Agility and Proximity Considerations in Supply Chain Design

Michael K. Lim

Department of Business Administration, University of Illinois, Urbana-Champaign, IL 61820, mlim@illinois.edu,

Ho-Yin Mak

Saïd Business School, University of Oxford, ho-yin.mak@sbs.ox.ac.uk

Zuo-Jun Max Shen

Department of Industrial Engineering and Operations Research, University of California, Berkeley, shen@ieor.berkeley.edu

Strategic supply chain design decisions are critical to the long-term success of a business. Traditional facility location models for supply chain design focus on the trade-offs between the costs and benefits of proximity, i.e., the distance between facilities and customers. These strategic-focused models do not consider the supply chain’s agility, i.e., its ability to quickly respond to unexpected fluctuations in customer needs. In this paper, we study the problem of designing a supply chain distribution network under demand uncertainty and analyze how the optimal design characteristics of proximity and agility depend on various input parameters. We are able to draw managerial insights on how agility considerations may invalidate well-established and widely-accepted qualitative results derived from traditional models. In particular, we show that it is optimal to increase the density of DCs when the shortage penalty cost increases, and to decrease the density of DCs when a certain unit transportation cost parameter increases. Through these findings, our work conveys the message that traditional, proximity-based facility location models can be inadequate for designing modern responsive supply chains, and calls for the need to develop a new class of models for the task.

Key words: Supply Chain Design, Facility Location, Inventory Sharing

History: This paper is accepted for publication in Management Science.

1. Introduction

The flexibility to respond to customer orders by dynamically adjusting network operations is crucial to supply chain success. IBM utilizes a “neighborhood” stocking strategy, in which an order may be allocated to any stocking location within a certain radius of the customer location, for its service parts logistics. With the aid of sophisticated information and optimization technology, this strategy has led to savings of approximately \$5 million per year, while maintaining or improving customer service (Gresh and Kelton [23]). Similar arrangements are commonly observed in the online retail industry. Amazon dynamically fulfills orders from different distribution centers (DCs) subject to

available stocking levels. This centralized order allocation process enables them to lower the logistics cost significantly (Xu *et al.* [43]). Dell also utilizes supply routing through lateral transshipments between facilities and diversion of in-transit inventory so as to quickly respond to its evolving customer demand (Foreman *et al.* [21]). In light of the increasing adoption of responsive inventory sharing arrangements, which contributes to the paradigm shift from efficient supply chains to *agile* supply chains, we study the implications of agility in designing modern supply chain network.

The agility of a supply chain network is the ability that its operations quickly adjust and react to changing environmental factors. In the case where customer demand quantities are uncertain, one crucial element of agility is to dynamically shift inventory among stocking locations to ensure fulfillment of demand with high probability and at low cost. Agility clearly brings benefits to the supply chain. Classical inventory theory (e.g., Eppen [20]) suggests that the benefits from inventory sharing, in the form of holding centralized stocks at a common location, increases when more parties share common stocks. Likewise, it is intuitive that total inventory costs will be lower if the degree of agility (e.g., the flexibility to transship inventory among DCs) is greater.

However, agility comes at a cost. Agility in inventory sharing typically increases the transportation cost since it requires quick response and short lead times. Increasing agility via greater degree of flexibility necessitates higher operational complexities to cope with (e.g., using flexible transportation modes and handling imports and exports for multiple countries). Furthermore, agility may entail redundancy in the system which typically goes against lean supply chain management principles. Therefore, it is important to carefully evaluate the trade-offs between the benefits and costs of agility in supply chain design. Our goal in this paper is to incorporate the agility dimension in addition to the traditionally-emphasized proximity dimension, and study its significance and implications on the optimal supply chain design. Our results shed light on how to design agile supply chains especially for the firms operating large-scale networks such as IBM, Amazon, and Dell.

Central to the theory and practice of supply chain design is the strategic choice of network consolidation (operation of a low density of large facilities) versus deconsolidation (high density of small facilities). The optimal configuration of density, or *proximity* to the market, involves the fundamental trade-off of balancing transportation costs and facility location costs. Most classical facility location models, such as the median, center, and fixed charge problems, are constructed based on this trade-off (Daskin [15]). Shen *et al.* [39] show that the classical models ignore the benefits of pooling inventory at the DCs and are thus biased towards deconsolidation, suggesting the location of a larger number of DCs than is truly optimal. They show that, by jointly considering

the tactical inventory control and strategic facility location decisions, such suboptimality can be reduced, and thereby point out the importance of incorporating inventory pooling considerations into the proximity trade-off. This result is also analytically confirmed by Naseraldin and Herer [32]. In a similar vein, we study how agility considerations may interact with the proximity trade-off. In this paper, we focus on the agility of a firm to meet random demand using responsive inventory sharing mechanisms. By enhancing the traditional proximity-focused decision models to incorporate the agility dimension, we show that the optimal choice over network density can be very different for agile and non-agile supply chains.

We refer to the physical inventory pooling at a single location (i.e., one individual DC) as *first-order* inventory sharing, under which the firm aggregates inventory for regional demands in a *centralized* DC. By pooling the demand, uncertainty in the aggregate level effectively decreases and thus the firm benefits from holding a reduced safety stock while maintaining the same service level. This is the well-known risk-pooling effect of physical centralization. On the negative side, physical centralization increases average shipping cost and response time, because inventory is located farther away from demand sources on average. In addition to first-order inventory sharing, we consider *second-order* inventory sharing, under which the firm employs information pooling by sharing information regarding availability of stocks in nearby DCs and subsequently reallocating orders that cannot be fulfilled by the nearest DC. Such type of second-order inventory sharing is referred to as *dynamic fulfillment* in our paper. In contrast to first-order inventory sharing that is proactive in nature, second-order inventory sharing is a reactive means, since it is only utilized if orders cannot be fulfilled by its primarily assigned DC (thus, served from farther location at extra shipping costs).

First- and second-order inventory sharing arrangements carry important implications on the proximity and agility of the supply chain. Clearly, optimally utilizing both modes of inventory sharing offers potential to better balance between inventory and transportation costs, while maintaining high service level. However, integration of physical centralization and dynamic fulfillment must be carefully studied, as implementation of one mode of inventory sharing will affect that of the other. In this paper, we formulate and analyze a network design model that determines jointly the proximity and agility design characteristics to minimize the overall costs. Following the integrated approach of supply chain design modeling (e.g., Shen *et al.* [39]), our objective function includes strategic, tactical, and operational cost components. This approach is non-myopic as the long-term impacts of the strategic decisions on tactical and operational performances of the supply chain are incorporated in the design objective.

To tackle complex supply chain design problems, operations researchers traditionally employed integer-programming-based facility location models (e.g., Shen *et al.* [39]). Although such approach produces a detailed implementable solution that is exact or close to optimal for a given set of input data, it fails to offer general design guidelines because structural properties of integer programming problems are too difficult to obtain. In contrast, the Continuum Approximation (CA) approach taken in this paper draws qualitative insights by utilizing concise data summaries and analytical models instead of detailed data and numerical algorithms (Daganzo [11]). Using this approach, it is possible to derive closed-form solutions of the given problem and obtain key insights resulting from the main trade-off features. Although our model is not intended for producing directly implementable supply chain design solutions, it is possible to convert solutions from continuous models into implementable solutions using the approach introduced by Ouyang and Daganzo [35]. However, this is beyond the scope of the current paper.

Our analysis leads to several managerial insights, including the following:

1. The design of modern responsive supply chains requires a careful analysis of the interactions between proximity and agility considerations. That is, the conventional proximity-based models do not capture the agility dimension with a widely implemented information pooling, and further research is needed to study modern supply chain designs. Our results point out a *need for refinement* in the conventional insights that are well-established and widely-accepted in facility location and network design practice and literature.
2. When the agility dimension is jointly optimized in supply chain design, we show that it is optimal to locate *more* DCs, each handling a *smaller* demand volume, as the unit shortage penalty cost increases. This result is different from the conventional understanding on the risk-pooling effect considering physical centralization only. The reasoning is that, in the presence of second-order inventory sharing (dynamic fulfillment), this effect offsets the first-order inventory sharing effect (conventional risk-pooling via centralization), resulting in a net force that pushes for smaller DCs to be located closer to customers as the unit penalty cost increases. Correspondingly, these smaller DCs share inventory with a more responsive dynamic fulfillment policy.
3. As the unit transportation cost increases, it can be optimal to deploy fewer facilities, locating facilities farther apart. This is another finding that contrasts with the conventional wisdom on the fundamental relationship in facility location models that higher transportation costs lead to a higher density of facility in the supply chain. The reason is that, as the transportation costs incurred by second-order inventory sharing becomes too large, it is optimal to favor first-order inventory sharing as a substitute strategy. Thus, to enhance risk-pooling benefits, DCs are located farther apart such that each one handles demand coming from a larger influence area.

2. Literature Review

Supply chain design problems have long been studied in the facility location literature, which an excellent overview is provided in Daskin [15], Drezner [18], and Drezner and Hamacher [19]. Due to the complexity of facility location models, supply chain design problems are typically solved using sequential approaches. First, a pure facility location model, such as the uncapacitated facility location model, is solved to determine the network structure in the strategic design phase. Given the network structure, then tactical level decisions such as inventory policy and control decisions are made. Finally, operational level decisions such as vehicle routing and scheduling decisions are made on a day-to-day basis. Under the sequential approach, strategic location decisions are made disregarding much of the future tactical and operational performance of the supply chain, often resulting in suboptimal decisions.

Recent effort in integrating different phases of decisions in supply chain includes Shen *et al.* [39] that study a location problem with inventory being held at facilities. The model captures the risk-pooling effect, since safety stock can be reduced by pooling demand from more retailers (customers) at a single DC. Their computational results show that, if location and inventory decisions are optimized separately, the number of facilities opened will be more than optimal and the cost difference is significant. Shen and Qi [40] further incorporate operational routing costs into the model. For surveys of using a continuum approximation for integrated supply chain design models, readers may refer to Shen [38] and Mak and Shen [29].

Ozsen *et al.* [36] consider the effect of storage capacity at facilities under multiple sourcing. They study a multiple sourcing procedure, under which each customer's orders are split randomly among multiple DCs. One limitation of the randomized multiple sourcing strategy is that inventory availability at DCs is not considered when demand is assigned. To address this issue, Mak [30] formulate a joint facility location and network design model in which the sourcing or fulfillment decisions are made dynamically subject to inventory availability and the sizes of realized demand. This strategy enables second-order inventory sharing among different DCs by switching demand assignments after observing demand. The model captures the trade-off between this benefit and the extra costs and disutility associated with this form of flexibility.

Mak's [30] paper considers both first-order and second-order inventory sharing. The former results from physical centralization of inventory at a single location to prepare for demand variability from different sources. In a classical paper of first-order inventory sharing by Eppen [20], it is shown that, under Gaussian demand distributions, a centralized system requires less safety

stock than does a decentralized system to achieve the same service level. The benefit of pooling decreases with the correlations of demand across locations. Corbett and Rajaram [8] extend the results to general demand distributions. The benefit of risk-pooling decreases with demand dependence, measured by the sum-convex stochastic order.

Second-order inventory sharing provides a trade-off between the benefits of inventory sharing and the resulting increased transportation costs due to centralization. One popular mode of second-order inventory sharing is lateral transshipments. The literature on transshipment operations and the potential benefits is extensive. An early paper on the topic was written by Krishnan and Rao [26]. The authors show that the optimal policy for a multiple-retailer system with identical costs is to order up to an equal fractile at all locations. Robinson [37] studies a more general setting, in which there can be multiple retailers with different cost parameters, and proves that a stationary base stock policy is optimal. Axsäter [3] studies the effect of transshipments on the stocking levels of spare parts supply chains. Dong and Rudi [17] investigate the benefits of individual retailers and the supplier in a transshipment setting. They show that whether and the degree to which retailers benefit from transshipments depend upon whether the supplier is a price taker or price setter. Foreman *et al.* [21] consider the optimization of Dell's inventory control policy, which involves second-order inventory sharing by performing transshipments and rerouting in-transit inventory.

Although the optimization and resulting benefits from transshipments are very well documented, the potential impact of second-order inventory sharing on the *geographical* supply chain network configuration has received less attention. Our goal is to address the relationships between the two modes of inventory sharing and their impact on the supply chain design. When both modes of inventory sharing are possible, there may be complementary or counteracting effects that make it unclear whether the same insights from the previous studies will continue to hold or not. To perform the analysis, we adopt the CA modeling methodology that is better catered for deriving qualitative insights compared to the traditional integer-programming approach.

The technique of CA was first introduced to transportation science literature by Newell [34], and has been applied to several interesting applications in the field. The book by Daganzo [11] explains this modeling methodology in detail and discusses a number of examples in logistics. The book also summarizes many fundamental results regarding the CA technique, such as approximate formulas for the optimal objective values of the vehicle routing problem (Daganzo [10]) and the transportation problem (Daganzo and Smilowitz [13]). Daganzo and Erera [12] study two problems with this approach: a vehicle routing problem with uncertain demand and a location-inventory-routing

problem. Smilowitz and Daganzo [41] formulate an analytical model for designing package distribution networks on a much larger scale than what can be typically handled by integer programming methods.

In the operations management literature, Cachon [5] uses a CA model to study the interactions between greenhouse gas emissions and retail store density. Using this stylized model, he uncovers several interesting results, including that ignoring greenhouse emissions in location planning leads to scenarios with undesirable emission levels, and that carbon taxes may not be an effective means to curb emissions. Dasci and Laporte [14] use similar technique to formulate an analytical model of multi-store competitive facility location. With the CA approach, they are able to study a more generalized problem than was previously done (e.g., allowing the number of facilities to locate to be a decision variable). Recently, Cui *et al.* [9] and Lim *et al.* [28] study location problems under the threat of disruptions.

A pioneer analytical study of integrated supply chain design is done by Naseraldin and Herer [32]. They model a continuous location-inventory problem on a line segment and analytically quantify the benefits of integrated modeling. While this model only considers first-order inventory sharing, Naseraldin and Herer [33] extend the problem to allow transshipments of inventory between facilities. They point out several interesting results, e.g., that it can be optimal to increase the density of facilities when unit holding cost increases. Their results highlight the importance of modeling other forms of inventory sharing than the classical risk-pooling. Our work differs from these previous works in several aspects. First, Naseraldin and Herer's work assumes unit transshipment costs to be independent of distance between DCs; whereas in our model, the cost incurred by second-order inventory sharing, i.e., the extra transportation costs incurred by dynamic fulfillment, can be dependent on the network design (i.e., density of DCs). The distance-independent transshipment cost assumption fits better with cases where large numbers of facilities are densely located, such as a retail network in a densely populated cities. In a setting where facilities (DCs in particular) are distanced far from one another, such a treatment tends to overestimate the benefit of such inventory sharing, because transshipping from far-away DCs is equally costly as from a nearby one. Therefore, for such supply chain networks, the optimal density of DCs is also influenced by the distance over which to perform second-order inventory sharing shipments. In our model, we attempt to capture this effect in the proximity-agility interactions by allowing a distance-dependent cost structure. Second, compared with Naseraldin and Herer's one-dimensional model setting, we study the problem in a more general two-dimensional setting and a different form of second-order inventory sharing. Specifically, in our dynamic fulfillment setting, shipments can be directly made from

different DCs to demand points subject to stock availability. This differs from the transshipment setting, in which items are first shipped from one DC to another, and then delivered to customers. Under the triangle inequality, the item always travels a longer distance, and thus incurs higher cost or longer response time, in the latter setting.

3. Model Development

3.1. Problem Description

We consider a supply chain design problem described as follows. A firm wishes to serve a number of geographically-dispersed customer demand points in a large service region. To do so, the firm will decide to locate a number of DCs at which inventory is held to satisfy stochastic demand of customers. We say that a customer is in the *primary influence area* of the DC that is nearest to it and refer to that DC as the *primary* DC of the customer. Furthermore, a set of DCs form an *inventory sharing group* within which other DCs with excess supply can satisfy customer demands via dynamic fulfillment in case the primary DC faces a supply shortage. We refer to the DCs in the same inventory sharing group as the *secondary* DCs of the customer. We assume that the inventory sharing groups are non-overlapping and identically sized, which collectively exhaust the entire service region.

Our primary focus is on the firm's strategic level decision in determining the number of DCs to locate along with the size of the information sharing group. In our problem, this decision encompasses both proximity and agility considerations. As discussed in Section 1, the fundamental trade-off regarding proximity is one of balancing transportation costs (which encourages deconsolidation) with fixed facility costs and first-order inventory sharing considerations (which encourage consolidation). However, as one introduces the agility dimension, it is important to also factor in the impact of second-order inventory sharing in supply chain design. In our model, the possibility of dynamic fulfillment alters the proximity trade-off, because as more DCs are located closer together, dynamic fulfillment can be done more efficiently at a lower cost, and thus the network can be made more agile.

As is typically assumed in multi-location inventory systems, the DCs replenish inventory periodically. The periodic review policy facilitates synchronization of replenishment and fulfillment decisions. Under some extra conditions (e.g., there is no fixed replenishment cost, and any shortage is filled in the subsequent period), it can be shown that a stationary base stock policy is optimal in this setting (Mak [30]). Therefore, we may consider the simplified case where inventory is managed for one single period since the expected cost of each additional period will be equivalent to that of the first period. Similar treatments are utilized by Lien *et al.* [27] and Mak [30]. We

only assume synchronized orders among DCs within the same inventory sharing groups (and the customer demands they cover), but not necessarily those that belong in different groups. Note that it is often difficult in practice to synchronize replenishment orders of facilities that are separated by long geographical distances and physical or virtual boundaries (e.g., DCs located in different countries with different time zones and work days). For a large-scale supply chain network consisting of thousands of geographically dispersed demand points, it is very costly, if not impossible, to collect all demand information in real time and optimize all operations in a centralized manner. In such a case, the arrangement of forming local coalitions to share inventory among geographically neighboring DCs (e.g., within the state of California) is a reasonable, albeit theoretically suboptimal, strategy.

