OPTIMIZATION MODELS FOR LAST-MILE SERVICE OPERATIONS

A Dissertation

by

YUE WANG

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Joseph Geunes
Co-Chair of Committee,	Xiaofeng Nie
Committee Members,	Hrayer Aprahamian
	Bahar Çavdar
	Alaa Elwany
Head of Department,	Lewis Ntaimo

August 2023

Major Subject: Industrial Engineering

ABSTRACT

Improving the quality of the last-mile service that comprises the movement of goods and people has been a recurrent theme in recent research. This dissertation aims to develop optimization models for addressing the emerging challenges encountered by three different last-mile service systems in the domains of inventory-delivery management, transportation and logistics, and emergency medical services, respectively.

The first chapter focuses on the domain of inventory-delivery management by investigating a two-echelon, single-product fulfillment system, in which the expected system-wide demand depends on the product's price, the committed delivery time, and the number of local distribution centers in the system. The proposed model maximizes the expected profit while accounting for the total expected revenue, product holding costs, and fixed facility costs.

The second chapter concentrates on the domain of transportation and logistics, in which a fleet composition problem is studied. The proposed models consider using both internal truckload capacity and external less-than-truckload shipments for fulfilling stochastic demand. The objective is to minimize the total expected cost per period while determining the number of trucks to own for each truck type given a set of truck classes.

The third chapter contributes to the field of emergency medical services by studying a routing problem in the aftermath of a disaster, wherein the proposed models aim to determine the optimal routing strategy while maximizing the total expected number of survivors. A single ambulance bus is used to transport a given number of casualties to a medical center, accounting for time-dependent survival probabilities of the casualties.

The models and solution approaches developed in this dissertation provide managerial insights and can serve as a starting point when making decisions on supply chain design, facility locating, asset procurement, and fleet management and operations.

DEDICATION

To my parents.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Joseph Geunes. I consider myself incredibly fortunate to have had the opportunity to be his student, learning from his expertise and wisdom. His guidance, expertise, and insightful feedback have been invaluable in shaping the direction and quality of our work. I am grateful for his support and patience throughout this five-year journey. I am also indebted to Dr. Xiaofeng Nie, my co-advisor, for his invaluable suggestions and encouragement. His commitment to rigor in research has taught me the essential qualities necessary to become a good researcher. I am deeply grateful for the mentorship and guidance I received from him. Furthermore, I would like to express my sincere gratitude to Dr. Bahar Çavdar. I want to thank her for the dedicated time she has devoted to our weekly research meetings. Her commitment and investment of time have been instrumental in the progress and success of our work. I am immensely grateful for the experience Dr. Hrayer Aprahamian has shared with me on teaching. His passion for teaching has had a tremendous influence on shaping my understanding of how to cultivate an effective, inclusive, and equitable learning environment. I would like to express my gratitude to Dr. Alaa Elwany for providing me with invaluable perspectives on fostering my professional growth.

Additionally, I would like to extend my heartfelt appreciation to my colleagues and friends who have offered their support and encouragement. It wouldn't be such an extraordinary journey without your presence. My deepest gratitude goes to Ximin. Thank you for being an incredible source of joy, strength, and support.

Finally, I am deeply grateful to my parents for their unwavering love, encouragement, and understanding. Their constant belief in me and their willingness to provide the necessary support throughout my academic pursuit has been the cornerstone of my success. I cannot thank them enough for their sacrifices and encouragement.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Joseph Geunes (Chair), Alaa Elwany, and Hrayer Aprahamian of the Wm Michael Barnes ’64 Department of Industrial & Systems Engineering, and Professors Xiaofeng Nie (Co-chair) and Bahar Çavdar of the Department of Engineering Technology and Industrial Distribution.

The data used for numerical experiments was randomly generated by the student independently.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by an assistantship from Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. OPTIMIZATION MODELS FOR INVENTORY PLACEMENT IN A TWO-ECHELON DISTRIBUTION SYSTEM WITH FULFILLMENT-TIME-DEPENDENT DEMAND *

1.1 Introduction

In recent years, technological advances and changing consumer preferences have led to substantial innovation in supply chain operations capabilities, as well as changes in the structure of many supply chains. Consumers have become conditioned to expect fast and reliable (and in some cases, free) delivery to their homes, which not only challenges last-mile delivery services, but places enormous pressure on inventory and production operations throughout the supply chain. This rapid shift has forced companies to invest in their logistics networks to strengthen their fulfillment capabilities and provide seamless, responsive, and reliable services. In order to fully take advantage of systems that provide last-mile delivery services, firms must carefully determine both price levels and inventory placement strategies, which combine with delivery and fulfillment costs to drive both customer responsiveness and profitability. Manufacturing and distribution firms should also ensure that the design and structure of their distribution systems provides the best fit with their operations capabilities and constraints. The modeling and solution approaches presented in this chapter provide an analytical framework for addressing this fit between distribution system structure and operations capabilities.

Inventory placement strategies, in particular, are vital in determining both the structure of the supply chain and its ability to quickly respond to customer requests. Staging inventory close to customers, for example, requires a network of local distribution centers (LDCs) and enables fast customer response times. This leads to an increase in both facility and inventory costs, while providing added value to consumers, which may enable suppliers to command higher prices. As a result, determining the best combination of product price and customer response time capability is a highly complex problem that is inextricably bound to the design and structure of the distribution

*Reprinted with permission from "Optimising inventory placement in a two-echelon distribution system with fulfillment-time-dependent demand" by Yue Wang, Joseph Geunes, and Xiaofeng Nie, 2022. International Journal of Production Research, 60:1, 48-72, Copyright [2022] by Taylor & Francis Group.

network. This chapter seeks to shed some light on the key decision drivers and structural properties of optimal solutions in such settings using a stylized distribution network model that simultaneously determines product pricing, inventory placement, and delivery response time decisions.

We consider a two-echelon fulfillment system in which a single upstream fulfillment center (FC) echelon serves multiple LDCs at a downstream echelon, which in turn provide last mile delivery to customers. Adding a new LDC to the system may, for example, effectively improve average customer delivery response time while also reaching new customers. This may, in turn, stimulate new demand, leading to higher revenue, while at the same time leading to increased system costs. The distribution system model we propose characterizes customer demand as a function of the number of distribution center locations (and the implied customer delivery response times) and product price, in addition to determining stock allocation levels across both echelons that either ensure meeting the prescribed delivery times with a very high probability, or utilize expediting to ensure that all demand is fulfilled within a prescribed amount of time. This model uses an expected profit objective, which accounts for revenue that depends on product pricing and delivery response time decisions, as well as inventory holding costs, fixed location costs, and variable delivery costs.

This chapter provides several contributions to the literature, including: (i) a proposed general demand growth model that characterizes total system demand as a function of the number of LDCs, and can capture a number of ways in which demand may respond to shorter delivery times in addition to price; (ii) the derivation of key structural equations that determine the boundaries between distinctly different distribution system structures and safety stock placement strategies; (iii) analysis and characterization of key system parameters that drive system performance and optimal distribution system structure; and (iv) a set of modeling and analytical tools that facilitate visualization of the firm's placement on a map of optimal safety stock placement strategies and corresponding distribution system structures. While we limit our analysis to systems with a single product due to space constraints, the analysis, modeling paradigm, and visualization tools can be extended in practice to multi-product systems.

The rest of this chapter is structured as follows. Section 1.2 reviews related literature in the area of multi-echelon distribution systems with guaranteed response times and safety stock considerations. In Section 1.3, we begin by describing a two-echelon system within a single market or region, followed by a definition of the market's expected demand function and a demand growth model that is dependent on the number of LDCs. We then elaborate on the approach for setting base-stock levels, and the implied safety stock and holding cost. At the end of Section 1.3, we provide a model formulation. Section 1.4 provides methods to determine optimal price and guaranteed service time values. We conduct numerical experiments and summarize our findings in Section 1.5. Parametric analyses are provided in Section 1.6. In Section 1.7, we explore a generalization of our single-market model to account for multiple, non-identical markets. Section 1.8 presents some conclusions and potential extensions for future research.

1.2 Literature review

Our research is rooted in studies on inventory optimization in multi-echelon/stage systems with guaranteed service times. The guaranteed service model was first introduced by Simpson [1], who models a sequence of production operations as a serial chain, wherein each stage fills the order placed by its downstream stage within a specified time while minimizing overall inventory cost. This paper proves that the optimal inventory cost occurs when each stage holds either zero stock or a full stock level that can meet the downstream demand up to a certain level, which holds under both uncorrelated and correlated demand. Since this publication, extensive research has considered this problem class under various modeling assumptions and system structures. Inderfurth [2] extends these results to general divergent systems with consideration of final-stage demand correlation that affects the size and distribution of safety stock. Graves and Willems [3] develop a strategic safety stock model (which we will refer to as the GW model) for general networks that can be modeled as a spanning tree, with the assumption of bounded demand and a 100% service level (i.e., all demand is met within a prescribed guaranteed service time). They later extend this model to nonstationary demand (see [4]) by applying a constant service time policy, which leads to a near-optimal solution that is beneficial in practice. Neale and Willems [5] also emphasize the need for nonstationary solutions, based on which the GW model is extended for practical implementation. Funaki [6] develops a guaranteed-service model for due-date-based demand using a so-called make-to-plan scheme, in which the actual inventory level at each node is derived based on the production and demand quantities at a specified time point using a 'backward explosion' procedure. This model is applicable to either stationary or nonstationary demand.

A variety of extensions of the guaranteed service model for multi-echelon inventory systems have been explored with different assumptions on the service level definition and supply process cost structure and constraints. Inderfurth and Minner [7] present a further extension of the approach used by Simpson [1] that considers two different types of service measures related to the occurrence, size, and duration of stockouts, and derive properties of optimal stock policies for serial, divergent, and convergent systems, respectively. The model developed by Sitompul et al. [8] assumes that a processing capacity exists at each stage, and defines a 'correction factor' representing a linking relationship between excess capacity and demand uncertainty in order to measure the variation in the safety stock level. This approach is tested numerically using Monte Carlo simulation. Graves and Schoenmeyr [9] extend the classical guaranteed-service safety stock model under production capacity constraints for both single-stage and multi-stage supply chains using a modified constant base-stock policy. Chen and Li [10] develop a guaranteed-service model for serial systems with fixed order costs and operating flexibility costs; assuming continuous- instead of periodic-review, they determine optimal parameters for an (R, Q) (reorder point, order quantity) policy at each stage. We refer the reader to Eruguz et al. [11] for a comprehensive survey of categorized extensions of guaranteed-service models for safety stock optimization in multi-echelon systems.

Another extensive collection of studies on multi-echelon inventory models integrates safety stock decisions within a broader set of supply chain design decisions. These studies argue that holding cost should be an integral factor in determining an optimal number of stock locations, since the number of locations has substantial impacts on inventory levels and demand allocation decisions (Ballou [12]). Shen, Coullard, and Daskin [13] formulate a location-inventory model as a mixed integer nonlinear program to determine safety stock levels and facility locations in a single-supplier, multi-retailer system, where a (sub)set of retailers may also function as distribution centers. This model is reformulated as a set covering model and solved using a column generation algorithm. Graves and Willems [14] incorporate the core of the GW model within a supply chain configuration problem that determines optimal options for satisfying functional requirements at each stage in the supply chain. This model minimizes the overall configuration cost comprised of safety stock cost, pipeline stock cost, and delivery cost, and can be solved by dynamic programming when the network has a spanning-tree structure. A similar approach is used by Funaki [6], who combines the safety stock placement model with supply chain design decisions, proposing a stepwise procedure, wherein a threshold function is introduced to improve search efficiency by limiting the number of location combinations. Schuster Puga, Minner, and Tancrez [15] later formulate an integrated facility location and inventory planning model with guaranteed service times as a conic, quadratic, mixed-integer program with two different delivery strategies, where binary variables determine which strategy is chosen and whether or not each candidate facility is opened. These previous models typically take a cost minimization approach for a given set of customer demand distributions. You and Grossmann [16] consider both total facility and inventory costs within a bi-objective location and inventory planning problem with guaranteed service times, using the two objectives of total cost and customer lead times in order to address the tradeoff between total cost and responsiveness. In contrast, our approach focuses on characterizing optimal system profit using a demand growth model that depends on the number of LDCs in the system under various assumptions on the relationship between the number of LDCs and total demand. This analysis then permits gaining some insight on how safety stock and supply chain structural decisions evolve over time as the distribution system and its corresponding total demand grow.

To solve these guaranteed-service models, dynamic programming (see e.g., [2, 3, 17]) and heuristic algorithms (see e.g., [18]) are commonly used. Other solution approaches, such as Benders' decomposition (see e.g., [19] and mixed integer programming techniques (see e.g., [20]), are also used for systems containing special structures. Although a significant number of methods have been developed to solve guaranteed-service problems, few existing works have analytically characterized model solutions. Most recently, Hua and Willems [21] employ the original GW model within a two-stage serial line supply chain and characterize optimal safety stock policies in terms of per unit holding costs and leadtimes. Their model minimizes total holding cost by considering the percentage of total unit cost (i.e., cost allocation) and percentage of total leadtime (i.e., leadtime allocation) allocated to one of the stages; the decision variable, the net replenishment time allocation, is then simply restricted to a value between zero and the maximum value of the total supply chain leadtime. The optimal safety stock is consequently determined by varying cost and leadtime allocation.

Our work is motivated by the idea of analytically characterizing the tradeoff between revenue and total cost in last-mile delivery systems, using the GW model as a base operations cost model within a two-echelon inventory-distribution system consisting of a single FC and multiple identical LDCs (where the number of LDCs plays a vital role in enabling fast customer response times). Our results provide quantitative threshold values at which a transition occurs between different safety stock placement strategies, and corresponding supply chain structural design decisions. As we will see, these threshold values can be determined on the basis of pairwise comparisons of key problem variables and parameters.

The operations cost model we apply uses a similar approach to the original GW model, assuming bounded demand and 100% service levels to determine appropriate base-stock levels. We note, however, that a structurally identical model is obtained by assuming that the supplier always fulfills all of the LDC's demand within a promised lead time using, for example, delivery expediting from its external supplier when insufficient stock exists (see, e.g., [22]; note that in this model, the LDC incurs a shortage cost per unit backlogged, while the FC incurs a cost per unit expedited). Thus, our model applies under either of these conventions. We assume the expected value of demand per period in a single market depends on both the product price and the committed delivery time from the LDC to customers. Instead of minimizing overall cost as in the majority of existing studies, we optimize safety stock placement using an objective of maximizing the expected system profit, while determining the optimal price, the optimal committed service time from the FC to LDCs, and the optimal service time from each LDC to its corresponding market (end customers), respectively. Parametrically varying the number of LDCs in the system facilitates both determining an optimal number of LDCs for a given set of revenue, cost, and demand parameters, and considering the impact of demand and logistics system growth on optimal safety stock placement strategies.

1.3 Problem statement and formulation

We first consider a two-echelon fulfillment system consisting of a single regional FC that serves N independent and identical downstream LDCs in a given market or region with a single product (Section 1.7 later considers the impact of multiple, non-identical markets). Both echelons use a periodic-review inventory policy, ordering up to some base-stock level after observing demand in each period. The physical leadtime for replenishment orders placed by the FC to an external supplier equals T_F, which may include production and transportation time. Shipments from the FC to an LDC require a physical leadtime of T_L, while deliveries from an LDC to its customers require a physical leadtime of T_C. We use the term *physical leadtime* to refer to the time required for production and/or transportation between stages, assuming that physical leadtimes between stages are determined by operational constraints.

Each LDC serves a subset of the market with periodic stochastic demand, where the mean demand per period in the market depends on the product's price, denoted by p, and a *committed delivery time*, denoted by ℓ, to customers. For any given p and ℓ, demand in a market is stationary, and periodic market demands are independent and identically distributed (IID). In addition to setting a committed delivery time to end customers, the FC also sets an internal *committed supply time*, denoted by s, for orders placed by any LDC within the market to the FC. The FC guarantees that any order placed by the LDCs will arrive within the quoted committed supply time s. Figure 1.1 illustrates the structure of the system we consider.

Observe that if no safety stock is held at the FC and LDCs, the committed delivery time ℓ cannot be smaller than $s + T_C$, while s must be at least as large as the sum of the two physical leadtimes, $T_F + T_L$. Similarly, if safety stock is held only at the FC, ℓ cannot be smaller than $s + T_C$,

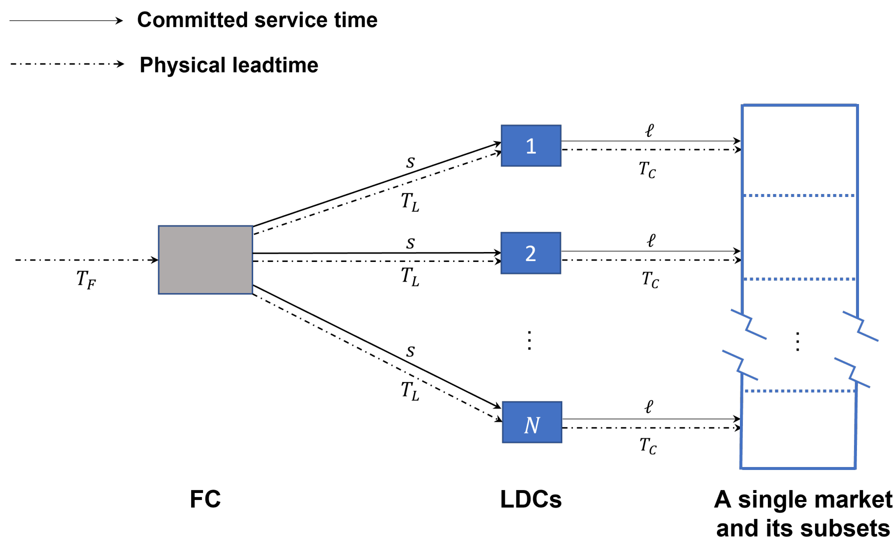


Figure 1.1: Two-echelon system structure (single market).

although s may take any value greater than or equal to T_L. Holding safety stock at the LDCs and not at the FC permits choosing a committed delivery time as small as T_C, while s must take a value at least $T_F + T_L$. The objective is to maximize total expected system profit by determining the committed supply time, s, between the FC and any LDC, and the committed delivery time, ℓ, to customers, as well as the product's price, p, and the number of LDCs, N.

1.3.1 Demand

A primary goal in studying this problem is to understand how strategic stock placement decisions are affected by demand growth and the system's ability to quickly satisfy demands, the latter of which depends in large part on the proximity of customers to LDCs, and therefore depends on the number of LDCs in the system, N. Studying these effects requires a model that characterizes the relationship between system demand and the number of LDCs serving these demands. We therefore propose a demand model that can accommodate various assumptions on the interaction between total system demand and the number of LDCs serving this demand. At one extreme, this demand model can account for strict demand growth, e.g., the addition of a new LDC may cover a geographic region that was not previously covered by existing LDCs, thus introducing new demand to the system. At the other extreme, the addition of a new LDC within a previously covered geographic region may not introduce any new demand to the system, although a subset of the customers previously covered by existing LDCs may now be served by the new LDC (assuming, e.g., each customer is served by its closest LDC). Our demand model also accounts for scenarios between these extremes, where the introduction of a new LDC may result in new demands and the new LDC may also cover customers previously served by existing LDCs.

1.3.1.1 Demand in a market with a single LDC

We begin by defining a single-LDC market (a market only served by a single LDC) demand model in which the expected market demand in a period is determined by the product's net price, p,[*] and the committed delivery time, ℓ. We assume that this market contains n independent customers uniformly distributed over a service region of area A, where each customer has an associated probability of ordering the product that is dependent on the price and committed delivery time, $P(p, \ell)$, in any period. The resulting market demand thus follows a binomial distribution with expected total demand per period of $\mu(p, \ell) = nP(p, \ell)$ and variance $\sigma^2(p, \ell) = nP(p, \ell)(1 - P(p, \ell))$. We further assume that the expected demand per period (and therefore the product of n and $P(p, \ell)$) depends on p and ℓ according to the function $\mu(p, \ell) = a - bp^u\ell^v$, where $0 < a \leq n$ is the maximum possible value of the mean periodic demand in the single market, $b > 0$ is a sensitivity coefficient of the price-delivery time interaction term, and u and v are positive values representing the elasticities of price and committed delivery time, respectively. As discussed in depth by Huang, Leng, and Parlar [23], many different specifications of the functional relationship between price, lead time, and demand are possible, where demand is decreasing in both price and delivery lead time. We opted to utilize one such functional form that is both reasonably general and leads to analytical tractability, which is a generalization of the power model suggested by Chen, Ray, and Song [24].

[*]We assume that expected demand can be expressed as a function of the difference between the product's price and any variable processing and delivery costs, which we define as the product's net price.

Note that $P(p,\ell) = \frac{a-bp^u\ell^v}{n}$ is a fraction between 0 and 1. For certain ranges of the value $P(p,\ell)$, the quantity $Q(p,\ell) = P(p,\ell)(1 - P(p,\ell))$ is relatively stable. For example, when $0.4 \leq P(p,\ell) \leq 0.6$, $Q(p,\ell)$ varies between 0.24 and 0.25, and for $0.3 \leq P(p,\ell) \leq 0.7$, $Q(p,\ell)$ varies between 0.21 and 0.25. Thus, price and committed lead time do not substantially influence the value of the variance for a range of probability values centered at 0.5. Observe that when $P(p,\ell) = 0.5$, this variance achieves its maximum value of $\bar{\sigma}^2 = 0.25n$. Because the value of this probability does not substantially affect standard deviation within a nontrivial range of values, and for analytical tractability reasons, our model uses a fixed value of variance equal to the upper bound on $\sigma^2(p,\ell)$, that is, $\bar{\sigma}^2$ (under a normal approximation to a system with bounded demand and a 100% service level, using this upper bound on variance leads to greater accuracy of the approximation of safety stock holding costs as well).

1.3.1.2 Demand growth model

We next consider a market containing N LDCs, and define a nondecreasing demand growth multiplier function of N that takes a value between 1 and N, denoted as $J(N)$, with $1 \leq J(N) \leq N$, to model market demand growth as the number of LDCs in the market increases from 1 to N. The function $J(N)$ scales the demand distribution to account for the possibility of demand (or market share) growth within a market as LDCs are added and the firm gains additional presence and visibility within the market. If $\mu(p,\ell)$ is the base expected demand per period in the market when a single LDC exists in the market, then we assume that the total expected market demand per period with N LDCs equals $J(N)\mu(p,\ell)$, with each LDC serving a set of customers with expected periodic demand of $\frac{J(N)}{N}\mu(p,\ell)$. This permits modeling a wide range of demand growth scenarios in a market as LDCs are added to the system. At one extreme, when $J(N) = 1$, the addition of LDCs leads to no demand growth, and a situation in which each of the N LDCs serves $\frac{1}{N}$ of the base demand. At the other extreme, when $J(N) = N$, the addition of each new LDC to the market increases expected demand per period by the base demand level of $\mu(p,\ell)$, for a total expected periodic market demand of $N\mu(p,\ell)$, with each LDC serving an expected demand per period of $\mu(p,\ell)$. Values of $J(N)$ between 1 and N then permit modeling demand growth levels between these extremes.

Figure 1.2 illustrates various scenarios as the number of LDCs increases from 1 to N, which depend on the market growth assumptions and the corresponding definition of $J(N)$. Observe that the extreme case of $J(N) = 1$ corresponds to a scenario in which all N LDCs are allocated an identical fraction (that is, $\frac{1}{N}$) of a fixed market of area A (corresponding to the case in which no new demand results from the addition of LDCs to a fixed market size), while $J(N) = N$ implies that each LDC serves a market of size A (corresponding to the strict market growth case with no market overlap).

Letting $D(p, \ell, N)$ denote the mean of the total market demand per period when the number of LDCs serving the market equals N, we characterize this mean value as $D(p, \ell, N) = J(N)\mu(p, \ell) = J(N)\,(a - bp^u\ell^v)$. The corresponding variance of the system demand per period then becomes $J(N)\bar{\sigma}^2$. Because this total system demand corresponds to the periodic demand seen by the FC, letting $\mu_F(p, \ell, N)$ and $\sigma_F(N)$, respectively, denote the mean and standard deviation of the periodic demand seen by the FC, we have $\mu_F(p, \ell, N) = J(N)\,(a - bp^u\ell^v)$ and $\sigma_F(N) = \sqrt{J(N)}\bar{\sigma}$.

The resulting demand seen by an LDC in any period follows a binomial distribution with parameters $\frac{J(N)n}{N}$ and $P(p, \ell)$, and thus has expected value $\frac{J(N)nP(p,\ell)}{N} = \frac{J(N)\mu(p,\ell)}{N}$ and variance $\frac{J(N)\bar{\sigma}^2}{N}$. Under these assumptions, we observe that the expected value and standard deviation of periodic demand at a single LDC are characterized by $\mu_L(p, \ell, N) = \frac{J(N)}{N}\,(a - bp^u\ell^v)$ and $\sigma_L(N) = \sqrt{\frac{J(N)}{N}}\bar{\sigma}$, respectively.

More generally, for $1 \leq J(N) \leq N$, we might consider a non-decreasing functional form $J(N) = N^\gamma$, where $0 \leq \gamma \leq 1$. A larger value of γ indicates a slower growth rate of system demand as new LDCs are added to the system. An alternative way to model a non-decreasing function that takes a value between 1 and N is by including an exponential term, i.e., $J(N) = 1 + f(N)(1 - e^{1-N})$, where $f(N)$ is a function of N. To ensure the value of $J(N)$ is between 1 and N, we define $f_{\max}(N) = (N-1)/(1 - e^{1-N})$ and choose $f(N)$ such that $0 \leq f(N) \leq f_{\max}(N)$.

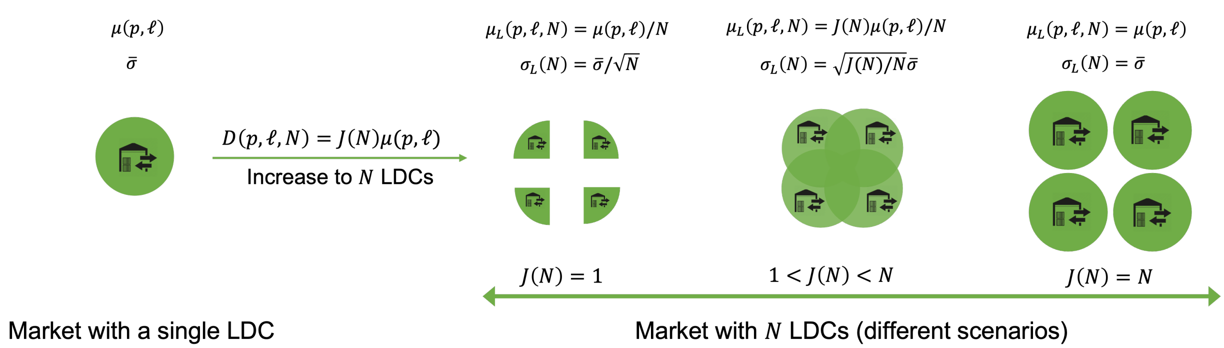


Figure 1.2: Potential distribution system and market structure changes as the number of LDCs grows from 1 to N.

1.3.2 Base-stock levels and holding costs

As noted at the beginning of this section, the FC and the LDC use a periodic review base-stock policy. Our model assumes that in a given period t, after observing demand, the LDC initiates demand satisfaction for all demands that occurred in period $t - (\ell - T_C)$, ensuring that these demands are satisfied within ℓ periods of their occurrence (in period $t + T_C$). Similarly, the FC ships out all orders that were placed by LDCs $s - T_L$ periods prior to period t. Clearly, the value of s should be at least T_L, as this is the physical lead time of a shipment from the FC to the LDC, and need not be greater than $T_F + T_L$, the time required for a shipment to be sent to the FC from its external supplier (who has unlimited capacity) and then to the LDC. Similarly, ℓ should be at least as great as the physical lead time to customers, T_C, and need not be greater than $s + T_C$, the time required to receive an order from the FC and send it to the customer. Ensuring that in period t the FC always ships all demand that occurred in period $t-(s-T_L)$ while the LDC initiates fulfillment of all orders in period $t - (\ell - T_C)$ can be achieved through one of two modeling approaches proposed in the literature.

The first of these approaches assumes that any difference between the amount due to be shipped from the FC or LDC and the on-hand stock at the corresponding location in any period can be immediately expedited to the location at a unit shortage cost. This is the approach used in the two-echelon distribution model proposed by Barnes- Schuster, Bassok, and Anupindi [22]. This approach requires explicitly defining a per-unit cost associated with expediting at each stage, and directly addresses the tradeoff between holding and shortage costs in setting the base-stock levels at the FC and LDCs.

An alternative approach, which produces a structurally identical mathematical model under a normally distributed demand assumption, is the use of the guaranteed service model approach as an approximation method when demand may be unbounded. The guaranteed service approach of Graves and Willems [3] assumes bounded demand in each period, and that an LDC is able to guarantee meeting all of its customers' demands within ℓ periods. The authors recognize that this 100% service approach is not exact under a normal demand assumption, and is unable to directly account for the common inventory pooling effects due to temporal aggregation of orders. Thus, they suggest using a demand bound that exceeds expected demand by some $k > 0$ standard deviations in the normal distribution case, and suggest combining standard deviations of independent demand streams in order to derive approximate bounds that account for inventory pooling effects. The value of the safety factor k determines the probability that actual demand may exceed the approximate bound (this probability is, of course, strictly decreasing in k, implying that the approximation strictly improves in k).

To determine the required base-stock levels at each stage, we use a similar approach developed in the GW model [3]. Letting B_F and B_L denote the base-stock levels at the FC and at a single LDC, respectively, the GW model sets $B_L = (s + T_C - \ell)\,\mu_L(p, \ell, N) + k_L \sigma_L(N)\sqrt{s + T_C - \ell}$ and $B_F = (T_F + T_L - s)\mu_F(p, \ell, N) + k_F \sigma_F(N)\sqrt{(T_F + T_L - s)}$, where k_F and k_L correspond to safety factors that determine the implied demand bounds at the FC and LDC, respectively (applying the modeling approach and interpretation of [22] leads to structurally identical base-stock levels, although we require redefining the values of the constants k_F and k_L in order to reflect their dependence on the optimal tradeoff between holding and shortage costs).

The safety stock level at a single LDC will then be equal to $SS_1(s, \ell, N) = B_L - (s + T_C - \ell)\mu_L(p, \ell, N) = k_L \sigma_L(N)\sqrt{s + T_C - \ell}$. The total safety stock held at N identical LDCs is thus equal to $SS_L(s, \ell, N) = NSS_1(s, \ell, N) = Nk_L \sigma_L(N)\sqrt{s + T_C - \ell}$, while the safety stock at the FC will be approximately equal to $SS_F(s,N) = B_F - (T_F + T_L - s)\mu_F(p,\ell,N) = k_F\sigma_F(N)\sqrt{T_F + T_L - s}$. Given a unit holding cost per period at the FC equal to h_F and at the LDC equal to h_L, the total system holding cost is expressed as $HC(s,\ell,N) = h_F k_F \bar{\sigma}\sqrt{J(N)\left(T_F + T_L - s\right)} + h_L k_L \bar{\sigma}\sqrt{J(N)N(T_C + s - \ell)}$.

1.3.3 Model formulation

The expected total system revenue can be expressed as the product of the product's price, p, and the total expected system demand, that is,

$$R(p,\ell,N) = pD(p,\ell,N) = J(N)\left(ap - bp^{u+1}\ell^{v}\right). \tag{1.1}$$

Our profit-maximizing model seeks the value of committed supply time from the FC to any LDC, s, the committed delivery time to customers, ℓ, as well as the product's price, p, and the number of LDCs, N.

The objective is to maximize the expected total system profit per period, denoted by $\mathit{\Pi}(p,s,\ell,N)$, which equals total net revenue minus the sum of total system costs, which include safety stock holding costs and fixed facility costs. Let K denote the fixed cost per period associated with operating an LDC. The resulting model is formulated as follows:

$$[\mathbb{P}] \quad \max \mathit{\Pi}(p,s,\ell,N) = J(N)\left(ap - bp^{u+1}\ell^{v}\right) - KN - h_F k_F \bar{\sigma}\sqrt{J(N)\left(T_F + T_L - s\right)}$$
$$- h_L k_L \bar{\sigma}\sqrt{J(N)N(T_C + s - \ell)} \tag{1.2}$$
$$\text{s.t. } T_L \le s \le T_F + T_L, \tag{1.3}$$
$$T_C \le \ell \le s + T_C, \tag{1.4}$$
$$1 \le N \le N_{UB}, \tag{1.5}$$
$$s,\ell,p \ge 0, N \in \mathbb{Z}^{+}. \tag{1.6}$$

The objective function (1.2) maximizes the expected total system profit per period. Constraint (1.3) requires the FC's committed service time to take a value at least the physical leadtime required to replenish an LDC (T_L), but no greater than the maximum physical leadtime of $T_F + T_L$. Constraint (1.4) similarly requires the committed delivery time to take a value greater than or equal to the physical leadtime of T_C, but no larger than the FC's committed service time to an LDC plus the physical leadtime to customers. Constraint (1.5) limits the total number of LDCs in the system and requires at least one operational LDC. This upper bound may arise due to, for example, a limit on capital or geographical limits on the total number of markets.

1.4 Solution approach

This section characterizes the structure of optimal solutions and defines parameter-based regions within which each among a set of candidate safety stock strategies serves as an optimal strategy. We characterize threshold values at which the optimal safety stock placement strategy switches from one form to another. Section 1.4.4 considers a model extension in which the physical leadtime to customers depends on the number of system LDCs.

1.4.1 Optimal price

In the objective function of problem $[\mathbb{P}]$, the price decision variable p only appears in the revenue term (the first term). Observe that the second-order derivative of the objective function with respect to p is given by

$$\frac{\partial^2 \Pi(p, s, \ell, N)}{\partial p^2} = -J(N) bu(u+1) p^{u-1} \ell^v. \tag{1.7}$$

Because we assume that $1 \leq J(N) \leq N$ and both b and u are positive, for any given positive ℓ, the right-hand side of Equation (1.7) is negative, implying that the objective function is concave in price. To obtain an optimal value of price, we set the first-order derivative of the objective function in p to zero, that is, $J(N)\left(a - b(u+1)p^u \ell^v\right) = 0$, which gives an optimal price of the form $p^*(\ell) = \rho \ell^{-v/u}$, where $\rho = \left(a/(bu+b)\right)^{1/u}$.

1.4.2 Derivation of candidate optimal solutions

We can therefore substitute $p^*(\ell)$ into the objective function (1.2) and obtain a function that only depends on the committed supply time s, the committed delivery time ℓ, and the number of LDCs N, which takes the form

$$\begin{aligned}\mathit{\Pi}(s,\ell,N) =& J(N)\beta\ell^{-\frac{v}{u}} - KN - h_F k_F \bar{\sigma}\sqrt{J(N)\left(T_F + T_L - s\right)} \\ & - h_L k_L \bar{\sigma}\sqrt{J(N)N(T_C + s - \ell)},\end{aligned} \tag{1.8}$$

where $\beta = a\rho - b\rho^{u+1}$. It is straightforward to show that for any given value of N, the Hessian matrix of (1.8) is positive definite in the variables s and ℓ (see Appendix A.1), which implies that $\mathit{\Pi}(s,\ell,N)$ is convex in s and ℓ. Note that in the (s,ℓ) space, problem [$\mathbb{P}$] has a closed, convex feasible region with four extreme points at $(T_L, T_L + T_C)$, (T_L, T_C), $(T_F + T_L, T_C)$, and $(T_F + T_L, T_F + T_L + T_C)$. Accordingly, for any given N, an optimal solution for problem [$\mathbb{P}$] occurs at one of these extreme points. Table 1.1 displays the functional form of each candidate optimal solution in the s and ℓ variables, as well as the corresponding implied safety stock placement strategy.

Table 1.1: Extreme point solutions in the (s,ℓ) space and corresponding safety stock placement strategy

Extreme point (s,ℓ)	Objective function	Safety stock placement strategy
$X_F = (T_L, T_L + T_C)$	$\mathit{\Pi}_F(N) = J(N)\beta(T_L + T_C)^{-v/u} - h_F k_F \bar{\sigma}\sqrt{J(N)T_F} - KN$	F (at the FC only)
$X_{FL} = (T_L, T_C)$	$\mathit{\Pi}_{FL}(N) = J(N)\beta T_C^{-v/u} - h_F k_F \bar{\sigma}\sqrt{J(N)T_F} - h_L k_L \bar{\sigma}\sqrt{J(N)NT_L} - KN$	FL (at both the FC and the LDCs)
$X_L = (T_F + T_L, T_C)$	$\mathit{\Pi}_L(N) = J(N)\beta T_C^{-v/u} - h_L k_L \bar{\sigma}\sqrt{J(N)N(T_F + T_L)} - KN$	L (at the LDCs only)
$X_O = (T_F + T_L, T_F + T_L + T_C)$	$\mathit{\Pi}_O(N) = J(N)\beta(T_F + T_L + T_C)^{-v/u} - KN$	O (make-to-order (MTO), zero safety stock)

The extreme point solution denoted by X_F corresponds to a policy of holding safety stock at the FC but not at the LDCs, while using a committed supply time of T_L and a committed delivery time of $T_L + T_C$. We refer to this solution as strategy F. Solution X_{FL} holds safety stock at both the FC and LDCs and quotes a committed supply time to LDCs of T_L and a committed delivery time of T_C to the customers, which corresponds to strategy FL. Solution X_L also uses a committed delivery time of T_C but applies a committed supply time of $T_F + T_L$, while no safety stock is held at the FC, resulting in strategy L. The policy implied by extreme point X_O corresponds to strategy O, where each echelon places orders in response to demand, with no safety stock held either at the FC or any of the LDCs. One of the extreme points $X_F, X_{FL}, X_L,$ and X_O, and, therefore, one of the strategies in the strategy set $S = \{F, FL, L, O\}$, provides an optimal solution in the s and ℓ variables for any given value of N, with the optimal strategy determined by the associated parameter values that determine the objective function value at each point.

The ability to determine the optimal number of LDCs, N, strongly depends on the form of the market growth function $J(N)$. In the extreme cases in which $J(N) = 1$ and $J(N) = N$ we can show that the expected profit at each extreme point solution in Table 1.1 is convex in N. In the former case we can also show that expected profit is strictly decreasing in N at every extreme point solution (and one LDC is, therefore, optimal),[†] while in the latter case, an optimal solution arises at one of the extreme solutions $N = 1$ or $N = N_{UB}$ (in this case, we can show that if a profitable solution exists, then at each extreme point solution, some threshold value of $N = N_T$ exists such that expected profit is strictly increasing in N for $N > N_T$; thus, if N_{UB} is sufficiently large, $N = N_{UB}$ at optimality). Analytically characterizing the optimal value of N may be difficult or impossible under more general functional forms of $J(N)$. For example, when $J(N) = N^\gamma$ as defined in Section 1.3.1.2 with $0 < \gamma < 1$, $J(N)$ is strictly concave in N. In this case, both $\sqrt{J(N)}$ and $\sqrt{NJ(N)}$ are also strictly concave, and problem [$\mathbb{P}$] maximizes a difference of concave functions in N, and thus falls in the class of $\mathcal{NP}$-Hard optimization problems in general (see, e.g., [25]). Because the problem is easily solved for any given value of N, and because the upper bound N_{UB} is likely to be a manageable value (e.g., less than 1,000), we can efficiently solve the problem by enumerating candidate integer values of N.

[†]Note that the optimality of a single LDC when $J(N) = 1$ occurs because of the assumption that the physical customer delivery time, T_C, is independent of N; the same does not hold in general for the case in which $J(N) = 1$ and T_C is decreasing in N, which we consider in greater detail in Section 1.4.4.

1.4.3 Optimality regions and threshold values

The objective function value at each extreme point in Table 1.1, and therefore, the structure of an optimal solution, depends on the number of LDCs, N. Both the number of LDCs, N, and the physical leadtime for delivery to customers, T_C, (which may, in turn, depend on N) play a vital role in determining an optimal safety stock placement strategy. We are, therefore, interested in characterizing how these two factors interact to determine the best safety stock placement strategy. To streamline our analysis, given a set of problem parameters, we create a two-dimensional mapping of T_C and N values to optimal extreme point solutions that permits visualizing and characterizing the way in which these values affect the optimal stock placement strategy.

To define the optimality regions for each of the stock placement strategies, we perform pairwise comparisons between the four objective function values in Table 1.1, assuming that T_C has a fixed value independent of N (Section 1.4.4 later considers a generalization in which T_C may depend on N). These pairwise comparisons lead to a series of inequalities that define threshold values of N and T_C such that the dominant strategy differs. This permits characterizing the subset of (positive) values of (T_C, N) corresponding to the optimality of each of the extreme point solutions in Table 1.1. We first perform comparisons for $J(N) = N$, in which case the optimality regions can be analytically expressed in terms of N and T_C. Appendix A.2 provides a detailed derivation and the resulting inequalities. The resulting optimality regions are summarized in Table 1.2. Note that R_S corresponds to the region in which strategy S is optimal, while $N_{S_1,S_2}(T_C)$ corresponds to a threshold function such that for a given a value of T_C, when $N = N_{S_1,S_2}(T_C)$ we are indifferent between strategies S_1 and S_2, i.e., $\Pi_{S_1}(N) = \Pi_{S_2}(N)$ for strategies $S_1, S_2 \in \{F, FL, L, O\}$. Similarly, $T_{S_1,S_2}(T_C)$ is a function such that when $T_{S_1,S_2}(T_C) = 0$, $\Pi_{S_1}(N) = \Pi_{S_2}(N)$ holds for any value of N. The functional forms of $N_{S_1,S_2}(T_C)$ and $T_{S_1,S_2}(T_C)$ are provided in Appendix A.2 for all strategy pairs. For the given set of parameters in Table 1.3, we illustrate the threshold value curves for N and T_C, as well as the corresponding optimal strategy regions in Figure 1.3.

Figure 1.3(a) shows the threshold value curves for T_C and N, where $N_{O,F}(T_C), N_{O,FL}(T_C)$, and $N_{F,L}(T_C)$ are nonlinear functions of T_C, while $N_{FL,L}$ is independent of T_C. Observe that at

Table 1.2: Optimality regions for each optimal solution candidate when $J(N) = N$

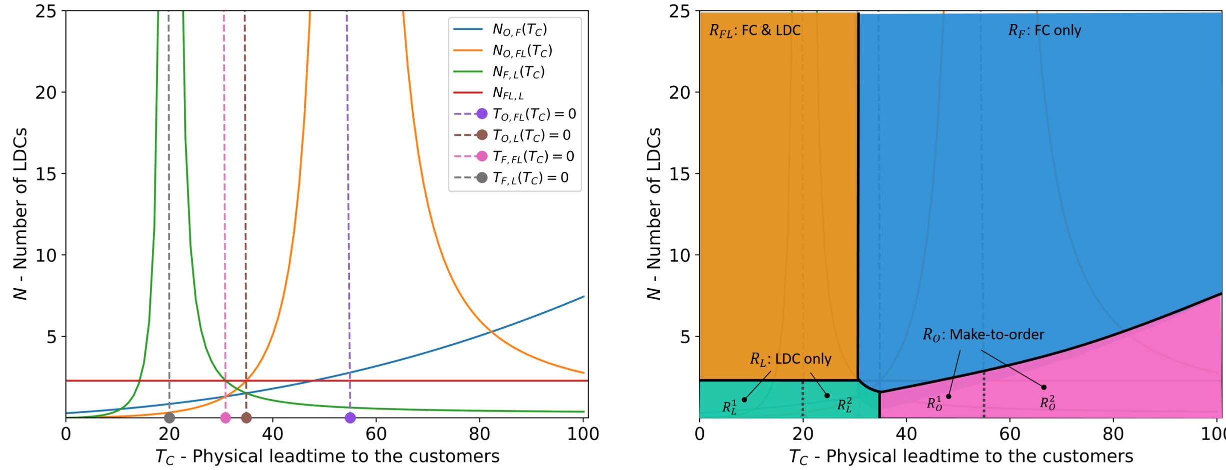
Objective value	Stock placement strategy	Optimality region
$\mathit{\Pi}_F(N)$	F (at the FC only)	$R_F = \Big\{(T_C, N) : N \geq \max\{N_{O,F}(T_C), N_{F,L}(T_C)\}, T_{F,FL}(T_C) \leq 0\Big\}$
$\mathit{\Pi}_{FL}(N)$	FL (at both the FC and the LDCs)	$R_{FL} = \Big\{(T_C, N) : N \geq \max\{N_{O,FL}(T_C), N_{FL,L}\}, T_{F,FL}(T_C) \geq 0\Big\}$
$\mathit{\Pi}_L(N)$	L (at the LDCs only)	$R_L = \Big\{(T_C, N) : N \leq N_{FL,L}, T_{F,L}(T_C) \geq 0\Big\} \bigcup \Big\{(T_C, N) : N \leq \min\{N_{FL,F}, N_{F,L}(T_C)\}, T_{F,L}(T_C) < 0, T_{O,L}(T_C) \geq 0\Big\}$
$\mathit{\Pi}_O(N)$	O (MTO)	$R_O = \Big\{(T_C, N) : N \leq \min\big\{N_{O,F}(T_C), N_{O,FL}(T_C)\big\}, T_{O,L}(T_C) \leq 0, T_{O,FL}(T_C) > 0\Big\} \bigcup \Big\{(T_C, N) : N \leq N_{O,F}(T_C), T_{O,FL}(T_C) \leq 0\Big\}$

Table 1.3: Parameters used for visualizing threshold values and optimality regions

a	b	$\bar{\sigma}$	h_L	h_F	k_L	k_F	T_L	T_F	u	v
10,000	100	100	1.5	1	3	3	25	30	2	0.5

the value of T_C such that $T_{F,FL}(T_C) = 0$, $N_{O,F}(T_C)$ (the blue curve) intersects with $N_{O,FL}(T_C)$ (the orange curve), and $N_{F,L}(T_C)$ (the green curve) intersects with $N_{FL,L}$ (the red horizontal line). Similarly, at T_C such that $T_{O,L}(T_C) = 0$, $N_{O,F}(T_C)$ (the blue curve) intersects $N_{F,L}(T_C)$ (the green curve), and $N_{O,FL}(T_C)$ (the orange curve) has the same value as $N_{FL,L}$ (the red horizontal line).

Figure 1.3(b) illustrates the optimality regions corresponding to each safety stock strategy, i.e., the regions formed by the resulting inequalities in the (T_C, N) space such that each extreme point solution is optimal, assuming $J(N) = N$. As the figure shows, when physical leadtimes are relatively long (e.g., $T_C = 60$), then if there are two LDCs, the optimal strategy does not hold any safety stock in the system but uses a strict make-to-order policy throughout the system, using a committed delivery time of $T_F + T_L + T_C$. If the system is expanded to five or more LDCs with the same customer physical leadtime of 60, then the optimal strategy transitions to holding safety stock at the FC only, using a committed delivery time of $T_L + T_C$. Thus, when physical leadtimes are relatively long ($T_C = 60$) and with a small number of LDCs ($N = 2$), the safety stock holding cost required to ensure a smaller committed delivery time outweighs the revenue increase associated with faster delivery times. With more markets ($N \geq 5$) and thus greater demand volume, it becomes optimal to reduce the committed delivery time from $T_F + T_L + T_C$ to $T_L + T_C$ and to hold pooled safety stock for all markets at the FC, as the pooled safety stock cost at the FC is less than the increased revenue resulting from shorter delivery times.



(a) Threshold values of N and T_C.

(b) Optimality regions for each safety stock strategy.

Figure 1.3: Threshold values and optimality regions when $J(N) = N$.

Similarly, when the physical leadtime to the customers requires a relatively short amount of time, e.g., $T_C = 10$, then as Figure 1.3(b) shows, for any number of LDCs greater than 2, the optimal solution holds safety stock at both the FC and the LDCs, as the associated safety stock cost is small relative to the demand potential associated with a fast committed delivery time (only when $N \leq 2$ do we hold stock at the LDC only; however, in a system with an FC and one LDC, the FC is wholly unnecessary).

We can consider a similar analysis when the number of LDCs (N) is fixed. For example, observe that the system with 10 LDCs and a physical customer leadtime of 15 leads to an optimal strategy that holds stock at both the FC and the LDCs, while the combination of 10 LDCs and $T_C = 40$ only holds stock at the FC. Because higher physical leadtimes increase the safety stock costs associated with reducing committed delivery times, and physical leadtimes constrain the ability to reduce committed delivery times, the revenue from increased demand due to shorter delivery times is not offset by the associated safety stock holding costs at the LDCs. Observe that if we increase T_C to a sufficiently high value, then even with 10 LDCs we eventually reach a threshold value that leads to no safety stock at the LDCs or the FC, i.e., a strict make-to-order system.

The visualization provided via Figure 1.3(b) can serve as a potentially powerful tool for logistics system managers. Given existing values of physical customer leadtime and the number of existing LDCs, this figure can be used to define a path of the projected values of these parameters over time, which can then be used to determine an appropriate future point at which to change safety stock placement strategy and distribution system structure.

1.4.4 Model generalization

In practice, adding LDCs may reduce the physical leadtime to customers (T_C) since a high facility density reduces physical delivery distance. Therefore, we consider the implications of allowing T_C to depend on the number of LDCs, N. Let us define $T_C(N) = T_0/N^{\theta}$, where T_0 is an initial physical leadtime to customers before adding any new LDCs, and $\theta \geq 0$ permits characterizing the rate of decrease in physical leadtime to customers as N increases. Substituting $\frac{T_0}{N^{\theta}}$ for T_C in $[\mathbb{P}]$ leads to a more general model with a more complex form of the objective function and a nonlinear form of Constraint (1.4). Observe that setting $\theta = 0$ is equivalent to setting T_C to a fixed value independent of N as in the previous analysis.

The derivation of optimal price in Section 1.4.1 and the resulting solution, $p^*(\ell)$, continue to hold under this generalized model, as do the convexity properties of model $[\mathbb{P}]$ for a fixed value of N. As a result, we can apply a similar analysis of the optimal strategy regions, although closed-form expressions for threshold function values of N are no longer available in general. Instead we define the functions $g_{S_1,S_2}(T_0, N)$ such that when $g_{S_1,S_2}(T_0, N) = 0$ we have $\Pi_{S_1}(T_0, N) = \Pi_{S_2}(T_0, N)$ for $S_1, S_2 \in \{F, FL, L, O\}$. Appendix A.3 provides the form for each $g_{S_1,S_2}(T_0, N)$ function, as well the resulting six inequalities that define the optimal strategy regions in the (T_0, N) space for each of the four safety stock strategies. Table 1.4 characterizes the resulting optimal strategy regions.

When $T_C(N) = T_0/N^{\theta}$ and $\theta > 0$, characterization of the behavior of $\Pi_S(N)$ ($S \in \{F, FL, L, O\}$) as a function of N becomes substantially more difficult than for the case in which $\theta = 0$ discussed in Section 1.4.2, even under the simpler functional forms of $J(N) = 1$ and $J(N) = N$. In particular, we can no longer claim that $\Pi_S(N)$ is convex in N when $J(N) = 1$ or when $J(N) = N$

and $\theta > 1$ (convexity in N continues to hold, however, when $J(N) = N$ and $0 < \theta \leq 1$). Thus, for example, when $J(N) = 1$ (all LDCs split a market of a fixed size) and the physical delivery time to customers depends on the number of LDCs, it is no longer necessarily optimal to serve the market with a single LDC, as was the case when physical delivery time was independent of the number of LDCs. In this case, therefore, an optimal solution will often locate more LDCs closer to customers, on average, in order to stimulate demand via decreased customer delivery times.

Table 1.4: Optimality regions for $T_C(N) = T_0/N^\theta$ and any $J(N)$

Objective value	Stock placement strategy	Optimality region
$\Pi_F(N)$	F (at the FC only)	$R_F = \Big\{(T_0, N) : g_{O,F}(T_0, N) \geq 0, g_{F,FL}(T_0, N) \leq 0, g_{F,L}(T_0, N) \leq 0\Big\}$
$\Pi_{FL}(N)$	FL (at both the FC and the LDCs)	$R_{FL} = \Big\{(T_0, N) : g_{O,FL}(T_0, N) \geq 0, g_{F,FL}(T_0, N) \geq 0, N \geq N_{FL,L}\Big\}$
$\Pi_L(N)$	L (at the LDCs only)	$R_L = \Big\{(T_0, N) : g_{O,L}(T_0, N) \geq 0, g_{F,L}(T_0, N) \geq 0, N \leq N_{FL,L}\Big\}$
$\Pi_O(N)$	O (MTO)	$R_O = \Big\{(T_0, N) : g_{O,F}(T_0, N) \leq 0, g_{O,FL}(T_0, N) \leq 0, g_{O,L}(T_0, N) \leq 0\Big\}$

1.5 Numerical analysis

Thus far, we have shown that one of the four extreme point solutions leads to an optimal solution, and have derived a series of inequalities that form the optimal strategy regions in the (T_C, N) space (when T_C is fixed) for each of the four candidate optimal strategies (in addition to analytical expressions for threshold values between these regions for certain $J(N)$ functions). The resulting optimal safety stock placement strategy is primarily affected by four factors: the number of LDCs, N, the total system demand growth function, $J(N)$, the internal committed supply time, s, and the committed delivery time to customers, ℓ. In this section, we are interested in exploring the properties of safety stock placement optimization by performing a numerical analysis based on these factors. Since the optimal values of s and ℓ are expressed in terms of T_C and T_L, we can therefore conduct our numerical tests by investigating the effects of N, T_C, and T_L under different forms of the function $J(N)$.

