Global Initiative of Academic Networks (GIAN)

UNDERSTANDING THE INCREASING RELEVANCE OF CHOICE MODELS FOR ADVANCING TRANSPORTATION MODELLING IN SMART CITIES

MAY 22 -26, NAGPUR, INDIA

Module 3

Naveen Eluru, *University of Central Florida*

COURSE MODULES

Introduction	• Introduction to Smart City Technologies, their impact on Transportation
Stated Preference Module	• Background on Data Collection Approaches • Stated Preference Design and application
Traditional Discrete Choice Models	• Binary logit, multinomial logit, ordered logit, and count models
Advanced Discrete Choice Models	• Nested logit, mixed logit, maximum simulated likelihood estimation, regret minimization, discrete continuous models
Transportation Planning	• Current state of the art and recent advances

IN THIS MODULE

I will introduce choice modeling approaches for data analysis including binary logit, multinomial logit, ordered logit and count models

RECAP

- Yesterday we learned
  - Choice theory
  - Binary logit models
  - Probit/logit
  - Ran models with R codes for BL model using sample datasets
- Today we will
  - Build on this with MNL model
  - OL model
  - Count models

5

MULTINOMIAL LOGIT MODEL

MULTINOMIAL LOGIT MODEL

- When we have only two alternatives
  - Individual n, alternatives i and j
    - Probability of *i* is: $P_n(i) = Pr(U_{in} \geq U_{jn})$
    - Probability of *j* is : $P_n(j) = 1 - P_n(i)$
  - $U_{in} = V_{in} + \varepsilon_{in}$; $U_{jn} = V_{jn} + \varepsilon_{jn}$
  - $P_n(i) = Pr(U_{in} \geq U_{jn})$
  - $P_n(i) = Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn})$
  - $P_n(i) = Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn})$
- Making the assumption on the error terms as gumbel we arrive at the binary logit.
- Now we will explore cases with more than two alternatives

MULTINOMIAL LOGIT MODEL

- For a choice context with J alternatives, the alternative i is chosen if $U_{in} \geq U_{jn}$;
  - where $U_{in} = V_{in} + \varepsilon_{in}$
  - $U_{jn} = V_{jn} + \varepsilon_{jn}$ for all alternatives except i
- Now the probability of choosing i is given by
- $P_{in} = Pr\ (U_{in} \geq U_{jn}) = Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn})$ for all j ($\neq$i)
- $P_{in} = Pr(\varepsilon_{jn} \leq V_{in} - V_{jn} + \varepsilon_{in})$ for all j ($\neq$i)
- i.e., we want ε_{jn} to be less than $V_{in} - V_{jn} + \varepsilon_{in}$ for all j ($\neq$i)
- i.e., it's a multivariate cumulative distribution of J-1 dimensions (from $-\infty$, $V_{in} - V_{jn} + \varepsilon_{in}$)

MULTINOMIAL LOGIT MODEL

- To compute $\Pr(\varepsilon_{jn} \le V_{in} - V_{jn} + \varepsilon_{in})$ lets assume ε_{in} is known
- In this case the probability is nothing but the cdf function $f(\varepsilon_{1n}, \varepsilon_{2n} \ldots . \varepsilon_{Jn})$ for j = 1,2…J and ≠ i.
- $\int_{-\infty, j \neq i}^{V_{in} - V_{jn} + \varepsilon_{in}} f(\varepsilon_{1n}, \varepsilon_{2n} \ldots . \varepsilon_{Jn})\, d\varepsilon_{1n}, d\varepsilon_{2n} \ldots\, d\varepsilon_{Jn}$
  - Note $f(\varepsilon_{1n}, \varepsilon_{2n} \ldots . \varepsilon_{Jn})$ and $d\varepsilon_{1n}, d\varepsilon_{2n} \ldots\, d\varepsilon_{Jn}$ does not have ε_{in}
- Now ε_{in} varies from -∞ to +∞, so add integral for that
- $\Pr(\varepsilon_{jn} \le V_{in} - V_{jn} + \varepsilon_{in}) =$ $\int_{-\infty}^{\infty} \int_{-\infty, j \neq i}^{V_{in} - V_{jn} + \varepsilon_{in}} f(\varepsilon_{1n}, \varepsilon_{2n} \ldots . \varepsilon_{Jn})\, d\varepsilon_{1n}, d\varepsilon_{2n} \ldots\, d\varepsilon_{Jn}\ f(\varepsilon_{in}) d\varepsilon_{in}$

MULTINOMIAL LOGIT MODEL

- Lets assume the error terms are independent
- Then the joint probability is nothing but product of marginal probabilities
- $\int_{-\infty}^{\infty} \int_{-\infty, j\neq i}^{\mathrm{V_{in}}-\mathrm{V_{jn}}+\varepsilon_{\mathrm{in}}} f(\varepsilon_{1n}, \varepsilon_{2n} \dots . \varepsilon_{Jn})\, d\varepsilon_{1n}, d\varepsilon_{2n} \dots\, d\varepsilon_{Jn}\ \ d\varepsilon_{in}$
- $= \int_{-\infty}^{\infty} f(\varepsilon_{in}) d\varepsilon_{in}\ \prod_{j\neq i} \int_{-\infty}^{\mathrm{V_{in}}-\mathrm{V_{jn}}+\varepsilon_{\mathrm{in}}} f(\varepsilon_{jn})\, d\varepsilon_{jn}$
- $= \int_{-\infty}^{\infty} f(\varepsilon_{in}) d\varepsilon_{in}\ \prod_{j\neq i} F(\mathrm{V_{in}} - \mathrm{V_{jn}} + \varepsilon_{\mathrm{in}})$
- F represents cumulative gumbel probability
- Now when we integrate this we get
- $\mathrm{P_{in}} = \frac{\exp(Vi)}{\sum_{\forall j} \exp(Vj)}$

DERIVATION OF MULTINOMIAL LOGIT

- $\int_{-\infty}^{\infty} f(\varepsilon_{in}) d\varepsilon_{in} \prod_{j\neq i} F(\mathrm{V}_{\mathrm{in}}-\mathrm{V}_{\mathrm{jn}}+\varepsilon_{\mathrm{in}})$
  - Using Gumbel cdf
  - $\prod_{j\neq i} F(\mathrm{V}_{\mathrm{in}}-\mathrm{V}_{\mathrm{jn}}+\varepsilon_{\mathrm{in}}) = \prod_{j\neq i} e^{-e^{-(\mathrm{V}_{\mathrm{in}}-\mathrm{V}_{\mathrm{jn}}+\varepsilon_{\mathrm{in}})}}$
  - Using Gumbel pdf
  - $f(\varepsilon_{in}) = e^{-e^{-(\varepsilon_{\mathrm{in}})}} * e^{-(\varepsilon_{\mathrm{in}})}$
- $=\int_{-\infty}^{\infty} \prod_{j\neq i} e^{-e^{-(\mathrm{V}_{\mathrm{in}}-\mathrm{V}_{\mathrm{jn}}+\varepsilon_{\mathrm{in}})}} e^{-e^{-(\varepsilon_{\mathrm{in}})}} * e^{-(\varepsilon_{\mathrm{in}})} d\varepsilon_{in}$
- $=\int_{-\infty}^{\infty} \prod_{j\neq i} e^{-e^{-(\mathrm{V}_{\mathrm{in}}-\mathrm{V}_{\mathrm{jn}}+\varepsilon_{\mathrm{in}})}} e^{-e^{-(\varepsilon_{\mathrm{in}})}} * e^{-(\varepsilon_{\mathrm{in}})} d\varepsilon_{in}$
- $=\int_{-\infty}^{\infty} \prod_{j} e^{-e^{-(\mathrm{V}_{\mathrm{in}}-\mathrm{V}_{\mathrm{jn}}+\varepsilon_{\mathrm{in}})}} * e^{-(\varepsilon_{\mathrm{in}})} d\varepsilon_{in}$
  - We can do this because $\mathrm{V}_{\mathrm{in}} - \mathrm{V}_{\mathrm{in}} = 0$;

DERIVATION OF MULTINOMIAL LOGIT

- $=\int_{-\infty}^{\infty} exp\left[-\left(\sum_j e^{-(\mathrm{V_{in}}-\mathrm{V_{jn}}+\varepsilon_{\mathrm{in}})}\right)\right] * e^{-(\varepsilon_{\mathrm{in}})}\ d\varepsilon_{in}$
- $\int_{-\infty}^{\infty} exp\left[-e^{-(\varepsilon_{\mathrm{in}})}\left(\sum_j e^{-(\mathrm{V_{in}}-\mathrm{V_{jn}})}\right)\right] * e^{-(\varepsilon_{\mathrm{in}})}\ d\varepsilon_{in}$
- Now let t = $e^{-(\varepsilon_{\mathrm{in}})}$; then dt = -$e^{-(\varepsilon_{\mathrm{in}})}\ d\varepsilon_{in}$
- We need to change limits;
  - $\varepsilon_{\mathrm{in}} = -\infty$ => $t = \infty$; $\varepsilon_{\mathrm{in}} = \infty$ => $t = 0$
- $\int_{\infty}^{0} exp\left[-t\left(\sum_j e^{-(\mathrm{V_{in}}-\mathrm{V_{jn}})}\right)\right] * (-dt)$
- $=\int_{0}^{\infty} exp\left[-t\left(\sum_j e^{-(\mathrm{V_{in}}-\mathrm{V_{jn}})}\right)\right] * dt$
- $\frac{exp\left[-t\left(\sum_j e^{-(\mathrm{V_{in}}-\mathrm{Vjn})}\right)\right]}{-\sum_j e^{-(\mathrm{V_{in}}-\mathrm{Vjn})}}\ to\ be\ computed\ at\ t=\infty\ and\ t=0$

DERIVATION OF MULTINOMIAL LOGIT

- $\frac{exp\left[-t\left(\sum_j e^{-(\mathbf{V}_{\text{in}}-\mathbf{V}_{\text{jn}})}\right)\right]}{-\sum_j e^{-(\mathbf{V}_{\text{in}}-\mathbf{V}_{\text{jn}})}}$ $at\ t = \infty$
  - = 0
- $\frac{exp\left[-t\left(\sum_j e^{-(\mathbf{V}_{\text{in}}-\mathbf{V}_{\text{jn}})}\right)\right]}{-\sum_j e^{-(\mathbf{V}_{\text{in}}-\mathbf{V}_{\text{jn}})}}$ $at\ t = 0$
  - $=\frac{1}{-\sum_j e^{-(\mathbf{V}_{\text{in}}-\mathbf{V}_{\text{jn}})}}$
- Thus, the integral $=\frac{1}{\sum_j e^{-(\mathbf{V}_{\text{in}}-\mathbf{Vjn})}} = \frac{e^{\mathbf{V}_{\text{in}}}}{\sum_j e^{\mathbf{V}_{\text{jn}}}}$

