Modelling Activity-Diary Data: Complexity or Parsimony?

Elke Moons

Transportation Research Institute -
Instituut voor Mobiliteit (IMOB)
Universiteit Hasselt - campus Diepenbeek
Belgium
Overview

• Activity-Based Models
  ○ History
  ○ Albatross Model
  ○ Complexity or Parsimony?
  ○ Application

• Mode Choice Models
  ○ Tree-based Lack-of-Fit Test
  ○ Semi- & Nonlinear Models
History: Transport Modelling

- 50’s: car use increased ⇒ Need for models to predict travel demand ⇒ Trip-based (Four Step) Model

- Drawback: focus on individual trips, relationship between trips is ignored ⇒ mid 70’s: Tour-based Models

- Home- & work-based chains ⇒ Drawback: no relationship between travel and non-travel aspects ⇒ 90’s: Activity-based Models

- Shift from travel demand to activity schedules: travel is derived demand ⇒ purpose is to predict which activities are conducted where, when, for how long, with whom and which transport mode was used to arrive there
Trip-Based Model

- Home
  - Sport
- Car
- Public Transport
- Work
- Walk
- × 2
- Restaurant

Elke Moons, 30th January 2006 – p. 5/54
History: Transport Modelling (3)

Activity-Based Model

and much more information...

Elke Moons, 30th January 2006 – p. 5/54
**History: Activity-Based Models**

- **Constraints-based Models**
  - Travel time
  - Duration & timing
  - Sequencing

- **Utility-maximising or Simultaneous Models**
  - One day pattern
  - Simultaneous
  - Econometric

- **Computational Process or Sequential Models**
  - How does one arrive at the sequence of activities?
  - ‘Optimal’ choice
  - IF - THEN rules
**Albatross Model**

- A Learning BAased TRansportation Oriented Simulation System
- Only fully operational computational process model to date
- Dutch Ministry of Transportation

**Data**
- Collected in Hendrik-Ido-Ambacht and Zwijndrecht (South-Rotterdam region)
- In February 1997, random sample of 1649 respondents
- Activity Diary: nature, location, day, begin & end time, transport mode (chain), travel times, accompanying persons
- General characteristics: age, gender, type of household, children, car availability (ratio), etc.

(Arentze and Timmermans, 2000)
Albatross Model (2)
Albatross Model (3)

- Purpose: Schedule activities
- Interactions between individuals in household
- Constraints
  - Situational
  - Institutional
  - Household
  - Spatial
  - Time
- 9 dimensions or choice facets that need to be modelled!
## Model Comparison

- **Choice Facet Level: Complexity & Accuracy**

<table>
<thead>
<tr>
<th>Dimension</th>
<th># cases</th>
<th># alternatives</th>
<th># variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode for Work</td>
<td>858</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>Selection</td>
<td>14190</td>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>Travel Party</td>
<td>2970</td>
<td>3</td>
<td>39</td>
</tr>
<tr>
<td>Duration</td>
<td>2970</td>
<td>3</td>
<td>41</td>
</tr>
<tr>
<td>Start Time</td>
<td>2970</td>
<td>6</td>
<td>63</td>
</tr>
<tr>
<td>Trip Chaining</td>
<td>2651</td>
<td>4</td>
<td>53</td>
</tr>
<tr>
<td>Mode Other</td>
<td>2602</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td>Location1</td>
<td>2112</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>Location2</td>
<td>1027</td>
<td>6</td>
<td>28</td>
</tr>
</tbody>
</table>
**Model Comparison (2)**

- **Activity Pattern Level:** Sequence Alignment Methods (Joh *et al.*, 2001, 2002)
  - Dissimilarity measures between observed and predicted sequences (type, location, mode & travel party)
  - Effort required to make two sequences identical
    \(\Rightarrow\) lower SAM measures are better
  - Insertion, deletion and substitution operators
  - 4 uni-dimensional SAM, UDSAM & MDSAM

- **Trip Matrix Level:** correlation coefficient
  - Origin-Destination matrix: trip is basic unit
  - Disaggregation on day, primary activity and transport mode
Complexity or Parsimony

- Why (or why not)?
  - High predictive performance: Complex ↔ Strongest effects, no (disturbing) details: Simple - Parsimonious Models
  - Middle ages: Occam’s razor (but beware!)
  - Psychology: Human Behaviour
  - What is best in Activity-Based Transportation context?

