Assortment Optimization under the Multi-Choice Rank List Model – Practical Application at CurveCatch

Guilherme Oliveira

Dissertation written under the supervision of professor
Joren Gijsbrechts

Dissertation in partial fulfilment of requirements for the
MSc in Business Analytics,
at the Universidade Católica Portuguesa,
December 31, 2022.

Assortment Optimization under the Multi-Choice Rank List Model – Practical Application at CurveCatch

Guilherme Oliveira

December 31, 2022

Abstract

In today's highly competitive market, retailers are under significant pressure to determine which products will most effectively satisfy the needs and preferences of their customers to maximize profits given strategical and operational limitations. Most of the assortment planning approaches proposed to help businesses understand customer behaviour are based on discrete choice models. However, many choice models assume that a customer can only purchase at most one product, which in some cases is not an accurate reflection of the real-world purchasing behaviour. In this paper I quantify the benefit of accounting for multi-choice behaviour in rank-based choice models and measure the impact that business requirements have on the optimal assortment. Based on the numerical experiment using secondary data provided by CurveCatch, an e-commerce lingerie retailer, I demonstrate that multi-choice modelling significantly improves the revenue generated by the assortment. Furthermore, I provide insight into the implementation of strategic and operational constraints and their impact on the optimal assortment.

Keywords: Assortment Optimization, Assortment Planning, Multi-choice Behavior, Non-parametric Choice, Choice Models, Product Assortment, Demand Substitution, Consumer Choice, Preference List

Assortment Optimization under the Multi-Choice Rank List Model – Practical Application at CurveCatch

Guilherme Oliveira

December 31, 2022

Abstract

Num mundo atual extremamente competitivo, os retalhistas estão sob uma pressão significativa para selecionar os produtos que vão satisfazer as necessidades e as preferências dos seus consumidores da forma mais eficaz de forma a maximizar os lucros dadas as limitações estratégicas e operacionais do seu negócio. Grande parte das abordagens propostas para ajudar as empresas a compreender o comportamento dos seus clientes baseia-se em modelos de escolha discreta. No entanto, a maior parte dos modelos de escolha parte do pressuposto que cada cliente pode apenas comprar no máximo um produto, o que em alguns casos não reflete de forma realística os comportamentos dos consumidores no mundo real. Nesta tese, eu quantifico o benefício associado em permitir que um cliente compre mais que um produto em modelos de escolha baseados em rankings e para além disso, meço o impacto que as limitações de negócio têm sobre a receita associada à gama de produtos ótima. Através da simulação numérica com base em dados fornecidos pela CurveCatch, uma empresa retalhista de roupa interior focada no comércio eletrónico, eu demonstro que permitir que um cliente compre mais que um produto melhora significativamente a receita gerada pela gama de produtos. Paralelamente, demonstro o impacto que a imposição dos requisitos estratégicos e operacionais pode ter na gama de produtos ótima.

Keywords: Assortment Optimization, Assortment Planning, Multi-choice Behavior, Non-parametric Choice, Choice Models, Product Assortment, Demand Substitution, Consumer Choice, Preference List

Table of Contents

List of Tables

List of Figures

1 Introduction

In today's highly competitive market, retailers face a multitude of complex challenges. In order to successfully navigate this complexity, it is imperative that businesses have a thorough understanding of consumers' dynamic needs and expectations, and are able to adapt accordingly. In recent years the increasing heterogeneity of consumer preferences coupled with the proliferation of stock-keeping units have forced retailers to re-evaluate their approach to the product mix. Traditionally, retailers have reviewed the selection of products periodically using historical sales data for each stock-keeping unit (SKU) whilst forecasting the demand for each SKU independently. Unfortunately, this approach is rather limited since it neglects the potential impact that the availability of other products can have on the overall demand of a particular product.

Given the importance of understanding substitution behaviour, there has been a significant amount of research focused on developing discrete choice models that capture this behaviour. However, often due to tractability reasons, the vast majority of discrete choice models are constructed on the underlying assumption that each consumer will purchase at most one product within a category (single-choice behaviour) which may not accurately reflect their behaviour. Lin et al. (2022), provides a rare example where multi-choice behaviour within a category is taken into consideration. Through the multi-choice rank list model (MC-RLM), Lin et al. (2022) demonstrates the benefits of considering the multi-choice behaviour using numerical experiments on real-world data for 11 different categories.

I explore the MC-RLM model proposed by Lin et al. (2022), and analyse the impact of accounting for customers' multi-choice behaviour on the revenue. Furthermore, I focus on formulating the MC-RLM for CurveCatch, an e-commerce lingerie retailer, in order to establish the optimal product assortment for a selling season under multiple strategic and operational constraints. CurveCatch is based in Belgium and operates with a try-before-you-buy business model that is set around a questionnaire to gather detailed information about a customer and a proprietary library to describe each product as a vector of fit and style. I use secondary data (product information and transaction history) provided by the lingerie retailer to establish realistic premises for the numerical experiments carried out. I limit the scope of my thesis to the assortment optimization problem, assuming a pre-established rank-based demand model over the set of available products, and a customer type (number of products each customer is willing to purchase). Through the numerical experiment I show that multi-choice modelling can have a significant positive impact on the revenue generated and I demonstrate how influential the strategic and operational constraints can be to both revenue and computational efforts in the assortment optimization problem.

This thesis makes several key contributions. It is the first work, to my knowledge, to implement Lin's et al. (2022) novel approach and to quantify the impact of multi-choice modelling through a numerical experiment. Additionally, this work provides insights into the implications of enforcing practical managerial decisions on the assortment. This thesis adds to a growing academic literature of multi-choice modelling as a means to capture consumer behaviour. I extend on this literature by carrying out a numerical study where consumers' behaviour is consistent with reality. The results of this study provide managerial insights into the impact of implementing practical constraints individually and collectively which is beneficial for the lingerie retailer and similar businesses that need to optimize their assortment of products and services. The critical implication is that neglecting multi-choice behaviour will harm a business profitability and can lead to the failure of a selling season. Therefore, managers should consider taking multi-choice behaviour into account, especially if it is typical for their customers to purchase multiple products. Furthermore, the failure to enforce business constraints can result in a product assortment that is misaligned with the business's strategy and operational limitations, particularly for businesses that are on a scaling phase.

The remainder of this thesis is organized as follows. In section 2, I review the literature on assortment optimization. Section 3 describes the data, the specification of the MC-RLM, and the formulation of strategic and operational constraints used in the numerical experiments. I report the results of the numerical experiments in section 4 and discuss the results in section 5. I point out the limitations in section 6 and conclude in section 7.

2 Literature Review

Assortment planning refers to the process of deciding which products to offer customers such that revenue is maximized. This is one of the most important decisions a retailer conducts since inadequate planning can lead to excessive holding costs of unpopular products or missed revenue due to leading products selling out rapidly.

It is broadly acknowledged in the Operations Management (OM) field that a two-step approach should be followed to accomplish an efficient assortment. Firstly, retailers must determine an appropriate choice model that is able to capture customers purchasing behaviour when faced with a specific set of products. Secondly, retailers must leverage an assortment optimization algorithm that derives the optimal subset of products to maximize revenue/ profit considering various context-specific constraints (e.g., purchasing budget, storage capacity, and product diversity requirements).

2.1 Choice Models

There are two predominant streams of choice models: parametric and non-parametric models. In brief, parametric models assume a fixed utility Gumbel distributed, independent of the data structure, whilst non-parametric models adopt a more flexible structure shaped by data (Berbeglia et al., 2018).

2.1.1 Parametric Choice Models

One of the most prominent classes of parametric choice models is based on the framework of the random utility model first introduced by Thurstone (1927). Under the random utility model, a customer confers a certain amount of utility U for each product j. The utility of each alternative is decomposed into two components $U_j = v_j + \varepsilon_j$, where v_j is an observable and deterministic component based on factors such as product attributes whilst ε_j is an unobservable and random component due to the inability to observe the utility consumers confer to different alternatives.

