A Joint Tour-Based Model of Vehicle Type Choice and Tour Length

Ram M. Pendyala
School of Sustainable Engineering & the Built Environment
Arizona State University
Tempe, AZ
Introduction

• Growing interest in microsimulation approaches to travel demand modeling and forecasting
  • Simulate activity-travel choices recognizing the inter-dependence among activities, trips, and persons subject to different spatial, temporal and resource constraints
  • Based on activity-based or tour-based paradigms
Tour-based Microsimulation Models

- The basic unit of analysis is a tour
  - Account for inter-dependency of trips within a tour
- Most tour-based models consider (to varying degrees) several basic tour dimensions
  - primary activity type, number of stops and stop locations, sequencing of stops, mode choice at tour- and stop-level
  - some tour-based models also consider intra-household interactions in simulating tour characteristics
Tour-based Microsimulation Models (continued)

- Critical choice dimension often missing is “choice of vehicle type to undertake the tour”
- Vehicle type choice and distance travelled (location choices) are key factors in determining energy consumption and greenhouse gas (GHG) emissions
- Objective of this study to study inter-relationship between these two choice dimensions
Vehicle Type Choice and Tour Length

- Simultaneously model discrete and continuous choice dimensions to account for common unobserved factors
- Also determine nature of relationship between the two choices
- Consider alternative possibilities
  - Do travelers try to economize by driving smaller fuel-efficient car for longer tours?
  - Do travelers use the larger and less fuel-efficient vehicle for longer tours because it is more comfortable?
  - Do travelers choose locations close by (reduce distance traveled) when driving their gas guzzler in an attempt to economize?
  
  and so on...
Policy Implications of Choice Relationships

- Need to determine nature of relationship between vehicle type choice and tour length
  - Does the vehicle type choice affect tour length?
  - Does the tour length affect vehicle type choice?
  - Do both model specifications provide equally plausible results?
- In general, cost of driving large vehicle short distances is about equal to cost of driving small vehicle long distances
  - If land use density is increased, then people can access destinations by traveling short distances
  - This may reduce VMT, but does it reduce emissions and energy consumption?
  - Depends on the vehicle type chosen to undertake tours
Modeling Methodology

- A probit-based joint discrete continuous modelling approach was employed (Ye and Pendyala 2009)
  - simultaneously models multiple dimensions
  - accommodates potential error correlations
- Advantages over other discrete-continuous frameworks
  - does not make any distributional transformations
  - assumes a multivariate normal distribution
Model Formulation

- The system of equations can be formulated as:

\[
\begin{align*}
    u_1^* &= x_1 \beta_1 + \delta_1 d + \varepsilon_1 \\
    u_2^* &= x_2 \beta_2 + \delta_2 d + \varepsilon_2 \\
    u_3^* &= x_3 \beta_3 + \varepsilon_3 \\
    d &= z\theta + \lambda_1 y_1 + \lambda_2 y_2 + \omega
\end{align*}
\]

For continuous variable:
- \(d\) - continuous choice variable
- \(z\) - explanatory variables
- \(\theta\) - coefficient vector for \(z\)

For discrete choice \(i\):
- \(u_i^*\) - latent utility function
- \(x_i\) - exogenous variables
- \(\beta_i\) - coefficient vector for \(x_i\)
- \(\delta_i\) - coefficient of the continuous choice variable \(d\)
- \(y_i\) - indicator variable
- \(\lambda_i\) - coefficient of \(y_i\)
Model Formulation (continued)

• Indicator variables for choice alternatives are defined as follows:

\[ y_1 = I(u_1^* > u_2^* \text{ and } u_1^* > u_3^*) \]
\[ y_2 = I(u_2^* > u_1^* \text{ and } u_2^* > u_3^*) \]

For discrete choice \( i \): 
\( u_i^* \) - latent utility function
\( y_i \) - indicator variable
Model Formulation (continued)

