When Zeros Count:

Confounding in Preference Heterogeneity and Attribute Non-attendance

Narine Yegoryan[a,*] Daniel Guhl[a] Friederike Paetz[b]

[a] *Institute of Marketing, School of Business and Economics, Humboldt University Berlin, Spandauer Str. 1, 10178 Berlin, Germany.*

[b] *Institute of Management and Economics, Department of Management and Marketing, Clausthal University of Technology, Julius-Albert-Str. 6, 38678 Clausthal-Zellerfeld, Germany.*

[*] Corresponding author: narine.yegoryan@hu-berlin.de (Narine Yegoryan)

Achnowledgments: The authors would like to thank Garrett Glasgow, Michael Keane, Liu Qing, and Christian Schlereth, for sharing the data that was imperative for the empirical application. Financial support by German Research Foundation (DFG) through CRC TRR 190 is gratefully acknowledged.

When Zeros Count:
Confounding in Preference Heterogeneity and Attribute Non-attendance

Abstract

Identifying consumer heterogeneity is a central topic in marketing. While the main focus has been on developing models and estimation procedures that allow uncovering consumer heterogeneity in preferences, a new stream of literature has focused on models that account for consumers' heterogeneous attribute information usage. These models acknowledge that consumers may ignore subsets of attributes when making decisions, also commonly termed "attribute non-attendance" (ANA). In this paper, we explore the performance of choice models that explicitly account for ANA across ten different applications, which vary in terms of the choice context, the associated financial risk, and the complexity of the purchase decision. We systematically compare five different models that either neglect ANA and preference heterogeneity, account only for one at a time, or account for both across these applications. First, we showcase that ANA occurs across all ten applications. It prevails even in simple settings and high-stakes decisions. Second, we contribute by examining the direction and the magnitude of biases in parameters. We find that the location of zero with regard to the preference distribution affects the expected direction of biases in preference heterogeneity (i.e., variance) parameters. Neglecting ANA when the preference distribution is away from zero, often related to whether the attribute enables vertical differentiation of products, may lead to an overestimation of preference heterogeneity. In contrast, neglecting ANA when the preference distribution spreads on both sides of zero, often related to attributes enabling horizontal differentiation, may lead to an underestimation of preference heterogeneity. Lastly, we present how the empirical results translate into managerial implications and provide guidance to practitioners on when these models are beneficial.

Keywords: Choice modeling, Preference heterogeneity, Attribute non-attendance, Inattention

1 Introduction

Understanding consumer heterogeneity is of utmost importance for marketers in a wide variety of decisions, including market segmentation and targeting, new product development, as well as pricing (Allenby and Rossi 1998; Allenby, Brazell, Howell, and Rossi 2014).

Most of the efforts in marketing literature have focused on models that accommodate consumer preference heterogeneity. First, the latent class model by Kamakura and Russell (1989) and later, the mixed multinomial logit (MMNL) model (e.g., McFadden and Train 2000) superseded the multinomial logit (MNL) model (McFadden 1974) as the new standard. Further advances have been made to account for even more flexible forms of preference heterogeneity by using a mixture of normals distribution (e.g., Allenby, Arora, and Ginter 1998, Rossi, Allenby, and McCulloch 2005, Burda, Harding, and Hausman 2008), Dirichlet process prior, and more recently, Dirichlet process mixture (e.g., see Voleti, Srinivasan, and Ghosh 2017). These models aim to accommodate multi-modal parameter distribution.

Nevertheless, consumers may differ not only in how much they value specific product attributes (i.e., preference heterogeneity) but also in how they make decisions (Kamakura, Kim, and Lee 1996) and, particularly, whether they value these attributes at all. The latter is of particular importance for marketers to understand and leverage. Imagine a laptop manufacturer considering introducing a feature that lets the user switch off the camera in its new model. Privacy-concerned consumers would find this feature valuable. However, some consumers may not care about this feature and ignore it when choosing which laptop to purchase. Generally, consumers may ignore subsets of product attribute information when making purchase decisions. In transportation science and health economics literature, such behavior is commonly referred to as "attribute non-attendance" (ANA)[1]. In our example, let us assume that around 20% of consumers ignore the camera off switch (i.e., around 20% of ANA). We illustrate the potential distribution of preference parameters for this feature in the left-hand panel of Figure 1. The dashed black line represents the mean partworth utility (excluded zeros), and the solid black line – the density of the actual preference distribution.

If we neglect the fact that some consumers ignore this feature, the uncovered (normal) distri-

[1] As much of the relevant literature comes from these fields, we adopt this terminology for the rest of the paper.

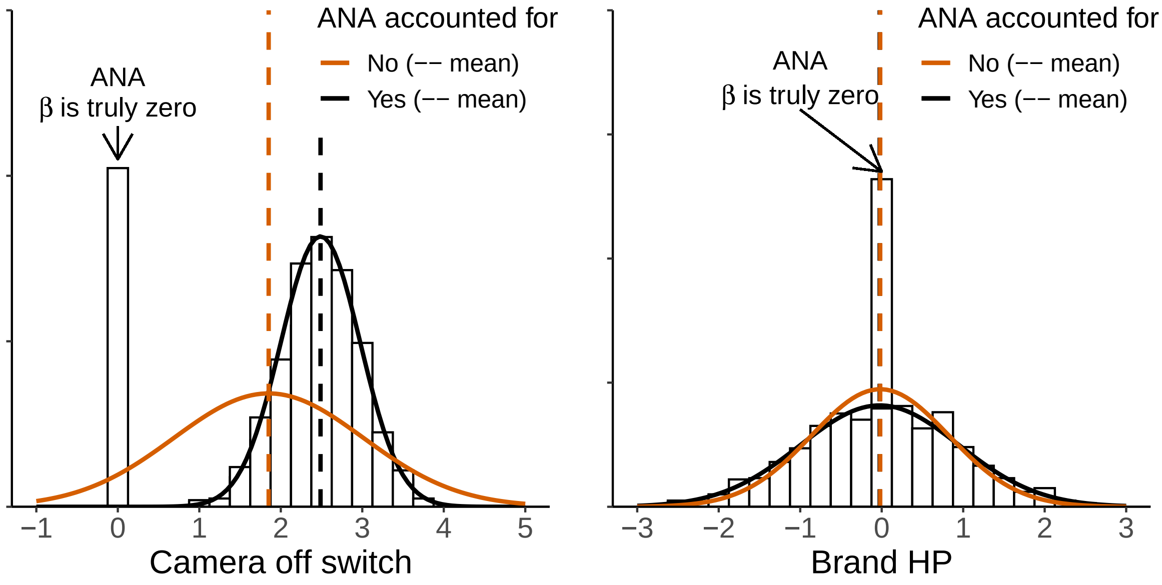


Figure 1: Potential biases when ignoring attribute non-attendance with 20% ANA

bution of preference parameters will be shifted, such that the mean will be biased towards zero and the variance – overstated (the red line in Figure 1). Were the actual parameter distribution located further to the right (i.e., further away from zero), we could expect an increase in the magnitude of the bias in both the mean and the variance when neglecting ANA. Similarly, an increase in the amount of ANA would lead to an increase in the magnitude of the bias.

Now consider that the laptop manufacturer is also interested in understanding preferences for display size of 14″ vs. 15″. Some consumers may prefer a smaller screen, while others – a larger one. We illustrate this example by plotting simulated preference values on both positive and negative domains in the right-hand panel of Figure 1. Again, let us assume around 20% of ANA, captured by precisely zero utility estimates. In this case, we do not observe much difference in the mean of the implied distribution that neglects ANA (the red line) and the actual distribution (the black line). However, the implied distribution results in smaller variance. With an increase in the amount of ANA, we can expect the variance to get even smaller.

ANA may arise due to various reasons. Consumers may ignore subsets of product attributes as they do not find them relevant (Gilbride, Allenby, and Brazell 2006), or they may be simplifying their decision due to its complexity and limited cognitive resources (Payne, Bettman, Coupey, and Johnson 1992). In any case, ANA violates one of the underlying assumptions we make when modeling choice behavior, the assumption of full information processing. As different consumers may care about different subsets of attributes, this leads to differences in the composition of the utility function on the individual level. Identifying ANA and accounting for it in choice models is important. As the simple example above illustrates, the estimated parameter distribution may be biased otherwise, leading to suboptimal marketing decisions.

Several approaches have been put forward that explicitly model ANA in marketing literature and adjacent fields of operation research, transportation science, and health economics. Some of these approaches only account for ANA but ignore preference heterogeneity (e.g., Hole 2011). Others simultaneously account for both ANA and preference heterogeneity (e.g., Gilbride et al. 2006; Hole, Kolstad, and Gyrd-Hansen 2013). For example, Gilbride et al. (2006) propose a heterogeneous variable selection model in the Bayesian estimation framework. Hole et al. (2013) and Hess, Stathopoulos, Campbell, and Caussade (2013) apply a latent class approach similar to the consideration set model by Swait and Ben-Akiva (1987). Here, latent classes represent potential combinations of subsets of attributes that consumers may ignore, i.e., each class describes a specific attribute processing strategy. Maldonado, Montoya, and Weber (2015) and Maldonado, Montoya, and López (2017) apply feature selection tools from machine learning, particularly support vector machine algorithm, for inferring ANA.

In all these applications, models that account for ANA outperform their "standard" counterparts that operate under the assumption of full information processing. Several applications have reported biases in the estimates that may arise when models neglect ANA (e.g., Gilbride et al. 2006; Hess et al. 2013; Hole et al. 2013). In particular, all the applications find that mean estimates of preference distribution are biased towards zero in line with our simple illustration in Figure 1. Regarding the biases in the variance, findings in previous literature do not allow any generalizations. In a marketing application, Gilbride et al. (2006) find that the amount of heterogeneity is predominantly understated. Due to the proprietary nature of the data, we do not have information on the specific product category or attributes in this application. By contrast, in applications in the field of transportation science, where ANA is a much more prominent topic, often overestimation of the amount of heterogeneity is reported (e.g., Hess et al. 2013; Collins 2012). These applications commonly deploy attributes for which one expects a clear preference direction (e.g., price or time). In marketing applications, these are often attributes (as our example of the camera off-switch feature for laptops) that allow firms to differentiate their products vertically. However, marketing applications also include such attributes as brand, color, or as in our example, display size for laptops, for which parameter distribution can span both positive and negative domains. Such attributes allow firms to differentiate their products horizontally. We expect that for these attributes, we would observe the variance and, therefore, the amount of heterogeneity to be underestimated when neglecting ANA, as our simple example for display size of laptops illustrates (the right-hand panel in Figure 1). Therefore, this paper examines the direction and the magnitude of biases in the estimated mean and variance of parameter distribution when neglecting ANA in choice models.

Furthermore, Hess et al. (2013) and Hole et al. (2013) outline that biases may arise when models account for ANA but neglect preference heterogeneity. In particular, they find that assuming homogeneous preferences across respondents may lead to an overstatement of ANA, as consumers with low sensitivity to an attribute might be treated as having zero sensitivity. Hess et al. (2013) and Hole et al. (2013), in their applications in the context of route choice and prescription drug choice, respectively, find that after accounting for preference heterogeneity, the amount of uncovered ANA either considerably decreases or completely diminishes. By contrast, Gilbride et al. (2006), in an unknown product choice context, as well as Yegoryan, Guhl, and Klapper (2020), in their application in the context of choices for coffee makers and laptops, find probabilities of ANA that exceed 45%. Due to the limited number of applications, it remains an open question how much of an issue ANA is in different marketing contexts. In particular, there is considerable variation when it comes to consumer involvement, the stakes or financial risk of the purchase, and the complexity of the decisions in marketing contexts. Hence, another objective of this paper is to understand the ANA's prevalence in different contexts, settings, and attributes.

In summary, in this paper, we aim at a deeper understanding of the potential confounding of preference heterogeneity and ANA. In particular, our objective is to shed more light on the direction and the magnitude of biases we can expect for different types of attributes. Building upon previous literature and as outlined in the illustrative example, we expect that the amount of ANA and the location of the actual preference distribution with respect to zero will affect the magnitude of the bias in the mean and both the direction and the magnitude of the bias in the variance of preference distribution. For this purpose, we set out to compare various models that account for neither preference heterogeneity nor ANA, only ANA or preference heterogeneity, or both in different empirical applications. Our primary focus is understanding the patterns of preferences different models can identify.

We use ten different datasets from choice-based conjoint (CBC) studies conducted in the context of durable products (e.g., laptops), fast-moving-consumer goods (FMCG; e.g., packaged orange juice), as well as entertainment and experiential goods and services (e.g., video-streaming services). Purchase decisions in these contexts carry different financial risks (choosing a package of orange juice is a low-stake decision compared to choosing a laptop). The datasets also include a different number of attributes and, therefore, vary in the complexity of the decision. Consumer involvement and the stakes of the choice decision have long been recognized to affect consumers' information search (e.g., Laurent and Kapferer 1985) and may drive ANA. Furthermore, complexity and information overload may also prompt consumers to ignore some information in the decision-making (e.g., Payne et al. 1992, Bettman, Luce, and Payne 1998, Orquin, Bagger, and Loose 2013). Hence, differences in these characteristics of the datasets also enable us to examine the prevalence of ANA in various purchase situations.

The rest of the paper is structured as follows. In the next section, we describe the methodology, including the models we are interested in comparing and the estimation procedure. In section 3, we start by presenting our ten empirical applications and their characteristics in more detail. We then discuss our findings, including the comparison of in- and out-of-sample model fit and the effects of ignoring either preference heterogeneity or ANA. Finally, we outline managerial implications and conclude with a summary and an outline of avenues for future research.

2 Methodology

We start this section by describing the standard MNL model (McFadden 1974). We then proceed to describe the models that build upon and extend the MNL model to account for (flexible) forms of unobserved preference heterogeneity and ANA (see Elshiewy, Guhl, and Boztuğ 2017 for a comprehensive review on MNL models in marketing). We conclude the section by discussing the estimation procedure.

2.1 Multinomial Logit Model

The utility individual i ($i = 1,\ldots,I$) obtains from alternative j ($j = 1,\ldots,J$) in choice task t ($t = 1,\ldots,T$) is given by:

$$U_{ijt} = \boldsymbol{x}_{ijt} \cdot \boldsymbol{\beta} + \varepsilon_{ijt}, \tag{1}$$

where $\boldsymbol{x}_{ijt}$ is a K-dimensional row vector of attribute values describing alternative j in choice task t for individual i, $\boldsymbol{\beta}$ is a column vector of corresponding preference parameters, which are homogeneous across consumers, and $\varepsilon_{ijt} \sim$ is i.i.d type I extreme value error term. Given the distributional assumptions of the error term, the probability of individual i choosing alternative j in choice task t has the following closed-form:

$$P_{ijt} = \frac{\exp(\boldsymbol{x}_{ijt} \cdot \boldsymbol{\beta})}{\sum_{j' \in J} \exp(\boldsymbol{x}_{ij't} \cdot \boldsymbol{\beta})}, \tag{2}$$

leading to the following likelihood function for individual i:

$$L_i = \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt}^{y_{ijt}}, \tag{3}$$

where y_{ijt} is a dummy indicating whether individual i chose the alternative j in choice task t.

2.2 Models Accounting for Preference Heterogeneity

To account for unobserved preference heterogeneity, we can extend the MNL model by assuming that preference parameters are individual-specific, denoted by adding the i index to $\boldsymbol{\beta}_i$ in Equation (1). Assuming that $\boldsymbol{\beta}_i$ are drawn from a particular continuous distribution (e.g., normal, log-normal, truncated; McFadden and Train 2000, Train 2009) and retaining the distributional assumptions on the error term, the mixed multinomial logit (MMNL) model is derived. As commonly done (Keane and Wasi 2013), we assume that $\boldsymbol{\beta}_i$ is distributed multivariate normal with mean $\bar{\boldsymbol{\beta}}$ and covariance $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\beta}_i \sim N(\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma})$ with coefficients being constant over T.

In the MMNL model, choice probabilities are defined as:

$$P_{ijt} = \int \frac{\exp(\boldsymbol{x}_{ijt} \cdot \boldsymbol{\beta}_i)}{\sum_{j' \in J} \exp(\boldsymbol{x}_{ij't} \cdot \boldsymbol{\beta}_i)} \phi(\boldsymbol{\beta}_i | \bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_i, \tag{4}$$

and the likelihood function can be written as:

$$L_i = \int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt}^{y_{ijt}} \phi(\boldsymbol{\beta}_i | \bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_i. \tag{5}$$

Note that by setting all the elements of $\boldsymbol{\Sigma}$ to zero, the MMNL model reduces to the standard MNL model with homogeneous preferences.

The first applications of the MMNL became only possible after the development of simulation methods (see McFadden and Train 2000 and Train 2009). Since then, the MMNL model has replaced the MNL model as the standard and become one of the most popular models used (Elshiewy et al. 2017).

Further extensions of the model have been proposed that allow capturing even more flexible forms of heterogeneity. We will mainly focus on the "mixed-mixed" multinomial logit (MM-MNL) model[2], which assumes that the mixing distribution of $\boldsymbol{\beta}_i$ is a discrete mixture of normals, i.e., $\boldsymbol{\beta}_i \sim N(\bar{\boldsymbol{\beta}}_q, \boldsymbol{\Sigma}_q)$ with w_q as the probability of class q ($q = 1, \ldots, Q$), $\sum_{q=1}^{Q} w_q = 1$ and $w_q > 0$ $\forall q$. Class probabilities w_q can be modeled as:

$$w_q = \frac{\exp(w_q^*)}{1 + \sum_{q=2}^{Q} \exp(w_q^*)}. \tag{6}$$

where w_q^* is a vector of class-specific intercepts. Note that this specification ensures that $\sum_{q=1}^{Q} w_q = 1$ (Keane and Wasi 2013).

In the MM-MNL model, the choice probability of individual i for alternative j in the choice task t is a weighted sum across classes:

$$P_{ijt} = \sum_{q=1}^{Q} w_q \cdot \int \frac{\exp(\boldsymbol{x}_{ijt} \cdot \boldsymbol{\beta}_{i|q})}{\sum_{j' \in J} \exp(\boldsymbol{x}_{ij't} \cdot \boldsymbol{\beta}_{i|q})} \phi(\boldsymbol{\beta}_{i|q} | \bar{\boldsymbol{\beta}}_q, \boldsymbol{\Sigma}_q) d\boldsymbol{\beta}_{i|q}, \tag{7}$$

where $\phi(\boldsymbol{\beta}_{i|q} | \bar{\boldsymbol{\beta}}_q, \boldsymbol{\Sigma}_q)$ is the normal density with mean $\bar{\boldsymbol{\beta}}_q$ and $\boldsymbol{\Sigma}_q$ in class q. Hence, the likelihood

[2]We adopt the terminology used by Keane and Wasi (2013).

function for individual i is given by:

$$L_i = \sum_{q=1}^{Q} w_q \cdot \int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt}^{y_{ijt}} \phi(\boldsymbol{\beta}_{i|q} | \bar{\boldsymbol{\beta}}_q, \boldsymbol{\Sigma}_q) d\boldsymbol{\beta}_{i|q}. \tag{8}$$

The MM-MNL model necessitates estimating additional $(Q-1)$ class parameters. If w_q for all but one class tends to zero, the MM-MNL model reduces to the MMNL model. The MM-MNL can generally approximate any distribution arbitrarily well (Keane and Wasi 2013). It has been shown to outperform the MMNL model both in- and out-of-sample (e.g., Keane and Wasi 2013, Rossi et al. 2005, Burda et al. 2008, as well as Voleti et al. 2017). Moreover, Keane and Wasi (2013) find that the MM-MNL model can capture more "extreme" patterns of consumer behavior, such as the use of lexicographic rules, as well as "random" behavior.

2.3 Model Accounting for Attribute Non-attendance

In this section, we describe the models that account for ANA. In particular, we start with the endogenous attribute attendance (EAA) model proposed by Scarpa, Gilbride, Campbell, and Hensher (2009) and Hole (2011). The EAA model is a confirmatory latent class approach (Hess et al. 2013). It builds upon the MNL model but introduces a structure by a priori defining latent classes, which describe all potential combinations of attribute processing strategies that can account for people ignoring some information. Given K attributes, 2^K latent classes are defined. In each class s ($s = 1, \ldots, S$), the specific attribute processing strategy can be described by a K-dimensional column vector $\boldsymbol{\lambda}_s = [\lambda_{s1}, \ldots, \lambda_{sK}]'$, where λ_{sk} is a dummy indicating whether class s includes attribute k ($\lambda_{sk} = 1$) or not ($\lambda_{sk} = 0$). Accordingly, the model implies a class-specific utility function:

$$U_{ijt|s} = \boldsymbol{x}_{ijt} \cdot \boldsymbol{\beta}_s + \varepsilon_{ijt}, \tag{9}$$

with $\boldsymbol{\beta}_s = \boldsymbol{\lambda}_s \circ \boldsymbol{\beta}$, τ_s as the probability of class s, $\sum_{s=1}^{S} \tau_s = 1$, and $\tau_s > 0\ \forall s$. As in Equation (1), $\boldsymbol{\beta}$ is a column vector of preference parameters. However, through the elementwise multiplication (denoted by $\circ$) with the indicator vector $\boldsymbol{\lambda}_s$, parameters for attributes that are not included in class s are set to zero, resulting in a class-specific vector of parameters $\boldsymbol{\beta}_s$. For accommodating categorical attributes, effects coding should be used[3]. Moreover, as now multiple elements in $\boldsymbol{x}_{ijt}$ will be related to an attribute, the $\boldsymbol{\lambda}_s$ vector should be extended and mapped onto the correct parameter dimensions. If the attribute k is ignored, all its m_k levels are ignored, and all the corresponding $\boldsymbol{\lambda}_s$ elements should be set to zero.

As parameters are switched on and off, a different linear (additive) utility function characterizes each class. The EAA model accommodates several decision rules, including full compensatory, when all attributes are attended, semi-compensatory, when two or more but not all attributes are attended, (a probabilistic version of) lexicographic rule, when only one attribute is attended, and the random choice, when none of the attributes is attended.

