Understanding the consequences of attribute non-attendance in discrete choice models

Authors:
Narine Yegoryan, Daniel Guhl and Daniel Klapper

Date: *1.12.2015*

Abstract:
Traditional choice models have been used for inferring the underlying product preference of individuals for more than 50 years. These model outcomes are widely used for efficient pricing, new product development and predicting consumer choice behavior. Increasing number of studies, though, streaming from behavioral economics point towards possible violations of assumptions of traditional choice models. In this paper we try to understand the possible consequences of the violation of full attribute-wise information processing and extend the existing standard model to accommodate the case of attribute non-attendance. For this purpose a latent class approach is utilized with further probabilistic inclusion of eye-tracking data. The results suggest that people do tend to limit their attendance to specific subsets of attributes, mostly two to three. Furthermore, not accounting for attribute non-attendance lead to substantial differences in willingness-to-pay estimates.

Keywords:
Attribute non-attendance, Eye-tracking, Discrete Choice Modeling

1 Introduction

In order to infer the underlying preference formation in traditional choice modeling literature products are described as a bundle of attributes, each of which provides a part-worth utility to the individual (linear additive utility function). In order to infer the specific values that individuals attach to each of the attributes certain assumptions are made regarding the choice process. In particular, it is assumed that consumers choose the alternative that provides maximum utility and deploy full compensatory decision rule. The later implicitly results into the assumption that all the provided information is relevant for all the individuals.

There is an increasing number of studies, though, suggesting that consumers tend to ignore information in choice situations (Orquin & Mueller Loose, 2013) alternative-wise and/or attribute-wise. While the literature on consideration set formation has extensively covered the case of ignoring information alternative-wise (e.g., Swait & Ben-Akiva, 1987), there are fewer studies that tackle attribute-wise information ignorance particularly in marketing.

As a result of partial and imperfect attribute-wise information acquisition only a subset of attributes will have a non-zero effect, a case which standard choice models are not sensitive enough to accommodate even when accounting for preference heterogeneity (Gilbride, Allenby, and Brazell, 2006). It is rather a special case of a "‘structural"’ heterogeneity in choice processes as Kamakura, Kim, and Lee (1996) refer to it.

It can certainly be argued that model outcomes under the assumption of full and perfect information processing nevertheless offer a good approximation (Ben-Akiva *et al.*, 1999), and, therefore, good predictions of average choice probabilities. But the estimated willingness-to-pay (WTP) measures may be biased (Hole, 2011) and can have significant implications on efficient pricing as well as new product development. Furthermore, better knowledge of different attribute attendance strategies applied by consumers can serve as a criteria for segmentation and targeting.

The objective of this paper is to extend the MNL model in a marketing context to incorporate the possibility of attribute non-attendance (ANA) and capture the specific structural heterogeneity. To this end we use a latent class model (Hole, 2011), which results in more accurate estimates and a better representation of the choice process. In addition we also utilize a concomitant variable latent class model (Kamakura, Wedel, and Agrawal, 1994), where eye-tracking data, in particular the number of fixations, is used as an objective measure of attention. This enables taking into account differences in cognitive processing of attributes.

In the next section we will elaborate more on the background of the particular choice of modeling approach. In the third section the methodology will be presented followed by the empirical application in section 4. In the last section implications of the suggested model will be discussed, the limitations and areas for further research will be pointed out.

2 Background

So far in the literature two main approaches of incorporating ANA can be identified. One approach is to set the coefficients of attributes which are not attended to zero. We refer to this as *exogenous* approaches. Non attended attributes are identified using additional survey information such as stated ANA (Hensher, Rose, & Greene, 2005) or eye-tracking data (Meißner, Scholz, & Decker, 2010). The main drawback of such approach is the deterministic nature of incorporation of additional information. For example, in case of stated ANA it can be argued that the stated not attended attribute can have low, but not zero importance.