The most popular family of models under the random utility framework is the multinomial logit model (MNL), initially proposed by Luce (1959). In the MNL model, the probability that a consumer will choose the alternative $j \in S \cup \{0\}$, where S is the set of products offered is:

$$P(j|S) = \frac{\exp(v_j)}{\sum_{i \in S \cup \{0\}} \exp(v_i)}$$

Where $\exp(v_0) = 1$. The MNL is favoured due to its tractability as it can be solved efficiently using standard nonlinear optimization methods. Talluri & van Ryzin (2004) established that in unconstrained settings the optimal assortment can be attained by greedily selecting the alternatives with highest revenues. Furthermore, Rusmevichientong et al. (2010) crafted a polynomial-time method to deal with cardinality constraints where retailers are limited to a number of alternatives.

However, the simplicity of the MNL is paired with limitations, due to the independence of irrelevant alternatives (IIA) property detailed by Ben-Akiva et al. (1985). The IIA property states that the relative likelihood of choice probabilities between two products is independent of the alternatives available in the set of products. Train (2009) demonstrated how the IIA property fails to capture product substitution through the red-bus-blue-bus paradox. As an illustration, let's consider a transportation problem with two products: a car and a red bus, each with a market share of $1/2$. Suppose a blue bus is now added to the mix. According to the IIA property each alternative would now have a market share of $1/3$, as the blue bus would take an equal share from the car and the red bus. In reality, it is more plausible to assume that the blue bus will only 'cannibalize' the market share of the red bus as consumers who intend to choose a car would have been unaffected by this new alternative, keeping a market share of $1/2$, whilst consumers who intend to choose a bus will now have to decide between a blue and red bus, changing their market share to $1/4$ each.

In the last decades, several generalizations of the MNL were developed to capture substitution effects. Ben-Akiva (1974) proposed the nested logit model where similar products are distributed across different nests, implying that alternatives concentrated in the same nest are closer substitutes than alternatives in the remaining nests. Davis et al. (2014) proved that when consumer is forced to choose an alternative within the nest selected the unconstrained optimization problem is tractable. On the contrary, when a consumer has the possibility to not make a purchase after selecting a nest then the optimization problem becomes NP-Hard.

Equally important, rather than assuming all consumers behave identically, Rusmevichientong et al. (2014) presented a mixture of MNL models to capture the heterogeneity across different consumer segments. Unfortunately, this formulation causes the assortment problem to become NP-Hard even with just two customer segments, as verified by Bront et al. (2009) and Rusmevichientong et al. (2014). Finally, Blanchet et al. (2016) proposed a Markov chain model where substitution effects between products are modelled as transitions of states which can be solved optimally in an unconstrained setting through a polynomial-time algorithm. Conversely, Désir et al. (2015) demonstrated that under cardinality constraints the assortment problem under a Markov chain becomes NP-Hard.

Ultimately, the optimization of these extensions is frequently nonlinear and nonconvex and requires a strong understanding of the market structure (Jagabathula, 2014) which makes parametric choice models precarious in realistic assortment scenarios.

2.1.2 Non-Parametric Choice Models

Non-parametric choice models have drawn significant attention from the OM community due to their flexible structure and data-driven nature. In recent years, rank-based choice models received outstanding contributions and emerged as the leading non-parametric choice model. Rank-based models first introduced in the Operations Management literature by Mahajan & van Ryzin (2001 a,b) and are based on the assumption that a customer's behaviour is described by a sorted preference list of available alternatives. The customer will then opt for the alternative that is ranked higher on his/ her preference list if the alternative is available.

Learning all different preference sequences can be computationally challenging. Honhon et al. (2010) solved special cases of the rank-based model by using the shortest path solution for the assortment optimization where each preference was restricted into belonging to a path of a binary tree. Jagabathula (2011) and Farias et al. (2013) made significant steps in avoiding the need to search in a factorial large space by obtaining worst-case revenues estimates among preference distributions. This strategy showed a 20% improvement in prediction accuracy over the MNL choice models (Farias et al., 2013), however choosing the best worst-case scenario is rather limited and suboptimal.

To avoid the exponentially large number of choice alternatives van Ryzin & Vulcano (2015) proposed a market discovery algorithm based on the maximum likelihood estimation method which is solvable through an integer programming heuristic. Bertsimas & Mišić (2015) also resort to the market discovery algorithm, however they suggest instead to minimize the absolute error between predicted and historical revenue probabilities through a novel mixed-integer programming model (MIP) to reach the optimal assortment. Despite that, both approaches are vulnerable to growing computing costs according to the size of the optimization problem. Recently, Bertsimas & Mišić (2019) addressed this vulnerability through a Benders decomposition algorithm which provided for a more scalable and faster optimization in line with retailers' reality. Concurrently, Jena et al. (2020) extended the concept of rank-based models with a novel partial ranking representation that allows a customer to be indifferent over a subset of alternatives and proving that theoretically a partial ranked-based model is equivalent to a fully rank-based model.

2.2 Multi-Choice Behaviour

All the aforementioned choice models include one critical assumption: each customer may select at most one alternative out of the assortment. Although this may be true with expensive products (e.g., purchasing a car) it fails to represent most scenarios where customers choose several alternatives (e.g., purchasing several styles/ colours of underwear at once).

The distinction between single and multi-choice behaviour can result in significant changes to the assortment. For instance, when purchasing only at most one product, adding a low-profit alternative might cannibalize the demand for higher-profit products in the assortment. Whereas, assuming a multi-choice behaviour customers could opt to purchase both products and generate higher profits.

In recent years practitioners incorporated multi-choice behaviour across both parametric and non-parametric streams of Operations Management. Tulabandhula et al. (2020) proposed a Bundle Multivariate Logit model which generalizes the MNL and allows consumers to purchase bundles of at most $K \geq 1$ alternatives, where K is given exogenously. Under this formulation the assortment problem becomes NP-Hard even when consumers purchase a bundle of two products. Feldman et al. (2021) suggested a Multi-Purchase MNL choice model which introduces a budget parameter that considers the number of products a consumer is predisposed to purchase. Under this structure, it is only possible to achieve an approximated solution through polynomial-time approximation schemes.

On the other hand, Lin et al. (2022) introduced a non-parametric framework extending the mixed-integer programming formulation proposed by Bertsimas & Mišic (2019) into mixed-integer linear program. In line with the research conducted by Lin et al. (2022), I examine the impact of multi-choice modelling on the revenue. Furthermore, I investigate the consequences of implementing strategic and operational requirements in a practical setting together with the computational complexity required.

3 Data and Methodology

In this section I discuss the background of the assortment optimization problem and portray the data provided by CurveCatch. Furthermore, I describe the single-choice formulation introduced by Bertsimas and Mišić (2019) followed by the multi-choice formulation proposed by Lin et al. (2022). In addition, I illustrate the strategic and operational constraints defined by the firm. Subsequent to that, I describe the numerical methods and practical assumptions made in order to understand the extent to which multi-choice modelling can improve the revenue generated by the assortment and the computational effort that is required.

3.1 Background

To provide further context to the relevance of the assortment optimization problem for businesses, I tailored a practical example original presented by Lin et al. (2022). Consider a retailer with two customers, each customer has a list in which they rank products by their propensity to make a purchase, where the product ranked in the first preference is preferred to the product ranked in the second preference.

