AIR CARGO REVENUE AND CAPACITY MANAGEMENT

A Thesis
Presented to
The Academic Faculty

by

Andreea Popescu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2006

AIR CARGO REVENUE AND CAPACITY MANAGEMENT

Approved by:

Dr. Ellis L. Johnson, Adviser
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Pinar Keskinocak, Adviser
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Hayriye Ayhan
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Julie L. Swann
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Dirk P. Günther
Department of Research and Development
Sabre Holdings

Date Approved: November 16, 2006

To my parents,

Ana and Cornel,

for their unconditional and unbreakable support

ACKNOWLEDGEMENTS

I would like to thank my family, especially my husband, who has never stopped believing in me, and has never complained being a mother and a father to our two children while I was working on my research. Without him I would have never become the person I am today.

I am very grateful I had the chance to meet my advisors, who have had so much patience and have never given up on me. The support I received from them was crucial to me, and their trust gave me the strength to finish what I had started. I would like to express my gratitude to all members of my committee for their continuous support and helpful suggestions.

I would like to thank my good friend, Ionut Porumbel, from the School of Aerospace Engineering, for sharing all the ups and downs with me, for helping me with his expertise, and for being there for me at any time.

Last but not least, I would like to thank my sister and my brother; their unconditional love is guiding me through life and is helping me be a better person.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The air cargo industry has substantially grown over the past few years, driving the need of a structured environment with the explicit goal of maximizing revenues. The air cargo supply chain is composed of shippers, freight forwarders, and airlines. The shippers send their shipment to freight forwarders, who are responsible for contacting the airlines and procuring space to ship the cargo according to the shippers' needs. Currently, the process takes more time than needed due to a lack of coordination between freight forwarders and airlines; it is said that an integrator, a freight forwarder which owns its own fleet, moves an international shipment two or three times faster than a traditional freight forwarder/airline team [10].

The scope of this thesis is to propose a structured methodology for improving the airlines' and freight forwarders' actions when confronted with accepting demand and acquiring capacity respectively. We develop methods to tackle two air cargo revenue management problems: space allocation and show-up rate estimation.

The space allocation problem is defined as distributing the available capacity for free sale among incoming cargo bookings over the booking horizon such that the revenue at the end of the booking period is maximized. We use bid price methods to accept/reject incoming bookings: if the rate of the booking is lower than the bid price value then the booking is rejected. We show that a good approach to deal with the demand lumpiness encountered in the cargo industry is to split the cargo into two categories: small, which contains small packages and mail, and large, which contains the bulk of commercial cargo. Whereas the small cargo demand behavior can be approximated with the passenger demand behavior, and techniques from the passenger sector could be adapted for the small cargo business, the large

cargo demand behavior shows similarities to wholesale retail and calls for different methods. We model the small cargo revenue management problem using a model from the passenger business and propose a new algorithm to solve it, which had a superior running time among the few algorithms known to solve the same model in the passenger business. The large cargo revenue management problem is solved via dynamic programming. In our simulations, when the demand is extremely lumpy, i.e., the cargo loads vary widely, the conjugated solutions from the two models result in up to 60% more revenue than the first come first serve method used in practice.

The second air cargo revenue management problem is estimating the cargo show-up rate, which is the ratio of cargo handed in at departure over bookings on hand. The show-up rate is used in the overbooking models to estimate the capacity available for free sale before the departure date. In the passenger business, the current practice is to estimate the show-up rate based on a Normal distribution. We show the Normal distribution is not suitable for the cargo business and propose a discrete distribution based on wavelet estimation. In a simulation study conducted for a set of real world demand date, the average yearly savings resulting from using the discrete estimator for a fleet of 300 flights per day and an average of cargo capacity per departure of 13,000 kilograms was $16,425,000.

Besides the airline's revenue management problems, we solve the capacity management problem for the freight forwarder. The freight forwarder bids for cargo space on flights offered by the airlines several months before the actual departure date of the aircraft. The committed capacity has to be confirmed a few days before the departure date. In spite of its importance, there are no known solutions to this problem. We propose modeling the problem as a perishable inventory problem with backlog and lead time. The lead time is the time between which the freight forwarder orders the capacity and the time the aircraft is scheduled to take off. During this time, new demand from shippers can materialize, and the freight forwarder should place the order

such that it accounts for the uncertainty related to shippers' demand. We propose a stochastic dynamic programming model to illustrate the freight forwarders' problem and show that the value function is a convex function in the state variable, which is either available capacity for the current period if it is positive, or backlog if it is negative, for a lead time of one and two periods. Furthermore, we show the optimal policy is a stationary policy, depending only on the value of the state variable and not on the period in which the order is placed. We also analyze special cases with subcontracting options when we have orders that have to be shipped immediately, and show the solution is a critical ratio solution under certain conditions.

CHAPTER I

INTRODUCTION

Revenue management has become more and more popular among industries of all sorts; traditionally applied to the airline industry, it extended to all business niches where the product offered is perishable, limited and price differentiable. One relatively new area, which became more and more important in the recent years, is the air cargo transportation. The product offered in this case, the cargo capacity, has all the features for revenue management techniques to be successful: it is lost after the plane takes off, it is limited, and it can be offered at different rates depending on the service offered - e.g., critical and speciality cargo, expedited, standard, etc. However, the transportation of freight by air was considered a premium transportation service by most companies. Only recently, due to globalization of trade, the rise of the e-commerce and increasing used of advanced logistics techniques, the air transportation of freight has become more affordable, and hence the airlines have started to show interest in managing their cargo capacity such that their expected revenue is maximized.

The air cargo supply chain is composed of three main players: the shippers, the freight forwarders, and the airlines. The freight forwarders are responsible for acquiring cargo space from airlines in order to satisfy the demand from shippers. The process of acquiring capacity from the airlines goes over two phases:

1. In phase one, six to twelve months before the actual departure, freight forwarders bid for cargo space the airlines have to offer; the cargo capacity committed during the auction process is called *allotted capacity*;

2. In phase two, a few days before the actual take off, the freight forwarders have

to confirm the allotted space, either returning unwanted space or confirming the need of the whole allotted capacity.

The remaining capacity available for free sale, which we refer to as *cargo capacity* throughout this thesis, is the object of the revenue management techniques developed in this thesis.

The air transportation supply chain poses several challenges: the freight forwarders and the airline should be coordinated and respect each others' needs: the airline should be able to honor its allotted capacity, and the freight forwarder should know approximately how much capacity it will actually need and release unwanted space in a timely matter for the airline, such that the latter is able to add the extra space to the pool of capacity available for free sale.

Unfortunately, this is not the case. It is said that due to the lack of coordination between the airlines and the freight forwarders the traditional airline/freight forwarder team needs two to three times more time than an integrator (like e.g., Fedex) to move an international shipment [10].

The objective of this thesis is to develop methods to improve both the freight forwarders' and the airlines' actions when dealing with air cargo. We have two problem categories:

1. **Air cargo revenue management** - which refers to the airlines' problem to manage the capacity available for free sale to generate more revenue. We show that applying revenue methods blindly from the passengers business to the cargo sector is not suitable, and we develop solutions for two problems in the area:

    (a) **Air cargo bid prices** - Bid prices are threshold values used to assist in the process of accepting/rejecting incoming bookings. If the incoming booking has a rate lower than the sum of the bid prices along the requested itinerary then the booking is rejected, otherwise it is accepted. We argue the unsuitability of the passengers techniques to determine bid prices for the cargo sector, and we tackle the problem of demand lumpiness existent in the air cargo industry. Whereas for passengers there is a clear matching between the request and the supply, which is exactly one seat, the demand for air cargo can vary from 0.001 kilograms to 100,000 kilograms. This disproportion introduces a new factor of difficulty besides two- or more dimensions of demand (weight, volume, container position), uncertainty of capacity available for free sale (depending on allotment utilization), and routing flexibility (cargo has to make its destination on time, no matter which route it takes), recorded in most of the air cargo papers (see e.g., [**?**]). To solve this problem, we split the cargo loads into two categories, small and large, based empirically on the demand features. The demand for small cargo shows similarities with the passenger demand, with a relatively high number of bookings during a quite wide booking window; on the contrary, the demand for large cargo comes from a relatively low number of customers, and it is booked during the last few days before the aircraft departure. We split the two streams of demand: we formulate the small cargo via a mathematical model from the passenger business, and we develop a new algorithm to solve it; we show that a dynamic program can be explicitly solved for large cargo if we decompose the network problem at the flight leg level.

(b) **Air cargo show-up rate estimation** - the air cargo show-up rate is the ratio of demand handed in at departure over the amount of bookings on hand at any point in time before the departure date; the show-up rate is used in the overbooking model. Overbooking is the technique used by airlines to sell more capacity than physically available to hedge against demand variability. In the passenger business the Normal distribution is used to determine overbooking levels. We show on real cargo data that this distribution is not suitable for cargo and define a discrete show-up rate distribution based on wavelet density estimation.

2. **Air cargo capacity management** - which refers to the freight forwarders' problem to confirm capacity with airlines a few days before the flight takes off such that the demand from shippers is satisfied at minimum cost. The freight forwarders' problem is modeled as a perishable inventory problem, where the perishable commodity is the capacity confirmed, which is lost after the flight departure and cannot be used for subsequent shipments. There is a lead time between the time the capacity confirmation is placed and the time the flight departs, which is usually three days. During these days, new demand materializes, and the freight forwarder has to backlog excess demand at extra cost if the capacity ordered cannot accommodate all demand. In the current air cargo industry the airlines do not penalize the freight forwarders for not using confirmed capacity; i.e., the perishing cost of capacity is the same as the ordering cost. We find the optimal solution for one and two period lead time, and for special cases when there are subcontracting options and the demand has due dates.

As the problems described above are disconnected from each other, we present them individually in Chapters 2 and 3. Chapter 2 contains the air cargo bid price problem in Section 2.1, and the air cargo show-up rate estimation in Section 2.2. Each section contains its own detailed description of the problem, of existing work, and of numerical experiments. The last chapter concludes the thesis and proposes directions for future reasearch.

CHAPTER II

AIR CARGO REVENUE MANAGEMENT

Revenue management focuses on maximizing profits from a limited capacity of a product by selling it to the right customer for the right price at the right time [40]. The fundamental decision in revenue management is whether to sell capacity when a request comes in, or to save it for a potential later sale at a higher price. For example, a seat on an airplane can be sold at different prices, depending on the capacity already sold and the time remaining until the departure of the aircraft. While revenue management practices have been widely used in the passenger segment of the airline industry, they received increased attention only recently in the cargo segment. Due to the globalization of trade and increased volumes of e-commerce, the worldwide demand for air cargo has been growing at a faster pace than passenger demand [38]. Industry forecasts predict a massive growth in cargo demand: world air cargo traffic will expand at an average annual rate of 6.2% for the next two decades, tripling over current traffic levels [6]. Hence, many airlines now recognize the revenue potential from cargo, and aim to make it a "vital component of their business rather than a sideline operation" [52].

We analyze two problems that arise within the air cargo revenue management area: space allocation using bid prices and estimation of the show-up rate with impact on the overbooking levels. In this chapter we present solutions to the two problems, with simulations and numerical results for each.

2.1 Bid prices for air cargo

The main players in the air cargo supply chain are: airlines, freight forwarders (FF), and shippers. FF satisfy the shippers' demand by securing cargo capacity from the airlines. Generally, airlines offer cargo space in two stages. In the first stage, a few months prior to a season, FF bid for cargo space over the next season; the cargo capacity committed during this bidding process is called *allotted capacity.* Out of the remaining cargo space, airlines usually allocate an additional amount to *contracts*, which is the space reserved for large customers at a fixed price. The remaining space, *the capacity available for free sale*, is open for booking in the second stage, within four weeks before the flight departs. The allocation of the capacity available for free sale (which we refer to as "cargo capacity" throughout the thesis) to the demand that arrives over time constitutes the object of our study. This problem is very similar to the seat inventory control problem in the passenger revenue management literature, which is defined as allocating the finite seat inventory to demand that occurs over time, such that at departure the plane is filled with the most profitable mix of passengers.

The decision of whether to accept or to reject an incoming booking request (for a seat on the plane or cargo capacity) can be based on different strategies. The most important types of control in the passenger segment are: booking limits, protection levels, and bid prices. Booking limits allocate a fixed amount of capacity to each fare class. Protection levels specify an amount of capacity to be reserved for a fare class or a set of fare classes. Bid prices are threshold values used to accept/deny incoming booking requests, i.e., the decision maker accepts the request if the sum of the bid prices along the itinerary is lower than the proposed fare. Unlike booking limits and protection levels, which are capacity based, bid price controls are revenue based and have the advantages of being simple, having a natural interpretation as the marginal value of a given resource, and have a very good revenue performance [53].

While the capacity allocation problems from the passenger and cargo segments have similarities, there are also some significant differences [**?**]: (i) For the passengers, the unit capacity is defined by a single dimension, i.e., seat; for cargo, capacity has two dimensions (weight and volume). (ii) Cargo capacity is often uncertain due to allotments, no-shows, and passenger luggage on combination carriers. (iii) Most passengers demand a specific itinerary whereas for cargo shipments customers often specify an origin and destination and accept any itinerary as long as the shipment arrives at the destination by the requested delivery time (itinerary-specific versus origin-destination-specific demand); hence, in cargo there is flexibility in routing. (iv) Most passengers demand one unit of capacity (seat) whereas customers request multiple units of capacity for cargo shipment (specified by weight and volume). Hence, cargo demand has a wide range of quantities[1] and can be "lumpy," which complicates matching demand to capacity.

In this thesis, focusing especially on the demand arrival characteristics (lumpiness) of cargo shipments, we develop efficient methods for determining bid prices for air cargo capacity allocation. We propose dividing the cargo bookings into two categories: "small" (packages and mail) and "large" cargo. The two categories exhibit similarities in booking behavior to retail and wholesale demand, respectively. Small packages are generated by a large number of independent customers and can be modeled by a a Poisson process, similar to that for passengers. By contrast, wholesale demand tends to come from a relatively small number of customers in large quantities and consumes the available capacity relatively quickly. For small cargo, we adapt existing passenger seat allocation methods and propose an efficient algorithm to solve the model from [13], based on a quadratic approximation scheme of the dual problem. To the best of our knowledge, the approach is novel and it proves superior computational efficiency as well as a significant increase in revenues over existing algorithms. For large cargo, we develop a dynamic programming model and decompose the problem into single leg problems. The main advantage of splitting the cargo bookings into a "small" and a "large" category is that the dynamic program for the latter becomes computationally

[1] The cargo weight booked on a single flight can range anywhere from 0.1 kilograms to 10-20 tons.

manageable due to the low number of bookings and their large sizes.

In Section 2.1.1 we provide an overview of the related literature. In Section 2.1.2 we describe the current air cargo booking process and motivate our approach, followed by the mathematical formulation of the small cargo problem and the details of the proposed algorithm in Section 2.1.3. Section 2.1.5 describes the dynamic program used for calculating the bid prices for large cargo. Numerical results are discussed in Section 2.1.6.

2.1.1 Literature review

Bid price methods were introduced by Smith and Penn [51], and extended by Simpson [50], and Williamson [57]. We provide an overview of the traditional mathematical models from the passenger literature to calculate network bid prices in Section 2.1.3.

Chen et. al. [13] addresses the issue of origin-destination-specific demand, which is common for low fare passengers and cargo, versus the itinerary-specific demand, widely used in the passenger models. They extend three popular models to incorporate origin-destination demand and introduce a routing algorithm tailored towards the special structure of the flight networks and their objectives. The simulation results report the superiority of an extended probabilistic model over a first come first serve (FCFS) policy applied to cargo revenue management. Rao [45] proposes a dual ascent scheme to solve the Lagrangian of the probabilistic model used in [13].

The air cargo revenue management literature focuses mainly on overbooking models. In practice, the relationship between airlines and FF is governed by long-term contracts that do not specify penalties for reserved but unused space. As a result, overbooking rather than space allocation has been the focus in the air cargo literature. Kasilingam [24] proposes a model that considers stochastic capacity as opposed to the passenger business where the capacity is certain. Luo et al. [27] extend Kasilingam's one dimensional model to two dimensions (weight and volume) under cost minimization. Moussawi and Cakanyildirim [29] consider profit maximization instead. Amaruchkul et al. [2] formulate the capacity allocation problem with random weight and volume as a Markov decision process. Arguing that the exact solution of the formulation is impractical due to its high-dimensional state space, they develop a few heuristics to approximate the value function. Extensive simulation experiments suggest that the value function approximation derived from solving separately for weight and volume for a single-leg model offers the best approach.

Xiao and Yang [58] and Pak and Dekker [39] address theoretical aspects of revenue management under multi-dimensional capacity. Xiao and Yang [58] model the problem as a continuous time stochastic control model, and derive structural properties for the case where the remaining capacities in two dimensions are equal or differ. When they are equal, they show that the optimal policy is not characterized by a nested price structure (if a fare class is open, then all classes with higher fares should also be open) as in the one-dimensional case. Pak and Dekker [39] model the problem as a multi-dimensional on-line knapsack problem and propose a heuristic to determine the bid prices based on a greedy algorithm proposed by Rinnooy et. al. [46]. A test case shows that the bid prices perform better than the traditional deterministic model used in the passenger business (see section 2.1.3 for details on different models used to approximate bid prices in the passengers literature).

All the different cargo characteristics mentioned earlier, multi-dimensionality, routing flexibility, uncertainty in capacity supply, have been addresses in the literature, but not the lumpiness of demand. To the best of our knowledge, we are the first to identify this feature, address it by splitting the cargo into two categories, and solve it. Our solution is shown to be better than two other methods widely used in practice.

2.1.2 Current air cargo booking process and motivation for our approach

Our motivation for dividing cargo into two categories (small and large) comes from the practices of the airlines and integrators who usually distinguish mail and packages from freight based on some threshold weight values. Integrators (such as FedEx) and carriers (such as Delta Airlines) consider everything under 150 lbs a “package,” and everything between 151 lbs and 2,200 lbs “freight”, which is attached to a skid (a forkliftable base used to support/elevate an object, typically made of hardwood or plastic) [19]. Raja Kasilingam, the Vice President of the Cargo Products at Sabre Holdings, confirmed that the cut off value between packages and freight is typically between 100 and 200 kilograms. Furthermore, airlines and integrators dedicate separate (non-overlapping) capacities to the two categories.

To further confirm the gap between small and large cargo bookings (and to use later in our computational study), we collected data (daily bookings by weight) from 4 combination carriers over a horizon of 29 departure dates. In the data, the average capacity dedicated to large cargo (freight) and small cargo is 5000 kilograms and 1000 kilograms, respectively. The average number of bookings (over the entire booking horizon of 30 days before departure) is 10 and 100 for large and small cargo, respectively. Based on these observations, we decompose the cargo capacity allocation problem into two, one focusing on large cargo and the other one focusing on small cargo.

The goal in air cargo capacity allocation is to maximize the expected revenues. In cargo, unlike in the passenger business, the revenues associated with different shipment categories are not as well structured.[2] They are often driven by customer relationship to the airline and the specific market, as well as the shipment features

[2] In the passenger business, the fare classes are well defined based on seat location (business, economy), and some additional features, such as flexibility of the departure and return dates, refundable versus non-refundable tickets, etc.

such as perishability, weight, and the time of booking. We follow [2] and define the shipment rates as a function of their "dimensional weight" and "class." Shipments belong to different classes depending on the types of goods (perishable, electronics) or arrival time.

We mentioned earlier that cargo has two dimensions, weight (w) and volume (v). However, in the cargo sector, booking volumes are usually not reported before the departure date, but very often derived by applying a standard "density" to the booked weight. The density is the amount of space a package occupies in relation to its actual weight, which is used in calculating the "dimensional weight" of a shipment.[3] The maximum between the actual weight and the dimensional weight is used as the billable quantity.

We use a revenue function to map the billable weight, $\hat{w} = \max\{w(kg.), \frac{v(cm^3)}{6000(\frac{cm^3}{kg.})}\}$ to revenue according to a schema such as the one in Table 1.

Table 1: Air cargo rates as a function of billable weight and cargo class

Class (in kg.)	**1**	**2**	...
$0 < \hat{w} \leq 20$	\$1.50	\$ 1.25	...
$20 < \hat{w} \leq 50$	\$1.45	\$ 1.2	...
...	...	...	...

The available data we have currently suggests that in practice the decision of whether to accept or reject a booking is mainly based on weight. Also, Amaruchkul et al. [2] showed that separate one-dimensional models for weight and volume gave better revenue performances in a simulation study than an approximate two-dimensional model; hence, our models focus on weight (i.e., they are one-dimensional). The air cargo classes defined as in Table 1 together with an itinerary request are used as the "fare classes" within the mathematical models that follow; although the air cargo

[3] Calculations of dimensional weight are based on the International Air Transport Association (IATA) volumetric standards. For the dimensional weight of a package in kilograms, the cubic size of the package in centimeters is divided by 6000.

industry is more flexible when the routing is concerned, the current practice is to book a specific cargo itinerary, as Raja Kasilingam from Sabre Holdings confirmed. Hence we developed our models with itinerary specific demand.

2.1.3 Small cargo capacity allocation

There are two commonly used mathematical programming models in the Passenger Revenue Management (PRM) literature for computing bid prices on a flight network: (i) The deterministic linear programming model (DLP) makes the assumption that the demand is deterministic and equal to its mean; (ii) the probabilistic nonlinear programming model (PNLP) maximizes the expected revenue assuming a randomly distributed demand. Recently, randomized linear programming models (RLP) have been proposed to incorporate the stochasticity of the demand into the DLP. A comprehensive analysis of the DLP and PNLP models can be found in Williamson [57], and an analysis of the DLP, PNLP, and RLP can be found in Talluri and Van Ryzin [53].

For PRM, DLP has shown to generate consistently more revenue than PNLP (see for example [18], [53], [57]). Both DLP and PNLP partition the network capacity into seat allocations to all possible combinations of itinerary-fare classes. This partitioned allocation does not reflect the real booking control policy where the seat allocations are nested, i.e., the highly profitable customers do not have access to less profitable itinerary-fare classes. PNLP suffers more from ignoring the nesting environment than its deterministic counterpart (see [18]). It tends to overprotect (i.e., allocates higher capacity) to high-margin classes than DLP; however, unless these classes are highly profitable, PNLP results in less expected revenue. On the other hand, in a non-nested environment PNLP yields higher profits than DLP, as Ciancimino et al. [15] have demonstrated for the railway industry, where the non-nesting assumption holds.

When applied to the air cargo industry, the bid prices derived from DLP have proven to be almost non-restrictive (see [39]), i.e., they reflect a first come first served (FCFS) capacity allocation policy. Clearly, allocating the capacity in a FCFS basis is, in general, not very profitable and goes against the fundamental premise of revenue management where some capacity is reserved for high-margin customers. This motivated our study for a probabilistic model for deriving bid prices for air cargo, considering the variability in demand. The research on developing efficient solution methods for PNLP has been very limited. Besides the specialized algorithm proposed by Ciancimino et al. [15] for the railway yield management problem, and the algorithm proposed by Rao [45] to solve a slightly different formulation, we are not aware of any other published work on algorithms for solving PNLP. In [15], the original constrained problem is transformed into an unconstrained minimization of a continuously differentiable merit function. Rao [45] formulates PNLP for cargo by nesting the objective function according to different costs of the routes, and uses Lagrangian relaxation schemes to solve it. The major difference between our work and the existing work is that we explicitly formulate and solve the dual PNLP. Our proposed method is shown to be highly efficient.

2.1.3.1 The probabilistic nonlinear program (PNLP)

An airline's flight network for a given departure date can be modelled by a graph $G = (V, E)$ where V denotes the set of nodes, representing departure/arrival times and cities, and $E = \{1, \ ..., L\}$ denotes the set of arcs, representing flight legs, where b_l is the (remaining) capacity (in kilograms, at a given time) on leg l.

An arriving customer specifies a weight w and an itinerary h for its shipment. Following the approach in Section 2.1.2, a rate r is assigned to the customer's shipment. A demand class is uniquely defined by a customer type $j = (r, w, h)$. The information about demand classes can be summarized in a connectivity matrix A, where each column A_j represents a demand class and each row a flight leg. An entry a_{ij} is equal to

Table 2: Notation for PNLP

A	:	$m \times n$ connectivity matrix with m number of legs and n demand classes
A_j	:	jth column of A (incidence vector for demand class j)
A^i	:	ith row of A (set of demand classes on leg i)
r_j	:	rate associated with demand class j
b	:	capacity vector for all legs
D	:	non-negative random variable for demand with known probability density function (pdf) $\phi(t)$ (continuous and differentiable) and cumulative distribution function (cdf) $\Phi(t)$

1 if leg i is a part of itinerary h associated with demand class $j = (r, w, h)$; otherwise, $a_{ij} = 0$. Note that identical columns are allowed but they differ in their contribution to the revenue in the objective function. Our notation is summarized in Table 2.1.3.1.

The goal is to find a partition $x = \{x_1, \ldots, x_n\}$ of the capacity such that x_j units of capacity is allocated to fare class j, the capacity b_l on any leg is not exceeded, and the expected revenue is maximized. Let $E[sales|x_j, D_j]$ and $f_j(x_j) = r_j \cdot E[sales|x_j, D_j]$ denote the expected sales and revenue, respectively, for demand class j under partition x. We model the cargo capacity allocation problem as follows:

$$\begin{array}{lll} & \max & \sum_{j=1}^{n} r_j \cdot f_j(x_j) \\ (PNLP) & \text{s.t.} & A \cdot x \leq b \\ & & x \geq 0 \end{array}$$

Note that $E[sales|x_j, D_j]$ is the minimum of the demand or the capacity allocation x_j for class j. Hence, we have the following expression for f_j:

$$f_j(x_j) = r_j \cdot E[sales|x_j, D_j] = r_j \cdot \int_0^{x_j} t \cdot \phi(t)dt + r_j \cdot \int_{x_j}^{\infty} x_j \cdot \phi(t)dt.$$

It is easy to show that the PNLP has a concave separable objective function (see for example Ciancimino et al. [15]): $f'_j(x_j) = r_j \cdot (1 - \Phi(x_j))$ and $f''_j(x_j) = -r_j \cdot \phi(x_j) \leq 0 \ \forall j$.