1.5.1 Results for different system demand growth models

In Section 1.3.1.2, we introduced the function $J(N)$ to describe system demand growth as the number of LDCs, N, increases. Recall that $J(N) = N$ leads to strict demand growth within a market by covering new service areas not covered by existing LDCs, while $J(N) = 1$ indicates that new LDCs do not introduce new demand but serve a subset of the existing customers instead. In addition to these two extremes, we also consider $J(N) = N^{\gamma}$ with $0 \leq \gamma \leq 1$ and $J(N) = 1 + f(N)(1 - e^{1-N})$ as more general cases where adding a new LDC generates some new demands but may also partially cover a subset of previously existing customers. In our numerical analysis we let $\gamma = 0.5$ and $f(N) = 1.5$, assume a fixed value of T_C, set $T_L = 25$ and $T_F = 30$, and use the cost and demand parameters in Table 1.3 to conduct numerical tests with $\{(T_C, N) : 1 \leq T_C \leq 41, 1 \leq N \leq 11\}$. Figure 1.4 illustrates the optimality regions for each safety stock placement strategy in the two-dimensional (T_C, N) space for different functional forms of $J(N)$. In our discussion of the results, we refer to the four different safety stock placement strategies as the MTO-strategy (no safety stock at either echelon), FL-strategy (holding safety stock at both echelons), LDC-strategy (holding safety stock only at the LDC) and FC-strategy (holding safety stock only at the FC).

Observe that for each of the forms of $J(N)$ tested, the border between the FL-strategy (orange region) and the LDC-strategy (turquoise region) corresponds to a horizontal line at $N = 2.28$, which is independent of T_C and the functional form for $J(N)$. Proposition 1.5.1 analytically proves this property.

Proposition 1.5.1. *For any functional form of $J(N)$, given a set of fixed parameter values, the transition from the FL-strategy to the LDC-strategy occurs at a fixed value of N, which is independent of T_C and the form of $J(N)$.*

All proofs of propositions and corollaries appear in Appendix A.4. The proof of Proposition 1.5.1 indicates that given all other parameters as fixed, the transition between the FL-strategy and the LDC-strategy only depends on the number of LDCs in the system and not on the market (total

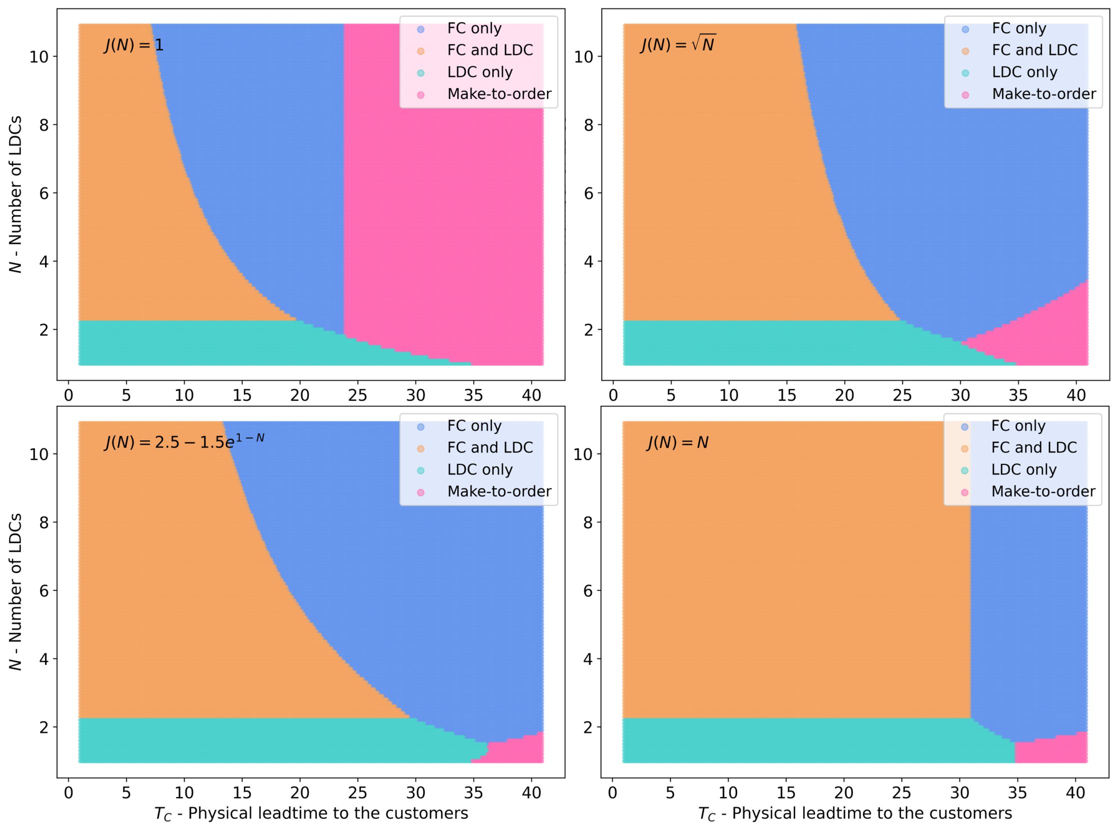


Figure 1.4: Optimality regions in the (T_C, N) space with different $J(N)$ functions.

demand) growth rate and the physical delivery time to the customers. Analytically, if we look at the expressions for $\Pi_{FL}(N)$ and $\Pi_L(N)$ in Table 1.1, $J(N)$ equally contributes to the growth in revenue in both objective functions when T_C is fixed; that is, the better stock strategy is solely determined by the overall holding costs (since each strategy also has the same fixed cost, KN). The function $J(N)$ serves as a scaling term (that is, $\sqrt{J(N)}$) in the holding cost for either policy, and will therefore cancel out when we compare $\Pi_{FL}(N)$ and $\Pi_L(N)$. Hence, the value of N determines which of these two strategies has a lower overall holding cost and a higher system expected profit. The value $N_{FL,L}$ corresponds to the points at which $\Pi_L(N_{FL,L}) = \Pi_{FL}(N_{FL,L})$, which can be visualized as a horizontal line in the (T_C, N) coordinate system.

Figure 1.4 also shows that the border between the FC-strategy (blue region) and the MTO-strategy (pink region) when $J(N) = 1$, and the border between the FC-strategy (blue region) and the FL-strategy (orange region) when $J(N) = N$, correspond to vertical lines independent of N. This is formalized in Proposition 1.5.2.

Proposition 1.5.2. *When T_C is independent of N, the threshold value of T_C at which the FC-strategy transitions to the MTO-strategy when $J(N) = 1$, and to the FL-strategy when $J(N) = N$, depends only on the physical delivery time to customers, T_C, and is independent of the number of LDCs, N.*

When $J(N) = 1$, the addition of a new LDC does not contribute to growth in total demand or revenue. Consequently, the total demand seen by the FC remains constant, which has no impact on FC holding cost. In this case, therefore, the value of T_C becomes the decisive determinant of the optimal values of demand and revenue. A larger T_C causes a lower demand and revenue. Thus, as we increase T_C, the loss in demand and revenue will reach a threshold and a transition from the FC-strategy to an MTO-strategy will happen, as the lower revenue no longer justifies the holding cost incurred by the FC.

At the other extreme, with strict market growth (that is, $J(N) = N$), for a system using an FC-strategy, the resulting optimal committed delivery time is $T_L + T_C$, and holding cost is only incurred at the FC. Under an FL-strategy, however, the resulting optimal committed delivery time is T_C, which implies a higher expected value of total demand (thus, higher expected revenue), although the system now absorbs holding cost at each LDC in addition to the FC. Thus, for a system with a small value of T_C, adding stock to the LDCs can lead to total revenue that offsets the holding cost incurred at each LDC, independent of the number of LDCs in the system, as Equation (A.36) in Appendix A.4 indicates.

1.5.2 Optimal committed service times

The optimal values of the committed supply time, s, and the committed delivery time, ℓ, are determined by the physical shipping time from FC to the LDCs, T_L, and the physical delivery time from LDCs to the customers, T_C, respectively, as shown in the first column of Table 1.1. To illustrate the effects of different physical leadtime values, we consider the optimal strategy for a system with a fixed number of LDCs and a range of physical leadtime values. In particular, we allow both T_L and T_C to vary from 1 to 41 under different forms of the growth function $J(N)$, using the same parameters from Table 1.3. Figure 1.5 compares the resulting optimal strategy regions in the (T_C, T_L) space when $N = 2$.

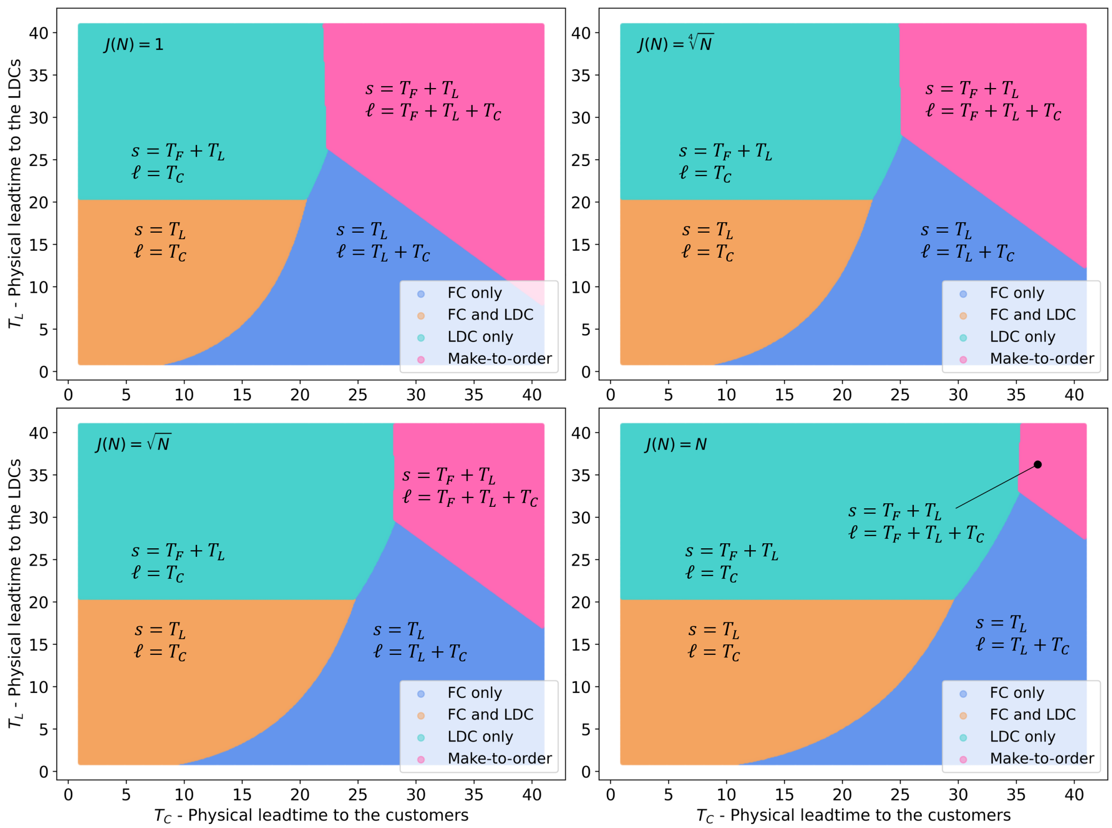


Figure 1.5: Optimality regions in the (T_C, T_L) space with different $J(N)$ functions.

Under each form of $J(N)$, if the system has a small value of T_C ($T_C \leq 10$) and a moderate value of T_L ($T_L \leq 20$), the optimal strategy is to hold safety stock at both the FC and LDCs (lower left orange region in Figure 1.5), where the FC commits a supply time of T_L to each LDC and each LDC guarantees its customers a delivery time of T_C periods. This confirms the intuition that shorter physical leadtimes result in smaller committed service times, which generate sufficient demand (higher revenue) to offset the corresponding system holding cost, which is relatively low due to the shorter physical leadtimes.

Suppose we begin with such a solution (in the orange region in Figure 1.5) and increase the value of T_L while holding T_C fixed. As we increase T_L (e.g., $T_L > 20$), we maintain an optimal delivery time of T_C periods to customers, but transition to using a longer internal supply time of $T_F + T_L$ periods to the LDCs, shifting the safety stock held at the FC to the LDCs (moving from the orange region to the turquoise region). In contrast, if we begin at a point in the orange region and hold T_L at a moderate value while increasing T_C (e.g., for $T_C > 25$), at maximum expected profit, the system is no longer able to commit a T_C-period delivery time to customers, instead providing a slower delivery service time equal to $T_L + T_C$ periods; this results in a shift of all safety stock held at the LDCs to the FC (from the orange region to the blue one). Intuitively, when T_C is small but T_L is large, the LDCs can commit a fast delivery time to increase revenue at the expense of additional safety stock that is shifted from the FC to the LDCs. On the other hand, if T_L is relatively small but T_C is large, holding safety stock at the LDCs becomes too expensive and the optimal guaranteed delivery time to the customers increases, attracting less demand (and lower revenue); this impact is mitigated by staging safety stock at the FC, which permits offering a delivery time of $T_C + T_L$, as opposed to the maximum possible value of $T_C + T_L + T_F$ that would apply under the MTO-strategy. As the figure illustrates, when both T_L and T_C take large values (e.g., $T_L \geq 30$ and $T_C \geq 30$), the resulting committed service time is maximal due to the high holding costs associated with both echelons, leading to the relatively low expected revenue resulting from an MTO-strategy (upper right pink regions).

Observe that as we progress through the subfigures in Figure 1.5, the area associated with the optimality region for an MTO-strategy (pink region) decreases as the magnitude of $J(N)$ increases from $J(N) = 1$ towards $J(N) = N$. To provide an intuitive explanation for this trend, note that in the $J(N) = 1$ extreme case, the total demand depends only on the optimal committed delivery time, and the addition of an LDC does not generate any new demand, but serves a subset of the existing market, and therefore, cannot create additional revenue to offset its holding cost and fixed cost. When $J(N) = 1$, we therefore hit the threshold values for the MTO-strategy at the smallest possible value of the optimal committed service time, because no additional revenue is gained to offset the corresponding costs. At the other extreme, when $J(N) = N$, we see the greatest potential for additional revenue with the addition of an LDC because each LDC covers a new independent market, which leads to maximal threshold values associated with the change to an MTO-strategy.

1.5.3 Results when T_C depends on N

Section 1.4.4 proposed a non-increasing functional form $T_C(N) = T_0/N^{\theta}$ to account for the possibility that adding new LDCs may improve the delivery speed to customers, where T_0 is an initial physical delivery time to customers before adding any new LDCs, and θ is a non-negative parameter that characterizes the relationship between the number of LDCs and the physical delivery time. To simulate the effects of adding LDCs on the change in the physical delivery time when physical lead time depends on the number of LDCs, we illustrate the optimal strategy regions in the (T_0, N) space for various values of θ in Figure 1.6 (where T_0 varies from 1 to 40 and N from 1 to 10, with $J(N) = N$, $\bar{\sigma} = 200$, and all other parameter values the same as those used in previous tests).

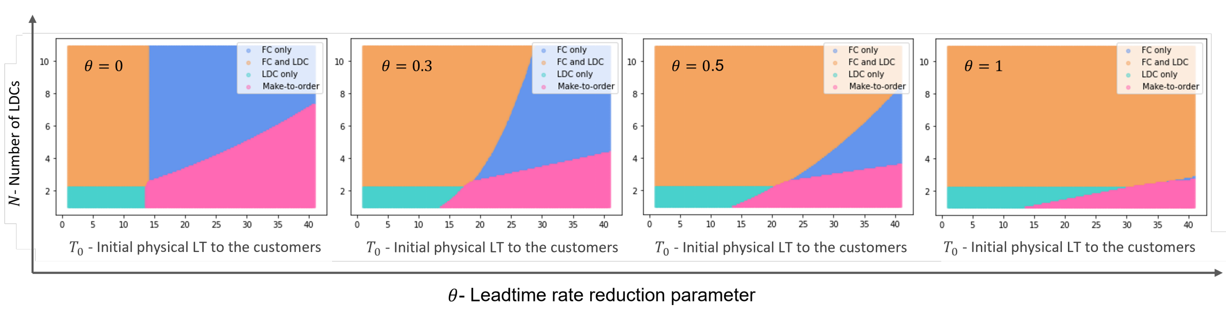


Figure 1.6: Optimality regions in the (T_0, N) space with different θ values.

The subfigures illustrate that as θ increases, the borders between the strategy regions become ‘stretched’ towards the lower right, and the areas of the FL- (orange) and LDC-strategy (turquoise) regions increase, while those associated with the FC-strategy (blue) and MTO-strategy (pink) decrease. Proposition 1.5.3 provides an analytical explanation for this trend.

Proposition 1.5.3. *For any given value of N, as θ increases, the threshold value of T_0 at which a transition occurs between safety stock placement strategies increases.*

Intuitively, adding new LDCs may decrease the average distance between a customer and its nearest LDC, so that a faster committed delivery service can be offered, and the value of θ determines the degree of this decrease in potential service time. Given a value of initial physical delivery time T_0, a larger θ implies that the system can offer a faster delivery time as N increases, reducing the threshold value of N at which a transition occurs between optimal strategies. As a result, given a fixed T_0, the borders between strategy regions move downward along the N-axis in Figure 1.6 (with the exception of the border between the FL- and LDC-strategy, which is independent of the physical lead time to customers).

1.6 Parametric analysis

In the previous section, we focused on analyzing the three decision variables (s, ℓ, and N), the demand growth model, and the associated key parameters that primarily determine the structure of an optimal solution. However, additional system parameters, such as the revenue function parameters, demand uncertainty, and relative holding cost values, are also crucial in determining the structure of an optimal solution. In this section, we provide a deeper analysis of the market size, demand uncertainty, and the unit holding cost parameters at the two echelons, and their impacts on the optimal solution structure.

1.6.1 Impacts of market size and demand uncertainty

The scale and uncertainty associated with customer demands serve as two key factors in inventory management. Our model uses a to denote maximum market demand per period, while $\bar{\sigma}$ corresponds to the standard deviation of periodic demand. We next consider how these two key parameters affect optimal inventory placement decisions. We tested the model under different combinations of small ($a = 10,000$) and large ($a = 20,000$) market sizes, as well as low ($\bar{\sigma} = 100$), medium ($\bar{\sigma} = 200$), and high ($\bar{\sigma} = 300$) values of demand uncertainty, assuming $J(N) = N$ and T_C is independent of N. Unless otherwise stated, all other parameters take the same values used in our previous numerical tests. Figure 1.7 illustrates the optimal strategy regions under different market demand attribute values.

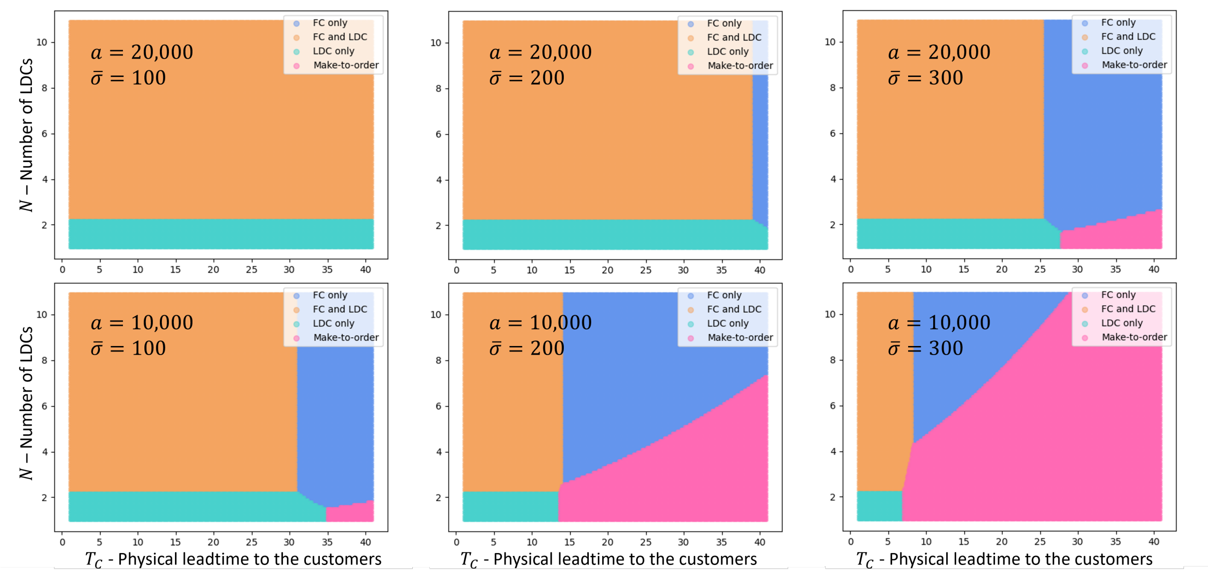


Figure 1.7: Optimal strategy regions in the (T_C, N) space under different market size (a) and demand uncertainty ($\bar{\sigma}$) values.

To understand the trends shown in Figure 1.7, we begin by considering the subfigure on the bottom left, which corresponds to a market with small size ($a = 10,000$) and low demand uncertainty ($\bar{\sigma} = 100$), in which the optimality region corresponding to the FL-strategy (orange region) occupies the most area in the (T_C, N) space. As the demand uncertainty increases (moving rightward through the subfigures in the bottom row), the area associated with the MTO-strategy (pink region) becomes larger, expanding from the lower right corner towards the upper left, implying a greater likelihood of the optimality of an MTO-strategy as uncertainty ($\bar{\sigma}$) increases. Next, consider the trends within the columns from bottom to top, as the market size (a) increases. It is notable that increasing a from 10,000 to 20,000 shifts the border between the orange and the blue regions to the right such that the area associated with the FL-strategy (orange) region is enlarged, while the regions corresponding to the FC-strategy (blue) and the MTO-strategy (pink) become narrower (and are eliminated at $\bar{\sigma} = 100$). This implies that for a fixed value of demand uncertainty, a larger market size leads to a larger range of T_C and N values such that the FL-strategy is optimal. Recall that in Section 1.4.2, we defined β as a parameter associated with the total system revenue, where $\beta = a\rho - b\rho^{u+1} = a^{(u+1)/u}b^{-1/u}u(u+1)^{-(1+u)/u}$. Therefore, increasing the value of a corresponds to increasing the value of β and, therefore, to additional revenue. The standard deviation $\bar{\sigma}$, on the other hand, only affects the system's holding costs. All else being equal, a larger standard deviation leads to greater safety stock, and thus, higher holding cost. A small market size with high demand uncertainty (as the bottom-right subfigure depicts) produces low revenue but requires high system holding cost to ensure short delivery times, which leads to a preference for an MTO-strategy. With a larger market size and smaller demand uncertainty (as shown in the subfigure on the top left), the model prescribes holding safety stock at the LDCs only (when the number of LDCs is small) or at both echelons (when the number of LDCs is large), which enables higher revenue and provides the smallest possible committed delivery time (that is, $\ell = T_C$), while maintaining a low system holding cost because of the relatively low value of $\bar{\sigma}$.

1.6.2 Impact of unit holding costs

The holding cost per unit per period at the FC, h_F, and at the LDCs, h_L, are also key parameters that affect optimal safety stock placement. Assuming $h_L \geq h_F$ (which typically holds in practice), we test four different combinations of the values of h_L and h_F and consider their impacts on the optimal solution structure using a demand growth function $J(N) = N$. The results are illustrated in Figure 1.8. A comparison of subfigures in the left column reveals that when we increase the value of h_F, the border between the FL-strategy region and the LDC-strategy region increases in N. We observe the same phenomenon when contrasting subfigures in the right column. This effect is formally stated in Corollary 1.6.1.

Corollary 1.6.1. *When T_C is fixed, for any valid $J(N)$ function, the threshold value of N at which $\Pi_{FL}(N) = \Pi_L(N)$ increases as the relative holding cost ratio h_F/h_L increases.*

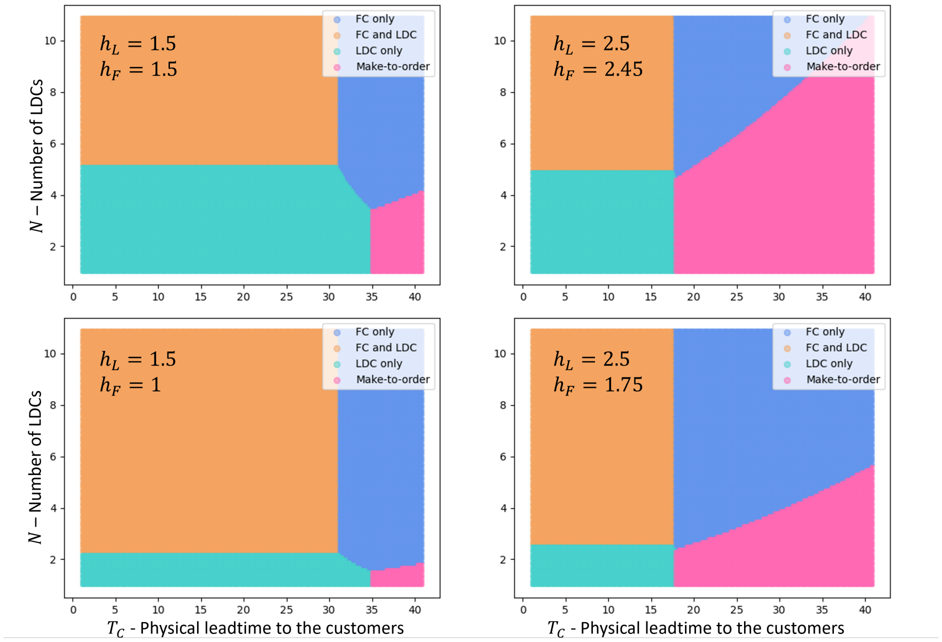


Figure 1.8: Optimality regions in the (T_C, N) space with different holding costs.

Corollary 1.6.1 implies that for a system with a small number of LDCs in which it is optimal to hold stock at both the FC and the LDCs, when the ratio of the periodic unit holding cost at the FC to that at an LDC increases, a shift of all stock to the LDCs may result at optimality. Another interesting property is revealed by comparing the subfigures within the same row. For example, by comparing subfigures in the top row, we see that when the value of h_L increases, the border between the FL-strategy region (orange) and the FC-strategy region (blue), and that between the LDC-strategy region (turquoise) and the MTO-strategy region (pink) decrease in T_C. This is shown analytically in Proposition 1.6.2.

Proposition 1.6.2. *When N is fixed, for any valid $J(N)$ function, the threshold values of T_C at which $\Pi_{FL}(N) = \Pi_F(N)$ and $\Pi_L(N) = \Pi_O(N)$ decrease as h_L increases.*

Proposition 1.6.2 implies that as the holding cost the LDCs increases, a system that holds stock at both echelons at optimality may transition to the FC-strategy if the system cannot offer a smaller committed delivery time. Similarly, if a system holds stock at the LDCs only at optimality, if the unit holding cost at LDCs increases, the system may require a transition to the MTO-strategy if it is not able to reduce the physical lead time to customers.

1.6.3 Impact of market growth rate and physical leadtime factor

Section 1.3.1.2 characterized a market growth function of the form $J(N) = N^\gamma$, where $0 \leq \gamma \leq 1$. We also proposed a model in which the physical leadtime to customers may depend on the number of LDCs in Section 1.4.4, using $T_C(N) = T_0/N^\theta$, where T_0 is a positive parameter and $\theta \geq 0$. The parameters γ and θ determine the rate of demand growth and delivery time decrease that are possible as N increases. Thus, we are interested in gaining some insight on how the optimal solution structure will be affected by these parameters. In this section, we consider the parameter values $\gamma \in \{0, 0.5, 1\}$ and $\theta \in \{0, 0.3, 0.5\}$, using the same values for all other parameters as in the previous tests. Figure 1.9 illustrates the results for all combinations of θ and γ in the (T_0, N) space.

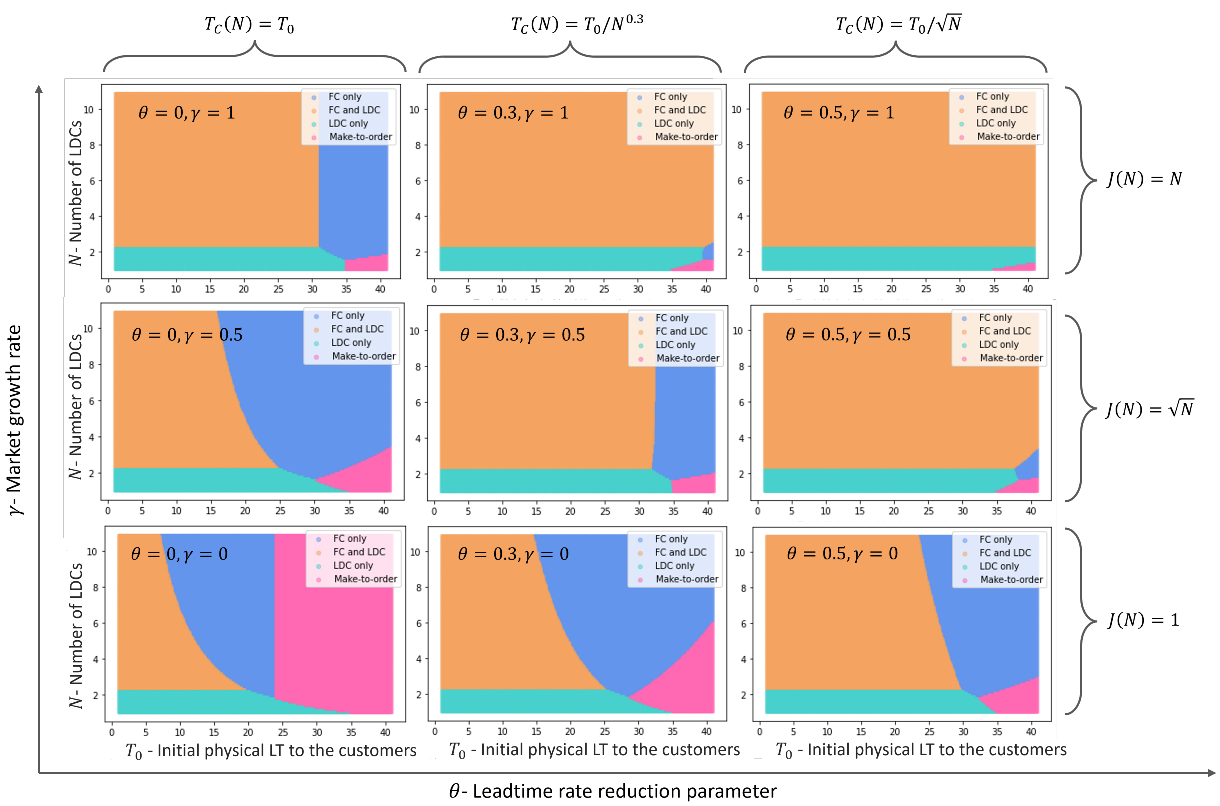


Figure 1.9: Optimality regions in the (T_0, N) space under different $\gamma - \theta$ combinations.

Figure 1.9 contains nine subfigures arranged in a 3×3 matrix. Each row includes three subfigures with the same value of γ but different values of θ, i.e., each row compares different physical leadtime reduction rates with the same market growth model. Similarly, each column compares three entries that have the same physical leadtime reduction rate, with different values of market growth rate as N increases. Within each column of subfigures, as γ increases (moving from the bottom figure to the top one), the orange-blue border and the blue-pink border are shifted toward the lower right, and the areas corresponding to blue and pink regions decrease. This phenomenon can be partially explained via Proposition 1.6.3.

Proposition 1.6.3. *When $T_C(N) = T_0/N^\theta$ and θ is fixed, using a market growth function $J(N) = N^\gamma$ with $0 \leq \gamma \leq 1$, for a given value of N, the threshold values of T_0 at which $\mathit{\Pi}_{FL}(N) = \mathit{\Pi}_F(N)$ and $\mathit{\Pi}_F(N) = \mathit{\Pi}_O(N)$ hold increase as γ increases.*

When θ is fixed, given a value of T_0, intuitively, if γ is larger, the total demand grows faster as we add LDCs to the system, which increases holding cost at the same time. This accelerates the change in expected profit with the addition of a new LDC, and, as a result, the transition between optimal stock strategies may occur at a smaller number of LDCs, N. We find a similar trend when comparing the subfigures within each row. With a fixed market growth model, as the value of θ increases, the borders shift to the right and the orange region increases in size, while the blue and pink regions are narrower. This trend has been explained in Section 1.5.3.

1.7 Model extension: Multiple, heterogeneous markets

This section generalizes the system to permit the FC to serve a set of heterogeneous markets. Let M denote the number of heterogeneous markets, where N_m is the number of LDCs serving market m for $m = 1, ..., M$. We assume that LDCs within a market are identical, providing the same level of service and responsiveness to customers in the market, while the LDCs that serve distinct markets are heterogeneous. Thus, all customers within a market have the same committed delivery time, ℓ_m (although different markets may have different delivery times), while the price, p, will be the same for all markets (this is not an unusual approach for companies that have web-based sales nationally).

Let s_m and $T_{L,m}$ denote the committed internal supply time and the physical leadtime to each LDC within market m, respectively, and let $T_{C,m}(N_m)$ denote the physical leadtime to customers in market m when there are N_m LDCs in the market. All LDCs are supplied by a single FC with inbound physical leadtime T_F. Figure 1.10 illustrates the system with market heterogeneity.

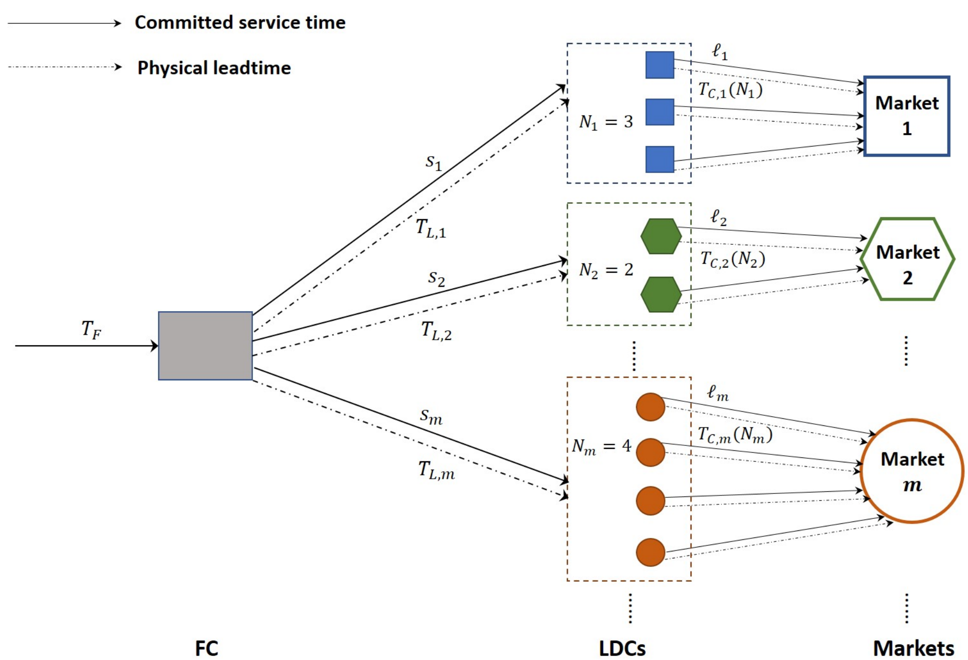


Figure 1.10: System with market heterogeneity.

1.7.1 Model formulation with heterogeneous markets

Let n_m denote the number of customers in market m when there is only one LDC serving the market. Assuming that each customer in the market has probability $P_m(p, \ell_m)$ of ordering the product, we have a mean per-period demand in market m (when it is only served by one LDC) equal to $\mu_m(p, \ell_m) = a_m - b_m p^u \ell_m^v = n_m P_m(p, \ell_m)$, and a variance of $\sigma_m^2(p, \ell_m) = n_m P_m(p, \ell_m)(1 - P_m(p, \ell_m))$. As discussed in Section 1.3.1.1, we use an upper bound on variance to approximate safety stock and holding costs within each market, e.g., setting $\sigma_m^2(p, \ell_m)$ as $\bar{\sigma}_m^2$, that is, $0.25n_m$, for $m = 1, ..., M$. Let $J_m(N_m)$ denote the demand growth function for market m. When market m contains N_m LDCs, the expected demand in the market equals $J_m(N_m)\mu_m(p, \ell_m)$, while the per-period expected demand and standard deviation seen by each LDC serving this market become $\mu_{L,m}(p, \ell_m, N_m) = \frac{J_m(N_m)}{N_m}\mu_m(p, \ell_m)$ and $\sigma_{L,m}(N_m) = \bar{\sigma}_m\sqrt{\frac{J_m(N_m)}{N_m}}$, respectively. The system revenue is therefore written as

$$R(p, \boldsymbol{\ell}, \mathbf{N}) = p\sum_{m=1}^{M} N_m\mu_{L,m}(p, \ell_m, N_m) = \sum_{m=1}^{M} J_m(N_m)\left(a_m p - b_m p^{u+1}\ell_m^v\right), \quad (1.9)$$

where $\boldsymbol{\ell} = (\ell_1, ..., \ell_M)$ and $\mathbf{N} = (N_1, ..., N_M)$. Within market m, the standard deviation of demand during effective leadtime seen by each LDC is $\sigma_{L,m}(N_m)\sqrt{T_{C,m}(N_m) + s_m - \ell_m}$, for $m = 1, ..., M$. Thus, the total safety stock holding cost at all LDCs across all markets equals

$$\begin{aligned} HC_L(\mathbf{s}, \boldsymbol{\ell}, \mathbf{N}) &= \sum_{m=1}^{M} N_m h_{L,m} k_{L,m} \sigma_{L,m}(N_m)\sqrt{T_{C,m}(N_m) + s_m - \ell_m} \\ &= \sum_{m=1}^{M} h_{L,m} k_{L,m} \bar{\sigma}_m \sqrt{J_m(N_m) N_m \left(T_{C,m}(N_m) + s_m - \ell_m\right)}, \end{aligned} \quad (1.10)$$

where $\mathbf{s} = (s_1, ..., s_M)$. Note that the standard deviation of effective lead-time demand seen by the FC from the LDCs serving market m equals $\sigma_{L,m}(N_m)\sqrt{N_m(T_F + T_{L,m} - s_m)}$, which leads to a pooled demand uncertainty across all markets seen by the FC equal to

$$\sigma_F(\mathbf{s}, \mathbf{N}) = \sqrt{\sum_{m=1}^{M} N_m \sigma_{L,m}^2(N_m)(T_F + T_{L,m} - s_m)} = \sqrt{\sum_{m=1}^{M} \bar{\sigma}_m^2 J_m(N_m)(T_F + T_{L,m} - s_m)}.$$

Letting $T_{C,m}(N_m) = \frac{T_{0,m}}{N_m^\theta}$, where $T_{0,m}$ is the base physical delivery time to customers in market m with a single LDC, the model formulation with multiple, heterogeneous markets can be written as follows:

$$
\begin{aligned}
[\mathbb{PG}] \quad \max \Pi(p, \mathbf{s}, \boldsymbol{\ell}, \mathbf{N}) = & \sum_{m=1}^{M} J_m(N_m)\left(a_m p - b_m p^{u+1} \ell_m^v\right) - \sum_{m=1}^{M} N_m K_m \\
& - h_F k_F \sqrt{\sum_{m=1}^{M} \bar{\sigma}_m^2 J_m(N_m)(T_F + T_{L,m} - s_m)} \\
& - \sum_{m=1}^{M} h_{L,m} k_{L,m} \bar{\sigma}_m \sqrt{J_m(N_m) N_m \left(\frac{T_{0,m}}{N_m^\theta} + s_m - \ell_m\right)} \\
\text{s.t.} \quad & T_{L,m} \le s_m \le T_F + T_{L,m}, \quad \forall m = 1, ..., M, \\
& \frac{T_{0,m}}{N_m^\theta} \le \ell_m \le s_m + \frac{T_{0,m}}{N_m^\theta}, \quad \forall m = 1, ..., M, \\
& s_m, \ell_m \ge 0, \quad \forall m = 1, ..., M, \\
& 1 \le N_m \le N_{UB,m}, \ N_m \in \mathbb{Z}^+, \quad \forall m = 1, ..., M, \\
& p \ge 0.
\end{aligned}
$$

1.7.2 Solution approach and convexity analysis

The approach discussed in Section 1.4.1 to determine an optimal price is also applicable to problem $[\mathbb{PG}]$. By setting $\frac{\partial \Pi(p,\mathbf{s},\boldsymbol{\ell},\mathbf{N})}{\partial p} = 0$, the optimal price with multiple heterogeneous markets can be written as

$$
p^*(\boldsymbol{\ell}, \mathbf{N}) = \left(\frac{A(\mathbf{N})}{(u+1)B(\boldsymbol{\ell}, \mathbf{N})}\right)^{\frac{1}{u}}, \tag{1.11}
$$

where $A(\mathbf{N}) = \sum_{m=1}^{M} J_m(N_m) a_m$ and $B(\boldsymbol{\ell}, \mathbf{N}) = \sum_{m=1}^{M} J_m(N_m) b_m \ell_m^v$. By replacing the price variable in the objective function of $[\mathbb{PG}]$ with Equation (1.11), the objective function becomes

$$\begin{aligned}
&\sum_{m=1}^{M} J_m(N_m) \left(a_m \left(\frac{A(\mathbf{N})}{(u+1)B(\boldsymbol{\ell}, \mathbf{N})} \right)^{\frac{1}{u}} - b_m \left(\frac{A(\mathbf{N})}{(u+1)B(\boldsymbol{\ell}, \mathbf{N})} \right)^{1+\frac{1}{u}} \ell_m^v \right) \\
&- \sum_{m=1}^{M} N_m K_m - h_F k_F \sqrt{\sum_{m=1}^{M} \bar{\sigma}_m^2 J_m(N_m)(T_F + T_{L,m} - s_m)} \\
&- \sum_{m=1}^{M} h_{L,m} k_{L,m} \bar{\sigma}_m \sqrt{J_m(N_m) N_m \left(\frac{T_{0,m}}{N_m^{\theta}} + s_m - \ell_m \right)}. \qquad (1.12)
\end{aligned}$$

For a fixed $\mathbf{N}$, i.e., given the number of LDCs in each market, we can show that the objective function (1.12) is convex in $(\mathbf{s}, \boldsymbol{\ell})$ when $0 < v \leq 1$, which leads to 4^M candidate extreme points (see Appendix A.5), implying that the number of extreme point solution candidates for a given $\mathbf{N}$ increases exponentially as the number of markets, M, increases. To determine the optimal $\mathbf{N}$, we may apply the same approach used in Section 1.4.2, by enumerating all possible integer values of N_m for each market.

However, if M is sufficiently large and/or the range of N_m for one or more markets is sufficiently broad, enumerating all possible combinations of N_m values and solving for the optimal extreme point solution by comparing all 4^M resulting objective values will be impractical. Heuristic algorithms will therefore be required to determine high-quality solutions in such cases, although space and scope considerations require us to leave this as a topic for future research.

1.7.3 Numerical results

In the case of multiple, heterogeneous markets, it is no longer straightforward to analytically derive simple threshold conditions characterizing transitions between optimal safety stock placement strategies due to the dimensional complexity. Instead, we conducted an additional set of numerical tests to illustrate how optimal safety stock strategies transition for multiple markets as system parameters change. To simulate a heterogeneous system, we randomly generate parameters associated with five heterogeneous markets. The parameters for the FC and each market are displayed in Table 1.5. With five markets, the total number of solution candidates for optimal $(\mathrm{s}, \boldsymbol{\ell})$ values is 1,024, where each solution candidate consists of (s_m, ℓ_m) for $m = 1, ..., 5$, and each pair of (s_m, ℓ_m) determines a safety stock strategy for the corresponding market m.

Table 1.5: Parameters used in numerical tests for the FC and each market

	a_m	b_m	$\bar{\sigma}_m$	$T_{L,m}$	h_L	k_L	$J_m(N_m)$	T_F	h_F	k_F
FC	-	-	-	-	-	-	-	30	2.202	4.5
$m = 1$	9,576	144	66.957	22	1.64	3	$J_1(N_1) = 1 + 6.015(1 - e^{1-N_1})$	-	-	-
$m = 2$	3,731	237	37.450	19	1.91	3	$J_2(N_2) = \sqrt{N_2}$	-	-	-
$m = 3$	10,043	137	59.410	25	2.98	3	$J_3(N_3) = N_3$	-	-	-
$m = 4$	7,225	283	44.677	18	2.98	3	$J_4(N_4) = 1$	-	-	-
$m = 5$	7,438	234	55.518	22	2.50	3	$J_5(N_5) = 1 + 8.003(1 - e^{1-N_5})$	-	-	-

To be consistent with our single-market computational tests, we considered various values of the physical leadtime to the customers ($T_{C,m}$) and the number of LDCs (N_m) in each market m (note that we assume $T_{C,m}(N_m) = T_{0,m}$, i.e., we set $\theta = 0$ for all $m = 1, ..., M$). To do this, we created 6 test instances that consider all combinations in which each N_m is restricted to the range of [1,5) or [5, 10), while each $T_{C,m}$ must take a value within [1,15), [15, 20), or [20,30), for every market in each instance. Table 1.6 displays the values of $T_{C,m}$ and N_m for each market, as well as the optimal solutions for each m for all instances.

Table 1.6: Numerical test instances and optimal strategies for systems with 5 heterogeneous markets

Instance	$[N_m]$	$[T_{C,m}]$	Optimal $[(s_m, \ell_m)]$ - Extreme point solution	Optimal stock strategy
1	3, 3, 1, 2, 2	8, 7, 5, 1, 12	(52, 8), (49, 7), (55, 5), (48, 1), (52, 12)	L, L, L, L, L
2	3, 3, 1, 2, 2	19, 15, 18, 16, 16	(52,71), (49, 15), (55, 73), (48, 64), (52, 68)	O, L, O, O, O
3	3, 3, 1, 2, 2	27, 26, 24, 20, 26	(52,79), (49, 75), (55, 79), (48, 68), (52, 78)	O, O, O, O, O
4	7, 7, 9, 6, 9	8, 7, 5, 1, 12	(22, 8), (19, 7), (25, 5), (18, 1), (22, 12)	FL, FL, FL, FL, FL
5	7, 7, 9, 6, 9	19, 15, 18, 16, 16	(22, 41), (19, 15), (25, 43), (18, 34), (22, 16)	F, FL, F, F, FL
6	7, 7, 9, 6, 9	27, 26, 24, 20, 26	(22, 49), (19, 45), (25, 49), (18, 38), (22, 48)	F, F, F, F, F

We can visualize the various safety stock strategies applied among the markets using a two-dimensional map of $T_{C,m}$ and N_m values as shown in Figure 1.11. We can see that when all markets

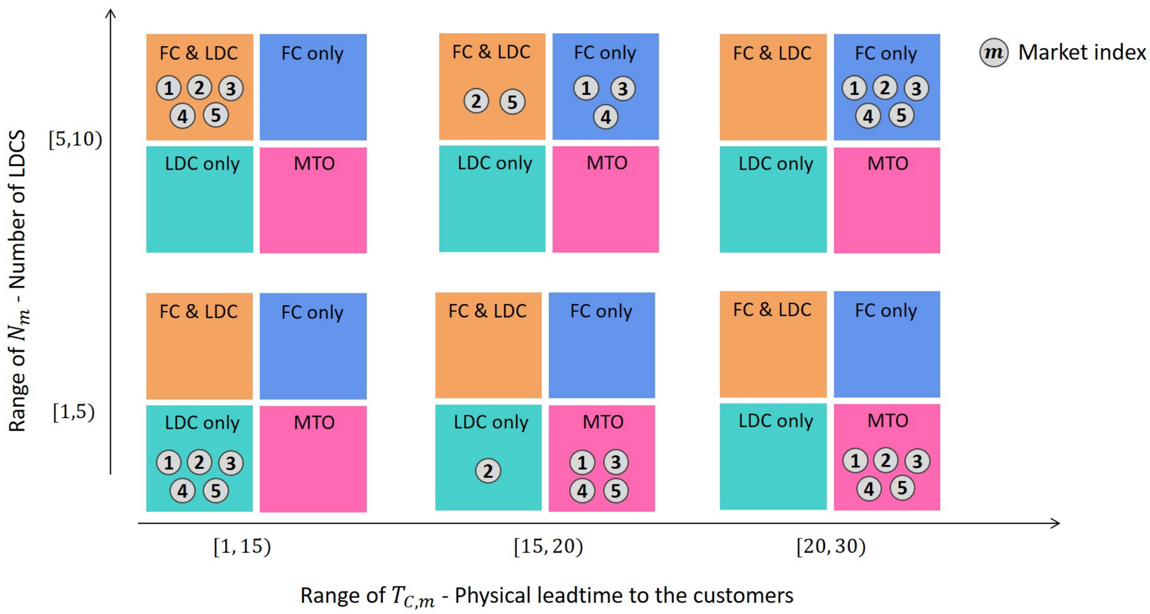


Figure 1.11: Optimal strategy distribution for heterogeneous systems.

have fewer than five LDCs and the customer delivery time is less than 15 periods, each market only holds safety stock at their respective LDCs (in the bottom-left corner of each four-color block, all markets are located in the turquoise region). As the customer delivery time increases to between 15 and 20 periods (moving to the right in the horizontal direction), some markets switch to a make-to-order stock strategy (e.g., markets 1, 3, 4, and 5 move to the pink MTO region, while market 2 continues to hold inventory at the LDCs). Eventually, all markets will use a make-to-order strategy when the customer delivery time in each market is greater than 20 periods. Similarly, if we start with small values of $T_{C,m}$ (1 to 14) and larger values of N_m (5 to 9), as displayed in the upper-left four-color block figure, all markets hold safety stock at both echelons (locating in the orange region). As the customer delivery times increase, while holding the N_m values fixed, markets will begin to move safety stock out of the LDCs and to the FC only (markets 1, 3, and 4 move to the blue region, holding safety stock only at the FC, while markets 2 and 5 continue their strategy of holding safety stock at both echelons), since the LDC holding costs are sufficiently high such that they are not offset by the system revenue available at lower delivery times in all markets. When the delivery time required by all markets is longer than 20 periods, the system will apply an FC-only strategy for each of its markets (all markets move to the blue region). These shifts in stock placement strategies among the markets display very similar trends to those illustrated in Section 1.4.3 for the single-market case.

1.8 Conclusions and future work

This chapter proposed a two-echelon, fulfillment-time-dependent delivery model that determines the product price, the internal supply time from the upstream stage to the downstream stage, and the delivery time committed to customers. The model's solution determines an optimal safety stock placement strategy. This model can be used to determine the service time, network expansion strategy, inventory placement, and pricing strategies for a single product under uncertain, fulfillment-time-dependent demand, while accounting for potential market growth based on the addition of new facilities. Our numerical results and parametric analysis illustrated tools for inventory placement decisions based on the number of facilities and the delivery time committed to customers.

This work provides some insights for real-world fulfillment systems in which demand depends on the committed service time and the price, and the market size and inventory placement decisions are impacted by the number of facilities and the scale of the system. Future work may consider including truckload transportation costs to account for transportation capacity limits and shipping cost structures. Future work may also consider a multiple-product network in which the leadtime between the upstream hub and each downstream is product-dependent, as well as the application of the modeling approach to networks with more complex structures.

2. OPTIMIZATION MODELS FOR FLEET COMPOSITION WITH TRUCKLOAD AND LESS-THAN-TRUCKLOAD SHIPPING OPTIONS

2.1 Introduction

In today's dynamic economic environment, logistics businesses are confronted with the mounting challenge of meeting ever-growing customer expectations. Customers now expect an elevated level of service that encompasses timeliness, reliability, and cost-effectiveness. Companies must strive to meet these heightened customer expectations while simultaneously dealing with rising operating costs, increased competition, and wage pressures stemming from inflation.

Many firms rely on private fleets to provide customers with reliable and efficient shipping services. By having a dedicated fleet of vehicles, they can have greater control over logistics operations and enhance the security of the goods during transportation, thereby mitigating the risks of delays and damages. However, due to the uncertainty of demand, it is challenging to maintain an in-house fleet of a fixed size that can consistently meet fluctuating demand while also ensuring high fleet utilization.

Forming a small fleet requires less initial capital investment and lower operating costs, but may not possess sufficient internal capacity to effectively meet the shipping requirements, resulting in hiring external carriers to complement existing capacities at a much higher shipping rate. On the other hand, having a very large fleet may lead to certain vehicles being idle for extended periods, resulting in a substantial opportunity cost and a waste of transportation resources. Therefore, companies must make careful decisions when it comes to choosing between insourcing and outsourcing, considering the trade-offs of acquiring and maintaining internal trucks with higher ownership and operating costs, versus outsourcing to external carriers with higher shipping expenses.

In this chapter, we address the challenge of establishing a new truck fleet for fulfilling uncertain demands using internal truckload (TL) capacity over a future time span, while taking into account the usage of less-than-truckload (LTL) services provided by external carriers within each period.

The fleet may consist of one or multiple truck types, and we must determine both the types of trucks to acquire and the number of trucks of each type, aiming to minimize the total expected costs in a period.

The major contributions of this work are as follows. (i) We consider stochastic demand for the fleet composition problem and develop a solution approach that allows using a mixture of internal fleet and external carriers even when the demand does not exceed the aggregate internal TL capacities. (ii) We analytically characterize optimal fleet sizing/composition strategies for both homogeneous and heterogeneous fleets, and obtain closed-form solutions for certain demand distributions. (iii) We propose a two-stage stochastic programming model to tackle the general fleet composition problem given multiple truck types and develop a decomposition-based heuristic algorithm for solving the proposed model. The performance of the algorithm is evaluated and compared with the results of using the sample average approximation method through extensive numerical experiments.

The remainder of this chapter is organized as follows. Section 2.2 provides a survey of the relevant studies. Section 2.3 includes the solution approach and analytical results for a homogeneous fleet. Section 2.4 presents the solution approach for fleet composition with two truck types, and includes the two-stage stochastic programming model for a heterogeneous fleet with multiple truck types, followed by the sample average approximation method and the decomposition-based heuristic. Computational experiments and sensitivity analysis results are included in Section 2.5. Section 2.6 provides the conclusions and future research directions.