Retaining the distributional assumptions of the error term, the choice probability of individual i of alternative j in choice task t is now conditional on class s:

$$P_{ijt|s} = \frac{\exp(\boldsymbol{x}_{ijt} \cdot \boldsymbol{\beta}_s)}{\sum_{j' \in J} \exp(\boldsymbol{x}_{ij't} \cdot \boldsymbol{\beta}_s)}. \tag{10}$$

We can, of course, define the submodel of class probabilities as in Equation (6). However, it would require estimating $(S-1)$ additional class parameters. As $S = 2^K$, it increases exponentially with the number of attributes. Already with $K = 6$ attributes, a typical case in CBC studies (Rao 2014), we would end up with 63 additional parameters. Such operationalization may reduce model stability and result in only marginal improvements in fit due to loss of parsimony (Hess et al. 2013).

In contrast, we follow Hole (2011) and make a more restrictive assumption that the probability of attending one attribute is independent of the probability of attending another attribute. While such an assumption, which Hole (2011) refers to as independence of attribute attendance (IAA) assumption, may seem restrictive, it results in a more parsimonious model. More specifically, by utilizing the IAA assumption, we can model the class probabilities τ_s as a mapping

[3]Note that dummy coding for categorical attributes is not appropriate in these types of models (Gilbride et al. 2006). The zero value of the preference parameter has a particular meaning in the EAA model, which is that the attribute is ignored. In the case of dummy coding, though, the estimate of the base level is automatically set to zero.

from attribute attendance probabilities π_k, which are parametrized as a logistic function:

$$\tau_s = \prod_{k=1}^{K} \pi_k^{\lambda_{sk}} \cdot (1-\pi_k)^{1-\lambda_{sk}}, \quad \text{with} \quad \pi_k = \frac{\exp(\gamma_k)}{1+\exp(\gamma_k)}, \tag{11}$$

where γ_k is an attribute-specific intercept for attribute k. The specification in Equation (11) is closely related to the model of choice set heterogeneity of Swait and Ben-Akiva (1987), as well as to concomitant latent class models of Kamakura, Wedel, and Agrawal (1994). Generally, it is possible to include individual-specific covariates in the submodel of class probabilities. For example, Hole et al. (2013) and Collins, Rose, and Hensher (2013) use respondents' stated measure of ANA, Yegoryan et al. (2020) use visual attention measure derived from eye tracking.

We can write the unconditional choice probability as a weighted sum of class-specific choice probabilities defined in Equation (10):

$$P_{ijt} = \sum_{s=1}^{S} \tau_s \cdot P_{ijt|s}, \tag{12}$$

and the likelihood of individual i as:

$$L_i = \sum_{s=1}^{S} \tau_s \cdot \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt|s}^{y_{ijt}}. \tag{13}$$

By setting $\tau_s = 1$ for the class where all attributes are attended, the EAA model reduces to the MNL. Hence, the EAA model nests the MNL at the boundary condition $\gamma_k \to \infty \quad \forall k$.

2.4 Model Accounting for Preference Heterogeneity and Attribute Non-attendance

The EAA model can be further extended to additionally account for preference heterogeneity. In particular, we follow Hess et al. (2013) and Hole et al. (2013), and assume that preference parameters are distributed multivariate normal (i.e., $\boldsymbol{\beta}_i \sim (\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma})$) across the latent classes.

While the parameter distribution is common across the latent classes, each class is related to a different subset of attended attributes and, therefore, has a different vector of parameters due to elementwise multiplication with $\boldsymbol{\lambda}_s$: $\boldsymbol{\beta}_{i|s} = \boldsymbol{\beta}_i \circ \boldsymbol{\lambda}_s$. Therefore, the choice probabilities can be written as:

$$P_{ijt} = \sum_{s=1}^{S} \tau_s \cdot \int \frac{\exp(\boldsymbol{x}_{ijt} \cdot \boldsymbol{\beta}_{i|s})}{\sum_{j' \in J} \exp(\boldsymbol{x}_{ij't} \cdot \boldsymbol{\beta}_{i|s})} \phi(\boldsymbol{\beta}_{i|s}|\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_i, \tag{14}$$

and the likelihood of individual i as:

$$L_i = \sum_{s=1}^{S} \tau_s \cdot \int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt|s}^{y_{ijt}} \phi(\boldsymbol{\beta}_{i|s}|\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_i. \tag{15}$$

By setting all elements of $\boldsymbol{\Sigma}$ to zero, the MEAA model reduces to the EAA model. By setting $\tau_s = 1$ for the class where all attributes are included, the MEAA model reduces to the MMNL model. Lastly, applying both restrictions reduces the MEAA model to the bare-bones MNL model, which neither accommodates preference heterogeneity nor ANA.

The MEAA model has some similarities to the MM-MNL model, as both should be able to capture lexicographic and random choice behavior. However, the MM-MNL and the MEAA models may excel at capturing different patterns in preference distribution. The MEAA model is specifically designed to capture and disentangle the genuinely zero estimates. On the other hand, the MM-MNL may be better at identifying cases where the preference distribution is multimodal. We summarize the differences in the models in Table 1.

Model	Accommodates		Likelihood function
	Pref. het.	ANA	
MNL	no	no	$L_i = \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt}^{y_{ijt}}$
EAA	no	yes	$L_i = \sum_{s=1}^{S} \tau_s \cdot \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt\|s}^{y_{ijt}}$
MMNL	yes	no	$L_i = \int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt}^{y_{ijt}} \phi(\boldsymbol{\beta}_i\|\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_i$
MM-MNL	yes	no	$L_i = \sum_{q=1}^{Q} w_q \cdot \int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt}^{y_{ijt}} \phi(\boldsymbol{\beta}_{i\|q}\|\bar{\boldsymbol{\beta}}_q, \boldsymbol{\Sigma}_q) d\boldsymbol{\beta}_{i\|q}$
MEAA	yes	yes	$L_i = \sum_{s=1}^{S} \tau_s \cdot \int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt\|s}^{y_{ijt}} \phi(\boldsymbol{\beta}_{i\|s}\|\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_i$

Table 1: Estimated models for comparison

2.5 Estimation Procedure

We split all the datasets into estimation (training) and holdout (validation) samples. More specifically, we randomly select two choice tasks as holdout tasks for each respondent in each dataset. We estimate five models, MNL, EAA, MMNL, MEAA, and MM-MNL, for each dataset using the corresponding training samples. For statistical inference, we employ maximum likelihood estimation with sample loglikelihood $LL(\boldsymbol{\theta}) = \sum_{i \in I} ln(L_i)$, where L_i denotes the likelihood of individual i given by Equation (3), (13), (5), (15), and (8) for the MNL, EAA, MMNL, MEAA, and MM-MNL models, respectively. $\boldsymbol{\theta}$ represents the vector of parameters to be estimated. To retain parsimony, we use a diagonal specification of $\boldsymbol{\Sigma}$ for heterogeneous models. For MMMNL models, we specify Q = 2 classes[4]. In the case of heterogeneous models, an integration over the density of taste parameters is required, for which we adopt the simulated maximum likelihood approach using 500 Halton draws and gradient-based Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Train 2009).

We tested multiple starting values for the models. We report the one with the best loglikelihood value. As MM-MNL models are prone to pick up extreme behavior and result in unrealistically large parameter values (Keane and Wasi 2013), we apply similar constraints to the parameter values as in Keane and Wasi (2013) post estimation. More specifically, we consider only the subset of the models estimated based on different starting values, in which absolute values of the utility estimates (both mean and standard deviation) do not exceed 20.

For the out-of-sample predictions, we use individual-level conditional estimates in all models except for the MNL. In the MNL model, all preference parameters are homogeneous across the sample. For the heterogeneous models, we employ Bayes' rule to condition on the observed choices and compute the posterior means of the individual-level preference parameters $\boldsymbol{\beta}_i^{\text{post.}}$ (see Train 2009, ch.11 for details). Correspondingly, for the MMNL model, individual-level parameters are given by:

$$\boldsymbol{\beta}_i^{\text{post.}} = \frac{\int \boldsymbol{\beta}_i \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt}^{y_{ijt}} \phi(\boldsymbol{\beta}_i | \boldsymbol{\beta}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_i}{\int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt}^{y_{ijt}} \phi(\boldsymbol{\beta}_i | \boldsymbol{\beta}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_i}, \tag{16}$$

and in the MM-MNL model by:

$$\boldsymbol{\beta}_i^{\text{post.}} = \sum_{q=1}^{Q} w_{iq} \cdot \frac{\int \boldsymbol{\beta}_{i|q} \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt|q}^{y_{ijt}} \phi(\boldsymbol{\beta}_{i|q} | \boldsymbol{\beta}_q, \boldsymbol{\Sigma}_q) d\boldsymbol{\beta}_{i|q}}{\int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt|q}^{y_{ijt}} \phi(\boldsymbol{\beta}_{i|q} | \boldsymbol{\beta}_q, \boldsymbol{\Sigma}_q) d\boldsymbol{\beta}_{i|q}}. \tag{17}$$

[4]We have tested the Q = 3 specification as well. However, we did not find improvements in BIC for any of the datasets. Also, Keane and Wasi (2013) find that in many other applications of the MMMNL, two-class specification usually results in better BIC.

In the case of the EAA and the MEAA models, the vector of preference parameters is (also) conditional on the individual's class allocation:

$$\tau_{is}^{\text{post.}} = \frac{\tau_s \cdot L_{i|s}}{\sum_{s' \in S} \tau_{s'} \cdot L_{i|s'}}. \tag{18}$$

As the central behavioral assumption of the ANA models (i.e., the EAA and the MEAA) is that each individual has a specific attribute processing strategy, we opt for a nonoverlapping segmentation and assign each individual to the class s^* with the highest value of $\tau_{is}^{\text{post.}}$ (cf. Desarbo, Ramaswamy, and Cohen 1995). We then compute the individual-level conditional estimates as:

$$\boldsymbol{\beta}_i^{\text{post.}} = \boldsymbol{\lambda}_{is^*} \circ \boldsymbol{\beta}, \tag{19}$$

for the EAA model and:

$$\boldsymbol{\beta}_{i\hat{s}}^{\text{post.}} = \frac{\int \boldsymbol{\beta}_{i|s^*} \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt|s^*}^{y_{ijt}} \phi(\boldsymbol{\beta}_{i|s^*} | \boldsymbol{\beta}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_i}{\int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt|s^*}^{y_{ijt}} \phi(\boldsymbol{\beta}_{i|s^*} | \boldsymbol{\beta}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_i}. \tag{20}$$

for the MEAA model. Yegoryan et al. (2020) show that such "crisp" segmentation works rather well: despite the high number of classes, the posterior probabilities usually strongly favor one class for an individual.

For the MMNL, MM-MNL, and MEAA models, we again employ simulation methods with Halton draws to approximate the integrals in Equations (16), (17), and (20), respectively.

3 Empirical Application

3.1 Data

For the empirical application, we utilize a broad set of ten different datasets from CBC studies on different product categories, including FMCG, such as smoothies and packaged orange juice, entertainment and experiential products and services, such as video-streaming services, basketball tickets, (student) parties, and holiday destinations, as well as consumer electronics, such as electric kettles, laptops, tablets, and cameras.[5] These studies differ significantly in terms of the financial risk or stakes of the decision in question. For example, in the case of electronic goods in our set, laptops, tablets, and cameras are more expensive: priced up to \$749, 550€, and \$279 in the studies, respectively. On the other hand, electric kettles are priced up to \$80 in the study. Hence, we classify the former three as relatively high-risk and the latter as a relatively low-risk category. Similarly, choosing holiday destinations, which costs up to \$1200 in the study, is a high-stakes decision.

In contrast, we consider smoothies and orange juices priced up to 1.99€ and 1.89€ in the studies, respectively, as relatively low-risk categories. Regarding entertainment goods, we consider the choice of a video-streaming service, with alternatives priced up to \$12.99, and which party to attend, with alternatives priced up to 7€, as a relatively low-stake decision. In contrast, basketball tickets are priced up to 30€ in the study. We, therefore, classify the latter as a relatively high-risk category.

Our classification gives us more context to investigate the patterns of ANA across different datasets. Moreover, price, which indicates the financial risk of the purchase, is related to product category involvement (Laurent and Kapferer 1985). Several of the developed scales for measuring involvement, e.g., in Laurent and Kapferer (1985), Jain and Srinivasan (1990), and McQuarrie and Michael (1986), include risk as an essential dimension.

Furthermore, within each (high or low) risk or stake classification, the studies vary in the number of attributes ranging from three to eight. The number of attributes serves as an indicator of the degree of complexity of the decision task (Dellaert, Donkers, and Van Soest 2012). Of course, other important factors contribute to the complexity of the decision task, such as the number of alternatives, the number of attribute levels, and the similarity of the alternatives (Dellaert et al. 2012). However, as using 3-4 alternatives in CBC studies is rather common (Rao 2014), we focus on the number of attributes as a primary driver of complexity[6].

In summary, the ten datasets allow us to descriptively investigate whether ANA prevails and how it varies across different product categories and settings. We summarize the main characteristics and the sources of the datasets in Table 2 (for details on the attribute levels, see

[5]Please note that except for Parties and Electric kettles, all other datasets have been formerly published. For more details on these studies, we refer the reader to the respective source outlined in Table 2.

[6]From hereon, we use complexity and the number of attributes as equivalents.

	Category	Source	Risk/Stake	Attributes (No. levels)		ANA-Classes	No. Alt.	Obs. = id × cs
1.	Smoothies	Paetz and Steiner (2017)	low	3 attr.:	brand (4), price (4), packaging (4)	8	3 + none	$8910 = 495 \times 18$
2.	Orange juice	Paetz and Guhl (2017)	low	4 attr.:	brand (4), price (4), packaging (2), fairtrade label (2)	16	3 + none	$5472 = 342 \times 16$
3.	Video-streaming services	Glasgow and Butler (2017)	low	5 attr.:	privacy policy (3), price (4), catalog size (3), fast content (2), commercials shown (2)	32	4 + none	$2860 = 260 \times 11$
4.	Parties	our dataset	low	6 attr.:	location (6), drinks (3), dress code (2), specials (3), music (4), price (2)	64	3 + none	$2120 = 212 \times 10$
5.	Electric kettles	our dataset	low	7 attr.:	brand (3), capacity (3), material (3), power (2), variable temperature (2), Amazon rating (3), price (4)	128	3 + none	$2624 = 164 \times 16$
6.	Basketball tickets	Schlereth and Skiera (2017)	high	3 attr.:	price category (4), price (4), additional features (3)	8	4 + none	$1920 = 160 \times 12$
7.	Laptops	Liu and Tang (2015)	high	4 attr.:	display size (3), memory (3), hard drive (3), price (3)	16	3	$1800 = 120 \times 15$
8.	Tablets	Schlereth and Skiera (2017)	high	6 attr.:	brand (3), price (4), display size (2), battery (2), resolution (2), storage capacity (3)	64	3 + none	$2484 = 207 \times 12$
9.	Cameras	Allenby et al. (2014)	high	7 attr.:	brand (4), pixel (2), zoom (2), video (2), swivel (2), WiFi (2), price (5)	128	4 + none	$5312 = 332 \times 16$
10.	Holiday destinations	Keane and Wasi (2013)	high	8 attr.:	airline (2), local tours (2), destination (2), length of stay (2), meal inclusion (2), peak season (2), price (2), accommodation (2)	256	2	$5296 = 331 \times 16$

Table 2: Sources and characteristics of the datasets

Appendix A).

Both CBC studies, on choices of parties and electric kettles, have been designed and administered using Sawtooth Software [7]. The CBC study on parties was conducted in 2010 at a large German university using a convenience sample of students. The CBC study in the context of electric kettles was conducted in December 2019 - January 2020, using a crowdsourcing platform Amazon Mechanical Turk [8]. To ensure the quality of the responses, in the case of the Electric kettles dataset, we used a qualification question, which required the respondents to solve a simple math problem at the very beginning of the questionnaire, as well as three attention questions evenly distributed across the questionnaire. We only included respondents who passed the qualification question, answered all three attention questions correctly, and took more than one minute (the 5%-quantile in the response time) to complete all the choice tasks.

For formerly published datasets, we used the data-pruning steps reported by the authors. As some but not all datasets include holdout tasks, we only use the main tasks to ensure consistency across datasets. Furthermore, in all datasets, we excluded respondents who selected the "none" option in all the main tasks[9] In Table 2, we report the number of respondents after the data pruning and the number of main choice tasks (initial holdout tasks excluded).

3.2 Model Comparison

We compare the in-sample performance of the estimated models in terms of the log-likelihood (LL) values and the Bayesian information criterion (BIC)[10], which penalizes for model complexity. We summarize the results in Table 3.

In all datasets, as expected, the MNL is the worst-performing model in terms of both LL and BIC. Accounting for either ANA or preference heterogeneity substantially increases the in-sample model performance: both EAA and MMNL models outperform the MNL model

[7] In both cases, we use complete enumeration as a method for generating random tasks, which ensures minimal overlap between concepts in a task and strives for the most nearly orthogonal design (Software 2017). Additionally, multiple versions have been generated and randomized between respondents.

[8] We recruited participants on this platform that resided in the US, had more than 1000 HITs approved with a 95% approval rating. The respondents were paid fair compensation for their time computed based on an \$8 hourly wage.

[9] Accordingly, we excluded one respondent in the Electric kettles and Parties datasets, two respondents in the Smoothie dataset, and seven respondents in the Tablets dataset.

[10] Note that BIC can be used for the comparison of non-nested models. As the MEAA model nests MNL, EAA, and MMNL at the boundary of the parameter space, and MEAA and MM-MNL are not nested, the log-likelihood ratio test is not applicable (McLachlan and Peel 2000).

Dataset		MNL	EAA	MMNL	MEAA	MMMNL
Smoothies	LL	−7931.84	−6945.04	−6233.87	−5863.34	**−5806.92**
	BIC	15935.46	13988.78	12611.31	**11897.18**	11909.96
		(8)	(11)	(16)	(19)	(33)
Orange juice	LL	−5083.52	−4575.24	−4049.22	−3880.15	**−3836.18**
	BIC	10226.37	9243.69	8217.08	**7912.83**	7918.10
		(7)	(11)	(14)	(18)	(29)
Video-streaming services	LL	−3591.15	−3278.25	−3264.93	−3167.17	**−3161.75**
	BIC	7244.36	6657.35	6654.00	**6497.27**	6579.51
		(8)	(13)	(16)	(21)	(33)
Parties	LL	−2016.36	−1924.86	−1771.78	−1750.12	**−1713.11**
	BIC	4144.27	4005.88	**3766.65**	3767.93	3879.82
		(15)	(21)	(30)	(36)	(61)
Electric kettles	LL	−2380.63	−1963.63	−1870.29	**−1753.58**	−1780.36
	BIC	4854.12	4074.29	3926.32	**3747.06**	3939.93
		(12)	(19)	(24)	(31)	(49)
Basketball tickets	LL	−1959.44	−1561.67	−1370.05	−1350.41	**−1303.80**
	BIC	3976.11	3202.03	2854.58	**2836.75**	2843.69
		(8)	(11)	(16)	(19)	(33)
Laptops	LL	−1239.08	−1034.99	−992.02	−979.90	**−969.00**
	BIC	2529.63	2150.86	**2086.98**	2092.14	2151.23
		(7)	(11)	(14)	(18)	(29)
Tablets	LL	−2414.47	−2142.53	−1952.79	−1926.38	**−1846.54**
	BIC	4897.65	4399.59	4043.01	4036.00	**3975.59**
		(9)	(15)	(18)	(24)	(37)
Cameras	LL	−5701.87	−4958.03	−4341.66	**−4131.95**	−4145.62
	BIC	11488.17	10059.61	8852.21	**8491.89**	8637.45
		(10)	(17)	(20)	(27)	(41)
Holiday destinations	LL	−2686.87	−2395.59	−2262.40	−2172.05	**−2160.81**
	BIC	5441.27	4926.23	4659.86	**4546.68**	4600.18
		(8)	(16)	(16)	(24)	(33)
Frequency:						
	Best LL	0	0	0	2	**8**
	Best BIC	0	0	2	**7**	1

Notes: The number of parameters in each model for each dataset is presented in parentheses. Values in bold indicate the best performing model for a given dataset based on a given criterion (LL or BIC).

Table 3: Model comparison: In-sample measures

indicated by both LL and BIC. However, accounting only for preference heterogeneity yields greater in-sample improvements than accounting only for ANA across our ten datasets: the MMNL model outperforms the EAA model for all datasets. Nevertheless, accommodating both ANA and preference heterogeneity is essential: the MEAA model outperforms both the EAA and MMNL models. In particular, the MEAA model outperforms the MMNL in terms of LL in all ten and BIC in 8 out of 10 datasets. For the datasets Parties and Laptops, the BIC of the MEAA model is only respectively 1.28 and 5.16 points higher than the BIC of the MMNL model. According to Raftery (1995), only a difference of more than 10 points provides evidence to favor the model with better BIC.

Even when considering the MM-MNL model, which accommodates a more flexible pattern of preference heterogeneity, we see that the MEAA model is the best based on BIC 8 out of 10 times and also based on LL 2 out of 10 times. Notably, while the MM-MNL model is mostly the best in LL, the difference between the MM-MNL and the MEAA models is small. Considering the minor improvements in LL and a rather substantial increase in the number of parameters in the MM-MNL model, it is no surprise that the MEAA model does much better on BIC, i.e., when we penalize for model complexity.

Dataset		MNL	EAA	MMNL	MEAA	MMMNL
Smoothies	hit rate	0.58	0.76	0.74	**0.79**	0.76
	hit probability	0.47	0.56	0.66	**0.70**	0.70
Orange juice	hit rate	0.57	0.66	0.73	**0.78**	0.75
	hit probability	0.44	0.53	0.65	**0.70**	0.69
Video-streaming services	hit rate	0.36	0.74	0.48	**0.76**	0.48
	hit probability	0.24	0.33	0.36	**0.41**	0.39
Parties	hit rate	0.50	0.59	**0.62**	0.59	0.60
	hit probability	0.36	0.43	0.52	**0.56**	0.53
Electric kettles	hit rate	0.54	0.68	0.73	**0.77**	0.73
	hit probability	0.42	0.54	0.64	**0.68**	0.65
Basketball tickets	hit rate	0.32	0.66	0.66	0.65	**0.68**
	hit probability	0.23	0.41	0.58	0.59	**0.61**
Laptops	hit rate	0.63	**0.77**	0.75	0.76	0.75
	hit probability	0.53	0.65	0.70	0.71	**0.71**
Tablets	hit rate	0.54	0.61	0.66	0.67	**0.68**
	hit probability	0.38	0.48	0.58	0.60	**0.61**
Cameras	hit rate	0.49	0.62	0.70	**0.73**	**0.73**
	hit probability	0.36	0.46	0.61	**0.64**	0.64
Holiday destinations	hit rate	0.70	**0.83**	0.80	0.82	0.81
	hit probability	0.60	0.69	0.75	**0.77**	0.77
Frequency:						
	Best hit rate	0	2	1	**5**	3
	Best hit probability	0	0	0	**7**	3

Notes: Values in bold indicate the best performing model for a given dataset based on a given criterion (hit rate or hit probability).

