Assortment Optimization under the Sequential Multinomial Logit Model

Alvaro Flores[*] Gerardo Berbeglia[†] Pascal Van Hentenryck[‡]

Thursday 30$^{\text{th}}$ August, 2018

Abstract

We study the assortment optimization problem under the *Sequential Multinomial Logit* (SML), a discrete choice model that generalizes the multinomial logit (MNL). Under the SML model, products are partitioned into two levels, to capture differences in attractiveness, brand awareness and, or visibility of the products in the market. When a consumer is presented with an assortment of products, she first considers products in the first level and, if none of them is purchased, products in the second level are considered. This model is a special case of the Perception-Adjusted Luce Model (PALM) recently proposed by Echenique, Saito, and Tserenjigmid (2018). It can explain many behavioral phenomena such as the attraction, compromise, similarity effects and choice overload which cannot be explained by the MNL model or any discrete choice model based on random utility. In particular, the SML model allows violations to *regularity* which states that the probability of choosing a product cannot increase if the offer set is enlarged.

This paper shows that the seminal concept of revenue-ordered assortment sets, which contain an optimal assortment under the MNL model, can be generalized to the SML model. More precisely, the paper proves that all optimal assortments under the SML are revenue-ordered by level, a natural generalization of revenue-ordered assortments that contains, at most, a quadratic number of assortments. As a corollary, assortment optimization under the SML is polynomial-time solvable. This result is particularly interesting given that the SML model does not satisfy the regularity condition and, therefore, it can explain choice behaviours that cannot be explained by any choice model based on random utility.

Keywords: Revenue management; Assortment optimization; assortment planning; discrete choice models; revenue-ordered assortments.

1 Introduction

The assortment optimization problem is a central problem in revenue management, where a firm wishes to offer a set of products with the goal of maximizing the expected revenue. This problem has many relevant applications in retail and revenue management (Kök, Fisher, and Vaidyanathan, 2005). For example, a publisher might need to decide the set of advertisements to show, an airline

[*]College of Engineering & Computer Science, Australian National University (alvaro.flores@anu.edu.au).

[†]Melbourne Business School, The University of Melbourne (g.berbeglia@mbs.edu).

[‡]Industrial and Operations Engineering & Computer Science and Engineering, University of Michigan, Ann Arbor (pvanhent@umich.edu).

must decide which fare classes to offer on each flight, and a retailer needs to decide which products to show in a limited shelf space.

The first consumer demand models studied in revenue management were based on the independent demand principle. This principle stated that customers decide beforehand which product they want to purchase: If the product is available, they make the purchase and, otherwise, they leave without purchasing. In these models, the problem of finding the best offer set of products is computationally simple, but this simplicity comes with an important drawback: These models do not capture the substitution effects between products. That is, they cannot model the fact that, when a consumer cannot find her/his preferred product, she/he may purchase a substitute product. It is well-known that choice models that incorporate substitution effects improve demand predictions (Talluri and Van Ryzin, 2004; Newman et al., 2014; van Ryzin and Vulcano, 2015; Ratliff et al., 2008; Vulcano, van Ryzin, and Chaar, 2010). One of the most celebrated discrete choice models is the *Multinomial Logit* (MNL) (Luce, 1959; McFadden, 1974). Under the MNL model, the assortment problem admits a polynomial-time algorithm (Talluri and Van Ryzin, 2004). However, the model suffers from the independence of irrelevant alternatives (IIA) property (Ben-Akiva and Lerman, 1985) which says that, when a customer is asked to choose among a set of alternatives S, the ratio between the probability of choosing a product $x \in S$ and the probability of choosing $y \in S$ does not depend on the set S. In practice, however, the IIA property is often violated. To overcome this limitation, more complex choice models have been proposed in the literature such as the Nested Logit model (Williams, 1977), the latent class MNL (Greene and Hensher, 2003), the Markov Chain model (Blanchet, Gallego, and Goyal, 2016), and the exponomial model (Daganzo, 1979; Alptekinoğlu and Semple, 2016). All these models satisfy the following property: The probability of choosing an alternative does not increase if the offer set is enlarged. Despite the fact that this property (known as regularity) appears natural, it is well-known that it is sometimes violated by individuals (Debreu, 1960; Tversky, 1972a,b; Tversky and Simonson, 1993). Recently, there have been efforts to develop discrete choice models that can explain complex choice behaviours such as the violation of regularity, one of the most prominent examples is the perception-adjusted Luce model (PALM) (Echenique, Saito, and Tserenjigmid, 2018).

While the PALM and the nested logit are both conceived as sequential choice processes, they have important differences. Probably the most important difference is that the nested logit model belongs to the family of random utility models (RUM)*, and therefore can't accommodate regularity violations. On the other hand the PALM does not belong to the RUM class, and allows regularity violations as well as choice overload. In terms of the choice process, in the nested logit model customers first select a nest, and then a product within the nest. In the perception-adjusted Luce's model products are separated by preference levels, so when a customer is offered a set of products, she first chooses among the offered products belonging to the lowest available level, and if none of them are chosen then she selects among the next available level, and keeps repeating this process until no more levels are available or until a purchase is made.

1.1 Our Contributions

In this paper, we study the assortment optimization problem for a two-stage discrete choice model model that generalizes the classical Multinomial Logit model. This model, which we call the *Sequential Multinomial Logit* (SML) for brevity, is a special case of the recently proposed model known as the perception-adjusted Luce model (PALM) (Echenique, Saito, and Tserenjigmid, 2018). In

*This is unless nest specific parameters are greater than one, a case rarely studied in the literature.

the SML model, products are partitioned *a priori* into two sets, which we call levels. This product segmentation into two levels can capture different degree of attractiveness. For example, it can model customers who check promotions/special offers first before considering the purchase of regular-priced products. It can also model consumer brand awareness, where customers first check products of specifics brands before considering the rest. Finally, the SML can model product visibilities in a market, where products are placed in specific positions (aisles, shelves, web-pages, etc.) that induce a sequential analysis, even when all the products are at sight. Our main contribution is to provide a polynomial-time algorithm for the assortment problem under the SML and to give a complete characterization of the resulting optimal assortments.

A key feature of the PALM and the SML, is their ability to capture several effects that cannot be explained by any choice model based in random utility (such as for example the MNL, the mixed MNL, the markov chain model, and the stochastic preference models). Examples of such effects include attraction, (Doyle et al., 1999), the compromise effect (Simonson and Tversky, 1992), the similarity effect (Debreu, 1960; Tversky, 1972b), and the paradox of choice (also known as choice overload) (Iyengar and Lepper, 2000; Schwartz, 2004; Haynes, 2009; Chernev, Böckenholt, and Goodman, 2015). These effects are discussed in the next section. In particular, the SML allows for violations of regularity. There are very few analyses of assortment problems under a choice model outside the RUM class.

Our algorithm is based on an in-depth analysis of the structure of the SML. It exploits the concept of revenue-ordered assortments that underlies the optimal algorithm for the assortment problem under the MNL. The key idea in our algorithm is to consider an assortment built from the union of two sets of products: A revenue-ordered assortment from the first level and another revenue-ordered assortment from the second level. Several structural properties of optimal assortments under the SML are also presented.

1.2 Relevant Literature

The heuristic of revenue-ordered assortments, consists in evaluating the expected revenue of all the assortments that can be constructed as follows: fix threshold revenue ρ and then select all the products with revenue of at least ρ. This strategy is appealing because it can be applied to assortment problems for any discrete choice model. In addition, it only needs to evaluate as many assortments as there are different revenues among products. In a seminal result, Talluri and Van Ryzin (2004) showed that, under the MNL model, the optimal assortment is revenue-ordered. This result does not hold for all assortment problems however. For example, revenue-ordered assortments are not necessarily optimal under the MNL model with capacity constraints (Rusmevichientong, Shen, and Shmoys, 2010). Nevertheless, in another seminal result, Rusmevichientong, Shen, and Shmoys (2010) showed that the assortment problem can still be solved optimally in polynomial time under such setting.

Rusmevichientong and Topaloglu (2012) considered a model where customers make choices following an MNL model, but the parameters of this model belong to a compact uncertainty set, i.e., they are not fully determined. The firm wants to be protected against the worst-case scenario and the problem is to find an optimal assortment under these uncertainty conditions. Surprisingly, when there is no capacity constraint, the revenue-ordered strategy is optimal in this setting as well.

There are also studies on how to solve the assortment problem when customers follow a mixed multinomial logit model. Bront, Méndez-Díaz, and Vulcano (2009) showed that this problem is NP-hard in the strong sense using a reduction from the *minimum vertex cover problem* (Garey and Johnson, 1979). Méndez-Díaz et al. (2014) proposed a branch-and-cut algorithm to solve the optimal assortment under the Mixed-Logit Model. An algorithm to obtain an upper bound of the revenue of an optimal assortment solution under this choice model was proposed by Feldman and Topaloglu (2015). Rusmevichientong et al. (2014) showed that the problem remains NP-hard even when there are only two customers classes.

Another model that attracted researchers attention is the *nested logit model* (Williams, 1977). Under the nested logit model, products are partitioned into nests, and the selection process for a customer goes by first selecting a nest, and then a product within the selectednest. It also has a dissimilarity parameter associated with each nest that serves the purpose of magnifying or dampening the total preference weight of the nest. For the two-level nested logit model, Davis, Gallego, and Topaloglu (2014) studied the assortment problem and showed that, when the dissimilarity parameters are bounded by 1 and the no-purchase option is contained on a nest of its own, an optimal assortment can be found in polynomial time; If any of these two conditions is relaxed, the resulting problem becomes NP-hard, using a reduction from the *partition problem* (Garey and Johnson, 1979). The polynomial-time solution was further extended by Gallego and Topaloglu (2014), who showed that, even if there is a capacity constraint per nest, the problem remains polynomial-time solvable. Li et al. (2015) extended this result to a d-level nested logit model (both results under the same assumptions over the dissimilarity parameters and the no-purchase option). Jagabathula (2014) proposed a local-search heuristic for the assortment problem under an arbitrary discrete choice model. This heuristic is optimal in the case of the MNL, even with a capacity constraint.