2.2 Literature review

Our work is grounded in the problem class known as fleet composition. Given a set of vehicle types, the fleet composition problem is to determine the optimal number of vehicles of each type for a heterogeneous fleet. Since it was first introduced by Kirby [26], the fleet composition problem has been widely extended in terms of various model assumptions and problem settings. Gould [27] proposes a linear programming model to determine the optimal fleet composition, and solves a real-world case given a set of vehicle options and deterministic, seasonal demand. Etezadi and Beasley [28] develop a mixed-integer programming model for a single-depot distribution system that serves multiple customers with known demand, using both hired vehicles and an owned fleet of composite vehicle types in terms of capacity. Wu et al. [29] solve a fleet composition and demand allocation problem with a linear programming model for truck-rental firms with known demand, in which vehicles are heterogeneous in capacity and age, and a two-phase approach is developed. Loxton et al. [30] propose a model that minimizes the total expected cost of owning and maintaining a fleet, when considering a discrete probability distribution for the number of vehicles required for each vehicle type per period, which determines the number of vehicles to procure for each type. They further develop an effective algorithm for the model based on dynamic programming and golden-section search that solves large-scale problems within a very short time. Konur and Geunes [31] present a mixed-integer-nonlinear programming model that integrates districting, fleet composition, and inventory management for a multi-retailer distribution system, in which the vehicle capacity can be shared within each district, assuming a known and constant demand rate at each retailer. The model is solved by a column generation based heuristic. More recently, Shehadeh et al. [32] study the fleet sizing and allocation problem for an on-demand last-mile service system under demand uncertainty, in which the vehicle allocation decision, the routing plan, and the customer assignment for each service region are determined. The authors propose a stochastic programming model and a distributionally robust model to address known and unknown demand distributions, respectively.

One vital research strand in relation to fleet composition is the *Fleet Size and Mix Vehicle Routing Problem* (FSMVRP), in which the fleet composition decision is integrated into optimization models for the vehicle routing problem. The most representative work of this problem type is introduced by Golden et al. [33], which optimizes the FSMVRP for a heterogeneous fleet by minimizing the total costs of acquisition and routing, based on a set of customers with known locations and demand. The authors discuss several efficient heuristic algorithms and provide procedures for assessing the quality of the solutions. Dell'Amico et al. [34] study the *Fleet Size and Mix Vehicle Routing Problem with Time Windows* (FSMVRPTW) based on known demands and develop an insertion-based constructive heuristic that outperforms previously published heuristic results. Jabali et al. [35] present a strategic model for fleet composition and routing decisions based on continuous approximation techniques, which determines the number of vehicles in different capacity levels and the optimal delivery routes when minimizing the total cost, considering a given set of customers distributed in a circular service region at a constant density. Hiermann et al. [36] solve the FSMVRPTW for a fleet of electric vehicles with restrictions on available recharge stations, which is formulated as a mixed-integer programming model and is solved by a branch-and-price algorithm as well as a hybrid heuristic. Alinaghian et al. [37] present a model for a time-dependent FSMVRP with a cost-minimization objective that encompasses greenhouse gas emissions, and develop an improved adaptive large neighborhood search algorithm that can effectively solve large-scale instances. The importance of combining fleet composition and routing is also recognized in various industries [38].

Models developed for FSMVRP primarily focus on making decisions at the operational level, specifically when having access to information regarding routing aspects (such as the features of the service region, customer locations, and service time constraints), whereas our work aims to determine the long-term fleet composition and develop a model that minimizes the total expected cost under demand uncertainty at the strategic planning level, without requiring detailed routing information.

The other stream of studies in which our work is rooted is the fleet composition problem with both internal fleet and external carriers. The problem, defined by Ball et al. [39], is one of the earliest studies that consider both shipping options, in which an internal truck fleet serves a subset of the demand while the remainder is outsourced to an external carrier, given constant and known demand at the beginning of each planning period. Klincewicz et al. [40] investigate the fleet planning problem by formulating it as a single-source capacitated facility location model that allocates a limited number of internal trucks and a given set of outside carriers to each service sector, considering stochastic daily demand and service time constraints, in which outside carrier services are requested when the demand exceeds the capabilities of the internal fleet. Hall and Racer [41] develop methods that determine whether to serve a customer by internal vehicles or by an external carrier as well as optimize shipment frequency based on a square-root route length approximation, accounting for both transportation and inventory costs. Chu [42] presents a cost-minimization integer programming model for routing a heterogeneous fleet of trucks with known customer demand, in which any demand in excess of the total capacity of owned trucks will be outsourced to an external LTL carrier. A three-phase heuristic algorithm is developed for solving the proposed model, which can provide near-optimal solutions. Improved heuristic algorithms are developed based on similar problem settings in [43] and [44]. Krajewska and Kopfer [45] extend the traditional vehicle routing problem to an integrated transportation planning problem by considering several subcontracting options while owning a private fleet, which is solved using a tabu search heuristic algorithm with real-life data. In a similar context, Gahm et al. [46] address a vehicle routing problem with an internal fleet and multiple external carriers using a mixed-integer program, considering volume discounts offered by the external carriers. Several heuristic approaches are developed to provide effective solutions. Kaewpuang et al. [47] study a resource cooperation problem that allows sharing of trucks among shippers who have their own fleets through a vehicle pool, in which a third-party LTL carrier provides supplemental capacity if the number of trucks in the pool is insufficient to fulfill the demand. The authors propose an integer linear programming model and a stochastic programming model to tackle deterministic and random demand, respectively.

The majority of existing literature in relation to both internal fleet and external shipping options either assumes deterministic demands or, when considering stochastic demands, that the external carriers are used only for demand in excess of the aggregate capabilities of the internal trucks. In contrast, and more consistent with practice, we permit a mixture of internal TL and external LTL shipping in any period under demand uncertainty, even when total demand does not exceed internal fleet capacity. This leads to a generally more complex class of fleet composition problems that more accurately reflects expected costs in practice. Moreover, we provide closed-form results for both homogeneous and heterogeneous fleets with certain demand distributions when accounting for external shipping services, which is rarely seen in previous literature.

2.3 Homogeneous fleet

We first consider the fleet composition optimization problem for a homogeneous fleet that consists of a single truck type (later in Section 2.4 a heterogeneous fleet is studied). Assume demand in a period is a random variable X, following a continuous distribution with PDF $f(x)$ and CDF $F(x)$. Let n denote the size of the fleet (i.e., the number of trucks owned). We assume each truck, with a capacity of W, can perform at most one shipment from the origin to the destination in a single period. The maximum internal TL capacity is therefore equal to nW. In practice, the freight charge quoted by a third-party carrier for an LTL shipment commonly depends on the weight of the load, the distance between the origin and the destination, the type of goods, and other additional fees, while the average cost of a TL shipment is relatively stable when using internal fleet capacity. In our model, we use a fixed dispatching cost of A for internal trucks, which includes the handling costs, the average fuel charges, and the driver's salary, and allows a truck to carry up to W units of weight per trip. As we are interested in investigating the average operational behavior, we consider a per-period fixed ownership cost per truck, denoted by K. One may obtain K by allocating the total costs of procuring and maintaining a truck over a planning horizon to each period. We use a to denote the LTL freight rate (i.e., the cost of per unit shipped via LTL) when using external shipping services, assuming $aW > A$. The objective is to minimize the total expected cost when using the combination of internal fleet and external LTL services.

2.3.1 The break-even point

Our shipping policy allows the use of LTL service not only for the scenario when demand is in excess of total internal TL capacity, but for cases in which the internal fleet is partially utilized. Understanding the criteria and achieving the benefit of such policy requires addressing two essential issues: (i) what are the conditions under which using LTL is more economical than using internal TL capacity and (ii) what is the optimal fleet size that minimizes the total expected cost when using a mixture of internal TL capacity and external LTL shipments.

We consider a fraction α of the truck capacity $(0 < \alpha < 1)$, and let $Y = \alpha W$ such that

$aY = A < aW$. Apparently, when the size of the shipment equals W, dispatching a full internal truck (at a cost of A) is more economical as it leads to a lower shipping cost than outsourcing to external LTL carriers (at a cost of aW); when the shipment size is less than W but equals Y, using LTL is equivalent to shipping via internal TL in terms of the shipping cost, although the shipped truck is partially loaded if one chooses using TL. For any amount less than Y, using an internal truck is no longer beneficial since the cost of shipping via LTL is less than the flat rate of A. When owning more than one truck, for any shipment size that is greater than W, we first dispatch W units using a full internal truck, then repeat the same analysis for the remaining units, and so on. Therefore, Y is the break-even point at which the cost of shipping using either option is equivalent. In fact, this pattern of transitioning between two shipping options as the shipping quantity varies, as illustrated in Figure 2.1, is analogous to *carload discount* defined for quantity discount schedules (see e.g., [48, 49, 50]). The original carload discount schedule provides a pricing strategy for carriers based on Y, while in our model we use Y to trigger the transition between sending internal TL shipments and requesting external LTL shipments.

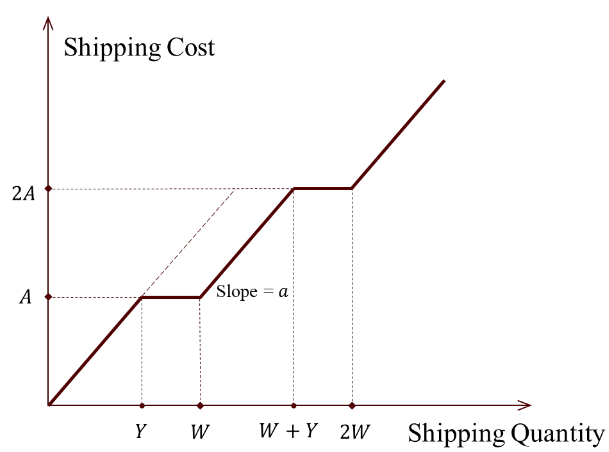


Figure 2.1: Discounted cost schedule for a homogeneous fleet.

2.3.2 The expected shipping quantities

We first characterize the expected number of TL shipments. Owning n trucks implies the number of TL shipments dispatched in any period is at most n. Observe that if demand is between Y and $W + Y$ with $n \geq 1$, we ship one truck that is loaded with up to W units and use LTL for any remaining amount; if it is between $W + Y$ and $2W + Y$ with $n \geq 2$, we ship two trucks that carry up to $2W$ units and outsource any leftovers to LTL carriers, and so on. Thus, the probability that we ship k trucks for any positive integer $k < n$ equals $F(kW + Y) - F((k-1)W + Y)$, while the probability that we ship all n internal trucks equals $1 - F((n-1)W + Y)$. As a result, the expected number of TL shipments using internal fleet in a period is

$$\begin{aligned} E_{TL}(n) &= \sum_{k=1}^{n-1} k\Big(F(kW+Y) - F\big((k-1)W+Y\big)\Big) + n\Big(1 - F\big((n-1)W+Y\big)\Big) \\ &= n - \sum_{k=0}^{n-1} F(kW+Y). \end{aligned} \tag{2.1}$$

Let x denote the observed demand in a period. If $0 < x - kW < Y$ and $x < nW$, it is optimal to ship the remaining $x - kW$ units via LTL after dispatching k full internal trucks. If $Y \leq x - kW < W$ and $x \leq nW$, the remaining amount $x - kW$ reaches the break-even point Y that allows dispatching an internal truck for the leftover amount at a lower shipping cost, although the shipped truck is partially loaded, and zero units are shipped via LTL. Note that when $x > nW$, any excess amount $x - nW$ must be shipped via LTL since all internal capacity is 100 percent utilized. This leads to three demand intervals that trigger using LTL service for minimizing the total expected cost. The corresponding units shipped via LTL are summarized in Table 2.1. The

Table 2.1: The demand intervals that trigger LTL shipments with $k = 0, \ldots, n-1$.

Demand intervals	$(kW, kW+Y)$	$[kW+Y, (k+1)W]$	(nW, ∞)
LTL units shipped	$x - kW$	0	$x - nW$

expected quantity shipped via LTL then equals

$$E_{LTL}(n) = \sum_{k=0}^{n-1} \int_{kW}^{kW+Y} (x - kW) f(x) dx + \int_{nW}^{\infty} (x - nW) f(x) dx. \tag{2.2}$$

Consequently, the cost-minimization model in terms of the fleet size is formulated as

$$\begin{aligned}
\underset{n \in \mathbb{Z}^+}{\text{Minimize}} \quad g(n) &= Kn + AE_{TL}(n) + aE_{LTL}(n) \\
&= Kn + A\left(n - \sum_{k=0}^{n-1} F(kW + Y)\right) \\
&\quad + a\left(\sum_{k=0}^{n-1} \int_{kW}^{kW+Y} (x - kW) f(x) dx + \int_{nW}^{\infty} (x - nW) f(x) dx\right). \quad (2.3)
\end{aligned}$$

2.3.3 Convexity analysis and the optimal solution

Since the value of n in the above model is a nonnegative integer, the objective function $g(n)$ in (2.3) is therefore defined at discrete values of n. However, we can show that $g(n)$ is *discretely convex* in n using second differences (see e.g., [51]), where $g(n)$ is discretely convex if and only if these second differences are nonnegative.

Suppose demand follows a continuous distribution with PDF $f(x)$ and CDF $F(x)$. Let $\ell(kW) = \int_{kW}^{\infty} (x - kW) f(x) dx$. With $aY = A$, we can rewrite the objective function in (2.3) as

$$g(n) = Kn + a\left(\sum_{k=0}^{n-1} \Big(\ell(kW) - \ell(kW + Y)\Big) + \ell(nW)\right). \tag{2.4}$$

Let $\Delta_n = g(n+1) - g(n)$ denote the first difference and $\Delta_{n+1} - \Delta_n$ denote the second difference. For any general continuous distribution, using the simplified objective function (2.4), we can obtain

$$\Delta_n = K + a\Big(\ell\big((n+1)W\big) - \ell(nW + Y)\Big), \tag{2.5}$$

and

$$\varDelta_{n+1} - \varDelta_n = a\Big(\ell\big((n+2)W\big) - \ell\big((n+1)W + Y\big) - \Big(\ell\big((n+1)W\big) - \ell(nW + Y)\Big)\Big). \quad (2.6)$$

Consider the function $\ell(r) = \int_r^{\infty}(x - r)f(x)dx$ and note that $\ell'(r) = F(r) - 1 \leq 0$, while $\ell''(r) = f(r) > 0$. Thus, $\ell(r)$ is a strictly convex and nonincreasing function of r. Let $h(k) = \ell\big(kW\big) - \ell\big((k-1)W + Y\big)$. Observe that $h'(k) = W\Big(F\big(kW\big) - 1\Big) - W\Big(F((k-1)W + Y) - 1\Big) = W\Big(F\big(kW\big) - F((k-1)W + Y)\Big) > 0$, $h(k)$ is therefore strictly increasing in k, which implies that

$$\varDelta_{n+1} - \varDelta_n = a\Big(h(n+2) - h(n+1)\Big) > 0. \quad (2.7)$$

This guarantees the discrete convexity of (2.4) in $n \in \mathbb{Z}^+$ for any continuous probability distribution.

It should be clear that for a one-dimensional discretely convex function $g(z)$ defined on the nonnegative integers $z \in \mathbb{Z}^+$, if z^* is the smallest nonnegative integer z such that $\varDelta_z = g(z+1) - g(z) \geq 0$, then z^* is a global minimizer of $g(z)$ on $\mathbb{Z}^+$. To see this, first note that by the definition of z^*, $g(z)$ is decreasing in z for $0 \leq z \leq z^*$ (because $g(z+1) - g(z) < 0$ for $0 \leq z \leq z^*$). For $z \geq z^*$, nonnegative second differences imply $\varDelta_{z^*+k+1} \geq \varDelta_{z^*+k}$ for any integer $k \geq 0$. When $k = 0$, this implies $\varDelta_{z^*+1} = g(z^*+2) - g(z^*+1) \geq \varDelta_{z^*} \geq 0$, so that $g(z^*+2) \geq g(z^*+1)$, while $\varDelta_{z^*} \geq 0$ is equivalent to $g(z^*+1) \geq g(z^*)$. By induction, we can show that $g(z^*+k+1) \geq g(z^*+k)$ holds for any $k \geq 0$. Hence, $g(z)$ is decreasing for $z < z^*$ and nondecreasing for $z \geq z^*$, implying that z^* is a global minimizer.

Next, note that a one-dimensional discretely convex function $g(z)$ defined for all $z \in \mathbb{Z}^+$ can easily be extended to a convex function $\tilde{g}(z)$ defined for all $z \in \mathbb{R}^+$. One such convex function is created by constructing a piecewise-linear function by connecting each pair of adjacent points, $(z, g(z))$ and $(z+1, g(z+1))$, with a line segment of slope $g(z+1) - g(z)$. Thus, for any z between some nonnegative integer c and $c+1$, we have $\tilde{g}(z) = g(c) + (z-c)(g(c+1) - g(c)) =$

$(1-\theta_z)g(c)+\theta_z g(c+1)$, where $\theta_z = z - c$ and $\theta_z \in [0,1]$. Clearly, if z^* minimizes $g(z)$, z^* also minimizes $\tilde{g}(z)$. Moreover, if z' minimizes $\tilde{g}(z)$ and $z' \in \mathbb{Z}^+$, then z' minimizes $g(z)$; because $\tilde{g}(z)$ is piecewise linear and convex with breakpoints at consecutive integers, if z' minimizes $\tilde{g}(z)$ and $z' \notin \mathbb{Z}^+$, then at least two alternative optimal solutions minimize $g(z)$, i.e., $\lfloor z' \rfloor$ and $\lceil z' \rceil$. Note that $\tilde{g}(z) = g(z)$ for any $z \in \mathbb{Z}^+$, while the piecewise linearity of $\tilde{g}(z)$ implies that an optimal solution always exists at some $z \in \mathbb{Z}^+$. Thus, where convenient, we can express expected costs in terms of the convex function $\tilde{g}(z)$ defined for $z \in \mathbb{R}^+$ rather than the discretely convex function $g(z)$ defined for $z \in \mathbb{Z}^+$.

2.3.3.1 Optimal solution for uniform demand distribution

It is straightforward to show that when the demand is uniformly distributed on $[l, \bar{D}]$, where l is the lower bound and $\bar{D}$ is the upper bound, the expected number of TL shipments equals

$$E^U_{TL}(n) = \frac{n(\bar{D}-Y)}{\bar{D}-l} - \frac{n(n-1)W}{2(\bar{D}-l)}, \tag{2.8}$$

and the expected number of units shipped via LTL equals

$$E^U_{LTL}(n) = \frac{nY^2 + (\bar{D}-nW)^2}{2(\bar{D}-l)}. \tag{2.9}$$

The resulting optimization model becomes

$$\underset{n\in\mathbb{Z}^+}{\text{Minimize}} \quad g_U(n) = Kn + An\left(\frac{\bar{D}-Y}{\bar{D}-l} - \frac{(n-1)W}{2(\bar{D}-l)}\right) + \frac{a\left(nY^2 + (\bar{D}-nW)^2\right)}{2(\bar{D}-l)}. \tag{2.10}$$

Consider the continuous function $\tilde{g}_U(n)$ defined in $n \in \mathbb{R}^+$ for the underlying discrete function $g_U(n)$ in $n \in \mathbb{Z}^+$. The second order derivative of $\tilde{g}_U(n)$ is $\tilde{g}''_U(n) = \frac{aW(W-Y)}{\bar{D}-l} > 0$ (because $Y < W$), implying $\tilde{g}_U(n)$ is strictly convex in $n \in \mathbb{R}^+$. Observe that $\tilde{g}_U(n)$ properly determines the expected total cost at each integer point, i.e., $\tilde{g}_U(\tilde{n}) = g_U(\tilde{n})$ holds for any integer $\tilde{n}$, and the discrete convexity of $g_U(n)$ thus naturally holds for all $n \in \mathbb{Z}^+$.

Because of strict convexity, at most one integer solution $\tilde{n}$ exists such that $g_U(\tilde{n}) = g_U(\tilde{n}+1)$, which implies for $n \in \mathbb{R}^+$, $\tilde{g}_U(n) > g_U(\tilde{n})$ for $n < \tilde{n}$ and $\tilde{g}_U(n) > g_U(\tilde{n}+1)$ for $n > \tilde{n}+1$, while $\tilde{g}_U(n) < g_U(\tilde{n})$ for $\tilde{n} < n < \tilde{n}+1$. We can show that

$$\tilde{n} = \frac{2\bar{D} - Y - W}{2W} - \frac{K(\bar{D} - l)}{aW(W - Y)} = n_U^* - \frac{1}{2}, \tag{2.11}$$

where $n_U^* = \frac{2\bar{D}-Y}{2W} - \frac{K(\bar{D}-l)}{aW(W-Y)}$ is the stationary point as well as the minimizer of $\tilde{g}_U(n)$. If $n_U^* - \frac{1}{2}$ has a integer value, i.e., $\tilde{n}$ is integer, then $\tilde{n}$ and $\tilde{n}+1$ serve as alternative optimal integer solutions. Otherwise, exactly one integer value falls between $\tilde{n}$ and $\tilde{n}+1$, and this integer value equal to $\lceil \tilde{n} \rceil = \lfloor \tilde{n} + 1 \rfloor$ is optimal.

2.3.3.2 Optimal solution for normal demand distribution

Suppose demand is normally distributed with mean μ and standard deviation σ. For the normal case, recall that the standard normal loss function is $L(z) = \int_z^\infty (u - z)\phi(u)du$. When X is normally distributed, it is well known that $\int_{kW}^\infty (x - kW)f(x)dx = \sigma L(z_{kW})$, where $z_{kW} = \frac{kW-\mu}{\sigma}$. According to Equations (2.1) and (2.2), we can write the expected number of TL and LTL shipments for normally distributed demand as follows:

$$E_{TL}^N(n) = n - \sum_{k=0}^{n-1} \Phi(z_{kW+Y}), \tag{2.12}$$

$$E_{LTL}^N(n) = \sum_{k=0}^{n-1} \Big(\sigma L(z_{kW}) - \sigma L(z_{kW+Y}) + Y\Phi(z_{kW+Y})\Big) - nY + \sigma L(z_{nW}), \tag{2.13}$$

where $\Phi(z)$ is the standard normal cumulative distribution function. The resulting optimization model becomes

$$\begin{aligned}\underset{n\in\mathbb{Z}^+}{\text{Minimize}} \quad g_N(n) =& Kn + a\left(\sum_{k=0}^{n-1}\Big(\sigma L(z_{kW}) - \sigma L(z_{kW+Y}) + Y\Phi(z_{kW+Y})\Big) - nY + \sigma L(z_{nW})\right)\\ &+ A\left(n - \sum_{k=0}^{n-1}\Phi(z_{kW+Y})\right). \end{aligned} \tag{2.14}$$

According to Equation (2.4), the objective function in (2.14) is equivalent to

$$g_N(n) = Kn + a\left(\sum_{k=0}^{n-1}\Big(\sigma L(z_{kW}) - \sigma L(z_{kW+Y})\Big) + \sigma L(z_{nW})\right). \tag{2.15}$$

We can easily verify that the second difference of function (2.15) is positive and thus $g_N(n)$ is discretely convex in $n \in \mathbb{Z}^+$ (see Appendix B.1). Due to the form of the normal distribution, it is not possible to obtain a closed-form expression for the optimal n that minimizes the value of $g_N(n)$. However, the maximum number of trucks a fleet may own is typically subject to some limitation (e.g., budget, facility capacity, etc.), and using binary search among a limited number of possible integer values can be quite efficient for determining the optimal fleet size.

2.3.4 Model adaptation to stochastic LTL rate

To be more consistent with practice, it is realistic to treat the LTL rate as a random variable. Suppose the LTL rate is random and independent of the demand X, with PDF $p(a)$ and mean value μ_a. Assume a nonnegative lower bound a_L on the LTL rate exists such that $a_L W \geq A$, and there is some upper limit a_U in the LTL market price. Based on Equation (2.4), with a stochastic LTL rate, the objective function for a general demand distribution can be modified to

$$g^a(n) = Kn + \int_{a_L}^{a_U} a\sum_{k=0}^{n-1}\left(\ell(kW) - \ell\left(kW + \frac{A}{a}\right)\right)p(a)da + \mu_a \ell(nW). \tag{2.16}$$

We can show (in Appendix B.2) that discrete convexity in $n \in \mathbb{Z}^+$ still holds for (2.16). Thus, our optimization model for a homogeneous fleet is amenable to incorporating a stochastic LTL rate, while we leave further extensions dealing with stochastic LTL rates for future study.

2.4 Heterogeneous fleet

This section characterizes the fleet composition model for a heterogeneous fleet given T different truck types that differ in capacity levels. For $t = 1, ..., T$, let n_t denote the decision variables, i.e., the number of type-t trucks to own. Let W_t and K_t denote the capacity level and the fixed ownership cost of type-t truck, respectively. We use a known LTL rate a, and the fixed dispatching cost for each truck type t equals A_t. As before, we assume $aW_t > A_t$ for all t. Obviously, a break-even point Y_t exists for each truck type such that $aY_t = A_t$. However, the transition behavior between using internal TL capacity and external LTL units is more complex than the case of the homogeneous fleet, especially when $T > 2$. In this section, we first analytically characterize the optimal shipping policies and the expected shipping quantities for a fleet that only considers two truck types, and then develop a general two-stage optimization model using stochastic programming for the heterogeneous fleet when considering multiple truck types.

2.4.1 Fleet with two truck types

Consider two potential truck types, which have capacity levels W_1 and W_2, respectively, and $W_1 = qW_2$, where q is a positive integer and $q \geq 2$. We assume the fixed ownership costs of the two truck types satisfy $K_2 < K_1 < qK_2$, implying owning q small trucks is more expensive than owning a big truck, although the truckload capacity provided by a big truck is replaceable by q small trucks. Considering the economies of scale in transportation, we assume $A_1/W_1 \leq A_2/W_2 < a$ and $A_2 < A_1$, which indicates that even though dispatching a big truck costs more than sending a small truck, the per-unit shipping cost becomes lower when shipping in larger batches. It is also straightforward to show that $A_1 \leq qA_2$. Similar to the homogeneous fleet case, there exist break-even points Y_1 (for type-1 trucks) and Y_2 (for type-2 trucks) such that $aY_1 = A_1$ and $aY_2 = A_2$, assuming $W_2 < Y_1$. Observe that $Y_1/W_1 \leq Y_2/W_2$. We aim to minimize the total expected costs of owning and shipping by determining the number of trucks to own of each type, denoted by n_1 and n_2, respectively. Letting n_2^* denote the optimal value of n_2, we show in Theorem 2.4.1 that the optimal value of n_2 is no more than q. Proofs of all theorems and propositions can be found in Appendix B.3.

Theorem 2.4.1. *When $W_1 = qW_2$, $K_1 < qK_2$ and $A_1 \leq qA_2$, at optimality, the maximum possible number of type-2 trucks to own is q, i.e., $n_2^* \in \{0, 1, ..., q\}$ with $q \in \mathbb{Z}^+$ and $q \geq 2$.*

2.4.1.1 *Special case of* $W_1 = 2W_2$

To investigate the optimal shipping policies and the effective discounted cost schedule, we first consider the case when $q = 2$, where $W_1 = 2W_2$, $K_1 < 2K_2$, and $A_2 < A_1 \leq 2A_2$. By Theorem 2.4.1, we know that the optimal number of type-2 trucks to own is either 0, 1 or 2. We can thus characterize the solution approach by only considering three cases where $n_2 = 0, 1$ and 2, respectively. We use the following notation for the expected number of TL shipments and the expected number of LTL units (which are functions of n_1) when owning i type-2 trucks, where $i \in \{0, 1, 2\}$.

- $E^{(i)}_{TL1}(n_1)$: the expected number of type-1 TL shipments when owning i type-2 trucks,
- $E^{(i)}_{TL2}(n_1)$: the expected number of type-2 TL shipments when owning i type-2 trucks,
- $E^{(i)}_{LTL}(n_1)$: the expected number of units shipped via external LTL services when owning i type-2 trucks.

If $n_2 = 0$, the problem becomes equivalent to one with a homogeneous fleet with only type-1 trucks. The approach discussed in Section 2.3 can be applied to this case through which one can characterize the following expected shipping quantities in terms of n_1.

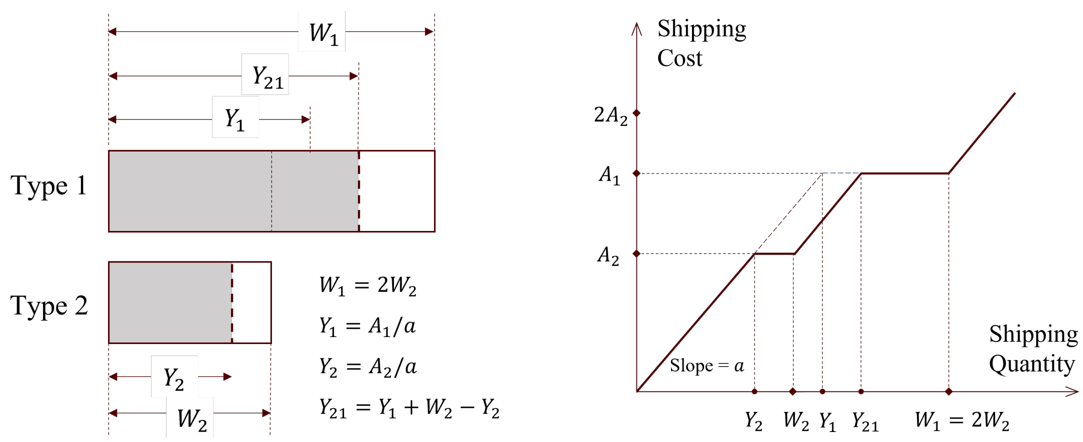
$$\begin{aligned} E^{(0)}_{TL1}(n_1) &= \sum_{k=1}^{n_1-1} k\Big(F\big(kW_1 + Y_1\big) - F\big((k-1)W_1 + Y_1\big)\Big) + n_1\Big(1 - F\big((n_1-1)W_1 + Y_1\big)\Big) \\ &= n_1 - \sum_{k=0}^{n_1-1} F(kW_1 + Y_1), \end{aligned} \tag{2.17}$$

$$E^{(0)}_{LTL}(n_1) = \sum_{k=0}^{n_1-1} \int_{kW_1}^{kW_1+Y_1} (x - kW_1) f(x) dx + \int_{n_1W_1}^{\infty} (x - n_1W_1) f(x) dx. \tag{2.18}$$

We then solve the following optimization model which retains discrete convexity in n_1:

$$\underset{n_1 \in \mathbb{Z}^+}{\text{Minimize}} \quad g_0(n_1) = K_1 n_1 + A_1 E^{(0)}_{TL1}(n_1) + a E^{(0)}_{LTL}(n_1). \tag{2.19}$$

If $n_2 = 1$, in a single period, when the amount of demand is between W_2 and W_1, one option is to dispatch a full type-2 (small) truck and then outsource any remaining units to an LTL carrier. This leads to a total shipping cost equal to the sum of A_2 and the cost of using LTL units. Alternatively, we can dispatch a type-1 (big) truck that provides sufficient capacity (up to W_1 units) for the same amount of demand at a shipping cost of A_1. Because $A_2 < A_1$, it is possible that the total shipping costs incurred in these two cases are equal when the LTL cost incurred in the first case equals $A_1 - A_2$. This implies that, when considering two truck types, there exists another break-even point, denoted by Y_{21}, at which $A_2 + a(Y_{21} - W_2) = A_1$ holds with $W_2 < Y_{21} < W_1$. At the amount of Y_{21}, dispatching a type-1 truck has an equivalent shipping cost to the combination of shipping a type-2 truck and using LTL. With $aY_1 = A_1$ and $aY_2 = A_2$, it is straightforward to show $Y_{21} = Y_1 + W_2 - Y_2$, implying $Y_{21} > Y_1$. Figure 2.2(a) illustrates the relationship between Y_1, Y_2 and Y_{21}.



(a) Break-even shipping points. (b) Discounted cost schedule.

Figure 2.2: Break-even points and discounted cost schedule for a heterogeneous fleet.

Let x be the observed demand in a period. If x lies within $[0, Y_2)$, shipping all units via LTL at a rate of a per unit costs the least; if $x \in [Y_2, W_2]$, dispatching a type-2 truck enables carrying all units within this interval at a constant rate A_2, which leads to the lowest shipping cost. If $x \in (W_2, Y_{21})$, the most economical option is to dispatch a full type-2 truck while outsourcing any remaining units to an LTL carrier. For any $x \in [Y_{21}, W_1]$, shipping all units using a type-1 truck at a cost of A_1 minimizes the shipping cost. Figure 2.2(b) illustrates this discounted cost schedule for $x \in [0, W_1]$. In fact, this behavior repeats cyclically for each demand interval $[kW_1, (k+1)W_1]$ with $k = 0, ..., n_1 - 1$. Table 2.2 summarizes the optimal shipping policies and the corresponding demand intervals for $x \in [0, n_1W_1]$. When $x > n_1W_1$, we first ship all n_1 type-1 trucks at capacity

Table 2.2: Optimal shipping policies for demand within $[0, n_1W_1]$ for $n_2 = 1$ or 2.

Demand intervals for $k = 0, 1, ..., n_1 - 1$	Optimal shipping policy
$[kW_1, kW_1 + Y_2]$	Ship k type-1 trucks at capacity and ship the remainder via LTL
$[kW_1 + Y_2, kW_1 + W_2]$	Ship k type-1 trucks at capacity and use one type-2 truck for the remainder
$(kW_1 + W_2, kW_1 + Y_{21})$	Ship k type-1 and one type-2 trucks at capacity, and use LTL for the remainder
$[kW_1 + Y_{21}, (k+1)W_1]$	Ship $(k+1)$ type-1 trucks

and have $x - n_1W_1$ units of leftover. Because we only have one type-2 truck, the optimal shipping policy for the remainder will be based on Y_2. If $x - n_1W_1 < Y_2$, we outsource all remaining units using LTL service. If $Y_2 \leq x - n_1W_1 \leq W_2$, dispatching a type-2 truck for the remainder is the best option. If the leftover is greater than W_2, the combination of a full type-2 truck plus LTL units is optimal. Therefore, the expected shipping quantities when $n_2 = 1$ can be characterized as

$$\begin{aligned} E^{(1)}_{TL1}(n_1) &= \sum_{k=1}^{n_1-1} k\Big(F\big(kW_1 + Y_{21}\big) - F\big((k-1)W_1 + Y_{21}\big)\Big) + n_1\Big(1 - F\big((n_1-1)W_1 + Y_{21}\big)\Big) \\ &= n_1 - \sum_{k=0}^{n_1-1} F(kW_1 + Y_{21}), \end{aligned} \tag{2.20}$$

$$E^{(1)}_{TL2}(n_1) = \sum_{k=0}^{n_1-1} \Big(F(kW_1 + Y_{21}) - F(kW_1 + Y_2)\Big) + \Big(1 - F(n_1W_1 + Y_2)\Big), \tag{2.21}$$

$$\begin{aligned} E^{(1)}_{LTL}(n_1) &= \sum_{k=0}^{n_1} \int_{kW_1}^{kW_1+Y_2} (x - kW_1) f(x)dx + \sum_{k=0}^{n_1-1} \int_{kW_1+W_2}^{kW_1+Y_{21}} (x - kW_1 - W_2) f(x)dx \\ &\quad + \int_{n_1W_1+W_2}^{\infty} (x - n_1W_1 - W_2) f(x)dx, \end{aligned} \tag{2.22}$$

and the optimization model becomes

$$\underset{n_1 \in \mathbb{Z}^+}{\text{Minimize}} \quad g_1(n_1) = K_1 n_1 + K_2 + A_1 E^{(1)}_{TL1}(n_1) + A_2 E^{(1)}_{TL2}(n_1) + a E^{(1)}_{LTL}(n_1). \tag{2.23}$$

If $n_2 = 2$, for any $x \leq n_1 W_1$, the optimal shipping policies are the same as summarized in Table 2.2. For any $x > n_1 W_1$, we can break down the remaining amount $x - n_1 W_1$ into five intervals and characterize the corresponding optimal shipping policies for each interval based on the discounted cost schedule in terms of Y_2, which allows using an additional type-2 truck when $x > n_1 W_1 + W_2 + Y_2$. Table 2.3 summarizes and compares the optimal shipping policies for the remaining amount $x - n_1 W_1$ when $n_2 = 1$ and $n_2 = 2$.

Table 2.3: Optimal shipping policies for quantities in excess of $n_1 W_1$.

	Intervals for the number of the remaining $x - n_1 W_1$ units	Optimal shipping policy
$n_2 = 1$	$[0, Y_2)$	Use LTL only
	$[Y_2, W_2]$	Ship one type-2 truck
	(W_2, ∞)	Ship a full type-2 truck, and use LTL for the remainder
$n_2 = 2$	$[0, Y_2)$	Use LTL only
	$[Y_2, W_2]$	Ship one type-2 truck
	$(W_2, W_2 + Y_2)$	Ship a full type-2 truck, and use LTL for the remainder
	$[W_2 + Y_2, 2W_2]$	Ship two type-2 trucks
	$(2W_2, \infty)$	Ship two full type-2 trucks, and use LTL for any leftovers

We next derive the expected shipping quantities for $n_2 = 2$. The optimal shipping policies for $x \leq n_1 W_1$ follow the same pattern for $n_2 = 1$ and $n_2 = 2$, thus the expected numbers of type-1 TL shipments are equal in both cases, and we have $E^{(2)}_{TL1}(n_1) = E^{(1)}_{TL1}(n_1)$. According to the results in Table 2.2 and Table 2.3, the expected number of type-2 TL shipments and the expected number of units shipped via LTL when $n_2 = 2$ can be obtained as follows:

$$
\begin{aligned}
E^{(2)}_{TL2}(n_1) &= \sum_{k=0}^{n_1-1} \Big(F(kW_1 + Y_{21}) - F(kW_1 + Y_2) \Big) + \Big(F(n_1W_1 + W_2 + Y_2) - F(n_1W_1 + Y_2) \Big) \\
&\qquad + 2\Big(1 - F(n_1W_1 + W_2 + Y_2) \Big), \\
&= \sum_{k=0}^{n_1-1} \Big(F(kW_1 + Y_{21}) - F(kW_1 + Y_2) \Big) - F(n_1W_1 + W_2 + Y_2) - F(n_1W_1 + Y_2) + 2, \\
E^{(2)}_{LTL}(n_1) &= \sum_{k=0}^{n_1} \int_{kW_1}^{kW_1+Y_2} (x - kW_1) f(x) dx + \sum_{k=0}^{n_1-1} \int_{kW_1+W_2}^{kW_1+Y_{21}} (x - kW_1 - W_2) f(x) dx \\
&\quad + \int_{n_1W_1+W_2}^{n_1W_1+W_2+Y_2} (x - n_1W_1 - W_2) f(x) dx + \int_{n_1W_1+2W_2}^{\infty} (x - n_1W_1 - 2W_2) f(x) dx.
\end{aligned} \tag{2.24}
$$

The resulting optimization model is

$$
\underset{n_1 \in \mathbb{Z}^+}{\text{Minimize}} \quad g_2(n_1) = K_1n_1 + 2K_2 + A_1E^{(2)}_{TL1}(n_1) + A_2E^{(2)}_{TL2}(n_1) + aE^{(2)}_{LTL}(n_1). \tag{2.25}
$$

Observe that in the case of $n_2 = 2$, only one of the type-2 trucks may be used for any $x \leq n_1W_1$, which is formalized in Proposition 2.4.2.

Proposition 2.4.2. *If $n_2 = 2$, then an optimal solution exists such that we do not dispatch both type-2 trucks unless all type-1 trucks have been dispatched.*

Consequently, one of the three functions $g_0(n_1)$, $g_1(n_1)$ and $g_2(n_1)$ provides the minimum objective value and leads to the best n_1, which permits easily determining the optimal value of n_2 (at 0, 1 or 2). By pairwise comparison of these three functions, we can obtain a system of inequalities that determines the expected shipping quantities, which can be used as threshold conditions for determining the optimal number of type-1 trucks, as shown in Table 2.4.

In particular, for uniformly distributed demand, it is possible to obtain closed-form threshold values for n_1 at which the optimal solution transitions among the three discussed scenarios (when $n_2 = 0, 1$ or 2); for a normal demand distribution, searching among a finite set of integer values for

Table 2.4: Comparison of $g_0(n_1), g_1(n_1)$ and $g_2(n_1)$.

Comparison	Resulting inequality
$g_0(n_1) \leq g_1(n_1)$	$a\big(E^{(0)}_{LTL} - E^{(1)}_{LTL}\big) \leq K_2 - A_1\big(E^{(0)}_{TL1} - E^{(1)}_{TL1}\big) + A_2E^{(1)}_{TL2}$
$g_0(n_1) \leq g_2(n_1)$	$a\big(E^{(0)}_{LTL} - E^{(2)}_{LTL}\big) \leq 2K_2 - A_1\big(E^{(0)}_{TL1} - E^{(2)}_{TL1}\big) + A_2E^{(2)}_{TL2}$
$g_1(n_1) \leq g_2(n_1)$	$a\big(E^{(1)}_{LTL} - E^{(2)}_{LTL}\big) \leq K_2 + A_2\big(E^{(2)}_{TL2} - E^{(1)}_{TL2}\big)$

n_1 can effectively locate the optimal solution. We include the equations for the expected shipping quantities for both distributions in Appendix B.4.

2.4.1.2 Special case of $W_1 = qW_2$

We can use a similar approach to characterize the optimal shipping policies for $W_1 = qW_2$ with any positive integer value of q, where $q \geq 2$, $K_1 < qK_2$, and $A_2 < A_1 \leq qA_2$. Letting $q_L = \lfloor A_1/A_2 \rfloor$, the break-even point Y_{21}, in this case, satisfies $q_LA_2 + a(Y_{21} - q_LW_2) = A_1$, which is equivalent to $Y_{21} = Y_1 + q_L(W_2 - Y_2)$. Figure 2.3 illustrats the discounted cost schedule for $2A_2 < A_1' < 3A_2$ and $3A_2 < A_1'' < 4A_2$ when $q = 4$.

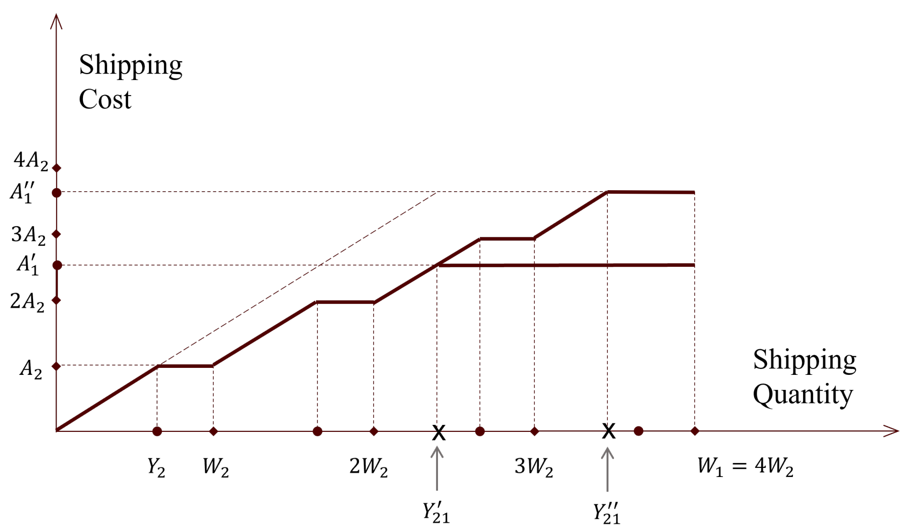


Figure 2.3: Discounted cost schedule for $W_1 = 4W_2$.

Proposition 2.4.3. *If an optimal solution has $n_2^* = \tilde{q}$ and $q_L \leq \tilde{q} \leq q$, the maximum possible number of type-2 trucks dispatched before dispatching all type-1 trucks is q_L. None of the remaining $\tilde{q} - q_L$ type-2 trucks will be used unless all type-1 trucks have been dispatched.*

According to Proposition 2.4.3, we can characterize the optimal shipping schedule, as summarized in Table 2.5, for any $x \leq n_1 W_1$.

Table 2.5: Optimal shipping policies for demand within $[0, n_1 W_1]$ when $W_1 = qW_2$.

Demand intervals for $k = 0, 1, ..., n_1 - 1$	Optimal shipping policy
$[kW_1, kW_1 + Y_2)$	Ship k full type-1 trucks and use LTL for the remainder
$[kW_1 + iY_2, kW_1 + iW_2]$ for $i = 1, ..., q_L$	Ship k full type-1 and $i - 1$ full type-2 trucks, and use an additional type-2 truck for any leftovers
$(kW_1 + q_L W_2, kW_1 + Y_{21})$	Ship k full type-1 and q_L full type-2 trucks, and use LTL for any leftovers
$[kW_1 + Y_{21}, (k+1)W_1]$	Ship $(k+1)$ type-1 trucks

For any demand that is greater than $n_1 W_1$, we first ship all n_1 type-1 trucks at capacity, and then use the following policies summarized in Table 2.6 for the remaining $x - n_1 W_1$ units.

Table 2.6: Optimal shipping policies for the remaining amount $x - n_1 W_1$ when $W_1 = qW_2$.

Intervals for the remaining units	Optimal shipping policy
$[iW_2, iW_2 + Y_2)$ for $i = 0, ..., n_2 - 1$	Ship i full type-2 trucks and then use LTL for the remainder
$[iW_2 + Y_2, (i+1)W_2]$ for $i = 1, ..., n_2 - 1$	Ship $i + 1$ type-2 trucks
$(n_2 W_2, \infty)$	Ship all type-2 trucks at capacity, and use LTL for any leftovers

2.4.2 General model for a fleet with multiple truck types

When considering more than two truck types (i.e., $T > 2$), deriving the break-even points that determine the transition between different shipping options becomes very complex, which makes it hard to analytically characterize the optimal shipping policies using the same approach to the one developed for $T = 2$. Instead, in this section, we create a two-stage stochastic programming model to tackle a wider range of truck types. We solve the model using the sample average approximation method and develop a decomposition-based heuristic that provides effective solutions.

2.4.2.1 Two-stage stochastic programming formulation

We consider the fleet composition problem as a two-stage process: in the first stage, prior to observing any demand, we must make a decision on the number of trucks to own given a set of truck types; after observing the demand, in the second stage, we need to decide the number of trucks to dispatch for each truck type, and the number of units to ship using LTL services, with an objective of minimizing the total expected cost. Therefore, it is natural to formulate the fleet composition problem as the following two-stage stochastic programming model:

$$[\mathbb{SP}] \qquad \underset{n_t \in \mathbb{Z}^+}{\text{Minimize}} \quad g(\mathbf{n}) = \sum_{t=1}^{T} K_t n_t + \mathbb{E}[Q(\mathbf{n}, X)], \tag{2.26}$$

where $\mathbf{n} = [n_t]$ are the first-stage decision variables representing the number of type-t trucks to own for $t = 1, ..., T$, and $Q(\mathbf{n}, X)$ is the optimal objective value of the second-stage problem in terms of the stochastic demand X, where

$$[\mathbb{SP}\text{-}2] \qquad Q(\mathbf{n}, X) = \quad \text{Minimize} \quad \sum_{t=1}^{T} A_t m_t + a\left(X - \sum_{t=1}^{T} y_t\right), \tag{2.27}$$

$$\text{Subject to:} \quad m_t \leq n_t, \qquad \forall t = 1, \ldots, T, \tag{2.28}$$

$$y_t \leq m_t W_t, \qquad \forall t = 1, \ldots, T, \tag{2.29}$$

$$\sum_{t=1}^{T} y_t \leq X, \tag{2.30}$$

$$m_t \in \mathbb{Z}^+, y_t \in \mathbb{R}^+, \qquad \forall t = 1, \ldots, T. \tag{2.31}$$

In the second-stage model, m_t and y_t denote the actual number of trucks dispatched and the number of units shipped using type-t trucks, respectively. The objective function (2.26) in the master (first-stage) problem aims to minimize the total expected cost, including the total ownership cost and the total expected shipping cost resulting from the second-stage problem. The objective function (2.27) in the second-stage problem is to minimize the total expected shipping cost of using internal TL shipments and external LTL units, with stochastic demand. Constraints (2.28) ensure that the number of trucks actually dispatched does not exceed the number of trucks owned for each truck type. Constraints (2.29) guarantee the accumulated number of units shipped via internal trucks is no more than the aggregated TL capacity for each truck type. Constraint (2.30) ensures the total number of units shipped via all internal trucks does not exceed the total demand.

We use the *Sample Average Approximation* (SAA) method to approximate the expected value function $\mathbb{E}[Q(\mathbf{n}, X)]$ by its sample average $\frac{1}{S}\sum_{i=1}^{S} Q(\mathbf{n}, d^i)$, where $\{d^1, d^2, ..., d^S\}$ is a demand sample of S scenarios generated according to a probability distribution $\mathcal{F}$. The master problem can therefore be approximately solved via the following deterministic equivalent model:

$$[\mathbb{DEP}] \qquad \text{Minimize} \quad g_S(\mathbf{n}) = \sum_{t=1}^{T} K_t n_t + \frac{1}{S}\sum_{i=1}^{S}\left(\sum_{t=1}^{T} A_t m_t^i + a\left(d^i - \sum_{t=1}^{T} y_t^i\right)\right) \tag{2.32}$$

$$\text{Subject to:} \quad m_t^i \le n_t, \qquad \forall t = 1, \ldots, T, i = 1, \ldots, S, \tag{2.33}$$

$$y_t^s \le m_t^i W_t, \qquad \forall t = 1, \ldots, T, i = 1, \ldots, S, \tag{2.34}$$

$$\sum_{t=1}^{T} y_t^i \le d^i, \qquad \forall i = 1, \ldots, S, \tag{2.35}$$

$$n_t \in \mathbb{Z}^+, \qquad \forall t = 1, \ldots, T, \tag{2.36}$$

$$m_t^i \in \mathbb{Z}^+, \qquad \forall t = 1, \ldots, T, i = 1, \ldots, S, \tag{2.37}$$

$$y_t^i \ge 0, \qquad \forall t = 1, \ldots, T, i = 1, \ldots, S, \tag{2.38}$$

in which m_t^i is the number of type-t trucks dispatched in scenario i, and y_t^i is the number of units shipped using type-t trucks in scenario i, for $t = 1, \ldots, T$ and $i = 1, \ldots, S$.

2.4.2.2 A decomposition-based algorithm

The formulation above for problem $\mathbb{DEP}$ is a *mixed-integer programming* (MIP) model and is $\mathcal{NP}$-hard (we show this in Appendix B.5). Classical algorithms for solving MIP models, like branch-and-bound, can provide exact solutions, while the computing time may escalate exponentially as the values of T and S increase. However, for a given solution $\hat{\mathbf{n}}$, problem $\mathbb{DEP}$ becomes separable for each scenario i, and we can then solve S subproblems for $Q(\hat{\mathbf{n}}, d_i)$ individually, each associated with m_t^j and y_t^i. As the value of T is usually chosen from a limited range of integers, even though the subproblems are also MIP models, the size of each subproblem is relatively small and can be solved to optimality relatively fast. Therefore, we propose the following *decomposition-based* (DB) heuristic.

Algorithm 1: The DB algorithm for solving problem $\mathbb{DEP}$

Initialization: Solve the linear relaxation of problem $\mathbb{DEP}$ based on a sample of size S, $\{d^1, \ldots, d^S\}$; obtain the relaxed (continuous) solution $\mathbf{n}^{RX} = [n_t^{RX}]$ for $t = 1, \ldots, T$.
Rounding: Round up each continuous variable n_t^{RX} to the nearest integer $n_t^{RD} = \lceil n_t^{RX} \rceil$, leading to the rounded solution $\mathbf{n}^{RD} = [n_t^{RD}]$ for $t = 1, \ldots, T$.
for $i \leftarrow 1$ **to** S **do**
    Solve the subproblem for each scenario i and obtain the resulting optimal objective value $Q(\mathbf{n}^{RD}, d^i)$ for $i = 1, \ldots, S$.
end
Averaging: Compute the mean of the objective values over all subproblems: $\bar{q}_S(\mathbf{n}^{RD}) = \frac{1}{S}\sum_{i=1}^{S} Q(\mathbf{n}^{RD}, d^i)$;
Result: Obtain the objective value for master problem
$g_S(\mathbf{n}^{RD}) = \sum_{t=1}^{T} K_t n_t^{RD} + \bar{q}_S(\mathbf{n}^{RD})$.

To obtain a first-stage solution, in the **Initilization** phase, we first solve the linear relaxation of problem $\mathbb{DEP}$ based on a sample of size S by setting $n_t, m_t^i \geq 0$, which can be solved within a negligible time, leading to a set of continuous variable values n_t^{RX} for each truck type t. We next **round** each n_t^{RX} to its nearest integer value n_t^{RD} such that $n_t^{RD} \geq n_t^{RX}$, which leads to an integer solution $\mathbf{n}^{RD} = \left[n_t^{RD}\right]$ and will be used for solving the subproblems. For each scenario i, we solve the subproblem to optimality and store the objective value $Q(\mathbf{n}^{RD}, d^i)$. After solving all subproblems, we compute the **average** of S objective values $\bar{q}_S(\mathbf{n}^{RD}) = \frac{1}{S}\sum_{i=1}^{S} Q(\mathbf{n}^{RD}, d^i)$, and then use this to obtain the objective value for the original $\mathbb{DEP}$ problem associated with $\mathbf{n}^{RD}$, which equals $\sum_{t=1}^{T} K_t n_t^{RD} + \bar{q}_S(\mathbf{n}^{RD})$. Since the time required for solving both the linear relaxation of problem $\mathbb{DEP}$ and each subproblem is small, this heuristic algorithm can provide a feasible integer solution to problem $\mathbb{DEP}$ within a short computing time even when the sample size is large.

2.5 Numerical experiments

In this section, we implement the SAA method by solving problem $\mathbb{DEP}$ based on a set of randomly generated instances (in Sections 2.5.1) and empirically show that using SAA can obtain high-quality solutions within acceptable computing time (in Section 2.5.2). We then execute the proposed DB algorithm and test its computational performance in Section 2.5.3, comparing the results with the SAA solutions. Finally, we perform sensitivity analysis to investigate the impacts of different parameters on the fleet composition decision, which we discuss in Section 2.5.4. We program and create random generators in the Python programming language to create data needed for all computational experiments. All test instances are solved using the Gurobi 10.0.1 solver on a PC with a 3.00 GHz Intel i5 Core and 8GB RAM.