Table 4: Model comparison: Out-of-sample measures

Furthermore, the out-of-sample predictive validity measured by hit rate and hit probability is presented in Table 4. The results uncover some interesting patterns. First, the models that account for ANA (i.e., the EAA and the MEAA) considerably outperform their counterparts that only account for preference heterogeneity (i.e., the MNL and MMNL). In particular, the EAA model has a much higher hit rate (mean difference of 17 percentage points (PP)) and hit probability (mean difference of 10PP) in all datasets compared to the MNL model. Also, the MEAA model mostly outperforms the MMNL: 8 out of 10 times in hit rate (mean difference of 4PP) and in all cases in hit probability (mean difference of 3PP).

Second, the EAA model outperforms all the models in two datasets in the hit rate (Laptops and Holiday destinations). Potentially, such a result indicates that in these cases understanding whether an attribute is ignored or not, i.e., whether the parameter is zero or not, is more informative than trying to estimate the precise values of non-zero parameters. This could happen if the actual amount of heterogeneity is relatively small.

Third, the MEAA model mostly remains the best-performing model out-of-sample: it outperforms all the models in hit rate and hit probability in 5 and 7 out of 10 cases, respectively. In six cases, when the MEAA is superior to the MM-MNL model, it offers an average of 6PP better hit rate – a considerably larger margin than the 1PP average improvement of the MM-MNL model in the other four cases. A similar pattern holds when comparing the MEAA and the MMNL models.

Furthermore, the models accounting for ANA are superior both in- and out-of-sample in categories we have classified as low-risk/stake, except for Parties, where the MMNL outperforms the MEAA model. For categories of high-risk/stake, the ANA models outperform in cases of higher complexity (Cameras and Holiday destinations). It appears that understanding which attributes are, in fact, in the utility function is more critical for low-stake as well as high-stake and high-complexity settings. This result is in line with our expectations, as consumers may search for less information in low-stake settings (Laurent and Kapferer 1985) and simplify their decisions by ignoring attribute information in high-complexity settings (Payne et al. 1992).

3.3 Effects of Ignoring Preference Heterogeneity

To investigate possible biases due to ignoring preference heterogeneity while accommodating ANA, we compare the average attribute attendance probabilities and the probability distribution of the number of attributes attended within the EAA and MEAA models computed based on Equation (11). These, along with corresponding confidence intervals, for both models are presented in Figure 2 and 3, respectively.[11]

First, from Figure 2, it is apparent that none of the attributes has exactly 0% or 100% attendance probability across all datasets. This is in line with the findings in Yegoryan et al. (2020) but in contrast to the results in Hess et al. (2013) and Hole et al. (2013). We observe some attributes with tiny attendance probabilities for more complex decisions (e.g., in Electric kettles and Holiday destinations datasets). Moreover, similar attributes (e.g., price or brand) in different applications have somewhat different attendance probabilities (i.e., they are application-specific). However, in most applications, price is one of the attributes with the highest attendance probabilities (50% or more according to the MEAA model).

Second, in line with previous literature (e.g., Hess et al. 2013, Hole et al. 2013, Yegoryan et al. 2020), ignoring preference heterogeneity results in a downward bias of attribute attendance probabilities. The average difference in attribute attendance probabilities between the MEAA and the EAA models across all datasets is 18.25PP ($p < 0.001$). However, in some cases, we do not observe a significant difference in particular attribute attendance probabilities between the two models (illustrated by very close mean values and overlapping confidence intervals). For example, the latter holds for all attributes except for price in the Video-streaming services dataset or three attributes (brand, capacity, and power) in the Electric kettles dataset. Nevertheless, in contrast to Hess et al. (2013) and Hole et al. (2013) (applications in the context of route and prescription medication choice, respectively), and, in line with Yegoryan et al. (2020) (application in the context of laptop and coffee-makers choice), in neither of our ten applications does ANA completely diminish after accounting for preference heterogeneity.

Also, the downward bias in the attribute attendance probabilities results in a shift of the probability distribution of the number of attributes attended to the right when we account for preference heterogeneity (see Figure 3). The only two cases where we do not observe such a shift are in the Video-streaming services and Electric kettles datasets. On average, the mode of the distribution increases from the EAA to the MEAA model by 1.1 ($p < 0.01$), and the ratio of the average number of attributes considered to the number of available attributes increases by 18.45PP ($p < 0.01$) across all datasets.

[11] We conduct parametric bootstrapping using 10.000 draws to generate the average attribute attendance and the average probability of attending a certain number of attributes based on the asymptotic distribution of class parameters $\hat{\boldsymbol{\gamma}}$ presented in Appendix A.

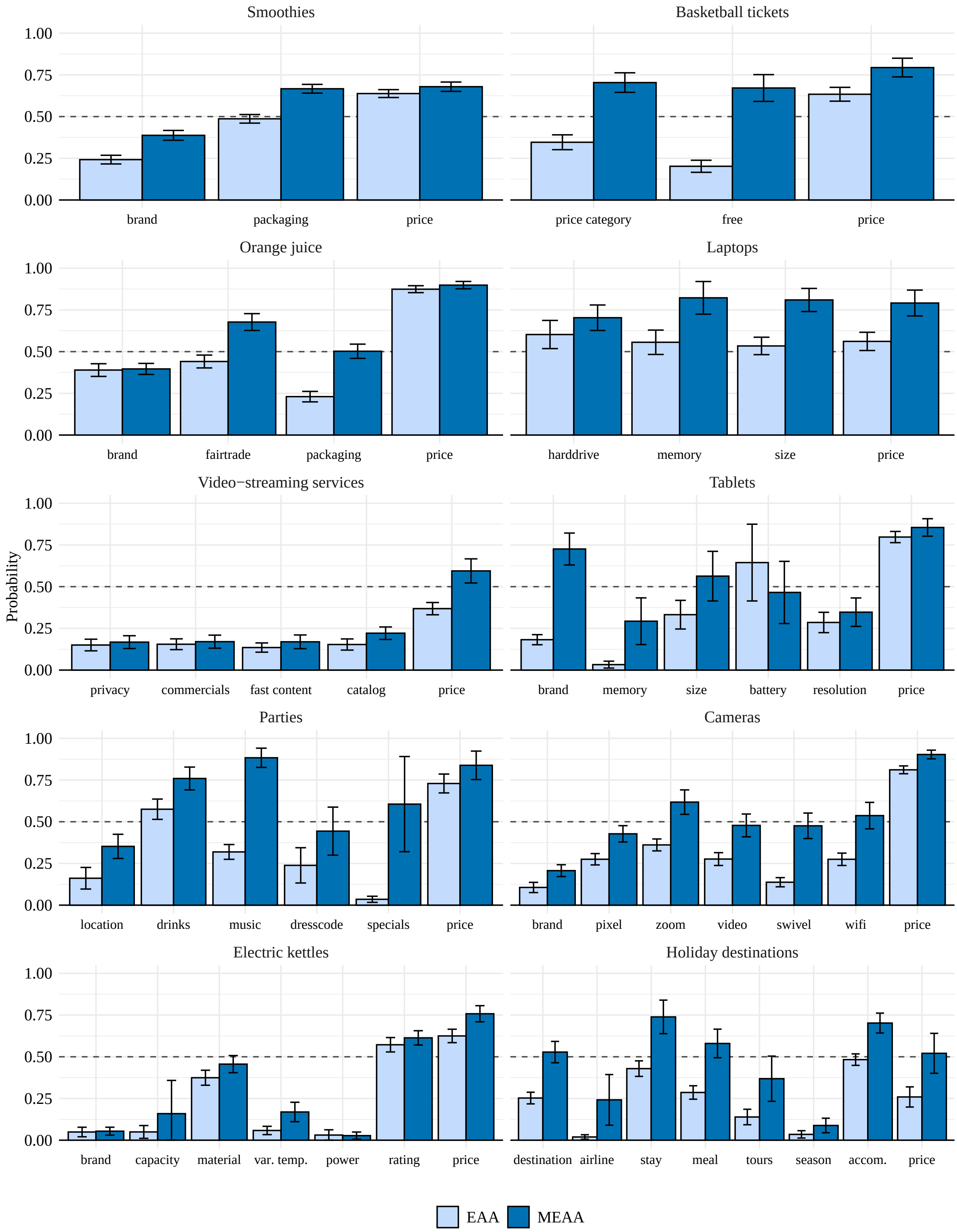


Figure 2: Average attribute attendance probabilities

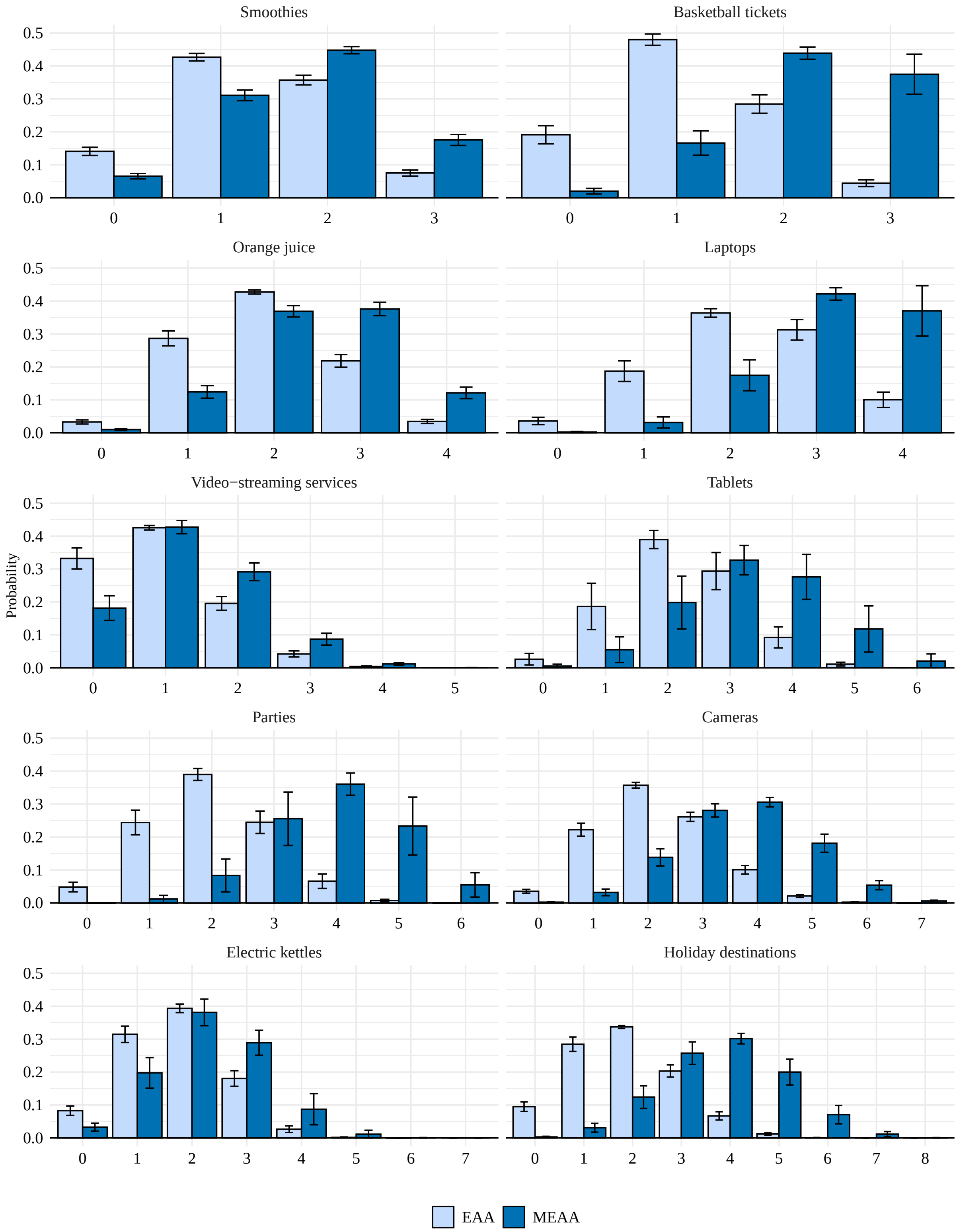


Figure 3: Probability of attending a certain number of attributes

As we have established that not accounting for preference heterogeneity results in biases in the amount of identified attribute non-attendance, we focus on the results of the MEAA model. Several critical observations can be made from Figure 3. First, across all datasets, only a very small proportion of respondents rely on random choice. The only exception is the Video-streaming services dataset, where we see around 18.1% probability of a random choice.

Second, a much larger share of respondents incorporates all the available attributes in the decision in datasets of low versus high complexity (average of 28.9% vs. 3.7%). We also see that the proportion of respondents incorporating all attributes into the decision-making is substantially higher for the low complexity and high-risk/stake categories (average of 37.3% across the datasets of Basketball tickets and Laptops) vs. low complexity and low-risk/stake categories (average of 12.1% across the datasets of Smoothies and Orange juice). However, we do not observe a substantial difference in the high-complexity settings.

Third, the average ratio of the number of attributes attended to the number of the available attributes substantially decreases in the high complexity settings (average difference of 70.8% vs. 47.7%, $p < 0.05$). On average, this ratio is also higher in high-risk/stake settings (60.3% vs. 46.2%). However, the difference in means is not statistically significant. The latter is mainly driven by the Parties dataset, for which the MMNL is the best-fitting model and for which the preference heterogeneity potentially plays a more critical role. We can observe this descriptively in the probability distributions of the number of attributes attended in Figure 3.

Lastly, a slightly larger proportion of respondents tends to use a lexicographic rule (i.e., consider only one attribute) in low- vs. high-risk/stake settings (right vs. left panel in Figure 3). The probability of attending only one attribute is particularly high in the Video-streaming services (42.7%) and Smoothies (31.1%) datasets, followed by the Electric kettles (19.8%), Basketball tickets (16.6%), and Orange juice (12.4%). For the rest of the datasets, the probability of attending only one attribute ranges between 1-5%.

3.4 Effects of Ignoring Attribute Non-attendance

Next, we turn to the de facto standard situation in current marketing literature when preference heterogeneity is accounted for while the analyst ignores ANA. To understand the potential consequences of ignoring ANA, we compare the mean and the standard deviation of the preference parameters implied by the MMNL and the MEAA model. Due to differing fit and, therefore, differing scales, we cannot directly compare the estimates of these models (Huber and Train 2001). To circumvent this issue, we regress the individual-level conditional estimates in the

MEAA model on those of the MMNL and MM-MNL models (similar to Frischknecht, Eckert, Geweke, and Louviere 2014) and use the slope parameter as a rescaling factor (for details, see Appendix B), which we apply for both population- and individual-level estimates.

First, we seek to examine the potential biases due to neglecting ANA at the population level. Here, we only compare the MEAA and the MMNL models, as the first moments of the MM-MNL distribution are not informative in such a comparison.[12] In particular, we check for how many cases we encounter biases in our hypothesized direction. Recall that we expect that the direction of the bias would depend on the share of ANA and the true preference distribution. We expect that failing to accommodate ANA, contingent upon ANA occurring (i.e., a share of respondents ignoring some attributes in the decision-making), would bias 1) the mean estimate of the preference distribution towards zero regardless of the characteristics of the true distribution, 2) the estimated variance (or standard deviation) upwards when the true distribution lies further away from zero, leading to an overestimation of the amount of preference heterogeneity in the population, 3) bias the estimated variance (or standard deviation) downwards when the true distribution includes zero (i.e., spans on both positive and negative domains), leading to an underestimation of the amount of preference heterogeneity in the population.

While we do not know the true preference distribution in the population, we acknowledge that the MEAA model, which accounts for both preference heterogeneity and ANA, outperforms the MMNL model in the majority of the empirical applications (see section 3.2). Subsequently, we treat the results of the MEAA model to be closer to the truth and use it as a benchmark. To understand the direction of the bias in the variance (or standard deviation), we use the estimates of the mean and the standard deviation of the MEAA model to classify where the zero lies with respect to the implied preference distribution: within the $(\beta - \sigma, \beta + \sigma)$ interval or outside of it.

We summarize the frequency and the percentage of occurrence of each of the expected biases in Table 5.[13] In line with the previous literature (e.g., Collins 2012), in the majority of cases, we do indeed see that the mean preference parameters in the MMNL are biased towards

[12]We include a comparison with the MM-MNL model when we examine the distribution of individual-level conditional estimates.

[13]The plots of the MEAA estimates against the MMNL estimates for each of the datasets are presented in Appendix C.

zero, i.e., underestimated (see column 3 in Table 5). This happens 100% of the time when zero is outside and 86.8% of the time when it is within the $(\beta-\sigma,\beta+\sigma)$ interval of the "true" MEAA distribution. Although we cannot generalize, we see some (descriptive) indication that the magnitude of the bias in the mean is larger the further the distribution from zero. In particular the mean absolute difference in the mean estimates of the MEAA and MMNL models is 0.90 vs. 0.52 when zero is outside vs. inside of the $(\beta-\sigma,\beta+\sigma)$ interval (see column 4 in Table 5). This also provides some explanation to the five exceptions we observe, for which the MMNL mean estimate is overestimated (see column 5 in Table 5). First, these exceptions only happen when the zero lies within the $(\beta-\sigma,\beta+\sigma)$ interval of the "true" (MEAA) distribution. Second, the magnitude of the difference in parameters is negligible (0.06 absolute difference on average). Third, our classification also ignores the amount of ANA, which impacts the magnitude of the bias. In these five cases, we notably observe relatively low amount of ANA ranging from 11.7% - 32.9%.

Turning to the bias in the standard deviation, related to the amount of heterogeneity, the results in Table 5 support at least one of our expectations. In 97.4% of the cases where zero lies within $(\beta-\sigma,\beta+\sigma)$ interval of the "true" (MEAA) distribution, the MMNL estimates of the standard deviations are underestimated, i.e., the amount of heterogeneity is understated (see columns 5 and 6 in Table 5). There is only one case of overstated standard deviation estimate in the MMNL model. However, again the difference in the MMNl and MEAA estimates is negligible (0.06).

	In MEAA, 0 lies...	Underestimated ($\lvert MMNL\rvert < \lvert MEAA\rvert$)		Overestimated ($\lvert MMNL\rvert > \lvert MEAA\rvert$)	
		N (%)	Mean abs(diff)	N (%)	Mean abs(diff)
Mean	...within $(\beta-\sigma,\beta+\sigma)$	33 (86.8%)	0.52	5 (13.2%)	0.06
	...outside $(\beta-\sigma,\beta+\sigma)$	46 (100.0%)	0.90	0 (0.0%)	0.00
SD	...within $(\beta-\sigma,\beta+\sigma)$	37 (97.4%)	0.93	1 (2.6%)	0.06
	...outside $(\beta-\sigma,\beta+\sigma)$	28 (60.9%)	0.24	18 (39.1%)	0.30

Table 5: Frequency and the direction of biases when ignoring attribute non-attendance

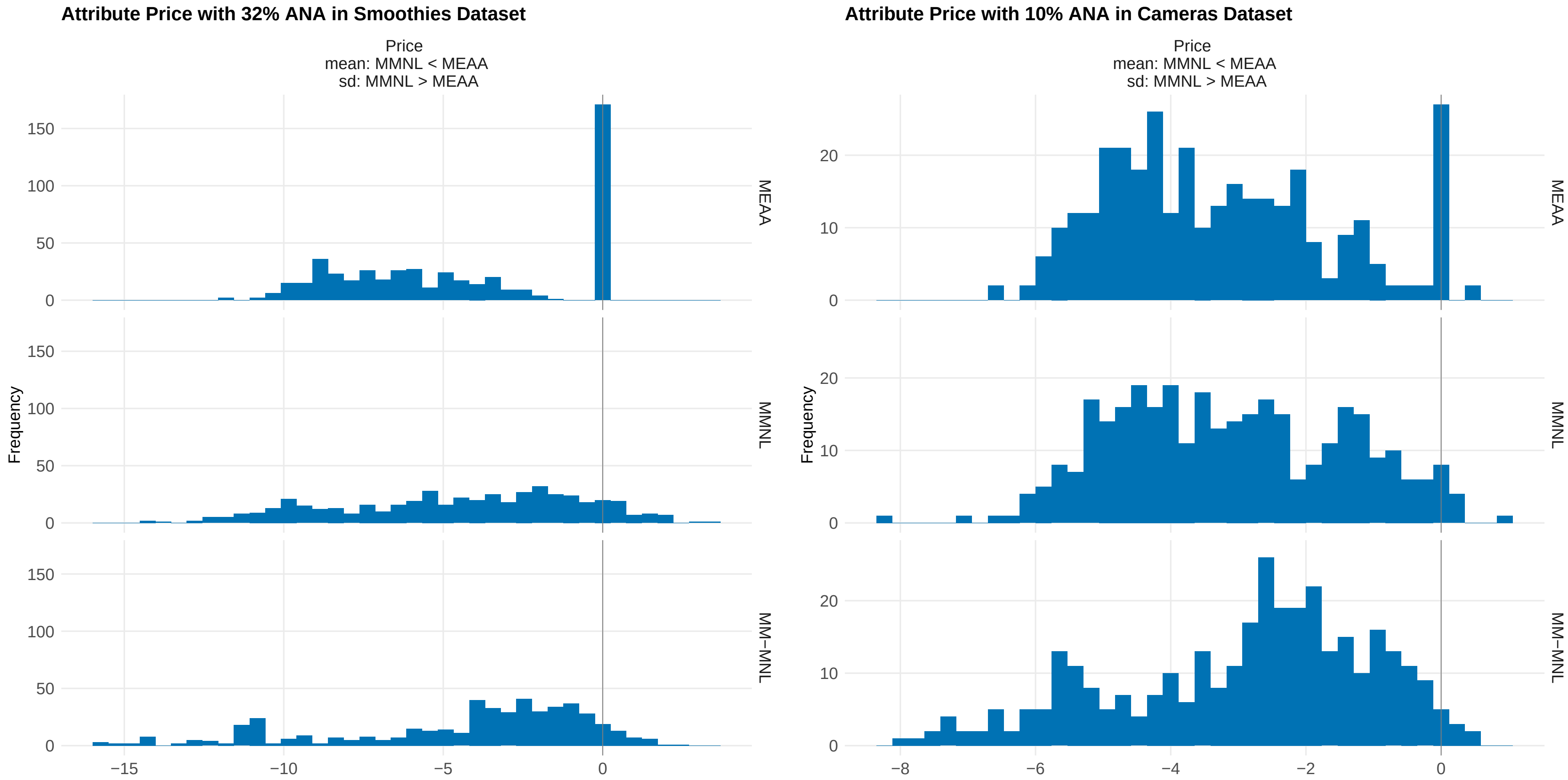
When the zero lies outside of the $(\beta - \sigma, \beta + \sigma)$ interval of the "true" (MEAA) distribution, the findings are somewhat split (see rows 3 and 4 in Table 5). While we expected to see the standard deviation overestimated in this case, that happens 39.1% of the time. For most parameters (60.9%), we still observe an understated standard deviation estimate in the MMNL model. Once again, consider that our classification is rather crude; several factors impact the bias that this classification ignores, particularly the amount of ANA. Another explanation could be some multimodality in the true parameter distribution. We are still fitting a normal distribution in the MEAA model, so multimodality may stretch the fitted distribution, resulting in a larger standard deviation estimate.

