# Are You The Beatles Or The Rolling Stones?

**Virginie Lurkin and Brian Sifringer**

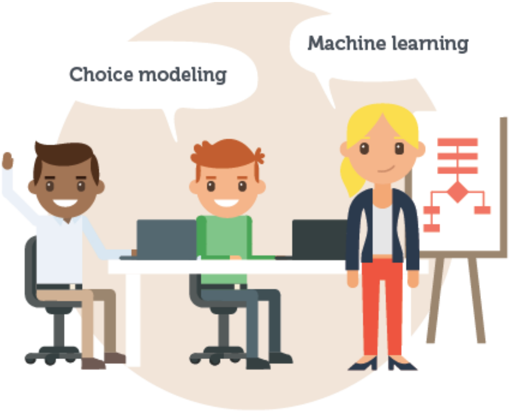
N. Ortelli, G. Lederrey, A. Alahi, M. Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil, and Environmental Engineering ENAC
École Polytechnique Fédérale de Lausanne

13th Workshop on Discrete Choice Models

June 21-23, 2018

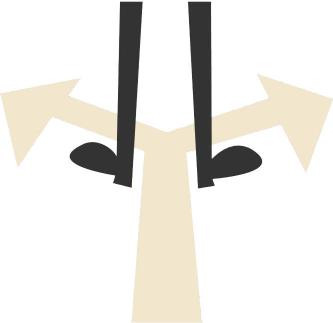

ARE YOU
THE BEATLES
OR
the Rolling Stones




*"The new and good ideas come from having a very broad and multidisciplinary range of interests."*

Robin Chase

# Discrete choice models



**Understand** and **predict** individual choice behavior
using mathematical models

# Discrete choice models

**Understand** and **predict** individual choice behavior
using mathematical models

# Discrete choice models

$$U_{in} = V_{in} + \varepsilon_{in}$$

# Discrete choice models

$$U_{in} = V_{in} + \varepsilon_{in}$$

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in \mathbb{C}_n} e^{V_{jn}}}$$

# Discrete choice models

$$U_{in} = V_{in} + \varepsilon_{in}$$

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in \mathbb{C}_n} e^{V_{jn}}}$$

$$V_{in} = ... + \beta_c Cost_{in} + \beta_t Time_{in} + ...$$

# Discrete choice models

$$U_{in} = V_{in} + \varepsilon_{in}$$

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in \mathbb{C}_n} e^{V_{jn}}}$$

$$V_{in} = \underbrace{... + \beta_c Cost_{in} + \beta_t Time_{in} + ...}_{\text{linear in parameters}}$$

# Discrete choice models

$$U_{in} = V_{in} + \varepsilon_{in}$$

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in \mathbb{C}_n} e^{V_{jn}}}$$

$$V_{in} = \underbrace{... + \beta_c Cost_{in} + \beta_t Time_{in} + ...}_{\text{linear in parameters}}$$

Prediction accuracy + Interpretability

# Machine learning

## Neural networks

# Machine learning

Neural networks

Very good prediction accuracy

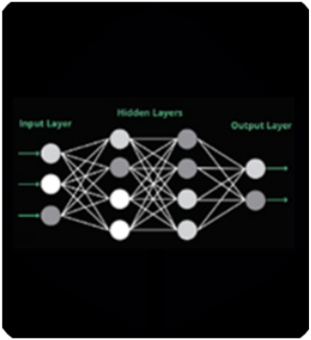
# Machine learning

Neural networks

Very good prediction accuracy

# Machine learning

# Machine learning



# Enhancing Discrete Choice Models with Neural Networks

# General idea

Bringing the **predictive strength** of Neural Networks, a powerful machine learning-based technique, to the field of Discrete Choice Models without compromising the **interpretability** of these models

# Multinomial Logit as Convolution NN

| MNL | CNN |
|---|---|
| Choice set $C_n = \{1,\ldots,J_n\}$ | Labels $(1 \times J_n)$ vector |
| Parameters $\beta_1,\ldots,\beta_p$ | Kernel weights $(p \times 1)$ |
| Variables $x_{ikn}$ | Features: $X$ $(p \times J_n)$ tensor |

# Multinomial Logit as Convolution NN

Activation Function: Softmax

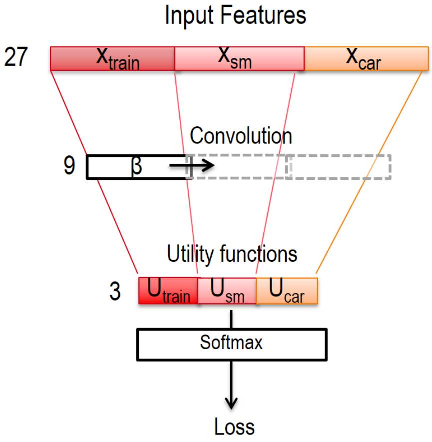
$$(\boldsymbol{\sigma}(\mathbf{V}_n))_i = \frac{\exp^{V_{in}}}{\sum_{j=1}^{p} \exp^{V_{jn}}}$$

