Australasian Transport Research Forum 2022 Proceedings
28-30 September, Adelaide, Australia
Publication website: http://www.atrf.info

A multi-objective optimization framework for assisted estimation of discrete choice models

Prithvi Bhat Beeramoole[1], Md. Mazharul Haque[1], Alban Pinz[2], Alexander Paz[1]

[1]School of Civil & Environment Engineering, Queensland University of Technology, Australia

[2] Economic Research and Analysis, Queensland Department of Transport and Main Roads, Australia

Email for correspondence: *alexander.paz@qut.edu.au*

1. Introduction

Discrete choice models are widely used in various fields including but not limited to economics, transportation planning, safety, and business, to study factors that likely influence behavior, reveal causality, and enable forecasting. Some applications extend to demand estimation and counterfactual analysis, which require model inferences to be adequately generalizable to the population or to an external dataset. However, most studies using discrete choice models have been found to heavily rely on in-sample goodness-of-fit measures to evaluate model quality, which often increases the risk of over-fitting (Parady et al., 2021).

Estimating specifications with adequate goodness-of-fit and predictive accuracy significantly increases estimation complexity but is likely to increase modeling realism. Traditional model development entails an iterative process often including a limited number of hypotheses tested based on knowledge, experience, and available resources. This often results in a limited and inefficient search for model specification (Paz et al., 2019) with partial or no validation testing performed (Parady et al., 2021). The complexity of considering both in- and out-of-sample performance is increased when important modeling aspects such as nonlinearities, correlations, and random parameters are also included/tested.

In this study, an optimization framework is proposed to generate diverse hypotheses to estimate multinomial and mixed-Logit specifications while seeking models with significant explanatory power and out-of-sample validity. The proposed specification problem is defined as a multi-objective non-linear mixed integer optimization problem, considering conflicting objective functions, which include minimization of in-sample BIC and out-of-sample MAE. While BIC evaluates the model's explanatory power with respect to the number of model parameters, MAE assesses the model's ability to accurately recover behavior from an external dataset. The competing characteristics of the two performance measures enable one to seek specifications that can capture important and reliable insights. A metaheuristic-based solution algorithm is developed to solve the proposed optimization problem, which involves a strategic and unbiased search for a set of non-dominant solutions that test many behavioral hypotheses, including the presence of nonlinear effects of attributes, heterogeneity in preferences and correlation. Assisting model estimation using an optimization framework enables to simultaneously consider all the above aspects and perform extensive hypothesis testing to enable the discovery of important insights that could potentially be missed using a conventional process. The proposed framework also enables the discovery of generalizable inferences while avoiding potential overfitting and minimizing cost.

2. METHODOLOGY

2.1. Problem formulation

An observed utility associated with alternative j for individual n is given by v_{nj} in eqn. 1. Depending on the contribution to fit and behavioral meaning, coefficients $\boldsymbol{\beta}$ for alternative attributes can be estimated as generic, alternative-specific, fixed, random, or random-correlated coefficients. For outcome-independent characteristics $\mathbf{Z}_{\boldsymbol{n}}$, the corresponding coefficients $\boldsymbol{\theta}$ are estimated as alternative-specific, and those for base-outcome are normalized to 0, ensuring that only $J-1$ are estimated.

$$v_{nj} = \sum_{m=1}^{M} \ddot{\alpha_m} \theta_{jm} z_{nm} + \alpha_{j0}\beta_{j0} + \sum_{k=1}^{K} \alpha_{jk}\beta_{jk} {x_{njk}}^{(\lambda_k)} \tag{1}$$

Our model development process is viewed as a multi-objective non-linear mixed-integer optimization problem. The objective functions are to minimize the Bayesian Information Criterion (BIC) (eqn. 2.) and minimize the out-of-sample Mean Absolute Error (MAE) (eqn. 3.) where $s_{test\ data,j}$ is the observed market share for alternative j in the testing/validation dataset, and $\hat{s}_{test\ data,j}$ is the corresponding alternative's market share in the testing dataset which is estimated using the model developed from the training dataset.