• For the model to be identified, either $\lambda$ or $\delta$ parameters must be restricted to zero

1. $\lambda = 0 \rightarrow$ model specification where continuous variable affects the discrete choice
2. $\delta = 0 \rightarrow$ model specification where discrete choice affects the continuous variable

\[
\begin{align*}
  u_1^* &= x_1 \beta_1 + \delta_1 d + \varepsilon_1 \\
  u_2^* &= x_2 \beta_2 + \delta_2 d + \varepsilon_2 \\
  u_3^* &= x_3 \beta_3 + \varepsilon_3 \\
  d &= z\theta + \lambda_1 y_1 + \lambda_2 y_2 + \omega
\end{align*}
\]
Model Formulation (continued)

• The error terms are assumed to be multivariate normally distributed with a variance-covariance matrix:

\[ \Sigma = \begin{bmatrix}
1 & 0 & 0 & \gamma_1 \\
0 & 1 & 0 & \gamma_2 \\
0 & 0 & 1 & \gamma_3 \\
\gamma_1 & \gamma_2 & \gamma_3 & \sigma^2
\end{bmatrix} \]

\( \gamma_i \) - covariance between \( \varepsilon_i \) and \( \omega \)
\( \sigma^2 \) - variance of \( \omega \)
Model Formulation: Varying Choice Sets

- Original model formulation applicable to situation with constant choice set across decision makers
- However, choice set may vary
  - Vehicle fleet composition varies across households
- Model formulation needs to be modified to accommodate varying choice sets
Model Formulation: Varying Choice Sets

• The utility expressions are modified as:

\[
\begin{align*}
    u_1^* &= k_1 x_1 \beta_1 + (1 - k_1)(-\infty) + \delta_1 d + \varepsilon_1 \\
    u_2^* &= k_2 x_2 \beta_2 + (1 - k_2)(-\infty) + \delta_2 d + \varepsilon_2 \\
    u_3^* &= k_3 x_3 \beta_3 + (1 - k_3)(-\infty) + \varepsilon_3 \\
    d &= z\theta + \lambda_1 y_1 + \lambda_2 y_2 + \omega
\end{align*}
\]

• If a choice alternative is present \(\rightarrow\) utility equation is the same as before

• If a choice alternative is absent \(\rightarrow\) the alternative is made unattractive by making the utility a large negative value

\(k_i\) - indicator variable for presence of alternative \(i\) in the choice set
Non-nested Hypothesis Test

- Two different model specifications based on whether parameter $\lambda$ or $\delta$ are set to zero
- More often than not, both structures provide behaviorally plausible results with similar goodness-of-fit measures
- Need a rigorous statistical hypothesis test to compare alternative specifications
Non-nested Hypothesis Test (continued)

- Likelihood ratio tests cannot be applied when the models are non-nested
- Previous literature on non-nested hypothesis test:
  - Cox (1961, 1962) proposed the first test
  - Horowitz (1983) presented a compact form of the Cox test for comparing non-nested discrete choice models
  - Ben-Akiva and Swait (1984) modified the test to accommodate comparison of standard goodness-of-fit statistics obtained from discrete choice model estimation results
Non-nested Hypothesis Test (continued)

- Ben-Akiva and Swait (1984)
  - Test was initially proposed to compare single equation model systems
  - Applicability for comparing simultaneous equations model systems is unknown
- A new test was proposed by Ye and Pendyala (2008) for comparing joint simultaneous model systems
- This was further modified to accommodate varying choice sets across decision makers
Non-nested Hypothesis Test (continued)

- According to Horowitz, the probability that goodness-of-fit statistic for a model B is greater than goodness-of-fit statistic for model A by a value $t > 0$, assuming model A is the true model is given by:

\[
\Pr \left[ \beta_B^2 - \beta_A^2 > t \right] \leq \Phi\left[-\sqrt{-2L^*} \right]
\]

- The likelihood ratio index for model $m$ is calculated as:

\[
\beta_m^2 = 1 - \frac{L_m - K_m}{2L^*}
\]