Following the spirit of CA approach, we consider the following setup. Demand points are evenly spread (with a homogeneous density δ points per square mile) on a two-dimensional infinite homogeneous plane. Each demand point faces a demand per period that follows an independent and identical Normal distribution with mean μ and standard deviation σ . For tractability, we assume that the DCs are located X miles apart, at grid points of the service region (see Figure 1). Each DC has a square-shaped primary influence area of X^2 square miles, in which the aggregate demand follows a Normal distribution with mean $\delta X^2 \mu$ and standard deviation $\sqrt{\delta} X \sigma$.

We assume that every M^2 neighboring DCs form a square-shaped inventory sharing group (where $M \geq 2$; otherwise, second-order inventory sharing is not utilized when $M = 1$), covering a combined influence area of $M^2 X^2$, in which the aggregate demand follows a Normal distribution with mean $\delta M^2 X^2 \mu$ and standard deviation $\sqrt{\delta} M X \sigma$. Intuitively, a square-shaped configuration yields lower transportation costs than does an elongated configuration with the same number of DCs. In this paper, as typically done for analytical logistics models, our analysis makes use of the fact that the service region in question is large relative to the primary influence area of each DC, such that we may formulate tractable models by considering densities instead of exact DC and customer locations. Although our model assumes $M \geq 2$, ideally, we focus our attention to cases in which inventory sharing groups consist of rather large numbers (e.g., in the order of dozens) of DCs, such that the agility dimension carries significant impact. It is then possible to form inventory sharing groups that are almost square-shaped, even if the number of DCs in the group is not a perfect square. Despite differences in model setting, Lim *et al.* [28] provides a validation of this assumption and show that the optimality gap is typically very small and thus the chosen approach offers a good approximation in terms of the total cost.

Figure 1 Segment of Service Region with Square-shaped Primary Influence Areas and Inventory Sharing Groups

The system, after the design phase, operates for a single period as follows. At the beginning of the period, before observing the demand realizations, each DC orders up to an inventory amount equal to the mean demand of its primary influence area (i.e., $\delta X^2 \mu$) plus a safety stock. The safety stock is equal to a factor Z_1 (> 0) times the standard deviation of demand in the primary influence area ($\sqrt{\delta} \sigma X$), where Z_1 is a decision variable. Each DC incurs a holding cost h per unit of safety stock held for the period. Then, an inventory sharing group will have a total safety stock of $M^2 Z_1 \sqrt{\delta} \sigma X$, which is equal to $Z_2 = M Z_1$ times the standard deviation of the aggregate demand within the group. Therefore, we can use decision variables Z_1 and $Z_2 = M Z_1$ to control the service levels within a primary influence area, in which first-order inventory sharing is performed, and within an inventory sharing group, in which second-order inventory sharing is carried out. Throughout the paper, we refer to Z_1 and Z_2 as the *local* and *group safety stock factors*, respectively. Note that our assumption of equal stocking levels for all DCs is not necessarily optimal even for our symmetric setup, because it is intuitively better to stock less at DCs near the boundary of an inventory sharing group as they are surrounded by fewer demand points. However, in a realistic setting, determining the optimal inventory levels at individual DCs requires solving a stochastic program with recourse and is not trivial. Moreover, if the sharing group consists of a large number of DCs, the boundary effect carries a relatively small impact on the entire group. Therefore, we opt to make a simplistic assumption of equal safety stock level that allows the model to retain tractability along with the basic trade-offs of the problem.

Under the dynamic fulfillment arrangement, shipment quantities from DCs within an inventory sharing group to demand points within their influence areas are optimized as recourse decisions, after the realization of demand. The part of demand that cannot be met with inventory within the inventory sharing group is assumed to be backordered, incurring a penalty of p per unit (in addition to shipping cost at the beginning of the next period). We assume that $p > 2h$, such that safety stock will be non-negative in the optimal solution. The transportation cost for a DC-customer

shipment is proportional to both the number of units shipped and the distance between the DC and the customer.

We summarize the notation below.

Unit Cost Parameters

f_1 = Fixed facility cost of locating a DC (dollars);

f_2 = Variable facility cost proportional to average demand handled at a DC (dollars per item);

κ = Discount factor, such that the amortized fixed cost per period is κf_1 (dimensionless);

g_1 = Fixed operations cost at a DC (dollars);

g_2 = Variable operations cost per unit of average demand handled at a DC (dollars per item);

h = Unit holding cost of safety stock (dollars per item);

p = Penalty cost per unit of shortage (dollars per item);

u, v = Unit transportation cost parameters to be defined below (dollars per item per mile)

Demand and Other Parameters

δ = Density of customers (points per square mile);

μ = Mean customer demand at each point (items per period);

σ = Standard deviation of customer demand (items per period)

Decision Variables

X = Separation between adjacent DCs (miles);

Z_1 = Local safety stock factor for a DC (dimensionless);

Z_2 = Overall safety stock factor for an inventory sharing group (dimensionless)

Recall that given $Z_2 = MZ_1$, the inventory sharing group size is given by $M^2 = (Z_2/Z_1)^2$.

Next, we develop the objective function for the supply chain design problem. Since the total number of customers in the entire service region is fixed, all cost functions will be expressed as the expected costs per customer demand point.

3.2. DC Location and Operation Costs

We consider the costs of both locating and operating DCs. In particular, the lifetime investment cost f_1 is amortized with the discount rate κ to obtain the per-period fixed cost. Also, we consider a per-period fixed operation cost g_1 , which may include overhead and fixed in-bound shipping (ordering) and handling costs. In addition to the fixed costs, a variable facility cost f_2 that is linear in the average demand volume handled by the facility (which may include the rental cost for storage space), and a variable component of operations and in-bound logistics cost proportional to the average unit handled at a DC, denoted as g_2 , are considered. In the long run, each DC serves an average demand volume equal to the aggregate mean demand of customers located within

its primary influence area. Since there are δX^2 demand points in the influence area, the average demand volume handled is equal to $\delta X^2 \mu$. Hence, we have:

$$\begin{aligned} \text{Cost of locating and operating one DC} &= \kappa f_1 + g_1 + (f_2 + g_2) \delta X^2 \mu, \\ \text{Facility cost per customer demand point} &= \frac{(\kappa f_1 + g_1)}{\delta X^2} + (f_2 + g_2) \mu. \end{aligned}$$

3.3. Inventory Costs

The safety stock factors Z_1 and Z_2 (and effectively, the inventory sharing group size M^2) are determined to minimize the ordering and holding costs for inventory (safety stock), penalty costs for backorders, and extra transportation costs incurred due to fulfillment of orders from secondary DCs. We will first discuss the modeling of the first two components, and the last will be discussed in the next subsection with the distribution-related cost components.

First, each DC holds a quantity of safety stock equal to Z_1 times the standard deviation of the demand in its primary influence area, incurring a holding cost which includes both the cost of capital and physical capacity (rental) cost associated with storage:

$$\begin{aligned} \text{Cost for holding safety stock at one DC} &= h Z_1 X \sqrt{\delta} \sigma, \\ \text{Holding cost per customer demand point} &= \frac{h \sigma Z_1}{\sqrt{\delta} X}. \end{aligned} \tag{1}$$

Furthermore, backorder penalty is incurred when there is insufficient stock within the inventory sharing group to meet demand. Denoting the total demand within the inventory sharing group by D_2 , the expected shortage will be given by $E[D_2 - M^2 \delta X^2 \mu - Z_2 \sqrt{\delta} M \sigma X]^+$, where $\{\cdot\}^+ = \max\{\cdot, 0\}$. Note that the expected shortage is given by $L(Z_2) M \sqrt{\delta} \sigma X$, where $L(z)$ is the standard Normal loss function $L(z) = \phi(z) - z \bar{\Phi}(z)$; and $\bar{\Phi}(\cdot)$ and $\phi(\cdot)$ are the complementary cumulative distribution function and density function of the standard Normal distribution, respectively. Hence, we have:

$$\begin{aligned} \text{Shortage penalty for one inventory sharing group} &= p L(Z_2) M \sqrt{\delta} \sigma X, \\ \text{Shortage penalty per customer demand point} &= \frac{p \sigma L(Z_2)}{\sqrt{\delta} X M} = \frac{p \sigma Z_1}{\sqrt{\delta} X} \frac{L(Z_2)}{Z_2}. \end{aligned} \tag{2}$$

3.4. Distribution Costs

The distribution-related decisions involve optimizing the fulfillment of realized customer demand using inventory of DCs within an inventory sharing group. In particular, after customer demand is realized, the distribution costs can be minimized by solving a linear program, i.e., the transportation problem. Unfortunately, despite its computational tractability, such linear optimization formulation

is not analytically tractable, making it very challenging to obtain managerial insights directly. Hence, in this subsection, we develop an approximate formula for the expected shipping distance incurred (in item-miles) from DCs to customers.

Our approximation is based on the *dimensional analysis* technique widely used in physics and related engineering disciplines for developing mathematical models to characterize complex physical systems. The fundamental idea of dimensional analysis is that meaningful relationships between dependent and independent variables for a problem should be *dimensionally homogeneous*, i.e., they should remain to hold when the variables are redefined in different units of measurement. As an example for illustration, consider the EOQ problem. We consider three independent variables (parameters) of fixed replenishment cost K , holding cost rate h and demand rate D . Further, we have the rate of total (holding and replenishment) cost C , under the optimal replenishment policy, as a dependent variable.

For the EOQ model, it is well known that the relationship $C \equiv \sqrt{2KDh}$ holds. Furthermore, this relationship is dimensionally homogeneous, i.e., it remains invariant if costs are measured in Euros instead of U.S. Dollars, or if time is measured in weeks instead of days. For example, measuring time in weeks instead of days would cause C, D and h to be scaled up by a factor of 7 each, and the identity remains unchanged. Specifically, dimensional homogeneity holds because the expression $\frac{C^2}{KDh}$ is dimensionless. This idea is formalized in Buckingham's celebrated π -Theorem (e.g., Bridgman [4]):

THEOREM 1 (Buckingham's π -Theorem). *Consider a problem that can be described in (dependent and independent) variables $q_1, \dots, q_n$, where a dimensionally homogeneous relationship in the form $f(q_1, \dots, q_n) \equiv 0$ holds. Then, the equation can be restated as $F(\pi_1, \dots, \pi_p) \equiv 0$, where $\pi_k, k = 1, \dots, p$ are dimensionless parameters in the form $\pi_k = q_1^{\alpha_{1k}} q_2^{\alpha_{2k}} \dots q_n^{\alpha_{nk}}$.*

To apply the π -Theorem, the first step is to identify the dimensionless parameters π_i , also known as the π -groups. Since they are products of powers of variables q_j , it is convenient to consider the *dimensional matrix* $\mathbf{A}$, whose rows correspond to dimensions, columns correspond to variables, and the (i, j) -th entry corresponds to the exponent of the i -th dimension in the j -th parameter. For the EOQ case, the three dimensions are items, time, and dollars, and the four variables are K, h, D , and C . Hence, we obtain the following dimensional matrix:

$$\mathbf{A} = \begin{bmatrix} 0 & -1 & 1 & 0 \\ 0 & -1 & -1 & -1 \\ 1 & 1 & 0 & 1 \end{bmatrix}.$$

For example, the second parameter h has dimensions dollars per item per unit time; hence the second column has entries of -1 (corresponding to items), -1 (time) and 1 (dollars). With the dimensional matrix, one can see that the π -groups can be formed by $\pi_k = q_1^{\alpha_{1k}} q_2^{\alpha_{2k}} \cdots q_n^{\alpha_{nk}}$, where $\alpha_k = (\alpha_{1k}, \dots, \alpha_{nk})'$ are solutions to the equation $\mathbf{A}\alpha = \mathbf{0}$. From the rank-nullity theorem, the number of linearly independent solutions to this equation is $n - \text{rank}(\mathbf{A})$. A set of linearly independent solutions corresponds to a collection of π -groups that are multiplicatively independent, i.e., none of which can be obtained as products of powers of others. For the EOQ case, we have $n = 4$ and $\text{rank}(\mathbf{A}) = 3$, and therefore only one multiplicatively independent π -group, $\frac{C^2}{KDh}$, can be obtained. Then, by the π -Theorem, there exists a function F such that $F\left(\frac{C^2}{KDh}\right) \equiv 0$, i.e., $\frac{C^2}{KDh} \equiv \gamma$ for some constant γ . This implies $C \equiv \sqrt{\gamma KDh}$, which reduces to the EOQ result when $\gamma = 2$.

Note from the EOQ example that dimensional analysis can help identify functional forms that relate the dependent and independent variables of problems, without using specific details of the problems. Therefore, in science and engineering domains, dimensional analysis is typically utilized to guide the construction of mathematical models to be calibrated with experimental data (e.g., [4]). This technique helps the modeler to restrict focus to certain functional forms and reduce the number of parameters to be estimated, and is particularly useful when the problem can be concisely represented using a small number of variables relative to the number of dimensions (such that the number of dimensionless groups is small).

There has been some recent work in transportation science literature that use dimensional analysis. Daganzo [11] derives an approximate formula for the traveling salesman problem on a homogeneous plane. Castillo *et al.* [6] use the same technique to investigate the validity of various traffic flow models proposed in the literature. Daganzo and Smilowitz [13] also use dimensional analysis to obtain asymptotic bounds to some transportation problems. They consider a setting in which supply points are located on a plane with a given homogeneous density, and are associated with zero-mean, Normally-distributed net supplies. This setting arises in problems of equipment (e.g., rail cars) repositioning in logistics. They show that, as the number of points in the plane, denoted by N , approaches infinity, the shipping distance asymptotically increases in the order of $N \log_2 N$. With the $N \log_2 N$ functional form, they fit an approximate formula for the optimal cost of the repositioning problem. In what follows, we obtain an approximate formula using similar techniques for the distribution costs in our supply chain problem.

We consider the following transportation problem. For tractability, we divide the primary influence area of each DC into K identical square subzones. We then approximate the problem by

Figure 2 Example of Aggregation of Customer Demand at Local Subzones, with $K = 4$

aggregating demand of all customers in each subzone to its central point. An example of demand aggregation of $K = 4$ subzones are illustrated in Figure 2.

Next, consider an inventory sharing group with M^2 DCs (indexed by j) and KM^2 aggregate demand points (indexed by i). The demand represented by each demand point follows a Normal distribution with mean $\hat{\mu} = \frac{\delta X^2}{K}\mu$ and standard deviation $\hat{\sigma} = \frac{\sqrt{\delta}X}{\sqrt{K}}\sigma$. Let the realization of demand at point i be denoted by e_i , while the inventory of DC j be denoted by y_j . With local safety stock factor Z_1 , each DC carries an equal amount of inventory $y_j = \delta X^2\mu + Z_1X\sqrt{\delta}\sigma$. Let d_{ij} be the ℓ_2 distance between i and j . We define two decision variables, w_{ij} and b_j , the shipment quantity from j to i and the backorder quantity at DC j , respectively. Finally, we obtain the following problem.

$$\begin{aligned}
 & \min_{w,b} \sum_{j=1}^{M^2} \left[\sum_{i=1}^{KM^2} d_{ij} w_{ij} + \bar{d} b_j \right] \\
 & \text{subject to: } \sum_{i=1}^{KM^2} w_{ij} \leq y_j + b_j, \text{ for } j = 1, \dots, M^2 \\
 & \sum_{j=1}^{M^2} w_{ij} \geq e_i, \text{ for } i = 1, \dots, KM^2 \\
 & w_{ij} \geq 0, \text{ for } i, j = 1, \dots, M^2.
 \end{aligned}$$

In the above, $\bar{d}$ is set to a sufficiently large value to ensure demand is fulfilled from some DC, instead of backordered, whenever there is available stock. We note that the quantity of our interest is the actual total transportation distance (in item-miles), which excludes the backorder penalty. Let $T = \sum_{i=1}^{KM^2} \sum_{j=1}^{M^2} d_{ij} w_{ij}^*$ denote this quantity, where w^* is the optimal solution to the above linear program. We then define the average expected transportation cost per DC, $\bar{T} = \frac{E[T]}{M^2}$, taken over the distribution of demand. Note that $\bar{T}$ has dimension item-miles.

To take advantage of the π -Theorem, we first identify the π -groups. Note that the two dimensions involved in the problem are items and miles. Next, the problem can be defined by six variables: the

independent variables $\hat{\mu}$ (in items), $\hat{\sigma}$ (in items), X (in miles), Z_1 (dimensionless), M (dimensionless), and the dependent variable $\bar{T}$ (in item-miles). Hence, we obtain the following dimensional matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

with $\text{rank}(\mathbf{A}) = 2$. Therefore, one can form four independent π -groups: $\frac{\bar{T}}{\hat{\mu}X}$, M , Z_1 , and $\frac{\hat{\sigma}}{\hat{\mu}}$. By the π -theorem, the relationship between the π -groups can be written in the form of:

$$\frac{\bar{T}}{\hat{\mu}X} = F\left(M, Z_1, \frac{\hat{\sigma}}{\hat{\mu}}\right), \text{ or equivalently } \bar{T} = \hat{\mu}X F\left(M, Z_1, \frac{\hat{\sigma}}{\hat{\mu}}\right).$$

In the above, $F(\cdot)$ is an unspecified function, which we identify via an extensive simulation experiment. This is a standard model identification approach with dimensional analysis [4]. Details of the simulation experiment is provided in Appendix B. For each of the 4,675 combinations of parameter values, we simulated for 30 demand realizations and took sample averages of optimal costs of the corresponding transportation problems. Of all parameter combinations, we choose a training set by including each combination with a probability of 0.5, and the remaining combinations constitute the validation set. Regression analysis on the training set shows that the following function is an excellent fit over the range of parameters tested, with r -squared value over 0.999:

$$\bar{T} = 1.427\hat{\mu}X + 0.329\hat{\sigma}L(Z_1)X \log_2 M. \quad (3)$$

The choice of this functional form is supported by the following factors. First, from Daganzo and Smilowitz [13], the expected cost of a transportation problem with random demand and supplies increases in $N \log_2 N$ as N , the number of demand/supply points, goes to infinity. This suggests that the expected transportation cost of unmet demand within an inventory sharing group with M^2 DCs increases in $M^2 \log_2 M^2$ (i.e., the cost per DC increases in $\log_2 M$). Second, the quantity of such shipments to be made is proportional to $\hat{\sigma}L(Z_1)$. Hence, the total expected transportation cost (per DC) to fill unmet demand within a group is proportional to $\hat{\sigma}L(Z_1) \log_2 M$.

