Overview: Customer Choice Modelling for Demand Management

There are several business decisions that need to be made before a product or service can be released to the public. At first glance, it may seem that the decisions made by a small business (e.g. a market trader) and the decisions made by a multinational airline would never have anything in common. However, the choices made by both of these businesses include, but are not limited to, pricing strategies, product supply and allocation, and year-round scheduling. A market trader looks to exploit the Christmas season at the same time that airlines would be preparing for one of the busiest travelling periods of the year. The complexity of these decisions scales up in accordance with the size of the business.

Business decisions can be either viewed from the perspectives of sales or from the perspective of demand. A sales-based perspective involves determining where and when to sell products and at what price. Whereas, a demand-based perspective involves understanding consumer preferences, estimating customer demand, and predicting ahead. In this report, we focus on understanding the latter viewpoint.

The goal of any business is to maximise profit. *Demand management* (also known as revenue management) refers to the collection of scientific practices used by businesses to manage demand for their products and services. As a discipline, demand management was first implemented in the US aviation industry during the 1970s. Over time, the practice has become enshrined within mainstream business, fuelled by an active interest within the operational research community. Today, demand management is actively used across several different sectors, including hospitality, advertising, and transportation.

Demand estimation is a central building block within demand management. Accurate demand forecasts can then be used as inputs for inventory control and pricing strategies. Over the past decade, there has been a shift from *independent demand* modelling to choice-based, *dependent demand* modelling. The traditional approach assumed that each product had its own demand segment. In this context, *demand segmentation* (also known as *market segmentation*) is the process of dividing a consumer base into subgroups according to certain characteristics, such as lifestyle or common interests.

In most cases, the *independent demand* assumption is unrealistic. To illustrate this, let's consider the demand for airline tickets with a specific operator. The probability of the airline selling a full-price ticket depends on whether or not they are offering a discount at the same time. Similarly, the probability that a potential customer buys a ticket at all from the airline depends on the cheapest fare offered. When a ticket discount runs out, the situation where customers choose to buy a more expensive ticket with the same airline is known as a *buy-up*. The situation where customers choose an alternative flight with the same airline is known as a *diversion*.

This kind of customer behavior has important consequences on the demand management practices of a business and should be considered carefully when making decisions. If we instead assume *dependent demand*, we are able to acknowledge the interactions and dependencies within products. Then, the demand for a product can be thought of as a function of the set of choices provided to customers.

A dependent demand framework must also be capable of anticipating and providing for a wide variety of customer preferences. To motivate this, we consider the simple example of a cinema chain. Let us suppose that there are two cinemas in a city, which each show different types of films. The first cinema is located in the city centre and consequently charges a premium price for its tickets. The second cinema, which is located in the suburbs of the city, offers cheaper prices. In a given week, the first cinema is showing *Thor: Ragnarok* and the second cinema is showing *Darkest Hour*. The population of customers is two-thirds students and one-third working professionals. Ultimately, the goal of the cinema chain is to maximise its revenue, by allocating its film showings to the screens with the most appropriate capacities. In order to do this, it needs a realistic idea of customer demand across the entire week.

In this report, we provide a technical overview of the different approaches available for modelling customer choice behaviour. We focus on three categories of techniques: *parametric, non-parametric* and *multi-stage*. Whilst parametric models are well-defined, they tend to make several assumptions about customer preferences. Recently, with the rise of business analytics, non-parametric models have been gaining more traction because of their flexible nature. We also discuss the use of choice modelling within the wider context of *availability control* problems.

Customer Choice Modelling in Demand Management

Author: Srshti Putcha   *Supervised by:* Dr. Catherine Cleophas

1 Introduction

In the OR community, *demand management* (alternatively known as *revenue management*) is the collective name for the family of scientific strategies that are used by businesses to manage demand for their products and services (van Ryzin and Talluri, 2005). As a discipline, demand management (DM) was first implemented in the US aviation industry during the 1970s. Over time, the practice has become enshrined within mainstream business, fuelled by an explosion in methodological research. Today, DM is actively used across several different sectors, including hospitality, advertising, and transportation (Strauss et al., 2018).

