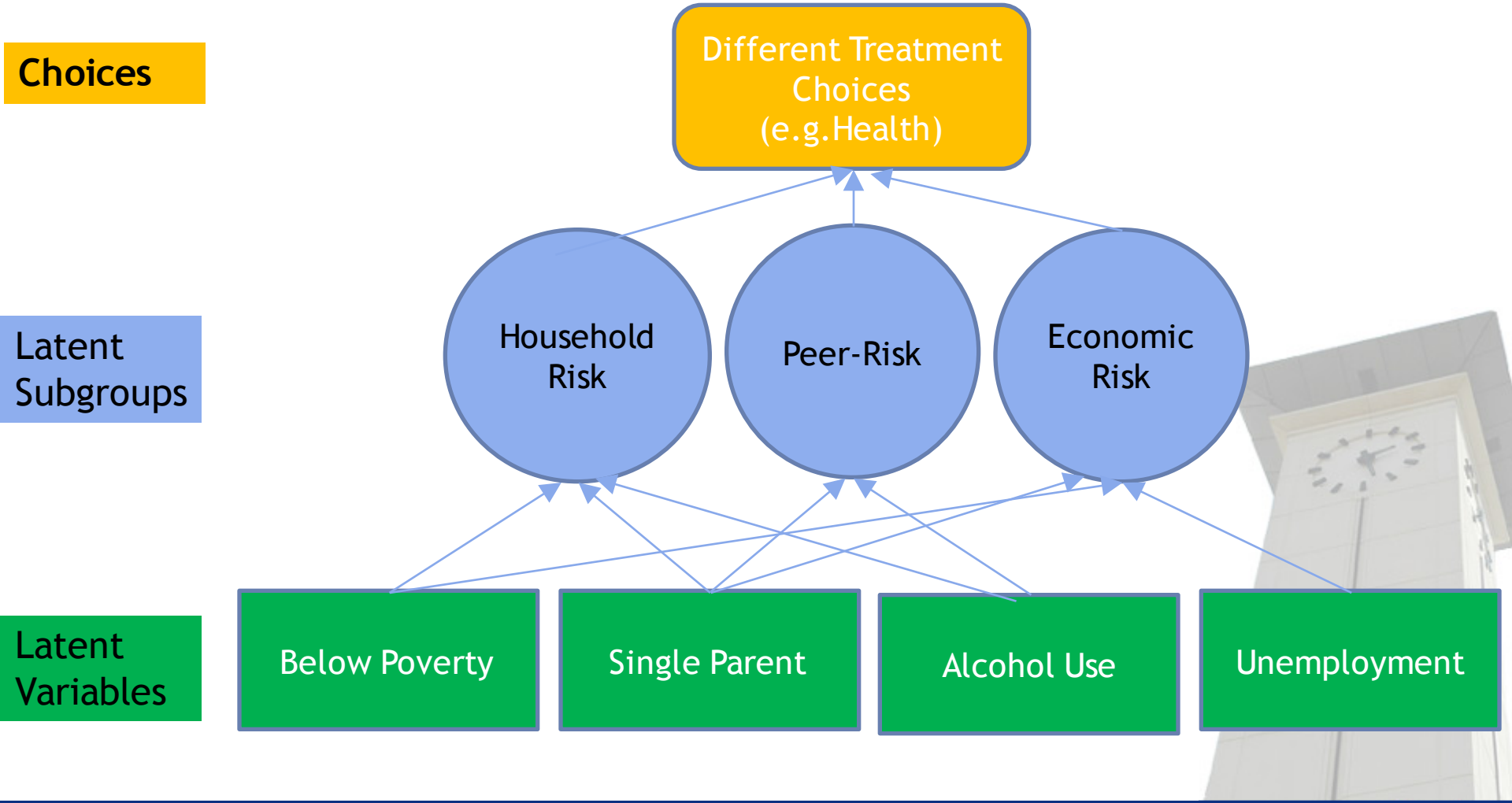
4. A third, unmeasured cause ?



