Analysis of Shipment Consolidation in the Logistics Supply Chain

by

Muhammed Ali Ülkü

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Management Sciences

Waterloo, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

M. Ali Ülkü

Abstract

Shipment Consolidation (SCL) is a logistics strategy that combines two or more orders or shipments so that a larger quantity can be dispatched on the same vehicle to the same market region. This dissertation aims to emphasize the importance and substantial cost saving opportunities that come with SCL in a logistics supply chain, by offering new models or by improving on the current body of literature.

Our research revolves around "three main axes" in SCL: Single-Item Shipment Consolidation (SISCL), Multi-Item Shipment Consolidation (MISCL), and Pricing and Shipment Consolidation. We investigate those topics by employing various Operations Research concepts or techniques such as renewal theory, dynamic optimization, and simulation.

In SISCL, we focus on analytical models, when the orders arrive randomly. First, we examine the conditions under which an SCL program enables positive savings. Then, in addition to the current SCL policies used in practice and studied in the literature, i.e. Quantity-Policy (Q-P), Time-Policy (T-P) and Hybrid Policy (H-P), we introduce a new one that we call the Controlled Dispatch Policy (CD-P). Moreover, we provide a cost-based comparison of those policies. We show that the Q-P yields the lowest cost per order amongst the others, yet with the highest randomness in dispatch times. On the other hand, we also show that, between the service-level dependent policies (i.e. the CD-P, H-P and T-P), H-P provides the lowest cost per order, while CD-P turns out to be more flexible and responsive to

dispatch times, again with a lower cost than the T-P.

In MISCL, we construct dispatch decision rules. We employ a myopic analysis, and show that it is optimal, when costs and the order-arrival processes are dependent on the type of items. In a dynamic setting, we apply the concept of time-varying probability to integrate the dispatching and load planning decisions. For the most common dispatch objectives such as cost per order, cost per unit time or cost per unit weight, we use simulation and observe that the variabilities in both cost and the optimal consolidation cycle are smaller for the objective of cost per unit weight.

Finally on our third axis, we study the joint optimization of pricing and time-based SCL policy. We do this for a price- and time-sensitive logistics market, both for common carriage (transport by a public, for-hire trucking company) and private carriage (employing one's own fleet of trucks). The main motivation for introducing pricing in SCL decisions stems from the fact that transportation is a service, and naturally demand is affected by price. Suitable pricing decisions may influence the order-arrival rates, enabling extra savings. Those savings emanate from two sources: Scale economies (in private carriage) or discount economies (in common carriage) that come with SCL, and additional revenue generated by employing an appropriate pricing scheme.

Throughout the dissertation, we offer numerical examples and as many managerial insights as possible. Suggestions for future research are offered.

Acknowledgments

Dedication

To loving memories of my late

mother, Fatma Ülkü,

sister, Esin Ülkü Kılıç,

father, R. Niyazi Ülkü.

You have the most influence on who I am now.

I will always miss you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation aims to emphasize the importance and substantial cost saving opportunities that come with transportation in a logistics supply chain. "We" (you, the reader and myself, the author) will specifically study the mechanics of shipment or freight consolidation, and will obtain as many managerial insights as possible.

A *Logistics Supply Chain* refers to the network of logistics parties in a supply chain. *Logistics supply chain management* is concerned with the flow of raw materials, parts, work-in-process, and finished products needed to ensure that the company's customers receive finished products at the correct time, in the correct location, and in the correct amount. A logistics supply chain may also describe a

company that has no manufacturing activities, such as a retailer or a third-party logistics provider (e.g. Shapiro, 2001).

Shipment Consolidation (SCL) is a logistics strategy that combines two or more orders or shipments so that a larger quantity can be dispatched on the same vehicle to the same market region. This may enable considerable economies of scale, greatly reducing the transportation cost per item, per order, or per unit weight. The challenge however is to determine a policy for shipping consolidation that still gives good service to the customers whose orders are among the first to be placed. Now, let us look at a simple consolidation problem at work.

Example 1.1. (Adapted from Bookbinder and Higginson, 2002) Suppose that, at the end of each day, 6,000 pounds of glass fibre roofing material is shipped from Waterloo to Ottawa. Assuming a constant production rate, an inventory holding cost of ten cents per hundred pounds (hundredweight) per day, and a transportation charge of \$2.95 per hundredweight (cwt), the total transportation and inventory holding cost for this shipment is [60 cwt×\$2.95/cwt]+[(60/2)cwt×\$0.10/cwt]=\$180/day. Then, the total weekly cost is 5×180/day =\$900. Now suppose that there is an all-unit discount scheme with volume rate of \$2.07/cwt for the weights of loads larger than 25,000 lbs. So, if we decide to consolidate (i.e. combine daily shipments) over a week, the total weekly transportation and inventory holding cost would be 300

cwt$\times$\$2.07/cwt+(300/2) cwt$\times$(\$0.10/cwt)$\times$5 days=\$696/week. Therefore, consolidating and then shipping at the end of the week rather than shipping daily creates a saving of \$204, which is almost 23% better than with daily shipments.

In the cost calculation of Example 1.1, we assumed that the shipper outsources the transportation service, as in *common carriage* (a public, for-hire trucking company). Alternatively, the shipper might have used *private carriage* (one's own truck). (A review of costing issues with respect to various carrier types can be found in Higginson, 1993). Because the cost of operating a vehicle in a private fleet largely depends on distance rather than on load size, the dispatch of a vehicle each week instead of each day could reduce transportation cost by as much as 80%.

SCL is not limited to just one product or to one destination. For instance, if several small orders, none qualifying individually for the lower freight rate, were destined for different customers in the same geographical region, the shipper could make one consolidated shipment under volume-freight (discount) rates to a central facility. There, the load would be disaggregated for local delivery to individual customers.

Moreover, suppose the manufacturer of glass fibre roofing material also sends small loads to a customer in Kingston. By combining these shipments with those to Ottawa, the resulting consolidated weight may be sufficiently large to qualify as

Figure 1.1: SCL Advertised on the Web

a volume shipment, even though a portion of the load will be removed in Kingston, with a stop-off before the vehicle reaches its final destination.

Another example of SCL is displayed in Fig. 1.1 (Source:www.bongous.com). That advertisement of a freight-forwarding company recognizes the fact that you can consolidate different type of items, and thus may obtain significant savings by having them sent to you in one box. The savings implied there is almost $265. i.e. more than 50% by the use of SCL.

Logistics planning includes three types of major consolidation methods: Vehicle, inventory and terminal consolidation, cf. Hall (1987) Vehicle consolidation involves the practice of aggregating two or more small shipments into a single truck. Inventory consolidation focuses on the number, type, and location of stocking points. Terminal consolidation brings items from different locations into a single facility

4
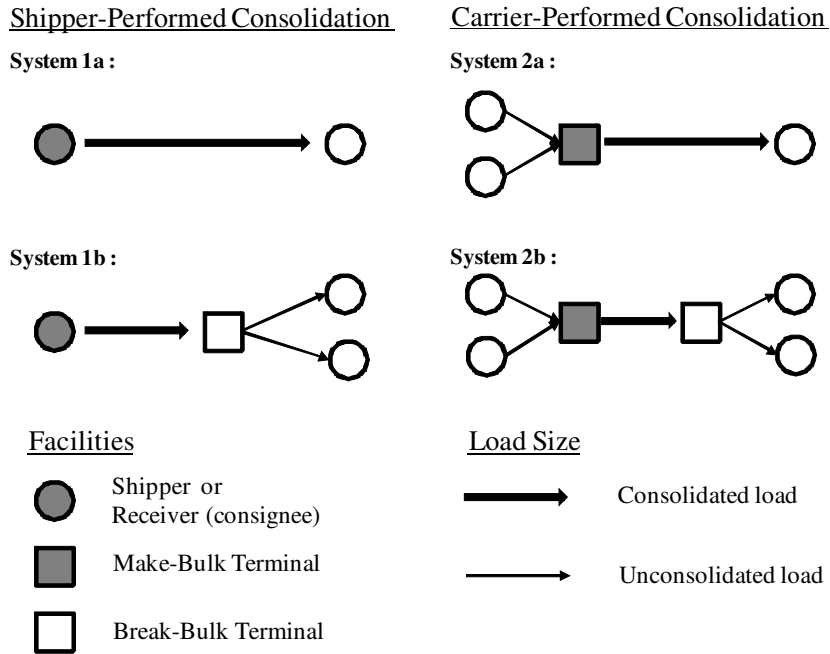
Figure 1.2: SCL Configurations

(a "terminal") where they are sorted, loaded into vehicles and taken to different destinations.

Fig. 1.2 (based upon Higginson, 1992) displays possible distribution system configurations in which SCL can be useful. For example, a shipper can consolidate its orders destined to a specific destination point, as in System 1a. Or in System 2b, a carrier may consolidate orders from different shippers at a make-bulk terminal, line-haul a particular lane, and then break-bulk that consolidated load at the destination terminal for the local deliveries to the end customers.

Transportation and inventory costs are greatly impacted by consolidation strate-
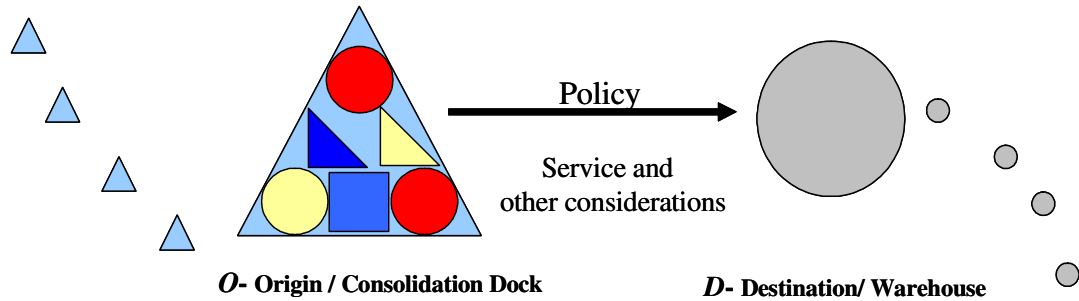
Figure 1.3: Single Stocking-point, Single-route Consolidation

gies within various supply chain configurations. To exemplify, one of many such consolidation configurations is displayed in Fig. 1.3

In that figure, the loads that are destined to a single point D, say a retailer, are consolidated at a single point O, say a warehouse. As depicted, the composition of the loads may be varying in size (volume, weight) or type. The best consolidation policy depends on various constraints such as management policies and objectives, topology of the logistics network, customer instructions, required transit time, product and transportation characteristics, and the cost parameters included in the model.

Goods may be moved to their destinations by several major modes, when applicable (i.e. by air, sea, rail carrier or motor carrier). The latter, transportation by truck, accounts for the largest share and it raises many opportunities for consolidating small shipments.

Small shipments, generally weighing less than 10,000 lbs, are an important part

of most businesses' total traffic volume. The majority of them, however, weigh less than 1,000 lbs., for example see Tyworth et al. (1987). The trucking industry is well suited to line-haul small loads and thus dominates this segment of the transportation market.

Many trucking firms use discounts to encourage shippers to consolidate freight into larger, but still less-than-truckload (LTL), weight groups above 5,000 lbs. That enables more efficient operations. Therefore, building full truckload shipments should not necessarily be the primary focus of consolidation programs, whether applied by the shipper or the carrier.

Foremost, any consolidation program imposes some administrative costs for program planning and management. Benefits are mainly derived from lowered transportation costs and better transportation operations. On the other hand, consolidation sometimes may lengthen the order cycle and thus adversely affect customer service.

Scale economies in transport operations make it possible for carriers to haul larger shipments at lower rates per unit. Shipment consolidation mostly favors the carrier's pickup, delivery and dock-handling costs. For example, a truckload shipment requires only two stops by the carrier: one for pickup at the origin and one at the destination. By contrast, small shipments require the *LTL* carrier to

make more stops for pickup and delivery. Moreover, it may not even be economical for the carrier to linehaul some operations in which shipments are so small, and pickup or delivery points so scattered.

Another advantage in consolidation shows up in transportation service. Consolidation may allow for faster and consistent transit times, which in turn would result in reduced inventories (safety or in-transit) without changing customer-service standards. Moreover, with faster transit times, capital is tied up in the consignment for a shorter time, and fast deliveries may generate earlier payments and speed cash flow, e.g. Masters (1980).

Higginson and Bookbinder (1994) generalize the types of decisions when developing a SCL program. They identify the major questions to be asked for a SCL program as:

- *What* will be consolidated? Which customer orders or products need be consolidated and which shipped alone?

- *When* will customer orders be released? What event(s) will trigger the dispatch of a vehicle containing a consolidated load?

- *Where* will the consolidation be done? Should consolidation take place at the factory, on a vehicle, at a warehouse or terminal?

- *Who* will consolidate? The manufacturer, shipper, customer, carrier, or a third party?

- *How* will consolidation be carried out? Which specific techniques will be employed?

Adequate answers to these questions boil down to understanding the basic trade-offs in applying consolidation policies, which impact virtually all areas of a firm's logistics system.

In practice, typical SCL policies comprise the *Time Policy* (T-P), *Quantity Policy* (Q-P), and *Hybrid Policy* (H-P) which is also known as the Time-and-Quantity Policy. A T-P dispatches each order at a predetermined shipping date (T^*), whether or not it is consolidated. In a Q-P, all orders are held and shipped when a minimum consolidated quantity (Q^*) is reached. Under the H-P, dispatching occurs upon attaining the earliest of "predetermined shipping date T^*" or "the accumulation of a minimum weight or volume Q^*." Because the Q-P uses a quantity-based dispatch policy, it requires continuous review of the accumulating load. One needs to be cautious in implementing this policy for two reasons: the orders might not be easily tracked, or the cost of tracking such orders might overwhelm the benefits of consolidation; and the order cannot be given a delivery-time guarantee. However, the T-P enables us to give a time-guarantee while not requiring the cost of monitoring the arrivals of orders.
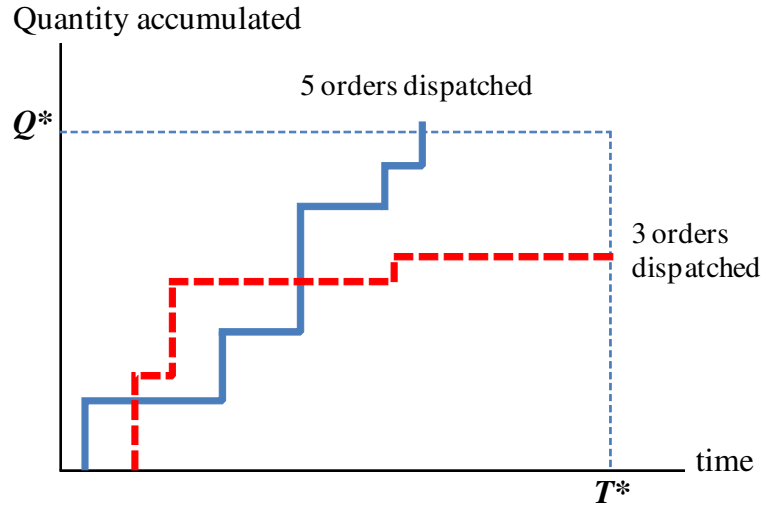
Figure 1.4: Sample Paths for the Q-P and T-P

As seen in the Fig.1.4, the time it takes to build up to Q^* orders (optimal dispatch quantity) might be less than T^* (optimal time at which load should be dispatched.) Also, we note that the quantity built up by time T^* might be less than Q^*.

This dissertation evolves around "three main axes" in SCL:

1. Analysis of shipment consolidation *policies* under random arrivals of orders, for a single type of freight

2. For various *objectives*, analysis of shipment consolidation for *multiple* item types and random arrivals of orders, with a distribution of order sizes

3. Analysis of shipment consolidation when the transportation-*pricing* decisions are made by the various *carrier types*

Next, we will give a **survey of literature** that has commonality for the rest of the thesis, and then outline the contributions of this dissertation. (Particular references to other studies will be given in the relevant chapters.)

The global supply chain relies on the effective and efficient use of transportation networks, which vary substantially, depending on the region, country, or function. Both intercontinental and regional transport become increasingly important in the context of global business, e.g. Skjott-Larsen et al. (2007). In that trend, shipment consolidation has again become an active research venue in logistics supply chain.

Besides general discussions in textbooks (for example Tyworth et al., 1987; Silver et al., 1998; Ballou, 2004), we may classify the literature on shipment consolidation as those articles usually published in the trade press (e.g. Newbourne and Barrett, 1972; Newbourne, 1976; Bookbinder, 1989) and those academic works published in scholarly journals.

There are successful applications of SCL in industry. For example, in the heavy petroleum industry by Mobil Oil, as reported by Bausch et al. (1995). Quinn (1997) reports that shipment consolidation enabled Nabisco Inc. to cut transportation costs by half, decrease inventory levels, and improve on-time delivery.

To our knowledge, a systematic approach to analyzing consolidation appears first in Beckmann et al. (1953), in the context of the uses of railway-switching

yards, where a train can transfer its small loads to another one, heading for the same destination. (Also see Beckmann et al., 1956). The academic literature on shipment consolidation may be divided into analytical models, simulation studies and heuristic methods. The pertinent analytical models are supported by various operations research techniques such as optimization, renewal processes, and queueing theory. The heuristic approaches, on the other hand, develop decision rules based on data analysis.

An important heuristic study is that of Schuster (1979), who presents an analysis of the economics of shipment consolidation. Drawing upon the transportation cost figures, he makes recommendations about how to revise rates to induce shippers to combine small loads, and to create effective shipment-consolidation-marketing strategies. An early simulation analysis appears in Jackson (1981), to investigate the major variables operating in an order-consolidating system. In another paper (Jackson, 1985), he surveys more than fifty US firms that apply freight consolidation. His survey results point out "cost reduction" to be the most important reason these firms engage in freight consolidation programs, and the biggest disadvantage to be the staff "effort" that is required to plan and operate these systems. Therefore, he emphasizes that the primary objective of managers of consolidation programs should be to increase consolidation-cost-savings *per employee*.

Using a stochastic clearing model, Gupta and Bagchi (1987) find the minimum cost-effective load which should be accumulated at a consolidation center, when the procurements are done just-in-time. In Buffa (1987), the effects (on the decision to consolidate particular items) of cost factors and transit times are investigated. Akaah and Jackson (1988) report on their empirical study of fitting various theoretical distributions to the order weights, and emphasize that they vary from industry to industry. Min and Cooper (1980) survey and classify the literature after 1980. Pooley and Stenger (1992) empirically test how factors such as level of LTL discount rate, order size, and geographic distribution of customer-demand influence the shipment consolidation and logistics cost performance. Jaruphongsa et al. (2007) provide a solution to an integrated dispatch problem when both LTL and FTL operations are allowed.

Based upon industry practice, Higginson and Bookbinder (1994) study the timing issue for shipment-release using simulation. Assuming Poisson order-arrivals and Gamma-distributed order weights, they analyze the relative cost and delay performance of the three shipment-release policies (Q-P, T-P, and H-P). For various combinations of holding time and arrival rates, those authors provide recommendations on when to use which policy, depending upon management's objectives with regard to customer service and cost.

Higginson (1995) distinguishes shipment-consolidation procedures as *recurrent* or *non-recurrent*. Assuming the choices of the optimum minimal quantity to dispatch and/or the optimal timing of dispatch have been decided in advance, a non-recurrent method proceeds in the same manner in each consolidation cycle. However, in a recurrent approach, as the arrival of each new order changes the consolidated load, a decision is made on whether to ship or continue to consolidate, based on that total quantity available and the length of the consolidation cycle to that point. Especially for consolidation cycles in which orders are too few or too small, dispatching early may ameliorate poor customer service. To investigate this situation, he uses marginal analysis.

Higginson and Bookbinder (1995) discuss the use of Markovian decision processes in a recurrent approach to freight consolidation programs. The dispatch rule is again based on the aggregated weight, rather than the individual orders that compose it. Optimal stationary policies are obtained, both for the private-carriage and the common-carriage cases. For common carriage, those authors show that shipment is the preferred action with very large *or very small* accumulated weights. On the other hand, for private carriage, they report an optimal policy of control-limit type. Several numerical examples are given.

Bookbinder and Higginson (2002) obtain practical decision rules for temporal

consolidation for the case of private carriage. Using a "nomograph," they relate the maximum holding time and desired dispatch quantity in a probabilistic setting. Tyan et al. (2003) evaluate different freight consolidation policies at an integrated global logistics company, using a mathematical programming model. The stochastic case for the Q-P and the T-P is analyzed in Çetinkaya and Bookbinder (2003). They employ renewal theory and obtain the optimal target weight before dispatch and the optimal length of each consolidation cycle. Those authors report that T-P is superior to Q-P on the basis of mean time before dispatch, hence may provide a better service.

Çetinkaya (2004) gives a review of shipment consolidation at large, in a coordinated setting with inventory decisions in supply chain management. In Chen et al. (2005), Q-P and T-P are analyzed in the context of a vendor who uses a lot-size, reorder policy to replenish the stock. They find that the Q-P scheme outperforms the T-P counterpart. Recently, Çetinkaya et al. (2006) investigate the cost- and customer-waiting-time impacts of using a hybrid policy, and show that H-P is superior to T-P. In sum, the preceding references illustrate that a given policy may exhibit its positive features to a greater extent in a particular setting or for different assumptions.

We also notice that some concepts and techniques for the consolidation problem

may be borrowed from queueing theory. The policies used for bulk or batch service queues (e.g. Deb and Serfozo, 1973) may prove useful in our analysis, although they are constructed for different operating environments. Consider a manufacturing setting with a single server where the arrivals follow a Poisson process, and each arrival seizes a general service time (i.e. M/G/1 system in queueing theory parlance). Yadin and Naor (1963), Heyman (1968), Sobel (1969), and Bell (1971) have studied the N-policy for such a system. An N-policy activates the server when there are N customers waiting for service, and deactivates the server when there are no customers in the system.

Balachandran (1973) and Balachandran and Tijms (1975) considered the D-policy, a control policy which turns the server on when the total work to be done reaches the value D, again continuing until the system is cleared. For a special case of service distribution, they show that the D-policy is superior to the N-policy (cf. Artalejo, 2002). However, to employ either of those policies, the server must continually monitor the queue for an arrival when the server is not active. For situations where this can not happen, Heyman (1977) proposes a T-policy, where the server scans the queue T time units after the end of the last busy period to determine if customers are present. For the minimum-cost-rate criterion, he shows that the optimal cost rate of the T-policy is worse than the one achieved by the comparable optimal N-policy.

According to the model used, the accumulation cycle may begin immediately after the dispatch of a consolidated load or when the first order of a new cycle arrives. If the quantity component is ignored, beginning a new cycle immediately upon load dispatch is similar to releasing shipments at set intervals, with the possibility of cancelling a dispatch if the accumulated weight is too small to be economical. Powell (1985) and Powell and Humblet (1986) investigate this case through bulk-queueing theory for passenger vehicle dispatching.

When we regard the second axis of our research, i.e. the multi-item case, the literature is not exhaustive. References that study multiple items generally fall into the category of deterministic or stochastic knapsack or lot-sizing problems in inventory management (Alp et al., 2003). For example, Anily and Tzur (2005) consider a system in which multiple items are transferred from a warehouse to a retailer through vehicles with identical capacities. In the context of time-varying deterministic demands for a fixed number of items, those authors propose a dynamic programming algorithm that finds a shipment schedule which minimizes the total cost, while satisfying demand on time. Brandimarte (2006), on the other hand, considers the stochastic version of the classical multi-item capacitated lot-sizing problem, where demand uncertainty is explicitly modeled by scenario trees.

Let us now give a brief review on joint pricing and consolidation decisions. The

interaction between the shipper and carrier has significant impacts on logistics costs and service (Sheffi, 1986). New paradigms in costing and pricing decisions for the carrier have evolved in recent years. One such paradigm, *revenue management,* is the practice of controlling the availability and/or pricing of resources in different demand classes, with the goal of maximizing expected revenues or profits. Pricing decisions, given that the carrier has an incentive on them, help him induce effects on the demand pattern of the customers.

Pricing resides at the conjunction of sales and production/inventory functions in a business. Dating back to Whitin (1955), the interactions between the pricing and inventory control decisions have been and still are being researched. Chan et al. (2004) give an excellent treatise on the coordination of pricing and inventory decisions. However, in logistics strategy, it is now being recognized that revenue generation is as important as cost reduction (Ballou, 2006). Transportation revenue-management, i.e. "yield management," has been on the research stage for the last two decades, yet there exists a large room for improvement in this field (McGill and Van Ryzin, 1999).

Although *service* might be viewed as the "product" of logistics systems, the literature on consolidation strategies with service considerations is limited and deserves a thorough investigation. There exist various approaches which incorporate

a service constraint into the tactical or operational level of logistic decisions. These depend on whether management's objective is to minimize cost, or to maximize profit or maximize utility of logistics service. For instance, in a physical distribution system with fixed demand and price, Bookbinder and Lynch (1997) maximize utility of the customers. Ray and Bookbinder (2005) work on profit maximization for the company, when customers now are price sensitive. In the context of customer lead-time management, Ray and Jewkes (2004) study the effects of a time guarantee on profit maximization when both demand and price are lead-time sensitive, and a particular service level (fill rate) is required. Some other studies pertinent to pricing in the transportation context and which relate to our work include Cavinato (1982), Lee and Rosenblatt (1986), Abdelwahab and Sargious (1990), and Abad and Aggarwal (2005).

Our **contributions** to the body of literature in SCL can be outlined as follows:

1. We add to the literature on single-item shipment consolidation by providing new results on the currently-used SCL policies, and by offering a new shipment consolidation policy, namely *Controlled Dispatch Policy.*

2. While proposing a new multi-item SCL model, we study the behavior of optimal dispatch decisions under various objectives.

3. We emphasize the importance of pricing to obtain better performance of SCL,

19

for an integrated logistics supply chain. In doing so, we propose various pricing schemes that can be used by a carrier.

4. Overall, we propose new models, or new results about the already existing SCL models, by employing various Operations Research methodologies.

In Chapter 2, we will focus on single-item shipment consolidation for SCL policies. The multiple-item shipment consolidation, and dispatch decision rules for various objectives will be the main concern in Chapter 3. Chapters 4 and 5 look closely at the case when the order arrival process is influenced by pricing decisions, for the case of common-carriage and private-carriage, respectively. And finally, we conclude with an overall view of the findings and possible research extensions.

Chapter 2

Single-Item Shipment

Consolidation: New Results

Single-Item Shipment Consolidation (SISCL), as the name suggests, focuses on the shipment of a particular product type. This is the basic version of SCL. For example, various retail stores place orders with a supplier of that item. A distribution center may combine those orders and dispatch them as a consolidated linehaul shipment, for eventual local deliveries to the stores.

In this chapter, we first investigate the conditions under which a consolidation program may enable savings compared to the case of non-consolidation. When these conditions hold, savings can be achieved both for private-carriage and common-

carriage transport. Later, we study the commonly used SCL policies (Quantity, Time, and Hybrid), assuming that orders of a standard (unit) size arrive according to a Poisson process. We also introduce the "Constrained Dispatch Policy." With the objective of minimizing expected cost per unit quantity, we compare those four policies based upon theoretical and numerical results.

2.1 When does SCL make sense?

Depending on the operating environment of the supply chain and the orders for the product or products, a consolidation program may not be at all beneficial to the shipper or carrier. This is especially true if the orders are very time sensitive (e.g. express delivery), or the product is fast-perishable (e.g. blood or some pharmaceutical or food items).

Based upon the Example 1.1, we consider a simple setting in which a consolidation program might potentially be applicable. For a stationary stochastic system, we can derive the conditions when consolidation is a viable strategy. To do so, let us first define Λ to be the constant arrival rate of orders per unit time where an order is charged a carrying cost of $\$h$ per unit time during the interval that it is held before dispatch. Also, let κ be the fixed cost to dispatch a vehicle of capacity W orders. Propositions 2.1 and 2.2 give conditions when a consolidation program

is beneficial, respectively for private carriage and common carriage.