$$\text{Min. BIC} = \delta \ln(\mathrm{N}) - 2\sum_{n}^{N}\sum_{j}^{J} y_{nj} \ln\left(\int \frac{e^{\sum_{m=1}^{M} \ddot{\alpha_m}\theta_{jm} z_{nm} + \alpha_{j0}\beta_{j0} + \sum_{k=1}^{K}\alpha_{jk}\beta_{jk} {x_{njk}}^{(\lambda_k)}}}{\sum_{j}^{J} e^{\sum_{m=1}^{M} \ddot{\alpha_m}\theta_{jm} z_{nm} + \alpha_{j0}\beta_{j0} + \sum_{k=1}^{K}\alpha_{jk}\beta_{jk} {x_{njk}}^{(\lambda_k)}}} \mathbf{f}(\boldsymbol{\beta})\mathbf{d}\boldsymbol{\beta}\right) \tag{2}$$

$$\text{Min. (MAE)} = \frac{1}{J}\sum_{j=1}^{J}\left|\hat{s}_{test\ data,j} - s_{test\ data,j}\right| \tag{3}$$

subject to constraints to ensure testing multiple modelling hypotheses including the existence of non-linearity, taste heterogeneity, and correlation. In addition, pre-specifications can be imposed to allow analysts to include any a priori knowledge, such as important factors, their functional forms, or distributional assumptions, into the model development.

2.2. Solution Algorithm

A multi-objective global-best harmony search (MOGBHS) solution algorithm was implemented based on Improved Global-best Harmony Search (Xiang et al., 2014) to solve the above problem. In the first step, the MOGBHS hyperparameters are initialized, followed by a harmony memory HM of size HMS, which contains an initial set of solutions $(M_1 \ldots M_{HMS})$ that are randomly generated subject to the proposed constraints. The analyst can also include known specifications in the initial harmony, such as those used in stated preference studies or based on knowledge, as starting points for the specification search. The solutions are then sorted from best to worst using Fast non-dominated sorting, proposed by Deb et al. (2002), which is based on dominance and crowding distance.

An iterative process of 'Improvising harmony' is then initiated, during which either m features from an existing solution M_{iter} in memory are randomly selected and considered for improvisation or a new solution is generated. A pitch adjustment step follows, in which the decision variables undergo perturbation based on a random number generator. Pitch adjustment allows testing of small changes in the specification, such as adding or removing an explanatory variable, or testing a nonlinear transformation, while seeking for an improvement. The objective functions are evaluated and included in HM followed by the non-dominant sorting of HM. A local search step is initiated as the iterations reach a pre-defined threshold. A solution in the Pareto-front (PF) is randomly selected for improvisation. The search terminates when no further improvement is observed in the PF solutions, or when the maximum number of iterations is reached. The final HM is returned which contains the best set of non-dominant solutions.

3. EXPERIMENT RESULTS

The proposed solution algorithm was applied to analyze transport mode choice behavior in Switzerland using a stated preference data collected by Bierlaire et al. (2001) in 1998. A detailed description of the dataset is provided by Antonini et al. (2007). Each respondent was presented with three transport mode alternatives (train, car, and Swiss metro), and nine hypothetical choice scenarios. Potential explanatory variables considered for the choice analysis included travel time (in minutes), travel cost (in CHF), headway for public transport modes (Train and Swiss metro), presence of luggage with traveler (no luggage, one, and more than one), seat configuration for Swiss metro (dummy variable indicating if the seats are arranged like airlines or not), dummy variable indicating if the traveler had an annual public transport ticket or not, traveler class (dummy variable to indicate first-class traveler), age, gender, income, and travel-cost bearer (self, employer, or both). **Table 1** presents the descriptive statistics for the considered variables. For the experiments, 80% of the total observations (10,395) were used as the training dataset, while the rest 20% were used to test out-of-sample prediction performance.

Table 1: Sample descriptive statistics of considered variables

Potential explanatory variables	Mean	Std.	Min	25%	50%	75%	Max.
Travel time (minutes)	126.1	81.0	0.0	71.0	112.0	171.0	1560.0
Travel cost (CHF)	424.6	1068.4	0.0	54.0	91.0	149.0	6720.0
Headway for public transport modes (minutes)	30.0	36.9	0.0	0.0	20.0	30.0	120.0
Presence of luggage with traveler	0.7	0.6	0.0	0.0	1.0	1.0	3.0
Annual public transport ticket	0.1	0.4	0.0	0.0	0.0	0.0	1.0
Traveler class	0.4	0.5	0.0	0.0	0.0	1.0	1.0
Age	2.9	1.1	1.0	2.0	3.0	4.0	6.0
Gender	0.7	0.4	0.0	0.0	1.0	1.0	1.0
Income	2.4	0.9	1.0	2.0	2.0	3.0	4.0
travel-cost bearer	1.8	0.9	1.0	1.0	1.0	3.0	3.0
seat configuration for Swiss metro	0.0	0.2	0.0	0.0	0.0	0.0	1.0