- Apply complex and parsimonious models within Albatross and compare results
Complexity or Parsimony (2)

- 2 Different Ways of Attaining Parsimony:
  - Simple Heuristics
    - One R
    - Naïve Bayes
  - Feature / Variable Selection: Relief-F

- Combination of Simple Models
  - Bagging
  - Boosting
Simple Heuristics

• One R(ule)
  ○ For each explanatory variable \( a \): find the majority class \( c \) of the response per value \( v \) in the domain of the predictor \( a \)
  ○ Rule: If \( a \) has value \( v \) then assign class \( c \)
  ○ Choose the rule with the highest accuracy
  ○ Example on Transport Mode
    Distance: Short \( \rightarrow \) Slow transport
    Long \( \rightarrow \) Car
    Accuracy of 70%
    Parking: Bad \( \rightarrow \) Public Transport
    Good \( \rightarrow \) Car
    Accuracy of 45%

  (Holte, 1993)
Simple Heuristics (2)

- Naïve Bayes
  - Bayes rule & naïve assumption of conditional independence

\[
P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad \text{by Bayes rule}
\]

\[\propto P(A_1, \ldots, A_n|C_i) \cdot P(C_i) \]
\[\propto \prod_{j=1}^{n} P(A_j|C_i) \cdot P(C_i)\]

- Prediction: Class with largest probability
- Robust to violations of independence assumption

- Zero R: choose the majority class

(Langley et al., 1992)
Decision Tree Induction

- Complex model
- Root node → Daughter nodes → Terminal node
- Stopping rule or pruning strategies
- Splitting criterion
  - Chi-square measure: CHAID (Kass, 1980)
  - Entropy: CART (Breiman et al., 1984)
  - Gain ratio: C4.5 (Quinlan, 1993)
- Binary or multi-way splits
Decision tree

example

Decision Tree Induction (2)
Feature / Variable Selection

- Too many variables: tree may over-fit the data (unstable)
- Tree structure: sensitive to highly correlated variables (problem of multicollinearity)
- Smaller tree (exclude redundant variables) $\Rightarrow$ higher probability of misclassification?
- Relief-F
  - Ranks variables according to their ability in distinguishing between the classes of the response, while mutually the variables should have a correlation as low as possible
  - Handles multi-class response variables

(Kononenko, 1994)
Application - Part I

- **Simple Heuristics & Trees after FS & Complex Models** in Albatross

- No dramatic differences at **Choice Facet Level**: CHAID and C4.5 perform best, but FS approach and simple heuristics come close (difference of 5-12% in accuracy)

- Important variables for:
  - **Mode for work**: Travel time by bike (!)
  - **Activity Selection**: Type of activity, day of the week
  - **Travel Party**: Household composition, day, type of activity
  - **Duration**: Travel party, type of activity
  - **Start Time**: Available time
  - **Trip Chain**: Enough time to include activity in schedule?
  - **Mode Other**: Household and transport characteristics
  - **Location 1 & 2**: Time-related and transport mode
### Activity Pattern Level

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean Distance</th>
<th>CHAID</th>
<th>Zero R</th>
<th>One R</th>
<th>NB</th>
<th>Full</th>
<th>FS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAM (AT)</td>
<td>2.777</td>
<td>3.130</td>
<td>3.027</td>
<td>3.022</td>
<td>2.903</td>
<td>2.929</td>
<td></td>
</tr>
<tr>
<td>SAM (L)</td>
<td>3.127</td>
<td>3.251</td>
<td>3.184</td>
<td>3.107</td>
<td>3.166</td>
<td>3.033</td>
<td></td>
</tr>
<tr>
<td>SAM (TM)</td>
<td>4.626</td>
<td>5.018</td>
<td>4.592</td>
<td>4.781</td>
<td>4.497</td>
<td>4.600</td>
<td></td>
</tr>
<tr>
<td>MDSAM</td>
<td>8.333</td>
<td>8.951</td>
<td>8.474</td>
<td>8.671</td>
<td>8.374</td>
<td>8.373</td>
<td></td>
</tr>
</tbody>
</table>

