International Journal of Forecasting 33 (2017) 433–441

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Structural forecasts for marketing data

Greg M. Allenby

Ohio State University, United States

ARTICLE INFO

Keywords:
Sparse data
Constraints
Statistical pooling

ABSTRACT

Marketing applications often require disaggregate forecasts of demand that pertain to subsets of individuals who are targeted for action. Examples include targeted price promotions that are made available through on-site couponing and forecasts of market segments for which new products have been developed. One challenge in the production of disaggregate forecasts of demand, and of consumer responses to marketing actions, relates to the limited amount of data that is available at the individual level. This paper discusses approaches to the improvement of marketing forecasts through the use of both parsimonious structural models of demand and random-effect models that pool data statistically across individual consumers.

1. Introduction

The functions of marketing within an organization are to represent existing and potential consumers and to help to guide product policy, including the development of new and existing goods and services, pricing, communication, and distribution activities within the firm. Marketing's interest in forecasting is related primarily to the prediction of the effects of various actions in terms of sales and demand. Examples of this include the effects of new product features on demand, changes in the pricing structure of an offering, advertising initiatives that aim to build brand recognition, and changes to channel policy, ranging from the use of existing dealerships to new electronic and digital venues. Because of the complexity of human behavior, marketing is also interested in predicting the effects of its expenditures on other variables that eventually lead to sales, such as brand recognition, recall, satisfaction and purchase intent.

Marketing forecasts are often context-dependent and disaggregate in nature. Forecasts pertain to specific brands in specific geographic regions, are designed to consider specific aspects of seasonality (e.g., fourth of July) and specific consumption occasions (e.g., backyard picnics), and are targeted at specific types of individuals. This is not to say that aggregate predictions of sales are unimportant to marketing, but that one of the primary goals of marketing is to forecast a demand that may not yet exist, which is generated from individuals who may not yet participate in the product category. It is important to understand the source of increased sales, because firms may have multiple offerings within a product category and wish to gain shares from specific competitors instead of from their own brands. Some marketing interventions are oriented toward growing the market, while others are merely reactions to the competition. The result of these factors is that assuming a naïve model for demand (e.g., an aggregate exponential smoothing model) is often inadequate. The micro-foundation of the forecast is important because the interventions being considered are thought to have specific effects on the consumer utility and sales.

Marketing forecasts are challenged by the nature of marketing data and the people from whom the demand is generated. People are heterogeneous in their preferences and sensitive to marketing variables such as prices. Their purchases are represented by 'lumpy' data, where the most frequently observed number is zero. The data cube

E-mail address: allenby.1@osu.edu.

http://dx.doi.org/10.1016/j.ijforecast.2016.09.003
0169-2070/

in marketing that corresponds to (respondents × products × time) is sparsely populated, in the sense that most people do not buy most products, visit most websites or attend to most attempts to get their attention. In fact, it is so costly to change an individual's nature that an overarching maxim in marketing is to "make what people will want to buy", not just to attempt to "make them want to buy". The issue of limited data at the individual level means that both point estimates and measures of the uncertainty are needed for forecasting, and the aim of understanding the determinants of the demand as well as their implications for market share and profitability means that traditional loss functions, such as the mean squared error of sales, are not always appropriate.

Moreover, the sparseness of marketing data at the individual level indicates that the data cannot speak for themselves, nor can accurate forecasts rely on the large-sample asymptotic properties of estimators based on descriptive models. The small-sample properties of estimators are more important, and there is a need to provide additional forecasting structure by using parsimonious and theoretically-grounded models of behavior and measurement. This paper describes constraints to model components, both at the individual level and across individuals, that "work", in the sense that they have been shown to lead to forecasting improvements (Allenby, Arora, & Ginter, 1995; Allenby & Rossi, 1998).

The organization of the paper is as follows. Individual-level models of demand based on a direct utility specification are presented in Section 2. These models distinguish utility generation from constraints, and are useful for understanding what is gained and what is given up in marketplace exchanges. Constraints include non-linear pricing, additional non-monetary constraints, and indivisibility due to packaging. Section 3 examines issues of heterogeneity, and contrasts Bayesian and non-Bayesian approaches to model specification. Covariates that are capable of describing the variation in preferences and sensitivities are discussed, along with models that are useful for detecting interactions. Section 4 concludes with a discussion of additional research topics that are useful for structuring models and data, and improving marketing forecasts.

2. Individual-level models of behavior

Over the last 20 years, empirical research in marketing has found that simple economic models of behavior work well at the individual level (Allenby, Fennell et al., 2005). There are several reasons for this conclusion. First, it is rare to have more than 20 or so observations about a particular construct of interest for an individual. Respondents can become fatigued when answering many questions about a single object (e.g., satisfaction scales, conjoint survey responses), and it is rare to have extended purchase histories of individuals in a particular category without there being some sort of change in the market, such as a new product intervention. Data limitations naturally restrict the complexity and parameterization of a model, and researchers have found that high-level interactions are rarely supported by the data once consumer heterogeneity is included in a model structure.

A second reason why simple models are favored is the real-world nature of forecasting. Marketing forecasts rely on individuals who have experience in a product category and are familiar with the offerings that are present. This is certainly true when using existing demand data, where consumers are paying for goods and services with their own money. In addition, forecasts based on stated preferences, as opposed to revealed preferences, are obtained using survey instruments in which respondents are screened so that they are included in the survey only if they are knowledgeable, interested and willing to purchase in the product category. This practice is in stark contrast to the behavioral studies in university settings that use undergraduate students without qualifications who participate for course credit. While many violations of simple economic models are documented in the academic literature, the expected effect sizes are usually difficult to determine because of the specialized nature of the sample (i.e., undergraduates) from which data are generated (see Frederick, Lee, & Baskin, 2014, Yang & Lynn, 2014).

