Assisted specification of discrete choice models

Nicola Ortelli Michel Bierlaire

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

October 24, 2021

Outline

1. Introduction
2. Assisted Specification of Discrete Choice Models [Ortelli *et al.*, 2021]
3. Experiments
4. Conclusion

Model development

1. Behavioral hypothesis.
2. Model specification.
3. Model estimation.
4. Test and validation.
5. Go to 1.

TRANSP-OR

EPFL

Context

Specifying DCMs is difficult

- No “absolute rule” and exhaustive testing is intractable.
- Usual approach: expert intuition and trial-and-error.

DCMs & ever-larger datasets

- Significant growth of collected data:
  - “Taller” data — more observations;
  - “Wider” data — more variables.
- Consequently, two major problems for DCMs: estimation & specification.

→ **Easy way out: use machine learning instead!**

DCMs vs. ML

ML $\succ$ DCMs

- Scalability — better at handling large datasets.
- Data-driven — no need for presumptive structural assumptions.

DCMs $\succ$ ML

- Extrapolation — need for theories to explore out of the data.
- Interpretability — important for trust in the model.

The Importance of Behavioral Realism

"Predictions made by the model are conditional on the correctness of the behavioral assumptions and, therefore, are no more valid than the behavioral assumptions on which the model is based. A model can duplicate the data perfectly, but may serve no useful purpose for prediction if it represents erroneous behavioral assumptions. For example, consider a policy that will drastically change present conditions. In this case the future may not resemble the present, and simple extrapolation from present data can result in significant errors. However, if the behavioral assumptions of the model are well captured, the model is then valid under radically different conditions."

[Ben-Akiva, 1973]

Motivation

Main premise

- Human expertise is fundamental to formulate plausible hypotheses and to verify their compliance with behavioral theories. Data — no matter how big — cannot replace such knowledge.
- Still, can we use data to mitigate the need for a model specification to be known *a priori*, without compromising model interpretability?

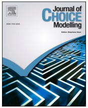
Outline

Journal of Choice Modelling 39 (2021) 100285

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: http://www.elsevier.com/locate/jocm



Assisted specification of discrete choice models

Nicola Ortelli [a,b,*], Tim Hillel [b], Francisco C. Pereira [c], Matthieu de Lapparent [a], Michel Bierlaire [b]

[a] School of Management and Engineering Vaud (HEIG-VD), University of Applied Sciences and Arts Western Switzerland (HES-SO), Switzerland
[b] Transport and Mobility Laboratory (TRANSP-OR), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
[c] Machine Learning for Smart Mobility Group (MLSM), Danmarks Tekniske Universitet (DTU), Denmark

A R T I C L E I N F O

Keywords:
Discrete choice models
Utility specification
Multi-objective optimization
Combinatorial optimization
Metaheuristics

A B S T R A C T

Determining appropriate utility specifications for discrete choice models is time-consuming and prone to errors. With the availability of larger and larger datasets, as the number of possible specifications exponentially grows with the number of variables under consideration, the analysts need to spend increasing amounts of time on searching for good models through trial-and-error, while expert knowledge is required to ensure these models are sound. This paper proposes an algorithm that aims at assisting modelers in their search. Our approach translates the task into a multi-objective combinatorial optimization problem and makes use of a variant of the variable neighborhood search algorithm to generate sets of promising model specifications. We apply the algorithm both to semi-synthetic data and to real mode choice datasets as a proof of concept. The results demonstrate its ability to provide relevant insights in reasonable amounts of time so as to effectively assist the modeler in developing interpretable and powerful models.

TRANSP-OR

EPFL

Disclaimer

What our method does

- **Assist** modelers.
- Save them as much time as possible.
- Provide a broader understanding of any dataset.
- Search more thoroughly than any human modeler would care to.

What our method does not

- Replace the analyst.
- Provide the best possible model for a given dataset.

Specification of DCMs...

Data

- For each individual n, we have:
  - a vector s_n of socioeconomic characteristics;
  - a vector x_{in} of attributes for each alternative i;
  - the chosen alternative i_n.

"Information set"

- The analyst provides:
  - a partition of attributes into K groups;
  - L potential transformations of the attributes;
  - S segmentations of the population;
  - M potential models $P_m(i|V;\mu)$, $m = 1, \ldots, M$.

→ What is the "best" specification we can come up with?

Specification of DCMs... as an Optimization Problem

Decision variables

- Five sets of **binary variables**:
  - $\delta_k = 1$ if group k is in the model;
  - $\gamma_k = 1$ if group k is associated with a generic coefficient;
  - $\phi_{k\ell} = 1$ if group k is associated with nonlinear transform ℓ;
  - $\sigma_{ks} = 1$ if the coefficients of group k are segmented based on s;
  - $\rho_m = 1$ if model m is used to calculate the choice probabilities.
- $\omega = (\delta, \gamma, \phi, \sigma, \rho)$ **unequivocally** describes a model specification.