Demand estimation is an important task within DM. Accurate demand estimates are needed as inputs for price optimisation and inventory control problems (van Ryzin and Vulcano, 2017). There has been a focus in the literature on *independent demand* modelling up until recently. It was popular to assume that demand for a particular product is independent of the demand for any other product. In other words, a product can only cater to the needs of a specific demand segment, without any overlap (see left-hand diagram in Figure 1). If a product is dropped, then the demand segment disappears.

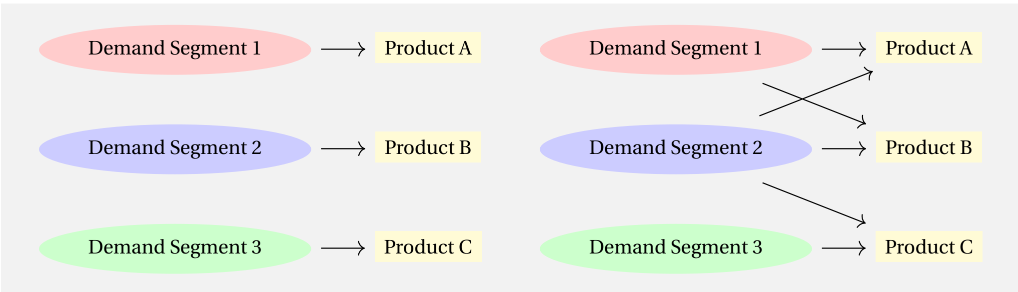

Figure 1.1: A comparison of the *independent demand* (L) and *dependent demand* (R) assumptions.

In general, the independent demand assumption is reasonable in quasi-monopolistic scenarios, where product offerings are not substitutable. For instance, airline fares used to be categorised by purchase type (advance purchase, refundable, cancellation etc.), in order to segment the customer base. However, with the advent of online travel and fare comparison sites, there is greater competition within the market and more information available to consumers, rendering the independence assumption invalid (Strauss et al., 2018).

Customers are now able to instantaneously compare flights across several different carriers, before making their final purchase decision. If the availability of the cheapest ticket type is limited, a customer may opt for the next-best ticket fare (buy-up) or they may switch to a different flight altogether (diversion) (van Ryzin and Talluri, 2005).

Therefore, it is far more practical to try to model the dependencies between products (see right-hand diagram in Figure 1). *Dependent demand* models incorporate customer preferences and behaviour into the framework of the problem. Then, the demand for a product can be thought of as a function of the set of choices provided to customers.

Heuristic (e.g. *buy-up factors*) and exact methods (e.g. *choice models*) have been developed for analysing and modelling customer behaviour (van Ryzin and Talluri, 2005). This paper largely focuses on reviewing discrete choice models for DM (see Section 3). Choice-based modelling is often considered to be a sub-problem of the wider topic of *availability control* over a finite time horizon. We provide a discussion of availability control in Section 2. Further details can be found in the review paper provided by Strauss et al. (2018). At the end of the paper, we provide a brief overview of the topic suitable for a non-technical audience.

Since the late 1990s, there has been a great deal of interest in customer-choice behaviour and demand management from both an OR perspective and a marketing perspective. Consequently, there is a vast number of relevant papers within the management science field. As the scope of the survey is limited, we have chosen to concentrate on introductory review papers and books to build the main points of our discussion. Our main literature sources are Strauss et al. (2018), van Ryzin and Talluri (2005), and Train (2009). Where appropriate, we reference specific journal papers that have made interesting and notable contributions to the body of literature. In particular, we discuss the contributions of Talluri and van Ryzin (2004), Jagabathula and Vulcano (2018) and van Ryzin and Vulcano (2015) in detail.

2 Choice-based availability control

In this section, we provide a general framework for the *choice-based network* problem of availability control within DM and then go on to briefly review the problem of *assortment optimisation.* For consistency, we mainly use the notation outlined in Strauss et al. (2018). The formulation of the simpler *single-resource* scenario, as proposed by Talluri and van Ryzin (2004), is discussed at the end.

2.1 General framework: choice-based network problem

2.1.1 Notation

Over the course of a *selling horizon* (also known as a *booking horizon*), we assume that a business provides a finite set of product choices to the customers arriving over time. When the firm is producing services for consumers, there will be constraints over both the resources that can be used and the capacities for delivering those services. It is generally assumed that there are large fixed costs for providing services. In this case, revenues (i.e. total income) can serve as an indicator for overall profit.