Table 1: Practical example – The relevance of the assortment optimization problem

Customer	*List of Preferences and Product Profit*		
	First Preference	*Second Preference*	*Third Preference*
Customer 1	Product A (14€)	Product B (10€)	Product C (9€)
Customer 2	Product C (9€)	Product B (10€)	Product A (14€)

In this hypothetical scenario, the goal is to determine the most profitable product assortment to offer customers. If the retailer assumes that each customer will behave rationally and purchase one product, then he should only offer product B, generating a profit of 20€. Additionally, in light of the single-choice behaviour, if he were to offer product B and product C, Customer 2 would purchase its first preference (Product C) instead of Product B, leading to the cannibalization of profits, as the assortment would only achieve 19€ instead of 20€ of profit. However, if the retailer assumes that both customers are interested in purchasing exactly two different products (multi-choice behaviour), then the assortment composed by Product B and Product C would originate a profit of 38€. This example serves to illustrate how neglecting multi-choice behaviour can lead to poor assortment planning and ultimately customer dissatisfaction.

3.2 Data

I use secondary data provided by CurveCatch for the purpose of assessing the extent to which multi-choice modelling affects the assortment coupled with the necessity to implement a scientific framework that supports the assortment planning of the lingerie retailer for each selling season.

The secondary data obtained from the lingerie retailer can be divided into two distinct categories: products and transactions. In the products category, the information describes the attributes of every SKU that is available to be in the assortment, as outlined in table 2 (Product characteristics). Regarding transactions, the information describes a subset of orders and details products that were actually purchased (due to the nature of the try-before-you-buy e-commerce business model) as outlined in table 3 (Purchase information).

Table 2: Product characteristics

SKU	*Vendor*	*Type*	*Size*	*Colour*	*Price*	*Cost*
1	Brand 1	Bralette	A 75	White	60	20
2	Brand 1	Bralette	B 85	Black	60	20
3	Brand 2	Plunge	D 90	Green	80	30
4	Brand 2	Plunge	J 65	Red	80	30
5	Brand 3	Push-up	H 70	Beige	100	40
6	Brand 3	Push-up	H 90	Ivory	100	40
7	Brand 3	Strapless	A 75	Blue	90	35
8	Brand 4	Strapless	A 75	Red	90	35
9	Brand 4	Balconette	A 75	Black	75	25
10	Brand 4	Strapless	N 95	Beige	125	45
11	Brand 5	Balconette	A 75	Black	85	30

Table 3: Purchase information

Order	*Customer*	*SKU*	*Purchased*
1	Customer 4	3	Yes
2	Customer 8	3	No
2	Customer 8	4	Yes
2	Customer 8	7	Yes
3	Customer 2	1	No
3	Customer 2	8	Yes
4	Customer 9	4	Yes
4	Customer 9	6	Yes
4	Customer 9	7	No
4	Customer 9	2	Yes
4	Customer 9	9	Yes

3.3 Single-Choice Model Formulation

To solve the optimal product assortment under the single-choice model, Bertsimas & Mišić (2019) propose a mixed-integer optimization model. As a starting point there are n products available to be included in the assortment coupled with the no-purchase alternative which is referred to as 0 in the set of available products $\{0,1,\dots,n\}$. Furthermore, there are K rankings over every alternative, in which there is an associated ranking σ^k. Where $\sigma^k(i)$ is the rank of product i and $\sigma^k(j)$ is the rank of product j. If $\sigma^k(i) < \sigma^k(j)$, then i is preferable to j for the ranking σ^k. In addition, λ is used to describe the probability mass function over the set of rankings $\{\sigma^1,\dots,\sigma^k\}$ with λ^k representing the relative size of segment k.

The mixed-integer optimization formulation states that x_i is defined as a binary decision variable for every product $i \in \{1,\dots,n\}$ that is 1 if product i is included in the assortment, and 0 otherwise and y_i^k is a binary decision variable that is 1 if product i is chosen in the k^{th} ranking and 0 otherwise. The proposed formulation relies on the assumptions that customers will purchase exactly one of the available alternatives (with the no-purchase alternative always being available), each customer will behave rationally and follow the ranking σ^k, and the choice probability is always well defined.

The mathematical formulation of the mixed-integer optimization proposed by Bertsimas & Mišić (2019) is as follows:

$$\underset{x,z}{maximize} \sum_{k=1}^{K} \sum_{i=1}^{N} \pi_i \cdot \lambda^k \cdot y_i^k$$

$subject\ to$:

(1a) $\sum_{i=0}^{N} y_i^k = 1, \ \forall\, k \in \{1,\dots,K\},$

(1b) $y_i^k \leq x_i, \ \forall\, k \in \{1,\dots,K\}, i \in \{1,\dots,n\},$

(1c) $\sum_{j:\, \sigma^k(j) > \sigma^k(i)} y_i^k \leq 1 - x_i, \ \forall\, k \in \{1,\dots,K\}, i \in \{1,\dots,n\},$

(1d) $\sum_{j:\, \sigma^k(j) > \sigma^k(0)} y_i^k = 0, \ \forall\, k \in \{1,\dots,n\},$

(1e) $\mathbf{Cx} \leq \mathbf{d},$

(1f) $x_i \in \{0,1\}, \ \forall\, i \in \{1,\dots,n\},$

(1g) $y_i^k \geq 0, \ \forall\, k \in \{1,\dots,K\}, i \in \{0,1,\dots,n\}.$

The objective function estimates the expected revenue generated by the assortment where π_i represents the revenue of product i. Constraint (1a) guarantees that exactly one choice is made under each ranking. Constraint (1b) certifies that under ranking k product i can only be chosen if product i is included in the assortment. Constraint (1c) assures that if product i is included in the assortment, then none of the alternatives that are less preferred to i under ranking σ^k can be chosen under ranking k. Constraint (1d) safeguards that those options that are less preferred to the no-purchase alternative cannot be chosen. Constraint (1e) ensures that the assortment respects the operational and strategic requirements of the firm through the matrix $\mathbf{C} \in \mathbb{R}^{m \times n}$ and vector $\mathbf{d} \in \mathbb{R}^m$ such that the set of admissible product assortments is encoded by all binary vectors **x** that fulfil $\mathbf{Cx} \leq \mathbf{d}$. Constraint (1f) guarantees that the decision variable x_i is binary. Constraint (1g) enforces that every y_i^k is non-negative.

3.4 Multi-Choice Model Formulation

Lin et al. (2022) refines the single-choice model proposed by Bertsimas & Mišić (2019) in order to portray a more accurate representation of consumers' behaviour.

As a point of departure, similarly to the single-choice formulation, there are n alternatives available to be included in the assortment coupled with the no-purchase alternative which is referred to as 0 in the set of available products $\{0,1,\ldots,n\}$. In the same fashion, Lin et al. (2022) designs $\boldsymbol{\sigma}$ as the collection of every possible preference ranking with $\sigma(i)$ being the rank of product i and $\sigma(j)$ being the rank of product j. If $\sigma(i) < \sigma(j)$, then i is preferable to j for the ranking σ.

Differently from the single-choice model, Lin et al. (2022) introduces the concept of intended purchase quantity as a non-negative integer q (it is 0 if the customer opts for the no-purchase alternative). This allows for a customer type to be identified as a pair $(q\,,\sigma_j^{(q)})$ in which at most q products will be purchased based under the preference ranking $\sigma_j^{(q)}$ (Note that for simplicity (q,j) will be used in the mathematical formulation to denote the customer type $(q\,,\sigma_j^{(q)})$ and $\mathcal{T} = \{\left(q\,,\sigma_{M_0}^{(0)}\right),\left(1\,,\sigma_1^{(1)}\right)\ldots,\left(1\,,\sigma_{M_1}^{(1)}\right),\ldots,\left(Q\,,\sigma_1^{(Q)}\right),\left(Q\,,\sigma_{M_Q}^{(Q)}\right)\}$, is defined as the collection of all possible customer types with $\lambda_{q,j}$ being the probability that a random customer belongs to the customer type $(q\,,\sigma_j^{(q)})$.