Latent Class Model Overview (1)

- First introduced in social science
  - To capture the effects of social factors
    - Economic development
    - Racial prejudice
    - Religious commitment
- Variable indicating **underlying subgroups of individuals**
  - "Latent" because factors can not be directly observed or measured

Latent Class Model Overview (2)



Use of Latent Class (1)

- Extremely popular in marketing research
  - Market segmentation
  - Market structuring

Source: http://www.wisegeek.org/what-is-market-segmentation.htm

Use of Latent Class (2)

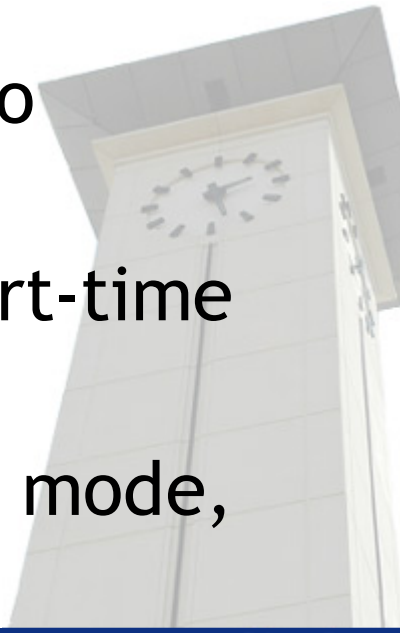
- Behavioral and health sciences
  - Identify differential treatment effects
  - By creating small subgroups based on e.g.
    - Age groups, alcohol use, household poverty, smoking behavior etc.
- Economics and Geography
  - Attitudinal measures of motivations
  - Heterogeneous preferences

Transportation Application (1)

- Personal attitudes and preferences are not observed in travel surveys
  - Greener life styles
  - Tech-savvy attitude

Transportation Application (2)

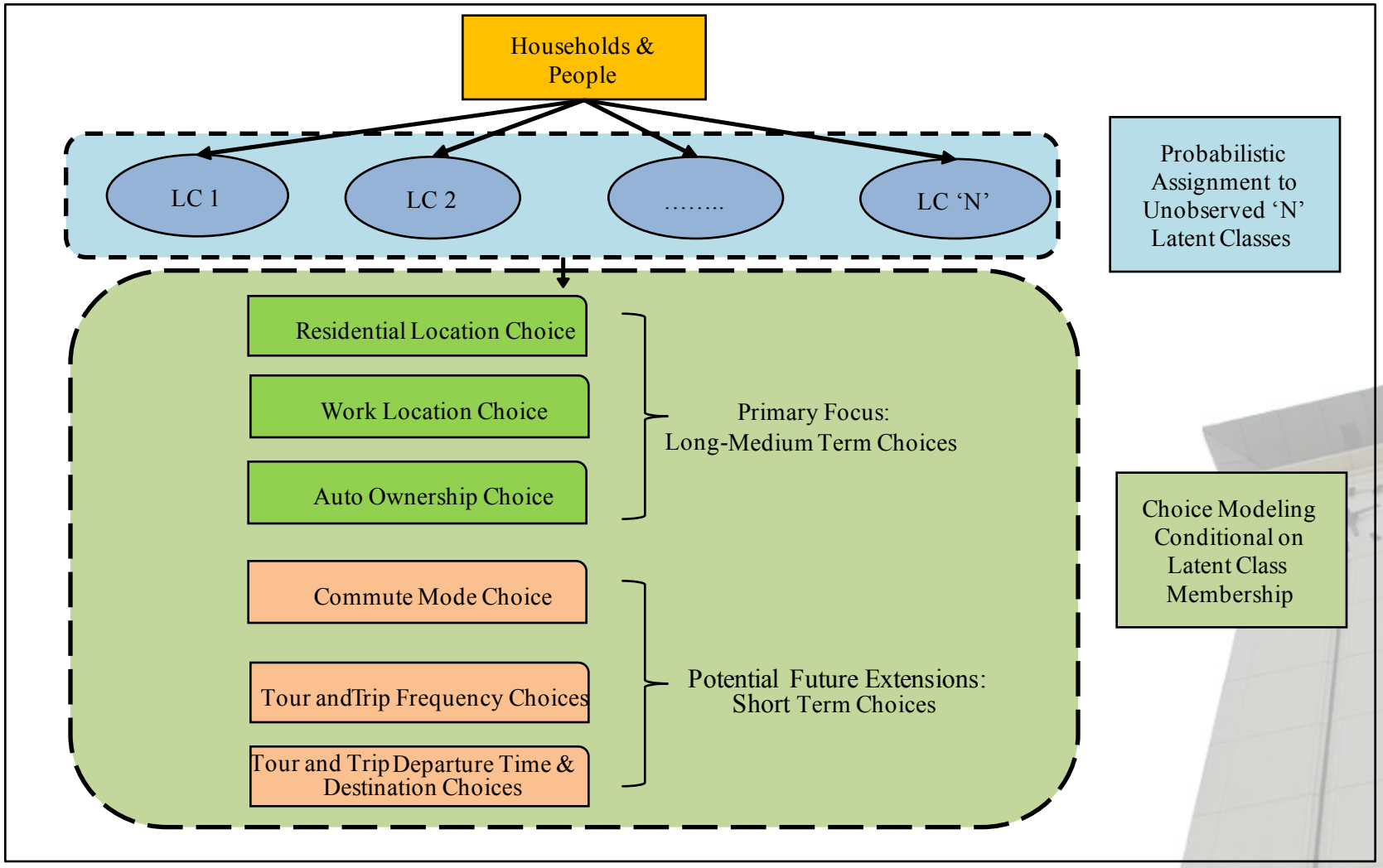
- Quantify effect of household & person factors
  - Joint household-workplace location choices
  - Travel behavior (commute mode, activity frequency etc.)
- Latent segmentation based on
  - Socio-demographic factors (income, auto ownership, educational level)
  - Work arrangements (work flexibility, part-time status, telecommuting option, industry)
  - Typical Travel Behavior (usual commute mode, non-motorized travel)