IMPLICATIONS OF THE MNL ASSUMPTIONS

- Independent errors
  - Consider mode choice model, with 4 alternatives car, shared ride, bus and metro. A person whose personality prefers transit modes will assign a higher value to both bus and metro… or a person who prefers car will assign higher utility to car and shared ride… So neglecting this might have implications for what we are trying to do
- $V_{in} + \varepsilon_{in}$

Remember ε affects the choice

IMPLICATIONS OF THE ASSUMPTIONS

- Non-identical variances
  - For auto alternatives the level of comfort for example are clear.. There is not as much variability. But depending on the transit line and no. of people travelling there is substantial variability in the level of comfort in public transportation modes. Hence.. The unobserved components have more variability… so assuming identical variances is incorrect!
- Please note that the reason we do complex models is because they allow us to incorporate complex interactions into the choice modeling process

MULTINOMIAL LOGIT MODEL

- Quickly recap our assumptions on error terms
  - Independent and identically distributed for all individuals
- Let us say we are considering different alternatives of mode choice
  - Car, car pool, bus, train, metro, walk and bike
  - Will the assumption hold?
  - Isn't it possible that car and car pool have errors coming from a distribution that is different from the distribution for other alternatives
  - Be careful with the assumptions

MULTINOMIAL LOGIT MODEL

- Strengths and Weaknesses of MNL
- The structure of the MNL lends itself to easy model estimation
  - Probability computation is free from integration or simulation
  - If you maintain linear utility specification irrespective of where we begin we will reach the optimal solution (concave)
  - Easy to interpret because of the utility structure
- There has to be a catch right?
- Taste Variation
  - Logit accommodates taste variation based on observed attributes (income or vehicle on mode choice)
  - Logit cannot accommodate taste variation based on unobserved attributes (social nature influence on mode choice)

IIA PROPERTY

- IIA property
  - MNL is saddled with "independence of irrelevant alternatives" property
- Consider the ratio of alternative probabilities for i and j.
- P_i / P_j = $\frac{\exp(Vi)}{\sum_{\forall j}\exp(Vj)} \Big/ \frac{\exp(Vj)}{\sum_{\forall j}\exp(Vj)} = \frac{\exp(Vi)}{\exp(Vj)}$ = exp(V_i-V_j)
- A function only of the alternatives i and j
- Consider that an individual has two options to get to work: (A) Auto and (R) Red Bus. Lets say the probability for choosing A and B are equal. Hence P(A) = P(R) = 0.5

IIA PROPERTY

- Now, a new bus service is introduced. The only difference from the existing bus service is that it is a Blue bus. Now since it is the exact same bus P(R) / P(B) = 1.
- However, P(A)/P(R)=1 and P(A)+P(R)+P(B) =1
- Hence P(A) = P(R)=P(B)=1/3
- In reality, we expect P(A) to remain same and the other two bus alternatives share the probability.
- It is not all bad – IIA has some advantages
  - Because of IIA, we can estimate the model on only a subset of alternatives and yet get consistent results

MARGINAL EFFECTS

- $P_{\text{in}} = \frac{e^{\beta' X_{in}}}{\left(\sum_{\forall j} e^{\beta' X_{jn}}\right)}$; Q $= \left(\sum_{\forall j} e^{\beta' X_{jn}}\right)$
- Let k^th variable be altered
- Self: $\frac{\partial P_{\text{in}}}{\partial \text{x}_{ik}} = \frac{Q \, * \, e^{\beta' X_{Tn} * \beta_{\,T\,k}} - e^{\beta' X_{Tn} *}\left(e^{\beta' X_{Dn} * 0} + e^{\beta' X_{Tn} * \beta_{\,T\,k}}\right)}{Q^2}$
- $= \beta_k[P_{in}]$ - $\beta_k\ [P_{in}\ ]^2$
- $= \beta_k[P_{in}]\ [1\ - P_{in}\ ]$
- Cross: $\frac{\partial P_{\text{in}}}{\partial \text{x}_{ik}} =$ - $\beta_k[P_{\text{in}}\ ]\ [1\ - P_{in}\ ]$

ELASTICITY EFFECTS

- Slightly different definition that marginal effects
- Self elasticity : $\frac{\partial P_{in}}{\mathrm{P}_{in}} \Big/ \frac{\partial \mathrm{x}_{ik}}{\mathrm{x}_{ik}} = \frac{\partial P_{in}}{\partial \mathrm{x}_{ik}} * \frac{\mathrm{x}_{ik}}{P_{in}}$
- $= \beta_k [P_{in}]\ [1 \ - P_{in}\ ] * \frac{\mathrm{x}_{ik}}{P_{in}}$
- $= \beta_k [1\text{-}P_{in}\ ]\ \mathrm{x}_{ik}$
- Cross-elasticity: $\frac{\partial P_{jn}}{\mathrm{P}_{jn}} \Big/ \frac{\partial \mathrm{x}_{ik}}{\mathrm{x}_{ik}} = -\beta_k [P_{in}\ ]\ \mathrm{x}_{ik}$

- Very similar to elasticity from binary logit models
- Cross and self exist only for variables that are related to alternative attributes
- For variables specific to individual we only have one effect

MARGINAL EFFECT OF INCOME

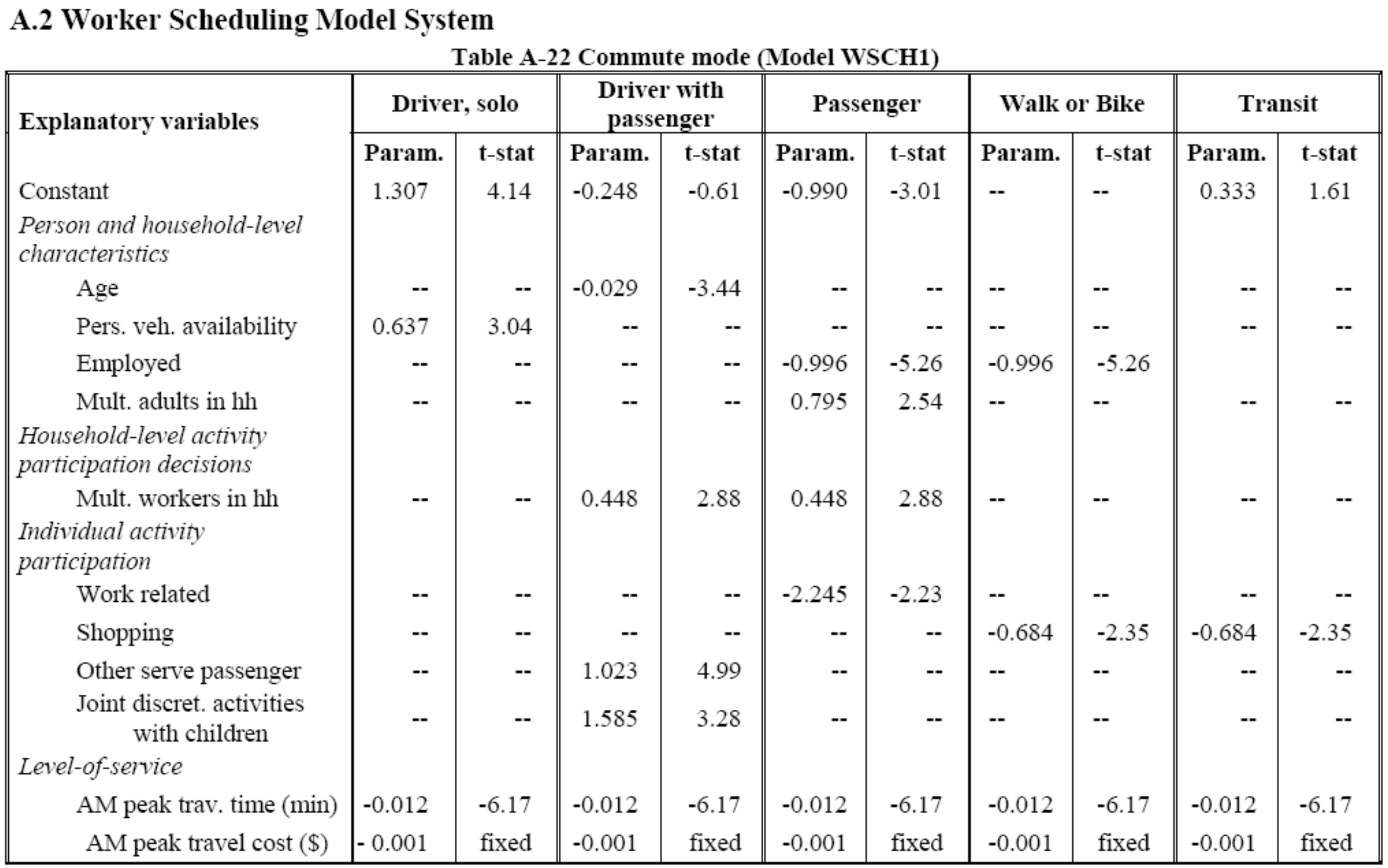
- Consider income effect on three alternative case - Let income coefficient is 0 for alt 1 (base)
- Consider prob for 2nd alternative
- $\mathrm{P}_{2\mathrm{n}} = \dfrac{e^{\beta' X_{2n} + \beta_{inc} Inc_n}}{e^{\beta' X_{1n}} + e^{\beta' X_{2n} + \beta_{inc} Inc_n} + e^{\beta' X_{3n} + \beta_{inc} Inc_n}}$
- D - $e^{\beta' X_{1n}} + e^{\beta' X_{2n} + \beta_{inc} Inc_n} + e^{\beta' X_{3n} + \beta_{inc} Inc_n}$
- $\dfrac{\partial P_{2n}}{\partial \mathrm{inc}} = \dfrac{D * \beta_{inc} * e^{\beta' X_{2n} + \beta_{inc} Inc_n} - \{e^{\beta' X_{2n} + \beta_{inc} Inc_n}\} * \left[\sum_{j \neq 1} \beta_{inc} (e^{\beta' X_{jn} + \beta_{inc} Inc_n})\right]}{D^2}$
- $= \mathrm{P}_{2\mathrm{n}} \left[\beta_{inc} - \sum_{j \neq 1} P_{jn} \beta_{inc}\right]$
- In general
- $\dfrac{\partial P_{in}}{\partial \mathrm{inc}} = \mathrm{P}_{\mathrm{in}} \left[\beta_{inc} - \sum_{j \neq 1} P_{jn} \beta_{inc}\right]$

