Volumetric demand and set size variation

by

Nino Hardt
Fisher College of Business
Ohio State University
hardt.8@osu.edu

Peter Kurz
bms marketing research + strategy
p.kurz@bms-net.de

February 24, 2021
click here for latest version

Abstract

To improve product planning, supply chain, and pricing decisions for packaged goods, brand managers must understand drivers of both overall category ('primary') brand-specific ('secondary') demand. A key reason for this is that their decision space can involve adding (or removing) not just one but several products to their current assortment. That is, they must consider not only composition, but assortment size. Although state-of-the-art Multiple Discrete Continuous Models (MDCM) can explain simultaneous demand for multiple varieties of products, they can unwittingly encode assumptions that hinder accurate demand forecasting across assortments of different sizes. Whereas classic MDCMs impose (in the absence of binding budget constraints) a monotonically increasing relationship between category demand and assortment size, empirical and behavioral research suggests that smaller assortments can often yield equal or higher sales; this in turn suggests a potentially positive effect of set size on the baseline utility of the outside good.

To that end, we develop a new model that retains the brand-level fidelity of MDCMs but enables a flexible relationship between assortment size and primary demand. Two large-scale choice experiments in disparate categories (chocolate bars and air fresheners) demonstrate the proposed model's ability to predict demand for market-like scenarios, while analogous MDCMs over-predict primary demand by 40%-80%. Moreover, the proposed model is computationally tractable using standard Bayesian machinery, allowing scalable inference for real-world category management.

Keywords: Choice Models, Volumetric Demand, Choice-set size, Multiple Discrete Continuous Models

1 Introduction

Modeling *demand quantities* is important for decisions in packaged goods. It is insufficient to only explain 'secondary' demand (i.e., market shares) because consumers may buy multiple varieties at the same time (Walsh, 1995). Offering more distinct products has the potential to increase 'primary' demand (overall category sales). For example, the introduction of an entire new line of yogurts might grow overall yogurt sales (i.e., increase primary demand). It is also possible that the new line primarily takes away sales volume from other offerings. The outcome depends on consumer budgets, preferences, satiation and the number of choice alternatives added. Yogurt manufacturers, other package goods companies or retail stores would like to conduct such policy simulations where multiple products are added to (or removed from) an assortment (i.e., a choice set).

Managers frequently face decisions that involve the addition or removal of several choice alternatives. It has been shown that competition from Costco or similar wholesale clubs may result in reduced assortment sizes in retail stores (Bauner and Wang, 2019). Manufacturer also need to decide how many varieties to offer. In this paper we assume that all products of an assortment are considered as viable choice alternatives, and therefore we do make a distinction between assortment size and choice-set size. Both refer to the number of alternatives that a consumer can choose from.

Extant demand models are unable to deal with choice-set size variation. In industry, it is common to use a 'quantity-then-choice' approach, where the quantity is independent of both composition and size of the choice set. This is similar in latent consumption occasion models (Dubé, 2004), where primary demand is largely unaffected by choice-set size. We will show that volumetric demand models, in particular models of multiple discreteness (MDC) (Dubé, 2019; Allenby et al., 2019), are over-identified and biased towards overestimating demand when assortment size is increased. Dealing with addition and removal from choice sets is relevant for practitioners (Eagle, 2018), but there is an apparent lack of clarity about the implications of choice-set size variation. This paper fills this void in the choice modeling literature and provides researchers with a tool for demand modeling when choice-set size varies.

Behavioral research suggests that set size variation influences decision-making processes (Meissner et al., 2019) and purchase probabilities (Scheibehenne et al., 2010). It has been suggested that larger assortments may result in less purchases (Chernev et al., 2015). For instance, Dhar (1997) find that larger assortments may lead to deferral of choice. Schwartz (2016) suggests that 'more is less'. Several explanations have been discussed in the literature, including changes in expectation regarding the inside goods (Diehl and Poynor, 2010), differences in motivation (Iyengar and Lepper, 2000) or search costs (Kuksov and Villas-Boas, 2010). Feinberg et al. (2016) find that participants in their study are less satisfied with choices from larger choice sets. Studying more versions of a choice experiment with a wider range of choice alternatives, Herzenstein et al. (2019) find an inverted U-shaped relationship between the number of choice alternatives and total allocation to inside options from a given budget. Previous research thus suggests the existence of (at least) two competing effects from increases in choice-set size: 1) A positive effect on demand based on providing more 'variety' and 2) a negative effect on demand. Many possible explanations for a negative effect on demand (or, conversely, a positive effect on demand for 'outside good') have been suggested.

Previous empirical studies of category reductions have shown mixed results, i.e. decreases and increases in sales in response to assortment size reductions (Boatwright and Nunes, 2001; Sloot et al., 2006; Borle et al., 2005). These mixed outcomes cannot be explained using extant volumetric demand models.

While extant volumetric demand models based on economic theory (Kim et al., 2002; Dubé, 2019; Allenby et al., 2017) are able to explain and predict demand quantities, they are not designed to handle set size variation. Once set size variation is present, their parameters are over-identified, and they are unable to explain increasing demand for outside good in response to set size increases.

To understand why, we need to review how these multiple discrete-continuous models (MDCMs) work. They allow for ‘corner solutions’ (products that are not purchased) and multiple ‘interior solutions’ (products that are bought, where the purchase quantity is continuous). Multiple interior solutions are possible because it assumed that there are diminishing marginal utilities to all goods. Consuming only the good with the highest baseline marginal utility might not be utility-optimal, since the marginal increase in utility decreases. Instead, consumers may choose to buy several different goods at the same time. When several additional product varieties are added to the choice-set, more products can be purchased at a given marginal rate of utility, resulting in increased overall demand for inside goods. It is therefore built into these multiple discrete-continuous models that primary demand is monotonically increasing with choice-set size. The relationship between set size and primary demand is governed by parameters already identified in the absence of set size variation. This means that these models are over-identified when choice-set size varies. These models are unable to explain negative effects of assortment size on primary demand.

We propose a parametric approach to incorporating choice-set size into volumetric demand models. The parameterization captures negative effects of choice-set size on inside good utilities (or positive effects on outside good utility). Apart from the behavioral reasoning for such effects, we believe there is a more simple reason that motivates a ‘negative’ set size effect. The maximum attainable utility from a product category increases when additional choice alternatives are added. In other words, the category is more valuable in terms of utils. Money that is allocated towards the category is thus also more valuable. Volumetric demand models define money that is dedicated to a category as the ‘budget’. This implies that unspent budget (outside good) is also more valuable. In this paper, we remain agnostic with respect to the true process behind negative set size effects. Our parameterization can only described the sum of positive effects of set size on outside good utility. We believe our approach would be helpful in further studying set size effects, but an integration of the vast behavioral literature and economic choice modeling literature is beyond the scope of this paper. Our model is also helpful for applying data fusion with volumetric demand models, because set sizes usually vary significantly between data from choice experiments and transaction histories. Failing to account for set size variation is likely going to result in poor-fitting models or worse, coming up with speculative explanations for discrepancies between the data sources involved.

We demonstrate the performance of our model using two empirical applications: Volumetric conjoint studies of chocolate bars in Germany and non-electric air fresheners in the US. We can show that parameters governing secondary demand ('part-worths') are largely unaffected by set size variation, however, ignoring set size variation leads to dramatic over-prediction of primary demand. Extant models are off by as much as 80% in our first study and 40% in our second study. We validate our model in different ways: Using self-reported 'last purchase' quantities and aggregate demand data (first study) and a virtual shelf task (second study).

The remainder of this paper is organized as follows: In the next section we introduce the economic model and statistical specification. Then, we describe our two empirical applications and results. We then discuss findings and illustrate consequences for policy simulations. The paper concludes with a discussion of implications and applications.

2 Model Development

In this Section, we first review an extant volumetric demand model and how it is over-identified in the presence of set size variation. We then develop our proposed model which incorporates a parameterization of set size. We describe the estimation procedure and how to generate predictions. Finally, we review related challenges for volumetric demand models.

2.1 Volumetric Demand Model

A common approach to modeling volumetric demand are Multiple Discrete Continuous Models (MDCMs). For an overview and the economic background of these models, see Dubé (2019). MDCMs can explain simultaneous demand for multiple distinct products. For each interior solution (i.e. each product that is bought), demand quantities are assumed to be continuous. This allows developing models based on the Karush-Kuhn-Tucker conditions which helps simplify estimation, particularly in higher dimensions. This assumption is unproblematic in many situations (Lee and Allenby, 2014), especially when bundling does not play a major role in category.

Following a simple direct utility specification with non-linear inside and outside goods, we assume that decision makers maximize utility $u(\mathbf{x}, z)$ subject to a budget constraint (Allenby et al., 2017). The utility maximization problem for a single choice occasion (i.e., a single choice task or shopping trip) can be expressed as:

$$\text{Max } u(\mathbf{x}, z) = \sum_{j=1}^{N} \frac{\psi_j}{\gamma} \ln(\gamma x_j + 1) + \psi_z \ln(z) \quad \text{s.t.} \quad \mathbf{p}'\mathbf{x} + z \leq E \tag{1}$$

Here, x_j is the purchased quantity of good j, and ψ_j represents the baseline preference for that good j. The rate of satiation of inside goods is controlled by γ, and p_j is the price of a unit of good j. The outside good z represents unspent money that the decision maker has been willing to allocate towards the focal category, but eventually did not end up spending on inside goods available in the choice set. We assume that there are diminishing returns to unspent money, and therefore use a nonlinear specification of z. This allows estimating the budgetary allotment E, which is identified through the functional form of the utility function.

Baseline marginal utility of good j is defined as follows, assuming multiplicative, independent error terms for each of the inside goods:

$$\psi_j = \exp\left(\mathbf{a}_j \beta + \varepsilon_j\right) \tag{2}$$

where β is the vector of 'part-worths' and $\mathbf{a}_j$ is the design vector for alternative j, and ε_j is a random term. $\mathbf{a_j}$ can be specified using dummy coding, in which case the first element of β, β_0, serves as an intercept capturing the baseline marginal utility of an inside good vs the outside good. Alternatively, effects coding can be used.

The corresponding likelihood function can be developed by exploiting the Karush-Kuhn-Tucker (KKT) conditions. The KKT conditions are derived by first forming the auxiliary function using Lagrangian multipliers:

$$\text{Max } L = u(\mathbf{x}_t, z_t,) + \lambda \left\{ E - \sum_{j=1}^{J} p_{jt} x_{jt} - z_t \right\} \tag{3}$$

Associating first-order conditions with observed demand yields:

$$u_{jt} = p_{jt} \cdot u_{zt} \qquad \text{if} \qquad x_{jt} > 0 \tag{4}$$

$$u_{jt} < p_{jt} \cdot u_{zt} \qquad \text{if} \qquad x_{jt} = 0 \tag{5}$$

where

$$u_j = \frac{\partial u\left(\mathbf{x}_t, z\right)}{\partial \mathbf{x}_j} = \frac{\exp\left(\mathbf{a}_j \beta + \varepsilon_j\right)}{\gamma x_j + 1} \tag{6}$$

$$u_z = \frac{\partial u\left(\mathbf{x}, z\right)}{\partial z} = \frac{\psi_z}{z} \tag{7}$$

Taking logarithms of 4 and 5, we have:

$$\varepsilon_j = g_j \qquad \text{if} \qquad x_j > 0 \tag{8}$$

$$\varepsilon_j < g_j \qquad \text{if} \qquad x_j = 0 \tag{9}$$

where

$$g_j = -\mathbf{a_j}\beta + \ln(p_j) + \ln(\gamma x_j + 1) - \ln(z) + \ln(\psi_z) \tag{10}$$

We assume that error terms are distributed *i.i.d.* Extreme Value:

$$\varepsilon \sim \mathrm{EV}(0, \sigma)$$

For the purpose of identification, and without loss of generality, it is common to constrain $\psi_z = 1$, which reduces the dimensionality of the system of equations defined by the KKT conditions by one. This approach, along with other consideration of identification, is described in more detail by Howell et al. (2016).