To better understand the patterns we observe in the population-level estimates, we now examine the distribution of the individual-level estimates across different models. This is particularly helpful in understanding the differences in the uncovered preference patterns across the models (Keane and Wasi 2013). As we are more specifically interested in understanding the different cases of under and overestimating preference heterogeneity when ignoring ANA, we plot the distribution of the individual-level estimates for selected cases in Figures 4 to 6. In addition to the MMNL and the MEAA models, we also include the MM-MNL results[14]. In the figures, we also report the average ANA for a given attribute and whether we identified under or overestimation of MMNL population-level parameters in comparison to MEAA. Attributes presented on the left panel of Figures 4, 5, and 6 have a relatively higher share of ANA (as identified by the MEAA model) relative to those on the right panel.

Figure 4 showcases the distribution of individual coefficients for price in two applications: Smoothies and Cameras. While, economically speaking, all else being equal, consumers should prefer lower prices, in the two presented cases, we observe some individuals with a positive price coefficient in the MMNL and the MM-MNL models. This is a common problem that can arise for several reasons, including design errors in the CBC study, use of price as a proxy for quality, or due to normality assumption coupled with only limited data available for each individual (Allenby et al. 2014). One possible way to circumvent the issue is to impose sign constraints (Allenby et al. 2014).

By contrast, instances of positive price coefficients are either completely diminished or sub-

[14]Both the MMNL and the MM-MNL estimates are rescaled. For details, see Appendix B.

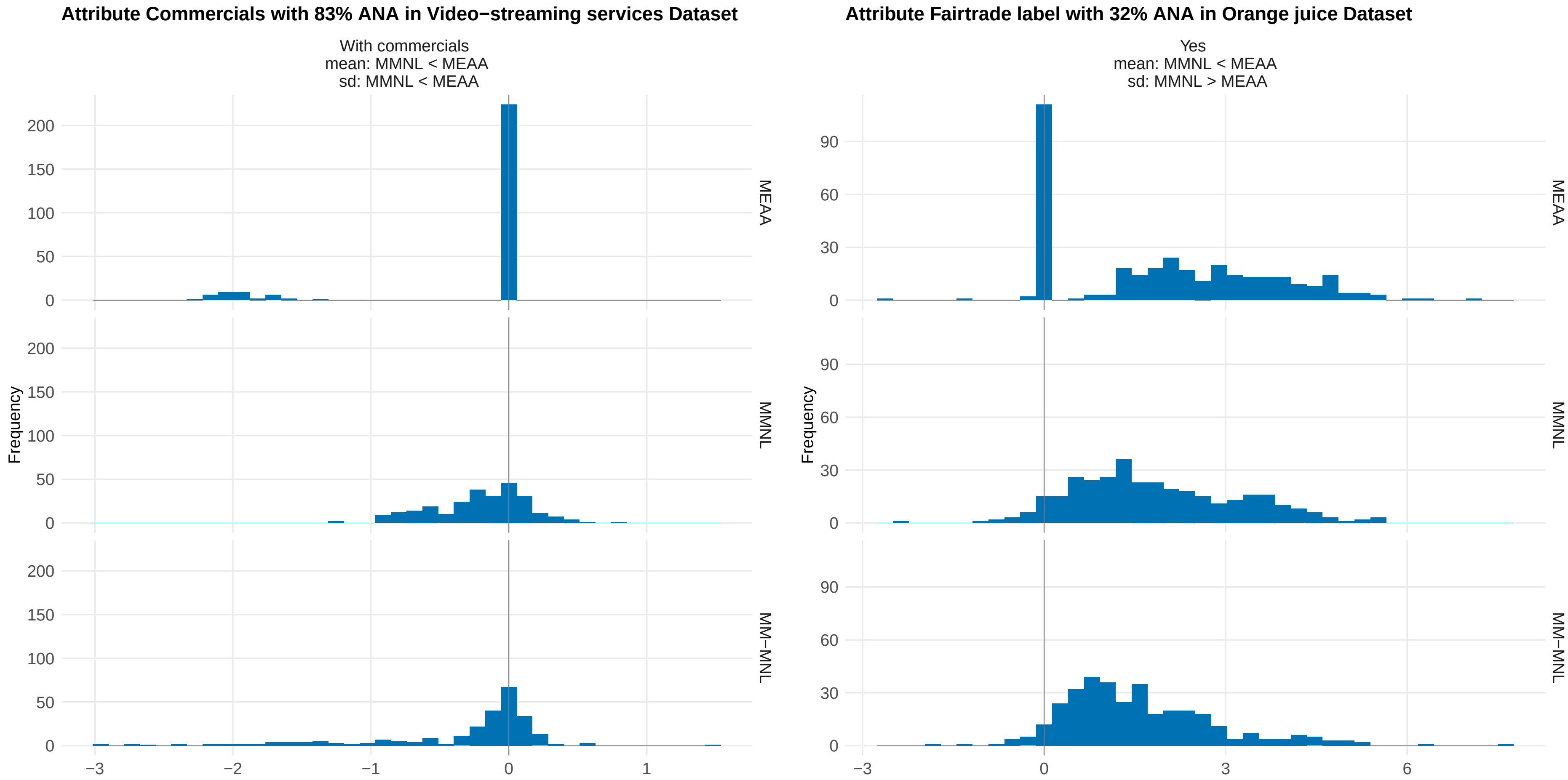


(a) Relatively high ANA

(b) Relatively low ANA

Figure 4: Distribution of individual coefficients for selected datasets – Example 1

Notes: The estimates of the MMNL and the MM-MNL models are rescaled.

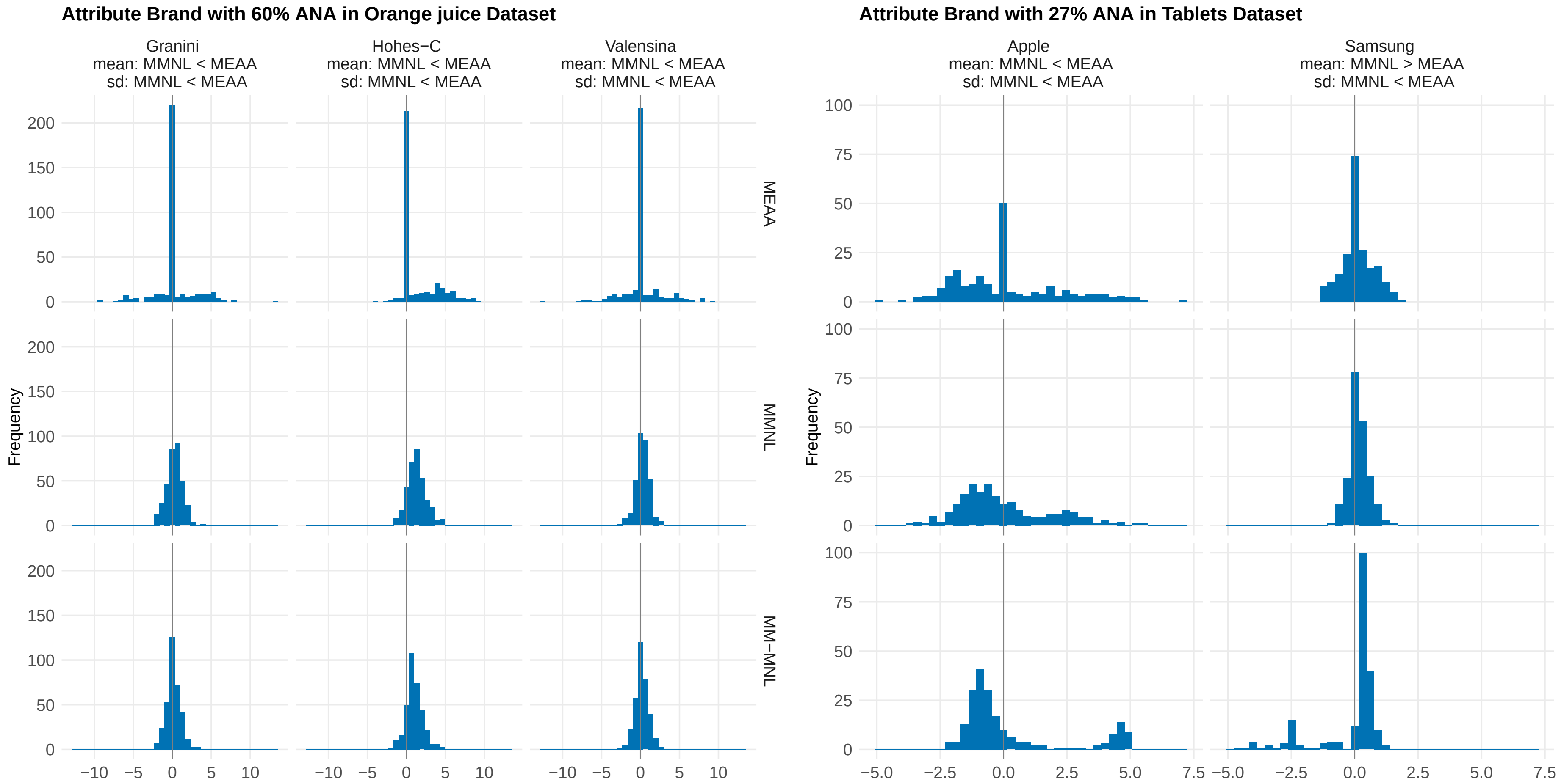


(a) Relatively high ANA

(b) Relatively low ANA

Figure 5: Distribution of individual coefficients for selected datasets – Example 2

Notes: The estimates of the MMNL and the MM-MNL models are rescaled.



(a) Relatively high ANA

(b) Relatively low ANA

Figure 6: Distribution of individual coefficients for selected datasets – Example 2

Notes: The estimates of the MMNL and the MM-MNL models are rescaled.

stantially reduced in the MEAA model. This is a general pattern across our ten applications. More specifically, the MMNL and the MM-MNL produce positive price coefficients in 9 out of 10, while the MEAA only in 5 datasets. On average, 6.94% and 7.78% of the sample have a positive price coefficient in the MMNL and the MM-MNL models, respectively, compared to only 1.07% in the MEAA model. Hence, some of the resulting positive price coefficients in models accounting only for preference heterogeneity (MMNL and MM-MNL) may be driven by people simply ignoring price when making choices in CBC settings. The MEAA model identifies, on average, 32% and 10% probability of ignoring price in the datasets of Smoothies and Cameras presented in Figure 4. In both cases, neglecting ANA results in an overestimation of the amount of heterogeneity in the MMNL and the MM-MNL models. We also observe a more substantial difference in the range of the distribution between the MEAA and the MMNL and MM-MNL models in Smoothies vs. Cameras. In the case of Smoothies, both ANA probability is higher, and the distribution is further away from zero. Both factors may drive the magnitude of the bias.

The pattern of the distribution in Figure 5 is very similar to the one discussed above. However, here we present non-price related product attributes with parameter distributions mostly lying on one side of zero. In the case of video-streaming services, we could generally expect consumers disliking commercials being shown (see the left panel in Figure 5). The MEAA model clearly identifies and aligns to a greater degree with this expected pattern. We find that a large proportion of respondent ignores this attribute altogether: the average probability of ANA is 83%. The rest who care about commercials being shown largely dislike it.

In contrast, the resulting preference distribution is different in the MMNL model. It is shifted towards zero with some mass on the positive side implying that some respondents derive positive utility from commercials being shown. If many respondents in fact ignore this attribute, which is what we identify with the MEAA model, this can explain the shift of the distribution towards zero in the MMNL model. Allowing for more flexibility also does not resolve the issue. The MM-MNL fits two distributions: for one larger class tightly distributed around zero (with $\beta = -0.03$ and $\sigma = 0.27$, see Table A3) and another smaller class with a flatter distribution further away from zero (with $\beta = -1.15$ and $\sigma = 1.33$). As a result, we still see a substantial amount of individual estimates on the positive side implying a preference for commercials being shown.

This is also an example where the population-level MMNL estimate of the standard deviation is smaller than the MEAA estimate, although we would expect it to be larger. The individual parameter distribution helps to understand why that may be the case. In this example, there is still a small mass of individual preference estimates close to but distinct from zero in the MEAA model. Such multimodality of the distribution could be the reason why the population-level standard deviation estimate may be inflated. Furthermore, the MEAA population-level standard deviation estimate is insignificant (see Table A3). As the share of ANA increases, the subsample based on which the mean and standard deviations in the MEAA model are estimated shrinks.

In the case of the fair trade label in the Orange juice dataset, the ANA probability is much lower (average of 32%), and the distribution in the MEAA is somewhat closer to zero (see the right panel in Figure 5). Here, we see a clearer upward bias, i.e., overstatement of the standard deviation estimate (i.e., heterogeneity) in the MMNL model compared to the MEAA.

In Figure 6, we present the individual preference distributions for brands in Orange Juice and Tablets datasets. Unlike previous examples, one wouldn't necessarily have a clear sign expectation for such an attribute as brand. The preference distribution may lie on both side of zero. This is typically the case for attributes allowing horizontal differentiation of products. For such cases, we found that failing to accommodate ANA results in understatement of the population-level standard deviation estimates, i.e., understatement of the amount of heterogeneity in preferences, in the MMNL model compared to the MEAA.

Once again, we present an example with a substantially high amount of identified ANA (around 60%) in the left panel and a relatively lower amount of ANA (around 27%) in the right panel of Figure 6. We see both positive and negative partworth utilities for various brands in all models in both cases. Nevertheless, the preference distribution in the MMNL model has fatter tails than the MEAA model and much more so for brands in the Orange juice (left panel, higher ANA) than in the Tablets datasets (right panel, lower ANA). In the example of the Orange juice dataset, the MM-MNL model does not necessarily perform better. In both classes, the implied distributions for brands do not differ much (see Table A2). As a result, the preference distribution in the MM-MNL model looks very similar to the MMNL.

In both cases, the population-level standard deviations were smaller in the MMNL than the

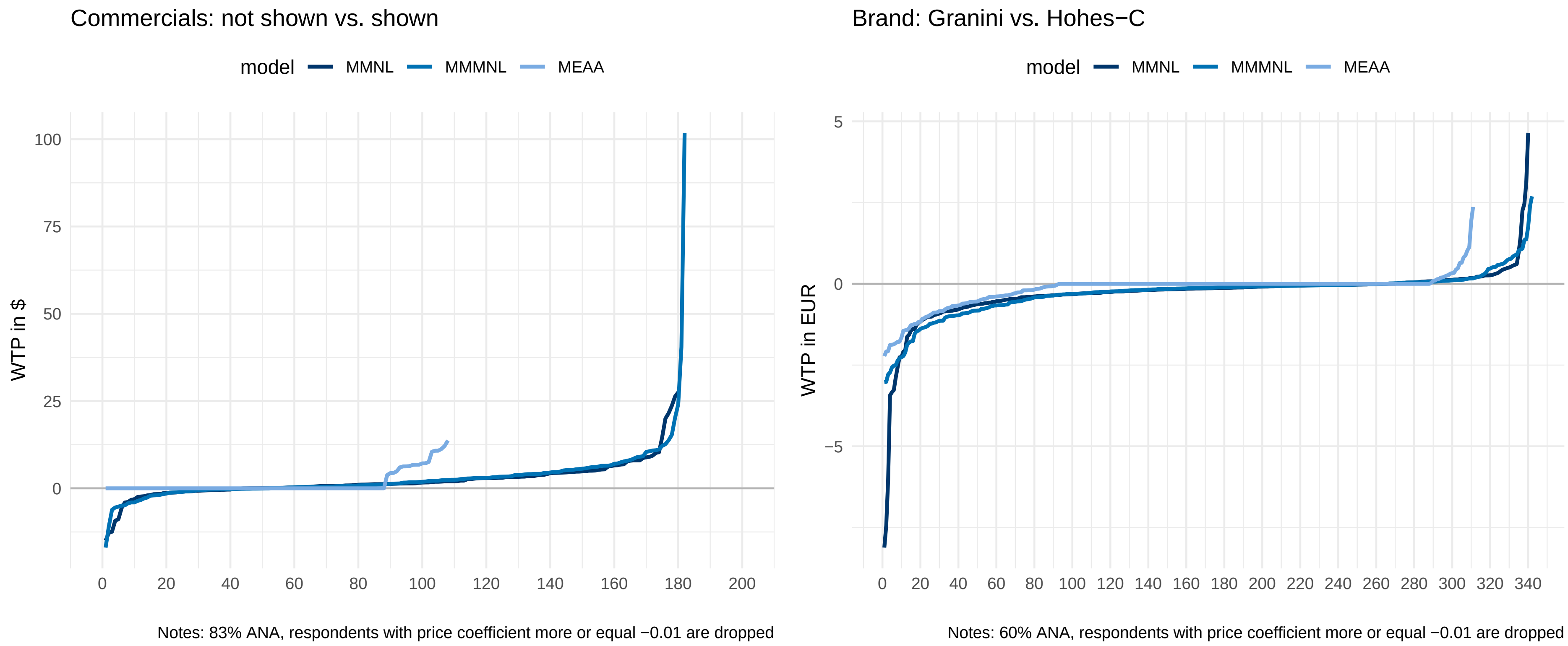
MEAA model, i.e., the amount of heterogeneity was understated. We also more clearly observe this based on the distribution of the individual preference estimates. Contrasting the right panel with a lower share of ANA vs. the left panel with a higher share of ANA, the understatement of preference heterogeneity (i.e., standard deviation) appears to be more problematic in the latter case. In general, we see similar patterns in many other cases of attributes allowing horizontal differentiation, including additional features (free public transport, parking, or VIP parking) in the Basketball tickets dataset, brand in Cameras and Smoothies, packaging in Orange juice and Smoothies, location, type of music, as well as specials in Parties, and catalog size (which included more TV or more movies content) in Video-streaming services.

Furthermore, unlike the example of brands in the Orange juice datasets, the MM-MNL model uncovers an interesting pattern in the preferences for brands in the Tablet dataset (the right panel in Figure 6). More specifically, the MM-MNL model identifies two classes with distinct preferences leading to a bimodal distribution. In one class, respondents have a very high preference for Apple and are less price-sensitive. In the second class, they prefer other brands and are relatively more price-sensitive (see Table A8). There are no substantial differences between these two classes on other attributes. The MEAA model cannot deal with such a bimodality of preference distribution, as after disentangling the zeros, it still enforces a normal distribution on the rest of the sample. As a result, while the MEAA slightly outperforms the MMNL both in- and out-of-sample, it is inferior to the MM-MNL (see Tables 3 and 4).

3.5 Managerial Implications

Overall, our analysis and detailed comparison of the different preference distributions implied by the models reveals that neglecting ANA in the model may result in considerable biases in the estimates. As a result, substantive decisions on optimal product design, market segmentation, demand estimation, (individual) pricing, and targeting may be impaired.

To illustrate, in Figure 7, we present a comparison of individual-level WTP across the models for not showing vs. showing commercials in the Video-streaming dataset and brand Granini vs. Hohes-C in the Orange juice dataset. For example, consider the potential decisions that can be made regarding offering consumers a video-streaming service with no commercials when ANA is neglected in the model. We find a larger mass with a positive valuation for this feature



(a) Video-streaming services

(b) Orange juice

Figure 7: Willingness-to-pay distribution for selected attributes

in the MMNL and the MM-MNL models, i.e., the overall demand is considerably overstated in both these models (see the left panel in Figure 7). Moreover, according to the MMNL and the MM-MNL models, there is a share of consumers willing to pay more than \$14 (the upper bound of WTP in the MEAA model) for a video-streaming service with no commercials shown. Hence, it may still seem profitable to introduce the option. In contrast, the MEAA model suggests that there is a much lower demand for this feature, and respondents are not willing to pay more than \$14 for it. It is unclear whether the revenue that can be made from this limited number of consumers would cover the opportunity cost that could have been made by showing commercials.

Similarly, if a considerable share of consumers does not care about attributes that allow differentiating products horizontally (e.g., brands), the firm's pricing decisions may be impaired. For example, consider the brand Granini vs. Hohes-C in the Orange juice dataset. We present the WTP distribution across models in the right panel in Figure 7. The analyst neglecting ANA and basing decisions on the MMNL or MM-MNL models again might overestimate the overall demand (larger mass with positive valuation compared to the MEAA model). Because of ANA, the average WTP in the MMNL and the MM-MNL models is lower compared to the MEAA model (zeros excluded). This may lead to suboptimal pricing decisions and loss of potential profits for the firm.