Research Objective

- Extensive synthesis of current and past literature on
  - Methodological aspects of latent class models
  - Applications in the context of transportation and land-use modeling
- Develop Alternate Behavioral Paradigms for modeling residential and work location choices
- Perform extensive post model estimation
- Demonstrate applicability by forecasting

Methodological Framework



Work Location Model in TN: Data

- Nashville Household Travel Survey data
- 5,164 households with 11,114 people
- 5,682 employed people

Data Assembly Steps

- Append Distance and Logsum information for each [Home TAZ, Sampled TAZ] zonal pair

- Append zonal employment information of the industry in which the person is employed for all the sampled alternatives
  - For a person employed in manufacturing industry, only manufacturing zonal employment must be used

- Append household and person explanatory variables to the estimation data set

Household Level Explanatory Variables

- Household Level Variables
  - Household Income
  - Housing Tenure (Own *versus* Rent)
  - Presence of Children
  - Household Auto Ownership
  - Highest Educational Attainment

Person Level Explanatory Variables

- Work Industry
- Work hours (part-time *versus* full-time)
- Work Flexibility
- Educational Attainment
- Gender
- Age
- License
- Student Status

Transportation Network Level Explanatory Variables

- Network Distance
  - Linear
  - Squared
  - Cubic
  - Logarithmic

- Mode Choice Logsum

Explanatory Variables Mode Choice Logsum

- Work Location Choice is a long term choice
- Choices also influenced by network accessibility conditions
  - Areas with better transit connectivity and less auto congestion may be preferred
- Actual travel conditions experienced depends on mode choice (short term decision)
- Expected utility or logsum (mode choice): Location choice models explanatory variable

Mode Choice Logsum

For each zonal pair, utility of the 10 modes was computed using the following coefficients:

Variable	Beta - Specific to Choice Alternatives									
	Auto	WCR	DCR	WUR	DUR	WEB	DEB	WLB	DLB	Walk/ Bike
In-vehicle Travel Time	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	
In-vehicle Travel Time=0		-999	-999	-999	-999	-999	-999	-999	-999	
Transit Fare		-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	
Distance	-0.015									-1.5
Distance>3 miles										-999
Initial Wait Time		-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	
Access Walk Time		-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	
Access Drive Time			-0.075		-0.075		-0.075		-0.075	
Number of Transfers		-0.03								
Transfer Wait Time		-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	
Transfer Walk Time		-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	
Egress Walk Time		-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	