Figure 1 presents the final Pareto front identified using the proposed solution algorithm with respect to the considered objective functions. A significant improvement in model quality can be observed between the initial and final Pareto solutions, which illustrates the efficiency of the proposed algorithm in finding solutions with improved behavioral insights along with adequate out-of-sample predictive performance. A substantial trade-off that exists between the two objective functions can be observed clearly within the non-dominant solutions in the Pareto front. A specification reported in literature for the same dataset (Bierlaire et al., 2001) that was estimated using a conventional specification process included in Figure 1 as a benchmark solution using a triangular marker. The proposed extensive hypothesis testing algorithm was able to identify significantly superior solutions for both objective functions in comparison to the benchmark solution.

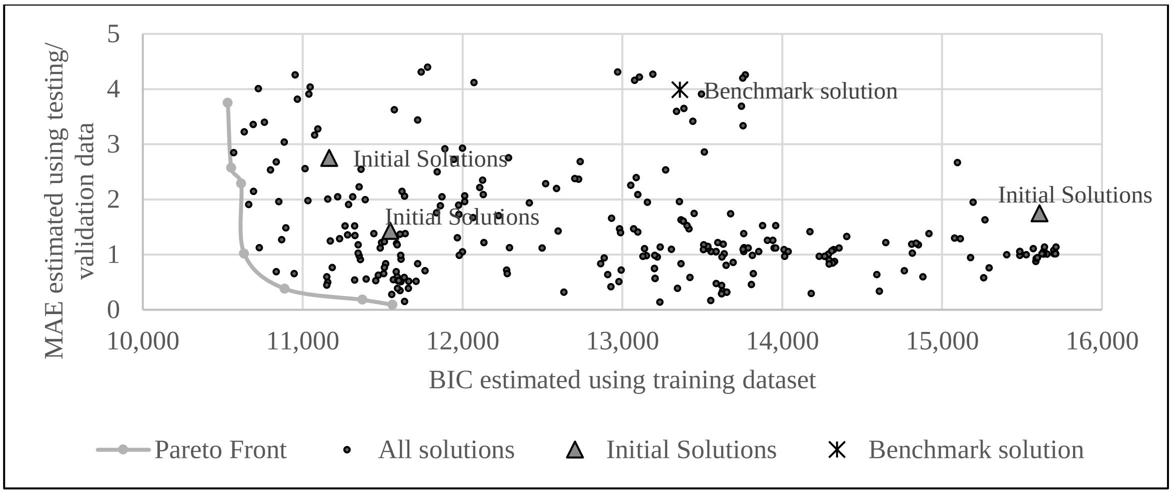

Figure 1- Pareto Front estimated using the proposed MOGBHS extensive hypothesis testing algorithm for Swissmetro data

Table 2- Specifications found by Bierlaire et al. (2001) and by the proposed solution algorithm

		Specification by Bierlaire et al. (2001)		**Estimated Specification by the proposed solution algorithm**		
Number of respondents: 924 Number of observations in training dataset: 8,316 Number of observations in testing dataset: 2,079						
Parameter		Estimate	t-ratio[1]	Estimate	t-ratio[1]	f[2]
For Swiss metro						
Seat configuration		0.16	2^{*}			
Travel cost	mean	-0.001	-19.6^{***}			
	s.d.					
Travel time		-0.01	-24.3^{***}	-0.04	-18.4	
Headway	mean	-0.01	-7.8^{***}	-0.01	-7.5	
	s.d.			0.04	8.1	*t*
Income	mean			-4.02	-9.6	
	s.d.			3.99	12.5	*ln*
For Car						
Mode-specific constant		0.062	1.2			
Age	mean					
	s.d.					
Luggage	mean	-0.12	-2.5^{**}	-3.18	-11.2	
	s.d.			6.87	16.1	*u*
Travel cost	mean	-0.001	-19.6^{***}			
	s.d.					
Travel time	mean	-0.01	-24.3^{***}	-0.02	-20.4	
	s.d.			0.02	15.1	*u*
For Train						
Mode-specific constant		-1.16	-10.4^{***}			
Luggage	mean			0.34	1.8	
	s.d.			2.53	8.6	*u*
Annual public transport ticket		7.49	21.9^{***}			
Age		0.19	6.1^{***}			
Travel cost for travelers without annual public transport ticket	mean			-0.04	-13.1	
	s.d.			0.07	13.1	*t*
Travel cost	mean	-0.001	-15.8^{***}			
	s.d.					
Travel time	mean	-0.01	-15^{***}	-0.02	-12.8	
	s.d.			0.03	15.9	
Headway	mean	-0.007	-7.8^{***}	-0.01	-7.5	
	s.d.			0.04	8.1	*t*
Log-Likelihood		**-6,565**		**-5,266**		
BIC		**13,211**		**10,633**		
Out-of-sample Log-Likelihood		**-1,940**		**-1,579**		
Out-of-sample MAE		**3.99**		**0.02**		