We plot the simulated versus approximated values of $\bar{T}$ for the validation set in Figure 3. The average absolute percentage error of this approximation (vs. the simulation result) is only 1.6%, and the maximum error is 14.6%. The absolute error is below 5% in 92% of the validation instances. Further, the error is positive and negative for 56 and 44% of cases, respectively, suggesting that the approximation is reasonably unbiased. Therefore, for the remainder of the paper, we use the approximation:

$$\text{Shipping distance per inventory sharing group} = k_1 \delta \mu X^3 + k_2 \sqrt{\delta} \sigma X^2 L(Z_1) \log_2 M,$$

Figure 3 Actual versus Fitted Values for Shipping Distances in Simulation Study

where $k_1 = 0.357$ and $k_2 = 0.082$ are dimensionless constants. We observe that $k_1 X$ is equal to the distance between a demand aggregation point and its primary DC. Therefore, shipping distance can be interpreted as the expected demand times the distance from the primary DC (the first term) plus the extra distance incurred due to shortage at the primary DC. The latter term is proportional to the expected excess demand that cannot be served at the primary DC ($\sqrt{\delta}\sigma XL(Z_1)$), the distance between DCs (X), and the logarithm of the inventory sharing group size. Therefore, as the inventory sharing group size increases, the extra distance incurred increases due to the possibility of shipping from a DC farther away. However, the marginal rate of increase is decreasing (in the order of $\log_2 M$) because the likelihood of shipping from the farthest DC diminishes as the group size expands. The two-part structure of the cost expression also allows for imposing an extra penalty cost on the additional shipping distance beyond that from the primary DC, by placing a larger weight on the second term. Such penalty may arise from, for example, inconvenience costs from extra administrative and communication procedures, and any expedited shipment costs to make up for additional lead time. To reflect this possibility, we use u and v ($\geq u$) as the unit costs per item-mile for first and second terms, respectively. Dividing the total cost by $M^2\delta X^2$ customers in the group, the expected transportation costs per demand point are:

$$\text{Transportation costs per customer demand point} = u\mu k_1 X + \frac{vk_2\sigma}{\sqrt{\delta}} L(Z_1) \log_2(Z_2/Z_1).$$

3.5. Total Cost for the Supply Chain Design Problem

Combining all cost components derived in the previous subsections, the supply chain design problem of our interest can be formulated as follows:

$$\min_{X, Z_1, Z_2} \mathfrak{C}(X, Z_1, Z_2)$$

$$= \min_{X, Z_1, Z_2} \frac{(\kappa f_1 + g_1)}{\delta X^2} + (f_2 + g_2)\mu + \frac{h\sigma Z_1}{\sqrt{\delta}X} + \frac{p\sigma Z_1}{\sqrt{\delta}X} \frac{L(Z_2)}{Z_2} + u\mu k_1 X + \frac{vk_2\sigma}{\sqrt{\delta}} L(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right). \quad (4)$$

The first two terms reflect the DC location and operation costs. The third and fourth term constitute the inventory (holding and stock-out penalty) costs at DCs. Finally, the last two terms combine to give the distribution costs. Note that the cost function involves three decision variables of managerial significance. First, the distance between adjacent DCs, X , determines the number of DCs to locate. The local safety stock factor Z_1 reflects the degree of first-order inventory sharing. It determines the safety stock level at each DC, which in turn determines the savings in the holding cost that can be achieved by pooling the demand within a primary influence area. Next, the ratio between Z_2 and Z_1 (or the equivalent decision variable of M) determines the inventory sharing group configuration and thus the degree of second-order inventory sharing. By varying these decision variables, different portfolios of first- and second-order inventory sharing can be obtained. The optimal supply chain network design corresponds to the portfolio that minimizes the total cost $\mathfrak{C}(X^*, Z_1^*, Z_2^*)$ in (4) which best balances the location, inventory, and transportation costs. Our goal is to gain understanding on the strategic trade-offs between the proximity and agility dimensions of the supply chain network by analyzing the interactions of these decision variables.

4. Model Analysis

In this section, we analyze the supply chain design problem described in (4). First, we obtain the optimal solution of the problem. Then, we provide managerial insights in the following subsections. In particular, we focus on those that draw interesting contrast with the conventional supply chain wisdoms in the absence of joint agility and proximity considerations.

Before proceeding, we first state the following technical assumption.

ASSUMPTION 1. *At the optimal solution to (4), $\frac{\phi(Z_1)}{L(Z_1)} \leq Z_2^2 + 1$ holds.*

Assumption 1 is a sufficient condition for some of our analytical results to hold. More specifically, this is a sufficient condition for the problem of optimizing over (Z_1, Z_2) to be convex, which is valid except when the inventory service levels are very low. For example, when $Z_2/Z_1 = 3$, the condition holds if $Z_2 \geq 0.45$; and when $Z_2/Z_1 = 6$, it holds if $Z_2 \geq 0.22$. Therefore, it is not a restrictive condition under practical parameter values, where it is typically optimal to hold a safety stock of at least one standard deviation of demand.

4.1. Optimal Solution for the Supply Chain Design Problem

We first study the interactions between the two modes of inventory sharing given a fixed value of X . The cost terms that depend on Z_1 and Z_2 , i.e., the third, fourth, and last terms in (4) correspond to

the safety stock holding, penalty, and extra transportation costs incurred by dynamic fulfillment, respectively. One may observe that these cost terms exhibit a similar trade-off structure as the newsvendor cost function in that there are overage and underage costs. Specifically, the overage cost is captured by the inventory holding cost in each DC. The underage costs consist of two levels. That is, demand is first filled with inventory at the primary DC; and when there is shortage at the primary DC, inventory from secondary DCs is used triggering the first level of underage costs (the extra transportation costs). If the entire inventory sharing group stocks out, the second level of underage costs (backorder) are incurred. It is intuitive that the first-level underage costs are proportional to the local shortages (reflected by $L(Z_1)$), while the second level are proportional to the total shortages for the sharing group (reflected by $L(Z_2)$). Furthermore, the first-level underage costs are increasing (slowly) with the size of the inventory sharing group (i.e., $M^2 = (Z_2/Z_1)^2$). The optimality conditions for the Z_1 and Z_2 variables, for fixed X , are provided in the following result.

LEMMA 1. *The optimal solution to (4), for given X , is given by the unique solution to the following equations:*

$$h = vk_2X \log_2 M \bar{\Phi}(Z_1) + p\bar{\Phi}(Z_2), \quad (5)$$

$$Z_1 = R^{-1} \left(\frac{\ln 2 p}{vk_2X} \frac{\phi(Z_2)}{Z_2} \right), \quad (6)$$

where $R(z) = L(z)/z$.

Proofs of all analytical results are provided in Appendix A. The optimality condition (5) can be interpreted as a generalization of the newsvendor critical fractile condition. In particular, the optimal stocking level is one that balances the holding cost with the unit costs of stocking out at a DC and for an inventory sharing group, multiplied by the respective probabilities of such stock-outs ($\bar{\Phi}(Z_1)$ and $\bar{\Phi}(Z_2)$, respectively). It is interesting to note the similarities between equation (5) and the optimality conditions for inventory systems with second-order inventory sharing in the literature, which typically generalize the newsvendor solution into conditions that balance two critical fractiles (e.g., for the transshipment problems studied in Zhang [44] and Naseraldin and Herer [33]). However, one key difference is that, in (5), the stock-out cost for local DCs (i.e., the first term on the right hand side) depends on the decision variables (X and $M = Z_2/Z_1$), while the corresponding terms in [44] and [33] are constants multiplied by the associated stock-out probabilities. This difference arises from our consideration of the extra transportation cost of fulfilling the order by secondary DCs, which increases with the size of the inventory sharing group as well as the separation between DCs.

The optimality condition (6) reflects the relationship between the local and group safety stock factors. One can observe that, while the optimal safety stock factor for the whole inventory sharing group (Z_2) increases as each DC stocks more (i.e., Z_1 increases, as $\frac{dZ_2^*}{dZ_1} \geq 0$), the optimal inventory sharing group size M decreases, as from the proof of Lemma 1:

$$\frac{dM^*}{dZ_1} = \frac{\phi(Z_1)}{Z_1^2 L(Z_1)} \left(\frac{Z_2^*}{Z_2^{*2} + 1} \right) - \frac{M^*}{Z_1} \leq 0.$$

The above inequality holds for all Z_1, Z_2 satisfying Assumption 1. This suggests first- and second-order inventory sharing exhibit a substitution effect, such that increasing the optimal degree of one will lead to a decrease in the optimal degree of the other.

Next, we study how the density configuration of the supply chain network, i.e., the distance between DCs (X), affects the optimal inventory sharing policy.

LEMMA 2. *Let (Z_1^*, Z_2^*) satisfy (5) and (6). Then, $\frac{dZ_1^*}{dX} > 0$ and $\frac{dZ_2^*}{dX} > 0$.*

The relationship highlighted in Lemma 2 is consistent with the classical risk-pooling theory. As X increases, each DC covers a larger demand volume. By first-order inventory sharing, it becomes less costly to maintain a higher safety stock factor at DCs, and thus the optimal value of Z_1 increases. While this effect possibly interferes with the second-order inventory sharing configuration, we find that the resulting group safety stock factor still increases. Next, we characterize the optimal density of facilities by studying the optimal value of X .

PROPOSITION 1. *The optimal solution to (4) is given by the unique solution to (5), (6), and*

$$\mu u k_1 = \frac{2(\kappa f_1 + g_1)}{\delta X^3} + \frac{\sigma Z_1}{\sqrt{\delta} X^2} (h + pR(Z_2)). \quad (7)$$

Condition (7) characterizes the optimal DC configuration. It states that the optimal separation (and thus density) of DCs should balance transportation costs, reflected by the term on the left of the equality sign, against fixed facility costs and inventory costs, reflected by the terms on the right hand side. Among these three components of costs, the consideration of transportation costs favors increasing DC density (decreasing X) such that transportation distances are shorter; while facility cost considerations favors decreasing density (increasing X) to achieve economies of scale. The consideration of inventory costs, however, complicates matters because it is not immediately clear whether first- and second-order inventory sharing jointly favor consolidation (location of fewer but larger DCs) or deconsolidation (location of denser and smaller DCs) of the distribution network. Next, we investigate this issue in more details.

4.2. Impact of Inventory Sharing on Network Design

To study the impact of two modes of inventory sharing in network design, we consider the first and second (cross-) partial derivatives of the total costs (4) with respect to the network structure (X) and service level (Z_1, Z_2) variables.

PROPOSITION 2. *The impacts of the two modes of inventory sharing on network design are as follows:*

$$\frac{\partial^2}{\partial X \partial Z_1} \mathfrak{C}(X, Z_1, Z_2) \leq 0 \quad \text{and} \quad \frac{\partial^2}{\partial X \partial Z_2} \mathfrak{C}(X, Z_1, Z_2) \geq 0.$$

We find from the first result of Proposition 2 that, given a fixed group safety stock factor (Z_2), a higher local safety stock factor (Z_1) decreases the marginal cost of increasing X . That is, a higher local safety stock factor encourages the consolidation of the network of DCs. This is consistent with the conventional result that, under first-order inventory sharing, a more consolidated network of DCs allows a larger demand volume to be pooled at each DC, and thus reduces inventory costs. Therefore, to enable first-order inventory sharing, each DC should cover a larger geographical area on average, i.e., the value of X should be larger. This also increases the average distance between each customer and its primary DC, thereby increasing transportation costs. These effects constitute the fundamental trade-off in the proximity dimension.

In contrast, the second result of Proposition 2 reveals that, given a fixed local safety stock factor Z_1 , increasing the group safety stock factor Z_2 increases the marginal cost of increasing X , i.e., a higher group safety stock factor encourages deconsolidation of the DC network. It is clear that, given a fixed Z_1 , increasing Z_2 (or equivalently, the inventory sharing group size) helps reduce shortage costs. The fundamental trade-off, taking into account the agility dimension, is to strike the right balance between such savings and the associated increase in transportation costs due to increased inventory sharing group size. These results are summarized in the following observation.

Observation 1 *While first-order inventory sharing favors the location of fewer and larger DCs at which larger volumes of demand can be pooled, the possibility of second-order inventory sharing produces a counteracting force against this effect.*

Although Observation 1 does not come at a surprise, it helps provide intuitive explanations for further results, which we discuss in the next section.

4.3. How Conventional Wisdom Breaks Down

The key message that we wish to deliver in this paper is the implications of considering the agility dimension in designing responsive supply chain networks that cannot be fully captured on the

conventional proximity-based model. In this subsection, we first discuss some conventional wisdoms in supply chain network design (obtained from the proximity-based models), and show how they may break down as the agility dimension enters the equation. For each case, we provide intuitive explanations based on the effect of second-order inventory sharing, and obtain relevant managerial insights. Our analysis points out that the conventional proximity-based results may not be sufficient to inform the design of agile supply chains that are becoming commonplace in modern days. We begin with a claim regarding the inventory service level.

Conventional Wisdom 1 *If the inventory service level becomes higher, it is optimal to consolidate the network of DCs.*

This result follows from the work of Shen *et al.* [39] in their study of inventory-location problems in which first-order inventory sharing produces economies of scale (i.e., centralization of larger demand volume at facilities lead to lower safety stock holding costs), and has been confirmed in subsequent research (e.g., Naseraldin and Herer [32]). The reasoning is that the economies-of-scale effect magnifies as the service level increases, such that inventory pooling is encouraged, shifting the balance in the proximity trade-off. To verify that this claim holds true, we consider the following supply chain design problem without the option of second-order inventory sharing:

$$\min_{X, Z_1} \mathfrak{C}(X, Z_1) = \min_{X, Z_1} \frac{(\kappa f_1 + g_1)}{\delta X^2} + (f_2 + g_2)\mu + \frac{h\sigma Z_1}{\sqrt{\delta}X} + \frac{p\sigma}{\sqrt{\delta}X}L(Z_1) + \mu uk_1 X. \quad (8)$$

We refer to the above as the *proximity-focused model* to contrast with the *agility-enhanced model* of (4). In what follows, we illustrate how this conventional wisdom may change in the presence of second-order inventory sharing. In our model, because service level is endogenously determined, we discuss this qualitative result in the context of increasing the penalty cost parameter p .

PROPOSITION 3. (a) *When the penalty cost parameter p increases, the optimal X value in (8) decreases.*

(b) *When the penalty cost parameter p increases, the optimal value of X in (4) decreases.*

Proposition 3(b) shows that, when the network design is jointly optimized under the agility and proximity considerations, the result of Conventional Wisdom 1 (i.e., Proposition 3(a)) is reversed. This can be logically explained using the trade-off between first- and second-order inventory sharing (Observation 1). As the penalty cost increases, there is incentive to enhance inventory sharing. Observation 1 suggests that the two modes of inventory sharing exert opposing forces on the optimal network design. Proposition 3(b) shows that the effect of second-order inventory sharing, which

favors the location of more DCs handling smaller demand volumes, dominates that of first-order inventory sharing, which favors the location of fewer and larger DCs.

In fact, Proposition 3(b) is consistent with popular retail location strategies in large cities. For example, the electronics store chain Fortress locates a high density of outlets in Hong Kong. In the district of Mong Kok, a busy shopping area less than one square kilometer in size, the chain operates six stores [22]. The outlets perform lateral transshipments as a mode of second-order inventory sharing. Due to the high response time requirement of shoppers, such inventory sharing is effective only when the outlets are very close to each other, e.g., even within walking distance. Therefore, although the high-density location strategy may be driven primarily by marketing considerations, our result shows that it can be desirable from the operations standpoint as well.

The next claim to be investigated is related to transportation costs.

Conventional Wisdom 2 *If the unit transportation cost for shipping from DCs to customers becomes higher, it is optimal to deconsolidate the network of DCs.*

This claim holds true for virtually all classical facility location problems where the number of DCs to locate is a decision variable. As the unit transportation cost increases, the balance in the proximity trade-off is shifted, so that it is optimal to incur more fixed location costs by locating more DCs, so as to decrease the average distance between DCs and customers to save on shipping costs. In what follows, we first show that this result holds in the proximity-focused model (8) where second-order inventory sharing is not considered. Then, we contrast the impact of the transportation cost change in the presence of second-order inventory sharing.

PROPOSITION 4. (a) *When the transportation cost parameter u increases, the optimal X value in (8) increases.*

(b) *When the unit transportation cost parameters change, the following hold:*

When u increases, the optimal X value in (4) decreases. In addition, when v increases, the optimal X value in (4) increases.

Proposition 4(a) confirms the conventional wisdom that it is optimal to locate DCs closer together as the unit transportation cost increases. However, this may not always hold when the agility dimension comes to play, as shown in Proposition 4(b). In particular, the effect of changes in v deserves some attention. Intuitively, if the unit transportation cost increases, one might be tempted to think that it is always optimal to increase the density of DCs, such that shipping distances (activated under dynamic fulfillment) become shorter. However, Proposition 4(b) points

out that the optimal strategy with regard to changes in u and v can be opposite under joint consideration of agility and proximity. In particular, when shipping from secondary DCs incurs high extra transportation costs, it is optimal to decrease the density of DCs. The reason is that, as second-order inventory sharing becomes more costly to perform, it is optimal to adjust the balance between first- and second-order inventory sharing to promote the former. One can see from (5) and (6) that, given a fixed value of Z_2 , Z_1 increases in v . This encourages facilities to be located farther apart, such that each one covers a larger demand volume, to enhance first-order inventory sharing at DCs.