Wang and Sahin (2018) has studied the assortment optimization in a context in which consumer search costs are non-negligible. The authors showed that the strategy of revenue-ordered assortments is not optimal. Another interesting model sharing similar choice probabilities to those of the PALM, is the one proposed in Manzini and Mariotti (2014) which is based on consider first and choose second process. Echenique, Saito, and Tserenjigmid (2018) showed that the PALM and the model by Manzini and Mariotti are in fact disjoint. The assortment optimization problem under the Manzini and Mariotti model has been recently studied by Gallego and Li (2017), where they show that revenue-ordered assortments strategy is optimal. Another choice model studied is the negative exponential distribution (NED) model (Daganzo, 1979), also known as the Exponomial model (Alptekinoğlu and Semple, 2016) in which customer utilities follow negatively skewed distribution. Alptekinoğlu and Semple (2016) proved that when prices are exogenous, the optimal assortment might not be revenue-ordered assortment, because a product can be skipped in favour of a lower-priced one depending on the utilities. This last result differs from what happens under the MNL and the Nested Logit Model (within each nest). Another recently proposed extension to the multinomial logit model, is the General Luce Model (GLM) (Echenique and Saito, 2015). The GLM also generalizes the MNL and falls outside the RUM class. In the GLM, each product has an intrinsic utility and the choice probability depends upon a dominance relationship between the products. Given an assortment S, consumers first discard all dominated products in S and then select a product from the remaining ones using the standard MNL model. Flores, Berbeglia, and Van Hentenryck (2017) studied the assortment optimization problem under the GLM. The authors showed that revenue ordered assortments are not optimal, and proved that the problem can still be solved in polynomial time.

Recently, Berbeglia and Joret (2016) studied how well revenue-ordered assortments approximate the optimal revenue for a large class of choice models, namely all choice models that satisfy regularity. They provide three types of revenue guarantees that are exactly tight even for the special case of the RUM family. In the last few years, there has been progress in studying the assortment problem in choice models that incorporate visibility or position biases. In these models, the likelihood of selecting an alternative not only depends on the offer set, but also in the specific positions at which each product is displayed (Abeliuk et al., 2016; Aouad and Segev, 2015; Davis, Gallego, and Topaloglu, 2013; Gallego et al., 2016).

We mentioned that PALM and SML are able to accommodate many effects that can't be explained by models the RUM class. We briefly describe each one of them in the following paragraphs.

The attraction effect stipulates that, under certain conditions, adding a product to an existing assortment can increase the probability of choosing a product in the original assortment. We briefly describe two experiments of this effect. Simonson and Tversky (1992) considered a choice among three microwaves x, y, and z. Microwave y is a Panasonic oven, perceived as a good quality product, and z is a more expensive version of y. Product x is an Emerson microwave oven, perceived as a lower quality product. The authors asked a set of 60 individuals ($N = 60$) to choose between x and y; they also asked another set of 60 participants ($N = 60$) to choose among x, y, and z. They found out that the probability of choosing y increases when product z is shown. This is a direct violation of regularity, which states that the probability of choosing a product does not increase when the choice set is enlarged, as described by McCausland and Marley (2013). Another demonstration of the attraction effect was carried by Doyle et al. (1999) who analyzed the choice behaviour of two sets of participants ($N = 70$ and $N = 82$) inside a grocery store in the UK varying the choice set of baked beans. To the first group, they showed two types of baked beans: Heinz baked beans and a local (and cheaper) brand called Spar. Under this setting, the Spar beans was chosen 19% of the time. To the second group, the authors introduced a third option: a more expensive version of the local brand Spar. After adding this new option, the cheap Spar baked beans was chosen 33% of the time. It is worth highlighting that the choice behaviour in these two experiments cannot be explained by a Multinomial Logit Model, nor can it be explained by any choice model based on random utility.

The compromise effect (Simonson and Tversky, 1992) captures the fact that individuals are averse to extremes, which helps products that represent a "compromise" over more extreme options (either in price, familiarity, quality, ...). As a result, adding extreme options sometimes leads to a positive effect for compromise products, whose probabilities of being selected increase in relative terms compared to other products. This phenomenon violates again the IIA axiom of Luce's model and the regularity axiom satisfied by all random utility models (Berbeglia and Joret, 2016). Again, the compromise effect can be captured with the PALM.

The similarity effect is discussed in Tversky (1972b), elaborating on an example presented in Debreu (1960): Consider x and z to represent two recordings of the same Beethoven symphony and y to be a suite by Debussy. The intuition behind the effect is that x and z jointly compete against y, rather than being separate individual alternatives. As a result, the ratio between the probability of choosing x and the probability of choosing y when the customer is shown the set $\{x, y\}$ is larger than the same ratio when the customer is shown the set $\{x, y, z\}$. Intuitively, z takes a market share of product x, rather than a market share of product y.

Finally, the choice overload effect occurs when the probability of making a purchase decreases when the assortment of available products is enlarged. To our knowledge, the first paper that shows the empirical existence of choice overload is written Iyengar and Lepper (2000). In their experimental setup, customers are offered jams from a tasting booth displaying either 6 (limited selection) or 24 (extensive selection) different flavours. All customers were given a discount coupon for making a purchase of one of the offered jams. Surprisingly, 30% of the customers offered the *limited selection* used the coupon, while only 3% of customers offered the *extensive selection* condition used the coupon. Another studies of choice overload are in 401(k) plans Sethi-Iyengar, Huberman, and Jiang (2004), chocolates (Chernev, 2003b), consumer electronics (Chernev, 2003a) and pens (Shah and Wolford, 2007). For a more in depth discussion of this effect the reader is referred to Schwartz (2004). Readers are also referred to Chernev, Böckenholt, and Goodman (2015) for a review and meta-analysis on this topic.

2 Problem Formulation

This section presents the sequential multinomial logit model considered in this paper and its associated assortment optimization problem. Let X be the set of all products and x_0 be the no-choice option ($x_0 \notin X$). Following Echenique, Saito, and Tserenjigmid (2018), each product $x \in X$ is associated with an intrinsic utility $u(i) > 0$ and a perception priority level $l(x) \in \{1, 2\}$. The idea is that customers perceive products sequentially, first those with priority 1 and then those with priority 2. This perception priority order could represent differences in familiarity, degree of attractiveness or salience of different products, or even in the way the products are presented. Let $X_i = \{x \in X : l(x) = i\}$ be the set of all products belonging to level $i = 1, 2$. Given an assortment $S \subseteq X$, we write $S = S_1 \uplus S_2$ with $S_1 \subseteq X_1$ and $S_2 \subseteq X_2$ to denote the fact that S is a partition consisting of two subsets S_1 and S_2.

The *Sequential Multinomial Logit Model* (**SML**) is a discrete choice model where the probability $\rho(x, S)$ of choosing a product x in an assortment S is given by:

$$\rho(x, S) = \begin{cases} \frac{u(x)}{\sum_{y \in S} u(y) + u_0} & \text{if } x \in S_1, \\ \left[1 - \frac{\sum_{z \in S_1} u(z)}{\sum_{y \in S} u(y) + u_0}\right] \cdot \frac{u(x)}{\sum_{y \in S} u(y) + u_0} & \text{if } x \in S_2. \end{cases}$$

where u_0 denotes the intrinsic utility of the no-choice option, which has a probability

$$\rho(x_0, S) = 1 - \sum_{i \in S} \rho(i, S)$$

of being chosen.

Observe that the probability of choosing a product $x \in S_1$ (which implies that $l(x) = 1$ and $x \in S$) is given by the standard MNL formula, whereas the probability of choosing a product y that belongs to the second level is given by the probability of not choosing any product belonging to the level 1 multiplied by the probability of selecting product y according the MNL again. Note that, if all the offered products belong to the same level, this model is equivalent to the classical MNL model. The SML corresponds to PALM restricted to two levels. We provide a full description of PALM in Appendix A.

Let $r : X \cup \{x_0\} \to \mathbb{R}^+$ be the revenue function which assigns a per-unit revenue to each product and let $r(x_0) = 0$. We use $R(S)$ to denote the *expected revenue* of an assortment S, i.e.,

$$R(S) = \sum_{x \in S} \rho(x, S) \cdot r(x). \tag{1}$$

The assortment optimization problem under the SML consists in finding an assortment S^* that maximizes R, i.e.,

$$S^* = \underset{S \subseteq X}{\operatorname{argmax}}\, R(S). \tag{2}$$

We use R^* to denote the maximum expected revenue, i.e.,

$$R^* = \max_{S \subseteq X} R(S). \tag{3}$$

Without loss of generality, we assume that $u(i) > 0$ in the rest of this paper. We use x_{ij} to denote the j^{th} product of the i^{th} level $(i = 1, 2)$, and m_i to denote the number of products in level i. Also, we assume that the products in each level are indexed in a decreasing order by revenue (breaking ties arbitrarily), i.e.,

$$\forall i \in \{1, 2\}, r(x_{i1}) \geq r(x_{i2}) \geq \ldots \geq r(x_{im_i}).$$

It is useful to illustrate how the SML allows for violations of the regularity condition, a property first observed by Echenique, Saito, and Tserenjigmid (2018). Our first example captures the attraction effect presented earlier.

Example 1 (Attraction Effect in the SML)**.** Consider a retail store that offers different brands of chocolate. Suppose that there is a well-known brand A and the brand B owned by the retail store. There is one chocolate bar a_1 from brand A and there are two chocolate bars b_1 and b_2 from Brand B, with b_2 being a more expensive version of b_1. When shown the assortment $\{a_1, b_1\}$, 71% of the clients purchase a_1 and 8.2% buy b_1. When shown the assortment $\{a_1, b_1, b_2\}$, customers select a_1 49.8% of the time and, surprisingly, bar b_1 increases its market share to about 10%, while bar b_2 accounts for 15% of the market. The introduction of b_2 to the assortment increases the purchasing probability of b_1, violating regularity. The numerical example can be explained with the SML as follows: Consider $A = \{a_1\}$, $B = \{b_1, b_2\}$ and $X = A \uplus B$. With $u(a_1) = 100, u(b_1) = 40, u(b_2) = 60$, and $u_0 = 1$ as the utility of the outside option, we have:

$$\rho(b_1, \{a_1, b_1\}) = \frac{40}{141} \cdot \left[1 - \frac{100}{141}\right] \approx 8.2\%.$$

and

$$\rho(b_1, \{a_1, b_1, b_2\}) = \frac{40}{201} \cdot \left[1 - \frac{100}{201}\right] \approx 10\%.$$

Hence $\rho(b_1, \{a_1, b_1\}) < \rho(b_1, \{a_1, b_1, b_2\})$ which contradicts regularity.

Our second example shows that the SML can capture the so-called *paradox of choice* or *choice overload effect* (e.g., Schwartz (2004); Chernev, Böckenholt, and Goodman (2015)): The overall purchasing probability may decrease when the assortment is enlarged. Once again, this effect cannot be explained by any random utility model and it is sometimes called the effect of "too much choice".