Objective function

- Goodness of fit, *i.e.*, log likelihood $\mathcal{L}(\omega)$?
- Parsimony, *i.e.*, number of parameters $\mathcal{Z}(\omega)$?
- **Why not both?**

Multi-Objective Optimization

Two objective functions!

- We want the best fit (maximize $\mathcal{L}$) with the least parameters (minimize $\mathcal{Z}$).
- **Conflicting objectives**: improving one deteriorates the other.
- AIC and BIC quantify the trade-off between $\mathcal{L}$ and $\mathcal{Z}$...
- ... but we don't need to!
- How do we rank solutions then?

Pareto Optimization

Dominance

- Solution ω_1 **dominates** ω_2 ($\omega_1 \succ \omega_2$) if:
  - not worse in any objective:

$$\mathcal{L}(\omega_1) \geq \mathcal{L}(\omega_2) \text{ and } \mathcal{Z}(\omega_1) \leq \mathcal{Z}(\omega_2),$$

  - strictly better in at least one objective:

$$\mathcal{L}(\omega_1) > \mathcal{L}(\omega_2) \text{ or } \mathcal{Z}(\omega_1) < \mathcal{Z}(\omega_2).$$

Pareto front $\mathcal{P}$

- Set of **non-dominated** solutions (if $\omega \in \mathcal{P}$, no feasible ω' such that $\omega' \succ \omega$).
- All solutions in $\mathcal{P}$ are considered equally good.
- Solution to a multi-objective problem.

Multi-Objective Variable Neighborhood Search

Metaheuristics

- Too many feasible solutions!
- Explore the search space efficiently.
- "Sufficiently good solutions in reasonable amounts of time."

Description

- **Iteratively** improve an **approximation** of the Pareto front.
- Start with one or several user-defined candidates.
- Three ingredients:
  - Exploration — how we move in the search space,
  - Intensification — how we find local optima,
  - Diversification — how we escape local optima,

Operators

Exploration — how we move in the solution space

- Generate “neighbors” of a model.
- Mimic what an experienced modeler would do.
- Each operator modifies a subset of the decision variables.
- The complexity of the modification depends on the “size” of the neighborhood.
- Operators are invoked randomly.

Operators

Change variables	$\delta_k \leftarrow 1 - \delta_k$
Change generic *	$\gamma_k \leftarrow 1 - \gamma_k$
Change non-linearity *	$\{(k, l) \mid \phi_{kl} = 1\} : \phi_{k\ell} \leftarrow 0,\ \phi_{k\ell'} \leftarrow 1$
Change linearity *	$\begin{cases} \text{If } \phi_{k0} = 1 : & \phi_{k0} = 0,\ \phi_{k\ell'} = 1 \\ \text{If } \phi_{k0} = 0 : & \phi_{k0} = 1,\ \phi_{k1} = \cdots = \phi_{kL} = 0 \end{cases}$
Change segmentation *	$\sigma_{ks} \leftarrow 1 - \sigma_{ks}$
Increase segmentation *	$\{(k, s) \mid \sigma_{ks} = 0\} : \sigma_{ks} \leftarrow 1$
Decrease segmentation *	$\{(k, s) \mid \sigma_{ks} = 1\} : \sigma_{ks} \leftarrow 0$
Change model	$\{m \mid \rho_m = 1\} : \rho_m \leftarrow 0,\ \rho_{m'} \leftarrow 1$

Generic VNS Algorithm

Intensification — how we find local optima

- Local search:
  - Select model $\omega \in \mathcal{P}$ at random;
  - Generate and estimate neighbor ω'.
- Update the Pareto front $\mathcal{P}$ when appropriate:
  - Add ω' to $\mathcal{P}$ if $\{\omega \in \mathcal{P} \mid \omega \succ \omega'\} = \emptyset$;
  - Remove from $\mathcal{P}$ all solutions dominated by ω'.

Diversification — how we escape local optima

- Systematic changes of **neighborhood size** $p = 1, \dots, P$.
- After Q unsuccessful iterations, $p \leftarrow p + 1$.
- After each successful iteration, $p \leftarrow 1$.

Ensuring Behavioral Realism

Consistency with behavioral intuition is important

- **User-defined** constraints to verify model validity.
- Two options:
  - Hard — reject all models that violate the constraints (post-estimation).
  - Soft — constrained maximum likelihood estimation. [Schoenberg, 1997]

Why constrained maximum likelihood?