Proposition 2.1. Consider a sequence of shipments over an interval of η periods (i.e. time units) by a private fleet that can supply multiple trucks as required. Denote by a "non-consolidation" program the dispatch of a vehicle each period (often containing a partial load). A "consolidation program" will wait, dispatching a full vehicle one or more times during the η-period interval. (A partially-full truck will be dispatched at the end of this interval if necessary.) Then, a sufficient condition for savings from the consolidation program over the η periods is

$$
\begin{aligned}
(\kappa D_{\mathrm{C}} - \eta \kappa D_{\mathrm{NC}}) + (hW^2/2\Lambda)[(D_{\mathrm{C}}-1) - \eta(D_{\mathrm{NC}}-1)] \\
+ (h/2\Lambda)[(\eta\Lambda - (D_{\mathrm{C}}-1)W)^2 - \eta(\Lambda - (D_{\mathrm{NC}}-1)W)^2] \\
+ (h/2\Lambda)[(\eta\Lambda - (D_{\mathrm{C}}-1)W)^2 - \eta(\Lambda - (D_{\mathrm{NC}}-1)W)^2] < 0
\end{aligned}
\tag{2.1}
$$

$$
\text{where} \quad D_{\mathrm{C}} = \lceil \eta\Lambda/W \rceil \text{ and } D_{\mathrm{NC}} = \lceil \Lambda/W \rceil
$$

Proof. For private carriage, the fixed cost of dispatching a load is independent of the amount shipped. However, if the accumulated load is more than a single vehicle's capacity, then the quantity remaining is to be shipped at the end of the period on another vehicle, only partially loaded. The total cost per dispatch would comprise the holding cost and the dispatch cost. The expected total number of vehicles employed each period is $D_{\mathrm{NC}} \triangleq \lceil \Lambda/W \rceil$, where subscript NC denotes "non-consolidation". Without loss of generality, suppose there are $D_{\mathrm{NC}}-1$ fully loaded, and one partially loaded vehicle. The expected time required to fill a ve-

23

hicle is W/Λ. (Naturally, in a period, if enough orders are accumulated before the end of that period, a vehicle is dispatched as soon as it is fully loaded.) For such an order, the average wait time is $W/2\Lambda$. Thus, the holding cost for one full truck is $hW(W/2\Lambda)$. Now, in a period, the average waiting time for an order that is shipped in a (possibly) partially loaded truck is $[1-(D_{\mathrm{NC}}-1)W/\Lambda]/2$ periods. Also, the total amount for the last (possibly partial) load to be dispatched at the end of the period is $\Lambda[1-(D_{\mathrm{NC}}-1)W/\Lambda] = \Lambda - (D_{\mathrm{NC}}-1)W$. Hence, the holding cost for a partial dispatch is $h[\Lambda - (D_{\mathrm{NC}}-1)W][1-(D_{\mathrm{NC}}-1)W/\Lambda]/2$. Including the cost of fixed cost of dispatching D_{NC} trucks in a period, the total cost of a "non-consolidation" program over η periods is found to be, $TC_{\mathrm{NC}}(\eta) \triangleq$

$$\eta\left(\kappa D_{\mathrm{NC}} + hW^2(D_{\mathrm{NC}}-1)/(2\Lambda)+h[\Lambda - (D_{\mathrm{NC}}-1)W][1-(D_{\mathrm{NC}}-1)W/\Lambda]/2\right).$$

Let us now denote by $TC_{\mathrm{C}}(\eta)$ the cost of a "consolidation" program over η periods. In such a program, a vehicle will be held until enough orders are received to fill it. (If the order arrival rate is less than the vehicle capacity, a vehicle will be held more than one period.) Define a dispatch cycle as the length of time to fully load a vehicle. The number of cycles, therefore the number of vehicles dispatched in η periods, is $D_{\mathrm{C}} \triangleq \lceil \eta\Lambda/W \rceil$. There will be for sure $(D_{\mathrm{C}}-1)$ fully loaded vehicles and (possibly) one partially loaded. Similar to $TC_{\mathrm{C}}(\eta)$, but using the concept of cycle, we obtain $TC_{\mathrm{C}}(\eta) = \kappa D_{\mathrm{C}}+hW^2(D_{\mathrm{C}}-1)/(2\Lambda)+h[\eta\Lambda - (D_{\mathrm{C}}-1)W][\eta-(D_{\mathrm{C}}-1)W/\Lambda]/2$.

There will be savings in using SCL only if η is chosen such that $TC_{\mathrm{C}}(\eta) <$ $TC_{\mathrm{NC}}(\eta)$. After some algebra, we obtain the condition (2.1). ∎

For the case of common carriage, i.e. when transport is conducted by a for-hire trucking company, the shipper pays only for the quantity of freight (in our case, the number of orders) to be sent, at a transportation rate or price per order set by the carrier. Because each order is assumed to be a standard size and weight, the vehicle capacity W can be measured in "orders." We note that this capacity is an issue not for the shipper, but for the carrier.

Let us define the "non-volume" transportation rate, in \$/order, as f_n. The carrier, however, offers a freight-rate discount to the shipper so that the carrier itself can benefit from scale economies. Denote the volume-freight rate as f_v, dollars per order transported; that rate is activated when the number of orders reaches or exceeds a particular break-point W_b. (We consider an all-unit discount with single break-point. Naturally, $f_n > f_v$.) As before, $\eta = \eta_0 \Upsilon$ is the time interval (and η_0 is the number of periods) in which a consolidation program is applied.

Proposition 2.2. Suppose the shipper utilizes common carriage and that $\Lambda\Upsilon < W_b$ (i.e., the expected total orders in a single period do not qualify for the volume rate). Then, there exist positive savings over a consolidation interval of η_0 periods,

if and only if the following conditions hold:

$$W_b/\Lambda\Upsilon \leq \eta_0 \leq 1+2(f_n-f_v)/h\Upsilon \qquad (2.2)$$

Proof. As before, the relevant costs of the dispatch will comprise holding costs plus transportation cost. For common carriage, we require a consolidated load of sufficient size to qualify for the volume rate. Hence, our first condition is $\Lambda\eta \geq W_b$, for $\eta_0 > 1$. (We note that the case $\Lambda\Upsilon \geq W_b$ is easy: The expected load available in a single period would exceed the order-break point W_b, and hence there is no need for a "purposeful" program of SCL.) Now, the total cost of "ship in each period" (i.e., whatever orders are received during a period, ship them at the end of that period) will be $\eta(\Lambda f_n+\Lambda h\Upsilon/2) = \eta\Lambda(f_n+h\Upsilon/2)$, whereas that of consolidation over the time interval η yields a total cost of $\eta(\Lambda f_v+\Lambda\eta h/2) = \eta\Lambda(f_v+\eta h/2)$.

It is now easy to see that if the difference between the former and the latter cost figures is strictly positive, inequalities (2.2) will hold. Conversely, if conditions (2.2) are satisfied, a purposeful SCL program yields positive savings. ∎

Note that if consolidation is sensible for the common-carriage case, the percent of savings with respect to non-consolidation can be obtained as

$$\% \text{ savings by consolidation} = [(f_n-f_v)+(h\Upsilon/2)(1-\eta_0)]/(f_n+h\Upsilon/2)$$

For instance, the savings in Example 1.1 can be found by this formula to be 23%, when the specific parameters are $f_n = \$2.95/\text{order}$, $f_v = \$2.07/\text{order}$, $h =$

$0.10/\text{order}/\text{day}$, and $\eta = 10$ days.

2.2 Policy Analysis in SISCL

In the previous section, we obtained conditions under which SCL is desirable. Now, assuming that there are positive benefits from such a program, we wish to derive structural results concerning the optimal consolidation *policies*.

As before, we model the dispatch of a consolidated load consisting of the accumulated orders of a single item. Transport is by private carriage. We analyze a simple but non-trivial consolidation setting where the product is unit-sized, and the arrival process is Poisson. We deliberately study unit-sized orders so that the accumulated quantity at any time is simply the number of orders.

The problem at hand can be cast as follows: A retailer is solely supplied by a single vendor for a specific product. The demand for that product follows a Poisson distribution with rate λ; thus the inter-demand times follow an exponential distribution with mean interarrival time $1/\lambda$. Aligned with our previous notation, let h be the holding cost per item per unit time. Let us define Q to be the random variable representing the total quantity consolidated. Orders are accumulated and then dispatched according to a particular release policy. Shipment of orders is done on a first-come, first-served fashion.

The admissible consolidation policies we analyze here include the Time Policy (T-P), Quantity Policy (Q-P) and Hybrid Policy (H-P). In this section, we introduce a new policy that finds the optimal quantity to dispatch, given a controlled dispatch time. We name this policy as *Controlled Dispatch Policy,* CD-P in short.

The main difference between CD-P and H-P is that H-P contains only a single dispatch time, the target holding time T^*. That is, for a consolidation cycle in which orders arrive somewhat infrequently, H-P will not dispatch until time T^* as long as $Q < Q^*$, the targeted quantity obtained from the optimization of Q-P. However, for CD-P, its optimal target quantity q^* (which is now a function of "controlled" or scheduled dispatch times) is allowed to decrease as time goes on: Dispatch will occur sooner in a "disappointing" consolidation cycle, rather than delay the inevitable shipment of a load smaller than q^* until a later dispatch time.

CD-P is intended to be near-optimal for a given *range* of dispatch times. CD-P in that way furnishes a more realistic service than H-P; this is the main motivation for CD-P.

Throughout our analysis, we assume that the delivery of each load is conducted by one's own truck with a constant cost κ per shipment. Hence, the consolidation costs only involve the inventory carrying and fixed transportation cost per load.

We now focus on the aforementioned shipment-release policies, beginning with

the Q-P. Its formulation assumes that the dispatcher ships the consolidated load whenever Q or more units are accumulated. The expected total cost $C(Q)$ of this policy is $C(Q) = \kappa + hQ(Q-1)/2\lambda$, and the expected cost per order is thus

$$\bar{C}(Q) = C(Q)/Q = \kappa/Q + h(Q-1)/2\lambda \qquad (2.3)$$

By treating the cost function as a smooth one, it is easy to show that $\bar{C}(Q)$ is convex in Q, i.e. $\frac{d^2\bar{C}(Q)}{dQ^2} \geq 0$. Let $x \wedge y \triangleq \min\{x, y\}$. Now, via first-order conditions, and constraining Q by the vehicle capacity W, we find the minimizer of $\bar{C}(Q)$ as

$$Q^* = \sqrt{2\kappa\lambda/h} \wedge W \qquad (2.4)$$

(In a *manufacturing* setting, Burns et al. (1985) consider the inventory holding cost at both origin and destination, and obtain a result similar to Eq. (2.4).) Excluding the case when $Q^* = W$, the optimal expected cost of the Q-P is

$$\bar{C}(Q^*) = \sqrt{2\kappa h/\lambda} - h/2\lambda$$

The quantity in this model setting is integer. Proposition 2.3 offers the following easy-to-use decision rule for finding Q^*_{int}, the cost-minimizing integer value of Q^*, cf. Ülkü and Bookbinder (2006).

Proposition 2.3. The optimal integral dispatch quantity for the Q-P is

$$Q^*_{\text{int}} = \begin{cases} \lfloor Q^* \rfloor \equiv Q^*_{\text{L}}, & \text{if } 2\kappa\lambda/h \leq Q^*_{\text{L}} Q^*_{\text{U}} \\ \lceil Q^* \rceil \equiv Q^*_{\text{U}}, & \text{o.w.} \end{cases}$$

Proof. $Q^*_{\text{int}} = Q^*_{\text{L}}$ if $\bar{C}(Q^*_{\text{L}}) \leq \bar{C}(Q^*_{\text{U}})$. Expanding this condition, and employing

29

the facts all the parameters are positive, $Q_L^* > 0$, and $Q_U^* - Q_L^* = 1$, the desired result is obtained. ∎

Example 2.1. Suppose $\kappa = 10, h = 1, \lambda = 0.5$. Then $Q^* = \sqrt{2\kappa\lambda/h} = \sqrt{10}$ and $Q_L^* = 3, Q_U^* = 4$. By Proposition 2.3, since $(3)(4) > 10$, the integral solution for the optimal dispatch quantity is $Q_{int}^* = Q_L^* = 3$. (Note that $\bar{C}(3) = 5.33 < \bar{C}(\sqrt{10}) = 5.32 < \bar{C}(4) = 5.5$.) However, suppose κ is changed to 32 instead. Then, $Q^* = \sqrt{32}$ and $Q_L^* = 5, Q_U^* = 6$. Again by Proposition 2.3, since $(5)(6) \not\geq 32$, we conclude that now $Q_{int}^* = Q_U^* = 6$.

Next, we turn our attention to the formulation of the T-P, in which our development follows that of Ülkü and Bookbinder (2006). Consider that such a policy is applied for a cycle length of T. Define X_n to be the time between the $(n-1)^{\text{st}}$ and the nth order, with $E[X] = 1/\lambda < \infty$. Without loss of generality, assume that the first arrival is at time 0. Then $S_n = \sum_{i=1}^{i=n} X_i$ will represent the arrival time of the nth order. Via renewal theory, and for Poisson arrivals with rate λ, the expected number of renewals in the time interval $[0, T]$, $m(T)$, can be shown to equal λT. Now, let us define $A(T) = T - S_{m(T)}$ as the age of the last order prior to or at time T. Using the fact that $E[A(T)] = 1/\lambda$ (See for example Çınlar, 1975) we can derive

the expected total cost of the T-P during a time interval T as

$$
\begin{aligned}
C(T) &= \kappa + E[hX_1 + 2hX_2 + \dots + [m(T)-1]hX_{m(T)-1}] + hm(T)E[A(T)] \\
&= \kappa + hE[X]m(T)[m(T)-1]/2 + hm(T)E[A(T)] \\
&= \kappa + h\lambda T^2/2 + hT/2
\end{aligned}
$$

Hence, the expected average cost per unit time is obtained as

$$
\bar{C}(T) = C(T)/m(T) = hT/2 + h/2\lambda + \kappa/\lambda T \tag{2.5}
$$

Define the maximum holding time (i.e. service level) by $\mathrm{T_{max}}$. Noting that $\bar{C}(T)$ is convex in T, the optimal cycle length of T-P is found to be

$$
T^* = \sqrt{2\kappa/h\lambda} \wedge \mathrm{T_{max}} \tag{2.6}
$$

where the optimal cost value of the T-P is simply $\bar{C}(T^*) = \sqrt{2\kappa\lambda/h} + h/2\lambda$, ignoring $T^* = \mathrm{T_{max}}$.

To formulate the H-P, let us define its minimal expected cost per order by

$$
\bar{C}(Q^*, T^*) = \bar{C}(\text{minimum time for } \{Q^*, T^*\}) = \bar{C}(Z(Q^*) \wedge T^*) \tag{2.7}
$$

where $Z(Q^*)$ is Gamma-distributed with mean time $1/\lambda$ and shape Q^*. This definition makes all three policies comparable on the same measure, the expected cost per order. Now, we are ready to prove

Proposition 2.4. The Q-P yields a lower cost per order than the T-P and H-P.

Proof. First we will show that Q-P gives lower cost per order than the T-P. Since

31

$h, \lambda > 0$, $\bar{C}(Q^*) = \bar{C}(T^*) - h/\lambda < \bar{C}(T^*)$. Now, by (2.7), we can rewrite

$$\bar{C}(Q^*, T^*) = \begin{cases} \bar{C}(Q^*), & \text{if } Z(Q^*) \leq T^* \\ \\ \bar{C}(T^*), & \text{o.w.} \end{cases}$$

Let $\theta = \Pr\{Z(Q^*) \leq T^*\}$, i.e. the probability that quantity policy portion of the H-P will be active before the T-P portion. Now, suppose the H-P gives a lower cost than Q-P. Then,

$$\bar{C}(Q^*, T^*) = \mathrm{E}\bar{C}(Z(Q^*) \wedge T^*) = \theta\bar{C}(Q^*) + (1-\theta)\bar{C}(T^*) = \theta[\bar{C}(Q^*) - \bar{C}(T^*)] + \bar{C}(T^*) <$$

$\bar{C}(Q^*)$ should hold. However, this inequality is satisfied only for $\theta > 1$, which is contrary to the very definition of a probability measure. ∎

Corollary 2.1. In terms of cost per order, H-P is superior to T-P.

Proof. For some θ such that $0 < \theta < 1$, let the cost difference between Q-P and H-P be Δ. Thus by Proposition 2.4, $\Delta = \bar{C}(Q^*, T^*) - \bar{C}(Q^*) \geq 0$. Then, the cost relations of these policies yield, $\bar{C}(T^*) = [\bar{C}(Q^*, T^*) - \theta\bar{C}(Q^*)]/(1-\theta) = [\bar{C}(Q^*, T^*) - \theta(\bar{C}(Q^*, T^*) - \Delta)]/(1-\theta) = \bar{C}(Q^*, T^*) + \theta\Delta/(1-\theta) > \bar{C}(Q^*, T^*)$. This completes the proof. ∎

Corollary 2.2. H-P gives the same expected cost as Q-P if $\sqrt{2\kappa\lambda/h} \to 0$, and gives the same expected cost as of T-P if $(2\kappa\lambda/h)(-1+h/\lambda) \geq 3$.

Proof. We investigate two extreme cases: Consider the case $\theta = 1$, i.e. the required time to reach optimal Q-P parameter is probabilistically always less than that of

the T-P. This case implies that $\theta = \Pr\{Q^* \text{ or more orders arrive before } T^*\} = 1 - \sum_{j=0}^{Q^*-1} \frac{(\lambda T^*)^j e^{-(\lambda T^*)}}{j!} = 1$. That condition is satisfied when $\lambda T^* = \sqrt{2\kappa\lambda/h} \to 0$, which can be interpreted as, "The best policy is to ship each order whenever received," if the fixed dispatch cost relative to the holding cost and the arrival rate are very small.

Now consider the case when $\theta = 0$, i.e. $\sum_{j=0}^{Q^*-1} \frac{(\lambda T^*)^j e^{-(\lambda T^*)}}{j!} = 1 - \theta = 1$. An exact and explicit expression for this term is not possible. Fortunately, we can approximate a Poisson distribution with rate λT^*, by a Normal distribution with mean λT^*, and variance λT^* when λT^* is "large" (generally when $\lambda T^* > 10$.) Now, from the characteristics of Normal distribution, $1 - \theta \to 1$ if $Q^* > 3\sqrt{\lambda T^*} + \lambda T^*$. Employing (2.4) and (2.6), and using the fact that $\sqrt{x} \geq (1/x)$ for $x \geq 1$, we obtain the condition $(2\kappa\lambda/h)(-1 + h/\lambda) \geq 3$. This completes the proof. ∎

In passing, we note that our analytical findings in Corollary 2.2 are congruent with those in Higginson and Bookbinder (1994). We would also like to note that results pertaining to the comparison of Q-P, T-P and H-P have been published in Ülkü and Bookbinder (2006).

Let us now investigate a policy when the dispatch time is set in advance. Beckmann et al. (1956) study, among other topics, the freight operations and the classification policy in railroad systems. In that work (pages 134-135), those authors
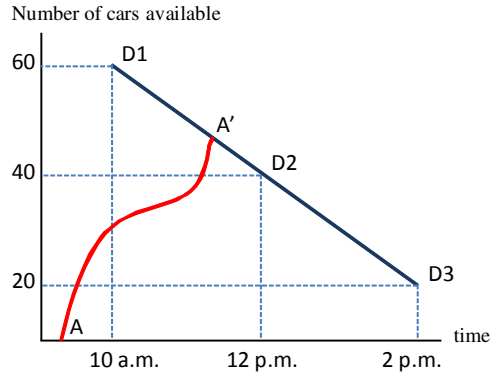
Figure 2.1: Dispatching Trains in a Yard

suggest a dispatching rule for the outbound trains in a yard. They *conjecture* that the relationship between required accumulation rate decreases as the dispatch time approaches. They write, "The important point in this discussion is not why the curve DD' should have such a slope only that it is possible to ensure against absurdly long trains (by making the schedule of departure times dependent on the numbers of cars of the proper kinds which have accumulated), and at the same time not completely abandon the timetable type of scheduling."

Fig. 2.1 is based upon Beckmann et al. (1956). The DD' line quoted in the preceding paragraph corresponds to line D1-D3 in this figure. The y-axis shows the number of outbound cars available to be attached to a train before its dispatch, and x-axis shows the possible times of departure (schedule). The curve depicts a sample path of accumulation; and suggests that the train should be dispatched *before* its scheduled departure time (12 p.m.) because for an economical dispatch, a sufficient

34

number of cars (i.e. 40 cars for 12 p.m.) has already been reached.

To our knowledge, there is no published research on why this is the case. Specifically, why is D1-D3 linear and sloped downward? This observation motivates us to introduce and study a new dispatch policy, what we name as "Controlled Dispatch Policy." The optimization of that policy forms the topic of the next section.

2.3 Controlled-Dispatch Policy

CD-P finds the optimal quantity (accumulation amount) *given* a controlled or scheduled dispatch time, e.g. for the dispatch of shuttles. It is similar to H-P in the sense that it considers the time dimension (service level) of the dispatch.

From (2.3) and (2.5), let us recall that the expected cost per order of the Q-P and T-P are $\bar{C}(Q)$ and $\bar{C}(T)$, respectively. Let us also recall that the H-P was a combination of these two policies, incorporating the optimal Q and T values that are found separately in those policies. Now, for CD-P, to distinguish its decision variables, we call them $q(\tau)$ and τ as opposed to Q and T in H-P. (From now on, for brevity, we will suppress τ in $q(\tau)$, and denote it simply by q. It shall be clear within the context.)

The CD-P dispatches the vehicle as soon as q orders are accumulated or τ time units (which is set previously) has elapsed since the first order arrived. In other

35

words, the consolidated shipments are released whenever q or τ is reached. Now, suppose the first shipment has arrived, at a time defined to be $t = 0$. With respect to our model setting, the probability $\mathrm{p}_{\geq q}$ that we do not have to wait τ time units until the departure is

$$\mathrm{p}_{\geq q} = \Pr\{q \text{ or more orders arrive in a time} \leq \tau\} = 1 - \sum_{j=0}^{q-1} \frac{(\lambda\tau)^j e^{-(\lambda\tau)}}{j!}. \text{ Also,}$$

$$\mathrm{p}_q = \Pr\{\text{ exactly } q \text{ orders arrive by time} = \tau\} = \frac{(\lambda\tau)^q e^{-(\lambda\tau)}}{q!}, \text{ and}$$

$$\mathrm{p}_{<q} = \Pr\{q-1 \text{ or less orders arrive by time} = \tau\} = \sum_{j=0}^{q-1} \frac{(\lambda\tau)^j e^{-(\lambda\tau)}}{j!}.$$

We want to find the variables for a CD-P that gives minimal cost per order for a single shipment type . Notice that q takes on nonnegative integer values, while τ is on the nonnegative real line. Assume that there is no capacity constraint. The expected cost per order when the CD-P is employed, $\bar{C}(q, \tau)$, can then be calculated as

$$
\begin{aligned}
\bar{C}(q, \tau) &= \mathrm{p}_{\geq q}\bar{C}(q) + \mathrm{p}_{<q}\bar{C}(\tau) = \bar{C}(q) + \mathrm{p}_{<q}[\bar{C}(\tau) - \bar{C}(q)] = \bar{C}(\tau) + \mathrm{p}_{\geq q}[\bar{C}(q) - \bar{C}(\tau)] \\
&= \frac{h\tau}{2} + \frac{h}{2\lambda} + \frac{\kappa}{\lambda T} + \left(1 - \sum_{k=0}^{q-1} \frac{(\lambda\tau)^k e^{-(\lambda\tau)}}{k!}\right)\left[\frac{h(q-1)}{2\lambda} + \frac{\kappa}{q} - \frac{h\tau}{2} - \frac{h}{2\lambda} - \frac{\kappa}{\lambda T}\right]
\end{aligned}
$$

Now, the optimality conditions for the CD-P can be stated by

Theorem 2.1. For a given τ, the corresponding optimal quantity q^* must satisfy the following conditions:

$$
\begin{aligned}
\text{i)} \quad & \mathrm{p}_{\geq q^*}\left(\frac{hq(q+1) - 2\lambda\kappa}{2\lambda q(q+1)}\right) - \mathrm{p}_q\left(\frac{h(q^2-1) + 2\lambda\kappa}{2\lambda(q+1)} - \frac{h\lambda\tau^2 + 2\kappa}{2\lambda\tau}\right) > 0, \text{ and} \\
\text{ii)} \quad & \left(\frac{q^*}{\lambda\tau}\mathrm{p}_{q^*} + \mathrm{p}_{\geq q^*}\right)\left(\frac{hq(q-1) - 2\lambda\kappa}{2\lambda q(q-1)}\right) - \frac{q^*}{\lambda\tau}\mathrm{p}_{q^*}\left(\frac{h(q^2-2q) + 2\lambda\kappa}{2\lambda q} - \frac{h\lambda\tau^2 + 2\kappa}{2\lambda\tau}\right) < 0
\end{aligned}
\tag{2.8}
$$

36

Proof. Via finite calculus, given τ, the optimal dispatch quantity q^* with respect to the CD-P need satisfy the condition that $\Delta\bar{C}(q^*-1,\tau) < 0 < \Delta\bar{C}(q^*,\tau)$, where Δ is the first-order difference operator, i.e. $\Delta f(x) = f(x+1)-f(x)$, $x \in \mathbb{Z}$. After some algebra, those conditions can be reduced to $p_{\geq q}\Delta\bar{C}(q)-p_q[\bar{C}(q+1)-\bar{C}(\tau)] > 0$ and $p_{\geq q-1}\Delta\bar{C}(q-1)-p_{q-1}[\bar{C}(q)-\bar{C}(\tau)] < 0$. Expanding these inequalities by further differencing $\bar{C}(q)$ as required, and using the substitutions $p_{q-1} = \frac{q^*}{\lambda\tau}p_{q^*}$, $p_{\geq q-1} = p_{q-1}+ p_{\geq q}$, and via (2.3) and (2.5), the desired conditions in (2.8) are obtained. ∎

Proposition 2.5. The (Q^*,T^*) pair is not necessarily equal to (q^*,τ^*).

Proof. (By contradiction). Assume $q^* = Q^*$. Consider $h = 2, \kappa = 200$, and $\lambda = 2$. From (2.4), $Q^* = \sqrt{2\kappa\lambda/h} = \sqrt{2(200)2/2} = 20$. The expected total time to accumulate Q^* units is $Q^*/\lambda = 10$. Hence, when $\tau = 10$, and $q^* = Q^* = 20$, while condition i) is satisfied (i.e. $0.0993 > 0$), condition ii) is violated (i.e. $0.0726 \not< 0$). Moreover, it turns that $q^* = 18$, contrary to what is assumed to equal $Q^* = 20$.∎

Theorem 2.2. On the basis of cost per order, for a controlled dispatch time τ, H-P is superior to CD-P, and CD-P is superior to T-P.

Proof. Let $\min_q \bar{C}(q,\tau{=}T^*) = \bar{C}(q^*,T^*)$. Since in the CD-P, the costs of unnecessary delays are minimized by q^*, $\bar{C}(T^*) \geq \bar{C}(q^*,T^*)$. However, because the minimization of the CD-P parameters is over a tighter region than for the H-P, $\bar{C}(q^*,T^*) \geq \bar{C}(Q^*,T^*)$. Now, by Proposition 2.4 and Corollary 2.1, we get

$\bar{C}(T^*) \geq \bar{C}(q^*, T^*) \geq \bar{C}(Q^*, T^*) \geq \bar{C}(Q^*)$, and hence the desired result. ∎

In the next section, we study a numerical example applying the CD-P. We then look at its sensitivity to various parameters used in the model.

2.4   Sensitivity of the CD-P

To get more insights about the Controlled Dispatch Policy, we focus on the case of private carriage and employ the base data set below:

$$\kappa = 200, \ h = 2, \ \lambda = 2, \ T_{\max} = 14, \ W = 20$$

First, let us note that the Q-P yields $Q^* = 20$, and the T-P yields $T^* = 10$. Without loss of generality, we will confine our analysis to $q \leq Q^*$ and $\tau \leq T^*$. Our optimization uses enumeration for this example: We assume that the controlled dispatch times are $1, 2, ..., T^*$. Then for fixed dispatch time, we find the cost minimizing q $(1 \leq q \leq Q^*)$ corresponding to that departure time. This procedure is applied to all dispatch times at hand.

