Never Enough Toys: Inventory Management Case Study

Seminar in Quantitative Logistics
FEB23014

Catalin Gheorghiu (505458)
Teona Banu (504450)
Sigita Lapina (505514)
Daan Derksen (498381)

May 2022

Abstract

This paper considers a finite horizon, multi-item capacitated inventory lot-sizing problem for bol.com, an e-commerce platform in the Netherlands and Belgium. We look for maximized revenue and its interaction with other key commercial objectives such as service level and relevance score. We solve the problem with (and without) service level constraints to discover that only the holiday season presents a logistical challenge, that the platform can allocate substantially less space to its toys section without disrupting its objectives, and that both revenues and relevance can be maintained at very high levels without compromising on service level. When solving the problem with stochastic demand, we find that the relevance score plays a bigger role than revenue.

1 Introduction

According to Jeff Bezos, "There are two ways to extend a business. Take inventory of what you're good at and extend out from your skills. Or determine what your customers need and work backward, even if it requires learning new skills" (Lyons, 2009). Online shopping has been growing tremendously in the past years, for instance, online store like bol.com saw a 69.9% increase in the last quarter of 2020 (NL Times, 2021). The case we are tackling sees bol.com pursue the second path to extending a business - building a new warehouse and reallocating products in order to meet customers' needs.

Already since industrialization, it has been one of the goals of practitioners to have cost-efficient production plans. Already in 1913, the first work in this area was published by Harris, namely, "How many parts to make at once?" (Harris, 1913). This is known as the economic order quantity (EOQ) model. Since then, several extensions to the basic model have been published, which is also what we are looking at.

This paper considers the inventory lot-sizing problem. To be more precise, following the literature review of the field by Ullah & Parveen (2010), in this paper we look at a capacitated inventory lot-sizing problem with finite planning horizon. We have multiple items and variable order quantity. We also have continuous review, zero lead time, deterministic dynamic demand in the beginning and then we move on to stochastic demand. Lastly, we have single-echelon stocking and in this case demand can be partially lost. We arrive at two research questions, namely, *"What is the maximized revenue with respect to the capacity constraints?"* and *"How do the commercial objectives of bol.com interact?"*.

Three different models are presented in this paper. We start by having deterministic dynamic demand and looking only into the revenue-maximization. The second model looks not only into the revenue-maximization but also the relevance, we also add the service level constraints. For this model, we still have deterministic dynamic demand. The third and final model looks into both, revenue and relevance when having stochastic demand. This paper contributes to the existing literature by building new models for the already well-discussed inventory lot-sizing problem.

The paper proceeds as follows. In Section 2 we review the relevant existing literature. Section 3 provides more information on the problem and the data. Section 4 explains the approach for solving the problem by giving details of the model. Section 5 provides our computational results based on our models. Section 6 explains the extension. Section 7 concludes and discusses the limitations of our research, and, lastly, in Section 8 we give the advice and re-order policy to bol.com.

2 Literature Review

As mentioned above, over the past decades, there has been extensive research into how to solve the inventory lot-sizing problem. Similarly to us, Parveen & Haque (2007) look into the multi-item single level capacitated dynamic lot-sizing problem. It includes arranging a number of items over a time horizon of finite time periods. However, unlike us, their objective is to minimize the setup and inventory holding costs (over the time horizon). Same as for us, they also consider the capacity constraint for each of the time periods. They used the Dixon Silver heuristic and improved its inclusion of setup times. That made the feasibility problem NP-complete (nondeterministic polynomial-time complete).

Same as many others, Absi & Sidhoum (2009) have proposed to use Lagrangian relaxation to solve the NP-hard lot-sizing problem. They use the relaxation of the resource capacity constraints. They also developed a dynamic programming algorithm to solve induced sub-problems. What is more, they proposed an upper bound using the Lagrangian heuristic as well as several smoothing algorithms. This gives an insight into new methods to solve the problem, however, after careful consideration, we take another approach as we believe it suits our problem better.

The multi-item, single-level, dynamic lot-sizing problem with single capacitated resource was discussed by Hindi (1995). They developed a strategy based on branch-and-bound search with precise lower bounds. After solving the multi-item lower bound problems by column generation with the capacity constraints, many single-item, uncapacitated lot sizing problems are obtained. They are solved as shortest path problems. To obtain upper bounds, the fitting minimum-cost network flow had to be solved at each of the nodes of the branch-and-bound tree. When looking at the computation time, one can see this solution is very efficient.

Miller, Nemhauser & Savelsbergh (2000) suggests using branch-and-cut algorithms as it is oftentimes very difficult to solve multi-item lot-sizing problems to optimality. We decided to follow Miller's approach and use the Java interface of CPLEX for a robust, repeatable, and tweakable execution of the branch-and-cut method. We deem this the most appropriate approach under the relatively lax problem constraints (zero lead time, no fixed order cost, deterministic demand); should these assumptions be relaxed, other approaches will be discussed correspondingly.

3 Problem Description and Data

During this research paper, data from bol.com, an e-commerce platform in the Netherlands and Belgium, is used.

The months of November and December are called the peak season because of holidays such as Black Friday, Christmas and Sinterklaas. Then, the logistic capacity is scarce. Bol.com has to choose which products to sell and keep in stock satisfying capacity constraints. What is more, the commercial objectives must be kept in mind. They are: keeping the relevance of customers as high as possible, maximizing the revenue, and maintaining a certain service level. The service level can be defined as the probability of not hitting a stock-out. For *General Toys* (*GT*), the minimal service level is 98% and for *Recreational and Outdoor Toys* (*ROT*) the minimal service level is 95%.

There are two warehouses A and B, respectively. Warehouse A is of size 15000 m^3 and warehouse B is of size 3000 m^3. Around 15% of the storage capacity in both warehouses is for the shop *Toys*, meaning, in total we can use 2700 m^3 of storage capacity. In warehouse A, we can use 2250 m^3 of storage and in warehouse B, we can use 450 m^3 of storage. We assume that all storage capacity can be used at perfect efficiency. More assumptions are presented in Sections 4.1.1, 4.2.1. and 4.3.1.

The main dataset shows consumer sales data of the shop *Toys* from the beginning of 2018 to and including the first week of 2020. The weekly consumer sales are consolidated per chuck. There are several size groups, varying between 3XS and 3XL, meaning, nine sizes in total. Depending on this size, the cost of storing in the warehouse is determined. There is also a relevance score for every chunk which is based on several data points such as sales, conversion, and views. The product groups of the shop *Toys* are *ROT* and *GT*.

3.1 Cleaning and Refactoring

We begin from an unprocessed dataset centered around chunks, the smallest subgroup of products considered for storage. For a chunk name, the same row contains the demanded product quantity in specific, non-ordered weeks of the years 2018, 2019, and 2020, as well as the average price, average volume, and the size of this chunk on a scale from 3XS to 3XL. The warehousing costs, incurred daily as a product is stored, depend solely on the size class of the chunk and are given in a separate table. The third and last table of the dataset links a relevance cost to each chunk, to be considered as a metric in determining the optimal inventory allocation.

In its current form, the data contains several issues which might impact the outcome and therefore relevance of our investigation. Before anything else, we perform the following adjustments and alterations to the content and structure of the dataset:

- Where the demand is listed as negative, we change it to zero. This allows the respective product to be considered, but the respective week is skipped.
- For each combination of chunk and size that has listings for some but not all weeks, we consider the demand to be zero in the unlisted weeks. Similarly, for all the sizes a chunk is missing, demand equal to zero is assigned.
- For entries with a demand equal to zero, mainly the newly modified/added entries mentioned above, if no price, volume, or holding cost is listed, we set the missing values to zero. This will not impact our solutions, since the demand is zero, therefore, there will be

no respective stock during that week, so the price at which sales could have happened, the holding cost of the stock, and the volume of the stock do not matter.

- Several entries have volumes equal to zero. In the interest of upholding the laws of physics, we change these to the chunk's average over the other weeks for the respective size, or outright delete the tuples if the specific size never appears again.
- Some chunk names appear in several tuples - sometimes in the same time period - due to having different size groups or belonging to different product categories. To guarantee the uniqueness of chunks, we center tuples around category-chunk associations instead of chunk name.
- We connect the warehousing cost table to the main one based on product size. We also connect the relevance cost table to the main one based on chunk name.
- A separate table is created for each relevant parameter (demand, price, volume, and warehousing cost). Chunk-category associations are listed on the rows, with a separate tuple for each of the nine sizes, and 52 columns representing weeks are created per chunk to create a unique entry for each size of a chunk from a specific category in each time period of 2018 or 2019.

3.2 Summary Statistics

Table 1 presents the minimum, maximum, median, mean, and standard deviation of the entire dataset (including 2019 and the first week of 2020), after setting the negative entries to zero and changing one single chunk's erroneous volume of almost 200000 m^3. The aforementioned refactoring of entries with a volume of zero does virtually nothing to our results, save for making mean demand higher by one unit. This is an encouraging result as it suggests these zero-volume entries were in rather low demand to begin with and insufficiently many to skew our volume analysis, therefore we only show the summary statistics after the refactoring was performed. This seems like the more intuitive approach, given that we would like this discussion of summary statistics to reflect the numbers we will be working with for modelling and further analysis down the line.

There are two very clear trends present in all three parameters in Table 1: strong skewness to the left (indicated by the vast differences between the maximum and the mean or median) and high volatility (indicated by the relatively large standard deviations compared to the means). We proceed to explain these in detail before moving on to the model formulation.

Table 1. *Summary statistics over all time periods after adding zeros in refactoring, after removing outliers*

Parameter / Statistic	**Demand (units)**	**Price (euro)**	**Volume (m^3)**
Minimum	1.00	1.23	5.65×10^{-5}
Maximum	60012.00	1486.78	0.80
Median	13.00	12.52	0.00
Mean	140.51	21.16	0.02
Std. Dev.	826.81	29.78	0.04

First, the high concentration of values on the left tail of the parameter distributions is apparent not only in the over 8000% gaps between mean and maximum values but also in the substantial gaps between mean and median values. Therefore the overall toys section not only contains predominantly small, cheap products, but it is clear that a high percentage of the supply is very close to the minimum. These observations are perfectly within expectations for a virtual toy store. Moreover, the small toys such as notebook stickers or painting brushes are likely to be sold in massive batches, accounting both for occasionally very high demand levels and the low medians compared to the means.

Second, the large volatility suggested by the standard deviations varying between 150% and 600% of the levels of the means is not wholly unexpected given both the previously observed large ranges of values and the nature of a toy store. This simply reflects that while there are many small cheap toys in very high demand as stated in the previous paragraph, there are also some large, expensive products such as video game arcades or outdoor trampolines that a few people demand over the course of a year.

4 Methodology

In the following sections, our models will be explained. In Section 4.1 we talk about the revenue-maximization model that maximizes the sales revenue in 2018. In Section 4.2 we talk about our second model, namely, model of expanded commercial objectives. In Section 4.3 we explain our final, third, model with stochastic demand.

4.1 Model 1: 2018 revenue-maximization

As the title suggests, the aim of this first model is to formulate a weekly ordering schedule that maximizes the sales revenue in 2018. We shall investigate the service level attained by focusing on this commercial objective and evaluate the performance of the optimal ordering schedule if it is maintained in 2019.

4.1.1 Assumptions and Notation

The variable order quantity, lack of a fixed order cost, and zero lead time allow for a straightforward stocking strategy. Bol.com can order, at the beginning of each week, the exact quantity needed to satisfy as much of the week's demand as possible in a way that maximizes revenue and respects the capacity constraints. Consequently, there is no need to keep track of inventory cross-period or differentiate between what is ordered and what is sold, since goods are only ordered under the guarantee that demand exists. Given the behavior of warehousing costs as a standard holding cost, there is no revenue-increasing strategy that ever holds stock.

The problem demands that we make a critical assumption on what constitutes a unique chunk entry. As the case specifies, the product category sits above the chunk in this classification hierarchy, hence we treat chunk names appearing under both ROT and GT as distinct entries and therefore we allow for a recreational toy variant to be in a different warehouse from the general toy variant. Note that later models may operate on modified versions of past assumptions or even make new assumptions altogether; we will signal this clearly and carefully.

The relevant sets, parameters, and variables of the model are defined and explained below. An explanation of the constraints and objective function follows after the model specification. We will first define the different sets.

- K: unique associations of chunk names and product groups; ROTBordspel and GTBordspel, for instance, are two different elements $k \in K$.
- T: relevant time horizon, in this case the 52 weeks of 2018; $t = 1, 2...52$.
- S: set of sizes, for instance "XL" is an element $s \in S = \{$3XS, XXS, XS, S, M, L, XL, XXL, 3XL$\}$.
- W: set of warehouses, i.e. warehouse A is an element $w \in W = \{$A, B$\}$.

Here we define the parameters.

- $p_{skt} \in \mathbb{R}$: average price of entry k of size s in week t.
- $d_{skt} \in \mathbb{N}$: demand (given by the recorded sales in 2018) for entry k of size s in week t.
- $v_{skt} \in \mathbb{R}$: volume in m^3 of entry k of size s in week t.
- $Q_w \in \mathbb{N}$: maximum volume that toys are allowed to occupy in warehouse w.

Lastly, we define our decision variables.

- $x_{skt} \in \mathbb{N}$: quantity ordered (and immediately sold within the same period) of entry k of size s in week t.
- $y_{kw} \in \mathbb{B}$: takes on value one if all products of entry k are stored in warehouse w, zero otherwise.
- $q_{kwt} \in \mathbb{R}^+$: volume of warehouse w occupied by entry k in week t.

4.1.2 Problem Formulation

The mathematical model is presented below. Further we will refer to this model as Model 1.

$$
\begin{aligned}
\max \quad & \sum_{(s,t):d_{skt}>0} \sum_{k \in K} p_{skt} x_{skt} && && \text{(1a)}\\
\text{s.t.} \quad & x_{skt} \leq d_{skt}, && \forall s \in S, k \in K, t \in T, && \text{(1b)}\\
& \sum_{w \in W} y_{kw} = 1, && \forall k \in K, && \text{(1c)}\\
& q_{kwt} \leq y_{kw} Q_w, && \forall k \in K, w \in W, t \in T, && \text{(1d)}\\
& \sum_{s:d_{skt}>0} x_{skt} v_{skt} = \sum_{w \in W} q_{kwt}, && \forall k \in K, t \in T, && \text{(1e)}\\
& \sum_{k \in K} q_{kwt} \leq Q_w, && \forall w \in W, t \in T, && \text{(1f)}\\
& x_{skt} \in \mathbb{N}, && \forall s \in S, k \in K, t \in T, && \text{(1g)}\\
& y_{kw} \in \mathbb{B}, && \forall k \in K, t \in T, && \text{(1h)}\\
& q_{kwt} \in \mathbb{R}, && \forall k \in K, w \in W, t \in T. && \text{(1i)}
\end{aligned}
$$

The objective function (1a) maximizes revenue, derived by adding up the total price-driven gains over all sizes that a given category-chunk combination can have in every period.

Constraint (1b) ensures that what is ordered and sold in a week cannot exceed the demand; the reasons why holding stock across periods will never be profitable in an optimal solution have already been presented. Constraint (1c) ensures that every chunk within a given product category has all its items stored in exactly one of the available warehouses and that it is not possible to change the storage location between periods, since the y_{kw} decision variables do not have a time component.