Economic models of behavior assume that individuals are goal-oriented in their purchases (Chandukala, Kim, Otter, Rossi, & Allenby, 2008). That is, they are assumed to be engaged in some form of constrained (utility) maximization. This assumed process is supported widely by marketing data, with zero being the most frequently observed response by consumers. The fact that most people do not consume most offerings in a product category, coupled with the observation that consumers are often price sensitive, gives credence to the assumption of constrained utility maximization and the use of economic models of demand.

2.1. Economic models of demand

A simple model specification for demand is that of a discrete choice model:

$$\begin{aligned} \text{Maximize} \quad & u(x, z) = \psi' x + z \\ \text{subject to} \quad & p' x + z \leq E, \end{aligned} \tag{1}$$

where ψ_k is the marginal utility of consumption ($\partial u / \partial x_k$) of the kth alternative, which is assumed here to be constant, x is the vector of demand quantities, p is the price vector, E is the budgetary allotment and z is the unused budget, also known as the "outside" good. The marginal utility of the outside good, z, is normalized to one in order to identify the other parameters of the utility function. The price of the outside good is also assumed to be one (dollar), and can be thought of as a no-choice option when a consumer decides to save some of their budget.

Eq. (1) assumes that the marginal utility of an offering does not change with the quantity consumed, or that there is no satiation. Indifference curves, or isoquants, for this model are shown in Fig. 1, along with the budget constraint. Since both the utility and the budget are linear, the economic maximizing solution resides at the corner in which it is optimal to consume only one good. This 'corner' solution corresponds to a multinomial outcome that is consistent with traditional logit and probit

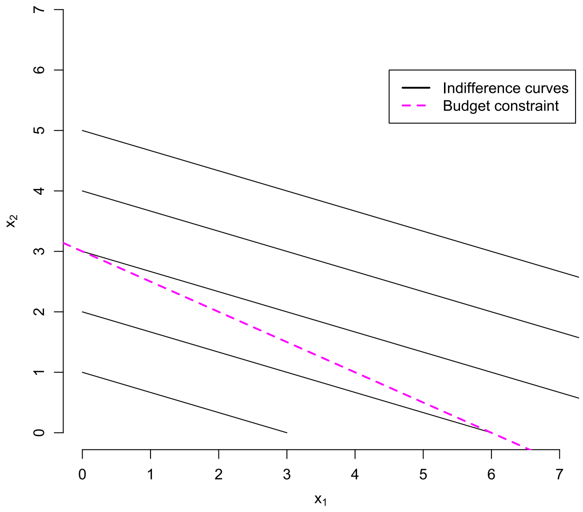

Fig. 1. Linear utility and linear budget.

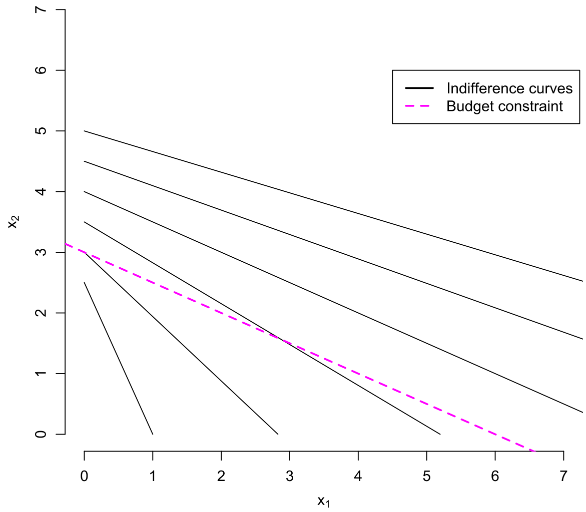

Fig. 2. Nonhomothetic utility and linear budget.

models of demand that are used extensively in marketing (McFadden, 2001).

The advantage of considering the economic underpinning of the logit/probit model is that it provides a basis for model extensions. One empirical finding in marketing demand forecasting is the asymmetric effect of prices, where a price discount of a superior good is observed to have a greater effect than that of an inferior good (Blattberg & Wisniewski, 1989). That is, more people respond to price discounts when an expensive good is on sale. One rationale for this observation is the idea that price discounts can be decomposed into what economists call substitution and income effects, with an income effect reflecting an increase in utility due to a new choice. Substitution effects, on the other hand, assume that any redistribution of demand because of changes in price resides along the same indifference curve.

One way of allowing economic-based forecasts to exhibit response asymmetries is to allow the indifference curves in Fig. 1 to fan out in the positive orthant. This can be accomplished by specifying an implicitly defined, rather than an explicitly defined, utility function (Allenby & Rossi, 1991):

$$\begin{aligned} &\text{Maximize} \quad u\,(x,z) = \psi\,(u)'x + z \\ &\text{subject to} \quad p'x + z \le E, \end{aligned} \tag{2}$$

where $\psi\,(u) = \exp(\alpha_j - \kappa_j u)$. For $\kappa_j > 0$, it can be shown that the indifference curves of this utility function will not cross. Moreover, the estimates of κ for each offering correspond to the relative rates of rotation of the indifference curves, with smaller values of κ being associated with superior goods and larger values of κ being associated with relatively inferior goods. For larger values of κ, increases in utility are associated with larger decreases in the marginal utility.