The firm seeks to maximise revenue by performing demand management. In order to this, it must take heterogeneous customer preferences into account. Each individual has a different willingness-to-pay and will ultimately make different choices. Therefore, the firm has to prepare for the upcoming selling horizon carefully, by deciding which products to offer when.

Availability control refers to the process of determining the set of products offered to customers, assuming dependent demand. Within the independent demand paradigm, this problem is referred to as *capacity control.* Both types of control problem are classified as *quantity-based* DM decisions (Strauss et al., 2018).

The choice-based network problem can be set up as a stochastic dynamic program (see Section 2.1.2). Let us initially consider the supply-side of the problem. We define,

- the selling horizon of duration T time periods, indexed discretely from $t = 1,...,T$,
- the set of products, $\mathscr{J} = \{1,2,3,....,J\}$, with fixed revenue (price) r_j ($j = 0$ is the no-purchase alternative),
- the set of resources, $\mathscr{I} = \{1,...,I\}$,
- the resources needed for a product are defined in a matrix, $\mathbf{A} \in \mathbb{N}_0^{I\times J}$, where $a_{ij} = b$, if resource i is required for product j,
- the resource consumption of product j is the j-th column vector $\mathbf{A}_j$,
- the available inventory at the start of period t, $\mathbf{c}_t \in \mathbb{N}_0^I$,
- the set $\mathscr{J}(\mathbf{c}_t) = \{j \in \mathscr{J} | \mathbf{A}_j \leq \mathbf{c}_t\}$, containing all products that can be produced at time t given current capacity, and,
- the offer set, $S \subseteq \mathscr{J}(\mathbf{c}_t)$, containing all products offered to the customer at time t for purchase.

Now, let us consider the demand-side of the problem. Each customer is assumed to belong to a particular demand segment from the set $\mathscr{L} = \{1,...,L\}$. For $l \in \mathscr{L}$, we need to know the *arrival rate, consideration set* and a statistical model that determines a customer's choices from an offer set. We define,

- the probability of a customer arriving at time t, λ,
- the probability of a customer belonging to segment l, p_l,
- the segment-specific arrival rate, $\lambda_l = p_l\lambda$,
- the consideration set, $\mathscr{C}_l \subseteq \mathscr{J}$, the set of products allocated to segment l (consideration sets can be either overlapping or disjoint across segments), and,
- for an offer set S, a customer in segment l can choose product j with probability $P_j^l(S^l)$ if $j \in S_l = S \cap \mathscr{C}_{\updownarrow}$.

So, the *choice probability* of a firm selling a product $j \in S$ to a customer is $P_j(S) = \sum_{l\in\mathscr{C}_l} p_l P_j^l(S_l)$. Correspondingly, the probability of the no-purchase alternative is $P_0(S) = 1 - \sum_{j\in S} P_j(S)$. This distribution is usually determined by specifying a discrete choice model (see Section 3). The probabilities are then estimated and calibrated using historical data. For now, we assume a general choice model.

2.1.2 Stochastic dynamic programming for choice-based availability control

The goal of availability control is to identify the offer set, S, that maximises revenue at each time period of the selling horizon. The optimal policy for the entire problem can now be written down.

We define the value function $V_t(\mathbf{c}_t)$ to be the aggregated expected revenue that can be obtained over the time horizon $[t, T]$, given the current state, $\mathbf{c}_t$. The state space consists of all possible inventories. Thus, $\mathbf{V}_t(\mathbf{c}_t)$ must solve the Bellman equation,

$$\mathbf{V}_t(\mathbf{c}_t) = \max_{S \subseteq \mathscr{J}(\mathbf{c}_t)} \left\{ \sum_{j \in S} \lambda P_j(S)\Big(r_j + \mathbf{V}_{t+1}(\mathbf{c}_t - \mathbf{A}_j)\Big) + \Big(\lambda P_0(S) + 1 - \lambda\Big)\mathbf{V}_{t+1}(\mathbf{c}_t)\right\}, \forall t, \forall \mathbf{c}_t. \tag{2.1}$$