2.5.1 Data generation

We randomly generate 50 independent instances that provide 50 distinct sets of parameters, including the truck capacity levels (W_t), the fixed ownership costs (K_t), the TL dispatching costs (A_t), and the LTL rate (a). We use five truck classes for the experiments according to the U.S. commercial truck classifications based on the gross vehicle weight rating [52], which provides lower and upper bounds on the values of W_t created for each truck class. To obtain the ownership cost K_t for each truck type on a weekly basis, we generate a fixed cost of purchasing and maintaining a brand new truck based on its market price and average annual maintenance cost, and then allocate this to weekly expenses, assuming a 260-week (5-year) truck life cycle. Recall that Y_t is the break-even shipping point at which using a TL shipment is more economical than using LTL services, and $aY_t = A_t$. Letting $Y_t = \alpha_t W_t$ with $0 < \alpha_t < 1$, we therefore have $A_t = a\alpha_t W_t$, through which the values of A_t can be generated based on the generation of a, α_t and W_t. Table 2.7 includes the lower and upper bounds used when generating these parameters for each truck class.

The demand samples are generated as weekly observations. We use the normal distribution to generate the samples used for implementing the SAA method and the DB algorithm in Sections 2.5.2 and 2.5.3, and for analyses through Sections 2.5.4.1 - 2.5.4.3. To compare the solutions when

Table 2.7: Lower and upper bounds used for generating capacity and cost parameters.

Truck class	Type index (t)	W_t (lbs)	$260K_t$ ($)	a ($/lb)	α_t
Class 8	1	(33000, 40000]	[300000, 350000]		
Class 7	2	(26000, 33000]	[190000, 260000]		
Class 6	3	(19500, 26000]	[150000, 190000]	[0.22, 0.42]	[0.5, 0.8]
Class 5	4	(16000, 19500]	[110000, 130000]		
Class 4	5	(14000, 16000]	[90000, 110000]		

using different demand distributions, we generate extra samples based on uniform and gamma distributions, respectively, which are used in Section 2.5.4.4.

2.5.2 SAA results and solution evaluation

We set a 0.1% Gurobi gap (which is the value of the `MIPGap` parameter in Gurobi for solving MIP models) and a maximum time limit equal to 18000 seconds (5 hours) as alternative termination conditions, whichever is reached first. Gurobi will terminate (for minimization problems) when the gap between the primal (upper) and the dual (lower) objective bounds becomes less than 0.1% of the incumbent objective value within 18000 seconds; the solver will stop after solving the model for 18000 seconds otherwise.

We solve each instance using a sample of size 1000 with $T = 3$ (we use truck types 1, 2, and 3). The outputs for each instance include decision variable values, the number of units shipped using LTL services, the objective value at termination, the best-known lower bound on the objective value provided by Gurobi, the Gurobi gap at termination, and the total runtime required by Gurobi (including preprocessing time in addition to the actual computing time, which may lead to a total time that is greater than the maximum time limit for some instances). As the number of scenarios (the value of S) in each sample generated is large and leads to a massive number of second-stage variables, we store these locally and only output the mean values of m_t^s and y_t^s computed over S observations, that is, $\bar{m}_t = \frac{1}{S}\sum_{s=1}^{S} m_t^s$ and $\bar{y}_t = \frac{1}{S}\sum_{s=1}^{S} y_t^s$.

Table 2.8 shows the outputs for Instances 9 and 17, in which solving Instance 9 reaches the 18000s time limit (although the gap is still over 0.1% at termination), while Instance 17 is solved in around 10 seconds with a Gurobi gap less than 0.1%. The SAA results for all instances with $S = 1000$ and $T = 3$ are provided in Appendix B.6. In total, 38 among 50 instances are solved at the target Gurobi gap (less than or equal to 0.1%) with an average runtime of 67.2 seconds. The remaining 12 instances are solved when the maximum computing time limit (18000 seconds) is reached, with an average Gurobi gap of 0.16% at termination, which is not significantly different from the 0.1% target.

Table 2.8: SAA results for Instances 9 and 17 with $S = 1000$ and $T = 3$.

Instance	Truck Class (t)	W_t (lbs)	a ($/lb)	K_t ($)	A_t ($)	n_t	$\bar{m}_t^s$	$\bar{y}_t^s$	LTL units	Obj.Val.	Obj.LB.	GRB gap (%)	GRB time (s)
	Class 8 (1)	34000		1237.17	6416	32	27.89	947997.86					
9	Class 7 (2)	26100	0.37	934.28	5022	1	0.59	15218.43	27792.59	248981.52	248694.12	0.12	18028.98
	Class 6 (3)	23600		652.66	4977	7	2.36	55312.28					
	Class 8 (1)	39600		1169.39	7975	31	25.21	997951.73					
17	Class 7 (2)	31000	0.38	878.70	6597	1	0.52	15922.90	25402.39	254527.92	254301.96	0.09	10.06
	Class 6 (3)	22000		663.58	5016	1	0.52	10985.24					

To evaluate the quality of a given SAA solution $\hat{\mathbf{n}}$, we follow a statistical procedure referring to Shapiro and Philpott [53]. Let g^* denote the true optimal objective value of problem $\mathbb{SP}$, and let $g(\hat{\mathbf{n}})$ be the objective value associated with $\hat{\mathbf{n}}$. The corresponding optimality gap is obtained by

$$\delta(\hat{\mathbf{n}}) = g(\hat{\mathbf{n}}) - g^*, \tag{2.39}$$

which can be used to evaluate the quality of $\hat{\mathbf{n}}$. Since it is difficult to compute $g(\hat{\mathbf{n}})$ and g^* exactly, we instead estimate the value of $\delta(\hat{\mathbf{n}})$ by constructing a bound that has at least $(1-2\beta)$ confidence on it, with $0 < \beta < 0.5$.

The value of $g(\hat{\mathbf{n}})$ can be estimated by

$$g_{\bar{S}}(\hat{\mathbf{n}}) = \sum_{t=1}^{T} K_t \hat{n}_t + \frac{1}{\bar{S}} \sum_{j=1}^{\bar{S}} Q(\hat{\mathbf{n}}, d^j), \tag{2.40}$$

where $\{d^1, ..., d^{\bar{S}}\}$ is a demand sample with $\bar{S}$ scenarios. Note that this sample of size $\bar{S}$ is independent of the sample used for obtaining $\hat{\mathbf{n}}$. The value of $\bar{S}$ is chosen to be sufficiently large since solving each individual subproblem for $Q(\hat{\mathbf{n}}, d^j)$ in (2.40) is relatively fast with a given first-stage solution $\hat{\mathbf{n}}$. Letting $\bar{q}_{\bar{S}}(\hat{\mathbf{n}}) = \frac{1}{\bar{S}} \sum_{j=1}^{\bar{S}} Q(\hat{\mathbf{n}}, d^j)$, the variance of $g_{\bar{S}}(\hat{\mathbf{n}})$ is then estimated by

$$\sigma_{\bar{S}}^2(\hat{\mathbf{n}}) = \frac{1}{\bar{S}(\bar{S}-1)} \sum_{j=1}^{\bar{S}} \left(Q(\hat{\mathbf{n}}, d^j) - \bar{q}_{\bar{S}}(\hat{\mathbf{n}})\right)^2. \tag{2.41}$$

This leads to an approximate $(1-\beta)$ confidence upper bound on $g(\hat{\mathbf{n}})$ that equals

$$B_{\bar{S}}(\hat{\mathbf{n}}) = g_{\bar{S}}(\hat{\mathbf{n}}) + z_\beta \sigma_{\bar{S}}(\hat{\mathbf{n}}), \tag{2.42}$$

where $z_\beta = \Phi^{-1}(1-\beta)$ and $\Phi(z)$ is the CDF of the standard normal distribution.

To estimate the true optimal value g^*, we solve problem $\mathbb{DEP}$ M times by using M independently generated samples (based on the same probability distribution), each including L scenarios. At termination, Gurobi provides each solved problem associated with sample j with the incumbent objective value g_L^j as well as the best known lower bound on g_L^j, denoted by $\tilde{g}_L^j$. The average of the M best known lower bounds equals

$$\bar{g}_{L,M} = \frac{1}{M} \sum_{j=1}^{M} \tilde{g}_L^j, \tag{2.43}$$

and the variance of $\bar{g}_{L,M}$ can be estimated by

$$\sigma_{L,M}^2 = \frac{1}{M(M-1)} \sum_{j=1}^{M} \left(\tilde{g}_L^j - \bar{g}_{L,M}\right)^2. \tag{2.44}$$

We then construct a lower bound on g^* with at least $(1-\beta)$ confidence that equals

$$B_{L,M} = \bar{g}_{L,M} - t_{\beta,M-1}\sigma_{L,M}, \tag{2.45}$$

where $t_{\beta,M-1}$ is the β-critical value of the t-distribution with $M-1$ degrees of freedom.

Consequently, a probabilistic bound on $\delta(\hat{\mathbf{n}})$ associated with $\hat{\mathbf{n}}$, with at least $(1-2\beta)$ confidence, is obtained by

$$B_{\bar{S}}(\hat{\mathbf{n}}) - B_{L,M}. \tag{2.46}$$

Table 2.9 summarizes the evaluation results for the SAA solutions obtained in Appendix B.6 for Instances 10, 20, 30, 40, and 50, with $\bar{S} = 3000$, $L = 1000$, $M = 20$ and $\beta = 0.025$. According to the results, the estimated optimality gaps for all five instances are within 2% of the corresponding estimated upper bound values with at least 95% confidence, in which the estimated gaps for Instances 10 and 20 are below 1% of their estimated upper bounds. This implies that the SAA method leads to an effective solution for any of these instances, providing an interval with at least a 95% confidence level for the true optimal value, in which the difference between the lower and upper bounds is relatively small (less than 2%).

Table 2.9: Evaluation results for selected instances.

Instance	$B_{\bar{S}}(\hat{\mathbf{n}})$	$B_{L,M}$	gap = $\left(1 - \frac{B_{L,M}}{B_{\bar{S}}(\hat{\mathbf{n}})}\right) \times 100\%$
10	181126.81	179877.26	0.69
20	308286.95	305766.59	0.82
30	291090.81	286511.53	1.57
40	175124.14	171732.12	1.94
50	283355.06	280184.65	1.12

2.5.3 Performance of the DB algorithm

To implement the DB heuristic algorithm proposed in Section 2.4.2.2, we use Gurobi to solve the relaxation of problem $\mathbb{DEP}$ and the corresponding subproblems for each instance. The samples we use for the DB algorithm are identical to the samples used for obtaining the SAA solutions listed in Appendix B.6, each including 1000 scenarios. All instances are solved using the DB algorithm with $T = 3$. We compare the SAA results with the results obtained using the DB algorithm in terms of decision variable values, objective values, and the required runtime. Table

2.10 summarizes the results for six instances. The last two columns in the table compute the differences between the objective values and between the runtime in percentages (based on the SAA results), respectively, in which a negative percentage value indicates the DB algorithm leads to a larger value of objective/runtime than those obtained by using SAA. An instance is starred if its DB algorithm runtime is better than the corresponding SAA runtime.

Table 2.10: Comparison of the SAA and the DB results for selected instances.

	Instance	SAA $\hat{\mathbf{n}}$	SAA obj.val.	SAA time (s)	DB $\mathbf{n}^{RD}$	DB obj.val.	DB time (s)	$\left(1 - \frac{\text{DB obj.val.}}{\text{SAA obj.val.}}\right) \times 100\%$	$\left(1 - \frac{\text{DB time}}{\text{SAA time}}\right) \times 100\%$
*	25	[2, 44, 0]	254613.38	32.69	[0, 46, 0]	254895.54	21.08	-0.11	35.53
	26	[0, 40, 3]	224691.81	4.92	[0, 43, 0]	225265.65	20.05	-0.26	-307.77
*	27	[25, 1, 14]	226044.06	18045.75	[27, 0, 14]	226288.31	22.16	-0.11	99.88
*	28	[33, 0, 1]	240940.82	46.01	[34, 0, 0]	241423.22	21.47	-0.20	53.32
*	29	[2, 41, 0]	235531.11	111.54	[0, 44, 0]	235803.60	21.09	-0.12	81.09
*	30	[25, 12, 0]	289372.50	54.89	[26, 12, 0]	289531.71	23.68	-0.06	56.85

Observe that in each instance, the resulting first-stage solution obtained by using the DB algorithm ($\mathbf{n}^{RD}$) differs from the SAA solution ($\hat{\mathbf{n}}$) but not significantly, with a difference of up to three vehicles per truck type. The objective values obtained by the DB algorithm, however, are slightly higher than the SAA objectives in all instances, implying that the SAA method outperforms the DB algorithm in terms of the optimal objective values. On the other hand, the DB algorithm can solve each instance and provides a solution within 24 seconds, which substantially decreases the runtime by 30% or more for most instances, compared with the time required when directly solving problem $\mathbb{DEP}$ in SAA. This improvement in runtime is particularly evident for those instances that reach the maximum runtime limit (Instance 27, for example). Note that the DB algorithm does not necessarily outperform the SAA method in terms of runtime for those instances that can be quickly solved using SAA. For example, in Instance 26, the SAA method only takes 4.92 seconds, while the DB algorithm requires about 20 seconds.

Table B.3 in Appendix B.7 includes the comparison results for all 50 instances, in which the average runtime of the DB heuristic across 50 instances is 21.47 seconds, which is about 68% lower than the average runtime of SAA method (67.2 seconds) for those instances that are solved within the maximum time limit. In particular, using the DB algorithm improves the runtime for 29 instances, among which 23 instances see a reduction of more than 50% in the runtime (including instances that reach the maximum time limit), and 6 instances improve the runtime by at least 5%. Moreover, the average difference between the objective values obtained by the DB algorithm and those obtained by the SAA method across 50 instances is only 0.12%.

To investigate how the DB algorithm performs when increasing the value of T and the sample size S, we select 20 instances and create multiple variants for each instance by changing the value of T and increasing the sample size, while fixing all other parameters. We include comparison results for these instance variants in Tables B.4 through B.8 in Appendix B.7. Table 2.11 summarizes the resulting average runtimes and the average differences between objective values (in percentages) for each group of instance variants.

Table 2.11: Performance comparison with different values of T and S.

T	S	Avg. SAA time (s)	Avg. DB time (s)	Avg. $\left(1 - \frac{\text{DB obj.val.}}{\text{SAA obj.val.}}\right) \times 100\%$
3	1000	8.96	20.82	-0.13
	3000	68.29	71.69	-0.12
	5000	192.77	116.88	-0.14
4	1000	17.27	24.46	-0.12
	3000	142.2	75.57	-0.11
	5000	446.17	118.15	-0.14

Notably, for a fixed value of T, a larger sample size results in an increased runtime when using either method, while the SAA runtime exhibits a more noticeable increase when the sample size increases. For example, with $T = 4$, the DB algorithm requires an average runtime of 24.46 seconds for solving an instance with 1000 scenarios, while it requires nearly five times longer (118.15 seconds) for an instance with a sample size of 5000. The increase in the average runtime when employing the SAA method, however, becomes even more significant as the sample size increases from 1000 to 5000, which results in an approximate 26-fold rise (from 17.27 to 446.17

seconds). Furthermore, when keeping the sample size fixed (for example, when $S = 5000$), a noticeable increase in the average runtime (from 192.77 to 446.17 seconds) occurs as the value of T increases (from 3 to 4) when using the SAA method. Conversely, the runtime of the DB algorithm experiences only a minimal increase (from 116.88 to 118.15 seconds) when increasing the value of T from 3 to 4 with $S = 5000$. The average objective values achieved by the DB algorithm are marginally higher than the SAA objective values, with an average difference of less than 0.15% for each group. This indicates that when raising the values of T and S, using the SAA method may lead to a significantly extended runtime, whereas the DB algorithm can provide highly effective solutions that are nearly comparable to those obtained through the SAA method but within a much shorter runtime.

2.5.4 Sensitivity analysis

In this section, we investigate the sensitivity of the fleet composition solutions to several factors, including (i) the LTL freight rate (the value of a), (ii) the number of truck types (the value of T), (iii) the demand parameters (including the mean value and the standard deviation of the demand), and (iv) the demand distribution. We use the following metrics to delineate the variations associated with each factor in the solutions.

Fleet capacity level (FCL). The FCL is the average capacity level of a fleet, which depicts the composition and flexibility of a fleet. A lower value of FCL indicates that a fleet has more small trucks and thus higher shipping flexibility, which enables shipping more units in smaller batches via internal TL capacity. Because each instance we have generated and solved corresponds to a unique set of T truck types, in order to standardize the vehicle capacity for each truck type, we first compute the average vehicle capacity over $\mathcal{M}$ test instances for each truck type, denoted by ω_t, such that

$$\omega_t = \frac{1}{\mathcal{M}} \sum_{i=1}^{\mathcal{M}} W_t^{(i)},$$

where the index i is associated with Instance i. For a given instance solved with a solution $[n_t]$, the

FCL is obtained by

$$FCL = \frac{\sum_{t=1}^{T} \omega_t n_t}{\sum_{t=1}^{T} n_t}, \tag{2.47}$$

where $\frac{n_t}{\sum_{t=1}^{T} n_t}$ is the percentage of type-t trucks within the fleet.

Fleet utilization. The fleet utilization, denoted by U_F, is defined by the ratio of the total expected number of trucks actually shipped to the total number of trucks owned, which can be obtained using the following equation:

$$U_F = \frac{\sum_{t=1}^{T} \bar{m}_t}{\sum_{t=1}^{T} n_t}. \tag{2.48}$$

Note that the truck utilization for a given truck type t equals $\bar{m}_t/n_t$.

Internal capacity utilization. This is the overall capacity utilization of all trucks in the fleet, which is the ratio of the total expected number of units shipped via TL shipments to the total TL capacity owned. This is obtained by

$$U_C = \frac{\sum_{t=1}^{T} \bar{y}_t}{\sum_{t=1}^{T} n_t W_t}. \tag{2.49}$$

LTL units percentage. This is the ratio of the expected number of units shipped via LTL services to the total expected number of units shipped using both internal and external shipping options, which is obtained by

$$\text{LTL units } \% = \left(\frac{\text{The expected LTL units}}{\text{The expected LTL units} + \sum_{t=1}^{T} \bar{y}_t} \right) \times 100\%, \tag{2.50}$$

where $\bar{y}_t$ is the expected number of units shipped using type-t trucks.

Cost structure. The cost structure includes four metrics: the total expected cost (i.e., the objective value), and three percentages corresponding to each cost component (the total fixed cost, the total dispatching cost, and the total cost of using LTL services) as a percentage of the total expected cost.

We use the 50 instances generated in Section 2.5.1 as base instances and their SAA solutions solved in Section 2.5.2 as benchmarks, based on which we perform the sensitivity analyses by varying the LTL rate in Section 2.5.4.1 and the number of truck types in Section 2.5.4.2. We investigate the impact of demand parameters in Section 2.5.4.3, and analyze the solution sensitivity to different demand distributions in Section 2.5.4.4.

2.5.4.1 Impact of LTL freight rate

Based on the 50 base instances, we create two groups of 50 instance variants by changing the values of a (medium LTL rate). In group one we decrease the LTL rate in each base instance by 50 cents ($a - 0.05$, low LTL rate) and in group two we increase each base LTL rate by 50 cents ($a + 0.05$, high LTL rate) while fixing all other parameters. We use the same normal demand sample with $S = 1000$ for each instance index (associated with one base instance and two instance variants). Each instance index corresponds to an independently generated sample. In total, 150 instances are solved (using Gurobi) for analyzing the impact of the LTL rate. We consider 3 truck types ($t = 1, 2, 3$) for all instances. The resulting metrics for each instance when changing the LTL rate are attached in Appendix B.8. Table 2.12 summarizes the average (over 50 instances) of each metric for three LTL rate levels.

Table 2.12: The mean values of the metrics as the LTL rate varies.

	Low LTL rate	Medium LTL rate	High LTL rate
Avg. FCL (lbs)	33631.71	33273.31	33086.00
Avg. U_F (%)	84.08	79.49	76.83
Avg. U_C (%)	84.51	80.24	77.76
Avg. LTL units (%)	5.88	3.12	2.14
Avg. fixed cost (%)	15.77	16.95	17.54
Avg. dispatching cost (%)	76.69	78.47	78.88
Avg. LTL cost ((%)	7.55	4.58	3.59
Avg. total cost ($)	237174.07	239412.39	240755.41

We observe that as the LTL rate increases, there is a corresponding decrease in several metrics: the average FCL, the average fleet utilization, the average internal capacity utilization, and the average percentages of both the LTL units and the LTL cost. Conversely, the total expected costs and the percentages of fixed and dispatching costs experience an increase.

When the LTL rate increases, a fleet tends to acquire more internal trucks, especially more small trucks, to replace using external LTL services while maintaining the flexibility of shipping smaller batches and minimizing the total expected cost in a period. This leads to a larger initial capital investment and therefore a higher percentage of the fixed cost. At the same time, dispatching more internal trucks leads to an increased percentage of the total dispatching cost and a reduced percentage of LTL units, and thus a reduction in the percentage of LTL cost. The total expected cost, however, still increases as it is dominated by the growth of the fixed and dispatching costs.

On the other hand, although an increased LTL rate encourages owning and using more internal trucks, it does not lead to either an increased fleet utilization or an increased overall internal capacity utilization. This is because when acquiring more trucks to offset the rising cost of LTL services, even though the average percentage of units shipped via TL shipments increases due to fewer LTL shipments being used, the increase in the aggregate internal TL capacities overrides the increase in the demand of using TL shipments, leading to a lower overall internal capacity utilization. Similarly, the increase in the total number of trucks owned dominates the increase in the expected number of trucks actually shipped, which results in a declined fleet utilization.

2.5.4.2 Impact of the number of truck types

In our numerical experiments, the truck types are indexed from 1 (the largest truck class) to 5 (the smallest truck class). Starting from type-1 truck, as the value of T increases by one, we add a truck type that has a successively smaller capacity to the pool of the truck types. For example, when $T = 3$, we solve the fleet composition problem based on three truck types including type-1, type-2, and type-3 trucks; when $T = 4$, the truck options are type-1, type-2, type-3, and type-4. A larger value of T implies more options for smaller truck sizes. For each base instance, we create two instance variants by increasing the value of T from 3 to 4 and 5, respectively. We use the same normal demand sample of size 1000 for each base instance and its variants. We solve 50 instances for each value of T and thus 150 instances are solved in total. Table 2.13 computes the mean values of the metrics across 50 instances for each value of T. Results for all instances are included in Appendix B.9.

Table 2.13: The mean values of the metrics as T changes.

	$T = 3$	$T = 4$	$T = 5$
Avg. FCL	33265.61	32764.77	32681.53
Avg. U_F (%)	79.50	78.64	78.47
Avg. U_C (%)	80.24	80.04	79.96
Avg. LTL units (%)	3.12	3.01	2.97
Avg. fixed cost (%)	16.95	16.93	16.95
Avg. dispatching cost (%)	78.47	78.64	78.68
Avg. LTL cost (%)	4.58	4.43	4.37
Avg. total cost ($)	239412.24	239340.66	239326.33

As the value of T increases, most metrics demonstrate a clear downward trend, including the average FCL, the mean values of fleet utilization and internal capacity utilization, the percentages associated with LTL, and the total expected cost. Despite the overall decline, the average percentage of the dispatching cost slightly grows, whereas the average percentage of the fixed cost remains relatively stable.

These trends suggest that on average, when additional small truck options are available, a fleet composition solution tends to add more small trucks, or, replace a certain number of big trucks with an equivalent or a greater number of small trucks, which leads to a decrease in the average FCL. However, in some instances, adding more small truck options does not affect the current fleet composition solution and thus the number of trucks owned for each type does not change. This implies that adding more smaller trucks does not necessarily improve the cost-effectiveness of the fleet.

In addition, if a solution suggests considering more small truck types, the total number of trucks owned increases, while the allocation of the demand tends to be more dispersed among different truck types and thus the portion assigned to each truck type is lower, leading to lower values of average fleet and internal capacity utilization. The reliance on LTL services decreases when more truck types are considered, leading to a decrease in the percentage of LTL cost that dominates the reduction in the average total expected cost.

2.5.4.3 Impact of demand mean and standard deviation

To investigate the impact of the mean and standard deviation of demand, we create four sets of normal demand samples of size 1000 based on four groups of mean values and standard deviations, representing four different demand categories: low mean low uncertainty (L, L), low mean high uncertainty (L, H), high mean low uncertainty (H, L), and high mean high uncertainty (H, H). Each group includes 50 independently generated samples, and each sample corresponds to an instance index. Table 2.14 summarizes the parameters used for each demand category.

Table 2.14: Four groups of demand parameters.

	(L, L)	(L, H)	(H, L)	(H, H)
Mean ($\times 10^4$)	105	105	125	125
Standard deviation ($\times 10^4$)	16.25	26.25	16.25	26.25

We solve each base instance with $T = 5$ using its associated sample in each demand category, respectively. We solve 50 instances for each demand category and in total 200 problems are solved for all demand categories. The results of metrics for each instance are included in Appendix B.10. Table 2.15 compares the mean values of the metrics between four demand categories.

We first compare the results in group (L, L) and group (L, H). For a low-demand market, as the demand uncertainty increases, the mean values of FCL, fleet utilization, and internal capacity utilization decrease. This indicates that when demand uncertainty is high, more small trucks are acquired in order to fulfill any unexpected shipping requirements that require high shipping flexibility, while the randomness of the demand leads to a higher portion of idle internal capacities in a period, which causes the reduction in the values of U_F and U_C, as well as the decreasing percent-

Table 2.15: The mean values of the metrics with different demand parameters.

	(L, L)	(L, H)	(H, L)	(H, H)
Avg. FCL (lbs)	33076.68	32681.53	33175.00	32827.57
Avg. U_F (%)	85.33	78.47	87.46	81.22
Avg. U_C (%)	86.54	79.96	88.61	82.65
Avg. LTL units (%)	1.89	2.97	1.63	2.52
Avg. fixed cost (%)	16.26	16.95	16.03	16.64
Avg. dispatching cost (%)	80.90	78.68	81.50	79.62
Avg. LTL cost ((%)	2.84	4.37	2.47	3.73
Avg. total cost ($)	235350.75	239326.33	278723.41	282987.73

age of dispatching cost. At the same time, more LTL units are used for meeting excessive shipping requirements as the demand uncertainty increases, leading to significant growth in the percentage of LTL units. As the demand uncertainty increases, the total expected cost also demonstrates an increase due to the growth in both the LTL cost and the fixed cost (incurred by the addition of small trucks).

However, when comparing the results in group (L, L) and group (H, L), we observe opposite trends in the variation of the metrics. By fixing the standard deviation, increasing the mean of the demand leads to a clear upward trend in the mean values of the FCL, U_F, and U_C. This is because when the demand is high but relatively stable, the shipping requirements in a period become more predictable, which leads to a lower percentage of idle trucks during a period, and thus both the fleet utilization and the internal capacity utilization are improved. Considering economies of scale, it is more economical to acquire more big trucks to fulfill the growing demand, leading to an increased FCL. On the other hand, a higher value of fleet utilization implies a growing number of trucks shipped, which results in an increase in the dispatching cost that dominates the growth of the total expected cost.

When increasing both the demand mean value and the standard deviation, the average total expected cost demonstrates a noticeable increase, which is dominated by the growth of the fixed cost and the LTL cost. The average FCL, as well as the fleet utilization and the internal capacity utilization, undergo a reduction.

2.5.4.4 *Impact of demand distribution*

To investigate the impact of the distribution of the demand, we create three groups of demand samples based on normal, uniform, and gamma distributions, respectively. We use the mean value and the standard deviation for the (L, H) category listed in Table 2.14 for each distribution. We create 50 independent samples for each distribution, and use each sample to solve the corresponding base instance. Appendix B.11 includes the results for all instances. We summarize the average metrics for each distribution in Table 2.16.

Table 2.16: The mean values of the metrics with different demand distributions.

	Normal	Uniform	Gamma
Avg. FCL (lbs)	32681.53	32486.13	32596.04
Avg. U_F (%)	78.47	76.35	78.34
Avg. U_C (%)	79.96	77.92	79.94
Avg. LTL units (%)	3.07	2.04	3.69
Avg. fixed cost (%)	16.95	17.51	16.78
Avg. dispatching cost (%)	78.68	79.60	78.01
Avg. LTL cost ((%)	4.37	2.90	5.21
Avg. total cost ($)	239326.33	239320.19	240517.66

Among the three demand distributions, the uniform distribution results in the lowest average FCL, average fleet and internal capacity utilization, and the average percentage of LTL units. Although the average percentages of fixed and dispatching costs in the uniform distribution group are the highest, the corresponding total expected cost is lower than those associated with normal and gamma distributions. Overall, the mean values of the metrics in the normal and gamma distribution groups are relatively close, wherein the normal distribution group has higher values of the average FCL, the average fleet utilization, the average internal capacity utilization, and the average cost percentages associated with internal trucks. The gamma distribution group has the highest metric values associated with LTL, and the highest average total expected cost.

When the demand is uniformly distributed, a lower average FCL implies that the truck types in a fleet are more diverse because they need to fulfill demand at each level with equal probabilities, while trucks of each type can only be utilized with certain probabilities, leading to lower values of fleet and capacity utilization. When the demand distribution has a 'bell' shape like normal or gamma distribution, as the expected demand is centered around a fixed value, the demand can be satisfied by the main truck types most of the time, leading to a less diverse fleet and higher utilization. Overall, the fleet composition solutions based on the three demand distributions differ in utilization and cost percentages, while the resulting total expected costs are relatively close.

2.6 Conclusions and future research directions

In this chapter, we explore the fleet composition problem with stochastic demand by accounting for using both TL and LTL shipping options when allowing for LTL shipments even with internal capacity being available. We propose a solution approach that offers analytical solutions for a fleet with lower diversity, as well as a two-stage stochastic programming model to incorporate a wider range of truck types. We solve the proposed model using the SAA method and develop a decomposition-based algorithm that provides effective solutions within relatively fast computing time. Our numerical results reveal that the fleet composition decisions are affected by the LTL freight rate, the number of truck types, and the properties of demand.

Although we have briefly discussed the potential extension for a homogeneous fleet by adopting a stochastic LTL rate, incorporating a stochastic and dynamic LTL rate in models for the heterogeneous fleet represents an interesting and challenging avenue for future work. In practice, a fleet usually serves multiple customers located in different regions, and tackling the allocation of internal capacity to different destinations and determining shipping priorities will also contribute to the current research.

3. OPTIMIZATION MODELS FOR AMBULANCE BUS ROUTING IN DISASTER SITUATIONS

3.1 Introduction

Natural disasters, such as earthquakes, hurricanes, and tornadoes, can have the potential to wreak havoc and inflict catastrophic harm on human society. Along with loss of lives, a disaster can also lead to huge economic losses that are hard to recover from. The year 2022 witnessed 387 recorded natural calamities that lead to the loss of 30,704 lives and nearly $223.8 billion in total over the globe [54]. Effective and efficient rescue operations and relief activities in the immediate aftermath of a disaster are critical to minimizing the post-disaster damages to communities and harmful impacts on public health. Lack of coordination and timely response to the aftermath of a disaster can push casualties into health risks which can cause further death.

The Emergency Medical Services (EMS) system plays a crucial role in ensuring successful disaster management. One critical component of an EMS system in the aftermath of disasters is making quick decisions on allocating and routing ambulances. The post-disaster environment is extremely time-sensitive because the conditions of the casualties become worse as time passes. In order to maximize the number of survivors, routing decisions must be made quickly to ensure prompt responses. Typical optimization models formulated for ambulance routing problems at the operational level require detailed input information, such as the exact locations of casualties, and the distances between casualties and available ambulances, which may not be fully available when a routing decision is needed. To guide decision making strategies, we incorporate tour length approximations into routing models when the only available information is the total number of casualties, which gives insight into the structure of optimal routing solutions.

Additionally, during large-scale disasters or massive-casualty incidents, traditional ambulances may become overwhelmed due to high demand, and therefore alternative transportation resources are deemed necessary, such as ambulance buses. An ambulance bus is specifically designed and equipped to accommodate a group of casualties within a single vehicle, which enables efficient transportation for medical evacuation in mass casualty incidents. Ambulance buses have been examined in various disaster management research projects (see e.g., [55, 56]) and have proven to be effective in real-life emergency situations (see e.g., [57, 58]).

In this chapter, our main goal is to investigate using an ambulance bus instead of traditional ambulances in the aftermath of a disaster by developing strategic optimization models that maximize the expected number of survivors, considering time-dependent survival probabilities of the casualties. We aim to provide insights and identify basic principles that lead to effective routing plans in a short response time without the need for precise location and distance information.

Our contributions are summarized as follows. First, we propose optimization models that maximize the total expected number of survivors, using an ambulance bus instead of a fleet of traditional ambulances. Second, we incorporate tour length approximation in our models while considering time-dependent survival probabilities, which reduces the need for detailed information on locations and distances. Furthermore, we discuss both linear and nonlinear survival probability functions and achieve analytical results that lead to effective routing solutions under linear survival probability functions. Finally, we propose two heuristic algorithms that solve the ambulance bus routing problem with linear and nonlinear survival probability functions, respectively, with high-quality solutions in a very short computing time.

The rest of this chapter is organized as follows. We review relevant literature in Section 3.2, followed by a problem definition in Section 3.3. Section 3.4 proposes an optimization model with a linear survival probability function and develops the golden ratio heuristic algorithm. Section 3.5 characterizes the optimization model under more general forms of survival probability functions and develops a heuristic algorithm based on dynamic programming. We conduct numerical experiments and evaluate the performance of the two proposed heuristic algorithms in Section 3.6. We further perform sensitivity analysis in Section 3.6.3 to the behavior of the solutions in terms of vehicle capacity and the survival probability function.

3.2 Literature review

Transportation prioritization, casualty assignment to hospitals/vehicles, and vehicle routing are critical decisions in disaster responses [59]. Existing studies devoted to supporting these decisions for ambulances in the EMS fall in three streams: locating, dispatching, and routing [60, 61]. Our work is primarily related to ambulance dispatching (i.e., determining the assignments of ambulances) and routing (i.e., determining the visit sequence of ambulances) problems. We focus on research that addresses allocating casualties to ambulances and combine the discussions on both of these streams.

Gong and Batta [62] propose a deterministic model that determines the allocation and reallocation of ambulances to clusters of casualties while minimizing the completion time for all clusters. Talarico et al. [63] study an ambulance routing problem for disaster response given a fleet of ambulances, in which patients are prioritized based on severity of injuries. Two optimization models are proposed to obtain routing plans with the objective of minimizing the latest service completion time, and a metaheuristic is developed that solves large-scale instances in short computing time. Repoussis et al. [64] formulate a mixed-integer programming model that combines ambulance dispatching, patient-to-hospital assignment, and treatment ordering while minimizing the overall response time and the total flow time required to treat all patients. Based on an ant colony system algorithm, Mouhcine et al. [65] propose a distributed strategy that generates the optimal path dynamically for multiple ambulances according to different environmental factors while minimizing the travel time. Yoon and Albert [66] present a Markov decision process model that dynamically assigns two ambulance types to emergency calls while allowing more than one ambulance type to be sent to the same call. The objective is to maximize the expected total reward (including the utilities of serving calls and the coverage of patients) over a given time horizon.

The majority of the aforementioned papers use a minimization objective in terms of time, such as minimizing travel time, completion time, or response time. Although time is typically used to assess the effectiveness of an EMS, other measurements should be considered to ensure capturing the ultimate goal of the EMS, which is to save lives. In this respect, models aiming to maximize the expected number of survivors considering time-dependent survival probabilities are increasingly investigated in EMS resource allocation problems. Sacco et al. [67] develop a linear programming model that maximizes the expected number of survivors while considering the timing and availability of medical resources in the aftermath of a mass-casualty incident, in which the survival probabilities of the victims are predicted based on a set of vital signs and victim deterioration is estimated using the Delphi technique. Erkut et al. [68] incorporate a survival function into maximum covering models that allocate ambulance resources to emergency service stations while maximizing the expected number of survivors. Knight et al. [69] present an extension of the model in [68] by considering multiple patient classes, wherein patients in different classes have different survival functions. Mills et al. [70] construct a fluid model for the mass-casualty triage problem when accounting for time-dependent survival probabilities given a number of ambulances, with the objective of maximizing the expected number of survivors. Bandara et al. [71] propose a priority-based heuristic dispatching policy that attempts to increase the survival probability of patients. Their numerical results show that implementing the proposed dispatching policy can effectively improve patient survivability, the average response time, and the percentage of calls for higher priority calls. Sung and Lee [72] propose a mixed-integer programming model that determines the allocation of ambulances to a given number of victims while maximizing the number of expected survivals, in which victims are classified into two groups and have time-sensitive survival probabilities. Mills et al. [73] develop a Markov decision process and two heuristic policies to simultaneously allocate ambulances to casualty locations and select hospitals for transportation, given limited emergency vehicles and treatment capacity at the hospitals. The objective in their model is to maximize the expected total discounted reward, which can be represented by maximizing the expected number of survivors.

The survival-maximizing models developed in the aforementioned papers commonly allocate a fixed number of ambulances to a given set of casualties, assuming each vehicle serves one casualty at a time and the travel time (or the distribution of the travel times) only depends on the distance between the casualty's location and the destination. However, when using a single ambulance bus to pick up a group of casualties in sequence, the travel time required during a trip also relies on the number of casualties on that trip. Consequently, the number of trips and the overall completion time when all trips are finished depend on the number of casualties assigned to each trip. Therefore, when incorporating time-dependent survival rates in the application of a single ambulance bus, we must balance the number of trips and the size of each trip in order to maximize the expected survivals.

To estimate the travel time of transporting a given number of casualties using an ambulance bus, we refer to the tour length approximation models in literature for routing problems (see e.g., [74, 75, 76, 77, 78]). If the total number of casualties exceeds the capacity of the ambulance bus, the casualties must be split into multiple routes and each route should follow the shortest *traveling-salesman problem* (TSP) tour to maximize the casualties' survival probabilities. Beardwood et al. [74] have shown that for a set of n points distributed in a bounded area of size A, the optimal TSP tour length asymptotically converges to $\gamma\sqrt{nA}$, where γ is a constant. We assume that the travel time is proportional to the approximated tour length and thus use this tour length approximation to represent the estimated time required for each route.

3.3 Problem description

We consider the aftermath of a disaster in a compact area of size A, where there are N casualties geographically dispersed and a rescue team runs a single ambulance bus with a capacity C to pick up each casualty in sequence. One or more routes are needed to evacuate all casualties, and each route is a route starting from a medical center (i.e., each route originates from the medical center and returns while transporting a group of casualties). The team must make a quick decision on the number of routes needed and the size of each route before detailed location information becomes available for advanced routing schedules. Let $S(t)$ be the probability of survival for a casualty if she arrives at the medical center and receives treatment at time t, where $S(t)$ is a nonincreasing function of t. Our objective is to maximize the total expected number of survivors.

As time passes, the conditions of casualties become worse. The sooner the casualties are sent back to the medical center, the higher their chance of survival. Because the ambulance bus has limited capacity, running multiple routes is required if a single route cannot accommodate all casualties. If a small number of casualties is included on each route, the travel time needed for each route is small, so the casualties who are included on the first few routes can be transported to the medical center sooner, while casualties on the last few routes may arrive late because of the need to run a large number of routes. In contrast, if the team picks up more casualties on each route, it takes longer for the first few routes to travel back to the medical center, whereas the casualties on the last few routes can arrive relatively earlier as the team runs fewer routes. Therefore, the team must consider the trade-off between the number of casualties to include on each route and the total number of routes, determining whether to use the 'smaller route sizes, more routes' strategy or the 'larger route sizes, fewer routes' strategy.

Letting n_i denote the number of casualties to include on route i for $i = 1, ..., N$ such that $\sum_{i=1}^{N} n_i = N$, we assume without loss of generality the rescue team runs the routes in index order. Note that the team can run at most N routes with $n_i = 1$ for $i = 1, \ldots, N$, and at least one route with $n_1 = N$ and $n_i = 0$ for $i = 2, \ldots, N$. We assume the casualties are uniformly dispersed in the disaster region with an area of A. Let $A(n_i)$ denote the area of the subregion covered by route i when transporting n_i casualties. If the rescue team travels across the entire region for each route, we have $A(n_i) = A$; if the area A is divided equally into k ($k \leq N$) non-overlapping subregions, we have $A(n_i) = A/k$ for all $i = 1, ..., k$. In both cases, $A(n_i)$ is fixed and identical for all routes. We assume that the team cannot assign any priority to the casualties and needs to pick them up following the order the requests are received, and we use $A(n_i) = A$ for all routes.

Assuming the completion time of a route is proportional to the tour length of the route, we use the approximated tour length $\gamma\sqrt{An_j}$ to represent the estimated time required by route j for n_j casualties, where γ is a constant. Letting $\beta = \gamma\sqrt{A}$ and assuming the first route starts from time 0, the time at the completion of the i^{th} route equals

$$t_i = \beta \sum_{j=1}^{i} \sqrt{n_j}. \tag{3.1}$$

We will use this time approximation for characterizing our models in the following sections.

We assume the survival probability function of each casualty is identical. Based on the route completion time in Equation (3.1), the expected number of survivors on route i that includes n_i casualties equals $n_i S(t_i) = n_i S\left(\beta \sum_{j=1}^{i} \sqrt{n_j}\right)$. To maximize the total expected number of survivors over all routes, we solve the following model:

$$[\mathbb{P}] \quad \text{Maximize} \quad \sum_{i=1}^{N} n_i S\left(\beta \sum_{j=1}^{i} \sqrt{n_j}\right) \tag{3.2}$$

$$\text{Subject to:} \quad \sum_{i=1}^{N} n_i = N, \tag{3.3}$$

$$n_i \leq C, \qquad i = 1, \ldots, N, \tag{3.4}$$

$$n_i \in \mathbb{Z}^+, \qquad i = 1, \ldots, N. \tag{3.5}$$

The objective function (3.2) maximizes the total expected number of survivors. Constraint (3.3) ensures all casualties are visited. Constraints (3.4) ensure the number of casualties included on each route is subject to the vehicle capacity. The difficulty of solving this model depends in large part on the form of the survival probability function $S(\cdot)$.

3.4 Stylized model with linear survival probability function

In this section, we consider a linear survival probability function $S(t) = a - bt$ with $0 < a \leq 1$ and $b > 0$, where $a/b \geq \beta N$ such that each casualty has nonnegative survival probability when arriving at the medical center. With $S(t) = a - bt$, the objective function (3.2) in problem $\mathbb{P}$ becomes

$$aN - b\beta \sum_{i=1}^{N} n_i \left(\sum_{j=1}^{i} \sqrt{n_j} \right). \tag{3.6}$$

Because the values of a, b and β are constant, given N, the optimal solution can be obtained by solving the following equivalent problem:

$$[\mathbb{EP}] \quad \text{Minimize} \quad \textstyle\sum_{i=1}^{N} n_i \left(\sum_{j=1}^{i} \sqrt{n_j} \right) \tag{3.7}$$

$$\text{Subject to:} \quad \textstyle\sum_{i=1}^{N} n_i = N, \tag{3.8}$$

$$n_i \leq C, \qquad i = 1, \dots, N, \tag{3.9}$$

$$n_i \in \mathbb{Z}^+, \qquad i = 1, \dots, N. \tag{3.10}$$

We can show that an optimal solution for problem $\mathbb{EP}$ must have successively nonincreasing decision variable values, that is, $n_i \geq n_j$ for $i < j$. This is formalized in Proposition 3.4.1. We include the proofs for all propositions in Appendix C.1.

Proposition 3.4.1. *The route sizes in an optimal solution are successively nonincreasing. That is,* $n_i \geq n_j$ *holds in an optimal solution, where* $i, j \in \{1, \dots, N\}$ *and* $i < j$.

3.4.1 KKT solutions for uncapacitated routes

To gain some insight into this problem, we initially ignore the vehicle capacity by assuming $N \leq C$ and consider the following uncapacited problem:

$$[\mathbb{EP}\text{-}\mathbb{U}] \quad \text{Minimize} \quad \textstyle\sum_{i=1}^{N} n_i \left(\sum_{j=1}^{i} \sqrt{n_j} \right) \tag{3.11}$$

$$\text{Subject to:} \quad \textstyle\sum_{i=1}^{N} n_i = N, \tag{3.12}$$

$$n_i \in \mathbb{Z}^+, \qquad i = 1, \dots, N. \tag{3.13}$$

In the above model, one extreme case is when $n_1 = N$ and $n_2 = n_3 ... = n_N = 0$, where the objective value equals $N\sqrt{N}$; at the other extreme, if $n_i = 1$ for all i, the objective value equals $N(N+1)/2$. Because $(N+1)/2 > \sqrt{N}$ holds for any $N \geq 2$, including all casualties in a single route always provides a better solution than transporting one casualty at a time using N individual routes. We can show that in the uncapacitated case, at most $N-1$ routes are needed in an optimal solution. This is proven in Proposition 3.4.2.

Proposition 3.4.2. *Given N casualties, if $1 < N \leq C$, an optimal solution has at most $N - 1$ routes.*

To solve problem $\mathbb{EP}$-$\mathbb{U}$, we first consider its relaxation problem:

$$\text{Minimize} \quad \sum_{i=1}^{N} n_i \left(\sum_{j=1}^{i} \sqrt{n_j} \right) \tag{3.14}$$

$$\text{Subject to:} \quad \sum_{i=1}^{N} n_i = N, \tag{3.15}$$

$$-n_i \leq 0, \qquad i = 1, \ldots, N. \tag{3.16}$$

Letting α denote the dual variable associated with the equality constraint and λ_i denote the dual variable associated with the i^{th} nonnegative constraint, the necessary KKT conditions can be written as

$$\frac{3\sqrt{n_i}}{2} + \sum_{j=1}^{i-1} \sqrt{n_j} + \frac{\sum_{j=i+1}^{N} n_j}{2\sqrt{n_i}} - \alpha - \lambda_i = 0, \quad i = 1, \ldots, N, \tag{3.17}$$

$$\sum_{i=1}^{N} n_i = N, \tag{3.18}$$

$$\lambda_i n_i = 0, \qquad i = 1, \ldots, N, \tag{3.19}$$

$$n_i \geq 0, \qquad i = 1, \ldots, N, \tag{3.20}$$

$$\lambda_i \geq 0, \qquad i = 1, \ldots, N. \tag{3.21}$$

Observe that $n_i > 0$ implies $\lambda_i = 0$; thus, for any i such that $n_i > 0$, we have

$$\frac{3\sqrt{n_i}}{2} + \sum_{j=1}^{i-1} \sqrt{n_j} + \frac{\sum_{j=i+1}^{N} n_j}{2\sqrt{n_i}} = \alpha. \tag{3.22}$$

The solution $n_1 = N$, $n_i = 0$, $i = 2, ..., N$ is a KKT point with $\alpha = 3\sqrt{N}/2$, while the solution $n_i = 1$ for $i = 1, \ldots, N$ is not a KKT point because in this case condition (3.17) requires $2 + N = 2 + N - 1$.

As a result, if $n_i > 0$ and $n_\ell > 0$, we must have

$$\tfrac{3\sqrt{n_i}}{2} + \sum_{j=1}^{i-1} \sqrt{n_j} + \tfrac{\sum_{j=i+1}^{N} n_j}{2\sqrt{n_i}} = \tfrac{3\sqrt{n_\ell}}{2} + \sum_{j=1}^{\ell-1} \sqrt{n_j} + \tfrac{\sum_{j=\ell+1}^{N} n_j}{2\sqrt{n_\ell}}. \tag{3.23}$$

If an optimal solution contains k routes ($k < N$) in which $n_i > 0$ for $i = 1, \ldots, k$ and $n_j = 0$ for $j = k+1, \ldots, N$, the term $\frac{\sum_{j=k+1}^{N} n_j}{2\sqrt{n_k}}$ then becomes zero; then the above implies

$$\tfrac{3\sqrt{n_k}}{2} + \sqrt{n_{k-1}} = \tfrac{3\sqrt{n_{k-1}}}{2} + \tfrac{n_k}{2\sqrt{n_{k-1}}},$$

or

$$n_{k-1} - 3\sqrt{n_k}\sqrt{n_{k-1}} + n_k = 0.$$

If we fix n_k and solve the above quadratic equation for $\sqrt{n_{k-1}}$, we obtain

$$\sqrt{n_{k-1}} = \sqrt{n_k}\left(\tfrac{3\pm\sqrt{5}}{2}\right).$$

According to Proposition 3.4.1, an optimal solution must satisfy $n_{k-1} \geq n_k$, so we only need to consider $\sqrt{n_{k-1}} = \sqrt{n_k}\left(\frac{3+\sqrt{5}}{2}\right)$. Letting $\varphi = \frac{1+\sqrt{5}}{2}$, which is the *golden ratio*, we can then write $\frac{3+\sqrt{5}}{2} = 1+\varphi$, implying $\sqrt{n_{k-1}} = (1+\varphi)\sqrt{n_k}$, or, $\sqrt{n_k} = (2-\varphi)\sqrt{n_{k-1}}$, noting that $\frac{1}{1+\varphi} = 2-\varphi$.

We first consider an optimal solution with two routes ($k = 2$) with $n_1 + n_2 = N$. Based on the above discussion, an optimal solution must satisfy $\sqrt{n_2} = (2-\varphi)\sqrt{n_1}$, implying $n_1 = \frac{N}{1+(2-\varphi)^2}$. We therefore can conclude that for a two-route optimal solution, $n_1^* > n_2^*$ must hold, with

$$n_1^* = \frac{N}{1+(2-\varphi)^2} \approx 0.8727N,$$
$$n_2^* = \frac{(2-\varphi)^2 N}{1+(2-\varphi)^2} \approx 0.1273N.$$

More generally, for an optimal solution with k routes and $2 < k < N$, according to (3.23), we can write

$$\frac{3\sqrt{n_{k-1}}}{2} + \sqrt{n_{k-2}} + \frac{n_k}{2\sqrt{n_{k-1}}} = \frac{3\sqrt{n_{k-2}}}{2} + \frac{n_{k-1}+n_k}{2\sqrt{n_{k-2}}}. \tag{3.24}$$

We know that an optimal KKT point must satisfy $\sqrt{n_k} = (2-\varphi)\sqrt{n_{k-1}}$, and the above becomes

$$\frac{\sqrt{n_{k-2}}}{2} + \frac{(1+(2-\varphi)^2)n_{k-1}}{2\sqrt{n_{k-2}}} = \frac{(3+(2-\varphi)^2)\sqrt{n_{k-1}}}{2}, \tag{3.25}$$

which is equivalent to

$$n_{k-2} - (3+(2-\varphi)^2)\sqrt{n_{k-1}}\sqrt{n_{k-2}} + (1+(2-\varphi)^2)n_{k-1} = 0. \tag{3.26}$$

By Proposition 3.4.1, the above quadratic equation is solved at

$$\sqrt{n_{k-2}} = \sqrt{n_{k-1}}\left(\frac{3+(2-\varphi)^2 + \sqrt{(3+(2-\varphi)^2)^2 - 4(1+(2-\varphi)^2)}}{2}\right). \tag{3.27}$$

Letting $m_{k-1} = (1+\varphi)^2$ and $\frac{1}{m_{k-1}} = (2-\varphi)^2$, the above can be written as

$$n_{k-2} = m_{k-2}n_{k-1},$$

where

$$m_{k-2} = \left(\frac{3+\frac{1}{m_{k-1}} + \sqrt{\left(3+\frac{1}{m_{k-1}}\right)^2 - 4\left(1+\frac{1}{m_{k-1}}\right)}}{2}\right)^2.$$

We can apply a similar approach to n_{k-2} and n_{k-3}, and obtain

$$n_{k-3} = m_{k-3}n_{k-2},$$

where

$$m_{k-3} = \left(\frac{3 + \frac{1}{m_{k-1}} + \frac{1}{m_{k-1}m_{k-2}} + \sqrt{\left(3 + \frac{1}{m_{k-1}} + \frac{1}{m_{k-1}m_{k-2}}\right)^2 - 4\left(1 + \frac{1}{m_{k-1}} + \frac{1}{m_{k-1}m_{k-2}}\right)}}{2} \right)^2 \quad (3.28)$$

In fact, a continuing pattern exists in which $n_{k-\ell} = m_{k-\ell} n_{k-\ell+1}$ for $\ell = 2, \ldots, k-1$, where $m_{k-1} = \rho_{k-1} = (1+\varphi)^2$ and

$$m_{k-\ell} = \frac{1}{4}\left(3 + \sum_{i=1}^{\ell-1} \frac{1}{\rho_{k-i}} + \sqrt{\left(3 + \sum_{i=1}^{\ell-1} \frac{1}{\rho_{k-i}}\right)^2 - 4\left(1 + \sum_{i=1}^{\ell-1} \frac{1}{\rho_{k-i}}\right)}\right)^2, \quad (3.29)$$

with

$$\rho_{k-\ell} = \prod_{i=1}^{\ell} m_{k-i}.$$

Consequently, for an optimal solution with k routes, we can characterize m_i for $i = 1, \ldots, k-1$, where $n_{k-1} = m_{k-1} n_k, n_{k-2} = m_{k-2} n_{k-1}, \ldots, n_1 = m_1 n_2$. This implies that $n_j = n_k \prod_{i=1}^{k-j} m_{k-i} = \rho_j n_k$. Then we can write

$$\sum_{j=1}^{k} n_j = n_k \left(1 + \sum_{j=1}^{k-1} \rho_j\right),$$

which implies

$$n_k = \frac{N}{1 + \sum_{j=1}^{k-1} \rho_j},$$

as well as

$$n_j = N \frac{\rho_j}{1 + \sum_{i=1}^{k-1} \rho_i}, \quad j = 1, \ldots, k-1.$$

Note that with $m_{k-1} = \rho_{k-1} = (1+\varphi)^2$, the coefficient $\frac{\rho_j}{1+\sum_{i=1}^{k-1} \rho_i}$ on the right-hand side of the above equation only depends on the value of k. In particular, the coefficient $r_1 = \frac{\rho_1}{1+\sum_{i=1}^{k-1} \rho_i}$ associated with n_1 represents the percentage (among N casualties) to include on the first route. Table 3.1 lists the value of this percentage when the value of k varies. We observe that r_1 tends to converge to approximately 86.6% for $k \geq 5$.