Loss: Categorical Cross-Entropy

$$H_n(\boldsymbol{\sigma}, \mathbf{y}_n) = -\sum_{i=1}^{P} y_{in} \log\left[\sigma(\mathbf{V}_n))_i\right]$$

# Multinomial Logit as Convolution NN

A single convolution gives us the utility functions



# Model and Validation on Swissmetro

- Stated Preference Survey
- Simple utilities for validation purposes:

$$V_{car} = ASC_{car} + \beta_{cost} \cdot Cost_{car} + \beta_{time} \cdot Time_{car}$$
$$V_{train} = ASC_{train} + \beta_{cost} \cdot Cost_{train} + \beta_{time} \cdot Time_{train}$$
$$V_{SM} = ASC_{SM} + \beta_{cost} \cdot Cost_{SM} + \beta_{time} \cdot Time_{SM}$$

## Biogeme betas

| Name | Value | Std err | t-test | p-value |
|---|---|---|---|---|
| ASC_CAR | -0.993 | 0.0385 | -25.78 | 0.00 |
| ASC_TRAIN | -1.49 | 0.0515 | -28.86 | 0.00 |
| B_COST | -0.663 | 0.0473 | -14.03 | 0.00 |
| B_TIME | -0.153 | 0.0373 | -4.11 | 0.00 |

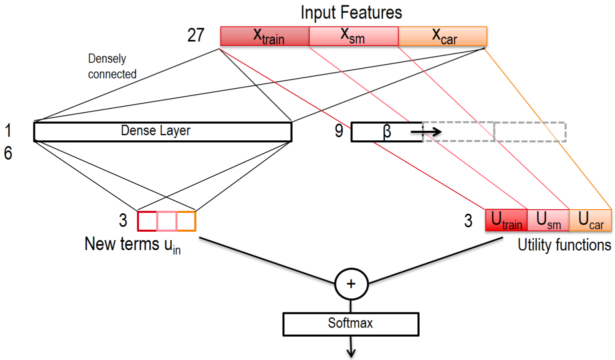
## Convolution Kernel weights

Convolution Kernel weights

- ASC_Car: -0.99298888
- ASC_Train: -1.48712599
- B_COST: -0.66329724
- B_TIME: -0.15334089

# Neural Network enhanced DCM

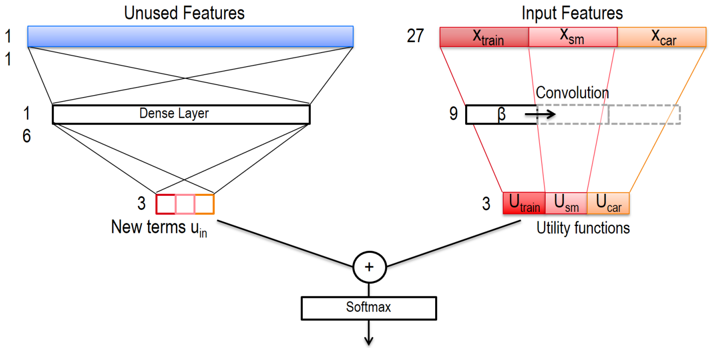
## Case 1: same inputs




The new term from the NN overruns the MNL linear parameters

# Neural Network enhanced DCM

## Case 2: different inputs




New term in the utility specification coming from a NN using all unused features

# Neural Network enhanced DCM

## Case 2: different inputs

$$U_{in} = ASC_i + \beta_{cost} \cdot Cost_{in} + \beta_{time} \cdot Time_{in} + \beta_{NN} \cdot NN_{in} + \varepsilon_{in}$$

- Interpretation: $NN_{in}$ = uncaptured information of MNL model
  - ✓ MNL estimates keep their significance
  - ✓ Increase in log-likelihood

# Swissmetro case

## Utility Functions

| Variable | | Alternative | | |
|---|---|---|---|---|
| | | Car | Train | Swissmetro |
| ASC | Constant | Car-Const | | SM-Const |
| TT | Travel Time | B-Time | B-Time | B-Time |
| Cost | Travel Cost | B-Cost | B-Cost | B-Cost |
| Freq | Frequency | | B-Freq | B-Freq |
| GA | Annual Pass | | B-GA | B-GA |
| Age | Age in classes | | B-Age | |
| Luggage | Pieces of luggage | B-Luggage | | |
| Seats | Airline seating | | | B-Seats |

# Swissmetro case

## Unused Features

| | |
|---|---|
| - Travel purpose: | Discrete value between 1 to 9 (Business, leisure, travel,... ) |
| - First class: | 0 for no or 1 for yes if passenger is a first class traveler in public transport |
| - Ticket: | Discrete value between 0 to 10 for the ticket type (One-way, half-day, ...) |
| - Who: | Discrete value between 0 to 3 for who pays the travel (self, employer, ...) |
| - Male: | Traveler's gender, 0 for female and 1 for male |
| - Income: | Discrete value between 0 to 4 concerning the traveler's income per year |
| - Origin: | Discrete value defining the canton in which the travel begins |
| - Dest: | Discrete value defining the canton in which the travel ends |