At $t = T+1$, we set $\mathbf{V}_{T+1}(\mathbf{c}_{T+1}) = 0, \forall \mathbf{c}_{T+1}$. We can also rewrite equation 2.1 in terms of the opportunity cost for a single product, $\triangle_j \mathbf{V}_{t+1}(\mathbf{c}_t) = \mathbf{V}_{t+1}(\mathbf{c}_t) - \mathbf{V}_{t+1}(\mathbf{c}_t - \mathbf{A}_j)$. This transformation yields,

$$\mathbf{V}_t(\mathbf{c}_t) = \max_{S \subseteq \mathscr{J}(\mathbf{c}_t)} \left\{ \sum_{j \in S} \lambda P_j(S)\Big(r_j - \triangle_j \mathbf{V}_{t+1}(\mathbf{c}_t)\Big)\right\} + \mathbf{V}_{t+1}(\mathbf{c}_t), \forall t, \forall \mathbf{c}_t. \tag{2.2}$$

In practice, equations 2.1 and 2.2 are often intractable. The value function $\mathbf{V}_t(\mathbf{c}_t)$ must be evaluated and stored for all time periods, t, and all possible inventories, $\mathbf{c}_t$. As the number of resources grows, the state space grows exponentially to accommodate. Consequently, the surrounding literature predominantly focuses on developing efficient methods for approximating the value function or the opportunity cost. A summary of these approximation techniques is provided in Section 5.3 of Strauss et al. (2018)

2.2 Assortment optimisation

At each state, $\mathbf{c_t}$, all possible actions have to be compared. Within the context of the availability control problem, an action refers to the specification of the offer set, S. In total, there are $2^{|\mathscr{J}(\mathbf{c}_t)|}$ actions to consider at time t. Naturally, this kind of comparison cannot be conducted using brute force. Consequently, an optimisation problem must be solved to determine the best action. This process is known as *assortment optimisation*.

If we assume that the opportunity cost for product j, $\triangle_j \mathbf{V}_{t+1}(\mathbf{c}_t)$, in equation 2.2 is known, every availability control problem in period t is reduced to an assortment optimisation problem. We can rewrite equation 2.2 as a revenue maximising assortment problem,

$$\max_{S \subseteq \mathscr{J}(\mathbf{c}_t)} \left\{ \sum_{j \in S} \lambda P_j(S) \tilde{r}_j \right\},$$

where $\tilde{r}_j = r_j - \triangle_j \mathbf{V}_{t+1}(\mathbf{c}_t)$ is the *displacement-adjusted revenue* that needs to evaluated $\forall j$.

Assortment optimisation problems often need to be solved in a short period of time. For instance, the process of deciding upon which fares to display on an airline website must be completed in real-time. The difficulty of finding the optimal assortment lies in determining the structure of the underlying choice model. The choice model is used to calculate the choice probabilities, $P_j(S)$, for all available products, $j \in S$. In the literature, assortment optimisation methods tend to either assume a general choice model structure or build directly upon an existing choice model. Please refer to Section 4 of Strauss et al. (2018) for further details.

2.3 Single-resource availability control

Talluri and van Ryzin (2004) focus on the situation where there is only one resource used for production (the so-called *single-leg* problem). In this case, the available inventory at time t, $\mathbf{c}_t$, is reduced to a scalar, c.

Talluri and van Ryzin develop a full theoretical framework for the problem of availability control. Assuming a general choice model, the fully characterised optimal policy of the single-resource case has a simple form. It first requires the identification of an ordered family of *efficient* subsets of $\mathscr{J}, S_1, ..., S_m$. Here, efficiency is determined by making a trade-off between the expected revenue and choice probability. The sets are indexed such that the revenue and the choice probabilities increase monotonically in the index. An optimal set, S_k, is opened at each time period, where k is the index of the set that maximises the Bellman equation in 2.2.

For a fixed time t, the optimal index, k, is increasing in the remaining capacity c. For a fixed capacity, c, the optimal index, k, is decreasing in the remaining time. In other words, the further along the optimal set, S_k, is in the ordered sequence, the more capacity (or less time) there is available.