The likelihood of the model parameters is proportional to the probability of observing demand for R of N goods, for which there exists a closed-form expression:

$$\begin{aligned}
\Pr(\mathbf{x}) = \quad & \Pr(x_{n_1} > 0, \quad x_{n_2} = 0, \quad n_1 = 1, \ldots, R, \quad n_{2,t} = R+1, \ldots, N) \\
= \quad & |J_R| \int_{-\infty}^{g_N} \cdots \int_{-\infty}^{g_{R+1}} f(g_1, \ldots, g_R,\, \varepsilon_{R+1}, \ldots, \varepsilon_N)\, d\varepsilon_{R+1}, \ldots, d\varepsilon_N \\
= \quad & |J_R| \left\{ \prod_{j=1}^{R} \frac{\exp(-g_j/\sigma)}{\sigma} \right\} \exp\left\{ -\sum_{i=1}^{N} \exp(-g_i/\sigma) \right\}
\end{aligned} \tag{11}$$

Contributions to the likelihood take the form of density and mass contributions, corresponding to the respective KKT condition. Transforming from random-utility error (ε) to the likelihood of the observed data (x), we need to consider the Jacobian $|J_R|$:

$$|J_R| = \prod_{j=1}^{R} \left(\frac{\gamma}{\gamma x_j + 1} \right) \left\{ \sum_{j=1}^{R} \frac{\gamma x_j + 1}{\gamma} \cdot \frac{p_j}{z} + 1 \right\} \tag{12}$$

The error scale parameter σ captures the price sensitivity of respondents.

The model yields estimates for E based on the assumption of nonlinear outside good utility and variation in inside good expenditures across tasks. While the lower bound for the budget estimate is given by the 'maximum ever spent' across all choice occasions, variation in expenditures among choice tasks contributes to estimates that a larger than the lower bound. This assumes that each all choice occasions are comparable, i.e. preferences, satiation and budget do not change between occasions.

Relative values of the 'part-worth' parameters β explain patterns of secondary demand, while the location or intercept of β (i.e., β_0), along with other parameters, relates to primary demand. Larger values of γ mean higher rates of satiation. This may reveal itself by demand patterns that are dispersed over a more distinct products. Everything else being equal, it would also lead to stronger increases in primary demand in response to set size increases.

2.2 Incorporating set size variation

The specification described in the previous Section (unwittingly) implies that, unless the budget constraint is binding, (1) primary demand is increasing in choice-set size and (2) the strength of this relationship is determined by parameters which are identified in the absence of set size variation. In other words, once set size variation introduced, the model is over-identified. While Ackerberg and Rysman (2005) describe over-identification in a discrete choice model with set size variation, we are not aware of a discussion of the consequences of over-identification for volumetric demand models.

Empirical and behavioral research suggests that the relationship between set size and primary demand can take on different directions and strengths. Major assortment reductions have been found to have positive, negative or no effect on overall sales (Boatwright and Nunes, 2001; Borle et al., 2005). In some cases, assortment reductions have resulted in (sometimes temporary) reductions in overall sales, in other cases overall sales remained largely unaffected. Therefore, we believe that over-identified MDCMs are potentially too restrictive if there is significant variation in set size. Our proposed model is able to describe such patterns in demand data by introducing an additional set size effect.

We re-write the maximization problem of Equation 1 to include a task-specific N_t:

$$\text{Max } u(\mathbf{x}, z) = \sum_{j=1}^{N_t} \frac{\psi_j}{\gamma} \ln(\gamma x_j + 1) + \psi_z \ln(z) \quad \text{s.t.} \quad \mathbf{p}'\mathbf{x} + z \leq E \tag{13}$$

Given variation in N_t, it is possible to identify a parameter in place of ψ_z if it is a function of N_t. Allowing the model to represent increasing appeal of the outside good (or decreasing returns from all inside goods) is consistent with previous empirical and behavioral research (constant or even increasing demand in response to assortment size reductions). While different combinations of β_0, γ, E can yield stronger relationships between set size and primary demand, the over-identified model is unable to represent weaker (or negative) relationships. This also means our parameterization of ψ_z will need to be constrained to only represent larger relative utility of the outside good.

A simple specification is shown in Equation 14, where outside good baseline marginal utility ψ_z is a function of N_t and a set size parameter ξ:

$$\psi_z = f(N_t; \xi) = \exp(0 + \ln(f(N_t; \xi))) \tag{14}$$
$$\text{where } \xi \geq 0$$

Here, ξ must be constrained to be positive, because stronger relationships between set size and primary demand can be represented by corresponding combinations of β_0, γ, E already. This constraint will ensure identification.

Depending on the amount of information available, $f(N_t; \xi)$ can be specified in more or less flexible ways, or even be estimated non-parametrically. Unless 'jumps' in demand are expected at specific set sizes, a first or second order polynomial parameterization should be flexible enough. However, higher-order specifications can be tested. This would be advisable if variance in demand and set size is expected to be large.

In our first empirical application, we only observe two discrete sizes of the choice set.

In this case, a simple linear specification can be fit:

$$f\left(N_t;\xi_1\right) = \xi_1 N_t + 1 \tag{15}$$

This parameterization implies that the marginal utility of the outside good is increasing in N_t. Relatively speaking, the utility of each inside good is decreasing in N_t. The resulting model nests the extant volumetric demand model from the previous section 2.1 when $\xi = 0$. Larger values of ξ mean that consumers do attach higher relative marginal utility to the outside good as N_t increases.

In our second empirical application, we observe three different choice-set sizes. In that case, we can also fit a 2nd order polynomial:

$$f\left(N_t;\xi = \{\xi_1,\xi_2\}\right) = \xi_1 N_t + \xi_2 N_t^2 + 1 \tag{16}$$

The appropriate order of the polynomial can be identified by comparing models based on the log-marginal likelihood.

While choice-set size variation has not been discussed in the volumetric choice literature, this has been studied for discrete choice models. Ackerberg and Rysman (2005) propose a discrete choice model that adjusts for variation in assortment size. They assume that the error term represents unobservable characteristics of choice alternatives. A nested Logit model thus implies that each additional product adds one additional dimension to the unobserved characteristic space. Our proposed model is different in that is explaining choice quantities instead of discrete choices and it is not motivated by 'crowding' in the error term, but the relative value of the outside good.

To demonstrate the effects of ignoring set size variation, we first re-arrange the model specification while keeping it likelihood-equivalent. Volumetric choices are determined only by relative utilities. Therefore, we can apply the adjustment to inside good marginal utilities, while maintaining the usual constraint of $\psi_z = 1$:

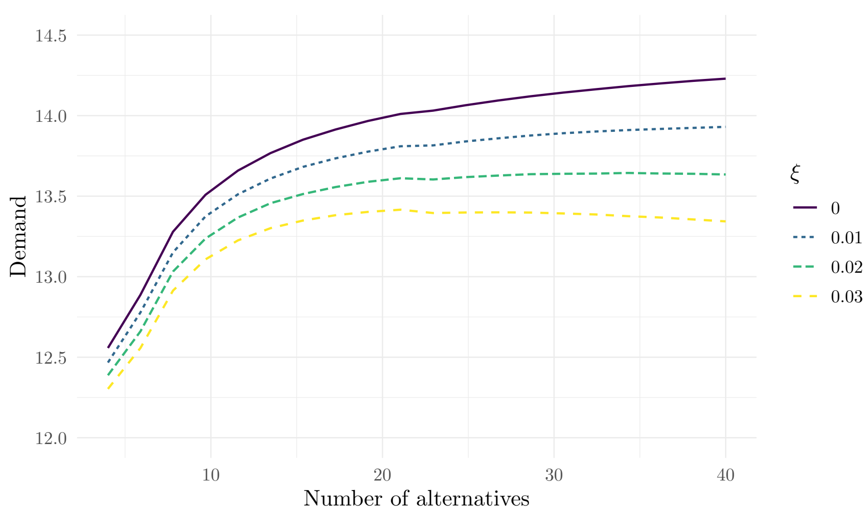
$$\psi_{kt} = \exp\left(\mathbf{a}_k\beta + \varepsilon - f\left(N;\xi\right)\right) \tag{17}$$

Here, we can see that the mean or location of the 'part-worth' vector (β) (or its intercept term β_0, if dummy coding is used) is a function of N_t if $\xi > 0$. Thus if the 'true' $\xi > 0$, the extant volumetric demand model will result in poor fit to data with choice-set size variation, because the location of β estimates will be inaccurate. This will also result in inaccurate predictions for counterfactual scenarios where N_t takes on different values. Based on this, we conjecture that for two (mostly) identical choice experiments, the one that involves larger choice-set sizes will result in estimates of β_{h0} that are smaller. We can also expect demand predictions for scenarios with more than k alternatives based on data with k choice alternative to be biased towards over-prediction.

Our proposed model specification (Equations 14, 15) may result in an inverted U-shape relationship between assortment size and primary demand. Initially, primary demand increases with increasing assortment size, because consumers are able to buy varieties that involve more unique products or more demand for inside goods overall. Here, larger values of ξ attenuate the relationship between set size and primary demand. As more alternatives are added to the set, and inflection point is reached and primary demand might actually decrease. Here larger values of ξ imply faster decreasing primary demand. Whether or not we end up observing decreasing demand in response to increasing assortment size depends on how many alternatives enter the consideration set. This shape is consistent with results described by Herzenstein et al. (2019), who find the same pattern for contributions to for-profit crowdfunding campaigns. It is also consistent with mixed results in empirical studies of retail assortments.

To further illustrate the influence of ξ on primary demand, we use a simple simulation exercise. To keep it simple, we use a linear specification of the set size adjustment. We compute expected demand for a single decision maker, varying N_t and ξ, while all choice alternatives have the same deterministic utility ($\mathbf{a}_j\beta = .25 \quad \forall j \in [1, ..., N_t]$) and prices ($\mathbf{p}_j = 1$), and we keep constant the rate of satiation ($\gamma = .75$), budget ($E = 15$), error scale ($\sigma = .25$) and prices ($\mathbf{p}_j = 1$). The resulting primary demand curves are shown in Figure 1. $\xi = 0$ corresponds to the simple volumetric demand model, while larger values of ξ show smaller increases in primary demand, or even decreasing primary demand. Comparing the different demand curves, we see that significant variation in the number of alternatives may be necessary to notice the relationship. An increase from 8 to 12 choice alternatives may only have a small impact on primary demand. However, once we consider demand in much larger assortments with 20 or more alternatives, there are considerable differences in predicted primary demand. Assuming identical baseline marginal utilities for all alternatives in this simulation, we can isolate the effect of variety on primary demand. We can see that larger values of ξ are able to offset the variety effect.

Figure 1: Number of alternatives, ξ, and primary demand (simulation)



In this paper we assume (for simplicity) that all choice alternatives available are also being considered, i.e. choice set size is controlled only by the vendor.

Exploring the processes behind the adjustment of marginal utility of the outside good is an interesting area for future research and is beyond the scope of this paper because we are not able to discern and isolate individual processes.