The Lagrangian dual of the PNLP is (we define π as a row vector for exposition purposes):

$$(DPNLP) \quad \min_{\pi \geq 0} \ q(\pi)$$

where

$$q(\pi) = \max_{x \geq 0} L(x, \pi) \tag{1}$$

and

$$L(x, \pi) = \sum_{j=1}^{n} f_j(x_j) + \pi \cdot (b - A \cdot x) \tag{2}$$

From Equations (1) and (2), we have:

$$\begin{aligned} q(\pi) &= \max_{x \geq 0} \sum_{j=1}^{n} f_j(x_j) + \pi \cdot (b - A \cdot x) \\ &= \sum_{j=1}^{n} \max_{x \geq 0} (f_j(x_j) - \pi \cdot A_j \cdot x_j) + \pi \cdot b \end{aligned}$$

We need to find the maximum of each $f_j(x_j) - \pi \cdot A_j \cdot x_j,\ x_j \geq 0$, to be able to state the DPNLP explicitly. Let x^* denote the optimal allocation. By the properties of the Lagrangian multipliers at optimality, we have:

$$f_j'(x_j^*) - \pi \cdot A_j \leq 0 \tag{3}$$

There are two cases to consider for finding x_j^*: (1) If $f_j'(0) = r_j > \pi \cdot A_j$, then (3) is satisfied as equality, i.e., $f_j'(x_j^*) - \pi \cdot A_j = 0$. (2) If $f_j'(0) = r_j \leq \pi \cdot A_j$, then (3) is satisfied at strict inequality and the corresponding capacity allocation is $x_j^* = 0$. In other words, if an incoming booking request for class j has the associated fare (r_j) greater than the sum of the bid prices for the corresponding itinerary $(\pi \cdot A_j)$, then the capacity allocation is strictly positive, and solves (3); otherwise, the capacity allocation is zero, i.e., the booking request is rejected (see Figure 1). This follows exactly the idea behind bid price control. Having the solution of the Lagrangian function, we can now restate the DPNLP:

$$\begin{array}{llll} & \min\limits_{\pi \geq 0} \quad q(\pi) = \pi \cdot b + \sum_{j=1}^{n} [f_j(x_j^*) - x_j^* \cdot f_j'(x_j^*)] & & \\ (DPNLP) & \text{s.t.} \quad f_j'(x_j^*) = \pi \cdot A_j & & \text{if } r_j > \pi \cdot A_j \\ & \qquad\quad x_j^* = 0 & & \text{if } r_j \leq \pi \cdot A_j \end{array}$$

From the DPNLP there is a 1-1 correspondence between the dual and the primal variables. That is, if we have a set of π's, we can calculate the corresponding capacity allocation as follows:

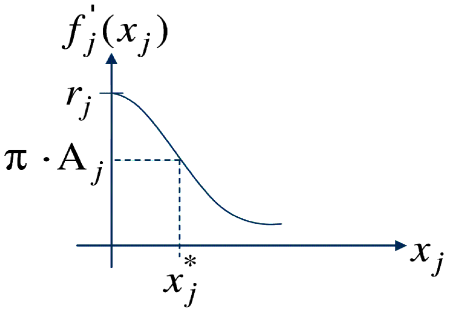

Figure 1: First derivative of jth term in the objective function

(i) $x_j^* = 0$ if $r_j \leq \pi \cdot A_j$.

(ii) $x_j^* = \Phi^{-1}\left(1 - \frac{\pi \cdot A_j}{r_j}\right)$ if $r_j > \pi \cdot A_j$. (Since $f_j'(x_j^*) = r_j \cdot (1 - \Phi(x_j^*)) = \pi \cdot A_j$.)

In the next section we discuss the details of our algorithm to solve the DPNLP.

2.1.3.2 The Newton method applied to DPNLP

DPNLP is a nonlinear programming model, constrained only by the non-negativity of the bid prices π. The Newton method (for both constrained and unconstrained problems) has a very high rate of convergence close to the optimal solution. Newton's method consists of the iteration:

$$x^{k+1} = x^k - \alpha^k \cdot (\nabla^2 f(x^k))^{-1} \cdot \nabla f(x^k) \tag{4}$$

assuming the Newton direction $d^k = -(\nabla^2 f(x^k))^{-1} \cdot \nabla f(x^k)$ is defined and it is a direction of descent (i.e., $d^k \cdot \nabla f(x^k) < 0$). The Newton's method converges very fast when near the optimal solution. However, far from such an optimum the Hessian may be singular or not positive definite. We will show that the DPNLP is convex, hence it has a positive semidefinite Hessian; our algorithm ensures the non-singularity of the

Hessian by maintaining a positive definite submatrix of the Hessian, which is always invertible.

Theorem 1 *The problem DPNLP is a convex problem*

Proof: We prove the convexity of the objective function by showing that the Hessian matrix is positive semidefinite. for calculating the Hessian, we first calculate the gradient. The gradient of the DPNLP is the constraint violation of the PNLP for a given solution x^*:

$\frac{\partial q}{\partial \pi_i} = b_i + \frac{\partial}{\partial \pi_i}[\sum\limits_{j=1}^{n}(f_j(x_j^*) - x_j^* \cdot f_j'(x_j^*))] = b_i + \sum\limits_{j=1}^{n}[f_j'(x_j^*) \cdot \frac{\partial x_j^*}{\partial \pi_i} - x_j^* \cdot \frac{\partial}{\partial \pi_i} f_j'(x_j^*) - f_j'(x_j^*) \cdot \frac{\partial x_j^*}{\partial \pi_i}] = b_i - \sum\limits_{j \in J} x_j^* \cdot \frac{\partial}{\partial \pi_i} f_j'(x_j^*)$, where $J = \{j : r_j > \pi \cdot A_j\}$.

By using $f_j'(x_j^*) = \pi \cdot A_j,\ j \in J$, we deduce:

$$\frac{\partial q}{\partial \pi_i} = b_i - \sum_{j \in J} x_j^* \cdot a_{ij} \tag{5}$$

which is the constraint violation of PNLP for a given solution x_j^*.

The Hessian of the DPNLP is defined as:

$$\frac{\partial^2 q}{\partial \pi_i \partial \pi_k} = -\sum_{j=1}^{n} a_{ij} \cdot \frac{\partial x_j^*}{\partial \pi_k}\ ,\ \forall\ i,k = 1..m \tag{6}$$

For $j \in J$, the following equality holds:

$f_j'(x_j^*) = \pi \cdot A_j$,

and if we take the derivative with respect to π_k, we obtain:

$f_j''(x_j^*) \cdot \frac{\partial x_j^*}{\partial \pi_k} = a_{kj}$.

By plugging in the expression for $f_j''(x_j^*)$, we derive:

$$\frac{\partial x_j^*}{\partial \pi_k} = -\frac{a_{kj}}{r_j \cdot \phi(x_j^*)} \tag{7}$$

and we conclude:

$$\frac{\partial^2 q}{\partial \pi_i \partial \pi_k} = \sum_{j \in J} a_{ij} \cdot \frac{a_{kj}}{r_j \cdot \phi(x_j^*)} \tag{8}$$

The objective function in DPNLP is convex, since all terms in the Hessian are non-negative.

■

We deal with a convex minimization problem, constrained only by non-negativity of π, for which the calculation of the gradient and Hessian is straightforward. However, there are a few difficulties:

- if $\pi \cdot A_j = 0, j \in J$, then the formula for calculating the corresponding capacity allocation $x_j^* = \Phi^{-1}(1 - \frac{\pi \cdot A_j}{r_j})$ would give a very big value. We can avoid this problem by defining an upper bound for the capacity allocation, as in Figure 2.

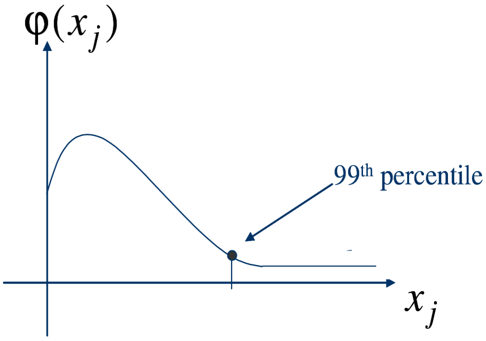

Figure 2: Demand distribution for PNLP

The revenue function $f_j(x_j)$ (see Figure 3) in the PNLP formulation becomes almost flat after a certain capacity is allocated, and it would not significantly improve if more capacity is allocated. Bounding the demand distribution will not significantly interfere with the scope of maximizing the revenue in the PNLP formulation.

- Once we have a set of π's, we can easily identify the set $\overline{J} = \{j : r_j \leq \pi \cdot A_j\}$; the corresponding variables $x_j^* = 0,\ j \in \overline{J}$, the columns $A_j,\ j \in \overline{J}$ can be deleted, since they have no contribution in the Hessian calculation.

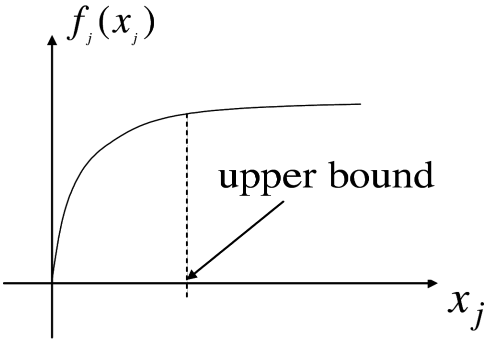

Figure 3: Revenue function for the jth term in the PNLP formulation

2.1.4 Solution methodology

The k-th iteration of the algorithm is:

1. For π^k calculate $x_j^{*^k}$:
   - if $\pi^k \cdot A_j = 0$ then set $x_j^{*^k}$ on the 99-th percentile of the demand distribution;
   - if $\pi^k \cdot A_j \geq r_j$ then set $x_j^{*^k} = 0$; delete column j from matrix A, $A = A_{reduced}$;
   - otherwise set $x_j^{*^k} = \Phi^{-1}(1 - \frac{\pi^k \cdot A_j}{r_j})$.
2. calculate the gradient based on (5);
3. if $||\nabla_{DPNLP}(\pi^k) \leq \epsilon||$ then STOP; else calculate the Hessian for the reduced problem based on (8);
4. find the descent direction via Newton $d^k = -(\nabla^2_{DPNLP}(\pi^k))^{-1} \cdot \nabla_{DPNLP}(\pi^k)$;
5. set $\pi^{k+1} = \pi^k + d^k$;
6. for all i with $\pi_i^{k+1} < 0$, set $\pi_i^{k+1} = 0$;
7. set $k = k + 1$ and go to the next iteration.

We use as an initial solution a set of 1's; however, the initial solution may also be the solution from one of the traditional models, like the deterministic model described before.

2.1.5 Large cargo capacity allocation

We model the large cargo problem as a dynamic program with state vector $S = (b, t)$, where b is the vector of available capacity when the remaining number of periods to departure is t. Let p_j^t denote probability of a request for class j and $p_0^t = 1 - \sum_j p_j^t$ be the probability of no arrival at time t when there are t periods remaining to departure. (We assume that the arrivals across the periods are independent.) We want to compute the maximum expected revenue $TR(b, t)$ at state (b, t), where

$$TR(b,t) = p_0^t \cdot TR(b,t-1) + \sum_j p_j^t \cdot \max(r_j \cdot w_j + TR(b-A_j, t-1), TR(b,t-1)) \ \forall t, b \quad (9)$$

and $TR(b, 0) = 0$ is the boundary condition for the departure day (end of horizon). It is known that the optimal policy is a threshold policy (see for example [53]), in which a booking request for w_j units of capacity is accepted if and only if its corresponding fare satisfies $r_j \cdot w_j > TR(b, t-1) - TR(b - A_j, t-1)$, where the right hand side is the marginal value of w_j units of capacity. The optimal policy can be explicitly included in the optimality equation (9), and since $p_0^t = 1 - \sum_j p_j^t$, we obtain:

$$\begin{aligned}
&TR(b,t) = (1 - \sum_j p_j^t) \cdot TR(b,t-1) + \sum_j p_j^t \cdot \max(r_j \cdot w_j + TR(b - A_j, t-1), \\
&TR(b,t-1)) = TR(b,t-1) + \sum_j p_j^t \cdot \max(r_j \cdot w_j + TR(b - A_j, t-1) - \\
&- TR(b,t-1), TR(b,t-1) - TR(b,t-1)) = TR(b,t-1) + \\
&+ \sum_j p_j^t \cdot \max(r_j \cdot w_j + TR(b - A_j, t-1) - TR(b,t-1), 0) \ \forall t, b
\end{aligned} \quad (10)$$

Given the large number of possible combinations of the remaining capacity on each leg, the exact solution of equation (10) cannot be found in a reasonable time (for 3 legs, each with capacity 7000, there are $7000^3 = 343,000,000,000$ possible combinations). A common approach to tackle this problem in the passenger literature is *dynamic programming decomposition* (see for example [53]), which decomposes the network problem into smaller single-leg problems using *fare proration* - adjusting the contribution of each revenue class at the leg level according to the leg's tightness at the network level. The leg's tightness is defined as the remaining capacity on that leg. For example, if we have a 2-leg itinerary, with capacity 100 and 10 respectively, the fare prorating mechanism will signal that the second leg is much more constrained in terms of capacity by assigning a high revenue for that class. We use the following prorating scheme:

1. Start with the bid price solution π_i from the PNLP model applied to big cargo;
2. for each class j and leg i, calculate the prorated cargo rate as: $\overline{r}_{ij} = \max\{0, r_j - \sum\limits_{k \in \beth(j), k \neq i} \pi_k\}$, where $\beth(j)$ is the set of legs on the requested itinerary for class j;
3. For each leg $i \in \beth(j)$, solve the following DP:
   $TR_i(b_i, t) = TR_i(b_i, t-1) + \sum\limits_j p_j^t \cdot \max(0, r_{ij} \cdot w_j + TR_i(b_i - w_j, t-1) - TR_i(b_i, t-1))\ \forall i \in \beth(j), t$;
4. For each leg $i \in \beth(j)$, calculate
   $BP_i = \sum\limits_{i \in \beth(j)} \Delta TR_i(b_i, t)$, where $\Delta TR_i(b_i, t) = TR_i(b_i, t) - TR_i(b_i - w_j, t)$;
5. The total bid price is approximated as before with the sum of bid prices on the legs requested on the itinerary: $BP = \sum\limits_{i \in \beth(j)} BP_i$.

2.1.6 Numerical results

The numerical experiments aim to answer the following questions about our proposed approach for air cargo capacity allocation:

1. How long does it take to solve real world instances?
2. What is the potential revenue improvement in comparison with current practices?

In the first set of experiments, we generated several problem instances of varying complexity. We assume the airline is servicing L locations out of a single hub, where $L \in \{2, 5, 10, 20\}$. This is a basic network structure of revenue management problems found in literature (see [1]). Each location is connected with two legs, to and out of the hub, such that the total number of resources is $m = 2 \cdot L$.

As described in Section 2.1.2, we define the cargo rates as a function of the billable weight and the shipment category. The total number of rates is the total number of billable weight breaks and the total number of shipment categories. We experiment with the following structure for each category, small and big:

Table 3: Air cargo shipment categories and weight breaks for small cargo

Shipment	Weight breaks		
category	10	20	50
5	Ex1sml	Ex4sml	Ex7sml
10	Ex2sml	Ex5sml	Ex8sml
15	Ex3sml	Ex6sml	Ex9sml

Table 4: Air cargo shipment categories and weight breaks for big cargo

Shipment	Weight breaks		
category	5	10	20
5	Ex1big	Ex3big	Ex5big
10	Ex2big	Ex4big	Ex6big

where Ex$i$sml are the different experiments ran for small cargo, and Ex$i$big for big cargo.

There are more shipment categories for small cargo, because the small packages might contain a higher variety of goods, such as fresh spices, fish, or money. Also, the higher variety of goods for the small cargo is correlated with a higher variety of billable weights. We generate the rates randomly as uniformly distributed between 0.5 and 20 for the small cargo, and 0.1 and 10 for the large cargo, and are descending as a function of the billable weight and shipment category. The number of shipment categories and weight break points influence the number of classes n on the different legs, the highest being 750 (15×50) for the small cargo, and 200 (10×20) for the big cargo. The total number of rate class itinerary combinations for each instance is $n \cdot (2 \cdot L + L \cdot (L - 1))$, and we have instance sizes ranging from 4 legs and 150 rate class itinerary combinations to 40 legs and 315,000 rate class itinerary combinations.

We set the capacity of each leg for small cargo at 200, and for large cargo at $10,000$. The small cargo demand is approximated with a gamma distribution. We generate the mean demand such that the load factor, i.e., the percentage of the total expected demand over available capacity on the network, is 60%, 100% and 160%. For big cargo, we choose the probability of arrival for each rate class itinerary combination such that the bigger sizes come closer to the departure date and the probability of no show is higher at the beginning of the time horizon. We experiment with three different time horizon (τ) lengths, i.e., $\tau \in \{5, 10, 15\}$. Our choices for the time horizon are motivated by the fact that there are no more than 13 big bookings per flight leg, as stated by Raja Kasilingam from Sabre Holdings.

We ran 50 instances for each experiment and averaged the results; the algorithms are coded in Matlab and run on a Mobile Intel(R) Pentium(R), 4 - M CPU 2 GHz, 1 GB of RAM. We report the average CPU time per instance. (The standard deviation or run time is less than 0.03 CPU seconds in all experiments, i.e., the average run time is an excellent representations for the computational performance.)

The results for small cargo are summarized in Table 5. The running time is primarily influenced by the number of demand classes in the instance, whereas the impact of the load factor is minimal. While the number of iterations is small across all instances (ranging from 4 to 8, indicating rapid convergence), the run time ranges from 3 seconds to almost 5 hours. We expect that the run time can be drastically reduced if the algorithm is implemented in a compiled language (such as C or Fortran).

The computational results for big cargo experiments are reported in Table 6. The running time is again mainly influenced by the number of classes in each instance; however, the largest instance is solved in under 2 minutes. The results of this first set of experiments for small and big cargo capacity allocation show that the proposed algorithms are computationally tractable and robust, hence, they can be used in practice.

Having shown the efficiency of the algorithms, next we turn to evaluate their effectiveness, i.e., the quality of the solutions. For this purpose, we develop a simulation of the cargo booking process, separately for small and big cargo.

For modeling small cargo booking quantities, we use a Non-Stationary Compound Poisson Process (NSCPP) (similar to the one in [22]), widely cited in the passenger literature (see for example [53]). NSCPP allows for batch arrivals (compound) and time-dependent arrival rates (non-stationary). In our setting, we have one arrival per unit time and the batch corresponds to the size of the booking, which varies from 0.1 kg to 200 kg. We model the small cargo booking arrival process for an itinerary h as a compound Poisson process with arrival rate $\lambda_h(t)$, where t denotes the reading interval. The reading interval is defined as the moment in time when the booking policy is updated - it can be every few days, every day (usual towards the end of the booking horizon), or after every booking, depending on each airline's choice. There is no widely accepted policy for how often to update bid prices. Therefore, we consider several alternative strategies in our simulation.

Table 5: CPU time for solving the small cargo problem with different load factors (LF)

	Number of classes	**Avg. CPU time in sec.**		
		LF 60 %	**LF 100 %**	**LF 160 %**
2 Locations				
Ex 1	300	3.05	4.75	5.5
Ex 2	600	7.34	8.74	10.64
Ex 3	900	10.68	11.6	13.08
Ex 4	600	7.33	8.73	10.63
Ex 5	1200	15.51	17.93	18.99
Ex 6	1800	29.82	31.93	33.04
Ex 7	1500	20.35	22.35	24.05
Ex 8	3000	40.44	42.74	44.78
Ex 9	4500	70.02	73.2	75.02
5 Locations				
Ex 1	1500	20.34	22.37	24.07
Ex 2	3000	40.46	42.76	44.76
Ex 3	4500	70.01	73.1	75.91
Ex 4	3000	40.44	42.74	44.74
Ex 5	6000	81.73	84.3	87.92
Ex 6	9000	147.01	149.12	151.2
Ex 7	7500	122.7	125.27	129.7
Ex 8	15000	200.09	206.00	216.01
Ex 9	22500	346.97	350.7	360.07
10 Locations				
Ex 1	5500	79.09	81.74	84.4
Ex 2	11000	165.47	171.53	179.3
Ex 3	16500	220.24	226.4	231.04
Ex 4	11000	165.57	171.57	171.57
Ex 5	22000	336.7	341.07	349.7
Ex 6	33000	508.4	515.02	523.2
Ex 7	27500	480.3	488.93	497.3
Ex 8	55000	1130.42	1139.43	1145.27
Ex 9	82500	2155.76	2163.6	2171.86
20 Locations				
Ex 1	21000	306.39	312.53	321.3
Ex 2	42000	939.22	946.84	957.4
Ex 3	63000	1756.24	1767.54	1779.63
Ex 4	42000	939.21	946.81	957.1
Ex 5	84000	2528.44	2537.84	2549.73
Ex 6	126000	4632.04	4640.67	4651.74
Ex 7	105000	3454.73	3469.3	3477.13
Ex 8	210000	10089.02	10099.2	10112.01
Ex 9	315000	16110.23	16121.3	16135.23

Table 6: CPU time for solving the big cargo problem for different time horizons (TH)

	Number of classes	**Avg. CPU time in sec.**		
		TH=5	TH=10	TH=15
2 Locations				
Ex 1	150	2.25	2.44	3.01
Ex 2	300	3.01	3.42	4.01
Ex 3	300	3.01	3.5	4.2
Ex 4	600	3.77	4.01	4.4
Ex 5	600	3.8	4.01	4.8
Ex 6	900	4.01	6.00	7.98
5 Locations				
Ex 1	750	3.85	4.05	5.00
Ex 2	1500	4.23	6.92	8.87
Ex 3	1500	4.18	6.98	8.9
Ex 4	3000	4.67	7.55	10.15
Ex 5	3000	4.67	7.55	10.01
Ex 6	6000	7.69	10.33	13.11
10 Locations				
Ex 1	2750	4.42	7.32	9.87
Ex 2	5500	7.44	9.78	12.01
Ex 3	5500	7.45	9.87	12.01
Ex 4	11000	10.52	15.22	19.26
Ex 5	11000	10.52	15.23	19.27
Ex 6	22000	20.01	26.13	30.01
20 Locations				
Ex 1	10500	9.16	14.88	17.38
Ex 2	21000	15.23	25.01	28.99
Ex 3	21000	15.23	25.01	28.99
Ex 4	42000	30.4	48.24	55.01
Ex 5	42000	30.32	48.12	54.89
Ex 6	84000	55.7	64.03	89.99

In every reading interval, which are simulated sequentially, the union of all arrival processes defines a compound process with arrival rate $\lambda_{total}(t) = \sum\limits_{h} \lambda_h(t)$. We use a random discrete variable with a probability of realization p of $\frac{\lambda_p(t)}{\lambda_{total}(t)}$ to assign types of arrivals (weight and shipment category). For each type of arrival, a size is assigned. The sizes are assigned according to an exponential distribution with mean $\mu_p(t)$, which is the mean of the weight category associated with the arrival type. It is common knowledge (see for example [21]) that a Gamma distribution $\Gamma(\alpha, \beta)$ (α integer) can be approximated by a sum of α Exponential distributions of mean $\mu = \beta$, so that the demand of the small cargo is approximated with the traditional passenger demand distribution.

The modeling of the big cargo booking process is similar to that of the small cargo, with the main difference being the sizes of the packages. The sizes of big cargo bookings vary from 200 kilograms to 6,000 kilograms, with the bigger loads more probable to arrive towards the end of the horizon. We experiment with different distributions for the big cargo demand: Binomial, Negative Binomial, and Gamma; each of these distributions can be seen as the result of summing random variables distributed Bernoulli, Geometric, or Exponential, respectively. As in the case of small cargo, we simulate reading intervals sequentially; however, given the relatively sparse arrivals of big cargo, in this case the reading interval is defined by each booking.

In both simulations, during each reading interval we decide whether to accept or reject the new booking based on the bid prices deduced according to each corresponding algorithm (PNLP for small cargo, dynamic programming for big cargo), and if accepted, we increment the total revenue by the value of the load, and decrease the available capacity by its size.

We compare the total revenue obtained from our proposed approach with two other approaches used in practice (using the same simulated arrivals): (1) First Come First Served (FCFS) policy, where the capacity is filled with incoming bookings until the limit is reached. (2) Obtaining bid prices by solving the deterministic model mentioned in Section 2.1.3 as an integer program. For each request we solve the model twice: first assuming the request is accepted, and then assuming the request is rejected. If the difference in the objective function is below the rate associated with the incoming booking, then the shipment is accepted. We call this approach the Deterministic Integer Program (DIP).

The simulations for the small cargo are run for example 8 in Table 4, for 20 locations and load factors of $60\%, 100\%$, and 160%. The booking horizon for small cargo spans 30 days, and we define 15 reading intervals following the practice from Sabre (see [41]). We use three approaches for refreshing the bid prices (re-optimization):

1. calculate at the beginning of the booking horizon, refresh last three days every day;

2. calculate at the beginning of the booking horizon, refresh last ten days every day;

3. refresh every day.

For big cargo, in our simulation we use the biggest instances, i.e., example 6 in Table 4, with 20 locations, the same load factors as for small cargo, and three different demand distributions: Binomial, Negative Binomial, Gamma.

The average revenue and standard deviation over 50 runs for the small and big cargo problems are reported in Tables 7 and 8, respectively. In Table 7, we take the revenue resulting from the FCFS policy as the base case. For DIP, we report the percentage improvement from using DIP over FCFS in parentheses. Similarly, for DPNLP, we report the percentage improvement from using DPNLP over FCFS or DIP. Both DIP and DPNLP result in significantly higher revenues than FCFS. The revenues resulting from DPNLP and DIP are very close, and we see a slight improvement from using DPNLP versus DIP when we refresh the bid prices every day. The bid price refreshing frequency has a bigger impact on the resulting revenue if we re-optimize every day for the last 10 days, but it does not result in a significant improvement if we refresh every day. Also, the higher the load factor, the higher the gain from using DIP or DPNLP over FCFS. Similar behaviors have also been observed in the passenger literature (see for example [49]).

For the big cargo, the revenue gain from applying the pro-rated dynamic program (DP) described in Section 2.1.5 over the other two methods is significant (see Table 8). The revenue gain from using DP versus FCFS ranges from 21% to 59%, depending on the load factor and the demand distribution. The influence of the load factor on the revenue gain is not as significant as in the case of the small cargo, but we observe a similar trend: the higher the load factor, the higher the gain from using revenue management. The revenue gain from using DP versus DIP ranges from 10% to over 18%, which is again a significant improvement. The highest revenue gains from using DP or DIP versus FCFS are observed for the Gamma and Negative Binomial distributions, which suggests that the gains from revenue management are more significant if the demand is lumpier.