# Results

## Benchmark - standard MNL

| Parameter number | Description | Coeff. estimate | Robust Asympt. std. error | $t$-stat | $p$-value |
|---|---|---|---|---|---|
| 1 | $ASC_{Car}$ | 1.20 | 0.183 | 6.58 | 0.00 |
| 2 | $ASC_{SM}$ | 1.19 | 0.182 | 6.53 | 0.00 |
| 3 | $\beta_{age}$ | 0.175 | 0.0512 | 3.41 | 0.00 |
| 4 | $\beta_{cost}$ | -0.00690 | 0.000577 | -11.97 | 0.00 |
| 5 | $\beta_{freq}$ | -0.00704 | 0.00116 | -6.09 | 0.00 |
| 6 | $\beta_{GA}$ | 1.54 | 0.168 | 9.17 | 0.00 |
| 7 | $\beta_{luggage}$ | -0.113 | 0.0479 | -2.36 | 0.02 |
| 8 | $\beta_{seats}$ | 0.432 | 0.115 | 3.76 | 0.00 |
| 9 | $\beta_{time}$ | -0.0129 | 0.000842 | -15.34 | 0.00 |

Number of observations = 7234

$\mathcal{L}(\hat{\beta}) = -5766.705$

# Results

## Benchmark - standard MNL

| Parameter number | Description | Coeff. estimate | Robust Asympt. std. error | $t$-stat | $p$-value |
|---|---|---|---|---|---|
| 1 | $ASC_{Car}$ | 1.20 | 0.183 | 6.58 | 0.00 |
| 2 | $ASC_{SM}$ | 1.19 | 0.182 | 6.53 | 0.00 |
| 3 | $\beta_{age}$ | 0.175 | 0.0512 | 3.41 | 0.00 |
| 4 | $\beta_{cost}$ | -0.00690 | 0.000577 | -11.97 | 0.00 |
| 5 | $\beta_{freq}$ | -0.00704 | 0.00116 | -6.09 | 0.00 |
| 6 | $\beta_{GA}$ | 1.54 | 0.168 | 9.17 | 0.00 |
| 7 | $\beta_{luggage}$ | -0.113 | 0.0479 | -2.36 | 0.02 |
| 8 | $\beta_{seats}$ | 0.432 | 0.115 | 3.76 | 0.00 |
| 9 | $\beta_{time}$ | -0.0129 | 0.000842 | -15.34 | 0.00 |

Number of observations = 7234

$\mathcal{L}(\hat{\beta}) = -5766.705$

# Results

## Hybrid model

| Parameter number | Description | Coeff. estimate | Robust Asympt. std. error | $t$-stat | $p$-value |
|---|---|---|---|---|---|
| 1 | $ASC_{Car}$ | 0.0652 | 0.179 | 0.37 | 0.71 |
| 2 | $ASC_{SM}$. | 0.327 | 0.171 | 1.92 | 0.06 |
| 3 | $\beta_{age}$ | 0.376 | 0.0464 | 8.12 | 0.00 |
| 4 | $\beta_{cost}$ | -0.0141 | 0.000595 | -23.63 | 0.00 |
| 5 | $\beta_{freq}$ | -0.00807 | 0.00123 | -6.55 | 0.00 |
| 6 | $\beta_{GA}$ | 0.130 | 0.181 | 0.72 | 0.47 |
| 7 | $\beta_{luggage}$ | 0.0153 | 0.0505 | 0.30 | 0.76 |
| 8 | $\beta_{seats}$ | 0.207 | 0.106 | 1.95 | 0.05 |
| 9 | $\beta_{time}$ | -0.0157 | 0.000952 | -16.53 | 0.00 |
| 10 | $\beta_{NN}$ | 1.24 | 0.0524 | 23.74 | 0.00 |

Number of observations = 7234

$$\mathcal{L}(\hat{\beta}) = -5008.996$$

# Results

## Hybrid model

| Parameter number | Description | Coeff. estimate | Robust Asympt. std. error | $t$-stat | $p$-value |
|---|---|---|---|---|---|
| 1 | $ASC_{Car}$ | 0.0652 | 0.179 | 0.37 | 0.71 |
| 2 | $ASC_{SM}$. | 0.327 | 0.171 | 1.92 | 0.06 |
| 3 | $\beta_{age}$ | 0.376 | 0.0464 | 8.12 | 0.00 |
| 4 | $\beta_{cost}$ | -0.0141 | 0.000595 | -23.63 | 0.00 |
| 5 | $\beta_{freq}$ | -0.00807 | 0.00123 | -6.55 | 0.00 |
| 6 | $\beta_{GA}$ | 0.130 | 0.181 | 0.72 | 0.47 |
| 7 | $\beta_{luggage}$ | 0.0153 | 0.0505 | 0.30 | 0.76 |
| 8 | $\beta_{seats}$ | 0.207 | 0.106 | 1.95 | 0.05 |
| 9 | $\beta_{time}$ | -0.0157 | 0.000952 | -16.53 | 0.00 |
| 10 | $\beta_{NN}$ | 1.24 | 0.0524 | 23.74 | 0.00 |