Table 3.1: The value of r_1 as k changes.

k	r_1	k	r_1
2	0.872678	7	0.866025
3	0.866352	8	0.866025
4	0.866041	9	0.866025
5	0.866026	10	0.866025
6	0.866025	11	0.866025

3.4.2 KKT solutions for capacitated routes

We next characterize the KKT solutions for capacitated routes with $N > C$. We consider the following relaxation of problem $\mathbb{EP}$:

$$[\mathbb{EPR}] \quad \text{Minimize} \quad \sum_{i=1}^{N} n_i \left(\sum_{j=1}^{i} \sqrt{n_j} \right) \tag{3.30}$$

$$\text{Subject to:} \quad \sum_{i=1}^{N} n_i = N, \tag{3.31}$$

$$n_i \leq C, \qquad i = 1, \ldots, N, \tag{3.32}$$

$$-n_i \leq 0, \qquad i = 1, \ldots, N. \tag{3.33}$$

As before, we use α to denote the KKT multiplier associated with the equality constraint, and use λ_i to denote the KKT multiplier associated with the i^{th} nonnegative constraint. Letting π_i denote the multiplier associated with the i^{th} capacity constraint, the necessary KKT conditions can be written as

$$\frac{3\sqrt{n_i}}{2} + \sum_{j=1}^{i-1} \sqrt{n_j} + \frac{\sum_{j=i+1}^{N} n_j}{2\sqrt{n_i}} - \alpha + \pi_i - \lambda_i = 0, \quad i = 1, \ldots, N, \tag{3.34}$$

$$\sum_{i=1}^{N} n_i = N, \tag{3.35}$$

$$n_i \leq C, \quad i = 1, \ldots, N, \tag{3.36}$$

$$\pi_i (C - n_i) = 0, \quad i = 1, \ldots, N, \tag{3.37}$$

$$\lambda_i n_i = 0, \quad i = 1, \ldots, N, \tag{3.38}$$

$$n_i \geq 0, \quad i = 1, \ldots, N, \tag{3.39}$$

$$\lambda_i \geq 0, \quad i = 1, \ldots, N. \tag{3.40}$$

Observe that $n_i > 0$ implies $\lambda_i = 0$, while $n_i < C$ implies $\pi_i = 0$. Therefore, for any i such that $0 < n_i < C$, we also obtain Equation (3.22). As a result, for any pair of n_i and n_ℓ such that $0 < n_i < C$ and $0 < n_\ell < C$, Equation (3.23) must hold; if $n_i = C$ and $0 < n_\ell < C$, as $\pi_i \geq 0$, a KKT solution must satisfy

$$\frac{3\sqrt{n_i}}{2} + \sum_{j=1}^{i-1} \sqrt{n_j} + \frac{\sum_{j=i+1}^{N} n_j}{2\sqrt{n_i}} \leq \frac{3\sqrt{n_\ell}}{2} + \sum_{j=1}^{\ell-1} \sqrt{n_j} + \frac{\sum_{j=\ell+1}^{N} n_j}{2\sqrt{n_\ell}}. \tag{3.41}$$

Consider an optimal solution that contains two routes, which can only occur if $N \leq 2C$. If $N = 2C$, $n_1 = n_2$ is the only feasible solution; we therefore consider cases with $C < N < 2C$. If $C \geq \frac{N}{1+(2-\varphi)^2}$, i.e., $N \leq (1 + (2-\varphi)^2)C$, based on the discussion for two routes in Section 3.4.1, this is equivalent to the uncapacitated case, and thus $n_1 = \frac{N}{1+(2-\varphi)^2} \leq C$ and $n_2 = N - n_1 = \frac{(2-\varphi)^2 N}{1+(2-\varphi)^2} < C$ provides a KKT solution. Otherwise, if $(1 + (2-\varphi)^2)C < N < 2C$, consider the solution with $n_1 = C$ and $n_2 = N - C$, which satisfies Equation (3.41) if and only if

$$\frac{3\sqrt{C}}{2} + \frac{N - C}{2\sqrt{C}} \leq \frac{3\sqrt{N - C}}{2} + \sqrt{C}, \tag{3.42}$$

which holds if and only if

$$N^2 - 9NC + 9C^2 \leq 0. \tag{3.43}$$

Note that the left-hand side of the above inequality is decreasing in N for $0 \leq N \leq 4.5C$. Observe that at $N = 2C$, the left-hand side equals $-5C^2$, while at $N = (1 + (2 - \varphi)^2)C$, the left-hand side equals zero, Thus, the above inequality holds for any N such that $(1 + (2 - \varphi)^2)C < N < 2C$, implying the solution $(n_1, n_2) = (C, N-C)$ is a KKT solution. The solution $(n_1, n_2) = (N-C, C)$ requires

$$\frac{3\sqrt{C}}{2} + \sqrt{N-C} \leq \frac{3\sqrt{N-C}}{2} + \frac{C}{2\sqrt{N-C}} \tag{3.44}$$

in order to be a KKT solution, which is equivalent to

$$N^2 - 9NC + 9C^2 \geq 0 \tag{3.45}$$

and cannot hold for $(1 + (2 - \varphi)^2)C < N < 2C$. Thus $(n_1, n_2) = (N - C, C)$ is not a KKT point.

More generally, consider a solution containing $K + 1$ routes, where $K = \lfloor N/C \rfloor$ and $N \leq (K + 1)C$ such that the first K routes are filled to capacity, i.e., $n_i = C$ for $i = 1, \ldots, K$ and $n_{K+1} = N - KC$. Consider any two successive routes i and $i + 1$, with $i + 1 \leq K + 1$. The KKT conditions require

$$\frac{3\sqrt{C}}{2} + \sum_{j=1}^{i-1} \sqrt{n_j} + \frac{\sum_{j=i+1}^{N} \sqrt{n_j}}{2\sqrt{C}} + \pi_i = \frac{3\sqrt{C}}{2} + \sum_{j=1}^{i} \sqrt{n_j} + \frac{\sum_{j=i+2}^{N} n_j}{2\sqrt{C}} + \pi_{i+1},$$

which is equivalent to

$$\pi_i - \pi_{i+1} = \frac{\sqrt{C}}{2}.$$

For the last two routes K and $K+1$, the KKT conditions require

$$\frac{3\sqrt{C}}{2} + \frac{N-KC}{2\sqrt{C}} + \pi_K = \frac{3\sqrt{N-KC}}{2} + \sqrt{C},$$

which is satisfied if and only if

$$N - KC - 3\sqrt{C}\sqrt{N-KC} + C \leq 0.$$

Letting $h = \sqrt{N-KC}$, we can write the above to

$$h^2 - 3\sqrt{C}h + C \leq 0. \tag{3.46}$$

We can show that condition (3.46) holds if and only if $\sqrt{C}\left(\frac{3-\sqrt{5}}{2}\right) \leq h \leq \sqrt{C}\left(\frac{3+\sqrt{5}}{2}\right)$, which is equivalent to $(2-\varphi)^2 C \leq N - KC \leq (1+\varphi)^2 C$ because $N - KC \leq C < (1+\varphi)^2 C$. Thus, a solution with K full routes and one addition route containing $N - KC$ casualties can be a KKT point if $(2-\varphi)^2 C \leq N - KC \leq C$.

If $N - KC < (2-\varphi)^2 C$, however, we have $N - (K-1)C < (1 + (2-\varphi)^2)C$, where $\frac{1}{1+(2-\varphi)^2} \approx 0.8727$, implying approximately 87% of the remaining casualties after running $K-1$ full routes is capacity feasible. Thus, the solution with $K-1$ full routes and two additional routes containing approximately 87% and 13% of the remaining casualties provides a KKT solution. We refer to this as the '87% rule' in further discussion (while the 87% rule applies to the problem's relaxation, we will use this term loosely in referring to integer solution in which a route contains approximately 87% of the remaining casualties).

3.4.3 The GR heuristic algorithm

The KKT solutions characterized in the previous section can only solve the relaxation of problem $\mathbb{EP}$. Because of the intractability of problem $\mathbb{EP}$ due to integer decision variables and the nonconvexity of the objective function, when the value of N is sufficiently large, using classical methods (such as branch-and-bound and cutting planes) to solve problem $\mathbb{EP}$ can be extremely costly. Fortunately, we can leverage the relaxation solution results to develop the following *Golden Ratio* (GR) heuristic algorithm that provides effective solutions to problem $\mathbb{EP}$.

Algorithm 2: The GR algorithm

Initialization: Given N casualties and the vehicle capacity C, calculate the number of full routes $K = \lfloor N/C \rfloor$;
if $K = 0$ **then**
  $R = N$;
else
  if $N - CK \geq (2 - \varphi)^2 C$ **then**
    Run K full routes, including C casualties on each route;
    $R = N - KC$;
  else
    Run $K - 1$ full routes, including C casualties on each route;
    $R = N - (K - 1)C$;
  end
end
Rounding: $R_\varphi = \lfloor \frac{1}{1+(2-\varphi)^2} R \rfloor$;
Completion: Pick up R_φ casualties on the next route, then include the remaining $R - R_\varphi$ on the last route.

Given the total number of casualties N and the vehicle capacity C, at the **Initialization** stage, we first compute the maximum number of full routes, which equals $K = \lfloor N/C \rfloor$. We next compute the value of $N - CK$ and compare it with the threshold value $(2 - \varphi)^2 C$. If $N - CK \geq (2 - \varphi)^2 C$, we run K full routes and leave the remaining value $R = N - KC$ for the next step; if $N - CK < (2 - \varphi)^2 C$, we run $K - 1$ full routes and the remainder becomes $R = N - (K - 1)C$. In the **Rounding** stage, we split the remainder R obtained in the previous step into two additional routes based on the 87% rule discussed in the previous section, and obtain the rounded value $R_\varphi = \lfloor \frac{1}{(2-\varphi)^2} R \rfloor$ ($\approx \lfloor 0.87R \rfloor$). Finally, in the **Completion** stage, we first include R_φ casualties in one of the additional routes, and then assign the remaining $R - R_\varphi$ to the other.

3.5 Dynamic programming model with general survival probability function

For a general survival probability function $S(t)$ which may have a complex nonlinear form, analysis of the problem's relaxation through characterization of the KKT conditions becomes prohibitively difficult. Instead, we use *dynamic programming* (DP) to enable incorporating a nonlinear $S(t)$ function. We follow the same problem description and use the tour length approximation technique discussed in Section 3.3. Suppose there are x casualties waiting for the rescue team and the current time is τ. If the team runs a single route, they can carry up to C casualties and return to the medical center at time $\tau + \beta\sqrt{\min\{x, C\}}$, and the total expected number of survivors for this route equals

$$\min\{x, C\}S(\tau + \beta\sqrt{\min\{x, C\}}).$$

If they run k routes ($k > 1$), they must first determine the number of casualties to include on the first route, denoted by y_k, and re-optimize the routing plan for the remaining $x - y_k$ casualties using $k - 1$ routes. The completion time for y_k casualties equals $\tau + \beta\sqrt{y_k}$. We consider each route as a *stage*, at which we need to take the *action* to transport y_k casualties on the current route, based on the current time τ and the number of the remaining casualties x. Let $V_k(x, \tau)$ be the optimal total expected number of survivors if there are k routes to run at time τ with x casualties remaining. Given the vehicle capacity of C, we can formulate the problem using the following recursive formulation:

$$V_k(x, \tau) = \begin{cases} \min\{x, C\}S(\tau + \beta\sqrt{\min\{x, C\}}) & \text{if } k = 1, \\ \max_{0 < y_k \leq \min\{x, C\}} \left\{ yS(\tau + \beta\sqrt{y_k}) + V_{k-1}(x - y_k, \tau + \beta\sqrt{y_k}) \right\} & \text{if } k = 2, \dots, x. \end{cases} \tag{3.47}$$

Given the total number of casualties N and a pre-determined value of k, the maximum expected number of survivors by running k routes from time τ is obtained by $V_k(N, \tau)$, with a resulting solution $[y_k, y_{k-1}, \dots, y_2, y_1]$, wherein y_k is the number of casualties included on the first route and $y_1 = N - \sum_{j=2}^{k} y_j$ is the number of casualties covered by the last route. Note that the above formulation only provides the optimal decision for a given k, it does not optimize the value of k, however. The global optimal solution is determined by $\max\{V_k(N,\tau), k = 1,\ldots,N\}$. Search techniques can be effective to solve for the optimal k when N lies within a reasonable range of integers.

We specifically consider a nonlinear survival probability function in the form of $S(t) = a - bt^c$, where $0 < a < 1$, $b, c > 0$, and $a/b \geq (\beta \sum_{i=1}^{k} \sqrt{y_i})^c$. Using the tour length approximation discussed in Section 3.3, the completion time of the first route becomes $\tau + \beta\sqrt{y_k}$, which is the start time of the next route. The completion time when all casualties are transported to the medical center equals $\tau + \beta \sum_{j=1}^{k} \sqrt{y_j}$. We can write the DP formulation as

$$V_k(x,\tau) = \begin{cases} \min\{x,C\}\left(a - b\big(\tau + \beta\sqrt{\min\{x,C\}}\big)^c\right) & \text{if } k = 1, \\ \max_{0\leq y_k \leq x}\left\{y_k\left(a - b(\tau + \beta\sqrt{y_k})^c\right) + (x - y_k)\left(a - b(\tau + \beta\sqrt{y_k} + \beta\sqrt{x - y_k})^c\right)\right\} & \text{if } k = 2,\ldots,x. \end{cases} \quad (3.48)$$

We next characterize the solution for the special case when $c = 1$, and discuss the solution approach for $c \neq 1$.

3.5.1 Alternative solution approach for linear survival probability function

If $c = 1$, the survival probability function becomes linear. For any $x \leq C$, if we run two routes starting from time τ, we have

$$\begin{aligned} V_2(x,\tau) &= \max_{0\leq y_2\leq x}\left\{y_2\big(a - b(\tau + \beta\sqrt{y_2})\big) + (x - y_2)\left(a - b(\tau + \beta\sqrt{y_2} + \beta\sqrt{x - y_2})\right)\right\} \\ &= \max_{0\leq y_2\leq x}\left\{x(a - b\tau) - bx\beta\sqrt{y_2} - b\beta\sqrt{(x - y_2)^3}\right\}. \end{aligned} \quad (3.49)$$

Consider the relaxed version of the above. Letting $z_2(y_2) = x(a - b\tau) - bx\beta\sqrt{y_2} - b\beta\sqrt{(x - y_2)^3}$, we can obtain the first-order derivative of $z_2(y_2)$ by fixing the value of x, which equals

$$\frac{\partial z_2(y_2)}{\partial y_2} = -\frac{b\beta x}{2\sqrt{y_2}} + \frac{3b\beta\sqrt{(x - y_2)}}{2}. \quad (3.50)$$

Letting the above equal zero, we have

$$9y_2^2 - 9xy_2 + x^2 = 0. \tag{3.51}$$

For a fixed value $\tilde{x}$, this univariate function of y_2 has two real roots at $(3-\sqrt{5})\tilde{x}/6$ and $(3+\sqrt{5})\tilde{x}/6$. This implies that $z_2(y_2)$ is nonincreasing in y_2 within the interval $[0,(3-\sqrt{5})\tilde{x}/6]$, then becomes nondecreasing within the interval $\left[(3-\sqrt{5})\tilde{x}/6,(3+\sqrt{5})\tilde{x}/6\right]$, after which it turns to nonincreasing again in $[(3+\sqrt{5})\tilde{x}/6,\tilde{x}]$. Therefore, $y_2^* = \left((3+\sqrt{5})\tilde{x}/6\right)$ is a local maximum. Comparing $z_2(y_2^*)$ with the extreme point value $z_2(0)$, we can show that $z_2(0) < z_2(y_2^*)$, implying that y_2^* is the global maximum solution for $y_2 \in [0,\tilde{x}]$. Thus, for any $x \leq C$ we have

$$V_2(x,\tau) = z\left((3+\sqrt{5})x/6\right) = x(a-b\tau) - \theta_2 b\beta\sqrt{x^3}, \tag{3.52}$$

where $\theta_2 = \sqrt{\frac{3+\sqrt{5}}{6}} + \sqrt{\left(\frac{3-\sqrt{5}}{6}\right)^3}$.

Similarly, when using three routes for the remaining x casualties starting from time τ, we have

$$\begin{aligned} V_3(x,\tau) &= \max_{0\leq y_3\leq x}\{y_3 S(\tau+\beta\sqrt{y_3}) + V_2(x-y_3,\tau+\beta\sqrt{y_3})\} \\ &= \max_{0\leq y_3\leq x}\left\{x(a-b\tau) - bx\beta\sqrt{y_3} - \theta_2 b\beta\sqrt{(x-y_3)^3}\right\}. \end{aligned} \tag{3.53}$$

Letting $z_3(y_3) = x(a-b\tau) - bx\beta\sqrt{y_3} - \theta_2 b\beta\sqrt{(x-y_3)^3}$ and setting $\frac{\partial z_3(y_3)}{\partial y_3} = 0$, we can show that $y_3^* = \left(\frac{1}{2} + \frac{\sqrt{9\theta_2^2-4}}{6\theta_2}\right)x$ provides the global maximum solution for $z_3(y_3)$. Using the same approach for $V_4(x,\tau)$, we can obtain the optimal solution at $y_4^* = \left(\frac{1}{2} + \frac{\sqrt{9\theta_3^2-4}}{6\theta_3}\right)x$, where $\theta_3 = \sqrt{\frac{1}{2} + \frac{\sqrt{9\theta_2^2-4}}{6\theta_2}} + \theta_2\sqrt{\left(\frac{1}{2} - \frac{\sqrt{9\theta_2^2-4}}{6\theta_2}\right)^3}$.

More generally, for an optimal solution for x remaining casualties with k routes, we can show that

$$y_k^* = \left(\frac{1}{2} + \frac{\sqrt{9\theta_{k-1}^2 - 4}}{6\theta_{k-1}}\right)x, \tag{3.54}$$

which leads to

$$V_k(x,\tau) = x(a - b\tau) - \theta_k b\beta\sqrt{x^3}, \tag{3.55}$$

for $k = 2, 3, ...$, where $\theta_k = \sqrt{\frac{1}{2} + \frac{\sqrt{9\theta_{k-1}^2-4}}{6\theta_{k-1}}} + \theta_{k-1}\sqrt{\left(\frac{1}{2} - \frac{\sqrt{9\theta_{k-1}^2-4}}{6\theta_{k-1}}\right)^3}$ and $\theta_1 = 1$.

Let us use q_k to denote the coefficient of x on the right-hand side of Equation (3.54), i.e., $q_k = \frac{1}{2} + \frac{\sqrt{9\theta_{k-1}^2-4}}{6\theta_{k-1}}$. The value of q_k indicates the percentage out of the remaining x casualties to include on the first route when there are k routes to run, which is equivalent to r_1 that is defined in Section 3.4.2. Table 3.2 shows the value of q_k when changing the value of k, which exhibits the same pattern of convergence as that demonstrated in Table 3.1.

Table 3.2: The value of q_k as k changes.

k	q_k	k	q_k
2	0.872678	7	0.866025
3	0.866352	8	0.866025
4	0.866041	9	0.866025
5	0.866026	10	0.866025
6	0.866025	11	0.866025

Therefore, for uncapacitated routes with a linear survival probability function, the DP formulation provides an optimal solution for the relaxed problem that is consistent with the KKT solution characterized in Section 3.4.1, in which all casualties are assigned to two routes and the first route always includes approximately 86.6% of the remaining casualties. To obtain a solution with integer decision variables, a straightforward heuristic is to round y_k to the nearest integer, and then consider the remaining number as the input for the next route until there are zero casualties remaining.

3.5.2 The DPB heuristic algorithm

If the exponent $c \neq 1$ in the survival probability function $S(t) = a - bt^c$, it becomes more complex to obtain closed-form solutions with larger values of k due to the intractability of the terms associated with the power c. For any $x \leq C$ with $k = 2$, however, we can obtain the optimal objective value by solving

$$V_2(x, \tau) = \max_{0 \leq y_2 \leq x} \left\{ y_2 \left(a - b(\tau + \beta\sqrt{y_2})^c\right) + (x - y_2)\left(a - b(\tau + \beta\sqrt{y_2} + \beta\sqrt{x - y_2})^c\right) \right\} \quad (3.56)$$

for y_2 within $[0, x]$. This can be solved relatively fast because the value of x is limited by the vehicle capacity C, and the possible values of C are in a small range of integers. For $x \leq C$, with the proposed nonlinear survival probability function, we can show that running two routes always leads to a better objective value than performing a single route. This is formalized in Proposition 3.5.1.

Proposition 3.5.1. *$V_2(x, \tau) \geq V_1(x, \tau)$ holds for any $x \leq C$.*

According to Proposition 3.5.1, for any number of remaining casualties that is greater than C, we can first run as many full routes as possible, and then assign the remaining casualties (which is capacity feasible) to two additional routes. Given the value of N, we propose the following *dynamic programming based* (DPB) heuristic algorithm in Algorithm 3, assuming we perform the first route starting at time τ_0.

Given N and C, we first compute the maximum number of full routes, which equals $K = \lfloor N/C \rfloor$. If $K = 0$, this implies $N < C$ and we can use two routes for all casualties N and solve for the optimal objective by searching the integer values for the number of the casualties on the first route in $[1, N]$.

If $K > 0$, one possible scenario is that the remainder satisfies $0 < N - KC < C$, implying the optimal strategy is to first run K full routes, and then use two additional routes for the remaining $N - KC$ casualties with a starting time of $\tau = \tau_0 + K\beta\sqrt{C}$. The other possible scenario is

Algorithm 3: The DPB algorithm

Initialization: Given N casualties and the vehicle capacity C, calculate the number of full routes $K = \lfloor N/C \rfloor$;
if $K = 0$ **then**
    $\tau = \tau_0$;
    Searching: Obtain $V_2(x, \tau)$ by searching integer values for y_2^* in $[1, N]$;
    Completion: The optimal solution is $[y_2^*, N - y_2^*]$;
else
    if $N - CK > 0$ **then**
        Run K full routes, including C casualties on each route;
        $\tau = \tau_0 + K\beta\sqrt{C}$ and $x = N - KC$;
    else
        Run $K - 1$ full routes, including C casualties on each route;
        $\tau = \tau_0 + (K-1)\beta\sqrt{C}$ and $x = N - (K-1)C$;
    end
    Searching: Obtain $V_2(x, \tau)$ by searching searching integer values for y_2^* in $[1, x]$;
    Completion: The optimal solution is $[y_k, y_{k-1} \ldots, y_3, y_2^*, x - y_2^*]$ where $y_k = y_{k-1} \ldots = y_3 = C$.
end

that N is a multiple of C and $N - KC = 0$, in which case we first run $K - 1$ full routes, and then perform two additional routes for the remaining $N - (K-1)C$ casualties starting from time $\tau = \tau_0 + (K-1)\beta\sqrt{C}$.

3.6 Numerical experiments

In this section, we solve the original problem $\mathbb{P}$ using the BARON solver for both linear and nonlinear survival probability functions and compare the solutions with the results obtained by the GR algorithm (for linear survival probability) and by the DPB algorithm (for nonlinear survival probability), respectively. The maximum time limit in BARON is set to 18000 seconds (5 hours). All tested problems are solved on a Windows 11 PC with an AMD Ryzen 7 5800X 8-Core processor and 16 GB memory. At the end of this section, we investigate the solution sensitivity to the vehicle capacity C and the survival reduction parameter c, respectively.

3.6.1 Performance of the GR algorithm

We use a linear survival probability function in the form of $S(t) = 1 - 0.005t$, and test each integer value of N in the range of $[10, 200]$ with a vehicle capacity of 22. All problems that are solved within the time limit have an optimality gap that is less than or equal to 0.001% of the incumbent optimal objective at termination. Note that the time limit parameter set in BARON does not include the pre-processing time, therefore, problems that reach the time limit may have a total runtime of more than 18000 seconds. We implement the GR algorithm on each problem and compare the results with the BARON solutions. Table 3.3 includes a subset of the results for $C = 22$. The objective values and the runtimes are rounded to two decimal places. The table only includes nonzero decision variables in the third column, while all other variables not displayed are equal to zero.

Table 3.3: BARON solutions and the GR algorithm results with $C = 22$.

			$C = 22$			
N	# of routes (k)	$n_i, i = 1, \ldots, k$	Obj.Val.	BARON gap (%)	BARON runtime (s)	GR runtime (s)
10	2	[9, 1]	9.85	<0.001	0.31	<0.01
20	2	[17, 3]	19.56		0.39	
40	3	[22, 16, 2]	38.69		1.79	
80	5	[22, 22, 22, 12, 2]	75.66		11.97	
100	6	[22, 22, 22, 22, 10, 2]	93.51		27.14	
150	8	[22, 22, 22, 22, 22, 22, 16, 2]	136.26		102.21	
200	10	[22, 22, 22, 22, 22, 22, 22, 22, 21, 3]	176.35	4.99	19820.43	

According to our experiment results, the GR algorithm provides exactly the same solutions as those obtained using BARON for all problems, while significantly improving the runtime, especially for large values of N. When $N = 200$, BARON reaches the maximum time limit and provides a solution with a 4.99% optimality gap at termination, while using the GR algorithm obtains the same solution within a negligible time.

Observe that when $N = 10$, although a single route can accommodate all casualties, the optimal solution suggests using two routes by allocating 9 casualties to the first route and leaving the remaining casualty to the second route. This is consistent with our discussion on the 87% rule.

3.6.2 Performance of the DPB algorithm

To test the performance of the DPB algorithm with a nonlinear survival probability function, we use a function in the form of $S(t) = 1 - 0.01\sqrt{t}$ and $\beta = 1$, assuming the first route is performed from time 0. We test each integer value of N in the range of $[10, 100]$ with different values of C in [16, 18, 20, 22], using both BARON and the DPB heuristic algorithm. We solve 91 instances for each capacity level and 364 instances in total. Two subsets of solutions for $C = 20$ and $C = 22$ are shown in Table 3.4, in which the problems are starred if the DPB solution is different from the BARON solution. The objective values are rounded to four decimal places and the runtime is rounded to two decimal places.

Table 3.4: BARON solutions and the DPB algorithm results with $C = 20$ and $C = 22$.

				C=20				
	N	DPB solution	DPB obj.val.	DPB runtime (s)	BARON solution	BARON obj.val.	BARON runtime (s)	BARON gap (%)
	15	[13, 2]	14.7083		[13, 2]	14.7083	0.35	
*	20	[17, 3]	19.5822		[16, 3, 1]	19.5822	0.69	
	25	[20, 4, 1]	24.4480		[20, 4, 1]	24.4480	0.44	
	30	[20, 9, 1]	29.3019		[20, 9, 1]	29.3019	1.54	
	35	[20, 13, 2]	34.1460	<0.001	[20, 13, 2]	34.1460	2.68	<0.001
	40	[20, 17, 3]	38.9822		[20, 17, 3]	38.9822	2.12	
	45	[20, 20, 4, 1]	43.8120		[20, 20, 4, 1]	43.8120	4.00	
	50	[20, 20, 9, 1]	48.6319		[20, 20, 9, 1]	48.6319	7.67	
	55	[20, 20, 13, 2]	53.4436		[20, 20, 13, 2]	53.4436	8.76	
	60	[20, 20, 17, 3]	58.2490		[20, 20, 17, 3]	58.2490	4.98	
				C=22				
	N	DPB solution	DPB obj.val.	DPB runtime (s)	BARON solution	BARON obj.val.	BARON runtime (s)	BARON gap (%)
	15	[13, 2]	14.7083		[13, 2]	14.7083	0.36	
*	20	[17, 3]	19.5822		[16, 3, 1]	19.5822	0.69	
*	25	[22, 3]	24.4475		[21, 3, 1]	24.4480	1.02	
	30	[22, 7, 1]	29.3051		[22, 7, 1]	29.3051	2.00	
	35	[22, 11, 2]	34.1509	<0.001	[22, 11, 2]	34.1509	2.14	<0.001
	40	[22, 15, 3]	38.9883		[22, 15, 3]	38.9883	2.33	
*	45	[22, 22, 1]	43.8175		[22, 19, 3, 1]	43.8191	5.18	
	50	[22, 22, 5, 1]	48.6438		[22, 22, 5, 1]	48.6438	10.36	
	55	[22, 22, 9, 2]	53.4588		[22, 22, 9, 2]	53.4588	11.82	
	60	[22, 22, 14, 2]	58.2663		[22, 22, 14, 2]	58.2663	16.72	

Our experimental results show that 49 out of 364 test instances have a DPB solution that is different from the BARON solution. However, the difference between the two resulting objective values is negligible. For example, in Table 3.4, when $N = 20$ with either capacity level, the DPB algorithm leads to a solution that suggests using two routes by including 17 casualties on the first route and assigning the remaining 3 to the second route, with an objective value of 19.5822. The BARON solution with the same values of N and C, however, provides a three-route solution that suggests including 16, 3, and 1 casualties on each route, respectively, while the resulting objective also approximately equals 19.5822. When $C = 22$ and $N = 25$, the DPB algorithm provides a two-route solution with an allocation of 22:3, whereas the BARON solution indicates using three routes by including 20, 4, and 1 casualties on each route, respectively, while improving the objective value by only 0.05%. We have a similar conclusion for the results with $N = 45$ and $C = 22$, in which the BARON solution suggests using four routes instead of the three-route solution obtained using the DPB algorithm, whereas the BARON objective is only 0.16% less than the DPB objective.

3.6.3 Sensitivity analysis

This section investigates how the optimal solution changes as the vehicle capacity C and the survivability reduction parameter c vary. We first analyze the impact of vehicle capacity by changing the value of C for problems with a linear survival probability function when fixing all other parameters. We then investigate the impacts of the survivability reduction rate c for problems with a nonlinear survival probability function when other parameters are fixed.

3.6.3.1 Impact of vehicle capacity

Using the linear survival probability function $S(t) = 1 - 0.005t$, we compare the BARON solutions for $N \in [10, 100]$ by changing the vehicle capacity in $[16, 18, 20, 22]$. Table 3.5 summarizes a subset of results for each capacity level.

Observe that when $N = 10$, the solutions for all capacity levels are equal, in which $n_1 = 9, n_2 = 1$ and $n_i = 0$ for $i = 3, \ldots, 10$, which is independent of the test capacity levels. This is because $N = 10$ is capacity feasible for all test capacity levels and thus the optimal strategy is to use two routes for all casualties following the 87% rule, resulting in a solution that includes 9 casualties on the first route and leaving the remaining one for the second. As the value of N increases, a smaller vehicle tends to perform more routes than a larger vehicle. For example, when

Table 3.5: Solutions and objective values at different capacity levels.

$C = 16$			
N	# of routes (k)	n_i for $i = 1, \ldots, k$	Obj.val.
10	2	[9, 1]	9.85
16	2	[14, 2]	15.69
20	3	[16, 3, 1]	19.56
32	3	[16, 14, 2]	31.05
40	4	[16, 16, 7, 1]	38.61
80	6	[16, 16, 16, 16, 14, 2]	75.21
100	8	[16, 16, 16, 16, 16, 16, 3, 1]	92.76
$C = 18$			
N	# of routes (k)	n_i for $i = 1, \ldots, k$	Obj.val.
10	2	[9, 1]	9.85
18	2	[16, 2]	17.63
20	2	[17, 3]	19.56
36	3	[18, 16, 2]	34.86
40	4	[18, 18, 3, 1]	38.65
80	6	[18, 18, 18, 18, 7, 1]	75.39
100	7	[18, 18, 18, 18, 18, 9, 1]	93.06
$C = 20$			
N	# of routes (k)	n_i for $i = 1, \ldots, k$	Obj.val.
10	2	[9, 1]	9.85
20	2	[17, 3]	19.56
40	3	[20, 17, 3]	38.67
50	4	[20, 20, 9, 1]	48.06
60	4	[20, 20, 17, 3]	57.33
80	5	[20, 20, 20, 17, 3]	75.54
100	6	[20, 20, 20, 20, 17, 3]	93.30
$C = 22$			
N	# of routes (k)	n_i for $i = 1, \ldots, k$	Obj.val.
10	2	[9, 1]	9.85
20	2	[17, 3]	19.56
22	2	[19, 3]	21.49
40	3	[22, 16, 2]	38.69
44	3	[22, 19, 3]	42.46
80	5	[22, 22, 22, 12, 2]	75.66
100	6	[22, 22, 22, 22, 10, 2]	93.51

$N = 20$, a vehicle with a capacity of 16 has to run 3 routes with an allocation of 16:3:1, while vehicles with larger capacity levels only require two routes by including one more casualty on the first route. When $N = 100$, only 6 routes are required if the vehicle capacity is 20 or 22, while an additional route is needed by a vehicle with a capacity of 18 and two more routes are required for a vehicle with a capacity of 16. Note that when the value of N is a multiple of C, the optimal solution always suggests running $\lfloor N/C \rfloor - 1$ full routes and then allocating the remaining casualties using two additional routes, following the 87% rule.

Given a fixed value of N, a larger vehicle leads to a higher objective value than that of using a smaller vehicle, even when they perform the same number of routes. For example, when $N = 100$, a capacity of 20 leads to a six-route solution with an expected number of 93.3 survivors, while increasing the capacity to 22 leads to an expected number of 93.51 survivors when also running 6 routes. For ease of comparison, we normalize all objective values by dividing by the objective value resulting at $C = 16$. Figure 3.1 illustrates the normalized objective values in N at different vehicle capacity levels.

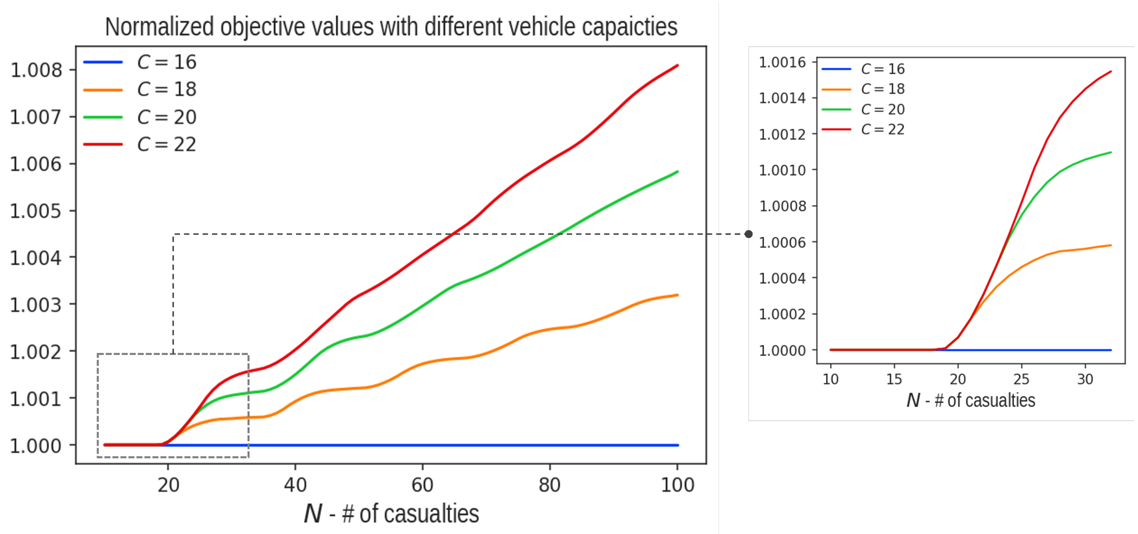


Figure 3.1: Normalized objective values as vehicle capacity varies.

As the figure illustrates, when $N \leq 18$, the four curves overlap, implying changing the vehicle capacity does not affect the optimal solution. This is because for any N value in this range, all casualties can be allocated to two routes following the 87% rule, and each route is capacity feasible for either vehicle capacity C. When $18 < N \leq 24$, at least one curve does not overlap the others because, for N in this range, the maximum number of casualties that can be transported on each route is restricted by the vehicle capacity, which leads to different solutions for some of the N values when using different values of C. For $N > 24$, the optimal solutions no longer overlap among the four capacity levels. In this interval, the optimal solution requires at least one route to be filled to capacity, and therefore different vehicle capacities will lead to different numbers of routes and thus distinct solutions. Notably, for a given N, a larger value of C leads to a higher objective value, that is, using an ambulance bus with a larger capacity increases the overall expected number of survivors. This improvement becomes more pronounced for larger values of N.

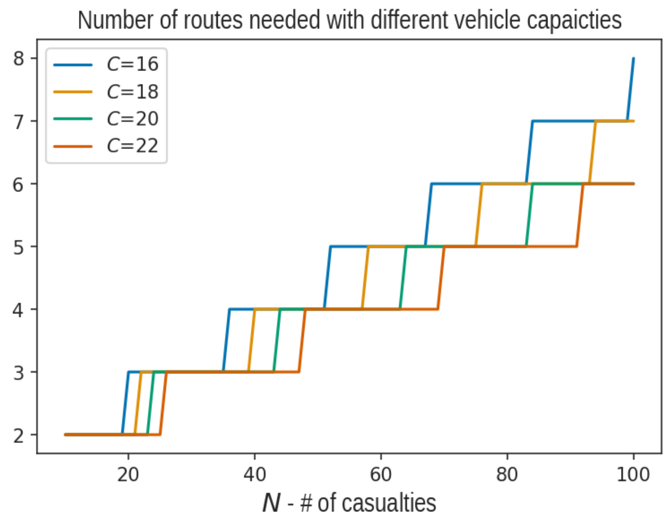


Figure 3.2: The number of routes needed as vehicle capacity varies.

In addition, the average number of routes required tends to be smaller when using a larger ambulance bus, which is demonstrated in Figure 3.2. The number of routes for each capacity level shows a step-wise growth trend. The more ‘steps’ in a curve, the larger the number of routes needed for certain intervals of N. When $C = 16$, the need to add another route occurs when the total number of casualties increases by 16. However, at $C = 18$, adding 20 casualties requires one additional route to maximize the total expected number of survivors. The frequency of adding routes is lower for $C = 20$ and $C = 22$, which occurs when the total number of casualties increases by 23-25.

Figure 3.3 compares the total completion time when all routes are finished. It is obvious that as the number of casualties increases, using a larger vehicle can transport all casualties to the medical center at an earlier time. It is also noticeable that each curve demonstrates an upward trend with some small ‘jumps’ as N increases. These are caused by the increase in the number of routes.

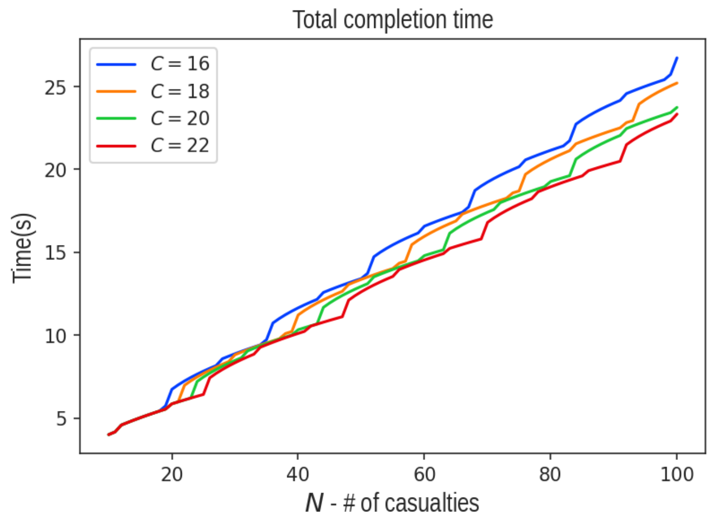


Figure 3.3: Total completion time for all routes with different vehicle capacity levels.

Every time a route is added, the total completion time experiences a sharp increase and then keeps increasing at a relatively stable rate until the next sharp rise.

All results imply that when the number of casualties is large, the best strategy is to use as large an ambulance bus as possible and perform fewer routes to maximize the total expected number of survivors.

3.6.3.2 Impact of survivability reduction rate

For a nonlinear survival probability function of the form $S(t) = a - bt^c$, the value of c primarily determines the reduction rate of a casualty's survival probability as time passes. To see the impact of varying c, we solve problem $\mathbb{P}$ using BARON with $a = 1$ and $b = 0.01$ for $N \in [10, 90]$ while changing the value of c among $[0.2, 0.5, 1, 1.5]$, using a vehicle with a capacity of 22. Table 3.6 includes the solutions for a subset of instances at each value of c.

As the table indicates, for a given N, the increase in the survivability reduction rate leads to a smaller expected number of survivors. Observe that as the value of c increases, a solution tends to include more casualties on the first route. This pattern is more obvious for smaller values of N. For example, when $N = 20$, solutions associated with $c = 0.2$ and $c = 0.5$ suggest including 16 casualties on the first route while splitting the remaining casualties into two additional routes with

Table 3.6: Solutions and objective values with different survivability reduction rates.

	$c = 0.2$		$c = 0.5$		$c = 1$		$c = 1.5$	
N	Solution	Obj.val.	Solution	Obj.val.	Solution	Obj.val.	Solution	Obj.val.
10	[8, 2]	9.87	[8, 2]	9.82	[9, 1]	9.69	[9, 1]	9.45
11	[9, 2]	10.86	[9, 2]	10.80	[10, 1]	10.64	[10, 1]	10.35
12	[10, 2]	11.85	[10, 2]	11.78	[10, 2]	11.59	[11, 1]	11.25
13	[11, 2]	12.83	[11, 2]	12.76	[11, 2]	12.54	[12, 1]	12.13
14	[11, 2, 1]	13.82	[12, 2]	13.73	[12, 2]	13.49	[13, 1]	13.01
15	[12, 2, 1]	14.80	[13, 2]	14.71	[13, 2]	14.43	[13, 2]	13.89
16	[13, 2, 1]	15.79	[13, 3]	15.68	[14, 2]	15.37	[14, 2]	14.75
17	[13, 3, 1]	16.78	[14, 3]	16.66	[15, 2]	16.31	[15, 2]	15.61
18	[14, 3, 1]	17.76	[15, 3]	17.63	[16, 2]	17.25	[16, 2]	16.47
19	[15, 3, 1]	18.75	[16, 3]	18.61	[17, 2]	18.19	[17, 2]	17.32
20	[16, 3, 1]	19.73	[16, 3, 1]	19.58	[17, 3]	19.12	[18, 2]	18.16
21	[17, 3, 1]	20.72	[17, 3, 1]	20.56	[18, 3]	20.06	[19, 2]	18.99
22	[17, 4, 1]	21.70	[18, 3, 1]	21.53	[19, 3]	20.99	[20, 2]	19.82
23	[18, 4, 1]	22.69	[19, 3, 1]	22.50	[20, 3]	21.92	[21, 2]	20.65
24	[19, 4, 1]	23.67	[20, 3, 1]	23.48	[21, 3]	22.85	[21, 3]	21.46
25	[20, 4, 1]	24.66	[21, 3, 1]	24.45	[22, 3]	23.78	[22, 3]	22.28
26	[21, 4, 1]	25.64	[21, 4, 1]	25.42	[22, 3, 1]	24.70	[22, 3, 1]	23.07
27	[22, 4, 1]	26.63	[22, 4, 1]	26.39	[22, 4, 1]	25.62	[22, 4, 1]	23.86
28	[22, 5, 1]	27.61	[22, 5, 1]	27.36	[22, 5, 1]	26.54	[22, 5, 1]	24.63
29	[22, 6, 1]	28.60	[22, 6, 1]	28.33	[22, 6, 1]	27.46	[22, 6, 1]	25.39
30	[22, 7, 1]	29.58	[22, 7, 1]	29.31	[22, 7, 1]	28.37	[22, 7, 1]	26.13

a ratio of 3:1. However, when the value of c increases to 1, the solution only contains two routes and assigns 17 casualties to the first route. The first route will include one more casualty if the value of c increases to 1.5. This is because when the reduction rate is large, the conditions of the casualties worsen very fast, and a solution tends to include as many as possible casualties on the first route to shorten the completion time (i.e., the time when the casualties arrive at the medical center for treatment).

For a larger value of N, however, all four reduction rates lead to the same solution. For example, when $N = 30$, the solution suggests filling the first route to capacity and splitting the remaining number into two routes with a ratio of 7:1. When N becomes larger, the best strategy is to first run as many as possible full routes, which implies the remaining number will be capacity feasible and thus it is possible to implement the 87% rule. In this case, the solution will not be affected by the value of c.

Figure 3.4 illustrates the objective function values when changing N from 10 to 90 at different

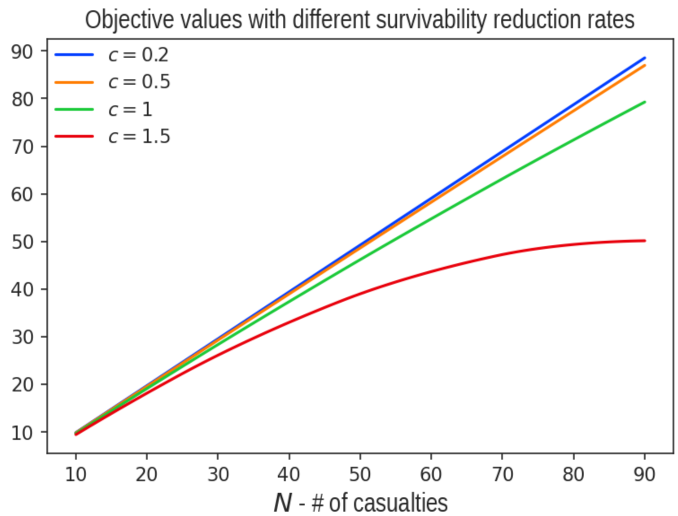


Figure 3.4: Objective values with different survivability reduction rates as N increases.

values of c. Notably, when fixing the value of N, a larger survivability reduction rate leads to a lower objective value. For any given N, the blue curve (when $c = 0.2$) leads to the largest total expected number of survivors, while the red curve (associated with $c = 1.5$) leads to the lowest objective value. When N is relatively small (in the range of $[10, 30]$, for example), the differences between the four curves are relatively small; when $N > 30$, the differences between the curves become more significant, in which the red curve ($c = 1.5$) opens up a substantial gap below the other three curves, while the green curve slightly diverges from the blue and the orange ones. The curves associated with $c = 0.2$ (blue) and $c = 0.5$ (orange) are very close. This is because when the value of c is large, the conditions of the casualties worsen faster as time passes. As the value of N rises, it requires more time to transport all casualties so that casualties who have waited for a longer time will have lower survival probabilities when arriving at the medical center, which leads to a reduced value of the expected number of survivors.

Figure 3.5 illustrates the number of routes needed for all routes as the value of c varies. Observe that the green curve associated with $c = 1$ is overlapped by the red curve corresponding to $c = 1.5$, implying the number of routes for each instance with these reduction rates is equal. For an interval of smaller values of N, a smaller value of c tends to require more routes. For example, when N

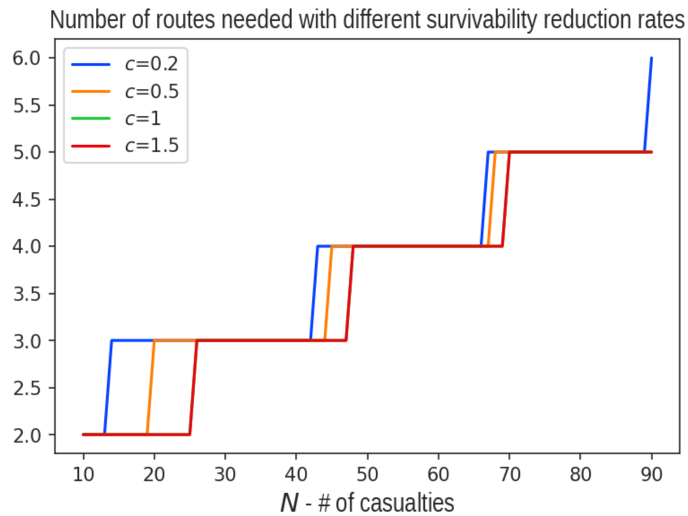


Figure 3.5: The number of routes needed with different survivability reduction rates.

is between 10 to 20, most instances with $c = 0.2$ need 3 routes, while the solutions for instances with larger values of c suggest 2 routes. This is caused by the increase in the number of casualties included on the first route as the value of c increases, as we have explained when discussing the observations in Table 3.6. Conversely, when N is between 70 and 80, most solutions, regardless of the value of c, suggest running 5 routes. This is because, for a large value of N, the solutions tend to have the same number of full routes plus two additional routes, which we discussed in interpreting the results in Table 3.6.

Figure 3.6 illustrates the total completion time for all routes when the value of c varies. We observe that the total completion time is considerably impacted by the total number of routes performed. The four curves come closer together at an N value that leads to the same number of routes in each solution. The curves completely overlap when solutions for all four values of c are the same. For example, when $N = 30$, the solution associated with each value of c includes one full route and two additional routes with a ratio of 7:1, and thus the resulting completion times for all routes are equal in each case.

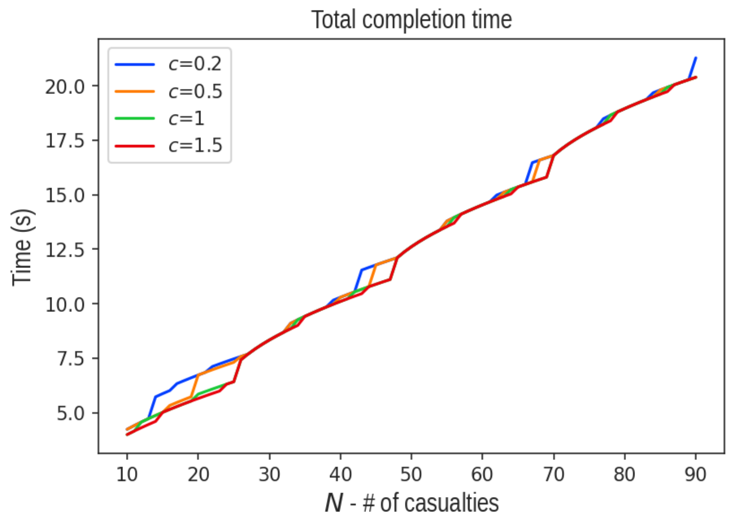


Figure 3.6: The total completion time for all routes with different survivability reduction rates.

3.7 Conclusions and future work

In this chapter, we investigate the routing problem in the aftermath of a disaster using a single ambulance bus, in which the survival probabilities of the casualties decline as time passes. We develop models that maximize the total expected number of survivors while determining the number of routes and the route sizes, accounting for both linear and nonlinear survival probability functions. For models with linear survival probability functions, we propose a solution approach that provides closed-form solutions. We also develop two heuristic algorithms for solving the models with linear and nonlinear survival probability functions. Our numerical results show that the heuristics provide effective solutions that are comparable to solutions provided by a commercial solver, while significantly improving the computing time. Our sensitivity analysis results imply that the vehicle capacity and the survival reduction parameters have considerable impacts on the optimal solutions. The approaches developed in this work can enhance the decision-making process, ultimately leading to more effective emergency responses.

We acknowledge that the proposed models are quite stylized and leave many other important issues outside the scope of this study. For example, transportation priorities based on the conditions of the casualties could be a critical factor to consider. Our discussion is limited to two specific forms of survival probability functions, whereas more practical survival functions should be investigated and incorporated into the models, such as log-logistic and log-normal functions used in past literature on disaster response. In real-life applications involving ambulance buses, it is possible to provide treatment to the casualties within the vehicle during transportation. Therefore, incorporating the in-transit treatment into the survival probability functions can be a practical extension.

4. SUMMARY AND CONCLUSIONS

In this dissertation, we investigate and develop optimization models for three different last-mile service systems in the domains of inventory-delivery management, transportation and logistics, and emergency medical services (EMS), respectively.

In the inventory-delivery system, we study a two-echelon, single-product fulfillment system where a regional fulfillment center (FC) replenishes multiple identical and independent local distribution center (LDC) orders within an internal committed resupply leadtime, and LDCs serve end customers within a committed demand fulfillment time, or committed delivery time. Expected system-wide demand depends on the product's price, the committed delivery time, and the number of LDCs in the system. Our proposed model determines the values of product price, committed resupply time, and committed delivery time that maximize expected system-wide profit per period, while accounting for product holding costs and fixed facility costs. We characterize key properties of optimal solutions that permit an efficient solution for a fixed number of identical LDCs, and consider the impacts of several proposed demand growth models as the number of LDCs increases. The results of a computational study provide interesting managerial insights on how operational constraints and the scale of the distribution system influence strategic stock placement and distribution system structure.

Within the domain of transportation and logistics, we study a fleet composition problem with stochastic demand, which must be fulfilled via a combination of internal truckload (TL) capacity and external less-than-truckload (LTL) shipments. Internal capacity costs include a fixed ownership cost per truck and a fixed dispatch cost per truck, while LTL costs are incurred per unit shipped. We characterize the expected total cost per period as a function of internal fleet size for both homogeneous and heterogeneous fleets. We propose a solution approach that provides optimal shipping policies with analytical solutions for homogeneous and certain types of heterogeneous fleets. For a fleet with a wider range of truck types, we create a two-stage stochastic programming model that minimizes the total expected cost when determining the number of trucks of each type to acquire. We further develop a decomposition-based algorithm that can effectively solve the proposed model, which provides solutions that are as good as the solutions obtained using a commercial solver, while significantly improving the computation time. Our numerical results reveal that the optimal fleet composition strategy is significantly affected by the LTL freight rate, the number of truck types, and the properties of the demand.