COEFFICIENT INTERPRETATIONS

- When multiple alternatives exist – interpretation is less straight forward
- Example income variable in a 3 alternative case (alt 1 is base)
    - Alt 2 Coeff = 0.25
    - Alt 3 Coeff = 0.55
- What will happen if income increases?
    - Alt 1 ?
        - reduces
    - Alt 3 ?
        - increases
    - Alt 2?
        - Depends
- The extreme cases are easy to predict – the intermediate ones are not so easy – they need to be computed using elasticity

MULTINOMIAL CHOICE MODELS

A mode choice model including traveler characteristics



A.2 Worker Scheduling Model System

Table A-22 Commute mode (Model WSCH1)

Explanatory variables	Driver, solo		Driver with passenger		Passenger		Walk or Bike		Transit	
	Param.	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**
Constant	1.307	4.14	-0.248	-0.61	-0.990	-3.01	--	--	0.333	1.61
Person and household-level characteristics										
Age	--	--	-0.029	-3.44	--	--	--	--	--	--
Pers. veh. availability	0.637	3.04	--	--	--	--	--	--	--	--
Employed	--	--	--	--	-0.996	-5.26	-0.996	-5.26		
Mult. adults in hh	--	--	--	--	0.795	2.54	--	--	--	--
Household-level activity participation decisions										
Mult. workers in hh	--	--	0.448	2.88	0.448	2.88	--	--	--	--
Individual activity participation										
Work related	--	--	--	--	-2.245	-2.23	--	--	--	--
Shopping	--	--	--	--	--	--	-0.684	-2.35	-0.684	-2.35
Other serve passenger	--	--	1.023	4.99	--	--	--	--	--	--
Joint discret. activities with children	--	--	1.585	3.28	--	--	--	--	--	--
Level-of-service										
AM peak trav. time (min)	-0.012	-6.17	-0.012	-6.17	-0.012	-6.17	-0.012	-6.17	-0.012	-6.17
AM peak travel cost ($)	- 0.001	fixed	-0.001	fixed	-0.001	fixed	-0.001	fixed	-0.001	fixed

SOURCE: http://www.ce.utexas.edu/prof/bhat/REPORTS/4080_8_draft_Dec11_2006.pdf

CASE STUDY

Eluru, N., V. Chakour, and A. El-Geneidy (2012), "Travel Mode Choice and Transit Route Choice Behavior in Montreal: Insights from McGill University Members Commute Patterns," Public Transport: Planning and Operations Vol. 4, No. 2, pp. 129-149

CASE STUDY: PART 1

- We investigate individual's decision framework to choose between transit and car mode of transportation for commuting to McGill University
- The sample consists of 1778 records
- Of these 1228 (69.1%) respondents commute using transit while 550 (30.9%) respondents commute by car
- We need to generate the LOS attributes for modes under consideration
- Car in-vehicle travel times for all individuals (irrespective of their choice) were generated using LOS matrices for postal code origin and destinations
- Google Maps were employed to generate the best transit alternative available to the individuals using car at the time of his/her departure to work
- For respondents choosing transit, the actual transit route alternative information compiled in the survey was employed to tag the chosen alternative

CASE STUDY: PART 1

Attributes	Parameter	t-stats
(Car alternative is the base)		
Constant	9.1685	8.691
Age	-0.2425	-6.062
Age squared	0.0022	5.453
Respondent status		
Staff member	0.6073	3.915
Student	0.8001	2.913
Full time member of the community	0.3433	1.735
Driver license status	-1.2406	-3.559
Household car ownership	-1.0623	-11.582
In-vehicle Travel time	-0.0594	-7.004
Transfers	-0.8143	-9.145
Walk time	-0.0145	-1.419
Initial Waiting Time	-0.0244	-5.054
Log-likelihood at Convergence	-685.7	
Log-likelihood at constants	-1099.8	
McFadden rho-square	0.37	

RESULTS: PART 1

- Age exerts a significantly negative influence on choosing the transit mode
  - Younger individuals of the McGill community (students and younger employees) are more likely to use the public transportation mode compared to older members of the McGill community
- The adoption of transit is the highest among students followed by staff members compared to faculty members
- Full-time employees and students are more likely to commute by transit compared to part time employees and students
  - The full-time members have a more definite work schedule, making it easier for them to commute to work by transit
- The license status of the individual affects the choice between transit and car
  - Within the student community it is possible a number of individuals do not have driver licenses and are captive to the public transportation mode

RESULTS: PART 1

- Household car ownership also has a strong negative effect on the choice of transit mode. Households with more cars are least likely to commute to work by transit
- LOS attributes including travel time, number of transfers, walking time, and initial wait time significantly influences the choice between auto and transit modes.
- Increasing travel time reduces the likelihood of choosing the alternative
- The increase in the amount of walking within the transit alternative significantly reduces the likelihood of the respondent using transit for commuting.
- Increase in the number of transfers for travelling by transit reduces the likelihood of using transit substantially
- The initial waiting time for the transit alternative exerts a strong influence of car versus transit choice

CASE STUDY: PART 2

- For 1228 respondents that commute using transit we studied their transit route choice alternatives
- Sample statistics

Transit route choice dataset	
Mean Travel Time	23.5
Mean Total Walking Time	17.0
Mean Total Waiting Time	3.7
Transit route alternatives comprising	
Bus	69.0
Metro	49.5
Train	14.8
Average travel time by mode (min)	
Bus	21.4
Metro	10.3
Train	24.3

RESULTS: PART 2

Attribute	Parameter	t-stats
Transit alternative has bus	-0.2375	-1.066
Transit alternative has metro	0.6378	2.145
Transit alternative has train	-1.5665	-2.142
The alternative with the earliest arrival time	0.2361	2.209
Travel time in bus	-0.2690	-5.997
Travel time in metro	-0.1616	-3.238
Travel time in train	-0.1737	-3.420
Standard Deviation	0.0496	2.000

RESULTS: PART 2

Attribute	Parameter	t-stats
Total Walking time	-0.3550	-7.806
Total Walking time squared	0.0013	1.441
Standard Deviation	0.1297	4.191
Number of transfers	-2.4985	-8.101
Standard Deviation	0.9752	2.293
Waiting Time per transfer	-0.0766	-2.341
Total travel time interactions with Socio-demographics		
Female	0.0688	2.955
Age	0.0012	1.584
Faculty	-0.0395	-1.465
Log-likelihood at Convergence	-681.7	
Log-likelihood at Equal shares	-1207.4	
McFadden rho-square	0.42	

RESULTS: PART 2 DISCUSSION

- The travel time coefficients clearly indicate the negative propensity towards travel for respondents.
- In the model, we introduced travel time by mode. The coefficient on each of these modes provides the sensitivity to travel time for respondents by that mode.
- The results indicate that individuals find travel time on the bus mode the most onerous while the sensitivity to travel time on metro and train are quite similar on average

RESULTS: PART 2 DISCUSSION

- The influence of walking time is along expected lines. Specifically, transit route alternatives with smaller walk times are preferred.
  - The model results indicate the presence of a non-linear relationship (linear and square terms).
- Further, the results indicate a substantial variation on the mean effect of the walking time variable. The result is quite intuitive, because, different individuals are likely to be differentially sensitive to walking time.
- The best statistical and intuitive fit was obtained for the specification that includes the transfer variable as well as the waiting time per transfer variable.
  - As expected, alternatives with fewer transfers were preferred.
- At the same time, individuals exhibited higher likelihood of choosing alternatives with smaller waiting time per transfer.

RESULTS: PART 2 DISCUSSION

- In a route choice model, it is not possible to evaluate the effect of socio-demographics directly.
- In the model we consider interactions of gender, age, employment status with total travel time (sum of travel time by all modes in a route).
- Travel time interacted with female gender results in a positive coefficient indicating that females are less sensitive to travel time compared to males.
  - To be sure, the overall sensitivity to travel time for females is still negative. However, it is lower than the sensitivity of travel time for males.
- The results corresponding to the interaction variable involving age and total travel time indicate that with increasing age of the respondent, there is a marginal reduction in the sensitivity of travel time.
- Faculty members are more sensitive to travel time compared to the students and staff members

POLICY ANALYSIS (ELASTICITY)

Attribute	Car	Transit
Travel time by Transit reduced by 5 minutes	-11.51	5.15
Travel time by Transit reduced by 10 minutes	-21.68	9.71
Travel time by Car increased by 5 minutes	-11.60	5.20
Travel time by Car increased by 10 minutes	-22.49	10.07
Walking time to transit reduced by 5 minutes	-2.88	1.29
Walking time to transit reduced by 10 minutes	-5.53	2.48
Initial Waiting Time reduced by 5 minutes	-3.66	1.64
Initial Waiting Time reduced by 10 minutes	-5.74	2.57
No. of transfers (for transit) reduced by 1	-18.75	8.39
Household vehicle ownership reduced by 1	-35.39	15.85

MARKET SEGMENTATION AND LATENT CLASS MODELS

MARKET SEGMENTATION

- In the discrete choice model framework we can estimate models for different segments
- Consider travel mode choice model: it is possible to estimate the models for different segments
  - Males vs Females
  - High income vs Low income
- Lets consider two segments. The pooled L can be shown to be = sum of L_1 and L_2.
- $L_1 + L_2 = \sum_{n=1}^{N} \sum_{\forall j} \delta_{jn} \sum_{s} \left(\Delta_s lnP_{jn}^{s} \right)$
- $= \sum_{s} \left(\sum_{n=1}^{N} \sum_{\forall j} \delta_{jn} \, lnP_{jn}^{s} \right) = L$

MARKET SEGMENTATION

- Example: three modes D, W, T
- Specification

Case	Alt	UnoW	UnoT	IVTT	UnoW * Male	UnoT * Male	IVTT * Male	UnoW * Female	UnoT * Female	IVTT * Female
1	D	0	0	7						
1	W	1	0	12						
1	T	0	1	18						
2	D	0	0	19						
2	W	1	0	45						
2	T	0	1	25						