The multi-choice formulation states that x_i is defined as a binary decision variable for every product $i \in \{1,\ldots,n\}$ that is 1 if product i is included in the assortment, and 0 otherwise and $z_{k,i}^{(q,j)}$ is a binary decision variable that is 1 if product i is chosen under the k^{th} ranking of a customer that belongs to the type $(q, \sigma_j^{(q)})$. The proposed formulation relies on the assumptions that each customer will behave rationally and follow the ranking σ, and the choice probability is always well defined.

The mathematical formulation of the multi-choice optimization proposed by Lin et al. (2022) is as follows:

$$\underset{x,z}{maximize} \sum_{q=1}^{Q} \sum_{j=1}^{M_q} \sum_{i=1}^{N} \sum_{k=1}^{q} \pi_i \cdot \lambda_{q,j} \cdot z_{k,i}^{(q,j)}$$

$subject\ to$:

$$\text{(2a)} \sum_{i=0}^{n} z_{k,i}^{(q,j)} = 1, \quad \forall (q, j) \in \mathcal{T}, k \in \{0,1,\dots,n\},$$

$$\text{(2b)} \sum_{k=1}^{q} z_{k,i}^{(q,j)} \leq x_i, \quad \forall (q, j) \in \mathcal{T}, i \in \{1,\dots,n\},$$

$$\text{(2c)} \sum_{e:\, \sigma_j^{(q)}(e) > \sigma_j^{(q)}(i)} z_{1,e}^{(q,j)} \leq 1 - x_i, \quad \forall (q, j) \in \mathcal{T}, i \in \{1,\dots,n\},$$

$$\text{(2d)} \sum_{e:\, \sigma_j^{(q)}(e) > \sigma_j^{(q)}(i)} z_{k,e}^{(q,j)} \leq 1 - x_i + \sum_{l=1}^{k-1} z_{l,i}^{(q,j)}, \quad \forall (q, j) \in \mathcal{T}, i \in \{1,\dots,n\}, k \in \{0,\dots,n\} \backslash \{1\},$$

$$\text{(2e)} \sum_{e:\, \sigma_j^{(q)}(e) > \sigma_j^{(q)}(0)} z_{k,e}^{(q,j)} = 0, \quad \forall (q, j) \in \mathcal{T}, k \in \{0,\dots,n\},$$

$$\text{(2f)} \ \mathbf{Cx} \leq \mathbf{d},$$

$$\text{(2g)} \ x_i \in \{0,1\}, \quad \forall\, i \in \{1,\dots,n\},$$

$$\text{(2h)} \ z_{k,i}^{(q,j)} \geq 0, \quad \forall (q, j) \in \mathcal{T}, i \in \{0\} \cup \{1,\dots,n\}, \ k \in \{0,\dots,n\}.$$

The objective function estimates the expected revenue generated by the assortment where π_i represents the revenue of product i. Constraint (2a) ensures that customers of type $(q, \sigma_j^{(q)})$ will choose exactly one product for each of the first q choice. Constraint (2b) certifies that customers of type $(q, \sigma_j^{(q)})$ will choose product i only if product i is available, and each product can at most be picked once. Constraint (2c) guarantees that customers of type $(q, \sigma_j^{(q)})$ will not choose products that are less preferred to product i in the preference ranking $\sigma_j^{(q)}$ if product i is offered. Constraint (2d) ensures that the k^{th} $(k \geq 2)$ choice of customers of type $(q\,, \sigma_j^{(q)})$ will not be products that are more preferred to the first $(k-1)^{th}$ chosen products, this constraint can generate three different scenarios: Firstly, if product i is not offered, the products that are less preferred still have an opportunity to be chosen. Secondly, if product i is offered and it is chosen previously, the products that are less preferred still have an opportunity to be picked. Finally, if product i is offered and it is not chosen previously, the products that are less preferred will not have the opportunity to be chosen. Constraint (2e) safeguards that the customers of type $(q\,, \sigma_j^{(q)})$ will not chose the products that are less preferred to the no-purchase alternative in the preference ranking $\sigma_j^{(q)}$. Constraint (2f) certifies that the assortment respects the operational and strategic requirements of the firm through the matrix $\mathbf{C} \in \mathbb{R}^{m \times n}$ and vector $\mathbf{d} \in \mathbb{R}^m$ such that the set of admissible product assortments is encoded by all binary vectors **x** that fulfil $\mathbf{Cx} \leq \mathbf{d}$. Constraint (2g) enforces that the decision variable x_i is binary. Constraint (2h) ensures that every y_i^k is non-negative.

It is important to note that if all customer types have a $q \leq 1$, then the multi-choice formulation is reduced to the formulation proposed by Bertsimas & Mišić (2019).

3.5 Strategic and Operational Constraints of CurveCatch

Due to the intricate nature of operating a business, it is necessary to consider practical constraints in order to support the execution of CurveCatch's business plan. For that reason, the lingerie retailer has identified a number of operational and strategic requirements to be enforced in the assortment.

In the first place, every size must be represented in the assortment. Secondly, it is mandatory to keep purchasing from all vendor groups (e.g., Brand1 and Brand2 are part of the VendorGroup1) meaning that at least one brand of each vendor group needs to be in the assortment. Thirdly, CurveCatch wishes to balance between the stability of permanent products (e.g., colour black, white, etc..) coupled with the novelty of fashion products (e.g., colour green, blue, etc...) and wants the assortment to be approximately 70% permanent and 30% fashion. Moreover, the lingerie retailer identifies price tiers (Low-Tier, Medium-Tier, High-Tier) and requires the assortment to be equally distributed between all of them. Additionally, every model type must be available in the assortment (e.g., strapless, plunge, etc…). Operationally, CurveCatch is currently outsourcing their fulfilment centre and it is indispensable that a capacity constraint is established (upper bound for the number of products available in the assortment) as well as budget constraint to limit the cost of the assortment.

To avoid the unnecessary clutter of mathematically formulating every strategic and operational constraint I will simply exemplify the budget constraint (3a):

$$(3a) \sum_{i=1}^{n} c_i \cdot x_i \leq Budget, \ \forall i \in \{1, \dots, n\}$$

in which c_i is the cost associated with acquiring product i and $Budget$ is a parameter that reflects the maximum amount of money that CurveCatch is willing to allocate to the assortment. Furthermore, I refer the reader to the general structure of the business constraints described in the constraint (1e) of the single-choice model proposed by Bertsimas and Mišić (2019) and in the constraint (2f) of the multi-choice model developed by Lin et al. (2022). For readers interested in the implementation of these formulations, the code is available in a repository on GitHub (Oliveira, 2022).

3.6 Numerical Methods

To investigate the extent to which multi-choice modelling can affect the assortment coupled with the potential impact of implementing strategic and operational constraints I resort to a numerical experiment where I simulate the optimal assortment for 500 customers. There are three main components to take into consideration to produce a realistic scenario: (1) product information, which is accessible and outlined in table 2 (Product Characteristics); (2) Customer types, which are obtainable through extrapolating the purchasing behaviour in table 3 (Purchase information); and (3) Customer preferences, which are reproducible based on the estimation of the demand for each product.

With regard to product information, some data cleaning and preparation was required to accommodate for the strategic and operational constraints, such as the appointment of price tiers and permanent/ fashion categories. In the light of the purchasing information (table 3), I randomly assigned a customer type to the 500 customers (assuming that each customer will not purchase more products than the customer type assigned) based on the distribution of products purchased per order (Figure 1), which was considered a suitable approximation of reality by CurveCatch.

On the other hand, extrapolating the demand for each product based on the subset of transactions available did not provide an accurate representation of the preferred products. As a result, inspired by Feldman et al. (2021), I opted to simulate weights from the log-normal distribution with location 0 and scale 1 to assign different levels of importance to different products and reflect the heterogeneity amongst preferences. Provided that the maximum number of products purchased per order is 9 (Figure 1), I randomly assigned 10 preferences to each customer. Moreover, every preference allotted ranks higher than the no-purchase alternative for the purpose of the experiment.