The authors further show that if the efficient subsets are nested, a *nested allocation policy* is optimal. In their tutorial paper, van Ryzin and Talluri (2005) state that a control policy is *nested* if there is a family of increasing subsets, $S_1 \subseteq S_2 \subseteq ... \subseteq S_m$, and an index $k_t(c)$ that is increasing in c, s.t. $S_{k_t(c)}$ is chosen at period t, for the remaining capacity c.

An estimation procedure, based upon the expectation-maximization (EM) algorithm, is also proposed in Talluri and van Ryzin (2004). The idea here is to jointly estimate the probability of arrival, λ, and the choice model parameters, assuming that non-purchase outcomes are unobservable.

3 Discrete choice models

In this section, we discuss the design of discrete choice models for DM, adopting the mathematical notation and framework used in Strauss et al. (2018). Broadly, choice models can be classified as *parametric* (also known as *random utility models*), *non-parametric*, or *multi-stage*. For a comprehensive introduction to the (econometric) theory of random utility models, we recommend the textbook by Train (2009).

The aim of this discussion is to illustrate the philosophy of each of the different approaches. In the operations literature, the focus has largely been on the simultaneous estimation and assortment optimisation of parametric models. Please refer to Strauss et al. (2018) for further information regarding the estimation procedures used for discrete choice models.

3.1 Parametric models

3.1.1 Notation

Parametric models draw heavily from random utility theory in economics. Here, the key assumption is that each individual has a quantifiable *utility* (i.e. benefit, value) associated with each available choice (i.e. product, service type). The goal of the individual is to then decide on the alternative that maximises his or her utility. We define,

- the set of products produced, $\mathscr{J} = \{1, 2, 3, ..., J\}$,
- the utility for choice j, $U_j = u_j + \epsilon_j$, where u_j is deterministic and ϵ_j is random with zero mean,
- the utility of not choosing anything, U_0 (non-purchase alternative),
- the set of choices offered to the customer for purchase, $S \subseteq \mathscr{J}$ (offer set), and,
- the probability of product j being chosen, $P_j(S) = P(U_j = \max\{U_{j'} | j' \in S \cup \{0\}\})$ (choice probability), where $P_j(S) \geq 0\ \forall j \in S$ and $\sum_{i \in S} P_i(S) + P_0 = 1$.

Remarks

1. The offer set, $\mathscr{J}$, must satisfy three key properties: (i) the alternatives in the choice set must be *mutually exclusive* from the customer's perspective; (ii) the choice set must be *exhaustive*; and, (iii) the choice set must be *finite* (Train, 2009).
2. The non-random component of utility, u_j, is typically expressed as a linear combination of attributes, $\boldsymbol{\beta}^T \mathbf{x}_j$, which are known to influence the choice probabilities. Depending on the model, different assumptions will be made about the distribution of the random component of utility, ϵ_j. Train (2009) state that the random component should capture any other unknown factors that influence the utility of a product.
3. The products in the choice set can be ranked according to their utility. In a *random utility model*, the decision rule is therefore: choose product i if $U_i > U_j\ \forall j \neq i$ (Train, 2009).

3.1.2 Multinomial logit

One of the most popular choice models in the literature is the *multinomial logit* (MNL) model. This model assumes that the preferences of an entire consumer base can be summarised with the parameters, $\boldsymbol{\beta}$. For each product j, the individual-level utility can be expressed as,

$$U_j = \underbrace{u_j}_{\text{non-random utility}} + \underbrace{\epsilon_j}_{\text{random utility}} = \boldsymbol{\beta}^T \mathbf{x}_j + \epsilon_j.$$

Here, the random utility components, ϵ_j ($\forall j \in S$), are i.i.d standard Gumbel (double exponential) random variables, i.e. for $x \in \mathbb{R}$, $P(\epsilon_j \leq x) = \exp(\exp(-x))$. We set $j = 0$ to be the no-purchase option. Since utility is ordinal in nature, we can assume w.l.o.g. that $u_0 = 0$ (Talluri and van Ryzin, 2004). For $j \in S \cup \{0\}$, we can express the choice probability as,

$$P_j(S) = \frac{\exp\left(\frac{u_j}{\mu}\right)}{\sum_{i \in S} \exp\left(\frac{u_i}{\mu}\right) + \exp\left(\frac{u_0}{\mu}\right)} = \frac{\exp\left(\frac{u_j}{\mu}\right)}{\sum_{i \in S} \exp\left(\frac{u_i}{\mu}\right) + 1}, \tag{3.1}$$

where μ is a scaling parameter. As $\mu \to 0$, the MNL becomes deterministic. Alternatively, for $\mu \to \infty$, the probability distribution becomes uniform, with $P_j(S) = \frac{1}{|S|+1} \forall j \in S$. In this case, the utility of a product loses interpretation. The inclusion of this scaling parameter improves the flexibility of the model.