Another stream of approaches, which we refer to as *endogenous*, allow a more realistic assumption of probabilistic nature of ANA strategies. Within this framework a stream of scholars, mainly in transportation and health economics, has utilized latent class models (e.g., Hensher, Rose, & Greene, 2012; Hole, 2011), where based on many or even all possible attendance/non-attendance combinations individuals are allocated to a priori defined Q latent classes. Hole, Kolstad, and Gyrd-Hansen (2013) have furthermore incorporated stated ANA for more precise estimation of class probabilities. In general, though, it is possible to use other individual level data as concomitant variables, e.g., socio-demographics, psychographics and eye-tracking data. Another endogenous approach, which, to the best of our knowledge, is the only one applied in marketing, is the variable selection model suggested by Gilbride *et al.* (2006). This model, though, is more complex and difficult for application, and leads to similar results as the latent class model (Scarpa, Gilbride, Campbell, and Hensher, 2009).

We suggest application of latent class model with incorporation of eye-tracking data as concomitant variable, which represents a more objective measure of attention and information acquisition (Schulte-Mecklenbeck *et al.*, 2011), as compared to stated ANA it is not dependent on the respondents' recall. Furthermore, considering the existing empirical evidence linking visual attention to the likelihood of choice (Pieters & Warlop, 1999), we introduce an additional knot in this causal line: visual attention drives attribute attendance probabilities, which in turn, might result in a higher likelihood of choice.

Therefore, the suggested method aims to fill the existing gap of attribute non-attendance studies in marketing as well as further improve the method by using eye-tracking information as explanatory variable in probabilistic mixing-function, indicating the prevalence of specific attribute attendance strategies.

3 Methodology

The suggested latent class framework extends the traditional multinomial logit model (MNL) in order to relax the assumption of full and perfect information processing. We, therefore, follow bounded rationality literature, which states that individuals can still act rationally in terms of maximizing their utility, but can do so based on partial and imperfect information (Rasouli & Timmermans, 2015). Latent classes are defined a priori covering all the possible attribute attendance combinations (i.e., we have $Q = 2^k$ classes, where k represents the number of attributes). Within each class (c) we specify for each alternative j a linear additive utility function, $U_j = \beta_c' \cdot x_j + \varepsilon_j$. While the parameters are common among the classes, each class defines a different subset of attributes and, therefore, has a different vector of parameters. For example in case of 2 attributes $Q = 2^2 = 4$ classes would be formed with the following vector of parameters $\beta_c:\ \beta_1 = (\beta^1, \beta^2),\ \beta_2 = (\beta^1, 0),\ \beta_3 = (0, \beta^2),\ \beta_4 = (0, 0)$.

The latent class model is, therefore, a more flexible modeling approach allowing incorporation of different types of decision rules including full compensatory, partial-compensatory (compensatory rule applies only within the subset of attributes), lexicographic decision rule (the best alternative on the most important attribute, is chosen) as well as random choice.

Within each class the traditional MNL model is utilized. Therefore the probability that decision-maker i chooses alternative j on choice occasion t conditional on the class c is:

$$P(y_{it} = j|c) = \frac{\exp(\beta_c' \cdot x_{ijt})}{\sum_{j' \in J} \exp(\beta_c' \cdot x_{ij't}')}. \qquad (1)$$

The marginal probability that i chooses j is:

$$P(y_{it}=j)=\sum_{c\in C}\pi_{ic}\cdot P(y_{it}=j|c) \tag{2}$$

Following Hole (2011) we specify the submodel for the class probability π_{ic} as:

$$\pi_{ic}=\prod_{a\in k}\frac{\exp\left(\gamma_a'\cdot z_{ia}\right)}{1+\exp\left(\gamma_a'\cdot z_{ia}\right)}\cdot\prod_{a\notin k}\frac{1}{1+\exp\left(\gamma_a'\cdot z_{ia}\right)}. \tag{3}$$

where z_{ia} contains class-/ attribute-specific *intercepts* and (possibly) *individual-level variables* such as demographics, eye-tracking, or survey variables. i represents an index for individuals and can be dropped in case individual-level data is not available.