The indifference curves associated with Eq. (2) are shown in Fig. 2. The offering depicted on the vertical axis (x_2) becomes relatively more favorable as the budgetary allotment is relaxed, while the offering depicted on the horizontal axis (x_1) becomes relatively less favorable. Price discounts also induce a budgetary effect, in that consumers can either afford greater quantities of products, or achieve higher levels of utility.

Estimates of κ can be obtained from the data, providing analysts with an objective measure of the product quality. Two goods of similar quality will have similar estimates of κ, while superior goods will have smaller estimates of κ. In Fig. 2, x_2 is the superior good and x_1 is relatively inferior. As the budget constraint is relaxed, higher levels of utility are obtained by increasing x_2 at the expense of x_1. It should be remembered that shifts in budget can come from relative price discounts as well as from relaxing the budgetary allotment E. Price discounts induce both substitution and income effects, with the income effect favoring goods of higher quality. The predictions from this model provide parsimonious predictions of asymmetric price effects.

Another variant of the simple economic model allows for diminishing marginal returns to utility, or satiation. Satiation involves the introduction of non-linearity into the utility function, such that the marginal utility, or the derivative of the utility function with respect to the demand quantity (x), is a positive function that approaches zero asymptotically (Kim, Allenby, & Rossi, 2002). As was discussed by Allenby, Kim, and Rossi (2016), a simple specification that accomplishes this with the class of additively separable utility is:

$$u(x,z) = \sum_k \frac{\psi_k}{\gamma} \ln(\gamma x_k + 1) + z. \tag{3}$$

Indifference curves for this utility function are shown in Fig. 3. The curves are convex to the origin, and intersect the axis because "+1" is added as an offset in order to translate the logarithmic function. Thus, the utility function allows for both corner and interior solutions as consumers engage in constrained utility maximization.

The estimation of the utility model parameters proceeds by employing the Kuhn–Tucker conditions that are associated with constrained maximization, where the ratio of marginal utility to price (i.e., bang-for-the-buck) for an offering is compared to those of other offerings in the

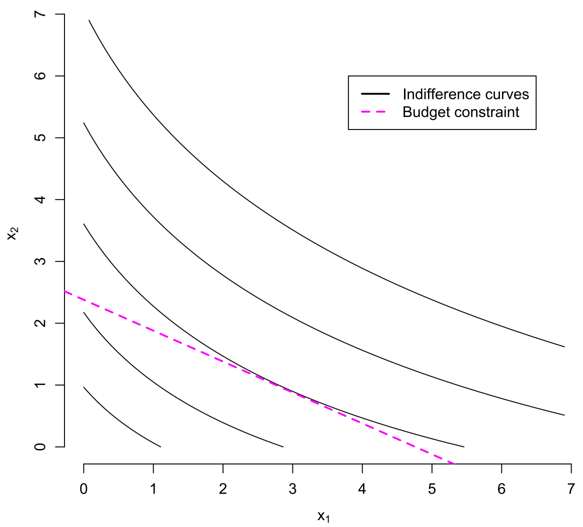

Fig. 3. Non-linear utility and linear budget.

choice set. The Kuhn–Tucker (K–T) conditions are related to the observed demand as follows:

$$\frac{u_i}{p_i} = \frac{u_k}{p_k} \quad x_i > 0 \text{ and } x_k > 0 \tag{4}$$
$$\frac{u_i}{p_i} > \frac{u_k}{p_k} \quad x_i > 0 \text{ and } x_k = 0,$$

where

$$u_k = \frac{\partial u(x, z)}{\partial x_k} = \frac{\psi_k}{\gamma x_k + 1} \tag{5}$$

is the marginal utility.

The Kuhn–Tucker conditions are intuitively plausible, in that, if consumers are constrained utility maximizers and are observed to purchase only one of the goods available, then the purchased good must offer a better marginal utility to price ratio (or bang-for-the-buck) than the other options available. If multiple goods are selected, then it is optimal for the bang-for-the-buck of the selected goods to be equal, and higher than those for the goods not selected. If the ratio of marginal utility to price were not equal, then a consumer could increase their demand for the good with the larger ratio, thus improving their satisfaction and utility.

Error terms can be introduced into the specification by regarding the baseline utility parameters, ψ, as random:

$$\psi_t = \exp[\psi + \varepsilon_t], \tag{6}$$

and this expression can be substituted into the K–T conditions. The presence of the budget constraint allows us to specify the outside good z without an error term, meaning that when the K–T conditions for each choice alternative are compared to those for the outside good, for which $u_z = 1$ and $p_z = 1$, we have:

$$\frac{u_k}{p_k} = \frac{\exp[\psi_k + \varepsilon_{kt}]}{p_{kt}(\gamma x_k + 1)} = 1 \tag{7}$$

or

$$\varepsilon_{kt} = g_{kt} \quad \text{if } x_k > 0 \tag{8}$$
$$\varepsilon_{kt} < g_{kt} \quad \text{if } x_k = 0,$$

where

$$g_{kt} = -\psi_k + \ln(\gamma x_k + 1) + \ln(p_{kt}). \tag{9}$$

The assumption that consumers are constrained utility maximizers leads to a likelihood relationship for the observed demand that is of the form:

$$\Pr(x_t) = |J_{R_t}| \left\{ \prod_{i=1}^{R_t} f(g_{it}) \right\} \left\{ \prod_{j=R_t+1}^{N} F(g_{jt}) \right\}, \tag{10}$$

where $i = 1, \ldots, R_t$ denotes the demand quantities that are positive, or in the interior, and $|J_{R_t}|$ is the Jacobian of the transformation from the model error term to x_i, the positive demand quantities (see Allenby et al., 2016).