- One R, FS, C4.5 (‘full’) and CHAID perform well
- 1 substitution and/or 1-2 insertion/deletion operations suffice
- MDSAM: reasonable amount of association between elements across dimensions
**Trip Matrix Level**
- Total number of trips from certain origin to certain destination

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Observed</th>
<th>Predicted</th>
<th>Zero R</th>
<th>One R</th>
<th>NB</th>
<th>CHAID</th>
<th>Full</th>
<th>FS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Day</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekday</td>
<td>2359</td>
<td>2572</td>
<td>2510</td>
<td>2414</td>
<td>2454</td>
<td>2564</td>
<td>2413</td>
<td></td>
</tr>
<tr>
<td>Saturday</td>
<td>356</td>
<td>276</td>
<td>277</td>
<td>380</td>
<td>416</td>
<td>331</td>
<td>262</td>
<td></td>
</tr>
<tr>
<td>Sunday</td>
<td>287</td>
<td>203</td>
<td>248</td>
<td>172</td>
<td>297</td>
<td>214</td>
<td>157</td>
<td></td>
</tr>
<tr>
<td><strong>Transport mode</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Car</td>
<td>1609</td>
<td>1580</td>
<td>1609</td>
<td>1465</td>
<td>1771</td>
<td>1573</td>
<td>1466</td>
<td></td>
</tr>
<tr>
<td>Slow</td>
<td>814</td>
<td>1020</td>
<td>1013</td>
<td>1031</td>
<td>999</td>
<td>1038</td>
<td>920</td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>79</td>
<td>83</td>
<td>81</td>
<td>102</td>
<td>83</td>
<td>113</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td>Car pas.</td>
<td>294</td>
<td>356</td>
<td>321</td>
<td>357</td>
<td>305</td>
<td>375</td>
<td>333</td>
<td></td>
</tr>
</tbody>
</table>
Application - Part I (4)

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Observed</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zero R</td>
<td>One R</td>
</tr>
<tr>
<td>Primary activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work out</td>
<td>970</td>
<td>901</td>
</tr>
<tr>
<td>Med. visit</td>
<td>44</td>
<td>32</td>
</tr>
<tr>
<td>Bring/get</td>
<td>538</td>
<td>485</td>
</tr>
<tr>
<td>Non-leis.</td>
<td>106</td>
<td>83</td>
</tr>
<tr>
<td>Non-groc.</td>
<td>251</td>
<td>287</td>
</tr>
<tr>
<td>Grocery</td>
<td>319</td>
<td>281</td>
</tr>
<tr>
<td>Leisure</td>
<td>466</td>
<td>329</td>
</tr>
<tr>
<td>Soc. visit</td>
<td>241</td>
<td>353</td>
</tr>
<tr>
<td>Service</td>
<td>59</td>
<td>243</td>
</tr>
<tr>
<td>Other out</td>
<td>18</td>
<td>63</td>
</tr>
</tbody>
</table>
• **Trip Matrix Level**

<table>
<thead>
<tr>
<th>Matrix</th>
<th>CHAID</th>
<th>Zero R</th>
<th>One R</th>
<th>NB</th>
<th>Full</th>
<th>FS</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.937</td>
<td>0.925</td>
<td>0.928</td>
<td>0.917</td>
<td>0.942</td>
<td>0.947</td>
</tr>
<tr>
<td>Mode</td>
<td>0.836</td>
<td>0.787</td>
<td>0.862</td>
<td>0.842</td>
<td>0.856</td>
<td>0.849</td>
</tr>
<tr>
<td>Day</td>
<td>0.944</td>
<td>0.925</td>
<td>0.937</td>
<td>0.919</td>
<td>0.950</td>
<td>0.946</td>
</tr>
<tr>
<td>Primary activity</td>
<td>0.830</td>
<td>0.766</td>
<td>0.801</td>
<td>0.800</td>
<td>0.861</td>
<td>0.840</td>
</tr>
</tbody>
</table>