An illustrating example for the insights from Proposition 4(b) is the region design of the US allocation system of cadaveric livers for transplants, studied by Kong *et al.* [25]. Because it is preferable to allocate a liver to a patient in close proximity to reduce transportation times, the system divides the country into a number of regions. When a liver is available, it is first offered within the local region, and then nationwide if it is not accepted by patients within the region. In this system, regions and the whole nation are akin to primary influence areas and inventory sharing groups, respectively. Due to the high degree of time sensitivity, it is desirable to complete the transplant within the same region instead of across regions. The optimization study by Kong *et al.* [25] suggests the consolidation of the current 11 regions into 6.

The results discussed in this section show that traditional supply chain models fail to capture some important interdependencies between the optimal network design and certain business environmental factors. We conclude the section by summarizing this message in the following observation.

Observation 2 *To design agile supply chains that are commonplace in the modern information-rich era, traditional proximity-focused models are insufficient and the network design should be optimized based on enhanced models capturing the trade-offs in both the proximity and agility dimensions.*

5. Numerical Illustration

To illustrate our model, we construct a numerical illustration using a set of realistic parameter values. Details of our parameter estimation, based on the fulfillment network of Amazon, are provided in Appendix C.

We first use the numerical example to contrast the differences between optimal network configurations under first- and second-order inventory sharing by solving the proximity-focused model in (4) and the agility-enhanced model in (8). In Figure 4, we present the total costs (per customer

demand point per year) against X for the two models, holding Z_1 and Z_2 to the optimal values under the corresponding values of X .

Figure 4 Total Costs as Functions of X

As expected, the total cost is reduced significantly when both modes of inventory sharing are present. At the respective optimal values of X , second-order inventory sharing leads to a 17% reduction in total cost. We observe that the optimal value of X is smaller when both modes of inventory sharing are active, suggesting an increase in number of SCs from 11 to 14. This is consistent with our analysis that second-order inventory sharing exerts an opposite force against the risk-pooling effect of first-order inventory sharing on the optimal density of DCs. We note that there have been recent expansions of fulfillment networks of major online retailers due to intensifying competition on delivery response time [42]. Our findings reflect that second-order inventory sharing may help align the operations and marketing motives on network design, in that a deconsolidation strategy is favored both in terms of reducing inventory costs and response times.

Next, in the following, we examine the effects of changes in the main parameters, location, inventory, and transportation costs, on the optimal network design. First, varying the cost of facility location corresponds to changes in the fixed costs (f_1, g_1) or discount factor (κ). The former may fluctuate with real estate, equipment, and labor costs, and the latter is a function of cost of capital which typically varies with the market interest rates. In Figure 5(a), we multiply the discounted fixed cost ($\kappa f_1 + g_1$) by a weight that varies from 1 to 5 in (4) and (8) (note that the variable cost term $(f_2 + g_2)\mu$ is a constant and does not affect the optimal decision variable values), and present the resulting optimal values of X for both models. Intuitively, under both models, increasing the fixed location cost promotes consolidation, i.e., increasing X . Such a change poses

the same impact on both models and does not affect the underlying trade-off between the modes of inventory sharing; and thus the two curves do not appear to diverge or converge.

Second, the safety stock costs may vary due to changes in the cost of capital and cost of sales (affecting h). This may also result from changes in penalty cost p (e.g., due to increase in customer service requirements) or from the difficulty in estimation. In Figure 5(b), we multiply the unit holding (h) and penalty (p) costs by a weight varying from 1 to 5, and present the corresponding optimal values of X under the two models. Interestingly, we find that the trends in the optimal DC density (hence, the optimal supply chain design) for each model diverge in opposite directions. Whereas the agility-enhanced model results in a mild decrease in X (i.e., deconsolidation) as the safety stock cost weight increases, the optimal X rapidly increases under the proximity-focused model. We note that, although the result shown Proposition 3(b) is relatively small in the given example (about 1% decrease in optimal X), the *contrast* in trends between the two models is stark and the disparity level in the optimal X increases with the safety stock cost weight.

This divergence between the two models can be explained by the interactions between the two inventory sharing modes discussed in Observation 1. In particular, while first-order inventory sharing promotes consolidation, second-order inventory sharing provides an opposite force. The numerical example shows that, under the current parameter settings, the two forces are almost balanced. We also note that, due to the offsetting effects of the two inventory sharing modes on the optimal density of DCs, the optimal solution when both inventory sharing modes are active tends to change less with respect to cost parameter changes than does that of the model with only the first-order mode. This implies that the optimal network configuration is more robust against changes or mis-estimations of safety stock cost parameters when second-order inventory sharing is active in addition to first-order sharing. This highlights another advantage of taking into account agility in supply chain planning.

Finally, the transportation costs may also vary significantly due to fluctuations of fuel costs. We multiply the unit transportation cost parameters u and v by a weight varying from 1 to 5, and present the resulting optimal X values in Figure 5(c). Intuitively, we observe that the optimal densities of DCs under the two models both increase as transportation costs become more expensive. We also observe that the optimal X values tend to become closer as transportation cost increases. This is due to the effect suggested by Proposition 4(b), that an increase in unit transportation cost v incurred when performing dynamic fulfillment promotes first-order inventory sharing and consolidation of DCs. As v becomes large, second-order inventory sharing becomes very costly to perform, and the optimal network designs under the two models tend to converge.

Figure 5 The Impact of Various Cost Weights on Optimal Supply Chain Design

6. Conclusion and Future Research

In this paper, we consider an integrated supply chain design problem with proximity and agility considerations. The goal is to design a responsive supply chain network, taking into account the agility of the system, defined by the inventory sharing arrangements through dynamic fulfillment. Using a stylized modeling approach, as opposed to the traditional integer-programming-based approach, we study the optimal network design with two modes of inventory sharing.

Our analysis draws several managerial insights on the optimal supply chain structure under demand variability. We show that under an agility-enhanced supply chain design model, some of the strategic trade-offs are greatly altered, compared with the case of a proximity-focused model. In particular, we show that it can be optimal to locate more DCs as shortage penalty costs increase. This is opposite to the conventional wisdom that the network of DCs should be consolidated due to risk-pooling. Under such circumstances, it is optimal to enhance second-order, instead of first-order, inventory sharing. Similarly, we show that as the unit transportation cost incurred by dynamic fulfillment increases, it is optimal to further consolidate the network of DCs. This is due to the fact that a higher degree of first-order inventory sharing is needed whereas second-order inventory sharing becomes more costly. Through a numerical example based on realistic parameters, we also highlight that the optimal network configuration can become more robust against the change or mis-estimation in some of the key parameters, highlighting another benefit of taking into account the agility dimension. These results point out our key message that the agility dimension must be accounted for properly to support the design of responsive supply chain networks.

We believe that there is ample room for further research. For example, while we consider a particular form of second-order inventory sharing based on inventory sharing groups, some other policies are used in practice by different firms as seen from examples discussed in Section 1. One future direction is to enhance our model to capture different features of such inventory sharing

policies. Furthermore, one important aspect related to agility in supply chain design is the threat of facility failures or disruptions. In fact, the stylized modeling approach has led to important insights for pure facility location problems (i.e., without tactical and operational considerations) under the threat of disruptions (Cui *et al.* [9], Lim *et al.* [28]). It is therefore a very promising direction to extend our analysis to consider potential failure of DCs. With such risks, the optimal network design decisions become more complex. It would be interesting to quantify the benefits of agility under the risk of DC disruptions.

Acknowledgments

The authors gratefully acknowledge Professor Gérard Cachon, Professor Yossi Aviv, the Associate Editor and three anonymous reviewers for their valuable comments that helped improve this paper.

References

- [1] Amazon.com (2013). *Annual Report (Form 10-K)*. Retrieved from US Securities and Exchange Commission EDGAR Database <http://www.sec.gov/edgar/searchedgar/companysearch.html>.
- [2] Amazon.com (2013). *Selling at Amazon.com - Perfect Order Percentage Score*. Retrieved from http://www.amazon.com/gp/help/customer/display.html/ref=im_xx_cont_custmetric?ie=UTF8&nodeId=200935610.
- [3] Axsäter, S. (1990). Modeling Emergency Lateral Transshipments in Inventory Systems. *Management Science*, 36(11), 1329-1339.
- [4] Bridgman, P.W. (1963). *Dimensional Analysis*, Yale University Press, Yale University, CT.
- [5] Cachon, G. (2013). Retail Store Density and the Cost of Greenhouse Gas Emissions. *Management Science*, Forthcoming.
- [6] Castillo, E., A. Calviño, M. Nogal and H.K. Lo (2014). On the Probabilistic and Physical Consistency of Traffic Random Variables and Models. *Computer-Aided Civil and Infrastructure Engineering*, Forthcoming.
- [7] Chopra, S. (2000). Seven-Eleven Japan, Co. Case, Kellogg Graduate School of Management, Northwestern University, Evanston, IL.
- [8] Corbett, C.J. and K. Rajaram (2006) A Generalization of the Inventory Pooling Effect to Nonnormal Dependent Demand. *Manufacturing and Service Operations Management*, 8(4), 351-358.
- [9] Cui, T., Y. Ouyang and Z.J. Shen (2010). Reliable Facility Location Design under the Risk of Disruptions. *Operations Research*, 58(4), 998-1011.

-
- [10] Daganzo, C.F. (1984). The Distance Traveled to Visit N Points with a Maximum of C Stops per Vehicle: An Analytic Model and an Application. *Transportation Science*, 18(4), 331-350.
 - [11] Daganzo, C.F. (2005). *Logistics Systems Analysis, Fourth Edition*. Springer-Verlag, Berlin.
 - [12] Daganzo, C.F. and A.L. Erera (1999). On Planning and Design of Logistics Systems for Uncertain Environments in *New Trends in Distribution Logistics*, (M.G. Speranza and P. Stahly, Eds.), Lecture Notes in Economics and Mathematical Systems, vol 480, Springer-Verlag, Berlin.
 - [13] Daganzo, C.F. and K. Smilowitz (2004). Bounds and Approximations for the Transportation Linear Program. *Transportation Science*, 38(3), 343-356.
 - [14] Dasci, A. and G. Laporte (2005). A Continuous Model for Multistore Competitive Location. *Operations Research*, 53(2), 263-280.
 - [15] Daskin, M.S. (2013). *Network and Discrete Location: Models, Algorithms, and Applications*. Second Edition. John Wiley, New York.
 - [16] Daskin, M.S., C. Coullard and Z.J. Shen (2002). An Inventory-Location Model: Formulation, Solution Algorithm and Computational Results. *Annals of Operations Research*, 110, 83-106.
 - [17] Dong, L. and N. Rudi (2004). Who Benefits from Transshipment? Exogenous vs. Endogenous Wholesale Prices. *Management Science*, 50(5), 645-657.
 - [18] Drezner, Z., Ed. (1995). *Facility Location: A Survey of Applications and Methods*. Springer, New York.
 - [19] Drezner, Z. and H.W. Hamacher, Eds. (2002) *Facility Location: Applications and Theory*. Springer, New York.
 - [20] Eppen, G.D. (1979). Effects of Centralization on Expected Costs on the Multi-location Newsboy Problem. *Management Science*, 25(5), 498-501.
 - [21] Foreman, J., J. Gallien, J. Alspaugh, F. Lopez, R. Bhatnagar, C.C. Teo and C. Dubois (2010). Implementing Supply-Routing Optimization in a Make-to-Order Manufacturing Network. *Manufacturing & Service Operations Management*, 12(4), 547-568.
 - [22] Fortress (2013). *Store Locator*. Retrieved from <http://www.fortress.com.hk/en/location/index.php>.
 - [23] Gresh, D.L. and E.I. Kelton (2003). Visualization, Optimization, and Business Strategy: A Case Study. *Proceedings of the 14th IEEE Visualization 2003*, 531-538.

- [24] Internet Retailer Portal to E-commerce Intelligence (2012). Amazon's Average Order Value Tops \$47 in June. Retrieved from <http://www.internetretailer.com/2012/07/26/amazons-average-order-value-tops-47-june>.
- [25] Kong, N., A.J. Schaefer, B. Hunsaker and M.S. Roberts (2010). Maximizing the Efficiency of the U.S. Liver Allocation System Through Region Design. *Management Science*, 56(12), 2111-2122.
- [26] Krishnan, K.S. and V.R.K. Rao (1965). Inventory Control in n Warehouses. *Journal of Industrial Engineering*, 16(3), 212-215.
- [27] Lien, R.W., S.M.R. Iravani, K. Smilowitz and M. Tzur (2011). An Efficient and Robust Design for Transshipment Networks. *Production and Operations Management*, 20(5), 699-713.
- [28] Lim, M., M. Daskin, A. Bassamboo and S. Chopra (2013). Facility Location Decisions with Random Disruptions and Imperfect Estimation. *Manufacturing & Service Operations Management*, 15(2), 239-249.
- [29] Mak, H.Y. and Z.J. Shen (2011). Integrated Supply Chain Design Models. *Wiley Encyclopedia of Operations Research and Management Science*.
- [30] Mak, H.Y. (2012). Supply Chain Network Design with Dynamic Inventory Sharing. Working Paper, Department of Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology.
- [31] MWPVL International (2013). *Amazon.com Distribution Network*. Retrieved from http://www.mwpvl.com/html/amazon_com.html.
- [32] Naseraldin, H. and Y.T. Herer (2008). Integrating the Number and Location of Retail Outlets on a Line with Replenishment Decisions. *Management Science*, 54(9), 1666-1683.
- [33] Naseraldin, H. and Y.T. Herer (2011). A Location-Inventory Model with Lateral Transshipments. *Naval Research Logistics*, 58(5), 437-456.
- [34] Newell, G.F. (1973). Scheduling, Location, Transportation and Continuum Mechanics: Some Simple Approximations to Optimization Problems. *SIAM Journal of Applied Mathematics*, 25(3), 346-360.
- [35] Ouyang, Y. and C.F. Daganzo (2006). Discretization and Validation of the Continuum Approximation Scheme for Terminal System Design. *Transportation Science*, 40, 89-98.
- [36] Ozsen, L., M. Daskin and C. Coullard (2009). Facility Location Modeling and Inventory Management with Multisourcing. *Transportation Science*, 43(4), 455-472.

-
- [37] Robinson, L.W. (1990). Optimal and Approximate Policies in Multiperiod, Multilocation Inventory Models with Transshipments. *Operations Research*, 38(2), 278-295.
- [38] Shen, Z.J. (2006). Integrated Supply Chain Design Models: A Survey and Future Research Directions. *Journal of Industrial and Management Optimization*, 3(1), 1-27.
- [39] Shen, Z.J., C. Coullard, and M.S. Daskin (2003). A Joint Location-Inventory Model. *Transportation Science*, 37, 40-55.
- [40] Shen, Z.J. and L. Qi (2006). Incorporating Inventory and Routing Costs in Strategic Location Models. *European Journal of Operational Research*, 179, 372-389.
- [41] Smilowitz, K. and C.F. Daganzo (2007). Continuum approximation techniques for the design of integrated package distribution systems, *Networks*, 50(3), 183-196.
- [42] USA Today (2013, Oct 6). Retail Stores Become Shipping Hubs to Battle Amazon. Retrieved from <http://www.usatoday.com/story/tech/2013/09/28/retailers-ship-from-store/2862405/>.
- [43] Xu, P.J., R. Allgor and S.C. Graves (2009). Benefits of Reevaluating Real-Time Order Fulfillment Decisions. *Manufacturing and Service Operations Management*, 11(2), 340-355.
- [44] Zhang, J. (2005). Transshipment and Its Impact on Supply Chain Members' Performance. *Management Science*, 51(10), 1534-1539.

Appendix A: Proof of Statements

Before we proceed, we note the following results which we often use throughout the proofs. First, we note that $d\phi(z)/dz = -z\phi(z)$ and $d\bar{\Phi}(z)/dz = -\phi(z)$. In addition, from the standard Normal loss function $L(z) = \phi(z) - z\bar{\Phi}(z)$, it follows that $dL(z)/dz = -\bar{\Phi}(z)$. Finally, for notational convenience, we let $R(z) = L(z)/z$. Then, we have $dR(z)/dz = (-\bar{\Phi}(z)z - L(z))/z^2 = -\phi(z)/z^2$.

Proof of Lemma 1

We collect the cost terms associated with Z_1 and Z_2 (third, fourth, and sixth terms in (4)), substitute $Z_2 = MZ_1$, and consider

$$C^X(Z_1, M) = \frac{h\sigma Z_1}{\sqrt{\delta}X} + \frac{p\sigma Z_1}{\sqrt{\delta}X} \frac{L(MZ_1)}{MZ_1} + \frac{vk_2\sigma}{\sqrt{\delta}} L(Z_1) \log_2 M.$$

Taking partial derivative with respect to M , we obtain the following first-order (necessary) condition for optimality:

$$\frac{\partial}{\partial M} C^X(Z_1, M) = -\frac{p\sigma Z_1^2}{\sqrt{\delta}X} \frac{\phi(MZ_1)}{(MZ_1)^2} + \frac{vk_2\sigma}{\sqrt{\delta}} \frac{L(Z_1)}{\ln 2 M} = 0.$$

By rearranging terms, we have

$$R(Z_1) = \frac{\ln 2 p}{vk_2 X} \frac{\phi(Z_2)}{Z_2}. \quad (9)$$

When $z > 0$, one can see that $R(z) = L(z)/z$ is a strictly decreasing and continuous function, and therefore its inverse $R^{-1}(z)$ is well-defined. This leads to the one-to-one relationship of

$$Z_1 = R^{-1}\left(\frac{\ln 2 p}{vk_2 X} \frac{\phi(Z_2)}{Z_2}\right)$$

given in (6).