Prediction is more difficult when using this model than when using models based on ARIMA or regression, where there is an explicit relationship between the input and output variables. The demand based on Kuhn–Tucker conditions requires an iterative algorithm that solves for the values of demand. This can be accomplished by using either pre-programmed optimization software such as "optim" in the statistical package *R*, or a recursive algorithm that iterates toward an optimal solution.

2.2. Constraints

Extensions to models involving constrained utility maximization can come from two different sources: extensions to the utility function or the constraints that people face during their pursuits. The field of marketing has tended to focus on the aspects of utility formation, without much consideration of the constraints. One of the main reasons for this is that, while the Kuhn–Tucker conditions provide a simple and elegant solution to constrained utility maximization for a simple price constraint, the incorporation of non-linear budget constraints and other realities that are present in the marketplace is more difficult. The actual constraints faced by consumers can be non-linear in nature, may come from multiple sources and are often related to the supply-side decision regarding what is available in the marketplace.

2.2.1. Non-linear pricing

Non-linear pricing is common both when a firm is attempting to induce trials by offering discounted prices up to a limited quantity, and when they are attempting to induce greater consumption by providing quantity discounts. In either situation, non-linear pricing creates kinks in the budget constraint at the point at which prices change (Howell, Lee, & Allenby, 2015). Fig. 4 shows the effect of a limited quantity discount, where the price is reduced for the first two units of the good, and the unit price is higher for additional purchases.

The limited quantity discount creates an outward kink at two units (i.e., $\tau = 2$), leading to a build-up of probability mass that is not located at a corner. That is, while the previous models were characterized by likelihoods that were a combination of density and mass point contributions, the density contribution to the likelihood was from interior solutions to the Kuhn–Tucker conditions (e.g., $x_i > 0$ and $x_k > 0$), whereas the mass point contribution to the

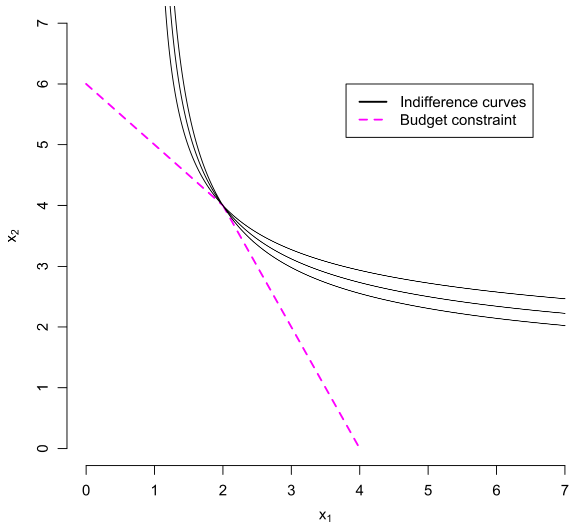

Fig. 4. Non-linear utility and a kinked budget.

likelihood was associated with corner solutions (e.g., $x_i > 0$ and $x_k = 0$). The presence of kinks in the budget leads to a mass buildup because there many different realizations of the error term lead to the same utility maximizing solution. Fig. 4 shows three indifference curves for which the kink point is the utility maximizing solution.

The budget constraint in the case of a limited price discount can be written as

$$\sum_i p_{li} \min(x_i, \tau_i) + p_{hi} \max(0, x_i - \tau_i) + z = E, \tag{11}$$

where "i" denotes the brand, p_{li} is the discounted price up to the kink point τ_i, and p_{hi} is the price for higher levels of demand.

The likelihood for choices where there are kinks in the budget set due to price discounts can be derived by dividing the positive orthant into regions in which the Kuhn–Tucker conditions can be applied, and then comparing the solutions in each region in order to obtain the maximum across regions. For the case of two goods and two prices, this leads to four regions that are defined by whether x_1 and x_2 are above or below the cutoffs τ_1 and τ_2. The Kuhn–Tucker conditions for this setting become:

$$\begin{aligned} \varepsilon_{kt} &< g_{\ell kt} && \text{if } x_k = 0 \\ \varepsilon_{kt} &= g_{\ell kt} && \text{if } 0 < x_k < \tau_k \\ g_{\ell k} < \varepsilon_{kt} &< g_{hkt} && \text{if } x_k = \tau_k \\ \varepsilon_{kt} &= g_{hkt} && \text{if } x_k > \tau_k, \end{aligned} \tag{12}$$

where:

$$\begin{aligned} g_{\ell kt} &= -\psi_k + \ln(\gamma x_k + 1) + \ln(p_{\ell kt}) \\ g_{hkt} &= -\psi_k + \ln(\gamma x_k + 1) + \ln(p_{hkt}), \end{aligned} \tag{13}$$

and the likelihood is a combination of mass points and densities that depend on the demand quantities.

2.2.2. Multiple constraints

In addition to the presence of non-linearities in the budget set, consumers may also be faced with non-monetary constraints. Additional constraints might be related to the amount of space in one's refrigerator or one's home, or the amount of salt in an offering; and performance constraints such as tradeoffs among product features. For example, the performance of a computer might be affected by both the amount of memory and the speed of the processor.