In addition to researching the movement of goods, in the domain of EMS, we investigate a multi-trip single-vehicle routing problem in the aftermath of a disaster where geographically-dispersed casualties require transportation to a medical center using a single ambulance bus. We consider time-dependent survival probabilities of the casualties, i.e., as patients wait longer, their survival probabilities decrease. We develop models that maximize the total expected number of survivors based on tour length approximations. For problems with a linear survival probability function, we characterize analytical results using KKT conditions and propose a heuristic approach that effectively solves the model within a negligible computing time. We also present a dynamic programming model that accommodates general forms of survival probability functions, and develop a heuristic algorithm that provides high-quality solutions. Our sensitivity analyses show that both the vehicle capacity level and the reduction rate in the survival probability function have considerable impacts on the optimal routing decisions.

REFERENCES

[1] K. F. Simpson, “In-process inventories,” *Operations Research*, vol. 6, no. 6, pp. 863–873, 1958.

[2] K. Inderfurth, “Safety stock optimization in multi-stage inventory systems,” *International Journal of Production Economics*, vol. 24, no. 1, pp. 103–113, 1991.

[3] S. C. Graves and S. P. Willems, “Optimizing strategic safety stock placement in supply chains,” *Manufacturing & Service Operations Management*, vol. 2, no. 1, pp. 68–83, 2000.

[4] S. C. Graves and S. P. Willems, “Strategic inventory placement in supply chains: Nonstationary demand,” *Manufacturing & Service Operations Management*, vol. 10, no. 2, pp. 278–287, 2008.

[5] J. J. Neale and S. P. Willems, “Managing inventory in supply chains with nonstationary demand,” *Interfaces*, vol. 39, no. 5, pp. 388–399, 2009.

[6] K. Funaki, “Strategic safety stock placement in supply chain design with due-date based demand,” *International Journal of Production Economics*, vol. 135, no. 1, pp. 4–13, 2012.

[7] K. Inderfurth and S. Minner, “Safety stocks in multi-stage inventory systems under different service measures,” *European Journal of Operational Research*, vol. 106, no. 1, pp. 57–73, 1998.

[8] C. Sitompul, E. Aghezzaf, W. Dullaert, and H. Van Landeghem, “Safety stock placement problem in capacitated supply chains,” *International Journal of Production Research*, vol. 46, no. 17, pp. 4709–4727, 2008.

[9] S. C. Graves and T. Schoenmeyr, “Strategic safety-stock placement in supply chains with capacity constraints,” *Manufacturing & Service Operations Management*, vol. 18, no. 3, pp. 445–460, 2016.

[10] H. Chen and P. Li, "Optimization of (*R,Q*) policies for serial inventory systems using the guaranteed service approach," *Computers & Industrial Engineering*, vol. 80, pp. 261–273, 2015.

[11] A. S. Eruguz, E. Sahin, Z. Jemai, and Y. Dallery, "A comprehensive survey of guaranteed-service models for multi-echelon inventory optimization," *International Journal of Production Economics*, vol. 172, pp. 110–125, 2016.

[12] R. H. Ballou, "Unresolved issues in supply chain network design," *Information Systems Frontiers*, vol. 3, no. 4, pp. 417–426, 2001.

[13] Z.-J. M. Shen, C. Coullard, and M. S. Daskin, "A joint location-inventory model," *Transportation Science*, vol. 37, no. 1, pp. 40–55, 2003.

[14] S. C. Graves and S. P. Willems, "Optimizing the supply chain configuration for new products," *Management Science*, vol. 51, no. 8, pp. 1165–1180, 2005.

[15] M. Schuster Puga, S. Minner, and J.-S. Tancrez, "Two-stage supply chain design with safety stock placement decisions," *International Journal of Production Economics*, vol. 209, pp. 183–193, 2019.

[16] F. You and I. E. Grossmann, "Balancing responsiveness and economics in process supply chain design with multi-echelon stochastic inventory," *AIChE Journal*, vol. 57, no. 1, pp. 178–192, 2011.

[17] S. Humair and S. P. Willems, "Optimizing strategic safety stock placement in supply chains with clusters of commonality," *Operations Research*, vol. 54, no. 4, pp. 725–742, 2006.

[18] J. Grahl, S. Minner, and D. Dittmar, "Meta-heuristics for placing strategic safety stock in multi-echelon inventory with differentiated service times," *Annals of Operations Research*, vol. 242, no. 2, pp. 489–504, 2016.

[19] H. Osman and K. Demirli, "Integrated safety stock optimization for multiple sourced stockpoints facing variable demand and lead time," *International Journal of Production Economics*, vol. 135, no. 1, pp. 299–307, 2012.

[20] J. Shu and I. A. Karimi, "Efficient heuristics for inventory placement in acyclic networks," *Computers & Operations Research*, vol. 36, no. 11, pp. 2899–2904, 2009.

[21] N. G. Hua and S. P. Willems, "Analytical insights into two-stage serial line supply chain safety stock," *International Journal of Production Economics*, vol. 181, pp. 107–112, 2016.

[22] D. Barnes-Schuster, Y. Bassok, and R. Anupindi, "Optimizing delivery lead time/inventory placement in a two-stage production/distribution system," *European Journal of Operational Research*, vol. 174, no. 3, pp. 1664–1684, 2006.

[23] J. Huang, M. Leng, and M. Parlar, "Demand functions in decision modeling: A comprehensive survey and research directions," *Decision Sciences*, vol. 44, no. 3, pp. 557–609, 2013.

[24] Y. Chen, S. Ray, and Y. Song, "Optimal pricing and inventory control policy in periodic-review systems with fixed ordering cost and lost sales," *Naval Research Logistics*, vol. 53, no. 2, pp. 117–136, 2006.

[25] H. Tuy, "Global minimization of a difference of two convex functions," *Mathematical Programming Study*, vol. 30, pp. 150–182, 1987.

[26] D. Kirby, "Is your fleet the right size?," *Journal of the Operational Research Society*, vol. 10, no. 4, pp. 252–252, 1959.

[27] J. Gould, "The size and composition of a road transport fleet," *Journal of the Operational Research Society*, vol. 20, no. 1, pp. 81–92, 1969.

[28] T. Etezadi and J. E. Beasley, "Vehicle fleet composition," *Journal of the Operational Research Society*, vol. 34, no. 1, pp. 87–91, 1983.

[29] P. Wu, J. C. Hartman, and G. R. Wilson, "An integrated model and solution approach for fleet sizing with heterogeneous assets," *Transportation Science*, vol. 39, pp. 87–103, Feb. 2005.

[30] R. Loxton, Q. Lin, and K. L. Teo, "A stochastic fleet composition problem," *Computers & Operations Research*, vol. 39, no. 12, pp. 3177–3184, 2012.

[31] D. Konur and J. Geunes, "Integrated districting, fleet composition, and inventory planning for a multi-retailer distribution system," *Annals of Operations Research*, vol. 273, pp. 527–559, 2019.

[32] K. S. Shehadeh, H. Wang, and P. Zhang, "Fleet sizing and allocation for on-demand last-mile transportation systems," *Transportation Research Part C: Emerging Technologies*, vol. 132, p. 103387, 2021.

[33] B. Golden, A. Assad, L. Levy, and F. Gheysens, "The fleet size and mix vehicle routing problem," *Computers & Operations Research*, vol. 11, no. 1, pp. 49–66, 1984.

[34] M. Dell'Amico, M. Monaci, C. Pagani, and D. Vigo, "Heuristic approaches for the fleet size and mix vehicle routing problem with time windows," *Transportation Science*, vol. 41, no. 4, pp. 516–526, 2007.

[35] O. Jabali, M. Gendreau, and G. Laporte, "A continuous approximation model for the fleet composition problem," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1591–1606, 2012.

[36] G. Hiermann, J. Puchinger, S. Ropke, and R. F. Hartl, "The electric fleet size and mix vehicle routing problem with time windows and recharging stations," *European Journal of Operational Research*, vol. 252, no. 3, pp. 995–1018, 2016.

[37] M. Alinaghian, M. Jamshidian, and E. B. Tirkolaee, "The time-dependent multi-depot fleet size and mix green vehicle routing problem: improved adaptive large neighbourhood search," *Optimization*, vol. 71, no. 11, pp. 3165–3193, 2022.

[38] A. Hoff, H. Andersson, M. Christiansen, G. Hasle, and A. Løkketangen, "Industrial aspects and literature survey: Fleet composition and routing," *Computers & Operations Research*, vol. 37, no. 12, pp. 2041–2061, 2010.

[39] M. O. Ball, B. Golden, A. Assad, and L. Bodin, "Planning for truck fleet size in the presence of a common-carrier option," *Decision Sciences*, vol. 14, no. 1, pp. 103–120, 1983.

[40] J. G. Klincewicz, H. Luss, and M. G. Pilcher, “Fleet size planning when outside carrier services are available,” *Transportation Science*, vol. 24, no. 3, pp. 169–182, 1990.

[41] R. W. Hall and M. Racer, “Transportation with common carrier and private fleets: system assignment and shipment frequency optimization,” *IIE Transactions*, vol. 27, no. 2, pp. 217–225, 1995.

[42] C.-W. Chu, “A heuristic algorithm for the truckload and less-than-truckload problem,” *European Journal of Operational Research*, vol. 165, no. 3, pp. 657–667, 2005.

[43] M.-C. Bolduc, J. Renaud, and F. Boctor, “A heuristic for the routing and carrier selection problem,” *European Journal of Operational Research*, vol. 183, no. 2, pp. 926–932, 2007.

[44] J.-F. Côté and J.-Y. Potvin, “A tabu search heuristic for the vehicle routing problem with private fleet and common carrier,” *European Journal of Operational Research*, vol. 198, no. 2, pp. 464–469, 2009.

[45] M. A. Krajewska and H. Kopfer, “Transportation planning in freight forwarding companies: Tabu search algorithm for the integrated operational transportation planning problem,” *European Journal of Operational Research*, vol. 197, no. 2, pp. 741–751, 2009.

[46] C. Gahm, C. Brabänder, and A. Tuma, “Vehicle routing with private fleet, multiple common carriers offering volume discounts, and rental options,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 97, pp. 192–216, 2017.

[47] R. Kaewpuang, D. Niyato, P.-S. Tan, and P. Wang, “Cooperative management in full-truckload and less-than-truckload vehicle system,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 5707–5722, 2016.

[48] J. V. Jucker and M. J. Rosenblatt, “Single-period inventory models with demand uncertainty and quantity discounts: Behavioral implications and a new solution procedure,” *Naval Research Logistics Quarterly*, vol. 32, no. 4, pp. 537–550, 1985.

[49] S. Elhedhli and Ö. Benli, “Optimal lot sizing under carload discount schedules,” *INFOR: Information Systems and Operational Research*, vol. 43, no. 4, pp. 361–370, 2005.

[50] S. Nahmias and T. Olsen, *Production and Operations Analytics: Eighth Edition*. Long Grove, IL: Waveland Press, 2020.

[51] U. Yüceer, “Discrete convexity: Convexity for functions defined on discrete spaces,” *Discrete Applied Mathematics*, vol. 119, pp. 297–304, 2002.

[52] Department of Energy, “Alternative fuels data center: Maps and data - vehicle weight classes & categories,” 2012. [Online]. Available: `https://afdc.energy.gov/data/10380`.

[53] A. Shapiro and A. Philpott, *A tutorial on stochastic programming.* 2007. [Online]. Available: `https://sites.gatech.edu/alexander-shapiro/files/2021/03/TutorialSP.pdf`.

[54] Centre for Research on the Epidemiology of Disasters, “Disasters in numbers 2022,” 2023.

[55] S. D. Smith, C. J. Dooley, and D. S. Burch, “Quantifying patient vibration patterns during ambulance bus (ambus) ground transport,” tech. rep., Infoscitex, 2020.

[56] G. Koutitas, K. S. Smith, G. Lawrence, V. Metsis, C. Stamper, M. Trahan, and T. Lehr, “A virtual and augmented reality platform for the training of first responders of the ambulance bus,” in *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, pp. 299–302, 2019.

[57] T. Lentz, C. Groizard, A. Colomes, A. Ozguler, M. Baer, and T. Loeb, “Collective critical care ambulance: an innovative transportation of critical care patients by bus in covid-19 pandemic response,” *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 29, no. 1, pp. 1–7, 2021.

[58] D. G. Barten, R. van Zijl, F. W. Körver, and N. A. Peters, “Simultaneous interfacility transfer of multiple non-critically ill covid-19 patients using a single vehicle: the ambulance bus experience,” *International Journal of Emergency Medicine*, vol. 15, no. 1, p. 10, 2022.

[59] R. Z. Farahani, M. Lotfi, A. Baghaian, R. Ruiz, and S. Rezapour, “Mass casualty management in disaster scene: A systematic review of or&ms research in humanitarian operations,” *European Journal of Operational Research*, vol. 287, no. 3, pp. 787–819, 2020.

[60] R. Aringhieri, M. E. Bruni, S. Khodaparasti, and J. T. van Essen, “Emergency medical services and beyond: Addressing new challenges through a wide literature review,” *Computers & Operations Research*, vol. 78, pp. 349–368, 2017.

[61] V. Bélanger, A. Ruiz, and P. Soriano, “Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles,” *European Journal of Operational Research*, vol. 272, no. 1, pp. 1–23, 2019.

[62] Q. Gong and R. Batta, “Allocation and reallocation of ambulances to casualty clusters in a disaster relief operation,” *IIE Transactions*, vol. 39, no. 1, pp. 27–39, 2007.

[63] L. Talarico, F. Meisel, and K. Sörensen, “Ambulance routing for disaster response with patient groups,” *Computers & Operations Research*, vol. 56, pp. 120–133, 2015.

[64] P. P. Repoussis, D. C. Paraskevopoulos, A. Vazacopoulos, and N. Hupert, “Optimizing emergency preparedness and resource utilization in mass-casualty incidents,” *European Journal of Operational Research*, vol. 255, no. 2, pp. 531–544, 2016.

[65] E. Mouhcine, Y. Karouani, K. Mansouri, and Y. Mohamed, “Toward a distributed strategy for emergency ambulance routing problem,” in *2018 4th International Conference on Optimization and Applications (ICOA)*, pp. 1–4, 2018.

[66] S. Yoon and L. A. Albert, “A dynamic ambulance routing model with multiple response,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 133, p. 101807, 2020.

[67] W. J. Sacco, D. M. Navin, K. E. Fiedler, R. K. Waddell II, W. B. Long, and R. F. Buckman Jr, “Precise formulation and evidence-based application of resource-constrained triage,” *Academic Emergency Medicine*, vol. 12, no. 8, pp. 759–770, 2005.

[68] E. Erkut, A. Ingolfsson, and G. Erdoğan, "Ambulance location for maximum survival," *Naval Research Logistics (NRL)*, vol. 55, no. 1, pp. 42–58, 2008.

[69] V. A. Knight, P. R. Harper, and L. Smith, "Ambulance allocation for maximal survival with heterogeneous outcome measures," *Omega*, vol. 40, no. 6, pp. 918–926, 2012.

[70] A. F. Mills, N. T. Argon, and S. Ziya, "Resource-Based Patient Prioritization in Mass-Casualty Incidents," *Manufacturing & Service Operations Management*, vol. 15, no. 3, pp. 361–377, 2013.

[71] D. Bandara, M. E. Mayorga, and L. A. McLay, "Priority dispatching strategies for ems systems," *Journal of the Operational Research Society*, vol. 65, no. 4, pp. 572–587, 2014.

[72] I. Sung and T. Lee, "Optimal allocation of emergency medical resources in a mass casualty incident: Patient prioritization by column generation," *European Journal of Operational Research*, vol. 252, no. 2, pp. 623–634, 2016.

[73] A. F. Mills, N. T. Argon, and S. Ziya, "Dynamic distribution of patients to medical facilities in the aftermath of a disaster," *Operations Research*, vol. 66, no. 3, pp. 716–732, 2018.

[74] J. Beardwood, J. H. Halton, and J. M. Hammersley, "The shortest path through many points," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 55, pp. 299–327, Cambridge University Press, 1959.

[75] D. M. Stein, "An asymptotic, probabilistic analysis of a routing problem," *Mathematics of Operations Research*, vol. 3, no. 2, pp. 89–101, 1978.

[76] P. Jaillet, "A priori solution of a traveling salesman problem in which a random subset of the customers are visited," *Operations Research*, vol. 36, no. 6, pp. 929–936, 1988.

[77] H. Ong and H. Huang, "Asymptotic expected performance of some tsp heuristics: an empirical evaluation," *European Journal of Operational Research*, vol. 43, no. 2, pp. 231–238, 1989.

[78] M. A. Figliozzi, "Planning approximations to the average length of vehicle routing problems with time window constraints," *Transportation Research Part B: Methodological*, vol. 43, no. 4, pp. 438–447, 2009.

[79] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

APPENDIX A

CHAPTER ONE APPENDIX

A.1 Proof of the convexity of $\Pi(s, \ell, N)$ for a given N

The objective function after substituting $p^*(\ell)$ is

$$\begin{aligned}\Pi(s, \ell, N) &= J(N)\beta\ell^{-\frac{v}{u}} - KN - h_F k_F \bar{\sigma}\sqrt{J(N)\left(T_F + T_L - s\right)} \\ &\quad - h_L k_L \bar{\sigma}\sqrt{J(N) N \left(T_C + s - \ell\right)}.\end{aligned}$$

We first derive the first-order derivatives for s and ℓ, respectively:

$$\begin{aligned}\frac{\partial \Pi(s, \ell, N)}{\partial s} &= \frac{1}{2} h_F k_F \bar{\sigma}\sqrt{\frac{J(N)}{T_F + T_L - s}} - \frac{1}{2} h_L k_L \bar{\sigma}\sqrt{\frac{J(N)N}{T_C + s - \ell}}, \\ \frac{\partial \Pi(s, \ell, N)}{\partial \ell} &= -\frac{v}{u} J(N)\beta\ell^{-\frac{v}{u}-1} + \frac{1}{2} h_L k_L \bar{\sigma}\sqrt{\frac{J(N)N}{T_C + s - \ell}}.\end{aligned}$$

Next, we derive the second-order derivatives:

$$\frac{\partial^2 \Pi(s, \ell, N)}{\partial s^2} = \frac{h_F k_F \bar{\sigma}\sqrt{J(N)}}{4\left(T_F + T_L - s\right)^{\frac{3}{2}}} + \frac{h_L k_L \bar{\sigma}\sqrt{J(N)N}}{4\left(T_C + s - \ell\right)^{\frac{3}{2}}} \equiv H_1, \tag{A.1}$$

$$\frac{\partial^2 \Pi(s, \ell, N)}{\partial s \partial \ell} = -\frac{h_L k_L \bar{\sigma}\sqrt{J(N)N}}{4\left(T_C + s - \ell\right)^{\frac{3}{2}}} \equiv H_2, \tag{A.2}$$

$$\frac{\partial^2 \Pi(s, \ell, N)}{\partial \ell \partial s} = -\frac{h_L k_L \bar{\sigma}\sqrt{J(N)N}}{4\left(T_C + s - \ell\right)^{\frac{3}{2}}} \equiv H_3, \tag{A.3}$$

$$\frac{\partial^2 \Pi(s, \ell, N)}{\partial \ell^2} = \frac{v}{u}\left(\frac{v}{u} + 1\right) J(N)\beta\ell^{-\frac{v}{u}-2} + \frac{h_L k_L \bar{\sigma}\sqrt{J(N)N}}{4\left(T_C + s - \ell\right)^{\frac{3}{2}}} \equiv H_4. \tag{A.4}$$

The Hessian matrix $\left[\begin{smallmatrix} H_1 & H_2 \\ H_3 & H_4 \end{smallmatrix}\right]$ is symmetric since $H_2 = H_3$. Note that $H_1 > 0$ and by the determinant test, $H_1 H_4 - H_2 H_3 > 0$. So, the Hessian matrix is positive definite, and $\Pi(s, \ell, N)$ is convex in (s, ℓ) for a given N.

A.2 Pairwise comparisons and optimality regions, $J(N) = N$

A.2.1 Optimality region for $\Pi_O(N)$

First, we suppose the optimal strategy is not holding any safety stock in the system, i.e., $\Pi_O(N)$ has the maximum value among the four candidate optimal solutions. We have

$$\Pi_O(N) \geq \Pi_F(N) : N \leq \frac{(h_F k_F \bar{\sigma})^2 T_F}{\beta^2 \left[(T_L + T_C)^{-\frac{v}{u}} - (T_F + T_L + T_C)^{-\frac{v}{u}}\right]^2} \equiv N_{O,F}(T_C); \tag{A.5}$$

$$\Pi_O(N) \geq \Pi_{FL}(N) : N \leq \frac{(h_F k_F \bar{\sigma})^2 T_F}{\left[\beta \left(T_C^{-\frac{v}{u}} - (T_F + T_L + T_C)^{-\frac{v}{u}}\right) - h_L k_L \bar{\sigma}\sqrt{T_L}\right]^2} \equiv N_{O,FL}(T_C) \tag{A.6}$$

$$\text{and } \beta \left(T_C^{-\frac{v}{u}} - (T_F + T_L + T_C)^{-\frac{v}{u}}\right) - h_L k_L \bar{\sigma}\sqrt{T_L} \equiv T_{O,FL}(T_C) > 0, \tag{A.7}$$

$$\text{or } T_{O,FL}(T_C) \leq 0; \tag{A.8}$$

$$\Pi_O(N) \geq \Pi_L(N) : \beta \left(T_C^{-\frac{v}{u}} - (T_F + T_L + T_C)^{-\frac{v}{u}}\right) - h_L k_L \bar{\sigma}\sqrt{T_F + T_L} \equiv T_{O,L}(T_C) \leq 0. \tag{A.9}$$

Because $T_{O,FL}(T_C) > T_{O,L}(T_C)$, the optimality region for $\Pi_O(N)$, denoted by R_O, is

$$R_O = R_O^1 \bigcup R_O^2, \tag{A.10}$$

$$\text{where } R_O^1 = \Big\{(T_C, N) : N \leq \min\big\{N_{O,F}(T_C), N_{O,FL}(T_C)\big\}, T_{O,L}(T_C) \leq 0, T_{O,FL}(T_C) > 0\Big\}, \tag{A.11}$$

$$R_O^2 = \Big\{(T_C, N) : N \leq N_{O,F}(T_C), T_{O,FL}(T_C) \leq 0\Big\}. \tag{A.12}$$

A.2.2 Optimality region for $\Pi_F(N)$

We next suppose that holding safety stock at the FC only is the optimal strategy, which implies $\Pi_F(N)$ is the maximum objective function value among the four optimal solution candidates. We compare $\Pi_F(N)$ with $\Pi_{FL}(N)$ and $\Pi_L(N)$, respectively. We have

$$\Pi_F(N) \geq \Pi_{FL}(N) : \beta \left(T_C^{-\frac{v}{u}} - (T_L + T_C)^{-\frac{v}{u}}\right) - h_L k_L \bar{\sigma}\sqrt{T_L} \equiv T_{F,FL}(T_C) \leq 0, \tag{A.13}$$

$$\Pi_F(N) \geq \Pi_L(N) : N \geq \frac{(h_F k_F \bar{\sigma})^2 T_F}{\left[h_L k_L \bar{\sigma}\sqrt{T_F + T_L} - \beta \left(T_C^{-\frac{v}{u}} - (T_L + T_C)^{-\frac{v}{u}}\right)\right]^2} \equiv N_{F,L}(T_C) \tag{A.14}$$

$$\text{and } \beta \left(T_C^{-\frac{v}{u}} - (T_L + T_C)^{-\frac{v}{u}}\right) - h_L k_L \bar{\sigma}\sqrt{T_F + T_L} \equiv T_{F,L}(T_C) < 0. \tag{A.15}$$

Additionally, we can simply take the opposite direction of Inequality (A.5) to obtain the threshold value of N for $\Pi_F(N) \geq \Pi_O(N)$, that is,

$$N \geq N_{O,F}(T_C). \tag{A.16}$$

$T_{F,FL}(T_C) \leq 0$ dominates $T_{F,L}(T_C) < 0$ since $T_{F,FL}(T_C) \geq T_{F,L}(T_C)$ (we assume that T_F and T_L are positive values). We can therefore write the optimality region for $\Pi_F(N)$ as

$$R_F = \Big\{(T_C, N) : N \geq \max\{N_{O,F}(T_C), N_{F,L}(T_C)\}, T_{F,FL}(T_C) \leq 0\Big\}. \tag{A.17}$$

A.2.3 Optimality region for $\Pi_{FL}(N)$

We take the opposite direction of Inequality (A.6) while restricting T_C to satisfy (A.7), and take the opposite direction of Inequality (A.13), to obtain the threshold values of N and T_C for the scenarios where $\Pi_{FL}(N) \geq \Pi_O(N)$ and $\Pi_{FL}(N) \geq \Pi_F(N)$. We have

$$\Pi_{FL}(N) \geq \Pi_O(N) : N \geq N_{O,FL}(T_C) \tag{A.18}$$

$$\text{and } \beta\left(T_C^{-\frac{v}{u}} - (T_F + T_L + T_C)^{-\frac{v}{u}}\right) - h_L k_L \bar{\sigma}\sqrt{T_L} \equiv T_{O,FL}(T_C) > 0, \tag{A.19}$$

$$\Pi_{FL}(N) \geq \Pi_F(N) : \beta\left(T_C^{-\frac{v}{u}} - (T_L + T_C)^{-\frac{v}{u}}\right) - h_L k_L \bar{\sigma}\sqrt{T_L} \equiv T_{F,FL}(T_C) \geq 0. \tag{A.20}$$

We now suppose $\Pi_{FL}(N) \geq \Pi_L(N)$ and compare the two objective function values. We have

$$\Pi_{FL}(N) \geq \Pi_L(N) : \; N \geq \left(\frac{h_F k_F}{h_L k_L}\right)^2 \frac{T_F}{\left(\sqrt{T_F + T_L} - \sqrt{T_L}\right)^2} \equiv N_{FL,L}. \tag{A.21}$$

It is straightforward to see that given T_C, $T_{F,FL}(T_C) < T_{O,FL}(T_C)$. So, $T_{F,FL}(T_C) \geq 0$ dominates $T_{O,FL}(T_C) > 0$. Moreover, $N_{FL,L}$ is a constant which does not depend on T_C. According to (A.19) to (A.21), the optimality region for $\Pi_{FL}(N)$ is given by

$$R_{FL} = \Big\{(T_C, N) : N \geq \max\{N_{O,FL}(T_C), N_{FL,L}\}, T_{F,FL}(T_C) \geq 0\Big\}. \tag{A.22}$$

A.2.4 Optimality region for $\mathit{\Pi}_L(N)$

To obtain the threshold values of N and T_C for scenarios where $\mathit{\Pi}_L(N) \geq \mathit{\Pi}_O(N)$ and $\mathit{\Pi}_L(N) \geq \mathit{\Pi}_{FL}(N)$, we can directly take the opposite directions of Inequalities (A.9) and (A.21), respectively. We have

$$\mathit{\Pi}_L(N) \geq \mathit{\Pi}_O(N): \ T_{O,L}(T_C) \geq 0, \tag{A.23}$$

$$\mathit{\Pi}_L(N) \geq \mathit{\Pi}_{FL}(N): \ N \leq N_{FL,L}. \tag{A.24}$$

For $\mathit{\Pi}_L(N) \geq \mathit{\Pi}_F(N)$, however, we can either take the opposite direction of Inequality (A.14) while satisfying (A.15), or solely take the opposite direction of Inequality (A.15). Thus, we have

$$\mathit{\Pi}_L(N) \geq \mathit{\Pi}_F(N): \ T_{F,L}(T_C) \geq 0, \text{ or } N \leq N_{F,L}(T_C) \text{ and } T_{F,L}(T_C) < 0. \tag{A.25}$$

Note that given T_C, $T_{O,L}(T_C) > T_{F,L}(T_C)$ implies $T_{F,L}(T_C) \geq 0$ dominates $T_{O,L}(T_C) \geq 0$. In accordance with (A.23), (A.24), and (A.25), the optimality region for $\mathit{\Pi}_L(N)$ in (T_C, N) is formed by

$$R_L = R_L^1 \bigcup R_L^2, \tag{A.26}$$

where $$R_L^1 = \Big\{(T_C, N) : N \leq N_{FL,L}, T_{F,L}(T_C) \geq 0\Big\}, \tag{A.27}$$

$$R_L^2 = \Big\{(T_C, N) : N \leq \min\{N_{FL,F}, N_{F,L}(T_C)\}, T_{F,L}(T_C) < 0, T_{O,L}(T_C) \geq 0\Big\}. \tag{A.28}$$

A.3 Pairwise comparisons and optimality regions when $T_C(N) = T_0/N^\theta$ and any $J(N)$

$$\Pi_O(N) \geq \Pi_F(N) : \sqrt{J(N)} - \frac{h_F k_F \bar{\sigma}\sqrt{T_F}}{\beta\left((T_L + \frac{T_0}{N^\theta})^{-\frac{v}{u}} - (T_F + T_L + \frac{T_0}{N^\theta})^{-\frac{v}{u}}\right)} \equiv g_{O,F}(T_0, N) \leq 0. \tag{A.29}$$

$$\Pi_O(N) \geq \Pi_{FL}(N) : \sqrt{J(N)} - \frac{h_L k_L \bar{\sigma}\sqrt{NT_L} + h_F k_F \bar{\sigma}\sqrt{T_F}}{\beta\left((\frac{T_0}{N^\theta})^{-\frac{v}{u}} - (T_F + T_L + \frac{T_0}{N^\theta})^{-\frac{v}{u}}\right)} \equiv g_{O,FL}(T_0, N) \leq 0. \tag{A.30}$$

$$\Pi_O(N) \geq \Pi_L(N) : \sqrt{\frac{J(N)}{N}} - \frac{h_L k_L \bar{\sigma}\sqrt{T_F + T_L}}{\beta\left((\frac{T_0}{N^\theta})^{-\frac{v}{u}} - (T_F + T_L + \frac{T_0}{N^\theta})^{-\frac{v}{u}}\right)} \equiv g_{O,L}(T_0, N) \leq 0. \tag{A.31}$$

$$\Pi_F(N) \geq \Pi_{FL}(N) : \sqrt{\frac{J(N)}{N}} - \frac{h_L k_L \bar{\sigma}\sqrt{T_L}}{\beta\left((\frac{T_0}{N^\theta})^{-\frac{v}{u}} - (T_L + \frac{T_0}{N^\theta})^{-\frac{v}{u}}\right)} \equiv g_{F,FL}(T_0, N) \leq 0. \tag{A.32}$$

$$\Pi_F(N) \geq \Pi_L(N) : \sqrt{J(N)} - \frac{h_L k_L \bar{\sigma}\sqrt{N(T_F + T_L)} - h_F k_F \bar{\sigma}\sqrt{T_F}}{\beta\left((\frac{T_0}{N^\theta})^{-\frac{v}{u}} - (T_L + \frac{T_0}{N^\theta})^{-\frac{v}{u}}\right)} \equiv g_{F,L}(T_0, N) \leq 0. \tag{A.33}$$

$$\Pi_{FL}(N) \geq \Pi_L(N) : N \geq \left(\frac{h_F k_F}{h_L k_L}\right)^2 \frac{T_F}{(\sqrt{T_F + T_L} - \sqrt{T_L})^2} \equiv N_{FL,L}. \tag{A.34}$$

A.4 Proofs of propositions and corollaries

Proof of Proposition 1.5.1. To prove this proposition, we simply set $\Pi_L(N) = \Pi_{FL}(N)$ by substituting the results in Table 1.1. The solution, denoted by $N_{FL,L}$, is

$$N_{FL,L} = (h_F k_F)^2 T_F / \left(h_L k_L(\sqrt{T_F + T_L} - \sqrt{T_L})\right)^2,$$

which is independent of T_C and $J(N)$. Therefore, when all other parameters are fixed, $N_{FL,L}$ is also fixed. ■

Proof of Proposition 1.5.2. To prove the former, let $J(N) = 1$ and set $\Pi_F(N) = \Pi_O(N)$; using the results in Table 1.1, we obtain

$$\beta\left((T_L + T_C)^{-\frac{v}{u}} - (T_F + T_L + T_C)^{-\frac{v}{u}}\right) = h_F k_F \bar{\sigma}\sqrt{T_F}. \tag{A.35}$$

For fixed parameter values, Equation (A.35) is satisfied at a fixed value of T_C that is independent of N. Similarly, to prove the latter, we set $\mathit{\Pi}_F(N) = \mathit{\Pi}_{FL}(N)$ and $J(N) = N$. The resulting equation is

$$\beta\left(T_C^{-\frac{v}{u}} - (T_L + T_C)^{-\frac{v}{u}}\right) = h_L k_L \bar{\sigma}\sqrt{T_L}. \tag{A.36}$$

Equation (A.36) is also satisfied at a fixed value of T_C that is independent of N when all other parameters are fixed. ■

Proof of Proposition 1.5.3. We first fix N such that $N = \hat{N}$. Let $T_C(\hat{N})$ denote the threshold value at which we transition between safety stock placement strategies (our previous methods can be used to determine $T_C(\hat{N})$ for any given $\hat{N}$). Next, note that $T_C(\hat{N}) = T_0/\hat{N}^\theta$, where the left-hand side of this equation corresponds to a fixed threshold value. As θ increases, the denominator of the right-hand side of the equation increases, while maintaining equality requires increasing T_0, which leads to the rightward movement of the borders along T_0 in Figure 1.6. Additionally, recall from Proposition 1.5.1 that the transition between the FL-strategy and the LDC-strategy is independent of T_C; thus, this transition threshold value remains independent of T_0. ■

Proof of Corollary 1.6.1. We show in Proposition 1.5.1 that the border between the optimality regions for the FL- and LDC-strategies occurs at $N_{FL,L} = \frac{(h_F k_F)^2 T_F}{\left(h_L k_L(\sqrt{T_F+T_L}-\sqrt{T_L})\right)^2}$. Fixing all other parameters, when h_F/h_L increases, the threshold value of $N_{FL,L}$ increases accordingly. ■

Proof of Proposition 1.6.2. When $\mathit{\Pi}_F(N) = \mathit{\Pi}_{FL}(N)$ holds, we can obtain the resulting equation:

$$\sqrt{\frac{J(N)}{N}} = \frac{h_L k_L \bar{\sigma}\sqrt{T_L}}{\beta\left(T_C^{-\frac{v}{u}} - (T_L + T_C)^{-\frac{v}{u}}\right)}. \tag{A.37}$$

At a given value of N such that $J(N)$ is fixed, the left-hand side of the above equation is fixed. When the value of h_L increases, in order to maintain equality, the denominator of the right-hand side must increase, which requires reducing the value of T_C (given fixed values of all other parameters) since the denominator is decreasing in T_C. Similarly, $\mathit{\Pi}_O(N) = \mathit{\Pi}_L(N)$ implies

$$\sqrt{\frac{J(N)}{N}} = \frac{h_L k_L \bar{\sigma}\sqrt{T_F + T_L}}{\beta\left(T_C^{-\frac{v}{u}} - (T_F + T_L + T_C)^{-\frac{v}{u}}\right)}, \tag{A.38}$$

which also requires reducing T_C in order to maintain equality if the value of h_L increases. ■

Proof of Proposition 1.6.3. To show this result, we can use the same approach used in the proof of Proposition 1.6.2. At $\mathit{\Pi}_{FL}(N) = \mathit{\Pi}_F(N)$, the equation representing the border is

$$\sqrt{N^{\gamma-1}} = \frac{h_L k_L \bar{\sigma}\sqrt{T_L}}{\beta\left((\frac{T_0}{N^\theta})^{-\frac{v}{u}} - (T_L + \frac{T_0}{N^\theta})^{-\frac{v}{u}}\right)}. \tag{A.39}$$

The left-hand side of the above equation increases as γ increases. To maintain equality with all other parameters fixed, the value of T_0 must increase (as the denominator is decreasing in T_0). Similarly, the equation representing the border by setting $\mathit{\Pi}_F(N) = \mathit{\Pi}_O(N)$ is

$$\sqrt{N^{\gamma}} = \frac{h_F k_F \bar{\sigma}\sqrt{T_F}}{\beta\left((T_L + \frac{T_0}{N^\theta})^{-\frac{v}{u}} - (T_F + T_L + \frac{T_0}{N^\theta})^{-\frac{v}{u}}\right)}. \tag{A.40}$$

As the value of γ grows, the left-hand side increases, which forces T_0 to take a larger value to maintain equality. ■

A.5 Proof of the convexity of $\mathit{\Pi}(\mathrm{s}, \boldsymbol{\ell}, \mathbf{N})$ in $(\mathrm{s}, \boldsymbol{\ell})$ for a given $\mathbf{N}$

Proof. Since we now aim on proving the convexity of the objective function in $(\mathrm{s}, \boldsymbol{\ell})$, we can rewrite function (1.12), by replacing $\mathbf{N}$ with a given vector, $\hat{\mathbf{N}} = (\hat{N}_1, ..., \hat{N}_M)$, as

$$\mathit{\Pi}(\mathbf{s}, \boldsymbol{\ell}, \hat{\mathbf{N}}) = R(\boldsymbol{\ell}, \hat{\mathbf{N}}) - \sum_{m=1}^{M} \hat{N}_m K_m - HC_F(\mathbf{s}, \hat{\mathbf{N}}) - HC_L(\mathbf{s}, \boldsymbol{\ell}, \hat{\mathbf{N}}), \tag{A.41}$$

where $R(\boldsymbol{\ell}, \hat{\mathbf{N}})$, $HC_F(\mathbf{s}, \hat{\mathbf{N}})$, and $HC_L(\mathbf{s}, \boldsymbol{\ell}, \hat{\mathbf{N}})$ correspond to the system expected revenue, the total holding cost at the FC, and the total holding cost at the LDCs in the system, respectively.

Therefore, it suffices to show that function (A.41) is a nonnegative weighted sum of convex functions.

Convexity of $R(\boldsymbol{\ell}, \hat{\mathbf{N}})$. By replacing $\mathbf{N}$ by $\hat{\mathbf{N}}$, we denote $A(\hat{\mathbf{N}}) = \sum_{m=1}^{M} J_m(\hat{N}_m)a_m$ and $B(\boldsymbol{\ell}, \hat{\mathbf{N}}) = \sum_{m=1}^{M} J_m(\hat{N}_m)b_m\ell_m^v$. Therefore, we have

$$\begin{aligned}
R(\boldsymbol{\ell}, \hat{\mathbf{N}}) &= \sum_{m=1}^{M} J_m(\hat{N}_m)\left(a_m\left(\frac{A(\hat{\mathbf{N}})}{(u+1)B(\boldsymbol{\ell}, \hat{\mathbf{N}})}\right)^{\frac{1}{u}} - b_m\left(\frac{A(\hat{\mathbf{N}})}{(u+1)B(\boldsymbol{\ell}, \hat{\mathbf{N}})}\right)^{1+\frac{1}{u}}\ell_m^v\right) \\
&= \left(\frac{A(\hat{\mathbf{N}})}{(u+1)B(\boldsymbol{\ell}, \hat{\mathbf{N}})}\right)^{\frac{1}{u}} A(\hat{\mathbf{N}}) - \left(\frac{A(\hat{\mathbf{N}})}{(u+1)B(\boldsymbol{\ell}, \hat{\mathbf{N}})}\right)^{1+\frac{1}{u}} B(\boldsymbol{\ell}, \hat{\mathbf{N}}) \\
&= \left(\frac{A(\hat{\mathbf{N}})}{u+1}\right)^{1+\frac{1}{u}}\left(\frac{u+1}{\left(B(\boldsymbol{\ell}, \hat{\mathbf{N}})\right)^{\frac{1}{u}}} - \frac{1}{\left(B(\boldsymbol{\ell}, \hat{\mathbf{N}})\right)^{\frac{1}{u}}}\right) \\
&= \left(\frac{A(\hat{\mathbf{N}})}{u+1}\right)^{1+\frac{1}{u}} u\left(B(\boldsymbol{\ell}, \hat{\mathbf{N}})\right)^{-\frac{1}{u}}. \qquad \text{(A.42)}
\end{aligned}$$

Given $\hat{\mathbf{N}}$, the value of $A(\hat{\mathbf{N}})$ is fixed. Thus, the convexity of (A.42) is dependent on that of $\left(B(\boldsymbol{\ell}, \hat{\mathbf{N}})\right)^{-\frac{1}{u}}$. It is straightforward that ℓ_m^v is concave in ℓ_m if $0 < v \leq 1$. So, $B(\boldsymbol{\ell}, \hat{\mathbf{N}})$ is concave in $\boldsymbol{\ell}$. Let $h(y) = y^{-\frac{1}{u}}$ $(u > 0)$ with domain $\mathbb{R}^+$. Since h is convex and its extended-value extension is nonincreasing, by composition [79], $R(\boldsymbol{\ell}, \hat{\mathbf{N}}) = \left(\frac{A(\hat{\mathbf{N}})}{u+1}\right)^{1+\frac{1}{u}} uh(B(\boldsymbol{\ell}, \hat{\mathbf{N}}))$ is convex in $\boldsymbol{\ell}$ for $0 < v \leq 1$.

Concavity of $HC_F(\mathbf{s}, \hat{\mathbf{N}})$ and $HC_L(\mathbf{s}, \boldsymbol{\ell}, \hat{\mathbf{N}})$. Let $g_F(\mathbf{s}, \hat{\mathbf{N}}) = \sum_{m=1}^{M} \bar{\sigma}_m^2 J_m(\hat{N}_m)(T_F + T_{L,m} - s_m)$, which is a linear function of s. Hence, it is also a concave function. Since the square root function with domain $\mathbb{R}^+$ is concave and its extended-value extension is nondecreasing, by composition, $HC_F(\mathbf{s}, \hat{\mathbf{N}}) = h_F k_F \sqrt{g_F(\mathbf{s}, \hat{\mathbf{N}})}$ is concave in s. Similarly, let $g_{L,m}(s_m, \ell_m, \hat{N}_m) = \sqrt{J_m(\hat{N}_m)\hat{N}_m\left(\frac{T_{0,m}}{\hat{N}_m^{\theta}} + s_m - \ell_m\right)}$, for $m = 1, ..., M$. For each m, by composition, $g_{L,m}(s_m, \ell_m, \hat{N}_m)$ is concave in (s_m, ℓ_m). Hence, $HC_L(\mathbf{s}, \boldsymbol{\ell}, \hat{\mathbf{N}}) = \sum_{m=1}^{M} h_{L,m}k_{L,m}\bar{\sigma}_m g_{L,m}(s_m, \ell_m, \hat{N}_m)$ is concave in $(\mathbf{s}, \boldsymbol{\ell})$. ■

APPENDIX B

CHAPTER TWO APPENDIX

B.1 Discrete convexity for homogeneous fleet with normal demand

The first difference of function (2.15) in n equals

$$\Delta_n = g_N(n+1) - g_N(n) = K + a\sigma\big(L(z_{(n+1)W}) - L(z_{nW+Y})\big),$$

and the second difference is written as

$$\Delta_{n+1} - \Delta_n = a\sigma\Big(L(z_{(n+2)W}) - L(z_{(n+1)W+Y}) - \big(L(z_{(n+1)W}) - L(z_{nW+Y})\big)\Big).$$

The first order derivative of the normal loss function $L(z) = \int_z^{\infty}(u-z)\phi(u)du$ equals

$$L'(z) = \phi'(z) + z\phi(z) - (1 - \Phi(z)).$$

Note that $\phi'(z) = -z\phi(z)$, and we have

$$L'(z) = -(1 - \Phi(z)) = \Phi(z) - 1 < 0, \tag{B.1}$$

which implies that $L(z)$ is strictly decreasing, and $L''(z) = \phi(z) > 0$, implying $L(z)$ is strictly convex.

Now consider the quantity

$$L(z_{(n+2)W}) - L(z_{(n+1)W+Y}) - \big(L(z_{(n+1)W}) - L(z_{nW+Y})\big)$$

and note that $L(z_{(k+1)W}) - L(z_{kW+Y}) < 0$ for any k. Let $h(k) = L(z_{(k+1)W}) - L(z_{kW+Y}) =$

$L\left(\frac{(k+1)W-\mu}{\sigma}\right) - L\left(\frac{kW+Y-\mu}{\sigma}\right)$, according to Equation (B.1), we have

$$\frac{dh(k)}{dk} = \frac{W}{\sigma}\left(\Phi(z_{(k+1)W}) - \Phi(z_{kW+Y})\right) > 0.$$

Thus, $h(k)$ is strictly increasing in k. This implies that $\Delta_{n+1} - \Delta_n = a\sigma\big(h(n+1) - h(n)\big) > 0$, i.e., the second difference is positive. Therefore, $g_N(n)$ is discretely convex in n.

B.2 Discrete convexity of $g^a(n)$ for homogeneous fleet

The first difference of function (2.16) in n is

$$\Delta_n^a = g^a(n+1) - g^a(n) = K - \int_{a_L}^{a_U} a\ell\left(nW + \frac{A}{a}\right)p(a)da + \mu_a\ell((n+1)W),$$

and the second difference equals

$$\begin{aligned}\Delta_{n+1}^a - \Delta_n^a &= \mu_a\ell((n+2)W) - \int_{a_L}^{a_U} a\ell\left((n+1)W + \frac{A}{a}\right)p(a)da \\ &\quad - \left(\mu_a\ell((n+1)W) - \int_{a_L}^{a_U} a\ell\left(nW + \frac{A}{a}\right)p(a)da\right).\end{aligned}$$

Let us define $\tilde{h}_a(k) = \mu_a\ell((k+1)W) - \int_{a_L}^{a_U} a\ell\left(kW + \frac{A}{a}\right)p(a)da$; we then have $\Delta_{n+1} - \Delta_n = \tilde{h}_a(n+1) - \tilde{h}_a(n)$. Note that

$$\begin{aligned}\frac{d\tilde{h}_a(k)}{dk} &= \mu_a W\ell'((k+1)W) - W\int_{a_L}^{a_U} a\ell'\left(kW + \frac{A}{a}\right)p(a)da \\ &= \mu_a W(F((k+1)W) - 1) - W\int_{a_L}^{a_U} a\left(F\left(kW + \frac{A}{a}\right) - 1\right)p(a)da \\ &= \mu_a WF((k+1)W) - W\int_{a_L}^{a_U} aF\left(kW + \frac{A}{a}\right)p(a)da \\ &= W\int_{a_L}^{a_U} a\left(F((k+1)W) - F\left(kW + \frac{A}{a}\right)\right)p(a)da \\ &\geq 0,\end{aligned}$$

where the last inequality holds because $\frac{A}{a} \leq \frac{A}{a_L} \leq W$ such that $F((k+1)W) \geq F\left(kW + \frac{A}{a}\right)$. This implies that $\tilde{h}_a(k)$ is nondecreasing in k, and that $\Delta^a_{n+1} - \Delta^a_n \geq 0$. Thus, $g^a(n)$ is discretely convex in n.

B.3 Proofs of theorems and propositions

Proof of Theorem 2.4.1. Given a demand observation $\tilde{x}$, we can solve the following capacity allocation problem formulated as a mixed integer programming (MIP) model:

$$\text{Minimize} \quad g(n_1, n_2, y_1, y_2) = (K_1 + A_1)n_1 + (K_2 + A_2)n_2 + a(\tilde{x} - y_1 - y_2) \tag{B.2}$$

$$\text{s.t.} \quad 0 \leq y_1 \leq W_1 n_1, \tag{B.3}$$

$$0 \leq y_2 \leq W_2 n_2, \tag{B.4}$$

$$y_1 + y_2 \leq \tilde{x}, \tag{B.5}$$

$$n_1, n_2, \in \mathbb{Z}^+, \tag{B.6}$$

$$y_1, y_2 \geq 0, \tag{B.7}$$

where the auxiliary variables y_1 and y_2 denote the number of units shipped via internal TL capacity using type-1 and type-2 trucks, respectively. Constraints (B.3) and (B.4) ensure the number of units shipped via TL shipments using each truck type does not exceed the aggregate internal capacity provided by that truck type. Constraint (B.5) ensures the total number of units shipped using internal trucks does not exceed the demand.

Suppose an optimal solution $(n_1^*, n_2^*, y_1^*, y_2^*)$ in which $n_2^* \geq q + 1$ exists, and the corresponding optimal objective equals $g^* = (K_1 + A_1)n_1^* + (K_2 + A_2)n_2^* + a(\tilde{x} - y_1^* - y_2^*)$. As $W_1 = qW_2$ and $A_1 \leq qA_2$, any amount shipped via q type-2 (small) trucks can be instead shipped using only one type-1 (big) truck at an equal or lower shipping cost, which leads to a new solution (n_1', n_2', y_1', y_2'), where $n_1' = n_1^* + 1$ and $n_2' = n_2^* - q$. Note that replacing trucks does not affect the total amount shipped via TLs, implying that $y_1' + y_2' = y_1^* + y_2^*$. The resulting objective value equals $g' = (K_1 + A_1)n_1' + (K_2 + A_1)n_2' + a(\tilde{x} - y_1' - y_2') = g^* + K_1 - qK_2 + A_1 - qA_2$.

Because $K_1 < qK_2$, it is straightforward to show that $g' < g^*$, implying g' is a better objective value than g^* for the observed demand $\tilde{x}$. We can further reduce the objective value by replacing q type-2 trucks with one type-1 truck until the number of remaining type-2 trucks is less than or equal to q. Therefore, $n_2^* \geq q+1$ cannot be optimal, and an optimal solution for any given demand observation must have $n_2^* \leq q$.

In fact, noting that the expected cost with stochastic demand is a weighted combination of costs weighted by probabilities associated with the corresponding demand observations, the above discussion is applicable to any demand observation and its associated MIP model. Because $n_2^* \leq q$ is optimal for all possible demand observations, it must therefore also hold for any solution that minimizes the expected cost. ■

Proof of Proposition 2.4.2. Suppose an optimal solution exists for a demand observation $\tilde{x} < n_1^* W_1$ in which two type-2 trucks are owned and both of them are dispatched without using all type-1 trucks. Let k_1^* denote the number of actual shipments via type-1 and $k_1^* < n_1^*$. The optimal objective then equals $g^* = n_1^* K_1^* + k_1^* A_1 + 2K_2 + 2A_2 + a(\tilde{x} - y_1^* - y_2^*)$. We can apply the same capacity replacement operation in the proof of Theorem 2.4.1 by substituting two type-2 trucks with one type-1 truck, which leads to a new objective value $g' = n_1^* K_1^* + (k_1^* + 1)A_1 + 2K_2 + a(\tilde{x} - y_1')$, where $y_1' = y_1^* + y_2^*$ (because the total number of units shipped via TLs does not change, and no type 2 truck is dispatched after capacity replacement). As $A_1 \leq 2A_2$, it is straightforward to show that $g' \leq g^*$, implying either g^* is an alternative optimal solution, or there is a contradiction. Therefore, for any demand $\tilde{x} < n_1^* W_1$, an optimal solution exists in which we do not dispatch both type-2 trucks before all type-1 trucks have been dispatched. ■

Proof of Proposition 2.4.3. Because $q_L = \lfloor A_1/A_2 \rfloor$, it is straightforward to show $q_L A_2 \leq A_1 < (q_L + 1)A_2$, implying $q_L Y_2 \leq Y_1 < (q_L + 1)Y_2$. Given $Y_{21} = Y_1 + q_L(W_2 - Y_2)$, we can show that $q_L W_2 \leq Y_{21} < (q_L + 1)W_2$. Thus, triggering the dispatch of an additional type-1 truck always occurs after shipping q_L type-2 trucks as long as there is at least one type-1 truck available. Once all type-1 trucks are dispatched (prior to this, q_L type-2 trucks have already been dispatched), we no longer need to consider Y_{21}. The minimum shipping cost is further determined by only considering the remaining $\tilde{q} - q_L$ type-2 trucks and the corresponding break-even point Y_2. ■

B.4 The expected shipping quantities for uniform and normal demand distributions

B.4.1 Uniform distribution

For a uniformly distributed demand variable in $[l, \bar{D}]$ with PDF $f(x) = 1/(\bar{D} - l)$ and CDF $F(x) = (x - l)/(\bar{D} - l)$, we can obtain the following expected shipping quantities:

$$E_{TL1}^{(0)}(n_1) = n_1 \left(\frac{\bar{D} - Y_1}{\bar{D} - l} - \frac{W_1(n_1 - 1)}{2(\bar{D} - l)} \right), \tag{B.8}$$

$$E_{TL1}^{(1)}(n_1) = n_1 \left(\frac{\bar{D} - Y_{21}}{\bar{D} - l} - \frac{W_1(n_1 - 1)}{2(\bar{D} - l)} \right), \tag{B.9}$$

$$E_{TL1}^{(2)}(n_1) = E_{TL1}^{(1)}(n_1), \tag{B.10}$$

$$E_{TL2}^{(1)}(n_1) = \frac{\bar{D} - n_1(W_1 + Y_2 - Y_{21}) - Y_2}{\bar{D} - l}, \tag{B.11}$$

$$E_{TL2}^{(2)}(n_1) = \frac{2\bar{D} - n_1(2W_1 - Y_{21} + Y_2) - W_2 - 2Y_2}{\bar{D} - l}, \tag{B.12}$$

$$E_{LTL}^{(0)}(n_1) = \frac{n_1 Y_1^2 + (\bar{D} - n_1 W_1)^2}{2(\bar{D} - l)}, \tag{B.13}$$

$$E_{LTL}^{(1)}(n_1) = \frac{(n_1 + 1)Y_2^2 + n_1(Y_{21} - W_2)^2 + (\bar{D} - n_1 W_1 - W_2)^2}{2(\bar{D} - l)}, \tag{B.14}$$

$$E_{LTL}^{(2)}(n_1) = \frac{(n_1 + 2)Y_2^2 + n_1(Y_{21} - W_2)^2 + (\bar{D} - n_1 W_1 - 2W_2)^2}{2(\bar{D} - l)}. \tag{B.15}$$

According to Table 2.4, we can obtain the threshold values of n_1 by performing pairwise comparison as follows:

$$g_0(n_1) \leq g_1(n_1) \iff n_1 \geq \frac{(W_2 - Y_2)(2\bar{D} - W_2 - Y_2) - 2K_2(\bar{D} - l)/a}{2(W_2 - Y_2)(W_1 - Y_1 + Y_2)} \equiv N_{01}, \tag{B.16}$$

$$g_0(n_1) \leq g_2(n_1) \iff n_1 \geq \frac{(W_2 - Y_2)(2\bar{D} - 2W_2 - Y_2) - 2K_2(\bar{D} - l)/a}{(W_2 - Y_2)(2W_1 - Y_1 + Y_2)} \equiv N_{02}, \tag{B.17}$$

$$g_1(n_1) \leq g_2(n_1) \iff n_1 \geq \frac{(W_2 - Y_2)(2\bar{D} - 3W_2 - Y_2) - 2K_2(\bar{D} - l)/a}{2W_1(W_2 - Y_2)} \equiv N_{12}. \tag{B.18}$$

If $n_1^* \geq \max\{N_{01}, N_{02}\}$, not owning type-2 trucks is optimal. If $N_{01} \geq N_{12}$ holds and $n_1^* \in [N_{12}, N_{01}]$, owning one type-2 truck leads to an optimal solution. If $n_1^* \leq \min\{N_{02}, N_{12}\}$, owning two type-2 trucks is optimal.