Number of observations = 7234

$\boxed{\mathcal{L}(\hat{\beta}) = -5008.996}$

++ Increased likelihood

# Results

## Hybrid model

| Parameter number | Description | Coeff. estimate | Robust Asympt. std. error | $t$-stat | $p$-value |
|---|---|---|---|---|---|
| 1 | $ASC_{Car}$ | 0.0652 | 0.179 | 0.37 | 0.71 |
| 2 | $ASC_{SM}$. | 0.327 | 0.171 | 1.92 | 0.06 |
| 3 | $\beta_{age}$ | 0.376 | 0.0464 | 8.12 | 0.00 |
| 4 | $\beta_{cost}$ | -0.0141 | 0.000595 | -23.63 | 0.00 |
| 5 | $\beta_{freq}$ | -0.00807 | 0.00123 | -6.55 | 0.00 |
| 6 | $\beta_{GA}$ | 0.130 | 0.181 | 0.72 | 0.47 |
| 7 | $\beta_{luggage}$ | 0.0153 | 0.0505 | 0.30 | 0.76 |
| 8 | $\beta_{seats}$ | 0.207 | 0.106 | 1.95 | 0.05 |
| 9 | $\beta_{time}$ | -0.0157 | 0.000952 | -16.53 | 0.00 |
| 10 | $\beta_{NN}$ | 1.24 | 0.0524 | 23.74 | 0.00 |

Number of observations = 7234

$\mathcal{L}(\hat{\beta}) = -5008.996$

++ Keeps important parameters significant

# Results

## Hybrid model

| Parameter number | Description | Coeff. estimate | Robust Asympt. std. error | $t$-stat | $p$-value |
|---|---|---|---|---|---|
| 1 | $ASC_{Car}$ | 0.0652 | 0.179 | 0.37 | 0.71 |
| 2 | $ASC_{SM}$. | 0.327 | 0.171 | 1.92 | 0.06 |
| 3 | $\beta_{age}$ | 0.376 | 0.0464 | 8.12 | 0.00 |
| 4 | $\beta_{cost}$ | -0.0141 | 0.000595 | -23.63 | 0.00 |
| 5 | $\beta_{freq}$ | -0.00807 | 0.00123 | -6.55 | 0.00 |
| 6 | $\beta_{GA}$ | 0.130 | 0.181 | 0.72 | 0.47 |
| 7 | $\beta_{luggage}$ | 0.0153 | 0.0505 | 0.30 | 0.76 |
| 8 | $\beta_{seats}$ | 0.207 | 0.106 | 1.95 | 0.05 |
| 9 | $\beta_{time}$ | -0.0157 | 0.000952 | -16.53 | 0.00 |
| 10 | $\beta_{NN}$ | 1.24 | 0.0524 | 23.74 | 0.00 |

Number of observations = 7234

$\mathcal{L}(\hat{\beta}) = -5008.996$

– Some parameters loose significance

# Results

## Simple hybrid model - only key parameters

| Parameter number | Description | Coeff. estimate | Robust Asympt. std. error | $t$-stat | $p$-value |
|---|---|---|---|---|---|
| 1 | $ASC_{Car}$ | 0.966 | 0.0977 | 9.89 | 0.00 |
| 2 | $ASC_{SM}$ | 1.13 | 0.0941 | 11.97 | 0.00 |
| 3 | $\beta_{cost}$ | -0.0165 | 0.000666 | -24.71 | 0.00 |
| 4 | $\beta_{freq}$ | -0.00820 | 0.00129 | -6.38 | 0.00 |
| 5 | $\beta_{time}$ | -0.0171 | 0.000853 | -20.05 | 0.00 |
| 6 | $\beta_{NN}$ | 1.25 | 0.0854 | 14.65 | 0.00 |

Number of observations = 7234

$$\mathcal{L}(\hat{\beta}) = -4894.539$$

# Results

## Simple hybrid model - only key parameters

| Parameter number | Description | Coeff. estimate | Robust Asympt. std. error | $t$-stat | $p$-value |
|---|---|---|---|---|---|
| 1 | $ASC_{Car}$ | 0.966 | 0.0977 | 9.89 | 0.00 |
| 2 | $ASC_{SM}$ | 1.13 | 0.0941 | 11.97 | 0.00 |
| 3 | $\beta_{cost}$ | -0.0165 | 0.000666 | -24.71 | 0.00 |
| 4 | $\beta_{freq}$ | -0.00820 | 0.00129 | -6.38 | 0.00 |
| 5 | $\beta_{time}$ | -0.0171 | 0.000853 | -20.05 | 0.00 |
| 6 | $\beta_{NN}$ | 1.25 | 0.0854 | 14.65 | 0.00 |