In order to ensure the validity of MNL, an important assumption must be made about the product choices. The *independence from irrelevant alternatives* (IIA) condition states that for any subsets S, S' of $\mathcal{J}$ and any two choices $i, j \in S \cap S'$,

$$\frac{P_i(S)}{P_j(S)} = \frac{P_i(S')}{P_j(S')},$$

allowing for proportional substitution of choices. In other words, IIA requires that if a new choice is offered, the probabilities for the original choices must adjust in precisely the amount necessary to retain the original ratio. Strauss et al. (2018) explain that this assumption may lead to the overestimation of the choice probabilities of products considered to be substitutable by a customer.

3.1.3 Finite-mixture logit / Latent class models

It is not uncommon for businesses to leverage the differences between customer preferences and behaviours, when making demand-management decisions. We can easily extend the MNL to a situation where there are multiple customer or demand segments to consider. Each group can have different membership specifications, requiring a different MNL model to be fitted to each one.

In practice, customer membership of a segment is unobservable. Instead, we can quantify the likelihood of an individual belonging to a specific class using *membership probabilities.* For every segment $l \in \mathcal{L}$, we need to jointly estimate the model parameters $\boldsymbol{\beta}_l$ and the membership probability q_l (s.t. $\sum_{l \in \mathcal{L}} q_l = 1$). The mixture of multiple MNL models is known as the *finite-mixture logit* or *latent class* model.

Each class l has an individual-specific, deterministic utility, $u_{lj} = \boldsymbol{\beta}_l^T \mathbf{x}_j$, where $\mathbf{x}_j$ is the vector of attributes associated with product j. We also define consideration sets of products, $C_l \subseteq \mathcal{J}$, at the class level. The choice probability for product $j \in S \cup \{0\}$ can be written as,

$$P_j(S) = \sum_{l \in \mathcal{L}} q_l \frac{\exp(u_{lj})}{\sum_{i \in S \cap C_l} \exp(u_{li}) + \exp(u_{l0})}. \tag{3.2}$$

Finite-mixture models fall under the wider class of *mixed multinomial logit* (MMNL) models. McFadden and Train (2000) demonstrate that any random utility model can be sufficiently approximated by an MMNL model. Strauss et al. (2018) point out that this makes finite-mixture logits better at modelling choices than a simple MNL (i.e. lower specification error), at the cost of additional complexity in parameter estimation.

3.1.4 Nested logit

To circumvent the issue caused by the IIA assumption, we can alternatively group product choices into *nests.* The *nested logit* model assumes that IIA holds within each nest, but not between nests. A nest comprises of a set of substitutes. Consumers must first choose a nest of substitutes and then pick a product out of that nest.

We define $\mathcal{K}$ to be the set of nests available and S_k to be the set of products available for the customer in nest k. For every nest $k \in \mathcal{K}$, the preference value for product j is set to be $v_{kj} = \exp\left(\frac{u_j}{\mu_k}\right)$. Here, u_j is the deterministic utility of product j and μ_k is a parameter measuring the degree of independence in unobserved utility over the choices in nest k (Train, 2009). For completion, the preference value for the non-purchase alternative is v_0. Therefore, the *overall preference* for a particular nest can be written as $V_k(S_k) = \sum_{j \in S_k} v_{kj}$.

