ACCOUNTING FOR THE CONSIDERATION SET IN DISCRETE CHOICE MODELS BASED ON HISTORICAL CHOICES

C. Angelo Guevara, Department of Civil Engineering, FCFM University of Chile, Institute of Complex Engineering Systems (ISCI), crguevar@ing.uchile.cl

Abstract

The consideration set represents the subset of alternatives that individuals evaluate when choosing becomes impractical when facing a large number of options, a challenge that is common to various transportation models of route or schedule choices. Since this set is latent to the researcher, explaining it adequately is crucial for achieving consistent parameter estimation in discrete choice models. While theoretically robust methods to construct the consideration set are often impractical, existing practical approaches lack theoretical foundation. This article aims to bridge that gap. This article provides a theoretical basis for constructing consideration sets using historical or cohort-based choices, leveraging data from passive sources such as mobile phones, travel smart cards, and credit card records. Building on McFadden's sampling of alternatives theorem, the study establishes the necessary conditions for consistent estimation of model parameters when using this historical cohort approach. Specifically, it demonstrates how assumptions of behavioral, consideration, and attribute invariability can link the latent consideration set problem to the sampling of alternatives problem, enabling consistent parameter estimation. Monte Carlo simulations illustrate the effects of varying sample sizes, the impact of correction methods, and the consequences of assumption violations. Additionally, a variation of Chamberlain's fixed-effects model is proposed to address scenarios where attribute values and choice probabilities vary over time. Practical recommendations are provided to balance the trade-off between data recency and stability of assumptions, offering a robust framework for latent consideration set modeling in diverse applications.

Keywords: Consideration Set, Sampling of alternatives, Discrete Choices

1. Introduction

The consideration set is the reduced set of alternatives that individuals actually evaluate when choosing among large options becomes impractical. The consideration setis latent (not observed) for the researcher and needs to be properly explained in order to achieve the consistent estimation of the parameters of a discrete choice model. Theoretically coherent approaches for the researcher to construct the consideration set are impractical, while the currently available practical methods lack theoretical support.

A practical approach is to form the consideration set from previous (or historical) choices of the same individual, or from choices made by a cohort of individuals who were faced with similar choice situations. Today's passive data sources, from cell phones, travel smart cards, credit cards, or loyalty records, make it possible to have the kind of data needed to build the consideration set in this way, making it an attractive method to use.

This article proposes a theoretical basis for the historical cohort approach by establishing the necessary conditions to achieve consistency in the estimation of model parameters, through the

reinterpretation of the sampling of alternatives theorem (McFadden, 1978). In addition, Monte Carlo tests are provided to illustrate the findings and practical recommendations.

2. Formulation

To explain the problem, consider a random utility model (RUM) configuration in which the utility *Uin* that an individual *n* obtains from alternative *i* can be written as the sum of a systematic part *Vin* and a random error term *εin*, as shown in Eq. (1)

$$U_{in} = V_{in} + \varepsilon_{in} = V(x_{in}, \beta^*) + \varepsilon_{in}, \tag{1}$$

where *Vin* depends on *xin* attributes and population parameters β^*.

Then, if *εin* is distributed *iid* Extreme Value I (0.μ), the probability that *n* chooses the alternative *i* will correspond to the Logit model shown in Eq. (2), where *Cn* is the true consideration set of *Jn* elements from which the individual *n* selects an alternative. The μ scale in Eq. (2) is not identifiable and is therefore usually normalized to 1.

$$P_n(i) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}} \tag{2}$$

Now suppose that the true consideration set*Cn* is latent to the researcher, but that the researcher can observe the individual making *R* choices from the set *Cn* in past instances. The researcher is interested in modeling the choice that occurs in the *R+1* instance. To do this, build a practical consideration set (or a sufficient set in Crawford's (2021) words) that includes all the alternatives that were observed in the previous R instances, plus the alternative chosen in the *R+1* instance, if it was not already included. Other sufficient sets of this type may be possible. These observations could correspond to choices of the same individual *or*, assuming some kind of group homogeneity, choices made by different individuals who faced the same choice situation. For example, Arriagada et al. (2021) use, in a public transport route choice model, the choices made by other individuals travelling in the same pair of ODs of stops and in the same period.

