Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Continuous Review Inventory Model with Dynamic Choice of Two Freight Modes with Fixed Costs

Aditya Jain, Harry Groenevelt, Nils Rudi,

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 12, No. 1, Winter 2010, pp. 120–139
ISSN 1523-4614 | EISSN 1526-5498 | 10 | 1201 | 0120

informs®
DOI 10.1287/msom.1090.0257

Continuous Review Inventory Model with Dynamic Choice of Two Freight Modes with Fixed Costs

Aditya Jain
Indian School of Business, Gachibowli, Hyderabad, 500032, India, aditya_jain@isb.edu

Harry Groenevelt
Simon School of Business, University of Rochester, Rochester, New York 14627, groenevelt@simon.rochester.edu

Nils Rudi
INSEAD, 77305 Fontainebleau, France, nils.rudi@insead.edu

We analyze a continuous review (Q, r) stochastic inventory model in which orders placed with a make-to-order manufacturer can be shipped via two alternative freight modes differing in lead time and costs. The costs of placing an order and using each freight mode consist of fixed components and hence exhibit economies of scale. We derive an optimal policy for using the two freight modes for shipping each order. This freight-mode decision is delayed until manufacturing is complete and the optimal policy uses information about the demand incurred in the meantime. Furthermore, given that the two freight modes are used optimally for shipping each order, we solve our model for reorder point and order quantity that minimizes cost. We analyze the cost savings achieved from postponing the freight-mode decision and provide analytical and numerical comparisons between the solutions to our two-freight model and single-freight models. Finally, we illustrate the properties of the solution to our model using an extensive set of numerical examples.

Key words: continuous review policy; lead times; freight mode selection; transportation economies of scale
History: Received: August 29, 2007; accepted: February 06, 2009. Published online in *Articles in Advance* June 12, 2009.

1. Introduction

In the last few decades, the manufacturing industry has witnessed an ever-growing trend of sourcing products from offshore locations. For a firm, one of the biggest challenges that accompanies global sourcing is managing timely shipment of inventory over a far-flung supply chain. The competitive edge attained by cost-effective sourcing can easily be eroded by the reduced ability to respond to surges and slumps in demand caused by long lead times, of which transportation time is a major part. In response, and aided by deregulation of the worldwide transportation industry, logistic service providers continue to develop innovative practices such as freight consolidation and multimodal shipping. This has resulted in an increased number of logistic services, often offered by the same provider, and typically differentiated by the trade-off between transportation time and cost. As transportation decisions are inherently linked with lead times and hence inventory-related costs, a firm needs to optimize them jointly with inventory decisions to gain the full advantage of low-cost sourcing. In this paper we analyze one such joint optimization for a firm that sources its product from a make-to-order supplier, can ship inventories using two alternative freight modes differing in lead time and costs, and faces fixed costs in placing orders as well as in shipping.

The practice of using two transportation modes for order fulfillment has not been uncommon; firms that use slower ocean freight for shipping on a regular basis resort to faster and more expensive air freight for emergencies. However, in recent years, under the increased pressure of meeting demand on time to stay competitive, several firms have increased their relative use of air freight (Bowman 1994, Sowinski 2004). The trend of combining different transportation modes is also evident in the variety of new multimodal services offered by logistic solution providers and the growing demand for time-critical expedited services (Krause 2000, Stanley 2004). Furthermore, there is now an increased emphasis on information technology (IT)-enabled logistics services that provide firms visibility across the supply chain and facilitate transportation decisions based on the latest information. For example, using the IT logistics solution provided by TradeBeam, clothing retailer Liz Claiborne Inc. can monitor its shipments in the pipeline and expedite their delivery online (Enslow 2006).

Several papers in the operations literature have addressed the issue of optimally managing inventory with two replenishment modes. Our work differs from this body of literature in two important aspects. First, most of these studies assume a simple quantity proportional cost of shipping, ignoring transportation economies of scale. Consequently, in the optimal solutions analyzed in these studies, orders are allowed to be split in small quantities. This is contrary to the observation made from real-life data by Thomas and Tyworth (2006) in their recent review paper. In contrast, we consider a model in which, in addition to the quantity proportional cost, the firm incurs a fixed cost on each use of both freight modes. Our paper thus incorporates previously ignored transportation economies of scale. Second, in our paper lead time is modeled to capture two phases—manufacturing and transportation. The freight-mode decision is made at the completion of manufacturing and is based on the latest demand information. Our model therefore captures the information-based nature of the transportation decision reflective of the evolving practices discussed earlier. This is in contrast to other models in the literature in which ordering and freight-mode decisions are made at the same time. The special case of our model with zero manufacturing lead time is equivalent to such a model.

The analysis of the optimal policy for placing orders in this problem is extremely complex, because of the challenges involved in handling order-crossing in a continuous review setting. We therefore limit our study to (Q, r) policies for placing orders. This is motivated by several factors. First, it is of a rather simple form and hence is implementable in real settings. From our experience with the implementation of scientific inventory models, managers trust and use only policies that are simple and transparent. Second, in our numerical investigations, we illustrate that our model can generate significant cost savings over using only one freight mode. Third, from our intuition, we expect it to be close to the actual optimal solution. In particular, it is optimal when the probability of order-crossing is negligible, and our extensive numerical investigation demonstrates this to be the case in a majority of settings.

After a brief literature review in §2, we present our model and derive the cost function in §3 assuming a (Q, r) policy. Next, we optimize our model for exogenous order size Q. In §4 we derive an optimal policy for shipping orders by the two freight modes for given values of Q and r; and in §5 we optimize the model over r. In §6 we address the question of determining the order quantity that minimizes the average cost. This problem is of independent interest, because often the order quantity cannot be varied arbitrarily. Section 7 discusses two special cases and numerical investigations to illustrate properties of our model. Finally, §8 concludes the paper. All proofs are provided in the appendix, and details of numerically solved examples can be found in the online supplement to the paper.

2. Literature Review

There is a vast body of literature on optimal inventory control when orders can be placed with two independent sources differing in lead times and costs. We refrain from providing a detailed survey of this body of literature and focus on the papers we deem most relevant. For detailed literature surveys, we refer readers to Minner (2003) and Thomas and Tyworth (2006) and the references therein.

Among the papers that consider multiple deliveries from a single supplier, closely related to our work are Groenevelt and Rudi (2003) and Jain et al. (2006). Both papers consider this problem in a periodic review setting with costs accounted for continuously in time—Groenevelt and Rudi (2003) in a base stock inventory policy setting and Jain et al. (2006) with fixed costs of ordering and using freight modes. Similarly, in a periodic review setting, Huggins and Olsen (2003) consider the problem of a manager who can use alternative sources such as production overtime and expedited delivery under the requirement of always meeting complete demand. Another notable work in the same direction is Lawson and Porteus (2000), who extend the classic multistage model of Clark and Scarf (1960) to incorporate the possibility of expediting units between two consecutive stages. The paper by Muharremoğlu and Tsitsiklis (2003) is a further generalization of Lawson and Porteus (2000). A few additional papers consider splitting an order from a single source into two shipments. In a continuous review setting, Moinzadeh and Lee (1989) consider a (Q, r) inventory model in which each order is randomly split and delivered in two shipments with different lead times (in our model the "split" for each order is a decision). The continuous review model of Chiang and Chiang (1996) illustrates reduction in inventory-holding costs resulting from splitting each order into multiple deliveries without any additional cost. Some recent studies analyze continuous review inventory models with the possibility of expediting complete shipments: Bookbinder and Çakanyildirim (1999) consider deterministic demand and stochastic lead time; Gallego et al. (2007) consider stochastic demand and deterministic lead times; and Çakanyildirim and Luo (2005) consider delayed expediting decisions that use the latest demand information.

The earliest inventory models with two supply sources that differ in lead times and costs are the periodic review models of Barankin (1961), Daniel (1962), Neuts (1964), and Fukuda (1964), and they show that the optimal inventory policy is of order-up-to nature. The study by Whittemore and Saunders (1977) is among the first few papers that allow the lead time of two supply sources to differ by an arbitrary number of review periods; the optimal policy is a complex function of the vector of outstanding order amounts scheduled to arrive between the faster and slower shipment for the current order. This policy is computationally cumbersome and too complex to implement in practical cases. Notable studies of the problem in the continuous review setting are Moinzadeh and Nahmias (1988) who incorporate fixed costs and propose a heuristic (Q, r) policy for placing orders with each source, and Moinzadeh and Schmidt (1991) who propose a heuristic consisting of an $(S, S-1)$ policy for the slower supply source and a complex policy based on the ages of various outstanding orders for the faster source. More recent studies focus on periodic review models: Tagaras and Vlachos (2001) propose a heuristic based on the age of the outstanding order; Veeraraghavan and Scheller-Wolf (2008) propose a "dual-index policy" that keeps track of two separate indices for placing orders with two different suppliers and establish that it is a "near optimal" heuristic; and Sheopuri et al. (2007) provide a further generalization of the dual-index policy.

Methodologically, our work is closely related to Axsäter (1990) and Zheng (1992). In his influential paper, Axsäter (1990) provides a novel perspective on traditional inventory management problems. This perspective is analyzed in detail by Muharremoğlu and Tsitsiklis (2008), who refer to it as the "single-unit decomposition approach." In this paper, we use the single-unit decomposition approach to derive an expression for the average cost of our model. Zheng (1992) offers an insightful geometric representation of the solution to a continuous review (Q, r) inventory model with a single lead time. We extend his approach to our more complex setting; the insights offered from the geometry of the solution are even more pronounced, as they shed light not only on the optimal reorder point and order quantity but also on the optimal use of freight modes.

3. Model and Cost Function

We consider a stochastic inventory system with a continuous review (Q, r) inventory policy. Define inventory level as on-hand inventory minus back orders and inventory position as the sum of inventory level and outstanding orders. The inventory position is continuously monitored, and as soon as it drops to the reorder point r, an order of size Q is placed with a make-to-order supplier. Placing a manufacturing order incurs fixed cost K_1 and triggers production at the manufacturing facility. In L_1 time units (manufacturing lead time), the order is completely manufactured and ready to be shipped. At that time, two freight modes are available for shipping the order: (i) regular freight, which has transportation lead time L_2 (regular freight lead time) and incurs fixed cost K_2 per shipment, and (ii) express freight, which has transportation lead time l_2 (express freight lead time, $l_2 < L_2$) and, in addition to a fixed cost k_2 per shipment, incurs variable cost c_2 per unit shipped. An order can be shipped completely by either of the freight modes or can be split across the two in any proportion. We denote $L = L_1 + L_2$ and $l = L_1 + l_2$, the total lead times for a unit shipped by regular and express freight modes, respectively. The system incurs holding cost h per time unit on each unit of physical inventory. The demand arriving to the system during stock-out is back ordered, and each unit in back order incurs penalty cost p per time unit.

We model the demand as a nondecreasing stochastic process with stationary and independent increments. In deriving an expression for average cost, we assume that demand takes place one unit at a time; i.e., it is a Poisson process. Subsequently, to simplify analysis, we approximate demand as a continuous process with the aforementioned properties. $D_{(t_1, t_2]}$ denotes cumulative stochastic demand incurred in the interval $(t_1, t_2]$, and $F_{(t_1, t_2]}(\cdot)$ denotes its cumulative density function. The infinitesimal mean and variance of the demand process are given by μ and σ^2, respectively. $\mathbb{E}$ denotes the expectation operator, and $\mathbb{P}(\omega)$ and $\mathbb{1}_{\{\omega\}}$ denote the probability and indicator functions, respectively, of an event ω.

Let $q(x)$ be the number of units shipped via express freight, if a realized manufacturing lead-time demand is $D_{(0, L_1]} = x$. The remaining $Q - q(x)$ units are shipped via regular freight. The function $q(\cdot)$ then defines a dual-freight policy. When $q(x) = 0$, $\forall x$ and $q(x) = Q$, $\forall x$, then our model simplifies to a single-freight model with regular and express freights, respectively.

To obtain an expression for the expected cost per time unit of our model with an arbitrary dual-freight policy $q(\cdot)$, we use the *single-unit decomposition approach* (see Axsäter 1990, Muharremoğlu and Tsitsiklis 2008 for the details of this approach), which requires the following assumption.

Assumption 1. *Product units are allocated to demand in the sequence in which they are ordered.*

Assumption 1 implies that even if for some reason product units are delivered out of the order sequence (i.e., orders cross), they are provided to customers in sequence anyway. Hence, under Assumption 1 it is possible that at a given time finished goods inventory and back orders coexist. In practice, one would reassign products to reduce both holding and penalty costs, so the cost expression derived below using the single-unit decomposition approach is an upper bound to the true cost. When order crossings are rare, this upper bound will be tight (we discuss this in detail toward the end of this section).

Consider our model as a two-stage inventory system. Stage 1 is the location of the inventory manager where the stochastic demand for single units of product arrives at a long-term average rate μ; Stage 2 is the manufacturing facility. A "virtual" product unit enters the system at Stage 2, when an order is placed for it by the inventory manager at Stage 1. After spending L_1 time at Stage 2, during which it becomes an "actual" product unit, the product unit is ready to be shipped to Stage 1. Product units and demand are matched in the sequence in which they arrive. A product unit satisfies the matched demand as soon as the product unit and the demand both have reached Stage 1. At that point, the product unit and the demand leave the system. A product unit incurs holding cost h for each time unit it waits for its matched demand at Stage 1, and it incurs a back order penalty cost p for each time unit its matched demand waits for it at Stage 1.