- Aggregated number of trips reflects in correlation coefficients
- Fit decreases with an increasing number of cells
- One R, FS and C4.5 give a good performance
Application - Part I (6)

- Conclusions - so far...
  - Simple models do not perform better, but are also not inferior to more complex models
  - Strong reduction in size of the trees can be obtained by first applying FS
  - Findings endorse primary belief that people do not rely on a complex series of rules to make a decision

- What is next?
Combination of Simple Models

• 2 powerful learning ideas introduced in the last decade

• Bagging
  ◦ Re-sample training set 50 times (sampling with replacement)
  ◦ Average the result of the classifier over the different data sets
  ◦ Reduces the variance of the prediction & improves the stability

• Boosting (AdaBoost)
  ◦ Build model on the data
  ◦ Increase influence of misclassified instances by giving them a weight
  ◦ Combine results of 10 different classifiers through a weighted majority vote (more accurate classifiers have a higher weight)

(Breiman, 1996; Freund and Schapire, 1997)
Combination of Simple Models (2)

- General: Feature Selection approach provided the best results of all the simple models

- Wickramaratna *et al.* (2001): use Bagging & Boosting only with weak classifiers

⇒ Application of Bagging & Boosting to One R and FS models
### Choice Facet Level

<table>
<thead>
<tr>
<th>Dimension</th>
<th>1R</th>
<th>1R Bag.</th>
<th>1R Bo.</th>
<th>FS</th>
<th>FS Bag.</th>
<th>FS Bo.</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode for work</td>
<td>0.595</td>
<td>0.605</td>
<td>0.640</td>
<td>0.595</td>
<td>0.611</td>
<td>0.614</td>
<td>0.648</td>
</tr>
<tr>
<td>Selection</td>
<td>0.677</td>
<td>0.678</td>
<td>0.734</td>
<td>0.669</td>
<td>0.672</td>
<td>0.673</td>
<td>0.724</td>
</tr>
<tr>
<td>With-whom</td>
<td>0.408</td>
<td>0.430</td>
<td>0.408</td>
<td>0.467</td>
<td>0.566</td>
<td>0.564</td>
<td>0.509</td>
</tr>
<tr>
<td>Duration</td>
<td>0.348</td>
<td>0.418</td>
<td>0.348</td>
<td>0.368</td>
<td>0.452</td>
<td>0.451</td>
<td>0.431</td>
</tr>
<tr>
<td>Start time</td>
<td>0.227</td>
<td>0.318</td>
<td>0.227</td>
<td>0.172</td>
<td>0.704</td>
<td>0.703</td>
<td>0.408</td>
</tr>
<tr>
<td>Trip chain</td>
<td>0.699</td>
<td>0.699</td>
<td>0.807</td>
<td>0.811</td>
<td>0.896</td>
<td>0.894</td>
<td>0.833</td>
</tr>
<tr>
<td>Mode other</td>
<td>0.413</td>
<td>0.724</td>
<td>0.413</td>
<td>0.508</td>
<td>0.931</td>
<td>0.930</td>
<td>0.528</td>
</tr>
<tr>
<td>Location 1</td>
<td>0.435</td>
<td>0.435</td>
<td>0.501</td>
<td>0.513</td>
<td>0.564</td>
<td>0.561</td>
<td>0.575</td>
</tr>
<tr>
<td>Location 2</td>
<td>0.234</td>
<td>0.296</td>
<td>0.234</td>
<td>0.312</td>
<td>0.708</td>
<td>0.710</td>
<td>0.372</td>
</tr>
</tbody>
</table>

- Clearly improvement in predictive performance per dimension!
- How about aggregate behaviour?
### Activity Pattern Level