Let us now consider a different (seemingly unrelated) challenge, which was originally addressed by McFadden (1978). In this case it is assumed that the individual can handle the true full consideration set C_n, but instead such a set is too large to be processed by the researcher in practice. This is solved by constructing a reduced practical set $D_n \subseteq C_n$ for estimation, using some conditional sampling protocol to the chosen alternative. Formally, $\pi(D_n \mid j)$ it corresponds to the conditional probability that the researcher will sample a reduced set D_n, given that alternative *j* was chosen by individual *n*. Under this scenario, McFadden (1978) showed that by maximizing a pseudo-log-likelihood, using the probabilities of choice shown in Eq. (3), consistent estimators of the population parameters can be *obtained* β^*.

$$\pi(i \mid D_n) = \frac{e^{V_{in} + \ln \pi(D_n \mid i)}}{\sum_{j \in D_n} e^{V_{jn} + \ln \pi(D_n \mid j)}} \tag{3}$$

The virtue of the pseudo-log-likelihood described in Eq. (3) is that it depends only on the alternatives of the reduced set D_n, transforming a problem of possibly millions of alternatives into one that has only a few. In addition, the resulting model has a closed form that corresponds to a

simple Logit model with an alternative correction term $\ln \pi(D_n \mid j)$ that only depends on the sampling protocol. This result is maintained thanks to the IIA property of the Logit model, but was extended to more flexible models such as MEV (GEV), Logit Mixture and RRM, by Guevara and Ben-Akiva (2013 a,b) and Guevara et al. (2016), respectively.

The link between McFadden's (1978) problem of sampling alternatives and the latent problem of the consideration set requires making some assumptions of invariability. First, if it can be assumed that the behavior of the choice and the consideration set do not vary across R *+1* choices, past choices could be understood as samplings with replacement of the true, but latent, consideration set.

Thus, the process of generating data for the latent consideration set problem can be interpreted as an estimation problem and the sampling of alternatives studied by McFadden (1978). To this end, the sampling protocol in this case will correspond to sampling by importance with replacement, in which the sampling probability of each alternative corresponds to the probability of choice $P_n(i \mid C_n, r)$ in instance *r*.

This result is based on two assumptions of invariability throughout the instances r. The first is behavioral invariability, in the sense that the pattern of choice behavior is the same, and the second is that the consideration set does not change. These assumptions seem easy to sustain in the case of historic elections, if past elections are "recent enough" not to be prone to contextual or behavioral changes. Instead, these assumptions may be more difficult to sustain in the case of cohort choice, but they do not seem implausible.

The problem that remains is to determine the sampling correction needed under this configuration, which is not trivial, not only because of the form it takes, but also because it depends on the probabilities of choice $P_n(i \mid C_n, r)$, which depend on the unknowns C_n and β^*. To solve this dilemma, we need to make an additional assumption.

Sampling correction can be very complicated depending on the protocol considered. McFadden (1978) explores four cases to show some practical examples in which neglect of this correction results in inconsistent estimators. This effort is later deepened by Ben-Akiva and Lerman (1985; Ch.9) and later refined by Ben-Akiva (1989). Among other cases, this last work derived the necessary correction for the following sampling protocol: Sample with replacement *R* times of the set of all alternatives *C* with selection probabilities q_j, with $\sum_{j \in C} q_j = 1$, and add the chosen alternative if it was not already sampled.