To clarify such a search process, Fig. 2.2 shows that when time to dispatch is 5 time units, the optimal dispatch quantity is 12. We also observe that as the accumulation amount increases, the costs of the Q-P, T-P, and CD-P converge. Also, we note that when $q \geq 9$, Q-P and CD-P give lower costs than the T-P.
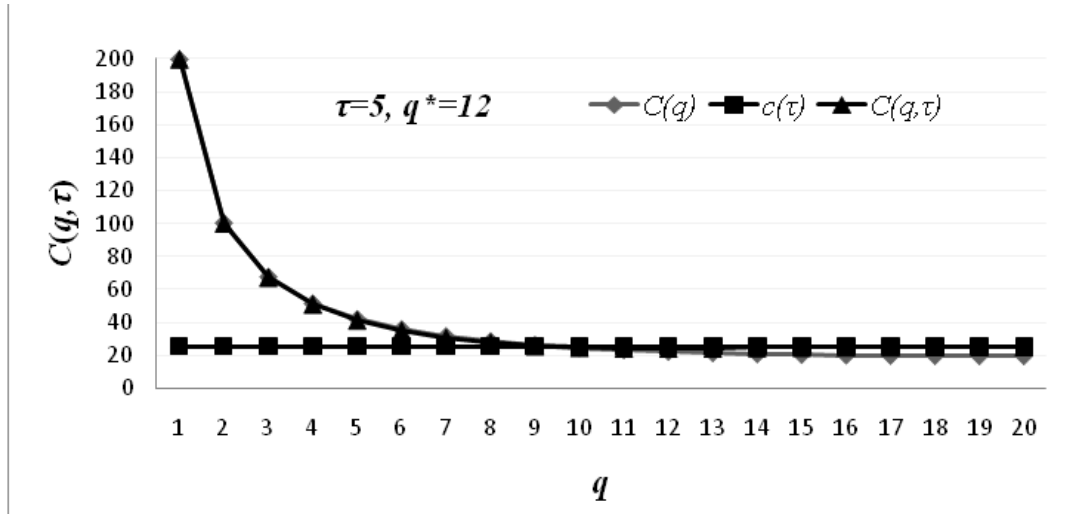
Figure 2.2: An Example of Optimization in the CD-P

Fig. 2.3 displays the complete optimal dispatch quantities for each dispatch time. We note that these values decrease as time to dispatch increases. That relation is *not necessarily linear*, unlike the linear relation conjectured in Beckmann et al. (1956).

It is observed in Fig. 2.4 that as the arrival rate increases, the optimal dispatch quantities for the CD-P and the H-P converge to the same value. However, for lower values of the arrival rate λ, CD-P suggests smaller dispatch quantities, and hence is more conservative than H-P. We also observe a similar relationship between the CD-P and H-P when the ratio of the fixed dispatch cost to that of holding is high enough in Fig. 2.5.

Overall, CD-P turns out to be a more conservative policy than H-P. Although,

39
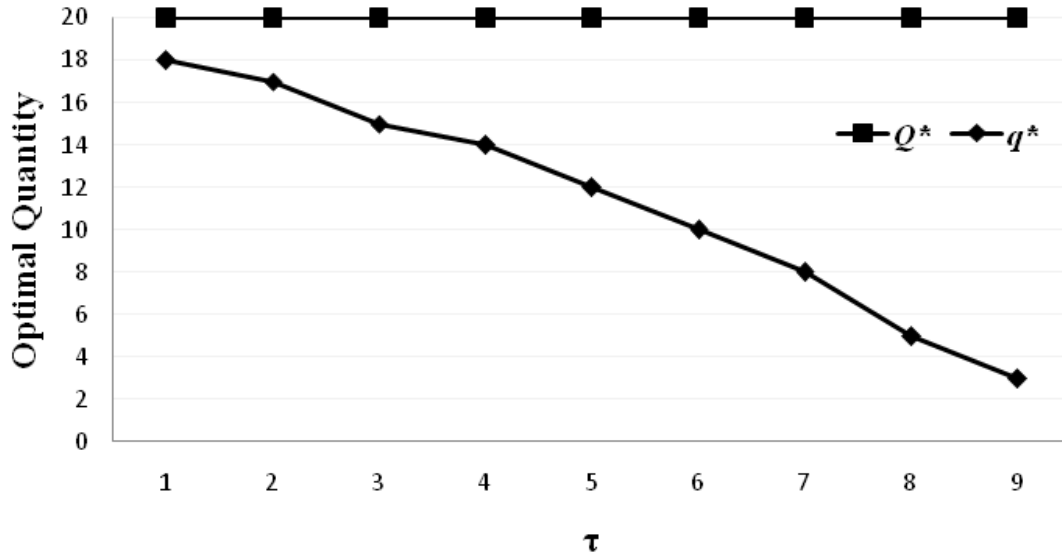
Figure 2.3: Optimal CD-P vs. H-P Dispatch Quantities

for fixed dispatch costs, it yields a larger cost per order for lower arrival rates, the performance of CD-P is less variable than that of the H-P. That is a desired quality to have streamlined operations in a logistics supply chain.

2.5 Concluding Remarks

This chapter investigated the conditions under which shipment consolidation would enable savings, both for the private and common carriage. We studied the commonly used SCL policies, for a single type of item and for random order arrivals. We introduced a new consolidation policy, CD-P. For the case of private carriage, and on the basis of cost per order, we found the cost relationships between all those
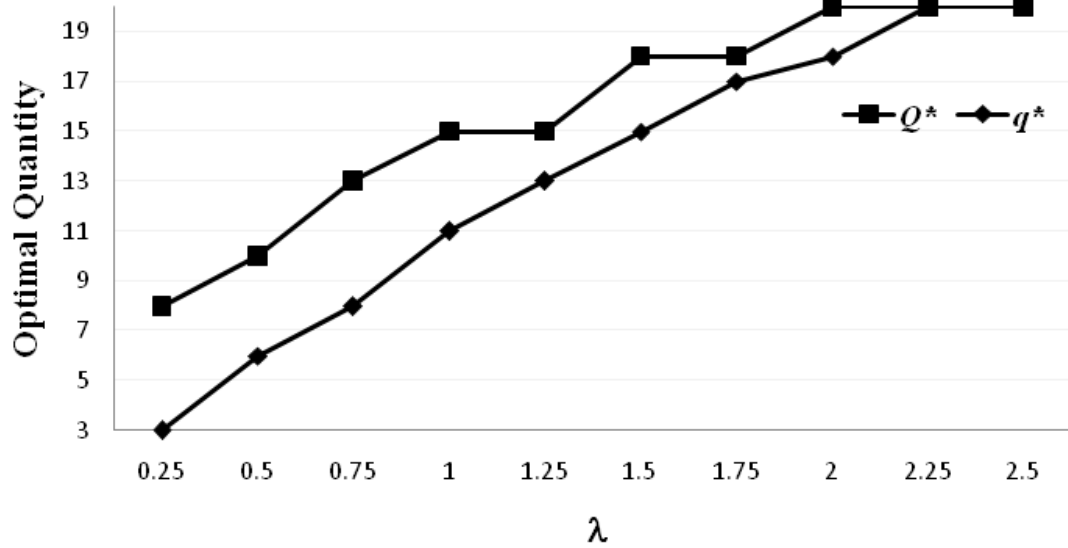
Figure 2.4: Sensitivities of the CD-P and H-P to Arrival Rate

policies, i.e. Theorem 2.2.

One extension to the current analysis herein would be to compare the SCL policies for the common-carriage case. Also the study can be extended to the other types of objectives, such as cost per unit time and cost per hundred weight, when the orders have random weights.

Those typical objectives will be employed for the analysis of the multiple item shipment consolidation, which is the topic of the next chapter.

41
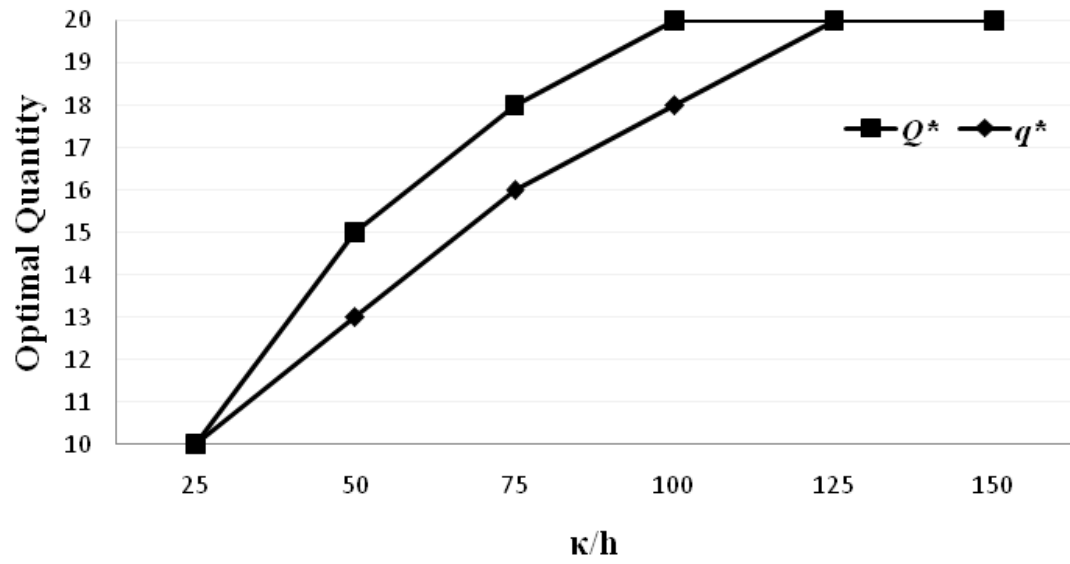
Figure 2.5: Sensitivies of the CD-P and H-P to Dispatch cost/Holding cost

Chapter 3

Multi-Item Shipment

Consolidation

Shipment Consolidation (SCL) proves most useful in LTL transportation, in which small loads with different characteristics are carried together. Therefore, *Multi-Item* Shipment Consolidation (MISCL) is a better representation of industrial practices, while requiring a more involved analysis. In this chapter, we introduce the Multi-Item Shipment Consolidation Problem (MISCLP), and study (cost-minimizing) consolidation decisions in which the small shipments, or orders, come from different types or classes of items. Our focus here is to model and then to analyze decisions on shipment dispatch ("shipment release") in the case of private carriage,

for typical objectives such as cost per order, cost per unit volume, cost per hundred weight and cost per unit time. We employ a myopic analysis in which the decisions of continuing to consolidate a load ("hold"), versus shipping it now ("dispatch"), are to be made at the *arrival* of each order. This would be the case where the resources are available to continually measure and record the weights of each arriving order, similar to a continuous review system in inventory management.

The literature on MISCL is scant. The work closest to ours is Higginson (1995). He employs a similar myopic approach (referred to as "recurrent decisions" in the paper) in the case of a single item with simpler costing mechanisms. He compares the findings with a deterministic Economic Shipment Quantity, and proposes practical insights on when a recurrent approach might be beneficial over a non-recurrent one.

Again for a single item, Bookbinder and Higginson (2002) obtain practical decision rules for temporal freight consolidation for a private carrier by employing results from stochastic-clearing systems. For a similar setting and including the case of common carriage, Çetinkaya and Bookbinder (2003), using renewal theory, propose stochastic models for the dispatch of consolidated shipments, again for one type of item. Brandimarte (2006) considers the stochastic version of the classical multi-item capacitated lot-sizing problem, where demand uncertainty is modeled

by *scenario trees*. Also, a multi-product stochastic dispatch problem is studied by Papadaki and Powell (2003). As opposed to myopic policy we devise for MISCLP, they apply an "approximate" dynamic programming algorithm to minimize the total cost, over a finite horizon of *discrete* periods.

On the other hand, Anily and Tzur (2005) study the *deterministic* version of the problem of shipping multiple items on vehicles, using dynamic programming. In the conclusion of their paper, they say "It is usually complex to generalize results obtained for the single-item case to settings with multiple items of different characteristics. For example, in capacitated shipments, if the items differ in their volume or weight specifications, or both, it is likely that the solution method for the problem would have to include or be combined with a bin-packing procedure." Our MISCLP model does bring in these issues, and in a *stochastic* setting. The justification of a myopic approach is offered in Section 3.3.

Another recent work on deterministic multi-item dispatch problem is that of Dror and Hartman (2007). Unlike our approach, those authors study the *cost allocations* for multiple items that are to be consolidated and shipped together. Using game theory, they show that if the portion of ordering cost common to all items is not too small, then the cost allocations are "fair," i.e. the core of the game is not empty.

The capacitated lot sizing problem and joint replenishment problem, a well-studied problem in the inventory literature (see Khouja and Goyal (2008) for a recent review), can be regarded as "close" to MISCLP. However, there are some fundamental differences: While lot-sizing problems emphasize savings from inventory costs, SCL focuses on transportation savings in the supply chain. Moreover, the objectives for MISCLP is well dependent on the weight, volume and the variety of the items (e.g., cost per hundred weight), whereas in lot-sizing models focus is more on minimizing the production and inventory costs, generally over multiple periods. Again, most of the lot-sizing models rely on periodic review systems while MISCLP, as we defined, requires "continuous" review.

3.1 Does load make-up matter in consolidation?

By the very nature of SCL, one would like to accumulate different types of items (in making up the smaller loads) on the same vehicle. Since now the cost structures are item-specific, we cannot compute the cost of a consolidation policy by simply averaging those costs over all items. For instance, depending on the type of the first-arrived item, the maximum holding time, thereby the transportation service level, might be affected. This observation is clarified by

Example 3.1. Item type i arrives with rate λ_i and incurs a holding cost of h_i

per unit time. Consider two types such that $\lambda_I > \lambda_{II}$, and $h_I < h_{II}$. Assume all the arrivals of products are on the same size pallets. Suppose all that is known is that we received one pallet of type-I item and 2 pallets of type-II item, since the last dispatch. How can we compute the expected total holding cost and how is the service level affected?

There are 3 pallets consolidated. The number of order-arrival patterns amounts to $3!/2!=3$ which are $\phi_1 =$(I,II,II), $\phi_2 =$(II,I,II), and $\phi_3 =$(II,II,I). Each consolidation cycle starts with the arrival of the first order. It is easy to compute the total holding costs of each pattern: $HC(\phi_1) = h_I(2\lambda_{II}^{-1}) + h_{II}(\lambda_{II}^{-1})$, $HC(\phi_2) = h_I(\lambda_{II}^{-1}) + h_{II}(\lambda_I^{-1} + \lambda_{II}^{-1})$, and $HC(\phi_3) = h_{II}(2\lambda_I^{-1} + \lambda_{II}^{-1})$. Thus, those costs yield different values. We should also note that the holding time for item-type I is longest for the first pattern (ϕ_1).

In the Example 3.1, we observed that the make-up or composition of the consolidated load *does* matter, and affects the SCL decisions. Now the problem pins down to precisely computing the total holding cost of a consolidated load comprised of different types of items, arriving at random times during the consolidation cycle.

3.2 MISCL Problem

The MISCLP tries to find the optimal decision rules that yield the cost-minimizing dispatch time or quantity. Here we try to model this problem at its generality. Then, in the coming sections, we focus on cases where some analytical results can be derived, or where computationally simple yet non-trivial results that may reveal some insights are obtained. Table 3.1 contains the nomenclature that will be used in the MISCLP.

The *realized* total cost of dispatching a load accumulated by time t, $TC_d(t)$, can now be calculated as:

$$TC_d(t) = \kappa + \sum_{i=1}^{N} \left[\kappa_i q_i(t) + \mathbf{I}_R(q_i(t)) \sum_{j=1}^{q_i(t)} h_i k_i(t - A_i^j) \right] \qquad (3.1)$$

$$\text{where } \mathbf{I}_R(q_i(t)) = \begin{cases} 1 & \text{if } q_i(t) \in R = \{1, .., n_i\} \\ 0 & \text{o.w.} \end{cases}$$

Our formulation of the holding cost includes a temporal and a fixed portion that is item-specific. A consolidated order of type i accrues a cost over time until it is dispatched (i.e. $h_i k_i(t - A_i^j)$) plus a fixed cost of handling (κ_i). This formulation gives us a flexibility in our analysis; individual costs are then aggregated over the number of item types. If there is no arrival of a specific type (which might well be the case) through the consolidation period, then there is no cost incurred. (3.1) reflects this particular situation by the indicator function, $\mathbf{I}_R(q_i(t))$. As observed

48

N	: total number of item types
i	: item type index, i=1,...,N
T	: length of a consolidation period
t	: current time since the last dispatch (time marker in a cycle [0,T])
L	: total loading and transportation time for the consolidated shipments
T_{max}	: maximum holding time for the first-arriving order
T^f	: forced dispatch time
T^*	: optimal dispatch time
C_V	: volume capacity of the vehicle
C_W	: weight capacity of the vehicle
κ	: fixed cost of dispatching the vehicle
κ_i	: handling cost for a type-i item
h_i	: holding cost per unit volume per unit time of item-type i
Y_i	: random variable for the interarrival times of item-type i
X_i	: random variable for the weight of a unit volume of item type-i
G_i and g_i	: cdf and pdf of Y_i
F_i and f_i	: cdf and pdf of X_i
λ_i	: Poisson arrival rate for item-type i
A_i^j	: arrival time of the jth item of type-i
k_i	: volume multiplier for type-i item, $k_i \in \{1, ..., C_V\}$
n_i	: max. # of items of type i that can fit the vehicle $(= C_V/k_i)$
V_i	: volume of order-type-i, $V_i = k_i$ (bounded by C_V)
W_i	: random variable representing the weight of type-i order, $W_i = k_i X_i$
W_i^j	: *realized* weight of the jth order of type i
$q_i(t)$	: total number of type-i orders accumulated by time t, $q_i \leq n_i$
Q(t)	: system state vector by time t, $[q_1(t), ..., q_N(t)]$
$Q(t)$	: total number of orders accumulated by time t
$V_r(t)$	: residual volume capacity at time t
$W_r(t)$	: residual weight capacity at time t
p_i	: probability that the next order arrival is of type i
p_d	: probability that a new item arrives before the forced dispatch time
$p^{V_i}(t)$	: probability at time t that a type-i item fits the residual volume
$p^{W_i}(t)$	: probability at time t that a type-i item fits the residual weight
$p_f(t)$	: probability at time t that an arriving item conforms to constraints

Table 3.1: Nomenclature for the MISCLP

in (3.1), we do not discount the total cost. This assumes that the dispatch time is short enough, e.g. a week, and that the items carried are not very expensive.

The uncertainty in our model comes from three sources: the arrival times of the orders, their weights, and their volumes. Below, we will derive the relevant formulas needed in our decision models. We begin with the probability that the weight of a randomly-arriving shipment will conform to the residual, or remaining, weight capacity.

3.2.1   Deriving $p^{W_i}(t)$

The weight constraint is especially important for heavy shipments. There might be plenty of room in the vehicle, yet the weight constraint might impede the loading of an additional quantity to achieve a higher capacity utilization. This is generally the case when the items are "dense." For example, heavy metal products fall into this category.

To derive the probability that an arriving shipment conforms to weight constraint, i.e. $p^{W_i}(t)$, let us first calculate the residual weight capacity by time t $W_r(t)$ by

$$W_r(t) = C_W - \sum_{i=1}^{N} \sum_{j=1}^{n_i} W_i^j \qquad (3.2)$$

50

Recall that W_i denotes the random weight of a shipment of type i. Then,

$$p^{W_i}(t) = \Pr\{W_i \le W_r(t)\}$$

We notice that our formulation of this probability value is time dependent, or in other words system-state dependent with respect to the model. In our setting, it is assumed that orders are weighed as they arrive. Data are recorded on the type of the item, arrival time and weight. (Weight is a single dimension, and thus is easier to measure than volume. Subsection 3.2.2 will expand on these differences.) Hence, upon arrival of each order (item), we have the realized value of its weight. The weight capacity remaining at time t is then deterministic.

We emphasize that the order (shipment) of type i itself is a *multiple* of the generic, unit volume weight of that type (X_i). This is a deliberate choice of modeling, so as to analyze the effects of packaging variability of the shipper on the vehicle packing efficiency of the carrier (the consolidating party). For example, the shipper may wish to pack his products in groups of, say, 12. We prefer not to deal with variability in terms of 12 i.i.d. random variables. That complication would divert the main goal of the analysis in this chapter. For this reason, we model the total weight W_i of an order of type i as the unit volume weight X_i (a random variable) multiplied by a scalar, a positive integer k_i. The shipment type can then easily be calibrated by industrial data: A light but bulky shipment type, for example, can

51

be modeled by a high value of k_i and a low mean (and narrower support) of X_i.

With that justification, we can compute $p^{W_i}(t)$ more explicitly as

$$p^{W_i}(t) = \Pr\{k_i X_i \leq W_r(t)\} \tag{3.3}$$

$$= F_i(W_r(t)/k_i)$$

We now note the following

Observation 3.1. Assuming s>0, t>0, and s+t≤ T, then $p^{W_i}(t) \geq p^{W_i}(t+s)$.

3.2.2 Deriving $p^{V_i}(t)$

What is the probability that an arriving shipment (of random size) will fit on a vehicle? This is not an easy-to-answer question, for the following reasons. First, even if the weight constraint were not violated, the shape of the arriving shipment is not known. The dispatcher might delay unnecessarily, attempting to include that order to include in the consolidated load, and hence leading to extra cost. Second, even in the case of regularly shaped or modular packages, the exercise of fitting a shipment into the truck depends not only on the composition of the accumulated load, but also on how the pallets would be stacked in the vehicle.

This observation leads us to another research area, what is called Vehicle Loading or Load Planning, to which we digress momentarily. The vehicle loading prob-

lem is a three-dimensional Bin-Packing Problem, and is computationally prohibitive to solve to optimality within a reasonable amount of time (Garey and Johnson, 1979). However, there are various heuristic approaches to load a vehicle. The main difference between these heuristics comes from the choice of an *off-line* or an *on-line* loading policy. An on-line packing approach requires each item to be packed into the vehicle as it arrives, whereas an off-line algorithm packs at once all the items that have accumulated over some time. In our model, we assume that an off-line loading mechanism is employed.

The vehicle loading problem becomes still more involved when there are loading restrictions,or if the items or vehicles do not always have regular shapes. For example, pallets of some items can be stacked on top of other shipments but not underneath. Moreover, we note that the allowable loading patterns may differ for a truck, rail car, or the various types of containers. These patterns are also affected by questions of "unit load", i.e. the use of pallets or boxes vs. loose freight. (By a case study, Attanasio et al. (2007) investigates the *issues* in joint dispatching and packing problem.)

Efficient packaging needs to account for loading rules and box orientations such as fixed, horizontal turn, and all-way. Even for the consolidation and packing of only two types of items, Fig. 3.1 shows how the packing problem affects the utilization

of the available space. In that figure, it is shown that when a 2-way orientation is allowed, at most 6 type-1 and 4 type-2 boxes are fit to the vehicle; thus yielding an 80% volume utilization. However, if instead 6-way stacking orientation is employed, the vehicle can be fully packed with 9 type-1 items and 4 type-2 items.

2-Way orientation
6 Type-1 & 4 Type-2 items fitted
Volume used : 80%

6-Way orientation
9 Type-1 & 4 Type-2 items fitted
Volume used : 100%
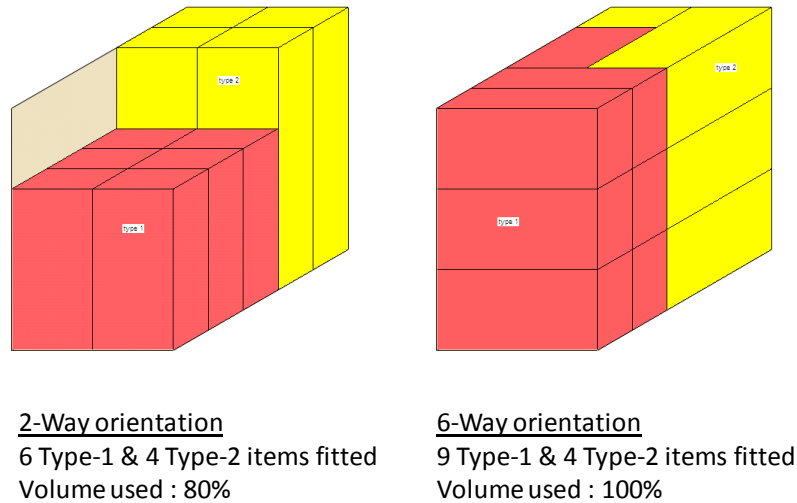
Figure 3.1: Effect of Loading Patterns on Volume Utilization

The preceding observations show that our probability depends not only on the system-state (i.e. composition of the load at time t), but also on how the current load could have been fit into the truck. It is thus extremely difficult to obtain a closed-form expression for the probability of fitting a newly-arrived item on the vehicle, given the item's characteristics and the loading restrictions. Hall (1989) derives approximations for some special cases of vehicle packing, but only for on-line algorithms. Those substantially differ from our problem setting, that of an off-line loading environment. Hence, while our probabilities include optimization

of the shipment load, his do not. (Lim et al. (2005) give a recent survey of three-dimensional packing heuristics.)

Therefore, in general, we will assume that the optimal loading of a given capacity can be done *a priori*, and will be tabulated for various shipment types. The state space of the system will be bounded because the number of the shipment types is finite, and there are only a few variations in truck capacity as well.

We will define the probability that an item of type i will fit in the remaining (residual) space on the vehicle, conditioning on the current load make-up, as

$$p^{V_i}(t) = \Pr\{V_i \leq V_r(t) \,|\, Q(t)\}, \ \text{i=1,...,N}\}$$

Naturally, this probability is defined only for $q_i(t) \leq n_i$. Also, we devise the following

Observation 3.2. Assuming s>0, t>0, and s+t≤ T, then $p^{V_i}(t) \geq p^{V_i}(t+s)$.

3.2.3 Service Level Considerations

In our analysis, T^f will stand for the "forced" dispatch time. That is, the vehicle has to be dispatched at or before T^f, regardless of the amount and the composition of the consolidated load at hand. T^f is related to the service level of the consolidation policy. Though it might be more economical, even optimal in the sense of cost, to wait longer and consolidate a larger load that enables greater utilization of the

truck, service to the first-arriving order might suffer. Hence, we apply a uniform maximum holding time (T_{max}) initiated by the arrival of the first order. We note that T_{max} is an important managerial parameter, and define the forced dispatch time by

$$T^f = T_{max} - L \tag{3.4}$$

where L is the time it takes to load, line-haul, and unload a vehicle. Rather than basing our formulas exclusively upon T_{max}, for our modeling purposes, the form in (3.4) enables us to separate the handling and delivery operations that lie at the heart of logistics. In what follows, we regard L, as well as the costs associated with it, to be deterministic.

T_{max} may be shorter or longer than the optimal dispatch time T^*. For a particular consolidation period, if T_{max} is larger than the optimal dispatch time, then we guarantee a cost-minimizing and timely delivery. If it is shorter, we trade off between the service level (delivery lead time) and the cost that could have been reduced by waiting longer. In our formulation, $T^* < T^f$. Hence, any improvement in loading or unloading will be reflected as a reduction in L. This in turn will increase T^f, a less-constrained value, possibly leading to a T^* of lower total cost.

Note that L itself may be a function of the composition of the load; certain types of items might be less time consuming to load than others. Yet, we will

assume L to be a constant value, determined by the dispatch practices. Also, since the loading operation is achieved in one shot, the arriving items are assumed to be consolidated at a storage area close to the loading dock, and we will assume that the cost of using this space is negligible.

3.3 Dispatch Decisions in MISCL

We have derived the *realized* cost of dispatching a consolidated load by time t in (3.1). However, upon arrival of the shipment at time t, we could have decided not to dispatch but to hold the consolidated load until another shipment arrives, and then dispatch all together. Let us recall that the total number of shipments that have arrived by time t is $Q(\mathrm{t}) = \sum_{\mathrm{i}=1}^{\mathrm{N}} q_{\mathrm{i}}(\mathrm{t})$, and the realized total cost of dispatching $Q(\mathrm{t})$ shipments together is $TC_{\mathrm{d}}(\mathrm{t})$. Suppose instead that this load is held until another shipment arrives at time s > t. Conditioning on the type of the next-arriving shipment, the expected waiting time $\bar{\mathrm{s}}$ before a state transition can be obtained by

$$\bar{\mathrm{s}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{p_i} EY_{\mathrm{i}} \tag{3.5}$$

where EY_{i} represents the expected interarrival time for shipment-type i. Again, conditioning on the type of shipment, the expected handling cost $\bar{\kappa}$ for the next

57

arriving shipment is simply

$$\bar{\kappa} = \sum_{i=1}^{N} p_i \kappa_i \tag{3.6}$$

Denote by p_d an upper bound for the probability that a new item arrives before the forced dispatch time. Then,

$$p_d = \Pr\{\min(Y_1, .., Y_n) \leq T^f\}$$

We can now obtain the probability that the arriving item will conform to the time, volume and weight constraints (i.e. fitting probability) by time t to be $p_f(t)$. Conditioning on the type of the item arriving, we get

$$p_f(t) = p_d \sum_{i=1}^{N} p_i p^{W_i}(t) p^{V_i}(t) \tag{3.7}$$

As will be clear in the development of dispatch-decision rules, our inclusion of $p_f(t)$ in this model possibly enables extra expected savings. For example, this happens when the rule implies to dispatch earlier than the forced dispatch time if the likelihood of a new item arriving within that limited time is zero.

Let us denote by $ETC_h(t)$ the *expected* cost of dispatching $Q(s) = Q(t)+1$ shipments together at time $s > t$. This is the expected cost of "holding" the load until the next arrival of an item. Noting that this cost figure is calculated at time t, we obtain

$$ETC_h(t) = TC_d(t) + EC(t) \quad \text{where } EC(t) = \bar{\kappa} + \bar{s} \sum_{i=1}^{N} h_i k_i q_i(t) \tag{3.8}$$

58

The structure of the decision rules will differ with respect to the objective employed. We investigate those rules for the most commonly used objectives, which follow.