The submodel in (3) is based on the underlying assumption of *independence of attribute attendance*, i.e., the probability of one attribute being considered is independent of other attributes being considered and is closely related the model proposed by Swait & Ben-Akiva (1987) for modeling parsimoniously choice set heterogeneity. This assumption allows estimation of a latent class model in case of many attributes by reducing the number of additional parameters to be estimated to the number of attributes plus the number of additional individual level variables in z_{ia}. We further refer to this model as IANA (Independent Attribute Non-Attendance). The model is estimated by the method of maximum likelihood.

4 Empirical Application

4.1 Data

The single cup coffee makers dataset obtained from Meißner, Musalem & Huber (2015)[1] is a combination of a choice-based conjoint study with eye-tracking, which was carried out at a large European university, and includes a sample of 59 respondents. The conducted *CBC study* consists of 12 choice tasks, within each of which respondents chose between three products and a *no-choice option*. The CBC was designed to include 6 attributes: Brand (*Braun, Krups, Philips, Severin*), Material (*stainless steel, plastic, aluminium*), System (*pad, capsule*), Design (*A, B, C, D*), Price per cup (*0.12, 0.22, 0.32 Eurocents*) and Price (*99, 129, 159, 189 Euros*).

From the simultaneous *eye-tracking* for each individual number of fixations on each attribute are calculated. Only number of fixations is used as an additional individual-level variable in z_{ia} as it is a common indicator of attention in the literature, and is highly correlated with other measures of attention (e.g., fixation duration). Moreover, the variable has been standardized within each respondent, as it is not the absolute number of fixations but the allocation of fixations among the attributes that is of interest. This further enables to avoid possible confounding with heterogeneity across individuals.

4.2 Results

In total three models have been estimated (using several randomized starting values): a simple MNL model, which serves as a benchmark, and two IANA models, one without any additional individual-level variables and one with incorporation of the standardized measure of number of fixations. In the Table 1 the results of three estimated models are presented.

---

[1] We cordially thank the authors for sharing the dataset with us.

utility parameters	MNL	IANA (w/o fixation)	IANA (with fixation)
no-choice	-0.195^{*}	0.333^{**}	0.300^{**}
Braun	0.061	1.513^{***}	0.620
Krups	0.011	0.694	0.354
Philips	0.225^{***}	0.994^{*}	0.601^{*}
stainless steel	0.508^{***}	1.519^{***}	1.423^{***}
plastic	-0.505^{***}	-1.668^{***}	-1.505^{***}
pad	0.218^{***}	1.735^{***}	1.492^{***}
design A	-0.292^{***}	-2.199^{***}	-1.764^{***}
design B	0.034	0.324	0.231
design C	0.134	1.400^{***}	1.241^{***}
price per cup	-8.031^{***}	-17.557^{***}	-17.313^{***}
price	-2.123^{***}	-3.722^{***}	-3.675^{***}
class parameters			
brand		-2.078^{***}	0.058
material		-0.152	0.262^{***}
system		-1.421^{***}	-1.828^{***}
design		-1.762^{***}	-2.162^{***}
price per cup		0.570	-0.147
price		1.604^{***}	1.195^{**}
fixation			1.994^{***}
LL	-745.114	-688.888	-655.656
BIC	1568.977	1495.901	1435.999
ρ^2	0.232	0.298	0.332

Signif. codes: $*$: $p < 0.10$ $**$: $p < 0.05$ $***$: $p < 0.01$

Table 1: Estimation results

Certain general patterns can be observed in all three models, such as negative utility parameters for price and price per cup, as well as overall preference of stainless steel material, pad over capsule system and Design C. There is a difference, though, in implied preferences of brands.

IANA models have clearly better model fit compared to MNL. First of all, Log-likelihood (and therefore ρ^2) significantly improves (LL of -688.888 and -655.656 compared to -754.114). Second, BIC decreases for IANA models indicating better fit even though more parameters are used. Third, larger scales of utility parameters of IANA models imply smaller variance of the logit error.