- Hard constraints may conceal severe specification issues.
- Fewer rejected models → more thorough search.
- All estimates guaranteed to be "behaviorally valid".

TRANSP-OR

EPFL

Outline

Data

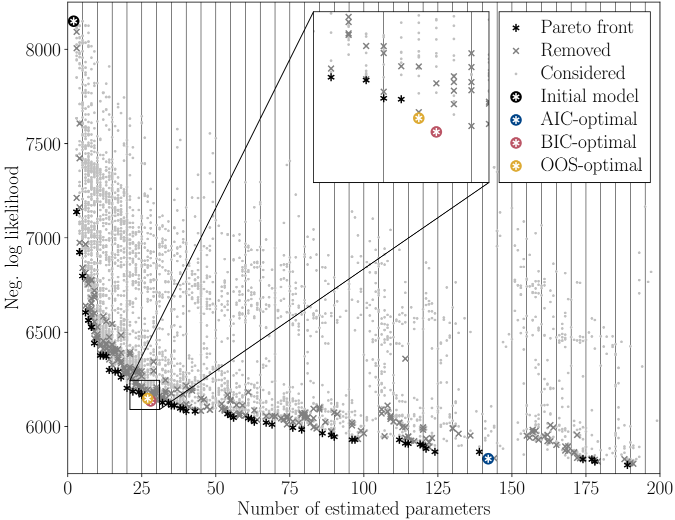
Swissmetro dataset [Bierlaire *et al.*, 2001]

- SP data.
- 3 alternatives.
- 10'710 observations, 20% as validation data.

Problem size

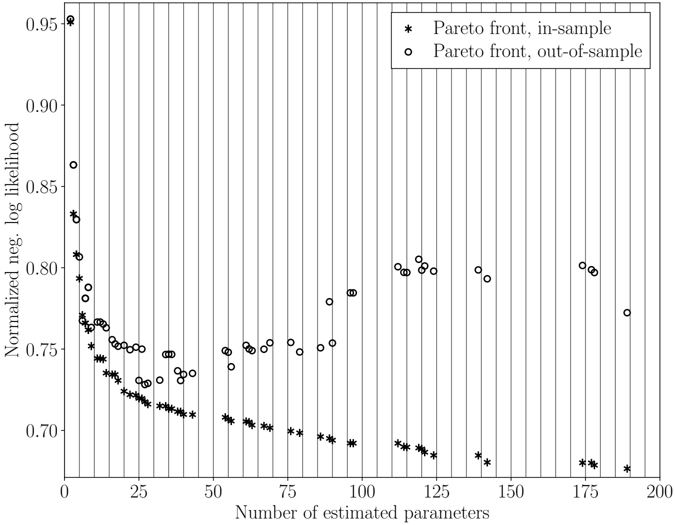
- $K = 3,\ L = 5,\ S = 7,\ M = 3.$
- $\approx 4.6 \times 10^8$ possible specifications.

In-Sample



Pareto front
Removed
Considered
Initial model
AIC-optimal
BIC-optimal
OOS-optimal
Neg. log likelihood
8000
7500
7000
6500
6000
0
25
50
75
100
125
150
175
200
Number of estimated parameters

Out-of-Sample



Data

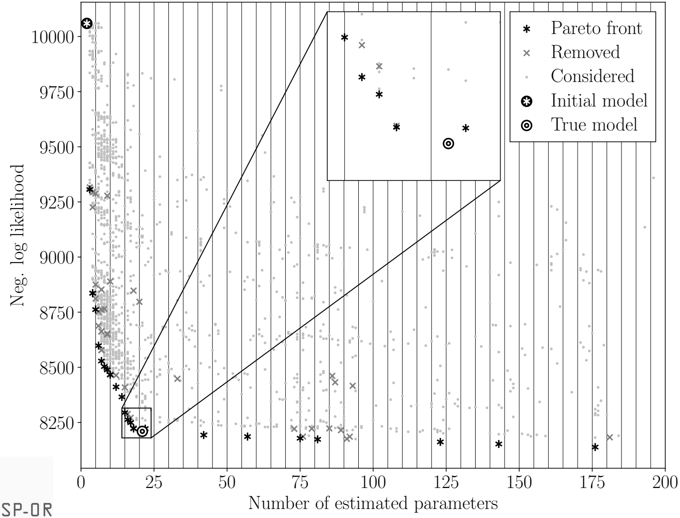
Synthetic choices

- Based on the Swissmetro data.
- We define a “true model”:
  - Estimated on the original data.
  - Used to sample new, synthetic choices.
  - **Can be reached by the algorithm!**

Problem size (same as previous)

- $K = 3, \ L = 5, \ S = 7, \ M = 3.$
- $\approx 4.6 \times 10^8$ possible specifications.