Constraints (1d), (1e), and (1f) collectively address the warehouse capacities. Constraint (1e) puts together the total warehouse volume occupied by a certain chunk by summing over sizes. All q_{kwt} terms to the right of the equality except for one will be zero thanks to constraint (1d), which for a given chunk k ensures that volume can only be occupied within the warehouse this chunk is assigned to and forces all other volumes to be zero. Now that we ensure chunks can only occupy volume in the warehouse they belong to, constraint (1f) enforces the capacity constraint itself against the summed volume of all chunks belonging to a given warehouse.

Constraints (1g), (1h), and (1i) ensure that the variables are natural, binary, and real respectively.

Recall that a substantial amount of zero-valued parameters have been inserted into the dataset to account for missing sizes and weeks of demand, so as to achieve a completed tabular dataset. A notable nuance of the formulation is that all the dummy variables created for the sake of the zero-valued parameters are actually removed by CPLEX in its pre-solve phase (since dummy demands are all equal to zero and the sales thus cannot exceed zero); moreover, we use only the relevant terms in the largest of sums (objective function and constraint 1e) in pursuit of the same computational efficiency as that of the solving algorithm. We thus solve the problem only for the relevant variables of the case, ignoring dummies.

4.1.3 Service Level Derivation

In a given year, the service level is defined as the fraction of the aggregate demand that bol.com satisfies. Using previous notation, satisfied demand is equivalent to the total amount of products ordered over the time horizon, as the problem formulation still makes it optimal for the company to order at the start of the week exactly what it plans to sell in that respective week. In order to get the service level, we divide the satisfied demand by total demand, which is deterministic and known before solving the model. Formally, using the orders and demands of the specific year (2018 or 2019), we denote:

$$Service_{year} = \frac{\sum_{s \in S} \sum_{k \in K} \sum_{t \in T} x_{skt}^{(year)}}{\sum_{s \in S} \sum_{k \in K} \sum_{t \in T} d_{skt}^{(year)}}, \quad year \in \{2018, 2019\}. \tag{2}$$

4.2 Model 2: Expanding Commercial Objectives

Assuming that products can be reallocated between warehouses at the beginning of every year, we can solve a separate problem that covers 52 weeks for each of the two years of activity (2018 and 2019). The reason why we choose to do this is twofold: first, a shorter time span halves the variables in our problem formulation and is therefore more computationally doable in the context of models already taking considerable time to finish. Second, without some idea of bol.com's costs of moving items between warehouses at the end of the year and how these costs would be prioritised compared to the revenue lost by having a less refined planning, we cannot reliably judge how good keeping the same warehouse allocation all throughout really is. Consequently, the model variant presented in this subsection is representative for a single year, formally defined as a time horizon of 52 periods $t = 1, 2..., 52$ as before.

4.2.1 Assumptions and Notation

While maintaining deterministic demand and instantaneous resupply, we investigate two extensions of the previous problem: imposing a minimum yearly service level and adding relevance scores to the objective function in order to seek a balance between raw revenue and high relevance. We do so, because after analysis the first model, we saw that even though the overall demand levels were higher in 2019 than in 2018, the total revenue turned out to be lower for 2019 than for 2018. What is more, the service levels were much lower for 2019 than those of 2018. These results are presented later in Table 3.

Service Level Constraint

It is required that each product category must guarantee a minimum service level: 98% for General Toys, 95% for Recreational and Outdoor Toys. This aspect is directly specified in the case description, therefore we first try to enforce a satisfactory level through model constraints and focus on maximizing other commercial objectives. Should this fail, we elect to remove the constraint, show whatever service level is attained by maximizing the other objectives and discuss possible reasons why this lower service level could not be improved.

Pursuing the constraint option first, we expand the notation to include two disjoint subsets of the category-chunk association set K. Let $c \in C = \{GT, ROT\}$ denote an arbitrary category, and let $\bigcup_{c \in C} K_c = K$ the disjoint sets of chunks belonging to C that completely cover K. Every entry k belongs to exactly one of these sets, and we denote $V_{GT} = 0.98$ and $V_{ROT} = 0.95$ to formally represent the minimum service levels.

With the newly-established notation, we can add the following two constraints to the model in order to address the minimum service level:

$$\sum_{s \in S} \sum_{k \in K_c} \sum_{t \in T} x_{skt} \geq V_c \sum_{s \in S} \sum_{k \in K_c} \sum_{t \in T} d_{skt}, \quad \forall c \in C. \tag{3}$$

Relevance Score

We seek to find a balance between maximizing revenues and the overall relevance score of what is sold. To make these measures comparable for a given set of planned sales x_{skt}, we make use of the fraction of the maximum possible revenue achieved by the ordering schedule, as well as the fraction of the maximum possible relevance score achieved. These maxima are derived by maximizing a specific objective function under all of the constraints discussed until this point: order size bounds, single-warehouse chunk allocation, capacity constraints and service level constraints. Denoting the maximum revenue as MAX_p and the maximum relevance score as MAX_r, MAX_p is thus defined as the optimal objective value of Model 1 and MAX_r is defined as the optimal objective value if the objective function of Model 1 is changed to $\sum_{s \in S} \sum_{k \in K} \sum_{t \in T} r_{skt} x_{skt}$, where r_{skt} is the relevance score of chunk k of size s in time period t.

We now treat these two values as constants and introduce the coefficient $\alpha \in [0, 1]$ to assign weights to the two commercial objectives. Intuitively, $\alpha = 1$ or $\alpha = 0$ will result in only one objective being pursued, reducing the problem to what was already solved to obtain MAX_p and MAX_r respectively. Similarly, setting $\alpha = 0.5$ would assign equal weights to the objectives.

A spectrum of values for α will be investigated in order to solve the final objective function of this model, which maximizes the weighted sum of an ordering schedule's relative performance to pursuing purely revenue or relevance. This objective function is formally written at the beginning of the problem formulation in the next subsection. This strategy of investigating the trade-off has the advantage of showing exactly how both revenue and relevance are affected should bol.com prioritise one more heavily than the other, all while ensuring good service. As it will become apparent in the next model, this is a rather computationally expensive approach (since every level of α requires a model re-solve) and therefore not feasible if the number of variables and constraints increases substantially; however, when possible, we prefer this approach for the ease of interpreting the trade-off.

4.2.2 Problem Formulation: One-Year Horizon

Note that the meaning of α, $MAX_{p,r}$, p_{skt}, r_{skt} in the objective value, as well as the meaning of $c \in C$, K_c and V_c in constraint (4g) below have been explained in the Section 4.2.1. All other constraints are directly copied from the formulation of Model 1. Because of how the objective function is set up, the optimal solution for a fixed α will be the ordering schedule that brings both the total relevance and total revenue as close as possible to the already-known maximum. We, therefore, propose the following model. Further we refer to this model as Model 2.

$$\max \quad \frac{\alpha}{MAX_p} \sum_{(s,t):d_{skt}>0} \sum_{k \in K} p_{skt} x_{skt} + \frac{1-\alpha}{MAX_r} \sum_{(s,t):d_{skt}>0} \sum_{k \in K} r_{skt} x_{skt} \tag{4a}$$

$$\text{s.t.} \quad x_{skt} \leq d_{skt}, \qquad \forall k \in K, (s,t) \in (S,T) : d_{skt} > 0, \tag{4b}$$

$$\sum_{w \in W} y_{kw} = 1, \qquad \forall k \in K, \tag{4c}$$

$$q_{kwt} \leq y_{kw} Q_w, \qquad \forall k \in K, w \in W, t \in T, \tag{4d}$$

$$\sum_{s:d_{skt}>0} x_{skt} v_{skt} = \sum_{w \in W} q_{kwt}, \qquad \forall k \in K, t \in T, \tag{4e}$$

$$\sum_{k \in K} q_{kwt} \leq Q_w, \qquad \forall w \in W, t \in T, \tag{4f}$$

$$\sum_{s \in S} \sum_{k \in K_c} \sum_{t \in T} x_{skt} \geq V_c \sum_{s \in S} \sum_{k \in K_c} \sum_{t \in T} d_{skt}, \quad \forall c \in C, \tag{4g}$$

$$x_{skt} \in \mathbb{N}, \qquad \forall k \in K, (s,t) \in (S,T) : d_{skt} > 0, \tag{4h}$$

$$y_{kw} \in \mathbb{B}, \qquad \forall k \in K, t \in T, \tag{4i}$$

$$q_{kwt} \in \mathbb{R}, \qquad \forall k \in K, w \in W, t \in T. \tag{4j}$$

4.3 Model 3: 2020

Compared to the previous models, there are quite some new assumptions for Model 3. Similarly, as for the previous models, we start by explaining our assumptions and notation in Section 4.3.1, then we show the problem formulation for Model 3.1 in Section 4.3.2. and, lastly, we show the problem formulation for Model 3.2 in Section 4.3.3.

4.3.1 Assumptions and Notation

The biggest difference between Models 1, 2, and the new Model 3 is that the demand is not known in advance. This calls for a forecasting strategy as well as substantial changes to the model. Furthermore, up to this point, we have assumed that bol.com can order from their suppliers every week. This remains the case for most of the year, except for November and December (weeks 44-52, nine weeks in total) when, for any specific category-chunk-size entry, two separate orders cannot be issued on two consecutive weeks. Consequently, additional variables and constraints are needed to model this buffer.

Exploring Size Aggregation

The main challenge for the upcoming forecasts of 2020 demand, price, and volume levels is the small sample of historical data we can infer information from. Focusing on demand, for the time being, the most straightforward way to forecast would only use an entry's 2018 and 2019 demands - two values that would result in a third. The most practical solution for expanding this pool would be to aggregate over size, thus obtaining up to 18 values (nine per size for two years) from which a more accurate forecast would - theoretically - be obtained. We investigate the effect of this size aggregation. To begin with, we use the previous notation explained in Section 4.1.1, and add the following:

- Demand aggregate over sizes, calculated as total demand for chunk k in week t:

$$D_{kt} = \sum_{s \in S} d_{skt} \qquad \forall k \in K, t \in T.$$

- Price aggregate over sizes, calculated as the *weighted* average price of chunk k in week t:

$$P_{kt} = \frac{\sum_{s \in S} d_{skt} p_{skt}}{\sum_{s \in S} d_{skt}} \qquad \forall k \in K, t \in T.$$

- Volume aggregate over sizes, calculated as the *weighted* average volume of chunk k in week t:

$$V_{kt} = \frac{\sum_{s \in S} d_{skt} v_{skt}}{\sum_{s \in S} d_{skt}} \qquad \forall k \in K, t \in T.$$

We choose between aggregating or not based on the resulting forecast's MSE when comparing them to the real demands of the first week in 2020 (the only one with available data). The mean squared errors are computed thus:

$$MSE = \frac{1}{count} \sum_{s \in S} \sum_{k \in K} (d_{sk1} - \widehat{d_{sk1}})^2;$$

$$MSE_{aggregate} = \frac{1}{count} \sum_{k \in K} (D_{sk1} - \widehat{D_{sk1}})^2.$$

For each formula, *count* represents the number of instances for each chunk (and for each size in the raw data formula) where the forecasted value, the real value, or both are larger than zero. Here, d_{sk1} and D_{sk1} represent specifically the 2020 raw and aggregated demand respectively, while $\widehat{d_{sk1}}$ and $\widehat{D_{sk1}}$ represent the forecasts. Note that we use only data corresponding to the first week since real data in 2020 is only available for week one. We use similar formulas to compute the MSEs for price and volume.

Parameter Forecasting

Equations 5 and 6 below show our chosen method for the 2020 demand estimation from either raw or size-aggregated data. In essence, we compute the relative change in demand between 2018 and 2019 in the fraction, apply this relative change to the demand of 2019 and finally add the result to the demand of 2019. This way, the trend in demand between 2018 and 2019 - whether positive or negative - is maintained when moving from 2019 to 2020. If a chunk's demand increased in 2019, it will keep increasing in 2020 by the same relative margin, with the reverse happening in case of a previous decrease.

We make two remarks on this method. First, it incorporates the inherent seasonality of 2018/2019 demand as the 2020 peak-season demand will of course be higher if it is derived exclusively from the higher demands of 2018 and 2019. Second, time series autocorrelation is incorporated to the extent that it already existed in 2018 and 2019; if 2018 sales in a given week depended on the previous weeks' sales, then the same will be true for the 2020 demands directly derived from it. Including some autocorrelation terms in the estimate of 2020 would just amount to overfitting by enforcing the use of redundant information.

$$\widehat{d_{skt}} = d_{skt}^{(2019)}\left(1 + \frac{d_{skt}^{(2019)} - d_{skt}^{(2018)}}{d_{skt}^{(2018)}}\right) = d_{skt}^{(2019)} \times \frac{d_{skt}^{(2019)}}{d_{skt}^{(2018)}} \qquad \forall\, s \in S, k \in K, t \in T \tag{5}$$

$$\widehat{D_{kt}} = D_{kt}^{(2019)} \times \frac{D_{kt}^{(2019)}}{D_{kt}^{(2018)}} \qquad \forall\, k \in K, t \in T \tag{6}$$

For price, volume and relevance score, since we do not expect these metrics to be as volatile as demand and a trend would be comparatively less likely to exist, we take a different approach to obtaining 2020 values. Rather than forecasting per se, the value of each parameter in 2020 is simply the parameter's value from 2019 if one exists, otherwise the 2018 value is used. It is impossible for both 2018 and 2019 values to be missing since the chunk would otherwise not be present, and this method ensures that the most recent values are used for the parameters. Note that relevance scores were defined by chunk in the first place, thus aggregation has no effect on them and the notation r_{kt} is kept.

Aggregation Decision

Table 2 contains the MSE results of forecasting the demand, price and volume of the first week in 2020 - using raw and aggregate data - compared to the real values. On all three parameters, the raw data forecast performs much better than the aggregate data forecast. Thus, we continue our analysis using the raw data, and we would ideally use the forecasted data separated by size in all future computations.

Table 2. *MSE results for raw vs aggregate data*

	Raw	Aggregate
Demand	785.00×10^3	590.69×10^6
Price	165.00×10^6	122.14×10^9
Volume	11.23	8.00×10^3

However, the model from the next subsection (henceforth dubbed Model 3.1) has proven too computationally expensive for our machines if chunks are separated by size. Since the raw-data forecasts are still more accurate, we adopt the following hybrid approach:

1. A model over all 52 weeks of 2020 (Model 3.1) is solved, taking inputs that are aggregated over size. We solve this model once with an objective value maximizing revenue, and then again with an objective value maximizing relevance score. The constraint set, which we discuss in-depth in the next subsection, is maintained between the two runs. We retain which warehouse each chunk is allocated to through the y_{kw} variables, and what weeks we order on in the last two months through the g_{kt} variables (which are explained below).

2. A weekly model (Model 3.2) is solved to determine the ordering schedule, this time using simulated demands to mimic a realistic setting and assess performance. The inputs of this model are separated by size in order to use the improved prediction accuracy and better mirror what 2020 parameters could actually be. Note that the simulations - and the weekly model that constitutes their core - are run separately for the revenue-maximizing and relevance-maximizing versions of the previously discussed 52-week model.

Additional Decision Variables

In order to formulate this problem, we introduce some new decision variables and redefine x_{kt}:

- $z_{kt} \in \mathbb{N}$, $\forall k \in K, t \in \{44, 45...52\}$: quantity *sold* of chunk k in week t belonging to the peak period.