a. ** = weakly significant ($p < 0.10$, $t > 1.645$), ** = significant ($p < 0.05$, $t > 1.96$), *** = strongly significant ($p < 0.01$, $t > 2.58$)*

b. *n = normal; u = uniform; t = triangular*

A detailed comparative analysis is performed between a relatively optimal solution selected from the elbow of the final pareto and the specification by Bierlaire et al., (2001). Table 2 presents the two specifications. The specification found by the proposed solution algorithm provides significantly better goodness-of-fit along with out-of-sample predictive accuracy compared to the specification by Bierlaire et al., (2001). In addition, the improved specification captured additional behavioral insights that are essential in understanding heterogenous preferences in choice behavior. For example, the specification tested and found that the effects of travel times for train and car modes, and headway for public transit modes significantly varied across the sample. In addition, socioeconomic characteristics, and trip-related attributes, including income, and presence of luggage were found to be significant factors influencing transport mode choices of the observed sample. For example, as the number of luggage pieces with the passenger increased, they were more likely to choose train over car, and the effect of luggage varied significantly across the sample for the train mode. The coefficients most likely capture the effect of journey length due to the wide range of travel times reported in the study, indicating that passengers with more luggage were possibly travelling longer distances and therefore prefer train over cars.

Figure 2 presents the model performance of solutions estimated during the extensive hypothesis testing with respect to their specification type. Three main clusters of solutions were found to emerge based on the type of specification. The first cluster mostly includes Linear mixed-Logit specifications with correlated parameters. Solutions in this cluster were found to provide the best in-sample goodness-of-fit compared to the other clusters, while maintaining prediction accuracy when tested on an external data. Most mixed-Logit specifications without a correlation structure were found in the second cluster, which performed relatively poorly in both objective functions, compared to the solutions in first cluster. The third cluster consists of solutions with multinomial-Logit specifications with fixed coefficients. These solutions were found to perform poorly in terms of both in-sample BIC and out-of-sample MAE. These observations further validate the limitations of using restricted specifications and subjective hypothesis testing. While nonlinear specifications were also tested, most of the solutions failed to converge due to the limitations of standard MLE. Therefore, specific insights regarding nonlinearity and model performance could not be derived.

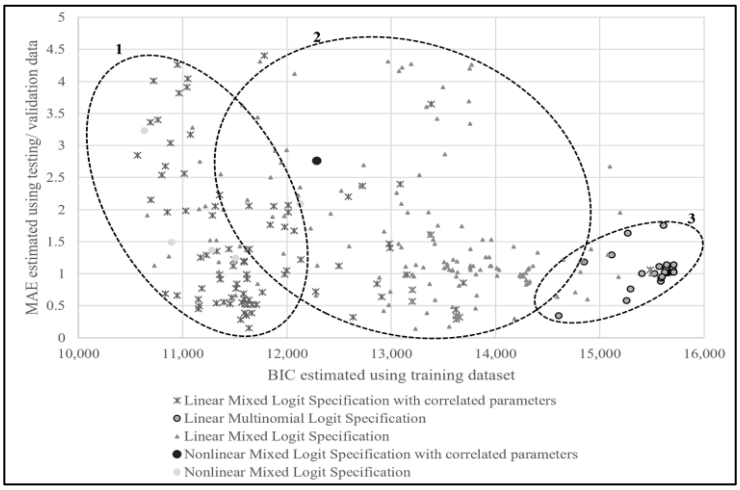

Figure 2: Clustering of Logit specifications estimated during the search based on performance

Figure 3 presents the overall search performance of the proposed MOGBHS algorithm for extensive hypothesis testing. A downward trend, indicating an improvement in BIC was observed during the search while maintaining out-of-sample prediction accuracy. The total estimation time was around 19 hours with more than 360 unique specifications tested.