Let $Z_2^* = M^* Z_1$ be the corresponding optimal value of Z_2 , given Z_1 . By implicit function differentiation on the first-order condition (9) with respect to Z_1 , we obtain:

$$\begin{aligned} \frac{\phi(Z_1)}{Z_1^2} &= \frac{\ln 2 p}{vk_2 X} \left[\frac{\phi(Z_2^*)}{Z_2^{*2}} + \phi(Z_2^*) \right] \frac{dZ_2^*}{dZ_1} \\ \Rightarrow \frac{dZ_2^*}{dZ_1} &= \frac{\phi(Z_1)}{Z_1 L(Z_1)} \frac{Z_2^*}{Z_2^{*2} + 1} \geq 0 \quad (\text{substituting (9)}) \\ \Rightarrow \frac{dM^*}{dZ_1} &= \frac{1}{Z_1^2} \left[\frac{\phi(Z_1)}{L(Z_1)} \frac{Z_2^*}{Z_2^{*2} + 1} - Z_2^* \right] \leq 0 \quad (\text{using } dZ_2^*/dZ_1 = Z_1 dM^*/dZ_1 + M^*). \end{aligned} \quad (10)$$

The last inequality in (10) holds for all Z_1, Z_2 satisfying Assumption 1. Then we consider

$$\begin{aligned} \frac{\partial}{\partial Z_1} C^X(Z_1, M^*) &= \frac{h\sigma}{\sqrt{\delta}X} + \frac{p\sigma}{\sqrt{\delta}X} \frac{L(M^* Z_1)}{M^* Z_1} - \frac{p\sigma Z_1}{\sqrt{\delta}X} \frac{\phi(M^* Z_1)}{(M^* Z_1)^2} \left(Z_1 \frac{dM^*}{dZ_1} + M^* \right) \\ &\quad - \frac{vk_2\sigma}{\sqrt{\delta}} \bar{\Phi}(Z_1) \log_2 M^* + \frac{vk_2\sigma}{\sqrt{\delta}} \frac{L(Z_1)}{\ln 2 M^*} \frac{dM^*}{dZ_1} \end{aligned}$$

$$= \frac{h\sigma}{\sqrt{\delta}X} + \frac{p\sigma}{\sqrt{\delta}X} \frac{L(M^*Z_1)}{M^*Z_1} - \frac{vk_2\sigma}{\sqrt{\delta}} \bar{\Phi}(Z_1) \log_2(M^*) - \frac{p\sigma}{\sqrt{\delta}X} \frac{\phi(M^*Z_1)}{M^*Z_1}$$

(by using (9) such that the $\frac{dM^*}{dZ_1}$ terms cancel out).

Taking the first-order condition by setting $\frac{\partial}{\partial Z_1} C^X(Z_1, M^*) = 0$, we obtain

$$h = vk_2X \log_2 M \bar{\Phi}(Z_1) + p\bar{\Phi}(Z_2),$$

which is given in (5). Finally, we check that $C^X(Z_1, M^*)$ is strictly convex in Z_1 by considering:

$$\frac{\partial^2}{\partial Z_1^2} C^X(Z_1, M^*) = \frac{p\sigma}{\sqrt{\delta}X} \phi(M^*Z_1) \frac{dZ_2^*}{dZ_1} + \frac{vk_2\sigma}{\sqrt{\delta}} \phi(Z_1) \log_2 M^* - \frac{vk_2\sigma}{\sqrt{\delta}} \frac{\bar{\Phi}(Z_1)}{\ln 2 M^*} \frac{dM^*}{dZ_1} > 0.$$

Since $\frac{dM^*}{dZ_1} \leq 0$ by (10), the above implies that $C^X(Z_1, M^*)$ is strictly convex in X and thus the optimal solution is unique.

Proof of Lemma 2

Differentiating both sides of (9) with respect to X , we obtain:

$$\begin{aligned} & -\frac{\phi(Z_1^*)}{Z_1^{*2}} \frac{dZ_1^*}{dX} = -\frac{\ln 2 p}{vk_2 X^2} \frac{\phi(Z_2^*)}{Z_2^*} - \frac{\ln 2 p}{vk_2 X} \left[\frac{\phi(Z_2^*)}{Z_2^{*2}} + \phi(Z_2^*) \right] \frac{dZ_2^*}{dX} \\ \Rightarrow & \frac{\phi(Z_1^*)}{Z_1^{*2}} \frac{dZ_1^*}{dX} = \frac{\ln 2 p}{vk_2 X} \frac{\phi(Z_2^*)}{Z_2^*} \left[\frac{1}{X} + \frac{1}{Z_2^*} \frac{dZ_2^*}{dX} + Z_2^* \frac{dZ_2^*}{dX} \right] \\ \Rightarrow & \frac{\phi(Z_1^*)}{Z_1^*} \frac{dZ_1^*}{dX} = \frac{L(Z_1^*)}{X} + L(Z_1^*) \left(Z_2^* + \frac{1}{Z_2^*} \right) \frac{dZ_2^*}{dX} \quad (\text{by using (9)}). \end{aligned} \quad (11)$$

Similarly, differentiating both sides of (5) with respect to X , we obtain:

$$\begin{aligned} 0 &= vk_2 \log_2 \left(\frac{Z_2^*}{Z_1^*} \right) \bar{\Phi}(Z_1^*) + vk_2 X \frac{Z_1^*}{\ln 2 Z_2^*} \bar{\Phi}(Z_1^*) \left[\frac{1}{Z_1^*} \frac{dZ_2^*}{dX} - \frac{Z_2^*}{Z_1^{*2}} \frac{dZ_1^*}{dX} \right] \\ &+ vk_2 X \log_2 \left(\frac{Z_2^*}{Z_1^*} \right) \left(-\phi(Z_1^*) \frac{dZ_1^*}{dX} \right) - p\phi(Z_2^*) \frac{dZ_2^*}{dX} \\ \Rightarrow & \log_2 \left(\frac{Z_2^*}{Z_1^*} \right) \frac{\bar{\Phi}(Z_1^*)}{X} = \left[\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) \phi(Z_1^*) + \frac{\bar{\Phi}(Z_1^*)}{Z_1^*} \right] \frac{dZ_1^*}{dX} + \left[\frac{Z_2^* L(Z_1^*)}{Z_1^*} - \frac{\bar{\Phi}(Z_1^*)}{\ln 2 Z_2^*} \right] \frac{dZ_2^*}{dX} \\ & (\text{multiplying both sides by } \frac{1}{vk_2 X} \text{ and by using (9)}) \\ &= \frac{Z_1^* L(Z_1^*)}{\phi(Z_1^*) X} \left(\frac{\bar{\Phi}(Z_1^*)}{Z_1^*} + \log_2 \left(\frac{Z_2^*}{Z_1^*} \right) \phi(Z_1^*) \right) + \left[\frac{Z_2^* L(Z_1^*)}{\ln 2 Z_1^*} - \frac{\bar{\Phi}(Z_1^*)}{\ln 2 Z_2^*} \right] \\ &+ \frac{Z_1^* L(Z_1^*)}{\phi(Z_1^*) X} \left(\frac{\bar{\Phi}(Z_1^*)}{Z_1^*} + \log_2 \left(\frac{Z_2^*}{Z_1^*} \right) \phi(Z_1^*) \right) \left(\frac{1}{Z_2^*} + Z_2^* \right) \frac{dZ_2^*}{dX} \\ & (\text{by using (11)}). \end{aligned}$$

By rearranging the terms and using $L(Z_1) = \phi(Z_1) - Z_1 \bar{\Phi}(Z_1)$, the above can be expressed as

$$\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) (\bar{\Phi}(Z_1^*) - Z_1^* L(Z_1^*)) - \frac{\bar{\Phi}(Z_1^*) L(Z_1^*)}{\phi(Z_1^*)}$$

$$= \left[\frac{Z_2^* X}{\ln 2 Z_1^*} \phi(Z_1^*) + \left(\frac{1}{Z_2^*} + Z_2^* \right) L(Z_1^*) \left(\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) Z_1^* + \frac{\bar{\Phi}(Z_1^*)}{\phi(Z_1^*)} - \frac{\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)} \right) \right] \frac{dZ_2^*}{dX}. \quad (12)$$

Now, we show that the terms in brackets on the both sides of (12) are positive. First, the sign of the LHS can be verified as follows:

$$\begin{aligned} & \log_2 \left(\frac{Z_2^*}{Z_1^*} \right) (\bar{\Phi}(Z_1^*) - Z_1^* L(Z_1^*)) - \frac{\bar{\Phi}(Z_1^*) L(Z_1^*)}{\phi(Z_1^*)} \\ & \geq \bar{\Phi}(Z_1^*) - Z_1^* L(Z_1^*) - \frac{\bar{\Phi}(Z_1^*) L(Z_1^*)}{\phi(Z_1^*)} \quad (\text{since } \log_2(Z_2^*/Z_1^*) \geq 1) \\ & = Z_1^* \left[Z_1^* \bar{\Phi}(Z_1^*) + \frac{\bar{\Phi}^2(Z_1^*)}{\phi(Z_1^*)} - \phi(Z_1^*) \right]. \end{aligned} \quad (13)$$

Hence, we will show the term inside the bracket in (13), $A(Z_1^*) := Z_1^* \bar{\Phi}(Z_1^*) + \frac{\bar{\Phi}^2(Z_1^*)}{\phi(Z_1^*)} - \phi(Z_1^*)$, is positive. Notice that $A(Z_1^*)$ is a monotone decreasing function in Z_1^* since

$$\frac{dA(Z_1^*)}{dZ_1^*} = -\frac{\bar{\Phi}(Z_1^*)}{\phi(Z_1^*)} L(Z_1^*) < 0.$$

Furthermore, one can show that $\lim_{Z_1^* \rightarrow \infty} A(Z_1^*) = 0$ (using the L'Hospital's rule) and $A(0) > 0$. Therefore, $A(Z_1^*)$ is positive for $Z_1^* > 0$ which confirms that the LHS in (12) is also positive.

The sign of the bracketed term in the RHS of (12) can be shown as follows:

$$\begin{aligned} & \frac{Z_2^* X}{\ln 2 Z_1^*} \phi(Z_1^*) + \left(\frac{1}{Z_2^*} + Z_2^* \right) L(Z_1^*) \left(\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) Z_1^* + \frac{\bar{\Phi}(Z_1^*)}{\phi(Z_1^*)} - \frac{\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)} \right) \\ & \geq \frac{Z_2^* X}{\ln 2 Z_1^*} \phi(Z_1^*) + \frac{1}{Z_2^*} \phi(Z_1^*) \left(\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) Z_1^* + \frac{\bar{\Phi}(Z_1^*)}{\phi(Z_1^*)} - \frac{\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)} \right) \quad (\text{by Assumption 1}) \\ & > \frac{Z_2^* X}{\ln 2 Z_1^*} \phi(Z_1^*) + \frac{\bar{\Phi}(Z_1^*)}{\ln 2 Z_2^*} \left(\ln 2 - \frac{\phi(Z_1^*)}{L(Z_1^*)} \right) \quad (\text{since } \log_2(Z_2^*/Z_1^*) Z_1^* > 0) \\ & > \frac{Z_2^* X}{\ln 2 Z_1^*} \phi(Z_1^*) + \frac{\bar{\Phi}(Z_1^*)}{\ln 2 Z_2^*} \left(1 - \frac{\phi(Z_1^*)}{L(Z_1^*)} \right) \quad (\text{since } \ln 2 = 1.44 > 1) \\ & \geq \frac{Z_2^* X}{\ln 2 Z_1^*} \phi(Z_1^*) + \frac{\bar{\Phi}(Z_1^*)}{\ln 2 Z_2^*} \cdot (-Z_2^{*2}) \quad (\text{using Assumption 1}) \\ & = \frac{Z_2^* X}{\ln 2 Z_1^*} (\phi(Z_1^*) - Z_1^* \bar{\Phi}(Z_1^*)) = \frac{Z_2^* X}{\ln 2 Z_1^*} L(Z_1^*) > 0. \end{aligned}$$

Therefore, we have $\frac{dZ_2^*}{dX} > 0$. Then, (11) implies that $\frac{dZ_1^*}{dX} > 0$.

Proof of Proposition 1

Consider the total cost function $\mathfrak{C}(X, Z_1^*(X), Z_2^*(X))$ given optimal $Z_1^*(X)$ and $Z_2^*(X)$ under X . Throughout the proof, we drop the dependency of Z_1^* and Z_2^* on X for notational brevity. Then, differentiating the cost function with respect to X , we have

$$\frac{\partial}{\partial X} \mathfrak{C}(X, Z_1^*, Z_2^*) = -\frac{2(\kappa f_1 + g_1)}{\delta X^3} - \frac{h\sigma Z_1^*}{\sqrt{\delta} X^2} + \frac{h\sigma}{\sqrt{\delta} X} \frac{dZ_1^*}{dX} - \frac{p\sigma Z_1^*}{\sqrt{\delta} X^2} \frac{L(Z_2^*)}{Z_2^*} + \frac{p\sigma}{\sqrt{\delta} X} \frac{L(Z_2^*)}{Z_2^*} \frac{dZ_1^*}{dX} - \frac{p\sigma Z_1^*}{\sqrt{\delta} X} \frac{\phi(Z_2^*)}{Z_2^{*2}} \frac{dZ_2^*}{dX}$$

$$\begin{aligned}
& +u\mu k_1 - \frac{vk_2\sigma}{\sqrt{\delta}}\bar{\Phi}(Z_1^*)\log_2\left(\frac{Z_2^*}{Z_1^*}\right)\frac{dZ_1^*}{dX} + \frac{vk_2\sigma}{\sqrt{\delta}}\frac{L(Z_1^*)Z_1^*}{\ln 2 Z_2^*}\left[\frac{1}{Z_1^*}\frac{dZ_2^*}{dX} - \frac{Z_2^*}{Z_1^{*2}}\frac{dZ_1^*}{dX}\right] \\
& = -\frac{2(\kappa f_1 + g_1)}{\delta X^3} - \frac{h\sigma Z_1^*}{\sqrt{\delta}X^2} - \frac{p\sigma Z_1^*}{\sqrt{\delta}X^2}\frac{L(Z_2^*)}{Z_2^*} + \mu u k_1.
\end{aligned}$$

The last equality can be obtained using the first-order conditions of Z_1 and Z_2 ; i.e., $\frac{dZ_1^*}{dX} = 0$ and $\frac{dZ_2^*}{dX} = 0$. Then, the first-order condition in X , $\frac{\partial}{\partial X}\mathfrak{C}(X, Z_1^*, Z_2^*) = 0$, yields (7).

Furthermore, we can verify the convexity of the cost function in X by considering:

$$\begin{aligned}
\frac{\partial^2}{\partial X^2}\mathfrak{C}(X, Z_1^*, Z_2^*) &= \frac{6(\kappa f_1 + g_1)}{\delta X^4} + \frac{2h\sigma Z_1^*}{\sqrt{\delta}X^3} - \frac{h\sigma}{\sqrt{\delta}X^2}\frac{dZ_1^*}{dX} + \frac{2p\sigma Z_1^*}{\sqrt{\delta}X^3}\frac{L(Z_2^*)}{Z_2^*} - \frac{p\sigma}{\sqrt{\delta}X^2}\frac{L(Z_2^*)}{Z_2^*}\frac{dZ_1^*}{dX} \\
&\quad + \frac{p\sigma Z_1^*}{\sqrt{\delta}X^2}\frac{\phi(Z_2^*)}{Z_2^{*2}}\frac{dZ_2^*}{dX} \\
&= \frac{6(\kappa f_1 + g_1)}{\delta X^4} + \frac{\sigma}{\sqrt{\delta}X^2}\left(h + p\frac{L(Z_2^*)}{Z_2^*}\right)\left[\frac{2Z_1^*}{X} - \frac{dZ_1^*}{dX}\right] + \frac{p\sigma Z_1^*}{\sqrt{\delta}X^2}\frac{\phi(Z_2^*)}{Z_2^{*2}}\frac{dZ_2^*}{dX}.
\end{aligned}$$

From Lemma 2, we know that $\frac{dZ_2^*}{dX} > 0$ and thus the last term in the above is positive. In addition, we also know that $\frac{dZ_1^*}{dX} > 0$. Therefore, to show $\frac{\partial^2}{\partial X^2}\mathfrak{C}(X, Z_1^*, Z_2^*) > 0$, it is sufficient to show that $\frac{dZ_1^*}{dX} - \frac{2Z_1^*}{X} \leq 0$. By (11), we have

$$\frac{dZ_1^*}{dX} - \frac{2Z_1^*}{X} = \frac{Z_1^*}{X}\left(\frac{L(Z_1^*)}{\phi(Z_1^*)} - 2\right) + \frac{Z_1^*L(Z_1^*)}{\phi(Z_1^*)}\left(\frac{1}{Z_2^*} + Z_2^*\right)\frac{dZ_2^*}{dX}. \quad (14)$$