Example 2 (Paradox of Choice in the SML)**.** Let $X_1 = \{x_{11}\}$, $X_2 = \{x_{21}, x_{22}\}$, $X = X_1 \uplus X_2$, $u(x_{11}) = 10$,$u(x_{21}) = 1$,$u(x_{22}) = 10$, and $u_0 = 1$. We have

$$\begin{aligned}\rho(x_0, \{x_{11}, x_{21}\}) &= 1 - \rho(x_{11}, \{x_{11}, x_{21}\}) - \rho(x_{21}, \{x_{11}, x_{21}\}) \\ &= 1 - \frac{10}{12} - \left(1 - \frac{10}{12}\right) \cdot \frac{1}{12} = 0.152\overline{7},\end{aligned}$$

and

$$\begin{aligned}\rho(x_0, \{x_{11}, x_{21}, x_{22}\}) &= 1 - \rho(x_{11}, \{x_{11}, x_{21}, x_{22}\}) - \rho(x_{21}, \{x_{11}, x_{21}, x_{22}\}) - \rho(x_{22}, \{x_{11}, x_{21}, x_{22}\}) \\ &= 1 - \frac{10}{22} - \left(1 - \frac{10}{22}\right) \cdot \frac{1}{22} - \left(1 - \frac{10}{22}\right) \cdot \frac{10}{22} = 0.\overline{27}.\end{aligned}$$

Hence $\rho(x_0, \{x_{11}, x_{21}\}) < \rho(x_0, \{x_{11}, x_{21}, x_{22}\})$.

In the following section we first focus in finding properties that any optimal solution of the assortment problem for the SML must satisfy. Then, in Section 4, we use those properties to show the optimality of an extension of the classical revenue-ordered assortments which we called revenue-ordered assortments by level.

3 Properties of Optimal Assortments

In this section we derive properties of the optimal solutions to the assortment problem under the SML. These properties are extensively used in the proof of our main result (Theorem 1) in Section 4. We establish bounds on the following: any product offered on any optimal solution, and also the assortments considered on an optimal solution on each level. We assume a set of products $X = X_1 \uplus X_2$ and use the following notations

$$U(S) = \sum_{x \in S} u(x), \quad \alpha(S) = \frac{\sum_{x \in S} u(x) r(x)}{\sum_{x \in S} u(x)} = \frac{\sum_{x \in S} u(x) r(x)}{U(S)} \quad \text{and} \quad \lambda(Z, S) = \frac{U(Z)}{U(S) + u_0} \tag{4}$$

where $Z \subseteq S$ and $Z, S \subseteq X$. Note that $\alpha(S)$ is the usual MNL formula for the revenue and, when $S = \{x\}$ for some $x \in X$, $\alpha(S) = r(x)$. With these notations, the revenue of an assortment $S = S_1 \uplus S_2$ is

$$\begin{aligned} R(S) &= \frac{\alpha(S_1)U(S_1)}{U(S_1) + U(S_2) + u_0} + \frac{\alpha(S_2)U(S_2)}{U(S_1) + U(S_2) + u_0} \cdot \left(1 - \frac{U(S_1)}{U(S_1) + U(S_2) + u_0}\right) \\ &= \frac{\alpha(S_1)U(S_1)}{U(S) + u_0} + \frac{\alpha(S_2)U(S_2)}{U(S) + u_0} \cdot \left(1 - \frac{U(S_1)}{U(S) + u_0}\right). \end{aligned} \tag{5}$$

The following proposition is useful to divide a set into disjoint sets, which can then be analyzed separately.

Proposition 1. *Let $S \subseteq X$ and $S = H \cup T$ with $H \cap T = \emptyset$. We have*

$$\alpha(S) = \frac{\alpha(H)U(H) + \alpha(T)U(T)}{U(S)}. \tag{6}$$

Proof.

$$\begin{aligned} \frac{\alpha(H)U(H) + \alpha(T)U(T)}{U(S)} &= \frac{\frac{\sum_{x \in H} r(x)u(x)}{U(H)} \cdot U(H) + \frac{\sum_{x \in T} r(x)u(x)}{U(T)} \cdot U(T)}{U(S)} && \text{/using definition of } \alpha(\cdot) \\ &= \frac{\sum_{x \in H} r(x)u(x) + \sum_{x \in T} r(x)u(x)}{U(S)} && \text{/cancelling } U(H) \text{ and } U(T) \\ &= \frac{\sum_{x \in S} r(x)u(x)}{U(S)} && \text{/using that } H \cup T = S \\ &= \alpha(S). && \text{/definition of } \alpha(S) \end{aligned}$$

□

The next proposition is useful to bound expected revenues.

Proposition 2. *Let $S_1, S_2 \subseteq X$. If $\forall x \in S_1, \forall y \in S_2, r(x) \geq r(y)$, then $\alpha(S_1) \geq \alpha(S_2)$.*

Proof. If $\forall x \in S_1, \forall y \in S_2 : r(x) \geq r(y)$, then $\min_{x \in S_1} r(x) \geq \max_{y \in S_2} r(y)$. We have

$$\alpha(S_1) = \frac{\sum_{x\in S_1} u(x)r(x)}{\sum_{x\in S_1} u(x)} \geq \min_{x\in S_1} r(x) \cdot \underbrace{\frac{\sum_{x\in S_1} u(x)}{\sum_{x\in S_1} u(x)}}_{1} \geq \max_{y\in S_2} r(y) \underbrace{\frac{\sum_{y\in S_2} u(y)}{\sum_{y\in S_2} u(y)}}_{1} \geq \alpha(S_2).$$

□

The following proposition bounds the MNL revenue of the products in the first level. We use $S_i^* = S^* \cap X_i$ to denote the products in level i in the optimal assortment, i.e., $S^* = S_1^* \uplus S_2^*$.

Proposition 3 (Bounding Level 1)**.** *$\alpha(S_1^*) \geq R^*$.*

Proof. The proof shows that the optimal revenue is a convex combination of $\alpha(S_1^*)$ and another term by using Equation (5) and multiplying/dividing the revenue associated with the second level by $\frac{U(S_2^*)+u_0}{U(S_2^*)+u_0}$. We have

$$\begin{aligned}
R^* =& \frac{\alpha(S_1^*)U(S_1^*)}{U(S^*)+u_0} + \frac{\alpha(S_2^*)U(S_2^*)}{U(S^*)+u_0} \cdot \left(1 - \frac{U(S_1^*)}{U(S^*)+u_0}\right) \\
=& \frac{\alpha(S_1^*)U(S_1^*)}{U(S^*)+u_0} + \frac{\alpha(S_2^*)U(S_2^*)}{U(S_2^*)+u_0} \cdot \underbrace{\frac{U(S_2^*)+u_0}{U(S^*)+u_0}}_{\left(1-\lambda(S_1^*,S^*)\right)\in(0,1)} \cdot \left(1 - \frac{U(S_1^*)}{U(S^*)+u_0}\right) \\
=& \alpha(S_1^*)\lambda(S_1^*, S^*) + R(S_2^*)\left(1 - \lambda(S_1^*, S^*)\right)^2.
\end{aligned}$$

R^* is a convex combination of $\alpha(S_1^*)$ and $R(S_2^*)(1 - \lambda(S_1^*, S^*))$. By optimality of R^*, $R(S_2^*) \leq R^*$ and hence $\alpha(S_1^*) \geq R^*$. □

We now prove a stronger lower bound for the value $\alpha(S_2^*)$ of the second level.

Proposition 4. *(Bounding Level 2) $\alpha(S_2^*) \geq \frac{R^*}{1-\lambda(S_1^*,S^*)}$.*

Proof. The proof is similar to the one in Proposition 3.

$$\begin{aligned}
R^* =& \frac{\alpha(S_1^*)U(S_1^*)}{U(S^*)+u_0} + \frac{\alpha(S_2^*)U(S_2^*)}{U(S^*)+u_0} \cdot \left(1 - \frac{U(S_1^*)}{U(S^*)+u_0}\right) \\
=& \frac{\alpha(S_1^*)U(S_1^*)}{U(S_1^*)+u_0} \cdot \underbrace{\frac{U(S_1^*)+u_0}{U(S^*)+u_0}}_{1-\lambda(S_2^*,S^*)} + \alpha(S_2^*) \cdot \underbrace{\frac{U(S_2^*)}{U(S^*)+u_0}}_{\lambda(S_2^*,S^*)} \cdot \underbrace{\left(1 - \frac{U(S_1^*)}{U(S^*)+u_0}\right)}_{1-\lambda(S_1^*,S^*)} \\
=& R(S_1^*) \cdot (1 - \lambda(S_2^*, S^*)) + (\alpha(S_2^*)(1 - \lambda(S_1^*, S^*))) \cdot \lambda(S_2^*, S^*).
\end{aligned}$$

R^* is a convex combination of and $R(S_1^*)$ and $\alpha(S_2^*)(1-\lambda(S_1^*, S^*))$. By optimality of R^*, $R(S_1^*) \leq R^*$ and $\alpha(S_2^*) \geq \frac{R^*}{(1-\lambda(S_1^*,S^*))}$. □

The following example shows that it not always the case that the inequality proved above holds if one considers the products in S_2^* separately. That is, the inequality $r(x) \geq \frac{R^*}{1-\lambda(S_1^*,S^*)}$ for all $x \in S_2^*$ is not always true.

Example 3. Let $X_1 = \{x_{11}\}$, $X_2 = \{x_{21}, x_{22}\}$, and $X = X_1 \uplus X_2$. Let the revenues be $r(x_{11}) = 10, r(x_{21}) = 9$, and $r(x_{22}) = 6$ and the utilities be $u(x_{11}) = u(x_{21}) = 1, u(x_{22}) = 3$, and $u_0 = 1$. The expected revenue for all possible subsets are given by

S	$R(S)$
$\{x_{11}\}$	5
$\{x_{21}\}$	4.5
$\{x_{22}\}$	4.5
$\{x_{11}, x_{21}\}$	$5.\overline{3}$
$\{x_{11}, x_{22}\}$	4.88
$\{x_{21}, x_{22}\}$	5.4
$\{x_{11}, x_{21}, x_{22}\}$	$5.41\overline{6}$

The optimal assortment is $S^* = \{x_{11}, x_{21}, x_{22}\}$ with an expected revenue of $R^* = 5.41\overline{6}$. By definition of $\lambda(\cdot)$, we have

$$\lambda(S_1^*, S^*) = \frac{U(S_1^*)}{U(S^*) + u_0} = \frac{1}{6} = 0.1\overline{6}.$$

It follows that $r(x_{22}) = 6 < \frac{R^*}{1-\lambda(S_1^*, S^*)} = \frac{5.41\overline{6}}{1-0.1\overline{6}} = 6.488$, showing that the bound does not hold for product x_{22}.

However, the weaker bound holds for every product, and more generally, we have the following proposition (a proof is provided in Appendix B).

Proposition 5. *In every optimal assortment S^*, if $Z \subseteq S_i^*$ $(i = 1, 2)$, then $\alpha(Z) \geq R^*$.*

The converse of Proposition 5 does not hold: Example 4 presents an instance where the optimal solution does not contain all the products with a revenue higher than R^*.