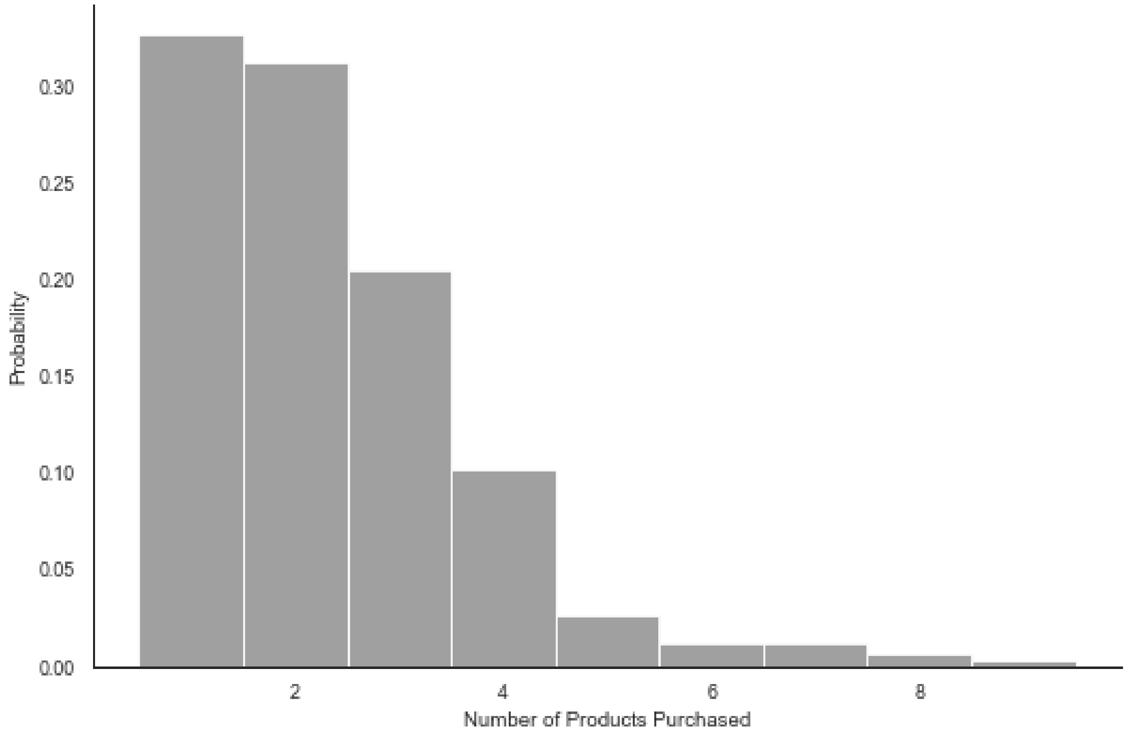

Figure 1: Distribution of products purchased per order at CurveCatch

To evaluate the potential impact of multi-choice modelling, a revenue comparison is made with multiple upper bounds for the number of products available on the assortment. On top of assessing the revenue generated by the single-choice and multi-choice model, I introduce and measure the revenue generated by the single-choice model with multi-purchase behaviour (under the assumption that every customer is able to purchase the products if the retailer includes them in their assortment). Furthermore, I conduct two sub experiments to expand upon the influence of customer preferences on the assortment. In the first sub experiment I assume that every product holds a similar demand in terms of customer preferences. Meanwhile, in the second sub experiment I devise a more realistic scenario in which customer preferences are modelled using the log-normal distribution previously described to ensure that products have different levels of demand.

To determine the influence that strategic and operational constraints have on the assortment, I select a capacity threshold and based on the premises of the second sub experiment I calculate both the individual impact of each constraint as well as the total impact if all constraints are enforced simultaneously.

Every evaluation is paired with the running time to provide insight into the computational effort required. Every optimization is made using the Gurobi Optimizer (version 9.5.1), on a computer with a processor 1.8 GHz Intel Core i7- 8550U, RAM of 8 GB, and Windows 11 as the operating system.

4 Experiment Results

4.1 The Benefit of Accounting for Multi-Choice Behaviour

I report the revenue generated by the single-choice model, single-choice model with multi-purchase behaviour, and multi-choice model with multiple upper bounds for the number of products available on the assortment, beginning with 5% (65 products) and incrementing by 5% on each subsequent iteration. The revenue of the single-choice model serves as reference, the improvement of multi-purchase behaviour reflects the actual revenue of the single-choice model under multi-purchase behaviour whilst the improvement of multi-choice model demonstrates the benefit of multi-choice modelling.

4.1.1 Sub Experiment 1 – Homogeneous Product Demand

In figure 2, I communicate the revenue at every capacity threshold where every product holds a uniform demand in terms of customer preferences (discrete uniform distribution). Similarly, in table 4, I disclose the percentage improvement for each model.

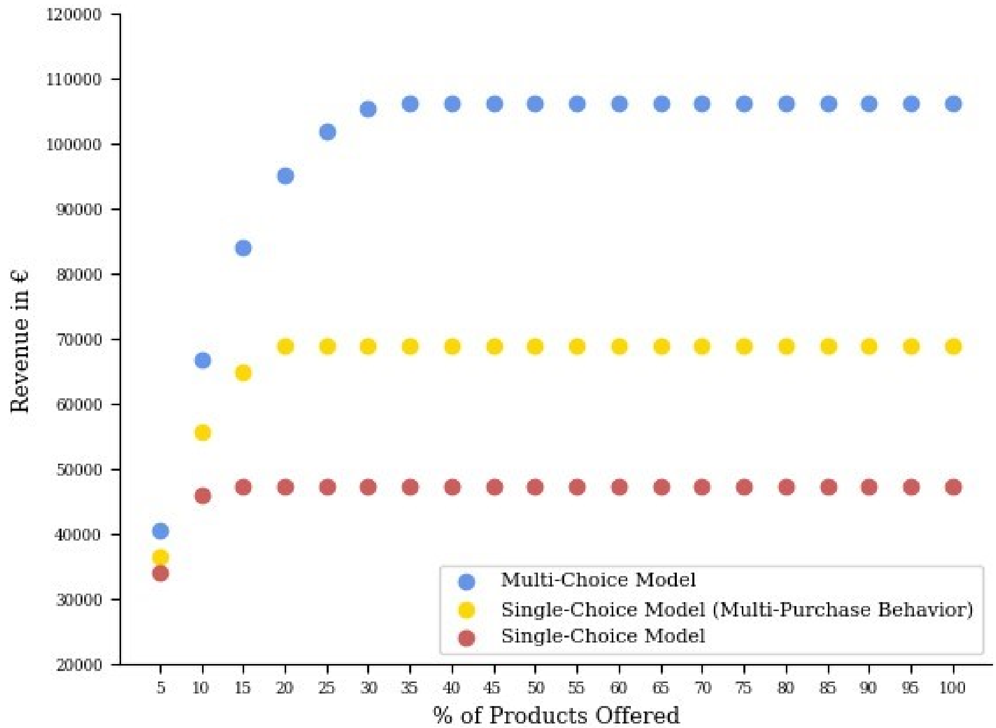

Figure 2: Sub experiment 1 – The change in revenue as the capacity threshold increases

At the most restrictive threshold (65 products) the single-choice model generates an expected revenue of 34038.50€. If multi-purchase behaviour is factored in revenue improves by approximately

6.96%. Concurrently, multi-choice modelling generates a revenue of 40363.80€, an improvement of 18.58% approximately. The revenue generated by the single-choice model reaches its maximum of 47258.50€ at the 20% threshold (260 products). However, if the multi-purchase behaviour is taken into account, the real revenue is approximately 45.62% higher. Comparatively, the revenue of the multi-choice model at the 20% threshold is 95248.50€, approximately 101.55% higher than single-choice model. In fact, the multi-choice model only reaches its maximum at the 45% threshold (586 products) with a total revenue of 106269.00€ representing an improvement of approximately 124.87% compared to the single-choice model. Beyond this threshold, increasing the number of available products does not translate into an increase in revenue since all customers have reached their maximum desired quantity of products.