Below, we show that the RHS of (14) is non-positive. Recall that the RHS bracketed term in (12) is strictly positive. Multiplying this term to the RHS of (14), we have the following:

$$\begin{aligned}
& \frac{Z_1^*}{X}\left(\frac{L(Z_1^*)}{\phi(Z_1^*)} - 2\right)\left[\frac{Z_2^*}{\ln 2 Z_1^*}\phi(Z_1^*) + \left(\frac{1}{Z_2^*} + Z_2^*\right)L(Z_1^*)\left(\log_2\left(\frac{Z_2^*}{Z_1^*}\right)Z_1^* + \frac{\bar{\Phi}(Z_1^*)}{\phi(Z_1^*)} - \frac{\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)}\right)\right] \\
& + \frac{Z_1^*L(Z_1^*)}{\phi(Z_1^*)}\left(\frac{1}{Z_2^*} + Z_2^*\right)\left[\frac{Z_2^*}{\ln 2 Z_1^*}\phi(Z_1^*) + \left(\frac{1}{Z_2^*} + Z_2^*\right)L(Z_1^*)\left(\log_2\left(\frac{Z_2^*}{Z_1^*}\right)Z_1^* + \frac{\bar{\Phi}(Z_1^*)}{\phi(Z_1^*)} - \frac{\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)}\right)\right]\frac{dZ_2^*}{dX} \\
& = \frac{Z_1^*}{X}\left(\frac{L(Z_1^*)}{\phi(Z_1^*)} - 2\right)\left[\frac{Z_2^*}{\ln 2 Z_1^*}\phi(Z_1^*) + \left(\frac{1}{Z_2^*} + Z_2^*\right)L(Z_1^*)\left(\log_2\left(\frac{Z_2^*}{Z_1^*}\right)Z_1^* + \frac{\bar{\Phi}(Z_1^*)}{\phi(Z_1^*)} - \frac{\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)}\right)\right] \\
& + \frac{Z_1^*L(Z_1^*)}{\phi(Z_1^*)}\left(\frac{1}{Z_2^*} + Z_2^*\right)\frac{1}{X}\left[\log_2\left(\frac{Z_2^*}{Z_1^*}\right)(\bar{\Phi}(Z_1^*) - Z_1^*L(Z_1^*)) - \frac{\bar{\Phi}(Z_1^*)L(Z_1^*)}{\phi(Z_1^*)}\right] \\
& \quad \text{(by replacing the RHS of (12) with its LHS expression)} \\
& = \frac{Z_1^*L(Z_1^*)}{X\phi(Z_1^*)}\left[\left(1 - \frac{2\phi(Z_1^*)}{L(Z_1^*)}\right)\left[\frac{Z_2^*}{\ln 2 Z_1^*}\phi(Z_1^*) + \left(\frac{1}{Z_2^*} + Z_2^*\right)L(Z_1^*)\left(\log_2\left(\frac{Z_2^*}{Z_1^*}\right)Z_1^* + \frac{\bar{\Phi}(Z_1^*)}{\phi(Z_1^*)} - \frac{\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)}\right)\right]\right. \\
& \quad \left. + \left(\frac{1}{Z_2^*} + Z_2^*\right)\left[\log_2\left(\frac{Z_2^*}{Z_1^*}\right)(\bar{\Phi}(Z_1^*) - Z_1^*L(Z_1^*)) - \frac{\bar{\Phi}(Z_1^*)L(Z_1^*)}{\phi(Z_1^*)}\right]\right] \\
& = \frac{Z_1^*L(Z_1^*)}{XZ_2^*\phi(Z_1^*)}\left[(1 + Z_2^{*2})\left[\log_2\left(\frac{Z_2^*}{Z_1^*}\right)(\bar{\Phi}(Z_1^*) - 2Z_1^*\phi(Z_1^*)) - 2\bar{\Phi}(Z_1^*) - \frac{\bar{\Phi}(Z_1^*)}{\ln 2} + \frac{2\phi(Z_1^*)\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)}\right]\right. \\
& \quad \left. + \frac{Z_2^{*2}\phi(Z_1^*)}{\ln 2 Z_1^*}\left(1 - \frac{2\phi(Z_1^*)}{L(Z_1^*)}\right)\right].
\end{aligned}$$

Next, we show that the term in brackets in the above is non-positive for all Z_1^* and Z_2^* , mainly using the relationships of $L(Z_1^*) = \phi(Z_1^*) - Z_1^* \bar{\Phi}(Z_1^*)$ and Assumption 1.

$$\begin{aligned}
& (1+Z_2^{*2}) \left[\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) (\bar{\Phi}(Z_1^*) - 2Z_1^* \phi(Z_1^*)) - 2\bar{\Phi}(Z_1^*) - \frac{\bar{\Phi}(Z_1^*)}{\ln 2} + \frac{2\phi(Z_1^*)\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)} \right] + \frac{Z_2^{*2}\phi(Z_1^*)}{\ln 2 Z_1^*} \left(1 - \frac{2\phi(Z_1^*)}{L(Z_1^*)} \right) \\
&= (1+Z_2^{*2}) \left[\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) (\bar{\Phi}(Z_1^*) - 2Z_1^* \phi(Z_1^*)) - 2\bar{\Phi}(Z_1^*) - \frac{\bar{\Phi}(Z_1^*)}{\ln 2} \right] \\
&\quad + \frac{2\phi(Z_1^*)\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)} + \frac{2(Z_2^*)^2\phi(Z_1^*)\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)} + \frac{Z_2^{*2}\phi(Z_1^*)}{\ln 2 Z_1^*} \left(1 - \frac{2\phi(Z_1^*)}{L(Z_1^*)} \right) \\
&= (1+Z_2^{*2}) \left[\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) (\bar{\Phi}(Z_1^*) - 2Z_1^* \phi(Z_1^*)) - 2\bar{\Phi}(Z_1^*) - \frac{\bar{\Phi}(Z_1^*)}{\ln 2} \right] + \frac{\phi(Z_1^*)}{\ln 2 L(Z_1^*)} \left(2\bar{\Phi}(Z_1^*) - \frac{Z_2^{*2}L(Z_1^*)}{Z_1^*} \right) \\
&\leq (1+Z_2^{*2}) \left[\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) (\bar{\Phi}(Z_1^*) - 2Z_1^* \phi(Z_1^*)) - 2\bar{\Phi}(Z_1^*) - \frac{\bar{\Phi}(Z_1^*)}{\ln 2} \right] + \frac{\phi(Z_1^*)\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)} \quad (\text{by Assumption 1}) \\
&\leq (1+Z_2^{*2}) \left[\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) (\bar{\Phi}(Z_1^*) - 2Z_1^* \phi(Z_1^*)) - 2\bar{\Phi}(Z_1^*) \right] - \frac{\phi(Z_1^*)\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)} + \frac{\phi(Z_1^*)\bar{\Phi}(Z_1^*)}{\ln 2 L(Z_1^*)} \\
&\quad (\text{by Assumption 1}) \\
&= (1+Z_2^{*2}) \left[\log_2 \left(\frac{Z_2^*}{Z_1^*} \right) \left(-\frac{L(Z_1^*)}{Z_1^*} - Z_1^* \phi(Z_1^*) \right) - 2\bar{\Phi}(Z_1^*) \right] \leq 0.
\end{aligned}$$

Since the above is non-positive, (14) in turn is also non-positive for all Z_1^* and Z_2^* . Therefore, the cost function is convex in X .

Proof of Proposition 2

By taking second derivatives of (4), the below follows immediately.

$$\begin{aligned}
\frac{\partial^2}{\partial X \partial Z_1} \mathfrak{C}(X, Z_1, Z_2) &= -\frac{\sigma}{\sqrt{\delta} X^2} \left(h + p \frac{L(Z_2)}{Z_2} \right) \leq 0, \\
\frac{\partial^2}{\partial X \partial Z_2} \mathfrak{C}(X, Z_1, Z_2) &= \frac{\sigma Z_1}{\sqrt{\delta} X^2} \frac{\phi(Z_2)}{Z_2^2} \geq 0.
\end{aligned}$$

Useful Lemmas for Subsequent Proofs

LEMMA 3. *Let $f(x, t)$ be a function that is continuous and increasing in $x \in \mathbb{R}$, differentiable in $t \in \mathbb{R}$, and has one root in x for any given value of t , denoted by $x^*(t)$. Then, $x^*(t)$ is increasing (decreasing) in t if $\frac{\partial}{\partial t} f(x, t)|_{x=x^*(t)} < 0$ (> 0).*

Proof. To prove this, we note that if $\frac{\partial}{\partial t} f(x, t)|_{x=x^*(t)} < 0$, there exists some $\beta > 0$ such that $f(x^*(t), t + \epsilon) < 0$ for any $\epsilon \in (0, \beta)$. Since there is only one root and the function is continuous in x , this implies that $x^*(t + \epsilon) > x^*(t)$. The argument for the decreasing case is analogous.

LEMMA 4. $\frac{L(Z)}{\phi(Z)}$ is decreasing in $Z > 0$.

Proof. Let $B(Z) := \frac{L(Z)}{\phi(Z)}$ and consider

$$\frac{dB(Z)}{dZ} = \frac{-\bar{\Phi}(Z)\phi(Z) + ZL(Z)\phi(Z)}{\phi^2(Z)} = \frac{-\bar{\Phi}(Z) + ZL(Z)}{\phi(Z)}.$$

Since $\phi(Z) > 0$, the above is non-positive if the numerator is non-positive. Denote the numerator by $C(Z) := -\bar{\Phi}(Z) + ZL(Z)$. Note that $C(Z)$ is increasing in $Z > 0$ since $\frac{dC(Z)}{dZ} = 2L(Z) > 0$. Also, one can verify that $C(0) = -0.5 < 0$ and $\lim_{Z \rightarrow \infty} C(Z) = 0$. Hence, $C(Z) \leq 0$ for all $Z > 0$, and in turn $\frac{L(Z)}{\phi(Z)}$ is decreasing in Z .

Proof of Proposition 3(a)

First, we note that the first-order condition for (8) yields the newsvendor critical fractile solution of $\bar{\Phi}(Z_1^*) = h/p$. Then, differentiating both sides with respect to p yields $\frac{\partial Z_1^*}{\partial p} = \frac{h}{p^2} \frac{1}{\phi(Z_1^*)} > 0$. Notice that Z_1^* is a constant given by $\bar{\Phi}^{-1}(h/p)$ and does not depend on X . Then, it follows that

$$\begin{aligned} \frac{\partial}{\partial X} \mathfrak{C}(X, Z_1^*) &= -\frac{2(\kappa f_1 + g_1)}{\delta X^3} - \frac{h\sigma Z_1^*}{\sqrt{\delta} X^2} - \frac{p\sigma}{\sqrt{\delta} X^2} L(Z_1^*) + \mu u k_1, \\ \frac{\partial^2}{\partial X^2} \mathfrak{C}(X, Z_1^*) &= \frac{6(\kappa f_1 + g_1)}{\delta X^4} + \frac{2h\sigma Z_1^*}{\sqrt{\delta} X^3} + \frac{2p\sigma}{\sqrt{\delta} X^2} L(Z_1^*) > 0, \end{aligned} \quad (15)$$

$$\begin{aligned} \frac{\partial^2}{\partial X \partial p} \mathfrak{C}(X, Z_1^*) &= -\frac{h\sigma}{\sqrt{\delta} X^2} \frac{\partial Z_1^*}{\partial p} - \frac{\sigma L(Z_1^*)}{\sqrt{\delta} X^2} + \frac{p\sigma}{\sqrt{\delta} X^2} \bar{\Phi}(Z_1^*) \frac{\partial Z_1^*}{\partial p} \\ &= -\frac{\sigma L(Z_1^*)}{\sqrt{\delta} X^2} < 0 \quad \left(\text{using } \bar{\Phi}(Z_1^*) = \frac{h}{p} \right). \end{aligned} \quad (16)$$

Let $f(X, p) := \frac{\partial}{\partial X} \mathfrak{C}(X, p)$. Then, (15) implies that $f(X, p)$ is increasing in X and (16) shows that $\frac{\partial}{\partial p} f(X, p)|_{X=X^*(p)} < 0$. Hence, from Lemma 3, the optimal X in (8), $X^*(p)$, is increasing in p .

Proof of Proposition 3(b)

For this proof, we omit the superscript $*$ for brevity, when there is no risk of confusion. In addition, we denote $Z_1^*(X, p)$ and $Z_2^*(X, p)$ (optimal Z_1 and Z_2 values for given X and p , from (5) and (9)) by simply Z_1 and Z_2 , respectively, for brevity.

Implicitly differentiating both sides of (9) with respect to p , we obtain

$$\begin{aligned} \frac{\phi(Z_1)}{Z_1^2} \frac{\partial Z_1}{\partial p} &= -\frac{\ln 2 \phi(Z_2)}{vk_2 X Z_2} + \frac{\ln 2 p \phi(Z_2)}{vk_2 X} \left(1 + \frac{1}{Z_2^2} \right) \frac{\partial Z_2}{\partial p} \\ \Rightarrow \quad \frac{\partial Z_1}{\partial p} &= \frac{Z_1^2}{\phi(Z_1)} \left[-\frac{L(Z_1)}{p Z_1} + \frac{L(Z_1)}{Z_1} \left(Z_2 + \frac{1}{Z_2} \right) \frac{\partial Z_2}{\partial p} \right] \quad (\text{by using (9)}). \end{aligned} \quad (17)$$

Similarly, differentiating both sides of (5) with respect to p , we have

$$0 = vk_2 X \frac{Z_1 \bar{\Phi}(Z_1)}{\ln 2 Z_2} \left(\frac{1}{Z_1} \frac{\partial Z_2}{\partial p} - \frac{Z_2}{Z_1^2} \frac{\partial Z_1}{\partial p} \right) - vk_2 X \phi(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) \frac{\partial Z_1}{\partial p} + \bar{\Phi}(Z_2) - p \phi(Z_2) \frac{\partial Z_2}{\partial p}.$$

Then, it follows that

$$\begin{aligned} \bar{\Phi}(Z_2) &= vk_2 X \left\{ \left[\frac{Z_2 L(Z_1)}{Z_1} - \frac{\bar{\Phi}(Z_1)}{\ln 2 Z_2} \right] \frac{\partial Z_2}{\partial p} + \left[\phi(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{\bar{\Phi}(Z_1)}{\ln 2 Z_1} \right] \frac{\partial Z_1}{\partial p} \right\} \quad (\text{by using (9)}) \\ &= vk_2 X \left[\frac{Z_2 L(Z_1)}{Z_1} - \frac{\bar{\Phi}(Z_1)}{\ln 2 Z_2} + Z_1 L(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) \left(Z_2 + \frac{1}{Z_2} \right) + \frac{\bar{\Phi}(Z_1) L(Z_1)}{\ln 2 \phi(Z_1)} \left(Z_2 + \frac{1}{Z_2} \right) \right] \frac{\partial Z_2}{\partial p} \end{aligned}$$

$$-\frac{vk_2X}{p} \left[Z_1 L(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{\bar{\Phi}(Z_1)L(Z_1)}{\ln 2 \phi(Z_1)} \right] \quad (\text{by using (17)}). \quad (18)$$

From (5), we know that

$$\bar{\Phi}(Z_2) = \frac{h}{p} - \frac{vk_2X}{p} \log_2 \left(\frac{Z_2}{Z_1} \right) \bar{\Phi}(Z_1).$$

Hence, (18) can be further rearranged as follows:

$$\begin{aligned} & \frac{h}{p} + \frac{vk_2X}{p} \left[\left(Z_1 L(Z_1) - \bar{\Phi}(Z_1) \right) \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{\bar{\Phi}(Z_1)L(Z_1)}{\ln 2 \phi(Z_1)} \right] \\ &= vk_2X \frac{\partial Z_2}{\partial p} \left[\frac{Z_2 L(Z_1)}{Z_1} - \frac{\bar{\Phi}(Z_1)}{\ln 2 Z_2} + Z_1 L(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) \left(Z_2 + \frac{1}{Z_2} \right) + \frac{\bar{\Phi}(Z_1)L(Z_1)}{\ln 2 \phi(Z_1)} \left(Z_2 + \frac{1}{Z_2} \right) \right] \\ \Rightarrow & \frac{h}{p} + \ln 2 \frac{Z_1 \phi(Z_2)}{Z_2} \left[\left(Z_1 - \frac{\bar{\Phi}(Z_1)}{L(Z_1)} \right) \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{\bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right] \\ &= \ln 2 p \frac{\partial Z_2}{\partial p} \left[\phi(Z_2) - \frac{Z_1 \phi(Z_2) \bar{\Phi}(Z_1)}{Z_2^2 L(Z_1)} + \frac{\phi(Z_2)}{Z_2} \left(Z_2 + \frac{1}{Z_2} \right) \left(Z_1^2 \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{Z_1 \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right) \right]. \quad (19) \end{aligned}$$

In the above, we obtain the last equality by using $\frac{vk_2X}{p} = \ln 2 \frac{Z_1 \phi(Z_2)}{Z_2 L(Z_1)}$ from (9). Now, we first verify that the bracketed term in right hand side is positive.

$$\begin{aligned} & \phi(Z_2) - \frac{Z_1 \phi(Z_2) \bar{\Phi}(Z_1)}{Z_2^2 L(Z_1)} + \frac{\phi(Z_2)}{Z_2} \left(Z_2 + \frac{1}{Z_2} \right) \left(Z_1^2 \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{Z_1 \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right) \\ \geq & \phi(Z_2) - \frac{Z_1 \phi(Z_2) \bar{\Phi}(Z_1)}{Z_2^2 L(Z_1)} + \frac{\phi(Z_2)}{Z_2} \frac{\phi(Z_1)}{Z_2 L(Z_1)} \left(Z_1^2 \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{Z_1 \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right) \quad (\text{by Assumption 1}) \\ = & \frac{Z_1 \phi(Z_2) \bar{\Phi}(Z_1)}{Z_2 L(Z_1)} \left[\frac{Z_2 L(Z_1)}{Z_1 \bar{\Phi}(Z_1)} - \frac{1}{Z_2} + \frac{Z_1 \phi(Z_1)}{Z_2 \bar{\Phi}(Z_1)} \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{1}{\ln 2 Z_2} \right]. \end{aligned}$$

The above is positive for all $Z_1 > 0$, because

$$\frac{Z_1 \phi(Z_2) \bar{\Phi}(Z_1)}{Z_2 L(Z_1)} > 0 \quad \text{and} \quad \left[\frac{Z_2 L(Z_1)}{Z_1 \bar{\Phi}(Z_1)} - \frac{1}{Z_2} + \frac{Z_1 \phi(Z_1)}{Z_2 \bar{\Phi}(Z_1)} \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{1}{\ln 2 Z_2} \right] > 0,$$

where the second part follows from the fact that

$$\frac{Z_2 L(Z_1)}{Z_1 \bar{\Phi}(Z_1)} - \frac{1}{Z_2} = \frac{Z_2^2 L(Z_1) - Z_1 \bar{\Phi}(Z_1)}{Z_1 Z_2 \bar{\Phi}(Z_1)} \geq \frac{\phi(Z_1) - L(Z_1) - Z_1 \bar{\Phi}(Z_1)}{Z_1 Z_2 \bar{\Phi}(Z_1)} = \frac{L(Z_1) - L(Z_1)}{Z_1 Z_2 \bar{\Phi}(Z_1)} = 0$$

using $Z_2^2 \geq \phi(Z_1)/L(Z_1) - 1$ of Assumption 1, and the remaining terms are positive.