- $g_{kt} \in \mathbb{B}$, $\forall k \in K, t \in \{44, 45...52\}$: equal to one if an order is placed for chunk k in week t, zero otherwise.

- $I_{kt}^{EOW} \in \mathbb{N}$, $\forall k \in K, t \in \{44, 45...52\}$: inventory of chunk k at the beginning of peak week t, computed after the week's orders and sales are processed.

4.3.2 Problem Formulation: Generating Warehouse Allocation and Order Timing (Model 3.1)

Using the forecasted demands, we try to predict the activity for the entire year at the beginning of the year with the end goal of fixing a warehouse allocation and peak period ordering weeks for the next phase. The distinguishing feature of this model is that in November and December (weeks 44 through 52), orders can no longer be placed every week but instead need at least one week of buffer time in between. This means that in those nine weeks, the stock may be purchased and held rather than immediately sold. In order to address this, the decision variables as explained before are introduced. Furthermore, let $T_p \subset T,\ T_p = \{44, 45...52\}$, be the set of peak periods discussed earlier, for ease of notation. It is important to note that, since a feasible solution cannot be obtained while the service constraints are enforced, we scrap them and instead discuss the service levels attained.

We explain our proposed model and then present it. The objective function (7a) computes total revenue or relevance score by considering what is ordered and thus sold in the first 43 weeks and what is sold in the peak period since only in the last weeks there is a difference between sales and orders. These are two separate models.

Constraints (7b) through (7d) limit what bol.com can order in a week. For the first 43 weeks, the limit is the demand of the respective week as in previous models. For weeks 44 through 51, we allow for bigger orders to compensate for the fact that there can be no new order in the week $t+1$ if an order was placed in week t. Since two weeks after order, it is possible to order again, there is no reason to order stock beyond what is demanded in the immediate next two weeks. Furthermore, since an order only happens after the inventory is exhausted, there is no need to constraint the inventory itself with demand levels. In week 52 the maximum order is once again the week's demand since there can be a new order immediately after (in the first week of 2021).

Note that the order can be greater than zero only if the g_{kt} variable allows for it; if no order is to be placed in a certain week, the order size is forced to be zero. For the first 43 weeks this g_{kt} is not added to the constraints, as it is optimal to order every week and all these g_{kt}'s would thus be equal to one.

Constraint (7e) enforces the order buffer, making it so that in two consecutive weeks of the peak period at most one order can be placed. This, of course, allows for multiple weeks to go by without ordering anything, if, for instance, it is not the season for the specific product.

Constraint (7f) manages the relation between sales and demand in the peak weeks, as sales can never exceed the demand of a certain week. Constraint (7g) keep the end-of-week inventory updated by adding the current week's order and subtracting sales from last week's ending inventory. In week 43, right before the peak period, the ending inventory is of course zero since orders and sales still coincide. Constraint (7h) ensures every chunk gets placed in the same warehouse throughout the year, as in the previous models.

Constraints (7i) though (7l) manage warehouse capacity. While (7i) and (7l) serve the same function as in the previous models and have already been explained, (7j) and (7k) reflect the new distinction between orders and sales. In the first 43 weeks, the occupied volume of a chunk is given entirely by the quantity ordered, as (7i) illustrates. In the peak weeks, however, the highest volume occupied by a given chunk in a period is given by the inventory level before subtracting sales, and it is this value that we consider most important when designing capacity constraints - hence the structure of constraint (7k).

$$
\begin{aligned}
\max \quad & \sum_{k \in K_c} \Big(\sum_{t \in T \setminus T_p} P_{kt} x_{kt} + \sum_{t \in T_p} P_{kt} z_{kt} \Big) \quad \textit{or} \quad \sum_{k \in K_c} \Big(\sum_{t \in T \setminus T_p} r_{kt} x_{kt} + \sum_{t \in T_p} r_{kt} z_{kt} \Big) && && \text{(7a)} \\
\text{s.t.} \quad & x_{kt} \le \widehat{D_{kt}}, && \forall k \in K, t \in T \setminus T_p, && \text{(7b)} \\
& x_{kt} \le (\widehat{D_{kt}} + \widehat{D_{k(t+1)}}) g_{kt}, && \forall k \in K, t \in T_p \setminus \{52\}, && \text{(7c)} \\
& x_{kt} \le \widehat{D_{kt}}\, g_{kt}, && \forall k \in K, t = 52, && \text{(7d)} \\
& g_{kt} + g_{k(t+1)} \le 1, && \forall k \in K, t \in T_p \setminus \{52\}, && \text{(7e)} \\
& z_{kt} \le \widehat{D_{kt}}, && \forall k \in K, t \in T_p, && \text{(7f)} \\
& I_{kt}^{EOW} = I_{k(t-1)}^{EOW} + x_{kt} - z_{kt}, && \forall k \in K, t \in T_p, && \text{(7g)} \\
& \sum_{w \in W} y_{kw} = 1, && \forall k \in K, && \text{(7h)} \\
& q_{kwt} \le y_{kw} Q_w, && \forall k \in K, w \in W, t \in T, && \text{(7i)} \\
& \sum_{w \in W} q_{kwt} = x_{kt} V_{kt}, && \forall k \in K, t \in T \setminus T_p, && \text{(7j)} \\
& \sum_{w \in W} q_{kwt} = (I_{kt}^{EOW} + z_{kt}) V_{kt}, && \forall k \in K, t \in T_p, && \text{(7k)} \\
& \sum_{k \in K} q_{kwt} \le Q_w, && \forall w \in W, t \in T, && \text{(7l)} \\
& x_{kt} \in \mathbb{N}, && \forall k \in K, t \in T, && \text{(7m)} \\
& z_{kt} \in \mathbb{N}, && \forall k \in K, t \in T_p, && \text{(7n)} \\
& g_{kt} \in \mathbb{B}, && \forall k \in K, t \in T_p, && \text{(7o)} \\
& y_{kw} \in \mathbb{B}, && \forall k \in K, t \in T, && \text{(7p)} \\
& q_{kwt} \in \mathbb{R}, && \forall k \in K, w \in W, t \in T, && \text{(7q)} \\
& I_{kt}^{EOW} \in \mathbb{N}, && \forall k \in K, t \in T_p && \text{(7r)}
\end{aligned}
$$

4.3.3 Problem Formulation: Weekly, Dynamic Model for the Ordering Schedule (Model 3.2)

Below we explain our weekly, dynamic model for the ordering schedule. We start by explaining how we randomize the simulated demand and then explain the weekly model.

Randomizing Simulated Demand

To test the performance of our allocation, we run 100 simulations with pseudo-randomized demands in which, while each chunk has its warehouse and peak ordering weeks fixed, we determine how much we order and sell from that chunk based on a dynamic, weekly model. Before the model itself, we show how noise is added to the forecasted demands in each simulation run to mimic real-world uncertainty. Recall that, in order to better reflect the 2020 parameter behavior we observe in the first week, we now return to using size-separated entries for all parameters: demand, price, volume, and relevance score. Their forecasted values - previously computed for the MSE analysis - are now used in the model and denoted $\widehat{d_{skt}}$, p_{skt}, v_{skt} and r_{skt} respectively as before Model 3.1.

Equation 8 shows how the average demand between 2018 and 2019 is computed, and Equation 9 shows how standard deviation is derived from this sample of two real observations on past demand (contrasting 2020 forecasted one, which essentially brings no new information beyond what 2018 and 2019 provide and is thus redundant in these formulas).

$$\mu_{skt} = \frac{d_{skt}^{(2018)} + d_{skt}^{(2019)}}{2} \qquad \forall\, s \in S, k \in K, t \in T \setminus \{1\} \tag{8}$$

$$\sigma_{skt} = \sqrt{(d_{skt}^{(2018)} - \mu_{skt})^2 + (d_{skt}^{(2019)} - \mu_{skt})^2} \qquad \forall\, s \in S, k \in K, t \in T \setminus \{1\} \tag{9}$$

Below in Equation 10 we compute the demand in a simulation run of 2020. The first term in this equation is the forecasted demand as computed in Equation 5 and the second term is to generate noise when simulating the data. γ is a parameter that has a 50% chance of being zero and a 50% chance of being one and thus determines if the standard deviation is subtracted or added from the forecast. β is zero with a 35% chance, one with a 30% chance, two with a 30% chance and finally three with a 5% chance and therefore determines by how much the simulation deviates from the original forecast. These percentages are chosen such that simulations are distributed in a fashion similar to a normal distribution.

$$d_{skt}^{(sim)} = d_{skt}^{(2019)} \times \frac{d_{skt}^{(2019)}}{d_{skt}^{(2018)}} + (-1)^{\gamma} * \beta * \sigma_{skt} \qquad \forall\, s \in S, k \in K, t \in T \tag{10}$$

Weekly Model

Having solved the whole-year Model 3.1, we attribute the values of the variables y_{kw} to the eponymous parameter used in this model for every chunk k and warehouse w. We also create a parameter g_{skt} for each size s of chunk k in period t and assign to it the value of the variable g_{kt} of the previous model across all sizes.

The previously-obtained results x_{kt} and z_{kt} are just beginning-of-year expectations of these values based on forecasted demand. However, the actual sales z_{skt} depend on the actual demand, which we simulate. Consequently, depending on the beginning of week (BOW) inventory levels, which in turn depend on previous sales, we must adjust the amount ordered x_{skt}. The order in which information is revealed is thus assumed to be: $z_{sk(t-1)} \rightarrow x_{skt} \rightarrow d_{skt} \rightarrow z_{skt}$, meaning, first we know what is sold last week, then we decide what is ordered this week, then we know the demand of this week, then we can decide what to sell this week.

Thus, we must solve a weekly optimisation problem to determine x_{skt} and consequently I_{skt}^{BOW}, which is an additional decision variable representing the beginning of the week inventory level, i.e. after receiving ordered stock x_{skt} but before any sales z_{skt}. Next to that, we introduce the parameter P_t, which is equal to one if t is a peak week and zero otherwise.

$$
\begin{aligned}
\max \quad & \sum_{s\in S}\sum_{k\in K}\sum_{t\in T} p_{skt} I_{skt}^{BOW} \quad or \quad \sum_{s\in S}\sum_{k\in K}\sum_{t\in T} r_{skt} I_{skt}^{BOW} && && \text{(11a)}\\
\text{s.t.} \quad & x_{skt} \leq max\{0, g_{skt}(\widehat{d_{skt}} + P_t \widehat{d_{sk(t+1)}} - I_{sk(t-1)}^{BOW} + z_{sk(t-1)})\}, && \forall k \in K, s \in S, && \text{(11b)}\\
& I_{skt}^{BOW} = I_{sk(t-1)}^{BOW} - z_{sk(t-1)} + x_{skt} && \forall k \in K, s \in S, && \text{(11c)}\\
& q_{sktA} = q_{sk(t-1)A} - y_{kA}(z_{sk(t-1)} v_{sk(t-1)} + x_{skt} v_{skt}), && \forall s \in S, k \in K, && \text{(11d)}\\
& q_{sktB} = q_{sk(t-1)B} - y_{kB}(z_{sk(t-1)} v_{sk(t-1)} + x_{skt} v_{skt}), && \forall s \in S, k \in K, && \text{(11e)}\\
& Q_w \geq \sum_{s\in S}\sum_{k\in K} q_{skwt}, && \forall w \in W, && \text{(11f)}\\
& x_{skt} \in \mathbb{N}, && \forall s \in S, k \in K, && \text{(11g)}\\
& I_{skt}^{BOW} \in \mathbb{N}, && \forall s \in S, k \in K && \text{(11h)}
\end{aligned}
$$

Based on these results we compute z_{skt} as:

$$z_{skt} = min(I_{skt}^{BOW}, d_{skt}^{(sim)}) \tag{12}$$

This problem aims to maximise the expected sales or relevance, which are represented by the beginning-of-week inventory (11a). Constraint (11b) determines the upper bound for x_{skt} given by the forecasted demands and the existing inventory. We purchase stock to supplement the existing inventory $I_{sk(t-1)}^{BOW} - z_{sk(t-1)}$ up to the expected demand $\widehat{d_{skt}}$. In peak weeks ($P_t = 1$), this also includes the expected demand of the following week $\widehat{d_{sk(t+1)}}$, but only if the current week is an ordering week ($g_{skt} = 1$). Whether or not the week is an ordering week is decided for each chunk at the beginning of the year. Constraint (11c) defines I_{skt}^{BOW} as the inventory at the beginning of the previous week minus what is sold in that previous week plus the new stock x_{skt}.

Constraints (11d) through (11f) ensure that even after the new stock comes in, the capacity constraint is still satisfied. The chunk - warehouse allocation y_{kw} is determined at the beginning of the year. Note that the last set of constraints uses the actual elements of the set $W = \{A, B\}$ and an indicator $w \in W$ interchangeably, simply to more clearly display how constraints (11d) and (11e) update the total warehouse volume taken by the size of a certain chunk in the focused week. As a last clarification, we point out that all variables and parameters are set to zero in $t = 0$: inventory, sales, and volume.

Equation 12 computes what bol.com sells in a given week after the inventory is determined through a run of Model 3.2. The minimum between simulated demand (the "actual" demand bol.com would face that week) and current inventory is chosen in order to never sell more than what is demanded.

5 Results

Below we present results. In Section 5.1 we present results of Model 1, in Section 5.2 we present results of Model 2, and, lastly, in Section 5.3, we present results of Model 3.2.

5.1 Model 1: 2018 revenue-maximization

In Table 3, we present the revenues and per-category service levels that the optimal solution for 2018 achieves, as well as the performance of this solution if it is to be maintained in 2019. Note that only the chunks of 2019 that also existed in 2018 are included in 2019 computations; the chunks that exist only in 2018 or only in 2019 are, therefore, (intuitively) ignored. Furthermore, we address the issue that 2018 sales sometimes exceed 2019 by considering the minimum of the two when computing revenues.

Table 3. *Revenue and service level for 2018 and 2019 under 2018 schedule*

Result / Year	**Revenue (euro)**	**Service level ROT (%)**	**Service level GT (%)**
2018	113,741,381.25	98.99	97.58
2019	100,099,132.39	68.78	71.57

We obtain revenue of 113 million euros for 2018 and 100 million euros for 2019. We can see that the revenue for 2018 is 13.74 million euros higher than that of 2019 - hence revenue would drop by about 10% if the ordering schedule was maintained. What is more, the service level is higher in 2018 for both ROT (98.99% to 68.78%) and GT (95.68% to 71.57%). These gaps are significantly larger than what we see for revenue and will be elaborated upon in the discussion section.

In the Appendix, Figures 1 and 3 respectively show the occupation of the warehouses and the service levels across 2018. From Figure 1 one can deduce two things. Firstly, the warehouse occupation levels for both warehouses A and B are almost never at 100%. The only time the warehouses are fully occupied is around the holiday season. Secondly, in warehouse A, one can see that there is another peak at the beginning of the summer. Then over 70% of the full capacity is occupied. From Figure 3 we can see that there is a sharp decrease in the service levels of GT and ROT during the holiday season that accompanies the fully occupied warehouses.

5.2 Model 2: trade-off between objectives

Below in Tables 4 and 5 the objective values, total revenues, total relevances, and service levels for GT and ROT for increasing α are presented.