This protocol is closely related to the latent considerations set problem studied in this article, but in order to apply it, we need to add a third type of **invariability** assumption $P_n(i \mid r) = P_n(i) = q_j$, which would be achieved if **the attributes** do not vary throughout the past instances of choice. In this case, the necessary sampling correction will correspond to Eq. (4), where n_j is the number of times that alternative j was chosen in R+1 *instances*.

$$\pi(D_n \mid i) = \frac{n_i}{P_n(i)} \left(\frac{R!}{\prod_{j \in D} n_j!} \prod_{j \in D} P_n(j)^{n_j} \right) = \frac{n_i}{P_n(i)} K_D, \tag{4}$$

The assumption of attribute invariability is likely to hold for models based on supermarket data for sufficiently recent periods in which attributes will not change significantly. However, the

validity of this assumption may be more debatable in models based on transport data, which are prone to everyday variability in, for example, travel times unless the choice is considered to be made considering average times, which should be more stable.

Note now that the term $K_D = \frac{R!}{\prod_{j \in D} n_j!} \prod_{j \in D} P_n(j)^{n_j}$ on the right of Eq. (4), although complicated and dependent on $P_n(j)$, does not depend on alternative *i*, and can therefore be omitted from the analysis, since it would be cancelled out when considered in Eq. (3). Furthermore, as the number of instances passed *R* grows, the sampling error of the empirical probability will fade away, n_i/R it will get closer to the choice probability $P_n(i)$, resulting in the entire sampling correction in Eq. (4) not being dependent on alternative *i* and therefore canceled out in Eq. (3) and can be ignored.

$$\pi(D_n \mid i) = K_D \frac{n_i}{P_i(i)} \approx K_D \frac{n_i}{n_i/R} = K_D R \tag{5}$$

Moreover, it is not really necessary for *R* to be large for the correction to cancel out in the face of invariability in the probability of choice and, therefore, the estimation in the usual way results in consistent estimators. In fact, for each individual n it is necessarily true that

$$P(i \mid C_n) = \frac{n_i}{\sum_{j \in D_n} n_j + \delta} = \frac{n_i}{\phi} \quad , \tag{6}$$

where ϕ it is a constant that does not vary between alternatives, for each individual. In this way, replacing (6) in (5) must

$$\pi(D_n \mid i) = K_D \frac{n_i}{\frac{n_i}{\phi}} = \phi K_D \tag{7}$$

which does not vary between alternatives, and therefore, cancels out, even when R is very small.

Therefore, the historical cohort approach to the problem of latent consideration sets will achieve consistency if the assumptions of behavioral, consideration, and attribute invariability are met.

As is often the case, this involves compensation. For example, on the one hand, it would be advisable to construct a set sufficient with options recently enough to meet the invariability of the set of behavior, consideration, and attributes, but also with options old enough to ensure a good approximation of the probability of choice. How and when a sufficient set will adequately meet these conditions should be analyzed on a case-by-case basis.

The article also studies the possibility of establishing a correction, even for the case in which the attributes, and then the probabilities of choice, change throughout the instances, for which a variation of the fixed effects model of Chamberlain (1980) is proposed.

3. Monte Carlo Experiment

The article uses Monte Carlo evidence to illustrate the impact of considering different *R*'s, different approaches to accelerate correction, and the impact of failure of different assumptions.

This Monte Carlo experiment is designed to test methods for investigating the impact of using historical choices to construct the consideration set in choice models. The study is based on a logit model with 1000 individuals and three explanatory variables, examining scenarios with 100 total

alternatives and a consideration set of 10 alternatives. The variables are analyzed under both fixed (X fixed) and variable (X variable) conditions, breaking the invariability assumption. The analysis compares models considering all alternatives, the true model, and those with varying the number of historical choices.

Results are evaluated through boxplots of estimator ratios derived from 100 experiments and an assessment of the coverage of the consideration set. This approach aims to explore the implications of using historical choice data on the performance and validity of choice models.