True Model Nearly Recovered



Data

London Passenger Mode Choice [Hillel *et al.*, 2018]

- RP data.
- 4 alternatives.
- $54'766 + 26'320$ observations.

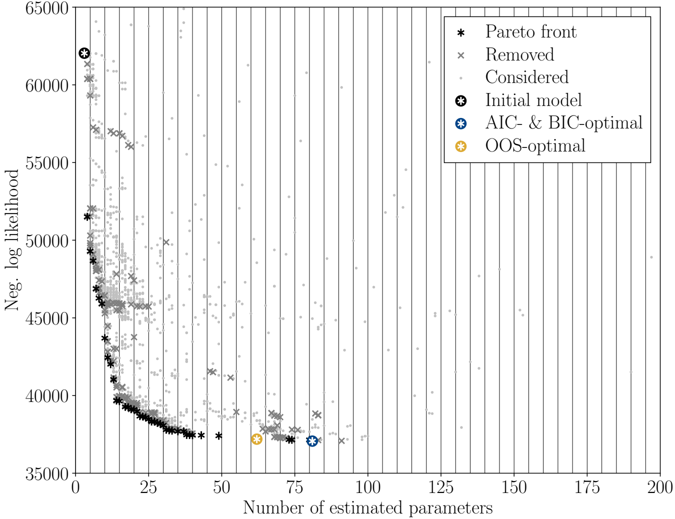
Problem size

- $K = 5,\ L = 3,\ S = 8,\ M = 4.$
- $\approx 4.7 \times 10^{10}$ possible specifications.

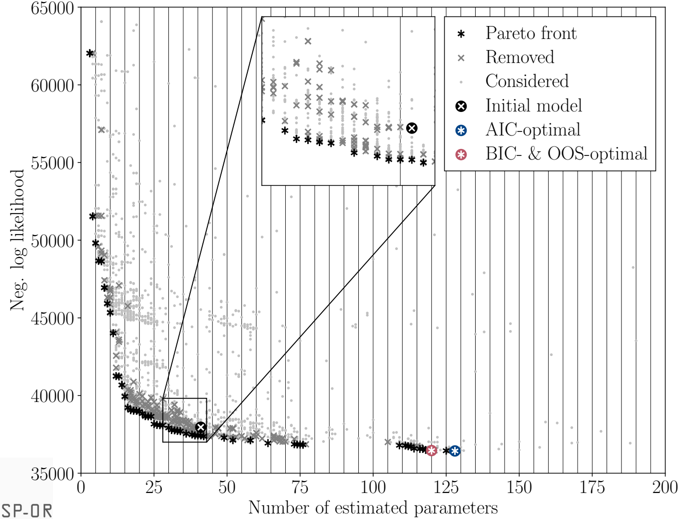
TRANSP-OR

EPFL

Initial Model: ASCs Only



Initial Model: Expert-Defined



Outline

Conclusion

Summary

- Flexible approach for assisted specification of DCMs.
- Appropriate use of data can **partially** relieve the modeler.
- Sets of high-quality specifications in reasonable amounts of time.
- Interpretable results!

Two major limitations

- Computational cost.
- Overfitting → This is why we should never replace the analyst!

Conclusion

Future work

- Try other objectives! Out-of-sample performance?
- Leverage the whole Pareto front. Model averaging?
- Faster estimation:
  - Heuristics based on partial estimation.
  - More complex optimization algorithms: HAMABS [Lederrey *et al.*, 2021]

TRANSP-OR

EPFL

References

Experiments

Ortelli, N., Hillel, T., Pereira, F.C., de Lapparent, M., Bierlaire, M., 2021. Assisted specification of discrete choice models. J.Choice.Model 39, 100285.

Datasets

Bierlaire, M., Axhausen, K., Abay, G., 2001. The acceptance of modal innovation: the case of Swissmetro. In: Proceedings of the 1st Swiss Transportation Research Conference.

Hillel, T., Elshafie, M., Jin, Y., 2018. Recreating passenger mode choice-sets for transport simulation: a case study of London, UK.Proc.Inst.Civ.Eng. Smart.Infrastruct.Construct. 171 (1), 29-42.

Quotation

Ben-Akiva, M.E., 1973. Structure of Passenger Travel Demand Models (Ph.D. Thesis). Dept. of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Faster estimation of DCMs

Lederrey, G., Lurkin, V., Hillel, T., Bierlaire, M., 2021. Estimation of discrete choice models with hybrid stochastic adaptive batch size algorithms. J.Choice.Model 38, 100226.

Questions?