The model likelihood is straightforward to derive. Applying the parameterization of ψ_z to the utility function in Equation 13, we can rewrite it as follows:

$$u\left(\mathbf{x_t}, z_t\right) = \sum_{j=1}^{N_t} \frac{\psi_{jt}}{\gamma} \ln\left(\gamma x_{jt} + 1\right) + \left(\xi N_t + 1\right) \ln\left(z_t\right) \tag{18}$$

Here, the marginal utility of the outside good is:

$$u_{zt} = \frac{\partial u\left(\mathbf{x}_t, z_t\right)}{\partial z_t} = \frac{\xi N_t + 1}{z_t} \tag{19}$$

This leads to an additional term in the expression for g_{kt}, while the Jacobian remains unchanged. The likelihood otherwise be computed similar to equation 11, with mass and density contributions corresponding to the new function g_{kt}:

$$g_{kt} = -\mathbf{a}_{kt}\beta + \ln\left(\xi N_t + 1\right) + \ln\left(p_{kt}\right) + \ln\left(\gamma x_{kt} + 1\right) - \ln\left(z_t\right) \tag{20}$$

It is important to remember that ξ is only identified when there is variation in N_t. The source of variation in N_t can be experimental (e.g., in a choice experiment) or natural (when a store changes assortments over time in purchase transaction or similar 'revealed preference' data).

The proposed model is based on a relatively simple utility specification (i.e., Equation 18). However, the need for including an adjustment term to the outside good marginal utility is not specific to that particular utility specification. Generally, volumetric demand models require some non-linear utility specification. This is necessary to describe simultaneous demand for varieties. Moreover, identification constraints are required, resulting in over-identification issues once choice-set size variation is introduced.

Based on Equation 17, we would expect that omitting our proposed set size adjustment, there would be significant differences in the intercept of the 'part worth' vector β_h. Intercept terms are likely going to be smaller given larger choice-sets.

2.3 Heterogeneity and Estimation

$\theta_h = \{\beta_h, \ln\gamma_h, \ln E_h, \ln\sigma_h, \ln\xi_h\}$ is subject/respondent h's vector of parameters of length M governing the individual-level demand model. We assume a simple Normal model for heterogeneity. It is straightforward to replace it with a Multivariate Regression or Mixture of Normals model of heterogeneity if deemed appropriate.

$$\theta_h \sim \text{Normal}(\bar{\theta}, \Sigma) \tag{21}$$

Hyperparameters are assumed to be weakly informative, i.e. $\bar{\theta} \sim N(0, 100I_M)$, $\Sigma \sim \text{IW}(M+4, (M+4)I_M)$. Estimation of this model is easy using standard Monte Carlo Markov Chain methods (Rossi et al., 2005). The Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) is used for sampling θ_h, and Gibbs-sampling steps are used for sampling the remaining parameters.

2.4 Demand predictions

Demand ($\mathbf{x}_{ht}$) of subject h at time t is a function of parameters of the demand model (θ_h), a realization of the vector of error terms (ε_{ht}), characteristics of the available choice set ($\mathbf{A}_t = \{\mathbf{a}_{t1}...\mathbf{a}_{tJ}\}$), and prices ($\mathbf{p}_t$). We call the demand function D. Given the configuration of available products (i.e., $\mathbf{A_t}$ and corresponding prices $\mathbf{p}_t$) it returns utility-maximizing demand for one realization of model parameters θ_h and one realization of the error term ε_{ht}:

$$\mathbf{x}_{ht} = \text{D}\left(\theta_h, \varepsilon_{ht} | \mathbf{A}_t, \mathbf{p}_t\right) \tag{22}$$

There is no closed form solution for D. However, D can be computed using an iterative procedure that at worst takes R_t iterations (see Appendix A). Finally, expected demand is obtained by integrating out the error term and posterior distribution of model parameters θ_h.

$$E\left(\mathbf{x}_{ht}\right) = \int_{\theta_h}\int_{\varepsilon_{ht}} \text{D}\left(\theta_h, \varepsilon_{ht} | \mathbf{A}_t, \mathbf{p}_t\right) p(\varepsilon_{ht}) d\varepsilon_{ht} d\theta_h \tag{23}$$

Numeric integration is computationally cheap, because draws of θ have already been produced in the process of estimating the model, and D can easily be computed. Posterior distributions of demand can be produced for groups of consumers.

2.5 Related modeling challenge

We have argued that in the presence of choice-set size variation, volumetric demand models should include a set size parametrization . We believe that it is intuitive to account for the change in choice context induced by different choice set sizes and to overcome an overidentified model. The resulting inverted U-shaped relationship between assortment size and primary demand is consistent with previous findings. In this section, we are discussion other model extensions that have been discussed in the context of choice models. However, we believe that the set size model is orthogonal to all of these extensions.

2.5.1 Multiple constraints

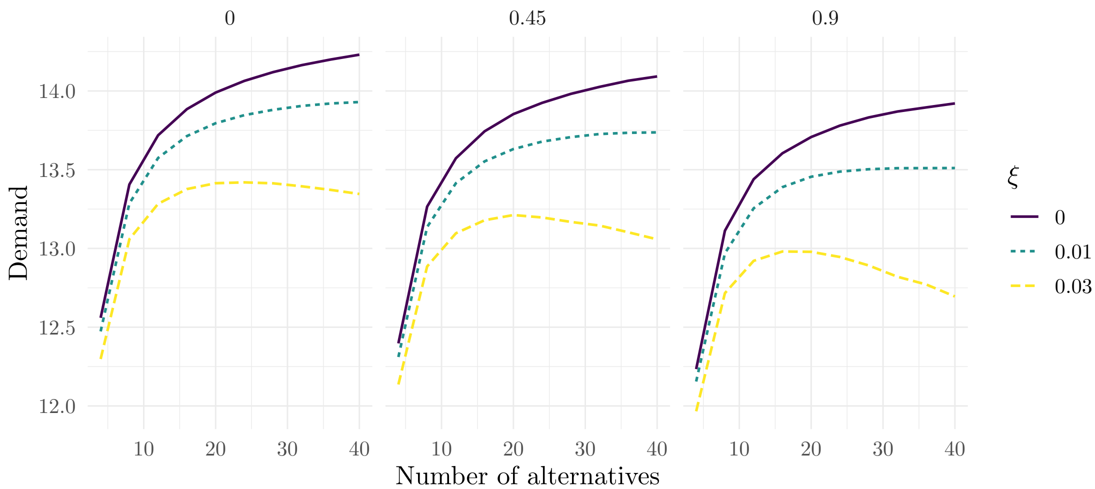
Satomura et al. (2011) propose a volumetric demand model that accounts for multiple constraints. In their application, they add a space constraint that prevents consumers from expanding their level of consumption beyond the capacity of their storage space. Binding space constraints can explain a ‘cap‘ on the curve describing the choice-set size / primary demand relationship. However, it cannot explain an inverted U-shape relationship between set size and primary demand. Moreover, we do not see expenditures across choice tasks ‘pile up‘ close to the budget constraint.

2.5.2 Correlated errors

As a choice set is populated with more choice alternatives, consumers may perceive more choice alternatives as ‘similar’. Researchers may argue that error terms of ‘similar‘ choice alternatives should be correlated. Given the same choice-set size, error correlation may lead to lower levels of primary demand. Dotson et al. (2018) propose a model that allows for correlated error terms in a discrete choice model, relaxing the independence of irrelevant alternatives (IIA) assumption, and allowing for higher rates of substitution among similar products. They parameterize the covariance structure as a function of similarity, where similarity is informed by similarity in overall (deterministic) utility. Conceptually, this could also be applied to volumetric demand models. However, estimating a (discrete/continuous) volumetric demand model with correlated errors is challenging, since the likelihood comprises mass and density contributions. Fortunately, it is straightforward to *simulate* volumetric demand given correlated errors.

We redo the simulation illustrated in Figure 1 for different levels of correlation (0. 0.45, .9) between the error terms. We generate error terms from a Multivariate Normal distribution (instead of Extreme Value), and then simulate demand following the algorithm described in Appendix A. Results are shown in Figure 2. We can see that even large levels of correlation between error terms (.9) only result in minor changes in primary demand relative to no error correlation (keeping all other parameters equal), while different values of ξ can explain different patterns of set-size-primary-demand relationships. Moreover, the proposed model is able to describe an inverted U-shaped relationship between assortment size and primary demand, while error correlation is only able to explain minor shifts of the assortment-size-primary-demand curve.

Figure 2: Set size and primary demand given different levels of error correlation and ξ



3 Empirical application

We use data from two studies to investigate the properties of our proposed demand model. Both datasets are collected from commercials panels. In both studies, we use experimental choice-set size variation which can help identify the proposed set size parameter(s). We use the estimated models to extrapolate from the relative small-N experimental world to market-like large-N scenarios. The following models will be applied:

- **vd** - an extant specification of a **v**olumetric **d**emand model (see Section 2.1)
- **vd-ss(o)** - our proposed model with **s**et-**s**ize adjustment (where o is the order of the polynomial, see Section 2.2)

To identify respondents with unrealistic or incoherent preferences, we first estimate simple volumetric demand models (vd) for each set-size, obtain individual log-likelihood values and remove about the 10% of worst-fitting respondents. We also remove respondents who never choose a single choice alternative.

The joint posterior distribution of the model parameters are obtained using the MCMC method for hierarchical Bayes models. We estimated all models using 500,000 iterations and used the last 100,000 draws to obtain parameter estimates.

3.1 Chocolate bars

Our first application is based on a volumetric choice experiment from a commercial panel in Germany. The design of our choice experiment follows common studies in the packaged goods industry, except for choice-set size variation. Chocolate bars are very popular in Germany, where people consume an average of about 9kg of chocolate products every year, half of which is traditional chocolate bars (Statista, 2018b). In order to test how accurately competing demand models predict market-level demand in a 'base case' scenario (i.e., current market demand at current market offerings), the study is designed to reproduce a set of typical market offerings available in supermarkets across Germany in 2018. Therefore, no new flavors or flavor combinations were added to the study.

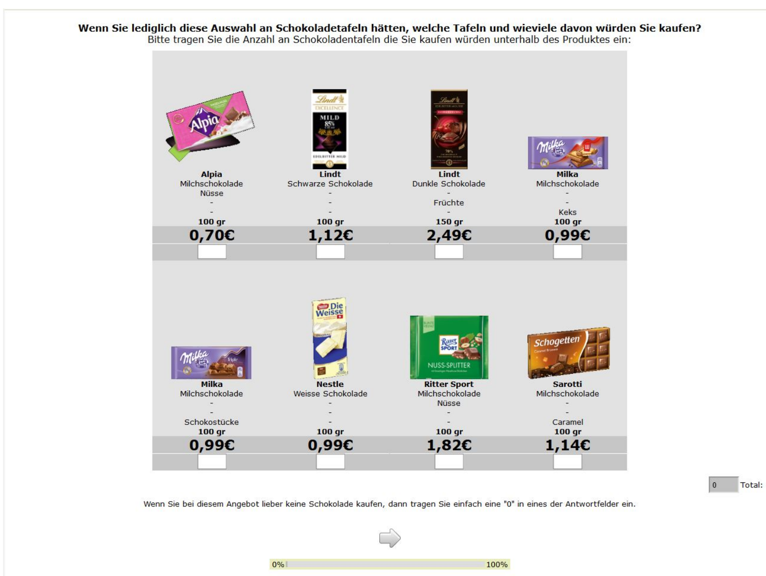
Table 1: Attributes and Levels (Chocolate bars)

Attributes	**Levels**
Brand	Alpia, Feodora, Kinder, Lindt, Merci, Milka, Nestle, Ritter, Sarotti, Schogetten, Suchard, Tobler, Trumpf, Ferrero/Yogurette
Chocolate	Milk, Dark, Black, White
Nut	Nut, No Nut
Fruit	Fruit, Berry, Grape, No Fruit
Filling	None, Yogurt, Choc Chunk, Coffee, Cookie, Black and White, Crisp, Nougat, Caramel, Milk Creme, Special, Marzipan

We characterized chocolate bars in terms of five key attributes: Brand name, Chocolate type, Nut content, Fruit or Berry content and Filling. An overview of attributes and levels in shown in Table 1. Using those attributes and levels we can map between product space (with about 100 unique products accounting for 80% of sales volume) and the lower-dimensional attribute space. Most bars are offered in 100g pack sizes, while some brands are traditionally associated with 80g or 120g bars. To account for this, we scale volumetric demand accordingly, i.e. 1 unit demand for a 120g bar is represented as demand for 1.2 units. Respondents were shown a glossary of 117 common product configurations before being presented with the choice tasks. The glossary is designed to help respondents understand the attributes used in the choice experiment, which is common in studied involving large numbers of attributes or levels (Rao, 2013).