4 Conclusion

IIn this paper, we set out to 1) investigate the confounding between preference heterogeneity and ANA, 2) gain a deeper understanding of potential biases in uncovered preference distributions when either preference heterogeneity or ANA is neglected, and 3) examine the prevalence of ANA in a broader set of applications. The model comparison across ten empirical applications indicates that significant biases arise when preference heterogeneity or ANA is not accommodated. In particular, ignoring potential differences in individual preferences results in an overstatement of the amount of ANA. In the opposite case, not accounting for ANA leads to biases in preference estimates (both mean and variance). The magnitude of the bias depends on the location of the true preference distribution and the amount of ANA. We find that there is a higher likelihood of overestimating the amount of preference heterogeneity in the presence of ANA for attributes that allow firms to differentiate among products vertically. On the other hand, an underestimation of the preference heterogeneity can be expected for attributes that allow horizontal differentiation, such as brands. Biases in the uncovered preference distribution lead to biases in demand estimation and may impair the firm's decisions on optimal product design, introductions of new products and features, segmentation and targeting, as well as pricing (e.g., Gilbride and Lenk 2010, Allenby and Ginter 1995).

Moreover, we find that explicitly accounting for ANA in choice models is critical in categories involving low risk/stakes and more complex settings requiring consumers to make trade-offs on many attributes. As both factors may drive consumers to ignore attribute information when making choices, accommodating such behavior becomes critical. Notably, models that impose more flexible forms of preference heterogeneity cannot necessarily deal with existing ANA patterns in the data.

Several limitations of the existing approach for modeling both preference heterogeneity and ANA should be outlined. First, the MEAA model is not equipped to deal with multi-modal preference distribution. While it may sufficiently identify consumers ignoring a particular attribute, it imposes one normal distribution for the rest of the sample. From this perspective, the MEAA and the MM-MNL model seem to be complementary. Future research could focus on extensions of the MEAA model to allow for the multi-modal preferences distribution. Second, we impose the assumption of independence of attribute attendance in the models that account for ANA. While this assumption is critical for retaining parsimony, it comes at a cost. Specifically, it does not allow for any correlation in the attribute attendance probabilities. On the other hand, relaxing this assumption results in a sharp increase in the number of parameters. Further research is necessary to find new approaches to accommodate this issue.

Another avenue for future research is to investigate the drivers of ANA. In our ten applications, we found that the risk/stakes and the complexity of the choice situation may lead to a higher ANA. Understanding how these factors affect consumers' attribute processing strategies would be essential. Furthermore, while we only looked at the risk/stakes of the decision, other facets of consumer involvement may affect the observed ANA patterns.

Appendices

Appendix A

In this Appendix, we present the estimation results of all the models for each of the ten datasets. Statistically significant estimates at a 5% significance level are indicated in bold. For all the datasets, we use dummy coding for the "none" option, linear coding for the price, and effect coding for other attributes. The omitted levels are indicated in a footnote of the corresponding table.

		MNL	EAA	MMNL	MEAA	MM-MNL	
						Class 1	Class 2
Utility paramaters							
None	β	**-4.06** (0.07)	**-6.08** (0.12)	**-9.55** (0.30)	**-11.81** (0.47)	**-16.17** (0.84)	**-5.49** (0.35)
	σ			**5.87** (0.28)	**8.70** (0.35)	**6.73** (0.45)	**3.06** (0.25)
Brand:							
A	β	**0.30** (0.02)	**1.13** (0.10)	**0.55** (0.04)	**1.49** (0.15)	**0.20** (0.06)	**1.21** (0.11)
	σ			**0.68** (0.06)	**1.49** (0.13)	**0.28** (0.09)	**1.22** (0.11)
B	β	**-0.03** (0.03)	**-0.63** (0.10)	**-0.06** (0.04)	**-0.28** (0.11)	**0.40** (0.07)	**-0.49** (0.07)
	σ			**0.35** (0.06)	**1.28** (0.11)	**0.46** (0.07)	0.12 (0.20)
C	β	**0.06** (0.03)	**1.02** (0.09)	**0.06** (0.05)	**0.60** (0.16)	**-0.43** (0.07)	**0.84** (0.10)
	σ			**0.92** (0.06)	**2.78** (0.21)	**0.58** (0.08)	**1.37** (0.11)
Packaging:							
Glass bottle	β	**0.54** (0.02)	**1.50** (0.06)	**0.79** (0.05)	**1.74** (0.11)	**-0.06** (0.10)	**1.74** (0.10)
	σ			**0.85** (0.05)	**1.37** (0.10)	**0.75** (0.10)	**0.85** (0.07)
Tetrapak	β	**-0.79** (0.03)	**-2.09** (0.09)	**-1.30** (0.06)	**-2.50** (0.14)	**-0.62** (0.11)	**-2.34** (0.16)
	σ			**1.06** (0.07)	**1.57** (0.11)	**0.93** (0.11)	**1.23** (0.11)
Plastic cup	β	**-0.42** (0.03)	**-0.87** (0.07)	**-0.63** (0.05)	**-1.14** (0.10)	**-0.37** (0.08)	**-0.98** (0.11)
	σ			**0.87** (0.06)	**1.48** (0.09)	**0.80** (0.09)	**1.08** (0.10)
Price	β	**-2.02** (0.05)	**-4.07** (0.09)	**-3.83** (0.16)	**-6.25** (0.25)	**-7.23** (0.38)	**-2.06** (0.18)
	σ			**3.53** (0.14)	**3.09** (0.21)	**5.23** (0.34)	**1.84** (0.16)
Class paramaters							
Brand			**-1.15** (0.14)		**-0.46** (0.12)		
Packaging			-0.05 (0.10)		**0.69** (0.12)		
Price			**0.57** (0.10)		**0.75** (0.13)		
Class 2							0.03 (0.03)
LL		−7931.84	−6945.04	−6233.87	−5863.34		−5806.92
BIC		15935.46	13988.78	12611.31	11897.18		11909.96
No. parameters		(8)	(11)	(16)	(19)		(33)

Note: The omitted levels are Brand: D, Packaging: plastic bottle.

Table A1: Estimation results for the dataset: Smoothies

		MNL	EAA	MMNL	MEAA	MM-MNL	
						Class 1	Class 2
Utility parameters							
None	β	**-3.63** (0.10)	**-5.12** (0.14)	**-8.23** (0.32)	**-9.31** (0.44)	**-18.13** (1.28)	**-3.12** (0.31)
	σ			**5.77** (0.33)	**5.95** (0.27)	**3.89** (0.39)	**2.81** (0.27)
Brand:							
Granini	β	**0.24** (0.06)	**1.28** (0.13)	**0.25** (0.10)	**1.10** (0.40)	**-0.08** (0.23)	**0.50** (0.17)
	σ			**1.35** (0.11)	**4.58** (0.53)	**1.52** (0.23)	**1.59** (0.16)
Hohes-C	β	**0.74** (0.05)	**2.58** (0.13)	**1.25** (0.11)	**3.13** (0.26)	**1.05** (0.22)	**1.44** (0.17)
	σ			**1.55** (0.11)	**3.16** (0.28)	**1.43** (0.21)	**2.09** (0.17)
Valensina	β	**0.23** (0.06)	**1.05** (0.13)	**0.24** (0.10)	0.31 (0.35)	0.01 (0.22)	0.24 (0.17)
	σ			**1.16** (0.11)	**4.13** (0.33)	**1.41** (0.26)	**1.57** (0.15)
Fairtrade label:							
Yes	β	**0.86** (0.04)	**2.20** (0.11)	**1.76** (0.10)	**2.67** (0.20)	**2.76** (0.28)	**1.46** (0.13)
	σ			**1.60** (0.11)	**1.73** (0.12)	**2.41** (0.21)	**1.42** (0.14)
Packaging:							
Carton	β	**0.38** (0.04)	**2.40** (0.15)	**0.78** (0.12)	**1.88** (0.25)	**1.20** (0.25)	**0.72** (0.18)
	σ			**2.07** (0.11)	**3.83** (0.27)	**2.42** (0.27)	**2.21** (0.15)
Price	β	**-2.69** (0.07)	**-4.05** (0.10)	**-5.06** (0.18)	**-6.50** (0.23)	**-12.22** (0.84)	**-2.49** (0.18)
	σ			**2.72** (0.15)	**2.35** (0.15)	**1.54** (0.20)	**0.87** (0.12)
Class parameters							
Brand			**-0.45** (0.16)		**-0.42** (0.14)		
Fairtrade label			-0.24 (0.16)		**0.75** (0.23)		
Packaging			**-1.22** (0.18)		0.01 (0.17)		
Price			**1.95** (0.19)		**2.20** (0.24)		
Class 2							**0.12** (0.04)
LL		−5083.52	−4575.24	−4049.22	−3880.15		−3836.18
BIC		10226.37	9243.69	8217.08	7912.83		7918.10
No. parameters		(7)	(11)	(14)	(18)		(29)

Notes: The omitted levels are Brand: Albi, Fairtrade label: No, Packaging: PET.

Table A2: Estimation results for the dataset: Orange juice

		MNL	EAA	MMNL	MEAA	MM-MNL	
						Class 1	Class 2
Utility paramaters							
None	β	**-1.42** (0.12)	**-2.61** (0.19)	**-4.17** (0.37)	**-4.81** (0.52)	**-5.15** (0.63)	**-4.22** (0.85)
	σ			**3.07** (0.37)	**4.21** (0.52)	**3.98** (0.55)	**1.92** (0.45)
Privacy policy:							
Share usage	β	0.05 (0.03)	**0.67** (0.20)	**0.10** (0.04)	**1.07** (0.38)	0.07 (0.04)	0.30 (0.18)
	σ			0.11 (0.17)	0.36 (0.53)	0.11 (0.10)	**0.65** (0.19)
Share all	β	-0.19 (0.03)	**-2.19** (0.39)	**-0.27** (0.05)	**-3.04** (0.74)	-0.06 (0.05)	**-1.77** (0.31)
	σ			**0.47** (0.06)	**0.86** (0.86)	**0.20** (0.09)	**1.71** (0.29)
Commercials:							
Shown	β	**-0.11** (0.02)	**-1.36** (0.17)	**-0.17** (0.04)	**-1.81** (0.32)	-0.03 (0.04)	**-1.15** (0.22)
	σ			**0.40** (0.04)	0.60 (0.42)	**0.27** (0.06)	**1.33** (0.25)
Fast content:							
Yes	β	**0.20** (0.02)	**1.70** (0.23)	**0.24** (0.04)	**2.08** (0.56)	**0.11** (0.03)	**1.59** (0.29)
	σ			**0.36** (0.05)	**1.09** (0.33)	0.11 (0.09)	**1.83** (0.31)
Catalog size:							
More TV	β	**-0.11** (0.03)	**-1.70** (0.32)	-0.09 (0.05)	**-0.85** (0.40)	-0.06 (0.05)	-0.35 (0.25)
	σ			**0.33** (0.07)	**2.24** (0.40)	0.00 (0.28)	**2.53** (0.42)
More movies	β	**0.22** (0.03)	**1.69** (0.21)	**0.28** (0.04)	**2.00** (0.29)	**0.13** (0.05)	**1.64** (0.31)
	σ			**0.35** (0.06)	**0.54** (0.25)	0.12 (0.13)	**1.85** (0.33)
Price	β	**-0.08** (0.01)	**-0.38** (0.02)	**-0.13** (0.02)	**-0.30** (0.05)	**-0.09** (0.03)	**-0.53** (0.10)
	σ			**0.26** (0.02)	**0.39** (0.04)	**0.27** (0.03)	**0.28** (0.06)
Class parameters							
Privacy policy			**-1.76** (0.28)		**-1.63** (0.28)		
Commercials			**-1.72** (0.24)		**-1.61** (0.28)		
Fast content			**-1.88** (0.24)		**-1.62** (0.29)		
Catalog size			**-1.73** (0.25)		**-1.27** (0.22)		
Price			**-0.54** (0.16)		0.39 (0.31)		
Class 2							**-0.94** (0.06)
LL		−3591.15	−3278.25	−3264.93	−3167.17		−3161.75
BIC		7244.36	6657.35	6654.00	6497.27		6579.51
No. parameters		(8)	(13)	(16)	(21)		(33)

Notes: The omitted levels are Privacy policy: No sharing, Commercials: Not shown, Fast content: No, Catalog size: 5000 movies, 2500 TV episodes.

Table A3: Estimation results for the dataset: Video-streaming services

		MNL	EAA	MMNL	MEAA	MM-MNL	
						Class 1	Class 2
Utility paramaters							
None	β	**-0.55** (0.07)	**-0.40** (0.08)	**-0.55** (0.16)	**-0.46** (0.18)	0.37 (0.71)	**-0.69** (0.18)
	σ			**1.61** (0.16)	**1.85** (0.20)	**9.07** (1.89)	**1.48** (0.18)
Location:							
Image	β	0.00 (0.07)	0.83 (0.43)	-0.06 (0.12)	0.42 (0.33)	**1.60** (0.62)	-0.19 (0.15)
	σ			0.33 (0.21)	0.80 (0.46)	**5.68** (1.26)	0.32 (0.24)
Cafe Madrid	β	**0.15** (0.07)	**1.36** (0.40)	**0.41** (0.12)	**1.87** (0.43)	**4.24** (1.00)	0.24 (0.15)
	σ			**0.41** (0.19)	**0.63** (0.32)	**2.23** (0.69)	0.25 (0.24)
B9	β	**-0.24** (0.08)	**-1.78** (0.67)	**-0.34** (0.12)	**-1.78** (0.46)	**-1.91** (0.79)	**-0.38** (0.15)
	σ			0.38 (0.32)	0.57 (0.72)	**3.19** (0.82)	0.28 (0.24)
Westbhf	β	**-0.19** (0.08)	**-1.72** (0.65)	**-0.31** (0.13)	**-1.27** (0.50)	**-4.36** (0.99)	-0.31 (0.16)
	σ			**0.57** (0.20)	**1.92** (0.57)	**3.57** (0.80)	**0.80** (0.24)
Apollo	β	0.13 (0.07)	0.31 (0.40)	**0.31** (0.13)	**0.82** (0.39)	**-2.72** (0.94)	**0.62** (0.15)
	σ			**0.78** (0.14)	**1.86** (0.43)	**8.73** (1.95)	**0.42** (0.21)
Drinks:							
Cheap prices	β	**0.63** (0.04)	**1.39** (0.13)	**1.17** (0.10)	**2.03** (0.23)	**2.57** (0.66)	**1.36** (0.13)
	σ			**1.02** (0.11)	**1.03** (0.18)	**4.68** (1.02)	**0.92** (0.14)
Normal prices	β	0.07 (0.05)	**0.27** (0.10)	**0.23** (0.07)	**0.35** (0.12)	-0.74 (0.44)	**0.33** (0.10)
	σ			**0.30** (0.13)	**0.45** (0.19)	**1.73** (0.52)	**0.51** (0.14)
Music:							
Mix	β	**0.28** (0.05)	**0.47** (0.15)	**0.80** (0.10)	**1.13** (0.16)	**2.47** (0.66)	**0.86** (0.12)
	σ			**0.44** (0.16)	**0.60** (0.20)	**4.61** (1.00)	**0.40** (0.18)
R&B/Hip hop	β	**-0.48** (0.06)	**-1.05** (0.23)	**-1.06** (0.16)	**-1.34** (0.24)	**-1.62** (0.73)	**-1.21** (0.17)
	σ			**1.65** (0.16)	**2.31** (0.28)	**11.76** (2.45)	**1.03** (0.19)
House	β	**0.36** (0.05)	**2.09** (0.19)	**0.60** (0.13)	**1.10** (0.23)	**6.50** (1.45)	**0.42** (0.17)
	σ			**1.59** (0.14)	**2.34** (0.28)	**10.69** (2.26)	**1.66** (0.20)
Dress code:							
No sneakers	β	**-0.18** (0.03)	**-0.90** (0.31)	**-0.33** (0.06)	**-0.90** (0.34)	-0.57 (0.30)	**-0.43** (0.08)
	σ			**0.47** (0.09)	**0.79** (0.19)	**1.84** (0.47)	**0.51** (0.11)
Specials:							
None	β	**-0.11** (0.05)	**2.32** (0.93)	-0.12 (0.07)	-0.23 (0.18)	**-2.19** (0.65)	-0.02 (0.09)
	σ			**0.39** (0.12)	**0.57** (0.27)	0.75 (0.46)	0.23 (0.17)
Happy hour	β	**0.17** (0.04)	**3.14** (1.03)	**0.24** (0.08)	0.50 (0.29)	**1.86** (0.54)	**0.26** (0.09)
	σ			**0.36** (0.12)	**0.91** (0.34)	**4.18** (0.87)	0.25 (0.14)
Price	β	**-0.17** (0.01)	**-0.31** (0.03)	**-0.31** (0.02)	**-0.45** (0.05)	**-0.35** (0.11)	**-0.39** (0.04)
	σ			**0.12** (0.03)	**0.16** (0.06)	**0.47** (0.12)	**0.18** (0.04)
Class paramaters							
Location			**-1.73** (0.48)		-0.62 (0.32)		
Drinks			0.31 (0.25)		**1.18** (0.38)		
Music			**-0.76** (0.20)		**2.14** (0.56)		
Dress code			**-1.24** (0.61)		-0.24 (0.64)		
Specials			**-3.44** (0.52)		0.65 (1.78)		
Price			**1.00** (0.29)		**1.79** (0.64)		
Class 2							**1.12** (0.07)
LL		−2016.36	−1924.86	−1771.78	−1750.12		−1713.11
BIC		4144.27	4005.88	3766.65	3767.93		3879.82
No. parameters		(15)	(21)	(30)	(36)		(61)

Notes: The omitted levels are Location: Abendrot, Drinks: expensive, Music: Rock/Alternative, Dress code: None, Specials: Go-go dancers.

Table A4: Estimation results for the dataset: Parties

		MNL	EAA	MMNL	MEAA	MM-MNL	
						Class 1	Class 2
Utility parameters							
None	β	**-2.21** (0.09)	**-3.14** (0.13)	**-4.61** (0.29)	**-4.99** (0.45)	**-1.76** (0.40)	**-6.07** (0.52)
	σ			**2.32** (0.19)	**3.53** (0.32)	**1.50** (0.57)	**3.14** (0.34)
Brand:							
Ovente	β	-0.03 (0.04)	**-1.96** (0.54)	-0.12 (0.06)	**-3.50** (0.74)	-0.08 (0.14)	**-0.14** (0.07)
	σ			**0.34** (0.08)	0.01 (0.47)	0.18 (0.16)	**0.34** (0.09)
Hamilton Beach	β	-0.01 (0.04)	**1.28** (0.45)	0.05 (0.06)	**2.12** (0.53)	-0.01 (0.13)	0.11 (0.07)
	σ			**0.17** (0.07)	0.22 (0.35)	0.26 (0.24)	0.15 (0.11)
Capacity:							
1.2 liter	β	**-0.09** (0.04)	**-2.22** (0.68)	**-0.19** (0.06)	-2.24 (2.53)	**-0.30** (0.14)	**-0.14** (0.07)
	σ			**0.16** (0.08)	0.22 (1.04)	**0.42** (0.13)	0.16 (0.11)
1.5 liter	β	**0.07** (0.04)	**1.38** (0.55)	0.08 (0.06)	1.18 (1.69)	0.15 (0.12)	0.07 (0.07)
	σ			**0.16** (0.07)	0.59 (0.54)	0.13 (0.16)	0.07 (0.12)
Material:							
Glass	β	**0.27** (0.04)	**1.19** (0.12)	**0.62** (0.08)	**1.78** (0.30)	**1.97** (0.27)	**0.19** (0.07)
	σ			**0.87** (0.08)	**1.80** (0.29)	**2.94** (0.80)	0.06 (0.14)
Stainless steel	β	**0.28** (0.04)	**1.14** (0.12)	**0.51** (0.08)	**1.74** (0.29)	**2.39** (0.28)	**0.20** (0.08)
	σ			**0.83** (0.08)	**1.55** (0.22)	**1.59** (0.21)	**0.28** (0.09)
Variable temp.:							
Yes	β	**0.12** (0.03)	**2.17** (0.38)	**0.23** (0.05)	**1.28** (0.38)	**0.73** (0.14)	0.06 (0.05)
	σ			**0.37** (0.06)	**1.19** (0.30)	**0.93** (0.16)	0.16 (0.10)
Power:							
1100 Watts	β	-0.01 (0.03)	**-1.38** (0.44)	-0.02 (0.04)	**-2.26** (1.13)	0.08 (0.11)	-0.03 (0.05)
	σ			**0.16** (0.07)	**3.33** (0.99)	**0.45** (0.11)	0.14 (0.09)
Amazon rating:							
3 stars	β	**-0.84** (0.05)	**-2.60** (0.18)	**-1.70** (0.12)	**-3.68** (0.32)	**-1.59** (0.22)	**-1.95** (0.18)
	σ			**1.47** (0.12)	**0.76** (0.21)	**1.35** (0.22)	**1.41** (0.15)
4 stars	β	**0.11** (0.04)	**0.62** (0.09)	**0.33** (0.06)	**0.92** (0.14)	**0.31** (0.14)	**0.39** (0.07)
	σ			0.17 (0.09)	**0.48** (0.14)	**0.30** (0.14)	0.02 (0.28)
Price	β	**-0.30** (0.01)	**-0.77** (0.03)	**-0.53** (0.04)	**-1.00** (0.08)	**-0.53** (0.08)	**-0.72** (0.07)
	σ			**0.64** (0.05)	**0.51** (0.05)	**0.36** (0.06)	**0.56** (0.05)
Class parameters							
Brand			**-3.11** (0.60)		**-2.96** (0.45)		
Capacity			**-3.20** (0.75)		-2.53 (1.85)		
Material			**-0.52** (0.19)		-0.18 (0.21)		
Variable temp.			**-2.87** (0.44)		**-1.65** (0.42)		
Power			**-3.83** (0.91)		**-3.82** (0.75)		
Rating			0.29 (0.18)		**0.47** (0.18)		
Price			**0.52** (0.17)		**1.16** (0.26)		
Class 2							**0.58** (0.05)
LL		−2380.63	−1963.63	−1870.29	−1753.58		−1780.36
BIC		4854.12	4074.29	3926.32	3747.06		3939.93
No. parameters		(12)	(19)	(24)	(31)		(49)

Notes: The omitted levels are Brand: Cuisinart, Capacity: 1.7 liters, Material: Plastic, Variable temperature: No, Power: 1500 Watts, Amazon rating: 5 stars.