Table 7: Average revenue and standard deviation using different methods for small cargo

Load Factor	Method	Bid price refreshing frequency					
		Last 3 days		Last 10 days		Every day	
		Avg. Rev.	**StDev.**	**Avg. Rev.**	**StDev.**	**Avg. Rev.**	**StDev.**
60%	**FCFS**	6471.75	770.99	5513.45	22.55	5846.78	779.4
	DIP	7341 (+13.43%)	253.88	7557.71(+37%)	405.60	7554.37 (+29.2%)	65.17
	DPNLP	7329.247 (+13.24%, -0.16%)	394.89	7554.54 (+37%, -0.001%)	396.65	7554.87 (+29.21%, +0.01%)	37.22
100%	**FCFS**	7705.09	1299.68	6513.45	22.56	6180.12	579.51
	DIP	9174.34 (+19.07%)	314.95	9611.04 (+47.56%)	219.88	9201.04 (+48.88%)	69.08
	DPNLP	9160.24 (+18.89%, -0.15%)	136.39	9609.54 (+47.53%, -0.02%)	396.66	9252.87 (+49.72%, +0.02%)	37.22
160%	**FCFS**	8038.42	776.78	7846.78	576.07	7513.45	22.56
	DIP	10174.34 (+26.47%)	314.95	12377.71 (+57.74%)	184.42	12344.37 (+64.29%)	516.12
	DPNLP	10155.91 (+26.34%, -0.18%)	85.82	12371.21 (+57.66%, -0.05%)	63.66	12347.54 (+64.34% +0.03%)	584.28

Table 8: Average revenue and standard deviation using different methods for big cargo

Load Factor	**Method**	**Demand distributions**					
		Binomial		**Gamma**		**Negative Binomial**	
		Avg. Rev.	**StDev.**	**Avg. Rev.**	**StDev.**	**Avg. Rev.**	**StDev.**
60%	**FCFS**	78666.52	10896.93	69931.86	10151.36	68401.61	9358.534
	DIP	96274.05 (+20.4%)	23269.6	87424.75 (+25%)	6175.534	89740.54(+31.2%)	18573.93
	DP	108321.51 (+37.7%, +12.5%)	17727.14	103395.4 (+47.85%, +18.27%)	6731.806	99537.42 (+45.5%) (+10.92%)	8248.95
100%	**FCFS**	122691.5	21771.3	109931.9	18215.79	95068.27	18235.72
	DIP	149524.1 (+21.9%)	27959.62	140758.1 (+28%)	13399.11	129740.5 (+36.47%)	18353.09
	DP	169821.5 (+38.42%, +13.57%)	17918.39	166728.7 (+51.66%, +18.45%)	1174.443	146204.1 (+53.79%, +12.69%)	9387.971
160%	**FCFS**	131916.5	29090.93	113265.2	12468.06	108401.6	18746.69
	DIP	162074.1(+22.9%)	37668.3	147424.8(+30.15%)	6175.534	149740.5 (+38.13%)	28348.46
	DP	183821.5 (+39.34%, +13%)	23708.32	176728.7 (+56%, +19.9%)	1174.443	172870.8 (+59.47%, +15.45%)	13392.5

2.2 Air cargo show-up rate estimation

The second air cargo revenue management problem solved in this thesis deals with the uncertainty related to tendering cargo at departure. The airlines do not know how much capacity they have available for free sale until departure. Freight forwarders intentionally bid on more capacity than they actually need to ensure space on constrained flights, since most airlines allow them to return unwanted space at no extra charge. The airlines add the released space to the pool of capacity available for free sale. They typically do not know how much allotted capacity will be unused in advance of the flight departure. In addition, for planes carrying cargo and passengers (combination carriers), the cargo space usually contains both, passengers' baggage and cargo in the same compartment. These factors plus weather (which affects the amount of fuel onboard the aircraft), and mail influence how much capacity is available for free sale (Figure 4). Finally, the cargo space is constrained by two dimensions, weight and volume, and the airline typically does not know which dimension is the most restrictive prior to departure.

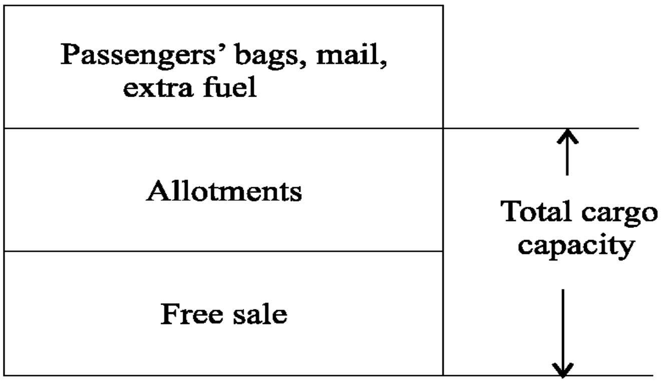

Figure 4: The combination passenger/cargo aircraft's cargo capacity

The booking processes for cargo and passengers are different. The time window during which the airline offers cargo capacity for free sale is shorter than that for passenger capacity; usually no longer than 30 days before departure. Cargo bookings, varying widely in size and volume, come from a relatively small number of customers. A booking may be canceled, rebooked to a different flight, canceled again, rebooked back to the original flight, several times until departure, since airlines typically do not charge for changing reservations.

To hedge against the variability in the amount of cargo actually handed in at departure (cargo tendered) and customers' cancellations, airlines commonly overbook their capacity. Air-cargo overbooking refers to the airlines' practices to sell more capacity than physically available to compensate for cargo that does not show up at departure. Two important considerations in overbooking are *spoilage* (demand turned away because the overbooking level was too low, leaving excess capacity at departure) and *off-loads* (booked demand that the airline cannot accommodate at departure because the overbooking level was too high). Airlines base their decisions on predictions of the *show-up rate*, the percentage of the demand booked that shows up at departure.

In the passenger sector, the common practice is to formulate the overbooking problem as a newsvendor problem, with the overbooking level selected to minimize the total expected costs of spoilage and off-loads (see e.g., [44]). Many airlines use the Normal distribution to model the cargo show-up rate, which is a good approximation for passengers (see [5]). We show that the Normal distribution is usually not a good fit for estimating the cargo show-up rate, and it can result in high lost revenue.

2.2.1 The cargo-booking process

The calculation of the overbooking levels is based on show-up rate estimates. *The cargo weight or volume show-up rate* is the percentage of cargo weight or volume that shows up at departure out of the total weight or volume of cargo booked at each *reading day* (RD). For cargo, the booking time window has 30 reading days, which are numbered backwards in time, from 0 (the departure day) to 30. For example, the show-up rate (in percentage) on reading day 21 before departure date x is the amount tendered at departure day x (in kilograms) out of the amount booked on reading day 21 (in kilograms), multiplied by 100.

Following the practice they use for the passenger business, most airlines estimate the cargo show-up rate separately for weight and volume as a normally distributed random variable at the flight-leg level (flight number and origin and destination airports). They feed the estimates to the overbooking module, which sets the level of capacity authorized for sale. In each reading day, the airline accepts demand if capacity remains after subtracting the *current bookings* (accepted bookings that have not been canceled) from the *authorized capacity* (the capacity available for free sale multiplied by the overbooking level). The show-up rate changes from reading day to reading day, and the airlines have to make sure they capture these changes and use correct levels for overbooking when selling cargo space. Usually, they do not monitor the booking process over the entire booking period, but rather only on specific reading days, which they consider as significant based on historical booking activity.

The cargo business is a relatively new candidate for providing additional revenue for the airline industry, and there is still a lot of manual handling involved; the orders come in through different channels (agencies, internet, freight forwarders), and most of them are not properly captured in the airlines' systems. Because of the nature of the data the airline collected, we focused on estimating the weight show-up rate only, having no data available for volume. Luo, Cakanyildirim, and Kasilingam ([28]) justified the use of a common distribution for the show-up rate for weight and volume by conducting statistical tests on real-world data.

We re-evaluate the level of the authorized capacity over the following pre-specified 15 reading days: RD30, RD28, RD21, RD14, RD10, RD9, ..., RD1, and RD0.

2.2.2 The data

For seven combination passenger and cargo flights, we have show-up-rate data for a 16-month period. For each flight number and departure date, we have 15 show-up rates, corresponding to each reading day. We used the first 12 months of the data to estimate the show-up-rate distribution, two months of data to forecast the fitted distribution, and the last two months to validate and compare our results with the normal distribution. Hence, we used 80 percent of the data for training and the remaining 20 percent for testing, a common practice among marketing and neural networks researchers.

2.2.3 Nonparametric distribution estimation and forecasting

For the flights we analyzed, we observed that the distribution of the show-up rate follows different shapes and skewness. We fitted the following continuous parametric distributions to all analyzed flights in the study for all reading days for one year of historical departures: normal, gamma, beta, weibull, lognormal, and exponential. We used three goodness-of-fit tests for each distribution: Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling. The hypothesis that the sample was from any of the specified distributions was rejected in approximately 90 percent of all cases.

These results motivated us to fit a nonparametric distribution to the show-up rate. If the probability distribution function were from a known parametric family (for example, Gaussian), we would have to estimate the finite-dimensional parameters which characterize that particular distribution (for example, the mean and the variance). Without the parametric assumption, the problem is known in statistics as the nonparametric estimation problem.

One of the easiest nonparametric estimators is the histogram, which is obtained by dividing the data range into equal intervals (bins) and counting the number of observations that fall into each interval (Figure 5). Each bin is represented by the

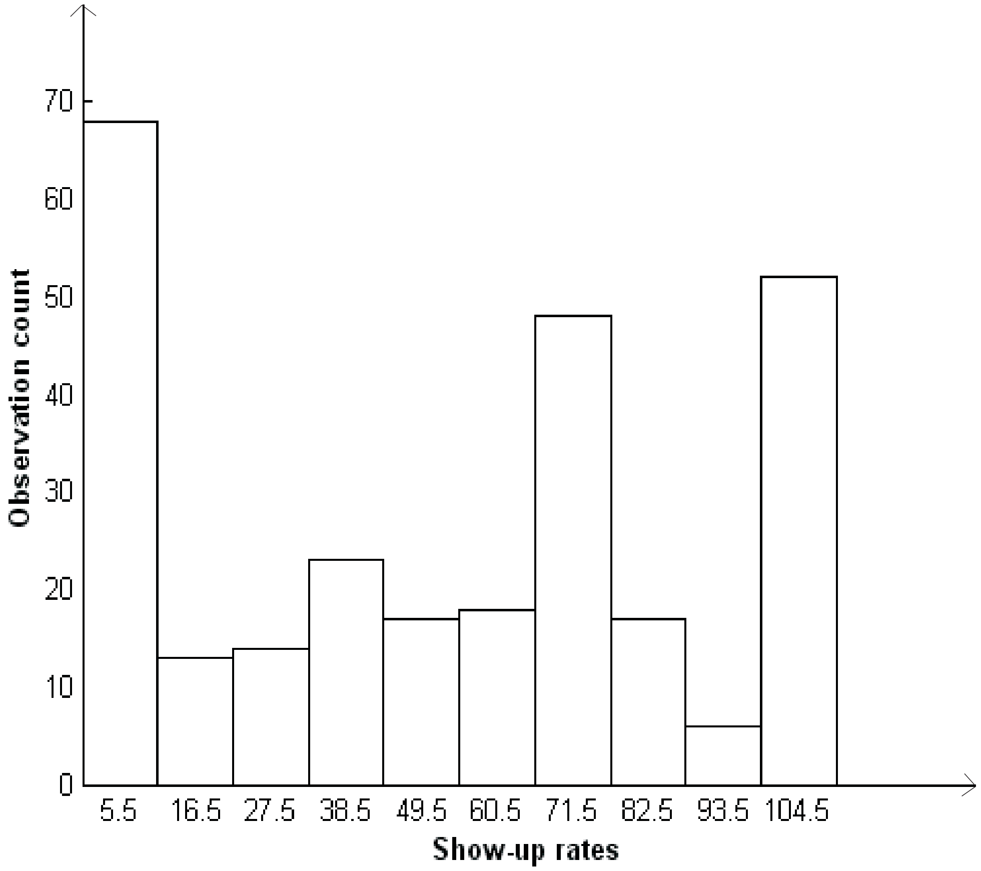

Figure 5: Histogram estimator

midpoint of the corresponding interval, with probability equal to the ratio of the number of observations it contains and the total number of observations. Figure 5 represents the show-up rates on reading day 6 for a period of one year for a given flight number, that is, the percentage of cargo tendered at departure for approximately 300 departure days out of the cargo booked six days before each departure. For cargo, the show-up rate can be higher than 100 percent, because of overtendering (handing in more cargo than booked).

The effectiveness of the histogram estimator depends on the number and the size of the bins. In general, the higher the number of bins, the higher the probability of capturing noise rather than the characteristics of data. On the other hand, a too low number of bins will represent the data poorly. We first developed an estimator with equal-size bins. To choose the optimal number of bins to minimize the error of mapping the data into the class intervals (bins), we implemented Birge and Rozenholc's method ([25]).

2.2.3.1 The histogram estimator with equally sized bins

The procedure proposed by Birge and Rozenholc ([25]) is limited for probability density functions with support on [0,1]. It is not a restrictive assumption. In our case we can replace the support by the data range, since there is no information available of what happens outside the range. The data however has to be normalized.

We use the following notation:

$X_1, \ldots, X_n$	is the data sample
D	the number of bins
$I_1, \ldots, I_D$	the equally sized intervals
f	is the unknown underlying probability density
$\overline{f}$	is the histogram estimator

The histogram estimator of f based on the regular partition with D bins, i.e., the partition of [0,1] consisting of D intervals $I_1, \ldots, I_D$ of equal length $1/D$ is given by:

$$\overline{f}_D = \overline{f}_D(X_1, \ldots, X_n) = \frac{D}{n} \cdot \sum_{j=1}^{D} N_j \cdot 1_{I_j} \text{ with } N_j = \sum_{i=1}^{n} 1_{I_j}(X_i) \qquad (11)$$

where $1_{I_j}(X_i)$ is the indicator function having a value of 1 if the data point X_i belongs to the interval I_j.

In order to measure the quality of such an estimator, a loss function l has to be chosen to compute its risk:

$$R_n(f,\overline{f},l) = E_f[l(f,\overline{f}_D(X_1,\ldots,X_n))] \tag{12}$$

The optimal value D^{opt} is given by $R_n(f,\overline{f}_{D^{opt}},l) = \inf_{D\geq 1} R_n(f,\overline{f}_D,l)$. Unfortunately, D^{opt} cannot be exactly computed because it depends on the unknown density f. The risk function is usually asymptotically evaluated.

The loss function l used by the authors is the squared Hellinger distance:
$h^2(f,g) = \frac{1}{2}\int_0^1 \left(\sqrt{f(y)} - \sqrt{g(y)}\right)^2$

Their choice is based on the fact that the Hellinger distance is the natural distance to use in connection with the maximum likelihood estimation and related procedures. The authors arrive at the following optimization problem from which the number of bins D can be determined:

$$\begin{array}{ll} \text{maximize} & L_n(D) - pen(D) \\ \text{s.t} & 1 \leq D \leq \frac{n}{\log n} \end{array}$$

where

$L_n(D) = \sum_{j=1}^{D} N_j \cdot \log(\frac{DN_j}{n})$ with N_j defined in (11), is the log-likelihood of the histogram with D bins

and $pen(D) = D - 1 + (\log D)^{2.5}$ is the penalty function for choosing D bins.

The variable is the number of bins D. The upper bound $\frac{n}{\log n}$ is chosen based on Castellan's [11] work. The bound is connected with the asymptotic evaluation of the risk function given in (12), when the Hellinger loss function is used.

The objective function is derived based on Castellan's [11] work. She has shown that a suitably penalized maximum likelihood estimator provides a data-driven method for selecting the number of bins, which results in an optimal value of the Hellinger risk. The objective function is such a penalized maximum likelihood estimator. The penalty function is proposed by the authors that have based their choice on intensive simulation studies. They compared their method with a number of existing methods. They show in their paper [11] that the method proposed by them outperforms all the other methods.

The show-up rate estimation algorithm that we implemented used the optimal number of bins D from the optimization problem mentioned above. The input is defined by the vector containing historical values for the show-up rate for a certain reading day. The algorithm goes through the following steps:

1. Normalize data
2. for $D = 1$ to $\frac{n}{\log n}$ do
    - divide [0,1] in D equal intervals
    - count data points from the normalized data that fall into each interval $\Longrightarrow N_j$
    - calculate $L_n(D) = \sum_{j=1}^{D} N_j \cdot \log(\frac{DN_j}{n})$
    - calculate $pen(D) = D - 1 + (\log D)^{2.5}$
    - calculate $Eval(D) = L_n(D) - pen(D)$
3. set D as the value for which $Eval(D)$ is minim
4. divide the range of data in D intervals
5. construct bins starting at the minimum data value by successively adding (maximum data value - minimum data value)/D
6. construct probability $p_j = \frac{N_j}{n}$ for each bin j
7. construct mid-points for each bin = (bin edge right - bin edge left)/2

The output of the algorithm is the show-up rate probability density distribution per reading day, consisting in mid-points for each bin and associated probability.

This equal-size-bin estimator performed slightly better than the normal estimator in terms of mean off-loads and spoilage. The main disadvantage was that it used equal-size bins, which led to bins containing no points for some of the reading days, where the show-up rates where clustered together around some values.

To capture the data distribution more accurately, we next used a histogram estimator with variable size bins, which uses the equal-size-bin estimator as a starting point.

2.2.3.2 The histogram estimator with varying size bins

Wavelet methods have been applied successfully to density estimation ([4]), because of their ability to filter out noise. Generally speaking, a wavelet basis is a collection of functions obtained as translations and dilations (shift and scale) of a *scaling function* ϕ and a *mother wavelet* ψ. Once the mother wavelet ψ is fixed, dilations and translations of the function ψ, $\psi_{jk}(x) = const \cdot \psi(2^j x - k)$, define an orthogonal basis in $L^2(R)$ (space of integrable functions) together with the scaling function ϕ; that is, any element of the space can be represented as a linear combination of the basis functions. Chui ([9]) provides a general exposition of the wavelet theory.

We chose ψ as the simplest of wavelets, the Haar wavelet, which is a step function taking values 1 and -1 on $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1)$. The scaling function for the Haar wavelet is the unity function on the interval $[0, 1)$: $\phi(x) = \mathbf{1}(0 \leq x < 1)$.

In general, for a data vector $y = [y_0, y_1, ..., y_{2^n-1}]$ of length 2^n associated with a piecewise constant function f on $[0, 1]$, the wavelet decomposition of f has the form

$f(x) = c_{00}\phi(x) + \sum_{j=0}^{n-1} \sum_{k=0}^{2^j-1} d_{jk}\psi_{jk}(x)$

with c_{00} and d_{jk} being the wavelet coefficients.

We chose the function f to be the observation count associated with the bins calculated by Birge and Rozenholc ([25]); if the number of bins from the procedure is not a dyadic (power of two) number, we set it to the closest higher dyadic number.

We used a quadratic variance-stabilizing transformation of the observation count per bin to improve the performance of the wavelet estimator ([4]).

We use the following notation:

N	number of data points.
$X_1, \ldots, X_N$	data sample.
D	number of bins calculated based on Birge and Rozenholc (2002) and adjusted to the closest higher dyadic number.
$ob = [ob_1, ..., ob_D]$	observation count per bin.

The steps of the procedure are as follows:

(1) Determine D; if $D = 2^n + c \leq 2^{n+1}$, with $c > 0$, set $D = 2^{n+1}$.

(2) Apply the following variance-stabilizing transformation to the bin count: $2 \cdot \sqrt{ob_i + \frac{3}{8}}$.

(3) Decompose the transformed observation count ob via forward wavelet transform.

(4) Threshold the wavelet coefficients to filter out noise.

(5) Recover the denoised signal $\underline{ob}$ via inverse wavelet transform.

(6) Calculate midpoints and probabilities based on $\underline{ob}$.

For Step 1, see Birge and Rozenholc ([25]); the method is fairly straightforward to implement. Step 3 and 5 refer to the Haar wavelet transform; most statistical packages have it already implemented.

Step 4 is the procedure used for denoising the original signal. The wavelet coefficients correspond to the details of the signal. The method considers the small details to be noise and deletes or smoothes them out without substantially affecting the main features of the original signal. The two types of thresholding are hard and soft. Hard thresholding is the usual process of setting to zero the elements whose absolute values are lower than the threshold. Soft thresholding is an extension of hard thresholding, first setting to zero the elements whose absolute values are lower than the threshold, and then shrinking the nonzero coefficients towards 0. In step 4, soft thresholding gave us better results, and we used it in the simulations.

For the threshold value, we had several choices, among them, the universal threshold, and the cross-validatory threshold. We chose the universal threshold value, that is, $\lambda_{UNIV} = \sqrt{2 \cdot \ln(D)} \cdot \sigma$, with σ^2 the noise variance estimated from the coefficients' standard deviation. The universal threshold is useful for obtaining a starting value when nothing is known of the signal condition.

We assumed a nonwhite noise in our signal (noise not having continuous and uniform frequency spectrum over a specified frequency band). As a consequence, we had to rescale thresholds using a level-dependent (within the wavelet decomposition) estimation of the level noise ([9]).

The denoised signal $\underline{ob}$ is of the form: $[ob_1, ob_1, ob_1, ob_2, ob_3, ob_3, ob_3, ..., ob_t]$. We calculated the new bins by clustering together adjacent bins (in the initial histogram) of equal observation count in the denoised signal $\underline{ob}$.

Because of such factors as seasonality, changes in demand patterns, and competition, the show-up rate and its underlying distribution may change over time. To update the show-up rate distribution, we use Murty's method ([23]), which is an extension of the exponential smoothing forecasting technique.

2.2.3.3 Updating the fitted distribution

In addition to the previous notation, we use the following:

$m_1, \ldots, m_t$	midpoints of the fitted bins.
$p_1, \ldots, p_t$	probability vector associated with the bins.
$y_1, \ldots, y_t$	probability vector for recent observations.
$x_1, \ldots, x_t$	updated probability vector
k	number of new observations.

The number of bins (t) remains unchanged throughout the process. We constructed the probability vector $y_1, \ldots, y_t$ by counting how many new observations fall into each bin and dividing this number by the total number of observations, k.

We used the weighted least squares to compute x from p and y. The optimization problem is

$$\begin{aligned} &min \quad \beta \cdot \sum_{i=1}^{t} (p_i - x_i)^2 + (1-\beta) \sum_{i=1}^{t} (y_i - x_i)^2 \\ &\text{s.t.} \qquad\qquad \sum_{i=1}^{t} x_i = 1 \\ &\qquad\qquad\qquad x_i \geq 0 \qquad\qquad \forall i = 1 \ldots t \end{aligned} \tag{13}$$

where β is a weight between 0 and 1. The quadratic optimization model minimizes the weighted sum of the squared forecasted errors over all value intervals, and when used periodically (every two months in our case), it has the effect of tracking gradual changes in the probability distribution of the random variable.

The unique optimum of the convex quadratic problem (13) is

$$x = \beta \cdot p + (1 - \beta) \cdot y \tag{14}$$

When the mean value of the random variable changes substantially, we refit the distribution before updating. Such situations could be encountered during high demand periods, such as Christmas, or when a lot of flights must be cancelled due to bad weather.

The method has two parameters that need to be carefully analyzed: β, the smoothing factor, and k, the number of new observations. We chose $\beta = 0.8$, based on the expert knowledge Sabre provided. Murty ([23]) also recommends a value of 0.8 or 0.9 for β: the influence of the second term in (14) should be small, since the vector y is based on only a small number of observations.

Murty ([23]) argues that one should use at least 50 observations to update the distribution. We used data for two months, which corresponds to approximately 60 observations.

2.2.4 Comparing the discrete distribution and the normal distribution

Using percentiles, we compared the proposed discrete distribution and the normal distribution used by the airline. Percentiles are position measures, describing where a specific data value falls within the data set or the distribution range. We computed nine percentiles (10, 20, ..., to 90), using the statistical tool SAS. While we could compute the percentiles directly for continuous distributions (for example, Normal), for the discrete distribution we used the value of the midpoint of each interval, for which the cumulative distribution function is closest to the considered percentile.

The mean absolute error for the discrete distribution ($MAE_{discrete}$) is consistently between 10 and 50 percent lower than the mean absolute error for the normal distribution (MAE_{normal}) for each reading day (Figure 6). The results encouraged us to proceed with studying the impact of the new estimator on overbooking.

The overbooking model some major airlines use is a newsvendor problem ([26]) with service level constraints and upper and lower bounds for the authorized capacity. The service level, or failure rate, is defined as the ratio between the expected value of the off-loads and the expected value of the show-ups. The airlines impose the failure rate constraint to discourage too high overbooking levels, so that they can meet the service levels promised to the customers.

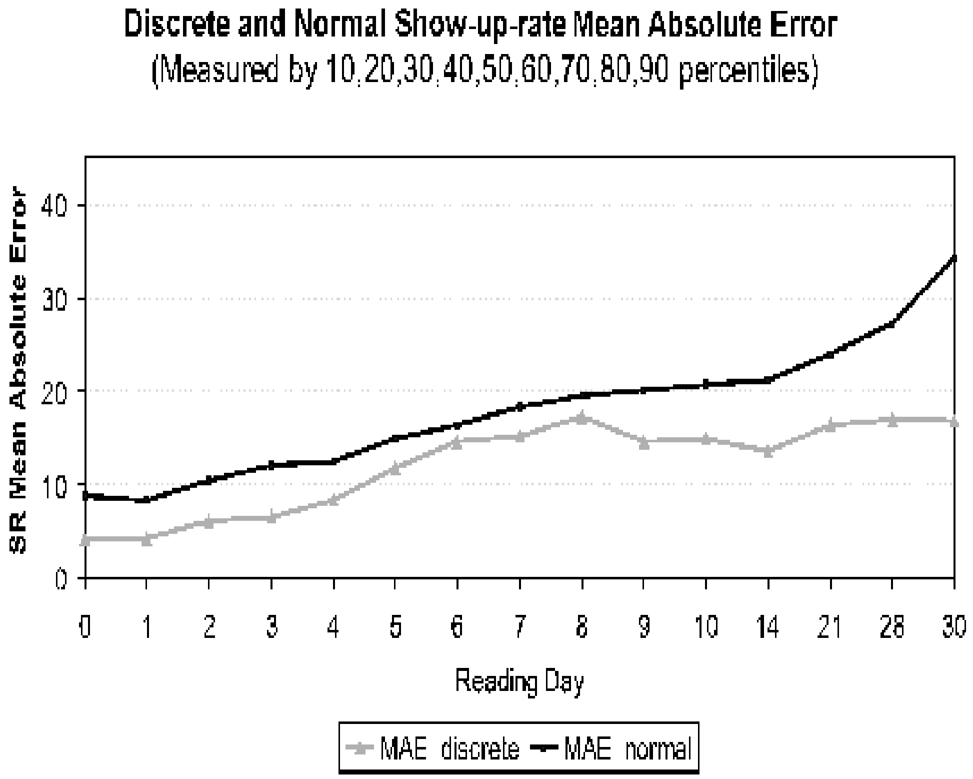

Figure 6: The discrete and the Normal show-up-rate mean absolute error for each reading day

2.2.4.1 The overbooking model

Most airlines do not address the issue of multidimensionality when adapting passenger models to cargo. Usually, they run the overbooking model for weight and for volume separately. We followed the same scheme and adapted the existing newsvendor-like overbooking model to the newly estimated (weight) show-up rate. New approaches to cargo overbooking are described in Luo and Cakanyildirim ([48])

We use the following notation:

SR	discrete random variable for the show-up rate.
$f_{SR}(x) = P(SR = x)$	probability mass function of the show-up rate.
v	authorized capacity.
$SU = SR \cdot v$	random variable corresponding to the show-ups.
$f_{SU}(u) = P(SU = u)$	probability mass function of the show-ups.
$SP = max\{0, c - SU\}$	random variable corresponding to spoilage.
$OF = max\{0, SU - c\}$	random variable corresponding to off-loads.