Number of observations = 7234

$$\boxed{\mathcal{L}(\hat{\beta}) = -4894.539}$$

Increased likelihood + significant parameters

# Results

## Comparison

| Parameter | MNL | Hybrid | Simple Hybrid |
|---|---|---|---|
| $\beta_{cost}$ | 100.0% | 204.3% | 239.1% |
| $\beta_{freq}$ | 100.0% | 114.6% | 116.5% |
| $\beta_{time}$ | 100.0% | 121.7% | 132.5% |
| | | | |
| Value of Time | 0.54 | 0.89 | 0.96 |
| Value of Frequency | 0.98 | 1.75 | 2.01 |
| | | | |
| Final Log-Likelihood | -5766.71 | -5009.00 | -4894.54 |
| Number or parameters | 9 | 10 | 6 |

# Conclusions

Enhancing DCM greatly increases likelihood

Same input or correlated input

- Breaks original parameter significance

Independent Input

- Best performances for likelihood
- All parameters have good sign and significance

# Future work

## Interpretability: what is really happening?

| Parameter | MNL | Hybrid | Simple Hybrid |
|---|---|---|---|
| $\beta_{cost}$ | 100.0% | 204.3% | 239.1% |
| $\beta_{freq}$ | 100.0% | 114.6% | 116.5% |
| $\beta_{time}$ | 100.0% | 121.7% | 132.5% |
| Value of Time | 0.54 | 0.89 | 0.96 |
| Value of Frequency | 0.98 | 1.75 | 2.01 |
| Final Log-Likelihood | -5766.71 | -5009.00 | -4894.54 |
| Number or parameters | 9 | 10 | 6 |

## What do you think?

# Automatic Utility Specification Using Machine Learning Techniques

# Credit

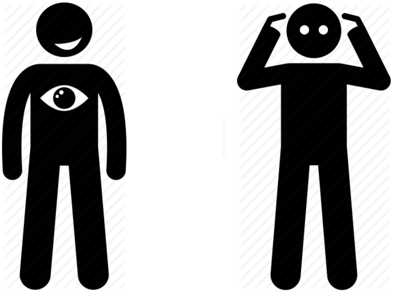
Nicola Ortelli

# Motivation

Determining the appropriate utility specification for a particular application is a difficult task

# Motivation

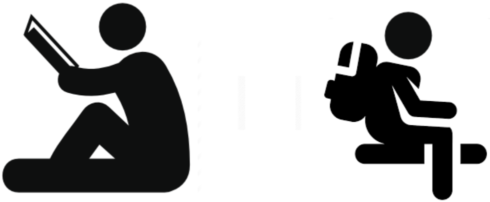
Determining the appropriate utility specification for a particular application is a difficult task



Expertise + Intuition

# Motivation

Determining the appropriate utility specification for a particular application is a difficult task



Inspiration + Experience

# Motivation

Determining the appropriate utility specification for a particular application is a difficult task

Trial-and-error

# Motivation

Determining the appropriate utility specification for a particular application is a difficult task

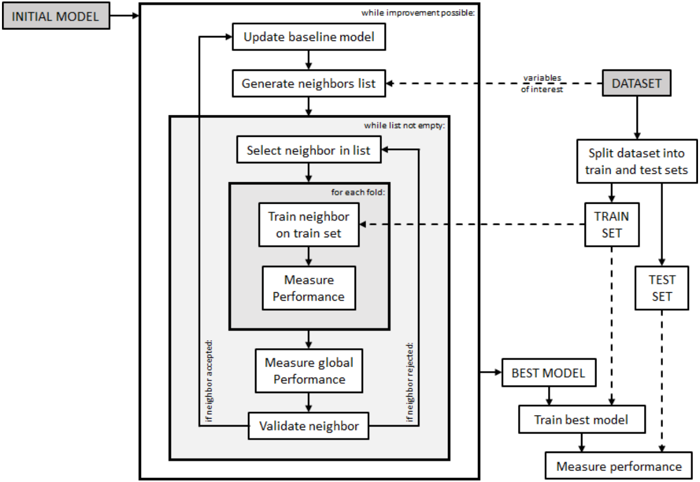
Time consuming

# Objective

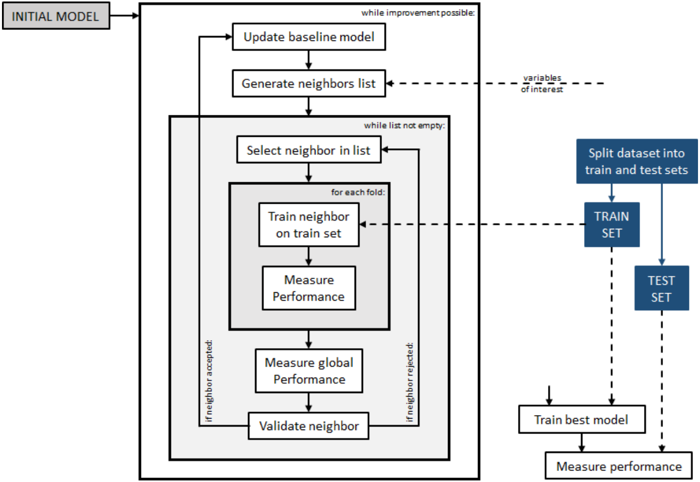
- Build a procedure that automatically finds a *good* utility specification based on the data