Figures 3, 4 show example choice tasks with 8 and 18 alternatives, respectively. Respondents were presented with a picture of the product and a description of the attribute levels associated with that particular product. Alternatives were arranged by manufacturer if multiple alternatives from the same manufacturer were included in a choice set. This resembles the typical shopping situation in a store, where products are arranged in the same way. Respondents are asked to type the number of bars they would buy from each variety. They are reminded that it is possible to not buy any of the offerings, in which case they need to type in '0' for any of the alternatives. The total number of bars chosen on a given task is shown on the bottom right corner of the screen.

Figure 3: Example choice task - 8 alternatives



In order to analyze the effect of choice-set size on demand, we created different versions of the volumetric choice experiment. Respondents were randomly assigned to one of the different versions of the conjoint experiment. Table 2 provides an overview of the different versions. Versions 1 and 2 are used to analyze commonalities and differences in demand model estimates between small and large set size scenarios. Our proposed methods is based on version 3, where the size of the choice set is changed across choice tasks. The

Figure 4: Example choice task - 18 alternatives

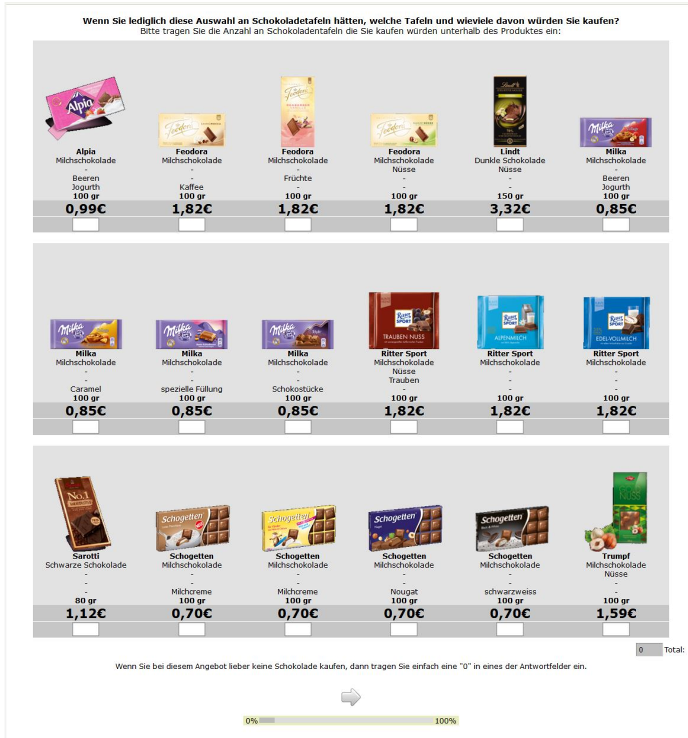


Table 2: Experimental set-up (Chocolate bars)

	Number of alternatives during		sample
Version	first 8 tasks	second 8 tasks	size
1	8	8	274
2	18	18	225
3	8	18	249

number of alternatives is increased from 8 to 18 after 8 tasks, providing variation needed to identify ξ, which is the effect of the choice-set size on marginal utility of a single alternative. In each version, respondents were presented with 16 choice tasks. In version 1 and version 2, 8 or 18 alternatives are presented across all 16 tasks, respectively.

In the remainder of this section, we show descriptive statistics, a comparison of model estimates based on version 1 and 2 of the choice experiment, and finally we compare marketplace predictions based on the proposed and competing models. Marketplace predictions are based on a typical offering that would be available at a typical German

Table 3: Descriptive Statistics (Chocolate bars)

Version	Number of Alternatives	Units per task		Varieties per task		$ spent per task		Maximum spent	
		mean	sd	mean	sd	mean	sd	mean	sd
1 (8 alternatives)	8	1.41	1.35	1.08	0.86	1.63	1.65	3.93	2.30
3 (8-then-18)	8	1.31	1.25	1.05	0.93	1.53	1.55	3.20	1.91
2 (18 alternatives)	18	2.16	2.09	1.51	1.19	2.49	2.61	5.48	4.06
3 (8-then-18)	18	1.95	2.08	1.54	1.61	2.29	2.56	4.44	3.27

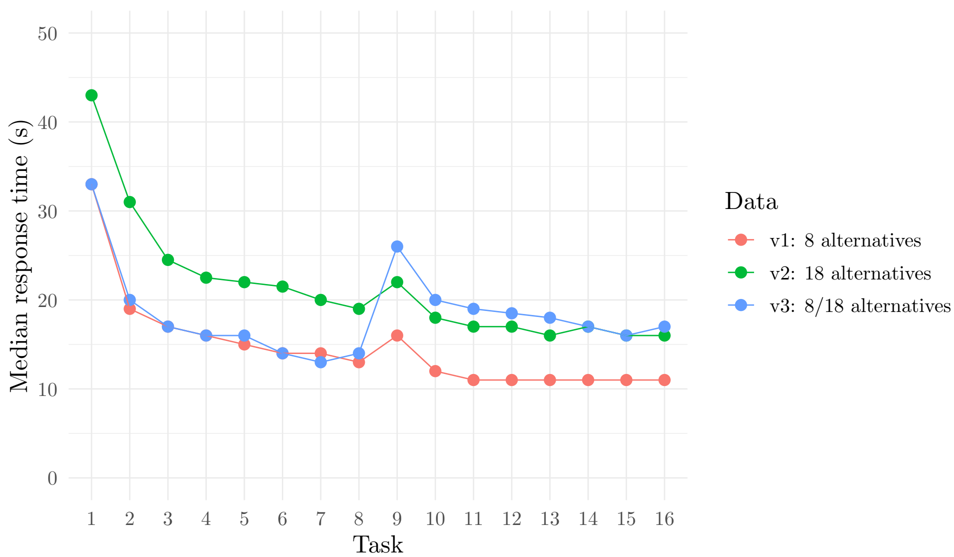
supermarket. This 'base case' scenario consists of 117 product and corresponding typical prices, which are mapped to the attribute space in our study.

3.1.1 Descriptive Analysis

Descriptive statistics of demand are summarized in Table 3. Summaries are broken down by the number of alternatives shown and choice experiment version (8 all the way, 18 all the way and 8-then-18 alternatives). Respondents choose larger quantities (around 2 instead of 1.4) and more varieties (around 1.5 instead of 1.1) when offered a larger assortment. Therefore, they overall spend more when a larger assortments are offered. Differences between versions are small, as respondents appear to quickly adjust to the size of the choice set being shown.

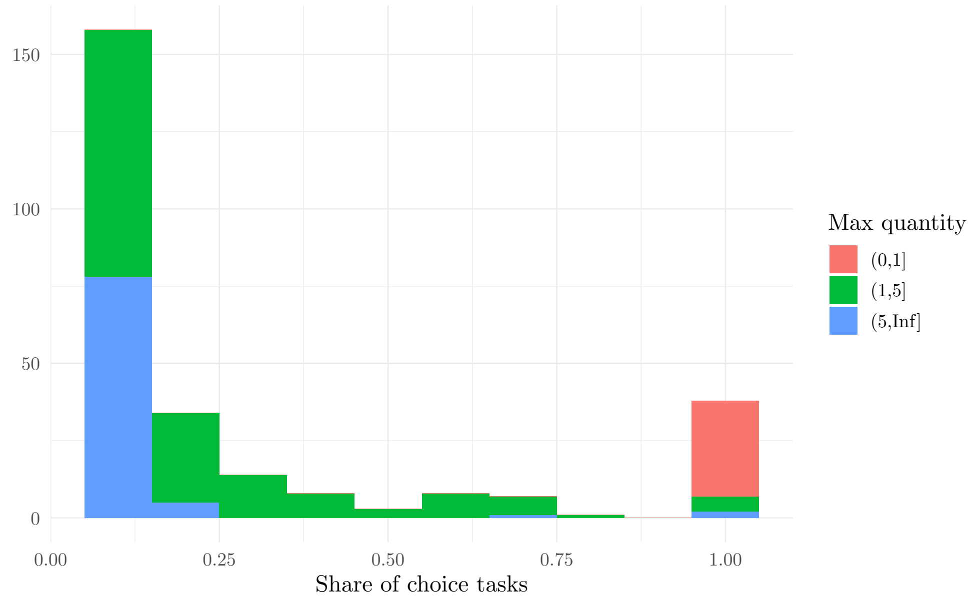
Choice-set size variation may increase burden on respondents. Therefore, we investigated response times. Their median response times are shown in figure 5. We see that (1) response times are longer when more alternatives are presented and (2) response time decrease over time, particularly over the first one to three tasks. After the first eight choice tasks all respondents saw an intermission screen. In version 3, they were told that the number of choice alternatives per task would increase. In version 1 and 2, the were only told that there would be 8 more choice tasks. The intermission screen lead to higher response times at task 9. In version 3, going from 8 to 18 alternatives, that increase in response time is larger. This is not surprising, as respondents may have needed some extra time to get used to the new look of the choice screen. However, they appear to adapt quickly. After just one task, their response times converge to those of respondents who have seen 18 alternatives throughout the entire experiment (version 2). This finding appears to be consistent with the notion of quick adaptivity to changes in set sizes (Meissner et al., 2019).

Figure 5: Median response times (Chocolate bars)



To investigate if space constraints play a significant role, we take a closer look at data from version 2 (18 choice alternative throughout the experiment). We compute the following proportion for each respondent: number of choice tasks in which the maximum observed quantity was chosen divided by number of choice tasks in which at least one product was chosen. Large shares could indicate a binding space constraint. Figure 6 shows a histogram of this proportion. We use color-coding to indicate maximum quantities chosen. Of those respondents who always choose their respective maximum quantity, that quantity is 1. Very few respondents choose a maximum quantity of between 2-5 for a majority of choice tasks. It is hard to believe that a space constraint is binding that would only allow for a single bar of chocolate. Therefore, we believe that space constraints are not a major concern in this experiment.