The airlines usually define the show-ups as the authorized capacity multiplied by the show-up rate. This definition is accurate if booking requests exceed the authorized capacity. When the booking requests are below the authorized capacity, the show-ups should be equal to the booking requests multiplied by the show-up rate, that is, the show-ups should be $SR \cdot \min\{v, B\}$, where B represents the booking requests.

However, the unavailability of data in practice force airlines to use $B \to +\infty$, that is, to use $SU = SR \cdot v$. Luo and Cakanyildirim ([48]) show that the two representations of the show-ups result in the same optimal solution for a one-dimensional model.

The known parameters are

c physical capacity,

c_s, c_o cost per unit spoilage and off-load,

r admissible service level, and

v_l, v_u lower and upper bound on the authorized capacity v.

Based on the definition of show-ups, we deduced its probability density function f_{SU}:

$f_{SU}(u) = P(SU = u) = P(SR \cdot v = u) = P(SR = \frac{u}{v}) = f_{SR}(\frac{u}{v})$.

The expected spoilage can be calculated as

$$E[SP] = E[max\{0, c - SU\}]$$
$$= \sum_{u=0}^{c}(c-u)\cdot f_{SU}(u) = \sum_{u=0}^{c}(c-u)\cdot f_{SR}(\tfrac{u}{v}) = \sum_{x=0}^{\frac{c}{v}}(c - x\cdot v)\cdot f_{SR}(x).$$

Similarly, we calculated the expected off-loads as

$$E[OF] = E[max\{0, SU - c\}]$$
$$= \sum_{u=c}^{+\infty}(u-c)\cdot f_{SU}(u) = \sum_{u=c}^{+\infty}(u-c)\cdot f_{SR}(\tfrac{u}{v}) = \sum_{x=\frac{c}{v}}^{+\infty}(x\cdot v - c)\cdot f_{SR}(x).$$

We deduced the expression for the expected total cost as a function of the authorized capacity v:

$$E[TC] = E[C_{SP}] + E[C_{SP}] = c_s \cdot \sum_{x=0}^{\frac{c}{v}}(c - x\cdot v)\cdot f_{SR}(x) + c_o \cdot \sum_{x=\frac{c}{v}}^{+\infty}(x\cdot v - c)\cdot f_{SR}(x).$$

We aimed to minimize the expected total cost as a function of the authorized capacity v under service level and upper and lower bound constraints for v. The overbooking optimization problem is

$$\begin{array}{ll} min & c_s \cdot \sum_{x=0}^{\frac{c}{v}}(c - x\cdot v)\cdot f_{SR}(x) + c_o \cdot \sum_{x=\frac{c}{v}}^{+\infty}(x\cdot v - c)\cdot f_{SR}(x) \\ \text{s.t.} & \frac{\sum_{x=\frac{c}{v}}^{+\infty}(x\cdot v - c)\cdot f_{SR}(x)}{v\cdot E[SR]} \le r \\ & v_l \le v \le v_u \end{array} \qquad (15)$$

The optimization problem should be solved for each reading day separately.

The optimal overbooking level with respect to the given problem is

$OB^{opt} = \frac{v^{opt}}{c} \cdot 100$

where v^{opt} represents the optimal solution to (15).

2.2.5 Impact of the show-up rate estimation on costs/profits

We compare the tendered amount of cargo at departure with the real capacity available on reading day zero (departure day). The closer the tendered amount was to the real capacity, the better the overbooking policy.

Two factors determine the tendered amount of cargo at departure:

(1) The overbooking levels per reading day, which are directly related to the show-up rate estimators, and

(2) The estimate of the capacity available for free sale.

To compare the influence of the show-up rate estimators on profits, we simulated the cargo-booking process, which can be summarized as follows. In each reading day,

(1) We calculate current bookings based on previous bookings and the cancellation rate,

(2) We calculate overbooking levels and hence the authorized capacity,

(3) Demand arrives,

(4) We accept demand according to the space available after subtracting current bookings from the authorized capacity, and

(5) We update current bookings to take account of the newly accepted demand.

We considered two demand scenarios. In the first scenario, we modeled the demand arrivals as normally distributed random variables. For cancellations, we used two random variables: a uniformly distributed random variable to model the probability of cancellations occurring on a certain day, and a normally distributed random variable to model the magnitude of cancellations.

In the second scenario, we used real-world demand data. Only truncated demand data was available, however, since most companies do not recorded lost sales. By truncated demand we mean the demand the airlines satisfied, not including the demand lost because of insufficient capacity. Hence, the truncated demand is a lower bound on the actual demand. Although not equal to the real demand, the truncated demand captures the dynamics of the booking process, that is, cancellations and still periods.

Truncated, or censored, data is common in the airlines' passenger business. Weatherford and Pölt ([43]) analyzed six methods used to uncensor passenger demand data: three so-called naïve methods, and three more sophisticated methods. These methods work on data that contain an indicator as to whether a particular fare class was open or closed to bookings at the specified time. The three naïve methods are:

- (N1) To use all data and ignore whether bookings were open or closed;
- (N2) To use only open observations and toss out the closed ones;
- (N3) To replace closed observations with the larger of the following, the actual observations or the average of the open observations.

The method we used to uncensor the data is close to (N3): for the days the capacity was completely utilized, we added a normally distributed random variable with a probability of 0.5, since we do not know which observations were open and which were closed. Although other methods to uncensor data exist, we would have had to test them empirically. (N3) is a reasonable trade-off between complexity and performance, as Weatherford and Pölt ([43]) pointed out.

To calculate the overbooking levels in both scenarios, we used a failure rate of 10 percent, a lower and upper bound of 100 percent and 200 percent of the physical capacity, respectively, and a ratio of 4 to 1 for spoilage and off-loads costs. In the air-cargo industry, spoilage is more costly than off-loads. At departure, airlines generally have a good mix of general and time-sensitive cargo. When they have capacity or over-show problems, airlines usually reroute the general cargo, for which there is no significant penalty or loss of goodwill. But when there is less cargo than capacity at departure, the aircraft flies partially empty, which translates into lost opportunity. Most airlines have a cost ratio of 1 to 3 or 4 for off-loads versus spoilage.

We implemented the overbooking policy used by several major airlines (see Section 2.2.4.1), using the Normal and the discrete show-up rate estimators. This resulted in two different authorized capacity levels and, consequently, in two different streams of accepted demand, that is, current bookings per RD, for each simulation run. We called the current bookings resulting from the Normal and discrete estimators CB_{normal} and $CB_{discrete}$. To obtain the tendered cargo from the Normal and the discrete estimators, T_{normal} and $T_{discrete}$, we applied the actual show-up rates for the validation period, SR_{actual}, to the current bookings per RD:

$T_{normal\,RD} = CB_{normal\,RD} * SR_{actual\,RD}$, and

$T_{discrete\,RD} = CB_{discrete\,RD} * SR_{actual\,RD}$.

We compared the tendered cargo ($T_{normal\,RD}$ and $T_{discrete\,RD}$) with an ideal solution and with the real capacity at departure. We obtained an ideal solution from the deterministic version of the process: if we knew all the demand that would show up in advance, then we would accept demand per reading day up to the estimated capacity at departure. The tendered demand in this case, $T_{ideal\,RD}$, is the accepted demand per reading day.

For the normally distributed demand scenario, we conducted experiments for all combinations of low, medium, and high mean demand as a percentage of the capacity and coefficient of variation (standard deviation over mean).

Mean demand Demand CV	low	medium	high
low	60% 0.2	80% 0.2	95% 0.2
medium	60% 0.4	80% 0.4	95% 0.4
high	60% 0.6	80% 0.6	95% 0.6

Figure 7: Demand scenarios: the upper right corner is the mean demand as percentage of capacity, the lower left is the coefficient of variation

We ran 500 simulations for each of the nine experiment settings (Figure 7) and obtained similar results. For all instances we have the following results:

(1) In a comparison with the ideal solution

- The mean absolute error of T_{normal} (compared to T_{ideal}) at departure was approximately seven percent higher than the mean absolute error of $T_{discrete}$;
- The standard deviation of the error was approximately two percent higher for T_{normal} than for $T_{discrete}$ (compared to T_{ideal}).

(2) In a comparison with the real capacity

- The mean absolute error of T_{normal} (compared to the real capacity at departure) was approximately four percent higher than the mean absolute error of $T_{discrete}$;
- The standard deviation of the error is approximately one percent higher for T_{normal} than for $T_{discrete}$ (compared to the real capacity at departure).

The differences between the comparisons with the real capacity and the ideal solution at departure result from the inaccuracy of the capacity estimate per reading day. Consider a simple example, in which the estimated capacity for any given reading day $j > 0$ exceeds the real capacity at departure and the demand is greater than the estimated capacity in any given reading day. In this case, even if we do not overbook and accept as much demand as the estimated capacity, we still end up with demand that cannot be accommodated at departure.

We reran the simulations assuming perfect forecast of cargo capacity at departure and found that the impact of a poor capacity forecast on the business is considerable. The mean absolute error between the tendered cargo and the real capacity at departure is on average 25 percent higher and the standard deviation of the error 10 percent higher for the Normal estimator for all instances.

The mean off-loads (accepted demand that cannot be accommodated at departure) are on average significantly higher (45 percent) for the Normal estimator, and the Normal estimator results in off-loads 10 percent more often than the discrete estimator. The discrete estimator results in spoilage about 25 percent more often than the Normal estimator, but the mean spoilage is about 10 percent lower for the discrete estimator. For cargo, the total quantity of spoilage, and not the frequency, is the leading factor for costs (or lost profits). Hence, the higher spoilage frequency does not affect the gain from its considerably lower mean.

When we ran the simulations using the altered real world truncated demand, the results were consistently better in terms of mean absolute error, mean spoilage, and frequency. The mean absolute error and spoilage were on average 14 percent and 22 percent higher, respectively, when we used the Normal estimator. The off-loads were statistically equal when the added normal variable for un-truncating demand had a high mean and variance, and the discrete estimator resulted in off-loads 5 percent lower in mean than the Normal estimator when the added normal variable had a low mean and variance.

For the real-world demand data, if we used a cost of \$1.6 for unit spoilage and

$0.4 for unit off-loads, typical for the South America to the United States market, the average savings from using the discrete estimator for a combination carrier with 300 flights per day and an average cargo capacity per departure of 13,000 kilograms was $16,425,000 per year. The estimated savings from the simulation may not be the same as savings to be realized in an actual implementation; however, the simulations indicate potentially substantial savings from using the discrete estimator.

CHAPTER III

AIR CARGO CAPACITY MANAGEMENT

This part of the thesis proposes and solves a model for the freight forwarders' problem. The freight forwarder is confronted with confirming the amount of allotted capacity that they need a few days prior to the flight departure. If they confirmed too much capacity, they lose it; if they did not confirm enough, they have to backlog the excess demand. This chapter of the thesis models and solves the problem, such that the costs of the freight forwarder are minimized.

3.1 Existing work

We model the freight forwarders' problem as a perishable inventory problem, the perishable commodity being the amount of capacity to confirm. The commodity perishes after one period, i.e., the flight at the beginning of the current period accommodates the demand coming in over the current period and it takes off at the end of the current period. The order for aircraft capacity is placed L periods of time before take off, and we assume no upper bound on the order capacity, i.e., we do not take into account the amount of capacity that has been allotted far in advance of the flight departure. The assumption is not restrictive, as the optimal order quantity can be capped at the value of the upper bound, in case there is an allotted quantity to take into account. We assume that the unsatisfied demand is backlogged. We assume linear ordering and backlogging costs.

Our work builds on the problem studied in Chew et al. [14]. They define a problem in which the FF has to decide whether to order additional space a few hours before the plane takes off. The decision is a function of the observed state (backlog on hand, carried forward from previous stages), the amount of cargo forecasted to arrive between the time the order is placed and the time the capacity is available (characterized by a known demand distribution function), and the allotted cargo capacity. They use a 6-period planning model with an estimated end of horizon cost. The problem is formulated as a stochastic dynamic programming model, where the state variable is the backlog and the decision variable is the additional space to be acquired. At every stage, the capacity for the next flight may be increased at a supplementary cost, to counterbalance the future penalty costs for having backlogged shipments. The authors showed that for a given state (backlog on hand), the return function is a convex function in the decision variable, and the optimal expected cost function for the remaining stages is a convex increasing function in the state variable. The problem is solved by recursively calculating the additional space to order for each of the 6 periods of the planning model.

We generalize the problem introduced in [14] for finite and infinite horizon, and for lead times of one and two periods. Unlike Chew et al. [14], who solved the problem numerically, we provide exact optimal solutions.

We analyze the problem by defining it as a perishable inventory problem with backlog and lead time. Our decision variable is the order quantity for the next period not taking into account the allotted capacity; however, this impacts only the model definition, since we can always subtract the allotted capacity from the optimal order quantity and decide whether to order additional capacity or not for the next period. If the optimal capacity is consistently under the amount of the allotted capacity, the freight forwarder can also draw conclusions on how allotments should be defined during the next year's bidding period. We analyze the structure of the optimal order quantity and we show that the optimal order quantity follows a stationary policy. We also analyze a few special cases, with subcontracting options for demand with due dates.

Generally, there has been very limited research done in capacity planning models for air cargo, despite its importance in the air cargo supply chain. However, there is a vast research on optimal perishable inventory policies. Table 9 synthesizes the main categories treated in the perishable inventory literature. One feature seen in almost all papers is the assumption that the order is received immediately. When there is a lead time, the solution is either myopic or numerical for a short planning horizon. Our work is characterized by a positive lead time, one period shelf life, backorder, finite and infinite horizon, and variable ordering costs. We assume the perishing cost is the unit ordering cost, as the freight forwarders are not penalized by the airline for not using ordered capacity. We find the optimal policy under these assumptions and show that the optimal expected cost function is a convex function with respect to the available and future capacity ordered L periods in advance, where $L = 1$, or $L = 2$.

We introduce the notation and the optimality equations in the next section. Section 3.3 describes our findings when the time lag is one period. Generalization of these results for two period time lag is given in Section 3.4. Numerical results are presented in Section 3.5. Section 3.6 presents solutions to a few special cases of the problem, when there are subcontracting options for demand with due dates.

Table 9: **Periodic review perishable inventory: summary of literature and work placement**

	L.T.	S.L.	B.O./L.S.	P.H.	Po.	S.T.	Costs
1.[8]	0	1	B.O.	F.H./I.H.	Optimal	Exact	V.O./H.C./B.C
2.[54]	0	2	L.S.	F.H./I.H.	Optimal	Exact	V.O./B.C
3.[36]	0	2	B.O./L.S.	One/F.H.	Optimal	Exact	V.O./B.C./H.C.
4.[20]	0	M.P.	B.O.	F.H./I.H.	Optimal	Exact	V.O./B.C./H.C/P.C.
5.[31]	0	M.P.	B.O.	F.H./I.H.	Optimal	Exact	V.O./B.C./H.C/P.C.
6.[7]	0	M.P.	B.O.	F.H./I.H.	F.O.Q.	Exact	V.O./B.C./H.C/P.C.
7.[30]	0	M.P.	B.O.	F.H./I.H.	B.S.,L.P.	Numer.	V.O./B.C./H.C/P.C.
8.[16]	0	M.P.	B.O.	F.H./I.H.	B.S.	Exact	V.O./B.C./H.C/P.C.
9.[32]	0	M.P.	B.O.	F.H./I.H.	Myopic,B.S.	Appr.	V.O./B.C./H.C/P.C.
10.[12]	0	M.P.	B.O.	F.H./I.H.	B.S.	Bounds	V.O./B.C./H.C/P.C.
11.[33]	0	M.P.	B.O.	F.H./I.H.	Optimal	Comp.	V.O./B.C./H.C/P.C.
12.[34]	0	M.P.	B.O.	F.H./I.H.	Mod.B.S.	Appr.	V.O./B.C./H.C/P.C.
13.[35]	0	M.P.	B.O.	F.H./I.H.	(s, S) type	Numer.	F.O./V.O./B.C./H.C/P.C.
14.[37]	0	M.P.	B.O.	F.H./I.H.	Myopic, B.S.	Bounds	V.O./B.C./H.C/P.C.
15.[55]	> 0	2	L.S.	F.H./I.H.	Myopic,S.P.	Exact	V.O./B.C./H.C/P.C.
16.[17]	0	M.P.	L.S.	F.H./I.H.	B.S.	Bounds	V.O./B.C./H.C/P.C.
17.[56]	1	2	L.S.	F.H./I.H.	Myopic,S.P.	Numer.	V.O./B.C./H.C/P.C.
18.[14]	1	1	B.O.	6 periods	Optimal	Numer.	V.O./B.C./H.C/P.C.
19. Our work	$1, 2$	1	B.O.	F.H./I.H.	Optimal	Exact	V.O./B.C.

The abbreviations in Table 9 refer to the following concepts:

B.O./L.S	:	Backorder or lost sales
B.C.	:	Backordering cost
Comp	:	Comparison
F.H.	:	Finite Horizon
F.O.	:	Fixed ordering cost
F.O.Q.	:	Fixed Order quantity policy
H.C.	:	Holding cost
I.H.	:	Infinite Horizon
L.P.	:	Linear policy (Proposed by [30] using nonperishable version of the problem)
L.T.	:	Lead time
M.P.	:	Multiple period
P.C.	:	Perishing cost
P.H.	:	Planning Horizon
Po.	:	Policy
S.C.	:	Shortage cost
S.L.	:	Shelflife
S.P.	:	Single Period
S.T.	:	Solution Technique
V.O.	:	Variable ordering cost

3.2 Notation and mathematical formulation

We define the FF problem as a periodic review perishable inventory problem with full backlogging and positive lead time. The perishable commodity is the capacity of the plane, which cannot be used after the aircraft's departure. The perishing cost is therefore equal to the cost paid to order the unused capacity, since there are no penalties imposed by airlines for not using committed capacity. There is a lag between the time the aircraft's capacity is confirmed/ordered and the time that the associated aircraft departs. If the demand exceeds the capacity at departure, it is fully backlogged and shipped with the next available opportunity; that is, we assume the demand is satisfied according to a first-in-first-out (FIFO) policy: ship backlog first, from oldest to newest, then new demand. The objective is to minimize the total discounted cost over the planning horizon.

The number of periods in which demands arrive is N and hence, the planning horizon is N periods, numbered from 1 to N. At any time t, the order is placed before the demand in the current period is realized, based on current capacity available or backlog on hand as well as the quantity of the outstanding orders. A snapshot of the process' time line for a two period time lag is given in Figure 8. At time t, we might have backlog from previous periods, denoted by B_{t-1}. Capacity ordered two periods before, q_{t-2}, is available to ship current backlog and incoming demand during period t. The demand materializes after the order q_t is placed. Since the time lag is 2 periods the capacity ordered at time t will only be available at the beginning of period $t+2$. The capacity of size q_{t-1}, which is ordered at time $t-1$, becomes available for shipping any backlog from the previous period and the demand during period $t+1$, D_{t+1}. We denote by $x_t = q_{t-2} - B_{t-1}$ the available capacity at time t, unrestricted in sign. If it is positive it represents remaining capacity after the backlog has been accommodated, and if it is negative it represents remaining backlog that could not be accommodated by q_{t-2}.

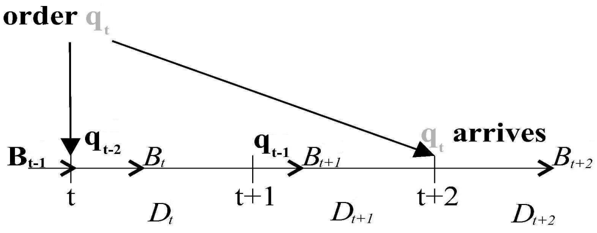

Figure 8: The time line for L=2

At time $N+1$, the end of the horizon, there is no demand to materialize anymore and we use a deterministic decision to clear up eventual backlog: if there is available capacity at $N+1$, we do not place any order; if there is some backlog at the end of the horizon, we place an order only if the backlog exceeds the sum of ordered capacities that have to arrive at $N+2, \ldots, N+L$.

We assume all costs are linear; Table 10 displays the parameters, variables and known quantities at time t. The density function of the demand D_t at any time t, $\varphi_t(y)$, is assumed continuous, differentiable, and bounded between m and M, with $m > 0$, and M sufficiently large; the demands in subsequent periods are assumed to be independent of each other. The cumulative distribution function for the demand at period t is defined as $\Phi_t(x) = \int_m^x \varphi_t(y)dy$.

Table 10: Problem parameters, variables and known quantities at time t for general lead time L

parameters	**known at t**	**decision**
c - ordering cost per unit capacity b - backorder cost per unit per period delay L - lead time	$\varphi_t(y)$,$\Phi_t(y)$ $x_t = B_{t-1} - q_{t-L}$	q_t

The optimal cost is not affected by the FIFO shipping policy. At period t, the capacity q_{t-L} has to accommodate the demand during period t and eventual backlog from previous periods. If $q_{t-L} \geq B_{t-1} + D_t$, then it is obvious that the FIFO rule does not affect optimality. If $q_{t-L} \leq B_{t-1} + D_t$ (see Figure 9), the cost for that period is: $\text{cost}_t = c \cdot q_{t-L} + b \cdot (B_{t-1} + D_t - q_{t-L})$, whether the backlog is shipped first or the demand, due to same the backlogging costs over the entire horizon. Therefore, the FIFO shipping policy is optimal with the underlying cost structures.

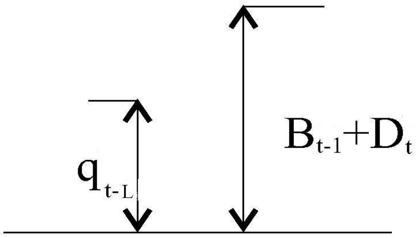

Figure 9: The capacity cannot accommodate demand and backlog for a given period

3.2.1 The general infinite horizon optimality equation.

The process is modeled as a discounted Markov Decision Process, with discount factor $0 < \alpha \leq 1$. The process is defined by:

- *the state space* $\mathcal{S} \subset \Re^L$*; the state* $s \in \mathcal{S}$ *is defined as the vector* $(x_t, q_{t-L+1}, \ldots, q_{t-1})$, where $q_{t-i}, 1 \leq i \leq L-1$ denotes the quantity of the order placed at period $t-i$;

- *the action space* $\mathcal{A} \subset \Re$*; the action* $a \in \mathcal{A}$ *is defined as the order quantity* q_t;

- *the objective function is to minimize the total discounted cost over the planning horizon.*

Let $v_t(x_t, q^*_{t-L+1}, \ldots, q^*_{t-1})$ be the minimum expected discounted cost from time t on, following an optimal policy, i.e., the quantities $q^*_{t-i}, 1 \leq i \leq L-1$ denote the optimal order quantities placed at period $t-i$. Then, the optimality equation can be written as follows:

$$
\begin{aligned}
&v_t(x_t, q^*_{t-L+1}, \ldots, q^*_{t-1}) = \min_{q_t \geq 0} \left\{ c \cdot q_t + b \cdot \int_{x_t}^{M} (y - x_t) \cdot \varphi_t(y) dy + \right. \\
&+ \alpha \cdot \left[\int_{m}^{x_t} v_{t+1}(q^*_{t-L+1}, \ldots, q^*_{t-1}, q_t) \varphi_t(y) dy + \right. \\
&\left. + \int_{x_t}^{M} v_{t+1}(x_t + q^*_{t-L+1} - y, \ldots, q^*_{t-1}, q_t) \varphi_t(y) dy \right]
\end{aligned} \tag{16}
$$

which can be rewritten as:

$$
\begin{aligned}
&v_t(x_t, q^*_{t-L+1}, \ldots, q^*_{t-1}) = \min_{q_t \geq 0} \left\{ c \cdot q_t + b \cdot \int_{x_t}^{M} (y - x_t) \cdot \varphi_t(y) dy + \right. \\
&+ \alpha \cdot \left[\Phi(x_t) \cdot v_{t+1}(q^*_{t-L+1}, \ldots, q^*_{t-1}, q_t) + \right. \\
&\left. + \int_{x_t}^{M} v_{t+1}(x_t + q^*_{t-L+1} - y, \ldots, q^*_{t-1}, q_t) \varphi_t(y) dy \right]
\end{aligned} \tag{17}
$$

The first term is the ordering cost, and the second term is the expected backlogging cost for the current period. The last term represents the minimum expected cost from time $t + 1$ on; depending on the realization of demand in the current period, period $t + 1$ starts with some positive available capacity or some excess demand is carried over so that no capacity exists.

The next section introduces the main results with proofs for $L = 1$. $L = 2$ is analyzed in Section 3.4. Although we present the optimality equations with period dependent demand, in the next sections we assume identically distributed demand and for simplicity of exposition, we drop the index t from the probability density and cumulative distribution function for the demand.

3.3 Main results and discussion for one period time lag

The problem consists of N sequential decisions on how much capacity to reserve on a plane that is scheduled to take off one period after we make the decision. We assume the decisions are continuous.