- Define neighborhood relations between specifications and use classical local search algorithms

# Automatic utility specification framework

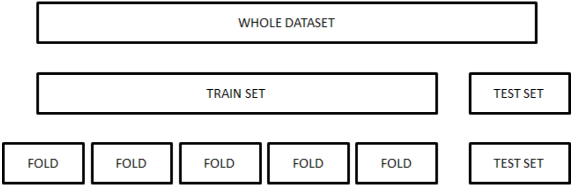


# Data partition

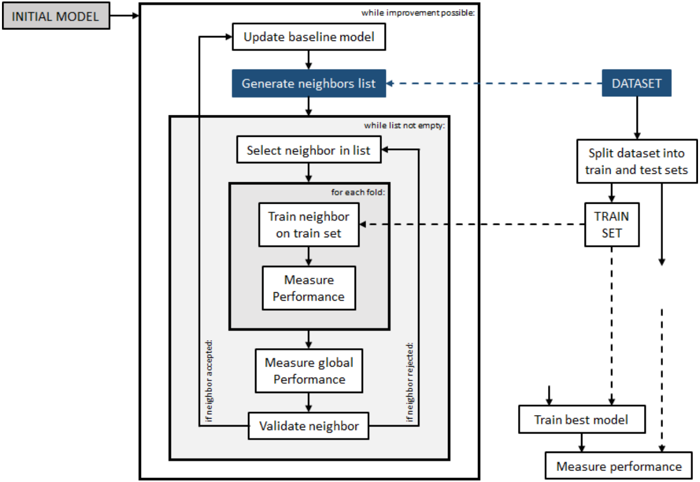


# Data partition

- Data set separated into train set and test set
- Train set separated into K folds



# Specifications and neighborhood structure



# Specifications and neighborhood structure

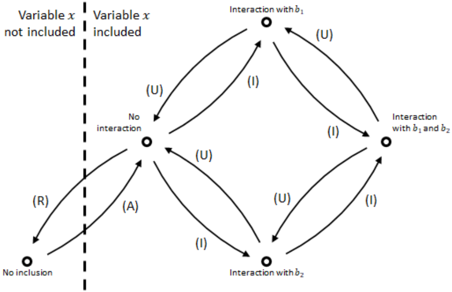
**Current assumptions**:

- Only continuous and binary variables
- Each continuous variable is included either on its own or in interaction with one binary variable
- All parameters are alternative specific
- Example:

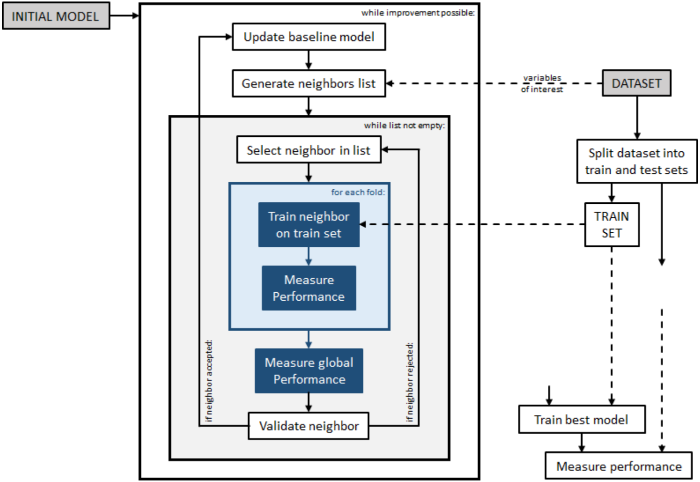
$$
\begin{aligned}
V &= ... + \beta_x \cdot x + ... \\
V &= ... + \beta_{x,b_1=0} \cdot x \cdot (1 - b_1) + \beta_{x,b_1=1} \cdot x \cdot b_1 + ... \\
V &= ... + \beta_{x,b_2=0} \cdot x \cdot (1 - b_2) + \beta_{x,b_2=1} \cdot x \cdot b_2 + ... \\
V &= ... + \beta_{x,b_1=0,b_2=0} \cdot x \cdot (1 - b_1) \cdot (1 - b_2) + \beta_{x,b_1=1,b_2=0} \cdot x \cdot b_1 \cdot (1 - b_2) \\
&\quad + \beta_{x,b_1=0,b_2=1} \cdot x \cdot (1 - b_1) \cdot b_2 + \beta_{x,b_1=1,b_2=1} \cdot x \cdot b_1 \cdot b_2 + ...
\end{aligned}
$$