Cost per Order

Using the cost-per-order criteria for multiple items is reasonable for two reasons: Foremost, besides order (item) dependent fixed cost (i.e. handling cost), there is a common fixed vehicle dispatch cost shared by each order. Second, further analysis for other cost criterion is easier to build on that objective. Hence, we begin to derive dispatch rules first on the basis of cost per order.

To minimize the expected cost per order, one needs to solve

$$\min_{d,h}\{TC_d^O(t), \ \mathrm{E}TC_h^O(t)\} \text{ where}$$

$$TC_d^O(t) = TC_d(t)/Q(t)$$

$$\mathrm{E}TC_h^O(t) = \mathrm{E}TC_h(t)/[p_f(t)Q(s)+(1-p_f(t))Q(t)]$$

Note that the expected number of orders will depend on whether the arriving order conforms to the feasibility conditions or not. Hence,

$$p_f(t)Q(s)+(1-p_f(t))Q(t) = p_f(t)(Q(t)+1)+(1-p_f(t))Q(t) = Q(t)+p_f(t) \quad (3.9)$$

The dispatching rule for the cost-per-order objective implies that if $TC_d^O(t) \leq$

$\mathrm{E}TC_\mathrm{h}^\mathrm{O}(\mathrm{t})$, the cost is minimized. That rule can be explicitly written as

$$\text{Dispatch,} \quad \text{if } Q(\mathrm{t}) \geq \mathrm{p_f}(\mathrm{t})TC_\mathrm{d}(\mathrm{t})/\mathrm{E}C(\mathrm{t})$$

$$\text{Hold,} \qquad \text{o.w.} \tag{3.10}$$

Now, we are ready to show

Theorem 3.1. The myopic decision rule to dispatch a consolidated load based on cost-per-order objective, given in (3.10) is optimal.

Proof. Rearranging the terms of the "dispatch" decision, i.e. $TC_\mathrm{d}^\mathrm{O}(\mathrm{t}) \leq \mathrm{E}TC_\mathrm{h}^\mathrm{O}(\mathrm{t})$ and incorporating (3.9), one can derive the simplified decision rule in (3.10). Since only two decisions are admissible, Hold (h) or Dispatch (d), the reverse of the inequality yields the other decision. Let $\Delta C(\mathrm{t})$ denote cost difference between the realized and expected cost per order at time t, i.e. $\Delta C(\mathrm{t}) = TC_\mathrm{d}^\mathrm{O}(\mathrm{t}) - \mathrm{E}TC_\mathrm{h}^\mathrm{O}(\mathrm{t})$. The decision d is meaningful when $\Delta C(\mathrm{t}) \leq 0$. Now, by expanding this term, we get

$$\Delta C(t) = TC_\mathrm{d}(\mathrm{t})[\mathrm{p_f}(\mathrm{t}) - \mathrm{E}C(\mathrm{t})]/[Q(\mathrm{t})(Q(\mathrm{t}) + \mathrm{p_f}(\mathrm{t}))]$$

From (3.8), we observe that $\mathrm{E}C(\mathrm{t})$ is non-decreasing in t since, $\bar{\kappa}$ and $\bar{\mathrm{s}}$ are time-invariant, and since $q_i(\mathrm{s}) \geq q_i(\mathrm{t})$, for $\mathrm{T}^\mathrm{f} \geq \mathrm{s} \geq \mathrm{t} > 0$. Moreover, from Observations 3.1 and 3.2, and using the fact that $\mathrm{p_d}$ is stationary, it easily seen that $\mathrm{p_f}(\mathrm{t})$ is decreasing in t. Now, due to the magnitude in increase of the value of $\mathrm{E}C(\mathrm{t})$ as t increases and the fact that $1 \geq \mathrm{p_f}(\mathrm{t}) \geq 0$, the term $\mathrm{p_f}(\mathrm{t}) - \mathrm{E}C(\mathrm{t})$ will decrease in

t. We note that there exists some t such that $EC(t) \geq 0$ and $p_f(t) = 0$, at which point $\Delta C(t) \leq 0$. Employing the facts that $Q(t)$ is non-decreasing in t, and that the ratio $TC_d(t)/[Q(t)(Q(t)+p_f(t))]$ is decreasing in t, we conclude that $\Delta C(t)$ is monotone decreasing and goes from positive to negative, and hence the myopic policy is optimal. ∎

Similar proofs to that of Theorem 3.1 hold for each of the other dispatch objectives that we consider in the remainder of this chapter. (For brevity, we directly claim the optimality of those related decision rules.) Those objectives show convexity. For instance, Fig. 3.2 exhibits such a behavior.
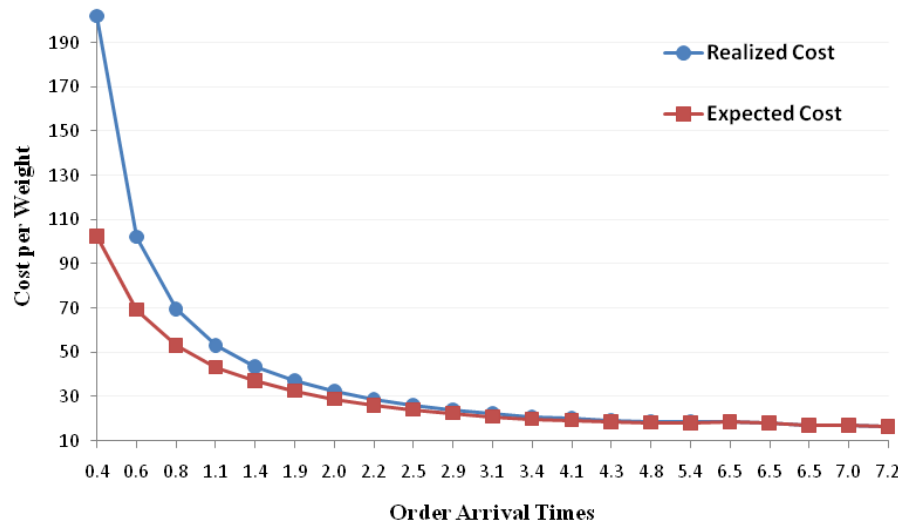
Figure 3.2: A Sample Path of Costs

Cost per Unit Time

To minimize the expected cost per unit time, the objective function is to be

61

modified as follows:

$$\min_{d,h}\{TC_d^T(t),\ ETC_h^T(t)\}\ \text{where}$$

$$TC_d^T(t) = TC_d(t)/t$$

$$ETC_h^T(t) = ETC_h(t)/[p_f(t)(t+\bar{s}) + t(1-p_f(t)]$$

We obtain the corresponding dispatch rule as

$$\begin{aligned} &\text{Dispatch,} \quad \text{if } t \geq \bar{s}p_f(t)TC_d(t)/EC(t) \\ &\text{Hold,} \qquad \text{o.w.} \end{aligned} \tag{3.11}$$

We note that when the probability of conformance is 0, due to time or the volume or weight constraint, the decision rule automatically implies to dispatch the current load.

Cost per Unit Weight

If we wish instead to minimize the expected cost per unit weight, the objective function and the corresponding dispatch rule are

$$\min_{d,h}\{TC_d^W(t),\ ETC_h^W(t)\}\ \text{where}$$

$$TC_d^W(t) = TC_d(t)/(C_W - W_r(t))$$

$$ETC_h^W(t) = ETC_h(t)/[p_f(t)(C_W - W_r(t) + \sum_{i=1}^{N} p_i \int dF_i) + (1-p_f(t)(C_W - W_r(t))]$$

$$\begin{aligned} &\text{Dispatch,} \quad \text{if } W_r(t) \leq C_W - [\sum_{i=1}^{N} p_i \int dF_i]p_f(t)TC_d(t)/EC(t) \\ &\text{Hold,} \qquad \text{o.w.} \end{aligned} \tag{3.12}$$

Cost per Unit Volume

Finally let us consider the minimization of expected cost per unit volume. The corresponding objective function and the dispatch rule now become

$$\min_{d,h}\{TC_d^V(t),\ ETC_h^V(t)\}\ \text{where}$$

$$TC_d^V(t) = TC_d(t)/(C_V - V_r(t))$$

$$ETC_h^V(t) = ETC_h(t)/[p_f(t)(C_V - V_r(t) + \textstyle\sum_{i=1}^{N} p_i k_i) + (1 - p_f(t)(C_V - V_r(t)))]$$

Hence, we

$$
\begin{aligned}
&\text{Dispatch,} &&\text{if } V_r(t) \le C_V - (\textstyle\sum_{i=1}^{N} p_i k_i) p_f(t) TC_d(t)/EC(t)\\
&\text{Hold,} &&\text{o.w.}
\end{aligned}
\tag{3.13}
$$

We now turn our attention to a simpler case of MISCLP.

3.4   2-Item Case

In this section, our focus is on two item-types, i= I, II (see Fig. 3.3). Suppose the arrival processes of each type of item to be Poisson, such that $\lambda_I > \lambda_{II}$. Denote the interarrival times as the random variables Y_I and Y_{II}, exponentially distributed with means $\beta_I = \lambda_I^{-1}$ and $\beta_{II} = \lambda_{II}^{-1}$. Without loss of generality, we will assume that the weight X_1 of a unit volume of a type-I order is a random variable with

63

mean μ_I and variance σ_I^2. Likewise, the weight of a unit volume of type-II order

has mean μ_{II} and variance σ_{II}^2. Let the pdf and cdf corresponding to those random

variables be f_1 and F_1, and f_2 and F_2, respectively. (The weight W_1 of a type-I

order is thus a random variable with mean $k_I\mu_I$ and variance $k_I^2\sigma_I^2$. The weight of

a type-II order, likewise, has $(k_{II}\mu_{II}, k_{II}^2\sigma_{II}^2)$, as its mean and variance.

Type-I item (Fast-moving)
Arrival rate: λ_I
Holding cost: h_I

Type-II item (Slow-moving)
Arrival rate: λ_{II} ($<\lambda_I$)
Holding cost: h_{II} ($>h_I$)

k_I Unit volume weight $\sim (\mu_I, \sigma_I)$

k_{II}

Unit volume weight $\sim (\mu_{II}, \sigma_{II})$
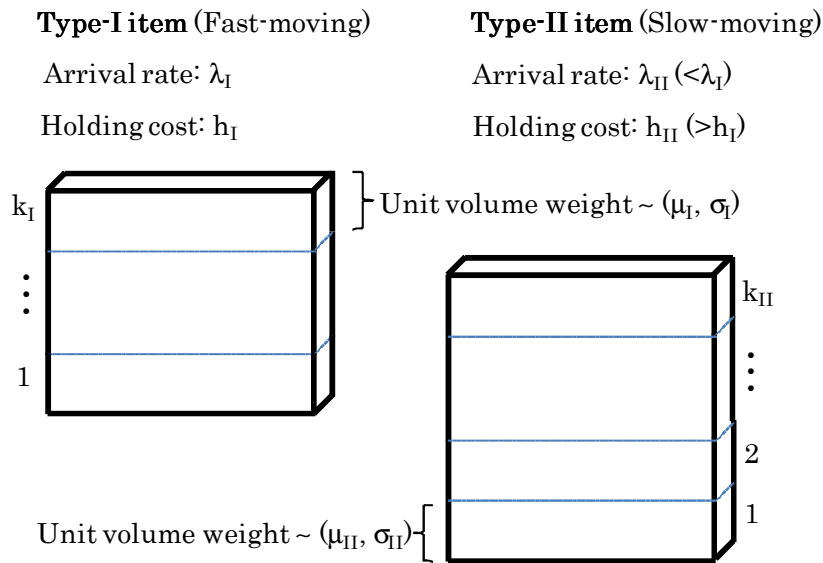
Figure 3.3: Characteristics of Items

In our model setting, type-I items are assumed to be fast-moving products that

might be considered as "typical" cargo. Type-II items are those that are shipped

less frequently. They are slow-moving (such as specialty electronics or a high-end

printer) but may need to be better taken care of, hence have a higher handling cost.

From the theory of Poisson processes (Çınlar, 1975), we first note that $p_I = \lambda_I/(\lambda_I+\lambda_{II})$ and that $p_{II} = 1-p_I$. Then, using (3.5)-(3.8), we can find the expected

cost of holding one shipment until the arrival of another, as

$$\mathrm{E}C = \mathrm{p_I}\kappa_\mathrm{I}+\mathrm{p_{II}}\kappa_\mathrm{II}+[(\lambda_\mathrm{I}+\lambda_\mathrm{II})^{-1}][h_\mathrm{I}\mathrm{k_I}q_\mathrm{I}(\mathrm{t})+h_\mathrm{II}\mathrm{k_{II}}q_\mathrm{II}(\mathrm{t})]$$

We also have, from (3.7),

$$\begin{aligned}
\mathrm{p_f(t)} &= \Pr\{\min(Y_1,..,Y_\mathrm{n}) \leq \mathrm{T^f}\}\sum_{\mathrm{i=1}}^{\mathrm{N}}\mathrm{p_i}\mathrm{p}^{W_\mathrm{i}}(\mathrm{t})\mathrm{p}^{V_\mathrm{i}}(\mathrm{t}) \\
&= (1-e^{-\mathrm{T^f}(\lambda_\mathrm{I}+\lambda_\mathrm{II})})[\tfrac{\lambda_\mathrm{I}}{\lambda_\mathrm{I}+\lambda_\mathrm{II}}\mathrm{p}^{V_\mathrm{I}}(\mathrm{t})\int_0^{W_\mathrm{r}(\mathrm{t})/\mathrm{k_I}}dF_\mathrm{I}+\tfrac{\lambda_\mathrm{II}}{\lambda_\mathrm{I}+\lambda_\mathrm{II}}\mathrm{p}^{V_\mathrm{II}}(\mathrm{t})\int_0^{W_\mathrm{r}(\mathrm{t})/\mathrm{k_{II}}}dF_\mathrm{II}]
\end{aligned}$$

Recall that the exact computation of $\mathrm{p}^{V_i}(\mathrm{t})$ requires an on-line optimization algorithm. Assuming that $\mathrm{k_{II}} \geq \mathrm{k_I}$, one can approximate this probability. Now, since type-II items are larger than type-I items, then $\mathrm{p}^{V_\mathrm{II}}(\mathrm{t}) \leq \mathrm{p}^{V_\mathrm{I}}(\mathrm{t})$. Introducing this condition implicitly forces $\mathrm{p_f(t)}$ to be zero. In turn, that results in an inevitable decision of dispatch, i.e. if there is not enough space to include even the type that has minimum volume, continuing to consolidate causes unnecessary costs.

To get more analytical insights, suppose that the weight of a unit volume of each type is uniformly distributed, i.e. $X_\mathrm{I} \sim \mathrm{Uni}(\mathrm{a_I},\mathrm{b_I})$ and $X_\mathrm{II} \sim \mathrm{Uni}(\mathrm{a_{II}},\mathrm{b_{II}})$. Then, we can explicitly write the probabilities that a type-i item fits the residual weight, for i=I, II, as

$$\begin{aligned}
\mathrm{p}^{W_\mathrm{i}}(t) &= \Pr\{W_\mathrm{i} \leq W_\mathrm{r}(\mathrm{t})\} = \int_0^{W_\mathrm{r}(\mathrm{t})/\mathrm{k_i}}dF_\mathrm{i} \\
&= \begin{cases} 0, & \text{if } W_\mathrm{r}(\mathrm{t}) < \mathrm{a_i}\mathrm{k_i} \\ \dfrac{W_\mathrm{r}(\mathrm{t})-\mathrm{a_i}\mathrm{k_i}}{\mathrm{k_i}(\mathrm{b_i}-\mathrm{a_i})}, & \text{if } \mathrm{a_i}\mathrm{k_i} \leq W_\mathrm{r}(\mathrm{t}) < \mathrm{b_i}\mathrm{k_i} \\ 1, & \text{if } W_\mathrm{r}(\mathrm{t}) \geq \mathrm{b_i}\mathrm{k_i} \end{cases}
\end{aligned} \quad (3.14)$$

C_V	C_W	$T_{\max}$	L	κ	$(\kappa_{\mathrm{I}}, \kappa_{\mathrm{II}})$	$(k_{\mathrm{I}}, k_{\mathrm{II}})$	$(h_{\mathrm{I}}, h_{\mathrm{II}})$
30	60	14	2	200	$(2,4)$	$(1,2)$	$(1,1)$

Table 3.2: Parameter Set for the 2-item SCLP

Assuming an all-way packing orientation, we can also compute the probability that a type-i item fits the residual volume simply by

$$
\begin{aligned}
\mathrm{p}^{V_i}(\mathrm{t}) \;&= \Pr\{V_i \le V_{\mathrm{r}}(\mathrm{t}) = C_V - \textstyle\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{k_i} q_{\mathrm{i}}(\mathrm{t})\} \\
&= \begin{cases} 0, & \text{if } V_{\mathrm{r}}(\mathrm{t}) < \mathrm{k_i} \\[2mm] 1, & \text{o.w.} \end{cases}
\end{aligned} \tag{3.15}
$$

3.5   A Numerical Example

In this section, a numerical example is solved and analyzed with respect to various parameters used in the model. MATLAB is employed to simulate the data and to obtain the optimal solution. The parameter values of our base case are exhibited in Table 3.2. There are two types of items, type-I and type-II, whose interarrival time distributions are exponential and for which the unit-volume weights follow a uniform distribution. That is, $Y_{\mathrm{I}} \sim \mathrm{Expo}(1), Y_{\mathrm{II}} \sim \mathrm{Expo}(0.5), X_{\mathrm{I}} \sim \mathrm{Uni}(1,2), X_{\mathrm{II}} \sim \mathrm{Uni}(3,4)$.

For these particular arrival rates and weight distributions, a possible sample path of cumulative weight is displayed in Fig. 3.4. Due to the choices of the base

Figure 3.4: Sample Path of Cumulative Weight
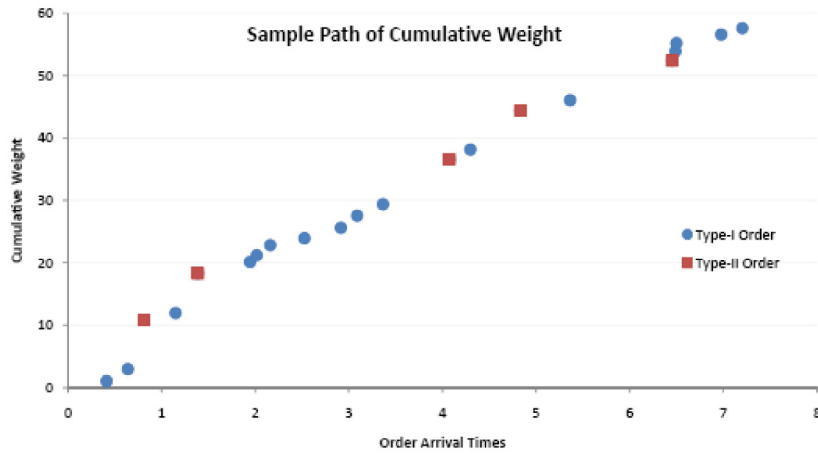
case parameters, the number of type-I arrivals happens to be larger than for type-II.

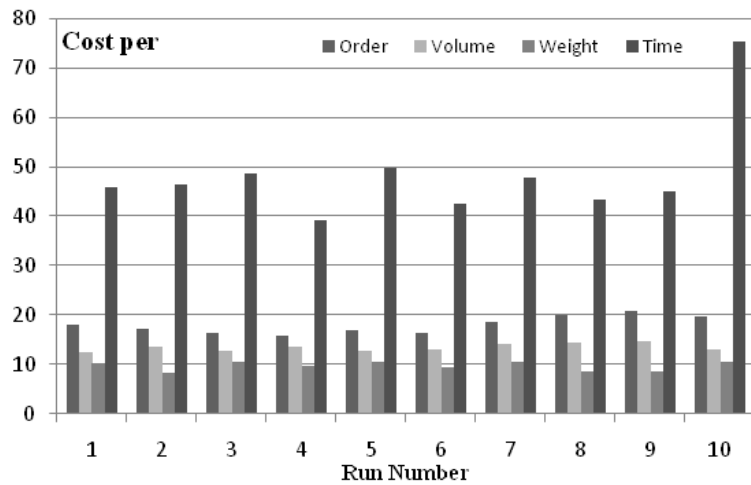We also note that the increase in total weight is higher for type-II.

Figure 3.5: Cost Variability in Objective Functions

A challenge in SCL is the appropriate determination of an objective function. To

see the effects of some possible objectives, we devise Fig.3.5. To enable reasonable

67
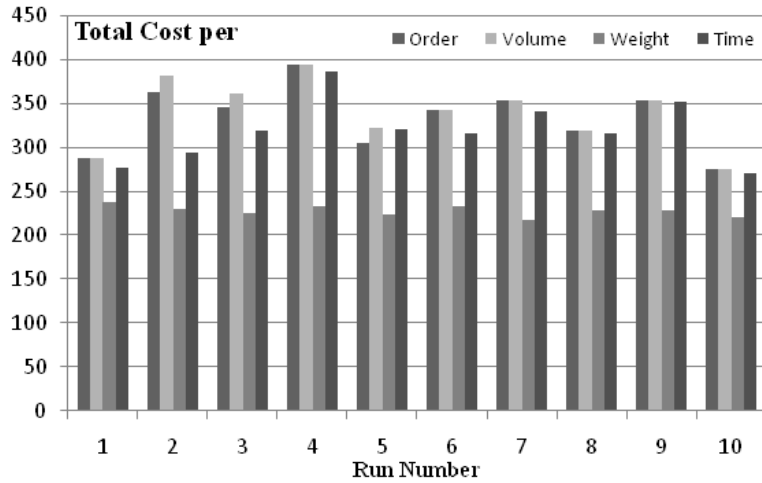
Figure 3.6: Cost-per-dispatch Variability in Objectives

comparisons, the technique of "common random numbers" is employed, e.g. Bratley et al. (1987). We observe from that figure that the variability is within $5 for the costs per unit weight. The corresponding range is higher, around $30, however, in the results on costs per unit time. The first observation is logical, based on an argument similar to that of a Central Limit Theorem. The second is also reasonable, since type-I items arrive twice as frequently per unit time.

We look now at the variabilities on the cost-per-dispatch for each objective considered (see Fig. 3.6). In our model, cost per dispatch for a particular objective, say cost per unit volume, is found by multiplying the realized (optimal) cost per unit volume by the volume accumulated in that SCL cycle. Similarly, we obtain other total cost variabilities. Though this example is not statistically conclusive, again we can at least observe that the total cost variability for cost-per-unit weight

68

is the smallest amongst the other objectives.
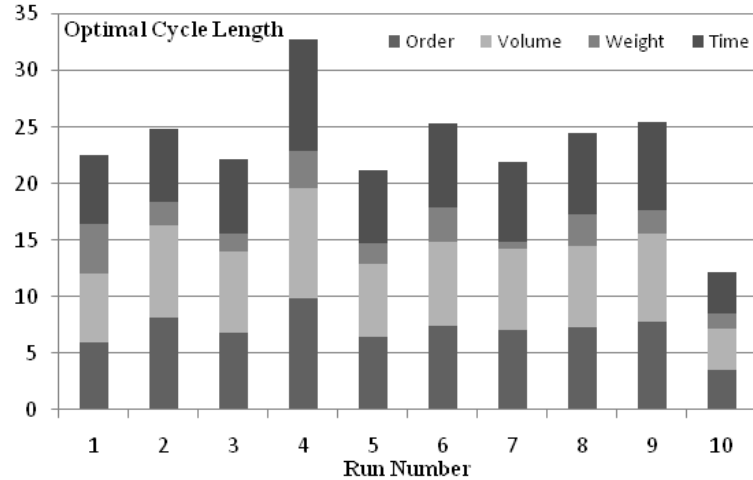
Figure 3.7: Variability of Optimal Cycle Length wrt. Various Objectives

In Fig. 3.7, the length of the optimal shipment consolidation cycle is compared for each objective. This comparison is quite important because the length of a cycle directly affects the delivery times of the consolidated load, and thereby the service level. The variability ranges in the length of the consolidation cycle for the objectives of cost-per-order, cost-per-unit volume, and cost-per-unit time are within 6.25 time units, as opposed to that of 3.9 time units for the objective of cost-per-unit weight. Again, we can observe that cost-per-weight is a more stable objective as justified here.

We now turn attention to the behavior of the probability of fit $p_f(t)$ when the volume ratio of type-I item to type-II item is decreased (see Fig. 3.8). In that figure, for example, the ratio 1:3 implies that a type-II item is three times as large
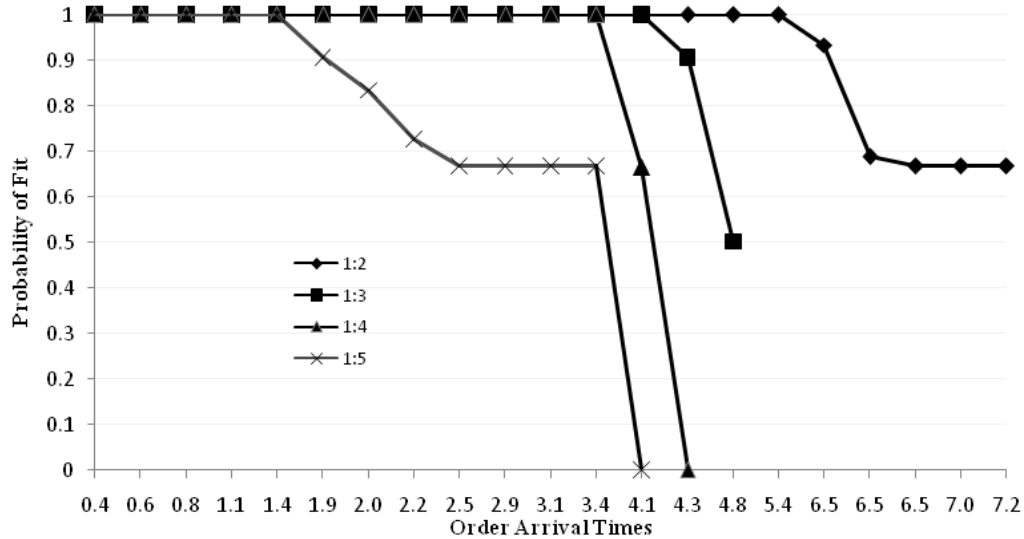
Figure 3.8: Probability of Fit for Various Ratios of Volumes

as a type-I item. Naturally, as the ratio decreases, i.e. as the volume of type-II item gets relatively larger, the fit probability decreases and actually vanishes earlier than for larger ratios. A simple insight can easily be derived from this graph: Larger volume differences increase the chances of earlier dispatch.

Similar to Fig. 3.8, in Fig. 3.9 we investigate the behavior of $p_f(t)$ as the maximum holding time T_{max} is varied. We note that as T_{max} is decreased, the optimal dispatching time is decreased automatically. This is because our modeling has explicitly incorporated the probability of not receiving a shipment between now and the required dispatch time, i.e. $p_d = \Pr\{\min(Y_1, .., Y_n) \leq T^f\}$. From Fig. 3.9, one can also deduce that for a certain threshold of T_{max} (for the base-case parameters, it is 9 time units), the probability of fit is stable. That is,
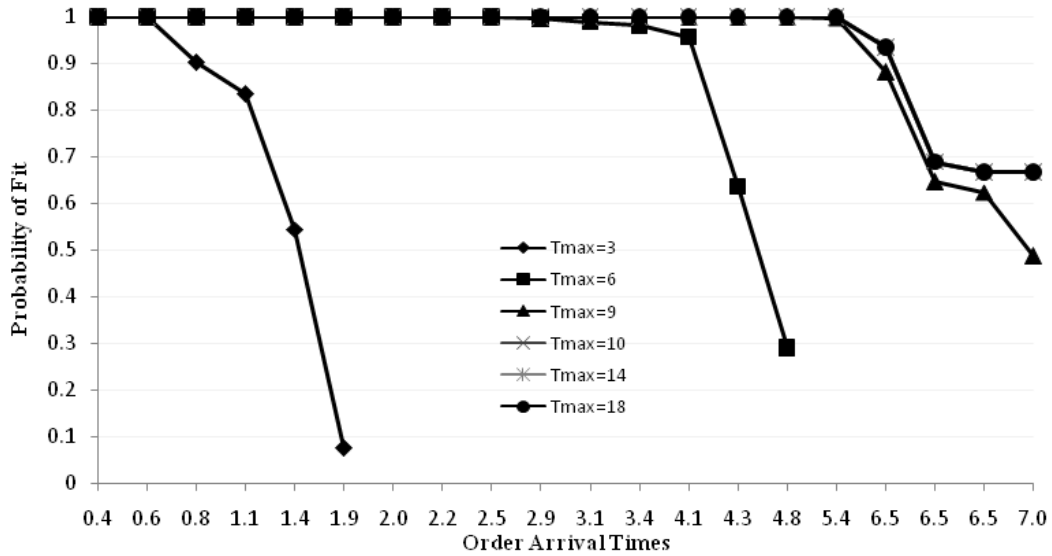
Figure 3.9: Probability of Fit for Particular Maximum Holding Times

$(p_f(t)|T_{max} = 9) = (p_f(t)|T_{max} > 9)$, which is a useful fact that can be employed to reduce the computational burden.