At time $t \in \{1, \ldots, N\}$, we have available capacity/backlog x_t to satisfy incoming demand, and we place order q^t, which arrives one period later. At the end of the horizon, time $N+1$, we assume we can clear any leftovers by ordering as much capacity as necessary. The capacity comes one period later, so we have to pay the backlogging cost for one period. The end of horizon cost is:

$$v_{N+1}(x_{N+1}) = \begin{cases} 0 & \text{if } x_{N+1} \geq 0 \\ -x_{N+1} \cdot (b+c) & \text{otherwise} \end{cases} \tag{18}$$

The finite horizon optimality equations are given as follows:

$$\begin{aligned} v_t(x_t) = \min_{q_t \geq 0} \Big\{ & c \cdot q_t + b \cdot \int_{x_t}^{M} (y - x_t) \cdot \varphi(y) dy + \alpha \cdot \Phi(x_t) \cdot v_{t+1}(q_t) + \\ & +\alpha \cdot \int_{x_t}^{M} v_{t+1}(x_t + q_t - y) \cdot \varphi(y) dy \Big\}, \ t \in \{1 \ldots N\} \\ v_{N+1}(x_{N+1}) = & \begin{cases} 0 & \text{if } x_{N+1} > 0 \\ -x_{N+1} \cdot (b+c) & \text{otherwise} \end{cases} \end{aligned} \tag{19}$$

For a negative state variable $x_t \leq 0$, the optimality equation becomes:

$$v_t(x_t) = \min_{q_t \geq 0} \left\{ c \cdot q_t + b \cdot \int_{m}^{M} (y - x_t) \cdot \varphi(y) dy + \alpha \cdot \int_{m}^{M} v_{t+1}(x_t + q_t - y) \cdot \varphi(y) dy \right\} \tag{20}$$

If the state is negative, any optimal policy will order more capacity than backlog on hand, since all demand has to be shipped by the end of the horizon and the objective is to minimize total cost. Thus, we can substitute $q_t = -x_t + Q_t$ in (20), with $Q_t \geq 0$, and obtain:

$$\begin{aligned} v_t(x_t) = \min_{Q_t \geq 0} & \left\{ -c \cdot x_t + c \cdot Q_t + b \cdot \int_{m}^{M} (y - x_t) \cdot \varphi(y) dy + \alpha \cdot \int_{m}^{M} v_{t+1}(Q_t - y) \varphi(y) dy \right\} = \\ & = -(b+c) \cdot x_t + v_t(0) \end{aligned} \tag{21}$$

We immediately see from Equation (21) that it is necessary to analyze the value functions only with nonnegative state variables. Moreover, as it will be stated in the subsequent parts of this section, when the state variable is negative, the optimal policy states that the optimal order quantity is the sum of the backorder quantity and the optimal quantity when the state variable is zero.

We assume $b \geq \frac{1-\alpha}{\alpha} \cdot c$, which is not a tight assumption since α usually takes values approaching 1. The assumption ensures the existence of a solution at the end of the horizon. In the inventory literature, such an assumption is often made (e.g., see Arrow et. al. [3]).

The Theorem below states the convexity of the value function in the state variable:

Theorem 2

$$v''_t(x_t) > 0 \ \forall x_t > 0, \ \textit{and} \ v''_t(x_t) = 0 \ \forall x_t \leq 0 \tag{22}$$

Proof:

The statement for negative state variables follows by taking the second derivative of equation (21) as a function of x_t.

The statement for negative state variables follows by taking the second derivative of equation (21) as a function of x_t. For positive state variables, we will use induction and the following general results presented in Lemmas 1-4:

Lemma 1 *The derivative of the value function at any time t is:*

$$v'_t(x_t) = -b - c + \Phi(x_t) \cdot (b - \alpha \cdot v'_{t+1}(q^*_t)) \tag{23}$$

where q^*_t is the solution of the following equation:

$$\frac{dv_t}{dq^*_t} = c + \alpha \cdot v'_{t+1}(q^*_t) \cdot \Phi(x_t) + \alpha \cdot \int_{x_t}^{M} v'_{t+1}(x_t + q^*_t - y)\varphi(y)dy = 0 \tag{24}$$

Proof of Lemma 1:

$$\begin{aligned} v_t'(x_t) &= c \cdot \frac{dq_t^*}{dx_t} - b \cdot (1 - \Phi(x_t)) + \alpha \cdot \Phi(x_t) v_{t+1}'(q_t^*) \cdot \frac{dq_t^*}{dx_t} + \\ &+ \alpha \cdot (1 + \frac{dq_t^*}{dx_t}) \cdot \int_{x_t}^{M} v_{t+1}'(x_t + q_t^* - y) \cdot \varphi(y) dy \end{aligned} \tag{25}$$

By factoring out $\frac{dq_t^*}{dx_t}$ in (25), and using (24), we get:

$v_t'(x_t) = -b \cdot (1 - \Phi(x_t)) + \alpha \cdot \int_{x_t}^{M} v_{t+1}'(x_t + q_t^* - y) \cdot \varphi(y) dy$

Using $\alpha \cdot \int_{x_t}^{M} v_{t+1}'(x_t + q_t^* - y) \cdot \varphi(y) dy = -c - \alpha \cdot v_{t+1}'(q_t^*) \cdot \Phi(x_t)$ from (24), we deduce (23). (24) is a straightforward derivation of the value function as a function of q_t^*.

■

The general second derivative of the value function is given in the following Lemma:

Lemma 2

$$v_t''(x_t) = \varphi(x_t) \cdot (b - \alpha \cdot v_{t+1}'(q_t^*)) - \alpha \cdot v_{t+1}''(q_t^*) \cdot \Phi(x_t) \cdot \frac{dq_t^*}{dx_t} \ , \forall x_t > 0 \tag{26}$$

which is a straightforward derivation of equation (23).

■

If we use induction and assume that $v_{t+1}''(x_{t+1}) \geq 0$, $\forall x_{t+1} > 0$, then we need two results for proving Theorem 2: $b - \alpha \cdot v_{t+1}'(q_t^*) \geq 0$, and $\frac{dq_t^*}{dx_t} \leq 0$. For the first result, we use:

Lemma 3 *The derivative of the value function has the following property:*

$$v_t'(x_t) \geq -b - c \textit{ for any } x_t; \textit{ furthermore, if } x_t < 0, \textit{ then } v_t'(x_t) = -b - c \tag{27}$$

Proof of Lemma 3: Derivating equation (21) as a function of x_t gives us the statement for negative state variables.

We use induction for positive state variables. Since $v_{N+1}(x_{N+1}) = 0$ for $x_N > 0$, we have $v'_N(x_N) = -b - c + b \cdot \Phi(x_N) \geq -b - c$ and $v'_{N-1}(x_{N-1}) = -b - c + \Phi(x_{N-1}) \cdot (b - \alpha \cdot v'_N(q^*_N)) = -b - c + \Phi(x_{N-1}) \cdot (b + \alpha \cdot (b + c - b \cdot \Phi(q^*_N))) = -b - c + \Phi(x_{N-1}) \cdot (b + \alpha \cdot (b + c)) \cdot (1 - \frac{\alpha \cdot b}{b + \alpha \cdot (b+c)} \Phi(q^*_N)) \geq -b - c$. where the last inequality comes from the fact that $1 - \frac{\alpha \cdot b}{b+\alpha\cdot(b+c)} \Phi(q^*_N) \geq 0$.

For the induction step, we assume $v'_{t+1}(x_{t+1}) \geq -b - c$, $\forall x_{t+1}$, and we need to show that the same holds for t. The induction assumption can be re-written as: $v'_{t+1}(x_{t+1}) = -b - c + \Phi(x_{t+1}) \cdot (b - \alpha \cdot v'_{t+2}(q^*_{t+1})) \geq -b - c$, which implies $b - \alpha \cdot v'_{t+2}(q^*_{t+1}) \geq 0$. Also, if the induction hypothesis holds for all periods $N + 1$ to t (counted backwards), then $b - \alpha \cdot v'_{t+2}(q^*_{t+1}) \leq b + \alpha \cdot (b + c)$.

From equation (23), we deduce:

$$v'_t(x_t) = -b - c + \Phi(x_t) \cdot (b - \alpha \cdot v'_{t+1}(q^*_t)) = -b - c + \Phi(x_t) \cdot (b - \alpha \cdot (-b - c + \Phi(q^*_t) \cdot \\ \cdot (b - \alpha \cdot v'_{t+2}(q^*_{t+1})))) = -b - c + \Phi(x_t) \cdot (b + \alpha \cdot (b + c)) \cdot \Bigg[1 - \frac{\alpha \cdot \Phi(q^*_t)}{b + \alpha \cdot (b + c)} \cdot \\ \cdot (b - \alpha \cdot v'_{t+2}(q^*_{t+1})) \Big]$$

We have $b + \alpha \cdot (b + c) \geq b - \alpha \cdot v'_{t+2}(q^*_{t+1}) \geq 0$ by the induction hypothesis. In order for (28) to be greater or equal to $-b - c$, we need $b - \alpha \cdot v'_{t+2}(q^*_{t+1}) \leq \frac{b+\alpha\cdot(b+c)}{\alpha \cdot \Phi(q^*_t)}$, which is true since $\alpha \cdot \Phi(q^*_t) \leq 1$.

Observation 1: An immediate consequence of the Lemma above is that $b + \alpha \cdot (b + c) \geq b - \alpha \cdot v'_{t+1}(q^*_t) \geq 0$, $\forall q^*_t$.

■

The first result to obtain (26) is proven. The second result is proven in the following:

Lemma 4

$$-1 \leq \frac{dq^*_t(x_t)}{dx_t} \leq 0 \ \forall x_t \tag{28}$$

Proof of Lemma 4: The quantity q_t^* is the solution of equation (24), of which we take the derivative as a function of x_t and obtain:

$$\alpha \cdot \varphi(x_t) \cdot v'_{t+1}(q_t^*) + \alpha \cdot \Phi(x_t) v''_{t+1}(q_t^*) \frac{dq_t^*}{dx_t} + \alpha \cdot \int_{x_t}^{M} v''_{t+1}(x_t + q_t^* - y)\varphi(y)dy \cdot$$
$$\cdot \left(1 + \frac{dq_t^*}{dx_t}\right) - \alpha \cdot \varphi(x_t) \cdot v'_{t+1}(q_t^*) = 0 \Longleftrightarrow \alpha \cdot \Phi(x_t) v''_{t+1}(q_t^*) \frac{dq_t^*}{dx_t} +$$
$$+ \alpha \cdot \int_{x_t}^{M} v''_{t+1}(x_t + q_t^* - y)\varphi(y)dy \cdot \left(1 + \frac{dq_t^*}{dx_t}\right) = 0$$

Under the induction assumption that $v''_{t+1}(x_{t+1}) \geq 0, \forall x_{t+1} > 0$, the equation does not hold unless $\frac{dq_t^*}{dx_t} \leq 0$ and $1 + \frac{dq_t^*}{dx_t} \geq 0$, since all other terms are positive or zero for negative state variables.

■

To conclude the induction, we calculate:

$v''_{N+1}(x_{N+1}) = 0$

$v''_N(x_N) = b \cdot \varphi(x_N) \geq 0 \; \forall x$

Theorem 2 is proven.

■

We next present a theorem that contains some structural results on the optimal ordering policy at any period t.

Theorem 3 *The optimal policy $q_t^*(x_t)$ has the following properties:*

1. *$q_t^*(x_t)$ is a continuous function of x_t;*

2. $\lim_{x_t \to \infty} q_t^*(x_t) > 0$;

3. $0 \geq \frac{dq_t^*(x_t)}{dx_t} \geq -1 \; \forall x_t$;

4. *if $x_t < 0$, then $q_t^*(x_t) = -x_t + q_t^*(0)$.*

Proof:

Point 1 immediately follows from Equation (21). To prove Points 2-3, we expand equation (24) using Lemma 3, which provides the quantity q_t^*.

$$\begin{gathered}c + \alpha \cdot \Phi(x_t) \cdot \big(-b - c + \Phi(q_t^*) \cdot (b - \alpha \cdot v'_{t+2}(q^*_{t+1}))\big) + \alpha \cdot \int_{x_t}^{M} (-b - c + \\ + \Phi(q_t^* + x_t - y) \cdot (b - \alpha \cdot v'_{t+2}(q^*_{t+1}))\big)\, \varphi(y)dy = c - \alpha \cdot (b + c) + \\ + \alpha \cdot (b - \alpha \cdot \cdot v'_{t+2}(q^*_{t+1})) \cdot \Phi(x_t) \cdot \Phi(q_t^*) + \alpha \cdot (b - \alpha \cdot v'_{t+2}(q^*_{t+1})) \cdot \\ \cdot \int_{x_t}^{M} \Phi(x_t + q_t^* - y)\varphi(y)dy = 0\end{gathered}$$

from which we deduce:

$$\begin{gathered}\Phi(x_t) \cdot \Phi(q_t^*) + \int_{x_t}^{x_t + q_t^*} \Phi(x_t + q_t^* - y)\varphi(y)dy = \frac{\alpha \cdot (b + c) - c}{\alpha \cdot \big(b - \alpha \cdot v'_{t+2}(q^*_{t+1})\big)} \geq \\ \geq \frac{\alpha \cdot (b + c) - c}{\alpha \cdot (b + \alpha \cdot (b + c))}\end{gathered} \tag{29}$$

The last inequality follows by the Observation 1 deduced from Lemma 3.

Hence: q_t^* solves:

$$\Phi(x_t) \cdot \Phi(q_t^*) + \int_{x_t}^{x_t + q_t^*} \Phi(x_t + q_t^* - y)\varphi(y)dy = \frac{\alpha \cdot (b + c) - c}{\alpha \cdot \big(b - \alpha \cdot v'_{t+2}(q^*_{t+1})\big)} > 0 \tag{30}$$

and since we assumed a continuous demand distribution, the solution to the equation above is a continuous function of the state variable x_t. Furthermore, if we take limits of equation (30), we obtain:

$$\lim_{x_t \to \infty} \left(\Phi(x_t) \cdot \Phi(q_t^*) + \int_{x_t}^{x_t + q_t^*} \Phi(x_t + q_t^* - y)\varphi(y)dy \right) = \lim_{x_t \to \infty} \Phi(q_t^*) > 0 \tag{31}$$

and by continuity and monotonicity (non-decreasing) of the cumulative distribution function, we can conclude that $\lim\limits_{x_t \to \infty} q_t^*(x_t) > 0$.

■

This theorem states that the optimal quantity ordered at any stage t is a continuous, nonincreasing function of the current available capacity. Interestingly, it is never optimal to order no capacity for the next period, no matter how large the current available capacity is. If at stage t we have no available capacity but backorders only, then for the next stage it is optimal to order the amount of backlog on hand plus what quantity we order when the current available capacity is zero. Furthermore, in Lemma 4, we also showed that a unit increase in the state variable results in less than a unit decrease in the optimal order quantity. This result is similar to the ones that are proved in [20] and [31] for zero lead time.

Before we discuss the stationary policy, we summarize the following theorem (see [42]):

Theorem: Under the assumptions that we have:

- Stationary rewards and transition probabilities - the cost/profits do not change from period to period; furthermore, the demand is identically distributed;
- Bounded rewards;
- Discounting;
- Discrete state space: the state space is finite or countably infinite

suppose there exists an optimal policy.

Then there exists a deterministic stationary policy that is optimal.

We know that our optimal policy is a deterministic (deterministic in the sense that once we know the state, the optimal quantity to order is known with certainty) stationary policy, i.e., the order placed far enough from the end of the horizon (to clear off the effects of finiteness of horizon on the order quantity) depends on the value of the state variable only, and not on the current period. We present the limiting equation in the following Theorem:

Theorem 4 *The optimal policy is a stationary policy when $t \to \infty$, and it solves the following equation:*

$$\Phi(x) \cdot \Phi(q^*) + \int_x^{x+q^*} \Phi(x+q^*-y)\varphi(y)dy = \frac{\alpha \cdot (b+c) - c}{(\alpha^2 \cdot (b+c) + \alpha \cdot b) \cdot (1-\mathcal{A}_x)} \tag{32}$$

where x is the available capacity/backlog during the current period, q^* is the order to be placed during the current period, and $1 \geq \mathcal{A}_x \geq 0$ is the limit of the alternating series: $S_{N-t} = \alpha \cdot \Phi(q^*_{t+1}) - \alpha^2 \cdot \Phi(q^*_{t+1}) \cdot \Phi(q^*_{t+2}) + \alpha^3 \cdot \Phi(q^*_{t+1}) \cdot \Phi(q^*_{t+2}) \cdot \Phi(q^*_{t+3}) - ... \pm \alpha^{N-t} \cdot \Phi(q^*_{t+1}) \cdot \dots \cdot \Phi(q^*_N)$. Furthermore, the value of the limit $\mathcal{A}_x$ can be approximated within an arbitrary error.

Proof: We have to show that $b - \alpha \cdot v'_{t+2}(q^*_{t+1})$ is a convergent series.

Lemma 5 *The series $b - \alpha \cdot v'_{t+2}(q^*_{t+1}) = (b + \alpha \cdot (b+c)) \cdot (1 - \alpha \cdot \Phi(q^*_{t+1}) + \alpha^2 \cdot \Phi(q^*_{t+1}) \cdot \Phi(q^*_{t+2}) - \alpha^3 \cdot \Phi(q^*_{t+1}) \cdot \Phi(q^*_{t+2}) \cdot \Phi(q^*_{t+3}) + ... \pm \alpha^{N-t} \cdot \Phi(q^*_{t+1}) \cdot \dots \cdot \Phi(q^*_N))$ is convergent.*

Proof of Lemma 5: The series in Lemma 5 is obtained by repeatedly using equation (23):

$$\begin{aligned}
&b - \alpha \cdot v'_{t+2}(q^*_{t+1}) = (b + \alpha \cdot (b+c)) \cdot (1 - \alpha \cdot \Phi(q^*_{t+1}) \cdot (1 - \alpha \cdot \Phi(q^*_{t+2}) \cdot (1- \\
&- \alpha \cdot \Phi(q^*_{t+3}) \cdot \cdot (1 - \dots)) \dots) = (b + \alpha \cdot (b+c)) \cdot (1 - \alpha \cdot \Phi(q^*_{t+1}) + \\
&+ \alpha^2 \cdot \Phi(q^*_{t+1}) \cdot \Phi(q^*_{t+2}) - \alpha^3 \cdot \Phi(q^*_{t+1}) \cdot \Phi(q^*_{t+2}) \cdot \Phi(q^*_{t+3}) + ... \pm \\
&\pm \alpha^{N-t} \cdot \Phi(q^*_{t+1}) \cdot \dots \cdot \Phi(q^*_N)
\end{aligned} \tag{33}$$

The condition for an alternating series to converge is that the terms are non-increasing in magnitude, with a limiting value of zero (see e.g., [47]). In the above series, the terms decrease by $\alpha \cdot \Phi(q_n^*) < 1$ and the series is a multiplication of terms less than 1, hence the nth term goes to 0 in time. The limit of the terms $b - \alpha \cdot v'_{t+2}(q_{t+1}^*)$ is $\alpha \cdot (b + \alpha \cdot (b+c)) \cdot (1 - \mathcal{A}_x)$, where $\mathcal{A}_x$ is the limit of the alternating series $\alpha \cdot \Phi(q_{t+1}^*) - \alpha^2 \cdot \Phi(q_{t+1}^*) \cdot \Phi(q_{t+2}^*) + \alpha^3 \cdot \Phi(q_{t+1}^*) \cdot \Phi(q_{t+2}^*) \cdot \Phi(q_{t+3}^*) - ... \pm \alpha^{N-t} \cdot \Phi(q_{t+1}^*) \cdot \dots \cdot \Phi(q_N^*))$. Furthermore, $b + \alpha \cdot (b+c) \geq b - \alpha \cdot v'_{t+2}(q_{t+1}^*) \geq 0$, by Observation deduced from Lemma 3, which implies $1 \geq \mathcal{A}_x \geq 0$.

■

The stationary equation is similar to the solution of a newsvendor problem. The order quantity q^* is placed such that the probability of ending the next period from current with no backlog is equal to the righthandside term in (32). The limit $\mathcal{A}_x$ in Theorem 4 above can be approximated by using the following theorem:

Theorem:

Suppose a_n is a monotone decreasing sequence that converges to 0, then:
$S = \sum_{n=0}^{n=\infty} a_n$
converges and has $|S_k - S| < a_{k+1}$.

As an example, we have the series: $\sum_{n=0}^{n=\infty} \frac{(-1)^n}{n}$, which converges conform Theorem above. To get the sum to within .001 will take 999 terms.

In our case, assessing the speed of convergence depends on the distribution shape Φ. Section 3.5 presents convergence rates for several distributions.

■

3.4 Main results and discussion for two period time lag

In this part, we again assume a planning horizon of N periods with active demand, and at the end of the horizon (the beginning of period $N+1$), we clear possible eventual backlogs. The order placed in the current period arrives in two periods. At the beginning of the current period we know the available capacity/backlog for the current period and the order placed one period before the current period. Hence, the state variable is described by the pair (x_t, q^*_{t-1}). As explained in the sequence of events above, the demand materializes after we place the order.

The quantity ordered at the end of the horizon $N+1$ is a deterministic function of the available capacity/backlog at $N+1$, and the quantity ordered at N, which is scheduled to arrive at $N+2$:

$$q^*_{N+1}(x_{N+1}, q^*_{N-1}) = \begin{cases} 0 & \text{if } x_{N+1} + q^*_{N-1} \geq 0 \\ -(x_{N+1} + q^*_{N-1}) & \text{otherwise} \end{cases} \tag{34}$$

If there is no backlog at time $N+1$ or if the backlog can be honored with the capacity ordered at N and scheduled to arrive at $N+2$, then we do not order any capacity, otherwise we order capacity to cover the backlog that exceeds the capacity to arrive at $N+2$. Hence, the corresponding cost is:

$$v_{N+1}(x_{N+1}, q^*_{N-1}) = \begin{cases} 0 & \text{if } x_{N+1} \geq 0 \\ -b \cdot x_{N+1} & \text{if } x_{N+1} < 0, \\ & \text{and } x_{N+1} + q^*_{N-1} \geq 0 \\ -b \cdot x_{N+1} - (c + \alpha \cdot b) \cdot (x_{N+1} + q^*_{N-1}) & \text{if } x_{N+1} + q^*_{N-1} < 0 \end{cases} \tag{35}$$

We pay backlogging costs for one period for any demand left unsatisfied at the end of the horizon, and if we have to place an additional order for demand that exceeds the capacity ordered for the next period we have to pay an extra period of backlogging costs until the placed order arrives.

The finite horizon optimality equation for $L = 2$ is:

$$v_t(x_t, q^*_{t-1}) = \min_{q_t \geq 0}\{c \cdot q_t + b \cdot \int_{x^t}^{M} (y_1 - x_t)\varphi(y)dy + \alpha \cdot \Phi(x_t) \cdot v_{t+1}(q^*_{t-1}, q_t) +$$
$$+ \alpha \cdot \int_{x^t}^{M} v_{t+1}(x_t + q^*_{t-1} - y_1, q_t)\varphi(y_1)dy_1\}$$
$$v_{N+1}(x_{N+1}, q^*_{N-1}) = \begin{cases} 0 & \text{if } x_{N+1} \geq 0 \\ -b \cdot x_{N+1} & \text{if } x_{N+1} < 0, \\ & \text{and } x_{N+1} + q^*_{N-1} \geq 0 \\ -b \cdot x_{N+1} - (c + \alpha \cdot b) \cdot (x_{N+1} + q^*_{N-1}) & \text{if } x_{N+1} + q^*_{N-1} < 0 \end{cases} \tag{36}$$

The optimal order quantity has to take into account the uncertainty in the incoming demand over the next two periods; we denote the demand to be realized in the current period with y_1, and the demand to be realized next period with y_2.

If $x_t < 0$ and $|x_t| < q^*_{t-1}$, then the value function becomes:

$$v_t(x_t, q^*_{t-1}) = \min_{q_t \geq 0}\{c \cdot q_t + b \cdot \int_{0}^{M} (y_1 - x_t)\varphi(y)dy + \alpha \cdot \int_{0}^{M} v_{t+1}(x_t + q^*_{t-1} - y_1, q_t)\varphi(y)dy\} \tag{37}$$

If $x_t < 0$ and $|x_t| > q^*_{t-1}$, i.e., the backlog on hand exceeds even the capacity to arrive next period, the order placed should be greater than the deficit $-(x^t + q^*_{t-1})$:

$$v_t(x_t, q^*_{t-1}) = \min_{q_t \geq |x^t + q^*_{t-1}|}\{c \cdot q_t + b \cdot \int_{0}^{M} (y_1 - x_t)\varphi(y)dy + \alpha \cdot \int_{0}^{M} v_{t+1}(x^t + q^*_{t-1} - y_1, q_t)\varphi(y_1)dy_1\} \tag{38}$$

We assume $\alpha \cdot b > \frac{1-\alpha}{\alpha} \cdot c$ to ensure the existence of the solution at the end of the horizon. The next results are a generalization of the results for one period time lead; for clarity of exposition we denote the state variables of the value function, x_t and q^*_{t-1}, as x_t and z_t $(\forall t)$:

Theorem 5 *The partial derivatives of the value function have the following properties:*

1. $\frac{\partial v_t}{\partial x_t} \geq -b - c - \alpha \cdot b$, $\forall x_t, z_t$;

2. $\frac{\partial v_t}{\partial z_t} \geq -c - \alpha \cdot b$, $\forall x_t, z_t$.

Proof: We carry our proofs for positive state variable x_t, as the most general case and use induction as well as the following general results for the partial derivatives presented in Lemmas 6-7.