For $v > 0$, define t_v as the random time it takes for v units of demand to arrive. For $v \leq 0$, let $\mathbb{P}(t_v = 0) = 1$. Define the index of a product unit at an instant as 1 *plus* the total number of product units present in the system that entered the system before the unit in question *minus* the number of demands waiting to be satisfied at Stage 1. The index of a product unit decreases as the system experiences demand, although the product unit may physically stay at the same location, and product units meet their demand in the sequence of their indices at Stage 1. The index of a product unit at an instant is an indicator of the total time for which either the product unit waits for its matched demand or the matched demand waits for the product unit at Stage 1. Define $g(y, \mathscr{L})$ as the future expected cost of a product unit indexed y, which at the current instant is getting shipped from Stage 2 to Stage 1 with deterministic lead time $\mathscr{L}$. If $y < 0$, then the demand matched to the product unit has already arrived in the system at Stage 1 and will wait for the product unit for additional time $\mathscr{L}$. If $y \geq 0$, then the demand matched to the product unit arrives at Stage 1 after a random time t_y; and if $t_y > \mathscr{L}$, then the product unit waits for the demand at Stage 1 for $(t_y - \mathscr{L})$ time. Otherwise, the demand waits for the product unit for $(\mathscr{L} - t_y)$ time. Assigning holding and back order penalty costs to these waiting times, we get

$$g(y, \mathscr{L}) = \begin{cases} p\mathscr{L} & \text{if } y < 0, \\ \mathbb{E}(h(t_y - \mathscr{L})^+ + p(\mathscr{L} - t_y)^+) & \text{if } y \geq 0. \end{cases} \quad (1)$$

Note that at any point in time, the highest index of a product unit in the system is the sum of inventory level and outstanding orders; i.e., it is the same as inventory position. Thus, with a (Q, r) policy, as soon as the highest index of a product unit in the system reduces to r, an order of size Q is placed, and a batch consisting of Q product units with indices $r+1$ to $r+Q$ enters the system.

In our model with two freight modes, consider a randomly chosen product unit that enters the system at index y. Clearly, the random number y is uniformly distributed over $\{r+1, \dots, r+Q\}$. When the product unit enters the system at index y, the demand matched to it has on average waited for $\mathbb{E}t_{(-y)}$ time at Stage 1. The product unit then spends L_1 time at Stage 2, during which its matched demand spends $(L_1 - t_y)^+$ time at Stage 1. The product unit thus incurs expected cost $p\mathbb{E}t_{(-y)} + p\mathbb{E}(L_1 - t_y)^+$ before it is shipped. At the time of shipment, the index of the product unit is $y - D_{(0, L_1]}$. If q units are to be shipped via express freight (which may depend on $D_{(0, L_1]}$), then the q product units with the smallest indices in a batch are assigned to express freight, as the demands for these product units arrive earlier than those for the rest. Thus, for $y \in \{r+1, \dots, r+q\}$, the product unit is shipped by express freight and for $y \in \{r+q+1, \dots, r+Q\}$ by regular freight. The expected cost (conditional on $D_{(0, L_1]}$) incurred on the product unit after shipping is $c_2 + g(y - D_{(0, L_1]}, l_2)$, if $y \le r+q$ and $g(y - D_{(0, L_1]}, l_2)$ otherwise. Hence, for a randomly selected unit, the expected inventory-related and variable express freight cost incurred is given by

$$\frac{1}{Q}\left(p\sum_{y=r+1}^{r+Q}(\mathbb{E}t_{(-y)}+\mathbb{E}(L_1-t_y)^+)+\sum_{y=r+1}^{r+q}(c_2+g(y-D_{(0,L_1]},l_2)) + \sum_{y=r+q+1}^{r+Q} g(y - D_{(0, L_1]}, L_2)\right). \quad (2)$$

The following fixed costs are incurred on a batch: K_1 of placing an order; K_2 of using regular freight if $Q - q > 0$; and k_2 of using express freight if $q > 0$. Amortizing these fixed costs over Q units, adding the resulting per unit cost to (2), and taking expectation over $D_{(0, L_1]}$ gives the total expected cost per product unit. Multiplying the resulting expression by the demand rate μ then leads to the following expression for $C_q(Q, r)$, the expected cost per time unit with an arbitrary dual-freight policy defined by $q(\cdot)$:

$$C_q(Q, r) = \frac{K_1 + p\sum_{y=r+1}^{r+Q}\{\mathbb{E}t_{(-y)} + \mathbb{E}(L_1 - t_y)^+\} + \mathbb{E}\Gamma(Q, r, q(D_{(0, L_1]}), D_{(0, L_1]})}{Q/\mu}, \quad (3)$$

where $\Gamma(Q, r, q(D_{(0, L_1]}), D_{(0, L_1]})$ is the expected cost conditional on $D_{(0, L_1]}$, incurred on a batch after shipping, if $q(D_{(0, L_1]})$ units are shipped using express freight. Its expression is

$$\Gamma(Q, r, q(D_{(0, L_1]}), D_{(0, L_1]}) = K_2\mathbb{1}_{\{q<Q\}} + k_2\mathbb{1}_{\{q>0\}} + c_2 q + \sum_{y=r+1}^{r+q} g(y - D_{(0, L_1]}, l_2) + \sum_{y=r+q+1}^{r+Q} g(y - D_{(0, L_1]}, L_2). \quad (4)$$

For a random variable d, define the cost-rate function $G(y, d) \overset{\text{def}}{=} \mathbb{E}(h(y-d)^+ + p(d-y)^+)$, where the expectation is taken only over the second argument d. Our next proposition simplifies the expression in (3) and states it in terms of cost-rate function $G(\cdot, \cdot)$.

Proposition 1. *The average cost of a (Q, r) policy with a dual-freight policy $q(\cdot) \in [0, Q]$ is*

$$C_q(Q, r) = \frac{\mathbb{E}\{\mu K(Q, q(D_{(0, L_1]})) + S(Q, r, q(D_{(0, L_1]}), D_{(0, L_1]})\}}{Q}, \quad (5)$$

where

$$K(Q, q) = K_1 + K_2\mathbb{1}_{\{q<Q\}} + k_2\mathbb{1}_{\{q>0\}}, \quad \text{and} \quad (6)$$

$$S(Q, r, q, x) = \sum_{y=r+1}^{r+q}(\mu c_2 + G(y - x, D_{(L_1, l]})) + \sum_{y=r+q+1}^{r+Q} G(y - x, D_{(L_1, L]}). \quad (7)$$

The numerator in the expression for $C_q(Q, r)$ *divided by* the demand rate μ is the expected aggregate cost incurred on each order, and it consists of *aggregate fixed cost* $K(Q, q)$, and *aggregate variable cost* $S(Q, r, q, x)/\mu$. Multiplying it by the ordering frequency μ/Q gives the expected cost per unit time. In the rest of the paper, we treat Q, r, and $q(\cdot)$ as continuous variables and use the following continuous approximation of $S(Q, r, q, x)$:

$$S(Q, r, q, x) = \int_{r}^{r+q} (\mu c_2 + G(y-x, D_{(L_1, l]}))\,dy + \int_{r+q}^{r+Q} G(y-x, D_{(L_1, L]})\,dy. \qquad (8)$$

This is analogous to the continuous demand approximation commonly used in the analysis of continuous review (Q, r) inventory models with a single freight (see Zheng 1992, Gallego 1998).

Remark 1. The cost function in Proposition 1 assumes that the inventory holding cost is incurred after the units have been received. Let h' $(<h)$ denote the cost incurred on each unit in transit per time unit. Then, on a unit shipped via regular freight, the transit inventory holding cost is $h'L_2$ and on a unit shipped by express freight is $h'l_2$. Clearly, on each unit shipped by express freight, transit inventory holding cost $h'(L_2 - l_2)$ is saved. We can accommodate this by adding $h'L_2$ to the cost per unit of product purchased and adjusting the variable shipping cost to $c_2' = c_2 - h'(L_2 - l_2)$.

Remark 2. The cost function in expression (5) reduces to the cost function for the model with a single freight when $q(x) = 0\ \forall x$ and when $q(x) = Q\ \forall x$. These expressions are the same as given by Equation (2) in Zheng (1992) with appropriate lead times and ordering costs.

A final issue worth discussing is to what extent (Q, r) policies are optimal for the problem we consider. If there were just a single freight mode in our model, there would exist an optimal stationary (Q, r) policy, because demands take place one unit at a time and orders cannot cross. For our model with two freight modes, this result remains true as long as we assume that orders do not cross. The argument for this is as follows: Consider the problem as a Markov decision problem. Costs are assigned as follows. When the action is not to place an order, the costs assigned are the expected holding and back order costs $L_1 + L_2$ time units into the future. Note that this depends only on the current inventory position. When an order is placed, the costs assigned are these same expected holding and back order costs $L_1 + L_2$ time units into the future (which now depend only on the current inventory position plus the order size) plus the fixed cost of ordering plus the expected fixed and variable costs of shipping minus the expected holding and penalty cost savings achieved by possibly shipping part of the order by express freight, assuming that the freight-mode decision is made optimally. Note that these costs depend only on the current inventory position and the order size, because the assumption of no order crossing implies that by $L_1 + l_2$ time units into the future all prior orders have been delivered. Hence, the inventory position constitutes the only relevant information for placing orders in the system. The decision epochs coincide with the demand epochs (because the state of the system remains unchanged between two consecutive demand epochs). Because demands take place one unit at a time and it is optimal to take the same decision every time a given state is entered, the optimal policy is of the form (Q, r).

Two orders cross when the following conditions are met simultaneously: (i) The second order is placed within $L_2 - l_2$ time interval after the first order, and (ii) the first order has some units shipped by regular freight and some of the second order units by express freight. The former condition provides an upper bound $\mathbb{P}(D_{(0, L_2 - l_2]} > Q)$ on the probability of order crossing. The latter implies that the actual probability of order crossing becomes larger (closer to the upper bound $\mathbb{P}(D_{(0, L_2 - l_2]} > Q)$) when both freight modes are used widely, i.e., when the dual-freight modes are of value in terms of cost reduction. In §7.3 of this paper and §A.1 of the online supplement, we investigate the upper bound on the probability of order crossing $\mathbb{P}(D_{(0, L_2 - l_2]} > Q)$, along with the value of dual-freight modes for the solution of our model. This investigation finds that for a large set of parameter values, the upper bound on the probability of order crossing $\mathbb{P}(D_{(0, L_2 - l_2]} > Q)$ is negligible. Moreover, the investigation suggests that for several cases where this probability is indeed significant, the value of dual-freight modes is negligible and the solution corresponds to dominant use of one of the freight modes. The investigation also sheds light on conditions under which a (Q, r) policy may not be reasonable for our model.

4. Optimal Dual-Freight Policy

This section derives the optimal dual-freight policy $q^*(\cdot)$ for given values of Q and r and characterizes the cost function, given that the optimal policy is used for shipping each order. Define $C(Q, r)$ as the expected cost per time unit of the model with the optimal dual-freight policy. Using (5),

$$C(Q, r) = \frac{\mathbb{E}[\min_{0\le q\le Q}\{\mu K(Q, q) + S(Q, r, q, D_{(0, L_1]})\}]}{Q}. \quad (9)$$

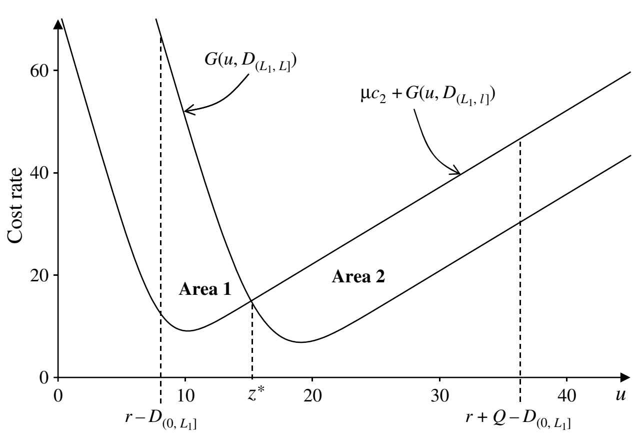
The optimal dual-freight policy is the solution q^* to the minimization problem inside the square brackets as a function of the realized value of $D_{(0, L_1]}$, given Q and r. To facilitate insights into the solution, we first consider the special case $K_2 = k_2 = 0$. This result is similar to the solution of Groenevelt and Rudi (2003) for the use of express freight in the periodic review setting.

