Quantifying errors in travel time and cost by latent variables in transport demand models

Juan Manuel Lorenzo Varela
16th July 2018
Contents

• Background (Motivation & Research questions)

• Methodology
  • Hybrid choice model (Model definition and assumptions)

• Case study (Commuting mode choice in Stockholm)
  • Data
  • Results

• Conclusions
Background / Motivation

- Discrete choice models to estimate the VTTS
Background / Motivation

- Discrete choice models to estimate the VTTS

RP

1. Small datasets / High correlation between variables
2. Measurement errors → endogeneity
Background / Motivation

- Discrete choice models to estimate the VTTS

**RP**
1. Small datasets / High correlation between variables
2. Measurement errors → endogeneity

**SP**
1. Short term / reference dependent
2. Hypothetical
3. Gain / loss asymmetries
4. Self-justification
5. ?
Background / Motivation

- Discrete choice models to estimate the VTTS

1. Small datasets / High correlation between variables
2. Measurement errors → endogeneity

1. Short term / reference dependent
2. Hypothetical
3. Gain / loss asymmetries
4. Self-justification
5. ?
Background / Motivation

- Discrete choice models to estimate the VTTS

RP
1. Small datasets / High correlation between variables
2. Measurement errors → endogeneity

SP
1. Short term / reference dependent
2. Hypothetical
3. Gain / loss asymmetries
4. Self-justification
5. ?

HCM

Modelling assumptions

Parameter estimates (Choice model; Structural and Measurement eq.)
Research Questions

- RQ1: Do parameter estimates depend on assumptions regarding the measurement error model and latent variable distributions?
- RQ2: What measurement error model should we use?
- RQ3: Which variables present more measurement errors?
Methodology - Hybrid Choice Model formulation

- Observable attributes of the alternatives
- Characteristics of the traveller
- Latent Variable
- Indicators:
  - Indicator 1
  - Indicator 2
  - ...
  - Indicator n
- Utility
- Choice: j
- Choice Model
- Latent attribute model
Methodology - Hybrid Choice Model formulation

\[ L(j, I | \beta, \theta, \lambda) = \int P(j | \beta, X, LV) f_M(I | LV, \lambda) f_{LV}(LV; \theta) dLV \]

- **Choice model**
- **Measurement eq.**
- **Structural eq.**

**Observable attributes of the alternatives**

**Latent Variable**

**Characteristics of the traveller**

**Indicators**
- Indicator 1
- Indicator 2
- ...
- Indicator n
Methodology - Hybrid Choice Model formulation – cont.

- Utility function

\[ U_i = \sum_{k \in S_i} \beta_{ik} \cdot x_{ik} + \sum_{l \in S_i} \gamma_{il} \cdot LV_{il} + ASC_i + \varepsilon_i, \]

- Structural equation (travel time or travel cost)

\[ LV_{il} = \mu_{il} + \sigma_{il} \cdot \phi \quad \text{with } \phi \sim N(0, 1^2) \]

- Measurement equation

\[ I_{ilm} = LV_{il} + \eta_{ilm} \]
\[ \eta_{im} = \sigma_{\varepsilon_i} \cdot \phi' \quad \text{with } \phi' \sim N(0, 1^2) \]
Methodology - Hybrid Choice Model formulation – cont.

- Utility function

\[ U_i = \sum_{k \in S_i} \beta_{ik} \cdot x_{ik} + \sum_{l \in S_i} \gamma_{il} \cdot LV_{il} + ASC_i + \varepsilon_i, \]

- Structural equation (travel time or travel cost)

\[ LV_{il} = \mu_{il} + \sigma_{il} \cdot \phi \quad \text{with} \quad \phi \sim N(0, 1^2) \]

- Measurement equation

\[ I_{ilm} = LV_{il} + \eta_{ilm} \quad \text{with} \quad \phi' \sim N(0, 1^2) \]

\[ \eta_{im} = \sigma_{\varepsilon i} \cdot \phi' \]
Methodology - Hybrid Choice Model formulation – cont.

- Utility function

\[ U_i = \sum_{k \in S_i} \beta_{ik} \cdot x_{ik} + \sum_{l \in S_i} \gamma_{il} \cdot LV_{il} + ASC_i + \varepsilon_i, \]

- Structural equation (travel time or travel cost)

\[ LV_{il} = \exp(\mu_{il} + \sigma_{il} \cdot \phi) \quad \text{with} \quad \phi \sim N(0, 1^2) \]

- Measurement equation

\[ I_{il} = LV_{il} \cdot \eta_{il} \rightarrow \text{Constrain} \quad E[\eta_{il}] = 1 \]

\[ \eta_{il} = \exp \left( -\frac{1}{2} \sigma_{\varepsilon i}^2 + \sigma_{\varepsilon i} \cdot \phi' \right) \quad \text{with} \quad \phi' \sim N(0, 1^2) \]

Multiplicative error formulation
Case Study – Commuting mode choice (Stockholm)

Data from NTS 2006

3777 observations

- 40.9% PT
- 48.8% Car (D + P)
- 10.3% Walk & bicycle
Case Study – Commuting mode choice (Stockholm)