Figure 6: Proportion of tasks where primary demand = max(primary demand)

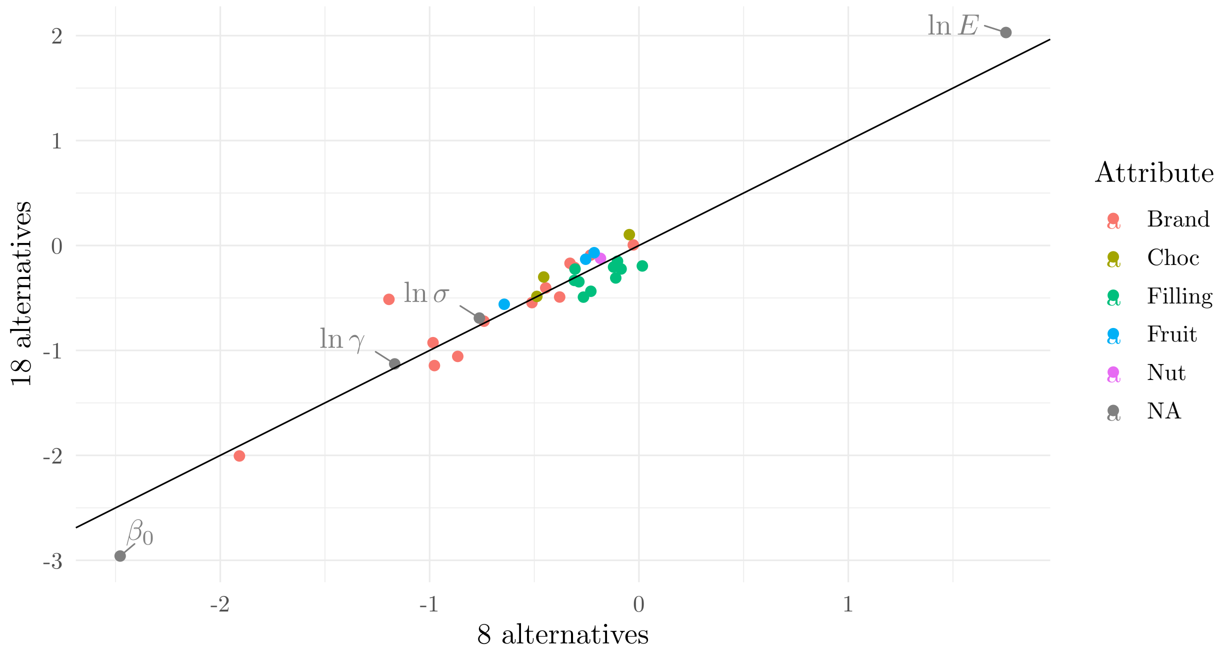


3.1.2 Estimating models on data without set size variation

In order to assess the effect of choice set size on estimated parameters of the demand model, we fit the standard volumetric demand model presented in section 2.1 to version 1 and 2 data from our VCE. The scatterplot in Figure 7 shows that estimates of $\bar{\theta}$ based on choice data with 8 and 18 alternatives.

Most parameters are estimated consistently, regardless of choice-set size. There are three exceptions: Estimates based on 18 alternatives include a larger budgetary allotment (exp(2.35) vs exp(1.75)), a smaller intercept (-3.18 vs -2.48) and a smaller error scale ($\exp -1.01$ vs -0.76). The effect on the budgetary allotment E is consistent with the descriptive summaries: respondents choose inside options more frequently and in higher quantities when more variety is offered. Since the lower bound of budget estimates is given by the maximum spent by a respondent across choice tasks, it is not surprising that budget estimates turn out larger in the 18 alternative condition. The intercept of the baseline marginal utility (β_0) of inside good is smaller in the 18 alternative condition. This is consistent with our conjecture in section 2.2: All else equal, larger choice sets

Figure 7: Comparing $\bar{\theta}$ estimated from 8 vs 18 alternatives (Chocolate bars)



imply smaller β_0 if set size effects exist. The difference in error scale may in part be explained by the more informative nature of higher-dimensional choice data.

3.1.3 Estimating models on data with set size variation

In version 3 of the VCE, respondents chose among 8 alternatives for the first 8 tasks, and then continued to choose among 18 alternatives for the last 8 choice tasks. We randomly select 1 choice task per respondent for out-of-sample fit statistic computation and estimate the proposed and benchmark models (vd-ss(1) and vd).

Table 4: Comparing Fit (Chocolate bars)

	In-sample	Out of sample	
Model	LMD	MSE	MAE
vd	-12,423.66	0.445	0.183
vd-ss(1)	-12,346.92	0.455	0.182

Fit statistics are presented in Table 4. It shows the log marginal density of the data

(LMD) for in-sample fit, and the mean squared error (MSE) and mean absolute error (MAE) for out-of-sample fit. They are computed as follows:

$$\text{MSE}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\sum_{i=1}^{n} (\hat{\mathbf{x}}_i - \mathbf{x}_i)^2}{n} \tag{24}$$

$$\text{MAE}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\sum_{i=1}^{n} |\hat{\mathbf{x}}_i - \mathbf{x}_i|}{n} \tag{25}$$

Out-of-sample fit statistics are based on predictions of demand for the hold-out tasks. These predictions are computed via numerical integration. We take 10,000 thinned draws from the individual-level posterior distribution of parameters, generate realizations for the error terms and compute demand conditional on both (see section 2.4). By averaging across the resulting conditional demands, we are integrating over the distribution of parameters and errors.

Comparing model fits in Table 4, two observations stand out: Firstly, there are no dramatic differences in model fit between the models. This is to be expected, because choice-set size variation is limited to 8 and 18 alternatives. A much better test of external validity is based on the ability of the model to predict actual purchase behavior, beyond the respective choice experiment. We do this in Section 3.1.4.

Estimates of $\bar{\theta}$ are summarized in Table 5. Overall, most estimated parameters are the same except for the introduction of ξ and a larger β_0 estimate in the proposed model. The shift in β_0 is to be expected, as laid out in Section 2.2.

3.1.4 Demand predictions for market scenarios

We use a 'base case' scenario that mimics an assortment available at a typical German supermarket in 2018. It consists of 117 products, including their configuration and typical price. This scenario provides a realistic approximation of actual choice sets experienced by out respondents when shopping in a grocery store. We use this scenario to conduct two assessments of the estimated demand models: (1) The ability to reproduce self-

Table 5: Estimates (Chocolate bars)

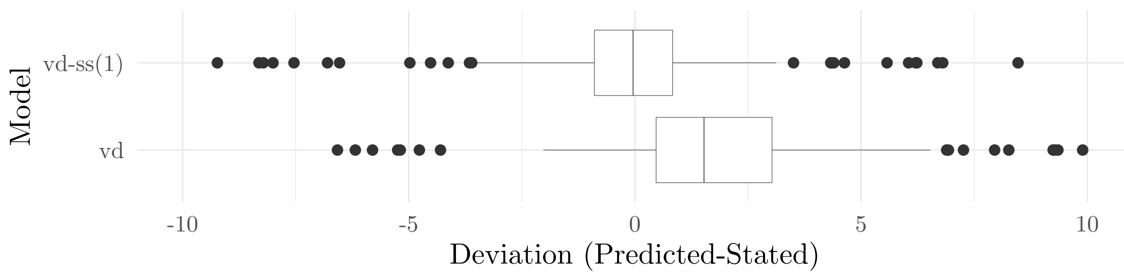
	vd		vd-ss(1)	
β_0	**-2.81**	(0.12)	**-2.51**	(0.13)
Brand (reference: Milka)				
Alpia	**-0.89**	(0.11)	**-0.89**	(0.12)
Feodora	**-2.16**	(0.19)	**-2.24**	(0.22)
Kinder	**-0.37**	(0.20)	**-0.30**	(0.18)
Lindt	**-0.30**	(0.14)	**-0.30**	(0.15)
Merci	**-0.65**	(0.12)	**-0.66**	(0.13)
Nestle	**-0.83**	(0.18)	**-0.85**	(0.21)
Ritter	-0.08	(0.09)	-0.08	(0.09)
Sarotti	**-0.96**	(0.12)	**-0.99**	(0.14)
Schogetten	**-0.61**	(0.09)	**-0.64**	(0.09)
Suchard	**-1.07**	(0.22)	**-1.11**	(0.21)
Tobler	**-0.41**	(0.12)	**-0.38**	(0.12)
Trumpf	**-0.68**	(0.17)	**-0.70**	(0.18)
FerreroYogu	**-0.23**	(0.13)	**-0.26**	(0.13)
Choc (reference: milk)				
dark	**-0.46**	(0.11)	**-0.48**	(0.11)
black	**0.24**	(0.13)	**0.31**	(0.14)
white	**-0.55**	(0.10)	**-0.56**	(0.11)
Filling (reference: none)				
yog	**-0.45**	(0.08)	**-0.46**	(0.08)
chocchunk	**-0.38**	(0.09)	**-0.37**	(0.08)
coffee	**-0.29**	(0.10)	**-0.28**	(0.12)
cookie	**-0.36**	(0.08)	-0.35	(0.08)
blackwhite	**-0.42**	(0.10)	**-0.39**	(0.09)
crisp	**-0.32**	(0.09)	**-0.34**	(0.10)
nougat	**-0.02**	(0.07)	**-0.02**	(0.07)
caramel	**-0.15**	(0.08)	**-0.16**	(0.07)
milkcreme	**-0.19**	(0.07)	**-0.20**	(0.07)
special	**-0.24**	(0.07)	**-0.21**	(0.07)
marzipan	-0.26	(0.09)	-0.27	(0.10)
Fruit (reference: none)				
fruit	**-0.78**	(0.13)	**-0.80**	(0.14)
berry	**-0.29**	(0.09)	**-0.29**	(0.08)
grape	**-0.12**	(0.07)	**-0.11**	(0.07)
Nut (reference: none)				
nut	**-0.13**	(0.08)	**-0.13**	(0.07)
Model				
$\ln \sigma$	**-0.73**	(0.06)	**-0.72**	(0.06)
$\ln \gamma$	**-1.17**	(0.08)	**-1.14**	(0.09)
$\ln E$	**1.92**	(0.06)	**1.94**	(0.07)
$\ln \xi$			**-3.74**	(0.21)

Standard deviations in parentheses; boldfaced parameters signify that the 95% posterior credible interval of the estimate does not include zero

reported choice quantities of our respondents and (2) the ability to extrapolate market-level demand. Like in most conjoint analysis studies, we do not have matched individual purchase histories of our respondents, and therefore self-reported quantities are the best approximation of real-life behavior available. Market extrapolation rests on additional assumptions: the number of potential buyers in the market and purchase frequency.

First, we will assess the ability of the models to predict self-reported demand quantities. We generate demand predictions for all respondents given the 'base case' scenario with 117 alternatives. We then compute the absolute error by subtracting the self-reported demand quantity. Figure 8 shows distributions of absolute error for the proposed and competing models. It is clear that the extant model is systematically biased, over-predicting self-reported quantities.

Figure 8: Individual primary demand prediction error (Chocolate bars)



In order to project marketplace demand, we need to make additional assumptions: The number of households in Germany that regularly shop chocolate bars is around 25,000,000 (Statista, 2018a). Germans shop for chocolate almost every week, for an average of 3 shopping trips per month during which they shop for chocolate bars. These are simplifying assumptions, ignoring both purchase dynamics (e.g., stockpiling) and consumption dynamics (e.g., consumers eating more because 'it is there'). Despite the assumptions, a useful model should produce predictions that are at least close to real demand.

Extrapolated marketplace demand estimates (in tons of chocolate) are shown in Table 6. For reference, we add an extrapolation based on stated quantity. From aggregate reports, we found that actual marketplace demand equals about 240,000t[1]. The extant

[1]The latest actual marketplace demand number we found is from 2016. We have not seen evidence for dramatic changes in primary demand (Planung&Analyse, 2016).

Table 6: Extrapolation: Annual chocolate bar demand (in kilotons)

Model	E(demand)	S(demand)	CI-5%	CI-95%
vd	434,730	9,989	419,298	450,810
vd-ss(1)	218,742	18,792	191,678	242,982
Based on stated quantity	215,422			
Actual	∼240,000			

model daramatically over-predict demand, while the proposed model produces a realistic prediction.