Given that a customer has chosen to buy from nest k, the probability of product $j \in S_k$ being chosen is,

$$\frac{v_{kj}}{V_k(S_k)}.$$

Further, the probability of a customer choosing nest k can be computed as,

$$\frac{V_k(S_k)^{\mu_k}}{v_0 + \sum_{h \in \mathcal{K}} V_h(S_h)^{\mu_h}}.$$

Thus, the choice probability for product j belonging to nest k is,

$$P_j(\{S_h\}_{h \in \mathcal{K}}) = \frac{v_{kj} V_k(S_k)^{\mu_k - 1}}{v_0 + \sum_{h \in \mathcal{K}} V_h(S_h)^{\mu_h}}. \tag{3.3}$$

The nested logit is a member of the larger family of *generalised extreme value* (GEV) models. Its mathematical formulation is simple compared to other types of GEV models, making it applicable to a wide variety of applications (i.e. housing, energy, transportation) (Train, 2009). Extensive theoretical development of GEV models is conducted in McFadden (1978) and McFadden (1981).

3.1.5 Alternative parametric models

Ordered generalised extreme value model: Small (1987) proposes a generalisation of the MNL model, by calculating the stochastic correlation between choices that are in close proximity. This is done over alternatives that have a predetermined ordering.

Exponomial models: Alptekinoğlu and Semple (2016) propose a negatively-skewed distribution for customer utilities. This is a departure from the traditional assumption made in MNL or NL models, where the willingness-to-pay distribution is positively-skewed. In this formulation, the choice probabilities can be expressed analytically as a linear function of exponential terms (*exponomial*).

3.2 Non-parametric models

The main advantage of parametric models is the ability to include various attributes, such as product characteristics and price, that can help explain customer preferences. It is also possible for these models to extrapolate predictions to products that have not been considered historically. However, the main drawback of any parametric framework is that assumptions must be made about the structural properties of customer preferences and the attributes that influence them.

van Ryzin and Vulcano (2015) point out that there is a trade-off to be made between specification and estimation error, when choosing a parametric model. In many cases, a more complex model (i.e. with more latent classes or variables) can capture choice behaviour better and lessen the impact of specification error. When faced with limited data, however, overfitting runs the risk of increasing estimation error. A parsimonious model, while less accurate, may combat estimation error. This means that, in most cases, practitioners must rely on their expert judgement or resort to trial and error to determine a suitable specification (Jagabathula and Vulcano, 2018).

Consequently, there is considerable interest in data-driven, non-parametric choice models. For each customer, the idea here is to construct a *preference list* or *ranking list* of all choices, $j \in \mathscr{J}$, together with the non-purchase alternative. A ranking list is typically referred to as a *customer type*. The demand is then modelled using a probability mass function, constructed over all customer types (Strauss et al., 2018). These models are known as *rank-based choice models*.

The number of customer types is factorial in the number of products offered. Therefore, it can be shown that rank-based choice models suffer from a *non-identifiability* problem (van Ryzin and Vulcano, 2015). In other words, two different probability mass functions could easily explain the same observational data.

3.2.1 Example: van Ryzin and Vulcano (2015)

Assuming a Bernoulli arrival process, van Ryzin and Vulcano (2015) propose a non-parametric approach that only uses product availability and sales transaction data. The authors' approach to selecting rankings is to start off with a simple, parsimonious collection of rankings and gradually add customer types to the collection that increase the likelihood.

Given the historic data, van Ryzin and Vulcano establish a market-discovery algorithm for augmenting the initial set of rankings. The procedure consists of iterating through two steps: (i) estimate the pmf to maximise the likelihood function, using the current set of customer types under consideration; and (ii) add a new customer type to the current set to improve the likelihood value.

This approach controls for overfitting by checking the log-likelihood value. The trade-off is now between more customer types and a higher chance of overfitting against fewer customer types and larger estimation error. In a technical extension paper, van Ryzin and Vulcano (2017) propose an EM method to estimate rank-based choice models. The method is designed to jointly estimate the arrival rate of customers and the pmf, using only sales transaction and product availability data.

3.3 Multi-stage models

The discrete choice models discussed so far have all been one-stage approaches. Given a set of product offerings, a one-stage approach outputs a purchase probability for each choice. Recently, there has been a shift in the marketing literature towards studying the *consider-then-choose* decision process. In this framework, a customer first formulates their consideration set and then sets about deciding between the available alternatives (Strauss et al., 2018).