While one would expect that, for $\alpha = 0$ and $\alpha = 1$, the objective value would be one, we are not concerned with model validity as the small discrepancy is the result of limited numerical precision common to both CPLEX and Java floating-point variables. The discussion section will thus focus on a graphically-fundamental discussion of the changes observed in total revenue and total relevance.

It is apparent from the fixed 98% value that only the GT service level constraint is binding. We do, however, note that ROT service is consistently lower in 2019, reinforcing the fact that there was comparatively more demand in this second year. We present objective value times 10^{-3} to be able to see the differences between the objective values as the first different number is the third or forth digit.

Table 4. *Output of Model 2 for 2018*

α	**Objective value x 10^{-3}**	**Total revenue**	**Total relevance**	**Service level GT %**	**Service level ROT %**
0.00	999.28	113,420,155.43	3,251,980.32	98.00	98.85
0.10	999.81	113,509,230.17	3,254,192.35	98.00	98.93
0.20	999.65	113,533,775.54	3,253,967.61	98.00	98.91
0.30	999.19	113,510,928.91	3,252,701.51	98.00	98.84
0.40	996.88	113,262,016.51	3,245,440.61	98.00	98.43
0.50	998.20	113,445,127.26	3,249,431.78	98.00	98.70
0.60	999.15	113,611,593.68	3,250,496.25	98.00	98.70
0.70	998.34	113,534,699.21	3,246,209.89	98.00	98.48
0.80	999.50	113,653,615.24	3,249,571.60	98.00	98.63
0.90	999.49	113,644,796.92	3,247,652.64	98.00	98.52
1.00	999.88	113,669,915.46	3,245,134.92	98.00	98.39

Table 5. *Output of Model 2 for 2019*

α	**Objective value x 10^{-3}**	**Total revenue**	**Total relevance**	**Service level GT %**	**Service level ROT %**
0.00	999.56	136,561,891.39	3,928,549.03	98.00	95.38
0.10	996.27	136,134,689.87	3,917,008.63	98.00	95.10
0.20	998.31	136,720,006.70	3,924,574.14	98.00	95.35
0.30	996.71	136,527,364.03	3,918,659.79	98.00	95.01
0.40	998.81	136,935,472.81	3,925,253.45	98.00	95.41
0.50	995.07	136,301,484.50	3,913,877.23	98.00	95.03
0.60	996.98	136,659,252.55	3,918,800.37	98.00	95.21
0.70	997.40	136,768,602.04	3,917,189.86	98.00	95.19
0.80	998.14	136,863,646.57	3,918,658.85	98.00	95.19
0.90	999.29	137,008,761.36	3,921,124.67	98.00	95.34
1.00	999.47	137,009,386.68	3,919,340.47	98.00	95.26

The values from the tables are also plotted in Figures 4 to 8, which can be found in the Appendix. We draw confidence in our results from the fact that Figures 7 and 8 show an overall increasing tendency of revenue as α grows; since higher values for α imply higher priority given to revenue, these results corroborate our findings so far. Similarly, these figures display passed sanity tests, as a decreasing tendency of relevance with α is fully understandable.

Furthermore, Figures 2 and 4 show the 2019 warehouse occupation and service levels respectively when α is set to one, i.e. revenue-maximization. Similar to the results we saw for Model 1 for 2018, occupation levels for both warehouses are only at 100% around the holiday season, accompanied by a sharp decrease in the service levels. What is to note, is the increased occupation of warehouse A around week 25. This hits 90% occupation and is accompanied by a slight but sharp drop of the GT service level.

5.3 Model 3: 2020

Tables 6 and 7 show the summary statistics for the simulated year 2020 using either the revenue model or the relevance model to generate the initial warehouse allocation and ordering policy as well as the dynamic inventory restocking. For both models we see very little variance in service levels although the relevance model achieves better ROT service levels. Furthermore, revenues achieved by the revenue model are on average 2.5% higher than those achieved using the relevance model. On the other hand, relevance levels achieved by the relevance model are on average 3.5% higher than those achieved by the revenue model.

That being said, the differences in these averages are much larger than the within-model variation. The difference in average revenue achieved by the two models is equal to 6.8 revenue model standard deviations and 7.8 relevance model standard deviations. Comparatively, the difference between the minimum and maximum values achieved is five and 4.5 standard deviations using the revenue and relevance models respectively. The difference in average relevance achieved by the models is equal to 15.1 revenue model standard deviations and 13.8 relevance model standard deviations. Comparatively, the difference between the minimum and maximum values achieved is 4.7 standard deviations for both models.

Table 6. *Summary statistics for 2020 revenue model simulations*

	Revenue	Relevance	Service GT	Service ROT
MIN	202,568,192.48	6,659,209.00	0.76	0.67
MED	204,586,620.26	6,702,183.00	0.77	0.68
AVG	204,486,791.02	6,699,820.00	0.77	0.68
MAX	206,260,522.59	6,731,606.00	0.78	0.68
STDEV	736,300.21	15416.56	0.38×10^{-2}	0.27×10^{-2}

Table 7. *Summary statistics for 2020 relevance model simulations.*

	Revenue	Relevance	Service GT	Service ROT
MIN	197,956,512.43	6,890,659.00	0.76	0.71
MED	199,499,210.85	6,933,882.00	0.77	0.72
AVG	199,506,377.01	6,933,158.00	0.77	0.72
MAX	200,851,619.87	6,969,776.00	0.78	0.73
STDEV	641,724.16	16,797.40	0.36×10^{-2}	0.30×10^{-2}

In Figures 22 - 29 in the Appendix, we see histograms of the revenue, relevance, GT service levels and ROT service levels that represent the frequency of the specified interval values. In pink, we can see the revenue models and in yellow we see the relevance models.

Figures 22 and 23 in the Appendix show the GT service levels for both models. The interval sizes on the x-axis are 0.2% for both of these figures. For the revenue-maximization model (Figure 22) the highest value is greater than 77.6%, the lowest is less than 76.0%, and the most common values are in the interval from 76.8 to 77.0%. The average value is 76.8%. For the relevance-maximization model (Figure 23), the highest value is greater lowest is lower than 76.0% and the most common values are also in the interval from 76.8 to 77.0%. The average value is 76.9%.

Figures 24 and 25 in the Appendix show the ROT service levels for both models. Again, the interval sizes are 0.2% for both figures. For the revenue-maximization model (Figure 24) the highest value is 68.2%, the lowest is less than 66.8%, and the most common values are in the interval from 67.6 to 67.8%. The average value is 67.5%. For the relevance-maximization model (Figure 25), the highest value is greater than 72.6%, the lowest is lower than 71.2% and the most common values are also in the interval from 71.8 to 72.0%. The average value is 71.8%.

Figures 26 and 27 in the Appendix show the relevance values for both models. In these figures, the intervals on the x-axis are of size 0.01 million. For the revenue-maximization model (Figure 26) the highest value is 6.74 million, the lowest is 6.66 million and the most common values are in the interval from 6.70 to 6.71 million. The average value is 6.70 million. For the relevance-maximization model (Figure 27), the highest value is 6.97 million, the lowest is lower than 6.9 million and the most common values are in the interval from 6.93 to 6.94 million. The average value is 6.93 million.

Figures 28 and 29 in the Appendix show the revenue values for both models. What can be seen in these figures, is that on the x-axis we have interval sizes of 0.5 million. For the revenue-maximization model (Figure 28) the highest value is 206.6 million, the lowest is 202.6 million and the most common values are in the interval from 204.6 to 205.1 million. The average value is 204.5 million. For the relevance-maximization model (Figure 29), the highest value is 201 million, the lowest is lower than 198 million and the most common values are in the double interval from 199 to 200 million. The average value is 199.5 million.

6 Extensions

Below we present our extensions. In Section 6.1, we look into how the change of the assigned capacity in the warehouses would affect the revenue and service levels. In Section 6.2, we take a closer look at the revenue, relevance and service levels during peak weeks. Section 6.3 discusses the performance measures assuming that we do not move chunks between years.

6.1 Reduced Capacities

After seeing that service levels are very high when using 15% of the warehouse capacities, we look at how the revenue and service levels would change if we changed the percentages. We look into the following combinations of the capacity constraints (the first number is for warehouse A and the second number is for the warehouse B): 15% and 15% (original problem); 15% and 10%; 15% and 5%; 10% and 15%; 10% and 10%; 10% and 5%; 5% and 15%; 5% and 10%; 5% and 5%; and 1% and 1%. What we obtain is that for 2018 only the first three possibilities give a feasible solution when using the service level constraints and for 2019, only the original capacity constraints give a feasible solution. Figures 16 to 21 in the Appendix give a clear overview of how revenues and service levels change.

While the decrease from 5% to 1% in both warehouses is evident in both revenues and service levels, the same cannot be said when comparing, for instance, 10% in warehouse A and 5% in warehouse B to the default 15% across the board. Corroborating the rest of our investigation, we note that service levels remain very stable in spite of successively shrinking total space, especially for the ROT product category. The revenues, on the other hand, are more apparently affected by reducing storage space, although more space can be saved than revenue is relatively lost: note, for instance, that allocating 5% of both warehouses is a 66% space saving that only causes a roughly 25% loss in revenue. Should this space have the opportunity to be filled with more profitable items, the possibility should be seriously considered.

A final observation on this extension revolves around the effect of the service constraints. As we already stated, 10% of the capacity of warehouse B can be saved while upholding the minimum service quotas and barely changing revenue. This indicates a recurrently-observed bias for warehouse A when placing products - likely a consequence of forcing items to remain in the same warehouse all year long. Lastly, it is clear that service level constraints are most impactful in 2019, when GT service level is raised by about 10%. The naturally lower service level in 2019 is of course connected back to the larger total demand - especially for General Toys.

6.2 Binding warehouse capacities in peak weeks

What can be seen in the Figures 9 and 10, is the trade-off between revenue and relevance for increasing α's. Figure 9 represents 2018, Figure 10 represents 2019. The y-axis represents the following: fraction of the revenue (relevance) which is calculated by dividing the revenue (relevance) corresponding to the specific value of α with the maximum revenue (relevance) which corresponds to $\alpha = 1$ ($\alpha = 0$). We do this in order to be able to compare the measures. As expected, revenue is increasing in α and relevance is decreasing in α.

For 2018, we can see that the intersection point is at $\alpha = 0.5$ which is what we would expect, assuming both measures to be equally important. We also see that revenue and relevance are relatively symmetric around $\alpha = 0.55$, meaning, the measures seem to be equally important. Decrease in one seem to have a proportional increase in the other one for the α values between 0.3 and 0.7. For 2019, the picture is not so clear. The intersection point is at $\alpha = 0.35$, making us believe that the relevance is more important as it is closer to $\alpha = 0.0$ than $\alpha = 1.0$.

What is more, the revenue line has a steep increase until $\alpha = 0.4$ and then it increases gradually. The relevance line is decreasing at a slower pace, also, the difference between the maximum and the minimum relevance is much smaller than that of the revenue. α values between 0.4 and 0.7 result in a proportional increase in revenue and decrease in relevance. What is more, from both Figures, we can see that even when paying full attention to only one measure, meaning, when $\alpha = 0.0$ or when $\alpha = 1.0$, the other measure is still taken care of as the fraction of the maximum is very high (above 95%).

From Figures 11 and 11 in the Appendix, we can see that the service levels during the peak season are not met no matter the α value. Service levels are higher in 2018 than in 2019. The worse service levels are for the GT in 2019, the best service levels are for the ROT in 2018. Lastly, it can be seen that the service levels are the highest when $\alpha = 0$.

6.3 Maintaining the chunk allocations across the years

For this part of the research, we look into the performance measures assuming that we do not move chunks between years. We start by solving Model 2 for 2018 with service constraints and used $\alpha = 1$ for revenue-maximization. This way we obtained the y_{kw} variables. We proceeded by solving the same model for 2019, however, we used the already obtained y_{kw} variables, meaning, this was one of our input parameters for the chunks that were the same as in 2018. Only new chunks obtained new y_{kw} variables. For 2019 we also used the service constraints. Finally, we solved the same model for 2020 again using the previously y_{kw} variables as given, however, we did not use the service level constraints as then the solution would be infeasible. The performance measures are obtained in Table 8 below.

Table 8. *Maintaining the chunk allocations across the years*

Result / Year	**Revenue (euro)**	**Relevance score**	**Service level GT (%)**	**Service level ROT (%)**
2019	136,921,330.20	3914379.60	98.00	95.07
2020	190,436,307.50	5414219.96	82.82	67.21

From the Table 8 above we see that the revenue and relevance increased from 2019 to 2020, however, the service levels decreased. If we compare these results of 2019 to the ones obtained in Table 5 when $\alpha = 1$, we see that all the measures are slightly smaller, however, the difference does not seem to be substantial. When comparing these results of 2020 with the ones presented in Section 5.3, the service level of GT is higher here, however, the service level of ROT, the revenue, and the relevance are slightly smaller here. Again, these results do not seem to differ much.

7 Discussion and Conclusion

At the beginning of this paper, we asked "What is the maximized revenue with respect to the capacity constraints?" and "How do the commercial objectives of bol.com interact?". Model 1 focused on the first such commercial objective - revenue - and found 113.7 million euro to be the maximum revenue. Since maintaining the same ordering schedule for 2019 seems intuitively sub-optimal given the different sets of demands and prices, Model 2 is instead employed to find the maximum revenue achievable in 2019: 137 million euro. It is through Model 2 that we explore the other two main commercial objectives we found relevant in the case's current form, namely service level and relevance score.

If we look at the Figures 13 to 15 in the Appendix, we see the overall demand, price, and volume of 2018. Figure 13 represents total demand across all category-chunk entries over time. It shows that the warehouses sit mostly empty until the holiday season when demand for toys understandably explodes. However, Figure 14 shows that the average price in euros explodes at the beginning of summer. This can be explained by the fact that usually at the beginning of summer big items such as trampolines are purchases which are very costly. However, most toys are relatively cheap and therefore in the holiday season, when gifts are purchased, one can observe a decrease in the average price. Similarly, as Figure 14, the Figure 15 has a peak average volume in the summer months. The explanation for this can also be the same as for the average prices.

Turning our attention back to Table 3, we first look at the revenues. While it is fully within expectations that 2019 would be less profitable than 2018 because the solution is simply not tailored for the specific demands, we find it noteworthy that 2019 revenue is only about 10% lower. On one hand, this result suggests that most 2018 chunks exist in 2019 as well and that demand for those chunks remains relatively steady (or perhaps even growing). On the other hand, this may also be indicative of more potential for revenue in 2019, likely a result of price increases.

The service levels, however, tell a different story. In 2018, we observe that the vast majority of demand is satisfied without having imposed any constraints to this end. This is indicative of warehouse capacities being non-prohibitive on average, with most weeks allowing for full demand satisfaction. We find evidence of this when plotting warehouse occupation levels over time (Figure 1) and total demand across all category-chunk entries over time (Figure 13). As a result of this observation, an extension will be devoted to analysing prohibitive weeks in particular.

Conversely to 2018, the 2019 service levels see quite a drop - about a third of demand goes unsatisfied in each category. This result works in parallel with the discussion of revenue, likely reflecting that demand levels in 2019 are substantially higher or that orders for various toys are simply shuffled around between various weeks. This difficulty in maintaining high service levels was shown to only occur in this case of misaligned variables and parameters. For all other models, it is clear that upholding service level constraints does not negatively impact revenue or relevance score in any meaningful way, particularly for the ROT product category. While the higher requirement of the GT category was never immediately fulfilled without imposing constraints, the service level was usually close enough to what is required in order for its improvement not to alter the commercial objectives.