B.4.2 Normal distribution

For normally distributed demand with mean μ and standard deviation σ, we can obtain the expected shipping quantities as follows:

$$E_{TL1}^{(0)}(n_1) = n_1 - \sum_{k=0}^{n_1-1} \Phi(kW_1 + Y_1), \tag{B.19}$$

$$E_{TL1}^{(1)}(n_1) = n_1 - \sum_{k=0}^{n_1-1} \Phi(kW_1 + Y_{21}), \tag{B.20}$$

$$E_{TL1}^{(2)}(n_1) = E_{TL1}^{(1)}(n_1), \tag{B.21}$$

$$E_{TL2}^{(1)}(n_1) = \sum_{k=0}^{n_1-1} \Big(\Phi(kW_1 + Y_{21}) - \Phi(kW_1 + Y_2)\Big) + \Big(1 - \Phi(n_1W_1 + Y_2)\Big), \tag{B.22}$$

$$E_{TL2}^{(2)}(n_1) = \sum_{k=0}^{n_1-1} \Big(\Phi(kW_1 + Y_{21}) - \Phi(kW_1 + Y_2)\Big) - \Phi(n_1W_1 + W_2 + Y_2) - \Phi(n_1W_1 + Y_2) + 2, \tag{B.23}$$

$$E_{LTL}^{(0)}(n_1) = \sum_{k=0}^{n_1-1} \Big(\sigma L(z_{kW_1}) - \sigma L(z_{kW_1+Y_1}) - Y_1\big(1 - \Phi(z_{kW_1+Y_1})\big)\Big) + \sigma L(z_{n_1W_1}), \tag{B.24}$$

$$\begin{aligned} E_{LTL}^{(1)}(n_1) = &\sum_{k=0}^{n_1} \Big(\sigma L(z_{kW_1}) - \sigma L(z_{kW_1+Y_2}) - Y_2\big(1 - \Phi(z_{kW_1+Y_2})\big)\Big) + \sigma L(z_{n_1W_1+W_2}) \\ &+ \sum_{k=0}^{n_1-1} \Big(\sigma L(z_{kW_1+W_2}) - \sigma L(z_{kW_1+Y_{21}}) + (W_2 - Y_{21})\big(1 - \Phi(z_{kW_1+Y_{21}})\big)\Big), \end{aligned} \tag{B.25}$$

$$\begin{aligned} E_{LTL}^{(2)}(n_1) = &\sum_{k=0}^{n_1} \Big(\sigma L(z_{kW_1}) - \sigma L(z_{kW_1+Y_2}) - Y_2\big(1 - \Phi(z_{kW_1+Y_2})\big)\Big) + \sigma L(z_{n_1W_1+W_2}) \\ &+ \sum_{k=0}^{n_1-1} \Big(\sigma L(z_{kW_1+W_2}) - \sigma L(z_{kW_1+Y_{21}}) + (W_2 - Y_{21})\big(1 - \Phi(z_{kW_1+Y_{21}})\big)\Big) \\ &- \sigma L(z_{n_1W_1+W_2+Y_2}) - Y_2\big(1 - \Phi(z_{n_1W_1+W_2+Y_2})\big) + \sigma L(z_{n_1W_1+2W_2}). \end{aligned} \tag{B.26}$$

B.5 Proof of NP-hardness of problem $\mathbb{DEP}$

Let us consider the extreme case when $a = \infty$ and $K_t = 0$ for $t = 1, \ldots, T$, which forces all units to be shipped via internal TL capacity. In this case, constraints (2.33) are no longer necessary due to zero fixed ownership costs. Thus, the resulting model for each i is equivalent to

$$\text{Maximize} \quad \sum_{t=1}^{T} v_t m_t^i \tag{B.27}$$

$$\text{Subject to:} \quad f_t m_t^i \leq h^i, \quad t = 1, \ldots, T, \tag{B.28}$$

$$m_t^i \in \mathbb{Z}^+, \quad t = 1, \ldots, T, \tag{B.29}$$

where $v_t = -A_t$, $f_t = -W_t$ for $t = 1, \ldots, T$ and $h^i = -d^i$. This model is an integer knapsack problem and is thus $\mathcal{NP}$-hard, and the $\mathbb{DEP}$ model is consequently $\mathcal{NP}$-hard.

B.6 SAA results for all instances with $S = 1000$ and $T = 3$

Table B.1 includes the SAA results for Instances 1 to 25. Results for Instances 25 to 50 are displayed in Table B.2. Instances that reach the maximum time limit are starred.

B.7 SAA solutions and the DB algorithm results

Table B.3 compares the SAA results with the results obtained by using the DB algorithm for all instances, in which instances that have improved runtime when using the DB heuristic are starred. Tables B.4 through B.8 include the comparison results for the variants of the selected 20 instances with different values of T and S.

B.8 Sensitivity analysis results for different LTL rates

Tables B.9 and B.10 summarize the metrics for all instances with different LTL rates.

B.9 Sensitivity analysis results for different numbers of truck types

Tables B.11 and B.12 summarize the metrics for all instances with different numbers of truck types.

B.10 Sensitivity analysis results for different demand parameters

Tables B.13 and B.14 summarize the results for all instances with different demand mean values and standard deviations.

B.11 Sensitivity analysis results for different demand distributions

Tables B.15 and B.16 include the metrics for all instances when changing the demand distributions.

Table B.1: SAA solutions for Instances 1 to 25 with $S = 1000$ and $T = 3$.

	Instance	Truck Class (t)	W_t (lbs)	a ($/lb) ($)	K_t ($)	A_t ($)	n_t	$\bar{m}_t^s$	$\bar{y}_t^s$	LTL units	Obj.Val.	Obj.LB.	GRB gap (%)	GRB time (s)
	1	Class 8 (1)	35600	0.31	1232.14	5849	32	26.91	957644.61	38969.50	219611.31	219460.03	0.07	9.95
		Class 7 (2)	28600		862.99	5054	3	1.61	45476.69					
		Class 6 (3)	24000		643.40	4687	0	0.00	0.00					
	2	Class 8 (1)	37600	0.32	1305.06	7099	2	1.32	49481.60	37108.39	241260.39	241021.49	0.10	789.33
		Class 7 (2)	30700		838.75	5894	38	31.48	965458.52					
		Class 6 (3)	24700		727.54	5217	0	0.00	0.00					
	3	Class 8 (1)	39700	0.39	1225.17	8361	26	22.50	893212.05	25738.05	268889.38	268620.50	0.10	811.34
		Class 7 (2)	32700		939.21	7014	8	4.47	145415.39					
		Class 6 (3)	20600		684.92	4579	0	0.00	0.00					
*	4	Class 8 (1)	33300	0.25	1273.44	4662	27	24.92	829467.96	57199.50	198683.05	198436.74	0.12	18000.15
		Class 7 (2)	30300		998.37	4469	9	5.56	168275.49					
		Class 6 (3)	23500		653.28	3936	0	0.00	0.00					
*	5	Class 8 (1)	34000	0.29	1180.66	5817	20	18.97	644702.61	38554.01	228355.43	227889.52	0.20	18003.90
		Class 7 (2)	29900		796.46	5376	19	12.67	378437.65					
		Class 6 (3)	21000		687.35	3898	0	0.00	0.00					
*	6	Class 8 (1)	36000	0.24	1243.81	4320	20	18.98	682910.18	42976.73	170019.04	169758.17	0.15	18001.16
		Class 7 (2)	27300		774.77	3473	18	11.20	305101.12					
		Class 6 (3)	19500		681.19	2574	0	0.00	0.00					
	7	Class 8 (1)	33400	0.35	1270.82	5962	0	0.00	0.00	25266.40	230334.67	230151.07	0.08	52.23
		Class 7 (2)	31700		863.98	5769	40	31.43	995283.51					
		Class 6 (3)	22000		595.48	4389	2	1.01	22071.80					
	8	Class 8 (1)	39600	0.38	1251.67	7674	32	25.44	1006455.74	22991.87	251774.80	251579.54	0.08	16.39
		Class 7 (2)	28400		857.16	5936	2	1.02	28600.87					
		Class 6 (3)	24800		655.34	6031	0	0.00	0.00					
*	9	Class 8 (1)	34000	0.37	1237.17	6416	32	27.89	947997.86	27792.59	248981.52	248694.12	0.12	18028.98
		Class 7 (2)	26100		934.28	5022	1	0.59	15218.43					
		Class 6 (3)	23600		652.66	4977	7	2.36	55312.28					
	10	Class 8 (1)	37000	0.23	1294.47	4681	31	26.57	980824.61	65650.48	179616.39	179461.84	0.09	0.54
		Class 7 (2)	32400		887.22	5067	0	0.00	0.00					
		Class 6 (3)	21600		712.49	3378	0	0.00	0.00					
	11	Class 8 (1)	35400	0.34	1322.22	6499	32	27.64	977692.72	36648.62	247775.15	247585.03	0.08	15.34
		Class 7 (2)	31900		925.31	6616	0	0.00	0.00					
		Class 6 (3)	24800		623.96	5396	5	1.90	46699.18					
	12	Class 8 (1)	37600	0.25	1168.53	5264	30	25.36	953320.11	47115.87	188483.94	188307.97	0.09	33.85
		Class 7 (2)	26400		858.21	3696	1	0.62	16313.12					
		Class 6 (3)	21100		649.43	3007	2	1.23	25742.47					
	13	Class 8 (1)	34900	0.36	1227.94	6282	33	27.69	965946.08	26091.82	237167.37	236964.90	0.09	10.74
		Class 7 (2)	28300		831.12	5603	4	1.25	35298.33					
		Class 6 (3)	20400		677.45	4113	1	0.56	11016.67					
	14	Class 8 (1)	37100	0.31	1204.92	5750	33	26.78	992097.15	34641.63	207812.61	207708.82	0.05	10.57
		Class 7 (2)	27100		833.20	4789	1	0.52	13578.48					
		Class 6 (3)	20600		655.88	4151	0	0.00	0.00					
	15	Class 8 (1)	36600	0.36	1290.59	6983	25	22.70	830658.75	27689.23	250140.20	249894.31	0.10	23.26
		Class 7 (2)	29100		959.73	5657	10	5.80	168498.45					
		Class 6 (3)	26000		736.78	5522	3	0.86	22368.41					
	16	Class 8 (1)	37300	0.32	1184.07	6565	34	27.51	1023238.02	37903.29	233010.28	232851.77	0.07	0.64
		Class 7 (2)	29100		976.77	6518	0	0.00	0.00					
		Class 6 (3)	23100		681.60	5322	0	0.00	0.00					
	17	Class 8 (1)	39600	0.38	1169.39	7975	31	25.21	997951.73	25402.39	254527.92	254301.96	0.09	10.06
		Class 7 (2)	31000		878.70	6597	1	0.52	15922.90					
		Class 6 (3)	22000		663.58	5016	1	0.52	10985.24					
	18	Class 8 (1)	33600	0.41	1277.28	7852	35	29.54	991572.17	34308.00	300175.38	300062.14	0.04	9.60
		Class 7 (2)	31600		802.45	8681	2	0.50	15819.71					
		Class 6 (3)	20600		734.60	5743	1	0.49	9699.68					
	19	Class 8 (1)	38600	0.38	1316.68	7921	0	0.00	0.00	22135.39	256423.62	256212.56	0.08	38.72
		Class 7 (2)	30900		818.43	6341	41	32.25	996003.42					
		Class 6 (3)	25500		659.52	5523	3	1.45	36628.92					
	20	Class 8 (1)	38600	0.41	1187.35	9654	32	25.78	994482.83	26764.32	307374.23	307124.27	0.08	3.74
		Class 7 (2)	26300		850.56	6685	2	1.17	30473.44					
		Class 6 (3)	22900		738.01	6572	0	0.00	0.00					
*	21	Class 8 (1)	34300	0.40	1236.12	7272	8	6.61	226427.68	22328.07	271845.04	270984.55	0.32	18008.08
		Class 7 (2)	31300		871.28	6886	33	25.59	800612.49					
		Class 6 (3)	20200		606.64	4848	0	0.00	0.00					
	22	Class 8 (1)	36100	0.27	1283.12	5166	0	0.00	0.00	36206.82	192179.59	192131.12	0.03	43.34
		Class 7 (2)	31000		876.84	4436	40	32.57	1009043.65					
		Class 6 (3)	25300		698.33	4099	1	0.53	12424.40					
*	23	Class 8 (1)	39900	0.41	1249.60	9488	3	1.969	78175.076	22615.32	291552.68	291170.66	0.13	18013.99
		Class 7 (2)	32400		799.17	7838	37	29.38	951521.15					
		Class 6 (3)	23100		713.87	6535	0	0	0					
	24	Class 8 (1)	39000	0.39	1243.50	8061	30	25.30	986239.84	21202.83	266643.81	266451.85	0.07	46.59
		Class 7 (2)	28300		765.79	6291	6	1.99	55744.79					
		Class 6 (3)	22900		653.12	5805	0	0.00	0.00					
	25	Class 8 (1)	33700	0.39	1249.92	6703	2	1.31	43690.91	22064.95	254613.38	254396.57	0.09	32.69
		Class 7 (2)	28000		906.32	5569	44	34.99	978715.43					
		Class 6 (3)	21000		723.77	4914	0	0.00	0.00					

Table B.2: SAA solutions for Instances 26 to 50 with $S = 1000$ and $T = 3$.

	Instance	Truck Class (t)	W_t (lbs)	a ($/lb)	K_t ($)	A_t ($)	n_t	$\bar{m}_t^s$	$\bar{y}_t^s$	LTL units	Obj.Val.	Obj.LB.	GRB gap (%)	GRB time (s)
	26	Class 8 (1)	33700	0.35	1328.16	6015	0	0.00	0.00	18533.67	224691.81	224477.48	0.10	4.92
		Class 7 (2)	32000		785.08	5712	40	31.13	995775.53					
		Class 6 (3)	23700		623.53	4396	3	1.62	37981.84					
*	27	Class 8 (1)	34900	0.29	1324.66	5870	25	23.63	824647.20	39827.62	226044.06	225799.06	0.11	18045.75
		Class 7 (2)	27600		993.68	4722	1	0.72	19849.40					
		Class 6 (3)	21900		606.96	4001	14	7.44	162656.12					
	28	Class 8 (1)	38900	0.38	1194.71	7391	33	25.62	995187.56	24563.02	240940.82	240804.93	0.06	46.01
		Class 7 (2)	27400		999.44	5727	0	0.00	0.00					
		Class 6 (3)	19500		710.76	4150	1	0.51	9252.23					
	29	Class 8 (1)	36600	0.36	1221.17	6720	2	1.15	42090.00	23475.81	235531.11	235304.25	0.10	111.54
		Class 7 (2)	30400		853.58	5581	41	32.60	988999.59					
		Class 6 (3)	24500		733.65	4851	0	0.00	0.00					
	30	Class 8 (1)	38000	0.41	1277.53	8725	25	22.71	862756.36	22217.45	289372.50	289083.17	0.10	54.89
		Class 7 (2)	30300		781.07	7330	12	5.57	168097.69					
		Class 6 (3)	20000		623.89	5494	0	0.00	0.00					
	31	Class 8 (1)	33700	0.34	1292.97	6531	0	0.00	0.00	32900.08	243162.73	242919.59	0.10	5.85
		Class 7 (2)	32600		920.76	6318	36	29.60	964355.79					
		Class 6 (3)	25700		741.03	5155	3	1.86	47446.40					
	32	Class 8 (1)	34300	0.38	1318.97	6517	1	0.58	19992.84	26138.76	247026.45	246791.37	0.10	102.61
		Class 7 (2)	26100		880.29	4959	48	38.26	996817.86					
		Class 6 (3)	21000		722.25	4309	0	0.00	0.00					
*	33	Class 8 (1)	37300	0.37	1299.38	8557	2	1.37	50942.09	37537.58	286786.59	286450.78	0.12	18000.12
		Class 7 (2)	26100		840.78	6084	0	0.00	0.00					
		Class 6 (3)	22700		705.79	5291	51	42.07	954630.93					
	34	Class 8 (1)	35400	0.23	1330.84	4641	27	24.64	872035.41	63641.42	185686.83	185538.63	0.08	10.96
		Class 7 (2)	29100		899.57	4150	6	3.02	87894.11					
		Class 6 (3)	23600		686.68	3528	1	0.61	14020.30					
	35	Class 8 (1)	33900	0.33	1275.62	5817	0	0.00	0.00	25991.48	219318.52	219118.55	0.09	3.55
		Class 7 (2)	30700		818.07	5369	42	32.85	1005564.96					
		Class 6 (3)	20200		622.28	4333	0	0.00	0.00					
	36	Class 8 (1)	38600	0.39	1215.41	8280	32	26.00	1003172.94	27879.51	272455.06	272214.99	0.09	14.50
		Class 7 (2)	28400		965.94	6424	1	0.51	14379.63					
		Class 6 (3)	21600		613.98	5223	1	0.48	10309.99					
*	37	Class 8 (1)	39400	0.33	1283.81	7151	22.00	20.22	796380.27	30659.84	237560.49	237170.30	0.16	18000.27
		Class 7 (2)	32100		826.91	6144	13.00	7.14	228656.83					
		Class 6 (3)	25200		701.28	5572	0.00	0.00	0.00					
	38	Class 8 (1)	36600	0.30	1227.90	5490	33	27.21	994989.28	37064.35	204269.02	204137.73	0.06	9.92
		Class 7 (2)	26100		787.00	4463	1	0.55	13318.49					
		Class 6 (3)	19600		747.19	3528	0	0.00	0.00					
	39	Class 8 (1)	39100	0.41	1173.41	8176	33	26.11	1018915.89	24589.89	265535.54	265397.47	0.05	42.72
		Class 7 (2)	27900		868.17	6635	0	0.00	0.00					
		Class 6 (3)	21000		655.08	5080	1	0.52	10861.57					
	40	Class 8 (1)	36000	0.23	1184.72	4388	29	25.76	926764.63	53529.16	171428.26	171274.54	0.09	16.14
		Class 7 (2)	27600		820.68	3555	4	2.37	65237.79					
		Class 6 (3)	23200		629.10	3202	0	0.00	0.00					
	41	Class 8 (1)	39900	0.29	1287.88	6711	1	0.67	26593.61	34624.92	214897.47	214721.28	0.08	43.31
		Class 7 (2)	31300		750.59	5355	39	31.71	991306.52					
		Class 6 (3)	24000		646.35	4733	0	0.00	0.00					
	42	Class 8 (1)	39100	0.32	1332.87	6882	31	25.48	995146.23	40989.27	233021.47	232953.01	0.03	12.37
		Class 7 (2)	31900		911.01	6431	0	0.00	0.00					
		Class 6 (3)	23500		683.51	4963	1	0.52	11632.02					
*	43	Class 8 (1)	34300	0.35	1244.12	6363	3	2.13	72914.32	31947.96	244980.43	244556.82	0.17	18000.11
		Class 7 (2)	28000		802.31	5390	42	33.92	949163.80					
		Class 6 (3)	25200		642.03	5645	0	0.00	0.00					
	44	Class 8 (1)	38300	0.23	1320.39	4669	24	22.41	858039.33	45010.00	175511.65	175336.71	0.10	23.39
		Class 7 (2)	31400		839.11	4189	9	4.56	142946.07					
		Class 6 (3)	20000		600.66	2714	1	0.58	11349.37					
	45	Class 8 (1)	40000	0.38	1285.91	7752	24	21.80	871291.40	27971.13	247553.72	247339.78	0.09	52.10
		Class 7 (2)	26900		801.68	5315	12	5.16	138276.44					
		Class 6 (3)	24500		727.21	5121	0	0.00	0.00					
	46	Class 8 (1)	35600	0.32	1175.27	5924	34	27.77	988346.68	31374.31	220507.30	220373.96	0.06	10.28
		Class 7 (2)	27300		853.01	4892	1	0.51	13729.19					
		Class 6 (3)	19600		605.41	3826	1	0.53	10109.28					
*	47	Class 8 (1)	37900	0.38	1211.27	8497	2	1.37	51594.88	32865.32	276844.97	276431.99	0.15	18000.15
		Class 7 (2)	29300		807.50	6680	40	32.64	955840.16					
		Class 6 (3)	20000		617.13	4712	0	0.00	0.00					
	48	Class 8 (1)	34700	0.39	1326.22	6902	32	27.47	952757.85	26883.00	261676.96	261464.56	0.08	26.59
		Class 7 (2)	27600		854.79	6028	6	1.83	50417.43					
		Class 6 (3)	19500		603.95	4411	1	0.54	10317.92					
*	49	Class 8 (1)	37300	0.42	1258.12	7833	26	23.05	859259.11	21313.02	269013.53	268707.55	0.11	18000.11
		Class 7 (2)	29700		853.97	6486	12	5.63	166919.32					
		Class 6 (3)	20200		594.77	4581	0	0.00	0.00					
	50	Class 8 (1)	38900	0.39	1301.82	8647	28	23.94	930850.51	29265.87	279543.29	279288.05	0.09	3.06
		Class 7 (2)	31100		854.59	7399	6	2.65	81677.41					
		Class 6 (3)	21400		738.83	6343	0	0.00	0.00					

Table B.3: SAA and DB results for Instances 1-50 with $T = 3$ and $S = 1000$.

	Instance	SAA $\hat{\mathbf{n}}$	SAA obj.val.	SAA time (s)	DB $\mathbf{n}^{RD}$	DB obj.val.	DB time (s)	$\left(1 - \frac{\text{DB obj.val.}}{\text{SAA obj.val.}}\right) \times 100\%$	$\left(1 - \frac{\text{DB time}}{\text{SAA time}}\right) \times 100\%$
	1	[32, 3, 0]	219611.31	9.95	[32, 3, 0]	219611.32	21.91	0.00	-120.19
*	2	[2, 38, 0]	241260.39	789.33	[0, 41, 0]	241409.87	19.37	-0.06	97.55
*	3	[26, 8, 0]	268889.38	811.34	[27, 8, 0]	269010.37	24.74	-0.04	96.95
*	4	[27, 9, 0]	198683.05	18000.15	[28, 9, 0]	198861.05	22.85	-0.09	99.87
*	5	[20, 19, 0]	228355.43	18003.90	[21, 19, 0]	228466.50	21.82	-0.05	99.88
*	6	[20, 18, 0]	170019.04	18001.16	[21, 17, 0]	170035.98	20.40	-0.01	99.89
*	7	[0, 40, 2]	230334.67	52.23	[0, 42, 0]	230760.12	21.39	-0.18	59.05
	8	[32, 2, 0]	251774.80	16.39	[33, 0, 0]	252305.08	21.38	-0.21	-30.45
*	9	[32, 1, 7]	248981.52	18028.98	[33, 0, 8]	249209.48	22.19	-0.09	99.88
	10	[31, 0, 0]	179616.39	0.54	[32, 0, 0]	179689.61	18.03	-0.04	-3208.82
	11	[32, 0, 5]	247775.15	15.34	[33, 0, 5]	247824.58	21.87	-0.02	-42.55
*	12	[30, 1, 2]	188483.94	33.85	[32, 0, 0]	189065.22	20.43	-0.31	39.66
	13	[33, 4, 1]	237167.37	10.74	[34, 5, 0]	237490.71	20.91	-0.14	-94.71
	14	[33, 1, 0]	207812.61	10.57	[34, 0, 0]	207991.27	19.41	-0.09	-83.61
	15	[25, 10, 3]	250140.20	23.26	[27, 9, 3]	250375.88	24.24	-0.09	-4.21
	16	[34, 0, 0]	233010.28	0.64	[35, 0, 0]	233132.95	20.93	-0.05	-3169.57
	17	[31, 1, 1]	254527.92	10.06	[33, 0, 0]	255278.64	21.51	-0.29	-113.80
	18	[35, 2, 1]	300175.38	9.60	[36, 4, 0]	300602.14	21.93	-0.14	-128.37
*	19	[0, 41, 3]	256423.62	38.72	[0, 44, 0]	256894.64	20.68	-0.18	46.58
	20	[32, 2, 0]	307374.23	3.74	[34, 0, 0]	308307.03	21.15	-0.30	-464.92
*	21	[8, 33, 0]	271845.04	18008.08	[0, 42, 0]	272218.64	21.34	-0.14	99.88
*	22	[0, 40, 1]	192179.59	43.34	[0, 41, 0]	192181.16	20.91	0.00	51.76
*	23	[3, 37, 0]	291552.68	18013.99	[0, 41, 0]	292020.33	20.70	-0.16	99.89
*	24	[30, 6, 0]	266643.81	46.59	[31, 6, 0]	266830.58	22.67	-0.07	51.33
*	25	[2, 44, 0]	254613.38	32.69	[0, 46, 0]	254895.54	21.08	-0.11	35.53
	26	[0, 40, 3]	224691.81	4.92	[0, 43, 0]	225265.65	20.05	-0.26	-307.77
*	27	[25, 1, 14]	226044.06	18045.75	[27, 0, 14]	226288.31	22.16	-0.11	99.88
*	28	[33, 0, 1]	240940.82	46.01	[34, 0, 0]	241423.22	21.47	-0.20	53.32
*	29	[2, 41, 0]	235531.11	111.54	[0, 44, 0]	235803.60	21.09	-0.12	81.09
*	30	[25, 12, 0]	289372.50	54.89	[26, 12, 0]	289531.71	23.68	-0.06	56.85
	31	[0, 36, 3]	243162.73	5.85	[0, 39, 0]	243553.97	21.47	-0.16	-267.04
*	32	[1, 48, 0]	247026.45	102.61	[0, 50, 0]	247338.04	21.37	-0.13	79.18
*	33	[2, 0, 51]	286786.59	18000.12	[0, 0, 55]	287160.73	21.19	-0.13	99.88
	34	[27, 6, 1]	185686.83	10.96	[28, 7, 0]	185866.66	20.01	-0.10	-82.59
	35	[0, 42, 0]	219318.52	3.55	[0, 42, 0]	219318.52	19.42	0.00	-446.52
	36	[32, 1, 1]	272455.06	14.50	[34, 0, 0]	273091.85	21.37	-0.23	-47.41
*	37	[22, 13, 0]	237560.49	18000.27	[24, 12, 0]	237642.43	22.85	-0.03	99.87
	38	[33, 1, 0]	204269.02	9.92	[35, 0, 0]	204659.92	20.77	-0.19	-109.30
*	39	[33, 0, 1]	265535.54	42.72	[34, 0, 0]	266083.82	21.93	-0.21	48.65
	40	[29, 4, 0]	171428.26	16.14	[30, 4, 0]	171442.73	21.34	-0.01	-32.15
*	41	[1, 39, 0]	214897.47	43.31	[0, 41, 0]	215012.56	20.64	-0.05	52.35
	42	[31, 0, 1]	233021.47	12.37	[32, 0, 0]	233282.65	20.74	-0.11	-67.73
*	43	[3, 42, 0]	244980.43	18000.11	[0, 47, 0]	245442.35	21.47	-0.19	99.88
*	44	[24, 9, 1]	175511.65	23.39	[25, 10, 0]	175751.47	22.20	-0.14	5.10
*	45	[24, 12, 0]	247553.72	52.10	[25, 12, 0]	247691.17	24.07	-0.06	53.80
	46	[34, 1, 1]	220507.30	10.28	[36, 0, 0]	221003.95	20.14	-0.23	-95.96
*	47	[2, 40, 0]	276844.97	18000.15	[0, 44, 0]	277453.54	21.48	-0.22	99.88
*	48	[32, 6, 1]	261676.96	26.59	[33, 7, 0]	262049.14	22.94	-0.14	13.71
*	49	[26, 12, 0]	269013.53	18000.11	[27, 12, 0]	269143.45	23.57	-0.05	99.87
	50	[28, 6, 0]	279543.29	3.06	[29, 6, 0]	279554.21	22.14	0.00	-623.05

Table B.4: SAA and DB results for selected instances with $T = 4$ and $S = 1000$.

Instance	SAA $\hat{\mathbf{n}}$	SAA obj.val.	SAA time (s)	DB $\mathbf{n}^{RD}$	DB obj.val.	DB time (s)	$\left(1 - \frac{\text{DB obj.val.}}{\text{SAA obj.val.}}\right) \times 100\%$	$\left(1 - \frac{\text{DB time}}{\text{SAA time}}\right) \times 100\%$
1	[31, 1, 0, 7]	219282.40	15.17	[32, 0, 0, 8]	219478.93	23.71	-0.09	-56.32
8	[32, 2, 0, 0]	251774.80	17.75	[33, 0, 0, 0]	252305.08	21.85	-0.21	-23.11
10	[31, 0, 0, 1]	179526.51	1.10	[32, 0, 0, 0]	179689.61	21.99	-0.09	-1896.46
11	[32, 0, 1, 6]	247515.04	22.10	[33, 0, 0, 7]	247661.37	21.94	-0.06	0.71
13	[33, 4, 0, 1]	237145.98	19.80	[34, 5, 0, 0]	237490.71	25.43	-0.15	-28.41
14	[33, 1, 0, 1]	207760.19	25.08	[34, 0, 0, 0]	207991.27	20.99	-0.11	16.31
16	[34, 0, 0, 0]	233010.28	0.74	[35, 0, 0, 0]	233132.95	20.73	-0.05	-2713.90
17	[31, 1, 1, 0]	254519.43	15.47	[33, 0, 0, 0]	255278.64	23.65	-0.30	-52.89
18	[35, 3, 1, 0]	300177.35	18.37	[36, 4, 0, 0]	300602.14	26.18	-0.14	-42.50
20	[32, 1, 0, 1]	307404.90	18.01	[34, 0, 0, 0]	308307.03	24.61	-0.29	-36.63
26	[0, 41, 1, 1]	224659.47	19.97	[0, 43, 0, 0]	225265.65	25.33	-0.27	-26.81
31	[0, 33, 0, 11]	242646.71	54.75	[0, 33, 0, 12]	242681.67	27.63	-0.01	49.54
34	[27, 3, 0, 6]	185546.44	15.35	[28, 3, 0, 6]	185603.96	26.42	-0.03	-72.16
35	[0, 42, 0, 0]	219318.52	3.68	[0, 42, 0, 0]	219318.52	24.26	0.00	-558.35
36	[32, 1, 0, 1]	272470.91	23.31	[34, 0, 0, 1]	272677.32	26.12	-0.08	-12.05
38	[33, 1, 0, 1]	204170.28	14.85	[35, 0, 0, 0]	204659.92	24.60	-0.24	-65.71
40	[29, 5, 0, 0]	171423.08	18.11	[30, 4, 0, 0]	171442.73	25.96	-0.01	-43.41
42	[30, 1, 0, 1]	232915.18	16.79	[32, 0, 0, 1]	233018.43	25.05	-0.04	-49.20
46	[34, 1, 0, 1]	220402.40	21.33	[36, 0, 0, 1]	220713.51	25.64	-0.14	-20.24
50	[28, 6, 0, 0]	279526.78	3.73	[29, 6, 0, 0]	279554.21	27.20	-0.01	-628.68
Avg.			17.27			24.46	-0.12	-313.01

Table B.5: SAA and DB results for selected instances with $T = 3$ and $S = 3000$.

Instance	SAA $\hat{\mathbf{n}}$	SAA obj.val.	SAA time (s)	DB $\mathbf{n}^{RD}$	DB obj.val.	DB time (s)	$\left(1 - \frac{\text{DB obj.val.}}{\text{SAA obj.val.}}\right) \times 100\%$	$\left(1 - \frac{\text{DB time}}{\text{SAA time}}\right) \times 100\%$
1	[32, 4, 0]	221898.62	59.27	[33, 4, 0]	222065.36	77.52	-0.08	-30.79
8	[32, 2, 0]	251916.64	115.02	[33, 0, 0]	252478.86	69.93	-0.22	39.20
10	[31, 0, 1]	181881.69	5.68	[32, 0, 0]	181885.15	63.59	0.00	-1019.48
11	[32, 0, 5]	246175.70	67.35	[33, 0, 5]	246263.50	70.82	-0.04	-5.15
13	[33, 4, 1]	239898.51	77.59	[34, 5, 0]	240143.50	76.77	-0.10	1.05
14	[33, 1, 0]	209637.14	86.59	[35, 0, 0]	209982.73	71.07	-0.16	17.92
16	[34, 0, 0]	230213.13	12.77	[34, 0, 0]	230213.13	69.27	0.00	-442.60
17	[31, 3, 0]	254916.50	80.00	[33, 0, 0]	255553.77	68.17	-0.25	14.80
18	[35, 3, 1]	301568.00	80.40	[36, 4, 0]	301910.99	77.25	-0.11	3.93
20	[31, 2, 0]	303982.12	52.89	[33, 0, 0]	304730.54	71.10	-0.25	-34.43
26	[0, 40, 2]	224003.18	51.66	[0, 42, 0]	224520.35	73.62	-0.23	-42.52
31	[0, 37, 3]	244632.36	52.63	[0, 40, 0]	245110.50	68.93	-0.20	-30.98
34	[28, 4, 1]	186947.97	55.14	[29, 6, 0]	187213.23	72.09	-0.14	-30.74
35	[0, 42, 0]	223644.94	114.11	[0, 43, 0]	223693.73	70.81	-0.02	37.95
36	[33, 0, 1]	271718.63	74.61	[34, 0, 0]	272266.78	69.33	-0.20	7.07
38	[33, 1, 0]	206986.14	86.22	[35, 0, 0]	207281.21	70.41	-0.14	18.34
40	[29, 5, 0]	172370.15	95.05	[30, 5, 0]	172442.32	76.43	-0.04	19.59
42	[31, 0, 1]	235071.21	83.55	[32, 0, 0]	235288.15	69.73	-0.09	16.53
46	[34, 2, 0]	223094.81	78.47	[36, 0, 0]	223482.00	69.85	-0.17	10.99
50	[29, 5, 0]	281368.25	36.84	[29, 6, 0]	281439.01	77.08	-0.03	-109.24
Avg.			68.29			71.69	-0.12	-77.93

Table B.6: SAA and DB results for selected instances with $T = 4$ and $S = 3000$.

Instance	SAA $\hat{\mathbf{n}}$	SAA obj.val.	SAA time (s)	DB $\mathbf{n}^{RD}$	DB obj.val.	DB time (s)	$\left(1 - \frac{\text{DB obj.val.}}{\text{SAA obj.val.}}\right) \times 100\%$	$\left(1 - \frac{\text{DB time}}{\text{SAA time}}\right) \times 100\%$
1	[31, 1, 0, 7]	221561.73	248.13	[32, 0, 0, 8]	221727.00	66.37	-0.07	73.25
8	[32, 2, 0, 0]	251916.64	128.34	[33, 0, 0, 0]	252478.86	71.27	-0.22	44.47
10	[31, 0, 0, 1]	181762.88	5.07	[32, 0, 0, 0]	181885.15	74.53	-0.07	-1369.44
11	[32, 0, 1, 6]	245938.33	129.22	[33, 0, 0, 7]	246097.23	82.69	-0.06	36.01
13	[33, 4, 1, 0]	239895.09	146.22	[34, 5, 0, 0]	240143.50	80.44	-0.10	44.99
14	[33, 1, 0, 1]	209591.59	242.41	[35, 0, 0, 0]	209982.73	74.76	-0.19	69.16
16	[33, 0, 0, 0]	230160.52	21.87	[34, 0, 0, 0]	230213.13	69.66	-0.02	-218.50
17	[31, 2, 0, 0]	254944.20	168.80	[33, 0, 0, 0]	255553.77	73.33	-0.24	56.56
18	[35, 3, 0, 1]	301533.46	163.34	[36, 4, 0, 0]	301910.99	78.68	-0.13	51.83
20	[32, 2, 0, 0]	303969.70	143.98	[33, 0, 0, 0]	304730.54	73.19	-0.25	49.17
26	[0, 40, 2, 0]	223988.40	214.26	[0, 42, 0, 0]	224520.35	81.18	-0.24	62.11
31	[0, 33, 0, 12]	244137.11	151.37	[0, 34, 0, 12]	244240.64	84.60	-0.04	44.11
34	[27, 4, 0, 4]	186825.79	146.50	[29, 3, 0, 5]	187008.07	83.22	-0.10	43.20
35	[0, 42, 0, 1]	223603.34	163.75	[0, 43, 0, 0]	223693.73	77.80	-0.04	52.49
36	[32, 1, 0, 2]	271647.45	143.27	[34, 0, 0, 1]	271831.70	69.04	-0.07	51.81
38	[33, 1, 0, 1]	206870.06	142.94	[35, 0, 0, 0]	207281.21	75.39	-0.20	47.26
40	[30, 4, 0, 0]	172374.94	135.66	[30, 5, 0, 0]	172442.32	77.61	-0.04	42.79
42	[31, 1, 0, 1]	234957.08	128.32	[32, 0, 0, 1]	235007.87	71.23	-0.02	44.49
46	[34, 1, 0, 1]	222957.23	173.12	[36, 0, 0, 1]	223182.33	68.24	-0.10	60.58
50	[28, 6, 0, 0]	281380.50	47.46	[29, 6, 0, 0]	281439.01	78.23	-0.02	-64.85
Avg.			142.20			75.57	-0.11	-38.93

Table B.7: SAA and DB results for selected instances with $T = 3$ and $S = 5000$.

Instance	SAA $\hat{\mathbf{n}}$	SAA obj.val.	SAA time (s)	DB $\mathbf{n}^{RD}$	DB obj.val.	DB time (s)	$\left(1 - \frac{\text{DB obj.val.}}{\text{SAA obj.val.}}\right) \times 100\%$	$\left(1 - \frac{\text{DB time}}{\text{SAA time}}\right) \times 100\%$
1	[32, 4, 0]	220429.73	175.71	[33, 4, 0]	220634.19	119.46	-0.09	32.02
8	[31, 2, 0]	249045.32	195.60	[33, 0, 0]	249668.70	113.83	-0.25	41.81
10	[31, 0, 0]	179582.61	51.11	[32, 0, 0]	179677.59	108.69	-0.05	-112.65
11	[32, 0, 5]	245106.79	202.23	[33, 0, 5]	245227.13	121.95	-0.05	39.70
13	[33, 4, 1]	238570.45	390.04	[34, 5, 0]	238890.68	114.82	-0.13	70.56
14	[33, 1, 0]	209722.84	229.30	[35, 0, 0]	210019.76	106.61	-0.14	53.51
16	[34, 0, 0]	231803.80	50.49	[34, 0, 0]	231803.80	115.39	0.00	-128.55
17	[31, 3, 0]	253783.21	181.65	[34, 0, 0]	254590.12	113.97	-0.32	37.26
18	[35, 2, 1]	298365.04	194.98	[36, 4, 0]	298858.16	125.44	-0.17	35.66
20	[32, 2, 0]	305674.56	186.02	[34, 0, 0]	306601.47	118.30	-0.30	36.41
26	[0, 40, 2]	222873.20	146.74	[0, 42, 0]	223381.31	118.60	-0.23	19.18
31	[0, 37, 3]	244162.06	119.12	[0, 39, 0]	244580.44	115.55	-0.17	3.00
34	[28, 5, 1]	186633.35	160.52	[29, 6, 0]	186894.43	120.55	-0.14	24.90
35	[0, 42, 0]	224399.14	299.82	[0, 43, 0]	224449.42	116.79	-0.02	61.05
36	[33, 0, 1]	271509.78	281.19	[34, 0, 0]	272095.22	117.03	-0.22	58.38
38	[33, 1, 0]	204839.48	236.30	[35, 0, 0]	205199.09	113.08	-0.18	52.15
40	[30, 4, 0]	172570.37	226.15	[31, 4, 0]	172667.28	122.72	-0.06	45.73
42	[31, 0, 1]	235026.85	212.33	[32, 0, 0]	235257.47	112.27	-0.10	47.13
46	[33, 3, 0]	221053.98	227.66	[36, 0, 0]	221501.71	116.78	-0.20	48.71
50	[29, 5, 0]	281810.25	88.49	[29, 6, 0]	281843.42	125.71	-0.01	-42.06
Avg.			192.77			116.88	-0.14	21.19

Table B.8: SAA and DB results for selected instances with $T = 4$ and $S = 5000$.

Instance	SAA $\hat{\mathbf{n}}$	SAA obj.val.	SAA time (s)	DB $\mathbf{n}^{RD}$	DB obj.val.	DB time (s)	$\left(1 - \frac{\text{DB obj.val.}}{\text{SAA obj.val.}}\right) \times 100\%$	$\left(1 - \frac{\text{DB time}}{\text{SAA time}}\right) \times 100\%$
1	[31, 1, 0, 7]	220074.38	275.26	[32, 0, 0, 8]	220272.31	127.54	-0.09	53.66
8	[31, 2, 0, 0]	249044.44	331.25	[33, 0, 0, 0]	249668.70	114.56	-0.25	65.42
10	[31, 0, 0, 1]	179505.29	110.85	[32, 0, 0, 0]	179677.59	113.95	-0.10	-2.80
11	[32, 0, 1, 6]	244865.08	491.37	[33, 0, 0, 8]	245168.38	131.21	-0.12	73.30
13	[33, 4, 1, 0]	238570.46	335.07	[34, 5, 0, 0]	238890.68	134.02	-0.13	60.00
14	[33, 1, 0, 1]	209665.81	496.95	[35, 0, 0, 0]	210019.76	116.54	-0.17	76.55
16	[33, 0, 0, 0]	231799.78	117.23	[34, 0, 0, 0]	231803.80	114.58	0.00	2.27
17	[31, 3, 0, 0]	253777.28	441.70	[34, 0, 0, 0]	254590.12	121.08	-0.32	72.59
18	[35, 2, 1, 0]	298359.54	366.43	[36, 4, 0, 0]	298858.16	129.19	-0.17	64.74
20	[32, 2, 0, 0]	305665.79	525.73	[34, 0, 0, 0]	306601.47	115.54	-0.31	78.02
26	[0, 40, 1, 1]	222831.41	872.03	[0, 42, 0, 0]	223381.31	108.49	-0.25	87.56
31	[0, 33, 0, 12]	243689.53	656.14	[0, 34, 0, 12]	243809.15	123.63	-0.05	81.16
34	[27, 4, 0, 4]	186491.33	449.82	[29, 3, 0, 6]	186802.38	120.96	-0.17	73.11
35	[0, 42, 0, 1]	224354.26	551.71	[0, 43, 0, 0]	224449.42	107.17	-0.04	80.57
36	[32, 1, 0, 2]	271406.17	705.11	[34, 0, 0, 1]	271660.07	115.41	-0.09	83.63
38	[33, 1, 0, 1]	204743.17	596.67	[35, 0, 0, 0]	205199.09	106.79	-0.22	82.10
40	[30, 4, 0, 0]	172570.21	357.21	[31, 4, 0, 0]	172667.28	117.72	-0.06	67.04
42	[31, 1, 0, 1]	234923.04	394.25	[32, 0, 0, 1]	234980.15	111.22	-0.02	71.79
46	[34, 1, 0, 1]	220903.41	678.60	[36, 0, 0, 1]	221203.49	113.16	-0.14	83.32
50	[29, 5, 0, 0]	281817.40	170.05	[29, 6, 0, 0]	281843.42	120.21	-0.01	29.31
Avg.			446.17			118.15	-0.14	64.17

Table B.9: Results for FCL, U_F, U_C, and LTL units with different LTL rates.

	FCL (lbs)			U_F (%)			U_C (%)			LTL units (%)		
Instance	Low LTL rate	Medium LTL rate	High LTL rate	Low LTL rate	Medium LTL rate	High LTL rate	Low LTL rate	Medium LTL rate	High LTL rate	Low LTL rate	Medium LTL rate	High LTL rate
1	36255.33	36071.94	35710.97	84.62	81.47	78.48	84.79	81.89	79.24	5.45	3.74	2.54
2	29394.00	29759.20	29741.81	86.45	82.00	79.04	86.38	81.73	78.81	6.52	3.53	2.41
3	34927.33	34979.41	35028.51	81.05	79.33	77.37	82.14	80.28	78.35	3.19	2.42	1.85
4	36194.28	34872.00	34275.23	93.17	84.66	80.29	93.29	85.15	81.42	13.79	5.42	3.44
5	33690.47	33139.64	33135.07	87.83	81.11	78.14	88.51	81.98	79.04	8.03	3.63	2.35
6	34187.25	33238.21	33228.60	86.80	79.42	76.09	88.33	81.56	78.57	9.02	4.17	2.89
7	29212.67	29057.24	29065.07	81.33	77.24	75.85	81.48	77.54	76.09	4.08	2.42	1.95
8	36469.75	36268.35	36280.63	81.04	77.82	76.04	81.21	78.18	76.33	3.55	2.17	1.64
9	34557.89	33999.60	34065.41	81.73	77.08	75.61	83.52	79.62	78.01	4.29	2.66	2.10
10	36698.00	36698.00	36275.18	93.84	85.72	81.19	93.79	85.51	81.35	15.43	6.27	3.50
11	36262.36	34755.30	33822.80	85.68	79.84	75.34	85.90	81.51	78.00	6.11	3.45	2.17
12	36184.57	35605.39	34701.33	89.66	82.46	78.43	89.84	83.18	79.70	10.07	4.52	2.88
13	36078.57	35550.84	35060.40	82.32	77.63	74.76	82.73	78.76	76.42	4.29	2.51	1.73
14	36698.00	36483.18	36280.63	85.10	80.30	78.90	84.93	80.36	79.19	5.94	3.33	2.70
15	34764.59	33640.95	33350.72	83.32	77.27	75.78	84.57	79.56	78.33	5.01	2.64	2.22
16	36698.00	36698.00	36298.67	84.02	80.92	77.91	83.90	80.68	78.06	5.51	3.57	2.29
17	36020.50	36041.03	36060.35	81.41	79.54	77.64	81.94	80.03	78.07	3.16	2.42	1.87
18	36698.00	35935.26	35608.20	84.60	80.33	77.24	84.48	80.73	77.75	5.38	3.26	2.19
19	29049.02	28911.82	28779.04	80.70	76.58	73.95	80.89	76.87	74.31	3.61	2.10	1.39
20	36241.50	36268.35	36280.63	82.98	79.26	77.46	83.16	79.59	77.79	4.20	2.54	1.91
21	30517.69	30641.02	30959.14	81.83	78.73	76.95	81.82	78.56	76.96	3.37	2.17	1.62
22	29394.00	29221.51	29229.53	86.04	80.72	77.97	85.92	80.73	77.91	6.60	3.42	2.26
23	29768.56	29941.80	29928.44	80.05	78.37	76.86	79.88	78.10	76.56	2.92	2.15	1.72
24	36034.00	35480.67	35316.16	80.76	75.81	74.52	81.69	77.77	76.53	3.52	1.99	1.53
25	29560.00	29711.57	29704.81	81.77	78.91	77.63	81.58	78.68	77.35	3.28	2.11	1.70
26	29040.40	28900.60	28911.82	80.43	76.16	74.74	80.57	76.51	75.06	3.22	1.76	1.35
27	33100.59	31483.80	31014.47	86.58	79.48	75.22	89.09	83.46	79.96	8.12	3.80	2.30
28	36248.75	36275.18	36287.26	80.46	76.85	75.09	80.65	77.07	75.26	3.89	2.39	1.85
29	29572.15	29733.72	29726.00	81.35	78.48	76.98	81.16	78.14	76.70	3.55	2.23	1.83
30	34611.14	34329.14	33888.77	79.54	76.42	73.80	81.33	78.48	76.00	3.20	2.11	1.39
31	29001.11	28850.00	28876.54	84.95	80.67	77.40	85.13	80.90	77.73	5.37	3.15	2.12
32	29549.40	29543.06	29398.55	81.66	79.27	76.87	81.45	79.00	76.79	3.52	2.51	1.77
33	23220.50	23427.85	22844.76	86.05	81.25	79.85	86.09	81.60	79.47	6.27	3.46	2.67
34	36698.00	34986.24	34532.81	95.52	83.14	78.96	95.51	84.40	80.56	21.07	6.13	3.71
35	29394.00	29394.00	29233.27	82.37	78.22	75.56	82.22	77.99	75.67	4.47	2.52	1.69
36	36248.75	36060.35	36287.26	82.78	79.40	77.44	83.08	79.98	77.63	4.08	2.64	1.93
37	34872.00	33985.09	33736.92	82.18	78.17	74.98	83.50	79.83	76.84	4.80	2.90	1.88
38	36469.75	36483.18	36095.78	84.86	81.65	78.61	85.04	81.72	79.13	5.46	3.55	2.37
39	36262.36	36275.18	36287.26	80.11	78.31	76.49	80.36	78.53	76.67	3.02	2.33	1.81
40	36698.00	35812.67	34742.76	91.74	85.25	78.79	91.66	85.93	80.88	10.92	5.12	2.60
41	29394.00	29576.60	29567.90	85.35	80.95	78.08	85.28	80.75	77.82	6.00	3.29	2.20
42	36698.00	36248.75	36060.35	86.00	81.23	77.74	85.87	81.48	78.20	6.86	3.91	2.52
43	29915.71	29880.93	30015.62	83.98	80.11	77.15	83.83	79.92	77.01	4.87	3.03	2.03
44	36136.15	34341.76	33940.59	91.86	81.03	76.41	92.19	82.86	78.60	13.06	4.26	2.22
45	34549.76	34263.33	34007.05	77.94	74.90	71.77	81.25	78.70	76.12	3.72	2.70	1.95
46	36476.67	36095.78	36112.05	84.39	80.04	78.32	84.50	80.50	78.80	5.41	3.01	2.38
47	29955.85	29741.81	29726.00	84.64	80.96	78.18	84.51	80.74	77.92	4.98	3.16	2.17
48	35690.00	35205.69	34922.20	81.07	76.52	73.59	82.10	78.23	75.68	4.23	2.58	1.76
49	34669.11	34391.47	34450.62	78.63	75.49	73.70	80.18	77.38	75.64	3.00	2.03	1.55
50	35556.75	35409.06	35445.89	81.47	78.18	76.38	82.50	79.36	77.57	4.44	2.81	2.12
Avg.	33631.71	33273.31	33086.00	84.08	79.49	76.83	84.51	80.24	77.76	5.88	3.12	2.14

Table B.10: Cost structures for all instances with different LTL rates.