- We can check if the full specification is required by comparing the incremental coefficients in the pooled model

SEGMENTATION MODELS

- Market segmentation is usually good if you have segmentation on 1 or 2 variables
- Market segmentation allows us to estimate different coefficients for different segments – thus allowing for coefficients to vary across the population i.e. we don't consider the entire population as one homogenous lump and allow for heterogenous variation
- How can we achieve this?
- We segment based on gender (2) and income (4) – total 8 segmented models

SEGMENTATION MODELS

- Now if I decide to consider the effect of location (downtown, urban and suburban)
- Now this will add 8*3 = 24 segments
- As you can see addition of one more variable will make it more cumbersome – further there are very few records in each of the segments – thus making the estimation process hard
  - You are trying to estimate a coefficient with very few records
- So the approach referred to as exogenous (deterministic) segmentation
  - The approach is mutually exhaustive segmentation
  - is feasible only for segmentation based on 2-3 variables
  - Results in a loss of efficiency

ENDOGENOUS SEGMENTATION MODELS

- So there are alternative ways of achieving segmentation : Endogenous segmentation approach
- In this approach, we allow decision makers to be part of different segments probabilistically
- To explain this, lets say there are two segments in the population; within each segment the population is assumed to be homogenous and we estimate the choice model for each segment
- There are two steps:
  - 1) segmentation
  - 2) discrete choice model for each segment
- We know how to do step 2. If we know how to do 1 and combine 1 and 2 we can develop latent segmentation model
- The question is how do we decide the segments
  - We do a probability model
  - So assign utility for decision makers to be part of a segment – we get a probability for each DM for every segment
  - So for individual p1 and p2 are probabilities of being part of segment 1 and segment 2 (p1+p2=1)

ENDOGENOUS SEGMENTATION MODELS

- Lets examine the mathematical structure
- Step 2
- Given an individual is part of segment s, the probability to choose alternative i is
- $P_{ni}(S) = \frac{\exp(Vi)}{\sum_{\forall j} \exp(Vj)}$
- However, we need to determine the probability to be part of segment S.
- Now $P(S) = \frac{\exp(Z_s)}{\sum_{\forall s} \exp(Z_t)}$ is also a logit probability where Zs represents individual utility for being part of segment 1
- The unconditional probability is obtained as $\sum_s P(S) * Pni(S)$
- For a two segment case:
  - Probability for i^{th} alternative is given by $[P(1) * Pni(1) + (1 - P(1)) * Pni(2)]$
- How do we decide on no. of segments?
  - We start with 2 segments and add segments until we improve the data fit; when additional segments do not add value to data fit we stop
- Approach allows us to determine segments based on a host of variables

CASE STUDY: ENDOGENOUS SEGMENTATION

Bhat, C.R., 1997. An endogenous segmentation mode choice model with an application to intercity travel. Transportation Science 31 (1), 34-48.

MODE CHOICE

- Intercity travel mode choice behavior.
- The data used in the current empirical analysis was assembled by VIA Rail in 1989 to develop travel demand models to forecast future intercity travel in the Toronto-Montreal corridor
- The data includes socio-demographic and general trip-making characteristics of the traveler, and detailed information on the current trip (purpose, party size, origin and destination cities, etc.).
- The universal choice set included car, air, train and bus).
- Level of service data were generated for each available mode and each trip based on the origin/destination information of the trip

SAMPLE CHARACTERISTICS

Mode	Frequency (departures/day)	Total cost (in Canadian $)	In-vehicle time (in mins.)	Out-of-vehicle time (in mins.)
Train	4.21 (2.3)	58.58 (17.7)	244.50 (115.0)	86.32 (22.0)
Air	25.24 (14.0)	157.33 (21.7)	57.72 (19.2)	106.74 (24.9)
Car	not applicable	70.56 (32.7)	249.60 (107.5)	0.00 (0.0)

MODEL ESTIMATION APPROACH

- Segmentation Model
  - The variables are: income, sex (female or male), travel group size (traveling alone or traveling in a group), day of travel (weekend travel or weekday travel), and (one-way) trip distance.
  - The segmentation variables were introduced as alternative-specific variables in the logit model with the last segment being the base.
- Mode choice Model
  - The level-of-service variables used to model choice of mode included modal level-of-service measures (frequency of service, total cost, in-vehicle travel time and out-of-vehicle travel time) and a large city indicator which identified whether a trip originated, terminated, or originated and terminated in a large city.

MODEL RESULTS: SEGMENTATION

Variable	Segment 1		Segment 2		Segment 3	
	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic
Constant	4.4227	7.62	1.5366	2.56	Base Segment	
Income	-0.0293	-5.73	-0.0447	-8.60		
Female	-0.7614	-3.46	0.9703	4.05		
Traveling Alone	-0.1657	-1.70	-0.7226	-4.07		
Weekend Travel	0.2423	0.65	1.5326	4.71		
Trip Distance	-0.0047	-5.91	-0.0030	-3.79		
Sample Share	0.4866		0.1220		0.3914	

SEGMENTATION DEMOGRAPHIC SUMMARY

Variable	Segment 1	Segment 2	Segment 3	Overall Market
Income (x 10^3 Can$)	52.16	44.09	60.28	54.36
Female	0.13	0.48	0.20	0.20
Traveling Alone	0.69	0.57	0.77	0.70
Weekend Travel	0.20	0.62	0.19	0.25
Trip Distance (km)	311.80	373.37	444.76	371.35

MODEL RESULTS: MODE CHOICE

Variable	Segment 1		Segment 2		Segment 3	
	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic
Mode Constants						
Train	-3.0617	-2.54	4.7763	2.12	1.1737	0.60
Air	-1.0516	-1.82	-1.3691	-1.01	4.3404	3.36
Large City Indicator						
Train	1.9273	2.20	0.2146	0.32	-0.0840	-0.12
Air	2.2240	3.46	-1.3691	-1.01	2.6892	2.37
Frequency of Service (deps./day)	0.1615	6.38	0.5784	3.49	0.1790	3.92
Travel Cost (Canadian $)	-0.0591	-4.53	-0.1728	-3.27	-0.0166	-0.54
Travel Time (minutes)						
In-Vehicle	-0.0254	-3.25	-0.0030	-1.20	-0.0657	-5.21
Out-of-Vehicle	-0.0436	-2.91	-0.0239	-1.84	-0.1627	-5.01

MODEL ESTIMATION

- Estimating these models is not easy – very unstable LL function
- Starting values are very critical
- EM algorithm is two stage model used to make the process easy

JOINT CHOICE MODELS

JOINT CHOICES

- We can use the multinomial logit model to study joint choices
  - Mode and departure time choice (2 distinct choices)
  - When people leave and how people leave are connected
    - Car – off peak, transit – peak etc.
  - We can generate joint alternatives by creating combinations of choice 1 and choice 2
  - Let us say we have 3 (n) mode combinations (D, T, W) and 2 (k) departure time combinations (Peak and Offpeak) – no. of posisble joint combinations is given by 3*2 (n*k) = 6
  - Alternatives: D-P, T-P, W-P, D-OP, T-OP, W-OP

JOINT CHOICES

- Let us examine the specification of say Vehicle ownership variable (#V) (D-P is base alternative)

Case	$\#V_{T\text{-}P}$	$\#V_{W\text{-}P}$	$\#V_{D\text{-}OP}$	$\#V_{T\text{-}OP}$	$\#V_{W\text{-}OP}$
D-P	0	0	0	0	0
T-P	#V	0	0	0	0
W-P	0	#V	0	0	0
D-OP	0	0	#V	0	0
T-OP	0	0	0	#V	0
W-OP	0	0	0	0	#V

- We can estimate (n*k-1) coefficients

JOINT CHOICES

- It is possible that we might have reason to believe #V only affects mode choice and not time choice
- How do we accommodate that?
- In this case we estimate 2 coefficients
- In cases where n and k are high (>5) – we start estimating across the dimensions i.e n-1 and k-1 are estimated and very important interactions are considered
- So instead of estimating (n*k-1) we end up estimating (n+k-2) parameters

Case	D	T	W
$V_{D\text{-}P}$	0	0	0
$V_{T\text{-}P}$	0	#V	0
$V_{W\text{-}P}$	0	0	#V
$V_{D\text{-}OP}$	0	0	0
$V_{T\text{-}OP}$	0	#V	0
$V_{W\text{-}OP}$	0	0	#V

CASE STUDY: ANALYSIS OF TEMPORAL AND SPATIAL FLEXIBILITY OF ACTIVITIES, VEHICLE TYPE CHOICE AND PRIMARY DRIVER SELECTION

Anowar S., N. Eluru, L. Miranda-Moreno, and M. Lee-Gosselin (2015), "A Joint Econometric Analysis of Temporal And Spatial Flexibility Of Activities, Vehicle Type Choice And Primary Driver Selection" Transportation Research Record Vol. 2495, Jan 2015, pp. 32-41

REGION AND DATA

- Quebec City Travel and Activity Panel Survey (QCTAPS)
  - Investigates how households and individuals organize their activities in space and time
  - Comprised of three waves, about one year apart
- Carried out from 2003-2006
  - Region: Quebec City, Canada
  - Number of households: 250
  - Retention rate: 67%

DEPENDENT VARIABLES

- Perceived temporal flexibility
  - Routine
  - Planned
  - Impulsive
- Perceived spatial flexibility
  - Routine
  - Planned
  - Impulsive
- Vehicle type
  - Compact sedan
  - Large sedan
  - Van and minivan
  - Sports Utility Vehicle (SUV)
  - Pick-up and truck
  - Other vehicles (walk, bike, transit)
- Primary driver
  - As many drivers as many adults (Maximum of 4)
- Total discrete alternatives 216 (3*3*6*4)

DATA

- A total of 46,730 activities
  - Out-of-home activities: 14,579
- The final sample
  - Out-of-home activities: 8,098
- Households: 234
- Individuals: 378
  - More than 90 percent owned at least one vehicle