Example 4. We show that some products with revenue greater or equal than R^* may not be included in an optimal assortment. Let $X_1 = \{x_{11}\}$, $X_2 = \{x_{21}\}$, and $X = X_1 \uplus X_2$. Let the revenues be $r(x_{11}) = r(x_{21}) = 1$ and the utilities be $u(x_{11}) = 10, u(x_{21}) = 1$, and $u_0 = 1$. Consider the possible assortments and their expected revenues:

$$\begin{aligned}
R(\{x_{11}\}) &= \tfrac{u(x_{11})r(x_{11})}{u(x_{11})+u_0} = \tfrac{10\cdot 1}{10+1} = 0.9\overline{09} \\
R(\{x_{21}\}) &= \tfrac{u(x_{21})r(x_{21})}{u(x_{21})+u_0} = \tfrac{1\cdot 1}{1+1} = 0.5 \\
R(\{x_{11}, x_{21}\}) &= \tfrac{u(x_{11})r(x_{11})}{u(x_{11})+u(x_{21})+u_0} + \left(1 - \tfrac{u(x_{11})}{u(x_{11})+u(x_{21})+u_0}\right) \cdot \tfrac{u(x_{21})r(x_{21})}{u(x_{11})+u(x_{21})+u_0} \\
&= \tfrac{10\cdot 1}{10+1+1} + \left(1 - \tfrac{10}{10+1+1}\right) \cdot \tfrac{1\cdot 1}{10+1+1} \\
&= \tfrac{10}{12} + \left(1 - \tfrac{10}{12}\right) \cdot \tfrac{1}{10+1} \\
&= 0.847\overline{2}
\end{aligned}$$

The optimal assortment is $S^* = \{x_{11}\}$. However, we have that $r(x_{21}) = 1 > R^*$, but x_{21} is not part of the optimal assortment.

The following corollary, whose proof is also in Appendix B, is a direct consequence of Proposition 5.

Corollary 1. *For any non-empty subset $S_0 \subseteq S^*$, where S^* is an optimal solution, we have $\alpha(S_0) \geq R^*$.*

The corollary above implies that $\alpha(\{x\}) = r(x) \geq R^*$ for all $x \in S^*$. Thus, every product in an optimal assortment has a revenue greater than or equal to R^*.

In the following example we show that the well known revenue-ordered assortment strategy for the assortment problem does not always lead to optimality.

Example 5 (Revenue-Ordered assortments are not optimal)**.** Let $X_1 = \{x_{11}\}$, $X_2 = \{x_{21}\}$, and $X = X_1 \uplus X_2$. Let $r(x_{11}) = 10$ and $r(x_{21}) = 12$. Let the utilities be $u(x_{11}) = 10, u(x_{21}) = 2$, and $u_0 = 1$. A direct calculation shows that the revenues for all possible assortments under this setting are:

S	$R(S)$
$\{x_{11}\}$	$9.\overline{09}$
$\{x_{21}\}$	8
$\{x_{11}, x_{21}\}$	8.12

The optimal assortment is $S^* = \{x_{11}\}$, yielding a revenue of $R^* = 9.\overline{09}$. However, the best revenue ordered assortment is $S' = \{x_{11}, x_{21}\}$, obtaining a revenue of $R' = 8.12$, this means that the revenue-ordered assortment strategy provides an approximation ratio of $\frac{R'}{R^*} \approx 89.3\%$ for this particular instance.

Up to this point, we understand some properties of optimal assortments, but we don't know the solution structure or an algorithm to calculate it. In the following section we propose a natural extension of the revenue-ordered assortment strategy, the *revenue-ordered by level assortments*, and show that guarantees optimality, and it can be computed in polynomial time.

4 Optimality of Revenue-Ordered Assortments by level

This section proves that optimal assortments under the SML are revenue-ordered by level, generalizing the traditional results for the MNL (Talluri and Van Ryzin, 2004). As a corollary, the optimal assortment problem under the SML is polynomial-time solvable.

Definition 1 (Revenue-Ordered Assortment by Level)**.** Denote by N_{ij} the set of all products in level i with a revenue of at least r_{ij} $(1 \leq j \leq m_i)$ and fix $N_{i0} = \emptyset$ by convention. A revenue-ordered assortment by level is a set $S = N_{1j_1} \uplus N_{2j_2} \subseteq X$ for some $0 \leq j_1 \leq m_1$ and $0 \leq j_2 \leq m_2$. We use $\mathcal{A}$ to denote the set of revenue-ordered assortments by level.

When an assortment $S = S_1 \uplus S_2$ is not revenue-ordered by level, it follows that

$$\exists k \in \{1, 2\}, z \in X_k \setminus S_k, y \in S_k : \quad r(z) \geq r(y).$$

We say that S has a gap, the gap is at level k, and z belongs to the gap. We now define the concept of *first gap*, which is heavily used in the proof.

Definition 2 (First Gap of an Assortment)**.** Let $S = S_1 \uplus S_2$ be an assortment with a gap and let k be the smallest level with a gap. Let $\hat{r} = \max_{y \in X_k \setminus S_k} r(y)$ be the maximum revenue of a product in level k not contained in S_k. The *first gap* of S is a set of products $G \subseteq X_k \setminus S$ defined as follows:

- If $\max_{x\in S_k} r(x) < \hat{r}$, then the gap G consists of all products with higher revenues than the products in assortment S, i.e.,

$$G = \{y \in X_k \setminus S_k \mid r(y) \geq \max_{x\in S_k} r(x)\}.$$

- Otherwise, when $\max_{x\in S_k} r(x) \geq \hat{r}$, define the following quantities:

$$r_M = \min_{\substack{x\in S_k \\ r(x)\geq\hat{r}}} r(x) \quad \text{and} \quad r_m = \max_{\substack{x\in S_k \\ r(x)\leq\hat{r}}} r(x). \tag{7}$$

  The set G contains products with revenues in $[r_m, r_M]$, i.e.,

$$G = \{y \in X_k \setminus S_k \mid r_m \leq r(y) \leq r_M\}.$$

We are now in a position to state the main theorem of this paper.

Theorem 1. *Under the SML, any optimal assortment is revenue-ordered by level.*

The proof is fairly technical, and is relegated to Appendix B. The intuition behind it is the following: Assume that S is an optimal solution with at least one gap as in Definition 2. Let G be the first gap of S, and that G occurs at level k. Define $S_k = H \cup T$ with $H, T \subseteq X_k$ and

$$H = \{x \in S_k \mid r(x) \geq \max_{g\in G} r(g)\}$$

and

$$T = \{x \in S_k \mid r(x) \leq \min_{g\in G} r(g)\}.$$

We call the set H as the *head* and the set T is called the *tail*. We prove that is always possible to select an assortment that is revenue-ordered by level and has revenue greater than $R(S)$ (contradicting optimality). The proof shows that such an assortment can be obtained either by including the gap G in S or by eliminating T from S. Figure 1 illustrates these concepts visually.

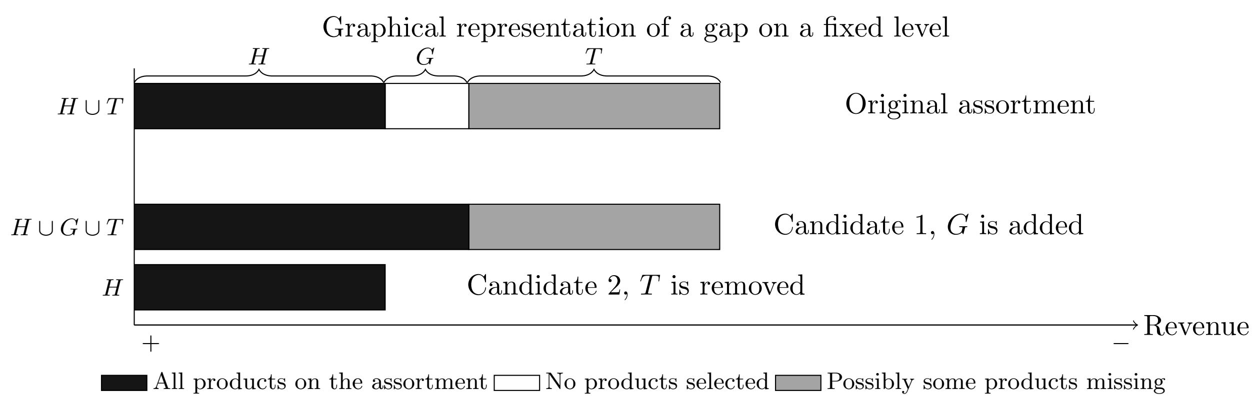

Figure 1: Representation of a level containing a gap G at the top, and the two proposed candidates fixing the gap by either adding G, or removing T.

The following corollary follows directly from the fact that there are at most $\mathcal{O}(|X|^2)$ revenue-ordered assortments by level and the fact that the revenue obtained from a given assortment can be computed in polynomial time.

Corollary 2. *The assortment problem under the sequential multinomial logit is polynomial-time solvable.*

5 Numerical Experiments

In this section, we analyse numerically the performance of revenue-ordered assortments (RO) against our proposed strategy (ROL) by varying the number of products, the distribution of revenues and utilities in each level, and the utility of the outside option. In our experiments with up to 100 products, we found that the optimality gap can be as large as 26.319%.

Each family or class of instances we tested is defined by three numbers: the number of products in the first and second level (n_1, n_2), and the utility of the outside option u_0. In total, we tested 20 classes or family of instances, each containing 100 instances. In each specific instance, revenues and product utilities are drawn from an uniform distribution between 0 and 10. We ran both strategies (RO and ROL) and we report the average and the worst optimality gap for the RO strategy, and the time taken for both strategies. These numerical experiments were conducted in Python 3.6, at a computer with 4 processors (each with 3.6 GHz CPU) and 16 GB of RAM. The computing time is reported in seconds is the average among the 100 instances in each class (or family).

(n_1, n_2)	u_0	RO			ROL
		Avg. Gap (%)	Worst Gap (%)	Avg. Time RO (s)	Avg. Time ROL (s)
(5,5)	0	0	0	0	0
(5,5)	1	1.811	18.098	0	0.001
(5,5)	2.5	3.43	17.416	0	0
(5,5)	5	3.413	13.183	0	0
(5,5)	10	1.923	12.406	0	0.001
(10,10)	0	0	0	0	0.004
(10,10)	1	3.23	20.669	0.001	0.004
(10,10)	2.5	5.613	20.359	0.001	0.004
(10,10)	5	5.975	15.521	0.001	0.004
(10,10)	10	5.347	15.694	0.001	0.004
(20,20)	0	0	0	0.003	0.04
(20,20)	1	4.331	19.427	0.003	0.04
(20,20)	2.5	8.523	21.873	0.003	0.039
(20,20)	5	9.776	19.719	0.003	0.038
(20,20)	10	9.682	18.771	0.003	0.039
(50,50)	0	0	0	0.016	0.564
(50,50)	1	6.315	26.319	0.016	0.549
(50,50)	2.5	11.662	24.117	0.016	0.55
(50,50)	5	14.94	24.445	0.016	0.551
(50,50)	10	15.543	22.232	0.016	0.547

Table 1: Numerical experiments comparing the revenue ordered assortment strategy (RO) and the revenue-ordered assortments by level (ROL, which is optimal). For each class of instances, we display the average optimality gap and the worst-case gap, as well as the computing time.