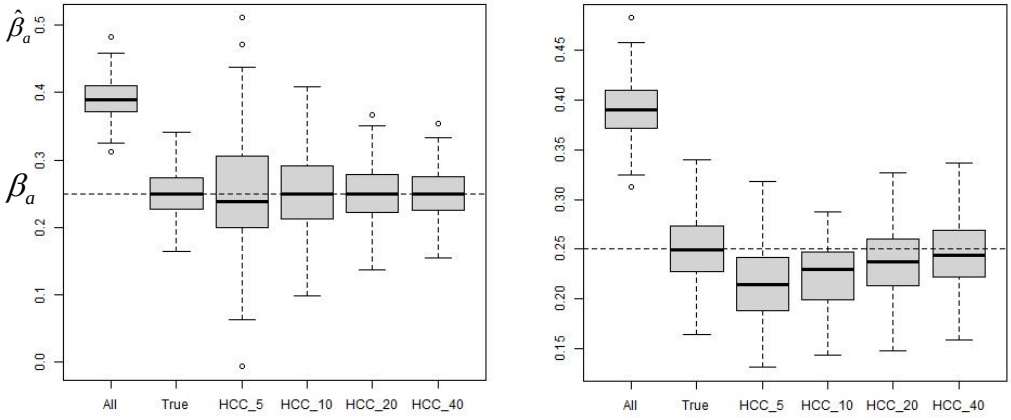

Figure 1: Monte Carlo Analysis (X fixed, X Varia ble)

When the consideration set is replaced by the universal set (as in the All model or Homo Economicus approach), the results demonstrate poor performance. Variance is notably reduced with the historical-consideration model, but this reduction does not hold when explanatory variables are variable. In scenarios where X is fixed, the model performs well even with as few as 5 historical choices, displaying large variance but no bias. However, when X is variable, the model introduces a bias that diminishes with the historical-choices approach, but only if the coverage of the consideration set closely approximates the true consideration set. This highlights the sensitivity of the model to variable explanatory data and the importance of accurately capturing the true consideration set.

4. Conclusión

The consideration set problem remains unresolved, but the historical approach demonstrates concrete advantages. Under reasonable assumptions, the proxy set can be interpreted as a subset of the true set, offering a transparent and interpretable framework by treating the problem as a special case of alternative sampling. This approach is consistent with unlabeled alternatives and performs well empirically in both simulated and real data scenarios. In the labeled case, the method works effectively under the invariability assumption. Despite these strengths, challenges remain: improving the solution when explanatory variables are variable and understanding the effects of using cohorts in the analysis. Addressing these challenges will be crucial to advancing methods for consideration set modeling.

5. References

Arriagada, J, Guevara, C.A. & Munizaga, M. (2021) Evaluating the Historical/Cohort approach for building the consideration-set in route choice modeling using smart card data from a large-scale public transport network. Working paper Universidad de Chile

Ben-Akiva, M. (1989). Lecture Notes of Large Set of Alternatives. Unpublished Manuscript, Massachusetts Institute of Technology.

Ben-Akiva, M., & Lerman, S. (1985). Discrete choice analysis: theory and application to travel demand (MIT press, Ed.; Vol. 9).

Chamberlain, G. (1980). Analysis of covariance with qualitative data. The review of economic studies, 47(1), 225-238.

Crawford, G. S., Griffith, R., & Iaria, A. (2021). A survey of preference estimation with unobserved choice set heterogeneity. Journal of Econometrics, 222(1), 4-43.

Guevara, C. A., Chorus, C. G., & Ben-Akiva, M. E. (2016). Sampling of alternatives in random regret minimization models. Transportation Science, 50(1), 306-321.

Guevara, C. Angelo, Ben-Akiva, Moshe E., 2013a. Sampling of alternatives in logit mixture models. Transp. Res. B 58, 185–198.

Guevara, C. Angelo, Ben-Akiva, Moshe E., 2013b. Sampling of alternatives in multivariate extreme value (MEV) models. Transp. Res. B 48, 31–52.

McFadden, D. 1978. Modeling the choice of residential location. In: Karlqvist, A., Lundqvist, L.,

Snickars, F., Weibull, J. (eds.), Spatial Interaction Theory and Planning Models, Vol. 1. North-Holland, pp. 75–96