DATA SUMMARY

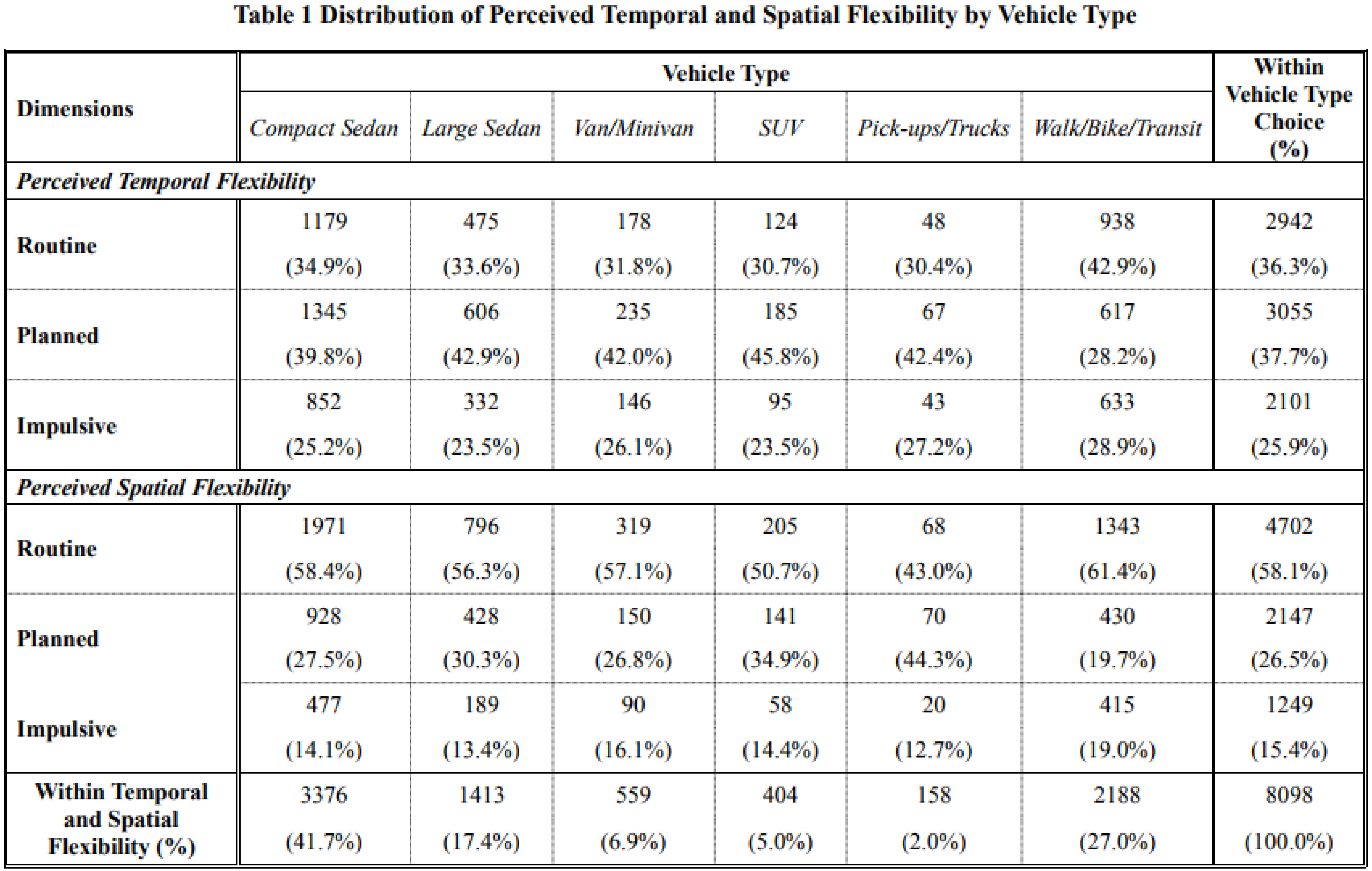


Table 1 Distribution of Perceived Temporal and Spatial Flexibility by Vehicle Type

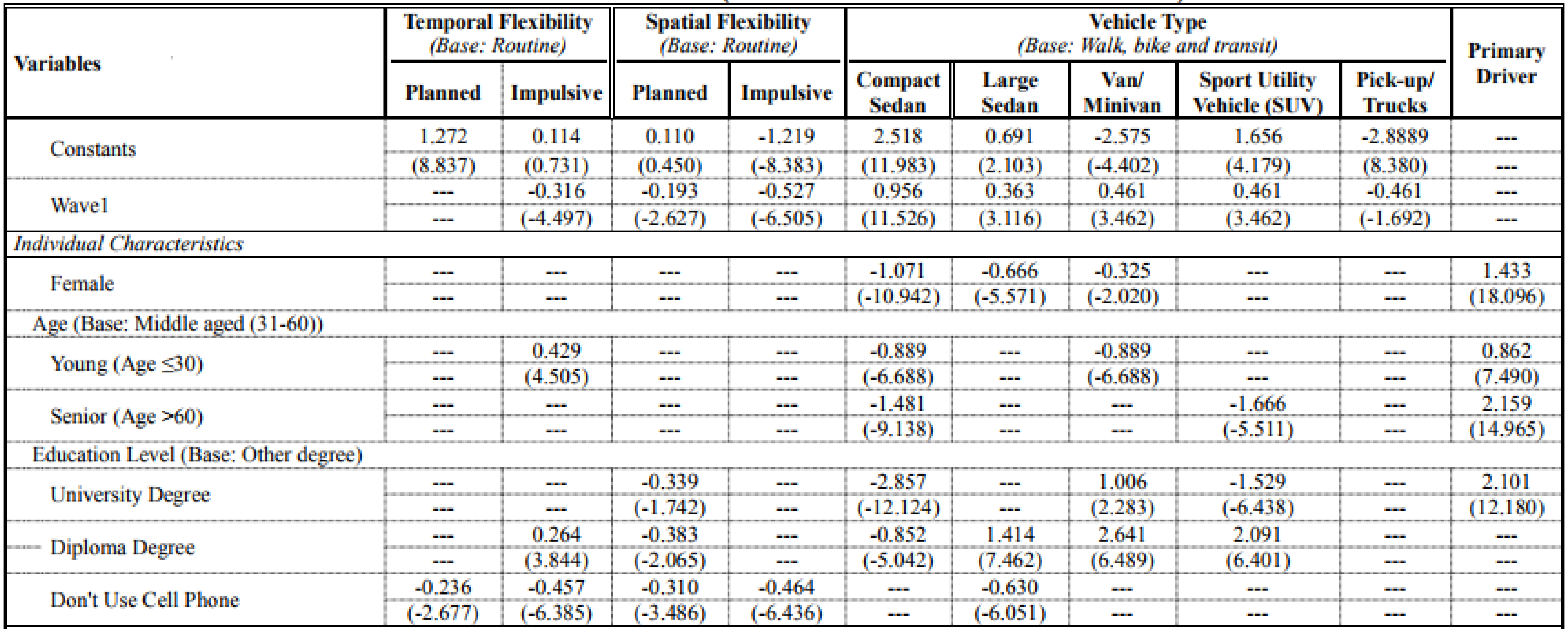
Dimensions	**Vehicle Type**						**Within Vehicle Type Choice (%)**
	Compact Sedan	*Large Sedan*	*Van/Minivan*	*SUV*	*Pick-ups/Trucks*	*Walk/Bike/Transit*	
Perceived Temporal Flexibility							
Routine	1179 (34.9%)	475 (33.6%)	178 (31.8%)	124 (30.7%)	48 (30.4%)	938 (42.9%)	2942 (36.3%)
Planned	1345 (39.8%)	606 (42.9%)	235 (42.0%)	185 (45.8%)	67 (42.4%)	617 (28.2%)	3055 (37.7%)
Impulsive	852 (25.2%)	332 (23.5%)	146 (26.1%)	95 (23.5%)	43 (27.2%)	633 (28.9%)	2101 (25.9%)
Perceived Spatial Flexibility							
Routine	1971 (58.4%)	796 (56.3%)	319 (57.1%)	205 (50.7%)	68 (43.0%)	1343 (61.4%)	4702 (58.1%)
Planned	928 (27.5%)	428 (30.3%)	150 (26.8%)	141 (34.9%)	70 (44.3%)	430 (19.7%)	2147 (26.5%)
Impulsive	477 (14.1%)	189 (13.4%)	90 (16.1%)	58 (14.4%)	20 (12.7%)	415 (19.0%)	1249 (15.4%)
Within Temporal and Spatial Flexibility (%)	3376 (41.7%)	1413 (17.4%)	559 (6.9%)	404 (5.0%)	158 (2.0%)	2188 (27.0%)	8098 (100.0%)

MODEL

- A MNL base structure is used
- In the paper a more advanced model is developed – Mixed MNL (to be discussed later)

RESULTS

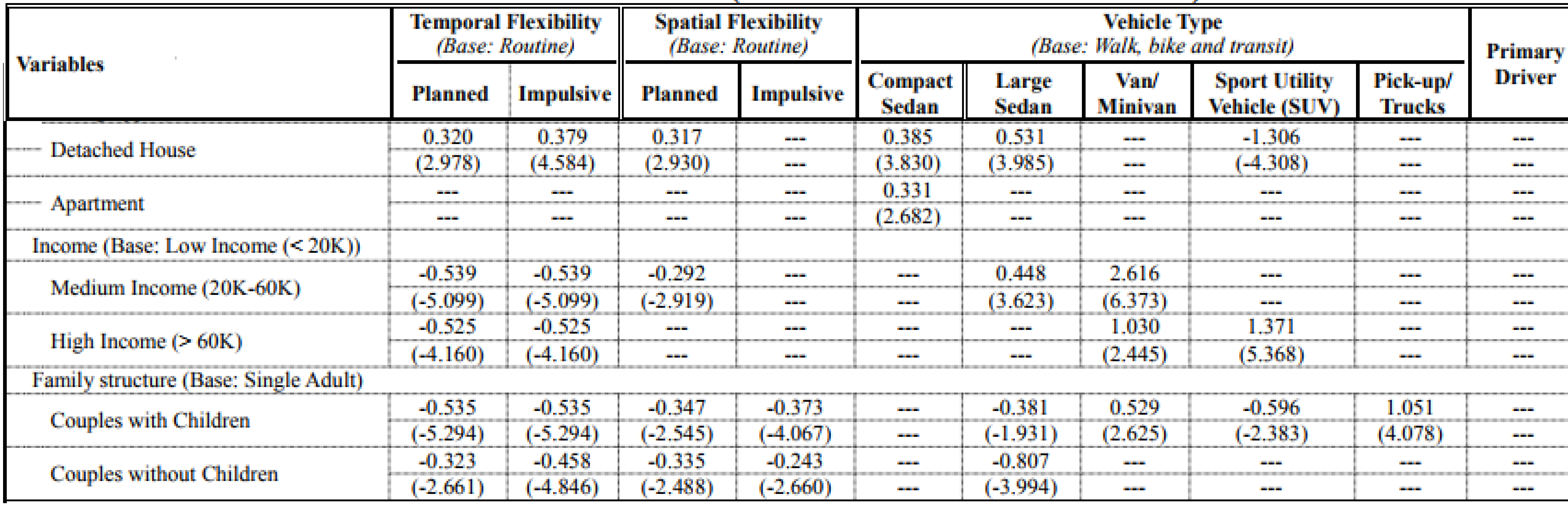
Table 3 Estimation Results (N=378 individuals and 8098 records)