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1R</td>
</tr>
<tr>
<td>SAM (AT)</td>
<td>3.027</td>
</tr>
<tr>
<td>SAM (TP)</td>
<td>3.312</td>
</tr>
<tr>
<td>SAM (L)</td>
<td>3.184</td>
</tr>
</tbody>
</table>

- One R Bagging and Boosting give overall the best performance at activity pattern level
### Trip Matrix Level

<table>
<thead>
<tr>
<th>Measure</th>
<th>(\rho(o,p))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1R</td>
</tr>
<tr>
<td>None</td>
<td>0.928</td>
</tr>
<tr>
<td>Mode</td>
<td>0.862</td>
</tr>
<tr>
<td>Day</td>
<td>0.937</td>
</tr>
<tr>
<td>Primary activity</td>
<td>0.801</td>
</tr>
</tbody>
</table>

- Best performance at aggregate level \(\neq\) best performance at dimensions separately
Conclusions

- Bagging and Boosting improve the performance of the Simple Models
- (Combinations of) Simple Models do not necessarily perform worse than more complex models
- They are able to capture the most important information in trying to predict activity-travel behaviour
Mode Choice Models

- Important issue: Modelling of Choice of Transport Mode

- Changes over the last decennium
  - Improvement in survey techniques ⇒ Larger data sets
  - Emphasis on activities ⇒ transportation characteristics + demographical variables + activity features
  - Variety of new techniques: semi- and nonlinear models

- Need for a test statistic that can assess the model fit of a logistic regression model in a high dimensional sample space

- How do (some of) these new techniques perform in this context?
Lack-of-Fit Test

- Selection of particular transport mode above others ⇒ Multiple Logistic Regression Setting

- Assessment of adequacy of model
  - Pearson test statistic (1900): all variables need to be categorical
  - Hosmer and Lemeshow (1980): continuous and categorical predictors
  - le Cessie and Van Houwelingen (1991): test based on smoothing
  - Aerts et al. (1999): statistic based on orthogonal series approximation
  - ...
Lack-of-Fit Test (2)

- Most of previous methods are faced with **curse of dimensionality** and have **practical difficulties** with implementation.

- **Hosmer and Lemeshow**
  - Pearson-like statistic \( \sim \chi^2_8 \) (simulations)
  - Form 10 equally sized groups (deciles of risk)
  - Grouping is based on the fitted probabilities under the null model.
  - Deal with dimensionality problem, **BUT** at cost of power.

- **Tree-Based Lack-of-Fit Test**
General Setting

- Null hypothesis:

\[ H_0 : \text{logit}(\pi) = \log\left(\frac{\pi}{1 - \pi}\right) = \beta \mathbf{X} \]

with \( \pi = P(Y = 1) \) and \( \mathbf{X} \) an \( n \times (p + 1) \) matrix consisting of the \( n \) measurements on the \( p \) variables

- Alternative hypothesis \( H_1 \): not a specific alternative model

- Omnibus test for \( H_0 \) against a wide range of alternative models
Tree-Based Lack-of-Fit Test

- Test statistic:

\[ T = \sum_{i=1}^{g} \frac{(y_i - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i (1 - \hat{\pi}_i)} \]

with \( \hat{\pi}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{\pi}_j^0 \) the average of the probabilities \( \hat{\pi}_j^0 \) for all covariate values \( (x_{j1}, \ldots, x_{jp}) \) in group \( i \), fitted under the null model

- Grouping is according to recursive partitioning algorithm underlying CART, which can be considered as nonparametric alternative model

- Number of groups will affect power characteristics

(Breiman et al., 1984)
Tree-Based Lack-of-Fit Test (2)

- Unbalanced groups

- Two variants
  - Weighted version
    
    $$T_W = \sum_{i=1}^{g} w_i \frac{(y_i - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}$$

    with weight $w_i = gn_i/N$, giving less weight to small groups
  
  - Cressie-Read version
    
    $$T_{CR} = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^{g} \left\{ y_i \left( \left( \frac{y_i}{n_i \hat{\pi}_i} \right)^\lambda - 1 \right) + (n_i - y_i) \left( \left( \frac{n_i - y_i}{n_i (1 - \hat{\pi}_i)} \right)^\lambda - 1 \right) \right\}$$

    with $-\infty < \lambda < \infty$
Tree-Based Lack-of-Fit Test (3)