Lemma 1. *If $K_2 = k_2 = 0$, then the optimal express-freight quantity q^* is characterized by: If $c_2 \ge p(L_2 - l_2)$, then $q^* = 0$; otherwise, $q^* = \min\{Q, (z^* - r + D_{(0, L_1]})^+\}$, where z^* solves*

$$\mu c_2 + G(z^*, D_{(L_1, l]}) = G(z^*, D_{(L_1, L]}). \quad (10)$$

The optimal dual-freight policy in the absence of fixed costs of freight modes recommends shipping up to z^* with express freight, if achievable. We geometrically illustrate the optimality of this policy in Figure 1, where $\mu c_2 + G(u, D_{(L_1, l]})$ and $G(u, D_{(L_1, L]})$ are plotted functions of u. In this plot, the aggregate variable cost $S(Q, r, q, D_{(0, L_1]})$ is the sum of the area under

Figure 1 $\mu c_2 + G(u, D_{(L_1, l]})$ **and** $G(u, D_{(L_1, L]})$ **Plotted as Functions of** u, **for** $c_2 < p(L_2 - l_2)$



$\mu c_2 + G(u, D_{(L_1, l]})$ from $u = r - D_{(0, L_1]}$ to $u = r + q - D_{(0, L_1]}$ and the area under $G(u, D_{(L_1, L]})$ from $u = r + q - D_{(0, L_1]}$ to $u = r + Q - D_{(0, L_1]}$. This sum is minimized when $r + q - D_{(0, L_1]}$ is the point z^* at which the two functions intersect. However, if $c_2 \ge p(L_2 - l_2)$, then $G(u, D_{(L_1, L]})$ lies below $\mu c_2 + G(u, D_{(L_1, l]})$ for all values of u, and hence it is optimal to use regular freight exclusively.

Remark 3. Note that $K(Q, 0) = K_1 + K_2$, $K(Q, Q) = K_1 + k_2$, and for $0 < q < Q$, $K(Q, q) = K_1 + K_2 + k_2$. Thus, if the optimal value of q (for the problem with nonzero K_2 and k_2) is such that $0 < q < Q$, then it minimizes $S(Q, r, q, D_{(0, L_1]})$ and is given by Lemma 1.

Remark 4. For $c_2 \ge p(L_2 - l_2)$, $S(Q, r, q, D_{(0, L_1]})$ is minimized at point $q = 0$. This, in conjunction with Remark 3, implies that for this case 0 and Q are the only candidates for the optimal value of q. Additionally, if $K_2 \le k_2$, then $q = Q$ is also eliminated as a potential optimum.

In managerial terms, Remark 3 states that if it is optimal to split an order across the two freight modes for shipping, then it is optimal to ship up to z^* with express freight. Remark 4, in contrast, states that if the maximum potential benefit of ordering a unit via express freight, namely $p(L_2 - l_2)$, is less than its marginal cost c_2, then it is never optimal to split an order.

The next lemma is useful in deriving the optimal dual-freight policy and the subsequent analysis.

Lemma 2. (a) *$G(u, D_{(0, L_2]}) - \mu c_2 - G(u, D_{(0, l_2]})$ is nonincreasing in u,*

(b) *$\int_r^{r+q}(G(u, D_{(0, L_2]}) - \mu c_2 - G(u, D_{(0, l_2]}))\,du$ is nonincreasing in r, and*

(c) $-(h(L_2 - l_2) + c_2)\mu q \le \int_r^{r+q}(G(u, D_{(0, L_2]}) - \mu c_2 - G(u, D_{(0, l_2]}))\,du \le (p(L_2 - l_2) - c_2)\,\mu q$.

The following proposition states the optimal dual-freight policy with a general fixed-cost structure.

Proposition 2. *Define*

$$Q_e = \frac{k_2 - K_2}{p(L_2 - l_2) - c_2} \quad \text{and} \quad Q_r = \frac{K_2 - k_2}{h(L_2 - l_2) + c_2}. \quad (11)$$

Define $z_p(Q)$ as the unique solution to

$$\int_{z_p(Q)}^{z_p(Q)+Q} (G(u, D_{(L_1, L]}) - \mu c_2 - G(u, D_{(L_1, l]}))\,du = \mu(k_2 - K_2); \quad (12)$$

define $\underline{z}$ as the unique solution to

$$\int_{\underline{z}}^{z^*} (G(u, D_{(L_1, L]}) - \mu c_2 - G(u, D_{(L_1, l]}))\, du = \mu k_2, \qquad \underline{z} \le z^*; \quad (13)$$

and define $\bar{z}$ as the unique solution to

$$\int_{z^*}^{\bar{z}} (\mu c_2 + G(u, D_{(L_1, l]}) - G(u, D_{(L_1, L]}))\, du = \mu K_2, \qquad \bar{z} \ge z^*. \quad (14)$$

For any r and Q, a value of the optimal express-freight quantity q^* that solves the minimization problem in (9) is given in Table 1 for different combinations of cost parameters.

Proposition 2 can be explained with the help of Figure 1. When $c_2 \ge p(L - l)$, there is no marginal benefit to ship a unit via express freight, and only one or the other freight mode is to be used for shipping the complete order depending on cost differences. For the case $c_2 < p(L_2 - l_2)$, plotted in Figure 1, the aggregate variable cost savings earned in shipping an order by optimally splitting it over shipping it by regular freight is given by "Area 1," which decreases with $D_{(0, L_1]}$ and equals μk_2 for $D_{(0, L_1]} = r - \underline{z}$.

Table 1 Optimal Dual-Freight Policy

Cases	$c_2 < p(L - l)$	$c_2 \ge p(L - l)$

$K_2 \le k_2$	If $Q \le Q_e$, then $q^* = 0$.
	If $Q_e < Q \le \bar{z} - \underline{z}$, then
	$\begin{cases} q^* = 0, & \text{if } D_{(0, L_1]} \le r - z_p(Q); \\ q^* = Q, & \text{otherwise.} \end{cases}$
	If $\bar{z} - \underline{z} < Q$, then
	$\begin{cases} q^* = 0, & \text{if } D_{(0, L_1]} \le r - \underline{z}; \\ q^* = Q, & \text{if } D_{(0, L_1]} \ge r + Q - \bar{z}; \\ q^* = z^* - r + D_{(0, L_1]}, & \text{otherwise.} \end{cases}$
		$q^* = 0$

$K_2 > k_2$	If $y - x \le Q_r$, then $q^* = Q$.
	If $Q_r < Q \le \bar{z} - \underline{z}$, then
	$\begin{cases} q^* = 0, & \text{if } D_{(0, L_1]} \le r - z_p(Q); \\ q^* = Q, & \text{otherwise.} \end{cases}$
	If $\bar{z} - \underline{z} < Q$, then
	$\begin{cases} q^* = 0, & \text{if } D_{(0, L_1]} \le r - \underline{z}; \\ q^* = Q, & \text{if } D_{(0, L_1]} \ge r + Q - \bar{z}; \\ q^* = z^* - r + D_{(0, L_1]}, & \text{otherwise.} \end{cases}$
		If $Q \le Q_r$, then $q^* = Q$.
		If $Q_r < Q < Q_e$, then
		$\begin{cases} q^* = 0, & \text{if } D_{(0, L_1]} \le r - z_p(Q); \\ q^* = Q, & \text{otherwise.} \end{cases}$
		If $Q_e \le Q$, then $q^* = 0$.

Similarly, "Area 2," which represents the aggregate variable cost savings earned in shipping an order by optimally splitting it over shipping it by express freight, decreases with $D_{(0, L_1]}$ and is equal to μK_2 for $D_{(0, L_1]} = r + Q - \bar{z}$. Because splitting an order incurs additional aggregate fixed costs μk_2 and μK_2, respectively, over use of only regular freight and only express freight, it is optimal if and only if Area 1 is at least as large as μk_2 and Area 2 is at least as large as μK_2. These two conditions can hold simultaneously only for $Q > \bar{z} - \underline{z}$, i.e., for orders smaller than $\bar{z} - \underline{z}$; the savings in inventory and variable shipping costs earned by splitting an order are never large enough to offset the higher fixed cost incurred. The aggregate variable cost savings earned in shipping by express freight over by regular freight is Area 1 *minus* Area 2, whereas the additional aggregate fixed cost incurred in doing so is $\mu(k_2 - K_2)$. Area 1 *minus* Area 2 is nondecreasing in $D_{(0, L_1]}$ and is equal to $\mu(k_2 - K_2)$ at $D_{(0, L_1]} = r - z_p(Q)$. Thus, for $D_{(0, L_1]} < r - z_p(Q)$, shipping by regular freight is optimal; for $D_{(0, L_1]} > r - z_p(Q)$, shipping by express freight is optimal. For smaller order sizes, when $K_2 \le k_2$ ($K_2 > k_2$), Q_e (Q_r) is the minimum order quantity needed to justify shipping an order by express freight (regular freight), which has a higher fixed cost. And for $K_2 > k_2$ and $c_2 \ge p(L - l)$, Q_e is the minimum order quantity needed to justify use of only regular freight, which has a higher fixed but a smaller per unit cost.

The optimal dual-freight policy derived in Proposition 2 is a significant departure from results in the existing literature (such as Lemma 1). First and foremost, it accommodates all four possible ways of combining the two freight modes: (i) q^* dynamically takes values 0, $z^* - r + D_{(0, L_1]}$, and Q; (ii) q^* dynamically takes values 0 and Q; (iii) $q^* = 0$; and (iv) $q^* = Q$. Of these, one is chosen depending on the order size and model parameters. It is also the first to propose that splitting of orders be restricted only to orders larger than a threshold value. Finally, it does not prescribe shipment of arbitrarily small quantities via any of the freight modes. Specifically the minimum quantity limit for regular freight is $\min\{\bar{z} - z^*, Q\}$ and for express freight is $\min\{z^* - \underline{z}, Q\}$.

Next, we formulate an expression for $C(Q, r)$ (defined in (9)) that encompasses all the cases of the optimal dual-freight policy in Proposition 2. This approach enables us to carry out an integrative analysis (instead of a case-by-case one) to determine values for r and Q that minimize $C(Q, r)$. Furthermore, it translates to a geometric representation that facilitates a better understanding of the optimal solution. To this end we define the following:

$$\tilde{G}(y, z) \stackrel{\text{def}}{=} \int_{-\infty}^{y-z} G(y-x, D_{(L_1, L]})\, dF_{(0, L_1]}(x) + \int_{y-z}^{\infty} (\mu c_2 + G(y-x, D_{(L_1, l]}))\, dF_{(0, L_1]}(x). \quad (15)$$

Recalling the single-unit decomposition approach described in §3, $\tilde{G}(y, z)/\mu$ is the expected cost incurred on a product unit that enters the system at index y and is shipped by express freight if $D_{(0, L_1]} > y - z$ and by regular freight otherwise. Note that as $z \downarrow -\infty$, $\tilde{G}(y, z)$ becomes $G(y, D_{(0, L]})$, the cost rate function with only regular freight. As $z \uparrow \infty$, $\tilde{G}(y, z)$ becomes $\mu c_2 + G(y, D_{(0, l]})$, the cost rate function with only express freight.

Definition 1. Define functions $z_1(\cdot)$ and $z_2(\cdot)$ such that with the optimal dual-freight policy an order is shipped by regular freight if $D_{(0, L_1]} < r - z_1(Q)$; it is shipped by express freight if $D_{(0, L_1]} > r + Q - z_2(Q)$; and it is otherwise optimally split across the two freight modes.

In the rest of the paper, we suppress the argument Q of $z_1(\cdot)$ and $z_2(\cdot)$ unless it is ambiguous.

Remark 5. It follows from Definition 1 and Proposition 2 that (i) when q^* takes values from 0, $z^* - r + D_{(0, L_1]}$, and Q, then $z_1(Q) = \underline{z}$ and $z_2(Q) = \bar{z}$; (ii) when q^* takes values from 0 and Q, then $z_1(Q) = z_p(Q)$ and $z_2(Q) = z_p(Q) + Q$; (iii) when $q^* = 0$, then $z_1(Q) = z_2(Q) \to -\infty$; and (iv) when $q^* = Q$, $z_1(Q) = z_2(Q) \to \infty$.

Remark 6. In each cases except the one in the upper-right quadrant of Table 1, the function $z_p(Q)$ is well defined at points inside a convex set. It follows from the implicit function theorem on (12) that inside such a set, $z_p(Q)$ is continuous and differentiable. Furthermore, it has the limiting values shown in Table 2.

The next lemma provides a unified expression for the average cost $C(Q, r)$ and its properties.

Lemma 3. (a) *An expression for the average cost with the optimal dual-freight policy is*

$$C(Q, r) = \frac{\mu \mathcal{K}(Q, r) + \int_r^{r+Q} G_d(y \mid Q, r)\, dy}{Q}, \quad (16)$$

Table 2 Limiting Values of $z_p(Q)$

Cases	$c_2 < p(L-l)$	$c_2 \geq p(L-l)$
$K_2 \leq k_2$	$\lim_{Q \downarrow Q_e} z_p(Q) = -\infty$ $\lim_{Q \uparrow \bar{z} - \underline{z}} z_p(Q) = z_p(\bar{z} - \underline{z}) = \underline{z}$	
$K_2 > k_2$	$\lim_{Q \downarrow Q_r} z_p(Q) = \infty$ $\lim_{Q \uparrow \bar{z} - \underline{z}} z_p(Q) = z_p(\bar{z} - \underline{z}) = \underline{z}$	$\lim_{Q \downarrow Q_r} z_p(Q) = \infty$ $\lim_{Q \uparrow Q_e} z_p(Q) = -\infty$

where $\mathcal{K}(Q, r) = K_1 + K_2 F_{(0, L_1]}(r + Q - z_2) + k_2 \cdot (1 - F_{(0, L_1]}(r - z_1))$, *and* $G_d(y \mid Q, r)$ *is a continuous and differentiable function of* y *given by the following: When* $c_2 < p(L_2 - l_2)$, *then*

$$G_d(y \mid Q, r) = \begin{cases} \tilde{G}(y, y - r + z_1), & \text{if } y \leq r + (z^* - \underline{z}), \\ \tilde{G}(y, y - r - Q + z_2), & \text{if } y \geq r + Q - (\bar{z} - z^*), \\ \tilde{G}(y, z^*), & \text{otherwise}, \end{cases}$$

and $G_d(y \mid Q, r) = \tilde{G}(y, y - r + z_1)$, *otherwise.*

(b) $C(Q, r)$ *is continuous and differentiable in* Q *and* r.