Furthermore, we can discuss the relevance - revenue tradeoff. Equation 9 shows the change in revenue and relevance attained in the weeks of 2018 *where the capacity constraints are binding, as percentages of the maximum values possible*, depending on α. For α larger than 0.7, the increase in revenue comes at the cost of a significantly higher percentage decrease in relevance. Conversely, the revenue can be increased by decreasing α to around 0.4 at little cost of relevance as the revenue graph is not steep at these values. The value of α depends on the importance bol.com places on revenues and customer relevance respectively. It is possible that higher relevance leads to higher profits in the long run, compensating the immediate loss in revenues.

The interaction between revenues and relevance should be further investigated in order to draw clearer conclusions on their trade-off. That being said, even when choosing to fully prioritise revenue, the cost in relevance is less than 1% its maximum possible value, and the same holds vice-versa. In 2019, in Figure 10 we see that for $\alpha = 0.4$ we get a one percentage point increase in revenues for basically no decrease in revenue. Revenues can be furthermore increased by another three percentage points by setting $\alpha = 0.5$ while still maintaining relevance above 98% its maximum value.

When turning our attention back to Model 3, we see that the smallest interval value of the revenue-maximization model is already higher than the highest value of the relevance-maximization interval for the revenue and it is the other way for relevance and the ROT service levels. The smallest and largest interval of the GT service levels are equal under both models. We do see that the service levels for both GT and ROT are below the expected service levels under both revenue- and relevance-maximization models. However, taking all into account, the service levels are slightly better under the relevance-optimization model, especially for ROT.

When it comes to the revenue-relevance trade-off, as we have noted in the results report section, the differences between models are much larger than the within-model differences. This effect is most significant for the relevance, where the difference between models is on average 3 times larger than the difference within model between the minimum and maximum values achieved by the simulator. For revenue, the inter-model difference is less than twice the within-model differences. Thus, one possible suggestion would be using the relevance model, as the loss of revenue resulting from doing so is proportionately smaller than the loss of relevance resulting from picking a revenue maximisation model. The revenue difference is further mitigated by the long term effect that relevance probably has on revenue.

That being said, there is no objective or definitive way to compare absolute values of revenue and relevance as we do not know how they influence each other. Furthermore, a more in depth analysis of this trade-off, similar to our alpha analysis in Model 2, would be useful to explore whether a combined objective function model could achieve a balance that is more in line with bol.com's preference for commercial objectives. Finally, it is this preference that will ultimately determine the superior model, and in spite of the proportionately larger loss of relevance resulting from picking a revenue model, due to the high inter-model differences, the better model will be the one the aligns with bol.com's priorities for commercial objectives, with the relevance model being suggested for service level and relevance maximisation, and the revenue model being best suited for revenue maximisation.

This papers suffers from a few limitations, the most prevalent one being a margin of error between $10^{-4}\%$ and $10^{-6}\%$ in our numerical analysis. This is owed to hardware limitations that do not allow the CPLEX algorithm to consider the entirety of the branch-and-cut tree before returning an optimal solution. Consequently, many of our results are potentially potentially estimates of the true maximum values of the various commercial objectives.

Some fault lies with the dataset as well. The plethora of missing or obviously incorrect values accumulated over time in the platform's system damage the validity and applicability of our results in a realistic setting.

What is more, we used revenue instead of profit. In the short run, a company can focus on revenue to grow its market share, but in the long run, depending on the companies' views, profit might be more important. However, improving our models thus would require obtaining more information about costs both fixed and variable, such as transportation costs, unit production costs, etc.

To further improve this research such that it resembles the real life, we could relax some of our assumptions presented in Sections 4.1.1, 4.2.1, and 4.3.1. For instance, delivery times for each supplier could be different, we never looked into multi- or single-item orders, we also assumed perfect efficiency of storage capacity which in life might not be the case. Relaxing these assumptions is a possible extension of this research.

8 Advice to bol.com

When using our Model 2, the only parameter needed to come up with is α. According to the decision of α, the importance of revenue or relevance score is determined. The higher the α, the higher the importance of revenue is assumed. We advice to choose α values between 0.4 and 0.7 as in between these values the increase in one measure and decrease in the other measure seem to be proportional according to our analysis performed on the data of 2018 and 2019.

Current assigned capacities in warehouse A and B are 15% in both for *Toys*. We do agree that the 15% capacity should remain in the warehouse A. However, there is no real need to have 15% of the warehouse B assigned to *Toys* as the revenue, relevance, and service levels of both GT and ROT barely decrease with allowing smaller capacity to be occupied by *Toys*.

If the objective for 2020 is to maximize the service levels, then using the relevance-maximization model is the better option as, according to our findings, this gives higher service levels for *ROT* and equal service levels for *RT* compared to the revenue-maximization model. However, the difference in the service levels is not big, therefore, if revenue is the main objective, using the revenue-maximization model would be the better option. Similarly, if relevance is the main objective, using the relevance-maximization model would be the better option. If revenue and relevance are equally important, it is better to use the relevance-maximization model as this gives a smaller proportional loss of revenue than bol.com would have in case of revenue-maximization model with regards to the relevance.

With regards to chuck allocation across years, we recommend to not move chucks at the end of a year. We do find some decreases for the revenue, relevance and service levels in the case that all chunks stay in their original warehouse, but these are all not substantial. When one would relocate the chunks, one would have to take into account transportation costs and other costs associated to redesigning the warehouse lay-out and these costs are likely to outweigh the slight decrease in revenue when chunks are not relocated.

9 References

Absi, N., & Kedad-Sidhoum, S. (2007). MIP-based heuristics for multi-item capacitated lot-sizing problem with setup times and shortage costs. *RAIRO - Operations Research, 41*(2), 171–192. https://doi.org/10.1051/ro:2007014

Absi, N., & Kedad-Sidhoum, S. (2008). The multi-item capacitated lot-sizing problem with setup times and shortage costs. *European Journal of Operational Research, 185*(3), 1351–1374. https://doi.org/10.1016/j.ejor.2006.01.053

Absi, N., & Kedad-Sidhoum, S. (2009). The multi-item capacitated lot-sizing problem with safety stocks and demand shortage costs. *Computers & Operations Research, 36*(11), 2926–2936. https://doi.org/10.1016/j.cor.2009.01.007

Harris, F. W. (1913). How Many Parts to Make at Once. *Operations Research, 38*(6), 947–950. https://doi.org/10.1287/opre.38.6.947

Hindi, K. S. (1995). Computationally efficient solution of the multi-item, capacitated lot-sizing problem. *Computers & Industrial Engineering, 28*(4), 709–719.

yons, D. (2009, December 24). *"We Start With the Customer and We Work Backward."* Slate Magazine. https://slate.com/news-and-politics/2009/12/jeff-bezos-on-amazon-s-success.html

Miller, A. J., Nemhauser, G. L., & Savelsbergh, M. W. P. (2000). Solving Multi-Item Capacitated Lot-Sizing Problems with Setup Times by Branch-and-Cut. *CORE DISCUSSION PAPER, 2000*(39).

NL Times. (2021, February 17). *Bol.com sees 70 percent sales increase in the past quarter.* NL Times. https://nltimes.nl/2021/02/17/bolcom-sees-70-percent-sales-increase-past-quarter

Parveen, S., & Haque, A. A. (2007). A heuristic solution of multi-item single level capacitated dynamic lot-sizing problem. *Journal of Mechanical Engineering, 38*, 1–7. https://doi.org/10.3329/jme.v38i0.893

Ullah, H., & S, P. (2010). A Literature Review on Inventory Lot Sizing Problems. *Global Journal of Research In Engineering, 10*(5).

10 Appendix

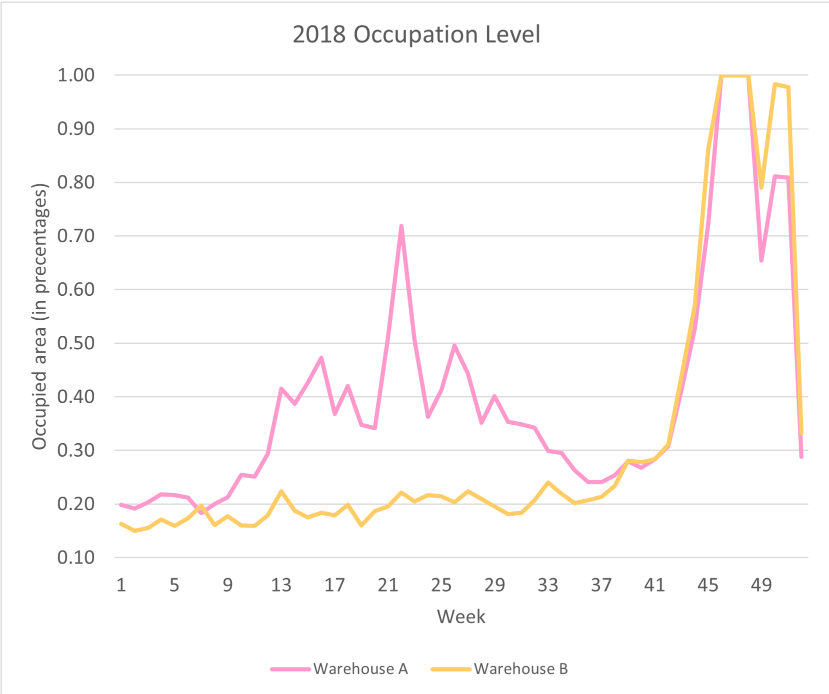

Figure 1: Warehouse occupation across 2018.

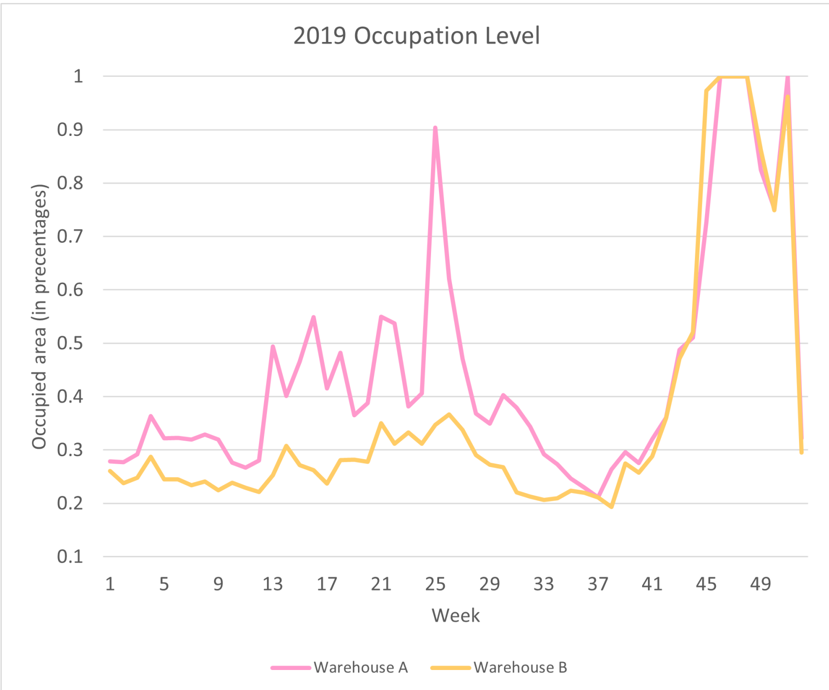

Figure 2: Warehouse occupation across 2019.

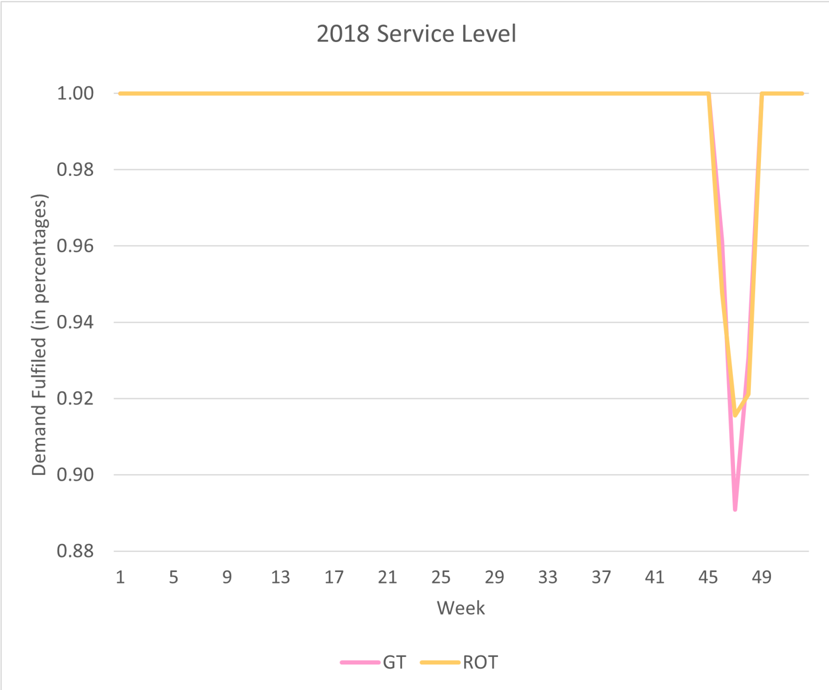

Figure 3: Service levels across 2018.

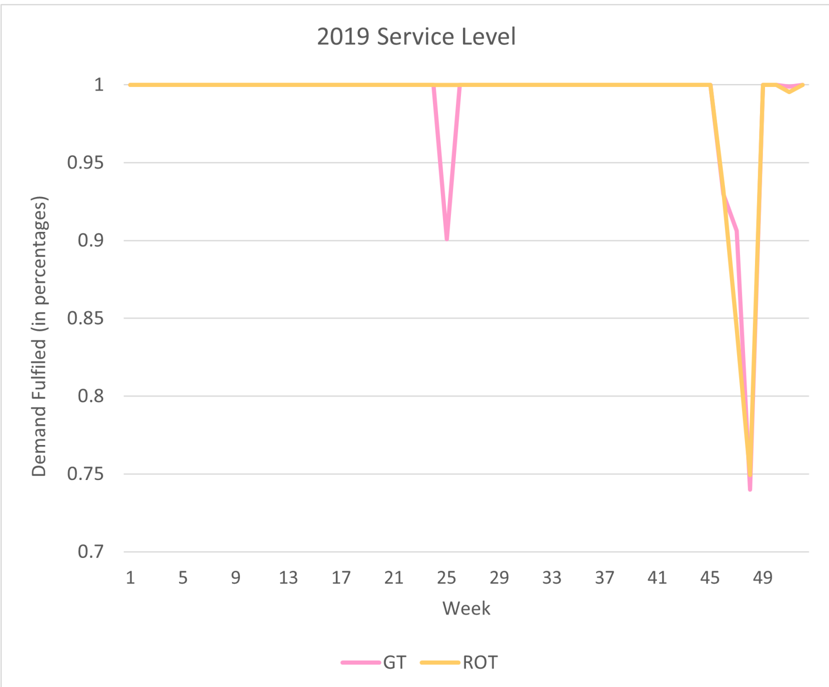

Figure 4: Service levels across 2019.