3.3.1 EXAMPLE: JAGABATHULA AND VULCANO (2018)

An interesting non-parametric, multi-stage approach is proposed by Jagabathula and Vulcano (2018). The authors assume that there is a set number of customers, m, who repeatedly purchase from a particular category of products (e.g. shampoo). Each customer under consideration belongs to a market segment and is represented by a set of partial preferences. Partial preferences are typically of the form, *'product 1 is preferred to product 2'*. A directed acyclic graph (DAG) is used to store the dependencies between products, but does not provide a fully specified ranking.

When a customer arrives in the system, they have access to the full assortment of products, along with the non-purchase alternative. The customer samples a ranking of the full product list that is consistent with their DAG and then selects the best option within their consideration set. In practice, the customer's DAG and their sampled list cannot be observed.

Jagabathula and Vulcano propose the following method for estimation: (i) build a DAG for each customer under consideration; (ii) cluster the DAGs into market segments to capture customer heterogeneity; and, (iii) derive marginal distributions for partial preferences under the multinomial logit model.

In their paper, panel data is used to train the model and test it. The authors use the first half of the data set to fine-tune the estimation procedure. The second half of the data set is then used to predict customer behaviour. Jagabathula and Vulcano report that their partial-order ranking system improves the accuracy of customer-level predictions, when compared to other state-of-the-art approaches.

4 CONCLUSION

In this paper, we have provided an overview of the different choice modelling approaches available, ranging from the structured models rooted in random utility theory to more data-driven techniques, such as Jagabathula and Vulcano (2018). We also discuss the place that discrete choice models hold within the wider problem of availability control. Indeed, many of the quantity-based, demand management decisions made by businesses are underpinned by the type of customer choice model used.

Choice-based demand management has benefitted greatly from a boost in methodological research in the past decade. In the past, research was hindered by the simplistic assumption of independent demand. Over time, this has been become an unrealistic assumption, due to the fast-changing nature of online retail. Consumers today have access to more information than ever and have more flexibility in how they choose and substitute products.

The shift to a dependent demand framework allows for a more nuanced analysis of consumer demand. In particular, non-parametric choice models are able to leverage a wealth of historic data for better predictions, without making any restrictive assumptions about customer preferences. Open problems in the area currently include, but are not limited to,

- modelling choice processes over several stages (perhaps more than two), incorporating add-on / ancillary service options,
- online learning in the network availability control problem, with the view to deal with the exploration-exploitation trade-off, and,
- personalisation of choice modelling, e.g. identifying the demand segment an incoming purchase is coming from and then offering segment-specific alternatives.

REFERENCES

Alptekinoğlu, A. and Semple, J. H. (2016), 'The exponomial choice model: A new alternative for assortment and price optimization', *Operations Research* **64**(1), 79–93.

Jagabathula, S. and Vulcano, G. (2018), 'A partial-order-based model to estimate individual preferences using panel data', *Management Science (Articles in Advance, 2017)* .
URL: *https://doi.org/10.1287/mnsc.2016.2683*

McFadden, D. (1978), 'Modeling the choice of residential location', *Transportation Research Record* (673).

McFadden, D. (1981), 'Econometric models of probabilistic choice', *Structural analysis of discrete data with econometric applications* **198272**.

McFadden, D. and Train, K. (2000), 'Mixed MNL models for discrete response', *Journal of applied Econometrics* pp. 447–470.

Small, K. A. (1987), 'A discrete choice model for ordered alternatives', *Econometrica* **55**(2), 409–424.
URL: *http://www.jstor.org/stable/1913243*

Strauss, A. K., Klein, R. and Steinhardt, C. (2018), 'A review of choice-based revenue management: Theory and methods'.

Talluri, K. and van Ryzin, G. (2004), 'Revenue management under a general discrete choice model of consumer behavior', *Management Science* **50**(1), 15–33.

Train, K. E. (2009), *Discrete choice methods with simulation*, Cambridge university press.

van Ryzin, G. J. and Talluri, K. T. (2005), 'An introduction to revenue management', *Tutorials in operations research* pp. 142–195.

van Ryzin, G. and Vulcano, G. (2015), 'A market discovery algorithm to estimate a general class of nonparametric choice models', *Management Science* **61**(2), 281–300.

van Ryzin, G. and Vulcano, G. (2017), 'An expectation-maximization method to estimate a rank-based choice model of demand', *Operations Research* **65**(2), 396–407.