Table 4: Sub experiment 1 – The effect of multi-choice behaviour on the revenue

Capacity Constraint (of Total)	*Revenue Single-Choice Model (€)*	*% Improve. Multi-Purchase Behaviour*	*% Improve. Multi-Choice Model*
65 Products (~ 5%)	34038.50	6.96	18.58
130 Products (~ 10%)	45939.80	21.12	45.38
195 Products (~ 15%)	47247.60	37.11	77.96
260 Products (~ 20%)	47258.50	45.62	101.55
325 Products (~ 25%)	47258.50	45.62	115.66
390 Products (~ 30%)	47258.50	45.62	122.90
456 Products (~ 35%)	47258.50	45.62	124.79
520 Products (~ 40%)	47258.50	45.62	124.86
586 Products (~ 45%)	47258.50	45.62	124.87
651 Products (~ 50%)	47258.50	45.62	124.87
716 Products (~ 55%)	47258.50	45.62	124.87
781 Products (~ 60%)	47258.50	45.62	124.87
846 Products (~ 65%)	47258.50	45.62	124.87
912 Products (~ 70%)	47258.50	45.62	124.87
977 Products (~ 75%)	47258.50	45.62	124.87
1042 Products (~ 80%)	47258.50	45.62	124.87
1107 Products (~ 85%)	47258.50	45.62	124.87
1172 Products (~ 90%)	47258.50	45.62	124.87
1237 Products (~ 95%)	47258.50	45.62	124.87
1303 Products (100%)	47258.50	45.62	124.87

4.1.2 Sub Experiment 2 – Heterogenous Product Demand

In figure 3, I communicate the revenue at every capacity threshold assuming a scenario where products have different levels of demand (log-normal distribution with location 0 and scale 1). Similarly, in table 5, I disclose the percentage improvement for each model.

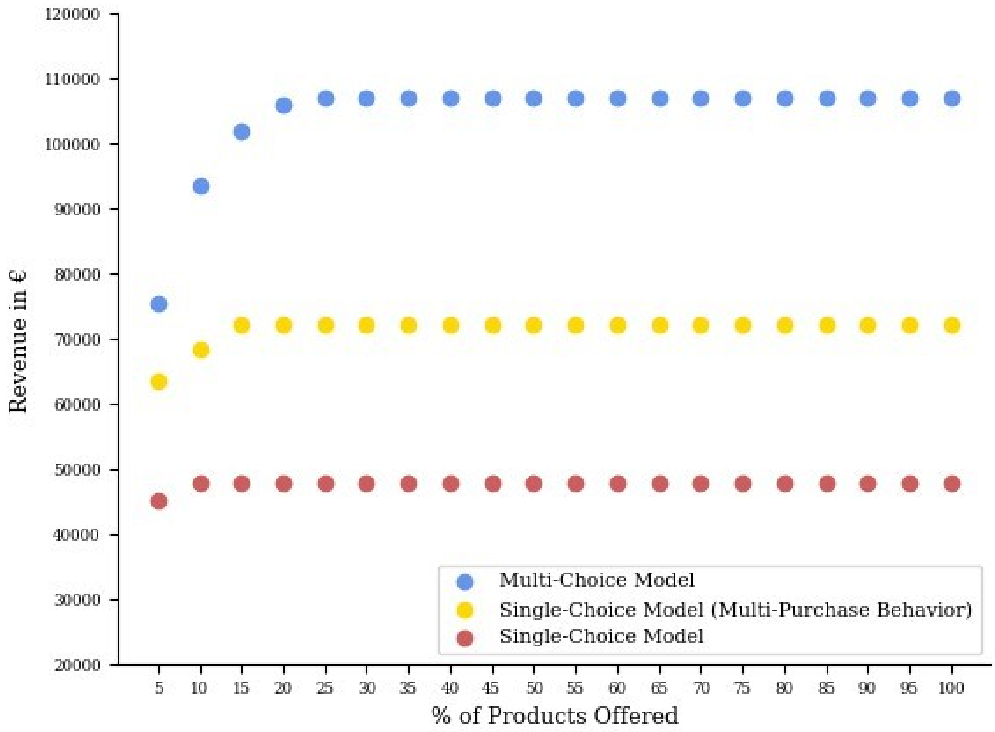

Figure 3: Sub experiment 2 – The change in revenue as the capacity threshold increases

At the most restrictive threshold (65 products) the single-choice model generates an expected revenue of 45019.20€. If multi-purchase behaviour is factored in the revenue improves by approximately 41.24%. Concurrently, multi-choice modelling generates a revenue of 75430.60€, an improvement of 67.55% approximately.

The revenue generated by the single-choice model reaches its maximum of 47779.50€ at the 15% threshold (195 products). However, if the multi-purchase behaviour is taken into account, the real revenue is approximately 51.15% higher. Comparatively, the revenue of the multi-choice model at the same threshold is 101791.00€, approximately 113.04% higher than single-choice model. In fact, the multi-choice model only reaches its maximum at the 30% threshold (390 products) with a total revenue of 106945.00€ representing an improvement of approximately 123.83% compared to the single-choice model. Beyond this threshold, increasing the number of available products does not translate into an increase in revenue since all customers have reached their maximum desired quantity of products.

In comparison with the first sub-experiment, the most notable difference is observed when only 65 products are allowed to be included in the assortment. In the first sub-experiment, where no products are clearly preferred by customers, the benefit of using the multi-choice model is approximately 18.58%. However, in sub experiment 2, a scenario where some products are more commonly purchased than others (e.g., certain bra models are more popular), the improvement of the multi-choice model increases to approximately 67.55%.

Table 5: Sub experiment 2 – The effect of multi-choice behaviour on the revenue

Capacity Constraint (of Total)	*Revenue Single-Choice Model (€)*	*% Improve. Multi-Purchase Behaviour*	*% Improve. Multi-Choice Model*
65 Products (~ 5%)	45019.20	41.24	67.55
130 Products (~ 10%)	47735.70	43.35	95.94
195 Products (~ 15%)	47779.50	51.15	113.04
260 Products (~ 20%)	47779.50	51.15	121.78
325 Products (~ 25%)	47779.50	51.14	123.80
390 Products (~ 30%)	47779.50	51.14	123.83
456 Products (~ 35%)	47779.50	51.14	123.83
520 Products (~ 40%)	47779.50	51.15	123.83
586 Products (~ 45%)	47779.50	51.15	123.83
651 Products (~ 50%)	47779.50	51.15	123.83
716 Products (~ 55%)	47779.50	51.15	123.83
781 Products (~ 60%)	47779.50	51.15	123.83
846 Products (~ 65%)	47779.50	51.15	123.83
912 Products (~ 70%)	47779.50	51.15	123.83
977 Products (~ 75%)	47779.50	51.15	123.83
1042 Products (~ 80%)	47779.50	51.15	123.83
1107 Products (~ 85%)	47779.50	51.15	123.83
1172 Products (~ 90%)	47779.50	51.15	123.83
1237 Products (~ 95%)	47779.50	51.15	123.83
1303 Products (100%)	47779.50	51.15	123.83

4.2 Computational Complexity of Multi-Choice Modelling

Together with the revenue generated for the choice models, I recorded the computational time necessary to reach the optimal solution at various capacity thresholds. As the only distinction between sub experiment 1 and sub experiment 2 is the structure of customer preferences, I only present the average and maximum computational time (in 5 runs) of sub experiment 2 in table 6.