Next, our sensitivity analysis focuses on various parameters for the objective of cost per order. The main interest is how the optimal length of SCL cycle is affected by the changes in those parameters, ceteris paribus. (Naturally, similar analyses could be carried out for other objectives as well.)

Fig. 3.10 exhibits the relationship between the optimal cycle length and the vehicle dispatch cost. As observed, for quite small values of dispatching cost, the optimal decision is to ship as items arrive, hence a consolidation program is not needed. However, as that cost is increasing, to enable scale economies, SCL proves useful. Hence, the optimal SCL cycle length is increasing in the dispatch cost.
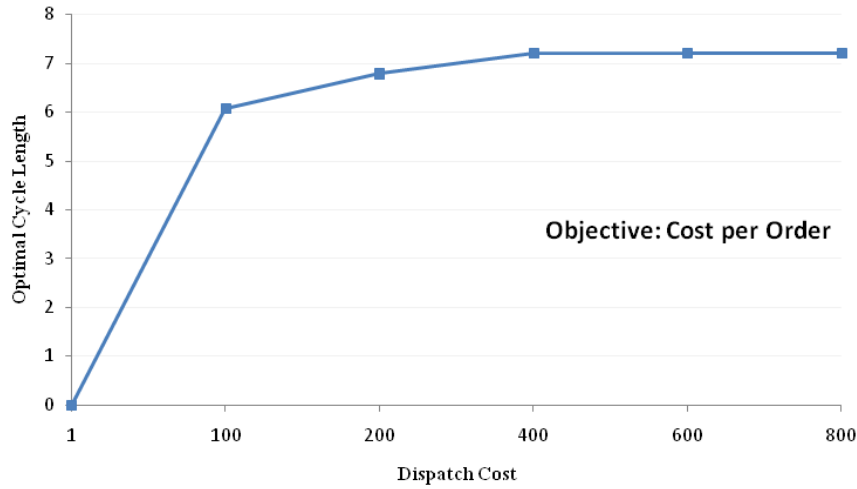
Figure 3.10: Dispatch Cost vs. Optimal Cycle Length

However, after a particular (larger) threshold value, i.e. 400 here, the optimal cycle lengths are the same. This result may be due to a binding constraint, either to volume, to weight or to maximum holding time constraint.

As the ratio of the arrival rates (λ_{II}/λ_I) increases (keeping $\lambda_I = 1$), i.e. as more and more type-II items arrive, the optimal length of the SCL cycle decreases somewhat (Fig. 3.11). Besides the increase in total arrival rate, that behavior is due to the fact that type-II items are larger than type-I items, both in weight and in volume.

From Figs. 3.12 and 3.13, the optimal cycle length is observed to be non-decreasing in both volume and weight capacity. Of course, one of the capacity constraints might be binding, as is the case here for volume. These graphs show
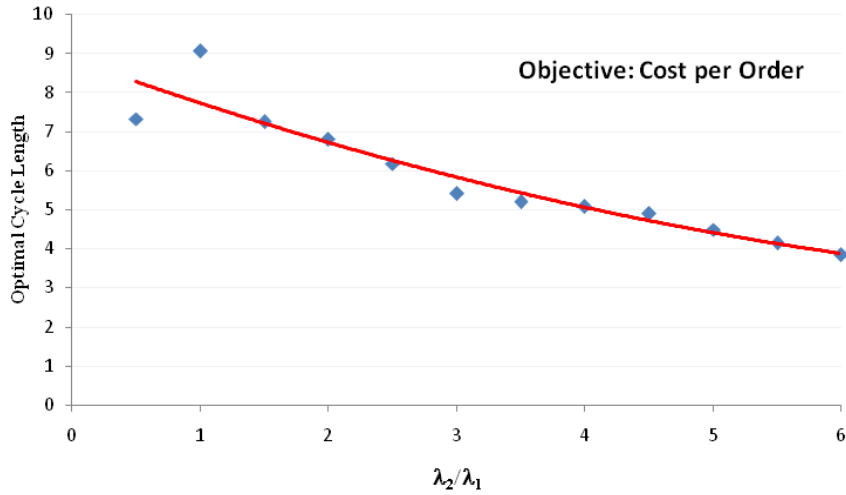
Figure 3.11: Ratio of Arrival Rates vs. Optimal Cycle Length

and justify the use of *modularization* in packaging. That is, if the LTL carrier knows exactly the weights and volumes of each type, then he can arrange his vehicle type in such a way that transportation costs are minimized, while vehicle utilization is increased and an acceptable service level is maintained.

Finally, in Fig. 3.14, the effect of varying T_{max} on the optimal cycle length is displayed. We first observe that as the maximum holding time increases, the optimal cycle length increases too. However, for a certain threshold of T_{max} and above, which is 10 time units for these base-case parameters, the optimal cycle length is the same.

73
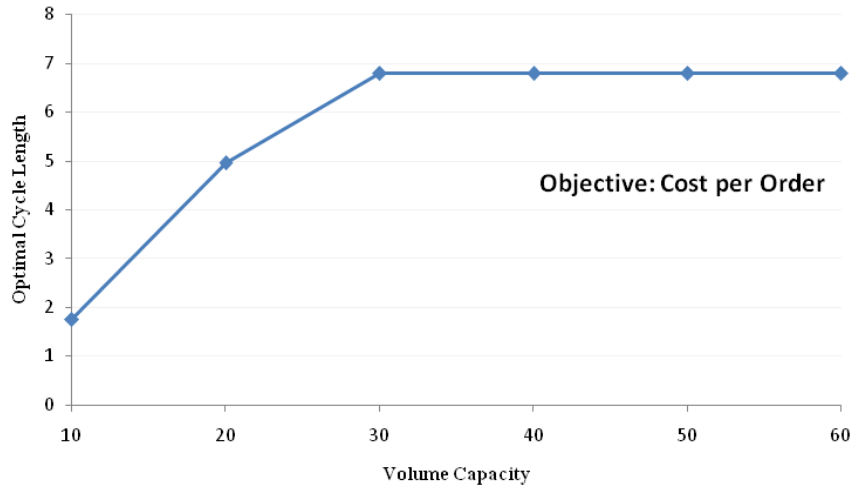
Figure 3.12: Volume Capacity vs. Optimal Cycle Length

3.6 Limitations and Extensions

Myopic analysis is easy to implement in MISCLP, given the fact that any LTL transport provider need only measure the weight and volume of each arriving shipment and record it. Our formulation assumed that we knew the relative rates of arrival of each item type, and their respective weights and volumes, at least probabilistically. Our modeling approach enabled us to explicitly incorporate a constraint on service (maximum holding time) in the decision rules by introducing the probability of fit $p_f(t)$. Generally, those timing decisions would have created major drawbacks in an alternative modeling framework.

However, in a situation where the dispatcher has no record of the weights and volumes of the arriving items, we face pure uncertainty. Hence, a renewal theoretic
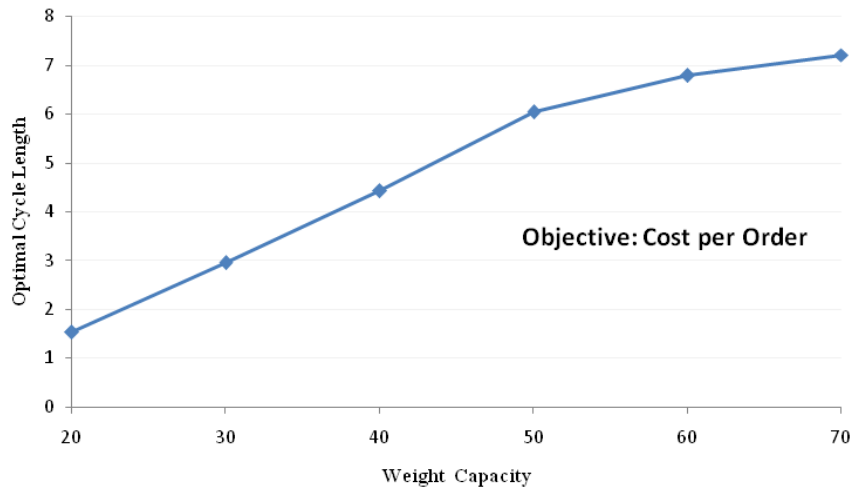
Figure 3.13: Weight Capacity vs. Optimal Cycle Length

approach may need to be employed, as opposed to a myopic one. Such a comparison could provide us the Expected Value of Perfect Information for MISCLP. This would be an interesting and challenging extension to our current model.

MISCLP can also be cast as a sequential decision of accepting or rejecting demands (loads to be dispatched) in the presence of a scarce resource (vehicle capacity), with random demand characteristics and no *recall* option. The case with recall option (i.e. having the chance to include the now-rejected order into the next batch) is another interesting extension. This approach shows quite a resemblance to stochastic knapsack problem. Hence, in the current setting, our model (with slight modifications) can be applied to other real-world problems, such as those in revenue management and scheduling. See for example Papastavrou et al. (1996),
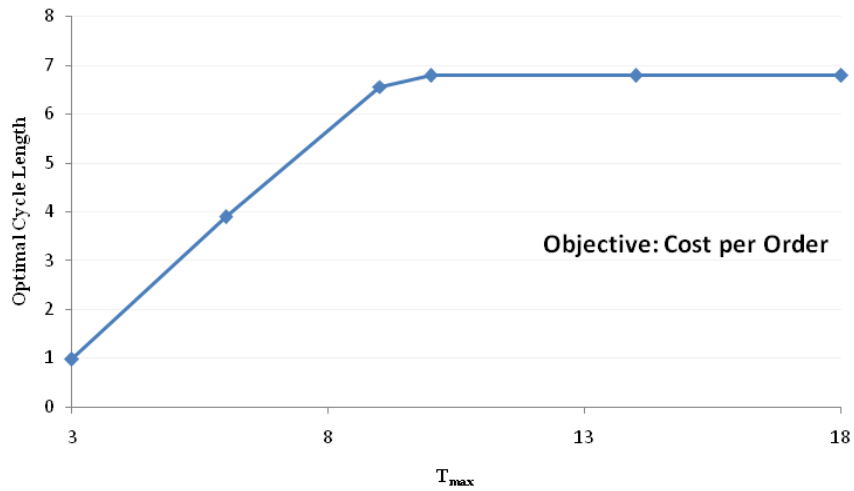
Figure 3.14: Maximum Holding Time vs. Optimal Cycle Length

Kleywegt and Papastavrou (2001), and Kavadias and Loch (2004).

We observed that the use and type of SCL depend not only on the cost parameters but also the variability in the order arrival process. In the next chapter, we will introduce pricing as a means to influence the rate of order arrivals. That is, along with the shipment consolidation decisions, we will now also consider pricing decisions for a transportation service.

76

Chapter 4

Pricing and Shipment

Consolidation-I

4.1 Introduction

Increased competition in business environments requires that firms provide a quality and timely service with minimal cost. Many companies quote delivery-time guarantees as a marketing weapon to attract customers in time-sensitive markets. Some customers are willing to pay a premium charge for fast or timely delivery, e.g. for Federal Express' next day delivery service e.g. see Lederer and Li (2000).

The shipper, the carrier and the consignee (buyer or receiver of the goods)

are the three main perspectives in freight transportation. The *shipper* refers to a company, e.g. a manufacturing firm, that owns or controls the freight. The shipper is arranging for transportation of these goods because they have been sold to the *consignee*. The *carrier* provides linkages between shippers and the respective consignees by offering transportation in their own vehicles.

Fig. 4.1 displays a schematic explanation of the problem at hand. In that figure, X, Y, Z, among many others, represent shippers, i.e. suppliers of products ordered by the consignee. For example, consignee could be the distribution center of a retail chain, who would pick up the goods at the destination terminal DT. Goods from them, destined to DT, will be accumulated (consolidated) in the carrier's origin terminal, OT. To offer a quality logistics service, the carrier offers a guaranteed delivery-time (i.e., time it takes to line-haul freight from OT to DT). Having negotiated a unit transportation cost and such a lead time before delivery, the carrier incorporates this information in designing the prices to be charged to the shippers. Then, the carrier solves a joint problem for the optimal length of consolidation period and the prices. The customers are then informed of the scheme for transportation pricing and of the delivery-time guarantee, i.e. the latest time that their shipment will be at the destination terminal DT. (At that time, their orders are ready for pickup, or de-consolidation and local delivery.)
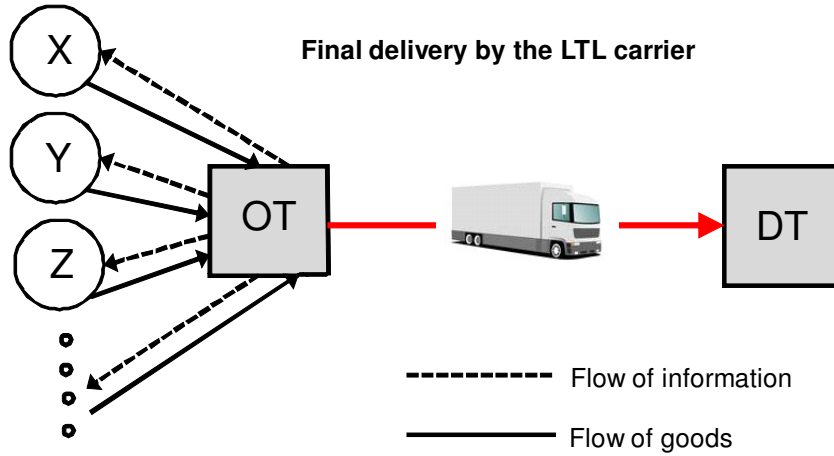
Figure 4.1: Flow of information and goods

The order arrival rates are *endogenous* to our problem formulation. A customer's decision on using a specific LTL carrier is affected by the price and the delivery-time guarantee offered by that carrier. The latter must determine the profit-maximizing consolidation time (the optimal timing for dispatch) and the price-time pattern for customers who are sensitive to both delivery time and price. We first seek the optimal decisions when there is a fixed cost per transit-delivery and no capacity restriction at the carrier. Then, we extend our model to the case of freight rates that incorporate an all-unit (all-weight) discount.

Our work improves on certain aspects in the classical joint pricing and inventory management literature, when demand is price- and time-sensitive, e.g. Rao et al. (2005). Foremost, here we model and study a logistics system that incorporates shipper and carrier integration. Second, we consider differential pricing with time-

varying guarantees as well as uniform guarantees. Finally, various easy-to-apply pricing schemes are proposed and compared.

To the best of our knowledge, the present chapter is first to address the following research questions: **i)** What is the optimal quoting for a 3PL provider in a price- and time-sensitive logistics market? **ii)** How can a discount schedule be integrated into this model?

In what follows, to alleviate the differences between pricing schemes used by the carrier, first we will assume that the transportation cost per order is fixed. Later, the inherent advantage of SCL, i.e. economies of scale, is integrated with this base model. We also assume that the LTL carrier is a common carrier. There is thus no capacity constraint from the perspective that an additional vehicle could be available if needed.

The remainder of this chapter is organized as follows. After a literature survey, the preliminaries and our model are presented in the next section. We examine four pricing schemes for a carrier in Section 4.3 and illustrate a numerical example and its sensitivity analyses in Section 4.4. Section 4.5 provides some additional structural results. The case of all-weight discount scheme is studied in Section 4.6. We conclude with our final thoughts and some future research directions.

4.2 Overview of Pertinent Literature

The pertinent literature on our specific problem is not exhaustive. However, there are relations between the existing research on some overlapping areas such as integrated pricing decisions in service and manufacturing industries, quoting lead-times or delivery times, and pure shipment consolidation.

In practice, the determination of prices for selling a product or provisioning a service influences the demand. Keeping other dimensions constant, the setting of low prices may trigger a higher demand. Studies in joint pricing and inventory management date back to early 50s. For example, Chan et al. (2004) give an excellent treatise on the coordination of pricing and inventory decisions.

With the advent of time-based competition, studies on quoting of lead times (in a manufacturing industry) or delivery time (in a service industry) have been increasing since the last decade. This line of research also appears under the general topic of "due-date management," e.g. Keskinocak and Tayur (2004). Existing models explicitly incorporate the delivery-time guarantee as well as the price in the demand function, and they mostly assume the delivery time to be random variable. Yet, there has been no consensus on the functional form of the expected demand; linear, multiplicative and exponential are such forms mostly employed (e.g. Lederer and Li, 1994; Palaka et al., 1998; Hill et al., 2000; Boyacı and Ray, 2003; Urban,

2009). So and Song (1998) includes the competition perspective into his models. Pekgün and Keskinocak (2008) investigate the coordination opportunities between production and marketing functions in a price- and lead-time-sensitive market.

The pricing strategies in delivery-time guarantees literature fall into two broad categories: *differential* pricing, in which each customer is provided with a distinct lead-time quotation, and *uniform* (or single) pricing, in which a single delivery-time guarantee is made to all of the customers. For example, Rao et al. (2005) model the impact of a uniform lead time guarantee on customer demand and production planning in a make-to-order environment, whereas Çelik and Maglaras (2008) consider differential pricing of multiple products in a dynamic version of such an environment.

4.3   Model Development

Details on the flow of goods and information in our problem now follow. Consider a single shipper who uses a time-based shipment release policy, i.e., the T-P. (We note that a quantity-based policy, under random arrivals, is unsuitable if a delivery time need be quoted.) That is, the shipper has some discretion on the timing of the load dispatch.

Fig. 4.2 depicts the sequence and timing of events occurring in a single consol-
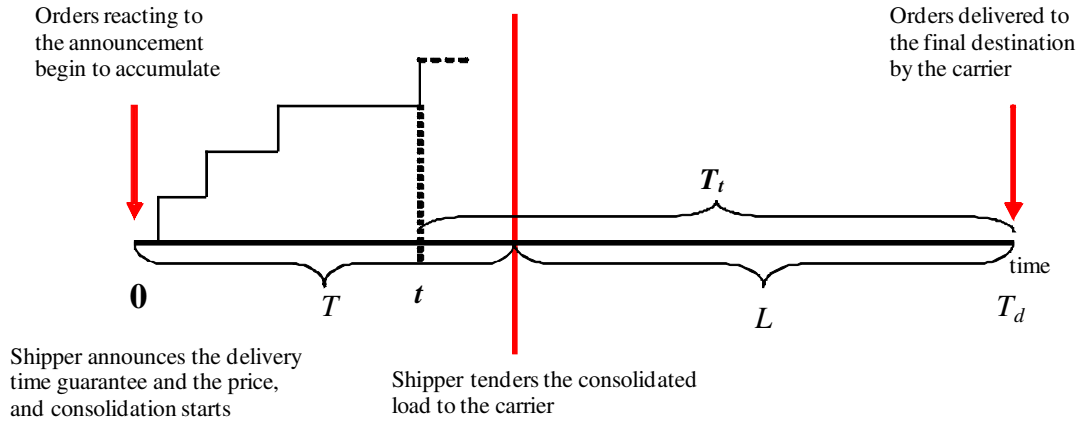
Figure 4.2: Sequence and Timing of Events

idation cycle. We consider a total planning horizon of length T_d, which comprises the consolidation period (T) plus the transit time (L). Since T is a decision variable, T_d implicitly is so. At time $t = 0$, the shipper announces the price per unit ordered with a guarantee of delivery by time T_d, *decided in advance*. X, Y, Z, ..., the customers, react to this announcement and their orders begin to build up at the shipper's site. Following the consolidation period, at time T, the shipper tenders the consolidated load to the carrier; the carrier hauls the load to its final destination within L time units, and hence $T_d = T + L$.

To establish some initial insights, we employ the following assumptions for modeling purposes: **i)** The orders come from a mix of a large number of potential customers, each with varying price- and time-sensitivities. (This assumption is important for the use of exponential inter-arrival times in our model.) **ii)** The arrival

83

rate is a decreasing function of both the price and the delivery-time guarantee. **iii)** The shipper incurs the holding and delivery costs, and pays the carrier a transportation rate per unit shipped which is decreasing in the transit time L. That transit time is guaranteed to the shipper by the carrier, and moreover, this arrangement between the two parties is done *a priori* with a contractual agreement.

For a given transit time, we are to determine the optimal interval T^* for consolidation and the optimal price per unit ordered, set in advance. Since L is fixed in our model, simply finding the optimal consolidation period will yield the time guarantee.

Let us now consider the orders destined for the same consignee R; these accumulate at the shipper according to a non-stationary Poisson process. For ease of analysis and to enhance clarity, let us mark the arrival time of an order to be t, where $0 < t < T < T_{\max}$ (see Fig. 4.2). Then we can define the maximum delivery-time guarantee for such an order to be

$$T_t = T + L - t \tag{4.1}$$

We employ $\lambda(p_{T_t}, T_t)$ as the rate of arrival of orders at time t for a given vector of price and delivery time, (p_{T_t}, T_t). We construct $\lambda(p_{T_t}, T_t)$ as a linear function of price and delivery-time guarantee, and define the unit transport price as another function of that guarantee. Further, linearity between price and time is widely used

84

in economics and operations management, and serves well to illustrate the trade-offs between price and delivery-time guarantees in the problem setting at hand, see for example Palaka et al. (1998) and LaFrance (1985). Explicitly, the order-arrival process and the price per unit are modeled as

$$\lambda(p_{T_t}, T_t) = a - s_p p_{T_t} - s_T T_t, \ 0 < t < T \tag{4.2}$$

$$p_{T_t} = \rho - \varepsilon T_t, \ \rho > 0, \ \varepsilon > 0, \ p > 0 \tag{4.3}$$

with **parameters**

a : maximum potential demand

s_p : sensitivity to price, in dimensions of demand per unit price

s_T : sensitivity to delivery-time guarantee (demand per unit time)

ρ : market-reservation price per unit

ε : delivery-time sensitivity of the price (dollars per unit time)

L : transit time guaranteed by the carrier

and **decision variables**

T : length of consolidation period

p_{T_t} : price per unit charged to an order arriving at time t, in dollars

$T_{\max}$: upper bound for delivery-time guarantee (i.e. $T_{t=0}$)

Regarding (4.3) and the definition of T_t (4.1), we note that the price is increasing in t (i.e. later arrivals pay a higher price), and decreasing in T and in L (a customer

is less willing to pay, the longer is the delivery-time guarantee). Besides tractability, our linear arrival (demand) function has several desirable properties. First, we observe that the price and time elasticities of demand are each increasing in both price and delivery time. For example, the price elasticity of demand, obtained as $[-s_p p_{T_t}/(a - s_p p_{T_t} - s_T T_t)]$ is increasing in the price (p_{T_t}) or the delivery-time quote (T_t) or both, i.e. the percentage change in demand is greater at higher prices and longer delivery times. In other words, customers experience greater relative sensitivity to long delivery times when they are paying more for the service, or vice versa. Second, we observe that the arrival-rate function is separable in price and time. This also is a desirable property because customers perceive time and money as substitutes, see for example Lederer and Li (1994)

We now express the arrival rate as a single-argument function and optimize over it. Combining (4.2) and (4.3), we obtain

$$\lambda(T_t) = \alpha - \beta T_t, \quad 0 < T_t < T_{\max}, \text{ where}$$
$$\alpha = a - s_p \rho, \text{ and } \beta = s_T - s_p \varepsilon$$

(4.4)

in which the sign of β in (4.4) will determine the type of market. If β is negative, the market is more "price sensitive" than "time sensitive."

In the remainder, we assume that α is positive, since otherwise, the arrival rate (if $\beta > 0$) would always be negative, which is not admissible. Our aim is to maximize

the profit rate of the shipper. Let us recall that h is the carrying cost per unit time for consolidating an order, and that c be the transportation cost per unit shipped via the carrier. Since the shipper is the consolidating party, there is no incentive on the *contracted* carrier's part to vary the shipment charges. Therefore the cost of the consolidation program, as seen by the shipper, comprises the total cost of holding plus that of transportation. Having the discretion on pricing, the shipper wants to maximize the expected total profit Π generated over a consolidation cycle of length T, which is

$$\Pi(\tau) = \int_0^T (p_{T_t} - c)\lambda(T_t)dt - h\int_0^T (T-t)\lambda(T_t)dt \tag{4.5}$$

Also let $N(T)$ be the total expected number of arrivals during T. Then

$$N(T) = \int_0^T \lambda(T_t)dt \tag{4.6}$$

In the next section, we focus on deriving the structures of optimal solutions for various pricing schemes based on (4.5) and (4.6).

4.4   Formulation of Pricing Schemes

We propose and consider four different pricing schemes, under the category of differential and uniform pricing. *Differential Pricing* (DF) would occur when each order is priced differently, with respect to its arrival time. *Uniform-First Pricing* (UF) is

the scheme whereby the shipper charges all orders the same as he would charge the first-arriving customer. *Uniform-Mean-Time Pricing* (UT) happens when every customer pays the amount that a customer arriving in the middle of the consolidation period would pay. And as the name suggests, *Uniform-Last Pricing* (UL) means that all customers are charged the same price as the "last customer," i.e. one who would have arrived just at the end of the consolidation period.

In the following subsections, we derive the objective functions for each pricing scheme, and obtain the corresponding optimal prices and consolidation lengths on the basis of the profit rate, i.e. profit per unit time. Profit rate for a particular pricing scheme refers to the ratio of the total expected profit generated during the consolidation cycle to the expected length of that cycle. We use the profit rate as our objective because it enables us to compare the pricing schemes on the same performance measure. Moreover, this is a commonly used objective function in the operations and marketing literature; for example see Kunreuther and Richards (1995) and Rajan et al. (1992).

4.4.1 Differential Pricing

The expected total profit function for DF, over a consolidation cycle of length T, is

$$
\begin{aligned}
\Pi_{\mathrm{DF}}(T) &= \int_0^T \left[(\rho-\varepsilon(T+L-t)-c)-h(T-t)\right]\left[\alpha-\beta(T+L-t)\right]dt \\
&= T^3 \tfrac{1}{3}\beta(\varepsilon+h)+T^2\left[\tfrac{1}{2}(hL+c-\rho)-\alpha(\varepsilon+h)+\varepsilon L\right]+T(\alpha-\beta L)(\rho-\varepsilon L-c)
\end{aligned}
$$

Let us define the *profit-rate* function for DF as

$$
\pi_{\mathrm{DF}} \triangleq \Pi_{\mathrm{DF}}(T)/T = T^2\delta_{1\mathrm{DF}}+T\delta_{2\mathrm{DF}}+\delta_{3\mathrm{DF}} \text{ where}
$$

$\delta_{1\mathrm{DF}} = \tfrac{1}{3}\beta(\varepsilon+h),\ \delta_{2\mathrm{DF}} = \beta\tfrac{1}{2}(hL+c-\rho)-\alpha(\varepsilon+h)+\varepsilon L,$ and $\delta_{3\mathrm{DF}} = (\alpha-\beta L)(\rho-\varepsilon L-c)$

We want to maximize π_{DF} over T under the following constraints.

   i) Consolidation length $T \geq 0$.

   ii) Arrival rate is to be nonnegative: $\alpha-\beta(T+L) \geq 0$, i.e. $T \leq -L+\alpha/\beta \triangleq \mathrm{U}_{1\mathrm{DF}}$.

   iii) Marginal revenue must be nonnegative: $\rho-c-\varepsilon(T+L) \geq 0$, and hence $T \leq$

      $-L+(\rho-c)/\varepsilon \triangleq \mathrm{U}_{2\mathrm{DF}}$.

Depending on the sign of $\delta_{1\mathrm{DF}}$, we have two optimization problems in which the objective function is either convex or concave. Accordingly the solutions differ for these two types of problems.

Case 1: $\delta_{1\mathrm{DF}} > 0$

Under this case, the objective function is strictly convex, thus a boundary solution is optimal. Let $U_{\mathrm{DF}} = \min\{U_{1\mathrm{DF}}, U_{2\mathrm{DF}}\}$ and define $x^+ = \max\{0, x\}$. Therefore, we choose T^* to be $(U_{\mathrm{DF}})^+$. We note that for $T^* = 0$, a consolidation program is *not* recommended.

Case 2: $\delta_{1\mathrm{DF}} \leq 0$

This condition implies that the objective function is concave and $\beta \leq 0$. Thus the second constraint is automatically satisfied.