Lemma 6 *The partial derivative of the value function as a function of z_t at any time t is:*

$$\frac{\partial v_t}{\partial z_t} = -c - \alpha \cdot b + \left(\Phi(x_t) \cdot \Phi(z_t) + \int_{x_t}^{x_t + z_t} \Phi(x_t + z_t - y_1) \varphi(y_1) dy_1 \right) \cdot \\ \cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}} \Bigg|_{\substack{x_{t+2} = q_t^* \\ z_{t+2} = q_{t+1}^*}} \right) \tag{39}$$

where q_t^* is the optimal order quantity and solves the equation:

$$c + \alpha \cdot \Phi(x_t) \cdot \frac{\partial v_{t+1}}{\partial z_{t+1}} \Bigg|_{\substack{x_{t+1} = z_t \\ z_{t+1} = q_t^*}} + \alpha \int_{x_t}^{M} \frac{\partial v_{t+1}}{\partial z_{t+1}} \Bigg|_{\substack{x_{t+1} = x_t + z_t - y_1 \\ z_{t+1} = q_t^*}} \varphi(y_1) dy_1 = 0 \tag{40}$$

Proof of Lemma 6: We calculate:

$$\frac{\partial v_t}{\partial z_t} = \alpha \cdot \Phi(x_t) \cdot \frac{\partial v_{t+1}}{\partial x_{t+1}} \Bigg|_{\substack{x_{t+1} = z_t \\ z_{t+1} = q_t^*}} + \alpha \cdot \int_{x_t}^{M} \frac{\partial v_{t+1}}{\partial x_{t+1}} \Bigg|_{\substack{x_{t+1} = x_t + z_t - y_1 \\ z_{t+1} = q_t^*}} \varphi(y_1) dy_1 \tag{41}$$

and

$$\frac{\partial v_t}{\partial x_t} = -b \cdot (1 - \Phi(x_t)) + \alpha \cdot \int_{x_t}^{M} \frac{\partial v_{t+1}}{\partial x_{t+1}}{}_{\substack{x_{t+1}=x_t+z_t-y_1 \\ z_{t+1}=q_t^*}} \varphi(y_1) dy_1 \tag{42}$$

and connect the two partial derivatives by their common term:

$$\frac{\partial v_t}{\partial z_t} = \frac{\partial v_t}{\partial x_t} + b \cdot (1 - \Phi(x_t)) + \alpha \cdot \Phi(x_t) \cdot \frac{\partial v_{t+1}}{\partial x_{t+1}}{}_{\substack{x_{t+1}=z_t \\ z_{t+1}=q_t^*}} \tag{43}$$

and by plugging it in into (40), we obtain:

$$\begin{aligned}
&c + \alpha \cdot \Phi(x_t) \cdot \left[\frac{\partial v_{t+1}}{\partial x_{t+1}}{}_{\substack{x_{t+1}=z_t \\ z_{t+1}=q_t^*}} + b \cdot (1 - \Phi(z_t)) + \alpha \cdot \Phi(z_t) \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}}{}_{\substack{x_{t+2}=q_t^* \\ z_{t+2}=q_{t+1}*}} \right] + \\
&+ \alpha \cdot \int_{x_t}^{M} \left[\frac{\partial v_{t+1}}{\partial x_{t+1}}{}_{\substack{x_{t+1}=x_t+z_t-y_1 \\ z_{t+1}=q_t^*}} + b \cdot (1 - \Phi(x_t + z_t - y_1)) + \right. \\
&\left. + \alpha \cdot \Phi(x_t + z_t - y_1) \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}}{}_{\substack{x_{t+2}=q_t^* \\ z_{t+2}=q_{t+1}^*}} \right] \varphi(y_1) dy_1 = c + \frac{\partial v_t}{\partial z_t} + \alpha \cdot \Phi(x_t) \cdot b \cdot \\
&\cdot (1 - \Phi(z_t)) + \alpha^2 \cdot \Phi(x_t) \cdot \Phi(z_t) \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}}{}_{\substack{x_{t+2}=q_t^* \\ z_{t+2}=q_{t+1}^*}} + \alpha \cdot b \cdot (1 - \Phi(x_t)) - \alpha \cdot b \cdot \\
&\cdot \int_{x_t}^{M} \Phi(x_t + z_t - y_1) \varphi(y_1) dy_1 + \alpha^2 \cdot \int_{x_t}^{M} \Phi(x_t + z_t - y_1) \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}}{}_{\substack{x_{t+2}=q_t^* \\ z_{t+2}=q_{t+1}^*}} \varphi(y_1) dy_1 = 0
\end{aligned}$$

from which equation (39) follows. The last equality follows by substituting the expression from (41) as $\frac{\partial v_t}{\partial z_t}$. ■

Lemma 7 *The partial derivative of the value function as a function of x_t at any time t is:*

$$\begin{aligned}\frac{\partial v_t}{\partial x_t} &= -b - c - \alpha \cdot b + \Phi(x_t) \cdot \left(b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+1}}{\partial z_{t+1}} \Bigg|_{\substack{x_{t+1}=z_t \\ z_{t+1}=q_t^*}} \right) + \\ &+ \int_{x_t}^{x_t+z_t} \Phi(x_t + z_t - y_1)\varphi(y_1)dy_1 \cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}} \Bigg|_{\substack{x_{t+2}=q_t^* \\ z_{t+2}=q_{t+1}^*}} \right)\end{aligned} \tag{44}$$

Proof of Lemma 7: We use the general formula (42) to deduce first:

$$\frac{\partial v_t}{\partial x_t} = -b \cdot (1 - \Phi(x_t)) + \alpha \cdot \int_{x_t}^{M} \frac{\partial v_{t+1}}{\partial x_{t+1}}{}_{\substack{x_{t+1}=x_t+z_t-y_1 \\ z_{t+1}=q_t^*}} \varphi(y_1)dy_1 = -b \cdot (1 - \Phi(x_t)) +$$

$$+ \alpha \cdot \int_{x_t}^{M} \left(-b \cdot (1 - \Phi(x_t + z_t - y_1)) + \alpha \cdot \int_{x_t+z_t-y_1}^{M} \frac{\partial v_{t+2}}{\partial x_{t+2}}{}_{\substack{x_{t+2}=x_t+z_t- \\ -y_1+q_t^*-y_2 \\ z_{t+2}=q_{t+1}^*}} \varphi(y_2)dy_2 \right) \cdot$$

$$\cdot \varphi(y_1)dy_1 = -b - \alpha \cdot b + b \cdot \Phi(x_t) + \alpha \cdot b \cdot \Phi(x_t) + \alpha \cdot b \cdot \int_{x_t}^{x_t+z_t} \Phi(x_t + z_t - y_1) \cdot$$

$$\cdot \varphi(y_1)dy_1 + \alpha^2 \cdot \int_{x_t}^{M} \int_{x_t+z_t-y_1}^{M} \frac{\partial v_{t+2}}{\partial x_{t+2}}{}_{\substack{x_{t+2}=x_t+z_t- \\ y_1+q_t^*-y_2 \\ z_{t+2}=q_{t+1}^*}} \varphi(y_2)dy_2\varphi(y_1)dy_1$$

We deduce the last term from equation (40), in which we plug in expression (41), and obtain:

$$\alpha^2 \cdot \int_{x_t}^{M} \int_{x_t+z_t-y_1}^{M} \frac{\partial v_{t+2}}{\partial x_{t+2}}{}_{\substack{x_{t+2}=x_t+z_t-y_1+q_t^*-y_2 \\ z_{t+2}=q_{t+1}^*}} \varphi(y_2)dy_2\varphi(y_1)dy_1 =$$

$$= -c - \alpha \cdot \Phi(x_t) \cdot \frac{\partial v_{t+1}}{\partial z_{t+1}}{}_{\substack{x_{t+1}=z_t \\ z_{t+1}=q_t^*}} - \alpha^2 \cdot \int_{x_t}^{x_t+z_t} \Phi(x_t + z_t - y_1) \frac{\partial v_{t+2}}{\partial x_{t+2}}{}_{\substack{x_{t+2}=q_t^* \\ z_{t+2}=q_{t+1}^*}}$$

and by substituting it back, we obtain the desired derivative. ■

We use induction and the general results from the Lemmas 6 and 7 for proving Theorem 5. We prove the two results simultaneously.

At the end of the horizon, $N + 1$, we have:

$$\frac{\partial v_{N+1}}{\partial x_{N+1}} = \begin{cases} 0 & \text{if } x_{N+1} \geq 0 \\ -b & \text{if } x_{N+1} < 0, \text{ and } x_{N+1} + q^*_{N-1} \geq 0 \\ -b - c - \alpha \cdot b & \text{if } x_{N+1} + q^*_{N-1} < 0 \end{cases} \tag{45}$$

$$\frac{\partial v_{N+1}}{\partial z_{N+1}} = \begin{cases} 0 & \text{if } x_{N+1} + q^*_{N-1} \geq 0 \\ -c - \alpha \cdot b & \text{if } x_{N+1} + q^*_{N-1} < 0 \end{cases} \tag{46}$$

where z_{N+1} is q^*_N, the quantity to arrive at $N + 2$.

At N, for $x_N > 0$:

$$v_N(x_N, z_N) = \min_{q_N \geq 0}\{c \cdot q_N + b \cdot \int_{x_N}^{M} (y_1 - x_N)\phi(y)dy + \alpha \cdot \Phi(x_N) \cdot v_{N+1}(z_N, q_N) + \\ + \alpha \cdot \int_{x_N}^{M} v_{N+1}(x_N + z_N - y_1)\phi(y_1)dy_1\} \tag{47}$$

where z_N is q^*_{N-1}. Its derivative as a function of x_N is:

$$\frac{\partial v_N}{\partial x_N} = -b \cdot (1 - \Phi(x_N)) - \alpha \cdot b \cdot (1 - \Phi(x_N + q^*_N)) - \alpha \cdot (c + \alpha \cdot b) \cdot (1- \\ \Phi(x_N + z_N + q^*_N)) = -b \cdot (1 - \Phi(x_N)) - \alpha \cdot b \cdot (1 - \Phi(x_N + q^*_N)) - c \geq -b - c - \alpha \cdot b \tag{48}$$

where the last equality resulted from q^*_N being such that $c - \alpha \cdot (c + \alpha \cdot b) \cdot (1 - \Phi(x_N + z_N + q^*_N)) = 0$.

Similarly, the derivative as a function of z_N is found as follows:

$$\frac{\partial v_N}{\partial z_N} = -\alpha \cdot b \cdot (1 - \Phi(x_N + q^*_N)) - \alpha \cdot (c + \alpha \cdot b) \cdot (1 - \Phi(x_N + z_N + q^*_N)) = \\ = -\alpha \cdot b \cdot (1 - \Phi(x_N + q^*_N)) - c \geq -c - \alpha \cdot b \tag{49}$$

using again the property of q^*_N.

The induction hypothesis assumes that $\frac{\partial v_{t+i}}{\partial x_{t+i}} \geq -b-c-\alpha \cdot b$, and $\frac{\partial v_{t+i}}{\partial z_{t+i}} \geq -c-\alpha \cdot b$, $\forall i \in \{1, 2, ..., N+1-t\}$. The induction hypothesis can be re-written using the general definition of the partial derivatives of the value function from Lemma 6 and 7:

$$\frac{\partial v_{t+i}}{\partial x_{t+i}} = -b - c - \alpha \cdot b + \Phi(x_{t+i}) \cdot \left(b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+i+1}}{\partial z_{t+i+1}}_{\substack{x_{t+i+1}=z_{t+i} \\ \\ z_{t+i+1}=q^*_{t+i}}} \right) +$$
$$+ \int_{x_{t+i}}^{x_{t+i}+z_{t+i}} \Phi(x_{t+i} + z_{t+i} - y_i)\varphi(y_i)dy_i \cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+i+2}}{\partial x_{t+i+2}}_{\substack{x_{t+i+2}=q^*_{t+i} \\ \\ z_{t+i+2}=q^*_{t+i+1}}} \right) \geq$$
$$\geq -b - c - \alpha \cdot b \tag{50}$$

as: $b + \alpha \cdot (b + c + \alpha \cdot b) \geq b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+i+1}}{\partial z_{t+i+1}}_{\substack{x_{t+i+1}=z_{t+i} \\ \\ z_{t+i+1}=q^*_{t+i}}} \geq 0$

and $\alpha \cdot (b + \alpha \cdot (c + \alpha \cdot b)) \geq \alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+i+2}}{\partial x_{t+i+2}}_{\substack{x_{t+i+2}=q^*_{t+i} \\ \\ z_{t+i+2}=q^*_{t+i+1}}} \geq 0, \forall i \in \{1, 2, ..., N+1-t\}$

and we have the same for the other derivative $\frac{\partial v_t}{\partial z_t} \geq -c - \alpha \cdot b$.

We need to show that the partial derivatives at t preserve the same properties.

The idea of the proof follows the same steps as for the one period time lag. We proceed with showing the result for $\frac{\partial v_t}{\partial z_t}$ first, since it is the most general.

We need to show that the parenthesis in the derivative:

$$\frac{\partial v_t}{\partial z_t} = -c - \alpha \cdot b + \left(\Phi(x_t) \cdot \Phi(z_t) + \int_{x_t}^{x_t+z_t} \Phi(x_t + z_t - y_1)\varphi(y_1)dy_1 \right) \cdot$$
$$\cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}}_{\substack{x_{t+2}=q_t^* \\ \\ z_{t+2}=q_{t+1}^*}} \right) \tag{51}$$

is positive, using the induction hypothesis. We expand the term as in the following:

$$\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}}_{\substack{x_{t+2}=q_t^* \\ \\ z_{t+2}=q_{t+1}^*}} = \alpha \cdot b - \alpha^2 \cdot [-b - c - \alpha \cdot b + \Phi(q_t^*) \cdot$$
$$\left(b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+3}}{\partial z_{t+3}}_{\substack{x_{t+3}=q_{t+1}^* \\ \\ z_{t+3}=q_{t+2}^*}} \right) + \int_{q_t^*}^{q_t^*+q_{t+1}^*} \Phi(q_t^* + q_{t+1}^* - y_3)\varphi(y_3)dy_3 \cdot =$$
$$\left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+4}}{\partial x_{t+4}}_{\substack{x_{t+4}=q_{t+2}^* \\ \\ z_{t+4}=q_{t+3}^*}} \right) \Bigg] = \alpha \cdot b + \alpha^2 \cdot (b + c + \alpha \cdot b) - \alpha^2 \cdot \Phi(q_t^*) \cdot$$
$$\cdot \left(b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+3}}{\partial z_{t+3}}_{\substack{x_{t+3}=q_{t+1}^* \\ \\ z_{t+3}=q_{t+2}^*}} \right) - \alpha^2 \cdot \int_{q_t^*}^{q_t^*+q_{t+1}^*} \Phi(q_t^* + q_{t+1}^* - y_3)\varphi(y_3)dy_3 \cdot$$
$$\cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+4}}{\partial x_{t+4}}_{\substack{x_{t+4}=q_{t+2}^* \\ \\ z_{t+4}=q_{t+3}^*}} \right)$$

By factoring out $\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b))$ we obtain:

$$\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}}\Bigg|_{\substack{x_{t+2}=q_t^* \\ z_{t+2}=q_{t+1}^*}} = \alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b)) \cdot \Bigg(1 - \alpha^2 \cdot \Phi(q_t^*) \cdot \frac{b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+3}}{\partial z_{t+3}}\Big|_{\substack{x_{t+3}=q_{t+1}^* \\ z_{t+3}=q_{t+2}^*}}}{\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b))} - \alpha^2 \cdot \int_{q_t^*}^{q_t^*+q_{t+1}^*} \Phi(q_t^* + q_{t+1}^* - y_3)\varphi(y_3)dy_3 \cdot \frac{\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+4}}{\partial x_{t+4}}\Big|_{\substack{x_{t+4}=q_{t+2}^* \\ z_{t+4}=q_{t+3}^*}}}{\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b))}\Bigg)$$

To show the parenthesis above is positive, we need to show that

$$\alpha^2 \cdot \Phi(q_t^*) \cdot \frac{b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+3}}{\partial z_{t+3}}\Big|_{\substack{x_{t+3}=q_{t+1}^* \\ z_{t+3}=q_{t+2}^*}}}{\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b))} + \alpha^2 \cdot \int_{q_t^*}^{q_t^*+q_{t+1}^*} \Phi(q_t^* + q_{t+1}^* - y_3)\varphi(y_3)dy_3 \cdot \frac{\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+4}}{\partial x_{t+4}}\Big|_{\substack{x_{t+4}=q_{t+2}^* \\ z_{t+4}=q_{t+3}^*}}}{\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b))} \leq 1$$

By the induction hypothesis, we have:

$b + \alpha \cdot (b + c + \alpha \cdot b) \geq b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+3}}{\partial z_{t+3}}\Big|_{\substack{x_{t+3}=q_{t+1}^* \\ z_{t+3}=q_{t+2}^*}} \geq 0$ and

$$\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b)) \geq \alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+4}}{\partial x_{t+4}}_{\substack{x_{t+4}=q^*_{t+2} \\ \\ z_{t+4}=q^*_{t+3}}} \geq 0$$

so we can write:

$$\alpha^2 \cdot \Phi(q^*_t) \cdot \frac{b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+3}}{\partial z_{t+3}}_{\substack{x_{t+3}=q^*_{t+1} \\ \\ z_{t+3}=q^*_{t+2}}}}{\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b))} + \alpha^2 \cdot \int_{q^*_t}^{q^*_t + q^*_{t+1}} \Phi(q^*_t + q^*_{t+1} - y_3)\varphi(y_3)dy_3 \cdot$$

$$\cdot \frac{\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+4}}{\partial x_{t+4}}_{\substack{x_{t+4}=q^*_{t+2} \\ \\ z_{t+4}=q^*_{t+3}}}}{\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b))} \leq \alpha^2 \cdot \Phi(q^*_t) + \alpha^2 \cdot \int_{q^*_t}^{q^*_t + q^*_{t+1}} \Phi(q^*_t + q^*_{t+1} - y_3)\varphi(y_3)dy_3 \leq 1$$

This inequality follows from the intuitive argument of the mathematical expression as follows: it represents the sum of the probability that the demand three periods from the current period is less than the available capacity plus the probability that the same demand is between the available capacity and the capacity that will be available next period. This probability should obviously be less than or equal to 1. Then, the induction is completed for $\frac{\partial v_t}{z_t}$.

Then the induction is completed for $\frac{\partial v_t}{z_t}$. $\frac{\partial v_t}{x_t}$ follows the same logic:

$$\frac{\partial v_t}{\partial x_t} = -b - c - \alpha \cdot b + \Phi(x_t) \cdot \left(b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+1}}{\partial z_{t+1}}_{\substack{x_{t+1}=z_t \\ \\ z_{t+1}=q^*_t}} \right) + \\ + \int_{x_t}^{x_t + z_t} \Phi(x_t + z_t - y_1)\varphi(y_1)dy_1 \cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}}_{\substack{x_{t+2}=q^*_t \\ \\ z_{t+2}=q^*_{t+1}}} \right) \tag{52}$$

We have already shown that $\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}} \Big|_{\substack{x_{t+2}=q_t^* \\ z_{t+2}=q_{t+1}^*}} \geq 0$, we need to show that the remaining parenthesis is positive. As before, we expand the parenthesis as follows:

$$b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+1}}{\partial z_{t+1}} \Big|_{\substack{x_{t+1}=z_t \\ z_{t+1}=q_t^*}} = b + \alpha \cdot b - \alpha \cdot \Bigg[-c - \alpha \cdot b + \Big(\Phi(z_t) \cdot \Phi(q_t^*) +$$

$$+ \int_{z_t}^{z_t+q_t^*} \Phi(z_t + q_t^* - y_2) \varphi(y_2) dy_2 \Big) \cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+3}}{\partial x_{t+3}} \Big|_{\substack{x_{t+3}=q_{t+1}^* \\ z_{t+3}=q_{t+2}^*}} \right) \Bigg]$$

and factor out $b + \alpha \cdot (b + c + \alpha \cdot b)$, where we denote by TERM1 $= \Phi(z_t) \cdot \Phi(q_t^*) + \int_{z_t}^{z_t+q_t^*} \Phi(z_t + q_t^* - y_2) \varphi(y_2) dy_2$:

$$b + \alpha \cdot b - \alpha \cdot \frac{\partial v_{t+1}}{\partial z_{t+1}} \Big|_{\substack{x_{t+1}=z_t \\ z_{t+1}=q_t^*}} = (b + \alpha \cdot (b + c + \alpha \cdot b)) \cdot$$

$$\cdot \left[1 - \frac{\alpha \cdot \text{TERM1} \cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+3}}{\partial x_{t+3}} \Big|_{\substack{x_{t+3}=q_{t+1}^* \\ z_{t+3}=q_{t+2}^*}} \right)}{b + \alpha \cdot (b + c + \alpha \cdot b)} \right]$$

TERM1 is the probability of ending period $t+2$ with no backlog, if we already placed the order z_t at $t-1$, and place the order q_t^* at t. Hence, it is less than or equal to 1. Again:

$$\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b)) \geq \alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+3}}{\partial x_{t+3}} \Bigg|_{\substack{x_{t+3}=q_{t+1}^* \\ z_{t+3}=q_{t+2}^*}} \geq 0 \tag{53}$$

by induction hypothesis. Theorem 5 is proven.

■

The result from Theorem 5 will be used in proving the positive definiteness of the Hessian, i.e., convexity of the value function in the two state variables. The following result prepares the proof of stationarity for the optimal order quantity, which will be eventually used for proving convexity as well.

Lemma 8 *The following equation holds at any time t:*

$$\begin{aligned}
&\Phi(x_t) \cdot \left(\Phi(z_t) \cdot \Phi(q_t^*) + \int_{z_t}^{z_t+q_t^*} \Phi(z_t + q_t^* - y_2)\varphi(y_2)dy_2 \right) + \\
&\int_{x_t}^{M} \left(\Phi(x_t + z_t - y_1) \cdot \Phi(q_t^*) + \int_{x_t+z_t-y_1}^{x_t+z_t-y_1+q_t^*} \Phi(x_t + z_t - y_1 + q_t^* - y_2)\varphi(y_2)dy_2 \right) \\
&\varphi(y_1)dy_1 = \frac{\alpha \cdot (c + \alpha \cdot b) - c}{\alpha \cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+3}}{\partial x_{t+3}} \Bigg|_{\substack{x_{t+3}=q_{t+1}^* \\ z_{t+3}=q_{t+2}^*}} \right)}
\end{aligned} \tag{54}$$

Proof of Lemma 8: By plugging in equation (39) into (40), we obtain:

$$c + \alpha \cdot \Phi(x_t) \cdot \Bigg[-c - \alpha \cdot b + \left(\Phi(z_t) \cdot \Phi(q_t^*) + \int_{z_t}^{z_t + q_t^*} \Phi(z_t + q_t^* - y_1) \varphi(y_2) dy_2 \right) \cdot$$

$$\left(\alpha \cdot b - \alpha^2 \cdot \left. \frac{\partial v_{t+3}}{\partial x_{t+3}} \right|_{\substack{x_{t+3} = q_{t+1}^* \\ z_{t+3} = q_{t+2}^*}} \right) \Bigg) \Bigg] + \alpha \cdot \int_{x_t}^{M} [-c - \alpha \cdot b + (\Phi(x_t + z_t - y_1) \cdot \Phi(q_t^*) +$$

$$+ \int_{x_t + z_t - y_1}^{x_t + z_t - y_1 + q_t^*} \Phi(x_t + z_t - y_1 + q_t^* - y_2) \varphi(y_2) dy_2 \Bigg) \cdot \left(\alpha \cdot b - \alpha^2 \cdot \left. \frac{\partial v_{t+3}}{\partial x_{t+3}} \right|_{\substack{x_{t+3} = q_{t+1}^* \\ z_{t+3} = q_{t+2}^*}} \right) \Bigg) \Bigg]$$

$$\varphi(y_1) dy_1 = c - \alpha \cdot (c + \alpha \cdot b) + \alpha \cdot \left(\alpha \cdot b - \alpha^2 \cdot \left. \frac{\partial v_{t+3}}{\partial x_{t+3}} \right|_{\substack{x_{t+3} = q_{t+1}^* \\ z_{t+3} = q_{t+2}^*}} \right) \cdot$$

$$\Bigg[\Phi(x_t) \cdot \Bigg(\Phi(z_t) \cdot \Phi(q_t^*) + \int_{z_t}^{z_t + q_t^*} \Phi(z_t + q_t^* - y_2) \varphi(y_2) dy_2 \ ) + \int_{x_t}^{M} (\Phi(x_t + z_t - y_1) \cdot$$

$$\Phi(q_t^*) + \int_{x_t + z_t - y_1}^{x_t + z_t - y_1 + q_t^*} \Phi(x_t + z_t - y_1 + q_t^* - y_2) \varphi(y_2) dy_2 \Bigg) \varphi(y_1) dy_1 \Bigg] = 0$$

from which the equation (54) follows.

■

We show subsequently that the optimal order quantity is a stationary policy.

Theorem 6 *The optimal policy is a stationary policy and solves the following equation:*

$$
\begin{aligned}
&\Phi(x)\cdot\left(\Phi(z)\cdot\Phi(q^*)+\int_{z}^{z+q^*}\Phi(z+q^*-y_2)\varphi(y_2)dy_2\right)+\\
&+\int_{x}^{M}\left(\Phi(x+z-y_1)\cdot\Phi(q^*)+\int_{x+z-y_1}^{x+z-y_1+q^*}\Phi(x+z-y_1+q^*-y_2)\varphi(y_2)dy_2\right)\varphi(y_1)dy_1=\\
&=\frac{\alpha\cdot(c+\alpha\cdot b)-c}{(\alpha^3\cdot(b+c+\alpha\cdot b)+\alpha^2\cdot b)\cdot(1-\mathcal{A}_{x,z})}
\end{aligned}
\tag{55}
$$

where x is the current available capacity/backlog, z is the capacity ordered one period prior to the current period, q^* is the capacity ordered during the current period, and $1-\mathcal{A}_{x,z})$ is the limit of the expanded term $\alpha\cdot b-\alpha^2\cdot\frac{\partial v_{t+3}}{\partial x_{t+3}}\Big|_{\substack{x_{t+3}=q^*_{t+1}\\ z_{t+3}=q^*_{t+2}}}$.

We again have an equation similar to the solution of a newsvendor problem. If the current period is t, the order quantity q^*_t is placed such that the probability of ending the period in which the order arrives (period $t+2$) with no backlog is equal to the righthandside term in (55).

Proof: The proof reduces to showing that the expanded term $\alpha\cdot b-\alpha^2\cdot\frac{\partial v_{t+3}}{\partial x_{t+3}}\Big|_{\substack{x_{t+3}=q^*_{t+1}\\ z_{t+3}=q^*_{t+2}}}$ is convergent.