In this lemma, $G_d(y \mid Q, r)/\mu$ is the sum of expected inventory holding, back order penalty, and variable express freight costs incurred on a product unit that enters the system at index $y \in (r, r + Q]$ and is shipped according to the optimal dual-freight policy. Let $\mathcal{S}(Q, r) \stackrel{\text{def}}{=} \int_r^{r+Q} G_d(y \mid Q, r)\, dy$; then $\mathcal{S}(Q, r)/\mu$ is the expected aggregate variable cost, with the optimal dual-freight policy. The total fixed cost incurred on an order of size Q is $K_1 + K_2$ if it is shipped by regular freight (when $D_{(0, L_1]} < r - z_1$); $K_1 + k_2$ if it is shipped by express freight (when $D_{(0, L_1]} > r + Q - z_2$); and $K_1 + K_2 + k_2$ if it is split across the two freight modes for shipping (when $r - z_1 < D_{(0, L_1]} < r + Q - z_2$). $\mathcal{K}(Q, r)$ is thus the expected aggregate fixed cost incurred on an order shipped by the optimal dual-freight policy. In comparing the cost function $C(Q, r)$ in (16) with the cost function of a single-freight model (for example Equation (1) in Zheng 1992), we note that the expected aggregate fixed cost $\mathcal{K}(Q, r)$ is the dual-freight model counterpart of fixed cost K, and the dual-freight cost rate function $G_d(y \mid Q, r)$ is the counterpart of $G(y)$. In contrast to their single-freight model counterparts, $\mathcal{K}(Q, r)$ and $G_d(y \mid Q, r)$ are functions of Q and r.

5. Reorder Point

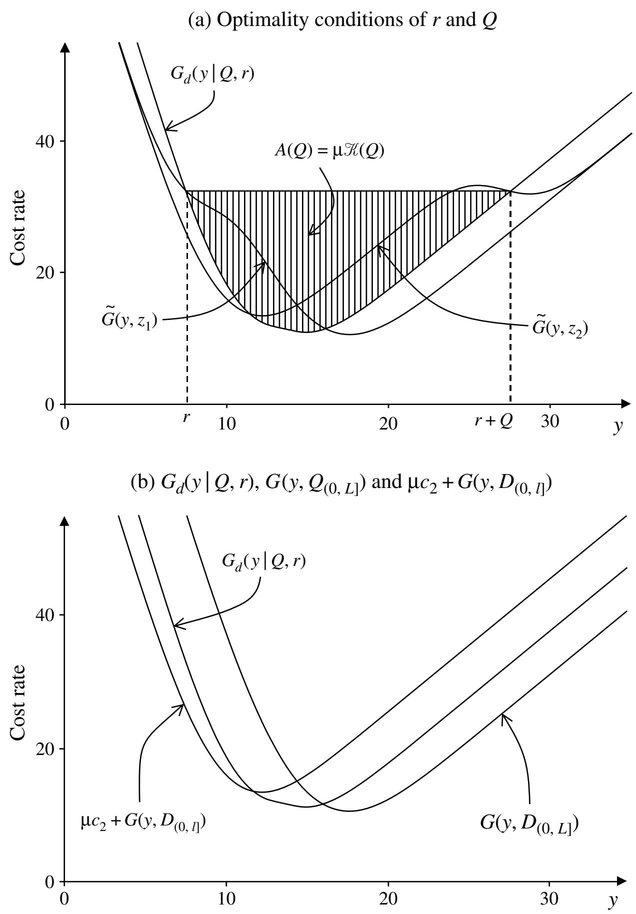
Let $r(Q) \stackrel{\text{def}}{=} \arg\min_r C(Q, r)$ denote the reorder-point solution for a given order size when each order is shipped with the optimal use of the two freight modes.

Proposition 3. *$r(Q)$ is a solution to*

$$G_d(r \mid Q, r) = G_d(r+Q \mid Q, r)$$
$$\Leftrightarrow \quad \tilde{G}(r, z_1) = \tilde{G}(r+Q, z_2). \tag{17}$$

The above optimality condition of r is geometrically illustrated in Figure 2(a). In this plot, the form of the function $G_d(y \mid Q, r)$ depends on the chosen values of Q and r. However, for a given Q, the loci of points $(r, G_d(r \mid Q, r))$ and $(r+Q, G_d(r+Q \mid Q, r))$, given by the functions $\tilde{G}(y, z_1)$ and $\tilde{G}(y, z_2)$, respectively, are

Figure 2 **Geometric Representation of Cost Function with Optimal Dual-Freight Policy**



fixed. Thus, the above optimality condition of r selects a point $y = r$ on $\tilde{G}(y, z_1)$ and a point $y = r + Q$ on $\tilde{G}(y, z_2)$ such that the functions' values are the same. The function $G_d(y \mid Q, r)$ then intersects the former at $y = r$ and the latter at $y = r + Q$.

In Figure 2(b), the dual-freight cost rate function $G_d(y \mid Q, r)$ is plotted along with single-freight cost rate functions $\mu c_2 + G(r, D_{(0,l]})$ and $G(r, D_{(0,L]})$. The geometric representations shown in Figure 2(a) and 2(b) effectively capture the dynamics of the optimal dual-freight policy. In particular, this reflects the usage of each freight mode when it is optimally used for chosen r and Q. When r is chosen such that r and $r+Q$ take small values, $\tilde{G}(r, z_1) \approx \mu c_2 + G(r, D_{(0,l]})$ and $\tilde{G}(r+Q, z_2) \approx \mu c_2 + G(r+Q, D_{(0,l]})$. The dual-freight cost rate, which passes through these points, is $G_d(y \mid Q, r) \approx \mu c_2 + G(y, D_{(0,l]})$, reflecting a dominant use of express freight for shipping. Intuitively, small values of r and $r+Q$ result in a large number of expected back orders, making it optimal to ship most units via express freight. Similarly, for large values of r and $r+Q$, $G_d(y \mid Q, r) \approx G(y, D_{(0,L]})$, reflecting that at such Q and r, shipping most of the units with regular freight is optimal, as it saves expected inventory holding and variable express freight costs without substantial risk of back orders. For intermediate r and $r+Q$ (such as the case shown in Figure 2(a)), for each y, $G_d(y \mid Q, r)$ closely follows the smaller of $\mu c_2 + G(r, D_{(0,l]})$ and $G(r, D_{(0,L]})$, reflecting significant use of each freight mode and the resulting cost savings in such cases.

There may exist multiple values of r satisfying the necessary optimality condition in (17). All such solutions, however, lie inside a finite interval (as shown in Lemma 4). The reorder-point solution r can thus be found by a bounded search. Moreover, the geometric representation of the optimality condition explains when its multiple solutions can be ruled out. Consider $c_2 < p(L_2 - l_2)$: Then for small values of the order quantity Q, z_1 and z_2 are either very large or very small numbers, and $\tilde{G}(r, z_1)$ and $\tilde{G}(r+Q, z_2)$ are both approximately equal to one of the single-freight mode cost rate functions. In such cases the solution to (17) is unique and in fact approximately equal to the optimal value of r for the equivalent single-freight model with the same Q. In contrast, for large values of Q, the reorder-point solution r is such that $\tilde{G}(r, z_1) \approx \mu c_2 + G(r, D_{(0, L_1+l_2]})$ and $\tilde{G}(r + Q, z_2) \approx G(r, D_{(0, L_1+L_2]})$. This implies that $\tilde{G}(y, z_1) \geq \tilde{G}(r, z_1)$ for $y < r$ and $\tilde{G}(y, z_2) \geq \tilde{G}(r + Q, z_1)$ for $y > r + Q$, and the solution to (17) is unique. For intermediate values of Q, an extensive numerical investigation indicates that there may exist at most three values of r satisfying (17). Of these three, the smallest and the largest are local minima corresponding to the greater use of express and regular freight, respectively, and the third is a local maximum. The reorder-point solution in such cases can be found by a simple comparison of costs at the two local minima.

Also noteworthy is the fact that $r(Q)$ is not a continuous function of Q in general. When there exist multiple solutions to the necessary optimality condition (17), each solution corresponds to a different mix of the freight modes. As the costs of freight modes exhibit economies of scale, sometimes as Q increases it becomes beneficial to switch from a solution that prescribes greater use of one freight to another solution that prescribes greater use of the other. This results in discontinuity of $r(Q)$ at the point where the switch takes place. Such discontinuity in $r(Q)$ exists only at the values of Q, for which there exist multiple solutions to (17). We clarify this further with the help of a numerical example in §A.4 of the online supplement.

Corollary 1 provides a managerial interpretation of the optimality condition of r.

Corollary 1. *Let $\mathcal{I}(Q, r)$ denote the inventory level at an arbitrary point in time with the optimal dual-freight policy. Then $r(Q)$ satisfies*

$$\mathbb{P}(\mathcal{I}(Q, r) > 0) = \frac{p}{p+h}. \tag{18}$$

The left-hand side of (18) is the time average probability of being in stock and is a measure of service level for the inventory system. According to Corollary 1, for a given order quantity the reorder-point solution sets the value of this service level to $p/(h+p)$. Analogous observations have been made by Gallego (1998) for a continuous review inventory model with single freight and by Groenevelt and Rudi (2003) for a base stock inventory model with single as well as two freights.

Let $r_s(Q) \stackrel{\text{def}}{=} \arg\min_r \mathbb{E}S(Q, r, 0, D_{(0, L_1]})$ and $r_f(Q) \stackrel{\text{def}}{=} \arg\min_r \mathbb{E}S(Q, r, Q, D_{(0, L_1]})$, the optimal reorder points for given Q, for single-freight models with regular and express freights, respectively. The next lemma bounds $r(Q)$ with $r_f(Q)$ and $r_s(Q)$.

Lemma 4. $r_f(Q) \leq r(Q) \leq r_s(Q)$.

Intuitively, for the same values of Q and r, the time average probability of being in stock with the optimal dual-freight policy is greater than that with the use of only regular freight, and is less than that with the use of only express freight. However, in all three cases, at the optimal reorder point for a given order size, the time average probability of being in stock is equal to the ratio $p/(h + p)$. Furthermore, for a given Q, the value of the time average probability of being in stock increases with the reorder point, implying the result in Lemma 4.

6. Order Quantity

This section addresses the usually longer-term issue of determining the order quantity. Let $\mathcal{S}(Q) \stackrel{\text{def}}{=} \mathcal{S}(Q, r(Q))$ and $\mathcal{K}(Q) \stackrel{\text{def}}{=} \mathcal{K}(Q, r(Q))$; then the average cost as a function of the order quantity, given that the jointly optimal reorder point and dual-freight policy for that order quantity are used, is

$$C(Q) \stackrel{\text{def}}{=} \min_r C(Q, r) = \frac{\mu\mathcal{K}(Q) + \mathcal{S}(Q)}{Q}. \tag{19}$$

Let $Q^* \stackrel{\text{def}}{=} \arg\min_Q C(Q)$ denote the order quantity solution to our model. Recall that when Q is not large, $r(Q)$ may be discontinuous for some values of Q, implying the nondifferentiability of $C(Q)$ at such points. This makes the determination of optimal Q somewhat complex. The next proposition provides the optimality condition of Q for the cases where $r(Q)$ is a continuous function.

Proposition 4. *Let $A(Q) \stackrel{\text{def}}{=} Q\tilde{G}(r(Q), z_1) - \mathcal{S}(Q)$. If $r(Q)$ is continuous at Q^*, then Q^* solves*

$$A(Q) = \mu\mathcal{K}(Q). \tag{20}$$

The analytical characterization of Q^* in Proposition 4 is somewhat restrictive. However, in all our numerical experiments we found that Q^* indeed satisfies (20). We provide an explanation for this: when $C(Q)$ is not differentiable at a point, it is because at that point the reorder-point solution $r(Q)$ switches between the two locally optimal values of r, satisfying (17). As illustrated in §A.4 of the online supplement, these two locally optimal reorder points constitute continuous functions of Q, one of which vanishes as Q becomes large. This implies that average cost $C(Q, r)$ evaluated at each locally optimal r is a continuous and differentiable function of Q wherever the locally optimal r exists. And $C(Q)$ is essentially the minimum of these two subfunctions. If Q^* does not satisfy the first-order condition in (20), then it is a point at which $C(Q)$ switches from one of these subfunctions to the other. Additionally, if any of these subfunctions has a nonzero first derivative at such Q^*, then a better Q could be chosen by moving in the direction in which the subfunction with the nonzero derivative decreases, contradicting the definition of Q^*. Hence, Q^* satisfies (20).