$L^* = \text{log-likelihood function value of model } m \text{ when all the parameters are assumed to be zero}$

$L_m = \text{log-likelihood value at convergence}$

$K_m = \text{number of parameters estimated}$
Non-nested Hypothesis Test (continued)

- In the original formulation, $L^*$ was defined as $-N \ln(J)$
- Modified $L^*$ to accommodate comparison of simultaneous model systems

$$L^*(\text{Joint Model}) = L^*(\text{Continuous Model}) + L^*(\text{Discrete Model})$$

where,

$$L^*(\text{Continuous Model}) = -\frac{N - 1}{2} - N \ln\left(\sqrt{2\pi} \hat{\sigma}\right)$$

$$L^*(\text{Discrete Model}) = -N \ln(J)$$

$N$ - number of observations
$J$ - number of choice alternatives
$\sigma$ - std. dev. of continuous variable
Non-nested Hypothesis Test (continued)

• To accommodate varying choice sets across observations:

\[ L^*(\text{Discrete Model}) = - \sum_{i=1}^{N} \ln(j_i) \]

\( j_i \) - number of choice alternatives for decision maker \( i \)

• Then, modified \( L^* \) for joint model with varying choice sets:

\[ L^*(\text{JointModel}) = - \frac{N-1}{2} - N \ln\left(\sqrt{2\pi} \sigma^\hat{}\right) - \sum_{i=1}^{N} \ln(j_i) \]

• Modified non-nested test:

\[ \Pr\left[ \frac{\bar{\rho}_B - \bar{\rho}_A}{\text{SE}} > t \right] \leq \Phi\left[-\sqrt{-2\left(- \frac{N-1}{2} - N \ln\left(\sqrt{2\pi} \sigma^\hat{}\right) + \sum_{i=1}^{N} \ln(j_i)\right)t}\right] \]
Data Description

- 2008-2009 National Household Travel Survey (NHTS) dataset
- Criteria to extract tours from travel diary records
  - Anchor locations - home or work
  - Same vehicle used on the entire tour
  - Only considered tours in households with multiple vehicles of different body types
  - Only considered pure auto tours
- The process resulted in a total of 102352 tours
  - generated by 64568 persons in 37938 households
  - average number of tours per person=1.59
  - average number of tours per household=2.70
Trip Chain Attributes: Tour Type

- Home-based Simple Work
- Home-based Complex Work
- Home-based Simple Non Work
- Home-based Complex Non Work
- Work-based
Data Description (continued)

- HBNW tours are of particular interest in the study
  - offer more flexibility in the choice of vehicle type and destination locations (tour length)
- A total of 66030 HBNW tours were selected
- A random subsample of just under 10 percent (6478) selected for the analysis
  - to avoid inflated t-statistics due to large sample sizes
  - for computational efficiency in simulation-based estimation
Tour Characteristics by Vehicle Type Chosen

Ignoring fleet composition variation across households

<table>
<thead>
<tr>
<th>Vehicle Body Type</th>
<th>Percentage</th>
<th>Tour Distance</th>
<th>Tour Travel Time</th>
<th>Number of passengers on Tour</th>
<th>Number of Stops on Tour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>41.9</td>
<td>16.0</td>
<td>37.7</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>Van</td>
<td>14.1</td>
<td>15.2</td>
<td>37.0</td>
<td>2.1</td>
<td>1.8</td>
</tr>
<tr>
<td>SUV</td>
<td>25.4</td>
<td>15.4</td>
<td>36.0</td>
<td>1.8</td>
<td>1.7</td>
</tr>
<tr>
<td>Pickup</td>
<td>18.6</td>
<td>15.6</td>
<td>36.5</td>
<td>1.5</td>
<td>1.6</td>
</tr>
</tbody>
</table>
## Tour Characteristics by Vehicle Type of Tour (continued)