We have:

$$\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+3}}{\partial x_{t+3}}_{\substack{x_{t+3}=q^*_{t+1} \\ \\ z_{t+3}=q^*_{t+2}}} = \alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b)) \cdot (1-$$

$$-\alpha^2 \cdot \Phi(q_{t+1}) \cdot \frac{b + \alpha \cdot b - \frac{\partial v_{t+4}}{\partial z_{t+4}}_{\substack{x_{t+4}=q^*_{t+2} \\ \\ z_{t+4}=q^*_{t+3}}}}{\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b))} -$$

$$\left. -\alpha^2 \cdot \int_{q^*_{t+1}}^{q^*_{t+1}+q^*_{t+2}} \Phi(q^*_{t+1} + q^*_{t+2} - y_4)\varphi(y_4)dy_4 \cdot \frac{\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+5}}{\partial x_{t+5}}_{\substack{x_{t+5}=q^*_{t+3} \\ \\ z_{t+5}=q^*_{t+4}}}}{\alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b))} \right)$$

For ease of exposition, we suppress the t index, and denote the constant multipliers by $\mathcal{T}$ (when they refer to the cumulative function) or $\mathcal{INT}$ (when they refer to the integral term). The constant multipliers are all between 0 and 1, being probabilities. We denote the terms in the expanded parenthesis by $\mathcal{G}$. The expression above becomes:

$$\mathcal{G}_(3) = \alpha\cdot(b+\alpha\cdot(b+c+\alpha\cdot b))\cdot\left(1-\mathcal{T}_1^1\cdot\frac{b+\alpha\cdot b-\frac{\partial v_4}{\partial z_4}_{\substack{x_4=q_2^*\\ z_4=q_3^*}}}{\alpha\cdot(b+\alpha\cdot(b+c+\alpha\cdot b))}-\right.$$

$$\left.-\mathcal{INT}_1^1\cdot\frac{\alpha\cdot b-\alpha^2\cdot\frac{\partial v_5}{\partial x_5}_{\substack{x_5=q_3^*\\ z_5=q_4^*}}}{\alpha\cdot(b+\alpha\cdot(b+c+\alpha\cdot b))}\right) = \alpha\cdot(b+\alpha\cdot(b+c+\alpha\cdot b))\cdot\Big[1-\mathcal{T}_1^1\cdot$$

$$\cdot\left(1-\frac{\mathcal{T}_2\cdot\left(\alpha\cdot b-\alpha^2\cdot\frac{\partial v_6}{\partial x_6}_{\substack{x_5=q_4^*\\ z_5=q_5^*}}\right)}{\alpha\cdot(b+\alpha\cdot(b+c+\alpha\cdot b))}\right)-\mathcal{INT}_1^1\cdot\frac{\alpha\cdot b-\alpha^2\cdot\frac{\partial v_5}{\partial x_5}_{\substack{x_5=q_3^*\\ z_5=q_4^*}}}{\alpha\cdot(b+\alpha\cdot(b+c+\alpha\cdot b))}\Bigg] =$$

$$= \alpha\cdot(b+\alpha\cdot(b+c+\alpha\cdot b))\cdot\big(1-\mathcal{T}_1^1+\mathcal{T}_1^1\cdot\mathcal{T}_2\cdot\mathcal{G}_(6)-\mathcal{INT}_1^1\cdot\mathcal{G}_(5)\big)$$

where we use the following notations (and suppress the index t):

$$\mathcal{T}_1^i = \alpha^2\cdot\Phi(q_{t+i}^*)$$

$$\mathcal{INT}_1^i = \int_{q_{t+i}^*}^{q_{t+i}^*+q_{t+i+1}^*}\Phi(q_{t+i}^*+q_{t+i+1}^*-y_{t+i+3})\varphi(y_{t+i+3})dy_{t+i+3}$$

$$\mathcal{T}_i = \Phi(q_{t+i}^*)\cdot\Phi(q_{t+i+1}^*)+\int_{q_{t+i}^*}^{q_{t+i}^*+q_{t+i+1}^*}\Phi(q_{t+i}^*+q_{t+i+1}^*-y_{t+i+3})\varphi(y_{t+i+3})dy_{t+i+3}$$

$$\mathcal{G}_{(i)} = \alpha\cdot b-\alpha^2\cdot\frac{\partial v_6}{\partial x_6}_{\substack{x_i=q_{i-1}^*\\ z_i=q_i^*}}$$

Using the notations above and expanding further, we deduce:

$$
\begin{aligned}
&\mathcal{G}_{(}3) = \alpha \cdot (b + \alpha \cdot (b + c + \alpha \cdot b)) \cdot (1 - \mathcal{T}_1^1 + \mathcal{T}_1^1 \cdot \mathcal{T}_2 \cdot (1 - \mathcal{T}_1^2 + \mathcal{T}_1^2 \cdot \mathcal{T}_3 \cdot \mathcal{G}_{(}9) - \\
&- \mathcal{INT}_1^2 \cdot \mathcal{G}_{(}8)) - \mathcal{INT}_1^1 \cdot (1 - \mathcal{T}_1^3 + \mathcal{T}_1^3 \cdot \mathcal{T}_4 \cdot \mathcal{G}_{(}8) - \mathcal{INT}_1^3 \cdot \mathcal{G}_{(}7))) = \alpha \cdot (b + \alpha \cdot (b + c + \\
&+ \alpha \cdot b)) \cdot (1 - \mathcal{T}_1^1 + \mathcal{T}_1^1 \cdot \mathcal{T}_2 - \mathcal{T}_1^1 \cdot \mathcal{T}_2 \cdot \mathcal{T}_1^2 + \mathcal{T}_1^1 \cdot \mathcal{T}_2 \cdot \mathcal{T}_1^2 \cdot \mathcal{T}_3 \cdot \mathcal{G}_{(}9) - \mathcal{T}_1^1 \cdot \mathcal{T}_2 \cdot \mathcal{INT}_1^2 \cdot \mathcal{G}_{(}8) - \\
&- \mathcal{INT}_1^1 + \mathcal{INT}_1^1 \cdot \mathcal{T}_1^3 - \mathcal{INT}_1^1 \cdot \mathcal{T}_1^3 \cdot \mathcal{T}_4 \cdot \mathcal{G}(8) + \mathcal{INT}_1^1 \cdot \mathcal{INT}_1^3 \cdot \mathcal{G}(7)) = \alpha \cdot (b + \alpha \cdot (b + \\
&+ c + \alpha \cdot b)) \cdot (1 - \mathcal{T}_1^1 + \mathcal{T}_1^1 \cdot \mathcal{T}_2 - \mathcal{T}_1^1 \cdot \mathcal{T}_2 \cdot \mathcal{T}_1^2 + \mathcal{T}_1^1 \cdot \mathcal{T}_2 \cdot \mathcal{T}_1^2 \cdot \mathcal{T}_3 \cdot \mathcal{G}_{(}9) - (\mathcal{T}_1^1 \cdot \mathcal{T}_2 \cdot \mathcal{INT}_1^2 + \\
&+ \mathcal{INT}_1^1 \cdot \mathcal{T}_1^3 \cdot \mathcal{T}_4) \cdot \mathcal{G}(8) - \mathcal{INT}_1^1 + \mathcal{INT}_1^1 \cdot \mathcal{T}_1^3 + \mathcal{INT}_1^1 \cdot \mathcal{INT}_1^3 \cdot \mathcal{G}(7))
\end{aligned}
\tag{56}
$$

The terms multiplied to the expanded parenthesis are probabilities and they will go to zero after a few more periods, depending on the distribution, leaving only constants, which define the limit of the term. As for $L = 1$, we know the limit has to be between 0 and 1, due to the bounds of the expanded parenthesis.

■

We still need to show that the order quantity that solves the equation (55) is indeed the optimal quantity by proving convexity of the value function. Before getting to the convexity proof, we need one more result.

Theorem 7 *The optimal policy $q_t^*(x_t, z_t)$ has the following properties:*

1. $q_t^*(x_t, z_t)$ is a continuous function of both variables, x_t and z_t;

2. $0 < \lim\limits_{x_t \to \infty} q_t^*(x_t, z_t)$, and $0 < \lim\limits_{z_t \to \infty} q_t^*(x_t, z_t)$;

3. $-1 \leq \frac{\partial q_t^*(x_t, z_t)}{\partial x_t} \leq 0$, and $-1 \leq \frac{\partial q_t^*(x_t, z_t)}{\partial z_t} \leq 0$, $\forall x_t, z_t$.

As for $L = 1$, the optimal quantity ordered at any stage t is a decreasing function of the state variables; it is never optimal to order no capacity, no matter how large the current available capacity or the capacity to arrive in the next period are.

Proof: The continuity from point 1 results again from the properties of the distribution function and the stationary equation (55), and the limits are deduced in the exact same manner as for $L = 1$. The last point can be deduced by taking the partial derivative of the stationary equation (55) as a function of x_t:

$$
\begin{aligned}
&\varphi(x_t)\cdot\left(\Phi(z_t)\cdot\Phi(q_t^*)+\int_{z_t}^{z_t+q_t^*}\Phi(z_t+q_t^*-y_2)\varphi(y_2)dy_2\right)+\\
&+\Phi(x_t)\cdot\left(\Phi(z_t)\cdot\varphi(q_t^*)\cdot\frac{\partial q_t^*}{\partial x_t}+\int_{z_t}^{z_t+q_t^*}\varphi(z_t+q_t^*-y_2)\cdot\frac{\partial q_t^*}{\partial x_t}\varphi(y_2)dy_2\right)+\\
&+\int_{x_t}^{M}\bigg[\varphi(x_t+z_t-y_1)\cdot\Phi(q_t^*)+\Phi(x_t+z_t-y_1)\cdot\varphi(q_t^*)\cdot\frac{\partial q_t^*}{\partial x_t}+\\
&+\int_{x_t+z_t-y_1}^{x_t+z_t-y_1+q_t^*}\varphi(x_t+z_t-y_1+q_t^*-y_2)\cdot\left(1+\frac{\partial q_t^*}{\partial x_t}\right)\varphi(y_2)dy_2-\varphi(x_t+z_t-y_1)\cdot\Phi(q_t^*)\bigg]\\
&\varphi(y_1)dy_1-\varphi(x_t)\cdot\left(\Phi(z_t)\cdot\Phi(q_t^*)+\int_{z_t}^{z_t+q_t^*}\Phi(z_t+q_t^*-y_2)\varphi(y_2)dy_2\right)=\\
&=\bigg[\Phi(x_t)\cdot\left(\Phi(z_t)\cdot\varphi(q_t^*)+\int_{z_t}^{z_t+q_t^*}\varphi(z_t+q_t^*-y_2)\varphi(y_2)dy_2\right)\\
&+\int_{x_t}^{M}\Phi(x_t+z_t-y_1)\cdot\varphi(q_t^*)\varphi(y_1)dy_1\bigg]\cdot\frac{\partial q_t^*}{\partial x_t}+\int_{x_t}^{M}\int_{x_t+z_t-y_1}^{x_t+z_t-y_1+q_t^*}\varphi(x_t+z_t-y_1+q_t^*-y_2)\cdot\\
&\left(1+\frac{\partial q_t^*}{\partial x_t}\right)\varphi(y_2)dy_2\varphi(y_1)dy_1=0
\end{aligned}
$$

which can be fulfilled only if $\frac{\partial q_t^*}{\partial x_t} < 0$, and $1+\frac{\partial q_t^*}{\partial x_t} \geq 0$, since all other terms are positive probabilities. Analogously, taking the derivative of (55) as a function of z_t, we obtain:

$$\Phi(x_t) \cdot \Bigg(\varphi(z_t) \cdot \Phi(q_t^*) + \Phi(z_t) \cdot \varphi(q_t^*) \cdot \frac{\partial q_t^*}{\partial z_t} + \int_{z_t}^{z_t+q_t^*} \varphi(z_t + q_t^* - y_2) \cdot \left(1 + \frac{\partial q_t^*}{\partial z_t}\right) \varphi(y_2) dy_2 -$$

$$-\varphi(z_t) \cdot \Phi(q_t^*)) + \int_{x_t}^{M} \Bigg[\varphi(x_t + z_t - y_1) \cdot \Phi(q_t^*) + \Phi(x_t + z_t - y_1) \cdot \varphi(q_t^*) \cdot \frac{\partial q_t^*}{\partial z_t} +$$

$$+ \int_{x_t+z_t-y_1}^{x_t+z_t-y_1+q_t^*} \varphi(x_t + z_t - y_1 + q_t^* - y_2) \cdot \left(1 + \frac{\partial q_t^*}{\partial z_t}\right) \varphi(y_2) dy_2 - \varphi(x_t + z_t - y_1) \cdot \Phi(q_t^*)\Bigg]$$

$$\varphi(y_1) dy_1 = \left(\Phi(x_t) \cdot \Phi(z_t) \cdot \varphi(q_t^* + \int_{x_t}^{M} \Phi(x_t + z_t - y_1) \cdot \varphi(q_t^*) \varphi(y_1) dy_1\right) \cdot \frac{\partial q_t^*}{\partial z_t} +$$

$$+ \Bigg(\Phi(x_t) \cdot \int_{z_t}^{z_t+q_t^*} \varphi(z_t + q_t^* - y_2) \varphi(y_2) dy_2 +$$

$$+ \int_{x_t}^{M} \int_{x_t+z_t-y_1}^{x_t+z_t-y_1+q_t^*} \varphi(x_t + z_t - y_1 + q_t^* - y_2) \varphi(y_2) dy_2 \varphi(y_1) dy_1\Bigg) \cdot \left(1 + \frac{\partial q_t^*}{\partial z_t}\right) = 0$$

and by the same argument as above, we have $\frac{\partial q_t^*}{\partial z_t} < 0$, and $1 + \frac{\partial q_t^*}{\partial z_t} \geq 0$.

■

Now we have all the results necessary to show convexity.

Theorem 8 *The Hessian of the value function is positive semidefinite.*

Proof: The convexity of the value function in the state variables can be shown using Sylvester's criterion, which states that a matrix is positive definite iff the determinants associated with all upper-left submatrices are positive. In our case we need to show $\frac{\partial^2 v_t}{\partial x_t^2} \geq 0$, and $\det H \geq 0$, where we denote by H the Hessian of the value function.

We note first that the Hessian is symmetric due to continuity of the value function's partial derivatives (Taylor theorem). We will use an inductive argument:

Assuming that $\frac{\partial^2 v_{t+1}}{\partial z_{t+1}^2} > \frac{\partial^2 v_{t+1}}{\partial z_{t+1} \partial x_{t+1}} > 0$, and $\frac{\partial^2 v_{t+1}}{\partial x_{t+1}^2} > \frac{\partial^2 v_{t+1}}{\partial z_{t+1} \partial x_{t+1}} > 0$, we show that the same holds at t.

We derivate (43) as a function of z_t:

$$\frac{\partial^2 v_t}{\partial z_t^2} = \frac{\partial^2 v_t}{\partial x_t \partial z_t} + \alpha \cdot \Phi(x_t) \cdot \left(\frac{\partial^2 v_{t+1}}{\partial^2 x_{t+1}}{}_{\substack{x_{t+1}=z_t \\ \\ z_{t+1}=q_t^*}} + \frac{\partial^2 v_{t+1}}{\partial x_{t+1} z_{t+1}}{}_{\substack{x_{t+1}=z_t \\ \\ z_{t+1}=q_t^*}} \cdot \frac{\partial q_t^*}{\partial z_t} \right) \tag{57}$$

but since $\frac{\partial^2 v_{t+1}}{\partial x_{t+1}^2} > \frac{\partial^2 v_{t+1}}{\partial z_{t+1} \partial x_{t+1}} > 0$ by the induction argument, we can deduce $\frac{\partial^2 v_{t+1}}{\partial x_{t+1}^2}{}_{\substack{x_{t+1}=z_t \\ \\ z_{t+1}=q_t^*}} + \frac{\partial^2 v_{t+1}}{\partial x_{t+1} \partial z_{t+1}}{}_{\substack{x_{t+1}=z_t \\ \\ z_{t+1}=q_t^*}} \cdot \frac{\partial q_t^*}{\partial z_t} > \frac{\partial^2 v_{t+1}}{\partial x_{t+1} \partial z_{t+1}}{}_{\substack{x_{t+1}=z_t \\ \\ z_{t+1}=q_t^*}} + \frac{\partial^2 v_{t+1}}{\partial x_{t+1} \partial z_{t+1}}{}_{\substack{x_{t+1}=z_t \\ \\ z_{t+1}=q_t^*}} \cdot \frac{\partial q_t^*}{\partial z_t} = \frac{\partial^2 v_{t+1}}{\partial x_{t+1} \partial z_{t+1}}{}_{\substack{x_{t+1}=z_t \\ \\ z_{t+1}=q_t^*}} \cdot \left(1 + \frac{\partial q_t^*}{\partial z_t}\right) > 0$ by the bounds on the derivatives of the optimal order quantity $-1 \leq \frac{\partial q_t^*}{\partial z_t} \leq 0$.

Now, we can show that $\frac{\partial^2 v_t}{\partial z_t^2} \geq 0$ and $\frac{\partial^2 v_t}{\partial x_t \partial z_t} \geq 0$ by derivating $\frac{\partial v_t}{\partial z_t}$ from (39) as a function of z_t:

$$\frac{\partial^2 v_t}{\partial z_t^2} = \frac{\partial \mathcal{P}}{\partial z_t} \cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}}{}_{\substack{x_{t+2}=q_t^* \\ \\ z_{t+2}=q_{t+1}^*}} \right) - \mathcal{P} \cdot \alpha^2 \cdot \frac{\partial^2 v_{t+2}}{\partial x_{t+2}^2}{}_{\substack{x_{t+2}=q_t^* \\ \\ z_{t+2}=q_{t+1}^*}} \cdot \frac{\partial q_t^*}{\partial z_t}$$

and as a function of x_t :

$$\frac{\partial^2 v_t}{\partial z_t \partial x_t} = \frac{\partial \mathcal{P}}{\partial x_t} \cdot \left(\alpha \cdot b - \alpha^2 \cdot \frac{\partial v_{t+2}}{\partial x_{t+2}}{}_{\substack{x_{t+2}=q_t^* \\ \\ z_{t+2}=q_{t+1}^*}} \right) - \mathcal{P} \cdot \alpha^2 \cdot \frac{\partial^2 v_{t+2}}{\partial x_{t+2}^2}{}_{\substack{x_{t+2}=q_t^* \\ \\ z_{t+2}=q_{t+1}^*}} \cdot \frac{\partial q_t^*}{\partial x_t}$$

In both expressions, the first term is positive by Theorem 5, the second by Theorem 7. We put all the results together: $\frac{\partial^2 v_t}{\partial^2 z_t} \geq 0$, $\frac{\partial^2 v_t}{\partial^2 z_t \partial x_t} \geq 0$, $\frac{\partial^2 v_t}{\partial^2 z_t} = \frac{\partial^2 v_t}{\partial x_t \partial z_t} + \Lambda$, where $\Lambda = \alpha \cdot \Phi(x_t) \cdot \left(\frac{\partial^2 v_{t+1}}{\partial^2 x_{t+1}} \Big|_{\substack{x_{t+1}=z_t \\ z_{t+1}=q_t^*}} + \frac{\partial^2 v_{t+1}}{\partial x_{t+1} z_{t+1}} \Big|_{\substack{x_{t+1}=z_t \\ z_{t+1}=q_t^*}} \cdot \frac{\partial q_t^*}{\partial z_t} \right) \geq 0$. We can conclude $\frac{\partial^2 v_t}{\partial^2 z_t} \geq \frac{\partial^2 v_t}{\partial x_t \partial z_t} \geq 0$.

Similarly, we derivate (43) as a function of x_t:

$$\frac{\partial^2 v_t}{\partial x_t^2} = \frac{\partial^2 v_t}{\partial x_t \partial z_t} + b \cdot \varphi(x_t) - \alpha \cdot \varphi(x_t) \cdot \frac{\partial v_{t+1}}{\partial x_{t+1}} - \alpha \cdot \Phi \cdot \frac{\partial^2 v_{t+1}}{\partial x_{t+1} \partial z_{t+1}} \cdot \frac{\partial q_t^*}{\partial x_t} \tag{58}$$

and derivate (42) as a function of x_t also:

$$\frac{\partial^2 v_t}{\partial x_t^2} = b \cdot \varphi(x_t) + \alpha \cdot \int_{x_t}^{M} \left(\frac{\partial^2 v_{t+1}}{\partial x_{t+1}^2} + \frac{\partial^2 v_{t+1}}{\partial x_{t+1} \partial z_{t+1}} \cdot \frac{\partial q_t^*}{\partial x_t} \right) \varphi(y) dy \tag{59}$$

Now, matching the terms in the two equations above, we obtain:

$$\frac{\partial^2 v_t}{\partial x_t \partial z_t} - \alpha \cdot \varphi(x_t) \cdot \frac{\partial v_{t+1}}{\partial x_{t+1}} - \alpha \cdot \Phi \cdot \frac{\partial^2 v_{t+1}}{\partial x_{t+1} \partial z_{t+1}} \cdot \frac{\partial q_t^*}{\partial x_t} = \alpha \cdot \int_{x_t}^{M} \left(\frac{\partial^2 v_{t+1}}{\partial x_{t+1}^2} + \frac{\partial^2 v_{t+1}}{\partial x_{t+1} \partial z_{t+1}} \cdot \right. \\ \left. \cdot \frac{\partial q_t^*}{\partial x_t} \right) \varphi(y) dy \geq 0$$

where the term on the right can be shown to be positive by the exact same argument as for equation (57). Since we already know that $\frac{\partial^2 v_t}{\partial^2 z_t \partial x_t} \geq 0$, we can conclude $\frac{\partial^2 v_t}{\partial^2 x_t} \geq \frac{\partial^2 v_t}{\partial x_t \partial z_t} \geq 0$.

Using the relationships, we can show that $\det H = \frac{\partial^2 v_t}{\partial^2 x_t} \cdot \frac{\partial^2 v_t}{\partial^2 z_t} - \left(\frac{\partial^2 v_t}{\partial x_t \partial z_t} \right)^2 \geq 0$.

To complete the proof, we present the derivatives of the value function at the end of the horizon:

At time N we have:

$$
\begin{aligned}
\frac{\partial v_N}{\partial x_N} &= -b\cdot(1-\Phi(x_N)) + \alpha\cdot b\cdot\Phi(x_N+z_N) - (c+\alpha\cdot b)\\
\frac{\partial v_N}{\partial z_N} &= -(c+\alpha\cdot b) + \alpha\cdot b\cdot\Phi(x_N+z_N)\\
\frac{\partial^2 v_N}{\partial x_N^2} &= b\cdot\varphi(x_N)) + \alpha\cdot b\cdot\varphi(x_N+z_N) \geq 0\\
\frac{\partial^2 v_N}{\partial z_N^2} &= \alpha\cdot b\cdot\varphi(x_N+z_N) \geq 0\\
\frac{\partial^2 v_N}{\partial x_N \partial z_N} &= \frac{\partial^2 v_N}{\partial z_N \partial x_N} = \alpha\cdot b\cdot\varphi(x_N+z_N) \geq 0
\end{aligned}
$$

and the induction assumption holds.

At $N-1$ we have:

$$
\begin{aligned}
\frac{\partial v_{N-1}}{\partial z_{N-1}} &= -c - \alpha\cdot b + +\alpha\cdot b\cdot\Big(\Phi(x_{N-1})\cdot\Phi(z_{N-1}) +\\
&+ \int_{x_{N-1}}^{x_{N-1}+z_{N-1}} \Phi(x_{N-1}+z_{N-1}-y_1)\varphi(y_1)dy_1\Big)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial v_{N-1}}{\partial x_{N-1}} &= -b - c - \alpha\cdot b + \Phi(x_{N-1})\cdot\big[\alpha\cdot(b+c+\alpha\cdot b) + b\cdot(1-\\
&-\alpha^2\cdot\Phi(z_{N-1}+q^{N-1*}))\big] + \alpha\cdot b\cdot\int_{x_{N-1}}^{x_{N-1}+z_{N-1}} \Phi(x_{N-1}+z_{N-1}-y_1)\varphi(y_1)dy_1
\end{aligned}
$$

The terms of the Hessian at $N-1$ are:

$$
\begin{aligned}
\frac{\partial^2 v_{N-1}}{\partial x_{N-1}^2} &= b\cdot\varphi(x_{N-1})\cdot(1-\alpha\cdot\Phi(z_{N-1}) - \alpha^2\cdot\Phi(z_{N-1}+q^*_{N-1})) +\\
&+ \alpha\cdot(b+c+\alpha\cdot b)\cdot\varphi(x_{N-1}) + \alpha\cdot b\cdot\int_{x_{N-1}}^{M} \Phi(x_{N-1}+z_{N-1}-y_1)\varphi(y_1)dy_1 -\\
&- \alpha^2\cdot b\cdot\Phi(x_{N-1})\cdot\varphi(z_{N-1}+q^*_{N-1})\cdot\frac{\partial q^*_{N-1}}{\partial x_{N-1}} \geq 0\\
\frac{\partial^2 v_{N-1}}{\partial z_{N-1}^2} &= \alpha\cdot b\cdot\Phi(x_{N-1})\cdot\varphi(z_{N-1}) + \alpha\cdot b\cdot\int_{x_{N-1}}^{M} \Phi(x_{N-1}+z_{N-1}-y_1)\varphi(y_1)dy_1 \geq 0\\
\frac{\partial^2 v_{N-1}}{\partial x_{N-1}\partial z_{N-1}} &= \alpha\cdot b\cdot\int_{x_{N-1}}^{M} \Phi(x_{N-1}+z_{N-1}-y_1)\varphi(y_1)dy_1 \geq 0
\end{aligned}
$$

and the induction assumption holds.

■

3.5 Numerical experiments for one period time lag

In this section, we analyze the behavior of the optimal order quantity for different discount factors, cost ratios, support and skewness for two distributions: the uniform and the triangular distributions, for one period time lag. We are interested in the speed of convergence of the optimal order quantity and its dependence on the above mentioned factors.

We use Matlab to implement the dynamic program, using a numerical approximation for the integrals. In order to keep the running times reasonable, we had to use a relatively large (discrete) grid for the search space (0.1). This resulted into a relatively coarse optimal order quantity, as can be seen in the subsequent graphs.

In the first experiment we vary the discount factor, keeping the same support for both distributions (the interval [1 5]), with the triangular distribution being symmetric (with mode $C = 3$). The backlogging cost is $b = 2$, and the ordering cost is $c = 3$. When we vary the discount factor between 0.9 and 1, keeping the other variables constant, we observe a lower optimal order quantity for a lower discount factor for both distributions, with a slightly higher impact of the discount factor variation on the uniform distribution. The discount factor variation is too small to have an effect on the convergence speed; both distributions stabilize 4 periods after the end of horizon (counting backwards from the end of horizon towards period number 1) for each of the discount factor.

If we keep the discount factor constant ($\alpha = 1$) and we vary the cost ratio ($\frac{b}{c}$), keeping all the other factors as in the previous experiment, the optimal order quantity is higher for a higher backlogging cost, and lower for higher ordering costs, as we can see in Figure 3.5. The convergence speed does not depend on the cost ratio, as we would expect from the limiting equation (32). The convergence speed should be influenced by parameters of the demand distribution and not by cost parameters.