Next, we comment on the uniqueness of the solution to (20). The second derivative of $C(Q)$ evaluated at a point satisfying the first-order condition of Q is

$$\begin{aligned}\frac{d^2C(Q)}{dQ^2} &= \frac{1}{Q}\frac{d\tilde{G}(r(Q), z_1)}{dQ}\\ &= \frac{1}{Q}\left(\frac{\partial \tilde{G}(y, z_1)}{\partial y}\bigg|_{y=r(Q)} \frac{dr(Q)}{dQ}\right.\\ &\qquad\left. + \frac{\partial \tilde{G}(r(Q), z)}{\partial z}\bigg|_{z=z_1} \frac{dz_1(Q)}{dQ}\right).\end{aligned}$$

Let $\tilde{G}'(y, z^*) = \partial\tilde{G}(y, z^*)/\partial y$; then for large values of Q, $r(Q)$ is uniquely characterized by (17) and satisfies $\tilde{G}'(r(Q), z_1) < 0$ and $\tilde{G}'(r(Q)+Q, z_2) > 0$, implying that

$$\frac{dr(Q)}{dQ} = -\frac{\tilde{G}'(r(Q)+Q, z_2)}{\tilde{G}'(r(Q)+Q, z_2) - \tilde{G}'(r(Q), z_1)}$$

is well defined and nonpositive. Thus, the first term inside the parenthesis is positive, and the second term vanishes for large values of Q (as z_1 takes a constant value). Consequently, for large K_1 (leading to a larger solution Q to the first-order condition), the optimality condition in (20) characterizes the unique minimum.

For smaller K_1, there may exist multiple local minima satisfying (20). Intuitively, each local minimum satisfying (20) is an order quantity that matches the rate at which average fixed costs $\mu\mathcal{K}(Q)/Q$ decrease with Q, with the rate at which the average variable cost $\mathcal{S}(Q)/Q$ increase with Q. Because of economies of scale in transportation costs, these rates are not uniformly monotone. More specifically, when Q becomes large enough so that orders can be split more often, $\mathcal{S}(Q)/Q$ does not increase with Q as rapidly as it does when one of the freight modes is dominantly used. This results in multiple ways of achieving the tradeoff between the fixed and variable costs by mixing the two freight modes in different ways. Numerical investigation suggests that there may exist at most two local minima; typically the smaller of these corresponds to the dominant use of one of the freight modes. In §A.4 of the online supplement, we illustrate this using a numerically solved example.

Revisiting Figure 2(a), when r is chosen optimally, the chord connecting the points $\big(r, G_d(r \mid Q, r)\big)$ and $\big(r+Q, G_d(r+Q \mid Q, r)\big)$ is a horizontal line, and the function $A(Q)$ is the area enclosed between the chord and the cost rate function $G_d(y \mid Q, r)$. Proposition 4 therefore states that at Q^* this area is equal to the expected aggregate fixed cost $\mathcal{K}(Q)$ *times* the demand rate μ. Although this geometric representation of the optimality conditions of r and Q is similar to the one shown in Zheng (1992) for a single-freight model, its dynamics are more complex in the current context. Indeed, the result of Zheng (1992) is a special case of our result. When the express freight is prohibitively expensive, then $z_1, z_2 \downarrow -\infty$, implying that $G_d(y \mid Q, r) = G(y, D_{(0, L]})$. Similarly, when the regular freight is prohibitively expensive, then $z_1, z_2 \uparrow \infty$, implying that $G_d(y \mid Q, r) = \mu c_2 + G(y, D_{(0, l]})$. In both these cases, our solution recommends using the cheaper freight mode all the time, and its geometric representation reduces to the one shown in Zheng (1992). However, for a less extreme case, when the solution allows significant use of both freights, the geometric representation becomes a combination of the two geometric representations corresponding to the two single-freight models.

7. Discussion

This section illustrates the properties of our model with the help of two special cases and a computational study. In §7.1, where we analyze the special case with negligible fixed costs of freight modes, we highlight that our model has properties similar to a single-freight model for large values of K_1 and provide analytical bounds on saving resulting from using two freight modes. Section 7.2 provides insights into the value of information resulting from delaying the freight-mode decision and its effects on the optimal cost and order size. Finally, in §7.3, we discuss the key observations from a computational study of our model.

7.1. Special Case Without Freight-Mode Fixed Costs

When $c_2 < p(L_2 - l_2)$ and the freight-mode fixed costs are negligible—i.e., $K_2 = k_2 = 0$—the expected aggregate fixed cost $\mathcal{K}(Q, r) = K_1$ and the dual-freight cost rate function $G_d(y \mid Q, r) = \tilde{G}(y, z^*)$ are considerably simpler. Note that for very large values of K_1 relative to K_2 and k_2, the dual-freight cost rate function asymptotically approaches the cost rate function of this special case. Hence, this special case is a good approximation for the general case, when K_1 is very large. Groenevelt and Rudi (2003) provide a comprehensive analysis of this special case for a base stock inventory model.

Lemma 5. (a) $\tilde{G}(y, z^*) \le \min\{G(y, D_{(0,L]}), \mu c_2 + G(y, D_{(0,l]})\}$.

(b) $\lim_{y\to-\infty}(\tilde{G}(y, z^*) + py) = \mu c_2 + \mu p l$ *and* $\lim_{y\to\infty}(\tilde{G}(y, z^*) - hy) = -\mu h L$.

(c) $\partial G(y, D_{(0,L]})/\partial y \le \partial\tilde{G}(y, z^*)/\partial y \le \partial G(y, D_{(0,l]})/\partial y$.

Remark 7. Let $y_s^0 = \arg\min_y\{G(y, D_{(0,L]})\}$ and $y_f^0 = \arg\min_y\{\mu c_2 + G(y, D_{(0,l]})\}$; then Lemma 5(c) implies that $\tilde{G}'(y, z^*) < 0$ for $y < y_f^0$, and $\tilde{G}'(y, z^*) > 0$ for $y > y_s^0$. Thus, for $K_2 = k_2 = 0$ and $c_2 < p(L_2 - l_2)$, when $r(Q)$ satisfies $y_s^0 - Q < r(Q) < y_f^0$ (which happens for sufficiently large $Q \gg y_s^0 - y_f^0$), then it is the unique solution r to the optimality condition $\tilde{G}(r, z^*) = \tilde{G}(r + Q, z^*)$.

The above remark leads to the following properties of the optimal solution for the special case.

Lemma 6. *For $K_2 = k_2 = 0$, $c_2 < p(L_2 - l_2)$ and value of Q such that $y_s^0 - Q < r(Q) < y_f^0$.*

(a) *$r(Q)$ is a continuous and differentiable function of Q.*

(b) $-1 < r'(Q) < 0$.

(c) *$A(Q)$ is a continuous, differentiable, and increasing function of Q.*

(d) *The solution for order quantity Q^* is increasing in K_1.*

The above lemma illustrates that the special case of our model has similar properties to the model with single-freight mode, as shown by Zheng (1992), and the managerial interpretations of the solution remain unchanged.

Let $S_s(Q) \stackrel{\text{def}}{=} \min_r \mathbb{E}S(Q, r, 0, D_{(0,L_1]})$ and $S_f(Q) \stackrel{\text{def}}{=} \min_r \mathbb{E}S(Q, r, Q, D_{(0,L_1]})$. Then $C_s(Q) \stackrel{\text{def}}{=} (\mu(K_1 + K_2) + S_s(Q))/Q$ and $C_f(Q) \stackrel{\text{def}}{=} (\mu(K_1 + k_2) + S_f(Q))/Q$ are the average costs of the single-freight models with regular and express freights, respectively, given that the reorder points are chosen optimally for order quantity Q. Our next result provides bounds on the cost savings of the two-freight model relative to the single-freight models.

Lemma 7. *Define* $\delta_s \stackrel{\text{def}}{=} (h/(h+p))(p(L_2 - l_2) - c_2)$, $\delta_f \stackrel{\text{def}}{=} (p/(h+p))(h(L_2 - l_2) + c_2)$, $C^* \stackrel{\text{def}}{=} \min_Q C(Q)$, $C^s \stackrel{\text{def}}{=} \min_Q C_s(Q)$ *and* $C^f \stackrel{\text{def}}{=} \min_Q C_f(Q)$. *Then for $K_2 = k_2 = 0$ and $c_2 < p(L_2 - l_2)$,*

(a) $0 \le C_s(Q) - C(Q) \le \mu\delta_s$ *and* $0 \le C_f(Q) - C(Q) \le \mu\delta_f$.

(b) $0 \le C^s - C^* \le \mu\delta_s$ *and* $0 \le C^f - C^* \le \mu\delta_f$.

7.2. Value of Information

Recall that the decision of allocating product units to the freight modes is postponed until L_1 time after placing an order. We refer to the cost reduction from this postponement as the *value of information*. To better understand the value of information, we parameterize the manufacturing lead time as Δ and keep the total regular freight lead time L and the total express freight lead time l fixed (where Δ satisfies $0 \le \Delta \le l$). For simplicity we focus on the cases with $K_2 = k_2 = 0$ and $c_2 < p(L - l)$. The expression for the average cost of this model is

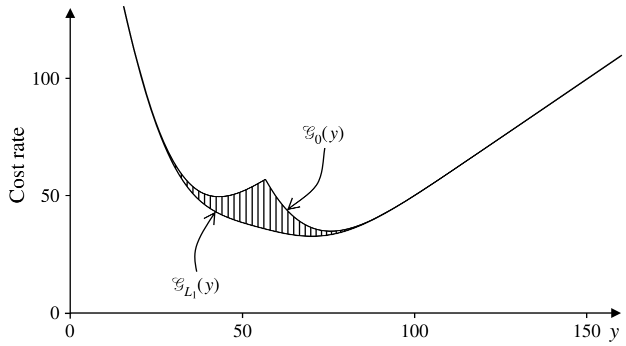
$$\mathscr{C}_\Delta(Q, r) = \frac{\mu K_1 + \int_r^{r+Q} \mathscr{G}_\Delta(y)\,dy}{Q}, \tag{21}$$

where the cost rate function $\mathscr{G}_\Delta(\cdot)$ is given by

$$\mathscr{G}_\Delta(y) = \min_z\left\{\int_{-\infty}^{y-z} G(y - x, D_{(\Delta,L]})\,dF_{(0,\Delta]}(x) + \int_{y-z}^{\infty} (\mu c_2 + G(y - x, D_{(\Delta,l]}))\,dF_{(0,\Delta]}(x)\right\}. \tag{22}$$

When $\Delta = L_1$ and $\mathscr{G}_{L_1}(y) = \tilde{G}(y, z^*)$, this is equivalent to the model studied in §7.1. In Figure 3, the cost rate

Figure 3 **Cost Rate Function $\mathscr{G}_\Delta(y)$ Plotted Against y for $\Delta = 0$ and $\Delta = L_1$**



function $\mathscr{G}_\Delta(y)$ is plotted for $\Delta = 0$ and $\Delta = L_1$. In this plot, the area between the cost rate functions $\mathscr{G}_0(y)$ and $\mathscr{G}_{L_1}(y)$ represents the value of information. In the next proposition, we characterize the effect of parameter Δ on expected cost and order quantity.

Proposition 5. *Let $\mathscr{C}^*_\Delta$ denote the minimum value of the cost function in* (21) *and r^*_Δ and Q^*_Δ the optimal values of reorder point and order quantity, respectively. Given $\Delta_1 < \Delta_2$,*

(a) $\mathscr{G}_{\Delta_1}(y) \geq \mathscr{G}_{\Delta_2}(y)$, $\forall\, y$;

(b) $\mathscr{C}^*_{\Delta_1} \geq \mathscr{C}^*_{\Delta_2}$; *and*

(c) *when conditions in Lemma* 6 *are satisfied,* $Q^*_{\Delta_1} \geq Q^*_{\Delta_2}$.

The intuitive explanation of parts (a) and (b) of the above proposition is quite straightforward: When Δ is large, the freight-mode decision is made based on more information, and consequently the cost is smaller. However, the effect of Δ on Q^*_Δ is more subtle: In our model with no postponement—i.e., $\Delta = 0$—the system experiences the highest cost of demand variance (i.e., the difference in stochastic and deterministic cost rates) when the inventory level is close to 0 (in Zheng 1992, this portion of the cost is referred to as "uncontrollable cost"). To incur this situation less frequently, the optimal order quantity with stochastic demand and zero manufacturing lead time Q^*_0 is higher than its deterministic counterpart. However, with a larger value of Δ, the optimal dual-freight policy allocates the manufacturing order across the two freight modes more effectively, thus decreasing the cost of uncertainty around inventory level 0. Therefore, with larger Δ, it does not cost as much to place more frequent orders, and Q^*_Δ is smaller.

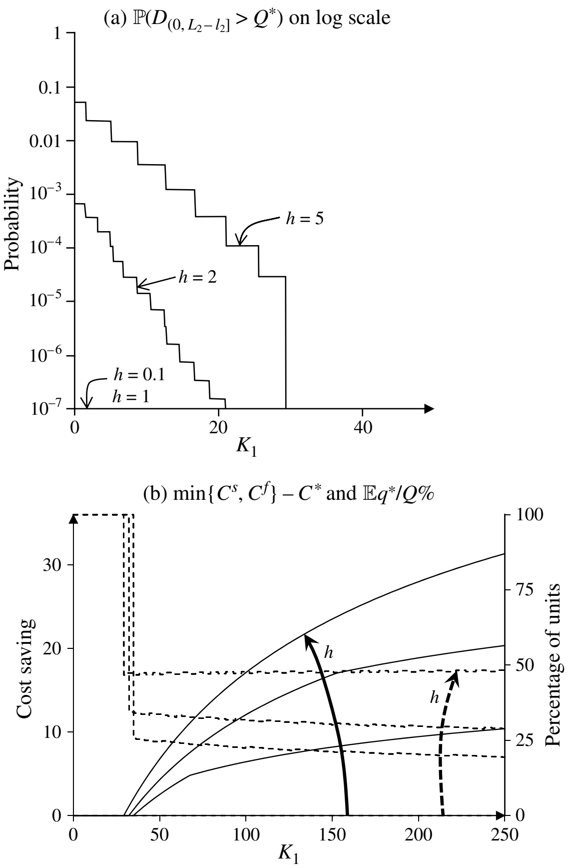
7.3. Observations from Numerical Investigations

We carried out an extensive computational study to supplement the analytical results in this paper. We discuss only the key observations in this subsection but provide a detailed description of the computational work in the online supplement.