### Considering the fleet composition

<table>
<thead>
<tr>
<th>Household Vehicle Fleet Composition by Body Type</th>
<th>Vehicle Body Type Selected for Tour</th>
<th>Percentage</th>
<th>Tour Distance</th>
<th>Tour Travel Time</th>
<th>Number of passengers on Tour</th>
<th>Number of Stops on Tour</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUV, Pickup</td>
<td>SUV</td>
<td>65.1</td>
<td>17.0</td>
<td>37.4</td>
<td>1.9</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>Pickup</td>
<td>34.9</td>
<td>15.6</td>
<td>37.4</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>Van, Pickup</td>
<td>Van</td>
<td>60.4</td>
<td>14.4</td>
<td>35.9</td>
<td>2.0</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>Pickup</td>
<td>39.6</td>
<td>15.9</td>
<td>37.5</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>Van, SUV</td>
<td>Van</td>
<td>56.8</td>
<td>17.2</td>
<td>39.6</td>
<td>2.1</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>SUV</td>
<td>43.2</td>
<td>16.4</td>
<td>40.3</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Van, SUV, Pickup</td>
<td>Van</td>
<td>39.4</td>
<td>15.4</td>
<td>31.9</td>
<td>1.9</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>SUV</td>
<td>43.7</td>
<td>17.9</td>
<td>38.4</td>
<td>1.9</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Pickup</td>
<td>16.9</td>
<td>17.4</td>
<td>61.4</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Car, Pickup</td>
<td>Car</td>
<td>64.5</td>
<td>17.1</td>
<td>39.3</td>
<td>1.6</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>Pickup</td>
<td>35.5</td>
<td>15.4</td>
<td>35.8</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>Car, SUV</td>
<td>Car</td>
<td>48.2</td>
<td>14.3</td>
<td>36.1</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>SUV</td>
<td>51.8</td>
<td>14.1</td>
<td>34.3</td>
<td>1.7</td>
<td>1.6</td>
</tr>
</tbody>
</table>
### Tour Characteristics by Vehicle Type of Tour (continued)

**Considering the fleet composition**

<table>
<thead>
<tr>
<th>Household Vehicle Fleet Composition by Body Type</th>
<th>Vehicle Body Type Selected for Tour</th>
<th>Frequency</th>
<th>Tour Distance</th>
<th>Tour Travel Time</th>
<th>Number of passengers on Tour</th>
<th>Number of Stops on Tour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car, SUV, Pickup</td>
<td>Car</td>
<td>34.1</td>
<td>16.6</td>
<td>36.9</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>SUV</td>
<td>42.0</td>
<td>16.0</td>
<td>37.4</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>Pickup</td>
<td>23.9</td>
<td>16.9</td>
<td>36.9</td>
<td>1.4</td>
<td>1.6</td>
</tr>
<tr>
<td>Car, Van</td>
<td>Car</td>
<td>46.6</td>
<td>15.2</td>
<td>36.5</td>
<td>1.7</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Van</td>
<td>53.4</td>
<td>15.0</td>
<td>37.5</td>
<td>2.1</td>
<td>1.8</td>
</tr>
<tr>
<td>Car, Van, Pickup</td>
<td>Car</td>
<td>39.3</td>
<td>15.8</td>
<td>35.6</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Van</td>
<td>40.5</td>
<td>17.0</td>
<td>38.9</td>
<td>2.1</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>Pickup</td>
<td>20.2</td>
<td>14.0</td>
<td>33.5</td>
<td>1.6</td>
<td>1.4</td>
</tr>
<tr>
<td>Car, Van, SUV</td>
<td>Car</td>
<td>32.9</td>
<td>17.2</td>
<td>41.2</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Van</td>
<td>35.0</td>
<td>11.1</td>
<td>28.8</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>SUV</td>
<td>32.2</td>
<td>15.4</td>
<td>36.4</td>
<td>1.5</td>
<td>1.9</td>
</tr>
<tr>
<td>Car, Van, SUV, Pickup</td>
<td>Car</td>
<td>22.0</td>
<td>20.3</td>
<td>42.4</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Van</td>
<td>24.0</td>
<td>21.7</td>
<td>43.8</td>
<td>2.3</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>SUV</td>
<td>34.0</td>
<td>18.3</td>
<td>40.4</td>
<td>2.1</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>Pickup</td>
<td>20.0</td>
<td>15.0</td>
<td>33.6</td>
<td>1.4</td>
<td>1.6</td>
</tr>
</tbody>
</table>
**Model Estimation**