Satomura, Kim, and Allenby (2011) propose a method for introducing multiple constraints into the basic model:

$$\text{Maximize} \quad u(x, z, w) = \sum_k \frac{\psi_k}{\gamma} \ln(\gamma x_k + 1) + z + w \tag{14}$$

$$\text{subject to} \quad p'x + z \le E \quad \text{and} \quad q'x + w \le Q,$$

where separate slack variables (z, w) are introduced for each constraint. The resulting Kuhn–Tucker conditions become:

$$\begin{aligned} \varepsilon_k &= g_k \quad \text{and} \quad x_k > 0 \\ \varepsilon_k &< g_k \quad \text{and} \quad x_k = 0, \end{aligned} \tag{15}$$

where

$$g_k = -\psi_k + \ln(\gamma x_k + 1) + \ln(p_k) + \ln(q_k). \tag{16}$$

It is important to point out that this model specification is meant to be illustrative, and has somewhat restrictive implications for demand quantities, in that the demand is a function of own price only, and does not involve cross-price effects. A more realistic specification with cross-price effects can be obtained by specifying the outside goods (z, w) with logarithms in the utility function. Non-linearities in the utility specification lead to a richer pattern of price competition (see Allenby et al., 2016, Satomura et al., 2011).

The constraints in an economic model of demand indicate choice boundaries that can be thought of in a manner similar to the presence of screening rules that lead to localized choice sets, also known as consideration sets. Consideration sets have been proposed as a device for narrowing down the set of goods for analysis from the universal set of alternatives (Gilbride & Allenby, 2004; Kohli & Jedidi, 2007). Many decisions involve too many decision options for them all to be included in a model, meaning that some mechanism is needed to reduce the domain from which choices are made. The consideration set comprises the choice alternatives that are not screened out from analysis and have some positive probability of being consumed.

Additional constraints can also be introduced into the model to serve as criteria for consideration. That is, goods that do not have the attribute or particular combinations of attributes would be screened from inclusion as options of choice. Thus, traditional models of consideration and choice can be thought of as special cases of a multiple constraint model that involves just one attribute or product feature at a time. It is assumed that consumers will not necessarily evaluate the marginal utility of alternatives that are ruled out because of constraints.

Ruling out specific items from the choice, by using either models of consideration or constraints in choice models, can have both advantages and disadvantages. The advantages arise from the choice set being reduced and the model likelihood being concentrated on a specific set of products. This is particularly helpful when one is modeling the demand for many items, a situation that is encountered in most product categories, due to the proliferation of choice options. Traditional forecasting models assign a separate error term to each of the choices available in the marketplace. When one is predicting the demand for items that are in competition with each other, the presence of many error terms (e.g., >50) degrades the model fit and diminishes the forecasting accuracy (Allenby, Brazell, Gilbride, & Otter, 2005). Ruling out specific choice options reduces the number of error terms in the model, but potentially at the expense of predicting zero demand if the model under consideration is inaccurate.

2.2.3. Indivisibility

A third form of constraint that impacts marketing forecasts is the fact that the demand, at least at the individual level, is often restricted to lie on a grid of available package sizes. It is not possible to purchase nine bottles of beer in most stores, although it is certainly possible to purchase a 6-pack or 12-pack package. Nearly all of the goods that are available in the marketplace come in pre-specified sizes, and demand estimates often ignore these constraints. While it is true that aggregated demand statistics often mask the coarse granularity of the individual-level demand, the generation of forecasts from individual-level data becomes problematic unless marketplace packaging is taken into consideration.

One issue that is at the core of deciding whether demand discreteness or indivisibility requires consideration is the question of whether to first aggregate the data and then conduct the analysis, or to conduct the analysis at the individual level and then aggregate the results. The answer to this question is determined largely by the purpose of the analysis and the decisions being contemplated. If the analysis informs a decision that is aggregate in nature and the goal of analysis is purely prediction, then analysis can often proceed using a model that is calibrated on aggregate data. However, if the analysis informs a disaggregate decision, the forecast must address the fact that demand in the marketplace is discrete.

To illustrate this issue, consider the need to predict the effects of a change in package design for a branded offering. Let us assume that the new package will be the same for everyone; that is, the same logo, brand claim, text and colors will be used regardless of the retailer that offers the item or the part of the country in which it is sold. In this case, consumer heterogeneity can be ignored because the unit of analysis is not the individual consumer but the market in general (Joo, Thompson, & Allenby, 2016). However, if the point of the analysis is to engage in some form of targeting at a disaggregate level, this disaggregate decision can be supported only by using a disaggregated model.

However, it should be remembered that a more disaggregate level of analysis always exists in any domain, and a certain level of aggregation must be made at some point in any analysis. Predictive models using ARIMA time series techniques (Box & Jenkins, 1976) can always be considered to arise from a less aggregate model in which the coefficients are "unpacked" or explained by some more fundamental story. The same is true of the models that do this unpacking; i.e., there is always a more fundamental story involving physical and psychological principles playing out that truly gives rise to the observed data.

At some point, the analyst needs to make an assumption of stationarity, or stability, which implies that the correlational structures of yesterday will also be present tomorrow. When conducting a regression analysis, stationarity assumptions are made about the error term (ε) but not about the dependent variable (y). The regression model is one example of a model that attempts to 'explain' the variation in y based on the variation in other independent variables X. In demand models, the error term is usually viewed as representing unobserved factors that affect the utility that consumers have for offerings. Thus, the implicit assumption is made that these factors that affect the demand are stationary over time; e.g., the context of usage, the motivation for consumption and the problems being solved by the individual that give rise to product demand.

It is important to deal with demand indivisibility when the decisions being considered are affected by it. For example, a firm might be interested in predicting the demand for the introduction of a new package size (e.g., a smaller container than that currently available). The presence of packaging constraints and other forms of indivisibility makes model estimation based on first-order conditions such as the Kuhn–Tucker conditions problematic. It can no longer be assumed that the Kuhn–Tucker conditions hold exactly at the observed levels of demand. Lee and Allenby (2014) propose an alternative estimator that does not rely on first-order conditions, and allows for the realized demand not necessarily adhering to the conditions described by Eq. (8). Instead, the selected offering is the best of the available offerings in the market. An aggregate model that does not consider the effects of package sizes explicitly cannot make such counter-factual forecasts.