While the focus of our investigation is primary demand, we also need to assess the ability of our approach to produce reasonable predictions of secondary demand. Since there are 117 products in our marketplace scenario, we aggregate demand to the brand-name level and report expected purchase quantities and shares per occasion in Table 7. While the proposed and extant model differ in terms of absolute quantity predictions, market share predictions are similar. This is to be expected since the drivers of secondary demand are also very similar between the models.

Table 7: Average demand by brand in units (Chocolate bars)

	Volume		Share	
	vd	vd-ss(1)	vd	vd-ss(1)
Milka	1.47	0.71	0.30	0.29
Ritter	0.95	0.49	0.20	0.20
Alpia	0.64	0.31	0.13	0.13
Lindt	0.47	0.24	0.10	0.10
Schogetten	0.38	0.19	0.08	0.08
Tobler	0.23	0.12	0.05	0.05
Sarotti	0.22	0.12	0.05	0.05
Merci	0.11	0.06	0.02	0.03
Kinder	0.11	0.05	0.02	0.02
FerreroYogu	0.10	0.05	0.02	0.02
Nestle	0.07	0.04	0.01	0.01
Feodora	0.04	0.02	0.01	0.01
Trumpf	0.02	0.01	0.00	0.01
Suchard	0.02	0.01	0.00	0.00

We use a sales prediction task an additional test of external validity. Most manufacturers do not disclose specific sales figures for chocolate bars. However, a preponderance of Ritter Sport's offerings are chocolate bar products, and most of their sales come from traditional chocolate bars. The majority of their 100g bars are sold for around EUR 1.19 (regular) and EUR 0.99 (sale). Retail prices include a 19% value-added tax. Accounting for typical retail margins, we assume that one bar generates a revenue of 50 cents for Ritter Sport. Based on these assumptions, we generate annual sales predictions (in millions of Euros). Results are shown in Table 8. Ritter Sport's published annual sales number is EUR 480m globally (WirtschaftsWoche, 2018), where about half of its business is domestic[2]. We expect an accurate sales prediction to be slightly lower than EUR 240m, accounting for the small share of Ritter Sport products that are not standard chocolate bars. The proposed model predict sales of around EUR 220m, which is in line with that expectation. In contrast, volumetric demand models without the proposed adjustment lead to sales predictions that are significantly too high.

Table 8: Domestic Annual Sales - Ritter Sport (in millions of Euros)

Model	E(Sales)	S(Sales)	CI-5%	CI-95%
vd	428,198	40,645	363,659	495,752
vd-ss(1)	221,398	33,589	167,129	275,976
Actual	∼240,000			

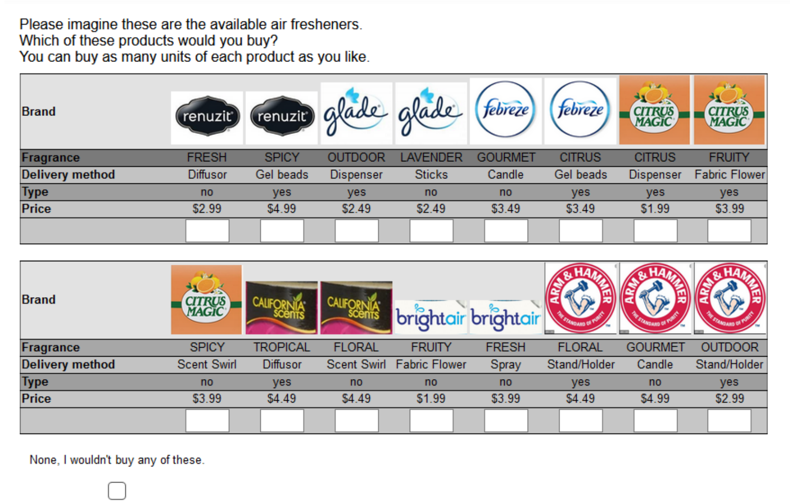
3.2 Air fresheners

In our second application, we conducted a volumetric choice experiment in the air 'NECA' freshener category. These are simple non-electric air fresheners available in regular retail stores. Respondents were recruited from a commercial panel in the United States. They were shown 8, 16 and 24 choice alternatives for a total of 15 choice tasks. An example choice task with 16 alternatives is shown in Figure 9.

In order to assess the ability to extrapolate demand to scenarios with more choice

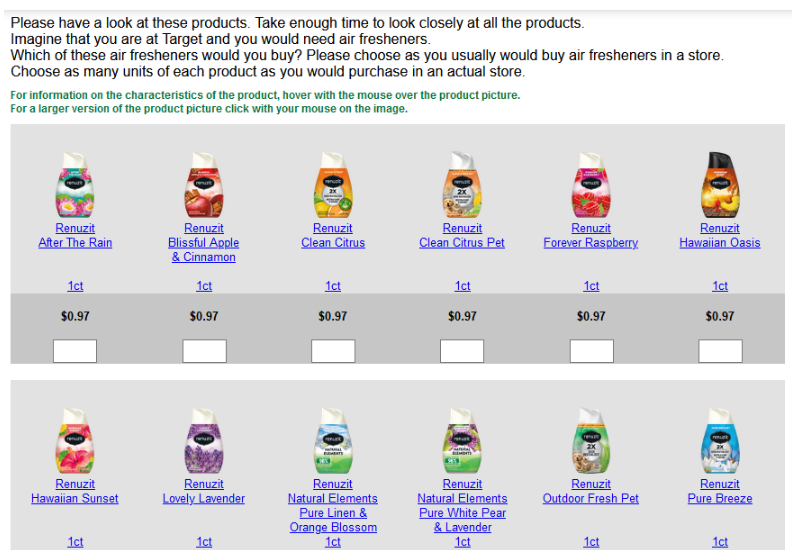
[2]It was at 60% in 2013, but the global share has increased steadily since (Hahn, 2013).

Figure 9: Air freshener example choice task - 16 alternatives



alternatives, we showed respondents an initial 'shelf task' with 57 choice alternatives. The assortment of 57 alternatives closely resembles a typical offering in a store. Moreover, this shelf task does not show an attribute grid, since it's meant to best mimic a real-world purchase decision in a store. The top portion of this task is depicted in Figure 10.

Figure 10: Air freshener validation shelf task



An overview of all attributes and levels is shown in Table 9. A brief glossary was provided to respondents. 'Brand name' and 'Scent' are straightforward attributes. 'Delivery'

Table 9: Attributes and Levels (Air fresheners)

Attribute	**Levels**
Brand	Citrus Magic, Glade, Arm & Hammer, California Scent, Febreze, Renuzit, BrightAir
Scent	Citrus, Gourmet, Fresh, Fruity, Lavender, Outdoor, Floral, Tropical Spicy
Delivery	Scent Swirl, Stand Holder, Candle, Fabric Flower, Diffusor, Spray, Dispenser, Gelbeads, Sticks
Adsorption	No, Yes

Table 10: Descriptive Statistics (Air fresheners)

Number of	Units per task		Varieties per task		$ spent per task		Maximum spent	
Alternatives	mean	sd	mean	sd	mean	sd	mean	sd
8	1.05	1.27	0.79	0.80	2.80	3.84	6.30	5.46
16	1.04	1.31	0.80	0.87	2.84	4.07	6.20	5.59
24	1.43	1.76	1.11	1.17	3.88	5.27	7.59	7.17
57	4.11	6.18	2.58	2.87	6.21	10.10	6.21	10.10

refers to the technique used to deliver the scent. While candles need to be lit to dispense a scent, Gelbeads only need to be placed in a room. 'Adsorbtion' refers to the ability of the product to remove unwanted smells from the air. While the actual process is called adsorbtion, it is sometimes sold as 'odor absorption' to consumers.

3.2.1 Descriptive Analysis

Descriptive statistics of demand are summarized in Table 10. Summaries are broken down by number of choice alternatives shown. Primary demand increases as the set size is increased beyond 16 alternatives. This supports the general idea that consumers respond to increased variety by increasing primary demand.

3.2.2 Model estimation and results

In-sample fit statistics for each of the models are shown in Table 11. The log marginal density (LMD) improves when adding the set-size specification. While accounting for set size variation improves fit, adding a quadratic term does not appear to further improve fit.

Table 11: In-sample fit (Air fresheners)

	vd	**vd-ss(1)**	vd-ss(2)
LMD	-32,529	-32,215	-32,233

Estimates of $\bar{\theta}$ are summarized in Table 12. Overall, we see consistency in the estimated 'part-worth' coefficients, error scale, satiation rate parameter and budget constraint. The estimated intercept coefficient changes as we control for choice-set size.

3.2.3 Model validation

Respondents were shown a realistic shelf task with 57 choice alternatives. This task was shown before respondents engaged with the choice experiments, in order to obtain a realistic approximation of each respondents' actual buying behavior. Table 13 shows the predictive accuracy of the competing models. Models with set size adjustment again outperform the extant model. Our proposed vd-ss(1) model produced the best in-sample fit and is able to generate more accurate predictions, with relative bias close to 0.

For Table 14, we aggregated demand for the 57 products to the brand-name level to facilitate comparisons. The proposed vd-ss(1) model predict overall demand of 2,196 units from our 516 respondents. Actual demand in the shelf task was 2,120. The extant model predicts a demand of 2,953 units, over-predicting demand by almost 40%.

Table 12: Estimates (Air fresheners)

	vd		vd-ss(1)		vd-ss(2)	
β_0	**-5.87**	(0.26)	**-3.85**	(0.17)	**-2.00**	(0.18)
Brand (reference: Citrus Magic)						
Glade	**-0.28**	(0.10)	**-0.26**	(0.08)	**-0.27**	(0.09)
ArmHammer	**-0.18**	(0.09)	**-0.17**	(0.08)	**-0.22**	(0.08)
CaliforniaScent	**-0.50**	(0.09)	**-0.45**	(0.08)	**-0.48**	(0.08)
Febreze	**-0.17**	(0.09)	**-0.22**	(0.08)	**-0.33**	(0.09)
Renuzit	**-0.31**	(0.08)	**-0.23**	(0.08)	**-0.33**	(0.09)
BrightAir	**-0.29**	(0.08)	**-0.33**	(0.08)	**-0.28**	(0.08)
Delivery (reference: Scent swirls)						
StandHolder	**-0.33**	(0.08)	**-0.26**	(0.08)	**-0.27**	(0.07)
Candle	**-0.16**	(0.09)	-0.21	(0.07)	-0.14	(0.07)
FabricFlower	**-0.13**	(0.08)	**-0.11**	(0.09)	**-0.13**	(0.08)
Diffusor	**-0.32**	(0.07)	**-0.24**	(0.09)	**-0.24**	(0.07)
Spray	**-0.21**	(0.08)	**-0.18**	(0.08)	**-0.23**	(0.07)
Dispenser	**-0.11**	(0.07)	**-0.02**	(0.08)	**-0.02**	(0.07)
Gelbeads	**-0.32**	(0.08)	**-0.30**	(0.07)	**-0.29**	(0.08)
Sticks	**-0.33**	(0.09)	**-0.34**	(0.07)	**-0.26**	(0.07)
Scent (reference: Citrus)						
Gourmet	**-0.34**	(0.09)	**-0.30**	(0.08)	**-0.26**	(0.08)
Fresh	**-0.30**	(0.09)	**-0.08**	(0.09)	**-0.11**	(0.08)
Fruity	**-0.37**	(0.09)	-0.26	(0.08)	**-0.24**	(0.08)
Lavender	**-0.61**	(0.08)	**-0.36**	(0.09)	**-0.43**	(0.09)
Outdoor	**-0.39**	(0.08)	**-0.24**	(0.08)	**-0.23**	(0.07)
Floral	-0.38	(0.08)	-0.19	(0.08)	-0.25	(0.08)
Tropical	**-0.33**	(0.08)	**-0.19**	(0.08)	**-0.27**	(0.08)
Spicy	**-0.32**	(0.08)	**-0.18**	(0.08)	**-0.14**	(0.07)
Adsorbent type						
yes	**-0.12**	(0.06)	**-0.13**	(0.05)	-0.05	(0.05)
Model						
$\ln \sigma$	**0.47**	(0.03)	**0.46**	(0.03)	**0.46**	(0.03)
$\ln \gamma$	**-0.31**	(0.04)	**-0.32**	(0.04)	**-0.32**	(0.04)
$\ln E$	**2.31**	(0.03)	**2.31**	(0.03)	**2.32**	(0.03)
$\ln \xi_1$			**-0.80**	(0.14)	**-6.01**	(0.17)
$\ln \xi_2$					**1.05**	(0.11)