# Specifications and neighborhood structure

- Neighbors of a particular specification = all specifications that are a single change away from it
- Four possible changes: add, remove, interact, un-interact



# Measure of performance



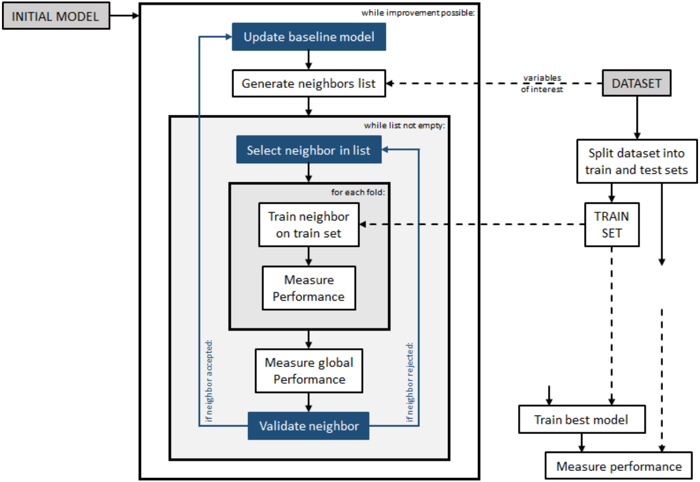
# Measure of performance

**Current assumptions**:

- Performance of a specification is measured as the log-likelihood it yields on new data
- Cross-validation is used to avoid overfitting or favorizing models with a large number of parameters
- The global performance $P_m$ of model $m$ is defined as the sum of the log-likelihoods obtained on each fold after estimation on the K-1 others:

$$P_m = \sum_{k=1}^{K} \mathbb{L}_{mk}$$

# Neighbor validation



# Neighbor validation

**Conditions of acceptance of a model $m$ :**

- A neighbor that performs better than the baseline is always accepted

$$P(\text{accept } m) = 1 \quad \text{if} \quad P_m \geq P_{m_b}$$

- One that performs worse might still get accepted, depending on the difference of performance $P_m - P_{m_b}$ and the current iteration number

$$P(\text{accept } m) = \frac{\exp(P_m - P_{m_b})}{T_t} \quad \text{if} \quad P_m < P_{m_b}$$

- The algorithm ends when all neighbors of a model are rejected

# Case study 1: Optima

# Case study: Optima

**Revealed preference survey conducted in Switzerland for CarPostal**:

- Three alternatives: public transports, car and soft modes
- 1'814 observations left after discarding incomplete ones
- Two binary and six continuous variables are considered
- 250'000 different specifications can be obtained

| Variable | Description |
|---|---|
| Gender | *Traveler's gender. 0 if female, 1 if male.* |
| UrbRur | *Area where the traveler lives. 1 rural, 0 if urban.* |
| TimePT | *Duration of a loop performed in public transport [minutes].* |
| MarginalCostPT | *Public transport cost, taking into account travel cards ownership [CHF].* |
| TimeCar | *Duration of a loop performed using the car [minutes].* |
| CostCar | *Total gas cost of a loop performed with the car [CHF].* |
| distance_km | *Total distance performed for a loop [km].* |
| NbTrajects | *Number of trips in a loop.* |

# Best specification

| Parameter | Final value | $t$-stat |
|---|---|---|
| $ASC_{PT}$ | 1.16 | 3.84 |
| $B_{MarginalCostPT}$ | –0.0651 | –7.75 |
| $B_{TimePT}$ | –0.0106 | –5.69 |
| $B_{PT\text{-}distance\_km}$ | 0.231 | 10.1 |
| $B_{PT\text{-}NbTrajects}$ | –0.759 | –4.69 |
| $ASC_{CAR}$ | 1.76 | 6.21 |
| $B_{TimeCar}$ | –0.0424 | –6.19 |
| $B_{Car\text{-}distance\_km}$ | 0.233 | 9.74 |
| $B_{Car\text{-}NbTrajects}$ | –0.66 | –4.47 |
| $ASC_{SM}$ | 0 | Fixed |
| $B_{SM\text{-}distance\_km}$ | 0 | Fixed |
| $B_{SM\text{-}NbTrajects}$ | 0 | Fixed |

# Comparison with an existing model

| Data | Statistic | Best encountered | Benchmark model |
|---|---|---|---|
| Train set | Log-likelihood | −952.51 | −843.29 |
| | Accuracy | 61.2% | 66.2% |
| | Correct guesses | 71.8% | 76.5% |
| Test set | Log-likelihood | −216.54 | −208.47 |
| | Accuracy | 60.3% | 65.1% |
| | Correct guesses | 69.0% | 74.7% |
| Number of estimated parameters | | 9 | 18 |
| Number of considered variables | | 5 | 13 |