Incorporation of the standardized measure of number of fixations not only improves the model fit (LL of -655.656 compared to -688.888), but also results into a significant and strong positive effect. This implies that (relatively) more fixations result in higher likelihood of incorporating the attribute in the decision making. This effect is further illustrated in Figure 1 (right panel, dots indicate avg. value over respondents). The large negative class parameters of attributes system and design have resulted in a shift of the curve to the right, implying lower attribute attendance probability for a given number of fixations. On the other hand, larger positive class parameters of price and material shift the curve to the left implying higher attribute attendance probabilities. In general, price, price per cup and material are the most attended attributes, while brand, system and design are the least attended. The avg. values are 89.82%, 64.89%, 50.77%, 20.61%, 16.59% and 7.75% attendance probabilities respectively.

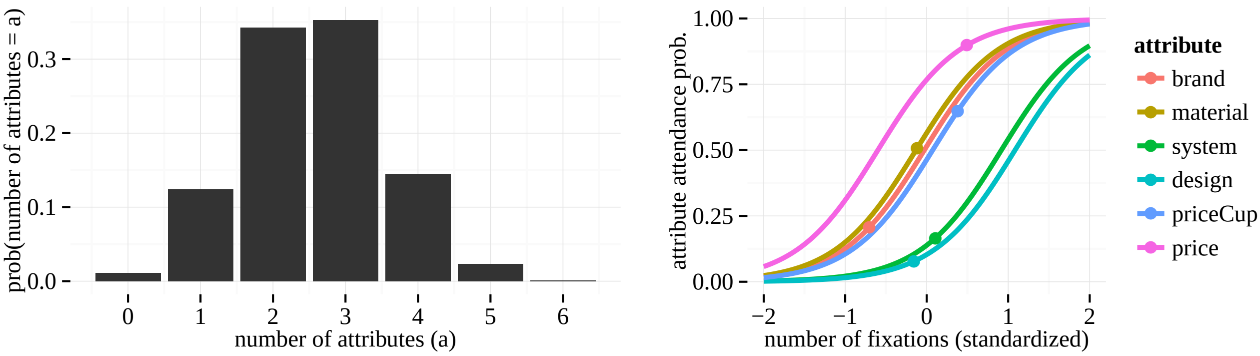

Figure 1: Distribution of the number of attributes (left) and attribute probabilities as a function of fixations (right)

Based on the class parameters class probabilities have been computed according to which only a very small proportion of individuals takes all the attributes into account (0.12%). Most of the respondents exhibit partial-compensatory behavior: 15.87% of individuals compensate among material, price per cup and price, and 15.42% only among price per cup and price (these are the two largest classes). The class with full attendance makes up only 0.12% of the whole sample, while the class, where random choice is applied 1.14%. In general, majority of individuals ($\approx$ 70%) attend only 2-3 attributes (see Figure 1, left panel).

The deviation from full attribute-wise information processing has an important implications on derived willingness-to-pay (WTP) values. To illustrate this we compare the derived WTP for brands in case of MNL and IANA (with fixation). In case of standard MNL the resulting WTP values are average values for the overall sample. In case of IANA, though, we follow Hole (2011) for deriving WTP values, and include only the individuals that attend both brand and price (23.6% of the sample). Otherwise, if brand is not attended than WTP is zero. As non-attendance of price can be due to the particular experimental setting, where not all the respondents' price range was covered, exclusion of these cases eliminates a possible bias.

The resulting WTP values are substantially different from those of standard MNL. In particular, in case of IANA 23.6% of the sample are willing to pay 59.24 € more for Philips than Severin compared to only 24.61 € in case of MNL. Moreover, according to MNL Philips is preferred over Braun by 7.72 €, while in case of IANA both brands have rather similar preferences with WTP difference for Braun of only 0.50 €. In case we calculate the weighted avg. WTP values for the IANA model, we have a 16.01 € difference between Philips and Severin. These results might have important managerial implications. In case firms are interested in targeting specific consumers that value brand, by using MNL model, it clearly underestimates the WTP of consumers, which can further result in inefficient pricing strategies. However, it overestimates avg. WTP values in the population.