Based on the results on Table 1 we can observe the following:

1. As expected, although ROL takes more time than RO, it takes a small amount of time to solve the instances.
2. When the utility of the outside option is $u_0 = 0$, the average and worst gap are identically zero. This is because in those cases, the optimal solution is simply selecting the highest revenue product and therefore both strategies coincide.
3. The average gap is generally increasing as the outside option utility increases. With a high outside option, we typically expect to select more products to counterbalance the effect of the no choice alternative. This can amplify the difference between ROL and RO as the likelihood that the optimal solution of the revenue-ordered by level is indeed a revenue ordered assortment decreases.

6 Conclusion and Future Work

This paper studied the assortment optimization problem under the *Sequential Multinomial Logit* (SML), a discrete choice model that generalizes the multinomial logit (MNL). Under the SML model, products are partitioned into two levels. When a consumer is presented with such an assortment, she first considers products in the first level and, if none of them is appropriate, products in the second level. The SML is a special case of the Perception Adjusted Luce Model (PALM) recently proposed by Echenique, Saito, and Tserenjigmid (2018). It can explain many behavioural phenomena such as the attraction, compromise, and similarity effects which cannot be explained by the MNL model or any discrete choice model based on random utility.

The paper showed that the seminal concept of revenue-ordered assortments can be generalized to the SML. In particular, the paper proved that all optimal assortments under the SML are revenue-ordered by level, a natural generalization of revenue-ordered assortments. As a corollary, assortment optimization under the SML is polynomial-time solvable. This result is particularly interesting given that the SML does not satisfy the regularity condition.

The main open issue regarding this research is to generalize the results to the PALM, which has an arbitrary number of levels. Note that one can easily extend the algorithm of revenue-ordered by level to the PALM model and it would take at most $\mathcal{O}(|X|^k)$ time where k is the number of levels [†]. We executed our algorithm over a series of PALM instances by varying the number of levels and the revenue-ordered assortment algorithm always returned the optimal solution. Our conjecture is that the optimality result of revenue ordered assortments by level holds for the general PALM, but the problem remains open. We note that our proof technique cannot be directly applied to answer this question as some of the bounds developed in Section 3 do not hold for $k > 2$.

A second interesting research avenue is to consider a new discrete choice model that allows decision makers to change the order in which the levels are presented to consumers. In the SML, the level ordering is intrinsic to products, but one may consider settings in which decision makers can choose, not only what to show, but also the priority associated with each of the displayed products. Another research direction is to study the assortment optimziation problem under the SML with cardinality or space constraints. Finally, it is important to develop efficient procedures to estimate the parameters of the SML model based on historical data (e.g., van Ryzin and Vulcano (2017)).

[†]Note that $\mathcal{O}(|X|^k)$ is the number assortments that are revenue-ordered by level.

References

Abeliuk, A.; Berbeglia, G.; Cebrian, M.; and Van Hentenryck, P. 2016. Assortment optimization under a multinomial logit model with position bias and social influence. *4OR* 14(1):57–75.

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.

Aouad, A., and Segev, D. 2015. Display optimization for vertically differentiated locations under multinomial logit choice preferences. *Available at SSRN 2709652*.

Ben-Akiva, M., and Lerman, S. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press series in transportation studies. MIT Press.

Berbeglia, G., and Joret, G. 2016. Assortment optimisation under a general discrete choice model: A tight analysis of revenue-ordered assortments. *CoRR* abs/1606.01371.

Blanchet, J.; Gallego, G.; and Goyal, V. 2016. A markov chain approximation to choice modeling. *Operations Research* 64(4):886–905.

Bront, J. J. M.; Méndez-Díaz, I.; and Vulcano, G. 2009. A column generation algorithm for choice-based network revenue management. *Operations Research* 57(3):769–784.

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

Chernev, A. 2003a. Product assortment and individual decision processes. *Journal of Personality and Social Psychology* 85(1):151.

Chernev, A. 2003b. When more is less and less is more: The role of ideal point availability and assortment in consumer choice. *Journal of consumer Research* 30(2):170–183.

Daganzo, C. 1979. *Multinomial probit: the theory and its application to demand forecasting*. Academic Press, New York.

Davis, J.; Gallego, G.; and Topaloglu, H. 2013. Assortment planning under the multinomial logit model with totally unimodular constraint structures. *Department of IEOR, Columbia University. Available at* `http://www.columbia.edu/~gmg2/logit_const.pdf`.

Davis, J. M.; Gallego, G.; and Topaloglu, H. 2014. Assortment optimization under variants of the nested logit model. *Operations Research* 62(2):250–273.

Debreu, G. 1960. Review of R. D. Luce Individual Choice Behavior. *American Economic Review* 50:186–188.

Doyle, J. R.; O'Connor, D. J.; Reynolds, G. M.; and Bottomley, P. A. 1999. The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases. *Psychology and Marketing* 16(3):225–243.

Echenique, F., and Saito, K. 2015. General Luce Model. California Institute of Technology, available at `http://www.its.caltech.edu/~fede/wp/generalluce.pdf`.

Echenique, F.; Saito, K.; and Tserenjigmid, G. 2018. The perception-adjusted luce model. *Mathematical Social Sciences* 93:67 – 76.

Feldman, J., and Topaloglu, H. 2015. Bounding optimal expected revenues for assortment optimization under mixtures of multinomial logits. *Production and Operations Management* 24(10):1598–1620.

Flores, A.; Berbeglia, G.; and Van Hentenryck, P. 2017. Assortment Optimization under the General Luce Model. Working paper, available at `https://arxiv.org/abs/1706.08599`.

Gallego, G., and Li, A. 2017. Attention, consideration then selection choice model. Working paper, available at SSRN: `https://ssrn.com/abstract=2926942`.

Gallego, G., and Topaloglu, H. 2014. Constrained assortment optimization for the nested logit model. *Management Science* 60(10):2583–2601.

Gallego, G.; Li, A.; Truong, V.-A.; and Wang, X. 2016. Approximation algorithms for product framing and pricing. Technical report, Submitted to Operations Research (2016).

Garey, M. R., and Johnson, D. S. 1979. *Computers and intractability : a guide to the theory of NP-completeness*. W. H. Freeman, first edition.

Greene, W. H., and Hensher, D. A. 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological* 37(8):681 – 698.

Haynes, G. A. 2009. Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction. *Psychology and Marketing* 26(3):204–212.

Iyengar, S. S., and Lepper, M. R. 2000. When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology* 79(6):995.

Jagabathula, S. 2014. Assortment optimization under general choice. *Available at SSRN* 1598–1620.

Kök, A.; Fisher, M. L.; and Vaidyanathan, R. 2005. Assortment planning: Review of literature and industry practice. In *Retail Supply Chain Management*. Kluwer.

Li, G.; Rusmevichientong, P.; ; and Topaloglu, H. 2015. The d-level nested logit model: Assortment and price optimization problems. *Operations Research* 63(2):325–342.

Luce, R. D. 1959. *Individual Choice Behavior*. John Wiley and Sons.

Manzini, P., and Mariotti, M. 2014. Stochastic choice and consideration sets. *Econometrica* 82(3):1153–1176.

McCausland, W. J., and Marley, A. 2013. Prior distributions for random choice structures. *Journal of Mathematical Psychology* 57(3–4):78 – 93.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics* 105–142.

Méndez-Díaz, I.; Miranda-Bront, J. J.; Vulcano, G.; and Zabala, P. 2014. A branch-and-cut algorithm for the latent-class logit assortment problem. *Discrete Applied Mathematics* 164:246–263.

Newman, J. P.; Ferguson, M. E.; Garrow, L. A.; and Jacobs, T. L. 2014. Estimation of choice-based models using sales data from a single firm. *Manufacturing & Service Operations Management* 16(2):184–197.

Ratliff, R. M.; Venkateshwara Rao, B.; Narayan, C. P.; and Yellepeddi, K. 2008. A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue and Pricing Management* 7(2):153–171.

Rusmevichientong, P., and Topaloglu, H. 2012. Robust assortment optimization in revenue management under the multinomial logit choice model. *Operations Research* 60(4):865–882.

Rusmevichientong, P.; Shmoys, D.; Tong, C.; and Topaloglu, H. 2014. Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management* 23(11):2023–2039.

Rusmevichientong, P.; Shen, Z.-J. M.; and Shmoys, D. B. 2010. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations research* 58(6):1666–1680.

Schwartz, B. 2004. *Paradox of Choice: Why More is Less*. HarperCollins.

Sethi-Iyengar, S.; Huberman, G.; and Jiang, W. 2004. How much choice is too much? contributions to 401 (k) retirement plans. *Pension design and structure: New lessons from behavioral finance* 83:84–87.

Shah, A. M., and Wolford, G. 2007. Buying behavior as a function of parametric variation of number of choices. *Psychological science* 18(5):369.

Simonson, I., and Tversky, A. 1992. Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research* 29(3):281. Last update - 2013-02-23.

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

Tversky, A., and Simonson, I. 1993. Context-dependent preferences. *Management Science* 39(10):1179–1189.

Tversky, A. 1972a. Choice by elimination. *Journal of Mathematical Psychology* 9(4):341 – 367.

Tversky, A. 1972b. Elimination by aspects: a theory of choice. *Psychological Review* 79(4):281–299.

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. Technical note—an expectation-maximization method to estimate a rank-based choice model of demand. *Operations Research*.

Vulcano, G.; van Ryzin, G.; and Chaar, W. 2010. Om practice—choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management* 12(3):371–392.

Wang, R., and Sahin, O. 2018. The impact of consumer search cost on assortment planning and pricing. *Management Science* 64(8):3649–3666.

Williams, H. C. 1977. On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and planning A* 9(3):285–344.

A The Perception-Adjusted Luce Model

In this section, for the purpose of completeness, we describe the perception-adjusted Luce model (PALM) proposed by in Echenique, Saito, and Tserenjigmid (2018). The authors recover the role of perception among alternatives using a weak order $\succsim$. The idea is that if $x \succ y$ then x tends to be perceived before y, and whenever $x \sim y$ then x and y are perceived at the same time.

A *Perception-Adjusted Luce Model* (PALM) is described by two parameters: a weak order $\succsim$, and a utility function u. She perceives elements of an offered set $S \subseteq X$ sequentially according to the equivalence classes induced by $\succsim$ (in our representation, we called them *levels*). Each alternative is selected with probability defined by μ, a function depending on u and closely related to Luce's formula.

Definition 3. A *Perception-Adjusted Luce Model* (**PALM**) is a pair $(\succsim, u)$, where the probability of choosing a product x when offering the set S is:

$$\rho(x, S) = \mu(x, S) \cdot \prod_{\alpha \in S/\succsim : \alpha > x} (1 - \mu(\alpha, S)), \tag{8}$$

where:

$$\mu(x, S) = \frac{u(x)}{\sum_{y \in S} u(y) + u_0}. \tag{9}$$

$S/\succsim$ corresponds to the set of equivalence classes in which $\succsim$ partitions S.