Table A5: Estimation results for the dataset: Electric kettles

		MNL	EAA	MMNL	MEAA	MM-MNL	
						Class 1	Class 2
Utility Parameters							
None	β	**-2.45** (0.31)	**-5.63** (0.45)	**-5.52** (0.69)	**-6.31** (0.67)	**-6.63** (1.16)	**-4.89** (1.00)
	σ			**2.80** (0.54)	**1.82** (0.42)	**1.43** (0.37)	**1.92** (0.50)
Price category:							
Category 1	β	**0.5** (0.13)	**2.76** (0.23)	0.62 (0.35)	**1.33** (0.45)	-0.06 (0.50)	**1.02** (0.48)
	σ			**3.01** (0.43)	**4.20** (0.47)	**1.78** (0.34)	**3.20** (0.41)
Category 2	β	**0.14** (0.07)	**1.53** (0.19)	**0.60** (0.21)	**1.77** (0.34)	0.23 (0.43)	0.56 (0.33)
	σ			**1.93** (0.22)	**2.60** (0.37)	**2.34** (0.50)	**3.12** (0.44)
Category 3	β	**-0.31** (0.08)	**-1.16** (0.19)	**0.45** (0.15)	0.31 (0.27)	**0.60** (0.27)	0.00 (0.28)
	σ			**0.76** (0.17)	**1.62** (0.28)	**1.16** (0.30)	**0.70** (0.22)
Additional features:							
Free parking	β	**-0.23** (0.06)	**-8.35** (2.33)	**-0.83** (0.19)	**-1.26** (0.35)	**1.25** (0.41)	**-2.40** (0.36)
	σ			**1.77** (0.18)	**2.92** (0.38)	**1.40** (0.30)	**0.90** (0.30)
Free VIP	β	0.08 (0.06)	**2.99** (0.82)	0.26 (0.14)	0.29 (0.24)	**-0.67** (0.28)	**0.98** (0.22)
parking	σ			**1.28** (0.16)	**1.77** (0.29)	**1.17** (0.28)	**1.23** (0.24)
Free public	β	0.04 (0.06)	**3.01** (0.79)	**0.20** (0.10)	**0.43** (0.20)	**-0.68** (0.25)	**0.81** (0.19)
transport	σ			0.32 (0.20)	**1.04** (0.21)	**0.42** (0.25)	**0.50** (0.19)
Price	β	**-0.11** (0.02)	**-0.36** (0.03)	**-0.25** (0.03)	**-0.39** (0.04)	**-0.33** (0.06)	**-0.26** (0.05)
	σ			**0.15** (0.02)	**0.11** (0.02)	**0.12** (0.03)	**0.18** (0.03)
Class Parameters							
Price category			**-0.64** (0.20)		**0.88** (0.28)		
Additional features			**-1.39** (0.23)		0.73 (0.37)		
Price			**0.55** (0.18)		**1.38** (0.34)		
Class 2							**0.47** (0.10)
LL		−1959.44	−1561.67	−1370.05	−1350.41		−1303.80
BIC		3976.11	3202.03	2854.58	2836.75		2843.69
No. parameters		(8)	(11)	(16)	(19)		(33)

Notes: The omitted levels are Price category: 4, Additional feature: None.

Table A6: Estimation results for the dataset: Basketball tickets

		MNL	EAA	MMNL	MEAA	MM-MNL	
						Class 1	Class 2
Utility parameters							
Hard drive:							
500GB	β	**-0.47** (0.05)	**-1.19** (0.15)	**-0.81** (0.11)	**-1.31** (0.16)	**-4.45** (1.94)	**-0.75** (0.12)
	σ			**0.66** (0.11)	0.31 (0.17)	**3.48** (1.39)	**0.56** (0.12)
750GB	β	0.03 (0.05)	0.08 (0.09)	0.06 (0.07)	0.08 (0.09)	**-1.04** (0.50)	0.15 (0.08)
	σ			0.01 (0.14)	0.17 (0.17)	1.66 (0.92)	0.03 (0.12)
Memory:							
4GB	β	**-0.71** (0.05)	**-1.77** (0.20)	**-1.22** (0.14)	**-1.66** (0.23)	**-8.51** (3.46)	**-1.04** (0.15)
	σ			**0.98** (0.12)	**1.00** (0.14)	**6.17** (2.48)	**0.82** (0.13)
6GB	β	**0.11** (0.05)	0.16 (0.10)	**0.21** (0.08)	**0.24** (0.10)	0.84 (0.54)	**0.20** (0.08)
	σ			**0.35** (0.11)	**0.35** (0.14)	**3.10** (1.44)	0.21 (0.14)
Dispaly size:							
12 inch	β	**-0.85** (0.06)	**-2.71** (0.21)	**-1.74** (0.17)	**-2.38** (0.30)	**-8.97** (3.74)	**-1.50** (0.20)
	σ			**1.39** (0.14)	**1.52** (0.19)	**7.32** (2.84)	**1.27** (0.18)
14 inch	β	**0.14** (0.05)	**0.59** (0.10)	**0.40** (0.09)	**0.56** (0.13)	**2.41** (1.14)	**0.37** (0.09)
	σ			**0.49** (0.09)	**0.58** (0.15)	**3.60** (1.34)	0.16 (0.16)
Price	β	**-0.79** (0.05)	**-2.12** (0.14)	**-1.51** (0.15)	**-2.02** (0.25)	**-2.42** (0.93)	**-1.72** (0.19)
	σ			**1.45** (0.16)	**1.56** (0.21)	**3.19** (1.23)	**1.31** (0.15)
Class parameters							
Hard drive			0.44 (0.36)		**0.89** (0.38)		
Memory			0.22 (0.30)		**1.68** (0.69)		
Display size			0.14 (0.21)		**1.51** (0.45)		
Price			0.25 (0.22)		**1.40** (0.47)		
Class 2							**1.06** (0.08)
LL		−1239.08	−1034.99	−992.02	−979.90		−969.00
BIC		2529.63	2150.86	2086.98	2092.14		2151.23
No. parameters		(7)	(11)	(14)	(18)		(29)

Notes: The omitted levels are Hard drive: 1 TB, Memory: 8 GB, Screen size: 15.6 inch.

Table A7: Estimation results for the dataset: Laptops

		MNL	EAA	MMNL	MEAA	MM-MNL	
						Class 1	Class 2
Utility paramaters							
None	β	**-1.53** (0.07)	**-1.88** (0.09)	**-2.81** (0.22)	**-3.08** (0.26)	**-3.46** (0.25)	0.12 (0.71)
	σ			**2.40** (0.22)	**2.72** (0.26)	**2.29** (0.21)	**3.70** (1.00)
Brand:							
Apple	β	0.02 (0.04)	**2.92** (0.22)	0.12 (0.13)	0.18 (0.27)	**-0.93** (0.11)	**5.53** (1.18)
	σ			**1.75** (0.13)	**2.58** (0.35)	**1.01** (0.11)	**0.92** (0.40)
Samsung	β	0.00 (0.05)	**-1.00** (0.21)	**0.16** (0.08)	0.13 (0.13)	**0.51** (0.08)	**-3.40** (0.90)
	σ			**0.61** (0.11)	**1.01** (0.21)	**0.44** (0.12)	**2.02** (0.69)
Display size:							
7 inch	β	**-0.30** (0.03)	**-1.04** (0.19)	**-0.52** (0.06)	**-1.11** (0.31)	**-0.70** (0.08)	-0.29 (0.25)
	σ			**0.60** (0.08)	**0.80** (0.16)	**0.64** (0.08)	**0.87** (0.33)
Battery:							
7 hours	β	**-0.17** (0.03)	**-0.35** (0.13)	**-0.33** (0.05)	**-0.76** (0.30)	**-0.39** (0.06)	-0.18 (0.19)
	σ			**0.24** (0.10)	**0.44** (0.16)	**0.43** (0.08)	0.36 (0.24)
Resolution:							
1280×800px	β	**-0.20** (0.03)	**-1.07** (0.15)	**-0.37** (0.06)	**-1.49** (0.33)	**-0.42** (0.07)	**-0.41** (0.20)
	σ			**0.56** (0.07)	**0.75** (0.25)	**0.60** (0.08)	0.48 (0.27)
Storage capacity:							
16GB	β	**-0.24** (0.05)	**-5.93** (2.42)	**-0.37** (0.08)	-1.37 (0.73)	**-0.53** (0.10)	-0.39 (0.34)
	σ			**0.38** (0.13)	**1.02** (0.48)	**0.52** (0.14)	**1.37** (0.41)
32GB	β	**0.22** (0.04)	**3.11** (1.27)	**0.21** (0.06)	**0.48** (0.23)	**0.29** (0.08)	-0.13 (0.24)
	σ			0.11 (0.14)	0.59 (0.37)	0.21 (0.15)	0.14 (0.39)
Price	β	**-0.47** (0.02)	**-0.80** (0.03)	**-0.87** (0.05)	**-1.12** (0.08)	**-1.06** (0.07)	**-0.52** (0.17)
	σ			**0.49** (0.04)	**0.38** (0.08)	**0.54** (0.05)	**0.84** (0.20)
Class paramaters							
Brand			**-1.52** (0.20)		**1.04** (0.50)		
Size			-0.72 (0.40)		0.28 (0.66)		
Battery			0.78 (1.26)		-0.15 (0.87)		
Resolution			**-0.94** (0.30)		-0.65 (0.38)		
Storage capacity			**-3.56** (0.61)		-0.98 (0.74)		
Price			**1.38** (0.21)		**1.84** (0.43)		
Class 2							**-1.42** (0.06)
LL		−2414.47	−2142.53	−1952.79	−1926.38		−1846.54
BIC		4897.65	4399.59	4043.01	4036.00		3975.59
No. parameters		(9)	(15)	(18)	(24)		(37)

Notes: The omitted levels are Brand: Smarttab, Display size 10 inch, Battery: 11 hours, Resolution: 2560×1600px, Storage capacity: 64GB.

Table A8: Estimation results for the dataset: Tablets

		MNL	EAA	MMNL	MEAA	MM-MNL	
						Class 1	Class 2
Utility parameters							
None	β	**-1.83** (0.06)	**-2.62** (0.08)	**-4.46** (0.19)	**-4.92** (0.27)	**-7.86** (0.52)	**-2.43** (0.23)
	σ			**3.50** (0.19)	**3.80** (0.27)	**4.54** (0.35)	**1.82** (0.17)
Brand:							
Sony	β	-0.04 (0.04)	**-1.20** (0.40)	-0.04 (0.05)	-0.13 (0.26)	0.02 (0.11)	-0.02 (0.08)
	σ			**0.44** (0.07)	**1.69** (0.23)	**0.73** (0.12)	**0.35** (0.14)
Nikon	β	**0.06** (0.03)	**1.54** (0.28)	0.11 (0.06)	**0.85** (0.28)	**0.39** (0.10)	-0.07 (0.08)
	σ			**0.58** (0.06)	**1.97** (0.24)	**0.88** (0.09)	**0.25** (0.11)
Panasonic	β	**-0.22** (0.04)	**-1.84** (0.32)	**-0.29** (0.05)	**-1.66** (0.34)	**-0.56** (0.12)	**-0.20** (0.08)
	σ			**0.28** (0.10)	**1.92** (0.28)	**1.07** (0.12)	0.16 (0.18)
Pixel:							
High	β	**0.38** (0.02)	**1.58** (0.11)	**0.60** (0.05)	**1.69** (0.17)	0.08 (0.06)	**1.18** (0.09)
	σ			**0.69** (0.05)	**0.86** (0.14)	**0.22** (0.08)	**0.85** (0.08)
Zoom:							
High	β	**0.42** (0.02)	**1.46** (0.09)	**0.68** (0.05)	**1.39** (0.15)	**0.20** (0.06)	**1.35** (0.09)
	σ			**0.71** (0.05)	**0.94** (0.12)	**0.23** (0.10)	**0.65** (0.07)
Video:							
Yes	β	**0.31** (0.02)	**1.27** (0.11)	**0.53** (0.04)	**1.25** (0.16)	**0.42** (0.07)	**0.73** (0.07)
	σ			**0.48** (0.05)	**0.54** (0.15)	**0.38** (0.10)	**0.65** (0.08)
Swivel:							
Yes	β	**0.17** (0.02)	**1.62** (0.17)	**0.25** (0.04)	**0.63** (0.13)	**0.19** (0.06)	**0.45** (0.08)
	σ			**0.49** (0.05)	**1.13** (0.15)	**0.46** (0.08)	**0.77** (0.09)
WiFi:							
Yes	β	**0.28** (0.02)	**1.29** (0.11)	**0.47** (0.04)	**1.06** (0.16)	**0.53** (0.08)	**0.57** (0.07)
	σ			**0.61** (0.06)	**0.94** (0.11)	**0.84** (0.11)	**0.61** (0.08)
Price	β	**-1.48** (0.03)	**-2.55** (0.06)	**-2.66** (0.11)	**-3.71** (0.17)	**-4.72** (0.35)	**-1.92** (0.12)
	σ			**1.80** (0.12)	**1.72** (0.14)	**2.49** (0.21)	**1.29** (0.10)
Class parameters							
Brand			**-2.17** (0.32)		**-1.36** (0.22)		
Pixel			**-0.97** (0.17)		-0.30 (0.20)		
Zoom			**-0.58** (0.16)		0.49 (0.32)		
Video			**-0.97** (0.19)		-0.09 (0.28)		
Swivel			**-1.86** (0.23)		-0.10 (0.32)		
WiFi			**-0.98** (0.19)		0.15 (0.33)		
Price			**1.46** (0.15)		**2.27** (0.30)		
Class 2							**0.08** (0.04)
LL		−5701.87	−4958.03	−4341.66	−4131.95		−4145.62
BIC		11488.17	10059.61	8852.21	8491.89		8637.45
No. parameters		(10)	(17)	(20)	(27)		(41)

Notes: The omitted levels are Brand: Canon, Pixels: Low, Zoom: Low, Video: No, Swivel: No, WiFi: No.

Table A9: Estimation results for the dataset: Cameras

		MNL	EAA	MMNL	MEAA	MM-MNL	
						Class 1	Class 2
Utility parameters							
Destination:							
Overseas	β	**0.09** (0.02)	**1.64** (0.16)	**0.18** (0.08)	0.29 (0.22)	**-0.96** (0.43)	**0.22** (0.05)
	σ			**1.16** (0.08)	**2.39** (0.32)	**14.56** (4.26)	**0.48** (0.07)
Airline:							
Virgin	β	-0.01 (0.02)	**2.44** (1.06)	-0.03 (0.04)	-0.24 (0.33)	0.49 (0.26)	-0.07 (0.04)
	σ			**0.35** (0.05)	1.11 (0.57)	0.70 (0.36)	**0.38** (0.06)
Length of stay:							
12 days	β	**0.27** (0.02)	**1.21** (0.12)	**0.53** (0.05)	**0.94** (0.16)	**1.49** (0.45)	**0.53** (0.05)
	σ			**0.58** (0.06)	**0.70** (0.09)	**3.95** (1.14)	**0.36** (0.07)
Meal:							
Included	β	**0.26** (0.02)	**1.53** (0.18)	**0.53** (0.05)	**1.09** (0.18)	**1.22** (0.38)	**0.54** (0.05)
	σ			**0.47** (0.05)	**0.57** (0.11)	0.42 (0.31)	**0.48** (0.06)
Local tours:							
Available	β	**0.09** (0.02)	**1.16** (0.25)	**0.19** (0.04)	**0.70** (0.28)	-0.13 (0.31)	**0.20** (0.04)
	σ			**0.32** (0.06)	**0.53** (0.15)	**3.02** (0.92)	**0.25** (0.07)
Peak season:							
Peak	β	**0.04** (0.02)	**1.97** (0.60)	0.05 (0.04)	0.50 (0.46)	**0.90** (0.37)	0.02 (0.04)
	σ			**0.32** (0.06)	**2.00** (0.62)	**2.22** (0.65)	0.06 (0.11)
Accomodation:							
4-star	β	**0.43** (0.02)	**1.56** (0.10)	**0.87** (0.06)	**1.53** (0.16)	**2.10** (0.79)	**0.91** (0.07)
	σ			**0.79** (0.06)	**0.88** (0.11)	**2.26** (0.82)	**0.77** (0.07)
Price	β	**-0.17** (0.02)	**-1.00** (0.17)	**-0.33** (0.04)	**-0.76** (0.17)	0.01 (0.21)	**-0.39** (0.04)
	σ			**0.34** (0.05)	**0.44** (0.11)	0.11 (0.21)	**0.39** (0.06)
Class parameters							
Destination			**-1.09** (0.19)		0.12 (0.26)		
Airline			**-4.14** (0.65)		-1.35 (0.91)		
Length of stay			-0.29 (0.19)		**1.11** (0.54)		
Meal included			**-0.92** (0.20)		0.33 (0.36)		
Tours			**-1.87** (0.39)		-0.60 (0.63)		
Season			**-3.49** (0.63)		**-2.45** (0.54)		
Accommodation			-0.07 (0.14)		**0.87** (0.29)		
Price			**-1.07** (0.32)		0.09 (0.51)		
Class 2							**1.09** (0.04)
LL		−2686.87	−2395.59	−2262.40	−2172.05		−2160.81
BIC		5441.27	4926.23	4659.86	4546.68		4600.18
No. parameters		(8)	(16)	(16)	(24)		(33)

Notes: The omitted levels are Destination: Australia, Airline: Qantas, Length of stay: 7 days, Meal: Not included, Local tours: Not available, Peak season: Off-peak, Accomodation: 2-star

Table A10: Estimation results for the dataset: Holiday destinations

Appendix B

In this Appendix, we plot the individual-level estimates in the MEAA model against those in the MMNL and MMMNL models along with the results of the fitted linear regression of the form: $\beta_i^{\text{MEAA}} = a + b \cdot \beta_i^{\text{(M)M-MNL}} + \varepsilon_i$. The estimated slopes $\hat{b}$ serve as a rescaling coefficient for the (M)M-MNL model, which we apply for both population- and individual-level estimates. For each, we report the estimates and the R-squared.