In the first part of our numerical study, we investigate conditions under which the (Q, r) policy is reasonable for our model. As noted in §3, the optimal policy is indeed a (Q, r) policy when orders do not cross. We therefore examine the likelihood of order-crossing for the solution to our model. Recall that a higher value of the probability $\mathbb{P}(D_{(0,\, L_2-l_2]} > Q^*)$ implies a higher likelihood of two consecutive orders placed within the $L_2 - l_2$ time interval, which in turn translates to a higher probability of order-crossing *given* both freight modes are employed in shipping each order (i.e., when the dual-freight model is of value in terms of cost reduction). In our model, where the freight mode usage is dynamic, the probability $\mathbb{P}(D_{(0,\, L_2-l_2]} > Q^*)$ provides an upper bound on the actual probability of order-crossing. This upper bound is more likely to be tight when the percent usage of express freight $\mathbb{E}q^*/Q^*\%$ is different from 0% and 100% and the value of the dual-freight model measured by cost savings over the cheaper of the two single-freight models $\min\{C^s, C^f\} - C^*$ is significant. To gain insight into the likelihood of order-crossing in our model, we examine the quantities $\mathbb{P}(D_{(0,\, L_2-l_2]} > Q^*)$, $\mathbb{E}q^*/Q^*\%$, and $\min\{C^s, C^f\} - C^*$ jointly.

In Figure 4(a) the upper bound on the probability of order-crossing $\mathbb{P}(D_{(0,\, L_2-l_2]} > Q^*)$ is plotted on a logarithmic scale as a function of $K_1 \in [0, 50]$. In Figure 4(b) the value of the two freight modes $\min\{C^s, C^f\} - C^*$ and the percent usage of express freight $\mathbb{E}q^*/Q\%$ are plotted as functions of $K_1 \in [0, 250]$ with solid and dotted lines, respectively. In Figure 4(b) the thick solid and dotted arrows indicate the direction in which h increases while keeping $h + p = 10$. The demand is a Poisson process with $\mu = 50$. In these figures, the upper bound on the probability of order-crossing $\mathbb{P}(D_{(0,\, L_2-l_2]} > Q^*)$ decreases sharply as K_1 increases. For small values of K_1 at which $\mathbb{P}(D_{(0,\, L_2-l_2]} > Q^*)$ is not negligible, $\min\{C^s, C^f\} - C^* \approx 0$ and $\mathbb{E}q^*/Q^*\%$ is either 0% (when $h = 0.1$, suggesting use of only regular freight) or 100% (when $h \in \{1, 2, 5\}$, suggesting

Figure 4 **Probability $\mathbb{P}(D_{(0, L_2-l_2]} > Q^*)$, $\mathbb{E}q^*/Q\%$, and $\min\{C^s, C^f\} - C^*$ as Functions of K_1**



Note. For $L_1 = 0.3$, $L_2 = 0.7$, $l_2 = 0.2$, $K_2 = 50$, $k_2 = 25$, $C_2 = 0.25$, $\mu = 50$, and $h \in \{0.1, 1, 2, 5\}$ with $h + p = 10$.

use of only express freight). Finally, for intermediate values of K_1, $\mathbb{P}(D_{(0, L_2-l_2]} > Q^*) \approx 0$ and $\min\{C^s, C^f\} - C^* \approx 0$ hold simultaneously. These observations suggest the following: When K_1 is not too small, the probability of order-crossing is negligible and (Q, r) policy is a reasonable assumption; when K_1 is small, the dual freight mode is of negligible value and the optimal policy is likely to use only one of the freight modes almost all the time, in which case a (Q, r) policy is optimal.

In §A.1 of the online supplement, we carry out analogous numerical investigations at values of model parameters chosen to put the model under high stress in terms of order-crossing. Even then we find that the (Q, r) policy is reasonable in most cases. Moreover, we find that the upper bound on the probability of order-crossing $\mathbb{P}(D_{(0, L_2-l_2]} > Q^*)$ and the value of the dual-freight model $\min\{C^s, C^f\} - C^*$ simultaneously take significant values, when demand rate μ and minimum fixed cost $K_1 + \min\{K_2, k_2\}$ take small values and $h/(h+p)$ is close to 0.5. In other words when these three conditions are met, a (Q, r) policy may not be a reasonable assumption for our model.

It is often not feasible to change the production batch size. A firm can nevertheless ship orders optimally using both freight modes and choose the reorder point optimally for a given batch size. An issue of interest, therefore, is the sensitivity of the cost function $C(Q)$ to the order quantity Q. To establish the insensitivity of the cost to the order quantity for a single-freight model, Zheng (1992) has analytically shown that using the optimal order quantity determined assuming deterministic demand, when the demand is actually stochastic, leads to an increase in cost no greater than 1/8 of the optimal cost. In his computational study the increase never exceeded 2.9%. In a similar test of our model, over a large set of examples solved (described in §A.2 of the online supplement), we found that the cost of our model is similarly insensitive to order quantity.

The next part of the study focuses on the effects of freight-mode costs K_2, k_2, and c_2 on the solution and cost of our model. The observations from these numerically solved examples suggest that when the fixed cost of placing an order K_1 is smaller than or comparable in magnitude to the freight-mode fixed costs K_2 and k_2, the solution is sensitive to fixed costs K_2 and k_2. In contrast, when K_1 is large as compared to K_2 and k_2, the solution becomes more sensitive to the variable cost of express freight c_2. When K_1 is small, the solution recommends shipping most of the units via the freight mode with smaller fixed costs and using the other freight mode only in extreme cases. The reorder-point and order-quantity solutions and corresponding costs in such cases are similar to the single-freight model with smaller cost. Similarly, in response to changes in express freight variable cost c_2, the solution can typically be characterized by three phases: For small values of c_2, the use of express freight dominates; for large values of c_2, the use of regular freight dominates; and for intermediate values of c_2, significant fractions of the order quantity are shipped by each freight mode. The intermediate phase vanishes when K_1 takes small values.

The numerical study also sheds light on the value of having two freight modes available. We found that whenever the difference in the optimal costs of the two single-freight models (i.e., C^s and C^f) is large, the cost with two-freight model C^* is marginally smaller than $\min\{C^s, C^f\}$. In such cases the solution with two freight modes corresponds to the dominant use of one of the freight modes and is similar to that of the single-freight model with smaller cost. However, when the difference between C^s and C^f is small, each freight mode is used often enough and C^* is substantially smaller than $\min\{C^s, C^f\}$. These cases correspond to parameter values $K_1 \gg \max\{K_2, k_2\}$ and $c_2 < p(L_2 - l_2)$. In such cases we also observe that $Q^* \geq \max\{Q^s, Q^f\}$. With parameter value $K_2 = k_2 = 0$ and $c_2 < p(L_2 - l_2)$, this phenomenon always occurred in our numerical examples. We discuss this further in §A.5 of the online supplement. Intuitively, when orders are split between the two freight modes, the average variable cost (consisting of inventory holding, back-order penalty, and variable cost of express freight) does not increase with Q as rapidly as it does when each order is shipped in a single shipment. Thus, with the greater possibility of splitting orders, the optimal trade-off between $\mu\mathcal{K}(Q)/Q$ and $\mathcal{S}(Q)/Q$ is attained at a larger value of Q. With the optimal dual-freight policy, the orders are split with higher probability when $c_2 < p(L_2 - l_2)$ and $K_1 \gg \max\{K_2, k_2\}$; hence, in such cases Q^* is larger than Q^s and Q^f.

8. Concluding Remarks

Our paper contributes to the rich literature on inventory management with multiple replenishment modes by incorporating economies of scale in transportation costs and modeling the demand-responsive nature of logistic decisions. In contrast to the existing results in which orders are always split, the optimal dual-freight policy derived in this paper allows for four different ways of mixing the two freight modes for shipping an order. Although the effect of freight-mode fixed costs on the optimal usage of the freight modes for a given order is relatively straightforward, its overall impact on the optimal reorder point and order quantity is not so. Our approach to analyzing this optimization problem integrates all four possible cases of the optimal freight-mode decision as well as its contingent nature. To elucidate this somewhat complex analysis involving nonconvex functions, we use a geometric representation of the cost function and the optimality conditions. The geometric representation of our analytical solution enables easier understanding of the effects of various cost parameters and decision postponement on the optimal solution and allows a comparison of our model with single-freight models. Our model illustrates that the fixed costs of freight modes play a significant role in determining freight-mode mix for shipping inventory when the fixed cost of placing an order is relatively small. However, at higher values of the fixed ordering cost, the variable cost of the express freight becomes more important. The availability of two freight modes is most beneficial when the fixed costs of freight modes are small relative to the fixed ordering cost and the per unit cost of the express freight is not too large.

Our work has several interesting extensions. Although in our paper the freight modes differ in lead times and costs, one can analyze situations where freight modes differ in reliability and costs. In addition to providing decision rules for mixing two such freight modes, such analysis would provide insights into the premium for lead-time reliability. The fixed and variable costs for the freight modes are typically part of a contract negotiated between the shipper and its logistics service provider(s). Our model can be used as a building block for analyzing such contracts and can help answer several interesting questions related to contract choices and their effects on the fleet use of the logistic service provider and the total supply chain costs.

Electronic Companion

An electronic companion, which contains the online supplement to this paper, is available on the *Manufacturing & Service Operations Management* website (http://msom.pubs.informs.org/ecompanion.html).

Acknowledgments

The authors thank Gregory Dobson and Robert Shumsky for their comments on an earlier draft of this paper. The authors also thank the associate editor and the referees for their thoughtful insights and suggestions.

Appendix

Proof of Proposition 1. The equivalence between expressions (3), stated in terms of expected waiting times, and (5), stated in terms of expected inventory levels, follows from Little's Law. We first apply Little's Law to a special case of our model and then extend it to the general case.

Consider the simple case of our model facing the same stochastic demand process, with single lead time $\mathscr{L}$, and one-for-one continuous review inventory policy with order-up-to level y. In this special case, on the arrival of each demand, a product unit indexed y enters the system and reaches Stage 1 after $\mathscr{L}$ time. For this system, the expected waiting time for a demand is $\mathbb{E}(\mathscr{L}-t_y)^+ + \mathbb{E}t_{(-y)}$, and the expected number of back orders or, equivalently, the number of demands waiting for fulfillment is $\mathbb{E}(D_{(0,\mathscr{L}]}-y)^+$. Applying Little's Law to this system, we get

$$\mu(\mathbb{E}(\mathscr{L}-t_y)^+ + \mathbb{E}t_{(-y)}) = \mathbb{E}(D_{(0,\mathscr{L}]}-y)^+. \tag{23}$$

Now consider the expression for $C_q(Q,r)$ in (3). Note that when the demands arrive in single units, $\mu\mathbb{E}t_{(-y)} = (-y)^+$. Along with (23) and the definitions of $g(y,\mathscr{L})$ and $G(y,D_{(0,\mathscr{L}]})$, this gives $\mu g(y,\mathscr{L}) = G(y,D_{(0,\mathscr{L}]}) - p(-y)^+$. Substituting for $g(y-D_{(0,L_1]},L_2)$ and $g(y-D_{(0,L_1]},l_2)$ in (4) and taking expectation over $D_{(0,L_1]}$ gives the following expression for $\mathbb{E}\Gamma(Q,r,q(D_{(0,L_1]}),D_{(0,L_1]})$:

$$\begin{aligned}
&\mathbb{E}\big(K_2\mathbb{1}_{\{q(D_{(0,L_1]})<Q\}} + k_2\mathbb{1}_{\{q(D_{(0,L_1]})>0\}} + c_2 q(D_{(0,L_1]})\big)\\
&\quad -\frac{1}{\mu}\bigg(\sum_{y=r+1}^{r+Q} p\mathbb{E}(D_{(0,L_1]}-y)^+\bigg)\\
&\quad +\frac{1}{\mu}\mathbb{E}\bigg(\sum_{y=r+1}^{r+q(D_{(0,L_1]})} G(y-D_{(0,L_1]},D_{(L_1,l]})\\
&\qquad\qquad + \sum_{y=r+q(D_{(0,L_1]})+1}^{r+Q} G(y-D_{(0,L_1]},D_{(L_1,L]})\bigg).
\end{aligned}$$