Standard deviations in parentheses; boldfaced parameters signify that the 95% posterior credible interval of the estimate does not include zero

Table 13: Validation task fit (Air fresheners)

Model	MSE	MAE	Bias
vd	1.19	0.17	0.03
vd-ss(1)	0.94	0.14	0.00
vd-ss(2)	0.94	0.14	0.00

Table 14: Brand-level demand predictions (Air fresheners)

Brand	*Actual*	vd	**vd-ss(1)**	vd-ss(2)
Renuzit	1,137	1,675	1,256	1,225
Glade	479	750	559	555
Febreze	311	245	177	174
BrightAir	63	39	28	29
CitrusMagic	51	105	77	77
ArmHammer	40	14	10	10
CaliforniaScent	39	126	88	89
Total	2,120	2,953	2,196	2,162
relative		139%	104%	102%

4 Discussion

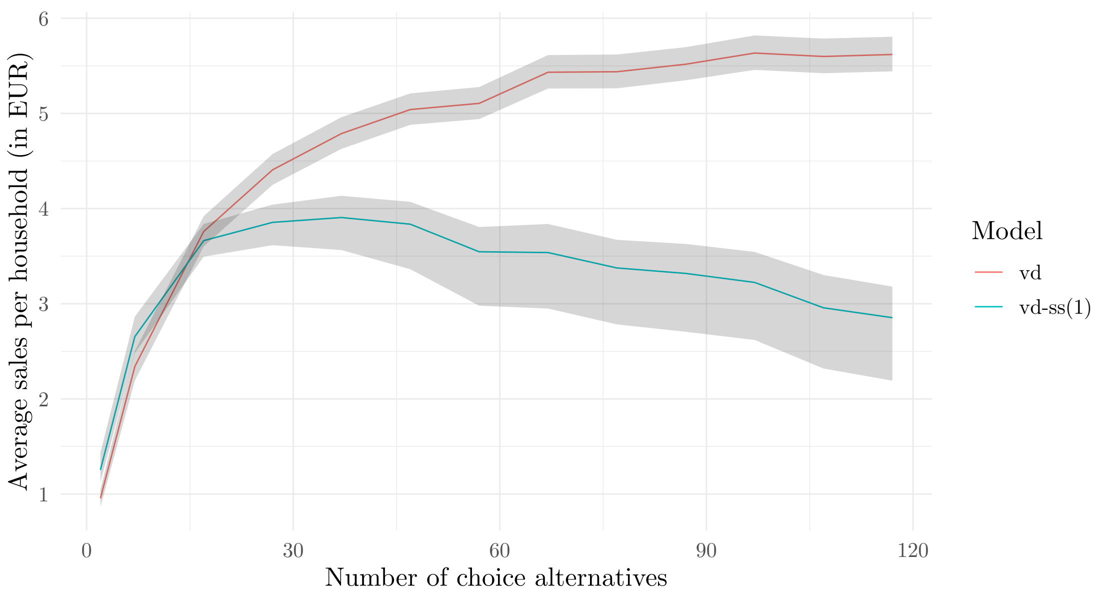
The following topics merit discussion: (1) implications for (volumetric) conjoint analysis, (2) policy simulations involving significant set size variation, (3) implications for data fusion, (4) implications for 'marketplace predictions' based on choice experiments and (5) the relationship to the behavioral literature on choice-set size.

Volumetric conjoint analysis with set size variation is a great tool for policy simulations that involve the addition or removal of several choice alternatives at the same time. In two applications, we have demonstrated the ability of our approach to produce accurate predictions while the extant model is prone to over-predictions (by 40%-80%). However, drivers of secondary demand (see Table 5 and 12) and market share predictions (see Table 7) seem largely unaffected by choice-set size variation. If predictions of demand quantities are not required, discrete choice conjoint may be sufficient. However, if demand quantities are important and significant changes to the number of offerings are considered, introducing choice-set size variation into the choice experiment may be advisable.

Using our model, we are able to predict primary demand in response to changes in set size. To illustrate that relationship, we use the vd-ss(1) from the chocolate bar study. We sort choice alternatives from our marketplace scenario by predicted sales (in EUR), and then predict primary demand for the best-selling 2....117 alternatives. The resulting average category sales per respondent (in EUR) are shown in Figure 11. The shaded ribbons show the 95% credibility intervals of the posterior distributions of expected demand. The plot does not look 'smooth' because we are removing choice alternatives with varying baseline marginal utilities and prices.

Results from the proposed model suggest that adding choice alternatives beyond 30 products does not result in significant increases in primary demand. The competing standard model predicts starkly increasing demand in the given range. If the number of alternatives was further increased, the standard model would converge towards a boundary as a result of the budget constraint. At some point, our model may predict decreasing primary demand. Retailers can use this model to optimize their assortment. Depending on the costs to maintain a larger assortment, there might be little to gain from adding more varities. In fact, stores might even be able to downsize their assortment by a few alternatives. Previous studies have found similar effects – reducing assortments often does not reduce sales, especially in the long run (e.g., Boatwright and Nunes, 2001).

Figure 11: Average category sales per customer and set size (Chocolate bars)

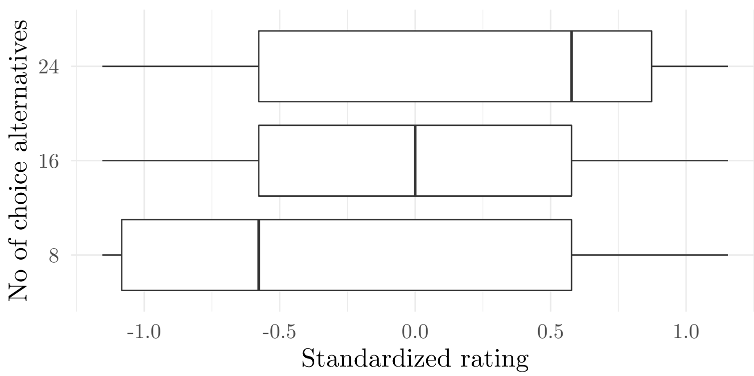


The proposed set size adjustment is indispensable for data fusion with volumetric choice experiments and demand data. Large differences in N_t will lead to poor fit for any $\xi > 0$. For instance, Ellickson et al. (2019) combine conjoint and transaction data. While they employ a discrete choice model, a volumetric version of their application would need to incorporate a set size adjustment function. Allenby et al. (2019) find that the intercept term β_0 based on transaction data is much smaller than based on conjoint data from the same subjects. In Section 2.2 we conjectured that set size differences would result in shifts of β_0. We believe that using our set size model would help to 'fuse' the choice experiment with transaction data.

The proposed model improves the ability of conjoint analysis as a tool for marketplace simulation and prediction by improving predictions for choice sets of different dimensions. Extrapolating from 18 to 100+ alternatives in our first study might appear like a huge leap. However, it is common practice to use a simple choice experiment, and then predict 'base case' demand given a set of common offerings available in stores. While this is straightforward for discrete choice models relying on the IIA, we have shown that volumetric models make this a more difficult task because we are also modeling primary demand. Our parsimonious model allows us to obtain realistic demand quantity predictions. Both per-task predictions and market-level quantities are in line with reality, while ignoring our proposed adjustment leads to predictions that are off by almost a factor of two. In contrast to common industry practice, we are able to obtain surprisingly accurate predictions of marketplace demand just from simulating demand for a market-like scenario, without additional calibration. There are numerous other challenges for marketplace predictions based on choice experiments, including purchase timing, response biases and competitive reactions. However, we are convinced that set size variation is one key aspect to be considered.

We view our proposed model as complementary to previous behavioral studies of the link between assortment size and demand. Most investigations on the topic are based on discrete choices, and we encourage researchers to consider quantity demand models, especially in packaged goods where consumers demand variety. One particular literature stream focuses on instances when increases in assortment size lead to lower purchase probabilities. This effect is sometimes referred to as 'choice overload' (Chernev et al., 2015). There are various explanations for the effect, including search costs (Kuksov and Villas-Boas, 2010), differences in motivation (Iyengar and Lepper, 2000), heightened expectations in larger choice sets (Diehl and Poynor, 2010), deferral of choice or increased appeal of choosing the outside option (Dhar, 1997). Based on the antecedents (complexity, difficulty, uncertainty, decision goal) of choice overload described by Chernev et al. (2015), we do not believe that the effect plays a major role in the chocolate bar application. Moreover, our model would allow to describe a 'choice overload' situation, since there are not additional constraints on $f(N_t;\xi)$. Our model suggests that German chocolate buyers are unlikely to experience choice overload in a typical assortment. The number of alternatives would have to be increased much more to see an inverse U-shaped relationship. In our air freshener study, we asked respondents to evaluate their level of satisfaction with their choice given after choosing from 8, 16 and 24 alternatives. Unlike Feinberg et al. (2016), we do not find that respondents are less satisfied given larger set sizes. Figure 12 shows boxplots of standardized satisfaction ratings (5-point ratings, mean-centered, divided by standard deviation) by set size, where no such relationship can be found.

Figure 12: Standardized satisfaction with choice by set size (Air fresheners)



5 Concluding Remarks

This paper proposes a volumetric demand model that accounts for choice-set size variation. It extends the multiple discrete-continous approach by allowing for a relationship between choice-set size and the utility of the outside good. We demonstrate that the proposed model fits volumetric demand data with set size variation better than previous models. Our findings suggest that policy simulations that involve varying choice-set sizes should be informed by data that also contains set size variation and a corresponding model.

Ignoring set size variation can lead to significant over-prediction of the demand effect of assortment increases.

This paper not only adds to the established literature on multiple discrete-continous demand models, but it also connects choice models with empirical and behavioral findings on choice-set size. Our model is able to explain an inverted-U-shaped relationship between the number of choice alternatives and primary demand. It turns the relationship into an empirical question. Even though we do not observe 'choice overload', our proposed model would be able to describe it.

The proposed model has implications for conjoint analysis, data fusion and modeling transaction data. All of these applications can involve significant variation in choice-set size. Transaction data can include assortment changes over time, or varying assortment sizes in different stores. Data fusion involving choice experiments and transaction data is likely to involve dramatically different set sizes.

The model also has implications for assortment optimization, which is typically based on discrete choice models (Kök et al., 2015). It is relevant for retailers, who need to manage costs of larger assortments and competition from wholesale clubs with limited assortments. It is also important for manufacturers who need to decide how many different varieties to offer in a product line. While the focus of this paper is on packaged goods, the model can also be applied to other categories where simultaneous demand for multiple varieties is common, including entertainment products and services or apparel.