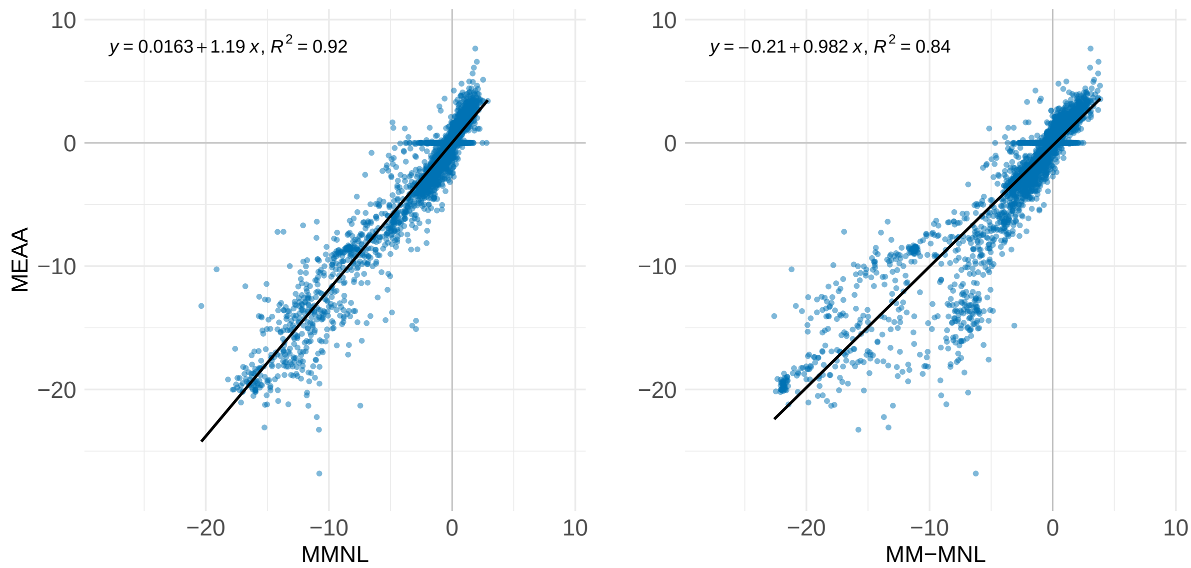


Figure B1: Individual-level estimates: Smoothies

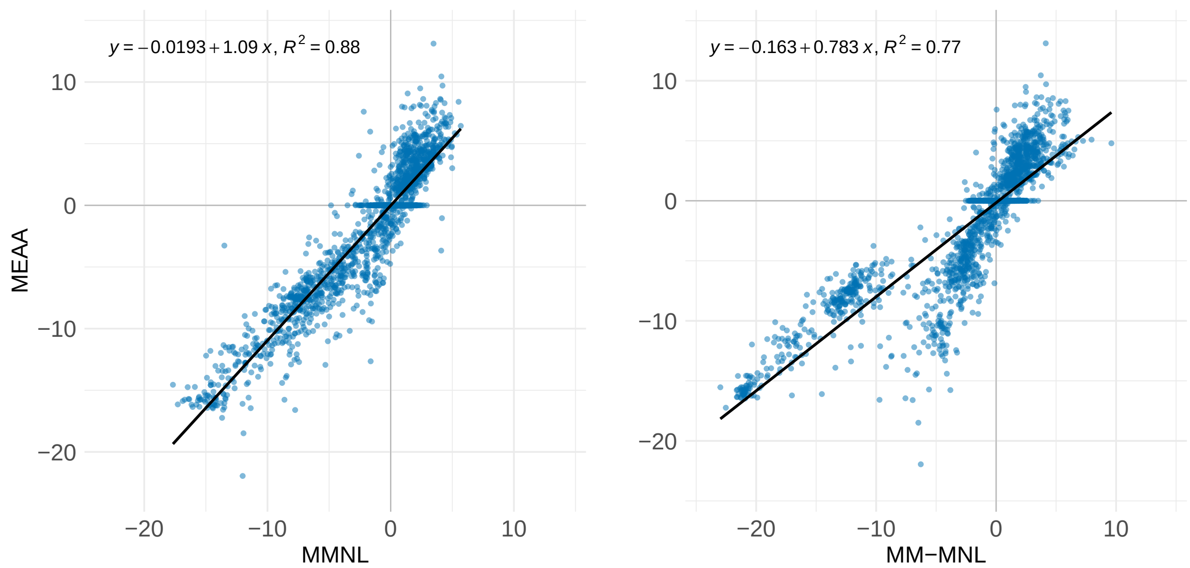


Figure B2: Individual-level estimates: Orange juice

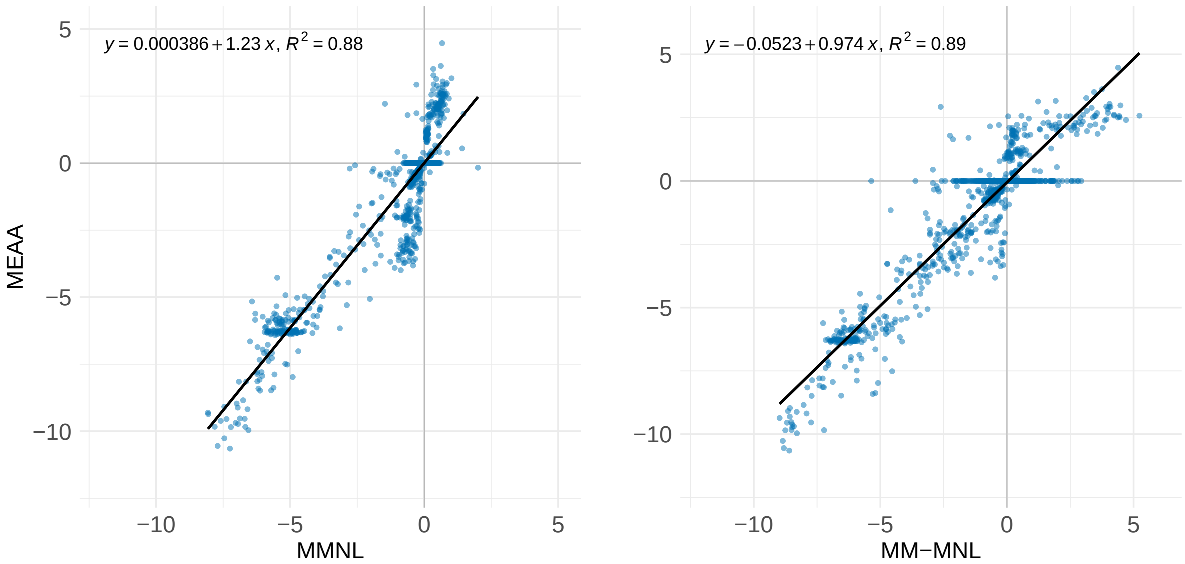


Figure B3: Individual-level estimates: Video-streaming services

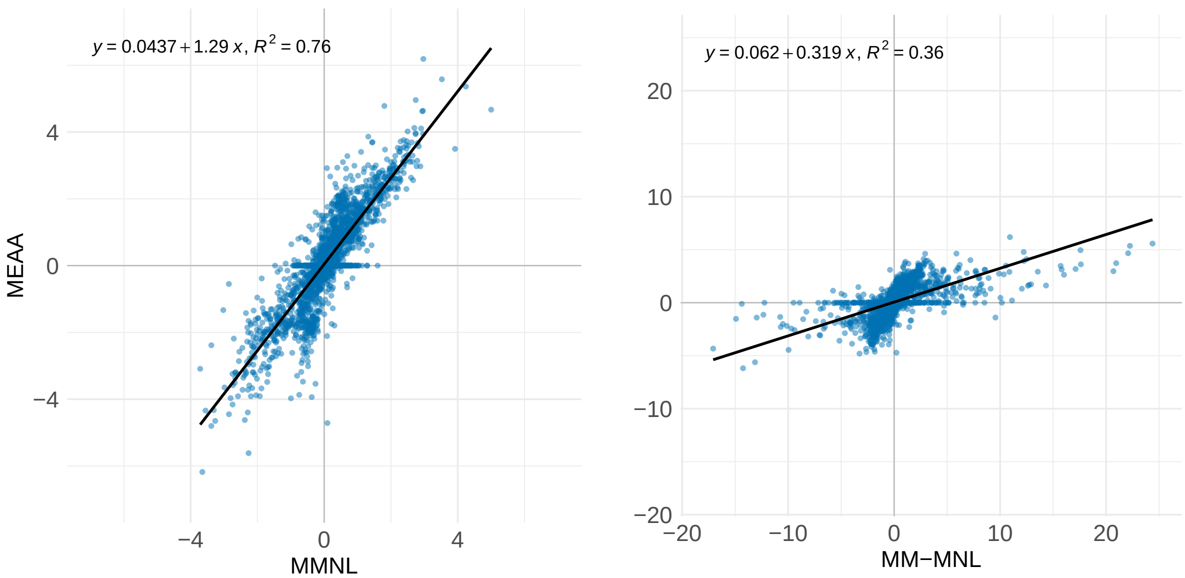


Figure B4: Individual-level estimates: Parties

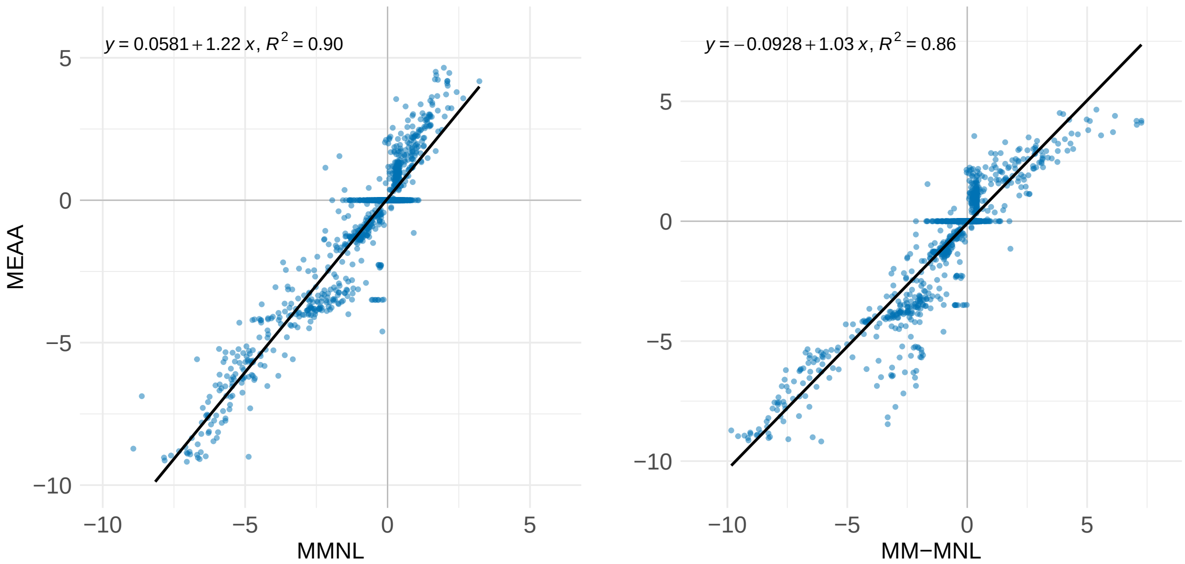


Figure B5: Individual-level estimates: Electric kettles

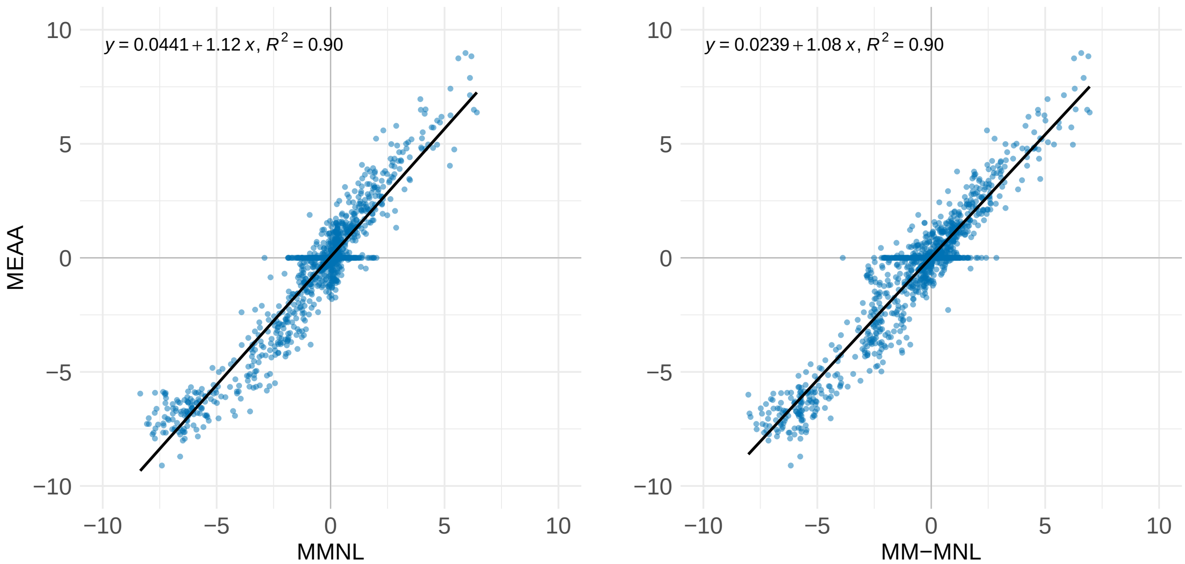


Figure B6: Individual-level estimates: Basketball tickets

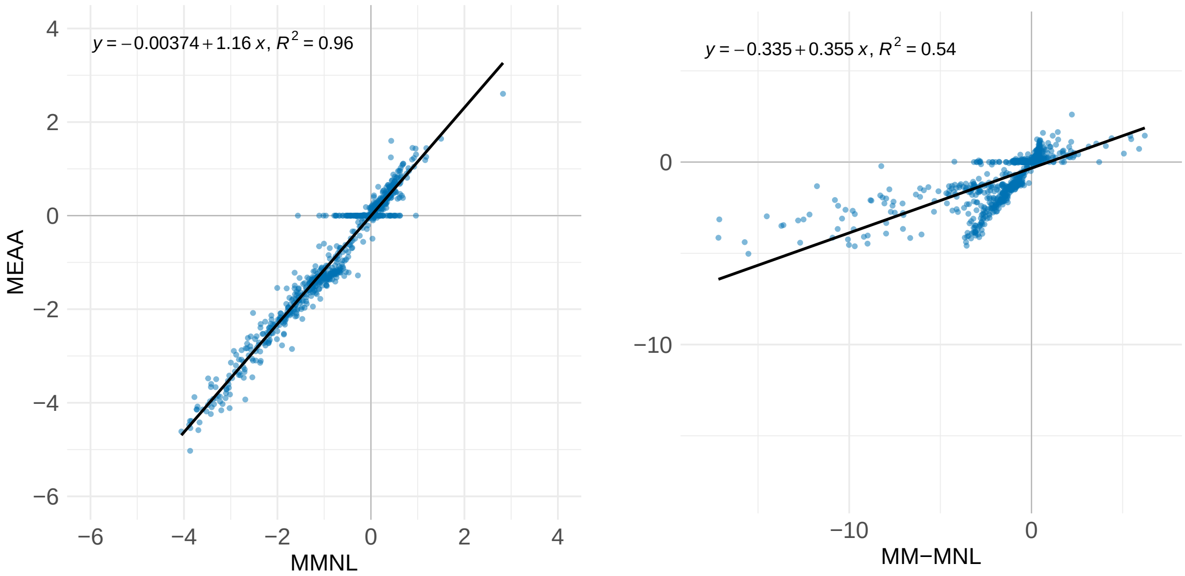


Figure B7: Individual-level estimates: Laptops

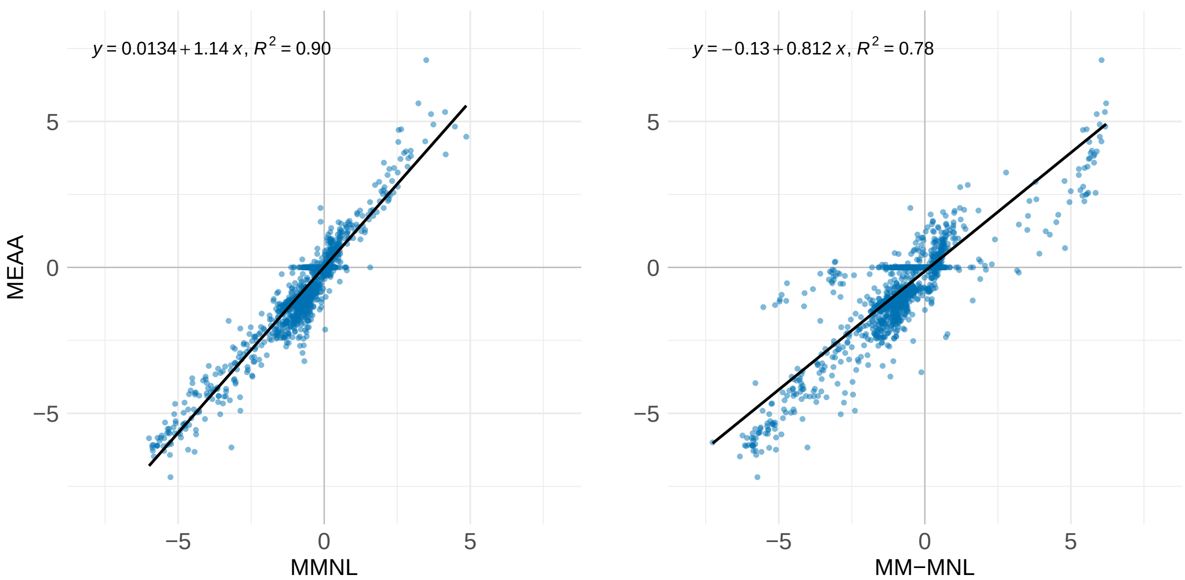


Figure B8: Individual-level estimates: Tablets

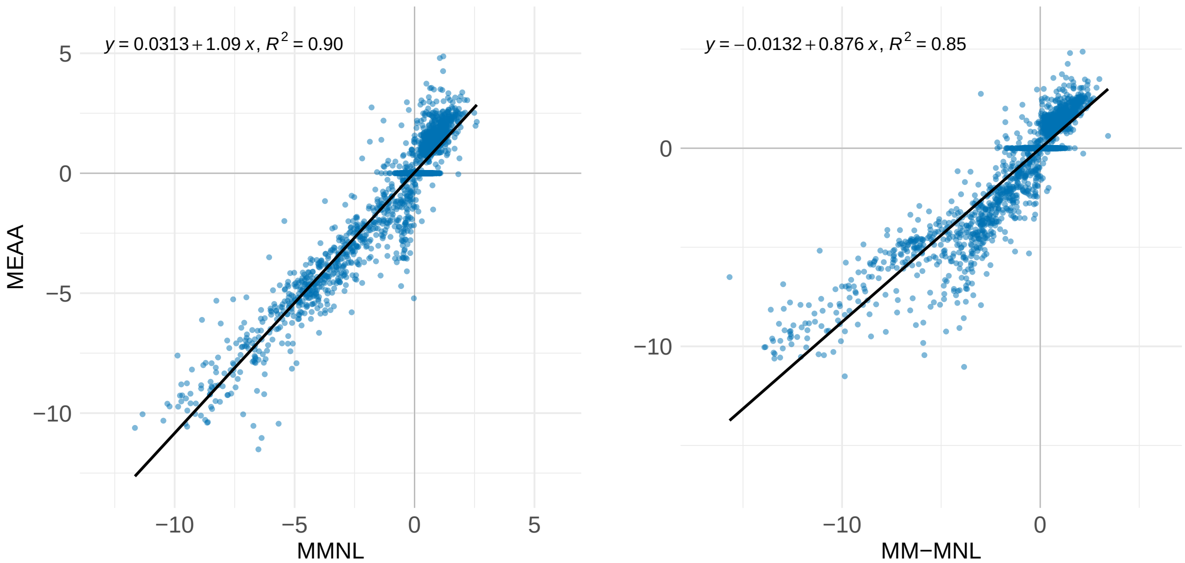


Figure B9: Individual-level estimates: Cameras

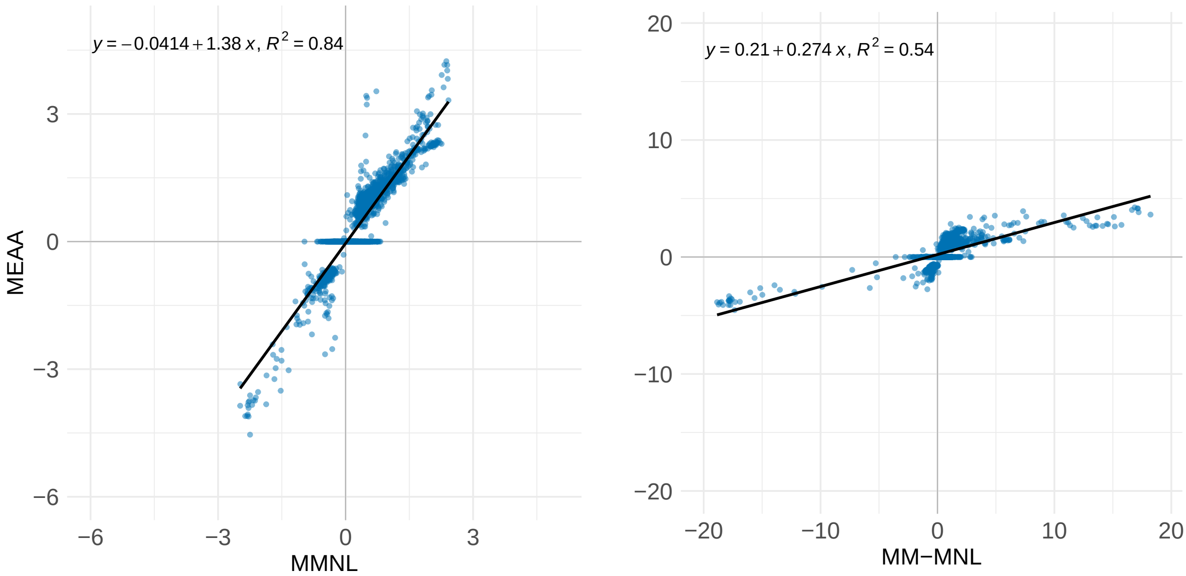


Figure B10: Individual-level estimates: Holiday destinations

Appendix C

In the following Appendix, we present scatterplots with the population-level estimates of the parameters for the MEAA and the MMNL models against each other. The MMNL estimates are rescaled using the slope parameters in Appendix B. The left panel plots the estimate of the mean, and the right panel – the estimates of the standard deviation.

The colors of the dots represent where the zero lies in the MEAA model according to our classification: dark blue – within the $(\beta-\sigma, \beta+\sigma)$ interval, light blue – outside the $(\beta-\sigma, \beta+\sigma)$ interval.