Variables	Temporal Flexibility *(Base: Routine)*		Spatial Flexibility *(Base: Routine)*		Vehicle Type *(Base: Walk, bike and transit)*					Primary Driver
	Planned	Impulsive	Planned	Impulsive	Compact Sedan	Large Sedan	Van/ Minivan	Sport Utility Vehicle (SUV)	Pick-up/ Trucks	
Constants	1.272	0.114	0.110	-1.219	2.518	0.691	-2.575	1.656	-2.8889	---
	(8.837)	(0.731)	(0.450)	(-8.383)	(11.983)	(2.103)	(-4.402)	(4.179)	(8.380)	---
Wave1	---	-0.316	-0.193	-0.527	0.956	0.363	0.461	0.461	-0.461	---
	---	(-4.497)	(-2.627)	(-6.505)	(11.526)	(3.116)	(3.462)	(3.462)	(-1.692)	---
Individual Characteristics										
Female	---	---	---	---	-1.071	-0.666	-0.325	---	---	1.433
	---	---	---	---	(-10.942)	(-5.571)	(-2.020)	---	---	(18.096)
Age (Base: Middle aged (31-60))										
Young (Age ≤30)	---	0.429	---	---	-0.889	---	-0.889	---	---	0.862
	---	(4.505)	---	---	(-6.688)	---	(-6.688)	---	---	(7.490)
Senior (Age >60)	---	---	---	---	-1.481	---	---	-1.666	---	2.159
	---	---	---	---	(-9.138)	---	---	(-5.511)	---	(14.965)
Education Level (Base: Other degree)										
University Degree	---	---	-0.339	---	-2.857	---	1.006	-1.529	---	2.101
	---	---	(-1.742)	---	(-12.124)	---	(2.283)	(-6.438)	---	(12.180)
Diploma Degree	---	0.264	-0.383	---	-0.852	1.414	2.641	2.091	---	---
	---	(3.844)	(-2.065)	---	(-5.042)	(7.462)	(6.489)	(6.401)	---	---
Don't Use Cell Phone	-0.236	-0.457	-0.310	-0.464	---	-0.630	---	---	---	---
	(-2.677)	(-6.385)	(-3.486)	(-6.436)	---	(-6.051)	---	---	---	---

RESULTS

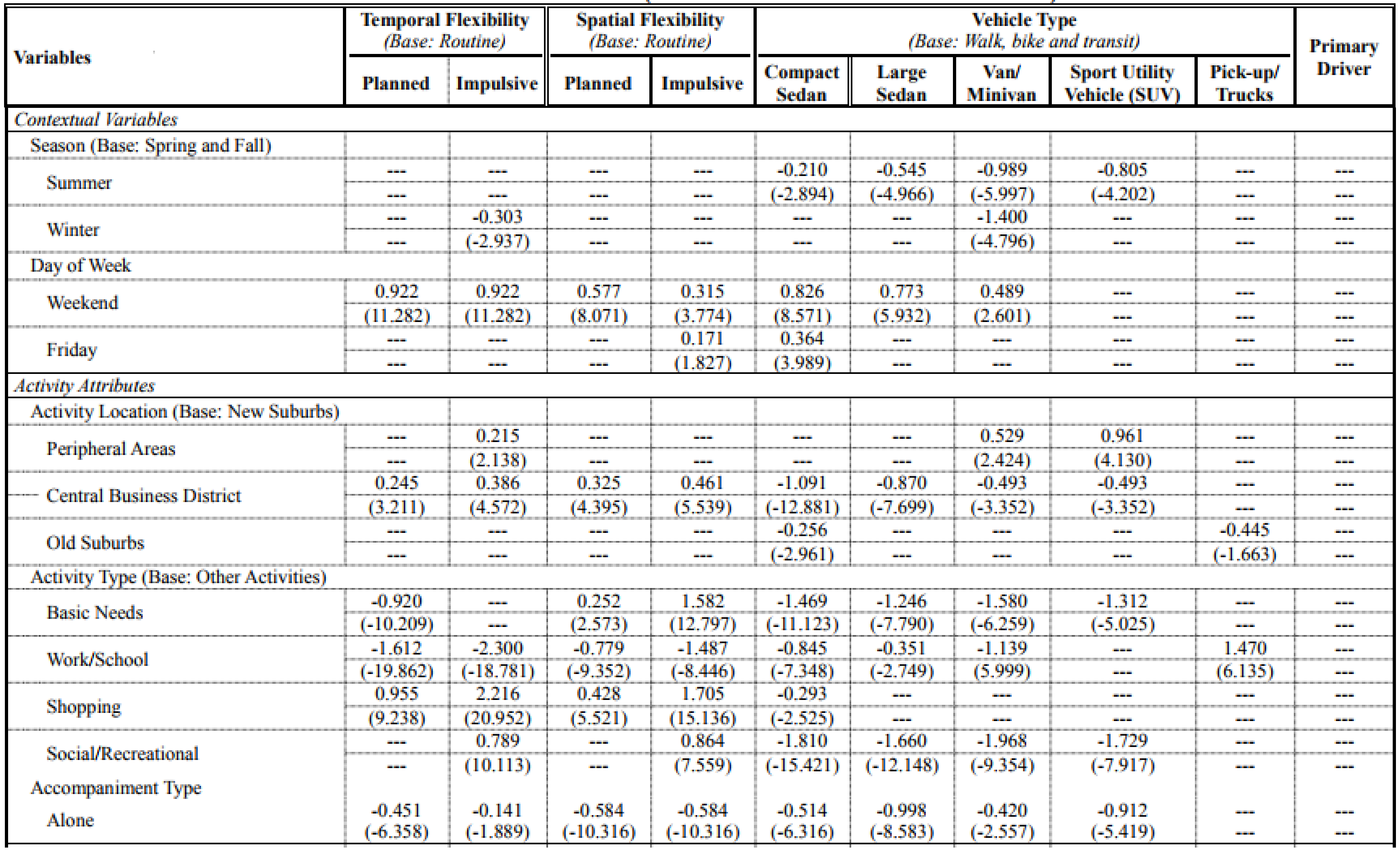
Table 3 Estimation Results (N=378 individuals and 8098 records)



Variables	Temporal Flexibility *(Base: Routine)*		Spatial Flexibility *(Base: Routine)*		Vehicle Type *(Base: Walk, bike and transit)*					Primary Driver
	Planned	Impulsive	Planned	Impulsive	Compact Sedan	Large Sedan	Van/ Minivan	Sport Utility Vehicle (SUV)	Pick-up/ Trucks	
Detached House	0.320	0.379	0.317	---	0.385	0.531	---	-1.306	---	---
	(2.978)	(4.584)	(2.930)	---	(3.830)	(3.985)	---	(-4.308)	---	---
Apartment	---	---	---	---	0.331	---	---	---	---	---
	---	---	---	---	(2.682)	---	---	---	---	---
Income (Base: Low Income (< 20K))										
Medium Income (20K-60K)	-0.539	-0.539	-0.292	---	---	0.448	2.616	---	---	---
	(-5.099)	(-5.099)	(-2.919)	---	---	(3.623)	(6.373)	---	---	---
High Income (> 60K)	-0.525	-0.525	---	---	---	---	1.030	1.371	---	---
	(-4.160)	(-4.160)	---	---	---	---	(2.445)	(5.368)	---	---
Family structure (Base: Single Adult)										
Couples with Children	-0.535	-0.535	-0.347	-0.373	---	-0.381	0.529	-0.596	1.051	---
	(-5.294)	(-5.294)	(-2.545)	(-4.067)	---	(-1.931)	(2.625)	(-2.383)	(4.078)	---
Couples without Children	-0.323	-0.458	-0.335	-0.243	---	-0.807	---	---	---	---
	(-2.661)	(-4.846)	(-2.488)	(-2.660)	---	(-3.994)	---	---	---	---

RESULTS

Table 3 Estimation Results (N=378 individuals and 8098 records)



Variables	Temporal Flexibility *(Base: Routine)*		Spatial Flexibility *(Base: Routine)*		Vehicle Type *(Base: Walk, bike and transit)*					Primary Driver
	Planned	Impulsive	Planned	Impulsive	Compact Sedan	Large Sedan	Van/ Minivan	Sport Utility Vehicle (SUV)	Pick-up/ Trucks	
Contextual Variables										
Season (Base: Spring and Fall)										
Summer	---	---	---	---	-0.210	-0.545	-0.989	-0.805	---	---
	---	---	---	---	(-2.894)	(-4.966)	(-5.997)	(-4.202)	---	---
Winter	---	-0.303	---	---	---	---	-1.400	---	---	---
	---	(-2.937)	---	---	---	---	(-4.796)	---	---	---
Day of Week										
Weekend	0.922	0.922	0.577	0.315	0.826	0.773	0.489	---	---	---
	(11.282)	(11.282)	(8.071)	(3.774)	(8.571)	(5.932)	(2.601)	---	---	---
Friday	---	---	---	0.171	0.364	---	---	---	---	---
	---	---	---	(1.827)	(3.989)	---	---	---	---	---
Activity Attributes										
Activity Location (Base: New Suburbs)										
Peripheral Areas	---	0.215	---	---	---	---	0.529	0.961	---	---
	---	(2.138)	---	---	---	---	(2.424)	(4.130)	---	---
Central Business District	0.245	0.386	0.325	0.461	-1.091	-0.870	-0.493	-0.493	---	---
	(3.211)	(4.572)	(4.395)	(5.539)	(-12.881)	(-7.699)	(-3.352)	(-3.352)	---	---
Old Suburbs	---	---	---	---	-0.256	---	---	---	-0.445	---
	---	---	---	---	(-2.961)	---	---	---	(-1.663)	---
Activity Type (Base: Other Activities)										
Basic Needs	-0.920	---	0.252	1.582	-1.469	-1.246	-1.580	-1.312	---	---
	(-10.209)	---	(2.573)	(12.797)	(-11.123)	(-7.790)	(-6.259)	(-5.025)	---	---
Work/School	-1.612	-2.300	-0.779	-1.487	-0.845	-0.351	-1.139	---	1.470	---
	(-19.862)	(-18.781)	(-9.352)	(-8.446)	(-7.348)	(-2.749)	(5.999)	---	(6.135)	---
Shopping	0.955	2.216	0.428	1.705	-0.293	---	---	---	---	---
	(9.238)	(20.952)	(5.521)	(15.136)	(-2.525)	---	---	---	---	---
Social/Recreational	---	0.789	---	0.864	-1.810	-1.660	-1.968	-1.729	---	---
	---	(10.113)	---	(7.559)	(-15.421)	(-12.148)	(-9.354)	(-7.917)	---	---
Accompaniment Type										
Alone	-0.451	-0.141	-0.584	-0.584	-0.514	-0.998	-0.420	-0.912	---	---
	(-6.358)	(-1.889)	(-10.316)	(-10.316)	(-6.316)	(-8.583)	(-2.557)	(-5.419)	---	---