Our paper has implications for conjoint analysis. If researchers are only interested in market shares, discrete choice conjoint may be sufficient. If the goal is to understand and predict demand quantities, it may be necessary to introduce choice-set size variation into the study. Deciding on the range and number of set sizes to use is an additional challenge, and respondent feedback should be collected to make sure that tasks remain manageable.

Careful model validation is important for new proposed demand models, however, many choice modeling papers rely on hold-out choice tasks for validation. We used two validation methods that are a stricter test of validity. In one study we predict demand for a marketplace scenario and compare with aggregate demand data. In the other study we use a virtual shelf task for validation.

We recognize some limitations of our study: We apply our proposed model to two choice experiments. Applying the model to transaction data or a combination of transaction and choice experiment data could provide further evidence for the validity of the model. Transaction data would also allow studying stockpiling and purchase timing. Controlling for set size variation might have implications for stockpiling models. We expect respondents to answer randomly at worst, and therefore see no need for 'incentive alignment' (Ding et al., 2005; Yang et al., 2018). Moreover, commercial panel providers do not allow implementing it.

There are many avenues for future research involving set size variation. Possible model extensions might include consideration set formation (Huang and Bronnenberg, 2018), volume discounts (Howell et al., 2016) or other extensions of volumetric demand models. Additional work could focus on combining purchase timing and stockpiling (Mela et al., 1998) with a model of set size variation. There would also be opportunities to apply volumetric conjoint experiments as part of behavioral research on choice set size in order to shed further light on the underlying processes. Finally, more work is needed to provide design recommendations for choice experiments that involve set size variation.

Appendix

A Demand prediction

Obtaining demand estimates given realizations of the error term is straightforward. Instead of using numerical optimization, we use a recursive algorithm requiring $\leq K$ iterations to find the utility-maximizing demand conditional on ε and θ. The algorithm makes it feasible to numerically integrate out ε, θ_h and τ. It can be applied when the utility derived from the outside good is non-linear, allowing to first find the optimal amount of the outside good z and then computing the corresponding inside good quantities x_k.

In order to apply the algorithm, we first define

$$\rho_i = \frac{p_i}{\psi_i \gamma} \text{ for } 1 \leq i \leq K \tag{26}$$

$$\rho_0 = 0 \tag{27}$$

$$\rho_{K+1} = \infty \tag{28}$$

and order the values ρ_i in ascending order so that $\rho_i \leq \rho_{i+1}$ for $1 \leq i \leq K$. Then $z > \rho_k$ implies $z > \rho_i$ for $i \leq k$. At the optimum, $x_i > 0$ for $1 \leq k \leq K$, $x_i = 0$ for $k < i \leq K$ and $\rho_k < z < \rho_{k+1}$. For non-considered alternatives in the screening rule model, we set the corresponding $\rho_k = 0$.

Taking the derivative of Equation 1 with respect to x_k yields the following optimal quantities for $x_k > 0$:

$$x_k = \frac{\psi_k z}{p_k} - \frac{1}{\gamma} \tag{29}$$

From equation 29 and $z = E - \sum_k x_k$, it follows that

$$z = \frac{\gamma E + \sum_k p_k}{\gamma + \sum_k \psi_k} \tag{30}$$

We can find the optimal z by following the algorithm:

1. $a \longleftarrow \gamma E$, $b \longleftarrow \gamma$, $k \longleftarrow 0$
2. $z \longleftarrow a/b$
3. while $z \leq \rho_k$ or $z > \rho_{k+1}$:
   (a) $k \longleftarrow k + 1$
   (b) $a \longleftarrow a + \rho_k$
   (c) $b \longleftarrow b + \psi_k$
   (d) $z \longleftarrow a/b$

Once the algorithm terminates, we can insert the optimal z quantity into equation 29 to compute the optimal inside good quantities x_k. Applying this algorithm to the set size model is straightforward – we only need to re-specify ψ_j:

$$\psi_j = \exp\left(\mathbf{a}_j\beta + \varepsilon_j - \ln\left(f(N_t, \xi)\right)\right) \tag{31}$$

The error term ϵ can only be integrated out numerically by simulating from its distribution.

References

Ackerberg, Daniel A., Marc Rysman. 2005. Unobserved product differentiation in discrete-choice models: Estimating price elasticities and welfare effects. *The RAND Journal of Economics* **36**(4) 771–788.

Allenby, Greg M., Nino Hardt, Peter E. Rossi. 2019. Economic foundations of conjoint analysis. *Handbook of the Economics of Marketing, Volume 1*. Elsevier, 151–192.

Allenby, Greg M, Jaehwan Kim, Peter E Rossi. 2017. Economic models of choice. *Handbook of Marketing Decision Models*. Springer, 199–222.

Bauner, Christoph, Emily Wang. 2019. The effect of competition on pricing and product positioning: Evidence from wholesale club entry. *International Journal of Industrial Organization* **67** 102525.

Boatwright, Peter, Joseph C. Nunes. 2001. Reducing assortment: An attribute-based approach. *Journal of Marketing* **65**(3) 50–63.

Borle, Sharad, Peter Boatwright, Joseph B. Kadane, Joseph C. Nunes, Shmueli Galit. 2005. The effect of product assortment changes on customer retention. *Marketing Science* **24**(4) 616–622.

Chernev, Alexander, Ulf Böckenholt, Joseph Goodman. 2015. Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology* **25**(2) 333–358.

Dhar, Ravi. 1997. Consumer preference for a no-choice option. *Journal of Consumer Research* **24**(2) 215–231.

Diehl, Kristin, Cait Poynor. 2010. Great expectations?! assortment size, expectations, and satisfaction. *Journal of Marketing Research* **47**(2) 312–322.

Ding, Min, Rajdeep Grewal, John Liechty. 2005. Incentive-aligned conjoint analysis. *Journal of Marketing Research* **42**(1) 67–82.

Dotson, Jeffrey P., John R. Howell, Jeff D. Brazell, Thomas Otter, Peter J. Lenk, Steve MacEachern, Greg M. Allenby. 2018. A probit model with structured covariance for similarity effects and source of volume calculations. *Journal of Marketing Research* **55**(1) 35–47.

Dubé, Jean-Pierre. 2004. Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Science* **23**(1) 66–81.

Dubé, Jean-Pierre. 2019. Microeconometric models of consumer demand. Tech. Rep. 25215, National Bureau of Economic Research.

Eagle, Thomas C. 2018. A comparison of volumetric models. *Proceedings of the Sawtooth Software Conference*.

Ellickson, Paul B., Mitchell J. Lovett, Bhoomija Ranjan. 2019. Product launches with new attributes: A hybrid conjoint-consumer panel technique for estimating demand. *Journal of Marketing Research* (forthcoming).

Feinberg, Fred M., Linda Court Salisbury, Yuanping Ying. 2016. When random assignment is not enough: Accounting for item selectivity in experimental research. *Marketing Science* **35**(6) 976–994.

Hahn, Uschi. 2013. Ritter sport auf wachstumskurs. URL *https://www.tagblatt.de/Nachrichten/Ritter-Sport-steigert-Umsatz-vor-allem-im-Ausland-132791.html*. [Accessed online: 13-Jan-2019].

Hastings, W. K. 1970. Monte carlo sampling methods using markov chains and their applications. *Biometrika* **57**(1) 97–109.

Herzenstein, Michal, Utpal M. Dholakia, Scott Sonenshein. 2019. How the number of options affects prosocial choice. *International Journal of Research in Marketing* .

Howell, John R., Sanghak Lee, Greg M. Allenby. 2016. Price promotions in choice models. *Marketing Science* **35**(2) 319–334.

Huang, Yufeng, Bart J. Bronnenberg. 2018. Pennies for your thoughts: Costly product consideration and purchase quantity thresholds. *Marketing Science* **37**(6) 1009–1028.

Iyengar, Sheena S., Mark R. Lepper. 2000. When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology* **79**(6) 995–1006.

Kim, Jaehwan, Greg M. Allenby, Peter E. Rossi. 2002. Modeling consumer demand for variety. *Marketing Science* **21**(3) 229–250.

Kök, A. Gürhan, Marshall L. Fisher, Ramnath Vaidyanathan. 2015. Assortment planning: Review of literature and industry practice. *Retail Supply Chain Management*. Springer US, 175–236.

Kuksov, Dmitri, J. Miguel Villas-Boas. 2010. When more alternatives lead to less choice. *Marketing Science* **29**(3) 507–524.

Lee, Sanghak, Greg M. Allenby. 2014. Modeling indivisible demand. *Marketing Science* **33**(3) 364–381.

Meissner, Martin, Harmen Oppewal, Joel Huber. 2019. Surprising adaptivity to set size changes in multi-attribute repeated choice tasks. *Journal of Business Research* .

Mela, Carl F., Kamel Jedidi, Douglas Bowman. 1998. The long-term impact of promotions on consumer stockpiling behavior. *Journal of Marketing Research* **35**(2) 250.

Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, Edward Teller. 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**(6) 1087–1092.

Planung&Analyse. 2016. Tafelschokolade vor schokoriegel. URL *https://www.horizont.net/planung-analyse/nachrichten/Tafelschokolade-vor-Schokoriegel-152303*. [Accessed online: 13-Jan-2019].

Rao, Vithala R. 2013. Methods for a large number of attributes. *Applied Conjoint Analysis*. Springer Berlin Heidelberg, 185–223.

Rossi, Peter E, Greg M Allenby, Robert McCulloch. 2005. *Bayesian Statistics and Marketing*. John Wiley and Sons Ltd.

Satomura, Takuya, Jaehwan Kim, Greg Allenby. 2011. Multiple-constraint choice models with corner and interior solutions. *Marketing Science* **30**(3) 481–490.

Scheibehenne, Benjamin, Rainer Greifeneder, Peter M. Todd. 2010. Can there ever be too many options? a meta-analytic review of choice overload. *Journal of Consumer Research* **37**(3) 409–425.

Schwartz, Barry. 2016. *The Paradox of Choice*. Harper Collins Publ. USA. URL *https://www.ebook.de/de/product/24767063/barry_schwartz_the_paradox_of_choice.html*.

Sloot, Laurens M., Dennis Fok, Peter C. Verhoef. 2006. The short- and long-term impact of an assortment reduction on category sales. *Journal of Marketing Research* **43**(4) 536–548.

Statista. 2018a. Anzahl der personen in deutschland, die innerhalb der letzten 14 tage tafelschokolade gekauft haben, von 2014 bis 2018 (in millionen). URL *https://de.statista.com/statistik/daten/studie/173658/umfrage/lebensmittel-schokoladenkonsum-der-bevoelkerung*. [Accessed online: 13-Jan-2019].

Statista. 2018b. Pro-kopf-absatz von schokoladenwaren in deutschland in den jahren 2010 bis 2021* (in kilogramm). URL *https://de.statista.com/statistik/daten/studie/72632/umfrage/pro-kopf-verbrauch-von-schokoladenwaren-in-deutschland*. [Accessed online: 13-Jan-2019].

Walsh, John W. 1995. Flexibility in consumer purchasing for uncertain future tastes. *Marketing Science* **14**(2) 148–165.

WirtschaftsWoche. 2018. Ritter sport auf wachstumskurs. URL *https://www.wiwo.de/unternehmen/industrie/starkes-auslandsgeschaeft-ritter-sport-auf-wachstumskurs/20895812.html*. [Accessed online: 13-Jan-2019].

Yang, Liu (Cathy), Olivier Toubia, Martijn G. de Jong. 2018. Attention, information processing and choice in incentive-aligned choice experiments. *Journal of Marketing Research* **55** 783–800.