Now, combining the results of Case 1 and 2 above and applying first-order conditions, we obtain the optimal length of consolidation and the corresponding price to be

$$T_{\mathrm{DF}}^* = (\min\{-\delta_{2\mathrm{DF}}/2\delta_{1\mathrm{DF}}, U_{\mathrm{DF}}\})^+ \tag{4.7}$$

$$p_{\mathrm{DF}}^*(t) = \rho - \varepsilon(T_{\mathrm{DF}}^* + L - t) \tag{4.8}$$

We emphasize that (4.7) and (4.8) will also hold for the *uniform* pricing schemes. Those cases are distinguished by setting t to a particular point within the optimal consolidation cycle.

4.4.2 Uniform-First Pricing

Assuming the arrival of the first order takes place at the start of the consolidation cycle, we fix $t = 0$ and thus $p_{\mathrm{UF}} = \rho - \varepsilon(T + L)$. Incorporating the modified price

function into that of the arrival rate, we obtain

$$\lambda(T_t) = a - s_p\rho + s_p\varepsilon L - s_T L + T(s_p\varepsilon - s_T) + s_T t = \alpha - \beta(T+L) + s_T t$$

The expected total profit function for the UF is

$$
\begin{aligned}
\Pi_{\mathrm{UF}}(T) &= \int_0^T \left[(\rho - \varepsilon(T+L) - c) - h(T-t)\right] \left[\alpha - \beta(T+L) + s_T t\right] dt \\
&= (\tfrac{1}{2}h\beta - \tfrac{1}{6}hs_T + \beta\varepsilon - \tfrac{1}{2}\varepsilon s_T)T^3 \\
&\quad + (c\beta - \tfrac{1}{2}h\alpha - \tfrac{1}{2}cs_T - \alpha\varepsilon - \beta\rho + \tfrac{1}{2}\rho s_T + \tfrac{1}{2}hL\beta + 2L\beta\varepsilon - \tfrac{1}{2}L\varepsilon s_T)T^2 \\
&\quad + \left(\alpha\rho - c\alpha + cL\beta - L\alpha\varepsilon - L\beta\rho + L^2\beta\varepsilon\right)T
\end{aligned}
$$

Similar to the derivation in Subsection 4.4.2, we obtain the profit per unit time as

$$\pi_{\mathrm{UF}}(T) = T^2\delta_{1\mathrm{UF}} + T\delta_{2\mathrm{UF}} + \delta_{3\mathrm{UF}}, \text{ where}$$

$$\delta_{1\mathrm{UF}} = \beta(\varepsilon + \tfrac{1}{2}h) - \tfrac{1}{6}s_T(h+3\varepsilon),$$

$$\delta_{2\mathrm{UF}} = \beta(2\varepsilon L + \tfrac{1}{2}hL - \rho + c) + \tfrac{1}{2}s_T(\rho - c - \varepsilon l) - \alpha(\varepsilon + \tfrac{1}{2}h),$$

$$\text{and } \delta_{3\mathrm{UF}} = (\alpha - \beta L)(\rho - \varepsilon L - c)$$

The upper bounds are $T \leq -L + \alpha/\beta \triangleq U_{1\mathrm{UF}}$ and $T \leq -L + (\rho - c)/\varepsilon \triangleq U_{2\mathrm{UF}}$. The optimal solution T_{UF}^* is found as in Section 4.1, but with U_{DF} replaced by $U_{\mathrm{UF}} = \min\{U_{1\mathrm{UF}}, U_{2\mathrm{UF}}\}$. Accordingly, the optimal price is derived as $p_{\mathrm{UF}}^* = \rho - \varepsilon(T_{\mathrm{UF}}^* + L)$, cf. (4.7) and (4.8).

4.4.3 Uniform-Mean-Time Pricing

Under this scheme, we set the price according to a customer that arrives in the middle of the consolidation period. With $p_{\mathrm{UT}} = \rho - \varepsilon(T/2+L)$, the arrival-rate function is obtained as $\lambda_{\mathrm{UT}}(T_t) = \alpha - \beta(T+L) - \varepsilon s_p T/2 + s_T t$. Then, the expected total profit function is given by

$$\Pi_{\mathrm{UT}}(T) = \int_0^T \left[(\rho - \varepsilon(L+T/2) - c) - h(T-t)\right]\left[\alpha - \beta(T+L) - \varepsilon s_p T/2 + s_T t\right]dt$$

Thus, we get the expected profit rate function as

$$\pi_{\mathrm{UT}}(T) = T^2 \delta_{1\mathrm{UT}} + T \delta_{2\mathrm{UT}} + \delta_{3\mathrm{UT}}, \ \ \text{where}$$

$$\delta_{1\mathrm{UT}} = h(\tfrac{1}{4}\varepsilon s_p + \tfrac{1}{2}\beta - \tfrac{1}{6}s_T) - \tfrac{1}{4}\varepsilon(s_T - 2\beta - \varepsilon s_p),$$

$$\delta_{2\mathrm{UT}} = \left[\beta(3L\varepsilon + hL + 2c - 2\rho) + \varepsilon\left(cs_p - \alpha - s_T L - \rho s_p\right) + s_T \rho + \varepsilon^2 L s_p - s_T c - \alpha h\right]/2,$$

$$\text{and } \delta_{3\mathrm{UT}} = (\alpha - \beta L)(\rho - \varepsilon L - c)$$

Now, the upper bounds for the consolidation cycle length in the UT scheme are $T \leq (\alpha - \beta L)/(\beta + \varepsilon s_p/2) \triangleq \mathrm{U}_{1\mathrm{UT}}$ and $T \leq 2[-L + (\rho - c)/\varepsilon] \triangleq \mathrm{U}_{2\mathrm{UT}}$. Again, the optimal solution, T_{UT}^* is in the same form of that in Section 4.1, but with U_{DF} replaced by $\mathrm{U}_{\mathrm{UT}} = \min\{\mathrm{U}_{1\mathrm{UT}}, \mathrm{U}_{2\mathrm{UT}}\}$ and the optimal price is $p_{\mathrm{UT}}^* = \rho - \varepsilon(L + T_{\mathrm{UT}}^*/2)$.

4.4.4 Uniform-Last Pricing

With respect to the last-arriving customer, we set $p_{\mathrm{UL}} = \rho - \varepsilon L$, i.e., when $t = T$, yielding the arrival-rate function

$$\lambda_{\mathrm{UL}}(T_t) = a - s_p(\rho - \varepsilon L) - s_T(T + L - t) = \alpha - \beta L - s_T(T - t)$$

The expected total profit over a SCL cycle of length T can then be derived as

$$
\begin{aligned}
\Pi_{\mathrm{UL}}(T) &= \int_0^T \left[(\rho - \varepsilon L - c) - h(T - t)\right]\left[\alpha - \beta L - s_T(T - t)\right]dt \\
&= T^3\left(\tfrac{1}{3}h s_T\right) - T^2\tfrac{1}{2}\left[h(\alpha - \beta L) + s_T(\rho - \varepsilon L - c)\right] + T(\alpha - \beta L)(\rho - \varepsilon L - c)
\end{aligned}
$$

Now, for the profit-rate function, we have

$$\pi_{\mathrm{UL}}(\tau) = T^2\delta_1 + T\delta_2 + \delta_3, \text{ where}$$

$$\delta_{1\mathrm{UL}} = h s_T/3, \delta_{2\mathrm{UL}} = [h(\beta L - \alpha) + s_T(c + \varepsilon L - \rho)]/2, \text{ and } \delta_{3\mathrm{UL}} = (\alpha - \beta L)(\rho - \varepsilon L - c)$$

The upper bounds are $T \leq (\alpha - \beta L)/s_T \triangleq \mathrm{U}_{1\mathrm{UL}}$, and $T \leq (\rho - c)/\varepsilon \triangleq \mathrm{U}_{2\mathrm{UL}}$. Since $\delta_{1\mathrm{UL}} > 0$, the second derivative of $\pi_{\mathrm{UL}}(T)$ is positive, hence π_{UL} is convex. Given $\rho \geq \varepsilon L$, the optimal length of consolidation cycle for UL is $T^*_{\mathrm{UL}} = (-L + \alpha/\beta)/s_T$, and the optimal price to be charged is $p^*_{\mathrm{UL}} = \rho - \varepsilon L$.

4.4.5 A Recap of Optimal Solutions

The preceding solutions, i.e. the optimal consolidation length and the prices to be charged by the shipper for *each* of these pricing schemes, may be interpreted in

various ways. One way is to choose the *best scheme* yielding the maximum profit per unit time for particular market parameters. On the other hand, if the shipper decided on the type of scheme to employ, then the optimal prices and the consolidation length can be obtained. Moreover, if the shipper commits herself for a predetermined delivery-time guarantee which may be induced by market competition, then we can find the corresponding optimal prices.

The following three observations need be emphasized: **i)** Our model is also valid for *pure* time- or price-sensitive markets. There one simply can set the price- or time-sensitivity parameter in the model to zero. **ii)** The optimality structures of each pricing scheme reveal the conditions under which consolidation is not desirable. This occurs when the optimal consolidation cycle length is found to be zero. For example, referring to the DF scheme, if $\min\{-\delta_{2DF}/2\delta_{1DF}, U_{DF}\} \leq 0$, then consolidation is not recommended for this regime. **iii)** Our design of the problem allows for a possible "surplus" of service satisfaction. Note that the customers arriving during the consolidation period are guaranteed a uniform delivery time. Suppose that customer A arrives at t_A, $0 < t_A < T$, during a particular consolidation period, and is quoted a delivery within T_d time units. Then, another customer, say B, arrives in the same consolidation cycle at time t_B, $0 < t_A < t_B < T$, and is also promised a delivery of T_d time units. Since these two orders are batched in the same cycle, customer B's item will be shipped at least $(t_B - t_A)$ time units earlier

94

than promised. This creates additional goodwill and positive impression about the shipper's operations.

4.5   Numerical Examples and Sensitivity Analysis

To illustrate differences between the pricing schemes used, we present a numerical example with the parameter values:

$$a = 20, \varepsilon = 0.2, s_p = 0.5, s_T = .9, \rho = 30, c = 8, h = 2, \text{ and } L = 2.$$

Table 4.1 presents, for each scheme, the *optimal* value of consolidation cycle length T^*, the corresponding optimal delivery-time guarantee T_d^*, and the optimal price p^* (or the price range for differential pricing scheme, DF). Also shown are the expected total number of arrivals $N(T^*)$ during the consolidation cycle, the profit-rate $\pi(T^*)$, total profit $\Pi(T^*)$ i.e. $T^* \times \pi(T^*)$, and profit per order $[\Pi(T^*)/N(T^*)]$.

We observe from Table 4.1, UF qualifies as the best scheme on the basis of profit rate and total profit per consolidation cycle; UL outperforms other schemes when the objective is profit per order. To investigate how the optimality results may be sensitive to changes in the parameters, these will be varied one at a time while keeping others constant.

	DF	UF	UT	UL
T^*	4.25	4.25	5.00	4.72
T_d^*	6.25	6.25	7.00	6.72
p^*	28.75-29.60	25.00	26.40	29.60
$N(T^*)$	7.23	8.13	7.00	6.02
$\pi(T^*)$	22.10	**34.27**	26.29	24.86
$\Pi(T^*)$	93.93	**145.63**	131.45	117.41
$\Pi(T^*)/N(T^*)$	13.00	17.92	18.78	**19.50**

Table 4.1: Optimal Solutions for the Pricing Schemes

We begin with the sensitivity to delivery-time guarantee, for $0 \leq s_T \leq 1$. In Fig. 4.3, we observe that for $s_T < 0.1$, no pricing scheme yields any profit, and thus consolidation is not an appropriate decision. We also notice that for $s_T < 0.5$, DF is superior to the others, but for greater values of s_T, UF dominates.

Fig. 4.4 depicts the effect of changes in s_T on the total profit generated over the optimal length of consolidation cycle. The choice of best pricing scheme differs with the magnitude of s_T. For delivery-time-sensitive orders, say $s_T \geq 0.6$, UF appears to generate the maximum profit over the optimal interval of consolidation. However, we observe that the profit difference amongst several pricing schemes becomes relatively small for a highly time-sensitive market, i.e. when $s_T > 0.8$.

Regarding the profit per order as the objective function and its sensitivity to s_T, the UF pricing scheme is often worst and never the best (see Fig. 4.5). However, for particular intervals of time sensitivity, each of the other schemes can dominate
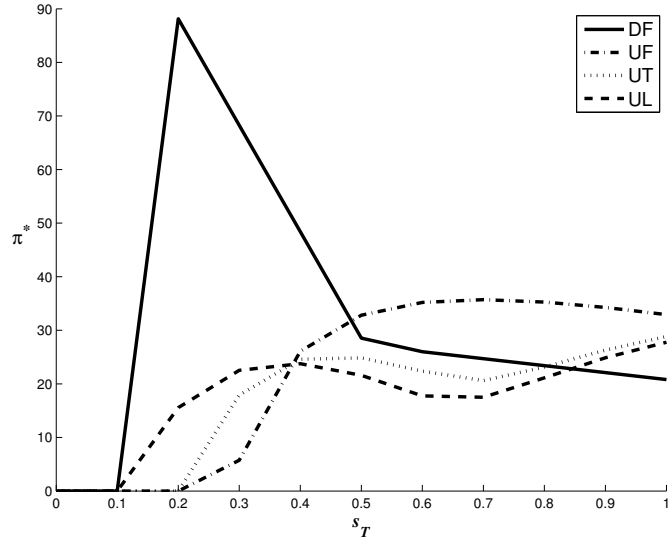
Figure 4.3: Time Sensitivity vs. Optimal Profit Rate

on this criterion. Hence, we have shown in this context that optimal policies can change with respect to the choice of management objectives. (An analytical proof of this observation is presented in Section 4.6.) We can also deduce from Fig. 4.6 that for medium to highly time-sensitive markets (e.g. $s_T \geq 0.4$), the uniform pricing schemes UL and UT are preferable to the DF regime, at least for these particular parameter values.

We notice in Fig. 4.6 that, as s_T increases above the value of 0.5, the optimal consolidation times decrease. (Indeed, they are non-increasing for $s_T \geq 0.3$.) That is true for *all* the pricing schemes at hand. This figure can also help eliminate schemes that are infeasible with respect to service level considerations. For example, if the shipper sets the maximum delivery time to be, say, 13 time units, then for
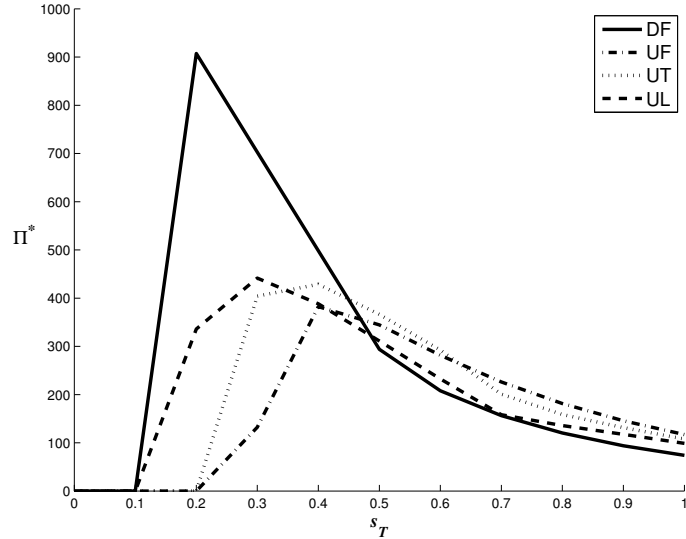
Figure 4.4: Time Sensitivity vs. Total Profit

$s_T < 0.5$, only DF is an admissible scheme.

Further insights are obtained by analyzing how the delivery lead time L, guaranteed by the carrier, impacts the optimal profit rate and the optimal prices. Fig. 4.7 indicates that, regardless of the pricing scheme used, all profit rates decrease as delivery lead time increases. Moreover, no scheme will yield profit if the transit requires 7 time units or more. Indeed, the demand vanishes when the delivery-time guarantee T_t can be promised at no better than 12 time units.

In Fig. 4.8, we notice that the optimal prices are non-increasing in the transit time. The gap between the maximum and minimum prices charged under DF vanishes after some particular delivery transit time. We also observe that for UT the optimal price decreases as L increases.

98
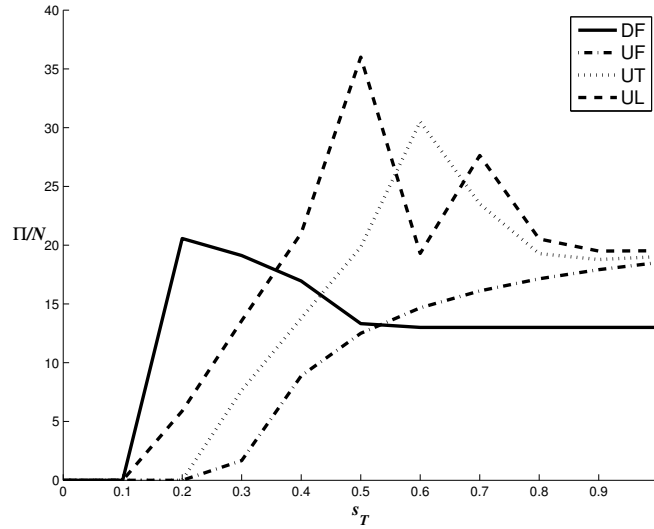
Figure 4.5: Time Sensitivity vs. Profit per Order

Thus, increased transit times negatively affect the demand process, and to be able to generate arrivals, lower prices are suggested by the model.

In sum, these graphs may be used for practical purposes. The manager (here, the shipper) can input the market parameters to the model and decide whether a consolidation program will enable savings or not, and if so at what price. These graphs may also be used as decision aids in choosing the admissible and then the best pricing regime, while seeking cost-saving opportunities in transportation operations. For example, the shipper can choose to work with another carrier if the currently quoted transit time exceeds the maximum time guarantee as required by the particular pricing scheme.
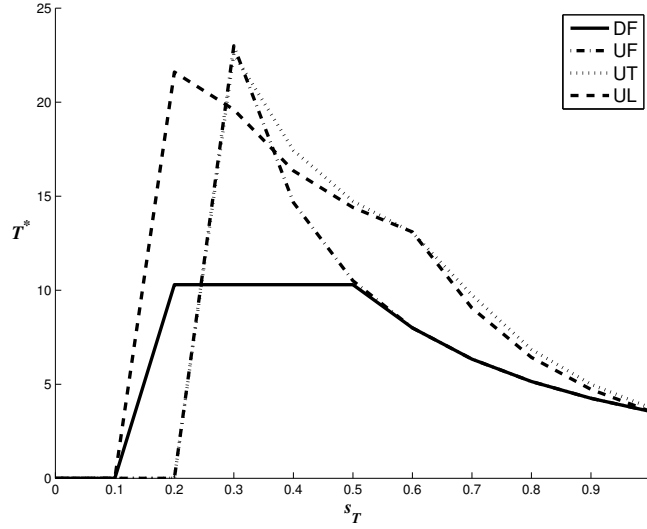
Figure 4.6: Time Sensitivity vs. Optimal Cycle Length

4.6 Additional Results

The numerical examples in the preceding section reveal that, depending on the range of parameters used, some pricing schemes outperform others, based on the choice of the objective function. We now present the conditions under which a particular pricing scheme can dominate the others.

For notational simplicity, let $i, j \in \{$DF, UF, UT, UL$\}$. The case $i \neq j$ thus denotes two distinct pricing schemes.

Proposition 4.1. Given $\delta_{1i} \neq \delta_{1j}$, scheme i dominates j in terms of *profit rate* if one of the conditions (i)-(iv) holds:

   i) $\delta_{1i} > \delta_{1j}$ and $\delta_{2i} > \delta_{2j}$
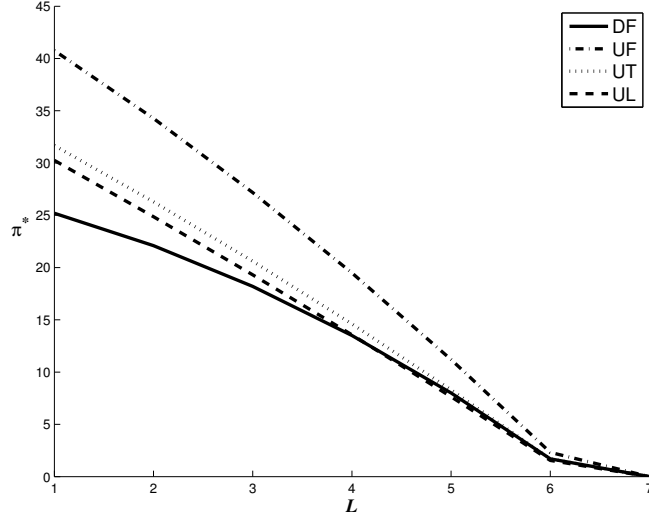
Figure 4.7: Transit Time vs. Optimal Profit Rate

ii) $\delta_{1i} < \delta_{1j}$ and $\delta_{2i} < \delta_{2j}$

iii) $\delta_{1i} > \delta_{1j}$, $\delta_{2i} < \delta_{2j}$, and $T > (\delta_{2j} - \delta_{2i})/(\delta_{1i} - \delta_{1j})$

iv) $\delta_{1i} < \delta_{1j}$, $\delta_{2i} > \delta_{2j}$, and $T > (\delta_{2j} - \delta_{2i})/(\delta_{1i} - \delta_{1j})$

Proof. We compare the quadratic profit rate functions in Section 4.4. Searching for the real roots T that satisfy the inequality $(\delta_{1i} - \delta_{1j})T^2 + (\delta_{2i} - \delta_{2j})T + (\delta_{3i} - \delta_{3j}) > 0$, the preceding conditions are obtained. ■

Proposition 4.1 implies that there exists a length of consolidation period such that any choice of pricing scheme can be dominated by others. Those conditions can also be used to obtain threshold values of consolidation length, above or below which a particular pricing scheme is preferable, if the objective is to maximize the profit per unit time. However, if the shipper wants to evaluate the schemes on the
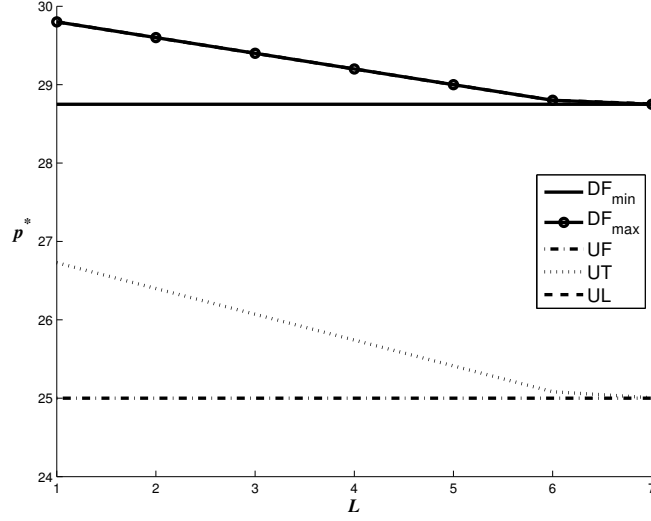
Figure 4.8: Transit Time vs. Optimal Prices

basis of profit per order, she can compare $\Pi_i(T)/N_i(T)$ versus $\Pi_j(T)/N_j(T)$, and then the analysis follows a similar reasoning to that of Proposition 4.1.

In practice, the number of orders processed affect transportation costs. Hence, we consider the ranking of the total number of order arrivals expected for a fixed but arbitrary length of SCL cycle. This is also important to reveal insights when the pricing schemes are compared under a *per-order* performance measure.

Proposition 4.2. For fixed T, the mean total number of arrivals for the respective pricing schemes satisfies $N_{\mathrm{UF}}(T) \geq N_{\mathrm{DF}}(T) = N_{\mathrm{UT}}(T) \geq N_{\mathrm{UL}}(T)$.

Proof. The total arrivals expected over a consolidation period is, for DF

$$
\begin{aligned}
N_{\mathrm{DF}}(T) &= \int_0^T [\alpha - \beta(T + L - t)] dt \\
&= T[\alpha - \beta(T/2 + L)]
\end{aligned}
$$

102

Now, we derive the expected number of order arrivals for the UF case. That is

$$
\begin{aligned}
N_{\mathrm{UF}}(\tau) &= \int_0^T [\alpha - \beta(T+L) + s_T t]dt \\
&= (\tfrac{1}{2}s_T - \beta)T^2 + (\alpha - \beta L)T \\
&= N_{\mathrm{DF}}(T) + \tfrac{1}{2}T^2(s_T - \beta) = N_{\mathrm{DF}}(T) + \tfrac{1}{2}T^2 s_p \varepsilon
\end{aligned}
$$

Below is the expected total number of arrivals for UT.

$$
\begin{aligned}
N_{\mathrm{UT}}(T) &= \int_0^T \lambda_{\mathrm{UT}}(T_t)dt \\
&= \alpha T - T^2\beta - \beta TL - \tfrac{1}{2}\varepsilon s_p T^2 + \tfrac{1}{2}s_T T^2 \\
&= N_{\mathrm{UF}}(\tau) - \tfrac{1}{2}\varepsilon s_p T^2 = N_{\mathrm{DF}}(T)
\end{aligned}
$$

Finally, we find the expected number of order arrivals for the case of UL as

$$
\begin{aligned}
N_{\mathrm{UL}}(T) &= \int_0^T \lambda_{\mathrm{UL}}(T_t)dt = -\tfrac{1}{2}s_T T^2 + \alpha T - \beta LT \\
&= N_{\mathrm{DF}}(T) + \tfrac{1}{2}T^2(\beta - s_T) \\
&= N_{\mathrm{UF}}(T) + T^2(\beta - s_T) = N_{\mathrm{UF}}(T) - T^2 s_p \varepsilon = N_{\mathrm{DF}}(T) - \tfrac{1}{2}T^2 s_p \varepsilon
\end{aligned}
$$

Given that T, s_p and ε are all non-negative, and benchmarking with N_{DF}, we obtain the ranking in Proposition 4.2. ■

Proposition 4.2 also agrees with intuition that the demand generated by UF is necessarily at least as large as that of DF: The price in UF is fixed and is less than that of the DF scheme. Hence, depending on the transportation market

structure, uniform pricing may trigger greater demand and profit compared to that of differential pricing.

To increase transport volume, however, a carrier will offer freight discounts when the order shipped has greater weight. Hence, we now extend our model to include a unit transportation cost that is non-increasing in the weight of load. Given the market conditions, the discount schedule of the carrier, and a particular objective function, what are the best pricing scheme and the optimal price and delivery time-guarantee?

4.7   Incorporating a Freight Discount

Freight discounts are mechanisms used by the carriers to enable them to regulate demand patterns. Various discount schemes e.g. Higginson (1993), are used in logistics management by a public, for-hire trucking company (common carrier). It is thus plausible for the shipper's profit function to account for the carrier's rate structure, and to investigate how those freight rates affect the shipper's decisions. Consider a prototype all-weight freight discount function, $c(w)$

$$c(w) = \begin{cases} c_0 & 0 \le w < W_0 \\ c_1 w, & W_0 \le w < W_1 \\ c_2 w, & W_1 \le w < W_2 \\ \vdots & \vdots \\ c_K w, & W_{K-1} \le w \end{cases}$$

where w is the accumulated weight, $c_1 > c_2 > ... > c_K$ stand for unit-weight freight rates, and $0 < W_0 < W_1 < ... < W_{K-1}$ denote the break-points for shipping larger quantities. We note that c_0 is a fixed charge applied to any order with a weight less than the minimum weight W_0. The rate structure given by $c(w)$ implies that if $c_2 W_1/c_1 \le w < W_1$, then $c(w) \ge c(y)$ for all y such that $W_1 \le y < c_1 W_1/c_2$. (For applications of various discount schemes, see for example. Burwell et al., 1997; Tersine and Barman, 1997; Wang and Wang, 2005; Altıntaş et al., 2008)

Because it is unreasonable to pay more for transporting a smaller weight than a larger one, shippers are allowed to over-declare their actual shipment weight. This means that the shipper has the opportunity to decrease total common-carrier charges by artificially inflating the actual shipping weight to the closest break point, when that makes sense. This strategy of declaring a "phantom weight" is known as taking advantage of the *bumping clause.*

The savings gained from economies of scale from such a tariff trades off with the loss in profit and/or service level that could otherwise have been achieved by setting higher prices and a longer consolidation period. We now demonstrate this

for orders on pallets of a standard shape that accounts for truck capacity.