Next, we show that the optimal X in (4), $X^*(p)$, is decreasing in p . Consider

$$\begin{aligned} \frac{\partial}{\partial X} \mathfrak{C}(X, Z_1, Z_2) &= -\frac{2(\kappa f_1 + g_1)}{\delta X^3} - \frac{h\sigma Z_1}{\sqrt{\delta} X^2} - \frac{p\sigma Z_1}{\sqrt{\delta} X^2} \frac{L(Z_2)}{Z_2} + \mu u k_1, \\ \frac{\partial^2}{\partial X^2} \mathfrak{C}(X, Z_1, Z_2) &= \frac{6(\kappa f_1 + g_1)}{\delta X^4} + \frac{2h\sigma Z_1}{\sqrt{\delta} X^3} + \frac{2p\sigma Z_1}{\sqrt{\delta} X^3} \frac{L(Z_2)}{Z_2} > 0. \end{aligned} \quad (20)$$

From (20), $\frac{\partial}{\partial X} \mathfrak{C}(X, Z_1, Z_2)$ is increasing in X . Further, it follows that

$$\frac{\partial^2}{\partial X \partial p} \mathfrak{C}(X, Z_1, Z_2) = -\left(\frac{h\sigma}{\sqrt{\delta} X^2} + \frac{p\sigma}{\sqrt{\delta} X^2} \frac{L(Z_2)}{Z_2} \right) \frac{\partial Z_1}{\partial p} - \frac{\sigma Z_1}{\sqrt{\delta} X^2} \frac{L(Z_2)}{Z_2} + \frac{p\sigma Z_1}{\sqrt{\delta} X^2} \frac{\phi(Z_2)}{Z_2^2} \frac{\partial Z_2}{\partial p}$$

$$\begin{aligned}
&= \frac{\sigma}{\sqrt{\delta} X^2} \left\{ \frac{hZ_1 L(Z_1)}{p\phi(Z_1)} + \frac{Z_1 L(Z_1)L(Z_2)}{Z_2\phi(Z_1)} - \frac{Z_1 L(Z_2)}{Z_2} \right. \\
&\quad \left. + p \frac{\partial Z_2}{\partial p} \left[\frac{Z_1\phi(Z_2)}{Z_2^2} - \left(\frac{h}{p} + \frac{L(Z_2)}{Z_2} \right) \frac{Z_1 L(Z_1)}{\phi(Z_1)} \left(Z_2 + \frac{1}{Z_2} \right) \right] \right\} \quad (21) \\
&\quad \text{(replacing } \frac{\partial Z_1}{\partial p} \text{ by using (17)).}
\end{aligned}$$

Finally, we show that the term in $\{\cdot\}$ of (21) is non-negative by using (19). Specifically, we multiply the bracketed term on the RHS of (19), which is positive, to the term in $\{\cdot\}$ of (21):

$$\begin{aligned}
&\left[\frac{hZ_1 L(Z_1)}{p\phi(Z_1)} + \frac{Z_1 L(Z_1)L(Z_2)}{Z_2\phi(Z_1)} - \frac{Z_1 L(Z_2)}{Z_2} \right] \left[\phi(Z_2) - \frac{Z_1\phi(Z_2)\bar{\Phi}(Z_1)}{Z_2^2 L(Z_1)} + \frac{\phi(Z_2)}{Z_2} \right. \\
&\quad \left. \left(Z_2 + \frac{1}{Z_2} \right) \left(Z_1^2 \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{Z_1 \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right) \right] + p \frac{\partial Z_2}{\partial p} \left[\frac{Z_1\phi(Z_2)}{Z_2^2} - \left(\frac{h}{p} + \frac{L(Z_2)}{Z_2} \right) \right. \\
&\quad \left. \frac{Z_1 L(Z_1)}{\phi(Z_1)} \left(Z_2 + \frac{1}{Z_2} \right) \right] \left[\phi(Z_2) - \frac{Z_1\phi(Z_2)\bar{\Phi}(Z_1)}{Z_2^2 L(Z_1)} + \frac{\phi(Z_2)}{Z_2} \left(Z_2 + \frac{1}{Z_2} \right) \left(Z_1^2 \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{Z_1 \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right) \right].
\end{aligned}$$

We further obtain the following using (19), i.e., replacing the RHS to its LHS:

$$\begin{aligned}
&\left[\frac{hZ_1 L(Z_1)}{p\phi(Z_1)} + \frac{Z_1 L(Z_1)L(Z_2)}{Z_2\phi(Z_1)} - \frac{Z_1 L(Z_2)}{Z_2} \right] \left[\phi(Z_2) - \frac{Z_1\phi(Z_2)\bar{\Phi}(Z_1)}{Z_2^2 L(Z_1)} + \frac{\phi(Z_2)}{Z_2} \right. \\
&\quad \left. \left(Z_2 + \frac{1}{Z_2} \right) \left(Z_1^2 \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{Z_1 \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right) \right] + \left[\frac{Z_1\phi(Z_2)}{Z_2^2} - \left(\frac{h}{p} + \frac{L(Z_2)}{Z_2} \right) \right. \\
&\quad \left. \frac{Z_1 L(Z_1)}{\phi(Z_1)} \left(Z_2 + \frac{1}{Z_2} \right) \right] \left[\frac{h}{\ln 2 p} + \frac{Z_1\phi(Z_2)}{Z_2} \left[\left(Z_1 - \frac{\bar{\Phi}(Z_1)}{L(Z_1)} \right) \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{\bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right] \right].
\end{aligned}$$

By using $\frac{h}{p} = \ln 2 \frac{Z_1\phi(Z_2)\bar{\Phi}(Z_1)}{Z_2 L(Z_1)} \log_2 \left(\frac{Z_2}{Z_1} \right) + \bar{\Phi}(Z_2)$ from (5) and (9), the above can be reexpressed as follows:

$$\begin{aligned}
&\left[\frac{\ln 2 Z_1^2 \phi(Z_2)\bar{\Phi}(Z_1)}{Z_2\phi(Z_1)} \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{Z_1 L(Z_1)\bar{\Phi}(Z_2)}{\phi(Z_1)} + \frac{Z_1 L(Z_1)L(Z_2)}{Z_2\phi(Z_1)} - \frac{Z_1 L(Z_2)}{Z_2} \right] \\
&\left[\phi(Z_2) - \frac{Z_1\phi(Z_2)\bar{\Phi}(Z_1)}{Z_2^2 L(Z_1)} + \frac{\phi(Z_2)}{Z_2} \left(Z_2 + \frac{1}{Z_2} \right) \left(Z_1^2 \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{Z_1 \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right) \right] \\
&+ \left[\frac{Z_1\phi(Z_2)}{Z_2^2} - \left(Z_2 + \frac{1}{Z_2} \right) \left(\frac{\ln 2 Z_1^2 \phi(Z_2)\bar{\Phi}(Z_1)}{Z_2\phi(Z_1)} \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{Z_1 L(Z_1)\bar{\Phi}(Z_2)}{\phi(Z_1)} + \frac{Z_1 L(Z_1)L(Z_2)}{Z_2\phi(Z_1)} \right) \right] \\
&\left[\frac{Z_1\phi(Z_2)\bar{\Phi}(Z_1)}{Z_2 L(Z_1)} \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{\bar{\Phi}(Z_2)}{\ln 2} + \frac{Z_1\phi(Z_2)}{Z_2} \left(\left(Z_1 - \frac{\bar{\Phi}(Z_1)}{L(Z_1)} \right) \log_2 \left(\frac{Z_2}{Z_1} \right) + \frac{\bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right) \right].
\end{aligned}$$

Upon algebraic expansion, we reexpressed the above by collecting terms with respect to $\log_2 \left(\frac{Z_2}{Z_1} \right)$ (and then, within, with respect to $(1 + Z_2^2)$), we obtain:

$$\begin{aligned}
&\log_2^2 \left(\frac{Z_2}{Z_1} \right) \frac{\phi(Z_2)Z_1^4}{\phi(Z_1)Z_2^3} (1 + Z_2^2) \left\{ \phi(Z_1) + \ln 2 \phi(Z_2)\bar{\Phi}(Z_1) + Z_2^2 \phi(Z_1) \right\} \\
&+ \log_2 \left(\frac{Z_2}{Z_1} \right) \frac{Z_1^2}{Z_2^3} \left\{ \frac{\ln 2 \phi^2(Z_2)\bar{\Phi}(Z_1)}{\phi(Z_1)} \left(Z_2^2 - \frac{Z_1 \bar{\Phi}(Z_1)}{L(Z_1)} \right) - Z_1 \phi^2(Z_2) \right\}
\end{aligned}$$

$$\begin{aligned}
& + (1 + Z_2^2) \left(\frac{Z_1 \phi^2(Z_2) \bar{\Phi}^2(Z_1)}{\phi^2(Z_1)} + Z_1 \phi^2(Z_2) \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right) \right) \\
& + (1 + Z_2^2)^2 \frac{1}{\ln 2} (2\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \Big\} \\
& + \frac{1}{\ln 2 Z_2^3} \left\{ Z_1 \phi(Z_2) \left(\frac{\ln 2 Z_2^2 \phi(Z_1) \phi(Z_2)}{L(Z_1)} \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right) - (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \right) \right. \\
& + (1 + Z_2^2) Z_1^2 \phi^2(Z_2) \bar{\Phi}(Z_1) \left(\frac{1}{\phi(Z_1)} \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right) - \frac{\ln 2}{L(Z_1)} \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right) \right) \\
& \left. + (1 + Z_2^2)^2 \left(\frac{\bar{\Phi}(Z_1)}{\ln 2} (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \right) \right\}.
\end{aligned}$$

To show that this value is positive, we show that expressions in the three sets of $\{\cdot\}$ are all positive (or non-negative).

It is straightforward to see that the first expression in $\{\cdot\}$, associated with the $\log_2^2(\frac{Z_2}{Z_1})$ term, is positive for $Z_2 > Z_1 > 0$. For the expression associated with the $\log_2(\frac{Z_2}{Z_1})$ term, we have the following:

$$\begin{aligned}
& \frac{\ln 2 \phi^2(Z_2) \bar{\Phi}(Z_1)}{\phi(Z_1)} \left(Z_2^2 - \frac{Z_1 \bar{\Phi}(Z_1)}{L(Z_1)} \right) - Z_1 \phi^2(Z_2) \\
& + (1 + Z_2^2) \left(\frac{Z_1 \phi^2(Z_2) \bar{\Phi}^2(Z_1)}{\phi^2(Z_1)} + Z_1 \phi^2(Z_2) \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right) \right) + (1 + Z_2^2)^2 \frac{1}{\ln 2} (2\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \\
& > \frac{\ln 2 \phi^2(Z_2) \bar{\Phi}(Z_1)}{\phi(Z_1)} \left(Z_2^2 - \frac{Z_1 \bar{\Phi}(Z_1)}{L(Z_1)} \right) - Z_1 \phi^2(Z_2) + (1 + Z_2^2) \left(\frac{Z_1 \phi^2(Z_2) \bar{\Phi}^2(Z_1)}{\phi^2(Z_1)} \right) \\
& \quad (\text{by } \frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} > 0 \text{ from Lemma 4 and dropping the } (1 + Z_2^2)^2 \text{ associated terms, which are positive}) \\
& \geq \frac{\ln 2 \phi^2(Z_2) \bar{\Phi}(Z_1)}{\phi(Z_1)} \left(\frac{\phi(Z_1)}{L(Z_1)} - 1 - \frac{Z_1 \bar{\Phi}(Z_1)}{L(Z_1)} \right) - Z_1 \phi^2(Z_2) + \frac{Z_1 \phi^2(Z_2) \bar{\Phi}^2(Z_1)}{\phi(Z_1) L(Z_1)} \\
& \quad (\text{by using Assumption 1 for both the first and the last terms}) \\
& = -Z_1 \phi^2(Z_2) + \frac{Z_1 \phi^2(Z_2) \bar{\Phi}^2(Z_1)}{\phi(Z_1) L(Z_1)} \quad (\text{the first term equals 0 using } L(Z_1) = \phi(Z_1) - Z_1 \bar{\Phi}(Z_1)) \\
& = \frac{Z_1 \phi^2(Z_2)}{L(Z_1)} \left(Z_1 \bar{\Phi}(Z_1) + \frac{\bar{\Phi}^2(Z_1)}{\phi(Z_1)} - \phi(Z_1) \right) > 0.
\end{aligned}$$

The last inequality holds from the fact that $A(Z_1) = Z_1 \bar{\Phi}(Z_1) + \frac{\bar{\Phi}^2(Z_1)}{\phi(Z_1)} - \phi(Z_1) > 0$ as shown in the proof of Lemma 2.

Finally, for the terms in $\{\cdot\}$ corresponding to the constant, we have the following:

$$\begin{aligned}
& Z_1 \phi(Z_2) \left(\frac{\ln 2 Z_2^2 \phi(Z_1) \phi(Z_2)}{L(Z_1)} \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right) - (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \right) \\
& + (1 + Z_2^2) Z_1^2 \phi^2(Z_2) \bar{\Phi}(Z_1) \left(\frac{1}{\phi(Z_1)} \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right) - \frac{\ln 2}{L(Z_1)} \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right) \right) \\
& + (1 + Z_2^2)^2 \left(\frac{\bar{\Phi}(Z_1)}{\ln 2} (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \right) \\
& = \frac{\ln 2 Z_1 \phi(Z_1) \phi^2(Z_2)}{L(Z_1)} \left(Z_2^2 + (1 + Z_2^2) \frac{Z_1 \bar{\Phi}(Z_1)}{\phi(Z_1)} \left(\frac{L(Z_1)}{\ln 2 \phi(Z_1)} - 1 \right) \right) \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right)
\end{aligned}$$

$$\begin{aligned}
& + \left((1 + Z_2^2)^2 \frac{\bar{\Phi}(Z_1)}{\ln 2} - Z_1 \phi(Z_2) \right) (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \quad (\text{by rearranging the terms}) \\
& \geq \frac{\ln 2 Z_1 \phi(Z_1) \phi^2(Z_2)}{L(Z_1)} \left(\frac{\phi(Z_1)}{L(Z_1)} - 1 + \frac{Z_1 \bar{\Phi}(Z_1)}{L(Z_1)} \left(\frac{L(Z_1)}{\ln 2 \phi(Z_1)} - 1 \right) \right) \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right) \\
& + \left((1 + Z_2^2)^2 \frac{\bar{\Phi}(Z_1)}{\ln 2} - Z_1 \phi(Z_2) \right) (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \quad (\text{by using Assumption 1}) \\
& \geq \frac{\ln 2 Z_1 \phi(Z_1) \phi^2(Z_2)}{L(Z_1)} \left(\frac{Z_1 \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right) \left(\frac{L(Z_1)}{\phi(Z_1)} - \frac{L(Z_2)}{\phi(Z_2)} \right) \\
& + \left((1 + Z_2^2)^2 \frac{\bar{\Phi}(Z_1)}{\ln 2} - Z_1 \phi(Z_2) \right) (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \quad (\text{by using } L(Z_1) = \phi(Z_1) - Z_1 \bar{\Phi}(Z_1)) \\
& \geq \left((1 + Z_2^2) \frac{\phi(Z_1) \bar{\Phi}(Z_1)}{\ln 2 L(Z_1)} - Z_1 \phi(Z_2) \right) (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \\
& \quad (\text{by dropping the first term, which is non-negative, and using Assumption 1 for the second term}) \\
& > \left(Z_2 \frac{\phi(Z_1) \bar{\Phi}(Z_1)}{\ln 2 L(Z_1)} - Z_1 \phi(Z_2) \right) (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \quad (\text{since } (1 + Z_2^2) > Z_2 \text{ holds for all } Z_2 > 0) \\
& > \left(\frac{Z_2 \phi(Z_1)}{\ln 2} - Z_1 \phi(Z_2) \right) (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \quad (\text{by using } \bar{\Phi}(Z_1) > L(Z_1) \text{ since } Z_1 > 0) \\
& > \left(\frac{Z_2 \phi(Z_1)}{\ln 2} - Z_2 \phi(Z_1) \right) (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) \quad (\text{by using } Z_2 > Z_1, \phi(Z_1) > \phi(Z_2)) \\
& = Z_2 \phi(Z_1) \left(\frac{1}{\ln 2} - 1 \right) (\phi(Z_2) \bar{\Phi}(Z_1) + Z_2 \bar{\Phi}(Z_2)) > 0.
\end{aligned}$$

Therefore, all terms in $\{\cdot\}$ are positive, and thus we obtain $\frac{\partial^2}{\partial X \partial p} \mathfrak{C}(X, Z_1, Z_2) > 0$. Then, the result follows from Lemma 3.

Proof of Proposition 4(a)

The proof follows similar to that of Proposition 3(a). From Proposition 3(a), we know that $\frac{\partial^2}{\partial X^2} \mathfrak{C}(X, Z_1^*) > 0$. Also, it is straightforward to show

$$\frac{\partial^2}{\partial X \partial u} \mathfrak{C}(X, Z_1^*) = \mu k_1 > 0.$$

Let $f(X, u) := \frac{\partial}{\partial X} \mathfrak{C}(X, u)$. Then, by Lemma 3, the optimal X in (8), $X^*(u)$, is decreasing in u .

Proof of Proposition 4(b)

For this proof, we omit the superscript $*$ for brevity, as there is no risk of confusion.