Data from NTS 2006

3777 observations

- 40.9% PT
- 48.8% Car (D + P)
- 10.3% Walk & bicycle

Variables
- Socio-economics
  (Car competition, Gender, etc.)
- Travel time
  (Indicators: software calculated & self-reported)
- Travel cost (log)
  (Indicator: Heuristic calculated cost)

Violin plots of reported and calculated travel times for car and public transport.
Case Study – Commuting mode choice (Stockholm)

Data from NTS 2006

3777 observations

- 40.9% PT
- 48.8% Car (D + P)
- 10.3% Walk & bicycle

Variables
- Socio-economics
  (Car competition, Gender, etc.)
- Travel time
  (Indicators: software calculated & self-reported)
- Travel cost (log)
  (Indicator: Heuristic calculated cost)

Violin plots of reported and calculated travel times for car and public transport.
Case Study – Results

RQ1: Do parameter estimates depend on assumptions regarding the measurement error model and latent variable distributions?
RQ1: Do parameter estimates depend on assumptions regarding the measurement error model and latent variable distributions?

- Define 2 models (Same utility function, but different assumptions for the structural and measurement equation)
- Focus on whether the different modelling assumptions modify the time and cost parameter estimates
- The scale of the model can be playing a role, therefore we should look at a ratio of parameters that is scale independent
Case Study – Results

RQ1: Do parameter estimates depend on assumptions regarding the measurement error model and latent variable distributions?

- Define 2 models (Same utility function, but different assumptions for the structural and measurement equation)
- Focus on whether the different modelling assumptions modify the time and cost parameter estimates
- The scale of the model can be playing a role, therefore we should look at a ratio of parameters that is scale independent

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$M_0\text{wtp}$</th>
<th></th>
<th></th>
<th>$M_1\text{wtp}$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>t-value</td>
<td>Value</td>
<td>t-value</td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{IVT PT-WTP}}$</td>
<td>0.078</td>
<td>10.25</td>
<td>0.073</td>
<td>10.44</td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{IVT CarD-WTP}}$</td>
<td>0.0433</td>
<td>3.60</td>
<td>0.047</td>
<td>3.93</td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{IVT CarP-WTP}}$</td>
<td>0.0470</td>
<td>3.42</td>
<td>0.058</td>
<td>3.74</td>
<td></td>
</tr>
</tbody>
</table>
Case Study – Results

RQ1: Do parameter estimates depend on assumptions regarding the measurement error model and latent variable distributions?

- Define 2 models (Same utility function, but different assumptions for the structural and measurement equation)
- Focus on whether the different modelling assumptions modify the time and cost parameter estimates
- The scale of the model can be playing a role, therefore we should look at a ratio of parameters that is scale independent

<table>
<thead>
<tr>
<th>Parameter</th>
<th>M0_wtp</th>
<th>M1_wtp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>t-value</td>
</tr>
<tr>
<td>( \beta_{IVT\ PT-WTP} )</td>
<td>0.078</td>
<td>10.25</td>
</tr>
<tr>
<td>( \beta_{IVT\ CarD-WTP} )</td>
<td>0.0433</td>
<td>3.60</td>
</tr>
<tr>
<td>( \beta_{IVT\ CarP-WTP} )</td>
<td>0.0470</td>
<td>3.42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Final log-likelihood</th>
<th>Number of parameters</th>
<th>LRT value</th>
<th>LRT P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted model</td>
<td>( M_0^{wtp} )</td>
<td>-70622.34</td>
<td>24</td>
<td>8.2</td>
</tr>
<tr>
<td>Unrestricted model</td>
<td>( M_0^{wtp} )</td>
<td>-70618.24</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

We can reject the hypothesis that the parameters are equal at a 95% confidence level.
Case Study – Results

RQ2: What measurement error model should we use?
Case Study – Results

RQ2: What measurement error model should we use?

**Additive error term**

QQplot of SOFTWARE
CALCULATED

QQplot of SELF-
REPORTED

**Multiplicative error term**

QQplot of SOFTWARE
CALCULATED

QQplot of SELF-
REPORTED
Case Study – Results

RQ2: What measurement error model should we use?

Density of SOFTWARE CALCULATED variables

Density of SELF-REPORTED variables
RQ2: What measurement error model should we use?

Kolmogorov-smirnov

\[ D = \text{distance between the empirical cumulative distribution function of the sample and the cumulative distribution function of the reference distribution} \]

<table>
<thead>
<tr>
<th>Model</th>
<th>( M_0 )</th>
<th>( M_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time variable (network calculated)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Transport</td>
<td>0.0234</td>
<td>0.0027</td>
</tr>
<tr>
<td>Car Driver</td>
<td>0.0201</td>
<td>0.0032</td>
</tr>
<tr>
<td>Car Passenger</td>
<td>0.0277</td>
<td>0.0099</td>
</tr>
<tr>
<td>Time variable (Self-reported)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Transport</td>
<td>0.0367</td>
<td>0.0100</td>
</tr>
<tr>
<td>Car Driver</td>
<td>0.0289</td>
<td>0.0119</td>
</tr>
<tr>
<td>Car Passenger</td>
<td>0.0433</td>
<td>0.0030</td>
</tr>
</tbody>
</table>
Case Study – Results