- Estimated probit-based joint discrete-continuous models and independent models (that ignore error correlations)
- **Vehicle type choice**
  - Multinomial probit model
  - Four categories, namely, car, van, SUV, pickup truck (base)
- **Tour length**
  - log-linear regression model
- **Both model specifications were explored**
  - Vehicle type choice affects tour length
  - Tour length affects vehicle type choice
Model Estimation (continued)

- Following independent variables included
  - Tour characteristics
    - Number of stops (to represent tour complexity)
    - Tour accompaniment (solo or joint)
  - Note that tour characteristics may not be exogenous
    - Representation of additional heterogeneity remains a future task
    - Multi-dimensional choice model systems complex to specify and estimate, although progress being made
  - Socioeconomic variables
Non-nested Hypothesis Test

• Both the model specifications provided plausible results
• Non-nested test was used to identify the appropriate model specification
  • Test indicated that the model specification where vehicle type choice affects tour length offered a better goodness-of-fit
  • Probability that this specification is erroneously chosen as the superior one was very small (0.007)
Results

- Results suggest that vehicle type choice affects tour length
- Finding consistent with notion of a temporal hierarchy to these choice dimensions
  - Vehicle type choice → a longer-term or medium-term choice dimension
  - Tour length choice → based on location choices for stops on tour, which constitute short-term choices
    - Function of vehicle type available/chosen, other tour attributes, socio-economic factors
## Results: Impact of Tour Attributes

Vehicle type affects tour length

<table>
<thead>
<tr>
<th></th>
<th>Joint Vehicle Type Choice Model</th>
<th>Joint Tour Length Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car</td>
<td>Van</td>
</tr>
<tr>
<td>Constant</td>
<td>1.491</td>
<td>1.625</td>
</tr>
<tr>
<td>Vehicle Type is Car</td>
<td>0.143</td>
<td></td>
</tr>
<tr>
<td>Vehicle Type is Van</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle Type is SUV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than one stop</td>
<td>0.290</td>
<td>-0.728</td>
</tr>
<tr>
<td>Solo tour</td>
<td>-0.055</td>
<td>0.234</td>
</tr>
<tr>
<td>Joint tour</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Results: Impact of Tour Attributes (continued)

<table>
<thead>
<tr>
<th>Tour length affects vehicle type</th>
<th>Joint Vehicle Type Choice Model</th>
<th>Joint Tour Length Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car</td>
<td>Van</td>
</tr>
<tr>
<td>Constant</td>
<td>1.259</td>
<td>6.5</td>
</tr>
<tr>
<td>Log of tour length in miles</td>
<td>0.093</td>
<td>1.3</td>
</tr>
<tr>
<td>More than one stop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solo tour</td>
<td>-0.267</td>
<td>-3.7</td>
</tr>
<tr>
<td>Joint tour</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Results: Impact of Socio-economic Attributes

<table>
<thead>
<tr>
<th>Vehicle type affects tour length</th>
<th>Joint Vehicle Type Choice Model</th>
<th>Joint Tour Length Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car</td>
<td>Van</td>
</tr>
<tr>
<td>Ratio of household size to number of vehicles</td>
<td>Coef</td>
<td>Coef</td>
</tr>
<tr>
<td>Male</td>
<td>-1.474</td>
<td>-21.4</td>
</tr>
<tr>
<td>Age 65 years or older</td>
<td>0.185</td>
<td>1.6</td>
</tr>
<tr>
<td>Number of children</td>
<td>-0.077</td>
<td>-2.8</td>
</tr>
<tr>
<td>Household in non-urban area</td>
<td>-0.179</td>
<td>-1.9</td>
</tr>
<tr>
<td>Education level (at least college)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can change start time of fixed activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household income less than $40k per year</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Results: Impact of Socio-economic Attributes (continued)