Let us denote the average weight of each pallet to be μ. Then the total expected weight of a load accumulated during a consolidation period of length τ for a particular pricing scheme $i \in I = \{\text{DF, UF, UT, UL}\}$ is $\bar{w}_i\,(T) \triangleq N_i(T)\mu$. For illustrative purposes, consider a single price break, although the procedure is applicable to additional price breaks. Let c be the non-volume rate and c' the volume rate $(c' < c)$ and W the minimum weight as stated by the carrier to obtain a volume discount. We observe that $\tilde{W} \triangleq (W)(c'/c)$ is the smallest weight where it is logical to employ the bumping clause with reduced volume rates. Hence, if the shipper's expected accumulated load is close to this value $\tilde{W}$, it might be reasonable to lengthen the SCL period and reduce the price, to increase the load to qualify for lower rates.

Recall that $T_{\max}$ is the maximum holding time. To comply with a minimum service level, the maximum holding plus transit time can be at most as long as the maximum delivery-time guarantee that is acceptable to customers. For any pricing scheme $i \in I$, the following procedure in Table 4.2 determines the optimal decision of the shipper when faced with an all-weight freight discount scheme offered by the common carrier.

0)	For the set of pricing schemes $I_0 \subseteq I$, initialize $i \in I_0$.
	Given c and c', find $\tilde{W}$. Choose a performance measure.
1)	Compute T_i^*. If $T_i^* \leq T_{\max}$, go to step 2. O.w., set $T_i^* = T_{\max}$ and calculate the corresponding p_i^* and performance measure. Go to step 3.
2)	Compute mean accumulated weight $\bar{w}_i(T_i^*)$. If $\bar{w}_i < \tilde{W}$, go to step 2b. Else,
2a)	Set c to c' and calculate the performance measure. Go to step 3.
2b)	Find $T_i^{\tilde{W}} \triangleq N_i^{-1}(\tilde{W})$, the expected length of consolidation required to accumulate $\tilde{W}$ orders. If $T_i^{\tilde{W}} > \tau_{\max}$, go to step 3. Otherwise, set c to c', T^* to $T_i^{\tilde{W}}$ and calculate the performance measure.
3)	Select max. performance measure amongst values found in 1, 2a, & 2b.
4)	Set i to the next pricing scheme in the set I_0. Go to step 1.
end)	For every $i \in I_0$, compare results in the corresponding step 3. Choose the best.

Table 4.2: Algorithm Incorporating a Freight Discount

4.8 Summary and Extensions

Transport pricing is an important component of the overall costing and pricing of any good or service. Without lowering the desired service level, it is a challenge to set a price acceptable to customers. In this study, we considered a shipper whose customers are sensitive to both price and delivery-time guarantee. Four pricing schemes were proposed: differential, uniform-first, uniform-mean-time, and uniform-last. For each of these schemes, we studied the problem of maximizing the shipper's profit per unit time through the optimal choice of consolidation period. In addition, through a numerical example, we studied the effects of using total profit per consolidation cycle and profit per order as alternative objectives.

We demonstrated that the problem is very parameter-sensitive, and the best choice of pricing scheme depends upon management's particular objective. Moreover, we have exhibited market conditions for which shipment consolidation might not be preferred at all. We reported on the optimality structures of the problem when distribution operations are provided by a public, for-hire trucking company. Contrary to intuition, charging according to an order's time of arrival is not necessarily the best pricing scheme. We reported on numerical examples with various sensitivity analyses, and proposed an algorithm that incorporates the all-weight freight discount schedule of the carrier.

The models developed in this chapter are applicable for any intermediary-distribution operation. Be it a shipper, freight forwarder or a third party logistics provider, the models can be adapted easily to the operating environment at hand, and may offer the decision maker a better understanding of shipment operations and pricing issues.

This stream of research can be expanded by studying cases in which the arrival process is a more general probability distribution. We can also consider a functional relationship between price and delivery-time quote that is not necessarily linear. Another interesting extension would be to study a uniform pricing scheme where the price is set with respect to the expected arrival time of the median customer

through the consolidation cycle, i.e. when $t = N^{-1}(N(T)/2)$.

Next, in Chapter 5, we consider the joint problem of shipment consolidation and pricing in the context of fleet management, i.e. for the case of private carriage.

Chapter 5

Pricing and Shipment

Consolidation-II

In this chapter, we study the problem of maximizing profit rate for a private carrier whose customers are sensitive both to delivery lead time and the price for that service. Although the demand generation model is the same as in Chapter 4, there are some major differences between the costing mechanisms of a private carrier and a common carrier. Foremost, an operational model for a private carrier need consider the capacity constraints and the fixed costs as opposed to the case of common carriage.

The problem at hand is visually represented in Fig. 5.1. The sequence of events
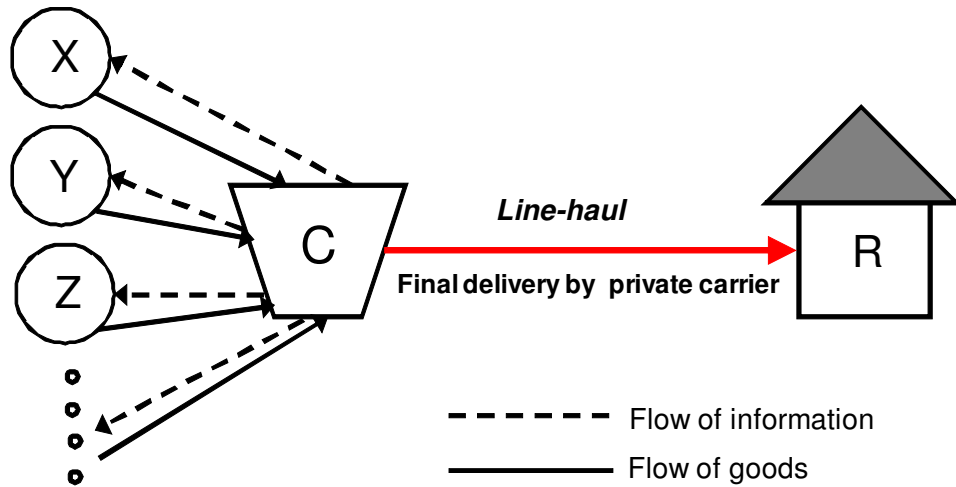
Figure 5.1: Flow of Goods and Information for Private Carriage

is as follows: Given the market conditions, the carrier optimizes the profit per unit time, and announces the resulting optimal prices and delivery-time guarantees. As a consequence of this announcement, the orders accumulate at the carrier's site according to a non-stationary Poisson process. At the end of the optimized consolidation period, the carrier transports the consolidated load to the final destination or the break-bulk terminal.

We model the order-arrival rate as a decreasing function of both price and the delivery time that is guaranteed. (The carrier applies a time-based shipment consolidation program, so that he can quote delivery times that are certain.) We propose two types of pricing mechanisms to be used: *Differential Pricing* (DF), in which the customers are charged with respect to their exact arrival time within the total consolidation period, and *Discrete Pricing* (DS), in which the order is charged

111

according to the sub-interval in which the arrival time lies, and is an easier-to-use mechanism.

Though there exists a vast body of scholarly research on pricing and ordering decisions, the literature on joint SCL and pricing is scant. The analysis in this chapter aims to shed light on this issue and to moderately fill the gap. To our knowledge, Cavinato (1982) is the only paper that addresses pricing issues for a private carrier, and he investigates the industry data while giving an overview of uses, problems and advantages of employing transfer pricing. Bookbinder and Higginson (2002) is the most recent work that gives analytical results for private carriage in the SCL problem, yet in a different setting where demand is exogenous.

In what follows, we address the following research questions:

i) What are the optimal price and consolidation time for a private carrier supplying customers that are price- and delivery-time sensitive?

ii) How are the profit, the optimal consolidation time, and the prices affected by model parameters such as cost and capacity factors?

iii) What is the benefit of discrete pricing as opposed to that of differential?

Our optimization model considers the revenue generated, the holding (or consolidation penalty) cost, and the fixed dispatch cost per truck, among others. We

112

assume that the carrier has an ample number of trucks to employ for these orders destined to a single location.

Here, the main modeling distinction from Chapter 4 is that the costing structure for a private carrier is different, plus the capacity constraint is explicitly included. Also, the fixed cost brings in extra difficulty. One assumption here is that the carrier works with a single capacity. However, this is not a shortcoming from the modeling perspective if that capacity is thought of as a combination of fleet capacity. Also from a practical point of view, once a vehicle is loaded (depending on the SCL policy employed), there is no point in waiting for the other vehicles. (The modification to that is straightforward.)

On the other hand, similar to the previous chapter, we employ a time-based shipment-release policy and choose our objective as the profit per unit time. We note that the CD-P, H-P and T-P are the only policies that a shipper may use as a means to quote delivery time guarantees. Yet, since the order arrival pattern follows a (nonstationary) Poisson process, as opposed to the deterministic settings, variability in the consolidation operations, unavoidably delaying shipments and requiring expediting, etc. are a part of this model. Our model attempts to give a guideline for a private carrier to (possibly) improve his operational efficiency and (perhaps) increase his profitability.

For our modeling purposes, we use the following

Parameters

a : maximum potential demand

s_p : sensitivity to price, in dimensions of demand per dollar

s_T : sensitivity to delivery-time guarantee (demand per unit time)

ρ : market-reservation price per unit

ε : delivery-time sensitivity of the price (dollars per unit time)

C : vehicle capacity

κ_1 : fixed order processing cost (independent of batch size)

κ_2 : fixed dispatching cost per vehicle

h_1 : handling cost per unit (e.g. loading, sorting, unloading)

h_2 : holding cost per unit per unit time

L : line-haul time

c : line-haul cost per unit time

$T_{\max}$: upper bound for delivery-time guarantee

Decision variables

T : length of consolidation period (or, the consolidation cycle)

T_t : delivery-time guaranteed for a customer arriving at time t

p_{T_t} : price per unit charged to an order arriving at time t, in dollars

As in Chapter 4, we use the following demand relationships between the delivery

114

time guarantee and price.

$$\lambda(p_{T_t}, T_t) = a - s_p p_{T_t} - s_T T_t, \ 0 < t < T$$

$$p_{T_t} = \rho - \varepsilon T_t, \ \rho > 0, \ \varepsilon > 0, \ p > 0 \tag{5.1}$$

For simplicity in the analysis, one can rewrite the demand function as follows:

$$\lambda(T_t) = \alpha - \beta T_t, \ 0 < L < T_t < T_{\max} \ \text{ where}$$

$$\alpha = a - s_p \rho \ \text{and} \ \beta = s_T - s_p \varepsilon \tag{5.2}$$

Due to our modeling assumptions, given the optimal length of consolidation T^*, the corresponding optimal prices for the DF case can be easily computed as

$$p_{\mathrm{DF}}^*(t) = \rho - \varepsilon(T_{\mathrm{DF}}^* + L - t) \tag{5.3}$$

The expected total profit per dispatch $\Pi(.)$ is

$$\Pi(.) = \text{revenue} - (\text{holding} + \text{transportation}) \ \text{cost} \tag{5.4}$$

Positive demand generation implies that $\alpha - \beta T_t \geq \alpha - \beta(T + L) \geq 0$, hence $T \leq (\alpha/\beta) - L$, and also $T \leq T_{\max} - L$. For profit generation, the condition that the prices need be positive requires that $\rho - h_1 - \varepsilon(T^* + L) \geq 0$, which in turn implies $T \leq (\rho/\varepsilon) - L$. Hence $T \leq U$, where we define U as the upper bound on the consolidation cycle as

$$U = \min\{(\alpha/\beta), (\rho - h_1)/\varepsilon, T_{\max}\} - L \tag{5.5}$$

In the remainder of this chapter, as in Chapter 4, we assume that α is positive,

since otherwise, the arrival rate (if $\beta > 0$) would always be negative, which is not admissible.

(5.4) denotes the objective function we want to optimize over T: It is the revenue generated less (holding cost + dispatch cost + line-hauling cost). Also, (5.5) ensures that the consolidation cycle is admissible. By imposing U, we implicitly introduce the natural requirements of this problem: There need be a positive demand generation, the price need be nonnegative, and the delivery time guarantee must be satisfied.

Model DF:

$$\sup_{T} \mathrm{E}[\pi_{\mathrm{DF}}(T)] = \mathrm{E}[\pi_{\mathrm{DF}}(T)]/T \ni 0 < T \leq \mathrm{U}\} \tag{5.6}$$

Below, the terms that frequently occur in the model are written explicitly. These terms are the expected profit rate $\pi_{\mathrm{DF}}(T)$, and the expected number of consolidated orders $Q_{\mathrm{DF}}(T)$. Since once the vehicle is dispatched, a new cycle begins, we naturally define $\pi_{\mathrm{DF}}(T)$ as the total *expected* cost $\mathrm{E}[\Pi(T)]$ in a cycle of length T divided by its length. (For a unique solution to the model DF, one can easily calibrate the data and the decision variable T such that 1 is set to be the lower bound.) After

some algebra, we find

$$
\begin{aligned}
\mathrm{E}\left[\pi_{\mathrm{DF}}(T)\right] &= T^{-1}\left[\int_0^T (p_{T_t}-h_1)\lambda(T_t)dt - h_2\int_0^T (T-t)\lambda(T_t)dt \right. \\
&\quad \left. -\kappa_1-(\kappa_2+cL)\left\lceil\int_0^T \lambda(T_t)dt/C\right\rceil\right] \\
&= \mathrm{a}_1 T^2+\mathrm{b}_1 T+\mathrm{c}_1 T^{-1}+\mathrm{d}_1-(\kappa_2+cL)\left\lceil(\mathrm{e}_1 T^2+\mathrm{f}_1 T)/C\right\rceil T^{-1} \\
\text{where} \quad &\mathrm{a}_1 = (1/3)\beta(\varepsilon+h_2),\ \ \mathrm{b}_1 = (1/2)[\beta(h_1+2L\varepsilon-\rho)-h_2(\alpha-L\beta)-\alpha\varepsilon] \\
&\mathrm{c}_1 = -\kappa_1,\ \ \mathrm{d}_1 = (\alpha-L\beta)(\rho-h_1-L\varepsilon),\ \mathrm{e}_1 = -\beta/2,\ \ \mathrm{f}_1 = \alpha-l\beta
\end{aligned}
$$

(5.7)

and where $\lceil x \rceil$ is the smallest integer greater than or equal to x. $Q(T)$ is a random variable representing the total number of orders received by time T. $\Lambda(t)$ is the intensity function, and $m(t)$ the mean value function. Then $Q(T)$ has a Poisson distribution with mean $m(T)$. Also let us note that the expected total number of orders *consolidated* by time T can be determined as

$$
Q_{\mathrm{DF}}(T) = \int_0^T \lambda(T_t)dt = -\beta/2T^2+(\alpha-L\beta)T \tag{5.8}
$$

To ease analysis, we can rewrite $\pi(T)$ removing the ceiling function, and thus the non-continuity in the function is removed. Call this modified (relaxed) profit rate function $\pi_{\mathrm{DF}}^{\mathrm{m}}(T)$ which is derived to be

$$
\begin{aligned}
\pi_{\mathrm{DF}}^{\mathrm{m}}(T) = \ &\mathrm{a}_1^{\mathrm{m}} T^2+\mathrm{b}_1^{\mathrm{m}} T+\mathrm{c}_1^{\mathrm{m}} T^{-1}+\mathrm{d}_1^{\mathrm{m}}\ \text{where} \\
&\mathrm{a}_1^{\mathrm{m}} = \mathrm{a}_1,\ \ \mathrm{b}_1^{\mathrm{m}} = \mathrm{b}_1+\beta\left(\kappa_2+Lc\right)/2C \\
&\mathrm{c}_1^{\mathrm{m}} = \mathrm{c}_1,\ \ \mathrm{d}_1^{\mathrm{m}} = \mathrm{d}_1-(\alpha-L\beta)\left(\kappa_2+Lc\right)/C
\end{aligned}
$$

(5.9)

5.1 The Optimality Analysis

Let us notice the obvious fact that, depending on the functional form of $\pi_{\mathrm{DF}}(T)$, the optimality conditions will be different when $\pi_{\mathrm{DF}}(T)$ is concave, and when it is convex We will first focus on the modified function. Based on results there, we will extend the findings to our original objective function. To this end, we give the following

Lemma 5.1. $\pi_{\mathrm{DF}}^{\mathrm{m}}(T)$ is concave if $T^3 \geq 3\kappa_1/\beta(\varepsilon + h_2)$, and is convex otherwise. Moreover, if $\beta < 0$, then $\pi_{\mathrm{DF}}^{\mathrm{m}}(T)$ is convex everywhere.

Proof. Simply checking the second derivatives, we obtain those conditions above. Also, since $T, \kappa_1 \geq 0$, and $\varepsilon, h_2 > 0$, $T^3 \geq 3\kappa_1/\beta(\varepsilon + h_2)$ when $\beta < 0$, and thus $\pi_{\mathrm{DF}}^{\mathrm{m}}(T)$ is convex. ■

The modified profit rate function $\pi_{\mathrm{DF}}^{\mathrm{m}}(T)$ can be concave or convex in its particular ranges. Accordingly, the (unique) optimal solution of T can be an interior or a boundary solution. This is reflected in (5.10). As noted earlier, for the case of convexity, we can impose a minimum value of T (here it is 1) for mathematical consistency.

$$T_0 = \arg\max \begin{cases} \pi_{\mathrm{DF}}(T_i) \ni \frac{d\pi_{\mathrm{DF}}^{\mathrm{m}}(T_i)}{dT_i} = 0,\, 0 < T_i \leq U,\ \text{if } \pi_{\mathrm{DF}}^{\mathrm{m}} \text{ concave} \\ \pi_{\mathrm{DF}}(T_i) \ni T_i = 1,\, \text{or } T_i = U,\ \text{if } \pi_{\mathrm{DF}}^{\mathrm{m}} \text{ convex} \end{cases} \tag{5.10}$$

Hence, T_i, $1 \leq i \leq 3$, are the feasible solutions for the relaxed model, and T_0

is the one that maximizes the objective function. We note that the solution of $\frac{d\pi^{\mathrm{m}}(T)}{dT} = 0$ requires us to find the real roots of the cubic function

$$(2\mathrm{a})T^3 + [\mathrm{b} - (\mathrm{e}\kappa_2/C)]T^2 - \mathrm{c} = 0$$

A cubic function with real coefficients has at least one, and at most three, real roots. Hence, the search over the candidate τ_i roots need not be exhaustive. We can now define the upper bound on the number of vehicles by

$$n_0 = \lceil Q(T_0)/C \rceil \tag{5.11}$$

A "candidate" solution for the optimal SCL length is the one that enables full utilization of the vehicle capacity. We can now define the feasible sets for the candidate solutions as

$$
\begin{aligned}
\mathrm{S_u} &= \arg\left\{Q(T) = n_0 C,\ 1 \le T \le \mathrm{U_1}\right\} \\
&= \left\{(1/2\mathrm{e})(-\mathrm{f} - \sqrt{\mathrm{f}^2 + 4\mathrm{e}n_0 C}),\ (1/2\mathrm{e})(-\mathrm{f} + \sqrt{\mathrm{f}^2 + 4\mathrm{e}n_0 C})\right\}
\end{aligned} \tag{5.12}
$$

$$
\begin{aligned}
\mathrm{S_l} &= \arg\left\{Q(\tau) = (n_0 - 1)C,\ 1 \le T \le \mathrm{U_1}\right\} \\
&= \left\{(1/2\mathrm{e})(-\mathrm{f} - \sqrt{\mathrm{f}^2 + 4\mathrm{e}(n_0 - 1)C}),\ (1/2\mathrm{e})(-\mathrm{f} + \sqrt{\mathrm{f}^2 + 4\mathrm{e}(n_0 - 1)C})\right\}
\end{aligned} \tag{5.13}
$$

Proposition 5.1. The T^* that solves (5.6) is either T_l, T_0 or T_{u}, where these quantities are given by the equations below and in (5.10). Moreover, T_0 is a candidate

119

optimal solution only if $T_l < T_0 < T_u$.

$$T_u = \arg\max\{\pi(T),\ T \in S_u\}$$

$$T_l = \arg\max\{\pi(T),\ T \in S_l\}$$

Proof. Since T_0 is the optimal solution to the relaxed problem, and since $\pi_{\mathrm{DF}}^{\mathrm{m}}(T)$ is the upper-bound function for $\pi_{\mathrm{DF}}(T)$, $\pi_{\mathrm{DF}}^{\mathrm{m}}(T_0) \geq \pi_{\mathrm{DF}}(T_0)$. Employing the upper bound on T in (5.5), one can easily show that $Q_{\mathrm{DF}}(T)$ is increasing in T. Hence, $T_l \leq T_0 \leq T_u$. Also since $\pi_{\mathrm{DF}}(T)$ is concave over $[T_l, T_u]$, $\pi_{\mathrm{DF}}(T_0) \geq \pi_{\mathrm{DF}}(T_l)$ and $\pi_{\mathrm{DF}}(T_0) \geq \pi_{\mathrm{DF}}(T_u)$. ∎

5.2   A Numerical Example

In this section, we wish to get some insights about the behavior of our model. Employing the base case data in Table 5.1, we first graph the expected total profit function and its cost components in Fig. 5.2.

In that figure, we observe that the total profit function displays a piecewise concave structure. When the costs are averaged over an SCL cycle, we obtain Fig.

a	ρ	ε	s_p	s_T	κ_1	κ_2	h_1	h_2	c	C	L	$T_{\max}$
80	20	1	0.9	0.5	30	150	3	2	15	100	4	14

Table 5.1:   Base Case Data Set for Private Carriage

5.3, in which the expected profit rate function and its upper bound, the modified profit function are depicted. Hence, we note that the expected profit rate function is bounded from above by the modified profit rate function, $\pi^{\mathrm{m}}_{\mathrm{DF}}(T)$.
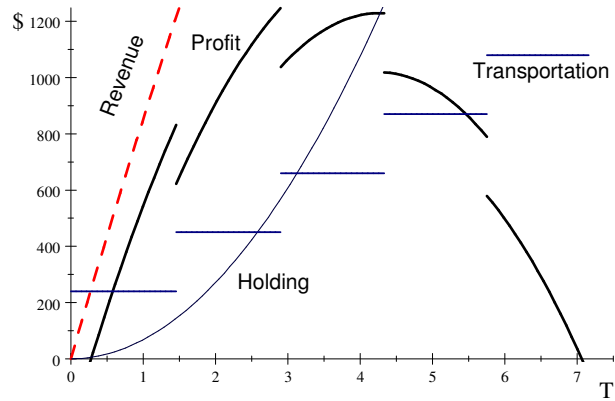

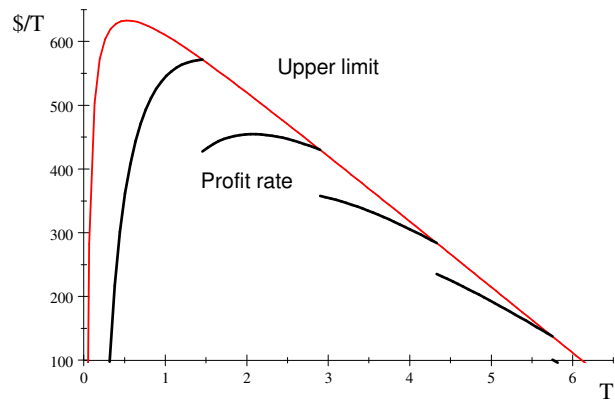Figure 5.2: Expected Total Profit and Its Components

Figure 5.3: Profit Rate and Its Upper Bound

121

5.3 Discrete Pricing

For any 3PL (e.g., the carrier) and the customers, it might be more convenient to give a uniform price for specific intervals of the consolidation period. Here, we analyze such a situation, which we call "discrete pricing."

In discrete pricing (DS), prices will be computed with respect to the interval the customers fall in. Assume $[0, T]$ is the consolidation period. Let us discretize this length into N equally spaced intervals, where each subinterval i has a length $\Delta_i = T/N$, $i = 1, .., N$. Hence, subinterval i will cover the portion of the consolidation period between $[T_{i-1}, T_i]$, $i = 1, .., N$, and $T_0 = 0$, $T_N = T$. We will assume that all customers arriving in this particular subinterval will be uniformly charged, as if arriving in the middle of the subinterval. Therefore, we can also define N as the number of price switches. This is similar to the "Mean-Time Pricing" scheme employed in Chapter 5. Accordingly, the prices and the demand function are affected by this modification. Setting the maximum delivery-time guarantee for an order arriving in i^{th} sub-interval $T_i = (N - i + 0.5)/(T/N)$, we find for $i = 1, ..., N$

$$\lambda_i = \alpha - \beta[(T/2N)(2(N-i)+1)+L]$$

$$p_i = \rho - h_1 - \varepsilon[(T/2N)(2(N-i)+1)+L]$$

We can now find the optimal expected total profit by the sum of the aggregated subinterval revenues less the holding costs. Since the transportation cost is common

to all orders consolidated, it will be included only in the final profit term.

The expected revenue less the holding cost, say R_i, for the orders accumulated in the interval i, can be calculated as

$$
\begin{aligned}
R_i \ &= \ \int_{T_{i-1}}^{T_i} p_i(t)\lambda_i(t)dt - h_2 \int_{T_{i-1}}^{T_i}(T-t)\lambda_i(t)dt \\[2mm]
&= \ \int_0^{\frac{T}{N}}[\rho - h_1 - \varepsilon(\tfrac{T}{2N}(2(N-i)+1)+L)][\alpha - \beta(\tfrac{T}{2N}(2(N-i)+1)+L)]dx \\[2mm]
&\quad - \int_0^{\frac{T}{N}}(T-x)(\alpha - \beta(\tfrac{T}{2N}(2(N-i)+1)+L))dx \\[2mm]
&= \ (T/4N^2)\left(2T\beta N^2 + (1-2i)\,T\beta N - 2\alpha + 2L\beta\right) \\[2mm]
&\quad \times \left(2Nh_1 - 2N\rho - Th_2 + 2NL\varepsilon + 2NTh_2 + (1-2i)\,N^2T\varepsilon + 2N^3T\varepsilon\right)
\end{aligned}
$$

The expected number of arrivals in interval i can be computed as

$$
\begin{aligned}
Q_i \ &= \ \int_{T_{i-1}=0}^{T_i=T/N}\lambda_i(t)dt \\[2mm]
&= \ -\tfrac{1}{2N}[NT^2\beta(2+(1-2i)\,N-2T(\alpha-L\beta)]
\end{aligned}
$$

and the expected total number of arrivals is

$$
Q_{\mathrm{DS}}(T) = \sum_{i=1}^N Q_i = \left(-\tfrac{1}{2}N^2\beta\right)T^2 + (\alpha - L\beta)\,T
$$

We note that when $N=1$, we obtain the same result as for DF pricing. Of course, for different values of N and β, the expected total accumulation of orders differs.

The profit rate in DS pricing case now can be found as

$$\pi_{\mathrm{DS}}(T) = T^{-1}\left(\sum_{i=1}^{N} R_i - \kappa_1 - (\kappa_2 + cL)\lceil Q_{DS}(T)/C\rceil\right)$$

$$= a_2 T^2 + b_2 T + c_2 T^{-1} + d_2 - (\kappa_2 + cL)\lceil (e_2 T^2 + f_2 T)/C\rceil T^{-1} \text{ where}$$

$$a_2 = (N\beta/12)[(\varepsilon(4N^3 - N) + 3h_2(2N-1)],$$

$$b_2 = \frac{N^2}{2}[(\beta(h_1 - \rho) - \varepsilon(\alpha - 2L\beta)] - \frac{h_2}{2N}(\alpha - L\beta)(2N-1),$$

$$c_2 = -\kappa_1, \quad d_2 = (\alpha - L\beta)(\rho - h_1 - L\varepsilon),$$

$$e_2 = -(1/2)N^2\beta, \quad \text{and } f_2 = \alpha - L\beta.$$

Now, we have the following model for the Discrete Pricing scheme, given the desired number of price switches.

Model DS:

$$\max_{T} E[\pi_{\mathrm{DS}}(T)] \ni \{0 < T \le U,\, N \in \mathbb{Z}^+\} \tag{5.14}$$

The procedure used to solve Model DS is similar to DF. For fixed N, relax the model and then follow the same procedure as above. However, it is interesting to see how the number of price switches affects the optimality decision.