Table 6: Average and maximum running time of the single and multi-choice models

Capacity Constraint (of Total)	*Average Model Running Time (s)*		*Maximum Model Running Time (s)*	
	Single-Choice	*Multi-Choice*	*Single-Choice*	*Multi-Choice*
65 Products (~ 5%)	242.25	9.29	251.31	10.69
130 Products (~ 10%)	7.92	25.23	8.49	25.80
195 Products (~ 15%)	7.69	22.03	7.96	24.08
260 Products (~ 20%)	7.60	15.17	8.28	16.01
325 Products (~ 25%)	7.60	16.75	8.02	17.02
390 Products (~ 30%)	7.55	15.63	8.03	16.81
456 Products (~ 35%)	7.48	14.48	7.83	15.31
520 Products (~ 40%)	7.76	14.37	8.38	14.68
586 Products (~ 45%)	7.46	14.85	7.72	15.89
651 Products (~ 50%)	7.61	14.30	8.07	14.70
716 Products (~ 55%)	7.61	14.19	8.32	14.68
781 Products (~ 60%)	7.22	14.27	7.56	14.49
846 Products (~ 65%)	7.61	14.28	8.93	14.71
912 Products (~ 70%)	7.77	14.26	8.64	14.45
977 Products (~ 75%)	7.70	14.25	8.04	14.58
1042 Products (~ 80%)	7.72	15.38	8.40	15.72
1107 Products (~ 85%)	7.51	14.97	7.78	15.14
1172 Products (~ 90%)	7.67	14.98	8.86	15.20
1237 Products (~ 95%)	7.43	14.95	8.26	15.22
1303 Products (100%)	7.37	15.06	7.54	15.23

At every capacity threshold but one, the multi-choice model requires longer computational efforts, particularly in the 10% (130 products) and 15% (195 products) thresholds where it takes approximately three times longer to compute the optimal solution compared to the single-choice model. Conversely, the 5% capacity threshold (65 products) requires longer computation time in the single-choice model. The average running time of the single-choice model at this threshold is 245.25 seconds, reaching upwards of 251.31 seconds on the slowest run. Meanwhile, the multi-choice model retrieves the optimal solution on average in 9.29 seconds.

4.3 Impact of Strategic and Operational Constraints on Optimal Assortment

4.3.1 Individual Impact on the Optimal Assortment

In table 7, I detail the individual impact of every constraint on the revenue coupled with the average and maximum computational complexity (in 5 runs) of the optimization model. As a reference, I allude to optimal revenue as the revenue generated by the multi-choice model of the sub experiment 2 at the 15% capacity threshold.

Table 7: Individual impact of strategic/ operational constraints on revenue and running time

Strategic/ Operational Constraints:	*% Gap Optimal Revenue*	*Average Running Time (s)*	*Maximum Running Time (s)*
Diversity in Sizes:			
Every Size must be Represented	3.61	81.0	84.1
Diversity in Colors			
Permanent (≥ 65%) & Fashion (≥ 25%)	1.31	484.2	515.2
Permanent (= 70%) & Fashion (= 30%)	1.31	2672.1	2844.7
Diversity in Vendors:			
All Vendor Groups must be Represented	0.01	36.3	39.1
Diversity in Price Tiers:			
Evenly Distributed (each 1/3 ± 5%)	9.95	15580.3	16745.3
Evenly Distributed (each 1/3)	11.81	25016.7	26332.4
Diversity in Model Types:			
All model types must be Represented	0.12	32.1	36.9
Maximum Budget allocated to Assortment:			
10000 €	72.38	16.2	16.8
20000 €	47.62	16.8	17.8
30000 €	23.27	56.8	60.9
40000 €	0.95	346.7	355.2
50000 €	0.00	30.7	32.1

Every requirement defined by CurveCatch influences the optimal reference revenue. The most significant factor that can impact the revenue is the maximum budget allocated to the assortment, as it indirectly limits the number of products that the lingerie retailer can purchase. For instance, a maximum budget of 10000€ results in a gap of approximately 72.38% to the optimal assortment, while a maximum budget of 20000€ results in a gap of approximately 47.62%.

However, from an optimization perspective, it is not necessary to enforce the representation of all vendor groups or model sizes in the assortment, as these requirements have almost no effect on the reference optimal revenue (approximately 0.01% and 0.12% respectively). Furthermore, the computational cost of two strategic requirements has been identified as being particularly high: the diversity of colours between permanent and fashion products, and the even distribution of price tiers. Both constraints have two formulations, and the chosen formulation can significantly impact the gap to the optimal revenue and the computational complexity required to optimize the assortment. For example, the diversity of price tiers can be enforced by requiring all tiers (Low-Tier, Medium-Tier, and High-Tier) to account for exactly 1/3 of the available products in the assortment, or alternatively to allow for a broader interval 1/3 $\pm$5%) for each tier. The more stringent formulation, besides being computationally more complex, results in a gap to the optimal revenue of approximately11.81% while the broader formulation generates a gap to the optimal revenue of approximately 9.95%.

4.3.2 Combined Impact on the Optimal Assortment

In table 8, I demonstrate the combined impact of enforcing multiple strategic and operational constraints on both revenue and computational time (in 5 runs), using the reference previously mentioned to report the individual impact.

Table 8: Combined impact of strategic/ operational constraints on revenue and running time

Strategic/ Operational Constraints:	*% Gap Optimal Revenue*	*Average Running Time (s)*	*Maximum Running Time (s)*
Diversity in Sizes: Every Size must be Represented Permanent (≥ 65%) & Fashion (≥ 25%) *Diversity in Vendors:* All Vendor Groups must be Represented *Diversity in Price Tiers:* Evenly Distributed (each 1/3 $\pm$ 5%) *Diversity in Model Types:* All model types must be Represented *Maximum Budget allocated to Assortment:* 20000 €	47.76	19.7	21.3

Through the simultaneous implementation of multiple strategic and operational constraints described in the table 8, the gap to the optimal revenue is approximately 47.76%. This result is similar to enforcing only the maximum budget of 20000€ allocated to the assortment, which was previously reported to result in a gap to the optimal revenue of approximately 46.62%. Enforcing all of the constraints simultaneously requires an average of 19.7 seconds of computational time, which is significantly faster than enforcing most of the constraints individually.

5 Discussion

The results of the numerical experiment indicate that incorporating multi-choice behaviour leads to changes in the optimal assortment and yields greater revenues when compared to the single-choice model. These findings align with those of Lin et al. (2022), who conducted experiments on real-world data across multiple categories. Additionally, this numerical experiment demonstrates that the effect of multi-choice modelling is more pronounced in situations where customer have identical preferences thus coinciding with the theoretical foundations of customer substitution behaviour which refers to the tendency of consumers to switch to an alternative product when a preferred product is not available.

The numerical experiment suggests that, on average, the computational time for multi-choice modelling is two to three times longer than the time required for single-choice modelling. This is due to the fact that multi-choice modelling involves optimizing a larger number of combinations of products, while single-choice modelling only considers the most preferred option available for each customer.

Furthermore, the numerical experiment shows that some of the strategic and operational requirements identified by CurveCatch have a minimal individual impact on the optimal assortment from an optimization perspective (e.g., ensuring representation of every size or model type) which might lead to the conclusion that these constraints are not necessary and can be omitted. However, since the lingerie retailer is a scaling business, it is important to consider that optimizing the assortment without them may result on losing potential outlier customers with uncommon preferences that are not yet represented. In addition, it is worth noting that certain requirements, such as the price range associated with each price tier or the colours that are deemed permanent or fashion products are based on human decision making, which can make them susceptible to variations if the underlying conditions for each category changes.

Equally important, the results of the numerical experiment demonstrate that the formulation of strategic and operational requirements can significantly impact both the revenue generated by the assortment and the computational complexity required (e.g., enforcing that permanent products account for exactly 70% of the assortment and fashion products to account for exactly 30% compared with allowing the permanent products to be between 65% and 75% of the assortment and the fashion products to be between 25% and 35%). It is therefore important for CurveCatch to refine these requirements and carefully consider their potential impact on the optimization process.