### Joint Vehicle Type Choice Model

<table>
<thead>
<tr>
<th>Tour length affects vehicle type</th>
<th>Joint Vehicle Type Choice Model</th>
<th>Joint Tour Length Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car</td>
<td>Van</td>
</tr>
<tr>
<td>Ratio of household to number of vehicles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-1.474</td>
<td>-21.4</td>
</tr>
<tr>
<td>Age 65 years or older</td>
<td>0.170</td>
<td>1.5</td>
</tr>
<tr>
<td>Number of children</td>
<td>-0.073</td>
<td>-2.5</td>
</tr>
<tr>
<td>Household in non-urban area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education level (atleast college)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can change start time of fixed activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household income less than $40k per year</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results: Significance of Error Covariance

- Considerable differences in model estimates between independent models and joint models
- Error covariance between van type choice and tour length was significant (value=-0.18, p-value=-2.6)
- Significant error covariance found only in specification where vehicle type choice affects tour length
- Behaviorally intuitive - unobserved factors that positively contribute to choice of van (e.g., household obligations, children) negatively contribute to tour length
## Results: Joint Models vs Independent Models

### Vehicle type affects tour length

<table>
<thead>
<tr>
<th></th>
<th>Independent Vehicle Type Choice Model</th>
<th>Independent Tour Length Model</th>
<th>Joint Vehicle Type Choice Model</th>
<th>Joint Tour Length Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Van</td>
<td></td>
<td>Van</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>Coef 2.045, t-stat 14.5</td>
<td>Coef 1.853, t-stat 30.8</td>
<td>Coef 1.625, t-stat 14.2</td>
<td>Coef 1.794, t-stat 24.2</td>
</tr>
</tbody>
</table>

### Tour Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Coef</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle Type is Car</td>
<td>0.079</td>
<td>2.3</td>
</tr>
<tr>
<td>Vehicle Type is Van</td>
<td>0.042</td>
<td>1.0</td>
</tr>
<tr>
<td>Vehicle Type is SUV</td>
<td>0.044</td>
<td>1.2</td>
</tr>
<tr>
<td>More than one stop</td>
<td>0.200</td>
<td>1.9</td>
</tr>
<tr>
<td>Solo tour</td>
<td>-0.917</td>
<td>-7.2</td>
</tr>
<tr>
<td>Joint tour</td>
<td>0.234</td>
<td>6.7</td>
</tr>
</tbody>
</table>

### Socio-economic Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Coef</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of household to number of vehicles</td>
<td>-0.058</td>
<td>-1.7</td>
</tr>
<tr>
<td>Male</td>
<td>-2.289</td>
<td>-18.3</td>
</tr>
<tr>
<td>Age 65 years or older</td>
<td>0.231</td>
<td>1.7</td>
</tr>
<tr>
<td>Number of children</td>
<td>0.146</td>
<td>3.0</td>
</tr>
<tr>
<td>Household in non-urban area</td>
<td>-0.227</td>
<td>-2.0</td>
</tr>
<tr>
<td>Education level (at least college)</td>
<td>0.047</td>
<td>1.8</td>
</tr>
<tr>
<td>Can change start time of fixed activities</td>
<td>-0.105</td>
<td>-3.1</td>
</tr>
<tr>
<td>Household income less than $40k per year</td>
<td>-0.066</td>
<td>-2.2</td>
</tr>
</tbody>
</table>
Summary