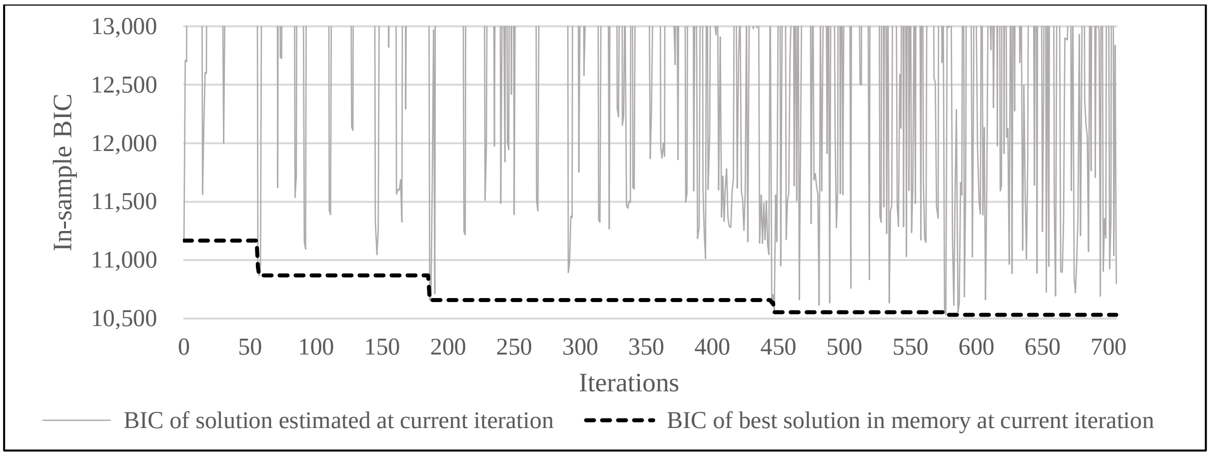

Figure 3: Search performance of the proposed MOGBHS hypothesis testing algorithm for Swissmetro data

4. Conclusion

This study proposed a framework to assist analysts in the estimation of discrete outcome models while evaluating their behavioral interpretability. The proposed formulation and associated solution algorithm enabled strategic testing of multiple modelling hypotheses including, the selection of potential explanatory variables, their functional forms, the distributional assumptions of coefficients, and correlations, thereby supporting estimation effort at a low cost. The Pareto front illustrates the efficacy of the proposed solution algorithm in providing multiple, unique, and acceptable solutions for the analyst to choose from based on the underlying modelling objective. A detailed comparison of specifications found by the proposed solution algorithm and the benchmarking specification reported in literature, illustrated the ability of the framework to conduct an extensive hypothesis testing to estimate solutions with improved predictive accuracy along with behavioral interpretability. Further, the proposed framework allows testing of multiple model performance measures, including likelihood estimates and number of model parameters to enable a specification search based on the study needs. The proposed framework can be particularly beneficial as a decision-support tool in behavioral analyses and demand estimation studies that involve highly dimensional datasets.

5. Reference

ANTONINI, G., GIOIA, C. & FREJINGER, E. 2007. Swissmetro: description of the data.

BIERLAIRE, M., AXHAUSEN, K. & ABAY, G. 2001. The acceptance of modal innovation: The case of Swissmetro.

DEB, K., PRATAP, A., AGARWAL, S. & MEYARIVAN, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation,* 6**,** 182-197.

PARADY, G., ORY, D. & WALKER, J. 2021. The overreliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: A review of validation practices in the transportation academic literature. *Journal of Choice Modelling,* 38**,** 100257.

PAZ, A., ARTEAGA, C. & COBOS, C. 2019. Specification of mixed logit models assisted by an optimization framework. *Journal of choice modelling,* 30**,** 50-60.

XIANG, W., AN, M., LI, Y., HE, R. & ZHANG, J. 2014. An improved global-best harmony search algorithm for faster optimization. *Expert Systems with Applications,* 41**,** 5788-5803.