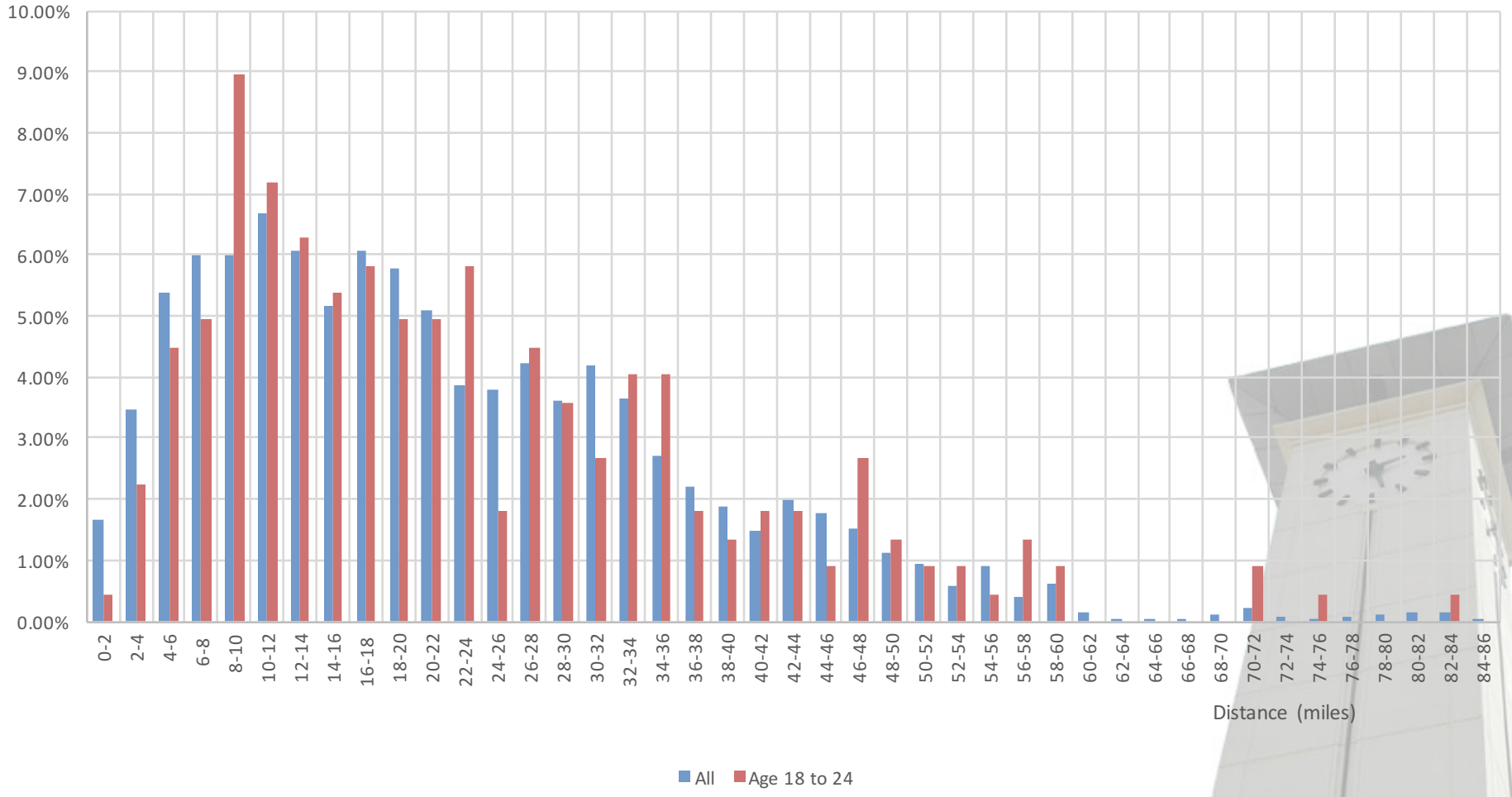
Final Estimation Dataset

- 3,635 workers

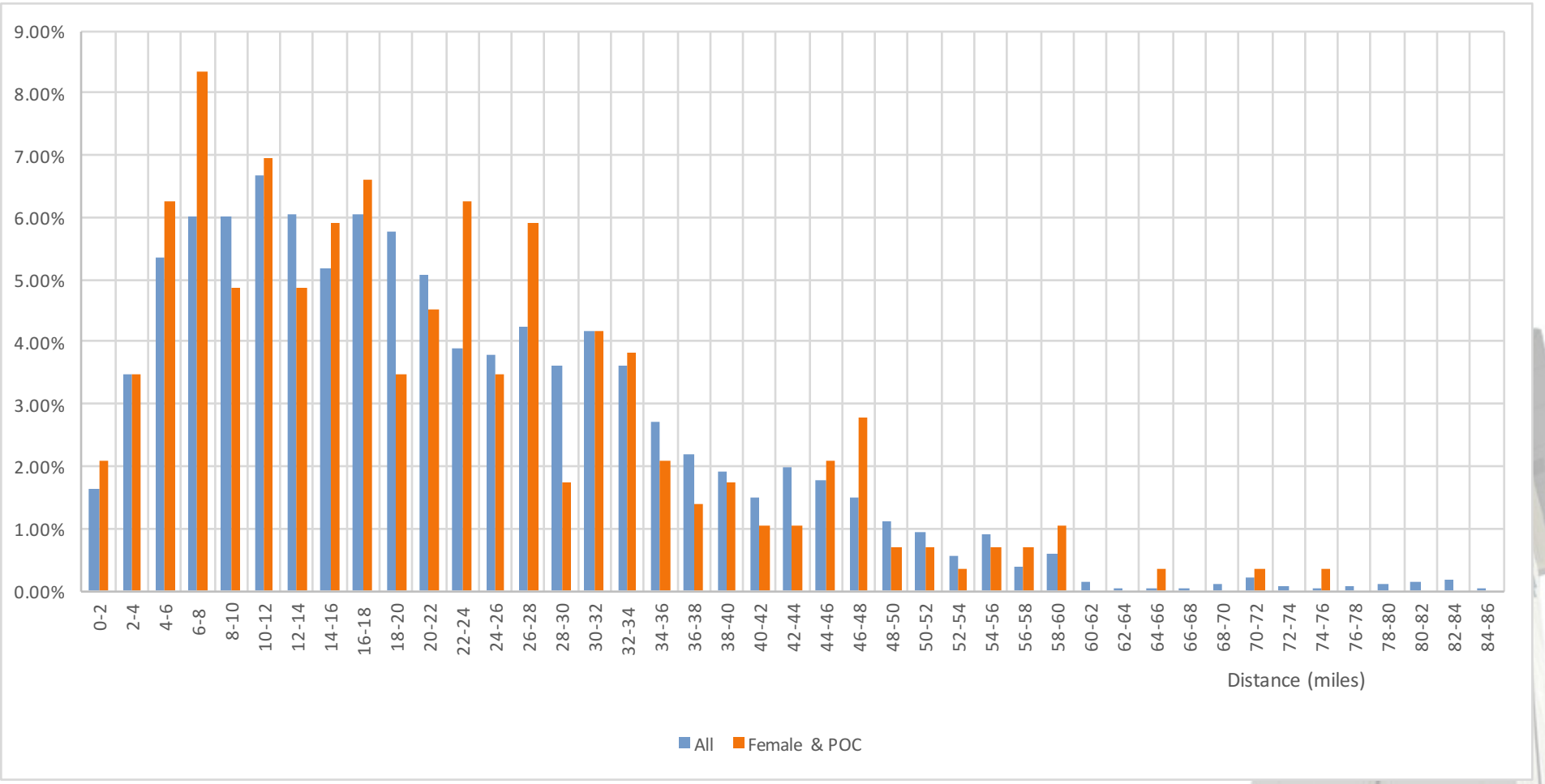
- 50 alternatives per worker

Industry	Frequency	Percentage
Agriculture	459	12.6
Manufacturing	133	3.7
Transportation	160	4.4
Retail	455	12.5
Office	2,428	66.5
Total	**3,635**	**100**