RESULTS

- The variable effects are considered across dimensions
  - Fairly parsimonious model specification
- Exogenous variable categories
  - Individual and household socio-demographics
  - Household residential location characteristics
  - Activity attributes
  - Contextual variables

RESULTS

- Individual socio-demographics
  - Females are less likely to drive sedans and vans/minivans and more likely to be the primary driver
  - Young individuals tend to undertake impulsive activities while being less inclined to use compact sedans or vans/minivans
  - Seniors are indifferent towards activity flexibility indicators and have a lower preference for compact sedans and SUVs
  - University degree holders prefer vans/minivans

RESULTS

- Household socio-demographics
  - Individuals from medium and high income households tend to perform routinized activities
  - Members from medium income households are more likely to opt for large sedans and vans/minivans
  - Vans/minivans and SUVs are the preferred vehicle type for individuals from affluent households
  - Individuals with kids are disinclined towards pursuing activities planned in a short period of time and tend to use vans/minivans

RESULTS

- Household residential location attributes
  - Residential location categories created using k-means cluster analysis using population density, land use mix and transit accessibility
  - Categories considered
    - Peripheral areas (lowest values of all 3 indices)
    - Old suburbs (medium land use mix and population density and served by main transit lines)
    - New suburbs (low to medium values of the 3 indices)
    - Central Business District (downtown cores with the highest population density, land use mix and transit accessibility)

RESULTS

- Household residential location attributes
  - Individuals living in peripheral areas have a higher propensity of getting involved in impulsive activities (temporal and spatial)
  - These individuals prefer large sedans, vans/minivans, and SUVs for activity participation
  - CBD residents also tend to engage in impulsive activities while choosing not to use sedans and SUVs for travel
    - Overall preference for non-auto oriented travel

RESULTS

- Contextual variables
  - Walking/biking/taking transit is preferred in summer
  - People are disinclined to undertake temporally impulsive activities in winter
  - In winter vans/minivans are less likely to be used
    - Increased heating leading to increased gas cost
    - Snow cleaning and parking difficulty
  - Pre-planned and impulsive activities are pursued in weekends
  - Sedans and vans/minivans are preferred vehicle type choice

RESULTS

- Activity attributes
  - Temporally impulsive activities are more likely to be pursued both in peripheral and Central Business Districts (CBD)
  - Contrasting vehicle type choices between
    - Larger vehicles preferred in peripheral areas; walk/bike/transit in CBDs
    - CBDs have diverse land use mix, increased number of easily accessible activity centres, pedestrian oriented urban form and parking restrictions

RESULTS

- Activity attributes
  - Activities involving basic needs are either routine or impulsive in time while the location is more likely to be pre-planned or selected impulsively
  - Temporal and spatial rigidity of work/school is confirmed
  - Both shopping and social/recreational activities are more likely to be impulsively undertaken
  - Individuals are disinclined to use sedans for shopping – presumably due to the grouped nature of the activity

ORDERED RESPONSE MODELS

ORDERED RESPONSE MODELS

- We examined choice scenarios that involved discrete variables that were unrelated
- In this class, we will examine a different paradigm of modelling for discrete variables that have an inherent ordering within them
- Let's begin with the binary models
- We examined binary models from the utility maximization
- Lets say we have alternatives i and j
  - $U_{in} = V_{in} + \varepsilon_{in}\,; U_{jn} = V_{jn} + \varepsilon_{jn}$
  - $U_{in} - U_{in} = V_{in} - V_{jn} + (\varepsilon_{in} - \varepsilon_{jn})$
- Now alternative i is chosen if $V_{in} - V_{jn} + (\varepsilon_{in} - \varepsilon_{jn}) \leq 0$ and j is chosen if $V_{in} - V_{jn} + (\varepsilon_{in} - \varepsilon_{jn}) > 0$
- This is same as selecting alternative with maximum utility

ORDERED RESPONSE MODELS

- In the ordered response we achieve this in a different fashion
- We use the index function formulation
- i.e. we assume there is a uni-dimensional index function (latent propensity) that determines the choice process
- The propensity is measured for the choice context
- However, there is no way to evaluate the propensity in the population -> so we connect propensity to an observed ordered variable

ORDERED RESPONSE MODELS

- Let us say we have two alternatives 0 and 1 [like a yes/no choice]
- There is a latent propensity for individual to choose either 0 or 1
- We can hypothesize that if the propensity value is >0 the individual chooses 1 and if the propensity is ≤ 0 the individual chooses 0
- It is similar to the utility being higher for the binary case
- The approaches becomes different when we have more than two alternatives

ORDERED RESPONSE MODELS

- Let us consider the following propensity for the individual's choice ($y = 0$ or 1)
- $y^* = \alpha + \beta \mathbf{x} + \varepsilon$
  - $y^* > 0 \Rightarrow y=1;$
  - $y^* \leq 0 \Rightarrow y=0;$
  - where y^* is the latent propensity and y is the observed choice
- Now the probability that $y^* > 0$ is given by
- $\text{Prob}(\alpha + \beta \mathbf{x} + \varepsilon > 0) = \text{Prob}(\varepsilon > -(\alpha + \beta \mathbf{x}))$

  $= 1 - \text{Prob}\ (\varepsilon < (-(\alpha + \beta \mathbf{x})));$

ORDERED RESPONSE MODELS

- So it follows that Prob($\alpha + \beta x + \varepsilon \leq 0$) = Prob ($\varepsilon < (-(\alpha + \beta x))$)
- Let us say ε is standard normally distributed then probability of choosing 1 is $1-\Phi(-(\alpha + \beta x))$ and probability of choosing 0 is $\Phi(-(\alpha + \beta x))$
- This yields the binary probit model (the same one we derived with maximum utility approach)
- Instead of the normal assumption we can assume a standard logistic error assumption to generate the binary logit model

ORDERED RESPONSE MODELS

- We can visualize the OR models as a horizontal partitioning scheme that divides the real line into components (for 0 and 1)
- Now what if we have more categories
  - (y = 0, 1, 2..K)
- The approach is the same, we have one index variable $y^* = \alpha + \beta x + \varepsilon$
- $y = 0$ if $y^* < 0$
- $y = 1$ if $0 < y^* < \psi_1$
- $y = 2$ if $\psi_1 < y^* < \psi_2$
- ….
- $y = K$ if $\psi_{K-2} < y^*$
- The ψ_i represent thresholds to be estimated