Thus, the probability of choosing a product is the probability of not choosing any products belonging to the previous levels and then selecting the product according a Luce's Model considering all the offered products. We can also write the probability of not choosing any product of the assortment:

$$\rho(x_0, S) = 1 - \sum_{y \in S} \rho(y, S), \tag{10}$$

or equivalently, as the probability of not choosing on each one of the equivalence classes.

$$\rho(x_0, S) = \prod_{\alpha \in S/\succsim} (1 - \mu(\alpha, S)) \tag{11}$$

In the Perception-Adjusted Luce Model, the customer makes her selection following a sequential procedure. She considers the alternatives in sequence, following a predefined perception priority order. Choosing an alternative is conditioned to not choosing any other alternative perceived before. If none of the offered alternatives is selected, then the outside option is chosen.

Is interesting to note that setting the outside option utility to zero does not result in zero choice probability for the outside option. As explained in Echenique, Saito, and Tserenjigmid (2018), there are two sources behind choosing the outside option in the PALM. One is the utility of the outside option, which is to the same extent as in Luce's model with an outside option. The second, and to our appreciation the one that differentiate this model from other models of choice, is due the sequential nature of choice. When a customer chooses sequentially following a perception priority order, it can happen that she checks all products in the offered set without making a choice. When this occurs, this seems to increase or bias the value of the outside option probability.

Another consequence of the functional form of the outside option probability (Equation (11)), is the ability to model choice overload. In addition, this model allows violations to regularity and violations to stochastic transitivity. The interested reader is referred to Echenique, Saito, and Tserenjigmid (2018) for more details.

B Appendix: Proofs

In this section we provide the proofs missing from the main text.

Proof of Proposition 5. The proof of this proposition relies on the following technical lemma.

Lemma B.1. *Consider an assortment $S = S_1 \uplus S_2$ and $Z \subseteq S_i$ for some $i = 1, 2$. $R(S)$ can be expressed in terms of the following convex combinations:*

- *If $Z \subseteq S_1$,*

$$\begin{aligned} R(S) =& R(S \setminus Z) \cdot (1 - \lambda(Z, S)) \\ &+ \left[\alpha(Z) - \frac{\alpha(S_2)U(S_2)(U(S_2) + u_0)(1 - \lambda(Z, S)}{(U(S) - U(Z) + u_0)^2} \right] \cdot \lambda(Z, S). \end{aligned} \tag{12}$$

- *if $Z \subseteq S_2$,*

$$\begin{aligned} R(S) =& R(S \setminus Z) \cdot (1 - \lambda(Z, S)) \\ &+ \left[\frac{\alpha(Z)(U(S_2) + u_0)}{U(S) + u_0} + \frac{\alpha(S_2 \setminus Z)(U(S_2) - U(Z))}{U(S) - U(Z) + u_0} \cdot \frac{U(S_1)}{U(S) + u_0} \right] \cdot \lambda(Z, S). \end{aligned} \tag{13}$$

Proof. For $Z \subseteq X$ and $Z_0 \subseteq Z$, we have:

$$\begin{aligned} \alpha(Z \setminus Z_0) &= \frac{\sum_{x \in Z \setminus Z_0} r(x)u(x)}{U(Z \setminus Z_0)} \\ &= \frac{\sum_{x \in Z} r(x)u(x)}{U(Z \setminus Z_0)} - \frac{\alpha(Z_0)U(Z_0)}{U(Z \setminus Z_0)} \\ &= \frac{\sum_{x \in Z} r(x)u(x)}{U(Z \setminus Z_0)} \cdot \frac{U(Z)}{U(Z)} - \frac{\alpha(Z_0)U(Z_0)}{U(Z \setminus Z_0)} \\ &= \underbrace{\frac{\sum_{x \in Z} r(x)u(x)}{U(Z)}}_{\alpha(Z)} \cdot \frac{U(Z)}{U(Z) - U(Z_0)} - \frac{\alpha(Z_0)U(Z_0)}{U(Z) - U(Z_0)} \\ &= \frac{\alpha(Z)U(Z) - \alpha(Z_0)U(Z_0)}{U(Z) - U(Z_0)} \\ &= \frac{\alpha(Z)U(Z) - \alpha(Z_0)U(Z_0)}{U(Z) - U(Z_0)}, \end{aligned} \tag{14}$$

which can be rewritten as

$$\alpha(Z \setminus Z_0)(U(Z) - U(Z_0)) = \alpha(Z)U(Z) - \alpha(Z_0)U(Z_0). \tag{15}$$

Note also that, when $Z_0 = Z$, $\alpha(Z \setminus Z_0) = \alpha(\emptyset) = 0$.

The rest of the proof is by case analysis on the level. If $Z \subseteq S_1$, $\lambda(Z, S) = \frac{U(Z)}{U(S)+u_0}$. We have:

$$\begin{aligned}
R(S) =& \frac{\alpha(S_1)U(S_1)}{U(S)+u_0} + \frac{\alpha(S_2)U(S_2)}{U(S)+u_0} \cdot \left(1 - \frac{U(S_1)}{U(S)+u_0}\right) \\
=& \frac{\alpha(S_1)U(S_1) - \alpha(Z)U(Z)}{U(S)+u_0} + \frac{\alpha(Z)U(Z)}{U(S)+u_0} + \frac{\alpha(S_2)U(S_2)}{U(S)+u_0} \cdot \left(1 - \frac{U(S_1)}{U(S)+u_0}\right) \\
=& \frac{\alpha(S_1)U(S_1) - \alpha(Z)U(Z)}{U(S)+u_0} \cdot \frac{U(S)-U(Z)+u_0}{U(S)-U(Z)+u_0} + \frac{\alpha(Z)U(Z)}{U(S)+u_0} + \frac{\alpha(S_2)U(S_2)}{U(S)+u_0} \cdot \left(1 - \frac{U(S_1)}{U(S)+u_0}\right) \\
=& \frac{\alpha(S_1)U(S_1) - \alpha(Z)U(Z)}{U(S)-U(Z)+u_0} \cdot (1 - \lambda(Z,S)) + \alpha(Z)\lambda(Z,S) \\
&+ \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)}{(U(S)-U(Z)+u_0)^2} \cdot \left(\frac{U(S)-U(Z)+u_0}{U(S)+u_0}\right)^2,
\end{aligned}$$

where we first add and subtract $\frac{\alpha(Z)U(Z)}{U(S)+u_0}$, and multiply and divide the first term by $(U(S) - U(Z) + u_0)$. The last step uses the definition of $\lambda(Z, S)$ and multiplies and divides the last term by $(U(S) - U(Z) + u_0)^2$. Now applying Equation (15) to S_1 and Z in the last equation, we obtain

$$\begin{aligned}
R(S) =& \frac{\alpha(S_1 \setminus Z)(U(S_1) - U(Z))}{U(S)-U(Z)+u_0} \cdot (1 - \lambda(Z,S)) + \alpha(Z)\lambda(Z,S) + \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)}{(U(S)-U(Z)+u_0)^2} \cdot (1 - \lambda(Z,S))^2 \\
=& \underbrace{\left[\frac{\alpha(S_1 \setminus Z)(U(S_1) - U(Z))}{U(S)-U(Z)+u_0} + \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)}{(U(S)-U(Z)+u_0)^2}\right]}_{R(S_1 \setminus Z \cup S_2)} \cdot (1 - \lambda(Z,S)) \\
&+ \left[\alpha(Z) - \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)\,(1 - \lambda(Z,S))}{(U(S)-U(Z)+u_0)^2}\right] \cdot \lambda(Z,S) \\
=& R(S_1 \setminus Z \cup S_2)\,(1 - \lambda(Z,S)) + \left[\alpha(Z) - \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)\,(1 - \lambda(Z,S))}{(U(S)-U(Z)+u_0)^2}\right] \cdot \lambda(Z,S) \\
=& R(S \setminus Z)\,(1 - \lambda(Z,S)) + \left[\alpha(Z) - \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)\,(1 - \lambda(Z,S))}{(U(S)-U(Z)+u_0)^2}\right] \cdot \lambda(Z,S).
\end{aligned}$$

If $Z \subseteq S_2$, the proof is essentially similar. It also uses $\lambda(Z, S) = \frac{U(Z)}{U(S)+u_0}$ and apply Equation (15) to S_2 and Z to obtain

$$\begin{aligned}
R(S) =& \frac{\alpha(S_1)U(S_1)}{U(S)+u_0} + \frac{\alpha(S_2)U(S_2)}{U(S)+u_0} \cdot \left(1 - \frac{U(S_1)}{U(S)+u_0}\right) \\
=& \frac{\alpha(S_1)U(S_1)}{U(S)-U(Z)+u_0} \cdot (1 - \lambda(Z,S)) + \frac{\alpha(Z)U(Z)(U(S_2)+u_0)}{(U(S)+u_0)^2} \\
&+ \frac{(\alpha(S_2)U(S_2) - \alpha(Z)U(Z))(U(S_2)+u_0)}{(U(S)+u_0)^2} \\
=& \frac{\alpha(S_1)U(S_1)}{U(S)-U(Z)+u_0} \cdot (1 - \lambda(Z,S)) + \frac{\alpha(S_2 \setminus Z)(U(S_2)-U(Z))(U(S_2)-U(Z)+u_0)}{(U(S)+u_0)^2} \\
&+ \frac{\alpha(S_2 \setminus Z)(U(S_2)-U(Z))U(Z)}{(U(S)+u_0)^2} + \frac{\alpha(Z)U(Z)(U(S_2)+u_0)}{(U(S)+u_0)^2},
\end{aligned}$$

where we multiply and divide the first term by $(U(S)-U(Z)+u_0)$ and use the definition of $\lambda(Z, S)$ and then add and subtract $\frac{\alpha(Z)U(Z)(U(S_2)+u_0)}{(U(S)+u_0)^2}$. The last step uses Equation (15) and adds and subtracts $\frac{\alpha(S_2 \setminus Z)(U(S_2)-U(Z))U(Z)}{(U(S)+u_0)^2}$. The goal of this manipulation is to form $R(S \setminus Z)\,(1 - \lambda(Z, S))$. We then obtain