- Partitioning of sample space is random $\Rightarrow$ Hard to determine null distribution theoretically

- As in Hosmer and Lemeshow’s approach: simulations learn that $\chi^2_{2 \times (g - p)}$ is a good approximation

- Null distribution can always be simulated by a parametric bootstrap method

- Tree may reveal a particular deviation from the null model
### Simulation Study

Null model:

\[
H_0 : \text{logit}(\pi(x_i, z_i)) = \beta_0 + \beta_1 x_i + \beta_2 z_i
\]

with \( Y_i \sim \text{Ber}(\pi(x_i, z_i)), i = 1, \ldots, 100, x_i, z_i \sim U(-6, 6), \beta_0 = 0.0, \beta_1 = 0.8 \) and \( \beta_2 = 0.3 \)

<table>
<thead>
<tr>
<th>Test</th>
<th>0.10%</th>
<th>0.05%</th>
<th>0.01%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( HL )</td>
<td>12.46</td>
<td>15.97</td>
<td>28.85</td>
</tr>
<tr>
<td>( \chi^2(8) )</td>
<td>13.36</td>
<td>15.51</td>
<td>20.09</td>
</tr>
<tr>
<td>( T )</td>
<td>20.86</td>
<td>25.13</td>
<td>32.45</td>
</tr>
<tr>
<td>( TW )</td>
<td>16.40</td>
<td>18.93</td>
<td>23.68</td>
</tr>
<tr>
<td>( TCR )</td>
<td>19.76</td>
<td>22.97</td>
<td>27.60</td>
</tr>
<tr>
<td>( \chi^2(12) )</td>
<td>18.55</td>
<td>21.03</td>
<td>26.22</td>
</tr>
</tbody>
</table>
First alternative model:

\[ H_1 : \logit(\pi(x_i, z_i)) = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_iz_i \]

with \( \beta_3 = 0.1 \) and \( 0.2 \)

<table>
<thead>
<tr>
<th>Interaction Model</th>
<th>( \beta_3 = 0.10 )</th>
<th>( \beta_3 = 0.20 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test(0.10)</td>
<td>Test(0.05)</td>
</tr>
<tr>
<td>( HL )</td>
<td>27.14</td>
<td>14.29</td>
</tr>
<tr>
<td>( T )</td>
<td>36.33</td>
<td>24.90</td>
</tr>
<tr>
<td>( TW )</td>
<td>45.71</td>
<td>32.04</td>
</tr>
<tr>
<td>( T_{CR} )</td>
<td>40.20</td>
<td>26.73</td>
</tr>
<tr>
<td>Oracle Test</td>
<td>77.02</td>
<td>73.08</td>
</tr>
</tbody>
</table>
**Simulation Study (3)**

Second alternative model:

\[ H_1 : \logit(\pi(x_i, z_i)) = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i^2 \]

with \( \beta_3 = 0.1 \) and 0.2

<table>
<thead>
<tr>
<th>Quadratic Model</th>
<th>( \beta_3 = 0.10 )</th>
<th>( \beta_3 = 0.20 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test(0.10) Test(0.05)</td>
<td>Test(0.10) Test(0.05)</td>
</tr>
<tr>
<td>( HL )</td>
<td>13.83 5.01</td>
<td>17.60 6.80</td>
</tr>
<tr>
<td>( T )</td>
<td>29.66 15.43</td>
<td>89.2  78.4</td>
</tr>
<tr>
<td>( TW )</td>
<td>47.70 32.06</td>
<td>95.40 91.40</td>
</tr>
<tr>
<td>( TCR )</td>
<td>36.47 23.85</td>
<td>94.40 86.60</td>
</tr>
<tr>
<td>Oracle Test</td>
<td>86.97 84.52</td>
<td>99.59 99.59</td>
</tr>
</tbody>
</table>
Other Examples