Substituting the above in (3) then leads to expression (5) for $C_q(Q,r)$. □

Proof of Lemma 1. When $K_2=0$ and $k_2=0$, $K(Q,q)=K_1$; minimizing $K(Q,q)+S(Q,r,q,D_{(0,L_1]})$ is equivalent to minimizing $S(Q,r,q,D_{(0,L_1]})$. It follows from the expression in (8) that $S(Q,r,q,D_{(0,L_1]})$ is convex in the target inventory position $z\equiv r+q-D_{(0,L_1]}$. Furthermore, when $c_2\ge p(L_2-l_2)$, the function is nondecreasing in z for all values. The unconstrained version of the problem of determining optimal z is therefore solved by $z^*\to-\infty$ when $c_2\ge p(L_2-l_2)$ and by z^*, the unique solution to the first-order condition in (10), otherwise. In the presence of the constraint $q\in[0,Q]$, the solution is $q^*=0$ when $r-D_{(0,L_1]}>z^*$, $q^*=Q$ when $r+Q-D_{(0,L_1]}<z^*$, and $q^*=z^*-r+D_{(0,L_1]}$, otherwise. □

Proof of Lemma 2. The proofs for (a) and (b) directly follow from evaluating the first derivatives of respective functions and noting that $D_{(0,L_2]}$ is stochastically greater than $D_{(0,l_2]}$. The bounds in (c) follow from applying the limits $r\to\infty$ and $r\to-\infty$ to the function. □

Proof of Proposition 2. It follows from Lemma 1 and Remark 3 that the optimal value of q is one of 0, $z^*-r+D_{(0,L_1]}$, and Q, where the second one is achievable only when it falls between 0 and Q and can be excluded when $c_2\ge p(L_2-l_2)$. The values of the objective function for these three cases are

$$\mu(K_1+K_2)+\int_{r-D_{(0,L_1]}}^{r+Q-D_{(0,L_1]}} G(u,D_{(L_1,L]})\,du, \tag{24}$$

$$\mu(K_1+K_2+k_2)+\int_{r-D_{(0,L_1]}}^{z^*}(\mu c_2+G(u,D_{(L_1,l]}))\,du + \int_{z^*}^{r+Q-D_{(0,L_1]}} G(u,D_{(L_1,L]})\,du, \quad\text{and} \tag{25}$$

$$\mu(K_1+k_2)+\int_{r-D_{(0,L_1]}}^{r+Q-D_{(0,L_1]}}(\mu c_2+G(u,D_{(L_1,l]}))\,du, \tag{26}$$

respectively. First consider $c_2<p(L_2-l_2)$. Then all three expressions need to be compared. The difference between expressions (24) and (25) is

$$-\mu k_2+\int_{r-D_{(0,L_1]}}^{z^*}(G(u,D_{(L_1,L]})-\mu c_2-G(u,D_{(L_1,l]}))\,du. \tag{27}$$

The integrand in the above expression is nonnegative for $u\le z^*$ (Lemma 2(a) and the definition of z^* in (10)). Consequently, the integral is nonnegative for $D_{(0,L_1]}\ge r-z^*$ and nondecreasing in $D_{(0,L_1]}$. Hence, (27) is nondecreasing in $D_{(0,L_1]}$ and vanishes at $D_{(0,L_1]}=r-\underline{z}\ge r-z^*$ (where $\underline{z}$ solves (13)). Clearly, between the choices $q=z^*-r+D_{(0,L_1]}$ and $q=0$, the former is optimal when $D_{(0,L_1]}>r-\underline{z}$, and the latter is optimal when $D_{(0,L_1]}<r-\underline{z}$. A similar comparison of expressions (25) and (26) implies that between the choices $q=Q$ and $q=z^*-r+D_{(0,L_1]}$, the former is optimal when $D_{(0,L_1]}>r+Q-\bar{z}$, and the latter is optimal when $D_{(0,L_1]}<r+Q-\bar{z}$ (where $\bar{z}$ solves (14)). In summation, if $D_{(0,L_1]}<r-\underline{z}$, then $q^*=0$; if $D_{(0,L_1]}>r+Q-\bar{z}$, then $q^*=Q$; and if $r-\underline{z}<D_{(0,L_1]}<r+Q-\bar{z}$, then $q^*=z^*-r+D_{(0,L_1]}$. However, when $Q\le\bar{z}-\underline{z}$, the last case never occurs and either $q=0$ or $q=Q$ is optimal. The difference between expressions (24) and (26) is

$$\mu K_2-\mu k_2+\int_{r-D_{(0,L_1]}}^{r+Q-D_{(0,L_1]}}\big(G(u,D_{(L_1,L]}) -\mu c_2-G(u,D_{(L_1,l]})\big)\,du \tag{28}$$

and is nondecreasing in $D_{(0,L_1]}$ (Lemma 2(b)) and vanishes at $D_{(0,L_1]}=r-z_p(Q)$, where $z_p(Q)$ solves (12). Thus, for $Q\le\bar{z}-\underline{z}$, if $D_{(0,L_1]}<r-z_p(Q)$, then $q^*=0$, and if $D_{(0,L_1]}>r-z_p(Q)$, then $q^*=Q$.

The expression in (28) is bounded below by $\mu K_2-\mu k_2-(h(L_2-l_2)+c_2)\mu Q$ and bounded above by $\mu K_2-\mu k_2+(p(L_2-l_2)-c_2)\mu Q$ (Lemma 2(c)). Thus, when $K_2<k_2$ and $Q\le Q_e$, the expression is nonpositive for all values of $D_{(0,L_1]}$, implying $q^*=0$; when $K_2>k_2$ and $Q\le Q_r$, the expression is nonnegative for all values of $D_{(0,L_1]}$, implying $q^*=Q$. This completes the proof of the proposition for the cases in the first column of Table 1.

For $c_2\ge p(L_2-l_2)$, using Remark 4, q^* can take only the value 0 or Q. Also, if $c_2\ge p(L_2-l_2)$, then $S(Q,r,Q,D_{(0,L_1]})>S(Q,r,0,D_{(0,L_1]})$. Furthermore, if $K_2<k_2$ (or, equivalently,

$K(Q,Q) > K(Q,0)$), then clearly $q^* = 0$. This completes the proof for the case in the upper-right quadrant of Table 1. For $K_2 > k_2$, the comparison of the two choices follows from analyzing expression (28): When $D_{(0,L_1]} < r - z_p(Q)$, then $q^* = 0$, and when $D_{(0,L_1]} > r - z_p(Q)$, then $q^* = Q$, and when $Q \le Q_r$, then $q^* = Q$. Additionally, as the expression in (28) is bounded above by $\mu K_2 - \mu k_2 - (c_2 - p(L_2 - l_2))\mu Q$, when $Q \ge Q_e$, $q^* = 0$ for all values of $D_{(0,L_1]}$. This covers the lower-right quadrant of Table 1 and completes the proof. $\square$

Proof of Lemma 3. (a) Equation (9) can be written as

$$C(Q,r) = \frac{\mu \mathbb{E}K(Q, q^*(D_{(0,L_1]})) + \mathbb{E}S(Q, r, q^*(D_{(0,L_1]}), D_{(0,L_1]})}{Q}.$$

Let $\mathcal{K}(Q,r) = \mathbb{E}K(Q, q^*(D_{(0,L_1]}))$; then it follows from Equation (6) that $\mathcal{K}(Q,r) = K_1 + K_2\mathbb{P}(q^*(D_{(0,L_1]}) < Q) + k_2\mathbb{P}(q^*(D_{(0,L_1]}) > 0)$, which on applying Definition 1 of $z_1(Q)$ and $z_2(Q)$, leads to the desired expression for $\mathcal{K}(Q,r)$. Substituting for $q^*(D_{(0,L_1]})$ from Proposition 2 in the expression for $S(Q, r, q^*(D_{(0,L_1]}), D_{(0,L_1]})$ in (8), taking expectation over all values of $D_{(0,L_1]}$, and subsequent algebraic manipulation of the resulting expression, leads to $\mathbb{E}S(Q, r, q^*(D_{(0,L_1]}), D_{(0,L_1]}) = \int_r^{r+Q} G_d(y \mid Q, r)\,dy$, where, (i) if $q^*(D_{(0,L_1]})$ takes values 0, $z^* - r + D_{(0,L_1]}$ and Q, then

$$G_d(y \mid Q, r) = \begin{cases} \tilde{G}(y, y - r + \underline{z}) & \text{if } y \le r + (z^* - \underline{z}), \\ \tilde{G}(y, y - r - Q + \bar{z}) & \text{if } y \ge r + Q - (\bar{z} - z^*), \\ \tilde{G}(y, z^*) & \text{otherwise;} \end{cases}$$

(ii) if $q^*(D_{(0,L_1]})$ takes values 0 and Q, then $G_d(y \mid Q, r) = \tilde{G}(y, y - r + z_p(Q))$; (iii) if $q^*(D_{(0,L_1]}) = 0$, then $G_d(y \mid Q, r) = G(y, D_{(0,L]})$; and (iv) if $q^*(D_{(0,L_1]}) = Q$, then $G_d(y \mid Q, r) = \mu c_2 + G(y, D_{(0,l]})$. Using Definition 1 of $z_1(Q)$ and $z_2(Q)$ and Remark 5, we can express $G_d(y \mid Q, r)$ for these four cases as is stated in the lemma.

(b) Equation (9) can be alternatively expressed as $QC(Q,r) = \mu K_1 + \mathbb{E}s(r - D_{(0,L_1]}, r + Q - D_{(0,L_1]})$, where, for a given realization x of $D_{(0,L_1]}$,

$$\begin{aligned} &s(r-x, r+Q-x) \\ &\quad = \min_{r-x \le z \le r+Q-x} \bigg\{ \mu K_2 \mathbb{1}_{\{z < r+Q-x\}} + \mu k_2 \mathbb{1}_{\{z > r-x\}} \\ &\qquad + \int_{r-x}^{z} (\mu c_2 + G(u, D_{(L_1,l]}))\,du + \int_z^{r+Q-x} G(u, D_{(L_1,L]})\,du \bigg\}. \end{aligned}$$

It follows that $s(v,w)$ is continuous and differentiable in v and w almost everywhere. This implies for nondeterministic and continuous $D_{(0,L_1]}$, $\mathbb{E}s(r - D_{(0,L_1]}, r + Q - D_{(0,L_1]})$ is continuous and differentiable in r and $r+Q$, and so is $C(Q,r)$ in Q and r. $\square$

Proof of Proposition 3. In the expression for $C(Q,r)$ in (16), only the term in the numerator depends on r. As $QC(Q,r)$ is a continuous and differentiable function of r, which $\uparrow \infty$ as $|r| \to \infty$, the first-order condition is a necessary condition for the optimal value of the reorder point. Denoting $f_{(0,L_1]}(x) = \partial F_{(0,L_1]}(x)/\partial x$, the probability density function of $D_{(0,L_1]}$, the first derivative of $QC(Q,r)$ with respect to r is,

$$\begin{aligned} \frac{\partial QC(Q,r)}{\partial r} &= G_d(r+Q \mid Q, r) - G_d(r \mid Q, r) \\ &\quad + \int_r^{r+Q} \frac{\partial G_d(y \mid Q, r)}{\partial r}\,dy + \mu K_2 f_{(0,L_1]}(r+Q-z_2) \\ &\quad - \mu k_2 f_{(0,L_1]}(r - z_1). \end{aligned} \tag{29}$$

Substituting for $G_d(y \mid Q, r)$ (from Lemma 3) in the third term of the above expression and calculating the resultant expression for each case in Remark 5, while applying Equations (12), (13), and (14), it follows that in all cases, the sum of the last three terms of the right-hand side of (29) vanishes. This leads to the first-order condition $G_d(r \mid Q, r) = G_d(r + Q \mid Q, r)$, or $\tilde{G}(r, z_1) = \tilde{G}(r + Q, z_2)$. The existence of at least one value of r satisfying (17) is ensured as $\tilde{G}(r+Q, z_2) - \tilde{G}(r, z_1) < 0$ for $r \to -\infty$, and $\tilde{G}(r+Q, z_2) - \tilde{G}(r, z_1) > 0$ for $r \to \infty$. $\square$

Proof of Corollary 1. Let $\mathcal{T}(Q,r)$ denote the time spent by an arbitrary product unit in inventory (i.e., available to meet its demand). Recall that the index $\mathcal{Y}$ of a product unit as it enter the system is $\mathcal{Y} \in (r, r+Q]$. Then, given a realization x of demand $D_{(0,L_1]}$, $\mathcal{T}(Q,r)$ takes value $(t_{\mathcal{Y}-x} - L_2)^+$ if the unit is shipped via regular freight and $(t_{\mathcal{Y}-x} - l_2)^+$ if it is shipped by express freight. Thus, with the optimal dual-freight policy, $\mathcal{T}(Q,r)$ satisfies

$$\begin{aligned} &\mathbb{P}(\mathcal{T}(Q,r) > 0) \\ &\quad = \frac{1}{Q}\bigg(\int_{-\infty}^{r-z_1} \int_r^{r+Q} \mathbb{P}(t_{y-x} > L_2)\,dy\,dF_{(0,L_1]}(x) \\ &\qquad + \int_{r-z_1}^{r+Q-z_2} \bigg(\int_r^{z^*+x} \mathbb{P}(t_{y-x} > l_2)\,dy \\ &\qquad\qquad + \int_{z^*+x}^{r+Q} \mathbb{P}(t_{y-x} > L_2)\,dy \bigg)\,dF_{(0,L_1]}(x) \\ &\qquad + \int_{r+Q-z_2}^{\infty} \int_r^{r+Q} \mathbb{P}(t_{y-x} > l_2)\,dy\,dF_{(0,L_1]}(x) \bigg). \end{aligned} \tag{30}$$