$$
\begin{aligned}
=&\frac{\alpha(S_1)U(S_1)}{U(S)-U(Z)+u_0}\cdot(1-\lambda(Z,S))+\frac{\alpha(S_2\setminus Z)(U(S_2)-U(Z))(U(S_2)-U(Z)+u_0)}{(U(S)+u_0)^2}\\
&+\frac{\alpha(S_2\setminus Z)(U(S_2)-U(Z))U(Z)}{(U(S)+u_0)^2}+\frac{\alpha(Z)U(Z)(U(S_2)+u_0)}{(U(S)+u_0)^2}\\
=&\frac{\alpha(S_1)U(S_1)}{U(S)-U(Z)+u_0}\cdot(1-\lambda(Z,S))+\frac{\alpha(S_2\setminus Z)(U(S_2)-U(Z))(U(S_2)-U(Z)+u_0)}{(U(S)-U(Z)+u_0)^2}\,(1-\lambda(Z,S))^2\\
&+\frac{\alpha(S_2\setminus Z)(U(S_2)-U(Z))U(Z)}{(U(S)+u_0)^2}+\frac{\alpha(Z)U(Z)(U(S_2)+u_0)}{(U(S)+u_0)^2}\\
=&\underbrace{\left[\frac{\alpha(S_1)U(S_1)}{U(S)-U(Z)+u_0}+\frac{\alpha(S_2\setminus Z)(U(S_2)-U(Z))(U(S_2)-U(Z)+u_0)}{(U(S)-U(Z)+u_0)^2}\right]}_{R(S_1\cup S_2\setminus Z)}\cdot(1-\lambda(Z,S))\\
&+\frac{\alpha(Z)U(Z)(U(S_2)+u_0)}{(U(S)+u_0)^2}+\frac{\alpha(S_2\setminus Z)(U(S_2)-U(Z))U(Z)}{(U(S)+u_0)^2}\\
&-\frac{\lambda(Z,S)\,(1-\lambda(Z,S))\,\alpha(S_2\setminus Z)(U(S_2)-U(Z))(U(S_2)-U(Z)+u_0)}{(U(S)-U(Z)+u_0)^2},
\end{aligned}
$$

where the second step multiplies and divides the second term by $(U(S)-U(Z)+u_0)$ and uses the definition of $\lambda(Z,S)$. The third step splits the second term by expressing $(1-\lambda(Z,S))^2$ as $(1-\lambda(Z,S))-\lambda(Z,S)(1-\lambda(Z,S))$ in order to identify the term $R(S_1\cup S_2\setminus Z)$. Now, putting together the remaining terms, we obtain:

$$
\begin{aligned}
R(S)=&R(S_1\cup S_2\setminus Z)\,(1-\lambda(Z,S))\\
&+\lambda(Z,S)\left[\frac{\alpha(Z)(U(S_2)+u_0)}{U(S)+u_0}+\frac{\alpha(S_2\setminus Z)(U(S_2)-U(Z))}{U(S)+u_0}\right.\\
&\left.\quad-\frac{(1-\lambda(Z,S))\,\alpha(S_2\setminus Z)(U(S_2)-U(Z))(U(S_2)-U(Z)+u_0)}{(U(S)-U(Z)+u_0)^2}\right]\\
=&R(S_1\cup S_2\setminus Z)\,(1-\lambda(Z,S))\\
&+\lambda(Z,S)\cdot\left[\frac{\alpha(Z)(U(S_2)+u_0)}{U(S)+u_0}+\frac{\alpha(S_2\setminus Z)(U(S_2)-U(Z))}{U(S)+u_0}\left[1-\left(1-\frac{U(S_1)}{U(S)-U(Z)+u_0}\right)\right]\right]\\
=&R(S\setminus Z)\,(1-\lambda(Z,S))+\left[\frac{\alpha(Z)(U(S_2)+u_0)}{U(S)+u_0}+\frac{\alpha(S_2\setminus Z)(U(S_2)-U(Z))}{U(S)-U(Z)+u_0}\cdot\frac{U(S_1)}{U(S)+u_0}\right]\cdot\lambda(Z,S).
\end{aligned}
$$

where the second line uses the definition of $\lambda(Z,S)$ to factorize the expression and the last step just simplifies the resulting expressions. □

Now we can continue with the proof of Proposition 5. Let $Z\subseteq S_i^*$ for some $i=1,2$. Consider first the case in which the optimal solution S^* contains only products from level i ($i\in\{1,2\}$).

Then,

$$
\begin{aligned}
R^* &= \frac{\sum_{y\in S^*} r(y)u(y)}{\sum_{y\in S^*} u(y)+u_0} \\
&= \frac{\sum_{y\in S^*\setminus Z} r(y)u(y)}{\sum_{y\in S^*\setminus Z} u(y)+u_0}\cdot\frac{\sum_{y\in S^*} u(y)-U(Z)+u_0}{\sum_{y\in S^*} u(y)+u_0}+\frac{\alpha(Z)U(Z)}{\sum_{y\in S^*} u(y)+u_0} \\
&= \underbrace{\frac{\sum_{y\in S^*\setminus Z} r(y)u(y)}{\sum_{y\in S^*\setminus Z} u(y)+u_0}}_{R(S^*\setminus Z)}\cdot(1-\lambda(Z,S^*))+\alpha(Z)\lambda(Z,S^*) \\
&= R(S^*\setminus Z)\cdot(1-\lambda(Z,S^*))+\alpha(Z)\lambda(Z,S^*).
\end{aligned}
$$

The optimal solution is a convex combination of $R(S^*\setminus Z)$ and $\alpha(Z)$. By optimality of R^*, $R(S^*\setminus Z)\le R^*$ and hence $\alpha(Z)\ge R^*$.

Consider the case in which the solution is non-empty in both levels, and suppose that $\alpha(Z)<R^*$. We now show that this is not possible. The proof considers two independent cases, depending on the level that contains Z.

If $Z\subseteq S_1^*$, by Lemma B.1, the revenue of S^* can be expressed as

$$
R(S^*) = R(S^*\setminus Z)\cdot(1-\lambda(Z,S^*))+\underbrace{\left[\alpha(Z)-\frac{\alpha(S_2^*)U(S_2)(U(S_2^*)+u_0)(1-\lambda(Z,S^*)}{(U(S^*)-U(Z)+u_0)^2}\right]}_{\Gamma_Z}\cdot\lambda(Z,S^*). \tag{16}
$$

R^* is a convex combination of $R(S^*\setminus Z)$ and Γ_Z. We show that $\Gamma_Z<R^*$.

$$
\begin{aligned}
\Gamma_Z &= \alpha(Z)-\frac{\alpha(S_2^*)U(S_2^*)(U(S_2^*)+u_0)\,(1-\lambda(Z,S^*))}{(U(S^*)-U(Z)+u_0)^2} \\
&= \alpha(Z)-\frac{\alpha(S_2^*)U(S_2^*)(U(S_2^*)+u_0)}{(U(S^*)-U(Z)+u_0)(U(S^*)+u_0)} && \text{/using definition of } \lambda(Z,S^*) \\
&= \alpha(Z)-\frac{\alpha(S_2^*)U(S_2^*)(1-\lambda(S_1^*,S^*))}{(U(S^*)-U(Z)+u_0)} && \text{/by definition of } \lambda(S_1^*,S^*) \\
&\le \alpha(Z)-\frac{R^*U(S_2^*)}{U(S^*)-U(Z)+u_0} && \text{/Using proposition 4} \\
&< R^*\left(1-\frac{U(S_2^*)}{U(S^*)-U(Z)+u_0}\right) && \text{/using the assumption } \alpha(Z)<R^* \\
&< R^*.
\end{aligned}
$$

Since $R(S^*/Z)\le R^*$, we have that $R^*<R^*$ and hence it must be the case that $\alpha(Z)\ge R^*$. Now, if $Z\subseteq S_2^*$, by Lemma B.1, the revenue of S^* can be expressed as

$$
\begin{aligned}
R(S^*) &= R(S^*\setminus Z)\,(1-\lambda(Z,S^*)) \\
&\quad+\underbrace{\left[\frac{\alpha(Z)(U(S_2^*)+u_0)}{U(S^*)+u_0}+\frac{\alpha(S_2^*\setminus Z)(U(S_2^*)-U(Z))}{U(S^*)-U(Z)+u_0}\cdot\frac{U(S_1^*)}{U(S^*)+u_0}\right]}_{\Gamma_Z}\cdot\lambda(Z,S^*).
\end{aligned} \tag{17}
$$

R^* is thus a convex combination of $R(S^* \setminus Z)$ and Γ_Z. Again, we show that $\Gamma_Z < R^*$:

$$\begin{aligned}
\Gamma_Z &= \frac{\alpha(Z)(U(S_2^*)+u_0)}{U(S^*)+u_0} + \frac{\alpha(S_2^* \setminus Z)(U(S_2^*)-U(Z))}{U(S^*)-U(Z)+u_0} \cdot \frac{U(S_1^*)}{U(S^*)+u_0} \\
&= \alpha(Z)(1-\lambda(S_1^*,S^*)) + \frac{\alpha(S_2^* \setminus Z)(U(S_2^*)-U(Z))}{U(S^*)-U(Z)+u_0} \cdot \lambda(S_1^*,S^*) && \text{/by definition of } \lambda(S_1^*,S^*) \\
&< \alpha(Z)(1-\lambda(S_1^*,S^*)) + \underbrace{\frac{\alpha(S_2^* \setminus Z)(U(S_2^*)-U(Z))}{U(S_2^*)-U(Z)+u_0}}_{R(S_2^* \setminus Z)} \cdot \lambda(S_1^*,S^*) && \text{/replacing } U(S^*) \text{ by } U(S_2^*) \\
&< R^*(1-\lambda(S_1^*,S^*)) + R(S_2^* \setminus Z) \cdot \lambda(S_1^*,S^*) && \text{/using that } \alpha(Z) < R^* \\
&< R^*(1-\lambda(S_1^*,S^*)) + R(S_2^* \setminus Z)\lambda(S_1^*,S^*) && \text{/Using the optimality of } R^* \\
&< R^*.
\end{aligned}$$

Hence, it must be the case that $\alpha(Z) \geq R^*$, completing the proof.

□

Proof of Corollary 1. By Proposition 5, we have $\alpha(\{x\}) = r(x) \geq R^*$ for all $x \in S^*$. For each set $S_0 \subseteq S^*$, we have:

$$\begin{aligned}
\alpha(S_0) =& \frac{\sum_{x \in S_0} u(x) r(x)}{\sum_{x \in S_0} u(x)} \\
\geq& R^* \frac{\sum_{x \in S_0} u(x)}{\sum_{x \in S_0} u(x)} \quad \text{/ by Proposition 5} \\
=& R^*.
\end{aligned}$$

□

Proof of Theorem 1. Assume that S is an optimal solution with at least one gap, G is the first gap of S, and G occurs at level k. Define $S_k = H \cup T$ with $H, T \subseteq X_k$ and

$$H = \{x \in S_k \mid r(x) \geq \max_{g \in G} r(g)\}$$

and

$$T = \{x \in S_k \mid r(x) \leq \min_{g \in G} r(g)\}.$$

In the following, the set H is called the *head* and the set T is called the *tail*. We prove that is always possible to select an assortment that is revenue-ordered by level and has revenue greater than $R(S)$. The proof shows that such an assortment can be obtained either by including the gap G in S or by eliminating T from S. Figure 1 illustrates these concepts visually. The proof is by case analysis on the level of G.