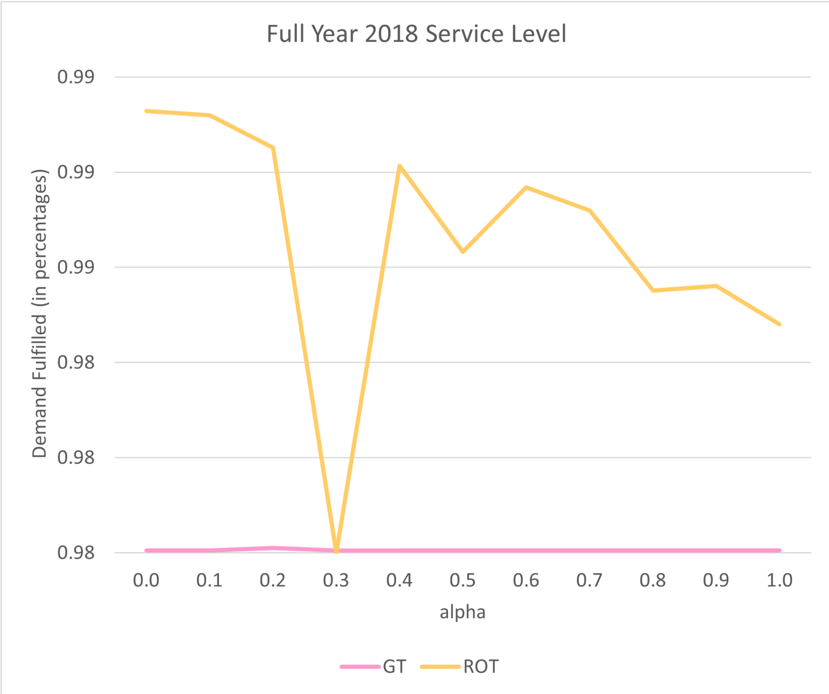

Figure 5: Total service levels in 2018 against α.

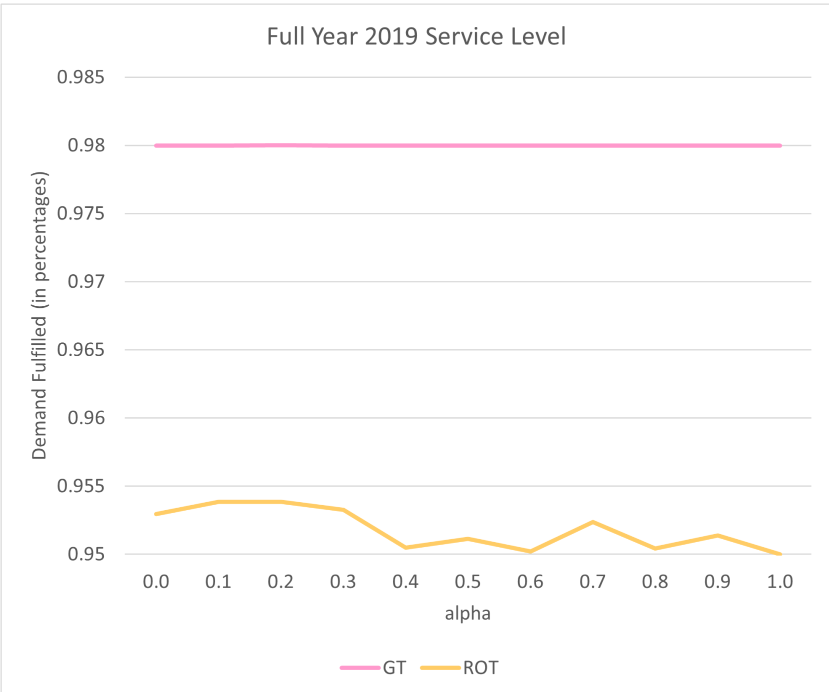

Figure 6: Service levels against α in 2019

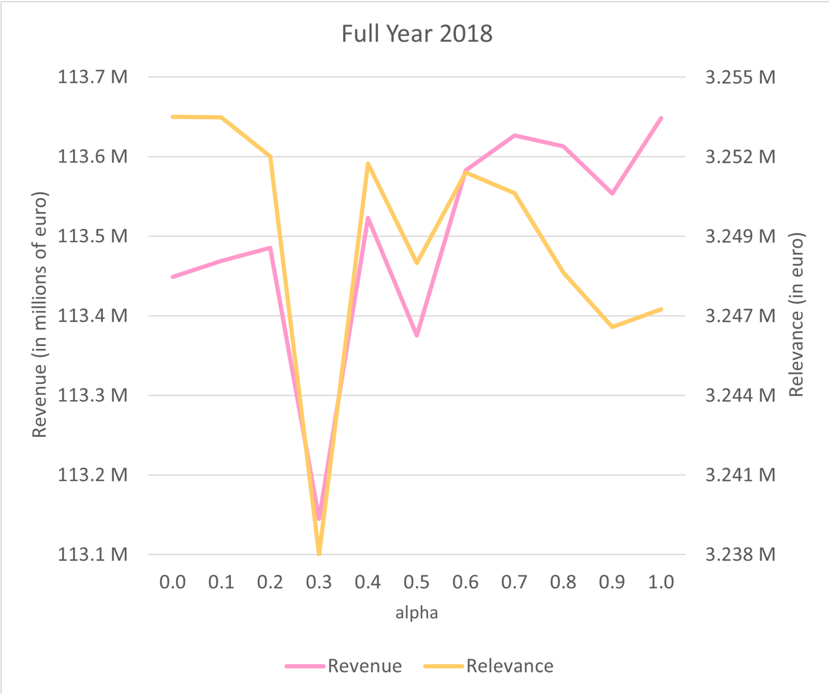

Figure 7: Total revenue and total relevance in 2018 against α.

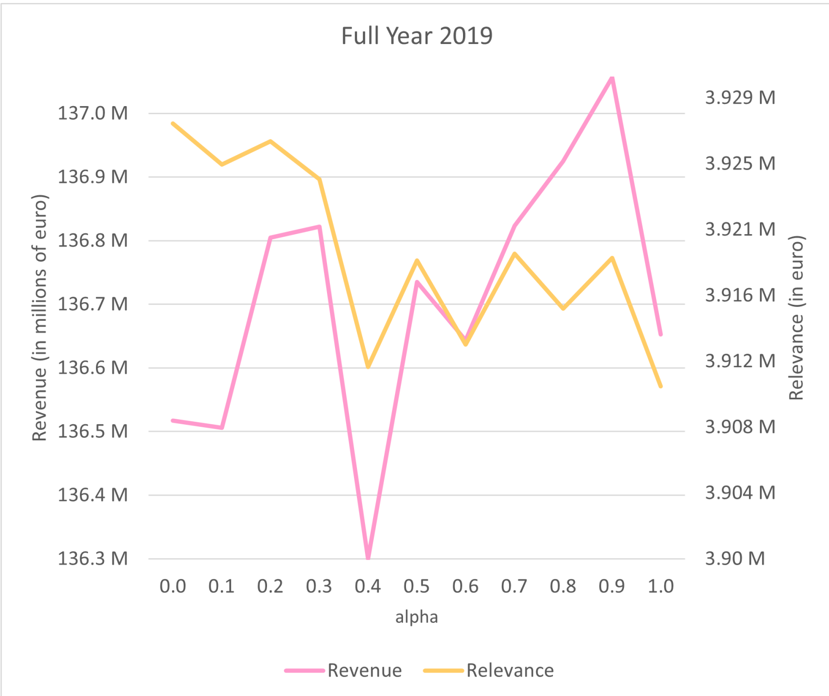

Figure 8: Total revenue and relevance in 2019 against α.

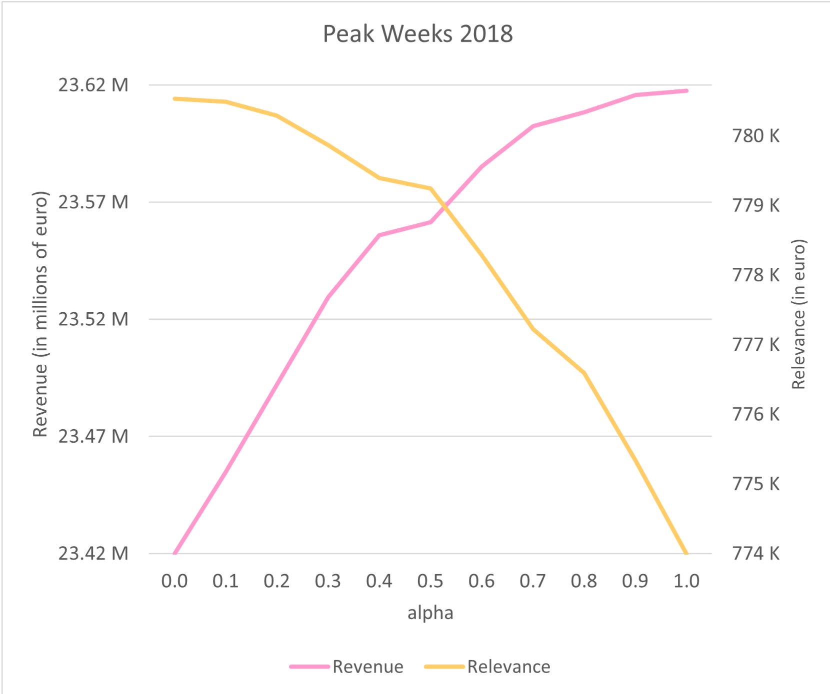

Figure 9: Trade-off between revenue and relevance for different values of the weight parameter α in the peak weeks of 2018.

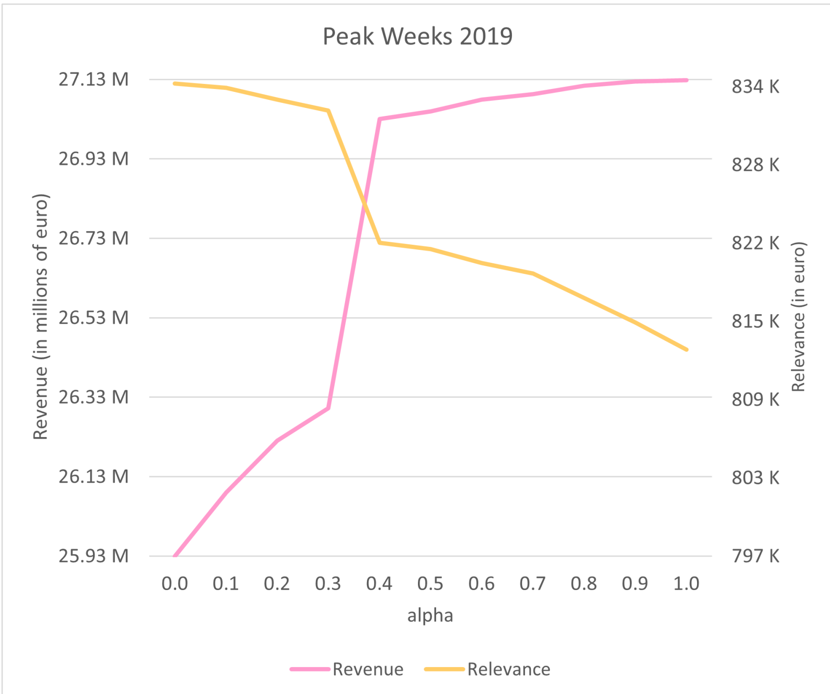

Figure 10: Trade-off between revenue and relevance for different values of the weight parameter α in the peak weeks of 2019.

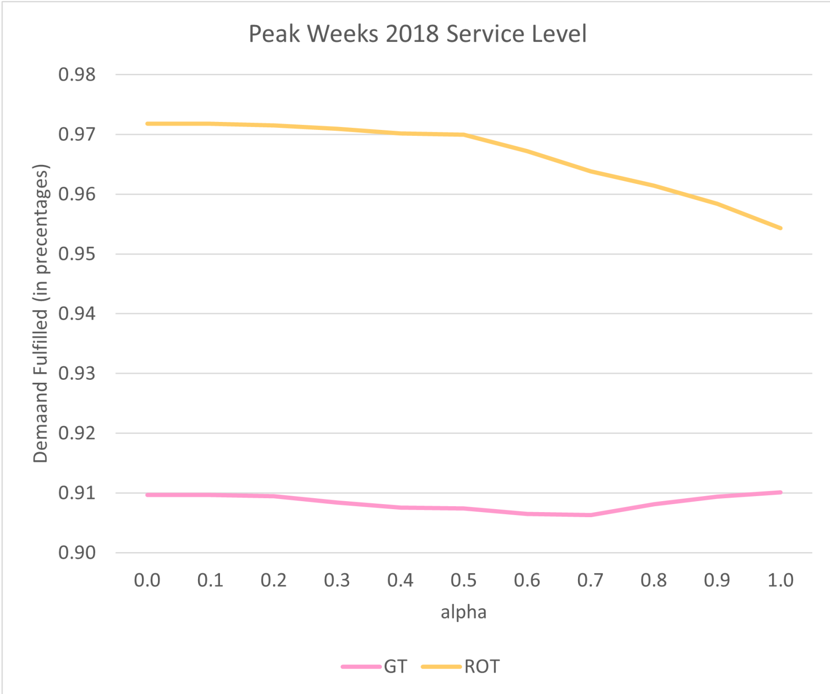

Figure 11: Service levels of GT and ROT in 2018 during the peak season.

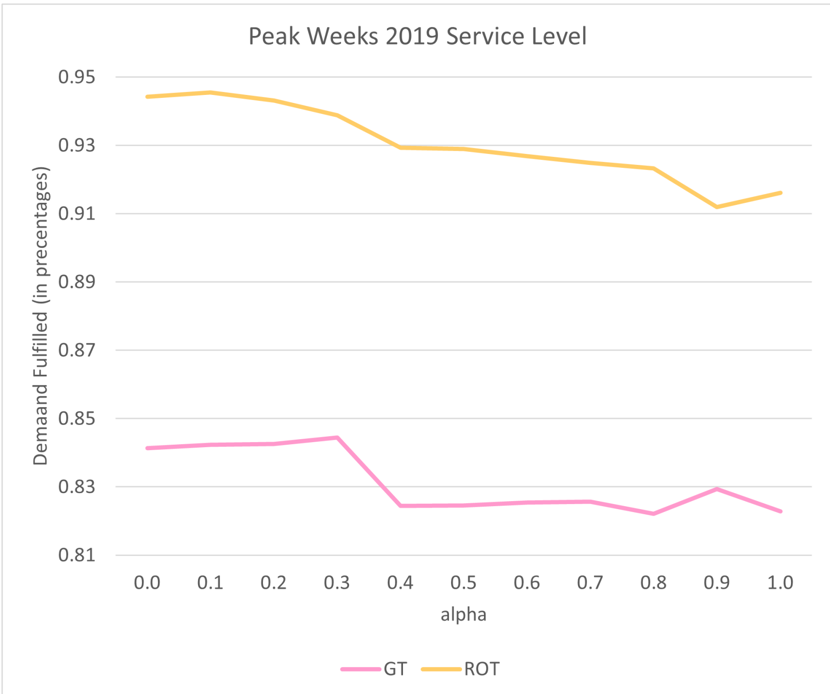

Figure 12: Service levels of GT and ROT in 2019 during the peak season

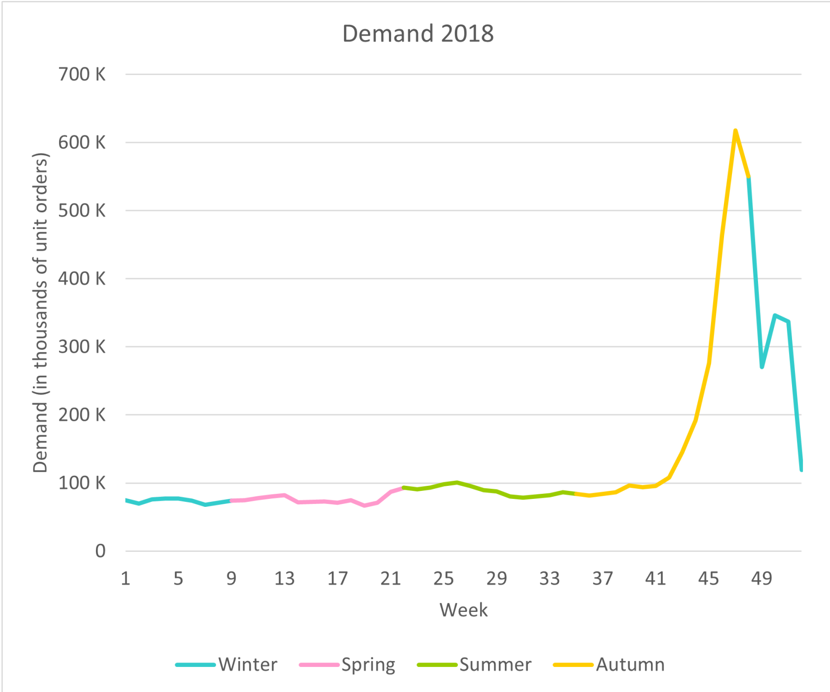

Figure 13: Total demand over the 52 weeks in 2018, separated by season.