3. Models of heterogeneity

Marketing as a field of study exists because of heterogeneity. If all people were the same in terms of their preferences and sensitivities to variables such as prices, there would be no need for market segmentation or brand positioning. There would be no reason for product lines because one brand would be capable of fulfilling the needs of all individuals. Heterogeneity gives rise to target offerings and niche products where firms attempt to establish local monopolies by tailoring their products to the needs of a specific group of consumers. Conversely, the variety of brands and sizes available for sale in nearly every product category demonstrates the diversity of tastes and preferences present in the market.

One of the things that distinguishes marketing from economics as a separate discipline is its interest in how people differ in their reasons for purchasing goods and services. Marketing is charged with the task of guiding product policy to make what people will want to buy, rather getting people to buy products that have already been made. Marketing operates at a disaggregate level in the marketplace, and embraces the distribution of heterogeneity of behavior in an ex-ante analysis of consumer behavior. In contrast, economics' ex-post analysis assumes the existence of the offering and tends to focus on expected outcomes for the economic system as a whole. Thus, economics tends to view heterogeneity as a nuisance that needs to be "controlled for" but not necessarily studied.

Marketing's interest in disaggregate groups of individuals, and even specific individuals in direct marketing applications, is challenged by the lack of data at the individual level. The field of marketing has coped with this information shortfall by embracing Bayesian hierarchical models of random effects that simultaneously assume that no individual is completely different from all others and that no two individuals are exactly the same (Allenby & Rossi, 1998). Instead, individuals' tastes, preferences and sensitivities can be summarized using some probability distribution among people:

$$\psi_i \sim Normal\left(\bar{\psi}, V_\psi\right), \tag{17}$$

where ψ_i denotes individual i's vector of parameters, $\bar{\psi}$ is the mean of the random-effects distribution, and V_ψ is the covariance matrix.

Denoting the likelihood of the data for individual i coming from one of the models in Section 2 by $\ell\left(x_{it}|\psi_i\right)$, a Bayesian analysis proceeds by specifying a prior distribution in order to arrive at the posterior. In a panel data setting, where there are $i = 1, \ldots, N$ individuals being modeled, the likelihood from N individuals takes the form:

$$\ell\left(\{x_{it}\} \mid \{\psi_i\}, \bar{\psi}, V_\psi\right) = \prod_i \prod_t \ell\left(x_{it}|\psi_i\right) \times \pi\left(\psi_i|\psi, V_\psi\right), \tag{18}$$

and priors are introduced for the hyper-parameters $\bar{\psi}$ and V_ψ. Bayes analysis proceeds by forming the posterior distribution of the data:

$$\pi\left(\{\psi_i\}, \bar{\psi}, V_\psi \mid \{x_{it}\}\right) \propto \prod_i \prod_t \ell\left(x_{it}|\psi_i\right) \times \pi\left(\psi_i|\bar{\psi}, V_\psi\right) \times \pi(\bar{\psi}, V_\psi). \tag{19}$$

Modern simulation-based estimation using Monte Carlo Markov chains yields a mechanism for generating draws from the entire posterior distribution. The availability of the posterior distribution allows the analysis to reflect all uncertainty in the parameters fully, not just to rely on point estimates, which are problematic in small samples. This model specification is in contrast to classical (non-Bayesian) estimators, which form the likelihood by integrating out the random effects:

$$\ell\left(\{x_{it}\} |\bar{\psi}, V_\psi\right) = \int \prod_i \prod_t \ell\left(x_{it}|\psi_i\right) \times \pi\left(\psi_i|\bar{\psi}, V_\psi\right) d\{\psi_i\}. \tag{20}$$

Bayesian analysis of the random-effects model retains the individual-level parameters $\{\psi_i\}$ for analysis. This is important for marketing because of its interest in understanding particular individuals and groups of individuals, which are not represented well by the parameters $\bar{\psi}$ and V_ψ.

Eq. (17) can be generalized to include covariates by specifying a multivariate regression model in place of the mean:

$$\psi_i \sim \text{Normal}(\Gamma' h, V_\psi), \tag{21}$$

where Γ is a matrix of regression coefficients and h denotes covariates that are of interest for forecasting and inference.

3.1. Covariates (h)

Previous research has shown that multi-level models such as those considered in this paper lead to improved forecasts when lower-level constructs, such as marginal utility, satisfaction and consideration, are incorporated into aggregate forecasts (Fok, Van Dijk, & Franses, 2005; Paap et al., 2005). However, one challenge of using Eq. (21) to help in understanding the correlates of the distribution of heterogeneity relates to the identification of appropriate covariates, h. It has been found that demographic and general psychographic variables are not very explanatory of brand preferences (Fennell, Allenby, Yang, & Edwards, 2003), although many of these variables could certainly be thought of as being independent variables that are consistent with the assumptions of a regression model.

One reason why demographic variables explain little of the distribution of heterogeneity is that they are often much broader in scope than the preferences and sensitivities they are attempting to explain. Demographic variables describe a person regardless of the context; for example, a person's age is exactly the same whether they are at work, at home, in the gym or on a golf course. On the other hand, preferences are context-dependent, and even the amount of money that a person is willing to spend in a product category will depend on whether the product is for themselves or others, whether it is for social or private consumption, and whether the individual is making a routine or new decision. The context of choice resides at a finer level of granularity than a demographic variable, which explains why demographics do a relatively poor job of predicting preferences.