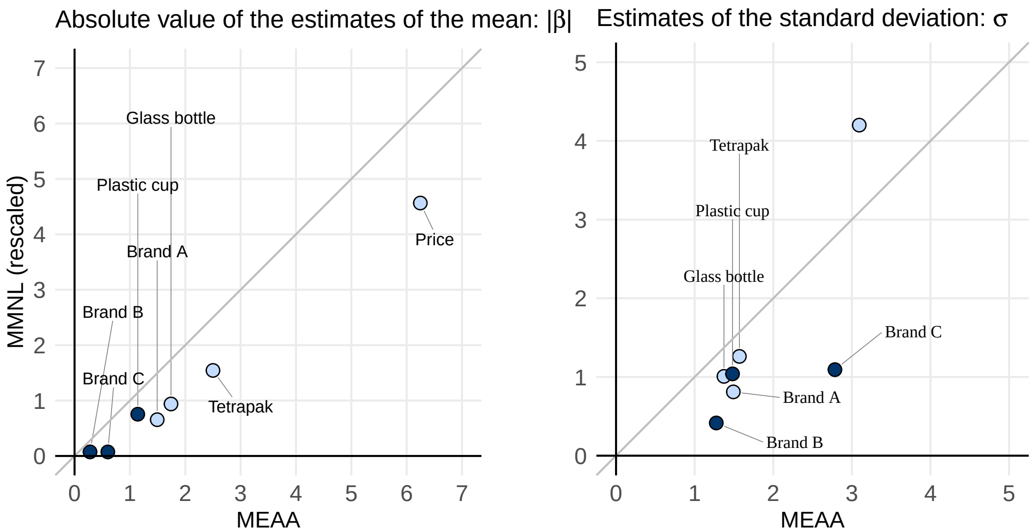


Figure C1: Population-level estimates: Smoothies

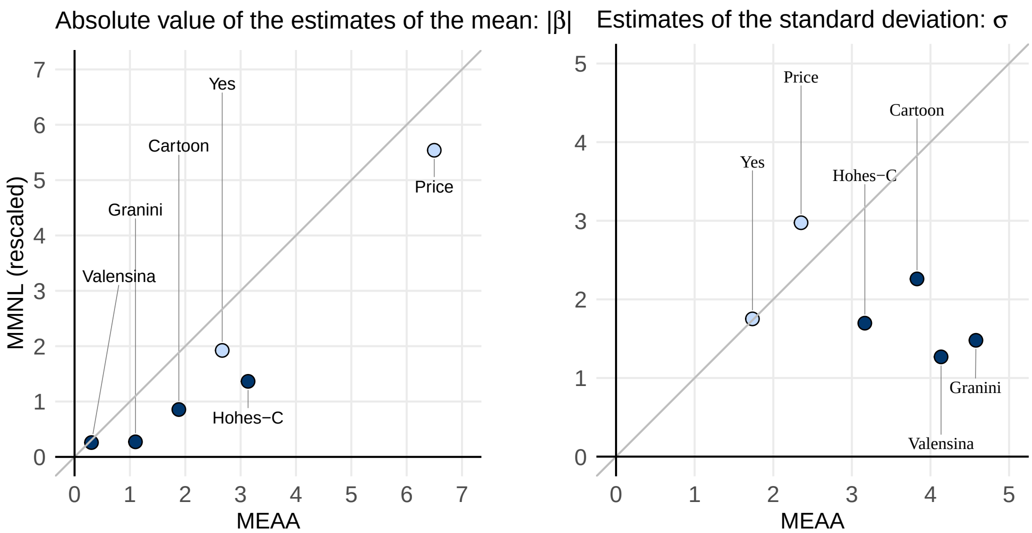


Figure C2: Population-level estimates: Orange juice

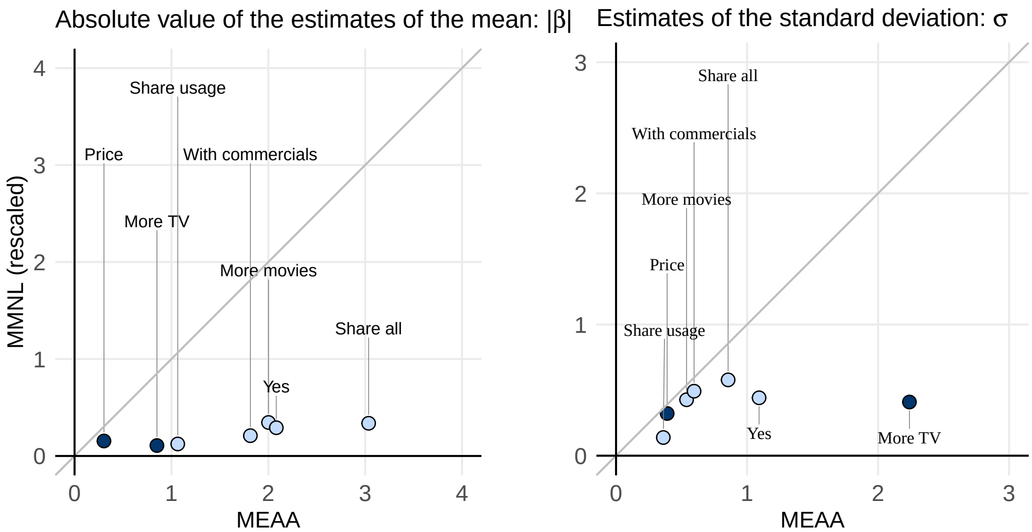


Figure C3: Population-level estimates: Video-streaming services

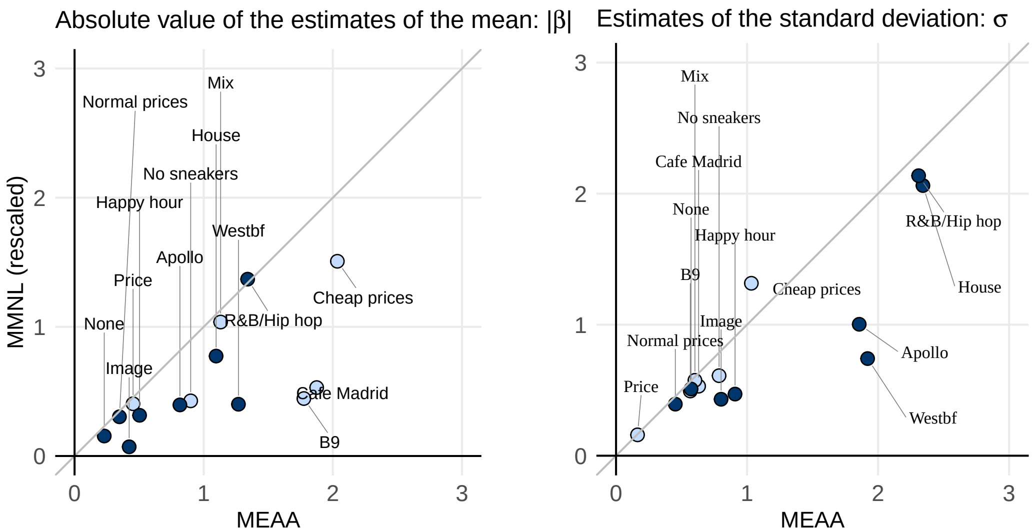


Figure C4: Population-level estimates: Parties

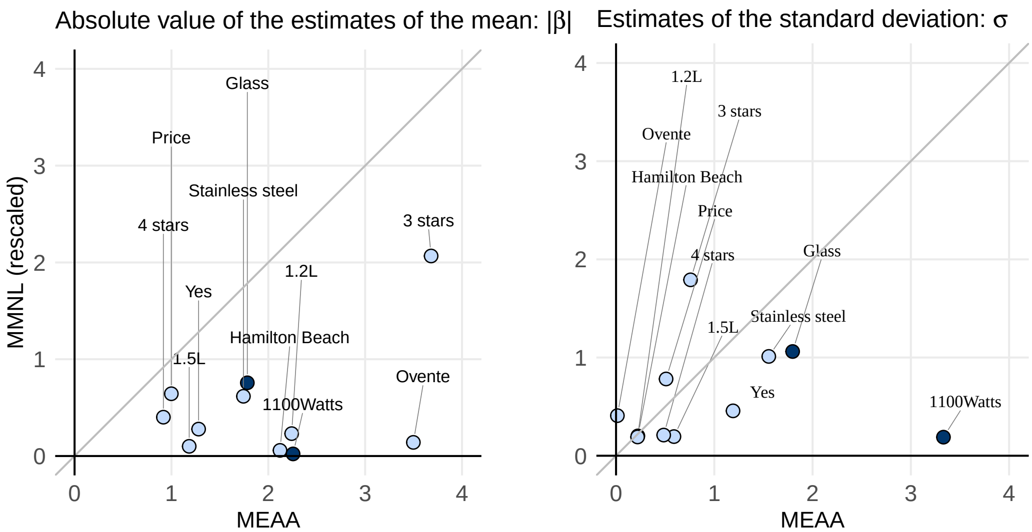


Figure C5: Population-level estimates: Electric kettles

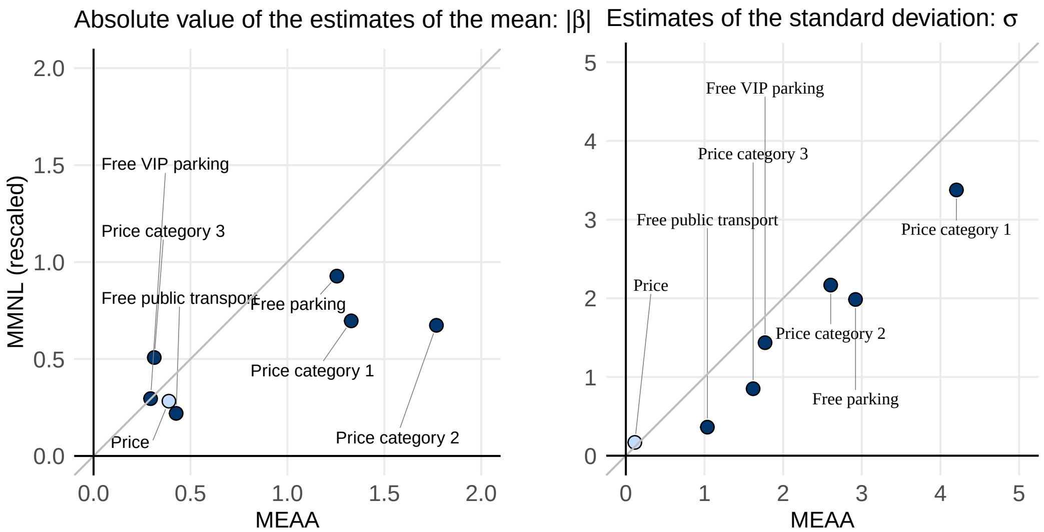


Figure C6: Population-level estimates: Basketball tickets

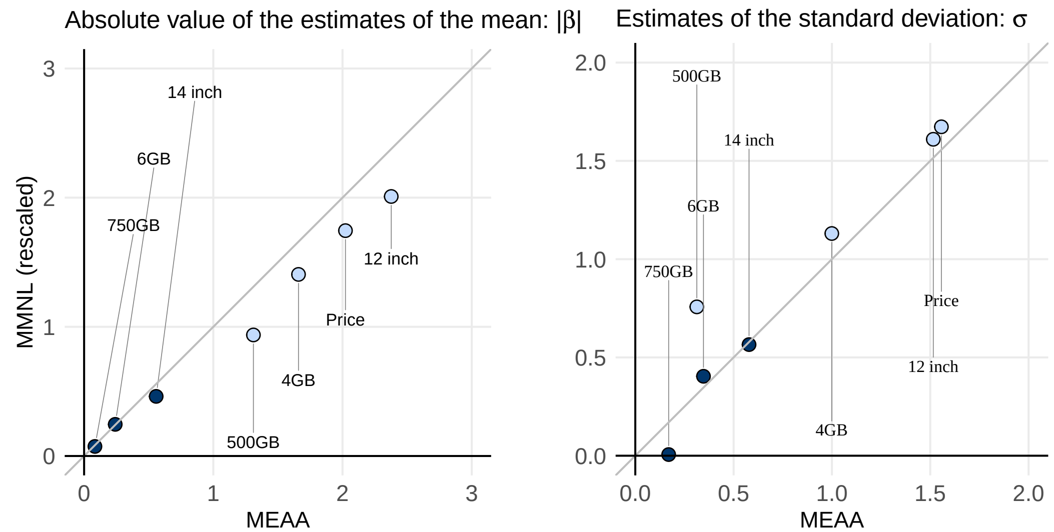


Figure C7: Population-level estimates: Laptops

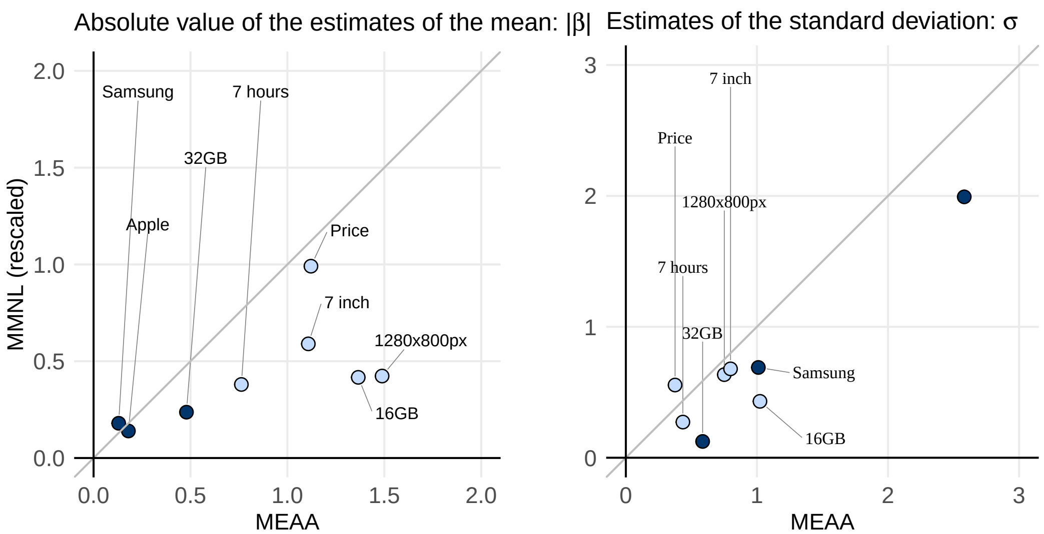


Figure C8: Population-level estimates: Tablets

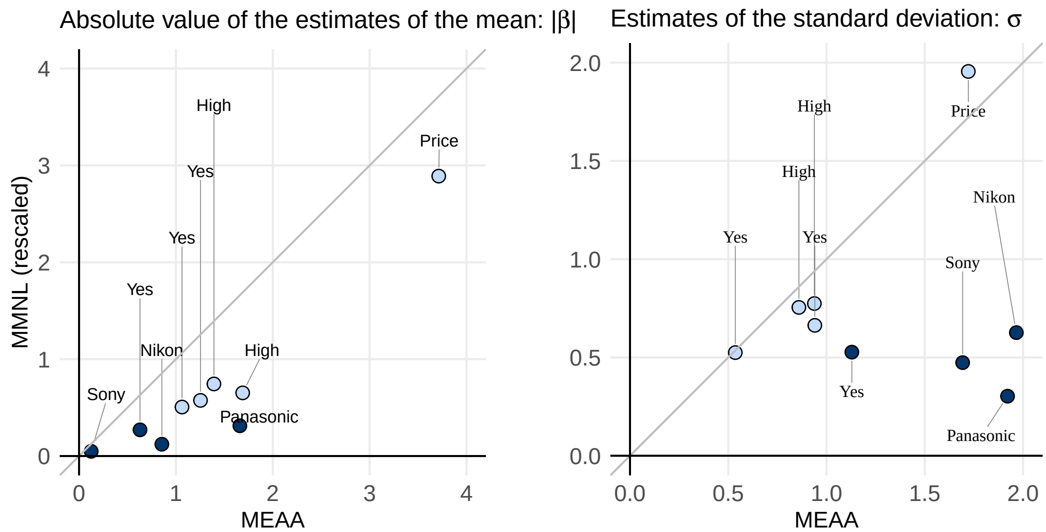


Figure C9: Population-level estimates: Cameras

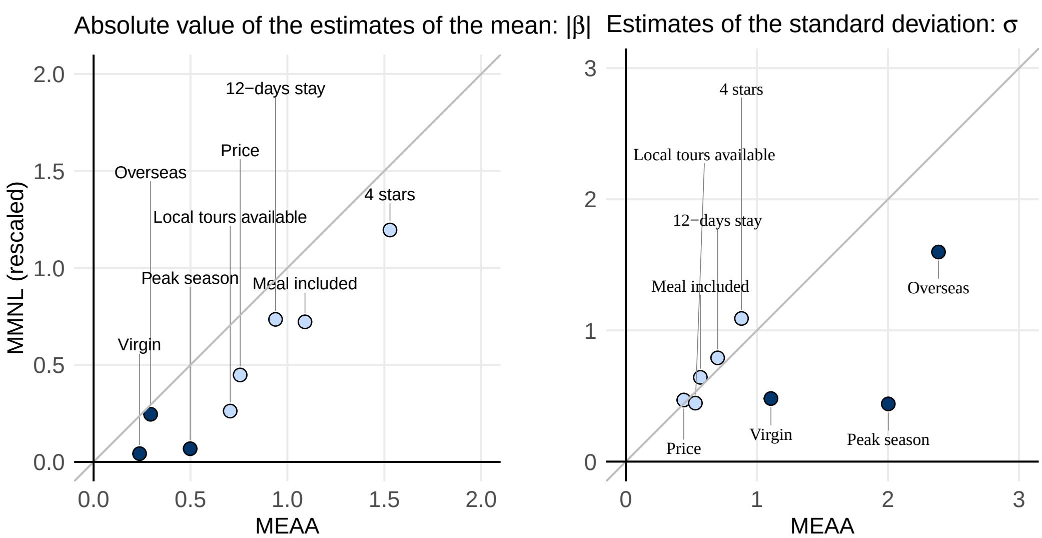


Figure C10: Population-level estimates: Holiday destinations

References

Allenby, G. M., Arora, N., & Ginter, J. L. (1998). On the heterogeneity of demand. *Journal of Marketing Research*, *35*(3), 384–389.

Allenby, G. M., Brazell, J. D., Howell, J. R., & Rossi, P. E. (2014). Economic valuation of product features. *Quantitative Marketing and Economics*, *12*(4), 421–456.

Allenby, G. M., & Ginter, J. L. (1995). Using extremes to design products and segment markets. *Journal of Marketing Research*, *32*(4), 392–403.

Allenby, G. M., & Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of Econometrics*, *89*(1), 57–78.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, *25*(3), 187–217.

Burda, M., Harding, M., & Hausman, J. (2008). A bayesian mixed logit – probit model for multinomial choice. *Journal of Econometrics*, *147*(2), 232–246.

Collins, A. T. (2012). *Attribute nonattendance in discrete choice models: measurement of bias, and a model for the inference of both nonattendance and taste heterogeneity* (PhD dissertation). Institute of Transportation and Logistics Studies, University of Sydney Business School.

Collins, A. T., Rose, J. M., & Hensher, D. A. (2013). Specification issues in a generalised random parameters attribute nonattendance model. *Transport Research Part B: Methodological*, *56*, 234–253.

Dellaert, B. G., Donkers, B., & Van Soest, A. (2012). Complexity effects in choice experiment-based models. *Journal of Marketing Research*, *49*(3), 424–434. doi: 10.1509/jmr.09.0315

Desarbo, W. S., Ramaswamy, V., & Cohen, S. H. (1995). Market segmentation with choice-based conjoint analysis. *Marketing Letters*, *6*(2), 137–147.

Elshiewy, O., Guhl, D., & Boztuğ, Y. (2017). Multinomial logit models in marketing – from fundamentals to state-of-the-art. *Marketing ZFP*, *39*(3), 32–49.

Frischknecht, B. D., Eckert, C., Geweke, J., & Louviere, J. J. (2014). A simple method for estimating preference parameters for individuals. *International Journal of Research in*

Marketing, *31*(1), 35–48.

Gilbride, T. J., Allenby, G. M., & Brazell, J. D. (2006). Models for heterogenous variable selection. *Journal of Marketing Research*, *43*(3), 420–430.

Gilbride, T. J., & Lenk, P. J. (2010). Posterior predictive model checking: An application to multivariate normal heterogeneity. *Journal of Marketing Research*, *47*(5), 896–909.

Glasgow, G., & Butler, S. (2017). The value of non-personally identifiable information to consumers of online services: evidence from a discrete choice experiment. *Applied Economics Letters*, *24*(6), 392–395.

Hess, S., Stathopoulos, A., Campbell, D., & Caussade, S. (2013). It's not that I don't care, I just don't care very much: Confounding between attribute non-attendance and taste heterogeneity. *Transportation*, *40*(3), 583–607.

Hole, A. R. (2011). A discrete choice model with endogenous attribute attendance. *Economic Letters*, *110*(3), 203–205.

Hole, A. R., Kolstad, J. R., & 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.

Huber, J., & Train, K. (2001). On the similarity of classical and bayesian estimates of individual mean partworths. *Marketing Letters*, *12*(3), 259–269.

Jain, K., & Srinivasan, N. (1990). An empirical assessment of multiple operationalizations of involvement. In M. Goldberg, G. Gorn, & R. Pollay (Eds.), *Advances in consumer research* (Vol. 17, pp. 594–602). Provo, UT: Association for Consumer Research.

Kamakura, W. A., Kim, B.-D., & Lee, J. (1996). Modeling preference and structural heterogeneity in consumer choice. *Marketing Science*, *15*(2), 152–172. doi: 10.1287/mksc.15.2.152

Kamakura, W. A., & Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, *26*(4), 379–390. doi: 10.1177/002224378902600401

Kamakura, W. A., Wedel, M., & Agrawal, J. (1994). Concomitant variable latent class models for conjoint analysis. *International Journal of Research in Marketing*, *11*(5), 451–464.

Keane, M., & Wasi, N. (2013). Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics*, *28*(6), 1018–1045.

Laurent, G., & Kapferer, J.-N. (1985). Measuring consumer involvement profiles. *Journal of Marketing Research*, *22*(1), 41–53.

Liu, Q., & Tang, Y. E. (2015). Construction of heterogeneous conjoint choice designs: A new approach. *Marketing Science*, *34*(3), 346–366.

Maldonado, S., Montoya, R., & López, J. (2017). Embedded heterogeneous feature selection for conjoint analysis: A svm approach using l1 penalty. *Applied Intelligence*, *46*, 775–787.

Maldonado, S., Montoya, R., & Weber, R. (2015). Advanced conjoint analysis using feature selection via support vector machines. *European Journal of Operational Research*, *241*(2), 564–574.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, *3*(4), 303–328. doi: 10.1016/0047-2727(74)90003-6

McFadden, D., & Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, *15*(5), 447-470.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

McQuarrie, E. F., & Michael, M. J. (1986). The zaichkowsky personal involvement inventory: Modification and extension. In P. Anderson & M. Wallendorf (Eds.), *Advances in consumer research* (Vol. 14, pp. 36–40). Provo, UT: Association for Consumer Research.

Orquin, J. L., Bagger, M. P., & Loose, S. M. (2013). Learning affects top down and bottom up modulation of eye movements in decision making. *Judgment and Decision Making*, *8*(6), 700–716.

Paetz, F., & Guhl, D. (2017). Understanding differences in segment-specific willingness-to-pay for the fair trade label. *Marketing ZFP - Journal of Research and Management*, *39*(4), 37–46.

Paetz, F., & Steiner, W. (2017). The benefits of incorporating utility dependencies in finite mixture probit models. *OR Spectrum*, *39*(3), 793–819.

Payne, J. W., Bettman, J. R., Coupey, E., & Johnson, E. J. (1992). A constructive process view of decision making: Multiple strategies in judgment and choice. *Acta Psychologica*, *80*(1), 107–141.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.

Rao, V. R. (2014). *Applied conjoint analysis*. Berlin, Germany: Springer.

Rossi, P., Allenby, G., & McCulloch, R. (2005). *Bayesian statistics and marketing*. Hoboken, NJ: Wiley.

Scarpa, R., Gilbride, T. J., Campbell, D., & Hensher, D. A. (2009). Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics*, *36*(2), 151–174.

Schlereth, C., & Skiera, B. (2017). Two new features in discrete choice experiments to improve willingness-to-pay estimation that result in sdr and sadr: Separated (adaptive) dual response. *Management Science*, *63*(3), 829–842.

Software, S. (2017). *The cbc system for choice-based conjoint analysis (version 9)* (Tech. Rep. No. Version 9). Orem, Utah: Sawtooth Software, Inc.

Swait, J., & Ben-Akiva, M. (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*, *21*(2), 91–102.

Train, K. E. (2009). *Discrete choice methods with simulation* (2nd ed.). Cambridge, UK: Cambridge University Press.

Voleti, S., Srinivasan, V., & Ghosh, P. (2017). An approach to improve the predictive power of choice-based conjoint analysis. *International Journal of Research in Marketing*, *34*(2), 325–335.

Yegoryan, N., Guhl, D., & Klapper, D. (2020). Inferring attribute non-attendance using eye tracking in choice-based conjoint analysis. *Journal of Business Research*, *111*, 290-304. doi: https://doi.org/10.1016/j.jbusres.2019.01.061