- Probit-based methodology for discrete-continuous model estimation
- Significant endogeneity in relationship between vehicle type choice and tour length
- Differences in model estimates between independent models and joint models
  - Significant error covariance pointing to need for modeling choice dimensions jointly
- Considered two model specifications and identified statistically superior model structure
  - Vehicle type choice affects tour length → presence of a temporal hierarchy
Conclusions

- If SUV is in vehicle fleet, it is the preferred vehicle type but SUV tour lengths are shorter
- Tour attributes and socio-economic variables impact the vehicle type choice and tour lengths
- Interesting policy implications
  - Incentives and rebates to entice consumers to purchase small fuel-efficient cars
  - But model shows that tours undertaken by smaller vehicles (cars) are longer in length
  - Any energy and environmental benefits from use of small cars may be partially offset by longer tour lengths (more miles of travel)
  - Consumers exercising trade-off in ownership and use of vehicle types
Model Formulation (continued)

• The notation is simplified into a deterministic component and an error component

\[
\begin{align*}
    u_1^* &= V_1 + \varepsilon_1 \\
    u_2^* &= V_2 + \varepsilon_2 \\
    u_3^* &= V_3 + \varepsilon_3 \\
    d &= U + \gamma_1 \varepsilon_1 + \gamma_2 \varepsilon_2 + \gamma_3 \varepsilon_3 + \sigma' \zeta
\end{align*}
\]

Eq. 4

• The error term ($\omega$) is parameterized as a linear combination of $\varepsilon_1^*$ and $\zeta$ where $\sigma'$ is assumed equal to ($\sigma^2 - \gamma_1^2 - \gamma_2^2 - \gamma_3^2$)
Model Formulation (continued)

The joint discrete-continuous probability for selecting a choice alternative conditional on $\varepsilon_i$:

$$
\Pr (y_1 = 1, d| \varepsilon_1) = \Pr (u_1^* > u_2^*, u_1^* > u_3^*, d| \varepsilon_1)
$$

$$
= \Pr (\varepsilon_2 < V_{12} + \varepsilon_1, \varepsilon_3 < V_{13} + \varepsilon_1, d| \varepsilon_1)
$$

$$
= \Pr (\varepsilon_2 < V_{12} + \varepsilon_1, \varepsilon_3 < V_{13} + \varepsilon_1| \varepsilon_1) \times
$$

$$
\Pr [(d| \varepsilon_2 < V_{12} + \varepsilon_1, \varepsilon_3 < V_{13} + \varepsilon_1)| \varepsilon_1]
$$

$$
= [\Phi (V_{12} + \varepsilon_1) \Phi (V_{13} + \varepsilon_1)] \times
$$

$$
\left\{ \frac{1}{\sigma'} \phi \left( \frac{d - U - \gamma_1 \varepsilon_1 - \gamma_2 \varepsilon_2 - \gamma_3 \varepsilon_3}{\sigma'} \right) \right\}_{\varepsilon_2 < V_{12} + \varepsilon_1, \varepsilon_3 < V_{13} + \varepsilon_1}
$$

$V_{12}$ - difference in deterministic components ($V_1 - V_2$)

$\varphi$ - probability density function

$\Phi$ - cumulative probability density function

Eq. 5
The unconditional probability can be derived by integrating over the distributional domains of $\varepsilon_i$:

- $\varepsilon_1$ extends from $-\infty$ to $+\infty$
- $\varepsilon_2$ extends from $-\infty$ to $V_{12} + \varepsilon_1$
- $\varepsilon_3$ extends from $-\infty$ to $V_{13} + \varepsilon_1$

The probability does not have a closed form solution.