Recent work on the explanation of consumer behaviors has focused on a search for other variables that are correlated, typically other purchases or actions of the consumer. This works well if the factors that caused the correlation yesterday are expected to be present tomorrow, but does not indicate how a consumer can best be appealed to or suggest any reason why a person may have acted in the first place. In general, if people who purchase one product also purchase another product, they may be qualified as candidates for cross-selling without knowing why the products are appealing. However, an understanding of why a person might prefer one brand to another requires a more in-depth analysis.

The search for more explanatory variables that can be used to predict consumer preferences using survey data is challenged by the lack of coherence in the survey responses. Fixed point ratings scales are often used to elicit respondent agreement and disagreement with statements in order to assess a person's beliefs and attitudes in relation to aspects of the purchase decision. The problem with the use of fixed point ratings data is that consumer responses reflect relative reactions not absolute reactions, and lack commonalities across people. That is, one person's "five" on a seven-point scale may be another person's "seven". Such differences in scale responses can be due to many factors, such as social norms and personal sensitivities. There is no gold standard for answers to the questions such as "How satisfied are you with your purchase?" or even "How much does your significant other love you?" regardless of the labels assigned to each of the response points (Büschken, Otter, & Allenby, 2013; Rossi, Gilula, & Allenby, 2001).

The lack of a common scale among respondents prevents the establishment of a predictive relationship in the upper level of the hierarchical Bayes model. There is evidence of success from using a simplified, two-point scale where respondents are asked whether or not a statement applies to them (Allenby & Brazell, 2016). A two-point scale minimizes the yea-saying and nay-saying that are associated with larger numbers of scale points. The average strength of the relationship between the independent (h) and dependent (ψ) variables is then measured by the coefficients in Γ.

3.2. Interactions

The identification of interactive effects is another source of frustration in predicting why some people are sensitive to product features and others are not. Interactions are costly to incorporate in the upper level of a HB model, where the coefficient matrix (Γ) is of dimension $\dim(\psi) \times \dim(h_i)$. For each interaction covariate that is added to the vector h, an additional $\dim(\psi)$ coefficients are added to Γ.

Recent advances in text analysis provide a possible solution to the investigation of high-level interactions in data. Text can be represented in a model using discrete multinomial distributions for word selection, with a vocabulary that is represented by a vector of word probabilities (Blei, Ng, & Jordan, 2003). The content of a document can be described further based on topics that are characterized by different word vectors, or vocabularies, that differ in their likelihoods of using particular words. A common model for text data is the latent Dirichlet association (LDA) model, which assumes that a document can be characterized by a probability distribution over topics, with each topic assigning its own probabilities to words. Words in a document are generated by first drawing a topic from the topic distribution, then drawing a word from the associated vocabulary with a topic-specific probability.

The LDA model introduces interactions to a model for discrete data in a parsimonious way. By allowing the word probabilities to be topic-specific, the model economizes on parameters to allow for interactions. A vocabulary of W words and T topics requires $W \times T$ parameters instead of W "choose two", or $W(W-1)/2$ for all two-way interactions, which is large when $W \gg T$. The same idea can be extended to a model of questionnaire responses that consists of responses to W pre-determined statements instead of words. Another model that is related to the LDA model, the "Grade of Membership" (GoM) model (Airoldi, Blei, Erosheva, & Fienberg, 2014), makes adjustments to the LDA model to allow for the fact that no statement can be selected more than once.

The grade of membership model derives its name from its model structure, where the probability masses associated with the discrete are modeled as a finite mixture across the membership profiles (λ):

$$Pr(w_{i,j} = l | g_i, \lambda) = \sum_{k=1}^{K} g_{i,k} \lambda_{j,k}(l), \tag{22}$$

where j indexes the questionnaire items, i indexes the respondent and k indexes a vector of archetype responses. Each respondent is assumed to comprised a mixture of these archetypes, with g indicating the probability of each. Thus, the GoM model introduces response interactions in a manner similar to the LDA model, by expanding the set of archetypes instead of searching for interaction effects directly. Dotson, Büschken, and Allenby (2016) explain heterogeneity by incorporating a GoM model into the upper level of a HB choice model.

4. Summary

Individual-level forecasts are of interest for marketing whenever an intervention that is being considered is directed toward a subset of individuals in a market for whom their predicted response is of interest. The primary challenge when producing marketing forecasts is in dealing with data sparsity at the individual level. While marketing data are frequently characterized as being large, their structure is actually sparse, in that the data cube defined by consumers, products and time is mostly empty. This paper reviews recent advances in the use of theory-based models of choice that help us to move away from descriptive models that tend to be overparameterized. In addition, hierarchical Bayes random-effect models provide an additional source of information for the production of market predictions through the sharing of information across respondents when making individual-level estimates.

The idea of imposing more structure on forecasts in order to address the issue of data sparseness is not new. Structure is imposed in even the most descriptive of models when deciding which covariates to include in the model specification and which to exclude, or when determining the functional form of a model. Structure is also imposed by selecting functional forms that restrict coefficients' algebraic signs, such as restricting a price coefficient to be negative. These aspects of modeling introduce strong prior information into an analysis. The individual-level models described above provide another class of models to consider, and the models of heterogeneity can be thought of as another way of introducing information through a hierarchical prior where information is shared across people.

Additional research is needed to compare forecasts from the hierarchical Bayesian models described above to the conventional models that are used in marketing. Having an understanding of the conditions under which a micro-foundation is and is not helpful for forecasting would help practitioners to identify when they should be used. High-dimensional forecasts are likely to benefit from the presence of strong theory, while lower-dimensional forecasts may be better with an aggregate model without a strong theoretical foundation. The establishment of boundary conditions for the proposed models in order to achieve improvements in forecasting and inference would be a welcome addition to the forecasting literature.