Expanding, the optimality condition of r in (17) and applying $\mu c_2 = G(z^*, D_{(L_1,L]}) - G(z^*, D_{(L_1,l]})$ from (10), and $G(r+Q, D_{(0,\mathcal{L}]}) - G(r, D_{(0,\mathcal{L}]}) = (h+p)\int_r^{r+Q} F_{(0,\mathcal{L}]}(y)\,dy - pQ$, we get

$$\begin{aligned} &\frac{1}{Q}\bigg(\int_{-\infty}^{r-z_1} \int_r^{r+Q} F_{(L_1,L]}(y-x)\,dy\,dF_{(0,L_1]}(x) \\ &\quad + \int_{r-z_1}^{r+Q-z_2} \bigg(\int_r^{z^*+x} F_{(L_1,l]}(y-x)\,dy \\ &\qquad + \int_{z^*+x}^{r+Q} F_{(L_1,L]}(y-x)\,dy \bigg)\,dF_{(0,L_1]}(x) \\ &\quad + \int_{r+Q-z_2}^{\infty} \int_r^{r+Q} F_{(L_1,l]}(y-x)\,dy\,dF_{(0,L_1]}(x) \bigg) = \frac{p}{h+p}. \end{aligned} \tag{31}$$

Because $\mathbb{P}(t_v > \mathcal{L}) = F_{(0,\mathcal{L}]}(v)$, the right-hand side of (30) is equivalent to the left-hand side of (31):

$$\mathbb{P}(\mathcal{T}(Q, r) > 0) = \frac{p}{h+p}.$$

Finally, as $\mathbb{P}(\mathcal{T}(Q, r) > 0) = \mathbb{P}(\mathcal{I}(Q, r) > 0)$, the desired result follows. □

Proof of Lemma 4. The first-order condition of r in (17) can be alternatively stated as

$$\begin{aligned}&G(r+Q, D_{(0,L]}) - G(r, D_{(0,L]})\\&+\Bigg[\int_{r+Q-z_2}^{\infty}\big[\{\mu c_2 + G(r+Q-x, D_{(L_1,l]}) - G(r+Q-x, D_{(L_1,L]})\}\\&\quad-\{\mu c_2 + G(r-x, D_{(L_1,l]}) - G(r-x, D_{(L_1,L]})\}\big]\,dF_{(0,L_1]}(x)\\&\quad-\int_{r-z_1}^{r+Q-z_2}\{\mu c_2 + G(r-x, D_{(L_1,l]})\\&\qquad\qquad-G(r-x, D_{(L_1,L]})\}\,dF_{(0,L_1]}(x)\Bigg]=0.\end{aligned}$$

Consider the expression inside the outer square brackets: The integrand of the first integral is always nonnegative (Lemma 2(a)) and so is the integral. The second integral vanishes for the cases $(z_1, z_2) = (z_p(Q), z_p(Q)+Q)$, $(z_1, z_2) \to (-\infty, -\infty)$ and $(z_1, z_2) \to (\infty, \infty)$. Finally, when $(z_1, z_2) = (\underline{z}, \bar{z})$, the integrand in the second integral is nonpositive between the limits of integration (Lemma 2(a) and the definitions of $\underline{z}$ and $\bar{z}$). It follows that the whole expression is always nonnegative. Thus, $r(Q)$ satisfies $G(r+Q, D_{(0,L]}) - G(r, D_{(0,L]}) \le 0$, where the expression on the left-hand side is nondecreasing in r and is equal to 0 when $r = r_s(Q)$, implying $r_s(Q) \ge r(Q)$. Similarly, it can be shown that $r(Q)$ satisfies $G(r+Q, D_{(0,l]}) - G(r, D_{(0,l]}) \ge 0$, which implies $r_f(Q) \le r(Q)$. □

Proof of Proposition 4. Given that $r(Q)$ is continuous in Q, $C(Q)$ is continuous and differentiable in Q. As $\lim_{Q\downarrow 0} C(Q) \to \infty$ and $\lim_{Q\uparrow\infty} C(Q) \to \infty$, the first-order condition is a necessary condition for the optimal order quantity. The first derivative of $C(Q)$ with respect to Q is

$$\frac{dC(Q)}{dQ} = -\frac{1}{Q^2}(\mu\mathcal{K}(Q) - A(Q)),$$

which implies the optimality condition in (20). □

Proof of Lemma 5. Two alternative expressions for $\tilde{G}(y, z^*)$ are

$$\begin{aligned}\tilde{G}(y, z^*) = G(y, D_{(0,L]}) - \int_{y-z^*}^{\infty}&(G(y-x, D_{(L_1,L]}) - \mu c_2\\&-G(y-x, D_{(L_1,l]}))\,dF_{(0,L_1]}(x) \quad\text{and}\end{aligned}\tag{32}$$

$$\begin{aligned}\tilde{G}(y, z^*) = \mu c_2 &+ G(y, D_{(0,l]})\\&-\int_{-\infty}^{y-z^*}(\mu c_2 + G(y-x, D_{(L_1,l]})\\&\qquad-G(y-x, D_{(L_1,L]}))\,dF_{(0,L_1]}(x).\end{aligned}\tag{33}$$

Because the two integrals on the right-hand sides of (32) and (33) are nonnegative (Lemma 2(a) and the definition of z^* in (10)), (a) follows. The two limits of (b) follow from letting $y \to -\infty$ in (33) and $y \to \infty$ in (32), respectively. Taking the derivative with respect to y of both sides of (32) and (33), and noting that $D_{(L_1,L]}$ is stochastically greater than $D_{(L_1,l]}$ then gives (c). □

Proof of Lemma 6. From Remark 7, if $r(Q)$ satisfies $y_s^0 - Q < r(Q) < y_f^0$, then it is the unique solution to $\tilde{G}(r(Q), z^*) = \tilde{G}(r(Q)+Q, z^*)$. Let $\tilde{G}'(y, z^*) = \partial\tilde{G}(y, z^*)/\partial y$, then taking the first derivative with respect to Q of this first-order condition, we get

$$\frac{dr(Q)}{dQ} = \frac{\tilde{G}'(r(Q)+Q, z^*)}{\tilde{G}'(r(Q), z^*) - \tilde{G}'(r(Q)+Q, z^*)}.$$

Because $y_s^0 - Q < r(Q) < y_f^0$, $\tilde{G}'(r(Q), z^*) < 0$ and $\tilde{G}'(r(Q)+Q, z^*) > 0$. Hence, the derivative is well defined and bounded in (0,1), proving parts (a) and (b). Parts (c) and (d) follow from the definition of $A(Q)$ and the first-order condition of Q in Proposition 4 and are analogous to Lemma 6 of Zheng (1992). □

Proof of Lemma 7. The inequalities $C(Q) \le \min\{C_s(Q), C_f(Q)\}$ and $C^* \le \min\{C^s, C^f\}$ follow immediately from noting that the use of only one of the freight modes is always also a feasible dual-freight policy. Comparing the use of only regular freight and the optimal dual-freight policy, the maximum cost saving incurred on a unit (for which demand would otherwise be back ordered) and caused by the availability of express freight is $p(L_2 - l_2) - c_2$. When the reorder point is chosen optimally, in the long term the demands are back ordered at the rate $\mu h/(h+p)$. Thus, the maximum reduction in the inventory holding and penalty cost over the use of only regular freight, if the freight modes are optimally used, is δ_s per unit of demand. Similarly, the maximum reduction in inventory holding and penalty costs over the use of only express freight, if freight modes are used optimally (delaying a part of replenishment and thus saving on holding and shipping cost), is δ_f per unit of demand. This proves the inequalities in (a). Considering the cost functions in (a) with their respective optimal order quantities gives (b). □

Proof of Proposition 5. In (22), better information (in case of Δ_2) results in a lower minimum value of the objective function, implying part (a). Part (b) follows by minimizing the cost functions and using part (a). The cost rate functions $\mathcal{G}_{\Delta_1}(y)$ and $\mathcal{G}_{\Delta_2}(y)$ approach each other asymptotically, and the latter has a smaller value of $\min_r\{\int_r^{r+Q}\mathcal{G}_\Delta(y)\}$ (from part (a)) for the same Q. Thus, it follows that $\mathcal{G}_{\Delta_2}(y)$ has a larger area enclosed between the cost rate function and the horizontal chord connecting the cost rate function at points r and $r+Q$ than $\mathcal{G}_{\Delta_1}(y)$ has. As the optimal r and Q are characterized by this area becoming equal to μK_1, the optimal Q is smaller for $\mathcal{G}_{\Delta_2}(y)$, or $Q^*_{\Delta_1} \ge Q^*_{\Delta_2}$, which proves part (c). □

References

Axsäter, S. 1990. Simple solution procedure for a class of two-echelon inventory problems. *Oper. Res.* **38** 64–69.

Barankin, E. W. 1961. A delivery-lag inventory model with an emergency provision (the single-period case). *Naval Res. Logist. Quart.* **8** 285–311.

Bookbinder, J. H., M. Çakanyildirim. 1999. Random lead times and expedited orders in (Q, r) inventory systems. *Eur. J. Oper. Res.* **115** 300–313.

Bowman, R. J. 1994. Cheaper by air? *World Trade* **7**(October) 88–90.

Çakanyildirim, M., S. Luo. 2005. (R, Q) policy with lead time option. Working paper, School of Management, University of Texas, Dallas.

Chiang, C., W. Chiang. 1996. Reducing inventory costs by order splitting in the sole source environment. *J. Oper. Res. Soc.* **47** 446–456.

Clark, A., H. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management Sci.* **6** 475–490.

Daniel, K. H. 1962. A delivery-lag inventory model with emergency. H. E. Scarf, D. M. Gilford, M. W. Shelly, eds. *Multistage Inventory Models and Techniques*. Stanford University Press, Stanford, CA, 32–46.

Enslow, B. 2006. Best practices in international logistics. Research study, Aberdeen Group, Boston.

Fukuda, Y. 1964. Optimal policies for the inventory problem with negotiable leadtime. *Management Sci.* **10** 690–708.

Gallego, G. 1998. New bounds and heuristics for (Q, r) policies. *Management Sci.* **44** 219–233.

Gallego, G., Y. Jin, A. Muriel, G. Zhang, T. Yildiz. 2007. Optimal ordering policies with convertible lead-times. *Eur. J. Oper. Res.* **176** 892–910.

Groenevelt, H., N. Rudi. 2003. A base stock inventory model with possibility of rushing part of order. Working paper, Simon School of Business, University of Rochester, Rochester, NY.

Huggins, E. L., T. L. Olsen. 2003. Supply chain management with guaranteed delivery. *Management Sci.* **49** 1154–1167.

Jain, A., H. Groenevelt, N. Rudi. 2006. Periodic review inventory management with contingent use of two freight modes with fixed costs. Working paper, Simon School of Business, University of Rochester, Rochester, NY.

Krause, K. S. 2000. FedEx launches new critical service. *Traffic World* **263**(July) 44–45.

Lawson, D. G., E. L. Porteus. 2000. Multi-stage inventory management with expediting. *Oper. Res.* **48** 878–893.

Minner, S. 2003. Multiple-supplier inventory models in supply chain management: A review. *Internat. J. Production Econom.* **81–82** 265–279.

Moinzadeh, K., H. L. Lee. 1989. Approximate order quantities and reorder points for inventory systems where orders arrive in two shipments. *Oper. Res.* **37** 277–287.

Moinzadeh, K., S. Nahmias. 1988. A continuous review model for an inventory system with two supply modes. *Management Sci.* **34** 761–773.

Moinzadeh, K., C. P. Schmidt. 1991. An $(S-1, S)$ inventory system with emergency orders. *Oper. Res.* **39** 308–321.

Muharremoğlu, A., J. Tsitsiklis. 2003. Dynamic leadtime management in supply chains. Working paper, Massachusetts Institute of Technology, Cambridge.

Muharremoğlu, A., J. Tsitsiklis. 2008. A single-unit decomposition approach to multi-echelon inventory systems. *Oper. Res.* **56** 1089–1103.

Neuts, M. F. 1964. An inventory model with an optional time lag. *J. Soc. Indust. Appl. Math.* **12** 179–185.

Sheopuri, A., G. Janakiraman, S. Seshadri. 2007. New policies for stochastic inventory control problem with two supply sources. Working paper, New York University.

Sowinski, L. L. 2004. Air cargo saves the bacon. *World Trade* **17**(September) 18–20.

Stanley, B. 2004. Express delivery firms expand their logistic services in China. *Wall Street J.* (*Eastern Edition*) (November 29) A16.

Tagaras, G., D. Vlachos. 2001. A periodic review inventory system with emergency replenishments. *Management Sci.* **47** 415–429.

Thomas, D. J., J. E. Tyworth. 2006. Pooling lead-time risk by order splitting: A critical review. *Transportation Res. Part E* **42** 245–257.

Veeraraghavan, S., A. Scheller-Wolf. 2008. Now or later: A simple policy for effective dual sourcing in capacitated systems. *Oper. Res.* **56** 850–864.

Whittemore, A. S., S. C. Saunders. 1977. Optimal inventory under stochastic demand with two supply options. *SIAM J. Appl. Math.* **32** 293–305.

Zheng, Y. S. 1992. On properties of stochastic inventory systems. *Management Sci.* **38** 87–103.