# Case study 2: SwissMetro

# Case study: SwissMetro

**Stated choice survey to analyze the impact of the Swissmetro**:

- Three alternatives: train, Swissmetro and car
- Nine different situations for each of the 1'192 respondents
- Two binary and eight continuous variables are considered
- 6'250'000 different specifications can be obtained

| Variable | Description |
|---|---|
| GENDER | *Traveler's gender. 0 if female, 1 if male.* |
| GA | *GA travel card ownership. 1 if the traveler owns one, 0 otherwise.* |
| $TRAIN_{TT}$ | *Train travel time [minutes]. Door-to-door, based on the car distance.* |
| $TRAIN_{CO}$ | *Train cost [CHF]. If the traveler owns a GA, equal to its annual price.* |
| $TRAIN_{HE}$ | *Train headway [minutes].* |
| $SM_{TT}$ | *Swissmetro travel time [minutes]. A speed of 500 km/h is considered.* |
| $SM_{CO}$ | *Swissmetro cost [CHF]. Equal to the rail fare multiplied by a fixed factor.* |
| $SM_{HE}$ | *Swissmetro headway [minutes].* |
| $CAR_{TT}$ | *Car travel time [minutes].* |
| $CAR_{CO}$ | *Car cost [CHF]. A fixed average cost per kilometer is considered.* |

# Best specification

| Parameter | Final value | $t$-stat |
|---|---|---|
| $ASC_{TRAIN,GA=0}$ | 0.472 | 3.23 |
| $ASC_{TRAIN,GA=1}$ | 4.31 | 9.29 |
| $B_{TRAIN\text{-}TT,GA=0,GENDER=0}$ | −0.00815 | −8.36 |
| $B_{TRAIN\text{-}TT,GA=0,GENDER=1}$ | −0.0213 | −21.9 |
| $B_{TRAIN\text{-}TT,GA=1,GENDER=0}$ | *0.000281* | *0.122* |
| $B_{TRAIN\text{-}TT,GA=1,GENDER=1}$ | *−0.0000779* | *−0.0516* |
| $B_{TRAIN\text{-}CO,GA=0}$ | −0.00746 | −8.76 |
| $B_{TRAIN\text{-}CO,GA=1}$ | −0.00102 | −8.72 |
| $B_{TRAIN\text{-}HE}$ | −0.00719 | −7.23 |
| $ASC_{SM,GA=0, GENDER=0}$ | 1.4 | 8.83 |
| $ASC_{SM,GA=0, GENDER=1}$ | 0.286 | 2.81 |
| $ASC_{SM,GA=1, GENDER=0}$ | 5.45 | 9.3 |
| $ASC_{SM,GA=1, GENDER=1}$ | 4.7 | 9.9 |
| $B_{SM\text{-}TT}$ | −0.0137 | − 19.4 |
| $B_{SM\text{-}CO,GA=0,GENDER=0}$ | −0.0058 | −7.75 |
| $B_{SM\text{-}CO,GA=0,GENDER=1}$ | −0.0078 | −16.1 |
| $B_{SM\text{-}CO,GA=1,GENDER=0}$ | −0.00103 | −7.6 |
| $B_{SM\text{-}CO,GA=1,GENDER=1}$ | −0.000589 | −6.74 |
| $B_{SM\text{-}HE}$ | −0.0071 | −2.46 |
| $ASC_{CAR}$ | 0 | Fixed |
| $B_{CAR\text{-}TT,GENDER=0}$ | −0.00383 | −4.07 |
| $B_{CAR\text{-}TT,GENDER=1}$ | −0.0125 | −17.8 |
| $B_{CAR\text{-}CO}$ | −0.00672 | −7.58 |

# Comparison with an existing model

| Data | Statistic | Best encountered | Bierlaire *et al.* (2001) |
|---|---|---|---|
| Train set | Log-likelihood | −6'431.72 | −6'759.69 |
| | Accuracy | 54.1% | 51.8% |
| | Correct guesses | 67.2% | 62.8% |
| Test set | Log-likelihood | −1'551.94 | −1'695.30 |
| | Accuracy | 54.9% | 51.5% |
| | Correct guesses | 68.3% | 62.1% |
| Number of estimated parameters | | 22 | 10 |
| Number of considered variables | | 10 | 12 |

# Conclusions

# Conclusions

- Despite a certain number of restrictions, results show that this topic is worth further investigation

- The procedure reaches very good specifications in a matter of minutes

- In some cases, the obtained model outperforms the benchmark with less variables under consideration

# Future work

VNS –> neighborhood structures + relax assumptions

What is a good utility specification ?

- Predictability
- Parameters significance
- Behavioral interpretation
- ....

# Thank you!

virginie.lurkin@epfl.ch

*The fact is I've always loved both bands :-)*