Simulation based techniques must be employed.
Model Formulation (continued)

- **Simulation approach**
  - draw $\varepsilon_1$ from a standard normal distribution
  - let $\varepsilon_2 = \Phi^{-1}[u_2 \Phi (V_{12} + \varepsilon_1)]$ and $\varepsilon_3 = \Phi^{-1}[u_3 \Phi (V_{13} + \varepsilon_1)]$ where $u_2$ and $u_3$ are draws from a standard uniform distribution
  - repeat the process $R$ times
  - The unconditional probability may be approximated as:

$$
\Pr(y_1 = 1, d) \approx \left\{ \sum_{r=1}^{R} \Phi(V_{12} + \varepsilon_{1r}) \Phi(V_{13} + \varepsilon_{1r}) \frac{1}{\sigma'} \phi\left( \frac{d - U - \gamma_1 \varepsilon_{1r} - \gamma_2 \varepsilon_{2r} - \gamma_3 \varepsilon_{3r}}{\sigma'} \right) \right\} / R
$$

---

**Eq. 6**
Normalization

• The joint discrete continuous system can be expressed in terms of differences in choice utilities as:

\[
\begin{align*}
    u_1^* - u_3^* &= V_1 - V_3 + \varepsilon_1 - \varepsilon_3 \\
    u_2^* - u_3^* &= V_2 - V_3 + \varepsilon_2 - \varepsilon_3 \\
    d &= U + \omega
\end{align*}
\]

Eq. 7

• The error covariance matrix can then be expressed as:

\[
\begin{bmatrix}
    2 & 1 & \gamma_1 - \gamma_3 \\
    1 & 2 & \gamma_2 - \gamma_3 \\
    \gamma_1 - \gamma_3 & \gamma_2 - \gamma_3 & \gamma_1^2 + \gamma_2^2 + \gamma_3^2 + \sigma^2
\end{bmatrix}
\]

Eq. 8
Normalization (continued)

- The normalization equations can be written as:

\[
\begin{align*}
\gamma_1 - \gamma_3 &= \gamma_{1N} - \gamma_{3N} \\
\gamma_2 - \gamma_3 &= \gamma_{2N} - \gamma_{3N} \\
\gamma_1^2 + \gamma_2^2 + \gamma_3^2 + \sigma'N^2 &= \gamma_{1N}^2 + \gamma_{2N}^2 + \gamma_{3N}^2 + \sigma'N^2
\end{align*}
\]

- As can be seen, there are three equations and four unknowns and one of them needs to be normalized.

\(\gamma_{iN}\) and \(\sigma'N\) are parameters after normalization

Eq. 9
γ Normalization

Approach: Set the parameter ($\gamma_N$) for the choice with lowest absolute $\gamma$ value to zero

$$\begin{align*}
\gamma_{1N} &= \gamma_1 - \gamma_3 \\
\gamma_{2N} &= \gamma_2 - \gamma_3 \\
\sigma'_N^2 &= \gamma_1^2 + \gamma_2^2 + \gamma_3^2 + \sigma'^2 - (\gamma_1 - \gamma_3)^2 - (\gamma_2 - \gamma_3)^2 \\
&= \sigma^2 - (\gamma_1 - \gamma_3)^2 - (\gamma_2 - \gamma_3)^2
\end{align*}$$

Eq. 10

Potential Issue: $\sigma'_N^2$ may be less than zero (e.g. $\gamma_1 = 0.55$, $\gamma_1 = 0.55$, $\gamma_3 = -0.55$)

Solution: Apply $\sigma$ Normalization
σ Normalization

Approach: Set the value of $\sigma'_N^2$ to a positive value

\[
\begin{align*}
\gamma_{1N} &= \gamma_1 - \bar{\gamma} \pm \sqrt{\gamma^2 + (\sigma'^2 - \sigma'_N^2)/3} \\
\gamma_{2N} &= \gamma_2 - \bar{\gamma} \pm \sqrt{\gamma^2 + (\sigma'^2 - \sigma'_N^2)/3} \\
\gamma_{3N} &= \gamma_3 - \bar{\gamma} \pm \sqrt{\gamma^2 + (\sigma'^2 - \sigma'_N^2)/3}
\end{align*}
\]

Eq. 11

This will always be a valid normalization so long as $\sigma'_N^2 < \sigma'^2$