	Total expected cost (%)			Fixed cost ($)			Dispatching cost (%)			LTL cost (%)		
Instances	Low LTL rate	Medium LTL rate	High LTL rate	Low LTL rate	Medium LTL rate	High LTL rate	Low LTL rate	Medium LTL rate	High LTL rate	Low LTL rate	Medium LTL rate	High LTL rate
1	217285.26	219611.31	221106.89	18.37	19.13	19.78	74.71	75.37	75.96	6.92	5.50	4.26
2	238689.98	241260.39	242667.10	13.00	14.29	14.90	78.99	80.79	81.28	8.01	4.92	3.82
3	267413.26	268889.38	269994.72	14.26	14.64	15.03	81.39	81.63	81.78	4.35	3.73	3.19
4	194006.33	198683.05	201036.27	18.75	21.83	22.86	64.79	70.97	71.85	16.46	7.20	5.30
5	225312.14	228355.43	229921.80	15.43	16.97	17.71	75.05	78.14	78.65	9.52	4.90	3.64
6	166785.60	170019.04	171817.30	20.77	22.83	23.77	68.07	71.10	71.27	11.16	6.07	4.96
7	228651.95	230334.67	231441.04	14.62	15.52	15.82	79.71	80.64	80.68	5.67	3.84	3.50
8	250327.35	251774.80	252906.75	15.84	16.59	17.01	79.14	79.94	80.06	5.02	3.47	2.93
9	247249.39	248981.52	250249.78	17.21	18.11	18.51	76.88	77.76	77.83	5.91	4.13	3.66
10	174387.43	179616.39	181899.12	18.56	22.34	23.88	62.97	69.25	70.64	18.47	8.41	5.48
11	245416.65	247775.15	249207.80	17.49	18.34	18.98	74.62	76.64	77.47	7.88	5.03	3.55
12	184909.41	188483.94	190428.19	17.41	19.74	20.73	70.53	74.01	74.63	12.06	6.25	4.64
13	235455.69	237167.37	238324.96	17.85	18.77	19.21	76.18	77.27	77.71	5.97	3.96	3.07
14	205556.08	207812.61	209258.71	18.17	19.53	19.80	73.80	75.30	75.40	8.03	5.17	4.80
15	248188.72	250140.20	251401.50	16.48	17.62	17.82	76.79	78.40	78.40	6.73	3.99	3.77
16	230512.81	233010.28	234499.85	16.44	17.28	17.96	76.57	77.52	78.25	6.99	5.21	3.78
17	253068.20	254527.92	255636.82	14.47	14.85	15.24	81.17	81.36	81.47	4.36	3.79	3.29
18	298041.34	300175.38	301455.37	15.00	15.67	16.14	78.01	79.64	80.39	6.99	4.69	3.47
19	254949.36	256423.62	257440.25	13.04	13.86	14.38	81.96	82.86	83.19	5.00	3.28	2.43
20	305681.21	307374.23	308535.05	12.21	12.91	13.25	82.50	83.52	83.78	5.29	3.57	2.97
21	270412.90	271840.00	272818.33	13.38	14.08	14.62	81.98	82.56	82.59	4.64	3.36	2.79
22	189530.98	192179.59	193637.62	17.12	18.61	19.38	74.51	76.30	76.72	8.37	5.09	3.90
23	290211.61	291552.68	292525.66	11.05	11.43	11.66	85.10	85.39	85.51	3.85	3.18	2.83
24	265183.13	266643.81	267587.90	14.93	15.71	15.94	80.19	81.19	81.40	4.88	3.10	2.66
25	253183.04	254613.38	255684.30	15.89	16.64	16.93	79.45	79.98	80.02	4.66	3.38	3.05
26	223467.47	224691.81	225520.31	13.91	14.81	15.10	81.48	82.30	82.39	4.61	2.89	2.50
27	223100.17	226044.06	227706.88	17.47	18.85	19.66	72.96	76.04	76.80	9.58	5.11	3.54
28	239379.12	240940.82	242140.31	15.77	16.66	17.07	78.62	79.47	79.56	5.61	3.87	3.37
29	233994.07	235531.11	236548.12	15.11	15.90	16.19	79.87	80.52	80.49	5.02	3.59	3.32
30	287967.65	289372.50	290360.30	13.80	14.28	14.59	81.94	82.58	83.11	4.26	3.15	2.29
31	241130.50	243162.73	244413.94	13.60	14.55	15.22	79.50	80.85	81.28	6.90	4.60	3.49
32	245392.89	247026.45	248093.29	17.04	17.64	18.21	77.98	78.34	78.61	4.99	4.02	3.18
33	284362.78	286799.61	288292.64	12.54	13.62	13.88	79.88	81.72	82.09	7.58	4.66	4.03
34	179355.55	185686.83	188096.24	17.07	22.63	23.77	56.85	69.49	70.64	26.09	7.88	5.59
35	217565.34	219318.52	220436.45	14.66	15.67	16.24	79.27	80.42	80.77	6.06	3.91	2.99
36	270798.59	272455.06	273731.28	14.14	14.85	15.32	80.37	81.15	81.43	5.49	3.99	3.25
37	235593.74	237560.49	238792.75	15.89	16.41	17.02	77.97	79.33	79.84	6.15	4.26	3.13
38	201999.48	204269.02	205734.53	19.23	20.22	21.04	73.55	74.33	74.79	7.21	5.44	4.17
39	264122.55	265535.54	266639.59	14.46	14.83	15.21	81.16	81.37	81.51	4.37	3.80	3.28
40	167771.64	171428.26	173207.27	19.77	21.96	23.50	67.18	70.86	72.22	13.05	7.18	4.28
41	212476.26	214897.47	216272.93	13.07	14.22	14.82	79.59	81.11	81.58	7.34	4.67	3.60
42	230302.68	233021.47	234665.23	16.78	18.03	18.86	74.53	76.35	77.04	8.69	5.63	4.10
43	243079.50	244980.43	246246.34	14.41	15.28	16.03	79.13	80.16	80.53	6.46	4.56	3.44
44	171639.11	175511.65	177126.04	19.44	22.70	23.91	64.61	71.40	72.44	15.95	5.90	3.64
45	245904.89	247553.72	248826.99	15.81	16.35	16.91	78.95	79.35	79.62	5.24	4.29	3.47
46	218449.63	220507.30	221905.43	17.61	18.78	19.19	75.24	76.66	76.69	7.15	4.55	4.12
47	275001.21	276844.97	278060.98	11.89	12.54	13.07	81.78	82.95	83.47	6.33	4.51	3.46
48	259872.05	261676.96	262848.90	17.55	18.41	18.98	76.59	77.58	77.99	5.86	4.01	3.03
49	267761.75	269013.53	269957.24	15.41	15.97	16.38	80.20	80.70	80.80	4.39	3.33	2.82
50	277804.88	279543.29	280663.52	14.19	14.87	15.28	80.05	81.04	81.28	5.76	4.08	3.44
Avg.	237174.07	239412.39	240755.41	15.77	16.95	17.54	76.69	78.47	78.88	7.55	4.58	3.59

Table B.11: Results for FCL, U_F, U_C, and LTL units with different values of T.

	FCL (lbs)			U_F (%)			U_C (%)			LTL units (%)		
Instance	$T=3$	$T=4$	$T=5$	$T=3$	$T=4$	$T=5$	$T=3$	$T=4$	$T=5$	$T=3$	$T=4$	$T=5$
1	36071.94	33122.72	33122.72	81.47	75.92	75.92	81.89	79.87	79.87	3.74	2.81	2.81
2	29759.20	29759.20	29759.20	82.00	81.91	81.91	81.73	81.70	81.70	3.53	3.56	3.56
3	34979.41	33501.89	33501.89	79.33	76.24	76.25	80.28	79.48	79.49	2.42	2.16	2.15
4	34872.00	33173.37	33588.27	84.66	82.40	83.79	85.15	84.67	85.55	5.42	4.96	5.51
5	33139.64	33139.64	33139.64	81.11	81.11	81.11	81.98	81.98	81.98	3.63	3.63	3.63
6	33238.21	33238.21	33238.21	79.42	79.42	79.42	81.56	81.56	81.56	4.17	4.17	4.17
7	29057.24	29057.24	28950.10	77.24	77.24	77.35	77.54	77.54	77.82	2.42	2.42	2.47
8	36268.35	36268.35	36268.35	77.82	77.82	77.82	78.18	78.18	78.18	2.17	2.17	2.17
9	33999.60	33999.60	33999.60	77.08	77.08	77.08	79.62	79.62	79.62	2.66	2.66	2.66
10	36698.00	36108.13	36108.13	85.72	84.50	84.49	85.51	84.80	84.80	6.27	5.54	5.55
11	34755.30	33425.38	33425.38	79.84	77.04	77.03	81.51	80.67	80.66	3.45	3.10	3.10
12	35605.39	36020.50	34583.88	82.46	83.73	82.02	83.18	84.15	83.23	4.52	5.11	4.52
13	35550.84	35432.42	35432.42	77.63	77.68	77.71	78.76	78.96	78.97	2.51	2.56	2.55
14	36483.18	35950.00	35869.03	80.30	78.84	78.96	80.36	79.61	79.78	3.33	2.88	2.93
15	33640.95	33640.95	33640.95	77.27	77.25	77.28	79.56	79.55	79.56	2.64	2.64	2.64
16	36698.00	36698.00	36698.00	80.92	80.92	80.92	80.68	80.68	80.68	3.57	3.57	3.57
17	36041.03	36041.03	36041.03	79.54	79.58	79.58	80.03	80.04	80.04	2.42	2.41	2.41
18	35935.26	35767.54	35767.54	80.33	78.74	78.74	80.73	79.22	79.22	3.26	2.69	2.69
19	28911.82	28970.27	28970.27	76.58	76.66	76.64	76.87	77.01	77.00	2.10	2.11	2.12
20	36268.35	35928.00	36268.35	79.26	79.32	79.23	79.59	79.94	79.58	2.54	2.67	2.55
21	30819.17	30819.17	31175.46	78.53	78.53	78.18	78.56	78.56	78.28	2.13	2.13	2.03
22	29221.51	29111.76	29111.76	80.72	80.94	80.94	80.73	81.12	81.12	3.42	3.49	3.49
23	29941.80	29941.80	29941.80	78.37	78.37	78.37	78.10	78.10	78.10	2.15	2.15	2.15
24	35480.67	35480.67	35480.67	75.81	75.81	75.81	77.77	77.77	77.77	1.99	1.99	1.99
25	29711.57	29301.22	29301.22	78.91	79.33	79.36	78.68	79.54	79.55	2.11	2.38	2.37
26	28900.60	28960.42	28960.42	76.16	76.03	76.03	76.51	76.38	76.38	1.76	1.71	1.71
27	31483.80	29014.55	28262.00	79.48	76.18	75.29	83.46	82.90	82.96	3.80	3.48	3.49
28	36275.18	35587.65	35504.29	76.85	76.87	76.99	77.07	78.08	78.32	2.39	2.71	2.74
29	29733.72	29131.00	29066.59	78.48	77.70	77.78	78.14	77.85	77.97	2.23	2.20	2.22
30	34329.14	34329.14	34329.14	76.42	76.42	76.42	78.48	78.48	78.48	2.11	2.11	2.11
31	28850.00	26501.00	26501.00	80.67	75.48	75.48	80.90	79.52	79.52	3.15	2.63	2.63
32	29543.06	29398.73	29692.12	79.27	79.35	78.98	79.00	79.27	78.66	2.51	2.56	2.31
33	22864.49	23531.23	20732.97	81.96	80.23	76.96	81.60	81.11	80.01	3.60	3.29	2.91
34	34986.24	32943.33	32534.65	83.14	81.03	79.75	84.40	84.34	83.53	6.13	5.99	5.38
35	29394.00	29394.00	29394.00	78.22	78.22	78.22	77.99	77.99	77.99	2.52	2.52	2.52
36	36060.35	35928.00	36060.35	79.40	79.31	79.45	79.98	80.08	79.99	2.64	2.71	2.62
37	33985.09	33985.09	34402.46	78.17	78.17	77.52	79.83	79.83	79.15	2.90	2.90	2.63
38	36483.18	35950.00	35950.00	81.65	80.15	80.15	81.72	80.98	80.98	3.55	3.13	3.13
39	36275.18	35747.94	35747.94	78.31	76.60	76.60	78.53	77.76	77.76	2.33	2.07	2.07
40	35812.67	35623.88	35623.88	85.25	83.66	83.66	85.93	84.63	84.63	5.12	4.33	4.33
41	29576.60	29111.76	29111.76	80.95	80.09	80.09	80.75	80.36	80.36	3.29	3.11	3.11
42	36248.75	35879.88	35468.12	81.23	81.47	79.82	81.48	82.09	81.12	3.91	4.11	3.60
43	29880.93	29880.93	29880.93	80.11	80.11	80.11	79.92	79.92	79.92	3.03	3.03	3.03
44	34341.76	34341.76	34341.76	81.03	81.02	81.06	82.86	82.84	82.86	4.26	4.27	4.25
45	34263.33	34263.33	34263.33	74.90	74.90	74.89	78.70	78.70	78.70	2.70	2.70	2.70
46	36095.78	35970.78	35970.78	80.04	80.10	80.10	80.50	80.57	80.57	3.01	3.02	3.02
47	29741.81	29741.81	29741.81	80.96	80.96	80.96	80.74	80.74	80.74	3.16	3.16	3.16
48	35205.69	35205.69	35018.41	76.52	76.53	76.65	78.23	78.23	78.57	2.58	2.58	2.70
49	34391.47	28510.26	28510.26	75.49	66.85	66.85	77.38	75.89	75.89	2.03	1.65	1.65
50	35409.06	35409.06	35623.88	78.18	78.15	77.93	79.36	79.36	78.98	2.81	2.82	2.69
Avg.	33265.61	32764.77	32681.53	79.50	78.64	78.47	80.24	80.04	79.96	3.12	3.01	2.97

Table B.12: Cost structures for all instances with different values of T.

Instances	Total expected cost ($)			Fixed cost (%)			Dispatching cost (%)			LTL cost (%)		
	$T=3$	$T=4$	$T=5$	$T=3$	$T=4$	$T=5$	$T=3$	$T=4$	$T=5$	$T=3$	$T=4$	$T=5$
1	219611.31	219282.40	219281.59	19.13	19.21	19.21	75.37	76.65	76.65	5.50	4.14	4.14
2	241260.39	241240.75	241240.75	14.29	14.29	14.29	80.79	80.74	80.74	4.92	4.97	4.97
3	268889.38	268862.93	268853.13	14.64	14.70	14.70	81.63	81.97	81.98	3.73	3.33	3.32
4	198683.05	198515.13	198507.67	21.83	21.65	21.41	70.97	71.76	71.27	7.20	6.59	7.32
5	228355.43	228355.43	228355.43	16.97	16.97	16.97	78.14	78.14	78.14	4.90	4.90	4.90
6	170019.04	170019.04	170019.04	22.83	22.83	22.83	71.10	71.10	71.10	6.07	6.07	6.07
7	230334.67	230334.67	230377.56	15.52	15.52	15.45	80.64	80.64	80.65	3.84	3.84	3.91
8	251774.80	251774.80	251774.80	16.59	16.59	16.59	79.94	79.94	79.94	3.47	3.47	3.47
9	248981.52	248981.52	248981.52	18.11	18.11	18.11	77.76	77.76	77.76	4.13	4.13	4.13
10	179616.39	179526.51	179531.86	22.34	22.62	22.62	69.25	69.95	69.94	8.41	7.43	7.44
11	247775.15	247515.04	247518.54	18.34	18.40	18.40	76.64	77.09	77.08	5.03	4.51	4.52
12	188483.94	188511.92	188526.78	19.74	19.40	19.70	74.01	73.54	74.06	6.25	7.06	6.24
13	237167.37	237145.98	237136.58	18.77	18.69	18.69	77.27	77.27	77.28	3.96	4.04	4.03
14	207812.61	207760.19	207704.29	19.53	19.77	19.74	75.30	75.75	75.71	5.17	4.48	4.55
15	250140.20	250141.94	250118.11	17.62	17.62	17.62	78.40	78.39	78.40	3.99	3.99	3.98
16	233010.28	233010.28	233010.28	17.28	17.28	17.28	77.52	77.52	77.52	5.21	5.21	5.21
17	254527.92	254519.43	254512.90	14.85	14.85	14.85	81.36	81.38	81.37	3.79	3.78	3.78
18	300175.38	300177.35	300176.69	15.67	15.94	15.94	79.64	80.19	80.19	4.69	3.87	3.87
19	256423.62	256343.75	256347.76	13.86	13.85	13.85	82.86	82.85	82.84	3.28	3.30	3.31
20	307374.23	307404.90	307365.03	12.91	12.79	12.92	83.52	83.47	83.50	3.57	3.74	3.58
21	271845.04	271845.04	271868.12	14.21	14.21	14.48	82.50	82.50	82.39	3.29	3.29	3.13
22	192179.59	192016.94	192016.94	18.61	18.52	18.52	76.30	76.30	76.30	5.09	5.19	5.19
23	291552.68	291552.68	291552.68	11.43	11.43	11.43	85.39	85.39	85.39	3.18	3.18	3.18
24	266643.81	266643.81	266643.81	15.71	15.71	15.71	81.19	81.19	81.19	3.10	3.10	3.10
25	254613.38	254523.50	254510.86	16.64	16.33	16.33	79.98	79.85	79.88	3.38	3.82	3.79
26	224691.81	224659.47	224646.59	14.81	14.80	14.80	82.30	82.39	82.39	2.89	2.81	2.80
27	226044.06	225767.69	225699.97	18.85	18.79	18.65	76.04	76.54	76.65	5.11	4.68	4.70
28	240940.82	240924.55	240885.16	16.66	16.24	16.21	79.47	79.36	79.34	3.87	4.39	4.45
29	235531.11	235438.63	235328.76	15.90	15.79	15.74	80.52	80.66	80.68	3.59	3.54	3.58
30	289372.50	289372.50	289372.50	14.28	14.28	14.28	82.58	82.58	82.58	3.15	3.15	3.15
31	243162.73	242646.71	242646.71	14.55	14.50	14.50	80.85	81.66	81.66	4.60	3.84	3.84
32	247026.45	247021.52	247013.83	17.64	17.58	17.82	78.34	78.31	78.47	4.02	4.11	3.71
33	286786.59	286761.46	286443.10	13.46	13.69	13.26	81.70	81.88	82.82	4.84	4.43	3.92
34	185686.83	185546.44	185558.50	22.63	22.38	22.64	69.49	69.92	70.44	7.88	7.70	6.92
35	219318.52	219318.52	219318.52	15.67	15.67	15.67	80.42	80.42	80.42	3.91	3.91	3.91
36	272455.06	272470.91	272438.09	14.85	14.80	14.86	81.15	81.11	81.19	3.99	4.09	3.96
37	237560.49	237560.49	237536.92	16.41	16.41	16.80	79.33	79.33	79.34	4.26	4.26	3.86
38	204269.02	204170.28	204170.28	20.22	20.44	20.44	74.33	74.75	74.75	5.44	4.81	4.81
39	265535.54	265500.87	265501.50	14.83	14.99	14.99	81.37	81.64	81.64	3.80	3.37	3.37
40	171428.26	171423.08	171423.08	21.96	22.44	22.44	70.86	71.49	71.49	7.18	6.07	6.07
41	214897.47	214935.62	214935.62	14.22	14.18	14.18	81.11	81.40	81.40	4.67	4.42	4.42
42	233021.47	232915.18	232896.20	18.03	17.75	18.09	76.35	76.34	76.73	5.63	5.91	5.18
43	244980.43	244980.43	244980.43	15.28	15.28	15.28	80.16	80.16	80.16	4.56	4.56	4.56
44	175511.65	175510.68	175499.03	22.70	22.70	22.70	71.40	71.38	71.41	5.90	5.92	5.89
45	247553.72	247553.72	247558.89	16.35	16.35	16.35	79.35	79.35	79.35	4.29	4.29	4.30
46	220507.30	220402.40	220402.40	18.78	18.74	18.74	76.66	76.69	76.69	4.55	4.57	4.57
47	276844.97	276844.97	276844.97	12.54	12.54	12.54	82.95	82.95	82.95	4.51	4.51	4.51
48	261676.96	261676.49	261675.42	18.41	18.41	18.23	77.58	77.59	77.58	4.01	4.00	4.19
49	269013.53	268097.46	268097.46	15.97	15.67	15.67	80.70	81.62	81.62	3.33	2.71	2.71
50	279543.29	279526.78	279509.92	14.87	14.87	15.04	81.04	81.03	81.06	4.08	4.09	3.91
Avg.	239412.24	239340.66	239326.33	16.95	16.93	16.95	78.47	78.64	78.68	4.58	4.43	4.37

Table B.13: Results for FCL, U_F, U_C and LTL units with different demand parameters.

	FCL (lbs)				U_F (%)				U_C (%)				LTL units (%)			
Instance	(L, L)	(L, H)	(H, L)	(H, H)	(L, L)	(L, H)	(H, L)	(H, H)	(L, L)	(L, H)	(H, L)	(H, H)	(L, L)	(L, H)	(H, L)	(H, H)
1	33873.44	33122.72	33827.52	32846.09	82.88	75.92	84.82	77.55	86.28	79.87	88.35	82.07	1.86	2.81	1.62	2.43
2	29778.42	29759.20	29726.00	29704.81	87.34	81.91	89.97	83.70	87.08	81.70	89.72	83.43	2.00	3.56	1.96	2.62
3	34004.67	33501.89	34246.74	33713.51	81.95	76.25	84.18	77.88	84.63	79.49	87.22	81.08	1.25	2.15	1.16	1.93
4	34784.63	33588.27	35064.63	33654.00	89.33	83.79	90.90	84.96	90.53	85.55	91.97	86.98	3.30	5.51	2.77	4.08
5	33654.67	33139.64	34089.43	33127.16	87.72	81.11	89.25	83.71	88.35	81.98	89.87	84.49	2.56	3.63	2.07	3.18
6	34402.46	33238.21	34872.00	33710.00	86.15	79.42	88.73	82.45	87.95	81.56	90.25	84.56	2.13	4.17	2.02	2.97
7	29031.33	28950.10	29236.84	29246.67	84.86	77.35	87.23	81.33	85.35	77.82	87.36	81.40	1.37	2.47	1.22	1.93
8	36226.77	36268.35	36292.22	36505.79	85.14	77.82	87.34	81.59	85.63	78.18	87.78	81.68	1.52	2.17	1.30	2.14
9	34557.89	33999.60	34921.72	33844.47	84.12	77.08	86.01	79.13	86.24	79.62	87.90	82.00	1.71	2.66	1.31	2.16
10	36089.10	36108.13	36173.67	36187.84	89.84	84.49	92.06	87.96	90.26	84.80	92.40	88.26	3.04	5.55	2.94	4.70
11	34669.31	33425.38	34505.80	33442.27	84.43	77.03	86.21	78.61	86.76	80.66	88.84	82.37	1.93	3.10	1.69	2.37
12	35350.25	34583.88	34806.42	35592.15	88.03	82.02	89.87	84.96	89.01	83.23	90.99	85.64	2.63	4.52	2.36	3.51
13	35486.28	35432.42	35812.93	35466.98	84.27	77.71	87.66	80.10	85.59	78.97	88.65	81.49	1.42	2.55	1.52	1.88
14	35818.79	35869.03	35934.47	35972.65	85.01	78.96	87.86	82.32	86.06	79.78	88.78	83.07	1.81	2.93	1.75	2.84
15	34979.41	33640.95	35211.64	34330.74	85.73	77.28	88.07	80.64	87.01	79.56	89.29	82.66	2.11	2.64	1.91	2.63
16	36698.00	36698.00	36698.00	36698.00	87.93	80.92	90.32	84.38	87.68	80.68	90.12	84.17	2.61	3.57	2.49	3.43
17	35998.65	36041.03	36095.78	36127.47	85.00	79.58	87.20	81.65	85.64	80.04	87.79	82.13	1.56	2.41	1.34	2.24
18	35689.17	35767.54	35833.29	36039.27	86.16	78.74	88.03	82.93	86.85	79.22	88.66	83.34	1.83	2.69	1.52	2.54
19	28927.90	28970.27	28988.70	29021.12	84.44	76.64	87.32	79.87	84.89	77.00	87.77	80.24	1.40	2.12	1.35	1.76
20	35879.88	36268.35	35990.43	35839.44	84.45	79.23	87.16	81.75	85.38	79.58	87.85	82.72	1.46	2.55	1.28	2.36
21	30931.68	31175.46	30367.87	29698.33	85.21	78.18	86.74	81.18	85.12	78.28	86.56	81.01	1.52	2.03	1.12	2.08
22	29089.47	29111.76	29136.84	29147.79	87.69	80.94	88.67	83.75	87.96	81.12	88.90	83.94	2.51	3.49	1.80	3.18
23	29986.22	29941.80	29903.58	29870.35	84.85	78.37	87.30	80.86	84.46	78.10	86.94	80.61	1.34	2.15	1.19	1.87
24	35785.00	35480.67	35908.38	35602.40	83.33	75.81	86.07	78.77	84.80	77.77	87.25	80.60	1.58	1.99	1.35	2.08
25	29297.00	29301.22	29310.31	29316.40	84.77	79.36	87.08	80.52	84.95	79.55	87.25	80.62	1.50	2.37	1.31	1.98
26	28843.28	28960.42	28916.71	28955.67	84.22	76.03	87.01	79.67	84.97	76.38	87.54	80.09	1.32	1.71	1.16	1.53
27	30428.25	28262.00	31184.43	29034.04	82.00	75.29	83.66	76.48	88.48	82.96	89.82	84.35	2.38	3.49	1.87	2.93
28	36108.13	35504.29	36187.84	35683.35	84.14	76.99	86.63	79.89	84.58	78.32	87.03	81.09	1.35	2.74	1.19	1.95
29	29042.63	29066.59	29242.89	29111.53	83.80	77.78	86.70	79.69	84.10	77.97	86.93	79.88	1.29	2.22	1.18	1.77
30	34764.59	34329.14	35199.74	34785.05	82.81	76.42	85.34	79.02	84.59	78.48	86.87	80.94	1.23	2.11	1.05	1.78
31	27368.90	26501.00	27178.09	26278.46	83.29	75.48	84.73	77.47	85.94	79.52	87.94	81.93	1.66	2.63	1.39	2.24
32	29399.04	29692.12	29531.81	29522.14	86.04	78.98	88.88	82.16	85.95	78.66	88.54	81.89	1.43	2.31	1.40	1.94
33	21255.24	20732.97	21405.25	21046.52	84.81	76.96	86.65	80.43	87.04	80.01	88.69	83.08	1.82	2.91	1.46	2.35
34	34388.00	32534.65	34445.20	33929.14	87.96	79.75	89.15	84.33	90.16	83.53	91.51	87.12	3.42	5.38	2.84	4.64
35	29394.00	29394.00	29142.43	29157.84	85.79	78.22	87.28	81.06	85.55	77.99	87.57	81.26	1.87	2.52	1.47	2.24
36	36020.50	36060.35	36112.05	35795.85	85.26	79.45	87.63	80.87	85.94	79.99	88.28	81.87	1.65	2.62	1.49	2.24
37	35284.32	34402.46	35480.67	34637.90	86.07	77.52	88.17	81.74	87.11	79.15	89.13	83.25	1.92	2.63	1.66	2.50
38	35904.67	35950.00	36009.05	36043.50	86.48	80.15	89.24	83.94	87.51	80.98	90.16	84.69	2.22	3.13	2.18	3.33
39	35879.88	35747.94	35722.70	35795.85	84.07	76.60	86.65	79.85	85.20	77.76	87.96	80.88	1.50	2.07	1.41	1.93
40	36241.50	35623.88	36121.37	35967.60	89.98	83.66	91.01	85.86	90.30	84.63	91.44	86.52	3.28	4.33	2.54	3.87
41	29089.47	29111.76	29136.84	29152.92	87.21	80.09	88.10	82.02	87.57	80.36	88.44	82.29	2.14	3.11	1.48	2.23
42	35388.77	35468.12	35970.78	36187.84	86.18	79.82	88.20	84.19	87.78	81.12	88.90	84.56	2.14	3.60	1.58	3.09
43	29915.71	29880.93	29841.18	29815.38	86.30	80.11	88.48	82.44	86.00	79.92	88.21	82.24	1.65	3.03	1.44	2.31
44	34879.25	34341.76	35003.41	34831.13	86.81	81.06	89.10	83.94	88.42	82.86	90.63	85.59	2.52	4.25	2.34	3.69
45	34927.33	34263.33	35160.32	34437.24	82.54	74.89	84.83	78.00	85.63	78.70	87.66	81.78	1.45	2.70	1.22	1.98
46	35928.00	35970.78	36026.72	36074.67	86.06	80.10	88.99	82.14	86.66	80.57	89.64	82.58	1.88	3.02	1.98	2.46
47	30330.41	29741.81	29704.81	29832.24	86.63	80.96	88.07	81.52	86.23	80.74	87.74	81.16	1.76	3.16	1.39	2.02
48	35690.00	35018.41	35660.10	35566.98	84.71	76.65	86.61	80.59	85.94	78.57	87.88	81.95	1.58	2.70	1.30	2.09
49	30380.70	28510.26	31502.09	29081.02	76.42	66.85	78.99	69.83	84.18	75.89	86.09	78.91	1.15	1.65	0.90	1.32
50	35991.16	35623.88	35908.38	35948.87	86.40	77.93	86.69	81.32	86.92	78.98	87.51	82.07	2.13	2.69	1.43	2.46
Avg.	33076.68	32681.53	33175.00	32827.57	85.33	78.47	87.46	81.22	86.54	79.96	88.61	82.65	1.89	2.97	1.63	2.52

Table B.14: Cost structures for all instances with different demand parameters.

	Total expected cost ($)				Fixed cost (%)				Dispatching cost (%)				LTL cost (%)			
Instance	(L, L)	(L, H)	(H, L)	(H, H)	(L, L)	(L, H)	(H, L)	(H, H)	(L, L)	(L, H)	(H, L)	(H, H)	(L, L)	(L, H)	(H, L)	(H, H)
1	216610.90	219281.59	256426.13	260824.18	18.47	19.21	18.18	18.84	78.74	76.65	79.37	77.55	2.79	4.14	2.45	3.61
2	236844.50	241240.75	280705.69	284235.22	13.85	14.29	13.48	14.20	83.32	80.74	83.74	82.12	2.83	4.97	2.78	3.68
3	260206.63	268853.13	308470.89	312205.92	14.18	14.70	13.88	14.53	83.87	81.98	84.29	82.48	1.96	3.32	1.83	2.99
4	194283.97	198507.67	229928.39	234059.33	21.55	21.41	21.53	21.62	73.98	71.27	74.69	72.91	4.47	7.32	3.78	5.47
5	223298.47	228355.43	264457.10	268513.54	16.45	16.97	16.57	16.64	80.04	78.14	80.58	79.05	3.50	4.90	2.84	4.31
6	168088.10	170019.04	198995.22	202318.51	22.83	22.83	22.64	22.88	73.99	71.10	74.32	72.74	3.18	6.07	3.03	4.38
7	227784.44	230377.56	269710.42	274153.60	14.56	15.45	14.32	15.03	83.23	80.65	83.70	81.88	2.21	3.91	1.98	3.09
8	244770.22	251774.80	289848.69	294324.50	15.53	16.59	15.27	16.03	82.00	79.94	82.60	80.53	2.47	3.47	2.12	3.44
9	245371.34	248981.52	290413.99	295481.32	17.34	18.11	17.24	17.82	79.94	77.76	80.67	78.79	2.72	4.13	2.09	3.39
10	176061.70	179531.86	208350.65	211959.98	22.33	22.62	21.97	22.21	73.50	69.94	73.97	71.40	4.17	7.44	4.06	6.39
11	239188.96	247518.54	283376.18	287342.26	17.94	18.40	17.63	18.30	79.20	77.08	79.85	78.22	2.86	4.52	2.52	3.48
12	186789.97	188526.78	221013.09	225165.50	19.19	19.70	18.92	19.55	77.10	74.06	77.73	75.55	3.72	6.24	3.36	4.90
13	236217.10	237136.58	279310.81	285117.24	17.86	18.69	17.40	18.42	79.85	77.28	80.13	78.59	2.29	4.03	2.46	2.99
14	204120.20	207704.29	241612.61	245553.38	18.91	19.74	18.47	19.15	78.23	75.71	78.74	76.40	2.86	4.55	2.79	4.45
15	246506.00	250118.11	291739.34	297027.45	16.73	17.62	16.42	17.20	80.03	78.40	80.63	78.80	3.24	3.98	2.95	4.00
16	224339.96	233010.28	265980.51	269072.13	16.36	17.28	16.03	16.72	79.77	77.52	80.25	78.23	3.87	5.21	3.72	5.05
17	249525.60	254512.90	295717.13	299464.78	14.21	14.85	13.97	14.57	83.31	81.37	83.89	81.90	2.48	3.78	2.14	3.53
18	294374.83	300176.69	348701.96	353667.63	14.99	15.94	14.86	15.47	82.33	80.19	82.91	80.85	2.67	3.87	2.23	3.68
19	250372.79	256347.76	296701.71	300583.08	12.87	13.85	12.52	13.45	84.90	82.84	85.32	83.77	2.22	3.31	2.16	2.78
20	300697.15	307365.03	356796.21	359716.34	12.28	12.92	12.01	12.48	85.65	83.50	86.16	84.18	2.07	3.58	1.82	3.33
21	267980.17	271868.12	317550.28	321777.24	13.44	14.48	13.04	13.22	84.17	82.39	85.21	83.55	2.38	3.13	1.76	3.23
22	187138.71	192016.94	221401.38	225432.88	17.59	18.52	17.64	18.10	78.60	76.30	79.62	77.12	3.81	5.19	2.74	4.77
23	286056.84	291552.68	339367.11	342525.52	10.81	11.43	10.52	11.13	87.18	85.39	87.69	86.09	2.01	3.18	1.78	2.78
24	256431.53	266643.81	304158.07	307212.11	14.77	15.71	14.50	15.26	82.74	81.19	83.37	81.49	2.49	3.10	2.14	3.25
25	253534.06	254510.86	299793.26	305939.39	15.68	16.33	15.38	16.25	81.89	79.88	82.47	80.57	2.43	3.79	2.15	3.18
26	220728.90	224646.59	261300.63	265550.63	13.60	14.80	13.29	14.26	84.19	82.39	84.76	83.21	2.21	2.80	1.94	2.52
27	223664.89	225699.97	264945.56	269207.05	18.52	18.65	18.60	18.78	78.23	76.65	78.83	77.26	3.24	4.70	2.57	3.96
28	240488.04	240885.16	284610.54	289324.65	15.59	16.21	15.27	15.97	82.17	79.34	82.74	80.83	2.24	4.45	1.99	3.20
29	230036.31	235328.76	272307.48	276677.73	14.99	15.74	14.68	15.55	82.90	80.68	83.37	81.58	2.11	3.58	1.95	2.87
30	283585.34	289372.50	336217.10	340096.96	13.74	14.28	13.64	14.17	84.40	82.58	84.76	83.16	1.86	3.15	1.60	2.67
31	240925.39	242646.71	285323.57	289754.98	13.88	14.50	13.64	14.18	83.66	81.66	84.28	82.51	2.46	3.84	2.08	3.31
32	243377.14	247013.83	288226.02	292992.18	16.75	17.82	16.34	17.28	80.90	78.47	81.35	79.58	2.35	3.71	2.31	3.14
33	285311.43	286443.10	338144.83	342770.83	12.66	13.26	12.56	13.03	84.85	82.82	85.43	83.78	2.49	3.92	2.01	3.18
34	185200.81	185558.50	218932.86	223666.97	22.36	22.64	22.37	22.51	73.14	70.44	73.87	71.44	4.50	6.92	3.76	6.04
35	217990.16	219318.52	258289.00	261575.98	14.64	15.67	14.44	15.20	82.41	80.42	83.23	81.29	2.96	3.91	2.33	3.51
36	268190.96	272438.09	317378.76	322911.49	14.18	14.86	13.90	14.61	83.29	81.19	83.80	82.00	2.53	3.96	2.30	3.40
37	231107.05	237536.92	273919.11	277358.16	16.03	16.80	15.87	16.24	81.09	79.34	81.63	80.06	2.87	3.86	2.49	3.70
38	201580.04	204170.28	238313.27	243261.31	19.49	20.44	19.06	19.68	77.03	74.75	77.51	75.16	3.48	4.81	3.43	5.16
39	260777.39	265501.50	308589.04	313782.48	13.99	14.99	13.65	14.55	83.52	81.64	84.00	82.29	2.49	3.37	2.35	3.16
40	170478.57	171423.08	201423.61	206064.35	21.81	22.44	21.81	22.29	73.50	71.49	74.54	72.24	4.68	6.07	3.65	5.47
41	211298.10	214935.62	250343.52	253843.66	13.36	14.18	13.38	14.08	83.55	81.40	84.48	82.72	3.09	4.42	2.15	3.20
42	227851.92	232896.20	269912.93	273791.97	17.32	18.09	17.29	17.69	79.53	76.73	80.38	77.81	3.15	5.18	2.33	4.50
43	238511.80	244980.43	282731.76	286225.28	14.68	15.28	14.37	15.04	82.79	80.16	83.40	81.45	2.53	4.56	2.22	3.51
44	171835.65	175499.03	203123.19	207480.75	22.49	22.70	22.19	22.62	73.94	71.41	74.48	72.22	3.57	5.89	3.33	5.16
45	246657.26	247558.89	291847.22	297271.70	15.63	16.35	15.42	16.05	82.01	79.35	82.59	80.77	2.35	4.30	1.99	3.18
46	216728.02	220402.40	256632.66	260710.93	17.97	18.74	17.47	18.55	79.13	76.69	79.45	77.70	2.90	4.57	3.08	3.75
47	273158.18	276844.97	324239.79	326791.22	12.27	12.54	11.95	12.73	85.19	82.95	86.02	84.37	2.54	4.51	2.03	2.90
48	260428.85	261675.42	307902.73	314467.70	17.51	18.23	17.24	18.00	79.98	77.58	80.68	78.73	2.51	4.19	2.08	3.27
49	263819.24	268097.46	312543.61	317019.14	14.80	15.67	14.78	15.37	83.28	81.62	83.71	82.45	1.92	2.71	1.51	2.18
50	277211.76	279509.92	328444.70	333115.28	14.07	15.04	14.12	14.70	82.79	81.06	83.75	81.69	3.14	3.91	2.13	3.61
Avg.	235350.75	239326.33	278723.41	282987.73	16.26	16.95	16.03	16.64	80.90	78.68	81.50	79.62	2.84	4.37	2.47	3.73

Table B.15: Results for FCL, U_F, U_C, and LTL units with different demand distributions.

	FCL (lbs)			U_F (%)			U_C (%)			LTL units (%)		
Instance	Normal	Uniform	Gamma	Normal	Uniform	Gamma	Normal	Uniform	Gamma	Normal	Uniform	Gamma
1	33122.72	32479.24	32268.30	75.92	73.08	75.23	79.87	77.52	80.04	2.89	1.82	3.71
2	29759.20	29741.81	29759.20	81.91	79.10	81.32	81.70	78.93	81.12	3.69	2.41	4.46
3	33501.89	33506.22	32774.67	76.25	73.71	74.52	79.49	76.32	78.64	2.20	1.30	2.80
4	33588.27	32701.90	33275.51	83.79	79.63	82.24	85.55	81.91	84.60	5.83	4.71	6.07
5	33139.64	32863.40	33430.42	81.11	78.21	82.59	81.98	79.25	83.47	3.77	2.99	5.01
6	33238.21	33228.60	33622.63	79.42	78.62	80.14	81.56	80.97	82.42	4.35	2.87	4.73
7	28950.10	29229.53	29225.62	77.35	76.87	77.88	77.82	76.92	77.97	2.53	1.24	3.21
8	36268.35	36280.63	36255.33	77.82	76.26	78.20	78.18	76.56	78.57	2.22	1.25	3.56
9	33999.60	33443.52	33426.88	77.08	74.44	75.64	79.62	77.34	78.83	2.73	1.59	2.94
10	36108.13	36126.00	36089.10	84.49	82.06	86.93	84.80	82.31	87.29	5.88	4.93	7.51
11	33425.38	33507.20	32313.61	77.03	76.13	73.85	80.66	79.58	78.73	3.20	2.02	3.71
12	34583.88	34644.29	34955.45	82.02	81.11	83.68	83.23	82.15	84.68	4.73	3.99	5.20
13	35432.42	34947.90	35392.97	77.71	74.63	77.05	78.97	76.38	78.39	2.62	1.63	2.93
14	35869.03	35892.06	35869.03	78.96	78.07	78.63	79.78	78.82	79.46	3.02	2.34	3.60
15	33640.95	33900.67	34138.00	77.28	76.00	78.71	79.56	78.17	80.59	2.71	1.78	3.80
16	36698.00	36698.00	36698.00	80.92	79.19	81.91	80.68	78.97	81.70	3.70	3.17	4.82
17	36041.03	36078.57	36078.57	79.58	76.92	76.51	80.04	77.35	76.94	2.47	1.33	2.54
18	35767.54	35973.40	35743.05	78.74	77.68	79.34	79.22	78.03	79.81	2.77	1.95	3.93
19	28970.27	28979.69	28950.10	76.64	75.79	77.81	77.00	76.09	78.20	2.16	1.09	2.54
20	36268.35	35970.78	35713.18	79.23	76.93	78.44	79.58	77.42	79.45	2.62	1.45	4.04
21	31175.46	30922.74	30641.02	78.18	75.21	77.96	78.28	75.36	77.94	2.07	1.19	2.87
22	29111.76	29118.48	29111.76	80.94	79.09	80.64	81.12	79.22	80.84	3.61	2.49	3.83
23	29941.80	29750.29	29941.80	78.37	76.69	77.64	78.10	76.50	77.36	2.20	1.37	2.56
24	35480.67	35513.57	35838.71	75.81	73.62	78.07	77.77	75.37	79.31	2.03	1.18	2.82
25	29301.22	29306.90	29303.19	79.36	76.08	78.04	79.55	76.21	78.13	2.43	1.16	2.39
26	28960.42	28905.86	28894.51	76.03	75.47	76.16	76.38	75.84	76.54	1.74	0.90	2.60
27	28262.00	26321.06	27506.30	75.29	70.66	73.03	82.96	79.43	81.60	3.62	2.52	3.86
28	35504.29	36158.69	36142.82	76.99	76.06	77.88	78.32	76.30	78.15	2.82	1.47	2.89
29	29066.59	28767.65	28916.04	77.78	75.41	75.66	77.97	76.26	76.49	2.27	1.21	2.64
30	34329.14	33514.21	34329.14	76.42	72.89	74.57	78.48	75.29	76.77	2.16	1.37	2.86
31	26501.00	25947.02	26123.65	75.48	72.23	74.36	79.52	76.70	79.09	2.70	1.60	3.03
32	29692.12	29537.22	29692.12	78.98	76.87	79.57	78.66	76.65	79.19	2.37	1.60	2.96
33	20732.97	20575.81	20518.56	76.96	75.13	76.63	80.01	78.31	80.05	2.99	1.98	3.41
34	32534.65	32452.00	33375.37	79.75	78.88	82.45	83.53	82.73	85.57	5.69	5.22	6.97
35	29394.00	29131.00	29118.48	78.22	77.30	78.99	77.99	77.50	79.23	2.58	1.49	2.42
36	36060.35	35171.28	36060.35	79.45	75.64	78.35	79.99	77.45	78.92	2.69	1.72	3.40
37	34402.46	33539.51	33985.09	77.52	75.84	77.58	79.15	77.71	79.40	2.70	1.65	3.15
38	35950.00	35990.43	35950.00	80.15	77.36	80.51	80.98	78.06	81.32	3.23	1.91	4.14
39	35747.94	35774.33	35666.97	76.60	74.12	75.89	77.76	75.23	76.99	2.11	1.12	2.63
40	35623.88	35284.22	35623.88	83.66	79.71	83.44	84.63	81.06	84.42	4.52	3.78	5.83
41	29111.76	29567.90	29111.76	80.09	78.52	79.55	80.36	78.31	79.82	3.21	1.94	3.82
42	35468.12	35928.00	35879.88	79.82	78.08	81.08	81.12	78.61	81.72	3.73	2.35	4.60
43	29880.93	29860.21	29880.93	80.11	77.67	79.92	79.92	77.53	79.74	3.13	1.69	3.37
44	34341.76	33991.71	34270.36	81.06	79.17	81.31	82.86	81.20	83.22	4.44	3.62	5.33
45	34263.33	33814.84	34060.44	74.89	73.51	75.81	78.70	77.53	79.84	2.78	1.58	3.08
46	35970.78	35990.43	35970.78	80.10	78.33	80.32	80.57	78.75	80.82	3.11	2.46	3.60
47	29741.81	29399.04	29399.27	80.96	76.80	79.13	80.74	76.86	79.18	3.26	1.60	3.57
48	35018.41	34922.20	35205.69	76.65	74.21	77.12	78.57	76.12	78.84	2.78	1.43	3.51
49	28510.26	27301.84	27894.33	66.85	65.50	66.51	75.89	74.77	75.97	1.68	0.56	1.90
50	35623.88	35654.57	35409.06	77.93	77.14	78.22	78.98	78.04	79.52	2.76	1.93	3.46
Avg.	32681.53	32486.13	32596.04	78.47	76.35	78.34	79.96	77.92	79.94	3.07	2.04	3.69

Table B.16: Cost structures for all instances with different demand distributions.

	Total expected cost ($)			Fixed cost (%)			Dispatching cost (%)			LTL cost (%)		
Instance	Normal	Uniform	Gamma	Normal	Uniform	Gamma	Normal	Uniform	Gamma	Normal	Uniform	Gamma
1	219281.59	223152.99	222797.34	19.21	19.82	18.74	76.65	77.56	76.00	4.14	2.62	5.26
2	241240.75	241565.60	242229.42	14.29	14.97	14.24	80.74	81.74	79.83	4.97	3.29	5.93
3	268853.13	263950.84	263051.15	14.70	15.40	14.62	81.98	82.62	81.19	3.32	1.98	4.19
4	198507.67	196072.18	195342.37	21.41	22.15	21.49	71.27	71.91	70.93	7.32	5.94	7.58
5	228355.43	224135.01	230718.81	16.97	17.47	16.61	78.14	78.62	76.96	4.90	3.91	6.42
6	170019.04	174166.18	174763.40	22.83	23.45	22.75	71.10	72.47	70.68	6.07	4.08	6.57
7	230377.56	231378.56	235609.59	15.45	15.94	15.29	80.65	82.10	79.80	3.91	1.96	4.92
8	251774.80	251023.46	250463.37	16.59	17.14	16.18	79.94	80.88	78.37	3.47	1.98	5.45
9	248981.52	247839.32	249720.88	18.11	18.72	18.12	77.76	78.84	77.46	4.13	2.44	4.43
10	179531.86	178899.26	181480.22	22.62	23.42	21.66	69.94	70.30	68.96	7.44	6.28	9.38
11	247518.54	247490.46	249644.68	18.40	18.94	18.31	77.08	78.16	76.50	4.52	2.90	5.18
12	188526.78	190526.00	188950.88	19.70	20.10	19.31	74.06	74.60	73.85	6.24	5.30	6.84
13	237136.58	236216.39	242749.69	18.69	19.30	18.68	77.28	78.17	76.85	4.03	2.53	4.47
14	207704.29	209584.87	208869.06	19.74	20.14	19.63	75.71	76.31	75.00	4.55	3.55	5.37
15	250118.11	250709.38	252287.92	17.62	18.18	17.27	78.40	79.17	77.23	3.98	2.65	5.51
16	233010.28	227167.63	232033.37	17.28	17.72	16.84	77.52	77.81	76.48	5.21	4.47	6.68
17	254512.90	257905.24	261570.69	14.85	15.56	15.34	81.37	82.39	80.80	3.78	2.05	3.85
18	300176.69	301233.96	299044.78	15.94	16.47	15.57	80.19	80.79	79.02	3.87	2.75	5.41
19	256347.76	255377.21	249265.26	13.85	14.22	13.59	82.84	84.08	82.54	3.31	1.70	3.87
20	307365.03	311458.87	308189.68	12.92	13.38	12.64	83.50	84.61	81.94	3.58	2.01	5.41
21	271868.12	271312.27	272209.30	14.48	15.02	14.06	82.39	83.17	81.65	3.13	1.81	4.29
22	192016.94	189888.89	192036.66	18.52	19.18	18.51	76.30	77.20	76.01	5.19	3.62	5.47
23	291552.68	288039.31	290615.67	11.43	11.69	11.46	85.39	86.31	84.85	3.18	2.00	3.69
24	266643.81	263881.39	262820.36	15.71	16.35	15.36	81.19	81.84	80.39	3.10	1.81	4.25
25	254510.86	256554.96	256164.55	16.33	17.26	16.58	79.88	80.90	79.69	3.79	1.84	3.72
26	224646.59	224887.74	227552.05	14.80	15.10	14.57	82.39	83.44	81.29	2.80	1.47	4.14
27	225699.97	226628.65	226834.86	18.65	19.06	18.69	76.65	77.65	76.33	4.70	3.30	4.98
28	240885.16	242069.87	245273.95	16.21	16.97	16.26	79.34	80.67	79.18	4.45	2.36	4.56
29	235328.76	234959.97	236773.20	15.74	16.27	15.94	80.68	81.79	79.93	3.58	1.93	4.13
30	289372.50	283866.90	286750.09	14.28	14.58	14.41	82.58	83.41	81.47	3.15	2.02	4.12
31	242646.71	241697.20	249653.81	14.50	15.10	14.44	81.66	82.60	81.27	3.84	2.30	4.29
32	247013.83	246945.14	250703.93	17.82	18.36	17.56	78.47	79.11	77.84	3.71	2.53	4.61
33	286443.10	289140.00	289877.44	13.26	13.64	13.12	82.82	83.74	82.45	3.92	2.62	4.43
34	185558.50	187640.75	185876.95	22.64	22.87	22.08	70.44	70.76	69.52	6.92	6.37	8.40
35	219318.52	222786.61	219680.68	15.67	16.01	15.49	80.42	81.70	80.84	3.91	2.29	3.68
36	272438.09	270219.88	272112.68	14.86	15.37	14.87	81.19	82.06	80.19	3.96	2.56	4.94
37	237536.92	237874.66	236947.20	16.80	16.90	16.46	79.34	80.72	79.06	3.86	2.38	4.48
38	204170.28	205844.94	207637.33	20.44	21.47	20.10	74.75	75.66	73.82	4.81	2.87	6.08
39	265501.50	261580.04	265141.07	14.99	15.66	15.00	81.64	82.53	80.85	3.37	1.81	4.15
40	171423.08	172358.18	174071.63	22.44	23.37	22.09	71.49	71.56	70.23	6.07	5.07	7.68
41	214935.62	215353.53	215472.47	14.18	14.89	14.15	81.40	82.40	80.65	4.42	2.71	5.20
42	232896.20	232829.15	233064.99	18.09	18.90	17.74	76.73	77.79	75.95	5.18	3.31	6.31
43	244980.43	244005.22	245391.83	15.28	16.00	15.25	80.16	81.50	79.85	4.56	2.51	4.90
44	175499.03	174230.75	172561.44	22.70	23.07	22.32	71.41	72.09	70.67	5.89	4.83	7.00
45	247558.89	247564.65	249154.55	16.35	16.80	16.05	79.35	80.71	79.19	4.30	2.48	4.76
46	220402.40	220374.42	222552.01	18.74	19.28	18.56	76.69	77.09	76.18	4.57	3.63	5.26
47	276844.97	279839.56	282323.82	12.54	13.35	12.66	82.95	84.39	82.43	4.51	2.26	4.91
48	261675.42	262016.38	266776.83	18.23	19.04	18.06	77.58	78.77	76.70	4.19	2.20	5.25
49	268097.46	269965.88	268894.58	15.67	15.88	15.47	81.62	83.20	81.48	2.71	0.92	3.06
50	279509.92	281799.05	282174.56	15.04	15.38	14.74	81.06	81.87	80.41	3.91	2.76	4.86
Avg.	239326.33	239320.19	240517.66	16.95	17.51	16.78	78.68	79.60	78.01	4.37	2.90	5.21

APPENDIX C

CHAPTER THREE APPENDIX

C.1 Proofs of propositions

Proof of Proposition 3.4.1. Let $g(n_i, n_j)$ be the term that contains two successive variables n_i and n_j in the objective function (3.7) for a solution $[n_1, \ldots, n_i, n_j \ldots, n_k]$, in which $n_i \geq n_j > 0$, where $i < j$. Consider another solution $[n_1, \ldots, n'_i, n'_j \ldots, n_k]$ with $n'_i = n_j$ and $n'_j = n_i$ such that $n'_j \geq n'_i > 0$, which has the term $g(n'_i, n'_j)$ associated with n'_i and n'_j. We then have

$$\begin{aligned}
g(n'_i, n'_j) - g(n_i, n_j) &= n'_i \left(\sum_{\ell=1}^{i-1} \sqrt{n_\ell} + \sqrt{n'_i} \right) + n'_j \left(\sum_{\ell=1}^{i-1} \sqrt{n_\ell} + \sqrt{n'_i} + \sqrt{n'_j} \right) \\
&\qquad - n_i \sum_{\ell=1}^{i} \sqrt{n_\ell} - n_j \left(\sum_{\ell=1}^{i} \sqrt{n_\ell} + \sqrt{n_j} \right) \\
&= (n'_i - n_i) \sum_{\ell=1}^{i-1} \sqrt{n_\ell} + n'_i \sqrt{n'_i} - n_i \sqrt{n_i} \\
&\qquad + (n'_j - n_j) \sum_{\ell=1}^{i-1} \sqrt{n_\ell} + n'_j \left(\sqrt{n'_i} + \sqrt{n'_j} \right) - n_j \left(\sqrt{n_i} + \sqrt{n_j} \right) \\
&= n_j \sqrt{n_j} - n_i \sqrt{n_i} + n_i \left(\sqrt{n_j} + \sqrt{n_i} \right) - n_j \left(\sqrt{n_i} + \sqrt{n_j} \right) \\
&= n_i \sqrt{n_j} - n_j \sqrt{n_i} \\
&= \left(\sqrt{n_i} - \sqrt{n_j} \right) \sqrt{n_i n_j}. \qquad \text{(C.1)}
\end{aligned}$$

The right-hand side of Equation (C.1) is a nonnegative value as $n_i \geq n_j$, implying the new solution with $n'_i \leq n'_j$ never improves the objective value. This argument can be applied to any pair of variables for $i = 1, \ldots, N-1$ and $j > i$. Thus, a solution with $n_1 \geq \ldots \geq n_N$ will lead to a minimum objective value. ■

Proof of Proposition 3.4.2. When $1 < N \leq C$, the maximum possible number of routes to run equals N with $n_1 = \ldots = n_N = 1$ and an objective value of $(N + 1)/2$, which can always be improved by another solution in which $n'_1 = 2$, $n'_2 = \ldots = n'_{N-1} = 1$ and $n'_N = 0$ with an objective value of $N\sqrt{2} + (N-2)(N-1)/2$ (it is straightforward to show that $N\sqrt{2} + (N-2)(N-1)/2 \leq (N+1)/2$ holds for any $N \geq 2$). The latter is optimal for $N = 2$, but can be further improved for $N \geq 3$. However, $n_1 = \ldots = n_N = 1$ will never lead to an optimal solution in either case. Thus, the maximum possible number of routes in an optimal solution is no more than $N - 1$. ■

Proof of Proposition 3.5.1. Let y_2^* be the optimal value of y_2 associated with $V_2(x,\tau)$. We have

$$\begin{aligned} V_2(x,\tau) - V_1(x,\tau) &= y_2^* \left(a - b(\tau + \beta\sqrt{y_2^*})^c\right) + (x - y_2^*)\left(a - b(\tau + \beta\sqrt{y_2^*} + \beta\sqrt{x - y_2^*})^c\right) \\ &\qquad - x\big(a - b(\tau + \beta\sqrt{x})^c\big) \\ &= by_2^*\left(\tau + \beta\sqrt{y_2^*} + \beta\sqrt{x - y_2^*}\right)^c - by_2^*\left(\tau + \beta\sqrt{y_2^*}\right)^c \\ &\qquad + bx\left(\tau + \beta\sqrt{x}\right)^c - bx\left(\tau + \beta\sqrt{y_2^*} + \beta\sqrt{x - y_2^*}\right)^c \\ &\geq by_2^*\left(\tau + \beta\sqrt{y_2^*} + \beta\sqrt{x - y_2^*}\right)^c - by_2^*\left(\tau + \beta\sqrt{y_2^*}\right)^c \\ &\qquad + bx\left(\tau + \beta\sqrt{x}\right)^c - bx\left(\tau + \beta\sqrt{y_2^* + x - y_2^*}\right)^c \\ &= by_2^*\left(\tau + \beta\sqrt{y_2^*} + \beta\sqrt{x - y_2^*}\right)^c - by_2^*\left(\tau + \beta\sqrt{y_2^*}\right)^c. \end{aligned} \tag{C.2}$$

Because $y_2^* \leq x$, the right-hand side of the last equation is nonnegative, and therefore $V_2(x,\tau) \geq V_1(x,\tau)$. ■