Figure 14: Total price over the 52 weeks in 2018, separated by season.
‘

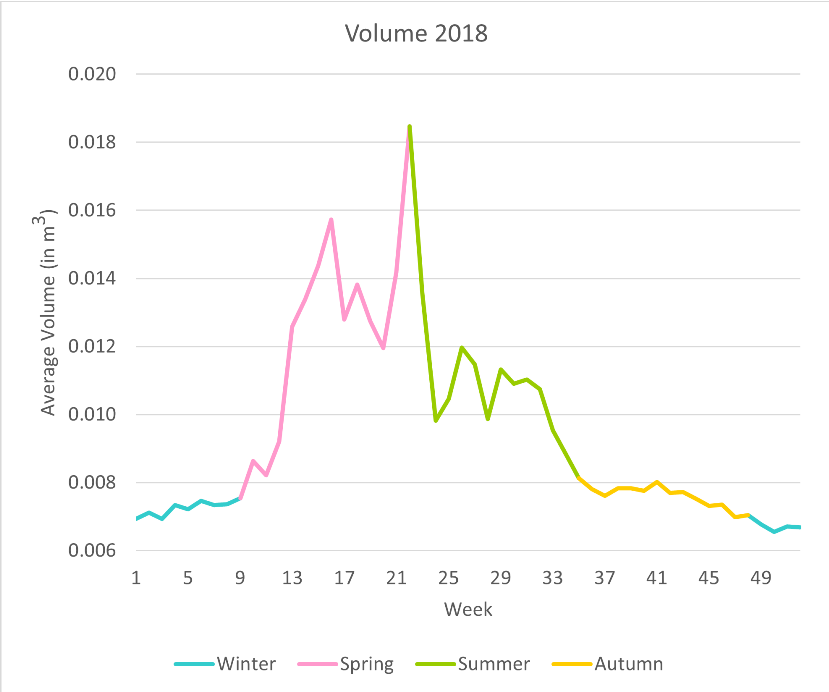

Figure 15: Total volume over the 52 weeks in 2018, separated by season.

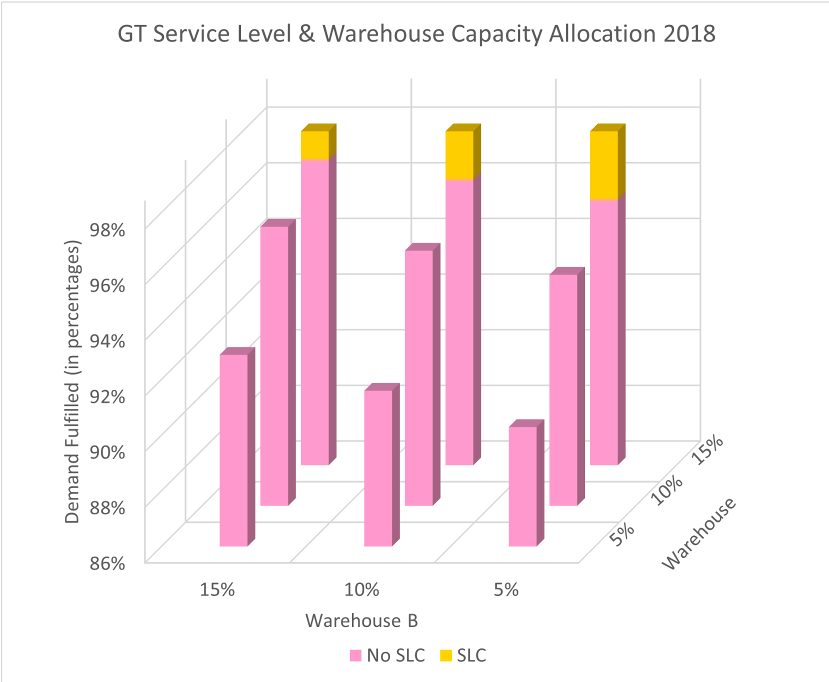

Figure 16: The comparison of the GT service levels in 2018 with and without service constraints for different warehouse capacity allocations.

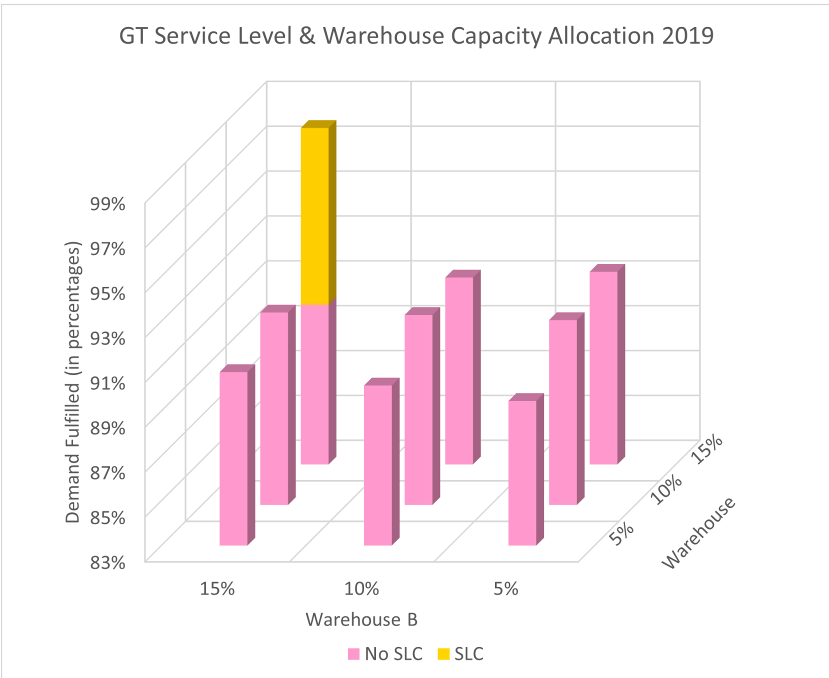

Figure 17: The comparison of the GT service levels in 2019 with and without service constraints for different warehouse capacity allocations.

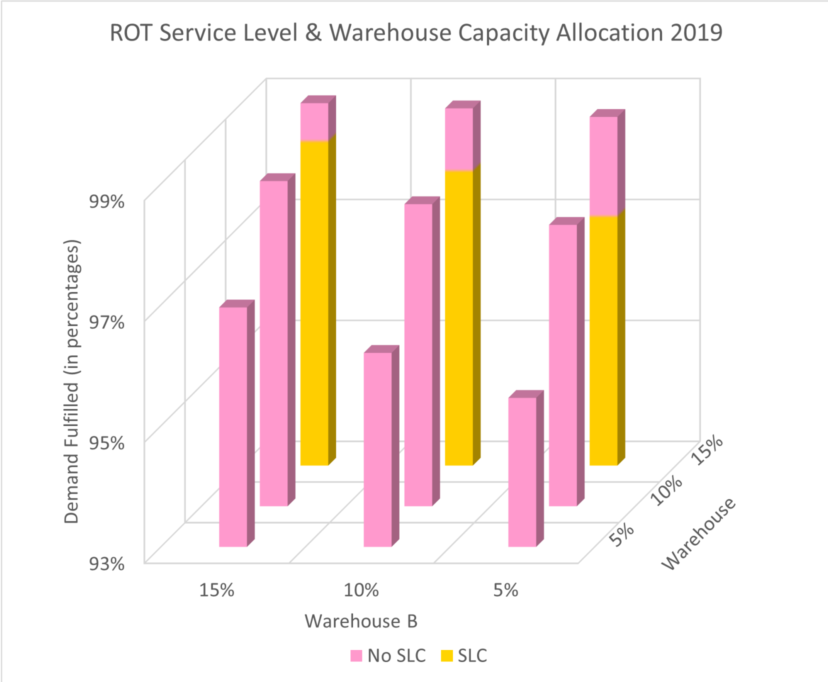

Figure 18: The comparison of the ROT Service levels in 2018 with and without service constraints for different warehouse capacity allocations.

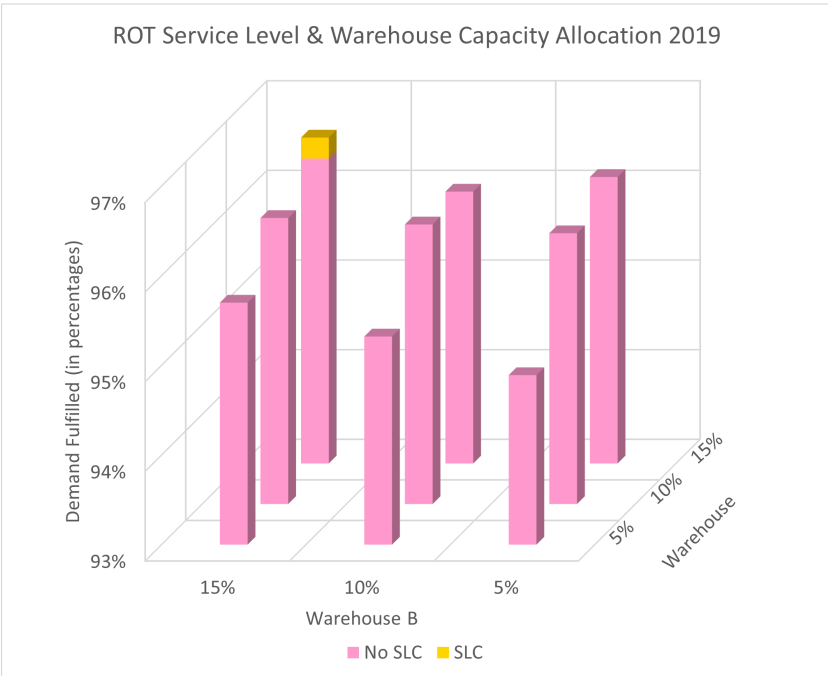

Figure 19: The comparison of the ROT Service levels in 2019 with and without service constraints for different warehouse capacity allocations.

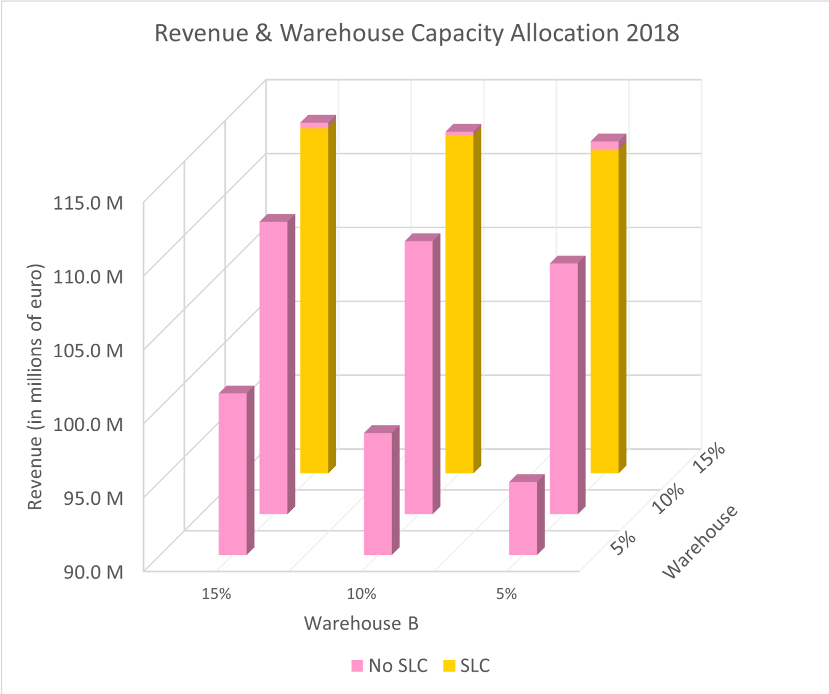

Figure 20: The comparison of the revenue in 2018 with and without service constraints for different warehouse capacity allocations.

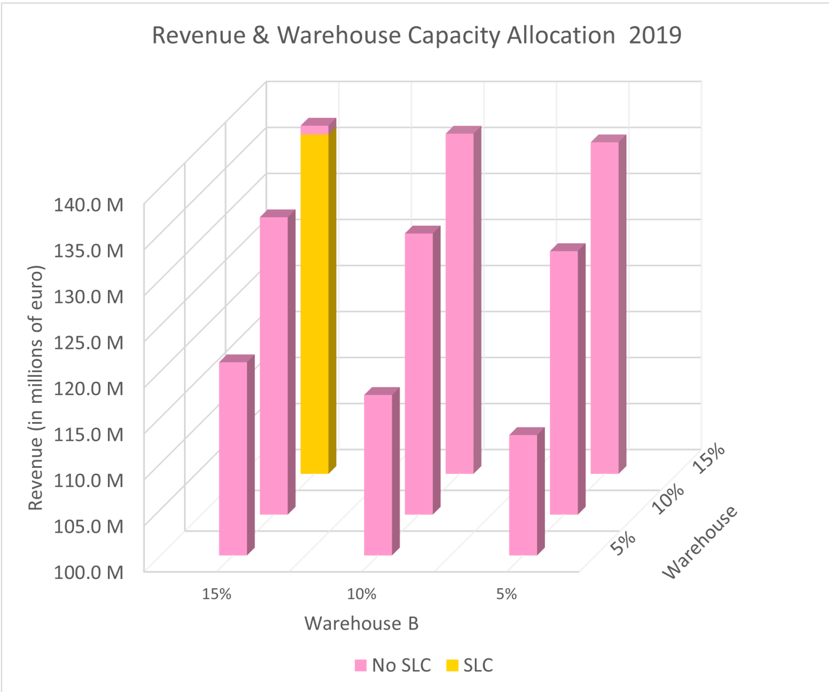

Figure 21: The comparison of the revenue in 2019 with and without service constraints for different warehouse capacity allocations.