Consider first the case where G is in the first level. We can define $S = S_1 \uplus S_2$, with $S_1 = H \cup T$ as defined above and $S_2 \subseteq X_2$. The revenue for S is

$$\begin{aligned}
R(S_1 \cup S_2) &= \frac{\alpha(S_1)U(S_1)}{U(S)+u_0} + \left(1 - \frac{U(S_1)}{(U(S)+u_0)}\right) \cdot \frac{\alpha(S_2)U(S_2)}{U(S)+u_0} \\
&= \frac{\alpha(H)U(H)}{U(S)+u_0} + \frac{\alpha(T)U(T)}{U(S)+u_0} + \frac{U(S_2)+u_0}{(U(S)+u_0)^2} \cdot \alpha(S_2)U(S_2)
\end{aligned}$$

where we used Proposition 1 on S_1 for deriving the second equality. We show that assortment $H \cup S_2$ or assortment $H \cup G \cup T \cup S_2$ provides a revenue greater than $R(S = H \cup T \cup S_2)$, contradicting our optimality assumption for S. The proof characterizes the differences between the revenues of S and the two considered assortments, adds those two differences, and shows that this value is strictly less than zero, implying that at least one of the differences is strictly negative and hence that one of these assortments has a revenue larger than $R(S)$. $R(H \cup S_2)$ can be expressed as

$$\frac{\alpha(H)U(H)}{U(H)+U(S_2)+u_0} + \frac{U(S_2)+u_0}{(U(H)+U(S_2)+u_0)^2} \cdot \alpha(S_2)U(S_2). \tag{18}$$

Let $\theta = U(H) + U(T) + U(S_2) + u_0$ (or, equivalently, $\theta = U(S) + u_0$). The difference $R(H \cup T \cup S_2) - R(H \cup S_2)$ is

$$\frac{U(T)}{\theta(\theta - U(T))} \left[-\alpha(H)U(H) + \alpha(T)(\theta - U(T)) - \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)(2(\theta - U(T)) + U(T))}{\theta(\theta - U(T))} \right]. \tag{19}$$

$R(H \cup G \cup T \cup S_2)$ can be expressed as

$$\frac{1}{U(S)+U(G)+u_0} \cdot [\alpha(H)U(H) + \alpha(G)U(G) + \alpha(T)U(T)] + \frac{U(S_2)+u_0}{(U(S)+U(G)+u_0)^2} \cdot \alpha(S_2)U(S_2).$$

The difference $R(H \cup T \cup S_2) - R(H \cup G \cup T \cup S_2)$ is given by

$$\frac{U(G)}{\theta(\theta + U(G))} \cdot \left[\alpha(H)U(H) + \alpha(T)U(T) - \alpha(G)\theta + \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)(2\theta + U(G))}{\theta(\theta + U(G))} \right]. \tag{20}$$

By optimality of S, these two differences must be positive. However, their sum, dropping the positive multiplying term on each difference, which must also be positive, is given by

$$\begin{aligned}
(19)+(20) &= \alpha(T)\theta - \alpha(G)\theta + \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)}{\theta} \cdot \left[\frac{2\theta + U(G)}{\theta + U(G)} - \frac{2(\theta - U(T)) + U(T)}{(\theta - U(T))} \right] \\
&= (\alpha(T) - \alpha(G))\theta + \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)}{\theta} \cdot \left[1 + \frac{\theta}{\theta + U(G)} - 2 - \frac{U(T)}{(\theta - U(T))} \right] \\
&= \underbrace{(\alpha(T) - \alpha(G))\theta}_{\leq 0, \text{ by Proposition 2 and } \theta \geq 0} + \frac{\alpha(S_2)U(S_2)(U(S_2)+u_0)}{\theta} \cdot \left[\underbrace{\left(\frac{\theta}{\theta + U(G)} - 1 \right)}_{<0} - \underbrace{\frac{U(T)}{(\theta - U(T))}}_{<0} \right] < 0
\end{aligned}$$

which contradicts the optimality of S.

Consider now the case where the gap is in the second level. Using the definition of the head and the tail discussed above, S can be written as $S = S_1 \uplus H \cup T$. The revenue $R(S)$ is given by

$$R(S_1 \cup H \cup T) = \frac{\alpha(S_1)U(S_1)}{\theta} + \frac{\alpha(H)U(H)}{\theta} + \frac{\alpha(T)U(T)}{\theta} - \frac{U(S_1)\alpha(H)U(H)}{\theta^2} - \frac{U(S_1)\alpha(T)U(T)}{\theta^2} \tag{21}$$

and the proof follows the same strategy as for the case of the first level. The revenue $R(S_1 \cup H)$ is given by

$$R(S_1 \cup H) = \frac{\alpha(S_1)U(S_1)}{\theta - U(T)} + \frac{\alpha(H)U(H)}{\theta - U(T)} - \frac{U(S_1)\alpha(H)U(H)}{(\theta - U(T))^2} \tag{22}$$

and the difference $R(S_1 \cup H \cup T) - R(S_1 \cup H)$ by

$$\begin{aligned}
&\frac{U(T)}{\theta(\theta - U(T))} \cdot \Bigg[-\alpha(S_1)U(S_1) - \alpha(H)U(H) + \alpha(T)(\theta - U(T)) \\
&\quad + \frac{\alpha(H)U(H)U(S_1)(2\theta - U(T))}{\theta(\theta - U(T))} - \frac{U(S_1)\alpha(T)(\theta - U(T))}{\theta} \Bigg]
\end{aligned} \tag{23}$$

The revenue $R(S_1 \cup H \cup G \cup T)$ is given by

$$\frac{\alpha(S_1)U(S_1)}{\theta+U(G)}+\frac{\alpha(H)U(H)}{\theta+U(G)}+\frac{\alpha(G)U(G)}{\theta+U(G)}+\frac{\alpha(T)U(T)}{\theta+U(G)}-\frac{U(S_1)}{(\theta+U(G))^2}\cdot[\alpha(H)U(H)+\alpha(G)U(G)+\alpha(T)U(T)] \tag{24}$$

and the difference $R(S_1 \cup H \cup T) - R(S_1 \cup H \cup G \cup T)$ by

$$\begin{aligned}&\frac{U(G)}{\theta(\theta+U(G))}\cdot\Bigg[\alpha(S_1)U(S_1)+\alpha(H)U(H)-\alpha(G)\theta+\alpha(T)U(T)\\ &-\frac{\alpha(H)U(H)U(S_1)(2\theta+U(G))}{\theta(\theta+U(G))}-\frac{\alpha(T)U(T)U(S_1)(2\theta+U(G))}{\theta(\theta+U(G))}+\frac{U(S_1)\alpha(G)\theta}{\theta+U(G)}\Bigg]\end{aligned} \tag{25}$$

Adding (23) and (25) and dropping the positive multiplying terms on each difference gives

$$\begin{aligned}&\theta\left(\alpha(T)-\alpha(G)\right)+U(S_1)\left(\alpha(G)-\alpha(T)\right)+U(S_1)\left[\frac{\alpha(T)U(T)}{\theta}-\frac{\alpha(G)U(G)}{\theta+U(G)}\right]\\ &+\frac{U(S_1)}{\theta}\left[\alpha(H)U(H)\left(1+\frac{\theta}{\theta-U(T)}\right)-\alpha(H)U(H)\left(1+\frac{\theta}{\theta+U(G)}\right)-\alpha(T)U(T)\left(1+\frac{\theta}{\theta+U(G)}\right)\right]\\ &=\left(\theta-U(S_1)\right)\left(\alpha(T)-\alpha(G)\right)+\frac{U(S_1)\alpha(H)U(H)}{\theta-U(T)}-\frac{U(S_1)}{\theta+U(G)}\cdot[\alpha(H)U(H)+\alpha(G)U(G)+\alpha(T)U(T)]\\ &=\underbrace{\left(\theta-U(S_1)\right)\left(\alpha(T)-\alpha(G)\right)}_{\leq 0,\text{ by Proposition 2 and }\theta\geq U(S_1)}+\underbrace{\frac{U(S_1)\alpha(H)U(H)(U(G)+U(T))}{(\theta-U(T))(\theta+U(G))}-\frac{U(S_1)}{\theta+U(G)}\cdot[\alpha(G)U(G)+\alpha(T)U(T)]}_{\Gamma}\end{aligned} \tag{26}$$

Γ cannot be greater or equal than zero, since otherwise

$$\begin{aligned}&\frac{U(S_1)\alpha(H)U(H)(U(G)+U(T))}{(\theta-U(T))(\theta+U(G))}-\frac{U(S_1)}{\theta+U(G)}\cdot[\alpha(G)U(G)+\alpha(T)U(T)]\geq 0\\ &\frac{U(S_1)}{(\theta+U(G))}\cdot\left[\frac{\alpha(H)U(H)(U(G)+U(T))}{(\theta-U(T))}-(\alpha(G)U(G)+\alpha(T)U(T))\right]\geq 0.\end{aligned} \tag{27}$$

The factor on the left is always positive, so Inequality (27) implies that the term between brackets is greater than zero. We now show that this contradicts the optimality of S. We do this by manipulating Inequality (27) and showing that, if this inequality holds, then $R(H) > R(S)$.

$$\begin{aligned}&\frac{\alpha(H)U(H)(U(G)+U(T))}{(\theta-U(T))}-(\alpha(G)U(G)+\alpha(T)U(T))\geq 0\\ &\frac{\alpha(H)U(H)}{(\theta-U(T))}\geq\frac{\alpha(G)U(G)+\alpha(T)U(T)}{(U(G)+U(T))}\\ &\frac{\alpha(H)U(H)}{(U(S_1)+U(H)+u_0)}\geq\frac{\alpha(G)U(G)+\alpha(T)U(T)}{(U(G)+U(T))}\\ &\underbrace{\frac{\alpha(H)U(H)}{(U(H)+u_0)}}_{R(H)}\cdot\left[1-\frac{U(S_1)}{U(S_1)+U(H)+u_0}\right]\geq\frac{\alpha(G)U(G)+\alpha(T)U(T)}{(U(G)+U(T))}\\ &R(H)\geq\underbrace{R(H)\cdot\frac{U(S_1)}{U(S_1)+U(H)+u_0}}_{>0}+\frac{\alpha(G)U(G)+\alpha(T)U(T)}{(U(G)+U(T))}>R(S)\cdot\frac{U(G)+U(T)}{U(G)+U(T)}>R(S).\end{aligned} \tag{28}$$

Inequality (28) follows from Proposition 5 applied to $T \subset S_2$, which implies $\alpha(T) \geq R(S)$, and from Proposition 2, which implies $\alpha(G) \geq \alpha(T)$ and hence $\alpha(G) \geq R(S)$.

$\square$