RQ3: Which variables present more measurement errors?
RQ3: Which variables present more measurement errors?
### Elasticities

<table>
<thead>
<tr>
<th></th>
<th>Benchmark model (MNL)</th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public transport</td>
<td>-0.26</td>
<td>-0.37</td>
<td>-0.41</td>
<td>-0.45</td>
</tr>
<tr>
<td>Car Driver</td>
<td>-0.23</td>
<td>-0.48</td>
<td>-0.55</td>
<td>-0.63</td>
</tr>
<tr>
<td>Car Passenger</td>
<td>-0.62</td>
<td>-0.62</td>
<td>-0.71</td>
<td>-0.71</td>
</tr>
</tbody>
</table>

*Aggregate direct price point elasticities*
Case Study – Results; Model properties and policy implications

**Elasticities**

<table>
<thead>
<tr>
<th>Mode</th>
<th>Benchmark model (MNL)</th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public transport</td>
<td>-0.26</td>
<td>-0.37</td>
<td>-0.41</td>
<td>-0.45</td>
</tr>
<tr>
<td>Car Driver</td>
<td>-0.23</td>
<td>-0.48</td>
<td>-0.55</td>
<td>-0.63</td>
</tr>
<tr>
<td>Car Passenger</td>
<td>-0.62</td>
<td>-0.62</td>
<td>-0.71</td>
<td>-0.71</td>
</tr>
</tbody>
</table>

Aggregate direct price point elasticities

**Values of time (SEK / h)**

<table>
<thead>
<tr>
<th>Mode</th>
<th>Benchmark model</th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>Latent average cost</th>
<th>Average cost with measurement errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car Driver</td>
<td>146</td>
<td>68</td>
<td>67</td>
<td>49</td>
<td>58</td>
<td>71</td>
</tr>
<tr>
<td>Car Passenger</td>
<td>55</td>
<td>23</td>
<td>29</td>
<td>24</td>
<td>58</td>
<td>29</td>
</tr>
<tr>
<td>Public transport</td>
<td>73</td>
<td>75</td>
<td>70</td>
<td>.*</td>
<td>58</td>
<td></td>
</tr>
</tbody>
</table>
Case Study – Results; Model properties and policy implications

Elasticities

<table>
<thead>
<tr>
<th>Mode</th>
<th>Benchmark model (MNL)</th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public transport</td>
<td>-0.26</td>
<td>-0.37</td>
<td>-0.41</td>
<td>-0.45</td>
</tr>
<tr>
<td>Car Driver</td>
<td>-0.23</td>
<td>-0.48</td>
<td>-0.55</td>
<td>-0.63</td>
</tr>
<tr>
<td>Car Passenger</td>
<td>-0.62</td>
<td>-0.62</td>
<td>-0.71</td>
<td>-0.71</td>
</tr>
</tbody>
</table>

Values of time (SEK / h)

<table>
<thead>
<tr>
<th>Mode</th>
<th>Benchmark model</th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car Driver</td>
<td>146</td>
<td>68</td>
<td>67</td>
<td>49</td>
</tr>
<tr>
<td>Car Passenger</td>
<td>55</td>
<td>23</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>Public transport</td>
<td>73</td>
<td>75</td>
<td>70</td>
<td>.*</td>
</tr>
</tbody>
</table>

VoT \( = \frac{\beta_{IVT}}{\beta_{Cost}} \cdot \text{Cost} \)
Case Study – Results; Model properties and policy implications

**Elasticities**

<table>
<thead>
<tr>
<th>Mode</th>
<th>Benchmark model (MNL)</th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public transport</td>
<td>-0.26</td>
<td>-0.37</td>
<td>-0.41</td>
<td>-0.45</td>
</tr>
<tr>
<td>Car Driver</td>
<td>-0.23</td>
<td>-0.48</td>
<td>-0.55</td>
<td>-0.63</td>
</tr>
<tr>
<td>Car Passenger</td>
<td>-0.62</td>
<td>-0.62</td>
<td>-0.71</td>
<td>-0.71</td>
</tr>
</tbody>
</table>

Aggregate direct price point elasticities

**Values of time (SEK / h)**

<table>
<thead>
<tr>
<th>Mode</th>
<th>Benchmark model</th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Latent average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>cost</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Average cost</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>with</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>measurement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>errors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Car Driver</td>
<td>146</td>
<td>68</td>
<td>67</td>
<td>49</td>
</tr>
<tr>
<td>Car Passenger</td>
<td>55</td>
<td>23</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>Public transport</td>
<td>73</td>
<td>75</td>
<td>70</td>
<td>.*</td>
</tr>
</tbody>
</table>

Conclusions

1. Estimated parameters depend on the modelling assumptions

2. A multiplicative error model allow us to compare the estimated measurement error distributions across indicators and variables

3. Cost indicators have proportionally larger errors than time indicators → Dilution of the cost coefficient which makes the model yield large values of time
Thank you
Questions?

jmlv@kth.se

https://www.kth.se/profile/jmlv