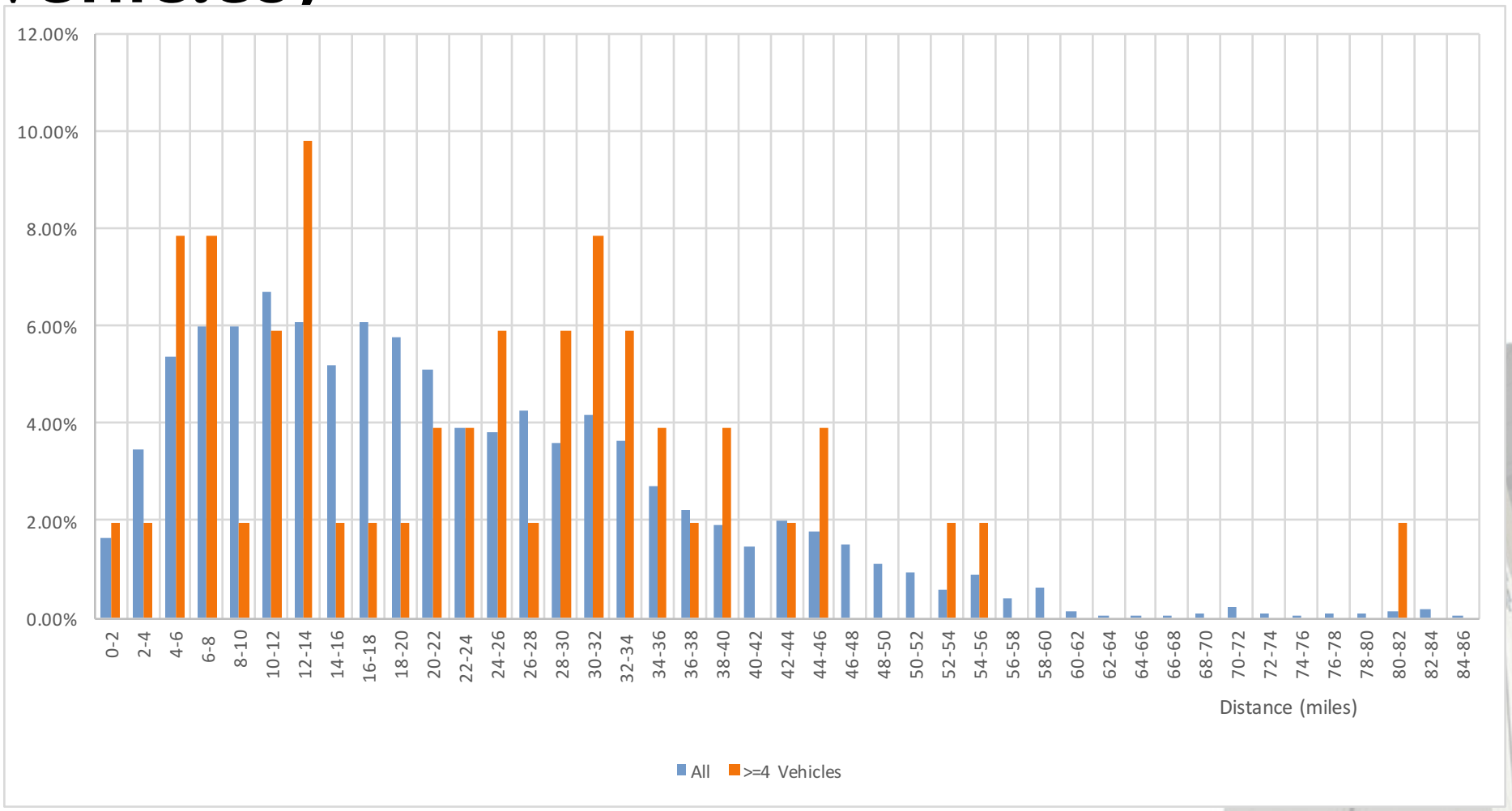
Commute Distance: Age 18 to 24 years



Commute Distance: Female and Presence of Young Child (0 to 5 years)



Commute Distance: Auto Ownership (>=4 vehicles)