5 Conclusion

In summary, the suggested latent class model allows relaxation of the restrictive assumption of full attribute-wise information processing and, therefore enables incorporation of a specific type of structural heterogeneity. The results strongly indicate existence of the deviation from this assumption and have significant implications on the derived willingness-to-pay. The more general latent class model, further results in better model fit and reduces the error in MNL model. Additionally probabilistic incorporation of standardized variable of number of fixations in modeling class probabilities allows further increase in model fit, as well as allows taking into account differences in cognitive processing of various attributes.

One of the main limitations is the small sample size. Additionally, application of the suggested approach to other product categories is required and versions of the model accounting for preference heterogeneity would be worth exploring. In addition incorporation of other exogenous information besides the number of fixations (e.g., stated ANA, socio-demographics) can be explored.

References

Ben-Akiva, M., McFadden, S., Gärling, T., Gopinath, D., Walker J., Bolduc, D., Börsch-Supan, A., Delquié, P., Larichev, O., Morikawa, T., Polydoropoulou, A., & Rao, V. (1999). Extended Framework for Modeling Choice Behavior. *Marketing Letters*, 10, 187-203.
Gilbride, T. J., Allenby, G. M., & Brazell, J. D. (2006). Models for Heterogeneous Variable Selection. *Journal of Marketing Research*, 43, 420-430.
Hensher, D. A., Rose, J., & Greene, W. H. (2005). The Implications on Willingness to Pay of Respondents Ignoring Specific Attributes. *Transportation*, 32, 203-222.
Hensher, D. A., Rose, J., & Greene, W. H. (2012). Inferring Attribute Non-Attendance from Stated Choice Data: Implications for Willingness to Pay estimates and a Warning for Stated Choice Experiment Design. *Transportation*, 39, 235-245.
Hole, A. R. (2011). A Discrete Choice Model with Endogenous Attribute Attendance. *Economics Letters*, 110, 203-205.
Hole, A. R., Kolstad, J., & Gyrd-Hansen, D. (2013): Inferred vs. stated attribute non-attendance in choice experiments: A study of doctors' prescription behaviour. *Journal of Economic Behavior & Organization*, 96, 21-31.
Kamakura, W. A., Kim, B., & Lee, J. (1996). Modeling preference and structural heterogeneity in consumer choice. *Marketing Science*, 15, 152-172.
Kamakura, W.A., Wedel M., & Agrawal J. (1994). Concomitant variables in latent class models for conjoint analysis. *International Journal of Research in Marketing*, 11, 451-464.
Meißner, M., A. Musalem, & J. Huber (2015). Eye-Tracking Reveals Processes that Enable Conjoint Choices to Become Increasingly Efficient with Practice. *Journal of Marketing Research* (forthcoming).
Meißner, M., Scholz, S. W., & Decker, R. (2010). Using Eye Tracking and Mouselab to Examine How Respondents Process Information in CBC. In Sawtooth Software (ed.) *Proceedings of the 15th Sawtooth Software Conference* (p. 151). Newport Beach, California.
Orquin, J.L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, 44, 190-206.
Pieters, R., & L. Warlop (1999). Visual Attention during Brand Choice: The Impact of Time Pressure and Task Motivation. *International Journal of Research in Marketing*, 16, 1-16.
Rasouli, S., & Timmermans, H. (2015). Bounded Rational Choice Behaviour: Applications n Transport. Emerald Group Publishing Limited.
Scarpa, R., Gilbride, T.J., Campbell, D., & Hensher, S.A. (2009). Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics*. (February), 1-24.
Schulte-Mecklenbeck, M., Kühberger, A., & R. Ranyard (2011). The Role of Process Data in The Development and Testing of Process Models of Judgment and Decision Making. *Judgment and Decision Making*, 6, 733-739.
Swait, J., & Ben-Akiva, M. (1987). Incorporating random constraints in discrete choice models of choice set generation. *Transportation Research*, 21B, 91-102.