• GVHD and POPS data: Results comparable to those obtained in the literature

• High-dimensional simulation settings: very low power for HL in detecting interaction ↔ reasonable results for tree-based statistic
**Other Examples (2)**

- Dutch Car Driver Data: 1025 cases and 39 variables
- Goodness-of-Fit Test as Model Selection Tool

<table>
<thead>
<tr>
<th>Variables used in model</th>
<th>HL Stat.</th>
<th>P-value</th>
<th>$T_W$ Stat.</th>
<th>P-value</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: 25</td>
<td>6.09</td>
<td>0.648</td>
<td>64.71</td>
<td>0</td>
<td>926.98</td>
</tr>
<tr>
<td>Model 2: 26</td>
<td>4.47</td>
<td>0.800</td>
<td>64.77</td>
<td>0.002</td>
<td>928.91</td>
</tr>
<tr>
<td>Model 3: 28</td>
<td>14.33</td>
<td>0.058</td>
<td>64.89</td>
<td>0.006</td>
<td>925.42</td>
</tr>
<tr>
<td>Model 4: 30</td>
<td>5.31</td>
<td>0.670</td>
<td>45.02</td>
<td>0.028</td>
<td>910.23</td>
</tr>
<tr>
<td>Model 5: 32</td>
<td>5.63</td>
<td>0.616</td>
<td>45.14</td>
<td>0.034</td>
<td>912.38</td>
</tr>
<tr>
<td>Model 6: 34</td>
<td>7.35</td>
<td>0.434</td>
<td>39.63</td>
<td>0.044</td>
<td>910.41</td>
</tr>
<tr>
<td>Model 7: 36</td>
<td>8.46</td>
<td>0.358</td>
<td>38.62</td>
<td>0.048</td>
<td>902.98</td>
</tr>
<tr>
<td>Model 8: 38</td>
<td>8.62</td>
<td>0.328</td>
<td>35.64</td>
<td>0.066</td>
<td>896.94</td>
</tr>
<tr>
<td>Model 9: 39</td>
<td>9.55</td>
<td>0.274</td>
<td>35.14</td>
<td>0.074</td>
<td>898.56</td>
</tr>
</tbody>
</table>
Conclusion

• Continuous and categorical explanatory variables

• Non-parametric method can be used to confirm or improve a parametric null model

• Simulation results: very promising power characteristics in detecting incorrectly modelled variables, omitted interaction or quadratic effects, . . .

• Further research on approximate null distribution
Nonlinear Models

- Field of Transportation Research: dominated by linear models

- The Right Choice to Make?
  - Semi-linear: Fractional polynomials (mfp)
  - Nonlinear
    - Support Vector Machines (SVM)
    - Classification and Regression Trees (CART)
Fractional Polynomial Approach

Fractional polynomial of degree $m$

$$\zeta_0 + \sum_{i=1}^{m} \zeta_i H_i(x)$$

for $(p_1, \ldots, p_m)$ vector of positive or negative integers or fractions

$$H_i(x) = \begin{cases} x^{(p_i)} & \text{if } p_i \neq p_{i-1} \\ H_{i-1}(x) \ln(x) & \text{if } p_i = p_{i-1} \end{cases}$$

Set $H_0(x) = 1$, $p_0 = 0$ and

Box-Tidwell: $x^{(p_i)} = \begin{cases} x^{p_i} & \text{if } p_i \neq 0 \\ \ln(x) & \text{else} \end{cases}$

(Royston and Altman, 1994)
Fractional Polynomial Approach (2)

Recommendations

• Restrict powers to $p_i \in \{-2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2, \max(3, m)\}$

• $m > 2$ rarely required in practice

Examples

• $(\frac{1}{2}, 1): \zeta_0 + \zeta_1 \sqrt{x} + \zeta_2 x$

• $(1, 1): \zeta_0 + \zeta_1 x + \zeta_2 x \ln x$

Setting

• Multiple Logistic Regression

• Iterative Procedure: Stepwise Regression
Support Vector Machines

- Optimal separating hyperplane
- Example

(Vapnik, 1996)
Support Vector Machines (2)