For the first part of the proof, it is straightforward to show that

$$\frac{\partial^2}{\partial X \partial u} \mathfrak{C}(X, Z_1, Z_2) = \mu k_1 > 0.$$

Also, from the proof of Proposition 3(b), we know that $\frac{\partial^2}{\partial X^2} \mathfrak{C}(X, Z_1, Z_2) > 0$. Then, by Lemma 3, we know that the optimal X in (4), $X^*(u)$, is decreasing in u .

For the second part of the proof, by implicitly differentiating both sides of (9) with respect to v , we obtain

$$\begin{aligned} -\frac{\phi(Z_1)}{Z_1^2} \frac{\partial Z_1}{\partial v} &= -\frac{\ln 2 p}{v^2 k_2 X} \frac{\phi(Z_2)}{Z_2} - \left[\frac{\ln 2 p}{v k_2 X} \frac{\phi(Z_2)}{Z_2^2} + \frac{\ln 2 p}{v k_2 X} \phi(Z_2) \right] \frac{\partial Z_2}{\partial v} \\ \Rightarrow \quad \frac{\partial Z_1}{\partial v} &= \frac{Z_1}{\phi(Z_1)} \left[\frac{L(Z_1)}{v} + L(Z_1) \left(Z_2 + \frac{1}{Z_2} \right) \frac{\partial Z_2}{\partial v} \right] \quad (\text{by using (9)}). \end{aligned} \quad (22)$$

Also, by implicitly differentiating (5) with respect to v , we obtain:

$$0 = k_2 X \bar{\Phi}(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) + v k_2 X \frac{Z_1 \bar{\Phi}(Z_1)}{\ln 2 Z_2} \left(\frac{1}{Z_1} \frac{\partial Z_2}{\partial v} - \frac{Z_2}{Z_1^2} \frac{\partial Z_1}{\partial v} \right) - v k_2 X \phi(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) \frac{\partial Z_1}{\partial v} - p \phi(Z_2) \frac{\partial Z_2}{\partial v}.$$

Then, by rearranging the terms, it follows that

$$\begin{aligned} \frac{\bar{\Phi}(Z_1)}{v} \log_2 \left(\frac{Z_2}{Z_1} \right) &= \left[\frac{\bar{\Phi}(Z_1)}{\ln 2 Z_1} + \phi(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) \right] \frac{\partial Z_1}{\partial v} + \left[\frac{Z_2 L(Z_1)}{\ln 2 Z_1} - \frac{\bar{\Phi}(Z_1)}{\ln 2 Z_2} \right] \frac{\partial Z_2}{\partial v} \\ \Rightarrow \frac{1}{v} \left[\left(\bar{\Phi}(Z_1) - Z_1 L(Z_1) \right) \log_2 \left(\frac{Z_2}{Z_1} \right) - \frac{L(Z_1) \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right] \\ &= \left[\frac{L(Z_1) Z_2}{\ln 2 Z_1} - \frac{\bar{\Phi}(Z_1)}{\ln 2 Z_2} + \left(\frac{L(Z_1) \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} + Z_1 L(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) \right) \left(Z_2 + \frac{1}{Z_2} \right) \right] \frac{\partial Z_2}{\partial v} \quad (\text{by using (22)}). \end{aligned} \quad (23)$$

We note that the bracketed term in the RHS of (23) is non-negative for all $Z_1 > 0$, as we show below. In particular, the third term is positive and first two terms are collectively non-negative since

$$\begin{aligned} \frac{L(Z_1) Z_2}{\ln 2 Z_1} - \frac{\bar{\Phi}(Z_1)}{\ln 2 Z_2} &= \frac{1}{\ln 2 Z_1 Z_2} (L(Z_1) Z_2^2 - Z_1 \bar{\Phi}(Z_1)) \\ &\geq \frac{1}{\ln 2 Z_1 Z_2} (\phi(Z_1) - L(Z_1) - Z_1 \bar{\Phi}(Z_1)) = 0. \end{aligned}$$

In the above, the inequality follows from $Z_2^2 \geq \phi(Z_1)/L(Z_1) - 1$ in Assumption 1.

Now, we consider

$$\frac{\partial^2}{\partial X \partial v} \mathfrak{C}(X, Z_1, Z_2) = -\frac{\sigma}{\sqrt{\delta} X^2} \left(h + p \frac{L(Z_2)}{Z_2} \right) \frac{\partial Z_1}{\partial v} + \frac{p \sigma Z_1}{\sqrt{\delta} X^2} \frac{\phi(Z_2)}{Z_2^2} \frac{\partial Z_2}{\partial v}.$$

By using (22), it follows that

$$\begin{aligned} \frac{\partial^2}{\partial X \partial v} \mathfrak{C}(X, Z_1, Z_2) &= -\frac{\sigma}{\sqrt{\delta} X^2} \left\{ \left(h + p \frac{L(Z_2)}{Z_2} \right) \frac{Z_1 L(Z_1)}{v \phi(Z_1)} \right. \\ &\quad \left. + \left[\left(h + p \frac{L(Z_2)}{Z_2} \right) \frac{Z_1 L(Z_1)}{\phi(Z_1)} \left(Z_2 + \frac{1}{Z_2} \right) - p \frac{Z_1 \phi(Z_2)}{Z_2^2} \right] \frac{\partial Z_2}{\partial v} \right\}. \end{aligned} \quad (24)$$

To show $\frac{\partial^2}{\partial X \partial v} \mathfrak{C}(X, Z_1, Z_2) \leq 0$, it is sufficient to show that the term in $\{\cdot\}$ of (24) is non-negative. We use (23) to show this. Specifically, we multiply the bracketed term on the RHS of (23), which is non-negative, to the expression in $\{\cdot\}$ in (24) to obtain the following:

$$\left[\frac{L(Z_1) Z_2}{\ln 2 Z_1} - \frac{\bar{\Phi}(Z_1)}{\ln 2 Z_2} + \left(\frac{L(Z_1) \bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} + Z_1 L(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) \right) \left(Z_2 + \frac{1}{Z_2} \right) \right] \left[\left(h + p \frac{L(Z_2)}{Z_2} \right) \frac{Z_1 L(Z_1)}{v \phi(Z_1)} \right]$$

$$+ \left[\frac{L(Z_1)Z_2}{\ln 2 Z_1} - \frac{\bar{\Phi}(Z_1)}{\ln 2 Z_2} + \left(\frac{L(Z_1)\bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} + Z_1 L(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) \right) \left(Z_2 + \frac{1}{Z_2} \right) \right] \\ \left[\left(h + p \frac{L(Z_2)}{Z_2} \right) \frac{Z_1 L(Z_1)}{\phi(Z_1)} \left(Z_2 + \frac{1}{Z_2} \right) - p \frac{Z_1 \phi(Z_2)}{Z_2^2} \right] \frac{\partial Z_2}{\partial v}.$$

We further obtain the following using (23), i.e., replacing the RHS to its LHS:

$$\left[\frac{L(Z_1)Z_2}{\ln 2 Z_1} - \frac{\bar{\Phi}(Z_1)}{\ln 2 Z_2} + \left(\frac{L(Z_1)\bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} + Z_1 L(Z_1) \log_2 \left(\frac{Z_2}{Z_1} \right) \right) \left(Z_2 + \frac{1}{Z_2} \right) \right] \left[\left(h + p \frac{L(Z_2)}{Z_2} \right) \frac{Z_1 L(Z_1)}{v \phi(Z_1)} \right] \\ + \frac{1}{v} \left[\left(h + p \frac{L(Z_2)}{Z_2} \right) \frac{Z_1 L(Z_1)}{\phi(Z_1)} \left(Z_2 + \frac{1}{Z_2} \right) - p \frac{Z_1 \phi(Z_2)}{Z_2^2} \right] \left[\left(\bar{\Phi}(Z_1) - Z_1 L(Z_1) \right) \log_2 \left(\frac{Z_2}{Z_1} \right) - \frac{L(Z_1)\bar{\Phi}(Z_1)}{\ln 2 \phi(Z_1)} \right].$$

By using $\frac{h}{p} = \ln 2 \frac{Z_1 \phi(Z_2) \bar{\Phi}(Z_1)}{Z_2 L(Z_1)} \log_2 \left(\frac{Z_2}{Z_1} \right) + \bar{\Phi}(Z_2)$ from (5) and (9), and reexpressing the above by collecting terms with respect to $\log_2 \left(\frac{Z_2}{Z_1} \right)$ (and then, within, with respect to $(1 + Z_2^2)$), we obtain:

$$\log_2^2 \left(\frac{Z_2}{Z_1} \right) \left\{ \frac{\ln 2 Z_1^2 \phi(Z_2) \bar{\Phi}^2(Z_1)}{v Z_2^2 \phi(Z_1)} (1 + Z_2^2) \right\} \\ + \log_2 \left(\frac{Z_2}{Z_1} \right) \frac{Z_1 \phi(Z_2)}{v Z_2^2} \left\{ -\bar{\Phi}(Z_1) + Z_1 L(Z_1) - \frac{Z_1 \bar{\Phi}^2(Z_1)}{\phi(Z_1)} - \frac{L(Z_1) \bar{\Phi}(Z_1)}{\phi(Z_1)} \right. \\ \left. + (1 + Z_2^2) \left(\frac{L(Z_1) L(Z_2) \bar{\Phi}(Z_1)}{\phi(Z_1) \phi(Z_2)} + \frac{Z_2 L(Z_1) \bar{\Phi}(Z_1) \bar{\Phi}(Z_2)}{\phi(Z_1) \phi(Z_2)} + \frac{L(Z_1) \bar{\Phi}(Z_1)}{\phi(Z_1)} \right) \right\} \\ + (1 + Z_2^2) \left(\frac{Z_1 Z_2 L(Z_1)}{\ln 2 v Z_2^2 \phi(Z_1)} \right) \left\{ \frac{\phi(Z_2) \bar{\Phi}(Z_1)}{Z_2} - \frac{L(Z_2) \bar{\Phi}(Z_1)}{Z_2} - \bar{\Phi}(Z_1) \bar{\Phi}(Z_2) + \frac{Z_2 L(Z_1) L(Z_2)}{Z_1} + \frac{Z_2^2 L(Z_1) \bar{\Phi}(Z_2)}{Z_1} \right\}.$$

To show that this value is positive, we show that expressions in the three sets of $\{\cdot\}$ are all positive.

It is straightforward to see that the first expression in $\{\cdot\}$, associated with the $\log_2^2 \left(\frac{Z_2}{Z_1} \right)$ term, is positive. For the expression associated with the $\log_2 \left(\frac{Z_2}{Z_1} \right)$ term, we have the following:

$$-\bar{\Phi}(Z_1) + Z_1 L(Z_1) - \frac{Z_1 \bar{\Phi}^2(Z_1)}{\phi(Z_1)} - \frac{L(Z_1) \bar{\Phi}(Z_1)}{\phi(Z_1)} \\ + (1 + Z_2^2) \left(\frac{L(Z_1) L(Z_2) \bar{\Phi}(Z_1)}{\phi(Z_1) \phi(Z_2)} + \frac{Z_2 L(Z_1) \bar{\Phi}(Z_1) \bar{\Phi}(Z_2)}{\phi(Z_1) \phi(Z_2)} + \frac{L(Z_1) \bar{\Phi}(Z_1)}{\phi(Z_1)} \right) \\ \geq -\bar{\Phi}(Z_1) + Z_1 L(Z_1) - \frac{Z_1 \bar{\Phi}^2(Z_1)}{\phi(Z_1)} - \frac{L(Z_1) \bar{\Phi}(Z_1)}{\phi(Z_1)} + \frac{L(Z_2) \bar{\Phi}(Z_1)}{\phi(Z_2)} + \frac{Z_2 \bar{\Phi}(Z_1) \bar{\Phi}(Z_2)}{\phi(Z_2)} + \bar{\Phi}(Z_1) \\ \text{(by using Assumption 1)} \\ = Z_1 L(Z_1) - \frac{(\phi(Z_1) - L(Z_1)) \bar{\Phi}(Z_1)}{\phi(Z_1)} - \frac{L(Z_1) \bar{\Phi}(Z_1)}{\phi(Z_1)} + \frac{L(Z_2) \bar{\Phi}(Z_1)}{\phi(Z_2)} + \frac{(\phi(Z_2) - L(Z_2)) \bar{\Phi}(Z_1)}{\phi(Z_2)} \\ = Z_1 L(Z_1) - \bar{\Phi}(Z_1) + \bar{\Phi}(Z_1) > 0.$$

Finally, for the terms in $\{\cdot\}$ corresponding to the constant, we take the first three terms:

$$\frac{\phi(Z_2) \bar{\Phi}(Z_1)}{Z_2} - \frac{L(Z_2) \bar{\Phi}(Z_1)}{Z_2} - \bar{\Phi}(Z_1) \bar{\Phi}(Z_2) \\ = \frac{\phi(Z_2) \bar{\Phi}(Z_1)}{Z_2} - \frac{\bar{\Phi}(Z_1) (\phi(Z_2) - Z_2 \bar{\Phi}(Z_2))}{Z_2} - \bar{\Phi}(Z_1) \bar{\Phi}(Z_2) = 0.$$

The remaining terms are all positive.

Therefore, we have $\frac{\partial^2}{\partial X \partial v} \mathfrak{C}(X, Z_1, Z_2) \leq 0$. Hence, from Lemma 3, we know that the optimal X in (4), $X^*(v)$, is increasing in v .

Appendix B: Details of Simulation Experiment for the Distribution Cost Approximation

In our simulation experiment, we varied the values of input parameters as follows: $X = \sqrt{50 + 100k}$ square miles for $k = 0, \dots, 4$, $\sqrt{50}$; M equals $2k + 1$ for $k = 1, \dots, 8$ (i.e., 9 to 289 DCs in the inventory sharing group), mean demand per aggregation point $\hat{\mu}$ equals $(3k + 1)/4$ for $k = 1, \dots, 10$ units, $\hat{\sigma} = \sqrt{\hat{\mu}}$, and the safety stock factors $Z_2 = 1 + 0.4k$ for $k = 0, \dots, 4$ (i.e., $Z_1 = Z_2/M$ ranges from 0.05 to 1.3). For computational tractability, we set $K = 4$. We note that increasing K only slightly affects the value of k_1 , but the functional form of approximation does not change. For each combination of parameter values, which adds up to 1925 test cases, we simulated for 15 demand realizations and took the average for the optimal costs of the corresponding transportation problems.

Appendix C: Parameter Values in Numerical Illustration

In the numerical study reported in Section 5, we set the parameter values based on public information regarding Amazon's distribution system. Although Amazon is known to carry out second-order inventory sharing at its fulfillment centers (Xu *et al.* [43]), we note that the purpose of this exercise is to provide a quantitative example of the various effects and trade-offs involved as a numerical illustration; we do not draw direct comparison to Amazon's current supply chain design, since other factors such as intensifying competition with brick-and-mortar retail chains (e.g., Walmart) will play significant role which we do not capture in our model.

First, we set the fixed cost of locating DCs, f_1 , as \$150 million, which is approximately the cost of new fulfillment center buildings (ranging from \$100 million to \$200 million [31]). To convert this to a per-period cost, we use a value of κ equal to the sum of Amazon's cost of capital and a depreciation factor. The cost of capital can be estimated based on the standard weighted cost of capital model. Because Amazon recorded a negative profit (i.e., no equity return) for the 2012-13 financial year, we may simply use the interest-debt ratio of about 3% as the cost of capital [1]. To obtain the depreciation rate, we assume straight-line depreciation over a 30-year life time for the building, i.e., 3.3% per year. Because the inventory turnover rate was 9, we consider a period to be 1/9 of a year, yielding a value of $\kappa = 0.063$. For the fixed operating costs of DCs g_1 , we consider the other operating expenses (i.e., those other than cost of sales, fulfillment, marketing, technology,

general & administrative expenses) of Amazon, divided among its current 18 fulfillment centers, yielding a value of \$8.8 million per year.

For the density of demand points δ , we consider each of the 48 contiguous states to have two demand points. Divided by the area of 3.1 million square miles, the value of δ is given by 2.6×10^{-5} points per square mile. We use a package to represent one unit of demand. This is reasonable because shipping costs are incurred per package. The mean demand per demand point μ is taken from the annual estimated North American gross sales (excluding media items), divided by the average package value of \$47 [24] ($\mu = 8.68 \times 10^6$). The gross sales are estimated from net sales figures adjusted for a gross margin of 57% [1]. We obtain the standard deviation of demand per period by considering the coefficient of variation of demand at each demand point to be 0.3 ($\sigma = 2.6 \times 10^6$).

The unit holding cost h is estimated based on the weighted cost of capital (estimated using data from the financial reports [1]) multiplied by the estimated item cost per unit of demand (i.e., the average sales value per package times one minus the gross margin ratio), which yields $h = \$0.6$. For the penalty cost p , it is difficult to obtain an estimate due to the difficulty to measure intangible costs such as loss of goodwill. Therefore, we attempt to estimate this figure using our model. Based on Amazon's current distribution network of 18 DC locations in the US, which gives X , as well as the other estimated parameter values, we consider the optimal Z_1 and Z_2 levels for different values of p . The final value of p is selected to match Amazon's "perfect order" (a performance metric that includes having the item in-stock) percentage of 95% that it recommends for third-party sellers [2], suggesting $p = \$20$.

Finally, the unit transportation cost v is estimated from the published shipping rates of Fedex, assuming an average shipping weight of 5 pounds per package. Because the rates are step functions (defined by shipping zones) of distance, we perform a linear regression to approximate cost as a linear function of distance. The r -squared value of the regression is 0.91. The variable cost coefficient is taken as the value of unit shipping cost u ($= \$2.23 \times 10^{-3}$). Finally, to reflect the inconvenience cost (e.g., time delay or additional administrative costs) of fulfilling an order from a secondary DC, we assume that the unit cost of the extra transportation cost term is three times that of a regular shipment ($\$6.70 \times 10^{-3}$).