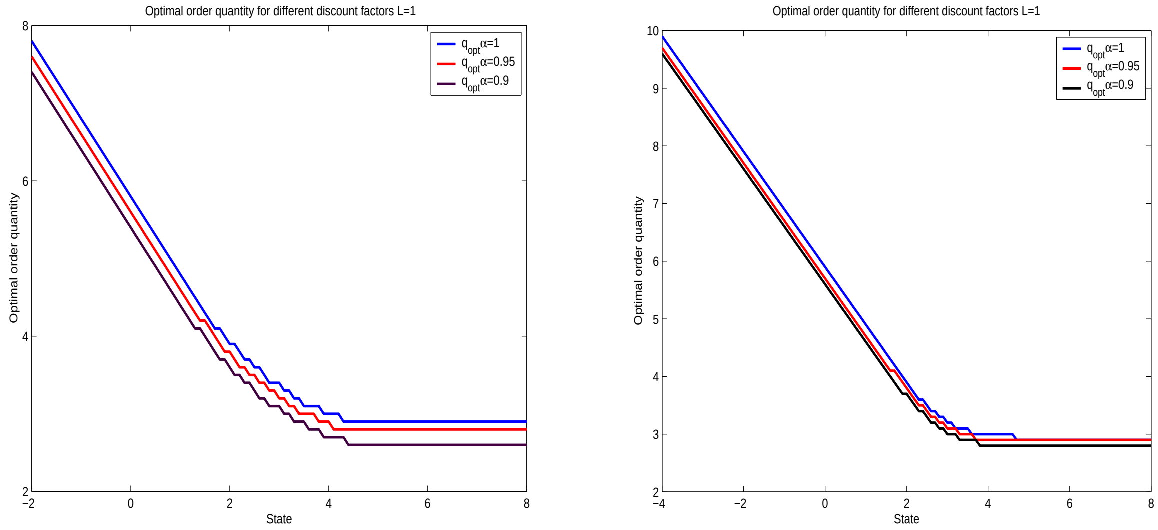

Figure 10: Dependency of the optimal order quantity on the discount factor: uniform (left) and triangular symmetric (right) distributions

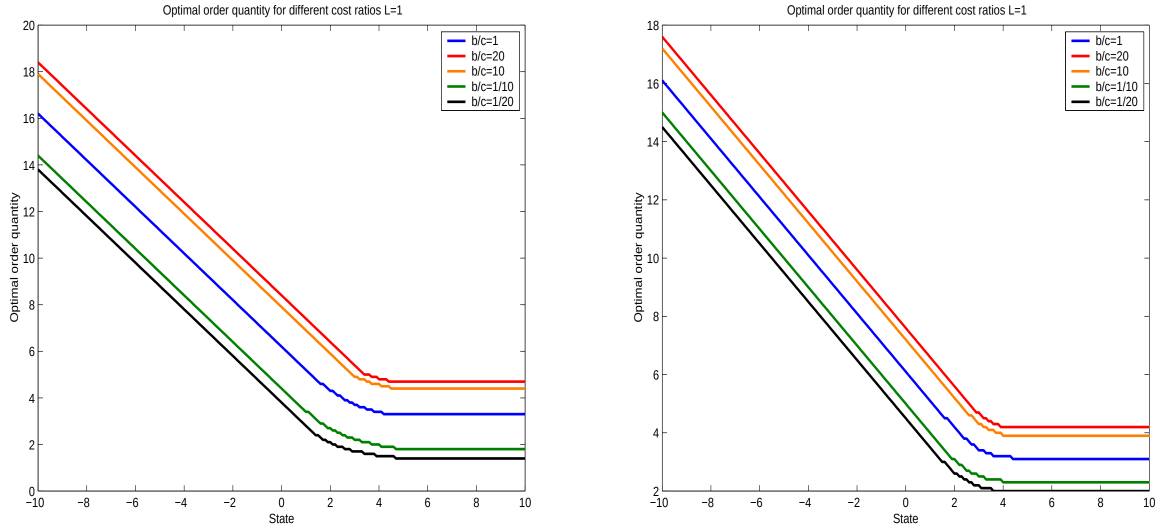

Figure 11: Dependency of the optimal order quantity on the cost ratio ($\frac{b}{c}$): uniform (left) and triangular symmetric (right) distributions

In both experiments, the optimal order quantity is higher for the triangular distribution than for the uniform. The distributions have the same mean (equal to 3), but different variances. The triangular distribution has half the variance of the uniform distribution on the same support. We can see that the optimal order quantity is more "abrupt" for the triangular distribution than for the uniform in all settings; this, together with the higher magnitude, are the effects of the smaller variance. In other words, the optimal order quantity for the triangular distribution has to address the same "uncertainty" on a tighter range, and it consequently places a slightly higher order.

If we vary the skewness of the triangular distribution, keeping the support and the cost ratio unchanged, between [1 7] and $\frac{2}{3}$ respectively, we observe a higher optimal order quantity the more the distribution is skewed to the right. The more the distribution is skewed to the right, the higher the probability to receive higher demand during the period, which translates into a higher optimal order quantity.

The last experiments that we conducted were to assert the dependency of the convergence speed on the support of the two distributions. We varied the support for both, the uniform and the triangular distributions, using the following intervals: [1 1.4], [1 2], [1 5], [1 10], and [1 20], using an increment of 0.1. The optimal order quantity for the uniform distribution has stabilized $3, 4, 4, 5$, and 5 periods, and for the triangular distribution $3, 3, 4, 4$, and 5 periods after the end of the horizon (counting backwards). The wider the support, the longer it takes for the optimal order quantity to stabilize, as we would again expect from equation (32). A wider support translates into higher optimal order values at each period, which is directly connected with higher values for the cumulative distribution function at each period; the terms in the limit of the stationary equation (32) diminish after more periods if the support of the distribution function is wider. It seems again that the uniform distribution is more impacted by the wideness of the support than the triangular distribution.

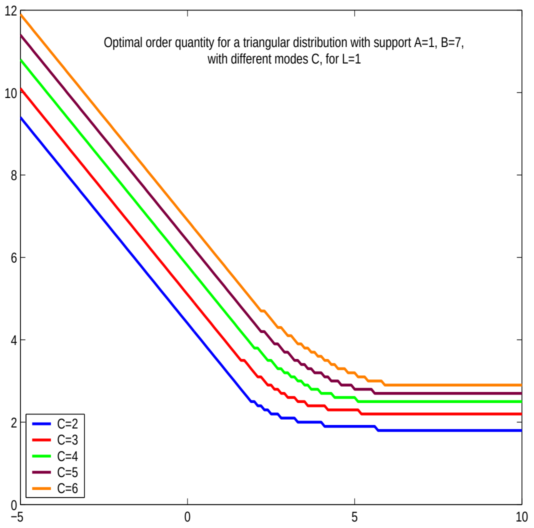

Figure 12: Dependency of the optimal order quantity on the skewness of the triangular distribution

3.6 Special cases

The problem can be extended by adding due dates to demand and subcontracting options for the excess demand. The demand is assumed to be shipped no later than l periods, i.e, the demand can be backlogged for at most l periods. The excess demand can be either backlogged (if it is not yet expired) or additional capacity can be subcontracted at a price $s > c$ to ship demand that expires soon.

Lemma 9 *When the orders are due immediately, $l = 0$, the capacity lead time is positive, $L \geq 0$, and subcontracting is an option, then the problem becomes a classic newsboy problem with the following solution:*

$$\Phi_{t+L}(q^t) = \frac{s-c}{s} \tag{60}$$

Proof: The capacity has to be confirmed at time t to satisfy demand at time $t + L$. Since the demand has to be shipped at the same time it occurs, there are no back orders from previous time periods. The problem can be modeled as a classic newsboy problem.

The cost incurred when ordering quantity q^t is denoted by v_t:

$$v_t = c \cdot q^t + s \cdot (D^{t+n} - q^t)^+$$

i.e., for the demand up to the level of the confirmed capacity the contract price is paid; subcontracting is used for the excess demand.

The expected value of the cost at time t is:

$E[v_t] = c \cdot q^t + s \cdot \int_{q^t}^{M} (y - q^t) \cdot \varphi_{t+n}(y) dy$

The optimal quantity for which the expected cost is minimized is deduced by setting $\frac{dv_t}{dq^t} = 0$, from which equation (60) follows. The underage cost, $c_u = s - c$, is the additional cost paid for not having ordered enough capacity. The overage cost, $c_o = c$, is the cost paid per unit unused capacity.

CHAPTER IV

CONCLUSIONS AND FUTURE WORK

We addressed two revenue management problems and one capacity management problem in this thesis. Our work is the basis for a more efficient management of revenues and capacity for the airlines and for the freight forwarders. For airlines, we provide solutions for accepting/rejecting incoming cargo bookings based on bid prices, such that their revenue at the end of the booking horizon is maximized, and for determining better overbooking levels based on a more accurate estimation of the cargo show-up rate at departure. For freight forwarders we provide solutions for making an optimal decision when confirming capacity with airlines. The advantages are mutual. The airlines can benefit from better capacity management from freight forwarders by being able to asses the degree of allotment utilization, and hence better forecast their capacity available for free sale. We have shown that a good forecast of capacity available for free sale has a crucial impact on overbooking levels, and consequently on profits. Better revenue management at the airline's end benefits the freight forwarders through higher sales and more efficient operations.

In the following sections we present a summary of our work by specific problem, together with directions for future research.

4.1 Air cargo revenue management

4.1.1 Air cargo bid prices

4.1.1.1 Contribution

In this thesis we propose a new method for determining bid prices for air cargo. We split the cargo bookings into two categories, namely, small and big cargo, and treat each category differently. The large bookings tend to be made close to the departure date of the airplane, and usually only a few bookings fill up the capacity dedicated to big cargo, whereas the small bookings are made throughout the booking period. We propose a probabilistic model to determine bid prices for the small cargo. Our contribution is the development of a novel algorithm to solve the traditional probabilistic nonlinear problem from the passenger side, which makes the problem tractable even for extremely large instances. The big cargo problem is modeled and solved as a dynamic problem decomposed by leg, with the revenue per leg pro-rated with the bid prices from the probabilistic network model.

We use extensive simulations (instances ranging from 4 legs and 300 classes to 40 legs and $315,000$ classes) to show that the proposed methods are both efficient and effective. The algorithm to solve the probabilistic model converges to a solution in less than 8 iterations even for the largest instances. The big cargo algorithm runs in less than 2 minutes even for large instances with 40 legs and $84,000$ classes.

We conducted an additional simulation study to asses the quality of our solution. We used simulated demand for both the small and the large cargo bookings. We simulated Gamma distributed demand for the small cargo, and used several strategies to update the bid prices: (1) once at the beginning of the booking period, and then every day 3 days before the departure; (2) once at the beginning of the booking period, and then every day 10 days before the departure; and (3) every day. For the large cargo, we used several demand distributions: (1) Binomial, (2) Negative Binomial, and (3) Gamma. We compare the total revenue obtained from our proposed approach with two other approaches used in practice (using the same simulated arrivals): (1) First Come First Served (FCFS) policy, where the capacity is filled with incoming bookings until the limit is reached. (2) Obtaining bid prices by solving the deterministic model mentioned in Section 2.1.3 as an integer program. For each request we solve the model twice: first assuming the request is accepted, and then assuming the request is rejected. If the difference in the objective function is below the rate associated with the incoming booking, then the shipment is accepted. We call this approach the Deterministic Integer Program (DIP).

The revenue gain over the FCFS method is substantial, reaching up to 60% depending on the simulated demand. The deterministic integer program from the passengers side gives almost the same results as our algorithm for the small cargo, slightly worse when we re-optimize more often. However, the deterministic integer program results in up to 18% less revenue than the dynamic problem in the big cargo setting. The deterministic integer program fails to produce better results when confronted with lumpier demand; our splitting strategy accounts for the lumpiness in the demand and results in higher overall revenue than any method that is currently used in practice.

4.1.1.2 Future directions

There are several directions of improving the air cargo revenue management problem:

1. Different prorating schemes: The prorating using the bid prices from the PNLP for big cargo might overestimate the tightness of the legs in the requested itinerary due to the linear relaxation (the integrality of the solution is relaxed in the PNLP model). A method to overcome this shortcoming is to prorate based on the sum of the bid prices along the requested itinerary, i.e., divide the bid price on the current leg by the sum of the bid prices along the requested itinerary.

2. Alternative routes: The small cargo model can be embedded with the findings in Chen et. al. [13] to extend the demand to origin destination specific; however, considering alternative routes in the DP may grow the space too much. For the beginning, we could restrict ourselves to a fixed set of preferred routes or to dynamic generation of a working set of routes.

3. Shipping Dates: If alternative shipping dates are considered for the big cargo problem, then the state space has to account for multiple days (potentially with a wrap-around). If the time frame contains enough days, then the problem might be stated as a a periodic problem and we could find the infinite time horizon (steady state) solution.

4. Concurrent capacity: We currently consider that the two problems for small and big cargo are perfectly separated. An interesting extension is to consider that the capacity is shared between the two categories. The idea is to dynamically change the capacity allocation between the two categories to improve revenues. An idea to tackle the problem would be to develop a master dynamic program to manage the capacity allocation such that the revenue is maximized; the master problem would pass different capacity allocations to two subproblems, which would be the problems solved in this thesis, and choose the allocation that gives the best revenue. The modeling of the master problem is challenging due the the high number of capacity allocation possibilities; approximation algorithms might be needed to solve the model.

4.1.2 Air cargo show-up rate estimation

4.1.2.1 Contribution

The show-up rate estimation is directly related to overbooking levels: better show-up rate estimation translates into better overbooking control, and hence in higher profits. We show that the Normal estimator used in the passengers business is not appropriate for the cargo business, and we develop a new discrete estimator based on wavelet density estimation. The discrete estimator outperforms the Normal estimator in various aspects. The overbooking levels using the discrete estimator prove a better approximation of the capacity at departure in terms of mean absolute error between the tendered cargo and the real capacity at departure, standard deviation of the error, spoilage, and off-loads. For a set of real world demand data, the average yearly savings from the discrete estimator for a combination carrier with 300 flights per day and an average cargo capacity per departure of 13,000 kilograms was $16,425,000. The discrete estimator resulted in significantly lower mean spoilage, that is, better utilization of capacity, and no increase in off-loads, leading to high savings in costs, increased profits, and improved customer satisfaction. Lower spoilage translates into more customers served promptly, and lower off-loads means that the airline turns down fewer customers. Hence, better utilization of the cargo capacity improves the service the airline offers to customers, which is important in the competitive market of air-cargo transportation.

We also found that forecasting capacity at departure plays an important role in cargo overbooking. If capacity estimates fluctuate over the reading period, spoilage or off-loads will occur even in the ideal setting when we know all demand in advance. Misestimation of capacity at departure results in poor utilization, which means unavoidable monetary losses because of the lost opportunity to satisfy more demand. Companies should invest in forecasting cargo capacity at departure, since, without accurate forecasts, any improved overbooking procedure would fail to improve the utilization of cargo capacity.

4.1.2.2 Future directions

The simulation used to asses the quality of the discrete estimator over the Normal estimator did not use any revenue management techniques to accept or reject the incoming bookings, but used a first come first serve policy. An immediate improvement is to employ the techniques developed in Chapter 2 in the simulation, and asses the quality of our estimator in an environment closer to the real world. The Normal estimator might perform better under these conditions than under a first come first serve acceptance policy.

Another shortcoming of our estimator is that it is one-dimensional. The discrete estimator should be generalized for two dimensions, weight and volume, using multivariate density estimates. The choice for a suitable estimator should be based on data analysis, and this is challenging due to the nature of the cargo business. In the current practice, the volume is deduced from the weight using a standard density, such that the data is correlated and cannot be used in such an analysis.

4.2 Air cargo capacity management

4.2.1 Contribution

To the best of our knowledge, we are the first to tackle the freight forwarders' capacity management problem. We defined the problem as an inventory perishable problem with backlogging options and time lag. We proposed¡ a model to this problem and solve it to optimality. We have shown that the problem is convex in the state variable for one and two period time lag, and we have proven that the optimal policy is a stationary policy. Furthermore, we provided the limiting equations for finding the stationary policy for one and two period time lag.

4.2.2 Future directions

The immediate future research direction is to extend the proofs for a general number of periods. We already have preliminary results that sustain the generalization, but the convexity might not hold for higher lead time periods. However, we could analyze the quality of several heuristics:

1. Develop a base stock heuristic ignoring perishability and compare with the optimal solution for L=1 and L=2, and with the deterministic version (demand is known with certainty upfront) for $L > 2$ (for higher lead time periods the numerical solution becomes intractable, so we need a different comparison basis);

2. Use the stationary equation, where the limit of the converging series is approximated by 0, and compare again with the optimal solution for L=1 and L=2, and with the deterministic version for higher lead times.

A shortcoming of our model is that we track capacity by flight, but we ignore the network effect. For example, if a shipment uses flights f_1 and f_2, then our forecast will show demand for flights f_1 and f_2. But if the shipment does not get on flight f_1, then it can't use free capacity on flight f_2. In general, backlog on one flight f_j affects demand for/backlog for other flights f_k.

REFERENCES

[1] Adelman, D., "Dynamic bid-prices in revenue management," *Operations Research*, to appear.

[2] Amaruchkul, K., Cooper, W., and Gupta, D., "Single-leg air-cargo revenue management." Working paper, Department of Mechanical Engineering, University of Minnesota. Available at: http://www.me.umn.edu/ amar0017/cargoTS3.pdf, retrieved July 2006, 2005.

[3] Arrow, K.J., K. S. and Scarf, H., *Studies in the mathematical theory of inventory and production.* Standford University Press, 3rd ed., 1958.

[4] B., V., *Statistical Modeling by Wavelets.* Wiley series in probability and mathematical statistics: Applied probability and statistics section, John Wiley & Sons, Inc., New York, 1999.

[5] Belobaba, P. P., *Air travel demand and airline seat inventory management.* Ph.d. dissertation, Flight Transportation Laboratory Massachusetts Institute of Technology, 1987.

[6] Boeing, "2004/2005 boeing world air cargo forecast," 2005. http://www.boeing.com/commercial/cargo/.

[7] Brodheim, E., Derman, C., and Prastacos, P., "On the evaluation of a class of inventory policies for perishable products such as whole blood.," *Management Science*, vol. 21, pp. 1320–1325, 1975.

[8] Bulinskaya, E., "Some results concerning optimum inventory policies," *Theory of Probability and Its Applications*, vol. 9, no. 3, pp. 389–402, 1964.

[9] C., C., *An Introduction to Wavelets*, vol. 1 of *Wavelet analysis and its applications.* Academic Press, oston, 1992.

[10] cargo management group, A., "International air freight and express industry performance analysis 2000," November 2000. http://www.cargofacts.com/annualstudies/index.htm.

[11] Catellan, G., "Modified akaike's criterion for histogram density estimation," tech. rep., Université Paris-Sud, Orsay, 1999.

[12] Chazan, D. and Gal, S., "A markovian model for a perishable product inventory," *Management Science*, vol. 23, no. 5, pp. 512–521, 1977.

[13] Chen, V. C. P., Guenther, D., and Johnson, E., "Routing considerations in airline yield management," *Proceedings 5th International Conference of the Decision Science Institute*, July 1999.

[14] Chew, E., Huang, E., E.L., J., G., N., J.S., S., and Leong, C., "Short-term booking of air cargo space," *European Journal of Operations Research*, vol. 174, pp. 1979–1990, 2006.

[15] Ciancimino, A., Inzerillo, G., Lucidi, S., and Palagi, L., "A mathematical programming approach for the solution of the railway yield management problem," *Transportation science*, vol. 3, pp. 168–181, May 1999.

[16] Cohen, M., "Analysis of single critical number ordering policies for perishable inventories," *Operations Research*, vol. 24, no. 4, pp. 726–741, 1976.

[17] Cooper, W., "Pathwise properties and performance bounds for a perishable inventory system," *Operations Research*, vol. 49, no. 3, pp. 455–466, 2001.

[18] De Boer, S., Freling, R., and Piersma, N., "Stochastic programming for multiple-leg network revenue management."

[19] Fedex, "Fedex ship manager at fedex.com - domestic and international shipping," 2006. http://www.fedex.com/us/.

[20] Fries, B., "Optimal ordering policy for a perishable commodity with fixed lifetime," *Operations Research*, vol. 23, no. 1, pp. 45–61, 1975.

[21] Grimmett, G. and Stirzaker, D., *Probability and Random Processes.* OXFORD University Press, Inc., New York, 3rd ed., 2001.

[22] Günther, D., *Airline Yield Management: Optimal Bid Prices, Markov Decission Process, and Routing Considerations.* PhD thesis, Industrial and Systems Engineering, Georgia Institute of Technology, 1998.

[23] K., M., "Histogram, an ancient tool and the art of forecasting." University of Michigan, 2002.

[24] Kasilingam, R., "An economic model for air cargo overbooking under stochastic capacity," *Comp. Ind. Eng.*, vol. 32, pp. 221–226, 1997.

[25] L., B. and Y., R., "How many bins should be put in a regular histogram." 2002.

[26] L., W. W., *Operations Research: Applications and Algorithms.* Duxbury Press, Belmont, California, 3rd ed., 1993.

[27] Luo, S., Cakanyildirim, M., and Kasilingam, R., "Two dimenional cargo overbooking models." Under review at Naval Research Logistics. Available at: http://som.utdallas.edu/faculty/working_papers/SOM200537.pdf, retrieved July 2006, 2005.

[28] Luo S., C. M. and R., K., "Two-dimensional cargo overbooking models," vol. Technical report, 2005.

[29] Moussawi, L. and Cakanyildirim, M., "Profit maximization in air cargo overbooking." Under review at Manufacturing and Service Operations Management. Available at: http://som.utdallas.edu/faculty/working_papers/SOM200562.pdf, retrieved July 2006, 2005.

[30] Nahmias, S., "A comparison of alternative approximations for ordering perishable inventory," *INFOR*, vol. 13, no. 2, pp. 175–184, 1975.

[31] Nahmias, S., "Optimal ordering policies for perishable inventory - ii," *Operations Research*, vol. 23, no. 4, pp. 735–749, 1975.

[32] Nahmias, S., "Myopic approximations for the perishable inventory problem," *Management Science*, vol. 22, no. 9, pp. 1002–1008, 1976.

[33] Nahmias, S., "Comparison between two dynamic perishable inventory models," *Operations Research*, vol. 25, no. 1, pp. 168–172, 1977.

[34] Nahmias, S., "Higher order approximations for the perishable inventory problem," *Operations Research*, vol. 25, no. 4, pp. 630–640, 1977.

[35] Nahmias, S., "The fixed-charge perishable inventory problem," *Operations Research*, vol. 26, no. 3, pp. 464–481, 1978.

[36] Nahmias, S. and Pierskalla, W., "Optimal ordering policies for a product that perishes in two periods subject to stochastic demand," *Naval Research Logistics Quarterly*, vol. 20, pp. 207–229, 1973.

[37] Nandakumar, P. and Morton, T., "Near myopic heuristics for the fixed-life perishability problem," *Management Science*, vol. 39, no. 12, pp. 1490–1498, 1993.

[38] O'Reilly, J., "Fare's fair: Making air cargo count." June 2003.

[39] Pak, K. and Dekker, R., "Cargo revenue management: bid prices for a 0-1 multiknapsack problem." ERIM Report Series Reference No. ERS-2004-055-LIS. Available at SSRN: http://ssrn.com/abstract=594991, retrievd December 2005, 2004.

[40] Pak, K. and Piersma, N., "Airline revenue management: An overview of or techniques 1982-2001." ERIM Report Series Reference No. ERS-2002-12-LIS. Available at SSRN: http://ssrn.com/abstract=370949, retrieved July 2004, 2002.

[41] Popescu, A., Keskinocak, P., Johnson, E., LaDue, M., and Kasilingam, R., "Estimating air-cargo overbooking based on a discrete show-up-rate distribution," *Interfaces*, vol. 36, pp. 248–258, May-June 2006.

[42] Puterman, M., *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley Series in Probabilities and Mathematical Statistics, John Wiley and Sons, Inc., 1994.

[43] R., W. L. and S., P., "Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues," *Journal of Revenue and Pricing Management*, vol. 1, no. 3, pp. 234–254, 2002.

[44] R., W. L. and S.E., B., "A taxonomy and research overview of perishable-asset revenue management: Yield management, overbooking and prizing," *Operations Research*, vol. 40, no. 5, pp. 831–844, 1992.

[45] Rao, B. V., "A convex programming model for cargo revenue-mix optimization." Internal Report, Sabre Holdings, 2000.

[46] Rinnooy, K., Stougie, A., and Vercellis, C., "A class of generalized greedy algorithms for the multi-knapsack problem," *Discrete applied mathematics*, vol. 42, pp. 279–290, 1993.

[47] Rudin, W., *Real and Complex Analysis.* McGraw-Hill Science/Engineering/Math, 3rd ed., 1986.

[48] S., L. and M., C., "Overbooking models for air cargo management," vol. Technical report, 2005.

[49] Secomandi, N., "An analysis of the control-algorithm re-solving issue in inventory and revenue management." July 2006.

[50] Simpson, R., "Using network flow techniques to find shadow prices for market and seat inventory control." MIT Flight Transportation Laboratory Memorandum M89-1, Cambridge, MA, 1989.

[51] Smith, B. and Penn, C., "Analysis of alternative origin-destination control strategies," *AGIFORS Annual Symposium Proceedings*, vol. 28, 1988. New Seabury, MA.

[52] Solutions, S. A., "Cargo management." 2006.

[53] Talluri, K. and Van Ryzin, G. J., *The theory and practice of revenue management.* International Series in Operations Research and Management Science, Boston/Dordrecht/London: Kluwer Academic Publishers, 2004.

[54] Van Zyl, G., "Inventory control of perishable commodities." Unpublished Ph.D. dissertation, 1964.

[55] Williams, C. L. and Patuwo, B. E., "A perishable inventory model with positive order lead times," *European Journal of Operations Research*, vol. 116, pp. 352–373, 1999.

[56] Williams, C. L. and Patuwo, B. E., “Analysis of the efect of various unit costs on the optimal incoming quantity in a perishable inventory model,” *European Journal of Operations Research*, vol. 156, pp. 140–147, 2004.

[57] Williamson, E., *Airline network seat control.* PhD thesis, MIT, Cambridge, MA, 1992.

[58] Xiao, B. and Yang, W., “Revenue management with multiple capacity dimensions.” Working paper, School of business, Long Island University, Brookville, NY, 2006.

VITA

Andreea Popescu received her bachelor degree in Images, Forms, and Artificial Intelligence from the Polytechnical University in Bucharest - Romania, School of Electrical Engineering. She received her Magister Degree in Operations Research from Rheinisch-Westfalische Hochschule in Aachen, Germany, and her Master's Degree in Operations Research from Georgia Institute of Technology in Atlanta, Georgia. She then continued her studies by pursuing a doctorate in Operations Research at the School of Industrial Engineering, at the Georgia Institute of Technology, in Atlanta. She is currently working for Turner Broadcasting System, Inc., as an Operations Research Analyst.