Outside the class of models discussed above, rational constraints are often found to lead to improved predictions and inferences. For example, while the analysis of textual data does not necessarily lend itself to analysis by models with an economic foundation, the forecasting abilities of such models can be shown to improve when they are constrained to be less flexible in the face of data limitations. For example, the standard topic model assumes that words are generated by topics that can change freely across words. Constraining the topics to be sentence dependent, or dependent on commas (,) and other punctuation present in the sentence, can be shown to lead to improvements in predicting ratings (Büschken & Allenby, in press). The structure imposed by any theory, even if it is not tightly derived, can provide valuable information for the improvement of forecasts in sparse-data environments such as marketing.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2016.09.003.

References

Airoldi, E. M., Blei, D., Erosheva, E. A., & Fienberg, S. E. (2014). *Handbook of mixed membership models and their applications*. CRC Press.

Allenby, G. M., Arora, N., & Ginter, J. L. (1995). Incorporating prior knowledge into the analysis of conjoint studies. *Journal of Marketing Research*, 152–162.

Allenby, G.M., & Brazell, J.D. (2016). Seven summits of marketing research.

Allenby, G., Brazell, J., Gilbride, T., & Otter, T. (2005). Avoiding IIA meltdown: Choice modeling with many alternatives. In *Proceedings of the Sawtooth Conference*, Vol. 2004, (pp. 209–218).

Allenby, G., Fennell, G., Huber, J., Eagle, T., Gilbride, T., Horsky, D., et al. (2005). Adjusting choice models to better predict market behavior. *Marketing Letters*, *16*(3–4), 197–208.

Allenby, G. M., Kim, J., & Rossi, P. E. (2016). Economic models of choice. In B. Wierenga, & R. van der Lands (Eds.), *Handbook of marketing decision models*. Springer.

Allenby, G. M., & Rossi, P. E. (1991). Quality perceptions and asymmetric switching between brands. *Marketing Science*, *10*(3), 185–204.

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

Blattberg, R. C., & Wisniewski, K. J. (1989). Price-induced patterns of competition. *Marketing Science*, *8*(4), 291–309.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control* (revised ed). Holden-Day.

Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, (in press).

Büschken, J., Otter, T., & Allenby, G. M. (2013). The dimensionality of customer satisfaction survey responses and implications for driver analysis. *Marketing Science*, *32*(4), 533–553.

Chandukala, S. R., Kim, J., Otter, T., Rossi, P. E., & Allenby, G. M. (2008). Choice models in marketing: Economic assumptions, challenges and trends. *Foundations and Trends® in Marketing*, *2*(2), 97–184.

Dotson, M.R., Büschken, J., & Allenby, G.M. (2016). Explaining preference heterogeneity with mixed membership modeling, Available at SSRN 2758644.

Fennell, G., Allenby, G. M., Yang, S., & Edwards, Y. (2003). The effectiveness of demographic and psychographic variables for explaining brand and product category use. *Quantitative Marketing and Economics*, *1*(2), 223–244.

Fok, D., Van Dijk, D., & Franses, P. H. (2005). Forecasting aggregates using panels of nonlinear time series. *International Journal of Forecasting*, *21*(4), 785–794.

Frederick, S., Lee, L., & Baskin, E. (2014). The limits of attraction. *Journal of Marketing Research*, *51*(4), 487–507.

Gilbride, T. J., & Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*, *23*(3), 391–406.

Howell, J. R., Lee, S., & Allenby, G. M. (2015). Price promotions in choice models. *Marketing Science*, *35*(2), 319–334.

Joo, M., Thompson, M.L., & Allenby, G.M. (2016). Optimal product design by sequential experiments in high dimensions, Available at SSRN.

Kim, J., Allenby, G. M., & Rossi, P. E. (2002). Modeling consumer demand for variety. *Marketing Science*, *21*(3), 229–250.

Kohli, R., & Jedidi, K. (2007). Representation and inference of lexicographic preference models and their variants. *Marketing Science*, *26*(3), 380–399.

Lee, S., & Allenby, G. M. (2014). Modeling indivisible demand. *Marketing Science*, *33*(3), 364–381.

McFadden, D. (2001). Economic choices. *The American Economic Review*, *91*(3), 351–378.

Paap, R., van Nierop, E., Van Heerde, H. J., Wedel, M., Franses, P. H., & Alsem, K. J. (2005). Consideration sets, intentions and the inclusion of "don't know" in a two-stage model for voter choice. *International Journal of Forecasting*, *21*(1), 53–71.

Rossi, P. E., Gilula, Z., & Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *Journal of the American Statistical Association*, *96*(453), 20–31.

Satomura, T., Kim, J., & Allenby, G. M. (2011). Multiple-constraint choice models with corner and interior solutions. *Marketing Science*, *30*(3), 481–490.

Yang, S., & Lynn, M. (2014). More evidence challenging the robustness and usefulness of the attraction effect. *Journal of Marketing Research*, *51*(4), 508–513.

Greg M. Allenby is the Kurtz Chair in Marketing at Ohio State University. He is a Fellow of the INFORMS Society for Marketing Science and the American Statistical Association. He is also the 2012 recipient of the AMA Parlin Award for his contributions to the field of marketing research and a recipient of the ISMS Long-Term Impact award. He is past editor of *Quantitative Marketing and Economics*, and past area/associate editor for *Marketing Science*, *Journal of Marketing Research*, and the *Journal of Business and Economic Statistics*.