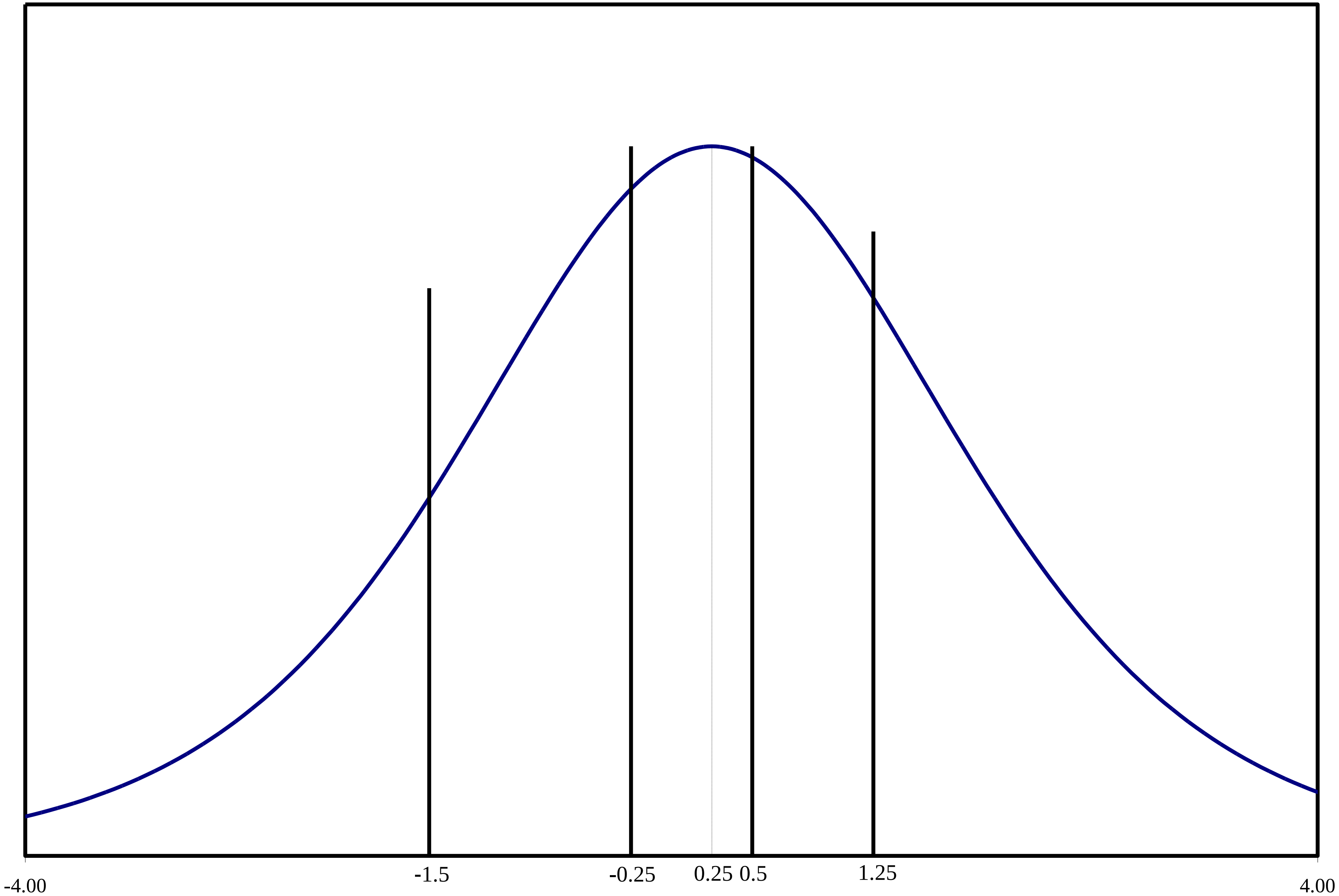
ORDERED RESPONSE MODELS

- The probability expressions are slightly complicated
- $P(y=0) = CDF[-(\alpha + \beta x)]$
- $P(y=1) = CDF[\psi_1 -(\alpha + \beta x)] -CDF[-(\alpha + \beta x)]$
- $P(y=2) = CDF[\psi_2 -(\alpha + \beta x)] -CDF[\psi_1 -(\alpha + \beta x)]$
- …
- $P(y=K) = 1 -CDF[\psi_{K-2} -(\alpha + \beta x)]$

- CDF could be normal or logistic based on your assumption
- The LL function setup and model estimation is exactly same as the MNL models
  - $\mathcal{L}(\beta, \psi) = \sum_{n=1}^{N} \sum_{\forall j} (\delta_{jn} lnP_j)$
    - We just have a different *Pj* term evaluation
    - Important aspect to note, we can either estimate a constant or set the first threshold to 0. We cannot do both

ORDERED RESPONSE MODELS

- Important aspect to note, we cannot estimate alternative specific coefficients in the OR regime
- We only have variables that are generic for all variables
  - i.e. a variable either increases the propensity or reduces the propensity
- Lets illustrate this through a figure
- Consider a propensity function ($y^* = \alpha + \beta x + \varepsilon$)
- If ε is normally distributed y^* will also be normally distributed
- Now if β is positive then the whole curve will move to the right and vice-versa


-4.00
-1.5
-0.25
0.25
0.5
1.25
4.00

COMPARISON WITH MNL

- MNL and OR models yield identical results for binary models
- For more than 2 alternatives they are different
- The utility maximization is in a way multidimensional partitioning scheme
- The MNL allows for effect of regressors to vary across the different alternatives
- Again, in OR scheme we only have one equation to represent behavior, whereas in the MNL scheme we have K-1 equations for utility
- So MNL might offer more as a model
- At the same time OR models are quite parsimonious and easy to estimate and understand

ELASTICITY EFFECTS

- The approach is similar to the multinomial logit models
- Since no alternative specific variables can be estimated no self and cross effects
- Marginal effect (change in probability of alternative i for a change in x) $= \frac{\partial P_{\mathrm{i}}}{\partial \mathrm{x}_k}$
- For ordered probit - alternative 1
  - $\frac{\partial P_0}{\partial \mathrm{x}_k} = \frac{\partial(\Phi[-(\alpha + \beta \mathrm{x})])}{\partial \mathrm{x}_n} = \phi[-(\alpha + \beta\mathrm{x})] * \beta_{\mathrm{k}}$
    - where Φ is the CDF function and ϕ is the pdf function of the standard normal distribution
    - Similarly marginal effects for other alternatives can be computed
- Elasticity effects - $\frac{\partial P_{\mathrm{i}}}{\mathrm{P}_{\mathrm{i}}} \Big/ \frac{\partial \mathrm{x}_k}{\mathrm{x}_k}$ - for computing the elasticity effects

COUNT MODELS

COUNT MODELS

- In the event of measuring
  - Travel trips
  - Traffic flows
  - Bicycle flows at intersections
  - No. of hospital visits in a year
  - Recreational travel visits in a year
- Potential approaches from our class we discussed so far?
  - Linear regression
  - Ordered response models
- Issues?
  - Regression assumes a continuous distribution which is not the case in count events
  - Ordered response models are suited only for groupings or bins rather than for every possible number

BIG PICTURE

- Now we have count data (i.e. dependent variable is counts)
- Our objective is to understand the relation between counts and the various variables related to the dependent variable
- For example
  - If we want to model bicycle flows at an intersection – we will try the impact of bicycle facility, proximity to downtown, land-use and transit access etc. as measures that affect the flows
  - Now, the objective of this exercise is to be able to replicate the observed flows through our model
  - How do we do that?
- Lets say for example we observed – 212 bicycle flows at an intersection
- The bicycle flows at an intersection can vary from 0 – 500 i.e. there is probability that 501 events could occur

BIG PICTURE

- We will try to maximize the probability for the chosen alternative (or the alternative we observed)
- So, we employ Maximum Likelihood such that Pr(212) is maximized
- Please note that because of the huge number of potential events the discrete approaches we used so far are not likely to be easily employed
  - Imagine using MNL for the 501 events for instance
- Hence we move to a different class of models often referred to as count models

COUNT MODELS - POISSON REGRESSION MODEL

- Poisson distribution
  - Pr[Y=y] = $^{e^{-\mu}\mu^{y}}\!/_{y!}$, y = 0, 1, 2, …,
  - μ is the intensity or rate parameter
  - μ represents the Mean and Variance of the distribution
- The expression allows us to model probability of each count for individual record
- Now how do incorporate the exogenous variables
- We do that by parameterizing μ (intensity or rate parameter)
- μ = exp(βx)
- Now the probability expression can be substituted with μ.

COUNT MODELS - POISSON REGRESSION MODEL

- The log-likelihood expression is given as
- L = Ln(Pr[Y=y]) = ln($^{e^{-\mu}\mu^{y}}/_{y!}$)
- = ln($e^{-\mu}\mu^{y}$) – ln(y!) = - μ + y ln(μ)-ln(y!)
- substitute μ = exp(βx) => = -exp(βx) +yβx - ln(y!)
- When we are trying to Maximize the function ln(y!) is a constant for every individual and hence can be dropped from the Log-likelihood for estimation purposes is:
- L = yβx -exp(βx)
- Readily available in most statistical software
- Same iterative process
- LL is used to determine whether the variables are significant or not (similar to discrete choice models)

COUNT MODELS - POISSON REGRESSION MODEL

- Model interpretations
- Quite simple to understand: we set the mean to be a function of regressors (μ = exp(βx)) and estimate the model
- To look at elasticity of mean
- $\frac{\partial \mu}{\partial x_j} = \beta_j \exp(\beta x)$
- So if the parameter coefficient is +ive it has a positive effect on the mean
- This relationship implies that say for 2 variables $\beta 1$ and $\beta 2$ and say $\beta 1 / \beta 2 = 4$ then effect of $\beta 1$ on μ will be 4 times that of $\beta 2$

COUNT MODELS - POISSON REGRESSION MODEL

- There is an implicit assumption within the assumption of employing the Poisson model
  - Mean and Variance of the distribution are same
  - This is often violated in the data
- When the variance > mean then data set is referred to have over-dispersion
- When variance < mean, the data has under-dispersion
- In both these cases the implicit assumption in Poisson model is violated and hence does not suit our needs

COUNT MODELS - NEGATIVE BINOMIAL REGRESSION MODEL

- If the distribution under examination does not have the same mean and variance then an approach to modelling such counts is Negative Binomial Model
- In this model, in addition to the μ we will also estimate another parameter
- The mean = μ; variance = $\mu(1+ \alpha\mu)$
- Even in this model μ = exp(βx) is used to examine the effect of various exogenous parameters
- In this model the variance has a quadratic term $\mu+ \alpha\mu^2$
- This is referred to as NB2 model - most commonly used model
- The pdf function for [Y=y] = $\frac{\Gamma(\alpha^{-1}+y)}{\Gamma(\alpha^{-1})\Gamma(y+1)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\mu+\alpha^{-1}}\right)^{y}$
- LL will be written based on the above pdf

COUNT MODELS – EXCESS ZEROES

- One approach to handle the problem with Poisson models is to account for too many 0s
- If your data has too many 0s, it is very unlikely that mean and variance are same
- Hence, we will try to model this scenario using different forms of Poisson models
  - Hurdle models
  - Zero-inflated models
- The hurdle models consider that the behavior behind the 0s and non 0s is quite different and needs to be explicitly considered
- The Zero-inflated model accommodates the same thing in a slightly different way

COUNT MODELS – HURDLE MODELS

- Zeros are determined by f1(.) and non-zeros through f2(.)
- Pr[Y=0] = f1(0)
- Pr[Y>0] = f2(Y|Y>0) = f2(Y) / (1-f2(0))
- To make sure the probabilities sum to 1 we also multiply Pr[Y>0] with (1-f1(0))
- To summarize
- g(Y) = $\begin{matrix} f1(0) \quad if\ Y = 0 \\ \frac{1-f1(0)}{1-f2(0)} f2(Y) \quad if\ Y > 0 \end{matrix}$
- Now we set μ_1 = exp(β_1x) and μ_2 = exp(β_2x)
- Write the new LL - Two terms
  - term for Y=0 and term for Y>0
- In the example we are discussing we are considering *f* to be Poisson or NB distribution, the models we examined will work any other distribution also

COUNT MODELS – ZERO-INFLATED MODELS

- This model takes a slightly different approach to the modeling 0s
  - Binary process f1(.) (logit model)
  - Count process f2(.) (poisson/NB model)
- g(Y) = $$\begin{array}{ll} f1(0)+\left(1-f1(0)\right)f2(0) & if\ Y=0 \\ \left(1-f1(0)\right)f2(Y) & if\ Y>0 \end{array}$$
- This process involves two terms in the LL
  - Similar to the hurdle models

REFERENCES

- Ben-Akiva, M. E. and S. R. Lerman (1985). Discrete choice analysis: theory and application to travel demand, The MIT Press.
- Ortuzar, J. D. and L. G. Willumsen (2011). Modelling transport, Wiley. 4th ed.