Location Choice Model Structure

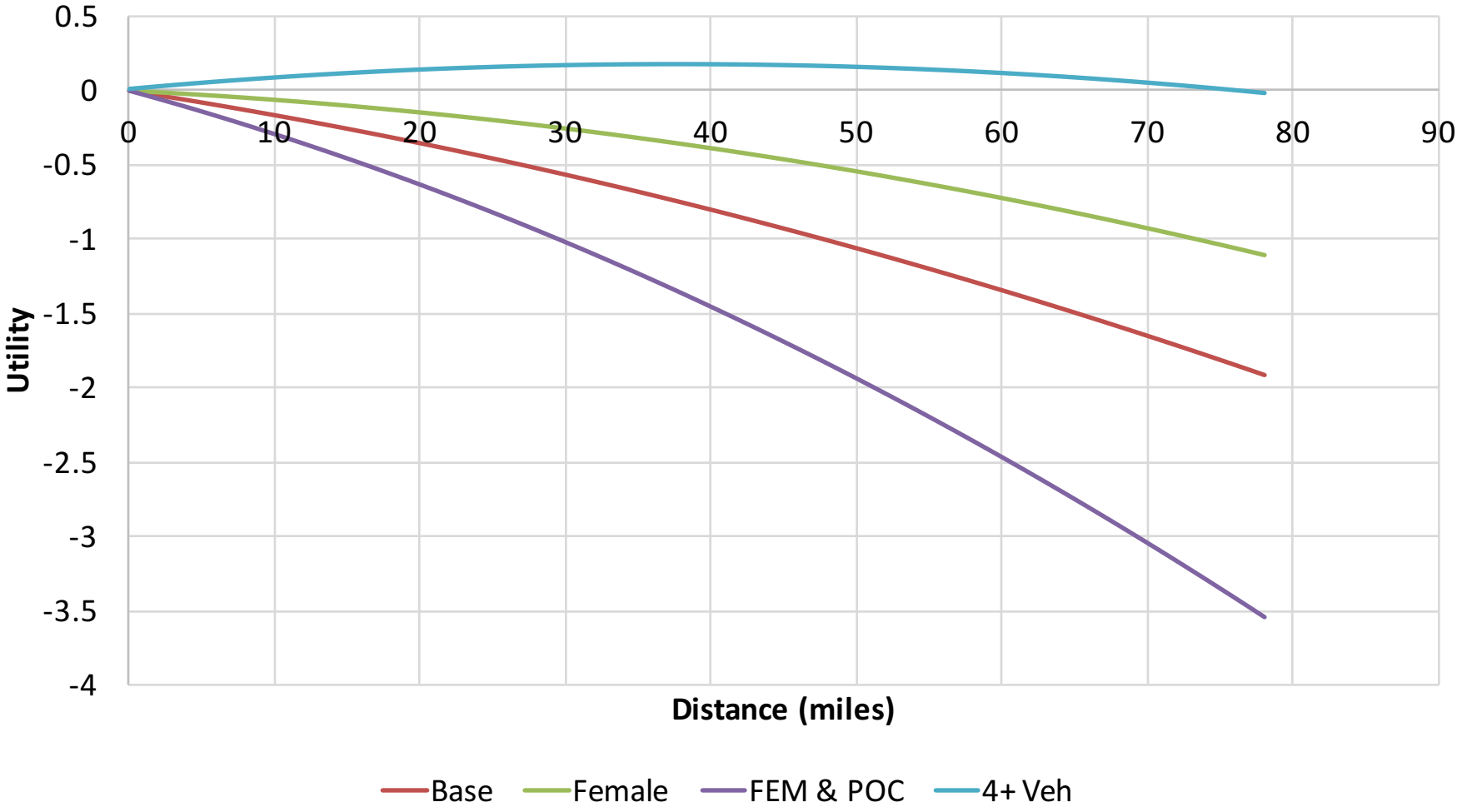
- Unlabeled MNL model
  - No alternate specific constants
- *i → Index of origin zone (home zone)*
- *J → Index of destination zones (work zone alternative)*
- *q → Index of the worker*
- *S → Size Variable (log of zonal employment in corresponding industry)*
- *LS → Logsum*
- *D → Distance between origin and destination zones*
  - *Utility can include several linear and non-linear distance effects*
- *X → Worker and household characteristics*

$$U_{i,j,q} = S_j + \alpha \times LS_{i,j} + \sum_k D_{i,j}^k + \sum_k \beta^k D_{i,j}^k X_q$$

Work Location Choice Model

Variables	Parameter	SE	T-Stat
Log(Zonal Employment)- (Size Variable)	1.0000 (fixed)	.	.
Mode Choice Logsum	1.0000 (fixed)		
Distance	-1.5378	0.436	-3.52
Distance * Part-Time	1.7894	0.382	4.68
Distance*Female	1.0434	0.465	2.24
Distance*Feamle*Presence of Child	-0.6537	0.556	-1.17
Distance*Four or more Vehicles	2.4295	0.400	5.65
Distance Squared	-1.1887	0.596	-1.99

Distance Effects



Latent Choice Modeling Update

- Developed the code for estimating latent choice models

- Currently, testing different latent models with varying number of latent classes and specifications
  - Estimation is done in two stages using Expectation Maximization algorithm
  - Model must be build gradually and can be time consuming!

Advantages of Latent Class Models

- Same data requirements as traditional choice models
- Can identify population segments with significantly different location patterns
  - Improved behavioral and forecasting accuracy
- These population segments can serve as good clusters for subsequent medium and short term choice models
  - Mode choice, daily activity patterns, tours and trips

Next Steps

- Explore important sampling methods (instead of random sampling)
- Estimate Latent Class Work Location model (Nashville)
- Demonstrate improved data fit
- Identify households and workers with significant differences in location choice preferences in Nashville
- Extend modeling method for other regions in TN

Thank you for your time

Q/A?