Proposition 5.2. For a fixed value of T, say τ, the number of price switches required in Model DS that gives the same optimal profit rate in Model DF can be found by solving for N in the following equality:

$$\pi_{\mathrm{DS}}(N)|_{T=\tau} = \pi_{\mathrm{DF}}(\tau),\, 0 < \tau \le U, N \in \mathbb{Z}^+$$

Proof. Since the procedures for solving both DF and DS are the same once the initial optimum is found, if it exists, the value of N can be obtained. ∎

To examine the optimality gap between the upper bounds of the solutions to the models DS and DF, we find the modified function for DF first, again by treating the ceiling function as a continuous one as in (5.9). We obtain

$$
\begin{aligned}
\pi_{\mathrm{DS}}^{\mathrm{m}}(T) = \quad & \mathrm{a}_2^{\mathrm{m}} T^2 + \mathrm{b}_2^{\mathrm{m}} T + \mathrm{c}_2^{\mathrm{m}} T^{-1} + \mathrm{d}_2^{\mathrm{m}} \text{ where} \\
& \mathrm{a}_2^{\mathrm{m}} = \mathrm{a}_2, \ \ \mathrm{b}_2^{\mathrm{m}} = \mathrm{b}_2 + N^2 \beta \left(\kappa_2 + Lc\right)/2C \\
& \mathrm{c}_2^{\mathrm{m}} = \mathrm{c}_2, \ \ \mathrm{d}_2^{\mathrm{m}} = \mathrm{d}_2 - (\alpha - L\beta)\left(\kappa_2 + Lc\right)/C
\end{aligned}
\tag{5.15}
$$

We now can show the following

Proposition 5.3. For a fixed value of T, say τ, the optimality gap between the upper bounds of the models DF and DS can be computed as $\pi_{\mathrm{DF}}^{\mathrm{m}}(\tau) - \pi_{\mathrm{DS}}^{\mathrm{m}}(\tau)$.

Proof. From (5.9) and (5.15), the optimality gap between the upper bounds of the models DF and DS can then be computed as

$$
\pi_{\mathrm{DF}}^{\mathrm{m}}(\tau) - \pi_{\mathrm{DS}}^{\mathrm{m}}(\tau) = (\mathrm{a}_1^{\mathrm{m}} - \mathrm{a}_2^{\mathrm{m}})T^2 + (\mathrm{b}_1^{\mathrm{m}} - \mathrm{b}_2^{\mathrm{m}})T + (\mathrm{c}_1^{\mathrm{m}} - \mathrm{c}_2^{\mathrm{m}})T^{-1} + (\mathrm{d}_1^{\mathrm{m}} - \mathrm{d}_2^{\mathrm{m}}). \ ∎
$$

5.4   Sensitivity Analysis

It is of interest to investigate how the various optimal decisions are affected by the changes in model parameters, ceteris paribus. Due to the heavy parametric nature

of our models, it is more practical and insightful to work with graphs, rather than comparative statics. Naturally, the data of the base case are employed to produce these graphs. (Analysis is only for the DF case; DS gives similar results.)

The models proposed in the preceding sections are solved by an algorithm coded in MATLAB, and run for parameters of interest in the desired ranges. As a major difference from similar models studied in Chapter 5, we first study the sensitivity of the optimal profit rate to vehicle capacity.

It is observed from Fig. 5.4 that the optimal profit rate shows an erratic behavior, with "plunges" at certain capacity levels. These downward spikes happen when the optimal SCL length requires a not-fully utilized vehicle. Generally, however, the behavior is approximately concave. The choice of vehicle capacity, as expected, affects the profitability level. One can claim that, for given parameters, there exists a certain range of capacity levels in which the profit rate is maximized. Also, there exists a threshold limit of capacity above which the profit rate vanishes, and hence it is not advised that the private carrier acquire or work with such types of vehicles. (One can easily notice that the term, *capacity*, needs to be calibrated with respect to the context of the operating environment. For example, if the focus of the carrier shifts to a smaller size of shipments compared to the current standard, his vehicle capacity is magnified automatically.)
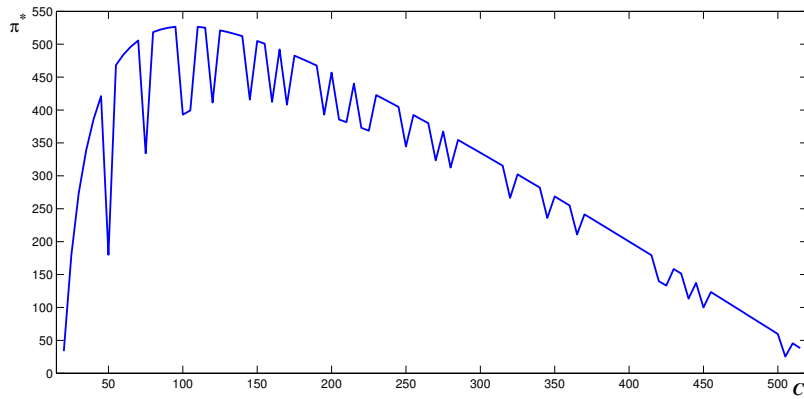
Figure 5.4: Vehicle Capacity vs. Optimal Profit Rate

The relationship between the optimal consolidation cycle lengths versus the capacity (see Fig. 5.5), reveals that as the vehicle capacity increases, it is optimal to wait longer for a higher level of load accumulation. Also, in line with this observation and as seen from Fig. 5.6, the optimal minimum decreases in the vehicle capacity.
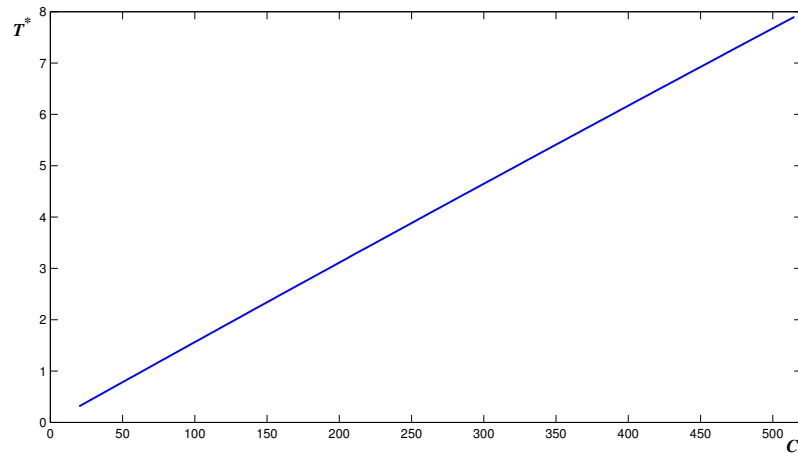

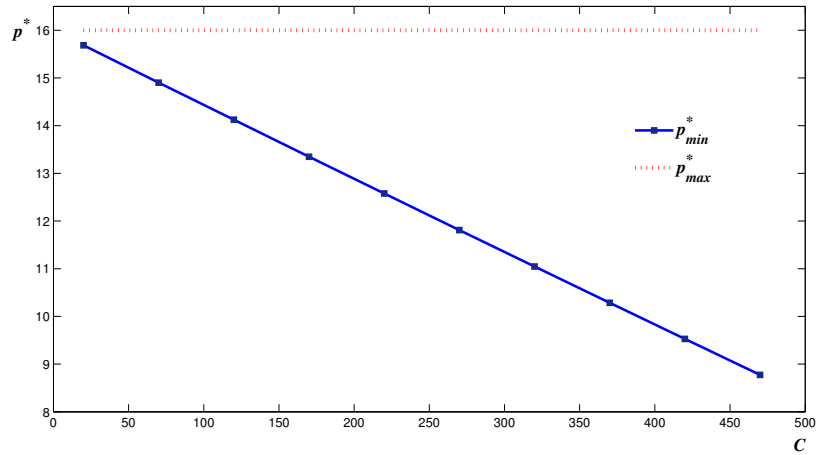Figure 5.5: Vehicle Capacity vs.Optimal Consolidation Cycle

127

Figure 5.6: Optimal price range vs. capacity

Since customers are sensitive to lead time, it is important to include the parameter line-haul time L in the sensitivity analysis. Fig. 5.7 depicts the relationship between the optimal profit rate and line-haul time. We observe that after a certain threshold of L, the profit vanishes. Hence, at the design stage, the carrier must be careful to find the most economical and fastest way to line-haul between the zone of customers and the final destination. Also, consistent with intuition, and the modeling assumptions, the profit rate is maximized when the line-haul time is minimized, which comes with the trade-off of higher cost.

Now, we consider the impact of (common) order processing cost on the optimal profit rate in Fig. 5.8. Approximately, the optimal profit rate decreases with the order processing cost. However, that relationship is neither linear nor concave. When κ_1 is around 300, the profit rate is maximized. That behavior can be explained by
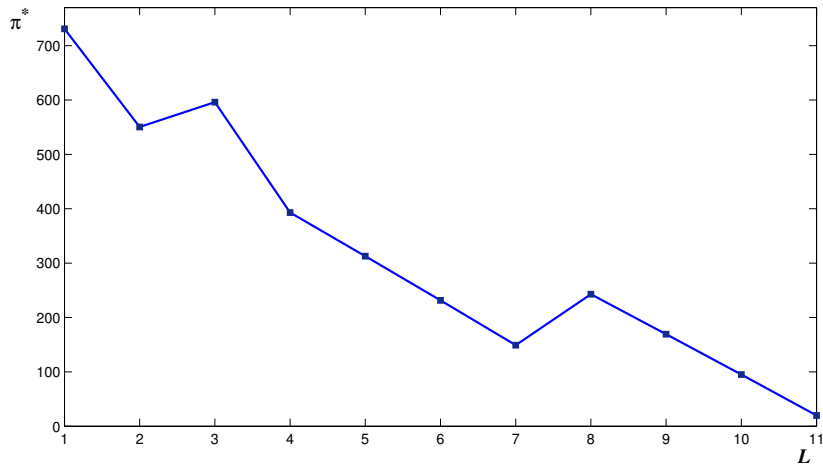
Figure 5.7: Line-haul Time vs. Optimal Profit Rate

the highly nonlinear form of our model. Similar nonlinearities show up when we investigate the effects of price sensitivities on the optimal profit rate, i.e. in Fig. 5.9. However, again, we can still observe that the optimal profit rate approximately decreases as the price sensitivity of the customers gets larger.
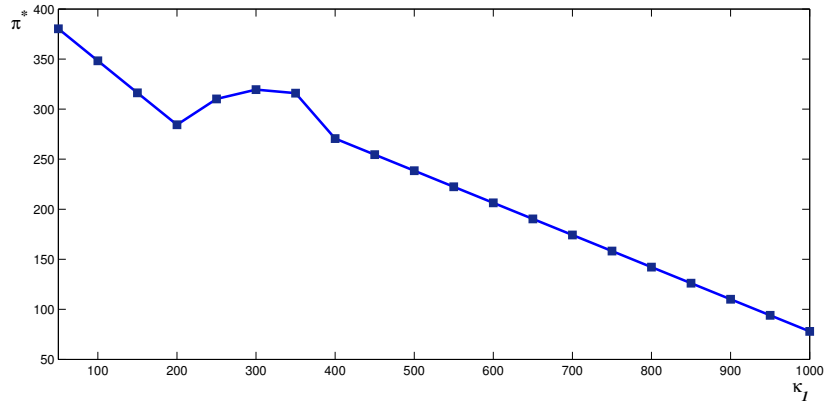
129

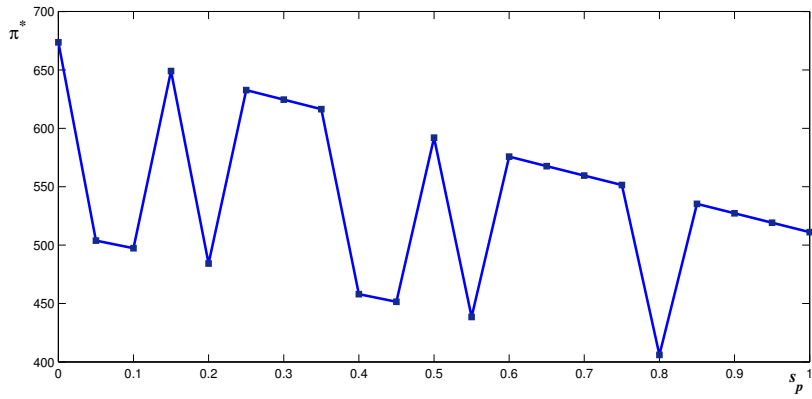Figure 5.8: Order Processing Cost vs. Optimal Profit Rate

Figure 5.9: Price Sensitivity vs. Optimal Profit Rate

130

Chapter 6

Final Thoughts and Future

Research

Shipment consolidation, apart from its substantial benefits in the logistics supply chain, is quite a rich research topic. Here, we aimed at improving on the current literature on SCL, and introduced new models. We used various Operations Research concepts or techniques such as Renewal Theory, Dynamic Optimization, and simulation.

The first research axis of this dissertation focused on analytical models specifically for single-item shipment consolidation, when the orders arrive randomly. Hence, in Chapter 2, we first examined the conditions under which an SCL pro-

gram enables positive savings. Then, in addition to the current SCL policies in the literature, we introduced a new one, i.e. Controlled Dispatch Policy. Moreover, we provided a cost-based comparison of those policies and showed that Quantity Policy yields the lowest cost per order amongst the other, yet with the highest randomness in dispatch times. On the other hand, in between the service-level dependent policies (i.e. the CD-P, H-P and T-P), while H-P provided the lowest cost per order, CD-P turned out to be more flexible and responsive to dispatch times, again with a lower cost than the T-P. An extension to SISCLP would be to compare those policies on other dispatch objectives as well, and investigate the cost variability thereof, if possible, *analytically.*

We offered a new problem in Chapter 3: Multi-Item Shipment Consolidation Problem. We employed and showed the optimality of myopic analysis when costs and the order-arrival processes were dependent on the type of items. In a dynamic setting, we employed a concept of time-varying probability to integrate the dispatching and load planning decisions. That model was analyzed for the most common objectives through simulation, and we observed that the cost and optimal SCL cycle variability was smaller for the objective of cost-per-unit weight, a result that justifies the industrial practices. Although our model in that chapter lay the foundation for it, we leave aside a technically very-challenging, yet quite useful topic of research; that is the "real-time" optimization of dispatch *and* loading decisions.

As the third axis, we studied joint optimization of pricing and SCL in a price-and time-sensitive logistics market, both for the common carriage (in Chapter 4) and private carriage (Chapter 5). The main motivation for introducing pricing in SCL decisions stems from the fact that transportation is a service and naturally demand is affected by the prices. Hence, if possible, influencing the order arrival rates by suitable pricing decisions enables the LTL carrier extra savings. Those savings emanate from two sources; scale economies (in private carriage) or discount economies (in common carriage) that comes with SCL, and revenue generated by employing an appropriate pricing scheme. We could show, contrary to the current literature, that a differential (or dynamic) pricing scheme does not necessarily yield higher profits than a uniform (or single) one. A challenging extension to our model for the common carriage would be to study that problem under a *general* price-time functional form. On the other hand, for the case of private-carriage, explicit incorporation of fleet management decisions into SCL remains another interesting topic.

Bibliography

P. L. Abad and V. Aggarwal. Incorporating transport cost in the lot size and pricing decisions with downward sloping demand. *International Journal of Production Economics*, 95:297–305, 2005.

W. M. Abdelwahab and M. Sargious. Freight rate structure and optimal shipment size in freight transportation. *Logistics and Transportation Review*, 26:271–292, 1990.

P. I. Akaah and G. C. Jackson. Frequency distribution of the weights of customer orders in physical distribution systems. *Journal of Business Logistics*, 9:155–164, 1988.

O. Alp, N. K. Erkip, and R. Güllü. Optimal lot-sizing / vehicle-dispatching policies under stochastic lead times and stepwise fixed costs. *Operations Research*, 51: 160–166, 2003.

N. Altıntaş, F. Erhun, and S. Tayur. Quantity discounts under demand uncertainty. *Management Science*, 54:777–792, 2008.

S. Anily and M. Tzur. Shipping multiple-items by capacitated vehicles - an optimal dynamic programming approach. *Transportation Science*, 39:233–248, 2005.

J. R. Artalejo. A note on the optimality of the N- and D- policies for the M/G/1 queue. *Operations Research Letters*, 30:375–376, 2002.

A. Attanasio, A. Fuduli, G. Ghiani, and C. Triki. Integrated shipment dispatching and packing problems: a case study. *Journal of Mathematical Modelling and Algorithms*, 6:77–85, 2007.

K. R. Balachandran. Control policies for a single server system. *Management Science*, 19:1073–1076, 1973.

K. R. Balachandran and H. Tijms. On the D-policy for the M/G/1 queue. *Management Science*, 21:1013–1018, 1975.

R. H. Ballou. *Business logistics / supply chain management*. Upper Saddle River, NJ, USA, 2004.

R. H. Ballou. Revenue estimation for logistics service offerings. *International Journal of Logistics Management*, 17:21–37, 2006.

D. O. Bausch, G. Brown, and D. Ronen. Consolidating and dispatching truck shipments of mobil heavy petroleum products. *Interfaces*, 25:1–17, 1995.

M. Beckmann, T. C. Koopmans, C. B. McGuire, and B. Winsten. The allocation of switching work in a system of classification yards. In *Proceedings of Winter Meeting of the Railway Systems and Procedures Association*, pages 94a–94j, University of Chicago, IL, 1953.

M. Beckmann, C. B. McGuire, and B. Winsten. *Studies in the Economics of Transportation*. Yale University Press, New Haven, USA, 1956.

C. E. Bell. Characterization and computation of optimal policies for operating an M/G/1 queuing system with removable server. *Operations Research*, 19:208–218, 1971.

J. H. Bookbinder. Dynamics of freight. *Materials Management and Distribution*, December:20–21, 1989.

J. H. Bookbinder and J. K. Higginson. Probabilistic modeling of freight consolidation by private carriage. *Transportation Research-E*, 38:305–318, 2002.

J. H. Bookbinder and M. E. Lynch. Customer service in physical distribution: a utility-function approach. *International Journal of Physical Distribution and Logistics Management*, 27:540–558, 1997.

T. Boyacı and S. Ray. Product differentiation and capacity cost interaction in time and price sensitive markets. *Manufacturing and Service Operations Management*, 5:18–36, 2003.

P. Brandimarte. Multi-item capacitated lot-sizing with demand uncertainty. *International Journal of Production Research*, 44:2997–3022, 2006.

P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation*. Springer-Verlag, NY, USA, 1987.

F. P. Buffa. Transit time and cost factors: their effects on inbound consolidation. *Transportation Journal*, Fall:50–63, 1987.

L. D. Burns, R. W. Hall, D. E. Blumenfeld, and C. F. Daganzo. Distribution strategies that minimize transportation and inventory costs. *Operations Research*, 33:469–490, 1985.

T. H. Burwell, S. D. Dinesh, K. E. Fitzpatrick, and M. R. Roy. Economic lot size model for price-dependent demand under quantity and freight discounts. *International Journal of Production Economics*, 48:141–155, 1997.

J. L. Cavinato. Pricing strategies for private trucking. *Journal of Business Logistics*, 3:72–84, 1982.

S. Çelik and C. Maglaras. Dynamic pricing and lead-time quotation for a multi-class make-to-order queue. *Management Science*, 54:1132–1146, 2008.

S. Çetinkaya. Coordination of inventory and shipment consolidation decisions: A review of premises, models, and justification. In J. Guenes, E. Akçalı, P. Pardalos, E. Romeijn, and M. Shen, editors, *Applications of Supply Chain Management and E-Commerce Research in Industry*, pages 3–51. Kluwer Series: Applied Optimization, vol 92, 2004.

S. Çetinkaya and J. H. Bookbinder. Stochastic models for the dispatch of consolidated shipments. *Transportation Research-B*, 37:747–768, 2003.

S. Çetinkaya, F. Mutlu, and C-Y. Lee. A comparison of outbound dispatch policies for integrated inventory and transportation decisions. *European Journal of Operational Research*, 171:1094–1112, 2006.

L. M. A. Chan, Z. J. M. Shen, D. Simchi-Levi, and J. L. Swann. Coordination of pricing and inventory decisions: A survey and classification. In D Simchi-Levi, S. D. Wu, and Z. J. M. Shen, editors, *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*, pages 335–392. Kluwer Academic Press, 2004.

F. Y. Chen, T. Wang, and T. Z. Xu. Integrated inventory replenishment and

temporal shipment consolidation: a comparison of quantity-based and time-based models. *Annals of Operations Research*, 135:197–210, 2005.

E. Çınlar. *Introduction to Stochastic Processes*. Prentice-Hall, NJ, USA, 1975.

R. K. Deb and R. F. Serfozo. Optimal control of batch service queues. *Advances in Applied Probability*, 5:340–361, 1973.

M. Dror and B. C. Hartman. Shipment consolidation: who pays for it and how much? *Management Science*, 53:78–87, 2007.

M. R. Garey and D. S. Johnson. *Computers and Intractibility: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York, USA, 1979.

Y. P. Gupta and P. K. Bagchi. Inbound freight consolidation under just-in-time procurement: application of clearing models. *Journal of Business Logistics*, 8: 74–94, 1987.

R. W. Hall. Consolidation strategy: inventory, vehicles and terminals. *Journal of Business Logistics*, 15:87–112, 1987.

R. W. Hall. Vehicle packing. *Transportation Research-B*, 23B:103–121, 1989.

D. P. Heyman. Optimal operating policies for m/g/1 queuing systems. *Operations Research*, 16:362–368, 1968.

D. P. Heyman. The t-policy for the m/g/1 queue. *Management Science*, 23:775–778, 1977.

J. K. Higginson. *Shipment Consolidation in Business Logistics Management, Ph.D. Thesis.* University of Waterloo, Waterloo, ON, 1992.

J. K. Higginson. Modeling shipper costs in physical distribution analysis. *Transportation Research-A*, 27:113–124, 1993.

J. K. Higginson. Recurrent deicison approaches to shipment-release timing in freaight consolidation. *International Journal of Physical Distribution and Logistics*, 25:3–23, 1995.

J. K. Higginson and J. H. Bookbinder. Policy recommendations for a shipment consolidation program. *Journal of Business Logistics*, 15:87–112, 1994.

J. K. Higginson and J. H. Bookbinder. Markovian decision processes in shipment consolidation. *Transportation Science*, 29:242–255, 1995.

A. V. Hill, J. M. Hays, and E. Naveh. A model for optimal delivery time guarantees. *Journal of Service Research*, 2:254–264, 2000.

G. C. Jackson. Evaluating order consolidation strategies using simulation. *Journal of Business Logistics*, 2:110–138, 1981.

G. C. Jackson. A survey of freight consolidation practices. *Journal of Business Logistics*, 6:13–34, 1985.

W. Jaruphongsa, S. Çetinkaya, and C-Y. Lee. Outbound shipment mode considerations for integrated inbentory and delivery lot-sizing decisions. *Operations Research Letters*, 35:813–822, 2007.

S. Kavadias and C. H. Loch. *Project Selection Under Uncertainty: Dynamically Allocating Resources to Maximize Value*. Kluwer Academic Publishers, Boston, MA, 2004.

P. Keskinocak and S. Tayur. Due date management policies. In D. Simchi-Levi, S. D. Wu, and Z. J. M. Shen, editors, *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*, pages 485–553. Kluwer Academic Press, 2004.

M. Khouja and S. K. Goyal. A review of joint replenishment problem: 1989-2005. *European Journal of Operational Research*, 186:1–16, 2008.

A. J. Kleywegt and J. D. Papastavrou. The dynamic and stochastic knapsack problem with random sized items. *Operations Research*, 49:26–41, 2001.

H. Kunreuther and J. F. Richards. Optimal pricing and inventory decisions for non-seasonal items. *Econometrica*, 39:173–175, 1995.

J. T. LaFrance. Linear demand functions in theory and practice. *Journal of Economic Theory*, 37:147–166, 1985.

P. Lederer and L. Li. Pricing, production scheduling, and delivery-time competition. *Operations Research*, 45:407–420, 1994.

P. Lederer and L. Li. Price and time competition for service delivery. *Manufacturing and Service Operations Management*, 2:392–409, 2000.

H. L. Lee and M. J. Rosenblatt. A generalized quantity discount pricing model to increase supplier's profits. *Management Science*, 32:1177–1185, 1986.

A. Lim, B. Rodgiues, and Y. Yang. 3-D container packing heuristics. *Applied Intelligence*, 22:125–134, 2005.

J. M. Masters. The effects of freight consolidation on customer service. *Journal of Business Logistics*, 2:55–74, 1980.

J. I. McGill and G. J. Van Ryzin. Revenue management: research overview and prospects. *Transportation Science*, 33:233–256, 1999.

H. Min and M. Cooper. A comparative review of analytical studies on freight consolidation and backhauling. *Logistics and Transportation Review*, 26:149–169, 1980.

M. J. Newbourne. *A Guide to Freight Consolidation for Shippers*. The Traffic Service Corporation, Washington, D.C., 1976.

M. J. Newbourne and C. Barrett. Freight consolidation and the shipper, parts 1-5. *Transportation and Distribution Management*, 22:Parts 1–5, 1972.

K. Palaka, S. Erlebacher, and D. H. Kropp. Lead-time setting, capacity utilization, and pricing decisions under lead-time dependent demand. *IIE Transactions*, 30: 151–163, 1998.

K. P. Papadaki and W. B. Powell. An adaptive dynamic programming algorithm for a stochastic multiproduct batch dispatch problem. *Naval Research Logistics*, 50:742–769, 2003.

J. D. Papastavrou, S. Rajagopalan, and A. J. Kleywegt. The dynamic and stochastic knapsack problem with deadlines. *Management Science*, 42:1706–1718, 1996.

P. M. Pekgün, P. Griffin and P. Keskinocak. Coordination of marketing and production for price and leadtime decisions. *IIE Transactions*, 40:12–30, 2008.

J. Pooley and A. J. Stenger. Modeling and evaluating shipment consolidation in a logistics system. *Journal of Business Logistics*, 13:153–174, 1992.

W. B. Powell. The bulk service queue with a general service control strategy:

theoretical analysis and a new computational procedure. *Operations Research*, 19:352–377, 1985.

W. B. Powell and P. Humblet. Analysis of vehicle holding and cancellation strategies in bulk arrival, bulk service queues. *Transportation Science*, 34:267–275, 1986.

F. J. Quinn. The payoff. *Logistics Management*, 36:37–41, 1997.

A. Rajan, R. Steinberg, and R. Steinberg. Dynamic pricing and ordering decisions by a monopolist. *Management Science*, 38:240–262, 1992.

U. S. Rao, J. M. Swaminathan, and J. Zhang. Demand and production management with uniform guaranteed lead time. *Production and Operations Management*, 14: 400–412, 2005.

S. Ray and J. H. Bookbinder. Logistics service management for price-and service-sensitive customers. In *Proceedings of the 10th International Symposium on Logistics*, pages 702–707, Lisbon, Portugal, 2005.

S. Ray and E. M. Jewkes. Customer lead time management when both demand and price are lead time sensitive. *European Journal of Operational Research*, 153: 769–781, 2004.

A. D. Schuster. The economics of shipment consolidation. *Journal of Business Logistics*, 1:22–35, 1979.

J. F. Shapiro. *Modeling the Supply Chain.* Duxbury Press, NY, USA, 2001.

Y. Sheffi. Carrier/shipper interactions in the transportation market: an analytical framework. *Journal of Business Logistics*, 27:1–27, 1986.

E. A. Silver, D. F. Pyke, and R. Peterson. *Inventory Management and Production Planning and Scheduling, 3rd ed.* J. Wiley and Sons, Ltd., NY, USA, 1998.

T. Skjott-Larsen, P. B. Schary, H. H. Mikkola, and H. Kotzab. *Managing the Global Supply Chain.* CBS Press, Liber, Copenhagen, Denmark, 2007.

K. C. So and J. S. Song. Price. delivery time guarantees and capacity selection. *European Journal of Operational Research*, 111:28–49, 1998.

M. J. Sobel. Optimal average-cost policy for a queue with start-up and shut-down costs. *Operations Research*, 17:145–162, 1969.

R. J. Tersine and S. Barman. Economic inventory/transport lot sizing with quantity and freight rate discounts. *Decision Sciences*, 22:1171–1179, 1997.

J. C. Tyan, F-K. Wang, and T. C. Du. An evaluation of freight consolidation policies in third party logistics. *Omega*, 31:55–62, 2003.

J. E. Tyworth, J. L. Cavinato, and C. J. Langley Jr. *Traffic management: Planning, operations and control.* Addison-Wesley, FIND, FIND, 1987.

M. A. Ülkü and J. H. Bookbinder. Policy analysis in shipment consolidation. In *Proceedings of the XXVI. IE/OR Conference*, pages 9–12, Kocaeli, Turkey, 2006.

T. L. Urban. Establishing delivery time guarantees. *European Journal of Operational Research*, 196:959–967, 2009.

Q. Wang and R. Wang. Quantity discount pricing policies for heterogeneous retailers with price sensitive demand. *Naval Research Logistics*, 52:645–658, 2005.

T. M. Whitin. Inventory control and price theory. *Management Science*, 2:61–68, 1955.

M. Yadin and P. Naor. Queuing systems with a removable service station. *Operational Research Quarterly*, 14:393–405, 1963.