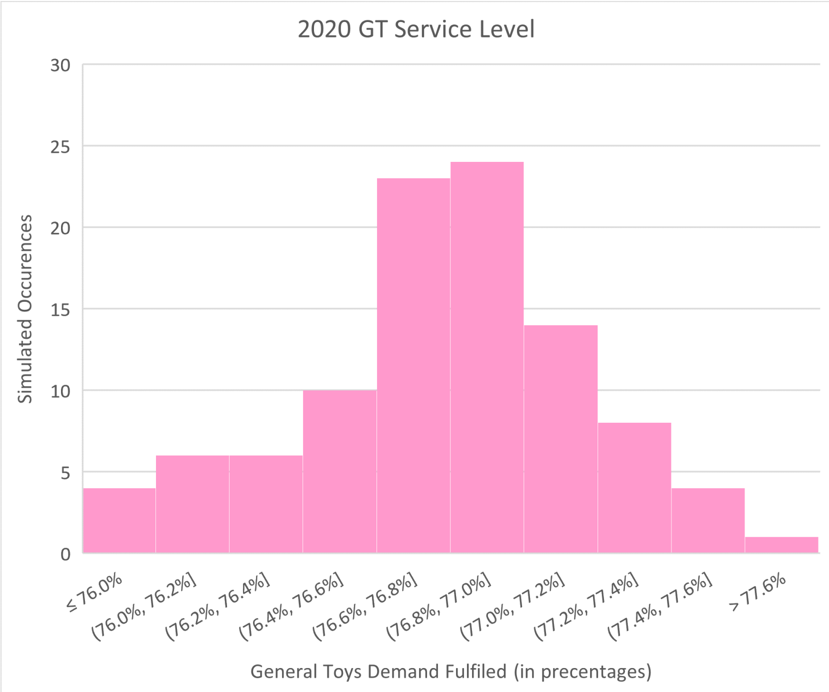

Figure 22: Simulated GT service levels occurrence under revenue-maximization model in 2020.

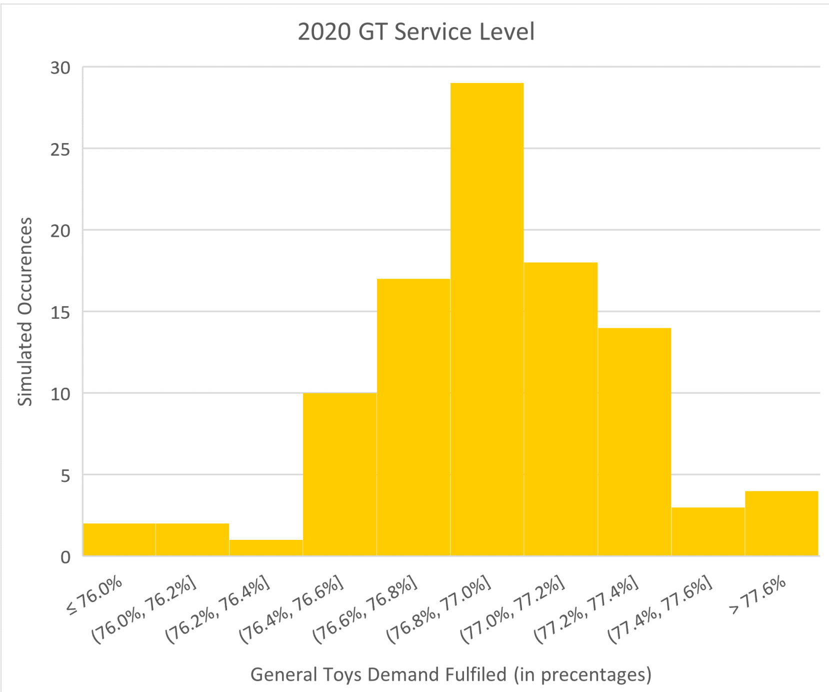

Figure 23: Simulated GT service levels occurrence under relevance-maximisation model in 2020.

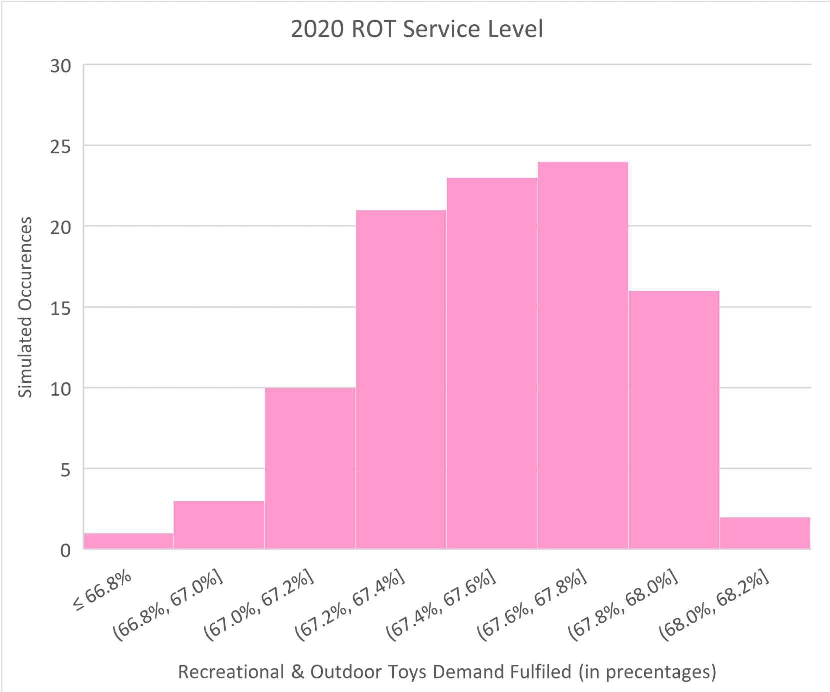

Figure 24: Simulated ROT service levels occurrence under revenue-maximization model in 2020.

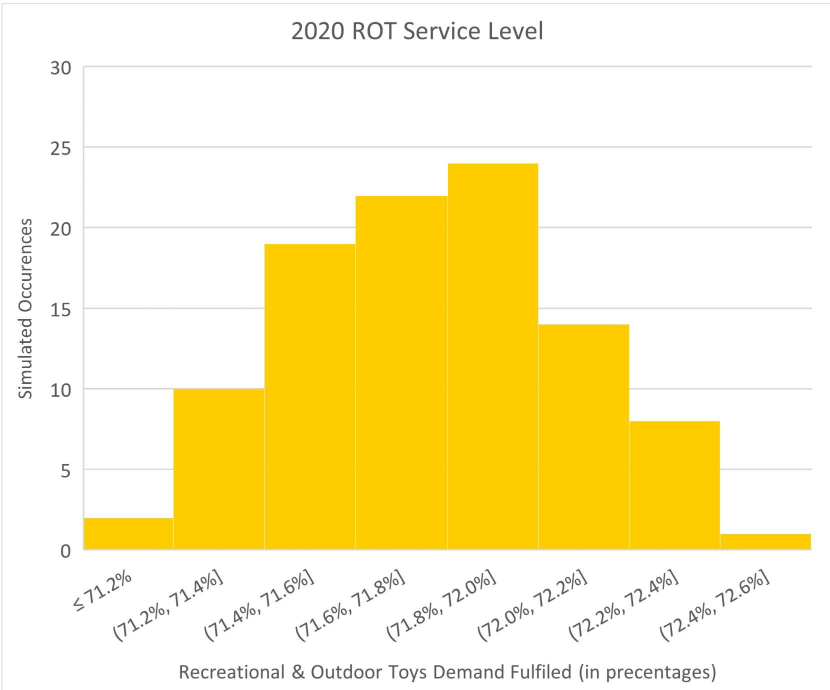

Figure 25: Simulated ROT service levels occurrence under relevance-maximisation model in 2020.

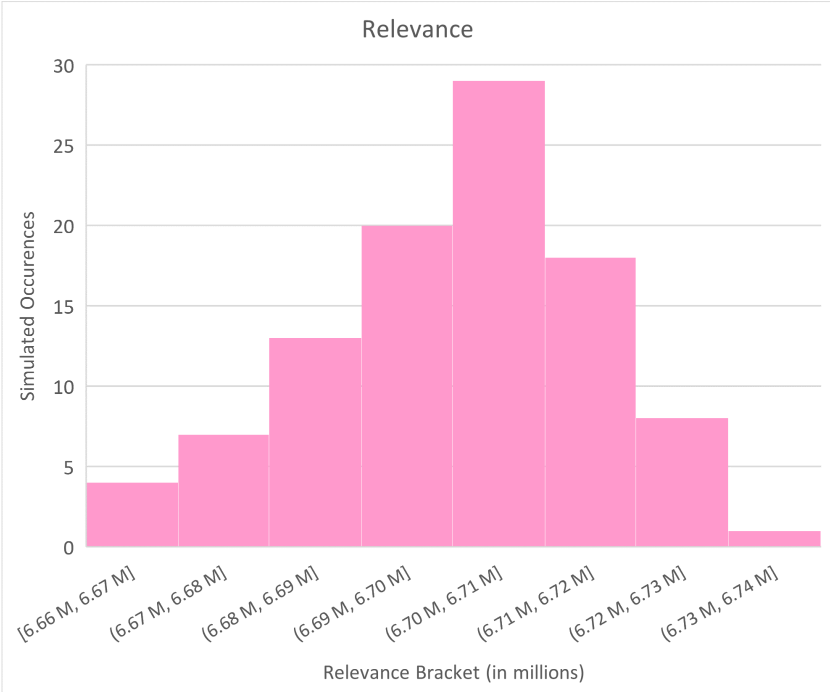

Figure 26: Simulated relevance occurrence under revenue-maximization model in 2020.

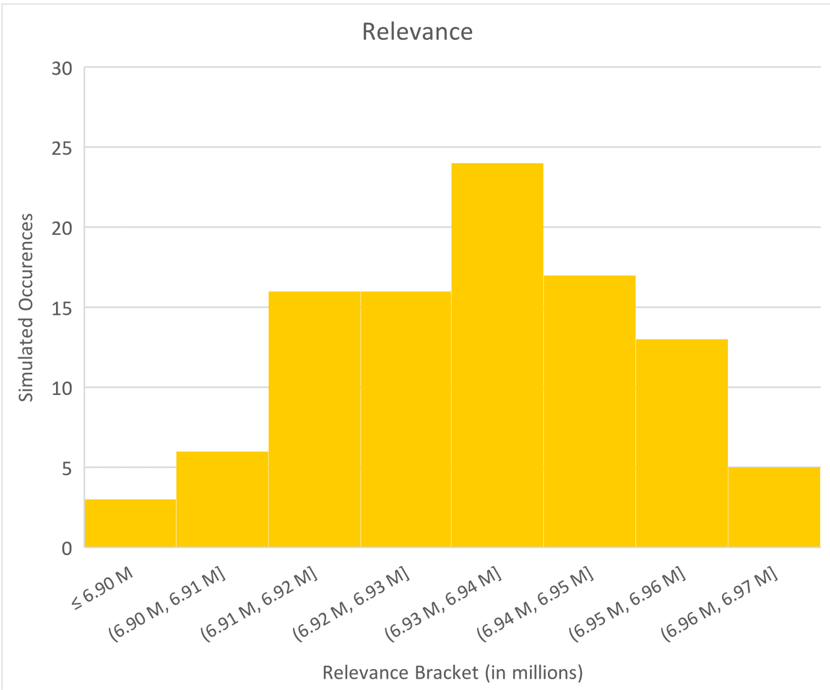

Figure 27: Simulated relevance occurrence under relevance-maximisation model in 2020.

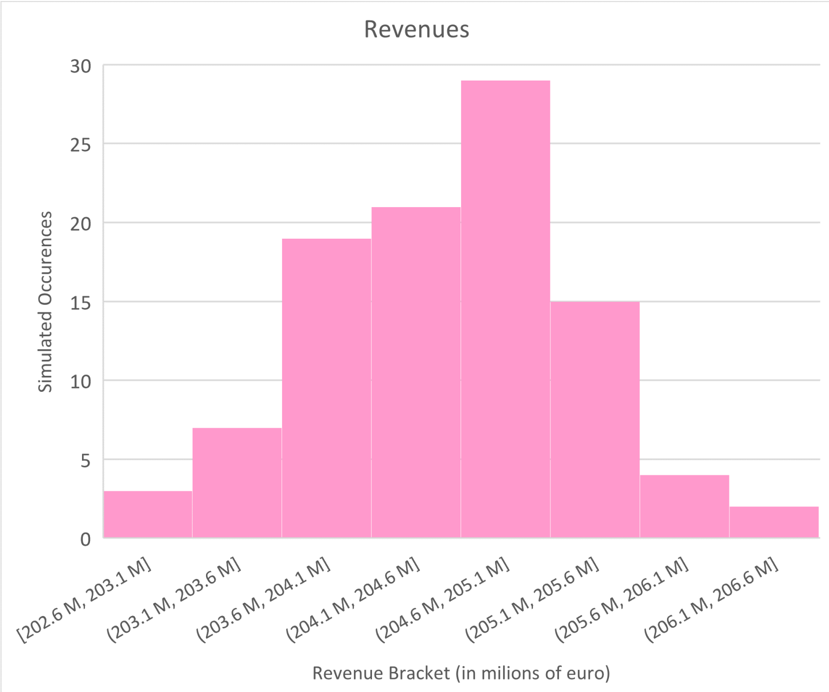

Figure 28: Simulated revenue occurrence under revenue-maximization model in 2020.

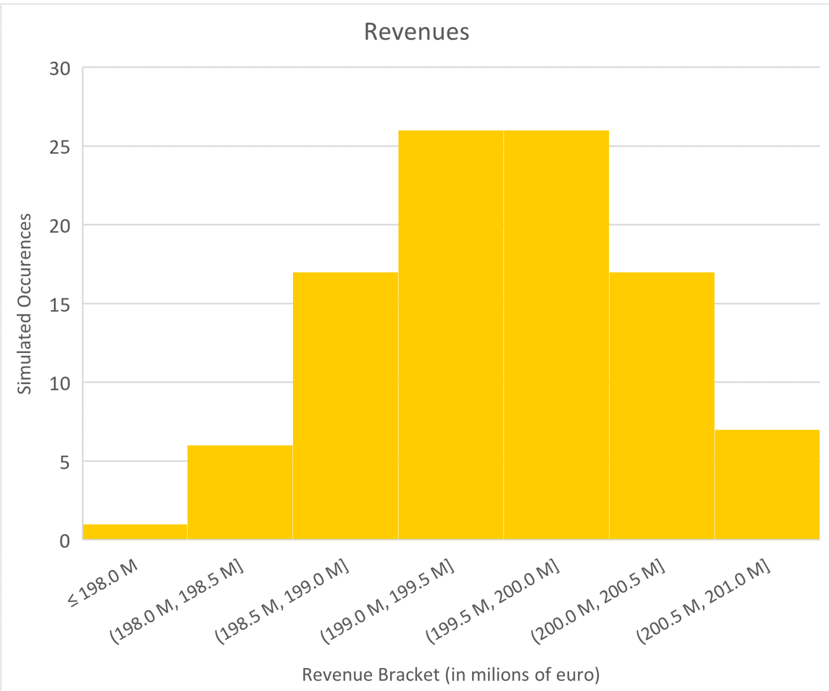

Figure 29: Simulated revenue occurrence under relevance-maximisation model in 2020.