- Non-separable Case
  - Maximise the margin
  - Allow for some points to be at the wrong side of the margin

- Nonlinear boundary: Transform original space and find linear boundaries in transformed space

- Kernel functions $K(x, x') = \langle h(x), h(x') \rangle$
  - $d$-th degree polynomial: $K(x, x') = (\kappa + \gamma \langle x, x' \rangle)^d$
  - Radial basis: $K(x, x') = \exp\left(-\frac{||x-x'||^2}{\gamma}\right)$
  - Neural network: $K(x, x') = \tanh(\kappa_1 \langle x, x' \rangle + \kappa_2)$
Data and Model Comparison

- Southeast Florida: 14527 cases and 15 variables
- Albatross: 1025 cases and 20 variables

<table>
<thead>
<tr>
<th></th>
<th>Slow</th>
<th>Public</th>
<th>Car driver</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch data</td>
<td>18.93%</td>
<td>12.29%</td>
<td>68.78%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Southeast Florida</td>
<td>2.71%</td>
<td>9.67%</td>
<td>87.13%</td>
<td>0.49%</td>
</tr>
</tbody>
</table>

- Slow and Public Transport will be hard to predict!
- Take Sensitivity $= P(y_{predicted} = 1 | y_{observed} = 1)$ as measure
- Accuracy will always be high, even if no ‘true’ case is predicted correctly!
Example: Dutch Data: Car Driver

\[
\text{logit}(\pi(x)) = \beta_0 + \beta_1(x_6 = 2) + \beta_2 x_7 + \beta_3(x_{14} = 1) + \beta_4(x_{24} = 1) + \beta_5(x_{26} = 2) + \beta_6(x_{26} = 8) + \beta_7 x_{30}^{-1} + \beta_8 x_{30}^3 + \beta_9 x_{31} + \beta_{10} x_{33}^{-1} + \beta_{11} x_{33}^{-\frac{1}{2}} + \beta_{12}(x_{38} = 1) + \beta_{13}(x_{39} = 1)
\]

• Mfp model: 2 fractional polynomials used; lowest AIC and BIC values among 3 ‘linear’ models

• Interpretation: Number of cars, at least 1 shopping, service, social, leisure, bring/get activity, travel time by bike, woman

• Classification tree: same variables

• Comparison of models occurs on the test set (70-30% split)
Examples: Summary

- Results of all the analyses on the test set

<table>
<thead>
<tr>
<th></th>
<th>D-CD</th>
<th>D-PT</th>
<th>D-ST</th>
<th>SEF-PT</th>
<th>SEF-ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>0.730</td>
<td>0.824</td>
<td>0.882</td>
<td>0.649</td>
<td>0.627</td>
</tr>
<tr>
<td>Mfp</td>
<td>0.811</td>
<td>0.029</td>
<td>0.882</td>
<td>0.000</td>
<td>0.659</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.788</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SVM - Linear</td>
<td>0.928</td>
<td>0.000</td>
<td>0.549</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>SVM - Polynomial</td>
<td>0.793</td>
<td>0.294</td>
<td>0.588</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SVM - Radial basis</td>
<td>0.946</td>
<td>0.000</td>
<td>0.510</td>
<td>0.052</td>
<td>0.008</td>
</tr>
<tr>
<td>SVM - Neural net</td>
<td>0.901</td>
<td>0.000</td>
<td>0.569</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>CART</td>
<td>0.864</td>
<td>0.118</td>
<td>0.549</td>
<td>0.097</td>
<td>0.056</td>
</tr>
</tbody>
</table>
Conclusions

• (Semi-)linear: Best interpretable

• SVM and CART: over-fit training data $\Rightarrow$ lower performance on test data

• Very skew data sets: better performance of (semi-)linear models

• More balanced data sets: nonlinear models
Thanks!

Everything should be made as simple as possible, but not simpler - Albert Einstein