6 Limitations

The research presented in this study aimed to examine the impact of multi-choice modelling and provide CurveCatch with a framework for generating an effective assortment of products that accounts for the retailer's context. While the findings of this study contribute to understanding the impact of multi-choice modelling on the revenue generated, there are several limitations that should be considered when generalizing the results of the numerical experiment. Since demand estimation is not within the scope of my thesis, the list of preferences for each customer was randomly assigned and two scenarios were constructed two reflect alternative realities, assuming that every product has a similar level of demand (sub experiment 1) and assuming some products are more commonly purchased than others (sub experiment 2). However, this is not a perfect solution, as it allows customers to have products on their preference list that in reality would not be compatible (e.g., Size A60 and N90 on the list of preferences of the same customer). In the same fashion, I limit my analysis by assuming that every customer has 10 products in their list of preferences that are more preferred than the option of not making a purchase. This assumption may potentially under- or over-estimate the number of preferred products in reality. Additionally, in order to depict the willingness of customers to purchase multiple products, customer types were randomly assigned based on the distribution of purchase transactions provided by the lingerie retailer, and while this method provides a useful approximation, it is important to recognize that the number of products a customer is willing to purchase can vary and is dependent on factors such as budget (e.g., considering two products, with a price of 50€ each and a customer who is willing to spend 100€, then the customer would purchase both products. If instead each product has a price of 80€, then the customer will only be able to purchase the most preferred product).

Furthermore, it is essential to recognize the limitations of the MC-RLM since it is the optimization model suggested to generate the assortment of products for CurveCatch. The effectiveness of the MC-RLM is reliant on the accuracy of both the demand estimation model, which predicts the product preferences for each customer, and the ability to estimate the number of products each customer intends to purchase (based on customer type or budget). If either of these tasks is not accurately performed, the effectiveness of the MC-RLM is compromised. Finally, it is important to note that the optimal assortment predicted by the MC-RLM is based on the assumption that all products have available in inventory and that there are no price variations. It is therefore important to consider the impact of changes in inventory levels and pricing when implementing the assortment.

7 Conclusion

This thesis analyses to what extent does multi-choice modelling affect the revenue generated by the assortment in a capacitated scenario coupled with measuring the impact that strategic and operational constraints have on the optimal assortment of CurveCatch, an e-commerce lingerie retailer.

To estimate the extent to which multi-choice modelling affects the revenue generated by the assortment I resort to a numerical experiment that compares the expected revenue under both single and multi-choice model formulations, while also considering multiple upper bounds on the number of products available on the assortment. I use secondary data provided by the lingerie retailer to design an experiment to closely reflects the actual circumstances of the lingerie retailer and the results indicate that multi-choice modelling can the revenue generated by the assortment over 120% when compared to the revenue generated by the single-choice model, and over 72% improvement when compared to the revenue generated by the single-choice model under the multi-purchase behaviour.

Additionally, in order to assess the impact of the strategic and operational requirements defined by CurverCatch on the assortment, I select a specific upper bound and evaluate the individual effect of each constraint on the optimal revenue, as well as the collective effect if every constraint is enforced simultaneously. The results indicate that individually, the maximum budget allocated to the assortment has the greatest potential to influence optimal revenue (with a gap up to 72.38%) followed by the requirement for an even distribution of price tiers (up to 11.81%) and the requirement for every size to be represented (3.61% gap). In contrast, the remaining constraints (diversity in model types and diversity in vendor groups) have a negligible impact on the optimal revenue from an optimization perspective. Finally, collectively, the gap to the optimal revenue is largely determined by the maximum budget constraint.

From an academic perspective, this thesis contributes to the existing body of research on multi-choice modelling, by analysing and quantifying the effect of multi-choice behaviour. Furthermore, the results of this study have implications for managerial decision making since the lingerie retailer must decide what strategic and operational constraints to enforce and how to formulate them.

References

Ben-Akiva, M. (1974). Structure of Passenger Travel Demand Models. *PhD Thesis, Department of Civil Engineering, MIT*.

Ben-Akiva, M. E., Lerman, S. R., & Lerman, S. R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. In *Google Books*.

Berbeglia, G., Garassino, A., & Vulcano, G. (2018). *A Comparative Empirical Study of Discrete Choice Models in Retail Operations*.

Bertsimas, D., & Mišić, V. (2015). *Data-driven assortment optimization*.

Bertsimas, D., & Mišić, V. (2019). *Exact First-Choice Product Line Optimization*.

Blanchet, J., Gallego, G., & Goyal, V. (2016). A Markov Chain Approximation to Choice Modeling. *Operations Research*, *64*(4), 886–905.

Bront, J. J. M., Méndez-Díaz, I., & Vulcano, G. (2009). A Column Generation Algorithm for Choice-Based Network Revenue Management. *Operations Research*, *57*(3), 769–784.

Davis, J. M., Gallego, G., & Topaloglu, H. (2014). Assortment Optimization Under Variants of the Nested Logit Model. *Operations Research*, *62*(2), 250–273.

Désir, A., Goyal, V., Segev, D., & Ye, C. (2015). Capacity Constrained Assortment Optimization Under the Markov Chain Based Choice Model. *SSRN Electronic Journal*.

Farias, V. F., Jagabathula, S., & Shah, D. (2013). A Nonparametric Approach to Modeling Choice with Limited Data. *Management Science*, *59*(2).

Feldman, J., Segev, D., Topaloglu, H., Wagner, L., & Bai, Y. (2021, June 14). *Assortment Optimization under the Multi-Purchase Multinomial Logit Choice Model*.

Honhon, D., Gaur, V., & Seshadri, S. (2010). Assortment Planning and Inventory Decisions Under Stockout-Based Substitution. *Operations Research*, *58*(5), 1364–1379.

Jagabathula, S. (2011). *Nonparametric choice modeling : applications to operations management*.

Jagabathula, S. (2014). *Assortment Optimization Under General Choice*.

Jena, S. D., Lodi, A., Palmer, H., & Sole, C. (2020). A Partially Ranked Choice Model for Large-Scale Data-Driven Assortment Optimization. *INFORMS Journal on Optimization*, *2*(4), 297–319.

Lin, H., Li, X., & Wu, L. (2022). *E-Commerce Assortment Optimization and Personalization with Multi-Choice Rank List Model*.

Mahajan, S., & van Ryzin, G. (2001a). Inventory Competition Under Dynamic Consumer Choice. *Operations Research*, *49*(5).

Mahajan, S., & van Ryzin, G. (2001b). Stocking Retail Assortments Under Dynamic Consumer Substitution. *Operations Research*, *49*(3).

Oliveira, Guilherme. (2022) https://github.com/guilherme-the-analyst/Assortment-Optimization---Thesis-Code

R Duncan Luce. (1959). *Individual choice behavior*. John Wiley & Sons.

Rusmevichientong, P., Shen, Z.-J. M., & Shmoys, D. B. (2010). Dynamic Assortment Optimization with a Multinomial Logit Choice Model and Capacity Constraint. *Operations Research*, *58*(6), 1666–1680.

Rusmevichientong, P., Shmoys, D., Tong, C., & Topaloglu, H. (2014). Assortment Optimization under the Multinomial Logit Model with Random Choice Parameters. *Production and Operations Management*, *23*(11).

Talluri, K., & van Ryzin, G. (2004). Revenue Management Under a General Discrete Choice Model of Consumer Behavior. *Management Science*, *50*(1).

Thurstone, L. (1927). *A law of comparative judgment.*

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Tulabandhula, T., Sinha, D., & Patidar, P. (2020). *Multi-Purchase Behavior: Modeling and Optimization*.

van Ryzin, G., & Vulcano, G. (2015). A Market Discovery Algorithm to Estimate a General Class of Nonparametric Choice Models. *Management Science*, *61*(2).